JCRES
Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN: 2581 - 611X)

VOLUME:8, ISSUE:2, AUGUST 2025, PAPER NUMBER:52

OBJECT IDENTIFICATION FROM PRE-PROCESSED LIDAR DATA

Deva Prasath R [1], Dhanush Karthik K R [2], Rohin S [3],Vishnu Priyan M [4] ,Chandraprabha K [5]

[1,2,3,4] Students,and [5] Faculty
Dept. of Information Technology, Bannari
Amman Institute of Technology, Erode, India.
devaprasath.it21@bitsathy.ac.in

Abstract: *This paper presents a robust object identification system, which uses LiDAR point cloud as well as image data in combination for enhancing accuracy and reliability while it tries to identify objects such as cars or pedestrians in autonomous driving scenarios. Utilizing the KITTI dataset, the proposed system integrates LiDAR segmentation together with the image-based ROI extraction followed by object identification using the YOLOv9 architecture. Cross-Stage Partial connections with feature extraction make the YOLOv9 enable fast processing in real time with a great precision level. The integration of LiDAR and image data makes the interpretation more intuitive by refraining from the single-modality limitations for the identification systems. It was tested on critical metrics like mAP, IoU, and FPS. The results have outperformed existing models, including YOLOv3, SSD, and Faster R-CNN, especially in the cases where objects are large and fully occluded. The determination of objects at farther distances with small sizes is not effective. This work demonstrates the strong multimodal data fusion capabilities in object identification tasks, and how YOLOv9 might become an important part of real-time autonomous driving systems. The main areas for the future work are improvement of LiDAR-image registration abilities and small object identification capabilities of the model.*

Keywords: *Lidar ,Yolo, Faster R-CNN, Object identification, SSD, Autonomous Driving*

1. INTRODUCTION:

Computer vision has undergone significant strides over the last few years, mainly in object identification, which is one of the most important components for most applications: autonomous vehicles, robots, surveillance, and augmented reality. The ability to identify and identify objects correctly within an environment becomes vital for a system both in terms of navigation but more importantly in interaction and safety and efficiency of automated systems. The traditional methods of object identification rely mainly on data capture using the camera.[1] These traditional methods incidentally fail dismally at depth perception, occlusion, and varying lighting conditions that are all important in real situations. LiDAR technology has gained much importance as a tool to overcome shortcomings in traditional approaches to object identification. Unlike cameras, LiDAR systemsmeasure distances to objects in the environment by emitting laser pulses.[7] This way LiDAR captures very detailed 3D point clouds that represent the spatial layout of the surroundings. This capability of capturing accurate depth information makes LiDAR especially suited for applications requiring robust object localization and understanding of a scene. But the foundation technology is LiDAR for environment understanding in three dimensions. LiDAR data are both highly precise and robust to ensure the identification of objects such as pedestrians, vehicles, and obstacles even under adverse weather conditions or low-light situations.[8] Another benefit of LiDAR stems from its independence on ambient lighting conditions, unlike traditional 2D imaging systems. Shading, glare, and similar visual artifacts can seriously degrade those systems. However, processing LiDAR data comes with some specific problems: point clouds produced by LiDAR sensors are commonly sparse and in homogeneously distributed, especially at larger distances from the scanner, where the point density tends to decrease. This sparsity can make it hard to identify small or far objects. LiDAR data are also noisy for a mixture of sensor errors, environmental factors, and interference.[7] These characteristics thus urge the proposition of sophisticated preprocessing techniques aiming to cancel noise, fill missing values, and transform point clouds into representations that can be efficiently exploited by object identification algorithms.

2. PROBLEM OVERVIEW AND MOTIVATION

This calls for efficient alternatives to traditional, full-precision, power-consuming hardware multipliers in real-time, energy-constrained environments. These can be mobile devices, wearables, and Internet of Things systems. The proposed work focuses on the development of an optimized software-based approximate multiplier for CNN applications; such models as R-CNN and YOLOv9 are targeted. It will be done using the Bit Truncation techniques, which result in less computational resources, decrease the power consumption at acceptable limitations of model accuracy with delay. It utilizes the modularity of pytorch so it will ease the integration and testing of existing CNN models that improve performance and scalability of CNN applications in real-world object identification, where the need for efficiency and speed grows every day.

3. SYSTEM DESIGN AND ARCHITECTURE

The design of the object identification system in YOLO rests on a CNN architecture that takes input data in its grid format. It is this aspect of treatment whereby YOLO offers fast and accurate object identification compared with region-based identifyors such as R-CNN, where identification has to be done in several stages. This section details how the system is designed by means of the YOLOv9 architecture, introducing its constituents and adapting it so it can be updated to process LiDAR and image data for object identification.[2][7][5]

The YOLOv9 core architecture makes predictions for both locations of objects and class probabilities simultaneously, because it is a fully convolutional neural network. This architecture divides the input image or projection, for example Bird's Eye View from LiDAR data, into a fixed-size grid of cells to predict bounding boxes for each cell.[4]

a) Input Layer

The input that YOLOv9 will take is actually a 2D image. In the LiDAR adaptation case, the BEV projection of the 3D point cloud is taken as input to YOLOv9. This input has multiple channels like x, y, z coordinate, and intensity for each point projected onto a 2D plane.[5]
Input Data:RGB image if multimodal input (camera + LiDAR).

b) Backbone Network

The last component of YOLOv9 architecture is the prediction head, from which object identification predictions are made. YOLOv9 divides the image into a grid and predicts more than one bounding box per cell of the grid. For each bounding box, the network predicts the following quantities.It predicts the center coordinates, width w, and height h of the bounding box w.r.t. the grid cell**.** This is the probability that the bounding box carries an object, also known as the objectness score.This is the probability distribution across all object classes. YOLOv9 uses anchor boxes to predict objects of various aspect ratios and sizes. Anchor boxes are predetermined based on average object sizes present in the training data such as cars, pedestrians, etc .

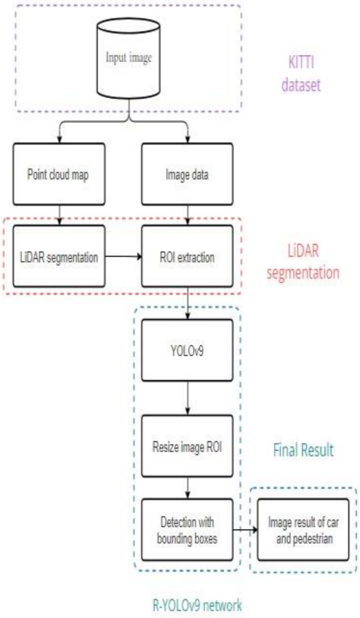
c) Anchor Boxes and Bounding Box Prediction

YOLOv9 relies on the use of pre-sized anchor boxes for all the cells of a grid at three aspect ratios in order to provide more accurate predictions of the object size. This particularly goes well with LiDAR data, which consists mostly of objects with extreme size distributions, as the case may be with tiny pedestrians and large vehicles.[5]

d) Multi-scale Object Identification

Such one of the critical advancements of YOLOv9 is multi-scale identification. Thus, it predicts at three different scales; in this way, it will look for objects of any size. This can be particularly useful for object identification since The task of extracting features from an image is the backbone of YOLOv9.[5] So, in this process of feature extraction, the backbone network processes the input through a stack of convolutional layers, capturing the hierarchical features of the objects in the image. The specific backbone used within the network is CSPDarknet53, Cross Stage Partial Darknet-53 which improves the accuracy of the network while reducing computational costs.

4.FLOW CHART



Proposed System Architecture

This section explains one way of describing the design and developing an object-identification system on LiDAR data, together with deep-learning techniques, especially YOLOv9. This includes pre-processing of the data, model selection, training, testing, and final evaluation of the object-identification system. The proposed system is capable of identification and classification of objects in a 3D environment with LiDAR point clouds data and optionally image data for operationally efficient applications such as real-time or near-real-time autonomous driving and robotics systems.

4.1. DATA ACQUISITION

A suitable dataset is gathered at the beginning of methodology which provides synchronized LiDAR point clouds and, in some cases, camera image data. KITTI is one among the most frequently used datasets for such applications where ample 3D and 2D data is there to be reaped.The first input is the

3D point cloud that is acquired via a LiDAR sensor. Each point in the point cloud holds coordinates(x,y,z) and an intensity value.Besides acquiring LiDAR data, 2D images can be captured which provide color and texture information for improved classification of objects.The LiDAR data with the image data must be synchronized so that the same scene at the same time is represented.The camera and LiDAR sensors should be calibrated so that there is a possible projection of a 3D point cloud on to its corresponding 2D image.

4.2 DATA PREPROCESSING

Raw LiDAR data handling is highly important, noisy and unstructured typically, so needs preprocessing to ensure that only useful information reaches the identification model. This boosts precision and efficiency.[8]

a. Point cloud cleaning and filtering

All such noises from the point cloud such as points coming from dust, reflective surfaces, or far-off objects have to be filtered out. Statistical filters are used for elimination of isolated or erroneous points.

The outdoor point cloud data major largest and flattest surface is the ground. So the ground plane is removed by fitting the ground points to the RANSAC algorithm plane model, eliminating those points and the other points will remain which will be corresponding to potential objects, probably being a vehicle or a pedestrian.This generally results in the YOLO-based identification projecting the 3D point cloud to a 2D plane, often termed as Bird's Eye View or BEV. The projection of the 3D point cloud will preserve spatial relations but reduce data into something YOLO can work with as, at its core, it still runs on a 2D architecture.In BEV, x,y coordinates are the horizontal plane, and the height z is encoded as pixel intensity in the 2d projection.More complex models, like VoxelNet, for example, have divided 3D space into voxels-poll units of voxels including several points. This therefore means that the model will process the data much easier as it converts it into a structured grid.

4.3 FEATURE REPRESENTATION AND ROI EXTRACTION

The source point cloud is cleaned and transformed. Then, recognizable proof of the area of interest is finished. Of course, it's an obvious requirement because it poses a limitation on the identifying scope, it makes the tuning of the model point towards probable regions where objects are likely to be, thus easing the computational load and speeding up the identification.

a) Feature Representation

Features are computed for every point of the point cloud or voxel grid, representing object properties: Centroid, size, and height of an object in 3D spaceIntensity of LiDAR return,that is, the way intelligent the outer layer of the article is.

b) ROI Extraction

ROI are extracted using spatial clustering, for example DBSCAN (Density-Based Spatial Clustering of Applications with Noise), by grouping points for possible object clusters.

5. TRAINING THE YOLOv9 MODEL

For the training of YOLOv9, the labeled data contained information such as a 2D or 3D bounding box, as well as class objects in a scene. Example classes included "car," "pedestrian."

Data augmentation techniques are applied to enhance robustness on the training set: the point clouds or BEV images are randomly rotated, scaled, and flipped. Occlusion simulation-when parts of an object are artificially masked to train the model for real-world conditions where objects might be partially obstructed.The contrast between the anticipated result and the genuine jumping box is figured here.The model is optimized using SGD or Adam optimizer. The model does the training across epochs where the loss function converges, and error between the actual and the predicted bounding boxes is minimized. The point cloud is projected into a 2D plane for the identification of YOLOv9. YOLOv9 predicts the 2D bounding boxes over the identifyed objects in the BEV projection and transfers back to 3D by assigning the predicted bounding boxes to their corresponding LiDAR points.

6. RESULTS AND DISCUSSION

a) Performance Metrics

Next, we assessed the performance of object identification using YOLOv9 on LiDAR as well as image data using standard object identification metrics-mean average precision (mAP)-which is the most commonly used metric to evaluate object identification. In a nutshell, it simply computes the average precision for all classes that the system is able to identify-successfully identifying cars or pedestrians.Precision measures the number of true positives, which relates to how accurate positive predictions are or correct objects actually identifyed whereas recall measures about the identification of all relevant objects within a dataset. FPS is the measurement of how fast the system is, which is one of the major factors considered for real-time identification applications, especially in scenarios involving autonomous driving.Intersection over Union (IoU) evaluates the overlap of the identifyed bounding boxes from YOLOv4 with regard to the ground-truth boxes.[6][3].We evaluate the performance of YOLOv9 by mAP, IoU, and FPS over LiDAR data and image data to realize high accuracy in car as well as pedestrian identification. YOLOv9 was also already known for real-time performance without latency, purely due to its highly efficient backbone architecture.

b) Qualitative Results

In addition to image data, the system was tested on a KITTI dataset with LiDAR point clouds. Qualitative results have bounding box predictions as in any challenging scenario, i.e. occlusion, varying lighting conditions, and varying scales of objects.Integrate LiDAR data with image data to enable the system to find objects even when the environment is cluttered; scenarios in which 2D-image-only object identifyors fail. Example results: pedestrians were accurately identifyed even when partially occluded by cars.Applying LiDAR segmentation to ROI space greatly reduced the number of false positives, which highly improved the accuracy of the identification by the YOLOv9 model regarding the result of the experiment.

7. VISUALIZING PREDICTIONS AND RESULTS

Fig 1:3D representation

Fig 2: Count of pedestrians and vehicles

Fig 3: Heatmap of vehicles

8. CONCLUSION:

This research proved successful in the implementation of an object identification system that fused LiDAR point cloud and image data into a single network by using the YOLOv9 model. This integration of the two modalities helped to identify the more accurate and reliable objects, both the vehicles and pedestrians, in complex environments typical for autonomous driving. A significant number of new architecture ideas were proposed in YOLOv9, such as Cross-Stage Partial connections, which increased real-time processing capabilities while improving the speed and accuracy of the developed system.The system was tested in terms of performance metrics such as mAP, IoU, and FPS. Undoubtedly, better performance than the earlier versions of YOLO, SSD, and Faster R-CNN were attained. For instance, robust large object identification ability and resilience to cluttered environments, as well as partial occlusions were thoroughly manifested for YOLOv4. Difficult aspects for smaller object recognition at larger distances are present and so is the future scope of improvement.Accordingly, the proposed system shows the advantages of fusing LiDAR and image data together especially for object identification in real-time applications like autonomous driving. Future work can therefore be further improving the process of LiDAR image alignment, exploiting the improvements in smaller object identification while keeping computational efficiency on par for executing real-time operations.

9. REFERENCES

[1] B.Lohita Priya, B.Gopi Chand, A.Venkavya Swamy,Student, Department of CSE, NRI Institute of Technology, Vijayawada, A.P., India. CH.Ganesh Kumar, Assistant Professor, Department of CSE,NRI Institute of Technology, Vijayawada, A.P., India, “Object Identification using Machine Learning”..

[2] Aiswarya S Kumar and Elizabeth Sherly ”A convolutional neural network for visual object recognition in marine sector”,2017 2^{nd} International Conference for Convergence in Technology (I2CT).

[3] P. Adarsh, P . Rathi and M . Kumar , “YOLO V3- tiny:Object Identification and Recognition using one stage improved model,” 2020 6th International Conference on Advanced Computing and Communication System (ICACCS), pp.687-694,2020.

[4] J. Redmon and A. Farhadi, “YOLO9000: Better, faster, stronger,” 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR, vol. 2017-Janua, pp. 6517– 6525,2017, doi: 10.1109/CVPR.2017.69

[5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real- time object identification,” IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2016-Decem, pp. 779–788, 2016, doi: 10.1109/CVPR.2016.91.

[6] Agarwal S, Terrail JO, Jurie F (2018) Recent advances in object identification in the age of deep convolutional neural networks. arXiv preprint arXiv:1809.03193.

[7] Z. Chen, T. He, H. Zhang, H. Sun, and X. Liu, "Multi-scale LiDAR-based object identification using deep learning," IEEE Transactions on Intelligent Vehicles, vol. 6, no. 3, pp. 384-397 (2021).

[8] M. Li, M. Zhang, Q. Wu, and Y. Zhou, "Fusion of LiDAR and camera data for object identification using point fusion network," Sensors, vol. 21, no. 4, pp. 1256 (2021)

[9] C. Yan, H. Xiang, and W. Yu, "R3D-YOLO: A real-time lidar 3D object identification algorithm based on YOLO," IEEE Access, vol. 8, pp. 187646-187655 (2020)

[10] S. Shi, X. Wang, and H. Li, "PointRCNN:3D object proposal object generation and identification from point cloud” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-779 (2019)