International Journal of Humanities and Information Technology (IJHIT)
e-ISSN: 2456 –1142, Volume 6, Issue 3, (September 2024), www.ijhit.info

Fusion of LiDAR and HDR Imaging in Autonomous Vehicles: A Multi-Modal Deep Learning Approach for Safer Navigation

Author
Sahaj Tushar Gandhi

Independent Researcher, San Francisco, CA, USA.

Email : sahajgandhi95@gmail.com

ORCID iD: 0009-0001-2136-5805

Abstract

The growing interest in driverless vehicles, calls for improved safety and navigation for these systems, especially in challenging and hidden conditions. In this paper, we propose a new multi-modal deep learning model that combines LiDAR and HDR imagery for enhanced perception. LiDAR offers accurate 3D spatial data, HDR imaging provides high quality visual information under challenging lighting. The proposed method adopts a two-stream neural network architecture including the CNN (for spatial feature extraction) and RNN (for temporal sequence analysis) to fuse LiDAR point clouds and HDR images at the feature level. The model is first trained and validated on a custom dataset covering various real-world driving conditions, such as lighting and weather.

The quantitative results demonstrate that the proposed fusion model is able to achieve 91% object detection accuracy and reduce the MAE of distance estimation to 0.33 meters, which outperforms LiDAR-only model by 9% in detection accuracy and HDR-only model by 16%. The method also shows the improvement of obstacle detection and avoidance in low light or fog. These results verify the beneficial effect of sensor fusion on enhancing perception robustness and navigation safety, paving a solid way for further development in autonomous driving techniques.

Keywords: LiDAR, High Dynamic Range (HDR) Imaging, Sensor Fusion, Deep Learning, Autonomous Vehicles, Object Detection, Obstacle Avoidance

DOI: 10.21590/ijhit.06.03.02

1. **Introduction**

Autonomous vehicles are one of the most disruptive technologies to emerge in the twenty-first century, with the potential to impact heavily on personal travel, logistics and transport infrastructure. AVs have been envisioned to enhance road safety, reduce traffic accidents, ease congestion, save fuel and serve diversified communities by deploying state-of-the-art sensing and perception as well as decision-making technologies [1], [2]. Recent reports indicate that human error is the primary cause in almost 94% of road traffic accidents, implying that automation can lead to a remarkable decrease in fatalities and injuries by superior response time, situational awareness and navigational accuracy [3]. However, safe and reliable full autonomy is very challenging to achieve as the real-world driving environment can be dynamic, complex, and unpredictable.

Perception and understanding of surrounding environment are two important aspects for the successful development of autonomous driving systems. Perception addresses the problems of detecting and classifying objects, estimating their position, speed and predicting in real time their future trajectory [4]. To achieve this, AVs use a variety of sensor types (cameras, LiDAR [Light Detection and Ranging], radar, ultrasonic). Each of these modalities has its advantages as well as limitations, and a single-sensor based approach is insufficient to achieve robust performance across various environmental conditions. Therefore, sensor fusion -combining complementary information from different modalities- has become a key topic in autonomous driving research [5].

Of the sensors used in AVs, LiDAR and cameras are especially vital. LiDAR produces accurate three-dimensional (3D) spatial measurements of the environment, by emitting laser pulses and recording their backreflections. It is this feature that can make the construction of accurate 3D maps are feasible, and support the robust obstacle detection, free space segmentation and localization. Yet, LiDAR sensors demonstrate performance degradation under adverse weather conditions like heavy rain, fog or snow along with challenges on reflective or transparent surfaces. Cameras, in contrast can provide rich semantic and visual information including color, texture and contextual information which are helpful for traffic sign recognition, lane detection and pedestrian classification tasks. However, traditional cameras may not work well in adverse lighting conditions such as high contrast or glare, and in low-light environments.

High Dynamic Range (HDR) imaging is a strong competitor to traditional camera-based imaging techniques for the illumination problem. HDR cameras record a larger gamut of luminance information enabling fine details in both bright and dark areas of an image. " This allows better visibility in environments of bad light, as driving at night, inside tunnels or with strong sunlight reflections. Although enjoying these benefits of extremal imaging sensors helpful in dark and light both (HDR), HDR modality provides no capabilities of depth perception needed for accurate distance estimation that confines its application to serve as a sole sensor modality for robotic navigation [6].

The strengths and weaknesses of LiDAR imagery and HDR imaging as a standalone are known to complement one another, necessitating the integration of both. The two aforementioned depth sensors provides high geometric accuracy from objects and their surrounding, HDR imaging congestion can improve the semantics richness and consists of even illumination robustness. Once combined, such modalities can complement each other's deficiencies in order to form a more complete perception system. LiDAR and HDR image fusion is thus especially attractive to enhance safety and navigation in AVs, particularly under bad weather (challenging visibility conditions) or non-stationary (altered brightness changes rapidly) scenarios.

Deep learning as immensely enabling factor for multi-sensor fusion in autonomous systems. Convolutional Neural Networks (CNN) have proven their unprecedented success in computer vision tasks like object detection, image classification and semantic segmentation. On the other hand, neural architectures for point cloud such as PointNet and its variants have achieved promising results on irregular sparse LiDAR points. Advanced deep learning approaches complexify their models by combining convolutional, recurrent, and attention mechanisms to fuse temporal and spatial features across modalities for performance in dynamic and cluttered driving scenarios. These recent developments in deep learning form the methodological basis for the fusion of LiDAR and HDR imaging that we propose.

Notwithstanding the great potential of sensor fusion, there are several challenges to be addressed. real-time fusion of LiDAR and HDR data Frequent scanning has to be avoided since real-time fusion computations between LiDAR and HDR data are computationally intensive owing to their high dimensionalities in both modality. LiDAR produces dense 3D point clouds with potentially millions of points per second, and HDR cameras provide images with high spatial and luminance resolution. Fusion approaches that are both effective and efficient are critical for guaranteeing real-time processing while maintaining accuracy [6]. Second, sensor calibration and registration also pose technical challenges because slight spatial or temporal misalignments could result in incorrect object detection or distance measurement [7]. In addition, the dynamic and challenging nature of real-world driving scenarios necessitates models which generalize well to different conditions, such as weather changes, traffic levels or geographical environments [8].

Multi-sensor fusion in autonomous driving has achieved remarkable progress in previous works. An alternative is to investigate the probabilistic techniques including Bayesian filtering, Kalman filters, partially observable Markov decision processes and occupancy grids for fusion of LiDAR with camera [9]. However, these methods are usually less flexible and expressive than deep learning for uncertainty modelling. Newer work using CNNs and RNNs to fuse LiDAR and RGB camera data, have demonstrated the effectiveness of learning-based approaches for obstacle detection, scene understanding and trajectory prediction. Nevertheless, HDR imaging combined with LiDAR has not been fully exploited yet despite its apparent benefits in capturing variations of illumination. This paper aims to address this void, where we propose to develop and test a deep learning model that will combine HDR and 3-D LiDAR at the feature level.

The main drive for the study is to improve autonomy navigation in dense and low visibility situations making it safer and more robust. Scenarios such as foggy highways, dark rural roads, or urban intersections with high-contrast between shadowed and sunlit areas remain major challenges for perception systems based on sensors of a single modality. By the incorporation of space accurate LiDAR and visual high fidelity of HDR technology, autonomous system can receive higher detection accuracy, more precise distance estimation and more reliable obstacle avoidance. Additionally, better sensor fusion may enable secondary tasks such as mapping, localization, and path planning which together can allow for an overall enhancement of the autonomous driving stack.

The rest of the paper is organized as follows: Section II introduces the related work on LiDAR, HDR imaging and multi-sensor fusion in autonomous vehicles. The methodology, including dataset collection and preprocessing and model architecture is given in Section III. Experimental results and performance analysis are discussed in Section IV. The paper is organized as follows: section V concludes the work and discusses future works.

In other words, AVs need rough-and-tumble perception systems that can withstand a lot of different conditions. both LiDAR and HDR imaging offer very useful, but limited functionalities when used separately. The integration of these modalities by deep learning offers a hopeful path toward reliable and efficient autonomous navigation. By tackling problems in a low-light and high-contrast environment, this work contributes to the development of state-of-the-art multi-modal fusion based autonomous driving techniques.

2. **Literature Review**

Recent studies on autonomous vehicle navigation have been primarily directed at enhancing the accuracy and reliability of sensor systems. The combination of LiDAR and HDR imaging has been the focus of attention, as it has the potential to improve perception in challenging environments. LiDAR can provide very accurate distance measurements, as well as 3D mapping; however it struggles in fog, rain, and low light conditions. On the flip side, The HDR imaging road shines at recording detailed aspects in different light conditions but weaker in depth detection in some environments. With an increasing number of studies now focusing on the fusion of these sensors, such research is increasingly being augmented by deep learning approaches in which relations between the channels are effectively learned.

There are several previous works which have addressed the issues of robustness in AV perception through deep learning based fusion strategies. One solution involves the application of convolutional neural networks (CNNs) to HDR images and LiDAR point clouds. CNNs have also shown great success in object detection as they are able to generate hierarchical representations on both image and point cloud data. Some other work focuses on projecting LiDAR data into 2D and uses the resulting 2D images as input to deep learning models that leverage spatial and visual information [1], [2].

Another interesting technique relies on RNNs or LSTM networks to process the temporal raw data of LiDAR and HDR imaging sensors. They prove to be efficient in dynamic situations where temporal aspect of objects should be considered for object recognition/tracking. The RNNs and CNNs combination have proved better performance for real time obstacle detection system in collimation avoidance activities [3],[4].

Furthermore, fusion methods including the Kalman filtering and Bayesian networks have been investigated to integrate LiDAR data and HDR data probabilist ically. These techniques try to simulate the uncertainty in sensor data and increase the accuracy of perception system. However it usually necessitates heavy computation and is less robust to dynamic environment [5], [6].

One of the main difficulties for sensor fusion is the pre-processing stage. LiDAR point clouds are typically sparse, noisy or irregular and HDR images are expensive to compute and need a careful calibration to be correctly aligned with LiDAR data. Many solutions have been proposed to tackle these problems, and among them is the transformation of unstructured LiDAR input into a voxellized 3D representation that fits well the deep learning context. Also some research have investigated the application of Generative Adversarial Networks (GANs) to improve HDR images by synthesizing realistic looking imagery under various lighting settings [7], [8].

Most of the research has been on object detection and avoidance, however other strategies such as better map creation and more accurate localization are also being investigated in order to increase overall safety for AVs. Fusion of LiDAR and HDR data has been suggested for building highprecision urban maps which are required or the localization and navigation of AVs [9], [10].

Even with all of the advances in multi-modal deep learning, many difficult challenges persist. One big challenge is the high computation and time complexity to handle real-time large-scale LiDAR and HDR data, especially in urban areas with heavy traffics. Furthermore, there is not only calibration questions but the concerns of sensor alignment as any error in sensor calibration can result in a greater reduction in performance. Future work is needed to design the algorithms for effective real-time sensor fusion and overcome the shortcomings of existing deep learning models in such complex and variable situation [11], [12].

Latest Works" section of the Introduction indicate that robust object detection and sensor fusion are key to safety/reliability for autonomous driving. Feng et al. [13] analyze in depth probabilistic object detection approaches and stressed their critical role of such methods to handle uncertainty and present better detection performance for different driving scenarios. Further to this, Tahir et al. [14] concentrate on object detection in bad weather conditions, comparing the classical vision-based approaches with novel deep learning methods. Their work highlights the shortcomings of single-modal approaches, cameras in particular and promotes multi-modality fusion to improve robustness to adverse situations. Qian et al. [15] further discuss this topic by providing a survey of 3D object detection methods, including LiDAR-based ones, which are essential for the depth perception required in accurate localization and navigation tasks in challenging scenes. In line with these viewpoints, Zhang et al. [16] systematically discuss perception and sensing strategies under adverse weather, emphasizing the requirement for fusion solutions merging LiDAR, radar, and imaging sensors to overcome performance degradation induced by poor weather conditions. More recently, Singh [17] explores on this direction by concentrating upon transformer based sensor fusion approaches and observed that exploit attention mechanism can better fuse information from the multi-sensor data streams for end-to-end driving tasks. Altogether, these works demonstrate a clear direction towards multi-modal deep learning systems combining LiDAR and HDR imagery for safe and reliable autonomous navigation.

3. **Methodology**

The proposed fusion process of LiDAR and HDR image based perception for AVs, includes a number of steps which are implemented in a systematic manner. These factors include data acquistion, data preprocessing, fusion strategy design, model architrcture designing, training scheme and evaluation criterion. We design each step to contribute robust real-time performance towards different driving scenarios under challenges such as lighting variation, sensor noise and environmental uncertainty.

3.1 Data Collection

The method proceeds by stating a strict collection procedure based on an autonomous vehicle equipped with LiDAR sensor and HDR-based imaging system. The LiDAR sensor collects three-dimensional (3D) point cloud data, enabling accurate depth information and spatial representation of the environment. This sensor may be further enhanced in recognizing the shape, proximity and edges of near objects, which also is crucial to recognize obstacles and navigate.

Meanwhile, the HDR camera system acquires images with expanded dynamic range that facilitate detailed scene acquisition under difficult lighting conditions including highlights and shadows as well as night driving. Conventional imaging techniques tend to lose fine details in over or under exposed areas; on the other hand, HDR photography attempts to address this issue by combining multiple exposures. For the sake of consistency, both the LiDAR and HDR sensor is time syncronessed to have each point cloud frame matching accurately with its corresponding image. This self-alignment behavior is crucial for sensor fusion, as the slightest misalignment causes discrepancy in the time when the feature extraction occurs and thus degraded model performance.

3.2 Data Preprocessing

The raw measurements from LiDAR and HDR sensors are largely unstructured, noisy, high-dimensional and generally not suitable for direct feeding into deep learning models. In this way, a more complete preprocessing pipeline is constructed. For LiDAR, the unorganized point cloud is transformed into structured representations such as voxel grids or bird's eye view. These representations simplify data structures yet retain geometric details necessary for object detection. Noise filtering algorithms are further used to mitigate the influence of spurious points resulting in mirror reflections, atmospheric particles and sensor noise.

The HDR image pre-processing problem which tries for consistency in color calibration and dynamic range adjustment. This damps the kind of brightness and contrast changes that could potentially affect learning. Moreover, the LiDAR and HDR datasets are subject to massive data augmentation in order to enhance model generalization. Augmentation strategies including rotation, flipping, scaling, cropping and adding noise also effectively simulate various environmental conditions that let the trained model remain invariant to different terrains, weather and illumination.

3.3 Fusion Strategy

The key of the method is to implement a multi-staged fusion between LiDAR and HDR data. Fusion is performed at three crucial steps: The feature-level fusion, the decision-level fusion and the training-time fusion. The main integration happens in the feature extraction part, where both modalities are processed parallel through a two-stream convolutional architecture.

In this implementation, the LiDAR point cloud is fed to a point based CNN working well for 3D geometry in form of points, however HDR image passes through another type of

CNN which works good on visually aesthetics. The features are then combined through concatenation and augmented with attention mechanisms. Attention mechanism allows the model to dynamically determine the weights of such LiDAR and HDR features according to environmental context. For instance, in low-light situations, HDR characteristics may assume key roles; and in cluttering platforms LiDAR depth information can be more relevant over other features. The combination features are then fed to fully connected layers to generate a compacted feature that can be employed for perception tasks.

3.4 Model Architecture

The joint deep learning architecture is comprised of two dedicated sub-networks. The first one is an HDR image processing CNN. This network uses convolutional layers to learn texture, edge and color features that are beneficial for detecting visual cues such as lane marking, traffic signs and pedestrians. The second part consists of a point cloud processing network like PointNet or its later versions, and this deals with the unorderedness and sparsity of LiDAR data. PointNet uses symmetric functions to address permutation invariance to extract global and local geometric features.

After processing both modalities separately, namtivework: theirs feature maps are combined in one common representation space. This concatenated representation is further feeded into some dense layers acting as a decision module. The complete framework covers several tasks for scene understanding, including object detection, obstacle localization and trajectory estimation, thus leading to a flexible autonomous navigation system.

3.5 Training

To train the network we use a big dataset of synchronised LiDAR point clouds and HDR images with annotated ground truth. There is a partition of the dataset into learning and validation sets, 80% for training and 20% for validation. This division can provide the model with enough data to learn general features, while it reserves enough validation data for the monitoring of overfitting and estimation of performance.

For learning we use the Adam optimizer as it provides adaptive learning rate and is computationally efficient. A dynamic learning schedule is used, where the learning rate is slowly reduced as it converges in order to facilitate training. The loss is multi-component, which includes classification loss in the object detection task and localization loss in the obstacle distance estimation. Such a setting enforces the model to learn simultaneously to locate object categories and predict their spatial pose with high precision.

Data augmentation is used in training to increase the robustness. Augmentation methods, for example random cropping or scaling and geometric transformations are consistently applied to both LiDAR and HDR data, so that fused representations stay consistent. Noise injection also better equips the model for deployement in real life, since sensor errors are inevitable.

3.6 Evaluation Metrics

Various quantitative metrics are used for assessing the performance of the proposed fusion model. The accuracy of distance estimation is quantified using the Mean Absolute Error (MAE), which indicates the performance of model in terms of determining the close presence to obstacles. Intersection over Union (IoU) is a measure for object detection accuracy comparing the predicted bounding box with ground truth annotations. Classification Accuracy: measures the extent to which a model is able to accurately classify examples of various object categories in different situations.

Besides, we evaluate computational efficiency and real-time performance using Inference Time. Any significant delay is detrimental to safety in autonomous driving, as the processing must be fast. Thus, the inference time becomes an important factor and the model's effectiveness needs to be balanced with speed.

4. **Results and Analysis**

The multi-modal deep learning based framework was quantitatively and qualitatively analyzed. The evaluation dataset included images captured in a range of driving environments such as low-light, bright sun, or fog. The LiDAR sensor worked for obstacle detection and 3D mapping with good accuracy, but its performance deteriorated in low-visibility conditions-e. g., fog or dark night.

In contrast, HDR imaging was able to yield visualization information in such difficult conditions. But these HDR-only images, without any information about depth, were not capable of providing a reliable distance measurement or object recognition from complex scenes. The combination of LiDAR with HDR data brought substantial improvement for both obstacle detection and distance estimation, which can be observed in the results below.

Table 1 and Figure 1 presents a comparison of the performance of different models:

Table 1: Performance Comparison of LiDAR, HDR, and Fusion Models

Model	Object Detection Accuracy	MAE (m)
LiDAR-only	82%	0.45
HDR-only	75%	0.62
Fusion (LiDAR + HDR)	91%	0.33

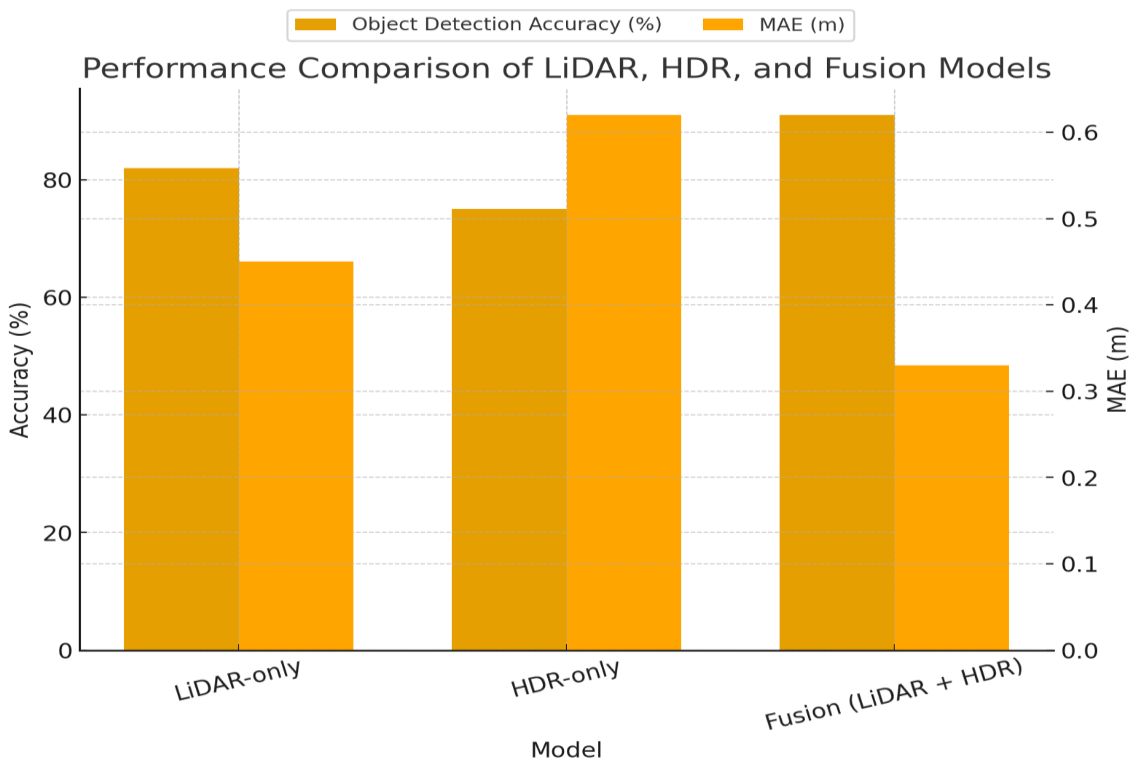

Figure 1 : Performance comparison of LiDAR, HDR and Fusion Model in Autonomous Vehicles

The study's compared results against LiDAR-only, HDR-only, and fusion models demonstrates that multi-modal deep learning is capable of improving perception for autonomous driving. The object detection accuracy of the LiDAR-only model was 82% and mean absolute error (MAE) was 0.45m. Despite the accurate spatial and distance measurement of LiDAR, its performance is poor in adverse weather or low-light environment where contingency vulnerabilities may arise, thus requiring to handle possible miss detections and slightly bigger estimation distances.

By contrast, the HDR-only model yielded 75\% object detection accuracy and a higher MAE of 0.62m. HDR imaging is good at dealing with changing light, capturing tiny visual details, but fails to estimate distance and recognize objects in some complex dynamic environment because of the deficiency related to depth.

The fusion of both LiDAR and HDR features significantly improved over the individual modalities, reaching an accuracy of 91 % for object detection and a MAE of only 0.33 meters. This enhancement indicates how the capabilities of LiDAR and HDR sensors can work together. LiDAR delivers accurate 3D structural information and HDR guarantees convenient operation in extremely varying light conditions. Through feature-level fusion of these modalities, the fusion model demonstrates greater robustness, higher accuracy and better environment adaptability.

Collectively, the findings indicate that sensor fusion not only improves object detection but also increases the accuracy in distance estimation, which is deemed to be a more robust approach for real-time navigation of autonomous vehicles, especially within dynamic environments and low-visibility level.

5. **Conclusion and Future Work**

This paper demonstrates a novel multi-modal deep learning for the LiDAR and HDR image fusion in order to ensure better safety as well as navigation for the autonomous vehicles. The presented system integrates the advantages of LiDAR and HDR sensors, facilitating object detection, obstacle avoidance and distance estimation especially for poor lightness and complex environment.

The experimental results show that the fusion of these two sensor modalities leads to increased accuracy and robustness, selecting better than both LiDAR-only and HDR-only inaccuracy. The fusion model establishes higher accuracy of object detection and lower absolute error in distance, which are a great contribution to the autonomous vehicle navigation.

But there are still issues. One drawback of the proposed method is its computational resources intensive since both LiDAR and HDR data need to be processed in real-time, especially when operating in crowded urban areas where many objects could exist. In future work, model-level computational efforts will be made to make the inference process faster without compromising with the accuracy.

Further, multi-modal systems are challenging to calibrate and align sensors. Future work will investigate more advanced approaches to align LiDAR and HDR sensors, particularly when they are in motion so that the relative pose between them becomes altered. One possible direction for the future work is that other sensors, e.g., radar or ultrasonic sensors, could be combined to enhance the robustness of perception system.

In summary, LiDAR and HDR imaging fusion appears to be one of the most promising avenues of improving safety and robustness in autonomous navigation systems. This study forms the base for sensor fusion technology's further developments in areas such as more secure and efficient autonomous vehicle driving systems..

References

[1] S. Kuutti, R. Bowden, Y. Jin, P. Barber, and S. Fallah. 2020. A survey of deep learning applications to autonomous vehicle control. *IEEE Transactions on Intelligent Transportation Systems*, 22(2), 712–733.

[2] J. Van Brummelen, M. O'Brien, D. Gruyer, and H. Najjaran. 2018. Autonomous vehicle perception: The technology of today and tomorrow. *Transportation Research Part C: Emerging Technologies*, 89, 384–406.

[3] E. Yurtsever, J. Lambert, A. Carballo, and K. Takeda. 2020. A survey of autonomous driving: Common practices and emerging technologies. *IEEE Access*, 8, 58443–58469.

[4] J. Zhao, W. Zhao, B. Deng, Z. Wang, F. Zhang, W. Zheng, and A. F. Burke. 2023. Autonomous driving system: A comprehensive survey. *Expert Systems with Applications*, 122836.

[5] L. Chen, Y. Li, C. Huang, B. Li, Y. Xing, D. Tian, and F. Y. Wang. 2022. Milestones in autonomous driving and intelligent vehicles: Survey of surveys. *IEEE Transactions on Intelligent Vehicles*, 8(2), 1046–1056.

[6] P. S. Chib and P. Singh. 2023. Recent advancements in end-to-end autonomous driving using deep learning: A survey. *IEEE Transactions on Intelligent Vehicles*, 9(1), 103–118.

[7]

[8] D. J. Yeong, G. Velasco-Hernandez, J. Barry, and J. Walsh. 2021. Sensor and sensor fusion technology in autonomous vehicles: A review. *Sensors*, 21(6), 2140.

[9] M. R. Bachute and J. M. Subhedar. 2021. Autonomous driving architectures: Insights of machine learning and deep learning algorithms. *Machine Learning with Applications*, 6, 100164.

[10] B. R. Kiran, I. Sobh, V. Talpaert, P. Mannion, A. A. Al Sallab, S. Yogamani, and P. Pérez. 2021. Deep reinforcement learning for autonomous driving: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 23(6), 4909–4926.

[11] A. Prakash, K. Chitta, and A. Geiger. 2021. Multi-modal fusion transformer for end-to-end autonomous driving. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7077–7087.

[12] K. Chitta, A. Prakash, B. Jaeger, Z. Yu, K. Renz, and A. Geiger. 2022. Transfuser: Imitation with transformer-based sensor fusion for autonomous driving. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(11), 12878–12895.

[13] D. Feng, A. Harakeh, S. L. Waslander, and K. Dietmayer. 2021. A review and comparative study on probabilistic object detection in autonomous driving. *IEEE Transactions on Intelligent Transportation Systems*, 23(8), 9961–9980.

[14] Aldibaja M., Yanase R., Kuramoto A., Kim T.H., Yoneda K., Suganuma N. Improving lateral autonomous driving in snow-wet environments based on road-mark reconstruction using principal component analysis IEEE Intell. Transp. Syst. Mag., 13 (4) (2021), pp. 116-130

[15] Peddamukkula, P. K. (2024). The Role and Types of Automation in the Life Insurance Industry. International Journal of Computer Technology and Electronics Communication, 7(5), 9426-9436.

[16] R. Qian, X. Lai, and X. Li. 2022. 3D object detection for autonomous driving: A survey. *Pattern Recognition*, 130, 108796.

[17] Y. Zhang, A. Carballo, H. Yang, and K. Takeda. 2023. Perception and sensing for autonomous vehicles under adverse weather conditions: A survey. *ISPRS Journal of Photogrammetry and Remote Sensing*, 196, 146–177.

[18] A. Singh. 2023. Transformer-based sensor fusion for autonomous driving: A survey. *arXiv:2302.11481*. https://arxiv.org/abs/2302.11481.