Deep Learning-based Sensor Fusion for Enhanced Perception in Autonomous Vehicle Environments

By Dr. Veronica Murillo

Associate Professor of Computer Science, Tecnológico de Costa Rica (TEC)

Introduction

In this paper, a thorough survey on the implementation of multimodal sensor-based knowledge of environmental elements using state-of-the-art deep learning and sensor processing advancements in HAVs and FAVs with respect to the vehicle-feature gathering and determining the perception level of vehicle detectors was carried out. Although on-the-shelves sensors provide a great amount of data for developing vehicle driver assistance systems, safety, reliability, security, and economic aspects of the system come into question for the proper application of these data for the typical vehicle perception systems, especially for the HAVs and FAVs. Even the high-end sensor systems are subject to uncertainties and vulnerabilities related to the environment and the sensor itself. The decrease in the cost of sensors, increasing robustness, for example for camera sensors against lighting conditions, more advanced deep learning methodologies, and computing capabilities are the most common reasons for the implementation of top-level perception systems not only around the obstacles in any environment over deep sensor fusion but also for the anticipation of their future actions.

[1] [2]Autonomous vehicles aim to provide a high level of safety, reduce driving costs and commutes, and improve individual mobility services. During the last decade, significant progress has been made in levels of vehicle autonomy - Assisted, Partial Automated, Highly Automated, and Fully Automated Driving. High-level perception systems are used by law enforcement officers, such as police officers and transport planners in smart cities. Urban traffic information infrastructure and related high-level perception systems are still under development, and the current vehicle detection systems mainly rely on the use of lidar and camera data [3]. The Dynamic Environment Analysis (DEA) and Digital Aided Conductor (DAC) summoning services are part of the European Railway Traffic Management System

(ERTMS) and have basic functionality for autonomous or semi-autonomous train navigation. A review of existing practices and their limitations in highly Automated Vehicles(HAVs) and fully Automated Vehicles (FAV) show the current evaluation system with discrete levels of fuzziness is not advanced, the uneven development of various sensors may lead to incompatibility between sensors, and the low-level fusion information is not clearly embodied in high-level products. Furthermore, raw data is not available for the decision module to greatly affect the early decision-making process during the capture of information with low-fidelity or highly occlusions.

Background and Motivation

In this article, authors present a comparison between conventional sensor fusion techniques and advanced deep learning-based sensor fusion techniques. The authors believe that combining the strengths of different sensors in an optimal manner could lead to better scene perception performance instead of conflicting or limited scene perception. Hence, this review article discusses the potentials and the current trending active research areas in deep learning-based sensor fusion algorithms, for enhanced perception modules for autonomous vehicles, ADAS, and robots. Common sensor fusion algorithms and deep learning sensor fusion-based methodologies are reviewed, compared, and detailed for perception modules in ADAS and autonomous vehicles system. This creative potential of sensor fusion algorithms with deep learning methods for enhancing a wide range of typical ADAS and autonomous vehicle perception systems. The potential for improvement in such systems shows the power of such algorithms.

Autonomous vehicles largely rely on various perception technologies such as data from LiDARs, object detection cameras, HD maps, and GPS systems to ensure safe driving under various environmental conditions [3]. Each sensor modality has its strength and weakness, whereas no single sensor can guarantee the autonomous operations in all the environmental conditions. Therefore, to achieve robust and safe autonomous navigation, the complementary characteristics of different sensor modalities are combined to develop a redundant and reliable sensor fusion perception module. Consequently, sensor fusion techniques are highly pertinent for safe and efficient ADAS and autonomous navigation application.

0e6377bf-8049-4911-9884-98d1fe280e89 8594840d-5099-4888-925a-a56a94cc923b

Research Objectives

We aim to create a training PRIORITIZATION algorithms for fusion applications. If such data is available, we will use the train-prioritization capabilities of our network to predict which data it is important to collect in order to improve model performance. The general goal of this work is to learn how to infer when additional domain confusion is required for meaningful integration to improve the fused or priority processed model.

[4] [3] This research investigates a multi-modal sensor fusion approach based on multi-layer deep learning architectures. The project aims to develop algorithms for the fusion of LiDAR and camera data to determine the 3D position, velocity and class of objects in scene. The project will culminate in a dataset consisting of the true labeled positions, velocities and classes of objects in autonomy-related scene. The project investigates fusion of: 1) Bounding boxes and lidar points with labels for 3D object detection and 2) Bounding boxes and camera images with labels for 2D object detection.

Scope and Organization of the Work

Scope and Organization of the Work: In this thesis, we focus mainly on sensor fusion. To this end, we study a few non-classical deep fusion architectures and propose new deep fusion layers and detection networks that can share features across different sensor modalities. Our main contribution, as part of the observation block, is the use of early fusion and late fusion of features computed from the shared deep sub-networks in combination. The use of the proposed deep sensor fusion-based perception algorithms will require a lot of sensor data streams, and issues such as data spatiality (cross-modal shift) and data modality scaling are crucial to address. To this end, we study them within the context of sensor fusion and propose solutions for these issues [5]. It is also crucial to have some robustness to some confounding factors related to the sensor dynamics of data. We study some of them in the thesis and propose novel strategies for addressing them. We partly tackle confounding factors hollistically through "Hydrafusion", an approach for context-aware autonomous vehicle sensor fusion to be presented in chapter 2. We have also proposed a novel DDRNet object detection network architecture which can process depth camera data to be presented in Chapter 3. DDRNet is expected to be useful for autonomous vehicle signaling and warning systems.

Current autonomous vehicle development is heavily driven by machine learning systems, particularly deep learning-based perception algorithms that use raw sensor modalities such as camera images, LiDAR, and Radar evolutions to perform crucial perception tasks such as clustering, tracking, and localization. Most solutions employ single deep nets to process different sensor inputs separately and follow this up with domain-specific processing layers for sensor fusion [6]. These shallow sensor fusion methods might not model inter-neural relationships and could lead to over-convergence on single modalities.

Fundamentals of Sensor Fusion

[ref: b05d2387-c614-4c98-bc43-c41b9e286531, 163560dc-f47d-4b64-a3e4-ffa113f1e966] 2.1. Introduction The basics of sensor fusion involve the instantaneous combination of data gathered by diverse sensors to enhance the overall perception of a scene. It finds its application in various domains such as robotics, surveillance, navigation, image parsing and image retrieval, to name a few. The primary goal with sensor fusion is to build a system with better understanding of the scene [7]. In general, exclusively using one sensor modality to understand the scene usually limits the system to effectively analyze other aspects of the scene such as detailed instances of objects, different categories of vehicles etc. To overcome this limitation, integrating information from different sensor modalities is of utmost importance. In recent years, deep learning techniques for recognition, detection, and segmentation have demonstrated great potential for producing highly detailed, accurate, and dense predictions on perceptually challenging tasks such as scene understanding. One of the successful aspects of deep learning is its generality which allows it to perform relatively well without the need for engineering a specific feature or feature set for a particular task. In the context of sensor fusion, feature set engineering is crucial because it becomes required to reason about the relationships and similarities between features of distinct input modalities with generality. Moreover, it is essential to decide when to use which sensor and how to maintain the generality of the system and at the same time capture the dependencies between different input elements [8].

Definition and Importance

Vision-based object detection, instance-aware segmentation and path planning and obstacle avoidance modules, among others, benefit from the high-resolution image sensor. However, due to the dot detection â€“ range detection type measurement mechanism of the LIDAR sensor and the inference of the targetâ€™s environment type from the intensity information, object detection and obstacle inference performance degrade under certain weather and environmental conditions. (Abu-Alhaija et al. [9]) In the case of radars, while they detect and generate a 3D representation of the objects (in some systems only at the front and rear) in all weather and environmental conditions, they necessitate other sensors and sensor fusion modules to improve the low level information (object class, type, color, etc.) they present to higher level decision making algorithms.

Sensor fusion, also known as sensor integration, is a vital component of the perception stack of self-driving vehicles. It combines information from multiple sensors including camera, radar, LIDAR and GPS receivers in order to enhance the perception of the vehicle. (Legendre et al. [10]) Sensor fusion allows a comprehensive understanding of the environment, compensates for the weaknesses of individual sensors, and offers robust and responsible decisions for the vehicle. Due to its ability to reliably detect objects across various environmental conditions, radar sensors play a crucial role in autonomous vehicles. (Zhang et al. [11]) Moreover, due to its high fidelity and rich spatial information, the camera sensor is the primary source of long range object classification and detection for AVs. In this study, sensor fusion at the perception level, particularly cameraradar fusion is our focus, and we take advantage of the qualitatively advanced performance of the two sensor modalities.

Traditional Sensor Fusion Methods

Before the prevalence of deep learning-based perception methods, the research for LIDAR, radar and camera fusion methods for traditional sensor, fusion techniques was very comprehensive. A large number of advanced reviews focused on the methods based on LIDAR, radar and camera respectively. A large number of discussions on SLAM and Bayesian network methods can be found in the literature for LIDAR and camera fusion. A multisensor data fusion-based vehicle localisation and positioning algorithm for autonomous vehicle in GNSS-denied environment was discussed by Wang et al. Moreover, proposed radar amalgamating particle filtering (RAPF) framework during multiple target tracking using the processed signals collected from detect-and-rescue radar systems. Yang et al. integrated LIDAR and radar data in a coarse-to-fine object detection method suitable for autonomous vehicle scenes. Meanwhile, radar was also integrated with more vision and positioning methods in recent related works, which was further reviewed. Because LIDAR and camera, LIDAR and radar and radar and camera integration belong to 2D-3D sensor fusion, 3D-3D sensor fusion and 2D-2D sensor fusion, respectively, the fusion of camera, radar and LIDAR formed the four-element fusion methods in theory, which have not been analysed in detail in recent review articles. All in all, traditional sensor fat t fusion method can compensate for the shortcoming of n single sensor, and has achieved good performance in the past decades. However, compared with 2D and 3D sensor fusion, they have high actual cost and more complex mounting form.

[[12] [11]] â€œTraditionally, German automobiles are built upon the core technology of driving safety, which mainly relies on the camera and radar to achieve comprehensive perception,â€ Yang et al. discussed the highdefinition map-based vehicle localisation technology using this method. In 2014, the Chinese government identified this technology as one of the focus areas in their 2025 China Automotive Industry Policy and started its full-scale promotion. Sensor fusion technology based on camera and radar is one of the most popular solutions for autonomous vehicle perception. Zhang et al. discussed the sensor fusion algorithm, combining three independent sensors, namely a multi-beam six-layer LIDAR, long-range SWORD radar and Omni-camera in a prediction and decision algorithm based on V2I communication. Although the CLRC system is applicable to a limited road range, its localisation performance has demonstrated the feasibility of combining the LIDAR, radar and camera with an acceptable level of cost and applicability. In general, sensor fusion based on camera and radar may be the most popular solution for vehicle and depth perception among the present fusion methods, especially for passenger vehicles. As far as the active package is concerned, radar has the advantages of high reliability, low cost and high angular resolution with which it has been prioritised in the active safety applications. However, Yang et al. has thoroughly discussed the monocular cameraâ€“LIDAR fusion method for autonomous vehicle scenes with the specificity of LIDAR and radar, considering that this combination is less common when it is applied in autonomous vehicles.

Challenges and Limitations

Moreover, mainly due to the above limitations with 2D, 3D and 1D sensors, which have no omnipotent sensors and thus a major limitation. The notion of weakly supplemented detection should be carefully weighed, and fruitful fruit solutions have not yet been established. Ensuring the stability and consistency of object detection sensors in complicated scenarios represents a crucial task that deep learning methods alone cannot always successfully address. Moreover, to obtain reliable detection results from perception modules, driver preference and velocity also need to be taken into account in some cases. It is difficult for 2D, 3D, radar sensors alone, to detect vulnerable objects (e.g. objects in shadow, objects resting/obscured by other objects, and very small bounding boxes) properly and effectively. And if only one of these sensors performs unsatisfactorily in these scenarios, the predictive performance of the fusion model in each perception system also deceases [4].

According to the limitations and challenges of each sensor in different scenarios, the world-level object detection of perception modules has the following main limitations [ref: 02bfeb78-84f0-465b-bc1c666623d20ab0; 163560dc-f47d-4b64-a3e4-ffa113f1e966]: 1. The camera sensor may perform poorly in low light situations and is easily dazzled by traffic lights or headlights at night. Even during daylight, obstacles such as lamp posts will often be detected because they merge with environmental boundaries. 2. The LiDAR sensor perceives well during daytime, and even in varying light conditions because light scattering will not have deviations in picture or pixel intensities. However, the LiDAR viewpoint could be easily blocked, which means the sensor cannot do well in detecting obstacles and traffic lights that are resting on the edge or behind a bigger and closer object in certain situations. 3. The Radar sensor is really good apart from the lack of details about the object and these building walls as well. Furthermore, a single radar point is usually only a far-distance small quantity sensing with really low resolution.

Deep Learning Fundamentals

It is based on artificial neural networks, which are models which are inspired by the human brainâ€™s neurons. Before deep learning became popular in the early 2000s, researchers mainly used shallow architectures with only a few linear layers because there was no efficient way to optimize the models. However, modern deep learning models can be effectively trained using the backpropagation algorithm, which is based on the chain rule of differentiation [1]. With backpropagation, gradients can efficiently be calculated for modern deep learning models that have hundreds of layers and millions or billions of parameters. As a result, it is common to use deep networks in many modern applications of machine learning as well as in autonomous systems.

In order to develop an in-depth understanding of deep learning-based sensor fusion for advanced perception in autonomous vehicles, this section begins by reviewing the fundamentals of deep learning [13]. Deep learning is a field that involves machine learning algorithms which can automatically learn hierarchical representations of input data to accomplish specific tasks, often involving large amounts of data. Although the field of deep learning has been around for decades, the term â€œdeep learningâ€ was only popularized in the early 2000s.

Neural Networks

Classical sensor fusion methods use geometry or logic-based algorithms for fusion tasks, most of them are based on Kalman Filters, which have some drawbacks and limitations like all of them require good system and sensor models, linearization and noise model estimation and doesnâ€™t handle well multimodal and nonGaussian distributions. Kalman filters and its variants need to know the likelihood function â€“ the function generating the measurements. This requirement can be quite restrictive, both due to the difficulty of capturing the model properly, and any mismatch between the model and real data. And there are different types of Multiple independent estimate and data association algorithms to handle multimodal data respectively, including Intersection over Union, center point distance matching, nearest neighbors, Hungarian assignment, and Munkreâ€™s Assignment and some maximum likelihood estimates, and most of handcrafted features are used at all. Such methods generally require feature engineered from sensors fusion for instance using color, intensity histogram, shape features, or target shape for multimodal data association. Because different sensors are designed to capture different modalities, some correspondences may be missing which causes large errors in the fusion. For example, some usual problems are: depth cameras can fail to perform well under strong sunlight or in a low-light environment making depth cameras facing 3D multi-object detection and tracking still a challenging issue. Although cameras map the scene with rich textures and semantic information, their tracking performance is severely undermined due to these issues. Deep learning-based sensor fusion algorithms apply deep learning to sensor fusion, most of them are well known Convolution Neural Network (CNN), Recurrent Neural Network (RNN), object detector with convolutional recurrent fusion layer, Feature Pyramid Network (FPN), Delayed Graph LSTM, CenterNet, and other widely used network. This type of fusion gets rid of the hand-crafted design of the sensor fusion algorithms, greatly simplifies the processing and extraction of features, and automatically catches the essential features of the multimodal data by stacked layers directly and learns a shared representation of features for all sensors, which extremely simplify the design of traditional sensor fusion for multimodal and provides strong generalization for different applications. However, this kind of method usually needs massive labeling data to achieve better results in some high complex tasks. And a variety of public datasets for monocular depth estimation take advantage of LiDAR sensors together with camera, which can be as a standard protocol for future competition tasks to take full advantage of distributed information of sensor fusion. On the other hand, deep learning-based sensor fusion algorithms are only applied in perception systems of autonomous vehicles. Their application to localisation has not received the same level of attention.

Sensor fusion can be performed in the sensor stream and feature spaces [13]. In sensor stream, raw data is directly fused, because sensors are kept separate, fusion is only possible in instances where the data is known to be beneficial separately, and at different abstraction levels of data, like raw, low level and high level. On the other hand, in feature space, data from sensors are passed through different stages of signal processing and information extraction, and then the fused data is used for higher-level perception tasks, this method is more geometry and semantic aware.

Sensor fusion is crucial in autonomous vehicle systems to ensure proper decision-making regarding the surrounding driving environment [4]. Data from cameras, LiDAR, radar, and other sensors are complementary in nature; fusing them together ensures minimal blind spots and improved understanding of the environment. This obtained 360-degree field of view also allows the system to capture information about vehicles and pedestrians from different perspectives, which can help with association and separate different road users. For example, fusing radar and LiDAR data together can help in acquiring information about the shapes of certain vehicles which are generally not present in LiDAR- only and radar-only frames, while fusing camera and radar can help in associating distant moving vehicles with their actual targets.

Convolutional Neural Networks (CNNs)

The development of feasible sensor fusion methods has been one of the advantages of the current deep learning regimens. The fusion of the viewpoint modalities by adopting the typical convolutional neural network (CNN) software package was proposed in intermediate representations by zhu et al.1 for deciding on only the visible and proposed regions (RoIs) after convolutions with the regional proposal network in the material-based methodology. It is therefore brought to attention that these methodologies are not promptly applicable to autonomous driving; this is attributable to their need to get different proposed plots for each visible view and every single selected object. The initial representations of the radar sensor and these convolved layers were merged before any regional suggestion treatment. Then this merged sensor network output is transformed into the last 3D development if the non-visible object localizes before the regional recommendation proposal stage. Similarly, the 2D and 3D prediction maps and the characteristic maps before the regional recommendation phase when a visual object is detected using the camera network are also merged. In autonomous driving, an alternative approach is proposed in the present work in which only one visual localizes the object on the independence of vehicle speed without the need for a transformation. For this purpose, only the typical viewpoint-based object localization is carried out using the camera network independently of other sensors in the intermediate convolutional layers of the network. As a result, only two instead of three network candidates are generated in the camera-only network at the end of the head layers (there is no need for the radar-only network in this framework). The last parts of these networks offer the classification of box predictions and the awareness of the object category. The visual backbone and intermediate characteristic maps between the radar head and the visual object detections are also merged.

[14] Sensor fusion in autonomous driving combines the benefits of both camera and lidar sensors, resulting in a synergistic effect. For example, each sensor captures different modality information, and therefore is sensitive to different problems. Cameras capture abundant texture and illumination changes for accurate performance in classifying objects. However, camera-based methods are generally affected by low-light and occlusion. On the other hand, Lidar sensors measure the 3D point clouds and provide point-wise depth information about objects. However, lidar sensors struggle with 2D segmentation, primarily due to semantic interpretations and reading traffic signs [15]. Therefore, two-dimensional object detections in local radar objects have been simultaneously carried out in camera images using BVLIR detectors. The collaboration of two different sensor feature maps is essential for battery-free object detections in both radial and horizontal sensors. However, the BVLIR network considers fan-beam projection for the LiDAR object detection that causes sparseness in the

low depth ranges. The localization and classification accuracy in the BVLIR model is impaired severely because of keeping a

limited number of foreground points in the 2Dradars concatenation scheme. For this purpose, the object detection and recognition of the same candidate objects must be performed by merging together the two feature maps of the two different sensors using fully connected layers. In the present study, a new joint network architecture named DenseBVP-Opt is also proposed in the network stage before the proposal region stage by their camera and LiDAR sensors.

Recurrent Neural Networks (RNNs)

Common RNNs generally have difficulty learning about events that are separated by multiple time steps. Alternatively, Long Short-Term Memory Networks (LSTMs) are capable of learning patterns that are so far apart in time that they had been forgotten by any common RNNs. Stacked LSTM network has been used to extracts and integrates features and information about the previous context, skipping frames to model long-term dependency. Several approaches were proposed for the task to predict the driverâ€™s intent to build a traffic scenario recognition, as well as predicting and generating the camera view with either convolutional or LSTM models. A novel Quaternion-Valued LSTM (QLSTM) is proposed for the monitoring task in the paper by using orientation data during the prediction.

[16] [17]Recurrent Neural Networks (RNNs), which have loops in them, allowing the network to persist information over time and have been for be used for the task of autonomous navigation. These models are formulated to predict the next action the autonomous vehicles is expected to take through sensor fusion with the capability to represent continuous-time dynamics, e.g. speed, heading angle etc. by employing Gated Recurrent Units (GRUs) combined with an input and output Recurrent parts to model motion and perception concatenately. In other navigation works, Convolutional and Recurrent Vision Integration Networks (CRVIN) were proposed to fuse visual and Lidar data from ego-motion estimation and scene interpretation by spatially progressive convolutional neural network and temporally progressive LSTM RNN, with on-policy training approach by RL algorithm. In the field of semantic scene perception, ConvLSTM network is applied for integrating the visual SLAM approach to handle rolling shutter effects and automatically learning the complex spatial and temporal patterns in a single end-to-end trainable architecture. On top of this,

fusing semantic information from visual SLAM with other sensor streams such as GPS, Inertial Measurement Unit (IMU), wheel odometry, and semantic segmentation, helped achieving high level of accuracy for pose reconstruction with VisSeq estimations for long image sequences. Semantic segmentation from a single-frame CNN model can be appended to multi-frame LSTM model to generate consistent perception over multiple frames. It is the first to study LSTM for perception in continuous integration framework where perception and mapping is just part of the task.

Deep Learning for Sensor Fusion

There are several challenges of multimodal data fusion, such as how to bring together different modalities and how to relate the features and decision making in the fusion layer for a good performance. In order to deal with this situation, fusion stages which can best represent the relationship between the modalities must be obtained by examining different modality fusion studies. The highest performance obtained from fusion can be achieved by using a multimodal sensor for each sensor type and combining the sensor values of different modes. As a result, deep learning seems to be the most appealing area of sensor fusion studies since it is able to deal with the redundant information in detail and is not affected by dimension problems for designing complex networks. Especially, creating a linear relationship between different modes to find the required values for the overall fusing stage one by one through different modality fusion types is highly advantageous. In the tests made by the recent literature, multimodal fusion techniques generally exhibit performance increase in the relevant object detection and classification [article_id: 86fc987d-9b1a-4def-981e-5afc0f3359be|].

Multi-sensor sensor fusion is an integral part of the sensing suite of modern autonomous vehicle technology [|article_id6dc7ad61-c5ae-4b89-98d1-f202fd4f51df|, |article_idc50a2324-58e7-4e40-83b2-05c16e50981b|]. Sensor fusion technologies lead to more accurate and reliable automotive surrounding perception, which is the foundation of many decision-based functions. Providing most types of sensor data, such as camera, LiDAR (Light Detection and Ranging), or radar, will lead to significant improvements in object detection and classification performance. Most modern autonomous vehicles are equipped with multiple sensors. The camera is a low-cost sensor, but its performance is highly dependent on environâ€‹mental brightness, and it is subjected to limitations such as changes in terrain color, shadowing, as well as difficulties in identifying objects from a distance. On the other hand, LiDAR is rather expensive, it may not be able to distinguish various types of objects very well, and it may not represent crossing or intersecting paths well. In spite of the fact that radar is not affected by environmental conditions, its main diffi- culty is that it causes quite sparse data [article_id: 57975e4b-59bf-4b7c-a372-bc4e6dcf0237||].

Sensor Technologies in Autonomous Vehicles

The long-range radar is responsible for monitoring the status of the far-away vehicle, pedestrian and so on, the detection range and resolution of which are generally 150 m and 0.05$\degree$ respectively. The mid-range radar typically monitors the side and rear of the vehicle, so its detection range and resolution are generally 50 m and 0.1$\degree$ respectively. Moreover, intersections and U-turns of dynamic obstacles are often the blind spots of the long- and medium-range radars, so in practice, the vehicle is generally also equipped with corner radars and centre forward radars to assist in detecting nearby vehicles, pedestrians and so on.

- Cameras: Cameras are visual sensors that provide rich information of the environment, such as color, shape, texture, etc.. Therefore, it is extremely important to obtain accurate and efficient visual information from the camera. Monoclar camera is very sensitive to illumination changes and suffers severe degradation, while stereo/multi-camera hardly perceive accurately with a deformed body. RGB-D camera involves the depth, but its detection range is limited; therefore, it can be only used as the sensor for near-field perception.
- LiDAR: LiDAR technology (Light Detection and Ranging) is also known as laser or laser radar. LiDAR can provide a set of two-dimensional (2D)/three-dimension (3D) point clouds, essentially visualizing the environment, but it is insensitive to colour information [18]. With the evolution of LiDAR technology, the scanning mode varies, mainly including â€˜Velodyneâ€™ , â€˜HDL-64â€™ and â€˜HDL-32â€™. However, more points in its point cloud usually not only increase the difficulty of filtering out redundant points but also make the subsequent algorithm extremely complex to process.

Autonomous vehicles are highly instrumented with various sensor modalities, such as camera, LiDAR, radar, and so on, which can be categorized into two types: environmental perception sensors and environmental map sensors [12]. The former are the main focus of this article. They are responsible for gathering information about the environment instantaneously and include the traditional midrange detection range sensors such as the camera, Long Range Radar (LRR) and Mid-Range Radar (MRR) and near-range detection range sensor, i.e., short-range radar [19]. Table 1 presents some widely adopted sensor technologies in autonomous vehicles. The technical principle and main features of these sensors are summarized as follows:

Camera Sensors

We describe that camera and radar sensors are the prevalent perception modules in autonomous vehicle systems. Where camera sensors provide high-resolution video data, radar sensors measure 3D localization and velocity of vehicles [5]. We therefore propose a data fusion processing pipeline that aggregates targets based on learned perception metrics (e.g., likelihood maps, object association, and tracking states) from the individual perception modules. We implemented the fusion methods using deep learning and prove the proposed pipeline under various scenarios, such as vehicle, pedestrians, or resultant perception data. We train and evaluate our data fusion approaches on different real-world datasets considering three real-world scenarios each.

Fusing data from a variety of sensors in autonomous vehicles leads to more robust predictions [10]. To aggregate sensor data more effectively, we implement 3D CNN in the target network. The network receives input feature maps from the first instance of ResNet blocks at various 2D positions and generates an aggregated representation, which we term as the latent feature map. This is followed by fully connected (FC) layers, which generate the final classification score. We implemented a similar 3D dataset-specific ResNet architecture for both camera and radar sensor data. We train the models using dataset-specific loss functions. In this section, we describe the architecture of our models implemented using spatial encoding [15]. We employ a multilevel fusion operation using a transformer-based architecture, which attends to both spatial and semantic information present in sensory data and generates multimodal features by interacting between different levels of features and modalities. We then utilize an object-centric feature fusion in the detection head that assigns each location a specific object identity, enabling the model to predict the object labels from the multimodal features and significantly improves the performance.

LiDAR Sensors

As mentioned earlier, for unified perception, we use a combination of sensors for increased perception robustness. In this particular work, we provide a deep learning-enabled sensor fusion system to look at the object detection task in a vehicular environment. For comparing the only modality results to the other available sensor range values within our dataset, we extract LiDAR-specific results using the same implementation across all range values. Depending on the choice of the number of classes to be object detected, the sensor range values are agnostic of the object detections [20].

An efficient data fusion strategy lies at the core of any modern perception system, as autonomous systems try to effectively incorporate sensor data to reason about the complex environment around it. The reasons to choose a specific data fusion strategy for autonomous vehicles is influenced by the sensor choice. There are generally four types of sensors used in the literature to enable LiDAR (Light Detection and Ranging) and sensor fusion-based object detection: LiDAR only [7], Camera Only, Radar only and all three modality sensors together. Among these, the only modality sensors are the most commonly employed in vehicle perception systems.

Radar Sensors

In addition to the valuable information and miscellaneous techniques, datasets are also an issue in sensor fusion. There are no standard datasets that contain radar data. To aid the researchers, we collected all the datasets that contain radar information, along with the papers that created the datasets. We utilized these datasets and found that the diverse collection provides the chance to compare and contrast methods and designs. In addition to the methods and datasets, the summary of evaluation matrices and general fusion strategies are collected as valuable references. We hope that this review will be useful for researchers and newcomers in deep learning and the autonomy field, specifically sensor fusion.

Radar fusion in autonomous driving is an essential topic. Many article review fusion strategies and methodologies, focusing on the visual and LiDAR sensors. However, in this paper, we provide a comprehensive review and summary on radar and camera fusion. We explicitly present a detailed review of the perception systems for autonomous driving and discuss how radar complements other sensors, specifically cameras. This survey covers not only the use of photos, videos, and intensity data, but also aims to address raw data from camera and radar. We present a detailed summary of the challenges and future directions.

The main information is extracted from the article [21].

Ultrasonic Sensors

<img src='https://d3i71xaburhd42.cloudfront.net/5d0969168147baa11391c934/5d097178011e1a5a0d6fbca2_Figure51.png'>

The ultrasonic sensor is a widely used sensor for autonomous vehicles because of its mature technique, shortrange response, and relatively lower cost [8]. The sensor output gives the distance information from the obstacle to the vehicle, but even it is used broadly in a car parking system. It has some disadvantages. The sensors are affected by the environmental noise and the bodywork of the car. So, in a vehicle, the ultrasonic sensor sometimes cannot identify the desired object, and the system may experience false alarms. Also, they may have a high probability of triggering false warnings next to reflective or absorptive surfaces, which will reduce passengersâ€™ trust [22]. This may cause unnecessary control signals and the decreased level of safety because it may lead the driver to stop an automobile willingly. For decrement of the probability on unnecessary stop decisions the ultrasonic sensor can be combined with additional sensors and also several sensors using under the different conditions. Besides the problems, the ultrasonic sensors have low contrast ratios resulting in detection impairment in various weather conditions, especially when it rains or snows.

Sensor Fusion Architectures

First, we provide a brief overview of the tasks involved in autonomous driving. The sensor architecture in autonomous driving is almost universally comprised of multiple types of sensors, such as cameras and LiDARs. The main challenge in sensor fusion is coping with the differences in data types, communication modalities, and data resolutions between the sensors. To generalize to different types of autonomous driving tasks, sensor fusion frameworks should be built in modular formats that can toss and aggregate different types of input data efficiently. This section first reviews general modular sensor fusion architectures in autonomous driving and then details camera-LiDAR, V-LiDAR, and mmWave-LiDAR fusion architectures [1].

Modular sensor fusion architectures are at the core of any deep learning-based sensor fusion method [15]. It is important to design these architectures based on the hierarchical representation of the input data and the semantic meaning of the target information in order to enable generalization to different types of data and to different autonomous driving tasks. This section reviews frameworks for sensor fusion techniques in autonomous driving and categorizes them according to sensor architecture, specifically, camera-LiDAR, VLiDAR, and mmWave-LiDAR.

Fusion Levels

The proposed system, HydraFusion, is designed to achieve dynamic sensor data fusion in order to robustify and increase the efficiency of the autonomous car perception system in highly varying driving contexts [23]. Conventional data fusion systems often detect environment objects inconsistently among the different driving contexts which often result in erroneous decision making by the perception system. Moreover, the computational parameter and memory requirements may cause the system to be overwhelmed or inefficient in such varying driving contexts. The extensive quantitative and qualitative comparisons demonstrated that adaptive fusion-based context-aware system outperforms all the competitor baselines.

Data fusion architectures have long been studied, and a key classification of data fusion systems comes from the taxonomy of fusion levels, i.e., early, late, and middle fusion. Sensor data fusion is a critical component for an autonomous vehicle's perception system across all data layers [4]. Despite the various advantages and limitations of different data fusion architectures, especially in the context of the level of abstraction in environment perception, utilizing fusion levels judiciously in different driving scenarios is nontrivial. The necessity and potential for utilizing a variety of fusion levels simultaneously are explored to robustify and adapt the autonomous vehicle perception system by being context-oriented. The classification of different driving scenarios as everyday driving and critical scenario driving have been proposed.

Fusion Strategies

The fusion process strives to extract and disseminate the standout features in a given sensor data modality and combine it with relevant information from other modalities to fill in the gaps using data-centric methodologies. The requirement of fusion process is to primarily generate perceptive systems capable of mimicking human recognition capabilities for identifying various features. From a driving automation perspective, fusion of sensor data modalities is important for decision making while inferring data about the driving environment. The aforementioned approaches suffer from some constraints like limited representation of the data or limited understanding of data and the necessity of engineering features in case of SC methods. The necessity of intricate design for specific perception blocks, and the absence of adaptation of system to scenarios if there are a few number of cross-modal edges/architectures. Navigating through these quandaries requires the invented mechanism to unravel the right amount of information across modalities while encapsulating the adequate abstraction. Also, the invention requires possessing the robustness against information degeneration if used in a real world situation of a data paucity and temporal dependencies, by balancing using readilypredicted information in the modality of one sensor and the act of back-lighting the other reciprocating sensor. Further criteria for intensity based evaluation are listed in Table 1. In the succeeding section, we will elaborate on methods which utilize MR model and attention and propose a deep learning based fusion architecture which uses low-level feature correspondence and high-level semantic correspondences to alleviate the challenges aforementioned. We propose a framework that harnesses the discriminative power of the scene encoding is elaborate and empowered with the attention mechanism to enhance sharp discriminative scene understanding and promote the right level of information flow between the two sensor modalities for enunciating the modality interdependencies [15].

The fusion process strives to extract and disseminate the standout features in a given sensor data modality and combine it with relevant information from other modalities to fill in the gaps using data-centric methodologies. In this proposal, we identified the need to infuse the multiple sensor output data and discussed the two distinct class of methodologies for sensor data fusion. For this end, we define the fusion operation as a simple operation: given a set of sensor datamodalities S = {Sm} which reports sensor data {Sp} for p sensormodality. The fusion process aims to integrate these data modalities in such a way that it results in a certain level of quality and reliability of the generated fused data. The data modalities relevant to our work are generally the sensor data modalities of RD and PC and numerous fusion research works are based on these sensor modalities [24].

Deep Learning-based Sensor Fusion Approaches

The integration of sensors can augment each otherâ€™s capabilities and enhance the overall perception performance significantly. In recent years, the focus of deep learning on multi-modality has evolved from late fusion and early fusion schemes to shared representation learning for the final layers to predict common decision tasks. With radar and camera sensor data, rotation invariant object detection was proposed in the YOLO head; a point-wise single anchor last upper 3D YOLO method for directly detecting a box that fully includes an object and a point-wise anchor on the 3D YOLO one-stage object detector in the comm area were used, which demonstrated the feasibility of helpless and accurate 3D structure object detection of such sensors with high robustness against adverse effects of other objects. It is observed by the competition of video and intensity space through a mutual attack loss Pseudo SIAMESE training is presented and achieved a consensus on the correlation of camera and radar modalities that significantly improved the understanding of the obstacles [25].

Multi-sensor fusion has become the standard approach for robust and safe object perception in autonomous vehicles. In this article, we focus on fusion approaches that integrate data from cameras and LiDAR sensors [15]. Those modalities differ in their inherent pros and cons: cameras yield rich information in terms of resolution and include color and semantic knowledge, but it also contains object appearance, which makes it hard to model occlusion and sparsely seen objects in these data. LiDAR provides the 3D geometrical shape and location in a compact form; in contrast to cameras, it has no strong issues in difficult lighting, as well as it can be used to detect and locate objects based on geometry. Other undersurface sensors like ultrasonic emit radiation while in contrast to the mentioned active sensors, short/visional distance cameras, and forwardlooking mono/stereo cameras do not send any radiation [10]. Also there are other active sensors like voxel lidar and radar that can work in HMI light or in the presence of rain, fog, DRIZZLE, and DRIZZLE, which, for example, lidar, can not work due to helophast.

Feature-level Fusion

The methods belonging to the class of DG have slightly less diversity, and gap of proposals referring to decisionmaking - almost entirely up to now to simple object detection tasks. Of course, each class of methods has its relevant selected articles, which according to the own authors are better than the others. In our survey, the selection of specific articles was left to the competence of the qualified reviewers. Using more variety in articles distribution for those subcategories, which less worked within the state-of-the-art approaches, would not have significantly influenced the boundary measurements of specificity or reproducibility of the results obtained on selected datasets. Our intention would have been to increase the informative aspect of our findings.

Regardless of the chosen sensor(s), the decision-making in all of the above techniques is generally based on two different stages. The main and local uv space point features are first generated separately from each sensorâ€™s point cloud to represent the 3D point cloud. For instance, the global uv space point features are generated by merging and projecting local uv space point features corresponding to predefined local neighborhood centroids [3]. In the second stage, all features are taken into account together or sequentially for efficient and accurate final decision-making. The literature includes a significant and growing share of works, which propose to apply SE to LVIS. In general, thematically diversified proposals of this type suggest covering processes of registration, object detection, segmentation, instance segmentation, probabilistic estimation of objectsâ€™ classes, and so on.

Decision-level Fusion

Given previously obtained sensor information, the different decision level fusion methods will be summarized next. Late fusion, also known as decision fusion, decision level fusion; Probability, evidence, or belief fusion; and Marginalization over sensor data, fusing directly the sensor output decision itself. Here the sensor data are processed separately, and their scores are merged to substitute/confuse/substantiate the final decision by the sensor fusion method. Feature level fusion, also known as productivity, selection level fusion fuses extracted features by the sensors. These extracted features are merged for decision-making and the merged features derived from the feature level fusion are used for generating and fusing sensor output decisions in the next step. Data level fusion directly fuses sensor data without explicitly considering the output from feature level fusion. Related sensor data are merged to create a new sensor data stream for the fusion variables.

[8] [4] It is worth noting that the ability of the sensor suite in a level 5 scenario in the current state-of-the-art is still unveried. One approach to reduce the uncertainty relating the quality of the perception results is to develop different localisations algorithms to reduce the dependence on a specific sensor modality. Additionally, redundancy of data can potentially be used for a system that follow the principle of negative redundancy, meaning to be able to

differentiate between a large range of normal situations and a small range of highly critical situations where the perception function has an out of specification performance. Late fusion allows also the reuse of simple algorithms for detecting corner cases. Therefore, this kind of fusion has the potential to increase to detectability of certain occurrences of corner cases measured by the number of runs of a given algorithm that detected an occurrence of a corner case. For this reason, the choice of the merging point could also be determined by additional settings where the decision for a certain position is also dependent on the usage probabilities of the various fusion parameters for the generated set of training data making the method more robust in its applicability in various scenarios [23].

Sensor-level Fusion

Lidar data provides us with local elevation grids for the position of the vehicle, which can be combined with the camera images as well as the semantic segmentation maps. The received raw laser range data, concatenated with camera images and the segmentation map depth that has been produced by a depth estimation network and in this way tried to enrich the laser data semantically. Learning an asymmetric network with shared weights enriched the laser data section and increased the depth estimation accuracy further. The slam pipeline fuses the enriched laser data and odometry data from the system in real time. To increase the performance the SLAM agent is used to transfer high-level features and by fine-tuning a third shallow network on the low-level features, state-of-the-art performance is reached. A proposed framework that uses an intuitive way to fuse the lidar and the visual odometry and and multi-modal learning to enrich the information of the lidar point cloud. It fuses the depth features of the lidar by cascading the shared ego-motion and SegNet encoder [26].

In autonomous systems that operate on the roads, sensors are responsible for perceiving local and global environments. The sensor data varies on data dimensionality, accuracy, speed, and different physical properties. Different kinds of deep learning models take all of the input features for fusion. Fusion of all of the input features may improve the fused sensor measure in various cases, but it has limitations. For instance, some of the sensed features with low accuracy may contribute negatively to the fusion. The model may work perfectly for recognizing a specific object when given various kinds of inputs. The model may not cope perfectly if some input sensor measures of the features show noisy information or do not have good information about matters. In this section, we define different kinds of inspector to give the importance to sensor measures based on different strategies. The main idea of this network is to let the network learn the importance of different input sensor measures based on a scale and take care of which kind of features of the sensor data is important and which one is not. Since it maps everything to the same tempered feature this IP network does not create any extra dimension that would cost us a longer working time. The reason we have created this IP network is that even with everything present in the input sensor measures it may be the case that some features provide more information than the other kind of features. This network trains the importance of each kind of feature in the feature space by considering all of the potential features during the learning process. The importance of sensor data presented in the input features may be special in different events. Sometimes, each one of the sensor data helps our model during the learning process to have a good decision by fusing the data in good manner [15].

Evaluation Metrics and Performance Analysis

In order to make further advancements in deep multi-modal object detection and classification methods, it may prove effective to take into consideration the range-image modality in future research efforts regarding 3Dtarget objects. Apart from this, few works on the handling of instances from an on-board perspective specifically deep multi-modal detections but if it does, it is usually evaluated on KITTI dataset to validate for autonomous driving. To help secure progress in this area, the researchers suggest developing not merely multi-modal, but multi-state data and deploying a complete state network on suitable datasets that span multiple techniques all at once. Nevertheless, the availability of larger scale 3D instance segmentation datasets or fine-grained datasets would likely benefit the domainâ€™s research. Consequently, the task not only includes the choice of the methods, but also the modification of the image fusion and network design depending on the method, leading to the problem of the limitation of data and overfitting of the model [8]. With deeper models more data are also required.

The challenges and performance evaluations of the sensor fusion methods are usually measured by using metrics like the Mean Average Precision (mAP), accuracy, Euclidean distances. Other metrics measure the precision and recall for the scenario both for the detection and segmentation, and if needed, coverage, since it captures missing detections in depth like classic metrics do for the detection. Since the inception of deep learning (DL) in

2006 by Krizhevsky et al., the performance of DL-object detection and multi-modal perception stack with sensor fusion for autonomous driving has seen remarkable progress. Based on this review, several open challenges in deep multi-modal object detection have been identified.

Applications in Autonomous Driving

Another study has been reported by Soltani et al. (2019) who suggested BCM-LSTMs for semantic fusion between camera images and LiDAR point clouds in order to use this fused data in object detection and imagebased localization in an autonomous vehicle localization and mapping scenario. Another work by Wan et al. (2019c) implemented a deep learning algorithm for fusing camera images and LiDAR point clouds for object detection in multiple object track Fusion Task Models (FTM) in the autonomous vehicle peripheral. The fusion of camera images and LiDAR data in combination with precise XYZ distance information in this context seems to be a promising sensor fusion technique for developing future perception systems in autonomous vehicles.

[4] [1] Autonomous driving applications for DL-based sensor data fusion and perception enhancement have been investigated in a number of recent works. One approach has been implemented by Hu et al. (2016) as a Residual Sensor Fusion Deep Reinforcement Learning (RSF-DRL) in a CARLA-based open test facility for exploring the regional advantages of sensors and for investigating how to apply the RSF-DRL scheme for autonomous vehicle localization and navigation. In this approach, the sensor measurements are treated as raw input for further processing inside the DRL network.

Challenges and Future Directions

Robustness of data model in context: In this paper related to various context-aware sensor fusion models including HydraFusion is, which refers that various problems in partial senor data in various environments including bad weather, self-shadow and occlusion. It proposes the dynamic adjustment between the early fusion and late fusion techniques of the multi-sensor data for the robust and efficient sensor data fusion, the UKFbased Bayesian nonparametric data association (DNA) approach to handle the corresponding short-term trackers within the proposed framework. Also [7] refers the challenges and discuss the advancement of earlier model and consider the semantic severance influence of the feature fusion in uncertainty status dataset. Also, it consider about the disassembling diverse feature branch extraction and improve the inference modules association from the filter level. In this direction, solutions of robustness and efficiency of model can be improved by bayesian probability and reinforced learning frameworks for deep learning-based sensors. Also improve the semantic segmentation and adversarial data generative network to generate contextual data for data model robustness and accuracy. These robust and efficient deep learning based sensor fusion propagation model is proposed by multimodal data detection and tracking for more accurate detection of moving object.

In this chapter that deep learning is used as the key technology to improve sensor data fusion in the autonomous vehicles. It is because the major challenge for autonomous vehicle (AV) perception is the robustness and accuracy of perception models in various driving scenarios, which can be mainly collected from the sensors. Although the deep learning-based perception model leads a solution to the problem, the robustness and the efficiency of the perception models have an even greater demand. Furthermore, the light-weight deep learning model and real-time data perception module have become a basic demand in practice, it is because the limitation computation resource of vehicles and that light-weight deep learning-based perception model can increase the real-time performance to minimize the computation time. Therefore, two words from the challenges (robustness and accuracy) lead our discussion related to two directions in future, deep learningbased sensor fusion in context which is discussed in [27] for robustness and light-weight deep learning-based detection and tracking module, which is discussed in [3]. In this chapter, the challenges and the future directions of the deep learning-based sensor fusion will be summarized and proposed.

Conclusion and Future Work

The rapid development in machine learning technology, including deep learning (DL), has greatly accelerated the growth of research into intelligent vehicle perception, which is essential for the safe and reliable deployment of autonomous vehicles [4].Important features are still being extracted from raw sensor data by hand-crafted heuristics, i.e., the various engineered rules, and then the machine learning models train to fit on these manual rules, which generally perform poorly in practice and for conditions not captured by algorithmic design. Meanwhile, end-to-end systems designed to learn self-driving policies directly from pixel-level camera inputs, such as the pioneering work by Bojarski et al. have shown this approach to be successful and promising key ingredient of a future end-to-end system in autonomous vehicles.

In this paper, we present a novel sensor fusion method and architecture for an end-to-end perception system in autonomous vehicles, which uses deep learning (DL) for both temporal and spatial semantic fusion, called DeepSTEP, to accurately and robustly process the input sensor data acquired from the noisy real-world environment [6]. The proposed DeepSTEP model can predict the semantic class, instance and trajectory information of surrounding objects simultaneously and accurately. Furthermore, the overall perception process in DeepSTEP can be highly optimized for Ivybridge, because it is based on a single-stage detector, MV3DHead, which directly takes raw radar and camera data as inputs. In this paper, we write a comprehensive, background research on vehicle perception and sensor fusion, and discuss the opportunities and challenges in learningbased 3D object detection and tracking. Furthermore, we present DeepSTEP, a deep fusion network to optimally combine deep features from all input sensors, including camera and radar data, in both spatial and temporal domains.

References

1. Y. Cui, R. Chen, W. Chu, L. Chen et al., "Deep Learning for Image and Point Cloud Fusion in AutonomousDriving: A Review," 2020. [PDF]
2. A. Jalal Aghdasian, A. Heydarian Ardakani, K. Aqabakee, and F. Abdollahi, "Autonomous Driving usingResidual Sensor Fusion and Deep Reinforcement Learning," 2023. [PDF]
3. Q. Zhang, X. Hu, Z. Su, and Z. Song, "3D car-detection based on a Mobile Deep Sensor Fusion Model andreal-scene applications," 2020. ncbi.nlm.nih.gov
4. M. Rahimi, H. Liu, I. Durazo Cardenas, A. Starr et al., "A Review on Technologies for Localisation andNavigation in Autonomous Railway Maintenance Systems," 2022. ncbi.nlm.nih.gov
5. T. L. Kim and T. H. Park, "Camera-LiDAR Fusion Method with Feature Switch Layer for Object DetectionNetworks," 2022. ncbi.nlm.nih.gov

6. S. Huch, F. Sauerbeck, and J. Betz, "DeepSTEP -- Deep Learning-Based Spatio-Temporal End-To-EndPerception for Autonomous Vehicles," 2023. [PDF]

7. D. Feng, C. Haase-SchÃ¼tz, L. Rosenbaum, H. Hertlein et al., "Deep Multi-modal Object Detection andSemantic Segmentation for Autonomous Driving: Datasets, Methods, and Challenges," 2019. [PDF]

8. J. Fayyad, M. A. Jaradat, D. Gruyer, and H. Najjaran, "Deep Learning Sensor Fusion for Autonomous VehiclePerception and Localization: A Review," 2020. ncbi.nlm.nih.gov

9. S. Yao, R. Guan, X. Huang, Z. Li et al., "Radar-Camera Fusion for Object Detection and SemanticSegmentation in Autonomous Driving: A Comprehensive Review," 2023. [PDF]

10. Tatineni, Sumanth. "Compliance and Audit Challenges in DevOps: A Security Perspective." *International Research Journal of Modernization in Engineering Technology and Science* 5.10 (2023): 1306-1316.

11. Vemori, Vamsi. "Evolutionary Landscape of Battery Technology and its Impact on Smart Traffic Management Systems for Electric Vehicles in Urban Environments: A Critical Analysis." *Advances in Deep Learning Techniques* 1.1 (2021): 23-57.

12. Mahammad Shaik. "Rethinking Federated Identity Management: A Blockchain-Enabled Framework for Enhanced Security, Interoperability, and User Sovereignty". *Blockchain Technology and Distributed Systems*, vol. 2, no. 1, June 2022, pp. 21-45, https://thesciencebrigade.com/btds/article/view/223.

13. Vemori, Vamsi. "Towards a Driverless Future: A Multi-Pronged Approach to Enabling Widespread Adoption of Autonomous Vehicles-Infrastructure Development, Regulatory Frameworks, and Public Acceptance Strategies." *Blockchain Technology and Distributed Systems* 2.2 (2022): 35-59.

14. F. Manfio Barbosa and F. Santos OsÃ³rio, "Camera-Radar Perception for Autonomous Vehicles and ADAS:Concepts, Datasets and Metrics," 2023. [PDF]

15. C. Cui, Y. Ma, J. Lu, and Z. Wang, "Radar Enlighten the Dark: Enhancing Low-Visibility Perception forAutomated Vehicles with Camera-Radar Fusion," 2023. [PDF]

16. D. Jong Yeong, G. Velasco-Hernandez, J. Barry, and J. Walsh, "Sensor and Sensor Fusion Technology inAutonomous Vehicles: A Review," 2021. ncbi.nlm.nih.gov

17. Y. Wang, Q. Mao, H. Zhu, J. Deng et al., "Multi-Modal 3D Object Detection in Autonomous Driving: aSurvey," 2021. [PDF]

18. Z. Wang, W. Zhan, and M. Tomizuka, "Fusing Bird View LIDAR Point Cloud and Front View Camera Image forDeep Object Detection," 2017. [PDF]

19. Q. V. Lai-Dang, J. Lee, B. Park, and D. Har, "Sensor Fusion by Spatial Encoding for Autonomous Driving,"2023. [PDF]

20. H. Xu, Y. Gao, F. Yu, and T. Darrell, "End-to-end Learning of Driving Models from Large-scale VideoDatasets," 2016. [PDF]

21. T. Lai, "A Review on Visual-SLAM: Advancements from Geometric Modelling to Learning-Based SemanticScene Understanding Using Multi-Modal Sensor Fusion," 2022. ncbi.nlm.nih.gov

22. D. Dworak, M. Komorkiewicz, P. Skruch, and J. Baranowski, "Cross-Domain Spatial Matching for Camera andRadar Sensor Data Fusion in Autonomous Vehicle Perception System," 2024. [PDF]

23. B. Shahian Jahromi, T. Tulabandhula, and S. Cetin, "Real-Time Hybrid Multi-Sensor Fusion Framework forPerception in Autonomous Vehicles," 2019. ncbi.nlm.nih.gov

24. M. Dibaei, X. Zheng, K. Jiang, S. Maric et al., "An Overview of Attacks and Defences on Intelligent ConnectedVehicles," 2019. [PDF]

25. Z. Wei, F. Zhang, S. Chang, Y. Liu et al., "MmWave Radar and Vision Fusion for Object Detection inAutonomous Driving: A Review," 2022. ncbi.nlm.nih.gov

26. A. Jafar Md Muzahid, S. Fauzi Kamarulzaman, M. Arafatur Rahman, S. Akbar Murad et al., "Multiple vehiclecooperation and collision avoidance in automated vehicles: survey and an AI-enabled conceptual framework," 2023. ncbi.nlm.nih.gov

27. F. Heidecker, J. Breitenstein, K. RÃ¶sch, J. LÃ¶hdefink et al., "An Application-Driven Conceptualization ofCorner Cases for Perception in Highly Automated Driving," 2021. [PDF]

28. D. Xu, H. Li, Q. Wang, Z. Song et al., "M2DA: Multi-Modal Fusion Transformer Incorporating Driver Attentionfor Autonomous Driving," 2024. [PDF]

29. F. Jibrin Abdu, Y. Zhang, M. Fu, Y. Li et al., "Application of Deep Learning on Millimeter-Wave Radar Signals:A Review," 2021. ncbi.nlm.nih.gov

30. C. Chen, S. Rosa, C. Xiaoxuan Lu, B. Wang et al., "Learning Selective Sensor Fusion for States Estimation,"2019. [PDF]

31. A. Vaibhav Malawade, T. Mortlock, and M. Abdullah Al Faruque, "HydraFusion: Context-Aware Selective Sensor Fusion for Robust and Efficient Autonomous Vehicle Perception," 2022. [PDF]