Deep Learning Architectures for Real-Time Sensor Fusion in Autonomous Vehicles: CNNs and Their Role in Camera and LIDAR Data Integration

2441638

Abstract

This report explores the use of Convolutional Neural Networks (CNNs) for real-time sensor fusion in autonomous vehicles, focusing on the integration of camera and LiDAR data. Autonomous vehicles depend on various sensors to perceive their surroundings and make decisions, with camera and LiDAR being central for capturing visual and spatial data. However, these sensors produce different types of data, which require effective fusion methods to ensure accurate environmental perception. Due to their ability to automatically learn and merge complex patterns from diverse data sources, CNNs are used for this task. Key fusion techniques (early, mid, and late fusion) along with recent advancements, such as hybrid CNN-transformer models, are discussed. The study concludes that CNN-based sensor fusion, optimised for speed and accuracy, is crucial for the development of reliable autonomous vehicles capable of navigating complex, real-world environments.

Keywords: Sensor Fusion, Autonomous Vehicles, Convolutional Neural Networks, LiDAR, Camera, and Real-Time Processing

Contents

1. Introduction

The use of automated vehicles (AVs) has been gaining worldwide popularity. AVs have the transportational capabilities of conventional vehicles and can perceive the environment and self-navigate with minimal human intervention (Yeong et al., 2021). This growing popularity is fuelled by the environmental, social, and economic advantages such as increased road safety, lower transportation costs, and reduced energy consumption (Geary and Danks, 2019). AVs utilise sensors that try to perceive the world in the way humans would perceive the world. The data from these sensors are combined using multi sensor fusion to enhance the perceived subject and create a more reliable and accurate perception of the vehicle's surroundings (Zhangjing Wang and Niu, 2020). There are three main levels of sensor fusion that work to enhance perceptions: **Low Level**, **Mid Level** and **High Level** fusion. LIDAR (Light Detection and Ranging) and cameras are two types of sensors that will be discussed in this report in the context of real-time sensor fusion.

A core problem in AV navigation is the incorporation of diverse sensor data streams that can recreate an accurate mapping of the vehicle's environment in real-time. Camera and LiDAR sensors provide different types of data. A struggle between the different formats, resolutions and temporal dynamics can complicate the unification of LIDAR and camera sensors that make decisions for navigation and obstacle avoidance.

Convolutional Neural Networks (CNNs), a DL (Deep Learning) model, is a primary technique used to combat this sensor fusion problem, particularly for camera and LiDAR data. CNNs can process image-like data, making them ideal for interpreting both 2D camera images and the spatial structure of 3D LiDAR point clouds (Melotti et al., 2020). Real-time processing requirements are important because autonomous systems must rapidly fuse and interpret data from cameras, LiDAR, and other sensors, to make split-second decisions in dynamic environments. This demands high computational efficiency and low-latency processing to avoid delays in decision-making.

Additionally, handling large data streams from high-resolution cameras and dense LiDAR point clouds in real-time can overwhelm traditional processing units. Efficient algorithms and hardware accelerators are needed to manage, compress and prioritise data (Melotti et al., 2020).

CNNs use DL to automatically learn complex patterns and correlations between the two data types without relying on manual engineering features (Zhaoa et al., 2020). This ability to generalise from large datasets is the key to handling the diverse scenarios that autonomous vehicles encounter.

2. Sensor Fusion

According to the World Health Organization, in the European Union, there were more than 40,000 road fatalities, of which 90% were caused by human error (Yeong et al., 2021). Sensor technology can be used to reduce this number by quickly reacting to changes in situations where the driver of the vehicle may be idle.

The use of sensors in AV is crucial because each sensor has a strength and a limitation. Cameras excel at visual recognition but perform poorly in low light, while LiDAR provides accurate 3D spatial data but may struggle in adverse weather 1. Combining the data from these sensors can compensate for weaknesses, improving perception and decision-making (Yeong et al., 2021).

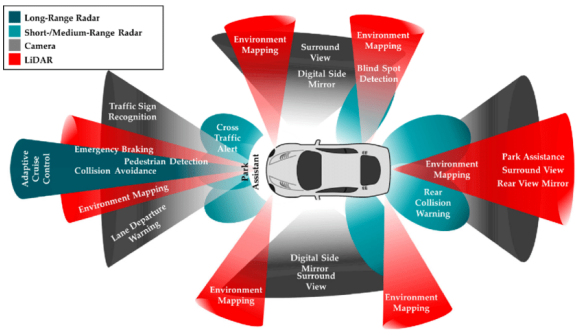

Figure 1: Type and positioning of sensors in an automated vehicle (Yeong et al., 2021)

Research describes sensor fusion as early-level, mid-level and late-level fusion (Levinson et al., 2007; Lejbølle et al., 2017). Early fusion merges raw sensor data from both modalities at the input stage, allowing a single CNN to process it. This method integrates information immediately but struggles with the drastically different data formats of camera and LiDAR, leading to inefficiencies in real-time performance (Levinson et al., 2007). Late fusion, in contrast, processes the camera and LiDAR data separately using two CNNs, before merging their outputs. This ensures that each data modality is handled optimally but can lead to the loss of cross-modal interactions (Lejbølle et al., 2017). Mid-level fusion is considered a balanced approach, it extracts features from each sensor independently and combines these features in intermediate layers of the network (Müller et al., 2008). This method has been successful in enhancing 3D object detection without sacrificing computational efficiency.

Sensor fusion quantitatively measures changes in the environment, and classifys them for processing into **proprioceptive sensors** or **exteroceptive sensors**. This classification is based on their operational principle (Levinson et al., 2007).

Proprioceptive, or internal state sensors, measure

the internal dynamics of a system, such as force, angular rate, or battery voltage. Examples include inertia measurement units , encoders, and global navigation satellite system receivers. In contrast, exteroceptive, or external state sensors, gather information from the environment, like distance or light intensity. Examples are cameras, radar, LiDAR, and ultrasonic sensors. Sensors can be passive (receiving energy from the environment, like vision cameras) or active (emitting energy and measuring the response, like LiDAR)(Yeong et al., 2021).

2.1. LiDAR Technology

LiDAR are range and motion sensors. They provide dense data that is composed of an inaccurate depth-of-field of surrounding objects and a 360° view of our surroundings (Roriz et al., 2022). Both machine learning and DL are used to improve the accuracy of the dense data and outputs 3D information about the environment by using the time-of-flight (TOF) principle 2 (Du et al., 2017; Royo and Ballesta-Garcia, 2019). TOF is the measurement of the time between events such as backscatter energy from a pulsed light beam. Using the speed of light in air, distances can be calculated or mapped (Royo and Ballesta-Garcia, 2019). Using real-world data from an entire point cloud sequence, DL calculates the distance to objects using the delay between emitting laser pulses and receiving the reflected signal (Du et al., 2017).

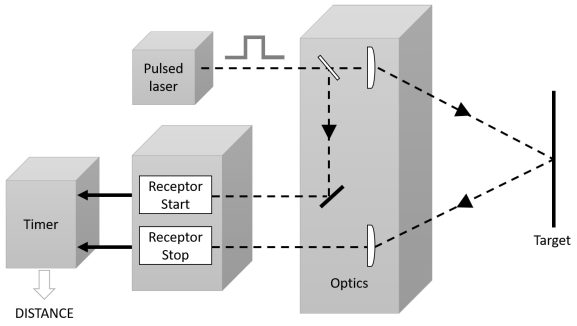

Figure 2: Pulsed time-of-flight (TOF) measurement principle (Royo and Ballesta-Garcia, 2019)

The system then generates a 3D point cloud by providing x, y, and z coordinates to output precise spatial data of the vehicle's surroundings. LiDAR excels at depth perception, making it ideal for detecting obstacles, estimating free space, and creating high-resolution maps (Li and Ibanez-Guzman, 2020).
Using a DL neural network like Pointpillar, objects can be detected in a environment. PointPillars exemplifies the power of CNNs in processing LiDAR data. They transform sparse 3D point clouds into a dense 2D grid (pseudo-image), which can then be processed by a CNN (Stanisz et al., 2020). This method has shown significant promise in reducing computational complexity while maintaining high detection accuracy. Furthermore, it allows for an extension to incorporate camera data, effectively using sensor fusion to improve the performance of object detection in autonomous vehicles (Tu et al., 2021; Stanisz et al., 2020).
[1]

There are different techniques used in LiDAR systems. The simplest method uses pulsed light. Short bursts of laser are emitted, and the echo's return time is used to calculate distances, achieving centimetre-level accuracy (Li and Ibanez-Guzman, 2020). The measured time presents as twice the distance to the object as light travels to the target and back. Therefore, it needs to be divided by two to get the actual range to the target.

$$R = \frac{c}{2} t_{oF},$$

Where R is the range to the target, c is the speed of light ($c = 3 \times 10^8\, m/s$) in free space, and t_{oF} is the time it takes for the pulse of energy to travel from its emitter to the observed object and then back to the receiver (Royo and Ballesta-Garcia, 2019).

Another approach, frequency-modulated continuous-wave (FMCW), uses frequency shifts to measure both distance and velocity, offering high precision and the ability to track moving objects (Wu et al., 2024).The signal is sent to the target, and the reflected signal that reaches the receiver t_{oF} is combined with a reference signal from the emitter (Wu et al., 2024).

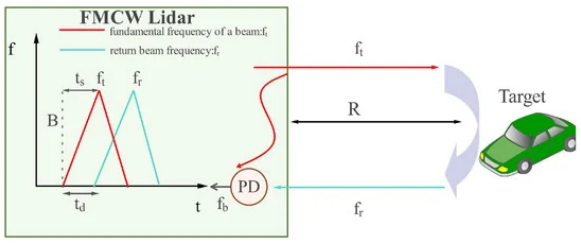

Figure 3: FMCW ranging method (Wu et al., 2024)

A third method is the amplitude modulation of continuous waves (AMCW), which compares the phase difference between the emitted and reflected signals to measure distance but is more limited in range. The detector collects the reflected light signal, and the distance R is determined from the phase shift $\Delta\Phi$ between the emitted and received signals (Li and Ibanez-Guzman, 2020).

[1] A link to the Pointpillar technique is available at:https://github.com/swordlidev/Vote3Deeplidar/blob/master/run.py

$$\Delta\Phi = kMd = \frac{2\pi f M c^2}{R} \Rightarrow R = \frac{c^2 \Delta\Phi}{2\pi f M}$$

The wave-number kM corresponds to the modulation frequency, d is the total distance travelled, f_M is the modulation frequency of the signal's amplitude, R and c are the range to the target and the speed of light respectively (Royo and Ballesta-Garcia, 2019).

FMCW method is generally considered superior for LiDAR applications that require both high-precision measurements and the ability to track moving objects. While the pulsed-light method offers centimetre-level accuracy in distance measurement, it only provides information on distance without velocity, which can be limiting in dynamic environments. FMCW, on the other hand, measures both distance and relative velocity by analysing frequency shifts, making it ideal for applications involving moving targets, such as autonomous driving or robotics (Wu et al., 2024). The ability to simultaneously track distance and velocity allows FMCW-based systems to make faster, more complex adjustments in response to environmental changes, giving them a functional edge over simpler pulsed-light methods.

AMCW lacks the range and precision achievable by FMCW. Although effective for short-range applications, AMCW's limitations in range make it less suitable for long-distance detection where more detailed data is required (Li and Ibanez-Guzman, 2020). An example highlighting FMCW's advantages is its usage in advanced automotive LiDAR systems, which must identify the speed and direction of other vehicles to make safe, autonomous driving decisions.

Most LiDAR technology operate in the 905 nm wavelength range, balancing safety with accuracy (Royo and Ballesta-Garcia, 2019). However, LiDAR has limitations, particularly in harsh weather conditions like fog or heavy rain, where its performance can degrade (Wu et al., 2024) Additionally, LiDAR lacks the ability to capture colour or texture, which makes it less effective at identifying specific objects like road signs or pedestrians (Li and Ibanez-Guzman, 2020). Despite these challenges, LiDAR remains a core component in autonomous vehicle sensor suites due to its accuracy and depth-sensing capabilities.

2.2. Camera Technology

Cameras are one of the most widely adopted technologies for perceiving the environment in autonomous vehicles (AVs). They work by detecting light from surroundings on a photosensitive surface (image plane) through a camera lens to produce clear images 4 (Wang et al., 2022). Cameras, when paired with appropriate software, can detect both moving and static obstacles within their field-of-view, providing high-resolution images of the surroundings (Ceccarelli and Secci, 2023). This makes them highly effective for recognising road signs, traffic lights, lane markings, and other objects in both road traffic and off-road environments.

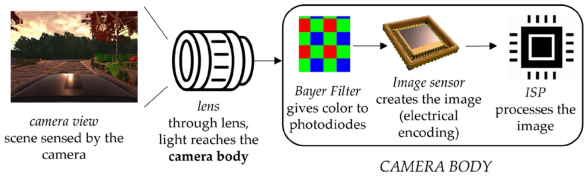

Figure 4: A camera consists of a lens and a body, which includes a Bayer filter, image sensor, and image signal processor (Ceccarelli and Secci, 2023).

Most cameras in AVs use a Bayer filter, a colour filter array placed over the image sensor to capture colour information (Xie et al., 2004). Each pixel on the image sensor detects only one of the three primary colours (red, green, or blue), and the Bayer filter alternates these colours in a mosaic pattern. The camera's image processing system then uses demosaicing algorithms to combine the information from neighbouring pixels and reconstruct a full-colour image (Xie et al., 2004). While this allows for high-quality colour capture, it limits each pixel to recording only one colour, affecting the fine details in some cases.

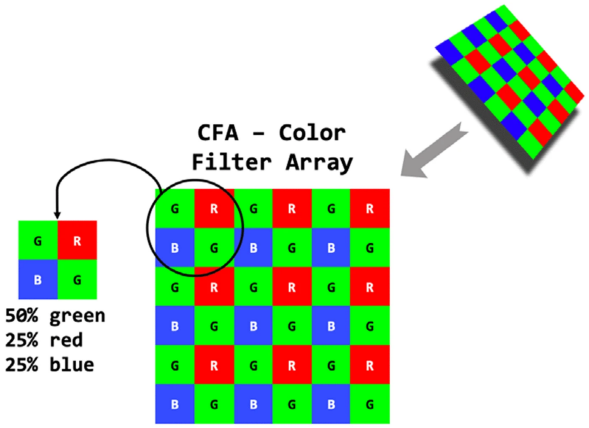

Figure 5: Bayer filter. On the left, a cell of 2x2 photosites (BGGR pattern). In digital imaging, a colour filter mosaic, is a mosaic of tiny colour filters placed over the pixel sensors of an image sensor to capture colour information (Northrup, 2016)

In AVs, two primary types of camera systems are used: **monocular** and **stereo cameras**. Monocular cameras use a single camera to capture a series of 2D images. Conventional red, green, blue monocular cameras are limited in that they do not natively capture depth information (Farah-

nakian and Heikkonen, 2020). Although depth can be estimated through complex algorithms, such as those using dual pixel autofocus technology, monocular systems often lack the spatial accuracy provided by other methods.

Stereo cameras, also known as binocular cameras, overcome this limitation by simulating depth perception. These systems use two cameras placed side-by-side to capture slightly different images, similar to how human eyes perceive depth (Farahnakian and Heikkonen, 2020). The disparity between the two images is then used to compute depth maps, enabling stereo cameras to perceive 3D spatial information. For example, Orbbec 3D cameras have been used in AVs, with their stereo vision processing aiding in simultaneous localisation and mapping (**SLAM**) for autonomous navigation (Yeong et al., 2021).

Other specialised cameras, like fisheye cameras, are often used for near-field perception tasks, such as parking assistance and traffic jam handling. These cameras provide a wide field of view and can create a 360° panoramic view when multiple fisheye cameras are combined (Sekkat et al., 2022). However, they suffer from optical distortions, such as barrel distortion, which can affect the accuracy of obstacle detection unless corrected through camera calibration.

The visual data provided by cameras, enhanced by Bayer filter processing, can be used for tasks such as object detection, lane recognition, and semantic segmentation. Unlike LiDAR, which measures distances directly, cameras capture 2D images but offer essential contextual information like colour, texture, and object appearance (Kocić et al., 2018). This makes them highly valuable for recognising traffic signs, detecting pedestrians, and identifying road markings which are vital for safe navigation.

However, cameras do have limitations. Their performance is highly dependent on external factors such as lighting, weather, and occlusion. In low-light conditions or during heavy rain or fog, the quality of images captured can degrade significantly, making it difficult to detect objects reliably (Kocić et al., 2018). Additionally, since cameras capture 2D images and lack intrinsic depth information, they must be paired with other sensors, such as LiDAR, to accurately perceive the spatial layout of the surroundings.

3. Convolutional Neural Networks

Convolutional Neural Networks (CNNs) play a key role in autonomous vehicles. They extract critical features from multi-modal sensor data, such as 2D images from cameras and 3D point clouds from LiDAR sensors. CNNs automatically identifies and fuses patterns like edges, shapes, and textures in camera data, and structural features in LiDAR data, making them highly effective for environment perception (Li et al., 2022). This fusion can happen early, by combining raw data or late, after each sensor's data has been processed by its respective CNN layers. For example, PointNet handles raw LiDAR point clouds directly, while VoxelNet converts point clouds into voxel grids for processing by 3D CNN layers.

The architecture of CNNs, inspired by biological visual perception, uses layers of filters, called kernels, to detect and process features without the need for manual feature selection (Malsburg, 1986). When processing multi-modal data, CNNs utilise convolution layers to extract features, activation functions (like Rectified Linear Units) to introduce non-linearity, and pooling layers to reduce the size of feature maps, mitigating overfitting and improving efficiency.

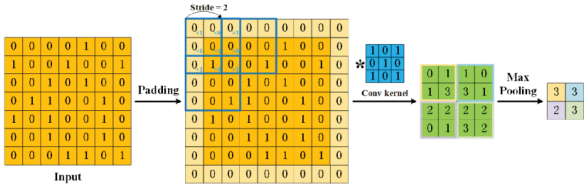

Figure 6: 2D CNN Model

For object detection, CNN-based fusion of camera and LiDAR data enhances the system's accuracy. One notable method is the Frustum PointNet, which relies on 2D camera-based object detection to create 3D frustums that focus on relevant LiDAR points (Wang et al., 2020). These points are then processed by the PointNet architecture to classify objects and predict 3D bounding boxes (Li et al., 2022). This approach leverages both the high resolution of camera images and the depth precision of LiDAR. However, its reliance on accurate 2D camera detections can be a limitation in poor lighting or occlusion.

Traditional methods like Haar cascade's object detection are also used in some image processing tasks. Haar cascades are a machine learning-based approach that can detect objects by scanning an image at various scales. The method works by identifying features, such as edges and changes in contrast, using Haar-like features (simple rectangular patterns) (Cuimei et al., 2017). Although less computationally intensive than CNNs, Haar cascades can be prone to errors in more complex scenes and may not generalise well to various lighting conditions or object orientations. However, they can still be effective for specific tasks such as face detection or identifying simple objects in controlled environments. In autonomous vehicles, Haar cascades could potentially serve as a supplementary technique for de-

tecting basic shapes or objects, but they are typically outperformed by CNNs in terms of flexibility and accuracy.

A more robust solution is seen in hybrid models like MVX-Net, which extracts features separately from both the camera and LiDAR data using independent CNNs, then fuses them at mid-level network layers (Kumar et al., 2024). This allows the model to balance efficiency and the richness of the features from both modalities, making it especially suitable for complex environments with a variety of objects and scene structures.

Despite their advantages, CNNs can be computationally expensive and require significant data and hardware resources. Platforms like NVIDIA Drive PX, Tesla's Full-Self Driving (FSD) chip, and Waymo's perception systems are optimised for real-time sensor fusion and decision-making in autonomous vehicles, supporting the high processing demands of CNN models handling both 2D and 3D data.

While CNNs are highly effective, methods like Haar cascade object detection can still complement CNNs for specific, less complex detection tasks, especially where real-time processing speed is crucial, and simpler features need to be recognised quickly. However, as autonomous systems become more advanced, CNN-based models are more commonly used due to their robustness, flexibility, and ability to handle a wide variety of conditions.

Recent advancements, such as deformable convolutions, improve CNNs' ability to handle irregular object shapes, further enhancing their performance. Studies by Dai et al., (Dai et al., 2015) and Hu et al. (Hu et al., 2018), continue to push the boundaries of how CNNs are structured and used for sensor fusion.

3.1. Commercial Implementations

Commercial implementations of sensor fusion in AVs focus on combining data from LiDAR and cameras using CNNs to enhance real-time environmental perception. Tesla's Full Self-Driving chip, for example, relies heavily on camera data, using CNNs to interpret images and identify objects, lane markings, and traffic signals. It largely avoids LiDAR due to cost and data processing demands, instead using multiple cameras to create a 3D perception of the vehicle's surroundings.

In contrast, Waymo's autonomous systems use a more comprehensive approach, integrating both LiDAR and cameras to deliver a precise, multi-dimensional understanding of the environment. Waymo's system, supported by NVIDIA Drive PX hardware, uses mid-level sensor fusion, combining CNNs for camera data and specialised networks for LiDAR data to improve object detection. Tesla's camera-first approach, which optimises cost and power whereas, Waymo's combined LiDAR and camera method, focuses on spatial accuracy. Both methods highlight the trade-off between cost and performance, where the choice of sensor fusion approach directly impacts vehicle capabilities in different driving scenarios.

4. Improvements

To improve performance in CNN-based applications, several strategies can be implemented. A promising approach is to combine CNNs with other DL models such as transformers or recurrent neural networks (RNNs). Hybrid networks can leverage the strength of CNNs in spatial data processing and the sequential capabilities of RNNs or transformers, making them better suited to handle time-series data or video streams. This integration not only enhances sequential data management but can also help reduce computational load, enabling faster decision-making in real-time systems.

Another potential advancement lies in developing new fusion algorithms that integrate multiple data streams, such as sensor fusion for autonomous vehicles or multi-modal systems. Techniques like graph-based learning or attention mechanisms could dynamically prioritise data sources depending on the environmental conditions. For example, in low-light driving conditions, attention mechanisms could prioritise LiDAR or infrared data over standard camera input, optimising for specific challenges based on real-time data analysis.

Optimising CNNs for faster performance on edge devices is a key challenge. Techniques like quantisation, pruning, and model compression are effective strategies. These methods allow CNN models to maintain high accuracy while operating on hardware with limited processing power, making them suitable for deployment in real-time applications like mobile devices or embedded systems.

To improve the synchronisation of multi-sensor inputs, such as camera and LiDAR data, better temporal alignment techniques are essential. Advanced calibration tools can help synchronise the timestamps and data streams, ensuring that data from different sensors are aligned accurately in time. This can be achieved by using algorithms that match features across modalities and correct for discrepancies in timing or spatial orientation, improving the overall performance of tasks like object detection and tracking.

These improvements would help CNN-based systems achieve greater efficiency, accuracy, and scalability, especially in time-sensitive or resource-constrained environments.

5. Conclusion

This report has examined the role of CNN-based DL architectures in sensor fusion for AV, particularly focusing on the integration of camera and LiDAR data. Autonomous vehicles rely on sensor fusion to interpret their surroundings reliably, enabling them to navigate complex environments with reduced human intervention. However, merging data from different sensor types presents challenges due to variations in data structure, resolution, and processing requirements. LiDAR provides precise 3D spatial data, while cameras capture valuable contextual information such as colour and texture, each complementing the other's limitations. CNNs have proven to be effective for this task, automatically learning complex correlations between data types and enabling accurate real-time perception and object detection. Approaches such as early, mid, and late fusion allow flexibility in balancing computational efficiency and the quality of environmental understanding. Recent advancements, including hybrid models combining CNNs with transformers, continue to push the boundaries of sensor fusion technology, achieving more adaptive and scalable solutions. Despite high computational demands, optimisation strategies like quantisation and model compression are essential to making these architectures feasible for real-time application. Future work focusing on improved fusion algorithms and alignment techniques can further enhance CNN-based sensor fusion, improving the safety, reliability, and adaptability of AV in dynamic and challenging conditions.

Bibliography

Andrea Ceccarelli and Francesco Secci. 2023. RGB Cameras Failures and Their Effects in Autonomous Driving Applications . *IEEE Transactions on Dependable and Secure Computing*, 20(04):2731–2745.

Li Cuimei, Qi Zhiliang, Jia Nan, and Wu Jianhua. 2017. Human face detection algorithm via haar cascade classifier combined with three additional classifiers. In *2017 13th IEEE International Conference on Electronic Measurement Instruments (ICEMI)*, pages 483–487.

Jifeng Dai, Kaiming He, and Jian Sun. 2015. Instance-aware semantic segmentation via multi-task network cascades. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3150–3158.

Xinxin Du, Marcelo H. Ang, and Daniela Rus. 2017. Car detection for autonomous vehicle: Lidar and vision fusion approach through deep learning framework. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 749–754.

Fahimeh Farahnakian and Jukka Heikkonen. 2020. Rgb-depth fusion framework for object detection in autonomous vehicles. In *2020 14th International Conference on Signal Processing and Communication Systems (ICSPCS)*, pages 1–6.

Timothy Geary and David Danks. 2019. Balancing the benefits of autonomous vehicles. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, page 181–186. Association for Computing Machinery.

Jie Hu, Li Shen, and Gang Sun. 2018. Squeeze-and-excitation networks. In *Computer Vision and Pattern Recognition (CVPR)*, pages 7132–7141.

Jelena Kocić, Nenad Jovičić, and Vujo Drndarević. 2018. Sensors and sensor fusion in autonomous vehicles. In *2018 26th Telecommunications Forum (TELFOR)*, pages 420–425.

Aakash Kumar, Ajmal Mian Chen Chen, Neils Lobo, and Mubarak Shah. 2024. Sparse points to dense clouds: Enhancing 3d detection with limited lidar data. *Computer Vision and Pattern Recognition*.

Aske R Lejbølle, Kamal Nasrollahi, and Thomas B Moeslund. 2017. Enhancing person re-identification by late fusion of low-, mid- and high-level features. *The Institute of Engineering Technology*.

Jesse Levinson, Michael Montemerlo, and Sebastian Thrun. 2007. Map-based precision vehicle localization in urban environments. In *Urban Environment*.

You Li and Javier Ibanez-Guzman. 2020. Lidar for autonomous driving: The principles, challenges, and trends for automotive lidar and perception systems. *IEEE Signal Processing Magazine*, 37(4):50–61.

Zewen Li, Fan Liu, Wenjie Yang, and Shouheng Peng. 2022. A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems*, 33(12):6999–7019.

C Van Der Malsburg. 1986. Frank rosenblatt: Principles of neurodynamics: Perceptrons and the theory of brain mechanisms. In *Brain Theory*, pages 245–248, Berlin, Heidelberg. Springer Berlin Heidelberg.

Gledson Melotti, Cristiano Premebida, and Nuno Gonçalves. 2020. Multimodal deep-learning for object recognition combining camera and lidar data. In *2020 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, pages 177–182.

M Himmelsbachand A Müller, Thorsten Luettel, and Hans-Joachim Wuensche. 2008. Lidar-based 3d object perception. In *Proceedings of 1st International Workshop on Cognition for Technical Systems*.

Tony Northrup. 2016. *Tony Northrup's Photography Buying Guide: How to Choose a Camera, Lens, Tripod, Flash, & More*. Mason Press, Waterford, Republic of Ireland.

Ricardo Roriz, Jorge Cabral, and Tiago Gomes. 2022. Automotive lidar technology: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 23(7):6282–6297.

Santiago Royo and Maria Ballesta-Garcia. 2019. An overview of lidar imaging systems for autonomous vehicles. *Applied Sciences*, 9(19).

Ahmed Rida Sekkat, Yohan Dupuis, Varun Ravi Kumar, Hazem Rashed, Senthil Yogamani, and Pascal Vasseur. 2022. Synwoodscape: Synthetic surround-view fisheye camera dataset for autonomous driving. *IEEE Robotics and Automation Letters*, 7(3):8502–8509.

Joanna Stanisz, Konrad Lis, Tomasz Kryjak, and Marek Gorgon. 2020. Optimisation of the pointpillars network for 3d object detection in point clouds. In *2020 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, pages 122–127.

Jiayin Tu, Ping Wang, and Fuqiang Liu. 2021. Pp-rcnn: Point-pillars feature set abstraction for 3d real-time object detection. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8.

Chaoyang Wang, Xiaonan Wang, Hao Hu, Yanxue Liang, and Gang Shen. 2022. On the application of cameras used in autonomous vehicles. *Archives of Computational Methods in Engineering*.

Leichen Wang, Tianbai Chen, Carsten Anklam, and Bastian Goldluecke. 2020. High dimensional frustum pointnet for 3d object detection from camera, lidar, and radar. In *2020 IEEE Intelligent Vehicles Symposium (IV)*, pages 1621–1628.

Zibo Wu, Yue Song, Jishun Liu, Yongyi Chen, Hongbo Sha, Mengjie Shi, Hao Zhang, Li Qin, Lei Liang, Peng Jia, Cheng Qiu, Yuxin Lei, Yubing Wang, Yongqiang Ning, Jinlong Zhang, and Lijun Wang. 2024. Advancements in key parameters of frequency-modulated continuous-wave light detection and ranging: A research review. *Applied Sciences*, 14(17).

Xiang Xie, GuoLin Li, XiaoWen Li, Chun Zhang, ZhiHuaWang, XinKaiChen, and HsiaoWeiSi. 2004. A new high quality image compression method for digital image sensors with bayer color filter arrays. In *IEEE International Workshop on Biomedical Circuits and Systems, 2004.*, pages S3/3–13.

De Jong Yeong, Gustavo Velasco-Hernandez, John Barry, and Joseph Walsh. 2021. Sensor and sensor fusion technology in autonomous vehicles: A review. *Sensors*, 21(6).

and Yu wu Zhangjing Wang and Qingqing Niu. 2020. Multi-sensor fusion in automated driving: A survey. *IEEE Access*, 8:2847–2868.

Xiangmo Zhaoa, Pengpeng Sun, Zhigang Xu, Haigen Min, and Hongkai Yui. 2020. Fusion of 3d lidar and camera data for object detection in autonomous vehicle applications. *IEEE Sensors Journal*, 20(9):4901–4913.