

How can deep learning improve lidar and radar data analysis in object detection for self-driving cars?

Krishay Bhople

Received January 17, 2024

Accepted April 29, 2024

Electronic access May 15, 2024

Autonomous driving encompasses many different technologies, including computer vision, radar, lidar, and deep learning. Each technology has its own strengths and limitations, which when compared, reveal synergy. Oftentimes, self-driving systems do not utilize radar and lidar simultaneously, which exposes weaknesses inherent in each technology. However, merging radar and lidar with deep learning allows autonomous vehicles to have a more comprehensive and accurate perception of their surroundings, especially due to their complementary characteristics, while reducing potential risks from each technology. In this paper, we discuss different aspects of deep learning, radar, and lidar, including their history and modern issues. We also describe methods for integrating the two sensor technologies using a form of deep learning known as multimodal learning in two main frameworks named M²-Fusion and ST-MVDNet. Prior research on the synthesis of lidar and radar is lacking given the ongoing development of both technologies with respect to deep learning, but several methods have been proposed for integration. This paper aims to provide a systematic review of lidar, radar, and deep learning to cohesively summarize development of these technologies and provide a critical analysis of the two proposed fusion algorithms M²-Fusion and ST-MVDNet. We reveal that the proposals are capable but have limitations that constrain their applicability in modern autonomous vehicles.

Introduction

In autonomous driving, computer vision is essential to the system in charge. Just as blindness results in an inability to drive, the absence of computer vision results in being unable to navigate the roads autonomously. Autonomous driving incorporates different facets of computer vision technology to efficiently drive without an individual in charge. Such pieces of technology include deep learning, radar, and lidar for object detection. Deep learning is a subset of machine learning that involves training neural networks to recognize complex patterns in data to perform tasks with high accuracy¹. Deep learning has taken great strides with the introduction of Large Language Models and generative image and video tools. Radar, or radio detection and ranging, is another technology utilized in computer vision, though it is much older than deep learning. It uses radio waves to obtain environmental information and measure the distance, speed, and direction of surrounding objects². Lidar is much newer than radar, and it utilizes laser light instead of radio waves to gather spatial data to map an environment to high accuracy³. While radar's performance in a multitude of weather conditions and its ability to identify objects at a distance have established usage in navigation systems, lidar's higher precision and ability to create detailed 3D maps have found applications in autonomous driving, robotics, and environmental monitoring. Additionally, although research dedicated to integrating radar and lidar is available, a detailed study associating correlations between each

technology to leverage strengths and mitigate weaknesses and providing a review on modern methods of integration is lacking. We do this by employing synergistic analysis to improve the domain expertise surrounding radar and lidar object detection in autonomous vehicles. By using artificial intelligence, it is possible to refine the interpretation of sensory data, allowing for reliable decision-making for autonomous navigation. In this paper, we discuss two proposed deep learning methods named M²-Fusion and ST-MVDNet to integrate lidar and radar data analysis in object detection for self-driving cars.

What is Deep Learning?

Deep learning is a sub-field of machine learning, focusing on using multilayered processing to send an output based on received input. Deep learning has a long history of development, beginning in 1943, with Walter Pitts and Warren McCulloch publishing their paper titled, "A Logical Calculus of the Ideas Immanent in Nervous Activity"⁴. Pitts and McCulloch compare the brain to a computing machine with neurons acting as processors. Both researchers proposed the first computational model of a neuron having an input g with an output f [Figure 1]. The structure is related to that of the neurons in the human brain with the inclusion of a dendrite to receive information and an axon to transmit the output.

Deep learning relies on the use of neural networks, which have a wide range of problem-solving abilities. Neural networks in-

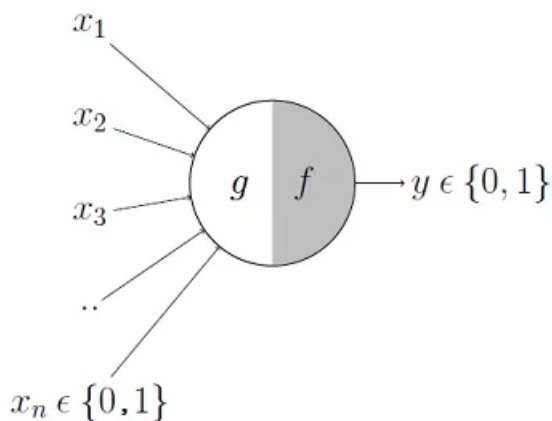


Fig. 1 The first computational model of a neuron, as proposed by Pitts and McCulloch⁵.

volve multiple layers of neurons that process information based on weights and biases [Figure 2]. Some specific instances of neural networks' use in self-driving include identifying street signs, lane detection and following, and parking assistance. Deep learning aims to take over many tasks that drivers would consider mundane, improving the driving experience overall.

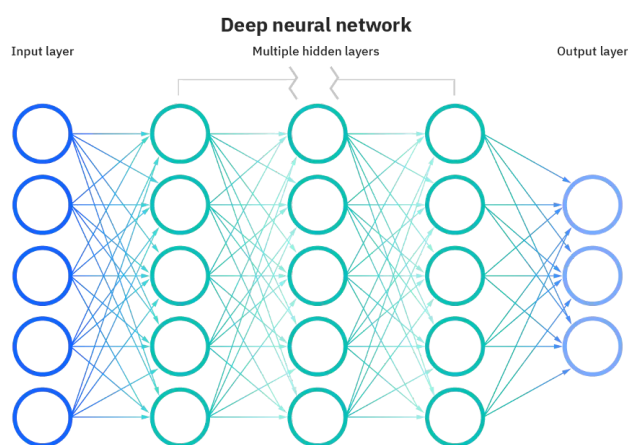


Fig. 2 The structure of a deep neural network involves an input layer (left), several hidden layers that process information (middle), and an output layer (right)⁶.

SAE International designed six tiers to categorize different states of autonomous driving named the Society of Automotive Engineers (SAE) Levels of Automation⁷. The initial three tiers, starting from level zero, encompass driver-assisting attributes. Level zero's automatic emergency braking, level one's lane centering, and the amalgamation of adaptive cruise control with lane centering evident in level two are key examples. The next

three levels take over entire tasks for drivers, ranging from the absence of pedals or steering wheels (level four) to being able to drive in all locations regardless of any hindrances (level five). As the tiers advance, the complexity of the tasks also rises.

Features that are described in the SAE Levels of Automation, such as “local driverless taxi” and “traffic jam chauffeur”, are currently undergoing testing. Main competitors in the self-driving field, such as Waymo and Tesla, are working to implement these functionalities. Waymo has developed a system to drive passengers unaccompanied by a human driver, effectively creating the driverless taxi described in level four. Additionally, Tesla has released their Full Self-Driving (FSD) technology involving the use of neural networks⁸, eliminating the need for human control over vehicles. As a result, it is reasonable to infer that deep learning has made significant advancements in the field of autonomous driving.

However, these systems have evidently not reached level five, as they bring limitations along with their innovation. Currently, one main way autonomous vehicles build spatial information is by utilizing lidar and camera technology⁹. Since these technologies can be easily obscured by fog, snow, and rain in the air, radar is often introduced to add an additional data modality that enhances reliability. In the proposed approaches, M²-Fusion and ST-MVDNet, deep learning is used to combine another form of data, radar input, to continue detecting objects despite having difficulties with currently used technology, allowing modern vehicles to be classified into levels four and five. In this way, deep learning proves its use in feature extraction for radar and lidar to advance the development of autonomous vehicles.

Computer vision and deep learning have been very closely tied in recent years due to the tremendous potential of convolutional neural networks (CNNs) in image recognition and segmentation tasks. CNNs were first proposed by Yann LeCun as a specialized architecture designed to process grid-like data structures. CNNs use multiple stacked layers, including pooling layers to reduce spatial dimensions, and convolutional layers that can detect features from local regions using learnable filters. When put together, these layered architectures can combine relative spatial information in image data for object detection and analysis. CNN-powered computer vision enables vehicles to recognize different types of objects, including partially hidden road signs and cones¹⁰. In M²-Fusion, a CNN is leveraged to create a feature map that links features of different objects in radar and lidar 3D spaces, allowing for improved object recognition. Figure 3 shows an example of such recognition capabilities from CNNs.

The addition of deep learning does pose challenges to autonomous driving systems. Firstly, implementing deep learning adds a layer of complexity. The huge variability and unpredictability on the road make it complicated to build deep learning models that maintain performance across a wide set of these environmental interactions¹¹. Secondly, with the increasing

popularity of deep learning, more adversarial attacks are possible¹¹. These perturbations to data may be imperceptible to humans but can cause decision-changing effects in models¹². Fortunately, autonomous vehicle engineers are aware of these issues and have looked at edge computing as well as privacy-aware knowledge sharing to combat each setback respectively. Using methods such as federated learning, client servers do not have to exchange large amounts of data with global servers. The model can be trained locally, increasing privacy and minimizing risk. Solutions such as these mitigate risks when utilizing deep learning in self-driving systems.

As deep learning continues to progress and make advancements in autonomous vehicles on the software side, other technologies including radar and lidar do the same on the hardware side as well.

Radar Methods

Invented in 1935 to guide pilots away from thunderstorms and still employed for various tasks today, radar is a widely respected technology.

Radar sends electromagnetic waves into an environment with a transmitter. The reflected waves in the scene are measured by a receiver. These devices are usually antennas, and oftentimes one antenna does both the jobs of transmitting and receiving. For each wave, the receiver measures signals with varying delays², as these delays help in determining the distances to different surfaces in the environment.

Radar has a multitude of benefits, including its ability to be used in tough weather conditions regardless of day or night¹⁴. Since its creation, it has been in constant development, perfecting itself with each iteration. By the modern day, radar has become reliable, albeit with certain limitations. For one, radar has susceptibility to ghost targets, which are reflections of a real object caused by the transmission of electromagnetic waves¹⁵. Ghost targets are detectable and can be avoided using neural networks, however. Additionally, radar suffers in extremely crowded and complex environments because of the relatively low resolution offered, with some single-chip radars having a resolution of two orders of magnitude (or one hundred times) lower than that of lidar¹⁶.

These challenges have prompted modern autonomous vehicle manufacturers, such as Tesla, to utilize technologies such as cameras rather than radar. As a result, most present-day autonomous vehicles have not been able to achieve higher SAE levels due to limitations with cameras, such as harsher environmental conditions that obscure visibility. These issues are unavoidable with the inappropriate sensor technology. However, Waymo's use of radar, along with lidar and cameras, has allowed them to be categorized as level four¹⁷, demonstrating that when integrated with other technologies, radar does pose great potential as an instrument to mitigate environmental constraints while

continuing to effectively navigate the roads.

The price of integration is a potential drawback, with some estimates for the manufacturing cost of a complete Waymo vehicle with state-of-the-art lidar, radar, and cameras reaching up to \$300,000¹⁸. Currently, the service is not profitable¹⁹, but as services such as Waymo are still in the research phase, there is much room for development.

Lidar Methods

Lidar technology investigates the use of light pulses, instead of electromagnetic waves, to map and understand its surrounding environment. In 1938, light pulses were first used to take atmospheric measurements, but it was not until 1953 that the term "lidar" was first introduced³. When the laser was invented in 1960, development of lidar sped up. Today, its abilities are already proficient for autonomous driving, and the technology continues to rapidly evolve.

Lidar emits light waves from a laser emitter, which are scattered and reflected from the environment. As light returns to the lidar sensor, it is passed through a photodetector. The device generates an electrical signal in response to the light waves to enable the next steps of processing. The Time-of-Flight Principle is then used to measure depth through the time intervals it takes for light to make a round trip to the sensor, similar to radar's distance measurements with electromagnetic waves. After some additional computations, the lidar system generates point cloud estimates to map and visualize information in 3D space.

While lidar is relatively advanced, the technology brings limitations with its use. The first involves reflections, which may be perceived as black holes, causing missing or distorted data points within the point cloud. Additionally, lidar equipment is expensive and high maintenance, making it impractical for mass production. Despite the existence of more sophisticated algorithms and sensors to mitigate reflection-based challenges, as well as cost-effective solutions that are becoming available to automakers, perhaps the most significant inconvenience is lidar's inability to be used in adverse weather due to light scattering and occlusion²⁰. On its own, lidar displays impressive capabilities in detailed 3D modeling of a vehicle's environment in ideal conditions. However, when expanding the range of autonomous driving, it is imperative for manufacturers to deal with a wide variety of situations, including poor weather that decreases visibility for lidar. Sensor fusion with radar is a niche solution to this issue that will allow modern autonomous vehicles to drive in a larger set of scenarios, thereby taking steps to reach SAE level five.

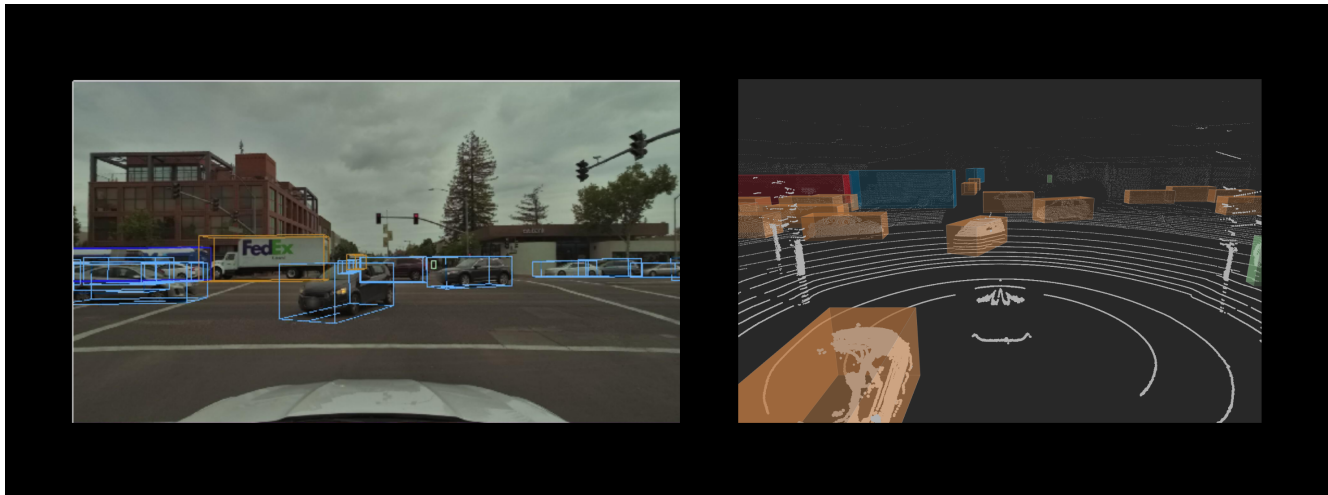


Fig. 3 Lyft’s lidar system detects different types of vehicles and categorizes them into separate classifications with CNNs; trucks and larger automobiles are shown differently in contrast to the orange cars in the right graphic¹³.

Contrasting Radar and Lidar

Table 1 summarizes the methodologies and weaknesses of radar and lidar and presents their potential for future development.

Radar and lidar are complementary technologies. For example, while lidar suffers in adverse weather conditions, such as fog, rain, and snow, radar maintains relatively consistent performance in such challenging environments. Additionally, at the expense of reduced range, lidar offers finer resolution data. In contrast, radar has a longer range than lidar, although at a lower resolution. Radar’s longer range compensates for its lower resolution in scenarios where detecting objects at greater distances is necessary, even if the details of those objects are less precise. This is crucial in collision avoidance systems, where early detection of potential hazards provides the autonomous driving system enough time to react. Meanwhile, the increased resolution of lidar allows for autonomous driving systems to accurately create 3D maps while radar serves as a safety redundancy measure to ensure lidar systems are functioning properly.

Integrating radar and lidar data through deep learning allows autonomous systems to leverage the strengths of each sensor while mitigating their respective limitations. Their complementary characteristics allow for a more robust perception system capable of adapting to diverse environmental conditions.

Deep Learning Integration

Multimodal learning is a sub-field of modern deep learning that is concerned with solving problems through integration of different forms of data. At the core, it aims to cohesively combine inputs sampled from varying processes. For instance, people visualize objects, taste flavors, and use several other senses to

perceive their surroundings. In a more technical example, a user may wish to integrate relevant image and textual data within a file system to improve their image search algorithm. The modalities of data being addressed in the context of this paper are lidar and radar, and the deep learning task involves object detection in autonomous driving. Utilizing multimodal learning for radar and lidar integration undertakes each technology’s challenges while simultaneously optimizing benefits. Other approaches to improving object detection in autonomous vehicles have also been proposed. One example is Tesla’s Tesla Vision, which utilizes camera input and deep learning processing to build spatial awareness²¹. However, using only one sensor makes autonomous driving systems vulnerable to many errors and creates a single point of failure. Multiple sensors in multimodal learning for autonomous driving achieve redundant safety and make the system fault tolerant, which is needed in the case of protecting human life.

Multi-Modal and Multi-Scale Fusion (M^2 -Fusion), among others, has been proposed as a fusion method for integrating lidar and radar data²². M^2 -Fusion consists of applying two main modules: the Center-based Multi-Scale Fusion (CMSF) and the Interaction-based Multi-Modal Fusion (IMMF). Both are utilized to improve feature expression capabilities of objects. In an autonomous driving setting, such an object may be as trivial as a cone or a curb, or another vehicle traveling on the freeway.

The first module in M^2 -Fusion, the CMSF, assists with object detection. The CMSF addresses the challenge of processing sparse radar data, since radar data resolution is not always up to par with that of lidar. Because there are often missing or inaccurate points, the raw radar data is evidently not useful on its own in autonomous driving where precision is critical, so the CMSF first extracts important details from the current data.

Table 1 A summary of key differences between lidar and radar.

	Radar	Lidar
Principle of Operation	Uses electromagnetic waves channeled from a transmitter and measures reflected waves with a receiver to determine distances to different surfaces based on signal delays.	Utilizes light pulses emitted by a laser emitter, which are scattered and reflected in the environment. The Time-of-Flight Principle generates point cloud estimates for 3D mapping based off the time for round trips of the light waves.
Detection Challenges	Susceptible to “ghost targets,” which are reflections of real objects caused by electromagnetic wave transmission. Low resolution in complex environments limits radar’s performance.	Faces challenges such as reflections causing “black holes” in the data and missing/distorted data points in the point cloud. These challenges require sophisticated algorithms and sensors to address.
Advancement & Development	Radar has reached a high level of reliability, but ongoing research aims to further enhance capabilities and overcome limitations above. Technologies such as millimeter wave radar provide finer resolution.	Continues to rapidly evolve, with advancements in algorithms, sensors, and cost-effective solutions driving lidar’s improvement.

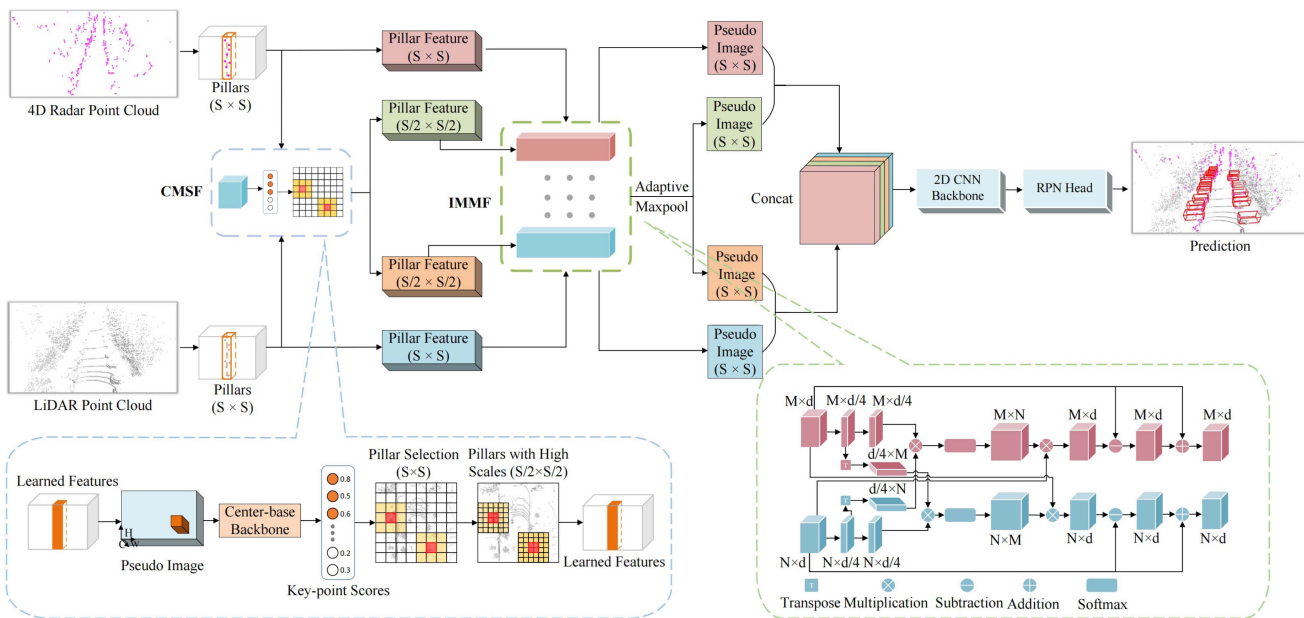


Fig. 4 The structure of M^2 -Fusion contains the CMSF and the IMMF acting as the two main components²².

Voxel-based frameworks are used to organize information and create voxel grids, where voxels refer to the 3D equivalent of a traditional 2D pixel. The framework divides the point cloud data into voxels, an example of which can be seen in Figure 5, but a challenge arises in determining the size of the voxel grids. When the grids are increased in size, processing speed is decreased, but finer details are missed. Conversely, smaller

grids provide more detail and noise, but more computational power is required. To compromise between the two cases, a variety of grid sizes should be used; however, this still takes a toll on computing resources, so the CMSF first finds important points in the data before voxels are selected around the points and to extract the most salient features.

After obtaining these features, the IMMF synthesizes specific



Fig. 5 NVIDIA wins the 3D Occupancy Prediction Challenge with FB-OCC, a technology designed for using camera data to generate 3D voxel bird's-eye maps. The bottom two maps exemplify the creation of voxel grids using sensor data from the six camera inputs above²³.

connections between the different data modalities regarding the features of such an object, including position, volume, and surface details. This is the key step used to link the lidar and radar input data. The IMMF proves to be valuable under tough weather conditions. For instance, in foggy environments, the features extracted by the CMSF from the lidar data might not meet the same quality standards as those extracted during clear weather conditions. The IMMF is crucial because it allows those object's features to be enriched by the stable features that the CMSF processed from the radar data. The results from M²-Fusion are depicted in Figure 6.

In extreme conditions where one of the two sensors experiences a complete failure, M²-Fusion will be unable to navigate the roads completely autonomously. The IMMF will no longer be able to synthesize data from both modalities, leading to a drastic reduction in performance. M²-Fusion relies on both sensors functioning, but natural causes such as aging, wear, manufacturing deficiencies, and mishandling can lead to sensor failure²⁴.

Other methods have been further explored to ensure autonomous driving can continue without interruption, as shown in Figure 7. One such system is the Self-Training Multimodal Vehicle Detection Network (ST-MVDNet)²⁵.

ST-MVDNet builds upon another model, the Multimodal Vehicle Detection Network (MVDNet)²⁶, while being trained using the Mean Teacher principle²⁷. As with M²-Fusion, MVDNet

was initially proposed to integrate lidar and radar data because of their complementary features with the assumption that both sensors are always available. MVDNet uses an intuitive method to extract and fuse important features from both sensors. The first stage of MVDNet identifies important features, or proposals, in both modalities of data. The second stage fuses these proposals using 3D Convolutions and attention mechanisms that pay attention to the most relevant parts of the data. The effectiveness of MVDNet is exemplified by Figure 8. However, both modalities of data are still needed to produce these results.

ST-MVDNet implements the Mean Teacher principle by utilizing a teacher and student pair, which is depicted in Figure 9. The teacher model generates predictions on a data set; these predictions are then used to calculate a consistency loss that shows the consistency of predictions made on the same input under different conditions or perturbations. The loss is then used to train the student network, which will adjust its parameters to align its predictions with that of the teacher. The communication between the two different models is key to the ST-MVDNet. While the teacher model accepts only clear modalities, the student model is trained to handle missing lidar or radar streams. This allows for stability and consistency while preventing overfitting, a scenario where a model is unable to generalize and instead fits too closely to a training data set.

The results from ST-MVDNet are compared in figure 10 along with a comparison to MVDNet. Without strong augmentation

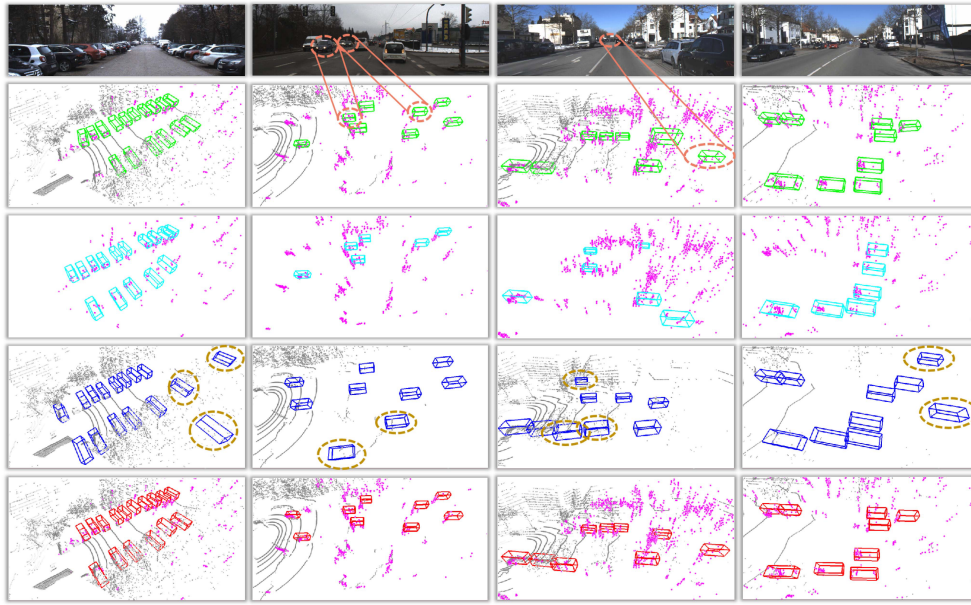


Fig. 6 The qualitative results from M^2 -Fusion are shown. The first row provides RGB images from the Astyx HiRes 2019 dataset depicting the scene. The second row is the ground truth, and the green bounding boxes surrounded with dotted orange ovals show missed detections. The third row shows detections for radar, and the fourth shows results for lidar, where the blue bounding boxes surrounded with dotted brown ovals show false positives. The last row shows results from M^2 -Fusion.

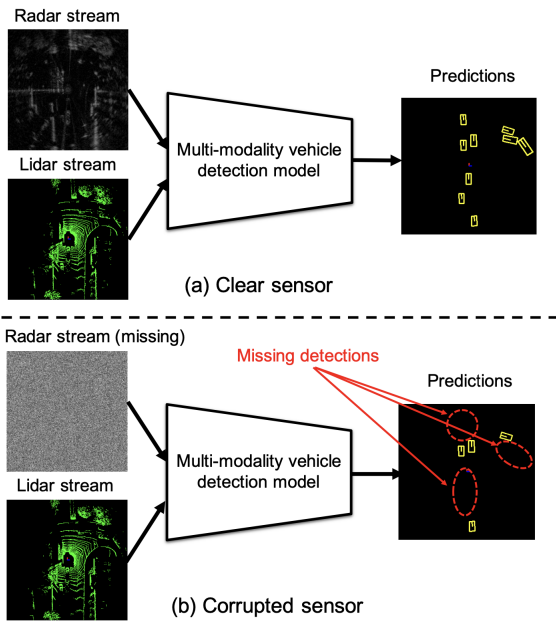


Fig. 7 When one data stream unexpectedly becomes inaccessible, the multimodal detection model is unable to identify all objects. ST-MVDNet has been proposed to mitigate such an issue²⁵.

to training data, ST-MVDNet produces false positives and is unable to detect all objects in the scene. The requirement of strong augmentation may cause over reliance on augmented data, which differs heavily from real-world scenarios, and an increase in the computational complexity of training. However, with the inclusion of augmentation, the model can accurately identify these objects. It is notable that ST-MVDNet recognizes objects without error even when a data stream is absent.

Conclusion

The aim of this paper was to explore different methods of integrating lidar and radar for object detection in autonomous vehicles. We first discussed radar, lidar, and deep learning, examining the development of each technology throughout history and present day. Due to lidar and radar's complementary characteristics, we showed that sensor fusion is a powerful way to improve computer vision technology in autonomous driving. As a result, we identified two distinct proposals for multimodal learning with lidar and radar in modern literature, namely M^2 -Fusion and ST-MVDNet, by exploring their methodology, limitations, and results with a critical analysis. Such an analysis is necessary for any new technologies when evaluating whether to utilize them in one's own products and manufacturing systems. In the broader context of autonomous driving, this study reveals that modern radar and lidar proposals need more development before

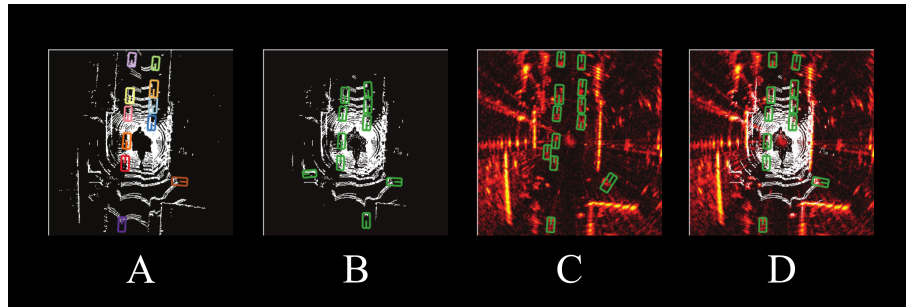


Fig. 8 The results of the MVDNet framework when used for object detection. Fig. A shows a top view of the 3D lidar point cloud data and is the ground-truth. Each colored box represents a different vehicle, with the vehicle equipped with lidar and radar in the center. Fig. B shows how MVDNet misses vehicles when used with only lidar at farthest range due to fog. Certain background points are also mistaken for vehicles. Fig. C demonstrates MVDNet using only radar; false positives are produced due to the noise in the radar data. Fig. D demonstrates that fusing lidar and radar with deep learning from MVDNet correctly detects vehicles; this matches the ground truth in Fig. A²⁶.

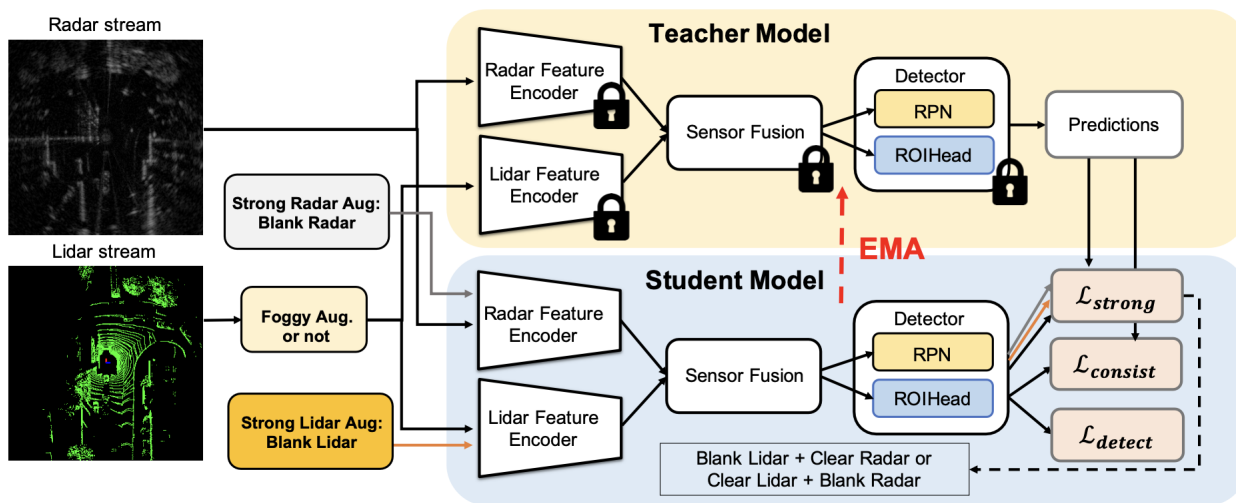


Fig. 9 The framework of ST-MVDNet consists of the Mean Teacher principle that employs a Teacher and Student pair of models. Each model in ST-MVDNet is architecturally identical with two feature encoders to transform raw lidar and radar data into a format suitable for processing, a sensor fusion model, and a detector that derives feature maps and generates proposals²⁵.

being fully integrated into self-driving systems. While current literature does present approaches that are able to successfully recognize objects, the circumstances and conditions necessary to do so are limited. When one sensor data stream fails or the training data is not properly augmented, M²-Fusion and ST-MVDNet will be unable to guarantee passenger safety due to an inability to detect possible surrounding dangers in all scenarios. However, if multimodal learning for object detection with lidar and radar is perfected, the technology does pose tremendous potential for benefiting a wide range of people and industries. Since the current work serves as a solid foundation to develop on, further research should focus on expanding the range of situations that integration proposals can apply to. Training models with larger or different data sets, model regularization, and

k-fold cross-validation are all viable methods beyond the scope of this paper to minimize the constraints that apply to these complex systems. In the present day, while the applicability to real-world situations needs to be further improved, the capability for these systems for object detection does exist. The research and innovations explored in this paper will continue to progress autonomous driving, paving the way for better road transportation.

References

1. I.B.M., *What is deep learning?* <https://www.ibm.com/topics/deep-learning>.
2. A. Srivastav and S. Mandal, *Radars for Autonomous Driving: A Review of Deep Learning Methods and Challenges*, arXiv: 2306.09304 [cs.CV].

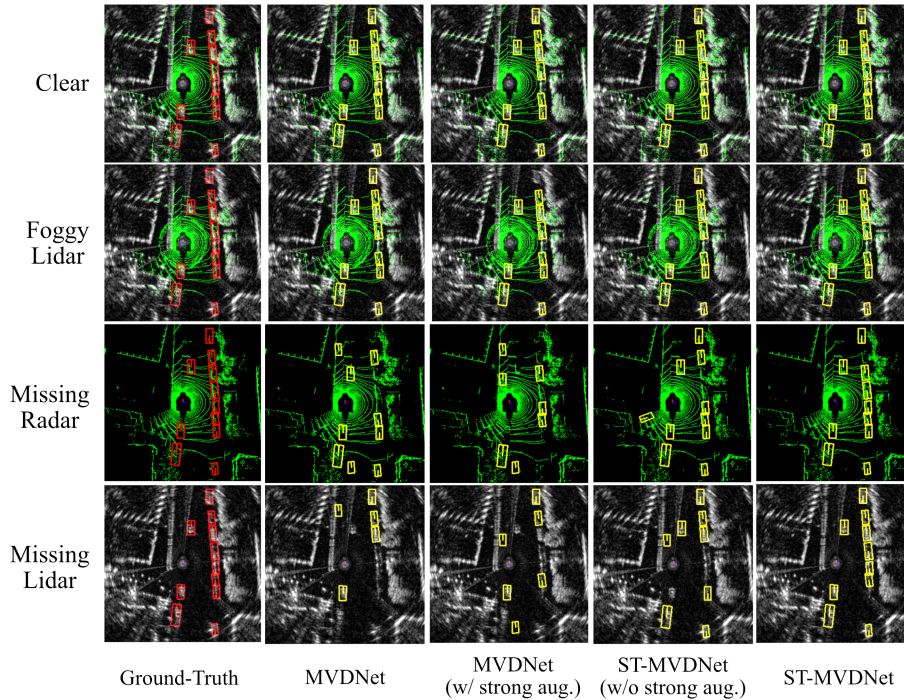


Fig. 10 Yellow bounding boxes are placed around detected objects in the scene. When testing with missing radar and missing lidar data streams, MVDNet appears to have missing detection as well as false positives, along with ST-MVDNet without strong augmentation. When strong augmentation is added, ST-MVDNet produces accurate results in each scenario where a sensor data stream is missing²⁵.

- 3 U. Wandinger, *Introduction to lidar*. In: *Lidar: range-resolved optical remote sensing of the atmosphere*, Springer, p. 1–18.
- 4 W. S. McCulloch and W. Pitts, *The bulletin of mathematical biophysics*, **5**, 115–133.
- 5 A. L. Chandra, *McCulloch-Pitts Neuron — Mankind’s First Mathematical Model Of A Biological Neuron*, <https://towardsdatascience.com/mcculloch-pitts-model-5fdf65ac5dd1>.
- 6 I.B.M., *What is a neural network?*, <https://www.ibm.com/topics/neural-networks>.
- 7 S. International, *SAE Levels of Driving Automation™ Refined for Clarity and International Audience*, <https://www.sae.org/blog/sae-j3016-update>.
- 8 Tesla, *AI Robotics*, <https://www.tesla.com/AI>.
- 9 A. Chougule, *IEEE Open Journal of Vehicular Technology*, **5**, 142–161.
- 10 N. Cristovao, *All Tesla FSD Visualizations and What They Mean*, <https://www.notateslaapp.com/tesla-reference/636/all-tesla-fsd-visualizations-and-what-they-mean>.
- 11 K. Muhammad, *IEEE Transactions on Intelligent Transportation Systems*, **22**, 4316–4336.
- 12 Dremio, *Adversarial Attacks in AI*, <https://www.dremio.com/wiki/adversarial-attacks-in-ai/>.
- 13 Lyft, *Dash Lyft Perception*, <https://dash.gallery/dash-lyft-explorer/>.
- 14 Waymo, *Waymo Safety Report*, <https://storage.googleapis.com/sdc-prod/v1/safety-report/2020-09-waymo-safety-report.pdf>.
- 15 T. Jeong and S. Lee, *IEEE Sensors Journal*, **22**, 14378–14386.
- 16 A. Prabhakara, 2023 IEEE International Conference on Robotics and Automation (ICRA).
- 17 N. Webb, *Waymo’s Safety Methodologies and Safety Readiness Determinations*, arXiv: 2011.00054 [cs.RO].
- 18 S. LeVine, *What it really costs to turn a car into a self-driving vehicle*, <https://qz.com/924212/what-it-really-costs-to-turn-a-car->.
- 19 B. Broadwater, *Waymo Stock - Is Waymo Publicly Traded in 2024?*, <https://www.wealthdaily.com/waymo-stock/>.
- 20 M. Dreissig, *Survey on LiDAR Perception in Adverse Weather Conditions*, arXiv: 2304.06312 [cs.RO].
- 21 Tesla, *Tesla Vision Update: Replacing Ultrasonic Sensors with Tesla Vision*, <https://www.tesla.com/support/transitioning-tesla->.
- 22 L. Wang, *IEEE Transactions on Vehicular Technology*, **72**, 5628–5641.

-
- 23 Z. Li, *FB-OCC: 3D Occupancy Prediction based on Forward- Backward View Transformation*, arXiv: 2307.01492 [cs.CV].
 - 24 S. Safavi, *Multi-Sensor Fault Detection, Identification, Isolation and Health Forecasting for Autonomous Vehicles* — [ncbi.nlm.nih.gov](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8038547/), <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8038547/>.
 - 25 Y.-J. Li, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, p. 918–927.
 - 26 K. Qian, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, p. 444– 453.
 - 27 A. Tarvainen and H. Valpola, *Advances in neural information processing systems* 30.