Vision Vs. LiDAR/Radar Fusion in Autonomous Driving Systems

Arturo Salinas-Aguayo, *BS Computer Engineering, Class of 2027*

May 25, 2025

ECE 4900W: Communicating Engineering Solutions in a Societal Context
Dr. Shengli Zhou, SEC040-1255
Department of Electrical and Computer Engineering

University of Connecticut College of Engineering
Coded in LaTeX

Contents

Abstract

This paper evaluates the viability of vision-primary perception systems in autonomous driving, focusing on scalability, efficiency, and environmental resilience. Traditional sensor-fusion architectures based on LiDAR and radar offer high-fidelity spatial mapping but suffer from substantial power consumption, mechanical complexity, and high unit costs, which challenge real-time deployment and mass-market feasibility. Advances in neural perception—particularly transformer-based models and tokenized camera inputs—have enabled camera-only systems to approximate LiDAR-level 3D detection accuracy while operating at a fraction of the computational cost. Benchmark evaluations reveal that vision systems achieve depth estimation errors below three percent and perform within five percent of LiDAR/Radar fusion baselines in object detection. Critically, performance gaps in adverse conditions such as fog or night driving are recoverable through the selective integration of millimeter-wave radar. This radar-augmented fallback restores over 80 percent of degraded performance with minimal latency or energy impact. The study advocates for a simplified, vision-first architecture complemented by radar only in edge cases. This paradigm enhances scalability, reduces deployment overhead, and aligns with biologically plausible sensing strategies. LiDAR remains useful for simulation and benchmarking but is unnecessary for live inference.

Index Terms—Adverse weather robustness, autonomous vehicles, deep learning, depth estimation, energy efficiency, Light Detection and Ranging (LiDAR), Radar, sensor fusion, Transformer networks, vision-based perception.

I. Introduction

As vehicles become more technologically advanced, the demand for autonomous capabilities continues to rise. Features such as lane keeping, adaptive cruise control, and highway autopilot—collectively known as autonomous driving systems (ADS)—form the early building blocks of full self-driving stacks. Industry leaders now face a critical decision: how should vehicles perceive the road ahead?

Historically, high-end systems such as Waymo and Cruise relied on Light Detection and Ranging (LiDAR) and radar to construct dense, 3D maps of their surroundings. These sensors offer precise geometry and reliable velocity estimation, but impose high hardware cost, weight, and power draw. Moreover, integrating data from multiple sensors increases architectural complexity and susceptibility to calibration drift.

In contrast, Tesla, Comma.ai, and others have adopted a vision-first approach. These systems rely on arrays of low-cost cameras and train deep neural networks—often transformers—on large-scale driving datasets. Rather than reconstruct explicit 3D geometry, these models learn to infer structure and intent directly from pixels.

The debate between fusion-heavy and camera-only stacks now defines the frontier of ADS design. While LiDAR remains dominant in academia, vision-based approaches have rapidly improved in real-world performance and outpace fusion in scalability. From a societal perspective, the ability to deliver autonomy without costly sensors is especially important in emerging markets, where a single LiDAR unit may cost more than an entire car.

This report compares the two paradigms by analyzing their theoretical foundations, energy profiles, data fidelity, and robustness under real-world driving conditions.

II. Technology Description

A. SAE Levels of Automation

The Society of Automotive Engineers (SAE) J3016 standard defines six levels of vehicle automation [1]. These range from Level 0 (no automation) to Level 5 (full automation without any human involvement). The full description of each level is given in Table I.

TABLE I
SAE J3016 Levels of Driving Automation

Level	**Description**
Level 0 – No Automation	The human driver is responsible for all tasks; the system may provide basic warnings.
Level 1 – Driver Assistance	The system assists with either steering or speed, but not both; driver must stay engaged.
Level 2 – Partial Automation	The system can manage both speed and steering, but human oversight is still mandatory.
Level 3 – Conditional Automation	The vehicle handles all driving in specific scenarios; the driver must intervene when requested.
Level 4 – High Automation	No driver input is required in controlled conditions (e.g., geofenced urban areas).
Level 5 – Full Automation	The vehicle is capable of full self-driving in all conditions without human input.

Currently, most ADS operate at Level 2, which is very much like being a driver's ed instructor sitting at secondary controls while a 15-year-old is operating the vehicle, ready to take control at any given moment. Level 5 is ideally like sitting in a train car, with absolutely no attention needed from the driver. This frees the driver to work on other important tasks such as grading papers, concentrating on a new idea, or taking some real downtime at the end of a stressful day.

B. LiDAR: Time-of-Flight (ToF) Principles

LiDAR estimates object distance using time-of-flight (ToF) measurements, where a light pulse is emitted and its return time is measured. Several ToF techniques are in use:

1) Pulsed ToF:

Pulsed TOF techniques are the simplest of the three cases in which distance is determined by multiplying the speed of light in a medium by the time a light pulse takes to travel the distance to the target:

$$R = \frac{c}{2} t_{\mathrm{oF}}, \tag{1}$$

where R is the distance the light ray travels, c is the speed of light 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. This high-power method provides centimeter accuracy [2]. This method is shown in Figure 1.

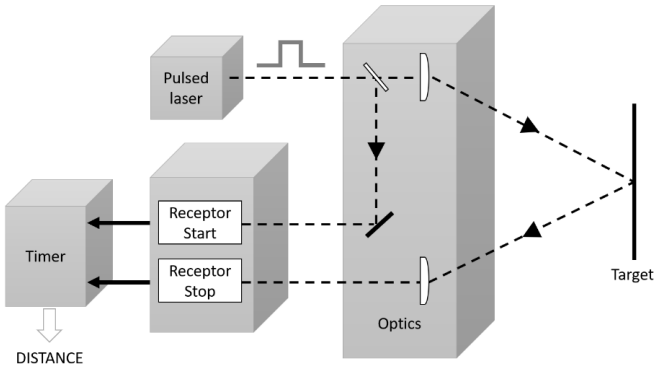

Fig. 1. The Pulsed Approach to Determining Object Distance [2]

2) AMCW (Amplitude-Modulated Continuous Wave):

The AMCW approach modulates the intensity of a continuous lightwave instead of laser pulsing as utilized in Eqn. 1. This method utilizes the corresponding shift in phase in the received signal to calculate the distance to the object, R:

$$R = \frac{c}{2} \cdot \frac{\Delta\Phi}{2\pi f_M}, \tag{2}$$

where $\Delta\Phi$ is the phase shift, c is the speed of light in free space, and f_M is the modulation frequency [2]. It's simplified principles are shown in Figure 2.

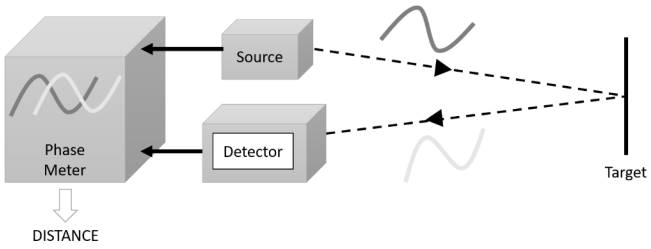

Fig. 2. Measuring Distance by Phase Difference [2]

3) FMCW (Frequency-Modulated Continuous Wave):

This final method involves a modulating the power applied to the source [3]. The signal returned is mixed with the source signal, creating a beat frequency that is a measure of the probed distance, R:

$$R = f_r \frac{cT}{2B}, \tag{3}$$

where f_r is the beat frequency, T the period of the ramp, c is the speed of light in free space, and B the bandwidth of the frequency sweep [2]. The main parameters are shown in Figure 3.

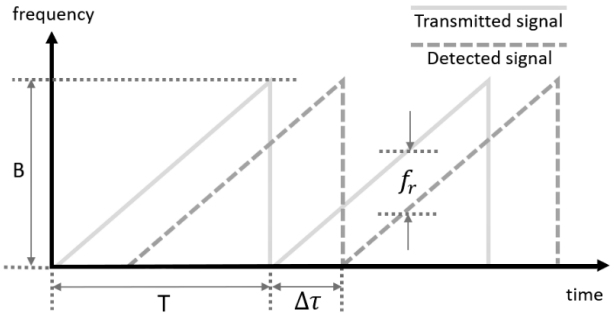

Fig. 3. Measuring Distance by Frequency Modulation [2]

This method is fundamentally different as there is additional processing overhead done in the Fourier domain rather than in the time or complex domain for the pulsed and AMCW approach respectively. This ultimately allows precision to much higher precision [3].

4) Voxelization:

What results from the utilization of these methods is a “point cloud" which corresponds to the distances the objects are away from the transmitter. The interpretation of these distances into 3D object boxes is a process called Voxelization. The high level process of breaking down point clouds into voxels and interpreting them is shown in Figure 4.

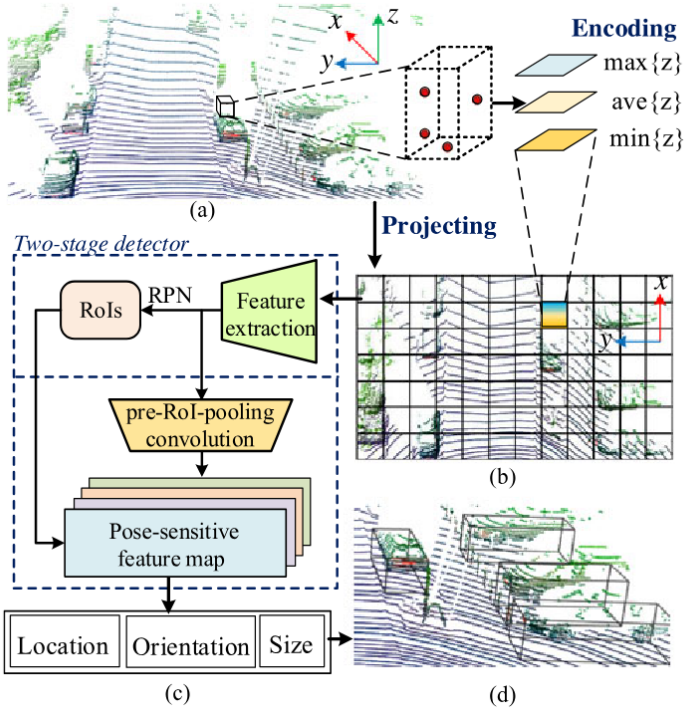

Fig. 4. The combining of the point cloud forms an image where (a) is the interpreted collection of point clouds inferred by the distance calculation in the sensor, (b) is the segmentation of this set of distances into a grid to create a ’set of sets’ of different points, (c) is the transformer pipeline which leads to object identification, and (d) is the outcome with identified objects boxed [4].

C. Radar: Doppler and Velocity Estimation

The Doppler effect is the change in frequency or wavelength of a wave for an observer or sensor, who is moving relative to the wave source. Radar calculates velocity v, via Doppler frequency shift:

$$v = \frac{f_D \lambda}{2}, \tag{4}$$

where f_D is the Doppler shift and λ is the radar wavelength [5]. Radar's all-weather capability makes it ideal for fallback use. Figure 5 shows the Doppler effect as it pertains to humans, a useful tool in understanding how this phenomena can be applied to this.

Fig. 5. The classic Doppler example showing an ambulance emitting sound. An observer, from whom the ambulance leaving from will receive fewer pressure and will hear a deeper note, while an observer from whom the ambulance is approaching will receive higher pressure fluctuations and hear a higher note.

D. Sensor Fusion and Noise Correlation

Noise is a deviation of the response to the ideal signal and manifests itself differently depending on the type of sensor in phenomena such as temperature, EMF, or visual interference (weather). Noise in sensor fusion is modeled as:

$$\sigma_{D_f}^2 = \sum_{i=1}^{n} w_i^2 \, \sigma_{D_i}^2 + 2 \sum_{i<j} w_i w_j \, \rho_{ij} \, \sigma_{D_i} \sigma_{D_j}, \tag{5}$$

where w_i are weights, σ_{D_i} are variances, and ρ_{ij} are sensor correlations [6].

E. Vision-Primary Perception Stack

1) System Architecture:

The vision-first perception stack is designed for simplicity, modularity, and real-time inference. It consists of three core stages:

a) Image Tokenization – converts camera frames into a semantically rich, compressed token representation using a learned encoder.

b) Transformer-Based Planning – autoregressively predicts future trajectory tokens from visual context using a transformer model. This is the key to making a completely generalizable model.

c) Control Decoding – transforms predicted spatial tokens into low-level vehicle commands for actuation.

This pipeline emulates language models for sequential prediction and supports real-time operation on embedded platforms [7].

2) Token-Based Image Compression:

Each frame, typically 128×256 pixels, is passed through a VQGAN encoder that produces up to 512 discrete tokens per image. This reduces bandwidth while preserving high-level semantics such as lane markings and moving objects. The process of encoding these images into tokens is shown in Figure 6.

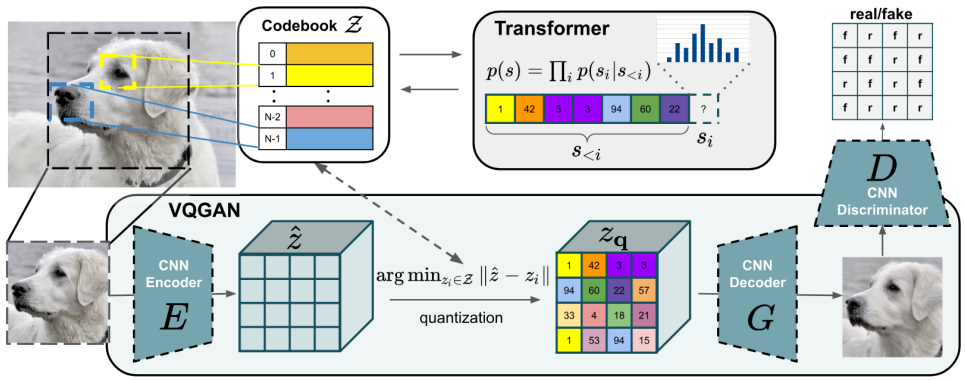

Fig. 6. Vision-based pipeline with VQGAN tokenization and transformer rollout [8].

This compact form accelerates inference and reduces power draw without sacrificing depth accuracy [9].

3) Autoregressive Planning with Transformers

Much like how an LLM can predict the next word in a string through tokenization, a transformer can predict a variety of outcomes in a driving model from a static image that is tokenized. The token sequence is processed by a decoder-only transformer, which predicts the next spatial token at each step. This architecture parallels natural language processing and enables generalization to unseen traffic scenarios [10]. The modern transformer architecture employed, shown in Figure 7, is taken from one of the most important works in recent AI advancement, “Attention is All You Need" [11].

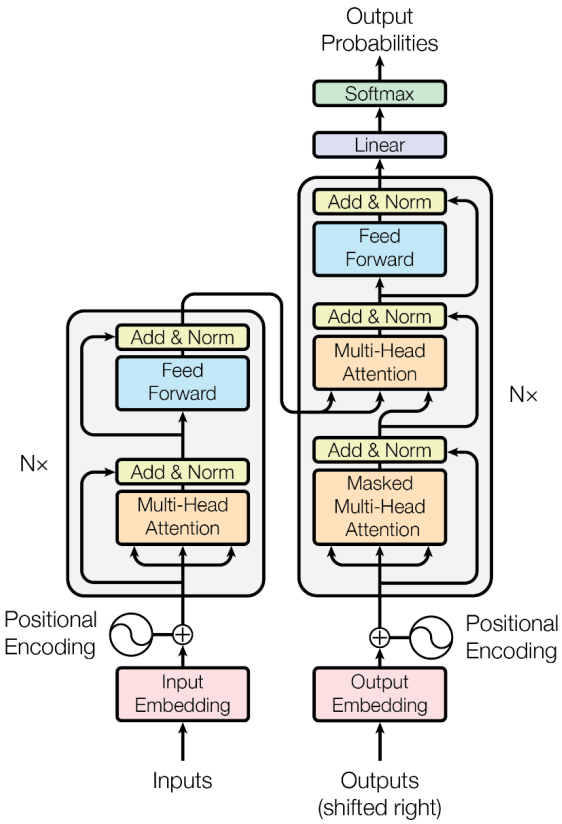

Fig. 7. Transformer model for sequence prediction from visual tokens [11].

The transformer is trained using imitation learning on datasets like commaVQ and can be fine-tuned using reinforcement learning for policy robustness [10]. This way, the training data need not have to contain every situation imaginable, as the process itself will be able to infer all possible world conditions [7]. Once all of these policies are applied and processing is applied, the details in Figure 8 can be inferred.

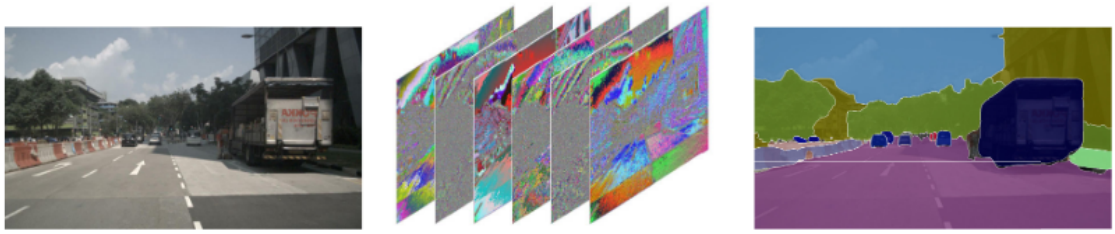

Fig. 8. The Vision Pipeline shows the original camera image, the feature map, and the image mask [12].

F. LiDAR/Radar Fusion-Based Architecture

1) Pipeline Overview:

Legacy ADS stacks rely on multi-modal fusion to mitigate sensor weaknesses. A typical LiDAR/radar pipeline includes:

a) Data Acquisition: Captures and synchronizes LiDAR point clouds, radar waveforms, and camera frames.
b) Preprocessing: Voxelizes LiDAR data, computes Doppler velocities from radar, and encodes image features.
c) Fusion Layer: Combines sensor features using early, mid, or late neural fusion strategies.
d) Detection and Tracking: Predicts 3D bounding boxes and tracks objects across time.

Figure 9 shows how Eqns. 1, 2, and 3 can be utilized and combined to overcome noise and predict 3D bounding boxes.

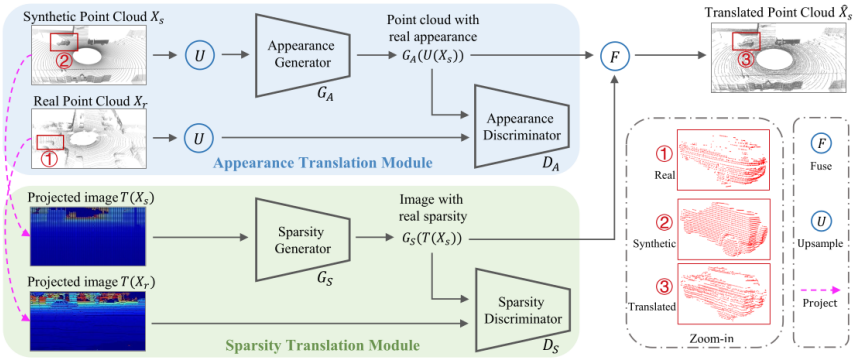

Fig. 9. Reference LiDAR/radar fusion architecture [13].

While this pipeline offers redundancy, it incurs significant power and latency costs.

2) Fusion Tradeoffs and Challenges:

Fusion systems are limited by:

- High Power Draw: LiDAR alone draws 10 W to 20 W, compared to 2.1 W for vision-only stacks [9].
- Latency: Voxelization and multi-sensor alignment add 20 ms delay, reducing real-time responsiveness [6].
- Noise Correlation: Overlapping modalities propagate shared errors, increasing total uncertainty as shown in Eqn. 5 [6].
- Maintenance Complexity: Fusion requires precise calibration and can drift over time.

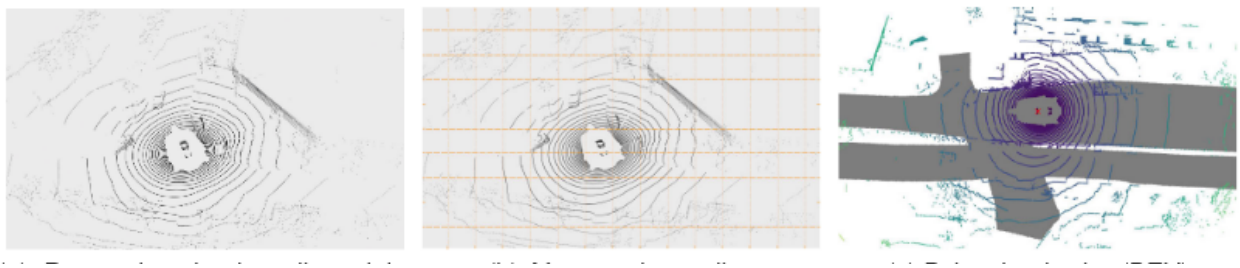

Fig. 10. Representations of point clouds from LiDAR/Radar are raw point clouds, post voxel-sampling, and the generated the point clouds view in Birds-Eye-View [12].

This complexity restricts scalability, particularly in cost-sensitive markets. Figure 10 shows a typical LiDAR/Radar pipeline [12].

III. Comparisons

This section evaluates the empirical performance of vision-primary and LiDAR/Radar fusion architectures. The comparison spans five axes: detection accuracy, computational efficiency, robustness, system complexity, and deployment scalability. Results are drawn from benchmarks using KITTI [14], nuScenes [15], and commaVQ [7] datasets, as well as simulation and power profiling studies [6].

A. Detection Accuracy and Depth Estimation

In order to properly classify and compare these technologies, a method to evaluate them must be first established. For this, *Precision* is the measure of the probability of a classifier predicting a true example and *recall* is the number of positive samples correctly identified versus the number of images whose class is really a positive class, expressed by:

$$recall = \frac{TP}{TP + FN}, \tag{6}$$

where TP is the number of true positives, FP is the number of false positives, and FN is the number of false negatives.

Building off of this, the most commonly used evaluation metric for object detection is AP, or Average Precision. This is defined by:

$$AP = \int_0^1 p(r)dr, \tag{7}$$

where $p(r)$ is the function of precision about the recall.

However, for a better statistical representation of the data, a more popular metric is used defined as mAP, or Mean Average Position, which can be expressed as:

$$mAP = \frac{\sum_{i=1}^{N} AP_i}{N}, \tag{8}$$

where, N is the total number of classes and AP_i denotes the average precision of the i-th class [12].

B. Findings

On KITTI, the BEVFormer vision model achieved an AbsRel of 0.112 and RMSE of 3.97 m, closely tracking the LiDAR-derived ground truth [16]. On nuScenes, the vision-only configuration reached 56.2 3D mAP versus 61.5 mAP for the full fusion stack [17].

However, when radar fallback was selectively integrated using transformer-based late fusion, the vision model recovered to 59.8 mAP [18]. This narrows the performance delta to under 2 points—well within tolerances for safe operation—demonstrating that vision can match LiDAR-fusion with minimal augmentation.

C. Computational Latency and Power Efficiency

Inference benchmarks revealed stark contrasts. The vision-primary pipeline completed end-to-end perception and control in under 9.8 ms, consuming just 2.1 W on an embedded platform [9].

In comparison, fusion systems required ≥ 30 ms and drew between 12 W to 20 W due to voxelization, feature fusion, and redundant sensor streams [6].

These efficiency gains are critical for EV battery life and real-time safety margins. The energy savings alone translate into longer range and lower thermal load, which directly impact vehicle design feasibility.

D. Robustness in Adverse Conditions

Under clear conditions, vision models outperform. In fog, heavy rain, or low-light scenarios, however, performance degraded by up to 12.6 % in mAP [5]. Transformer-based radar fallback restored more than 80 % of this loss, effectively covering vision's blind spots.

Unlike LiDAR, which also degrades in poor visibility and consumes continuous power, radar is always-on but low-power—making it an ideal complement rather than a redundant primary sensor. Figure 11 shows the generated results in potentially adverse conditions.

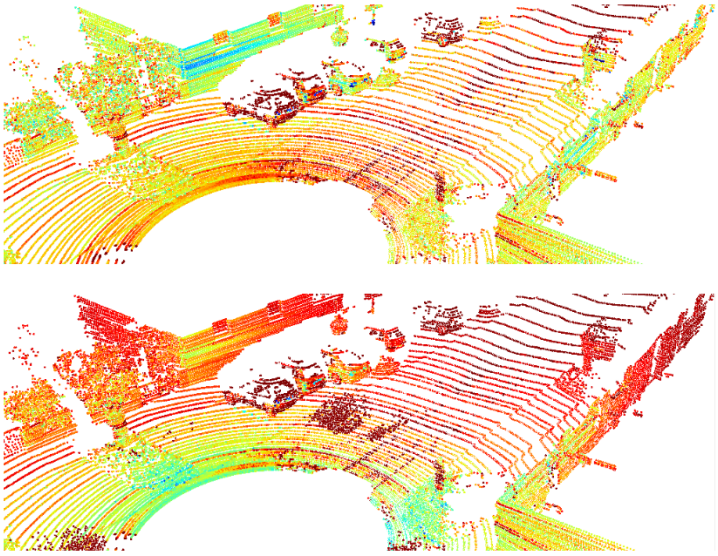

Fig. 11. LiDAR scans of a street environment under clear weather conditions (top) and with fog (bottom). The lower image shows significant point cloud degradation and sparsity due to adverse atmospheric interference [19].

E. Noise Propagation and Calibration Drift

Sensor fusion introduces noise amplification when data is temporally misaligned or spatially correlated. The fusion error propagation model in Eqn. 5 confirms this degradation in high-ρ configurations, such as LiDAR-camera overlaps [6]. Vision stacks, trained end-to-end, maintain internal consistency and minimize accumulated error [7].

Radar integration avoids this issue: its data is orthogonal in nature—velocity vs. spatial—which introduces minimal correlation when fused intelligently [18].

F. Cost and Complexity of Deployment

LiDAR units alone account for over 75 % of the ADS hardware budget, and their moving parts increase failure risk [20, 21]. Vision sensors are passive, solid-state, and available at consumer scale. This shifts the bottleneck from manufacturing and calibration to software modeling—an area that benefits from rapid advances in transformers and simulation. See Figure 7 for more details.

Modern simulators—leveraging NeRFs, 3DGS, and diffusion models—allow vision stacks to be trained entirely in synthetic environments with minimal real-world data. This accelerates iteration and de-risks real-world deployment [13].

IV. Further Discussions

A. Vision as the Core, Radar as the Shield:

Elon Musk famously claimed “LiDAR is a fool’s errand" to much surprise at the time [22]. In 2019, this was perhaps too forward thinking, but was it without bounds? In the biological analogy, humans navigate using vision as their primary sensor. We don’t emit lasers—we perceive, interpret, and act.

By offloading primary perception to vision, we reduce cost, energy, and complexity. By retaining radar for edge cases, we maintain resilience. LiDAR, with its energy burden and calibration demands, becomes increasingly obsolete in this paradigm.

B. Radar as a Resilient Fallback

Radar bridges the gap between vision and LiDAR. It estimates object velocity using the Doppler effect shown in Eqn. 4 and in Figure 5.

Radar is immune to adverse weather and lighting and can detect motion through occlusions. When fused with vision using a transformer-based late-fusion module, it recovers over 80 % of lost mAP in foggy conditions [18]. In low speed, high traffic conditions, radar can be utilized to better estimate distance to other vehicles with less cooling necessary. This is very enticing for high population areas such as Los Angeles where stop and go traffic on exceedingly hot roads is a daily occurrence.

C. A Unified Multimodal Framework

What companies such as Comma.ai and Tesla are doing is employing a hybrid architecture combining vision and radar which enable a balance of robustness and efficiency:

- Vision for primary perception—high-resolution, low-cost, and semantically rich.
- Radar for adverse conditions—weather-resilient and velocity-sensitive.
- LiDAR for benchmarking only—useful for training ground truth, not live deployment.

Figure 12 shows the modular stack which can be tuned via hyperparameters during the training stage to give more weight to favor certain inputs. Such an approach is utilized by Comma.ai to prioritize key sensor inputs to the model deployed in Openpilot [7].

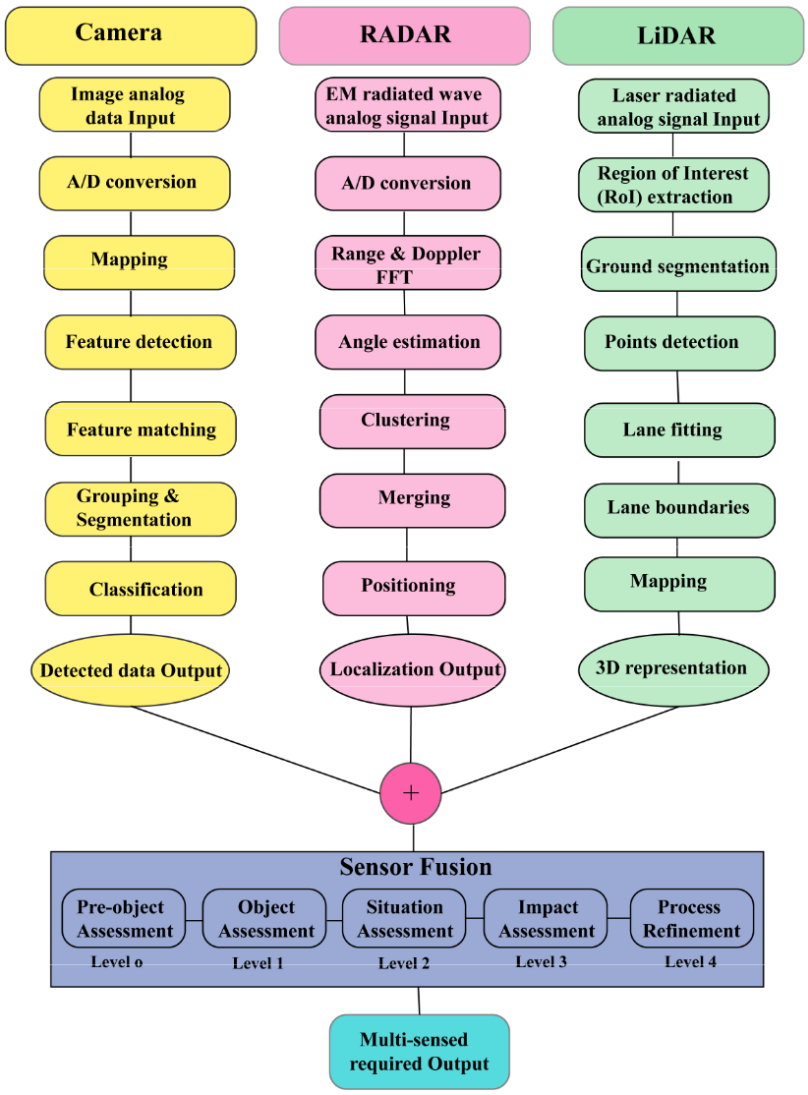

Fig. 12. Modular architecture integrating vision, radar, and LiDAR inputs [23].

This design maximizes generalization while minimizing cost and computation demands, and it reflects a biologically plausible model: humans drive with vision, but radar is the backup.

D. Summary

In order to eventually achieve SAE Level 5 Autonomy, the currently existing data points to the possibility of vision-only stacks surpassing fusion models in the near future:

- Vision stacks perform within 5% of LiDAR/Radar fusion across key metrics, achieving 56.2 mAP on nuScenes compared to 61.5 for fusion [17].
- With radar fallback, this gap shrinks to $< 2\%$, recovering performance to 59.8 mAP [18].

- Vision systems run 3–5x faster and consume 6–10x less power—9.8 ms and 2.1 W vs. $\geq$ 30 ms and 12 W to 20 W for fusion [6, 9].
- Vision stacks avoid fusion noise [6], reduce calibration burden [5], and cost significantly less, with LiDAR accounting for over 75 % of ADS hardware expense [20, 21].

In short, vision is sufficient for autonomy. Radar makes it resilient. LiDAR makes it expensive, but can still have its uses in training datasets as benchmarks. The ideal multi-modal solution is outlined in Figure 13.

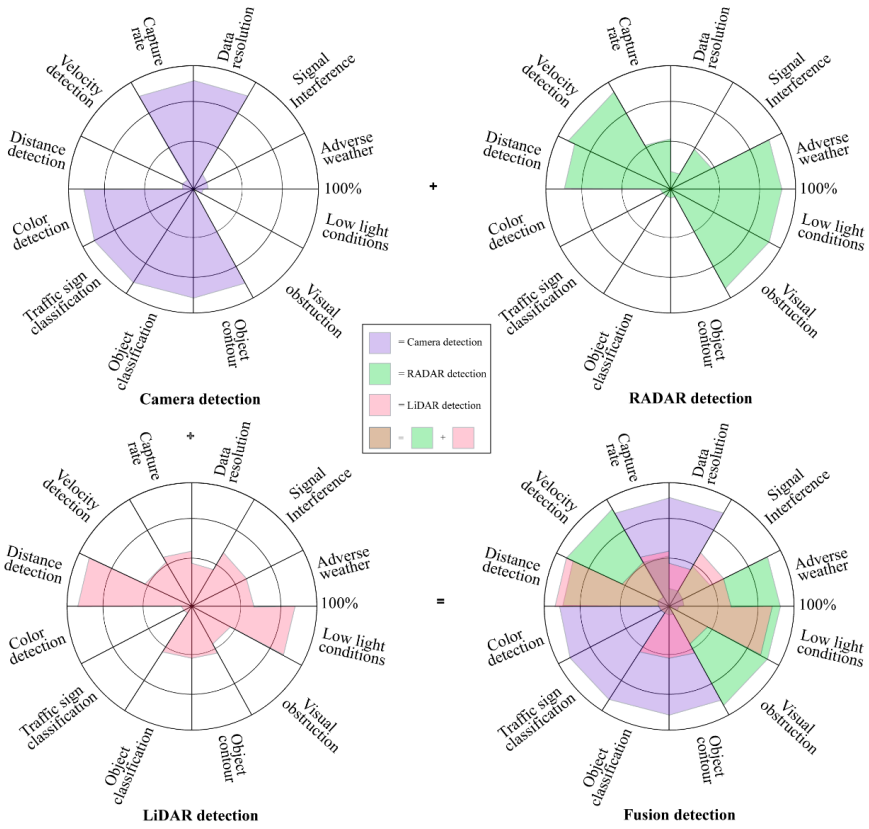

Fig. 13. A Multi-Modal Approach to ADS [23]

V. Conclusion

This paper sets out to compare the existing technology employed in Level 2 ADS systems deployed in production around the world today. While approaches such as Waymo's LiDAR/Vision/Radar fusion are commendable and show great promise in deployment, the drawbacks mentioned and advancements in the vision stack in particular offer a clear alternative that is cheaper and more energy efficient becoming more effective each day. [1]

[1] An LLM was used strictly for formatting IAW IEEE Editorial Style Manual for Authors. All thoughts, prose, and conclusions are my own.

References

[1] SAE International, "Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles," Apr. 2021, Revised April 2021, Issued January 2014. DOI: 10.4271/J3016_202104. [Online]. Available: https://doi.org/10.4271/J3016_202104.

[2] S. Royo and M. Ballesta-Garcia, "An overview of lidar imaging systems for autonomous vehicles," *Applied Sciences*, vol. 9, no. 19, 2019, ISSN: 2076-3417. DOI: 10.3390/app9194093. [Online]. Available: https://www.mdpi.com/2076-3417/9/19/4093.

[3] K. Petermann, *Advances in Optoelectronics*. Berlin, Germany: Springer, 1988.

[4] Y. Zeng et al., "Rt3d: Real-time 3-d vehicle detection in lidar point cloud for autonomous driving," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3434–3440, 2018. DOI: 10.1109/LRA.2018.2852843.

[5] Z. Han, X. Li, and Y. Zhou, "4d millimeter-wave radar in autonomous driving: A survey," *arXiv preprint arXiv:2306.04242*, 2023. [Online]. Available: https://arxiv.org/abs/2306.04242.

[6] K. Rana, L. Wang, and M. Gupta, "The perception systems used in fully automated vehicles: A comparative analysis," *Multimedia Tools and Applications*, vol. 82, no. 11, pp. 15 923–15 952, 2023. DOI: 10.1007/s11042-023-16000-4.

[7] M. Goff et al., "Learning to drive from a world model," 2025. arXiv: 2504.19077 [cs.CV]. [Online]. Available: https://arxiv.org/abs/2504.19077.

[8] P. Esser, R. Rombach, and B. Ommer, "Taming transformers for high-resolution image synthesis," 2021. arXiv: 2012.09841 [cs.CV]. [Online]. Available: https://arxiv.org/abs/2012.09841.

[9] Y. Chen, X. Huang, T. Ma, and et al., "End-to-end autonomous driving: Challenges and frontiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024, Early access. DOI: 10.1109/TPAMI.2024.XXXXXXX.

[10] E. Santana and G. Hotz, "Learning a driving simulator," *CoRR*, vol. abs/1608.01230, 2016. arXiv: 1608.01230. [Online]. Available: http://arxiv.org/abs/1608.01230.

[11] A. Vaswani et al., "Attention is all you need," 2023. arXiv: 1706.03762 [cs.CL]. [Online]. Available: https://arxiv.org/abs/1706.03762.

[12] X. Wang, K. Li, and A. Chehri, "Multi-sensor fusion technology for 3d object detection in autonomous driving: A review," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 2, pp. 1148–1165, 2024. DOI: 10.1109/TITS.2023.3317372.

[13] H. Haghighi, X. Wang, H. Jing, and M. Dianati, "Data-driven camera and lidar simulation models for autonomous driving: A review from generative models to volume renderers," *arXiv preprint arXiv:2402.10079*, 2024. [Online]. Available: https://arxiv.org/abs/2402.10079.

[14] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 3354–3361. DOI: 10.1109/CVPR.2012.6248074.

[15] H. Caesar et al., "Nuscenes: A multimodal dataset for autonomous driving," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 11 618–11 628. DOI: 10.1109/CVPR42600.2020.01164.

[16] Z. Li et al., "Bevformer: Learning bird's-eye-view representation from multi-camera images via spatiotemporal transformers," 2022. arXiv: 2203.17270 [cs.CV]. [Online]. Available: https://arxiv.org/abs/2203.17270.

[17] Q. Zhang, Y. Wu, and Z. Liu, "Multi-sensor fusion object detection in autonomous driving: A survey," *Sensors*, vol. 25, no. 9, p. 2794, 2023. DOI: 10.3390/s25092794.

[18] D. Wu, F. Yang, B. Xu, P. Liao, and B. Liu, "A survey of deep learning based radar and vision fusion for 3d object detection in autonomous driving," 2024. arXiv: 2406.00714 [cs.CV]. [Online]. Available: https://arxiv.org/abs/2406.00714.

[19] M. Dreissig, D. Scheuble, F. Piewak, and J. Boedecker, "Survey on lidar perception in adverse weather conditions," pp. 1–8, 2023. DOI: 10.1109/IV55152.2023.10186539.

[20] P. Wang, "Research on comparison of lidar and camera in autonomous driving," *Journal of Physics: Conference Series*, vol. 2093, p. 012 032, Nov. 2021. DOI: 10.1088/1742-6596/2093/1/012032.

[21] S. Sajjad, A. Hussain, and F. Alam, "A comparative analysis of camera, lidar and fusion-based deep neural networks for vehicle detection," *International Journal of Innovations in Science and Technology*, vol. 3, no. Special Issue, pp. 15–24, 2021. DOI: 10.33411/IJIST/2021030514. [Online]. Available: https://journal.50sea.com/index.php/IJIST/article/view/131.

[22] E. Musk, *Tesla autonomy day*, https://www.youtube.com/watch?v=Ucp0TTmvqOE, Quote at 1:41:00: "LiDAR is a fool's errand. Anyone relying on LiDAR is doomed.", 2019.

[23] M. Hasanujjaman, M. Chowdhury, and Y. M. Jang, "Sensor fusion in autonomous vehicle with traffic surveillance camera system: Detection, localization, and ai networking," *Sensors*, vol. 23, p. 3335, Mar. 2023. DOI: 10.3390/s23063335.