Proceedings of the 7th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/2027.39429

Millimeter-Wave Radar and Camera Multi-modal Fusion for 3D Object Detection in Autonomous Driving

Guohang Su

School of Shipping and Naval Architecture, Chongqing Jiaotong University, Chongqing, China
sgh2372977605@outlook.com

Abstract. Environment-perception modules underpin autonomous driving system operation for complex road scene comprehension. Three-dimensional (3D) object detection completes localization and classification work for road traffic participants including vehicles, pedestrians and cyclists. Camera sensors deliver rich texture and semantic data with performance degradation under low-light, fog and heavy-rain conditions. Millimeter-wave radar supports all-weather range and radial-velocity measurement with sparse point-cloud output and limited semantic discrimination capacity. Multi-modal fusion technology leverages complementary sensor characteristics for perception robustness improvement. This paper conducts millimeter-wave radar analysis covering historical evolution, working principle and core sensor characteristics. It compares sensor complementary attributes and three mainstream fusion paradigms, namely data-level, decision-level and feature-level fusion, to build the analysis basis for representative bird's-eye view (BEV) based Transformer cross-attention fusion networks. Technical bottlenecks covering space-time misalignment, modality imbalance and computing overhead are sorted out for subsequent discussion of practical engineering constraints and promising research directions in this field. Engineering-oriented suggestions are finally provided covering online calibration, model compression, physical-prior embedding and adverse-weather data augmentation.

Keywords: Autonomous driving, 3D object detection, millimeter-wave radar, multi-modal fusion, bird's-eye-view

1. Introduction

Environment perception constitutes the fundamental support for vehicle safety decision-making in fast-developing autonomous driving technology [1]. Three-dimensional (3D) object detection as one core perception task acquires location information of pedestrians vehicles and other traffic participants under complex road conditions to generate input data for path planning and motion control modules [2]. Cameras and millimeter-wave radar serve as low-cost mainstream onboard sensors for mass-produced intelligent vehicles. Rich texture and semantic information can be captured by cameras with obvious performance drop under poor illumination fog and heavy rain [3]. Millimeter-wave radar realizes all-weather ranging and radial velocity measurement with sparse point-cloud output and weak semantic classification capability [4]. Multi-modal fusion makes full use of complementary sensor advantages to raise perception robustness of real-world autonomous

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2027.39429

driving systems [5]. Data-level decision-level and feature-level fusion represent three major categories for existing radar-camera fusion schemes [6]. Bird's-Eye-View representation promotes extensive research interest for feature-level fusion. Transformer-based cross-attention modules are adopted by representative works CenterFusion and CRAFT to complete feature alignment of heterogeneous sensor data [7, 8].

Practical bottlenecks still exist for current technical solutions. Complete elimination of spatial-temporal misalignment between radar and image signals cannot be achieved in dynamic driving scenes [9]. Network training processes produce modality imbalance phenomena. Dense image features receive excessive model dependence and valid radar information gets ignored under visual input failure caused by adverse weather [10]. High computing overhead from multiple fusion models blocks deployment on automotive embedded hardware with limited computing resources.

Promising detection results have been obtained by existing studies on public benchmarks represented by nuScenes. Most research work focuses on indicator promotion with insufficient analysis of practical engineering constraints. Systematic sorting is absent for sensor characteristic influence over fusion performance and accuracy-computation cost trade-off relations. This paper carries out targeted research for millimeter-wave radar-camera multi-modal 3D object detection based on above-mentioned research gaps. Sensor strengths and defects are analyzed, mainstream fusion pipelines are compared, and existing technical challenges are discussed.

Section 2 covers millimeter-wave radar sensor principles and historical evolution. Section 3 implements analysis of sensor complementary characteristics and typical fusion paradigms. Section 4 completes comparison and strength-weakness analysis for mainstream bird's-eye view (BEV)-based fusion networks. Section 5 summarizes core open-ended problems of radar-camera fusion perception and engineering-oriented research directions. Section 6 draws the final conclusion.

2. Historical development and working principle of millimeter-wave radar

Millimeter-wave radar represents one type of active ranging sensor working in the 30 GHz-300 GHz electromagnetic frequency band corresponding to 1 mm-10 mm wavelength. 77 GHz frequency band gets widely adopted by automotive-grade radar for advanced driver assistance systems and autonomous-driving perception. This section completes review work for its historical development physical working principle and sensor characteristics with reference to overseas research outcomes [11-13].

2.1. Historical development

Military aircraft and ship monitoring requirements in World War II drive the birth of early radar technology with main working frequency set for long-wavelength microwave band. Millimeter-wave hardware component technological progress from the 1970s pushes millimeter-wave radar application scenario expansion from military reconnaissance to civilian industrial fields. First-generation automotive millimeter-wave radar devices emerge in the 1990s. Discrete semiconductor components lead to large hardware volume and high power consumption for early products.

Commercialization of frequency-modulated continuous-wave (FMCW) radar chips completes in the early-2000s. SiGe-BiCMOS and other semiconductor processes reduce radar sensor volume and cost for mass-produced vehicle realization of adaptive cruise control and automatic emergency-braking functions. Multiple-input multiple-output antenna array technology obtains wide automotive radar application after 2015. Angular resolution improvement and sparse target

DOI: 10.54254/2755-2721/2027.39429

point-cloud output get realized to supply hardware foundation for multi-modal fusion 3D-detection algorithms [11]. 4D millimeter-wave radar capable of simultaneous range azimuth elevation and velocity acquisition becomes core research object for next-generation low-cost autonomous-driving perception [12].

2.2. Working principle

FMCW constitutes the technical framework for most commercial automotive millimeter-wave radar. Transmitting antennas output linear frequency-modulated chirp signals. Receiving antennas capture echo signals reflected by surrounding objects. Mixing processing of transmitted signals and time-delayed echo signals generates intermediate-frequency beat signals for target parameter calculation.

Signal time-delay for round-trip electromagnetic wave propagation between radar and target is defined in equation (1):

$$\tau = \frac{2d}{c} \tag{1}$$

In equation (1), d stands for distance value between radar and target. c represents the speed of light (3×10^8 m/s). The coefficient value "2" accounts for the complete round-trip travel of electromagnetic waves from radar transmitter to target and back to radar receiver.

Relative radial motion of moving targets produces Doppler frequency shift for echo signals, as shown in equation (2).

$$f_d = \frac{2v_r}{\lambda} = \frac{2v\cos\theta}{\lambda} \tag{2}$$

In equation (2),

- v_r is radial relative velocity between radar and target.
- λ represents electromagnetic wavelength.
- θ records the included angle between target motion direction and radar sight-line direction.

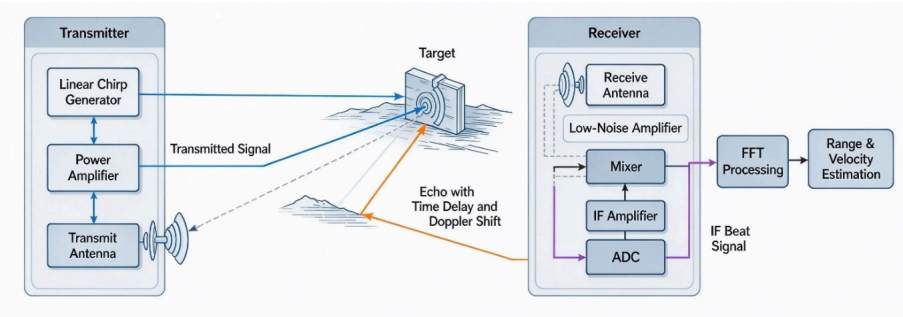

Figure 1. Schematic diagram of FMCW millimeter-wave radar signal transmission-reception workflow (picture: original)

Fast Fourier transform processing for beat-frequency signals supports simultaneous solving of target range radial velocity and angle-of-arrival. Phase difference information from multi-channel receiving terminals gets used by multiple-input multiple-output radar for target azimuth and elevation angle estimation to generate sparse 3D radar-point sets $\mathscr{R} = \{p_i = (x_i, y_i, z_i, v_i)\}_{i=1\ldots N}$.

DOI: 10.54254/2755-2721/2027.39429

Each radar point stores spatial coordinate and radial-velocity information without RGB texture data [11]. As shown in Figure 1, the FMCW millimeter-wave radar transmits a linear chirp signal through the transmitter. After being reflected by the target, the received echo contains time-delay and Doppler shift information. The mixer generates an IF beat signal, which is then processed by FFT to extract the target range and velocity.

3. Sensor characteristics and fusion paradigms

Strong complementary relations exist between camera and millimeter-wave radar. Overall perception performance gets improved by multi-sensor fusion mechanisms for mutual offset of respective sensor defects. This chapter implements analysis for sensor complementarity and three mainstream fusion-level classification schemes with foreign literature support [14, 15].

3.1. Complementary characteristics of radar and camera

Millimeter-wave radar carries unique advantages and physical limitations compared with visual cameras. Fog rain dust and night illumination changes cannot interfere millimeter-band electromagnetic wave transmission to guarantee stable sensor output [12]. Extra visual motion-estimation computation gets avoided by direct radial-velocity output for moving targets. Compact low-power and low-cost chip design supports large-scale vehicle installation.

Independent high-level autonomous-driving perception gets restricted by inherent sensor defects. Far lower point-cloud density compared with image pixels brings insufficient target geometric sampling results. Missing texture and color information causes poor semantic-classification performance. Multi-path electromagnetic reflection from guardrails road surfaces and buildings creates ghost points and clutter noise. Limited elevation-angle resolution of traditional automotive radar generates height-estimation uncertainty for 3D-object detection [13]. Passive working mode is adopted by cameras.Dense pixel data with texture color and semantic features get output for object category classification and contour description [14]. Direct accurate depth measurement cannot be completed by monocular cameras. Night backlight dense fog and heavy-rain scenarios produce sharp detection performance decline. Motion blur and over-exposure bring extra interference for image algorithm execution.

Active millimeter-wave radar outputs sparse point clouds with accurate range and radial-velocity measurement free from illumination and weather disturbance. Visual texture acquisition cannot be realized by radar sensors. Sparse sampling points cannot fully describe object contour details. Table 1 records qualitative comparison results for two kinds of sensors.

Table 1. Qualitative comparison of camera and millimeter-wave radar

Property	Camera	Millimeter-wave radar
Texture information	Abundant	Absent
Depth measurement	Indirect, noisy	Direct, accurate range
Velocity measurement	Need post computation	Direct radial velocity output
Adverse weather robustness	Poor	Excellent
Data density	Dense pixel	Sparse point cloud
Hardware cost	Low	Medium low

DOI: 10.54254/2755-2721/2027.39429

Visual modules provide semantic and contour cues. Radar supplies reliable distance and velocity constraints under visual observation degradation [14].

3.2. Classification of radar-camera fusion schemes

Data-level fusion decision-level fusion and feature-level fusion cover deep-learning-based radar-camera fusion approaches according to multi-sensor information merging stages .

3.2.1. Data-level fusion (early-fusion)

Feature-extraction network input receives raw or pre-processed multi-sensor data. Extrinsic calibration matrix completes radar-point projection onto image coordinate space. Original radar-point information gets appended to image channels as neural-network input tensor. Mathematical projection formula for coordinate transformation from radar ego-coordinate to image pixel coordinate is shown in equation (3):

$$s \cdot \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \cdot T_{radar2cam} \cdot \begin{bmatrix} x_r \\ y_r \\ z_r \\ 1 \end{bmatrix} \tag{3}$$

In equation (3), x_r, y_r, and z_r represent the radar-point coordinates in the radar coordinate frame. $\mathbf{T}_{radar2cam}$ means extrinsic transformation matrix. K is camera intrinsic matrix. (u,v) are image-plane pixel coordinates. s serves as scale factor.

Maximum original sensor information retention represents the main merit of data-level fusion. Strict sensor calibration time-synchronization and radar-point quality requirements restrict practical application. Sparse radar-point projection onto images creates unbalanced spatial distribution and introduces extra noise [14].

3.2.2. Decision-level fusion (late-fusion)

Independent detection pipelines run for radar and camera sensors. Separate 3D-detection results covering bounding boxes category and confidence scores get generated. Non-maximum suppression Kalman filtering or probability weighting complete post-processing fusion for final result output. Low coupling degree exists between two sensor branches. Independent debugging and upgrading can be finished for each subsystem [15]. Cross-modal feature interaction cannot be exploited for limited performance gain [14].

3.2.3. Feature-level fusion (middle-fusion)

Image branch and radar branch extract respective intermediate-depth features. Heterogeneous-modality feature maps complete alignment and fusion inside neural networks for detection-head prediction. BEV representation and Transformer cross-attention mechanism push feature-level fusion to become mainstream technical solution. Automatic cross-modal correlation learning balances camera semantic information and radar geometric measurement information. Higher computation overhead constitutes its primary defect [14]. As shown in Figure 2, three radar–

DOI: 10.54254/2755-2721/2027.39429

camera fusion paradigms are presented, namely data-level, feature-level, and decision-level fusion. In the data-level paradigm, raw radar and camera data are directly fused before state estimation. In the feature-level paradigm, feature vectors extracted from radar and camera data are fused to improve target state estimation. In the decision-level paradigm, independent detection and tracking results from radar and camera are combined at the decision level to obtain the final target state.

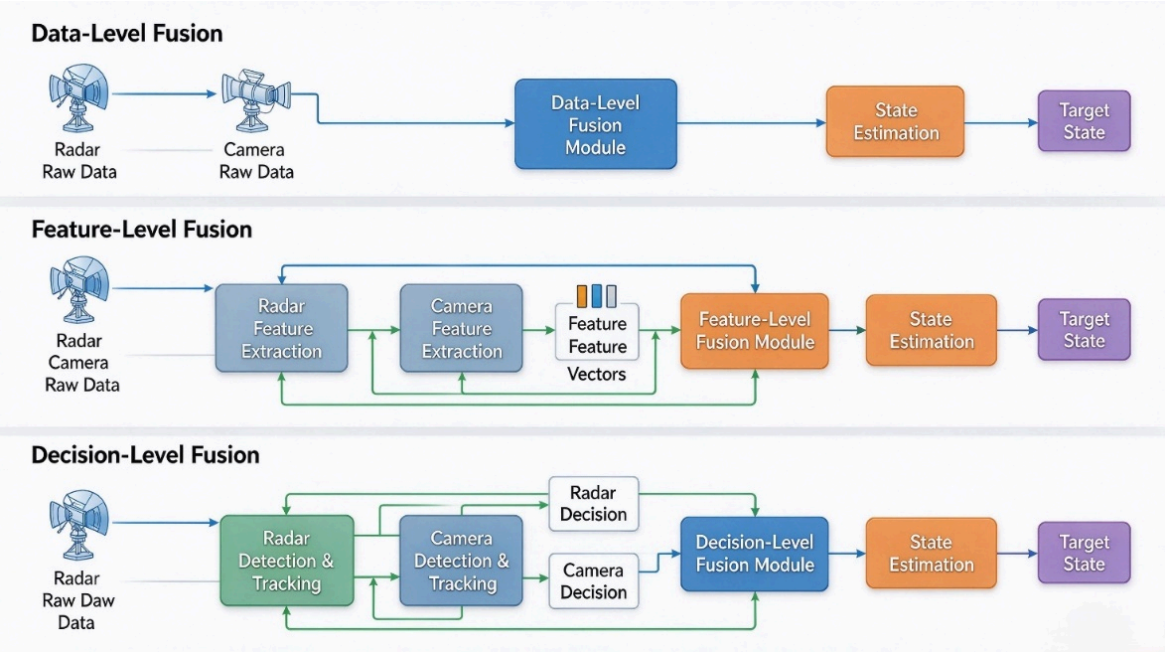

Figure 2. Schematic workflow of three radar-camera fusion paradigms (picture: original)

3.3. Overview of traditional fusion mechanisms

Hand-crafted features and filtering algorithms dominate pre-deep-learning traditional fusion mechanisms. Radar detection results generate image region-of-interest proposals. Hand-crafted visual feature extraction inside target ROI supports subsequent object classification. Kalman-filter-based track-level fusion completes visual-detection track and radar track association for target state estimation optimization.

Traditional fusion systems rely heavily on manual parameter tuning. Stable operation gets maintained under simple road scenarios. Complex crowded traffic scenes with occlusion and multi-target interaction produce poor adaptability. Deep-learning-driven feature-level fusion schemes gradually replace traditional hand-crafted mechanisms [14].

4. BEV-based radar-camera fusion detection networks

Top-down bird-eye perspective under vehicle ego-coordinate system gets adopted by Bird's-Eye-View representation for sensor-data projection. Heterogeneous-modality data unification on identical top-down feature map simplifies cross-sensor spatial alignment work.

4.1. Fundamental principle of BEV spatial representation

Lift-Splat-Shoot serves as widely-used paradigm for camera-to-BEV transformation. Depth-distribution prediction lifts image features from perspective-view into 3D space. Feature

DOI: 10.54254/2755-2721/2027.39429

information gets splattered onto BEV grid feature map. BEV feature map is defined as $\mathbf{F}_{bev} \in \mathbb{R}^{C \times H_b \times W_b}$. H_b and W_b represent BEV grid height and width.

C records feature-channel dimension. Real-world physical area around ego-vehicle corresponds to each grid cell.

Point-wise feature embedding and grid-aggregation operation encode sparse radar-point cloud R into BEV feature representation $\mathbf{F}_{radar}^{bev}$ which is given in equation (4):

$$\mathbf{F}_{radar}^{bev}(u_{bev}, v_{bev}) = \text{AvgPool}(\{\mathscr{E}(p_i) \mid p_i \in \mathscr{R},\ \text{grid}(p_i) = (u_{bev}, v_{bev})\}) \tag{4}$$

In equation (4), $\mathscr{E}(\cdot)$ means point-feature embedding network. grid$(\cdot)$ maps 3D radar point to BEV grid index coordinate (u_{bev}, v_{bev}). Cross-modal fusion operation executes on unified BEV plane after acquisition of camera-derived BEV feature $\mathbf{F}_{img}^{bev}$ and radar-derived BEV feature $\mathbf{F}_{radar}^{bev}$. As shown in Figure 3, the BEV feature generation pipeline takes radar and camera inputs as the initial data sources. Radar point clouds are processed to generate radar BEV features, while multi-view camera images are transformed into camera BEV features through feature extraction and view transformation. The radar and camera BEV features are then aligned and fused in the BEV space, and the output BEV feature map provides comprehensive representations for subsequent perception tasks.

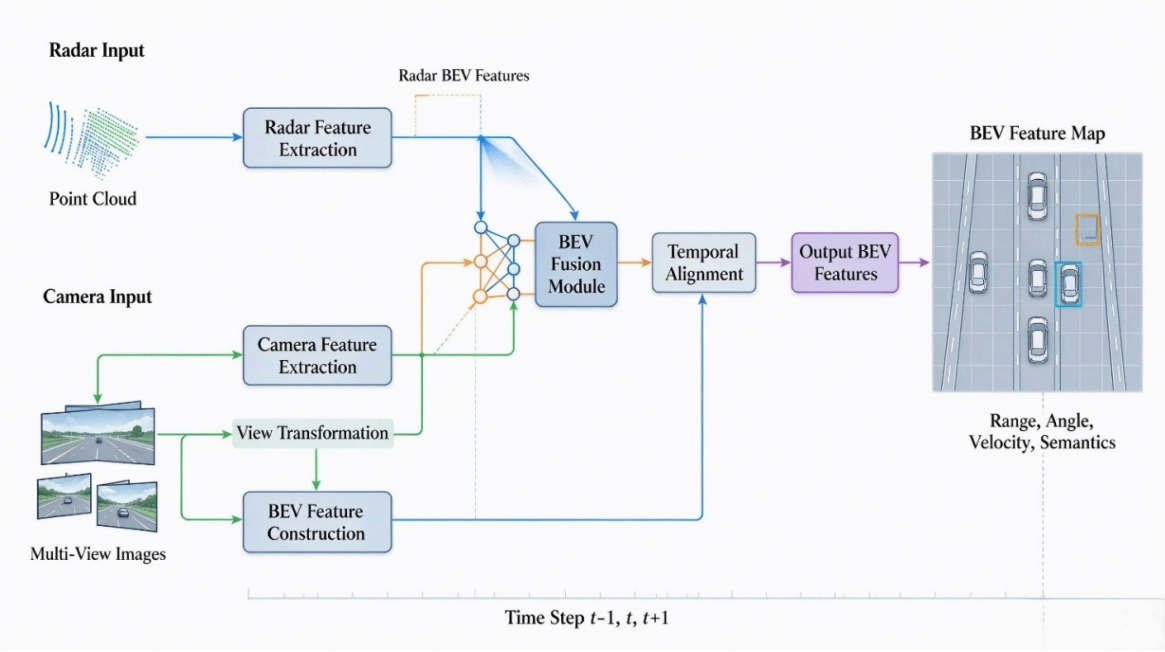

Figure 3. Illustration of BEV feature generation pipeline for radar-camera inputs (picture: original)

4.2. Typical transformer-based fusion methods

Feature query and information interaction between two modalities get completed by Transformer-based cross-attention modules as core components for BEV-based radar-camera fusion networks. Standard cross-attention calculation formula is written in equation (5):

DOI: 10.54254/2755-2721/2027.39429

$$\text{CrossAttn}(Q, K, V) = \text{Softmax}\left(\frac{QK^{\top}}{\sqrt{d_k}}\right)V \tag{5}$$

Equation (5) defines the cross-attention computation used for multi-modal feature interaction. Query Q comes from one-modality feature. Key-value pair (K, V) originates from another-modality feature. d_k represents feature dimension of key vector.

Radar points get projected onto image plane for image feature map enhancement in CenterFusion [4]. Fusion computation runs on perspective-view image plane instead of BEV space with inherent view-domain gap.

Spatio-contextual fusion transformer is proposed by CRAFT [3]. Cross-attention builds connection between image proposal features and radar features for 3D-box regression performance improvement. Good detection accuracy is achieved on public datasets. High-quality image proposal generation constitutes system precondition. Visual quality drop under adverse weather triggers overall performance degeneration.

Separate BEV encoders for radar and camera are constructed by RCBEVDet [9]. Deformable cross-attention module completes BEV-domain feature alignment. Feature content guides adaptive sampling-position offset for calibration error and vehicle-motion-caused spatial misalignment alleviation. Extra computation load from deformable-attention layers creates deployment pressure for automotive embedded chips.

Dual cross-attention structure is adopted by bidirectional cross-attention fusion network [8]. Radar features query image semantic information. Image features query radar geometric and velocity information. Bidirectional information flow mitigates modality-imbalance phenomenon. Model parameter quantity and inference latency get increased simultaneously.

5. Limitations and future outlook

5.1. Practical limitations of current fusion algorithms

The NuScenes and View-of-Delft datasets provide benchmark support for existing fusion algorithms with competitive test results. Three major bottlenecks appear for algorithm practical deployment.

Vehicle vibration sensor calibration drift and asynchronous sampling timestamps produce spatial offset and temporal desynchronization between radar-point cloud and image frames. Complete misalignment elimination cannot be realized by finely calibrated driving systems [9]. Feature alignment error misleads cross-attention modules and reduces detection accuracy.

Dense image-branch gradient signals dominate network training loss functions. Visual-prior knowledge receives main model learning focus and effective radar geometric and velocity cues get ignored. Model performance drops sharply under foggy or night visual-perception-failure conditions [10]. Adaptive-weight training strategy design for radar information activation under visual degradation remains unresolved.

Massive parameters exist in most Transformer-based BEV fusion models. High inference latency blocks deployment on low-power automotive embedded hardware. Detection accuracy and inference speed show obvious trade-off relations. Academic research prioritizes mean average precision improvement and ignores mass-production-vehicle embedded-deployment requirements.

Public datasets contain limited heavy-rain and dense-fog sample data. Model generalization capability gets weakened for harsh real-world environment adaptation.

DOI: 10.54254/2755-2721/2027.39429

5.2. Future research and engineering-optimization directions

Corresponding research ideas are summarized for above-mentioned bottlenecks.

Online dynamic calibration modules get embedded into network frameworks. Adaptive compensation completes extrinsic-parameter drift and vehicle-vibration-caused time-offset correction instead of dependence on offline fixed calibration parameters. Learnable gating-weight modules get constructed. Visual-quality estimation drives dynamic contribution-weight adjustment for radar branch and camera branch. Radar-feature contribution ratio gets automatically elevated under bad-weather image-signal corruption to guarantee perception robustness.

Network pruning quantization and knowledge-distillation process execute for Transformer cross-attention modules. Lightweight network-architecture design pursues balance between detection precision and automotive-chip-oriented real-time inference latency. Large-scale fog heavy-rain night-backlight radar-camera data get collected. Dataset bias gets repaired for complex-weather-oriented model generalization enhancement.

Millimeter-wave radar physical prior knowledge covering radar cross section characteristics and multi-path-clutter model gets introduced into neural-network design. Ghost-point interference gets restrained and sparse radar-information utilization efficiency gets promoted.

6. Conclusion

Obvious complementary sensing properties are owned by millimeter-wave radar and camera. Multi-modal fusion represents essential technical path for autonomous-driving system 3D-object-detection robustness promotion. Millimeter-wave radar historical review FMCW working- principle interpretation and sensor-characteristic analysis are finished in this paper. Comparative analysis covers three classic fusion levels of data-level decision-level and feature-level fusion. Transformer cross-attention acts as mainstream feature-level-fusion technical scheme under BEV unified representation. Space-time misalignment training-phase modality imbalance and high computational overhead constitute prominent practical obstacles for current approaches.

Engineering-oriented constraint analysis deserves more research focus besides benchmark-index pursuit. Modality-adaptive lightweight physics-informed radar-camera fusion algorithm development promotes multi-modal-perception-technology large-scale deployment on mass-production autonomous vehicles.

Three main findings emerge from this review. First, the complementarity between the two sensors is structural rather than incidental: radar supplies range and radial velocity that cameras cannot measure directly, while cameras supply the texture and semantic cues that radar point clouds lack. Second, the three fusion paradigms differ systematically in their cost-benefit profile. Data-level fusion retains the most raw information but is highly sensitive to space-time misalignment; decision-level fusion is robust and easy to deploy but discards cross-modal intermediate evidence; feature-level fusion offers the best accuracy-efficiency balance at the price of higher design complexity. Third, the combination of bird's-eye view representation and Transformer cross-attention has become the mainstream scheme because it places heterogeneous sensor outputs in one geometric frame and lets the network learn the alignment instead of relying on hand-crafted rules.

This review has two main limitations. Only qualitative comparison is carried out, because publicly available radar-camera datasets differ in sensor configuration, annotation format and evaluation protocol, which prevents a unified quantitative benchmark. The discussion is also restricted to perception-level fusion and does not cover downstream prediction and planning. Three directions are suggested for future engineering work. Online dynamic calibration modules should be

DOI: 10.54254/2755-2721/2027.39429

embedded so that sensor drift is compensated during inference. Pruning, quantization and knowledge distillation should be combined under explicit hardware latency budgets. Radar physical priors such as radar cross section and multi-path clutter models should be injected into network design so that sparse radar information is exploited more efficiently.

References

[1] Wang, L., Zhang, X., Song, Z., Bi, J., Zhang, G., Wei, H., Tang, L., Yang, L., Li, J., Jia, C., & Zhao, L. (2023). Multi-modal 3D object detection in autonomous driving: A survey and taxonomy. *IEEE Transactions on Intelligent Vehicles, 8*(7), 3781–3798. https://doi.org/10.1109/TIV.2023.3264658

[2] Wei, Z., Zhang, F., Chang, S., Liu, Y., Wu, H., & Feng, Z. (2022). MmWave radar and vision fusion for object detection in autonomous driving: A review. *Sensors, 22*(7), 2542. https://doi.org/10.3390/s22072542

[3] Kim, Y., Kim, S., Choi, J. W., & Kum, D. (2023). CRAFT: Camera-radar 3D object detection with spatio-contextual fusion transformer. *Proceedings of the AAAI Conference on Artificial Intelligence, 37*(1), 1160–1168. https://doi.org/10.1609/aaai.v37i1.25198

[4] Nabati, R., & Qi, H. (2021). CenterFusion: Center-based radar and camera fusion for 3D object detection. In *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 1526–1535). IEEE. https://doi.org/10.1109/WACV48630.2021.00157

[5] Li, Z., Wang, W., Li, H., Xie, E., Sima, C., Lu, T., Yu, Q., & Dai, J. (2025). BEVFormer: Learning bird's-eye-view representation from LiDAR-camera via spatiotemporal transformers. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 47*(3), 2020–2036. https://doi.org/10.1109/TPAMI.2024.3515454

[6] Wu, D., Yang, F., Xu, B., Liao, P., & Liu, B. (2024). A survey of deep learning based radar and vision fusion for 3D object detection in autonomous driving. arXiv:2406.00714. https://doi.org/10.48550/arXiv.2406.00714

[7] Lin, Z., Liu, Z., Wang, Y., Zhang, L., & Zhu, C. (2024). RCBEVDet++: Toward high-accuracy radar-camera fusion 3D perception network. arXiv:2409.04979. https://doi.org/10.48550/arXiv.2409.04979

[8] He, W., Deng, Z., Pan, P., & Ye, Y. (2025). Bidirectional radar–camera fusion with dual cross-attention for 3-D object detection. *IEEE Sensors Journal, 25*(19), 36526–36540. https://doi.org/10.1109/JSEN.2025.3597983

[9] Lin, Z., Liu, Z., Xia, Z., Wang, X., Wang, Y., Qi, S., Dong, Y., Dong, N., Zhang, L., & Zhu, C. (2024). RCBEVDet: Radar-camera fusion in bird's eye view for 3D object detection. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 14928–14937). IEEE. https://doi.org/10.1109/CVPR52733.2024.01414

[10] Wan, R., Meng, W., Zhao, T., & Lu, W. (2026). Overcoming radar sparsity and cross-view misalignment: A sparse-to-sparse fusion paradigm for robust 3D object detection. *Digital Signal Processing, 171*, 105831. https://doi.org/10.1016/j.dsp.2025.105831

[11] Patole, S. M., Torlak, M., Wang, D., & Ali, M. (2017). Automotive radars: A review of signal processing techniques. *IEEE Signal Processing Magazine, 34*(2), 22–35. https://doi.org/10.1109/MSP.2016.2628914

[12] Cai, Y., Bai, J., Shen, H., Huang, L., Rao, B., & Wang, H. (2025). Development of low-cost single-chip automotive 4D millimeter-wave radar. *Sensors, 25*(15), 4640. https://doi.org/10.3390/s25154640

[13] Scheiner, N., Kraus, F., Appenrodt, N., Dickmann, J., & Sick, B. (2021). Object detection for automotive radar point clouds – a comparison. *AI Perspectives, 3*(1), 6. https://doi.org/10.1186/s42467-021-00012-z

[14] Shi, K., He, S., Shi, Z., Chen, A., Xiong, Z., Chen, J., & Luo, J. (2026). Radar and camera fusion for object detection and tracking: A comprehensive survey. *IEEE Communications Surveys & Tutorials, 28*, 3478–3520. https://doi.org/10.1109/COMST.2025.3599596

[15] Deng, K., Xing, L., Wu, H., Ma, H., Ling, Y., & Gao, J. (2025). Advances in object detection for autonomous driving using mmwave radar and camera: A comprehensive survey. *Journal of King Saud University Computer and Information Sciences, 37*(10), 328. https://doi.org/10.1007/s44443-025-00302-0