4D Radar Perception Algorithms for Autonomous Driving: A Review

Xumin Wu, Jun Zhou, Jilin Mei, Chen Min, and Yu Hu

***Abstract*—Research on 4D millimeter-wave radar perception algorithms has flourished in recent years, extending from signal processing and object detection to semantic segmentation, motion estimation, occupancy prediction, and dynamic scene reconstruction. This review organizes the field according to the evolution of perception tasks and algorithms. It first introduces radar fundamentals, data representations, and quality-enhancement methods, and then reviews object-level perception, motion and localization, local and dense spatial perception, and dynamic scene understanding. Across these directions, we compare radar-only learning, multimodal fusion, and cross-modal supervision and knowledge distillation. Particular attention is paid to how elevation, Doppler measurements, and radar physical priors are exploited across tasks. We further summarize the task coverage, input data, annotations, and evaluation protocols of existing datasets, clarifying the empirical support for different research directions. Finally, we discuss the common challenges and future directions of 4D radar perception for autonomous driving. This review provides a task-oriented perspective on the transition from sparse object perception to dynamic spatial understanding.**

***Index Terms*—4D millimeter-wave radar, autonomous driving, Doppler velocity, dynamic scene understanding, multimodal perception, occupancy prediction, scene flow**

I. Introduction

AUTONOMOUS driving requires perception that remains reliable over long ranges, in dynamic traffic, and under adverse weather. Millimeter-wave radar is compact, cost-effective, robust under many adverse conditions, and directly measures radial velocity. Advances in multiple-input multiple-output (MIMO) arrays, imaging radar, and signal processing now enable elevation-resolved measurements, transforming radar from a target detector into a sensor that jointly observes 3D spatial structure and radial motion [1], [2].

With this increase in spatial resolution, 4D radar research has expanded from vehicle detection to point-cloud enhancement, semantic segmentation, tracking, scene flow, ego-motion estimation, localization, occupancy prediction, and dynamic scene reconstruction. These tasks operate at different output levels, from individual returns and objects to local maps, occupied space, and temporally evolving scenes. A review of 4D radar perception should therefore explain how algorithms have evolved across these levels rather than treating detection as the sole downstream task. The organization of this review and the chronological progression of representative algorithms are shown in Figs. 1 and 2, respectively.

Existing surveys cover FMCW principles and radar data representations [1], [52], [55], while broader reviews synthesize deep-learning radar perception and 4D radar applications [51], [53], [56]. General automotive perception provides wider context [54], and radar–vision fusion has been reviewed separately [57]. Prior 4D radar surveys focus mainly on fundamentals, representations, datasets, detection, and tracking [1], [51]; recent work on semantic segmentation, occupancy prediction, scene flow, and dynamic reconstruction has not yet been synthesized within a task-oriented framework. The cross-task use of elevation, Doppler, and radar physical priors, together with the empirical support provided by different datasets, also remains fragmented. This review addresses these gaps by organizing the progression from sparse object perception to dense and dynamic spatial understanding.

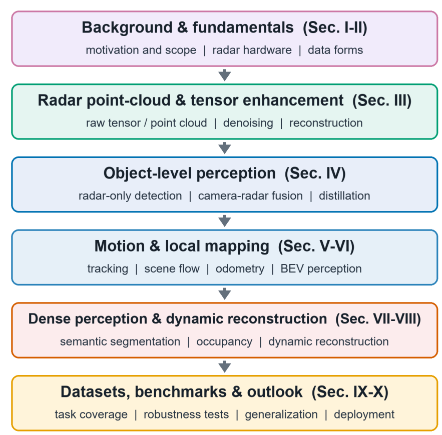

Fig. 1. Organization of this review. The paper introduces 4D radar fundamentals and data forms, then examines radar point-cloud and tensor enhancement, object-level perception, motion perception, local spatial perception, dense spatial perception, and dynamic dense scene understanding. It concludes with datasets, benchmark protocols, and future directions.

Xumin Wu, Jun Zhou, Chen Min, and Yu Hu are with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China.

Jilin Mei is with Beijing Jingwei Hirain Technologies Co., Inc., Beijing, China.

Corresponding authors: Chen Min (mincheng@ict.ac.cn) and Yu Hu (huyu@ict.ac.cn).

This review makes three contributions. First, it develops a task-oriented taxonomy that organizes 4D radar algorithms from point-cloud and tensor enhancement through object-level perception, motion and localization, local and dense spatial perception, and dynamic scene understanding; within these tasks, it compares radar-only learning, multimodal fusion, and cross-modal supervision, including knowledge distillation. Second, it analyzes how elevation, Doppler, RCS, radar measurement geometry, and motion consistency are used as physical priors or explicit constraints across perception tasks, and identifies where these cues remain underused. Third, it maps the task coverage, input data, annotations, and evaluation protocols of existing datasets, and identifies common challenges in data availability, physical fidelity, multimodal robustness, cross-platform generalization, and real-time deployment.

The literature search covered studies available through August 18, 2026. Google Scholar, IEEE Xplore, arXiv, and OpenAlex were searched, with backward reference tracing used to identify additional studies. We included automotive 4D imaging-radar research on data representation, object perception, motion estimation, occupancy prediction, and dynamic reconstruction, together with foundational measurement and dataset papers; non-automotive sensing was excluded. When a preprint and a published article reported the same study, the published version was retained.

II. 4D Radar Signals: Characteristics, Limitations, and Processing

A. Measurement Information and Signal Characteristics

4D automotive radar is commonly described by four measurement dimensions: range, azimuth, elevation, and radial velocity. Its key capability is sufficiently resolved elevation combined with range, azimuth, and velocity, enabling returns to be localized in 3D space while radial motion is measured. Many sensors additionally report amplitude or radar cross section (RCS), whose value varies with target material, geometry, pose, and incidence angle [1]. RCS has also been incorporated into Gaussian radar modeling and scan matching [14].

B. Processing Chain and Algorithmic Inputs

A minimal processing chain is ADC/I/Q samples → range and Doppler

→ CFAR or learned target extraction → point clouds or target lists. Range and Doppler fast Fourier transforms produce range-Doppler maps, while direction-of-arrival estimation or spatial Fourier transforms extend them to range-azimuth or range-azimuth-elevation tensors. CFAR, filtering, clustering, and postprocessing convert these dense responses into sparse point clouds. Downstream networks may further reorganize point clouds or tensors into bird's-eye-view (BEV), voxel, occupancy-grid, graph, or map representations [55].

Raw I/Q retains the most phase and weak-return information. These raw measurements contain the most complete information available to an algorithm, but they also depend strongly on waveform, antenna configuration, sampling strategy, and front-end calibration [55]. Range-Doppler and range-azimuth-elevation tensors preserve structured spectral and spatial information and support early or intermediate learning. Point clouds are compact and compatible with LiDAR-inspired models, but CFAR thresholding removes sub-threshold responses before downstream learning [55]. BEV and voxel features are task-oriented derived representations that facilitate fusion and planning. However, their construction can introduce quantization or completion errors.

Signal processing is mature in conventional automotive radar, whereas learning-based CFAR, angle super-resolution, and interference or clutter suppression remain rapidly developing directions for 4D imaging radar.

C. Algorithm-Relevant Limitations

Several limitations directly shape algorithm design. First, 4D radar returns remain sparse and nonuniform, and elevation accuracy is constrained by angular resolution, aperture, calibration, and sidelobes. Second, Doppler measures only radial velocity; lateral motion requires temporal association, geometry, or additional sensors. Third, multipath, clutter, sidelobes, interference, and target-dependent RCS can generate false, missing, or unstable returns. Radar is often more resilient than cameras or LiDAR in adverse weather [2]. It is nevertheless not immune to precipitation and other noise sources, which ... iability [79].

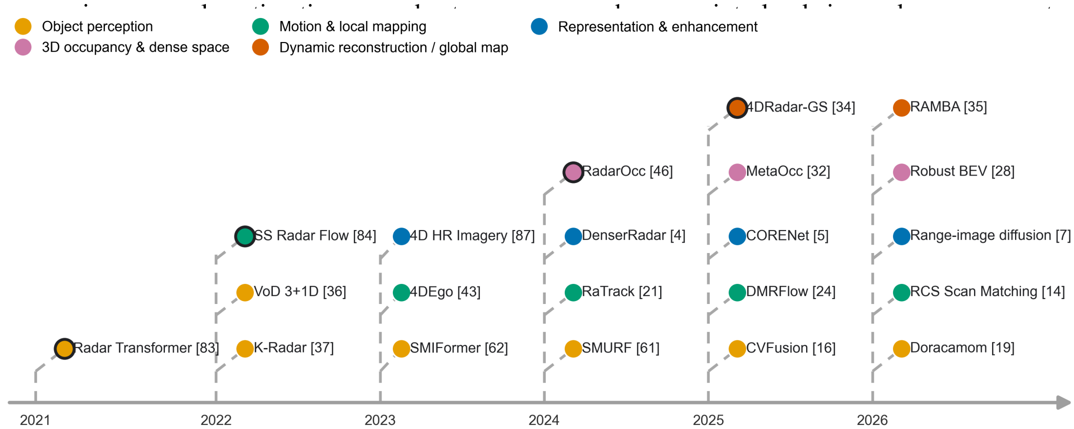

Fig. 2. Timeline of representative automotive 4D imaging radar algorithms (2021–2026); black outlines mark verified category origins.

Finally, radar inputs are not standardized across devices.

Antenna layout, waveform, sampling rate, CFAR settings, coordinate conventions, and proprietary postprocessing alter point density and feature distributions [15], [55]. Algorithms trained on one sensor may therefore learn its processing pipeline rather than general radar structure. Studies should report the radar configuration and front-end processing and evaluate cross-device generalization in addition to downstream accuracy.

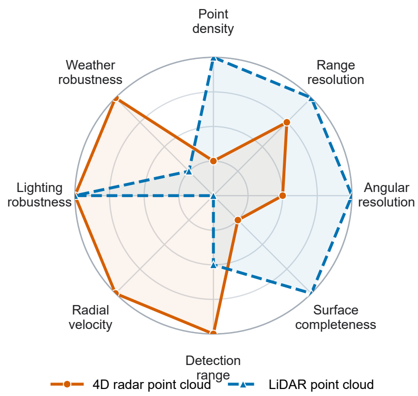

Fig. 3. Qualitative comparison of 4D-radar and LiDAR point clouds. Greater radial extent denotes a stronger characteristic. Sensor-level ratings are adapted from Table I in [51], while point density and surface completeness summarize general point-cloud trends rather than a common benchmark.

III. Radar Point-Cloud and Tensor Enhancement

Radar point-cloud and tensor enhancement is a moderately mature direction. As summarized in Fig. 3, 4D radar point clouds are typically sparser and less geometrically complete than LiDAR point clouds, but they offer longer detection range, direct radial-velocity and RCS measurements, and stronger adverse-weather robustness [1], [51], [55]. This complementarity motivates enhancement that improves geometric usability without discarding radar-specific motion and scattering evidence.

Current studies focus on denoising, densification, super-resolution, reconstruction, and synthetic radar-tensor generation. Their main learning mechanisms include cross-modal supervision, self-supervised representation learning, diffusion or adversarial generation, state-space modeling, and detection-guided reconstruction. Table I organizes representative methods by the position at which they enhance the radar representation.

Cross-modal supervision uses a denser sensor during training to provide geometric targets while retaining radar-only inference. DenserRadar constructs dense 3D occupancy labels from temporally stitched LiDAR point clouds and trains a radar point-cloud detector to recover valid returns [4]. The two-branch CORENet framework is shown in Fig. 4. Its training-only LiDAR branch converts aligned LiDAR points into a voxel-level target mask through fixed-radius KD-tree matching. In the radar branch, voxelized returns pass through the hierarchical multi-scale denoising network, which combines point-topology and sparse-voxel features to predict which radar voxels should be retained. The predicted mask is

TABLE I
Operational Taxonomy of 4D Radar Quality Enhancement

Level	Main operation	Representative methods
Input	Direct point-cloud or radar-tensor reconstruction and synthesis	R2LDM [6]; range-image diffusion [7]; L2RDaS [9]
Feature	Denoising or representation enhancement	CORENet [5]; Bootstrapping Radars [42]; Radar-Mamba [50]
Task	Enhancement optimized jointly with object detection	DenserRadar [4]; Dual-View [8]; SD4R [29]

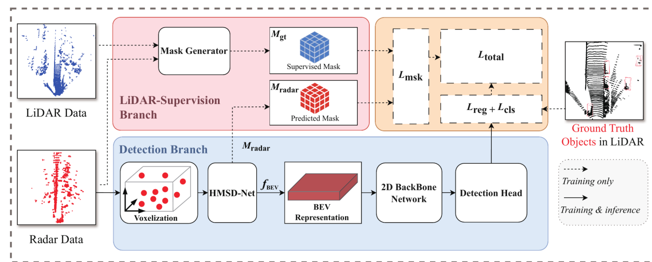

Fig. 4. Overall architecture of CORENet for LiDAR-supervised 4D radar denoising and object detection [5].

supervised by the LiDAR-derived target mask, while the retained radar features are jointly optimized by the classification and regression losses of the detector. Inference therefore uses only the radar branch [5].

Self-supervised representation learning reduces dependence on point-level annotations. For each radar measurement, Bootstrapping Autonomous Driving Radars independently samples two transformations from an augmentation set to generate two radar heatmaps. The transformations include horizontal flipping, rotation and center cropping in radar coordinates, and Radar MIMO Mask. The two heatmaps derived from the same measurement form a positive pair, whereas heatmaps derived from different measurements in the same training mini-batch serve as negatives. Its intra-radar contrastive loss brings the positive radar views together and separates the negative samples. Its radar-to-vision loss further aligns each radar representation with the embedding of its synchronized camera frame and separates mismatched radar–image pairs. Radar MIMO Mask applies antenna dropout and random phase perturbations to MIMO channels to simulate incomplete apertures and Doppler phase distortion. The pretrained radar encoder is then adapted to downstream detection [42].

Generative methods explicitly reconstruct dense point clouds or synthesize radar tensors. R2LDM encodes paired radar and LiDAR point clouds as latent voxel features, conditions reverse diffusion on the radar features, and predicts nonempty voxels and within-voxel point offsets to produce a dense cloud. It reports a six- to ten-fold increase in point count together with gains in registration and detection [6]. Range-image-driven diffusion instead projects radar points into a perspective range image and transfers priors from a pretrained image diffusion model before reconstructing the 3D cloud [7].

L2RDaS uses a conditional generative adversarial network and an object-information supplement module to synthesize spatial radar tensors from LiDAR for dataset expansion [9].

Other methods enhance features or couple reconstruction directly to detection. Radar-Mamba aligns radar and LiDAR occupancies during training, uses state-space scans to capture local and global spatiotemporal features, and fuses Doppler and elevation features for denoising [50]. Dual-View Radar Reconstruction predicts a BEV occupancy map and a perspective depth map from radar measurements, back-projects and fuses the two views under cross-view geometric consistency, and jointly encodes the original and reconstructed points for 3D detection [8].

Existing methods introduce concrete physical constraints primarily at the measurement and geometry levels. CORENet predicts a selection mask over measured radar voxels rather than synthesizing replacement coordinates, so retained points preserve their measured intensity and radial-velocity attributes [5]. Bootstrapping Autonomous Driving Radars uses radar-coordinate transformations and MIMO-channel perturbations that preserve scene geometry while modeling antenna loss and Doppler-phase corruption [42]. R2LDM reconstructs points on a shared spatial voxel grid, supervises nonempty-voxel masks, and bounds each predicted offset within its voxel; Dual-View Radar Reconstruction enforces geometric agreement between the back-projected BEV-occupancy and perspective-depth views [6], [8]. L2RDaS combines aligned real radar tensors, object-centered spatial anchors, and adversarial, feature-matching, and voxel-wise losses to constrain reflection distributions and spatial structure [9].

IV. Object-Level Perception

Object-level perception is one of the most mature and extensively studied 4D radar directions. Research has progressed from early radar object classification [83] to extensive 3D detection, with practical baselines established for both radar-only and fusion-based perception. Current work continues to improve recognition of vulnerable road users under sparse returns [86].

Current object-level studies follow three technical routes: radar-only learning, multimodal fusion, and cross-modal supervision. Radar-only learning uses radar during both training and inference; multimodal fusion uses cameras or LiDAR at both stages; cross-modal supervision uses auxiliary modalities during training but retains radar-only inference.

Radar-only detectors commonly encode point clouds as pillars, voxels, BEV maps, or multiple projected views. RadarNeXt combines a reparameterizable depthwise-convolution backbone with a multi-path deformable foreground-enhancement neck to suppress clutter and strengthen irregular foreground responses [11]. 4DRadDet first forms potential object clusters and then uses their features as queries in cross-attention, allowing clustered returns to guide detection [13]. SRFF fuses high- and low-level features with different sparsity to retain localization detail and semantic context for pedestrians and cyclists [86]. MVFAN, SMURF, and SMIFormer further improve sparse-radar

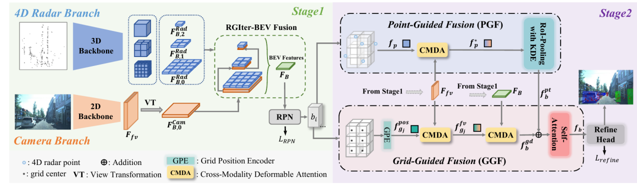

Fig. 5. Overall two-stage architecture of CVFusion for camera–4D-radar 3D object detection [16].

representation through multi-view or multi-representation interaction [59], [61], [62].

Multimodal fusion is currently the most active route in object-level perception, with camera–radar fusion accounting for the largest share of recent studies. According to sensor combination, these methods can be grouped into camera–radar, LiDAR–radar, and camera–LiDAR–radar fusion.

Camera–radar methods mainly use radar geometry to support image-view transformation and image semantics to compensate for sparse radar returns. GRC-Net extracts graph and pillar-based radar features and aligns them with image features in a common geometric space [63]. MSSF alternates point–image interaction and multi-stage sampling, then uses semantic-guided voxel reweighting to reduce feature blurring [60]. LXL predicts image-depth distributions and radar occupancy grids and uses the latter to guide image feature sampling [64]. R4Det combines panoramic depth fusion, pose-independent temporal fusion, and instance-guided refinement to improve depth and small-object detection [17].

Other camera–radar methods focus on region selection and feature alignment. RPGFusion injects radar priors into multimodal feature selection [18]; RCBEVDet++ performs dynamic alignment in BEV, whereas ZFusion explicitly models camera–radar feature interaction [65], [68]. Figure 5 shows CVFusion as a representative two-stage camera–radar detector. In the first stage, a radar-guided iterative BEV fusion module derives occupancy weights from radar features, uses them to progressively refine projected image BEV features, and generates high-recall proposals. In the second stage, point-guided and grid-guided branches aggregate radar-point, image-view, and BEV features within each proposal before final box refinement [16].

LiDAR–radar fusion directly combines complementary point-cloud measurements. M^2-Fusion uses interaction-based multimodal fusion to exchange intermediate radar and LiDAR features and center-based multi-scale fusion to extract object features at several resolutions [58]. L4DR introduces multimodal encoding, foreground-aware denoising, and gated fusion to adapt to unequal sensor degradation under adverse weather [66]. BSM-NET combines multi-bandwidth processing for point-density completion and denoising with multi-scale radar–LiDAR feature fusion [67].

Camera–LiDAR–radar fusion remains less developed in 4D radar research. LRVFNet separately encodes the three modalities, combines them through multi-scale attention, and applies a feature pyramid to multi-scale 2D object detection [90].

Cross-modal supervision transfers information from a denser modality without requiring that modality during deployment. HyperDet uses short-window radar accumulation, cross-sensor validation, and Doppler-guided motion compensation to construct enhanced radar inputs. LiDAR-guided pseudo-radar supervision enriches foreground geometry during training, while the measured radar background and radar-native attributes are retained; inference remains radar-only [12]. SCKD uses a radar–LiDAR fusion teacher and a radar-only student. LiDAR-to-radar and fusion-to-radar feature distillation transfer teacher representations through separate adapters, while semi-supervised output distillation converts confident teacher predictions into pseudo-labels for labeled and unlabeled radar data [20].

Across these routes, radar physics is incorporated through measured attributes and geometric constraints. Elevation is retained through 3D coordinates, voxel features, and multi-view projections, whereas RCS and Doppler are used to distinguish returns and motion; MVFAN explicitly exploits RCS and Doppler, and 4DRadDet uses Doppler-supported clustering to construct object queries [13], [59]. HyperDet further applies Doppler-guided motion compensation and Doppler-consistent object relocation during temporal accumulation and augmentation [12]. In fusion models, LXL uses radar occupancy to guide image-depth sampling, while CVFusion uses radar occupancy probabilities to constrain BEV features and proposals [16], [64].

Table II compares representative methods from the three technical routes. All listed methods address 3D object detection, with SRFF focusing specifically on pedestrians and cyclists.

TABLE II

COMPARISON OF REPRESENTATIVE 4D RADAR METHODS FOR 3D OBJECT DETECTION

Method	Technical route	Core mechanism	Code available
SMIFormer [62] (Sensors, 2023)	Radar-only learning	Multi-view interactive Transformer for spatial feature representation	×
SRFF [86] (Applied Sciences, 2024)	Radar-only learning	Fusion of high- and low-level features with different sparsity	×
RadarNeXt [11] (EURASIP JASP, 2025)	Radar-only learning	Reparameterized backbone and deformable foreground enhancement	✓
4DRadDet [13] (ICRA, 2025)	Radar-only learning	Candidate-object clustering and query-based detection	×
LXL [64] (IEEE T-IV, 2024)	Camera–radar fusion	Radar-occupancy-guided sampling of image-depth features	✓
MSSF [60] (IEEE T-ITS, 2025)	Camera–radar fusion	Point–image interaction and multi-stage sampling	✓
CVFusion [16] (ICCV, 2025)	Camera–radar fusion	Iterative BEV fusion and two-stage proposal refinement	×
M^2-Fusion [58] (IEEE TVT, 2023)	LiDAR–radar fusion	Intermediate feature interaction and multi-scale object fusion	×
L4DR [66] (AAAI, 2025)	LiDAR–radar fusion	Foreground-aware denoising and gated fusion	✓
SCKD [20] (AAAI, 2025)	Cross-modal supervision	Fusion-teacher and semi-supervised knowledge distillation	×

Note: ✓ indicates that a method-specific official implementation or configuration was publicly available and verified on September 9, 2026; × indicates that no such release was found.

V. OBJECT MOTION PERCEPTION

Motion perception comprises subdirections with unequal maturity. Multi-object and extended-target tracking form the most established branch. Moving-instance segmentation, ego-motion estimation, odometry, localization, and SLAM have reached an intermediate stage, whereas scene-flow estimation remains an early but active direction.

Object tracking is studied mainly through radar-only tracking and camera–radar fusion. Radar-only methods follow tracking-by-detection, extended-object filtering, or box-free moving-instance tracking. RaTrack combines motion segmentation, clustering, and motion estimation without category-specific boxes, whereas CenterRadarNet learns 3D detection and re-identification features from 4D radar tensors and uses the resulting embeddings for online association [21], [70]. Comparative studies evaluate tracking-by-detection frameworks, while extended-object filtering jointly estimates motion states and target contours from sparse radar measurements [71], [72]. In camera–radar tracking, CR3DT adds radar spatial and velocity features to a camera BEV detector, while disparity-domain depth recovery uses filtered radar points as metric anchors and performs cross-modal association at both detection and track levels [48], [82].

Moving-instance segmentation and ego-motion estimation are increasingly addressed with radar-only temporal models. RadarMOSEVE uses spatial–temporal attention to jointly segment moving points and estimate ego velocity, while Radar Instance Transformer preserves full-resolution sparse features and assigns class-agnostic identities to moving instances [45], [73]. 4DEgo instead regresses ego velocity from high-resolution radar heatmaps through three-dimensional convolution and attention [43].

Odometry and SLAM rely more strongly on inertial or visual assistance. 4D iRIOM combines scan-wise ego-velocity estimation, scan-to-submap registration, inertial filtering, and loop closure in a radar–inertial SLAM system [75]. DGRO and Go-RIO couple radar velocity with inertial measurements through graph-based preintegration and Gaussian-process continuous-time preintegration, respectively [44], [85]. Gaussian radar–inertial odometry models the scene with three-dimensional Gaussian distributions and evaluates multiple registration hypotheses [22]. RDynaSLAM follows a different route: it projects radar-derived dynamic clusters into camera images and removes the corresponding visual keypoints before pose estimation [23].

Scene flow estimates point-wise three-dimensional motion and is studied through three main routes. Radar-only methods include self-supervised learning for sparse point clouds and decoupled correspondence and refinement in DMRFlow [24], [84]. RadarSFEMOS performs diffusion-based joint flow estimation and motion segmentation; its pipeline is illustrated in Fig. 6 [25]. RadarMP jointly generates radar points and estimates flow from consecutive low-level radar echoes, using radar-derived self-supervision rather than point-wise flow annotations [27]. At inference time, RaLiFlow fuses radar and LiDAR with dynamic-aware bidirectional cross-modal interactions, whereas Hidden Gems uses colocated sensor cues only as cross-modal supervision for radar scene-flow training [26], [74].

The physics-based priors and constraints used by these algorithms are derived from Doppler and RCS observations, rigid-motion consistency, and temporally coupled poses. RadarMOSEVE incorporates measured radial velocity into radar self- and cross-attention, while DGRO estimates ego velocity from Doppler, removes dynamic returns, and weights scan-to-submap registration by RCS [44], [45]. DMRFlow decouples positional and radial-velocity correspondence before refining static and dynamic flow separately; RadarSFEMOS derives motion-segmentation supervision from the discrepancy between robust rigid ego-motion and predicted scene flow; and RadarMP constructs spatial and motion consistency losses from Doppler shifts and echo intensity [24], [25], [27]. For temporal pose estimation, 4D iRIOM combines robust single-scan velocity estimation with inertial filtering and loop closure, whereas Go-RIO uses continuous-time preintegration to accommodate asynchronous radar and IMU measurements [75], [85].

Table III compares representative motion-perception methods across tracking, motion segmentation, ego-motion estimation, odometry, SLAM, and scene flow.

TABLE III
COMPARISON OF REPRESENTATIVE 4D RADAR METHODS FOR MOTION PERCEPTION

Method	Task	Technical route	Core mechanism	Code available
RaTrack [21] (ICRA, 2024)	Moving-object tracking	Radar-only learning	Motion segmentation, clustering, and box-free tracking	✓
CenterRadarNet [70] (ICIP, 2024)	3D detection and tracking	Radar-only learning	Joint detection and re-identification embeddings for association	✓
CR3DT [82] (IROS, 2024)	3D detection and tracking	Camera–radar fusion	Camera BEV detection augmented by radar position and velocity	×
RadarMOSEVE [45] (AAAI, 2024)	Motion segmentation and ego velocity	Radar-only learning	Spatial–temporal Transformer for joint estimation	✓
Radar Instance Transformer [73] (IEEE T-RO, 2024)	Moving-instance segmentation	Radar-only learning	Full-resolution sparse features and class-agnostic instance identities	×
4D iRIOM [75] (IEEE RA-L, 2023)	Odometry and mapping	Radar–inertial fusion	Single-scan ego velocity, submap registration, filtering, and loop closure	×
DGRO [44] (Sensors, 2024)	Radar odometry	Radar–inertial fusion	Doppler–gyroscope preintegration and RCS-weighted registration	×
RDynaSLAM [23] (J. Intell. Robot. Syst., 2025)	Visual SLAM	Camera–radar fusion	Radar-derived dynamic masks for visual-keypoint removal	×
DMRFlow [24] (IEEE TCSVT, 2025)	Scene flow	Radar-only learning	Decoupled position–velocity matching and flow refinement	×
RadarSFEMOS [25] (IEEE RA-L, 2025)	Scene flow and motion segmentation	Radar-only learning	Rigid-motion self-supervision and diffusion-based refinement	✓
RaLiFlow [26] (AAAI, 2026)	Scene flow	LiDAR–radar fusion	Dynamic-aware bidirectional cross-modal interaction	✓
Hidden Gems [74] (CVPR, 2023)	Scene flow	Cross-modal supervision	Multisensor supervision for radar-only scene-flow learning	✓

Note: Code availability was verified against official repositories on September 9, 2026; symbol definitions are given in Table II.

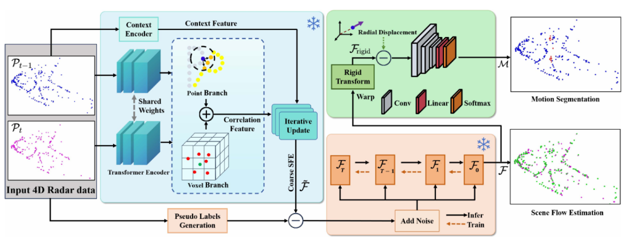

Fig. 6. Self-supervised scene-flow and motion-segmentation framework of RadarSFEMOS. Correlation features are constructed from adjacent 4D radar point clouds; coarse scene flow is first estimated and then refined by a diffusion model, while rigid-motion residuals are used to segment moving objects [25].

VI. LOCAL SPATIAL GEOMETRY AND DEPTH COMPLETION

Radar-guided depth completion and radar-based road-boundary perception remain early-stage 4D radar directions. Current work follows two main routes: camera–radar methods infer dense image-aligned metric depth from sparse radar ranges, whereas radar-only methods estimate local road-boundary geometry from temporal point clouds. Radar already measures range at observed returns; depth completion therefore fills unobserved image regions rather than re-estimating the measured radar range.

For depth completion, Semantic-Guided Depth Completion uses image semantics to diffuse sparse radar depth within category-specific regions. Its gradient-guided smooth-edge loss constrains consistency inside semantic regions and preserves depth discontinuities at their boundaries [69]. SA-RCD uses RGB structure to define targeted support regions for sparse radar returns, enhances the radar depth representation, and fuses it with monocular-depth features through a multi-scale structure-guided network [30]. In both methods, measured radar range establishes metric scale, while image structure supplies dense spatial support.

Radar-only local geometry is represented by 4DRadarRBD. It first removes unlikely returns using physical road-boundary constraints and then penalizes predicted points according to their distance from the annotated boundary. For temporal consistency, each radar point is augmented with its deviation from the motion-compensated boundary estimate of the previous frame before point-wise boundary segmentation [49].

VII. DENSE SPATIAL PERCEPTION

Within dense spatial perception, point-wise semantic segmentation remains an early-stage direction, whereas grid- and voxel-based occupancy prediction has developed rapidly into a major 4D radar task. Point-wise segmentation assigns semantic labels only to measured radar returns, while occupancy

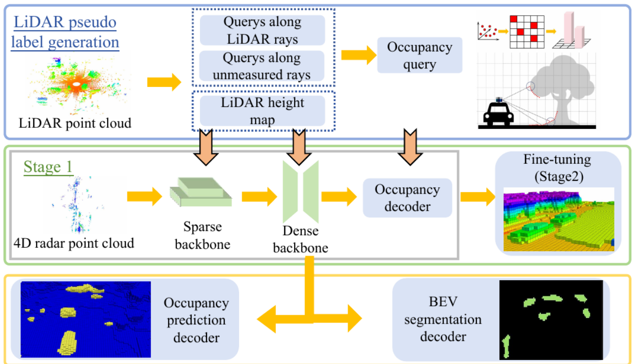

Fig. 7. Radar occupancy learning framework of 4D-ROLLS. LiDAR rays and height maps generate occupancy pseudo-labels; a sparse-to-dense backbone trains a radar-only occupancy network that jointly outputs occupancy prediction and BEV segmentation [10].

prediction estimates free, occupied, and semantic states on a regular 3D grid, including locations without a return.

Existing point-wise segmentation studies are dominated by cross-modal knowledge distillation. RaSS introduces ZJUSSet with point-level radar and LiDAR labels and trains a radar-only student from a LiDAR segmentation teacher. Adaptive Doppler compensation aligns motion cues before segmentation [31]. MKFusion likewise transfers knowledge from a multimodal teacher to improve radar point-cloud segmentation [81].

Occupancy methods differ mainly by sensor availability during training and inference. RadarOcc uses radar at both stages. It directly processes 4D radar tensors using Doppler-bin descriptors, sidelobe-aware spatial sparsification, and range-wise self-attention; spherical encoding preserves the measurement geometry before Cartesian voxel aggregation [46]. 4D-ROLLS uses LiDAR only during training and remains radar-only at inference; Fig. 7 summarizes its supervision and occupancy network [10].

Camera–radar occupancy methods differ in how radar evidence enters the 3D representation. 4DRC-OCC fuses both modalities in voxel space, combining radar range and angle cues with image semantics [33]. MetaOcc uses radar-height self-attention and hierarchical local–global fusion across modalities and time; its pseudo-label pipeline is a separate semi-supervised strategy for reducing annotation requirements [32]. Doracamom combines radar geometry and image semantics to initialize voxel queries, then refines temporal representations in BEV and voxel spaces for joint detection and semantic occupancy [19]. 4DR360 instead treats occupancy as a persistent scene state and updates it through state-guided BEV enhancement and Doppler-guided temporal fusion [88].

LiDAR-derived pseudo-labels in 4D-ROLLS and image semantics in camera–radar models are auxiliary spatial priors rather than radar physics. Measured Doppler, elevation, and spherical-coordinate structure instead provide radar-derived motion or geometric priors, while range and angle supply geometric anchors in voxel fusion. Encoding these quantities as features does not by itself constitute a hard physical constraint unless measurement consistency is explicitly enforced. Because the absence of a radar return does not imply free space, dense outputs should retain a distinction among directly observed, model-inferred, and unknown regions.

Table IV compares representative methods for point-wise semantic segmentation and dense occupancy prediction.

TABLE IV
COMPARISON OF REPRESENTATIVE 4D RADAR METHODS FOR DENSE SPATIAL PERCEPTION

Method	Task	Technical route	Core mechanism	Code available
RaSS [31] (Sensors, 2025)	Point-wise semantic segmentation	Cross-modal supervision	Knowledge distillation from a LiDAR teacher to a radar student	×
MKFusion [81] (KBS, 2026)	Point-wise semantic segmentation	Cross-modal supervision	Semantic knowledge transfer from a multimodal teacher	×
RadarOcc [46] (NeurIPS, 2024)	3D occupancy prediction	Radar-only learning	Doppler descriptors, sidelobe-aware sparsification, and spherical encoding	✓
4D-ROLLS [10] (IROS, 2025)	3D occupancy and BEV segmentation	Cross-modal supervision	LiDAR-ray pseudo-labels for a radar-only occupancy network	×
Doracamom [19] (IEEE TCSVT, 2026)	Detection and semantic occupancy	Camera–radar fusion	Voxel-query initialization and temporal BEV–voxel fusion	✓

Note: Code availability was verified against official repositories on September 9, 2026; symbol definitions are given in Table II.

VIII. DYNAMIC DENSE SCENE UNDERSTANDING

Dynamic dense scene understanding is an exploratory direction that aims to represent both spatial structure and temporal evolution. Existing studies provide partial solutions at three levels—point-wise motion, temporally updated voxel states, and cross-frame scene reconstruction. However, a unified representation connecting these outputs has not yet emerged. Figure 8 summarizes these three conceptual levels.

At the point level, scene-flow methods recover the 3D motion of sparse radar returns between adjacent frames. These estimates provide motion cues for dense temporal reasoning, but they do not themselves encode free, occupied, dynamic, and unknown regions in a common spatial grid.

At the voxel level, existing 4D-radar occupancy methods mainly estimate geometric or semantic states, even when temporal fusion is used, and do not assign explicit motion to each voxel. Among them, 4DR360 is closest to persistent scene modeling because it maintains occupancy as a state across frames; its outputs, however, remain object detection and semantic occupancy rather than a velocity-aware dynamic occupancy field [88].

At the reconstruction level, 4DRadar-GS and RF4D represent two distinct technical routes. 4DRadar-GS adopts camera–radar fusion: radar spatial and velocity measurements guide dynamic-object segmentation, monocular-scale recovery, and Gaussian initialization, while a velocity-guided PointTrack model trained with scene-flow supervision maintains dynamic trajectories across frames [34]. RF4D follows a radar-only route. Its time-conditioned neural field predicts scene-flow offsets to enforce temporal occupancy coherence and uses a

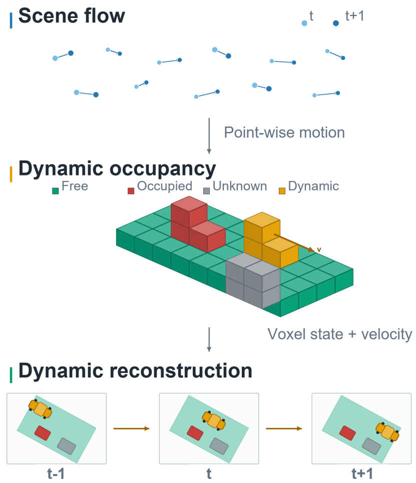

Fig. 8. Conceptual levels of dynamic dense scene understanding. Scene flow estimates sparse point-wise 3D motion, dynamic occupancy would associate voxel states with motion, and dynamic reconstruction maintains cross-frame trajectories and scene geometry.

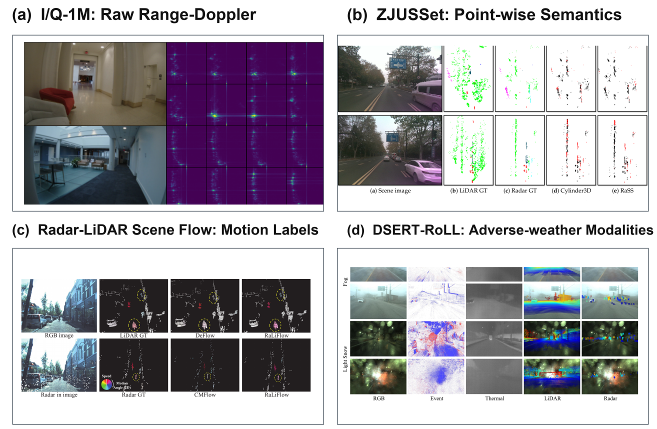

Fig. 9. Representative radar data forms and annotation strategies. Panels (a) and (b) show camera and range-Doppler frames from I/Q-1M and point-wise semantic labels from ZJUSSet, respectively [3], [31]. Panels (c) and (d) show VoD-derived radar-LiDAR scene-flow labels and synchronized adverse-condition modalities from DSERT-RoLL, respectively [26], [40]. Together, the panels contrast direct measurements with point-wise, generated-motion, and multimodal annotation strategies.

radar-specific power-rendering formulation derived from the measurement process [89].

The two reconstruction routes also differ in output space. 4DRadar-GS produces image-oriented dynamic Gaussian representations, whereas RF4D synthesizes radar observations and occupancy in a continuous field; their evaluation targets are therefore not directly comparable. Cross-modal supervision and teacher–student distillation have not yet become established routes for this task.

This completes our review of the principal 4D radar perception directions and their representative methods.

IX. Datasets and Benchmark Protocols

Early 4D radar datasets concentrated on object-level perception. VoD and TJ4DRadSet established point-cloud benchmarks for 3D detection, while K-Radar added radar tensors and adverse-weather sequences. RaDelft subsequently provided lower-level radar data for automotive radar detection [80].

Task coverage later expanded beyond detection. ZJUSSet provides point-wise semantic labels, and the Radar-LiDAR Scene Flow Dataset repurposes synchronized VoD sequences with generated motion labels. DIDLM and NTU4DRadLM target localization, odometry, and mapping, whereas OmniHD-Scenes adds semantic occupancy to a full-surround radar-camera-LiDAR configuration. Dual Radar and V2X-Radar broaden multi-radar and cooperative perception, while DSERT-RoLL provides synchronized modalities, track identities, and ego odometry under diverse conditions. Table V compares the official data forms and task annotations of the representative datasets retained for this analysis.

Official task coverage does not exhaust later use. RaLiFlow repurposes synchronized VoD radar/LiDAR sequences by pairing adjacent frames and generating scene-flow labels during preprocessing. Its radar-denoising and label-generation procedure provides more reliable point-wise flow supervision for radar returns, especially those outside annotated object boundaries, and supports joint radar-LiDAR scene-flow learning [26]. 4DRadar-GS uses synchronized radar-camera sequences from OmniHD-Scenes for dynamic reconstruction. It uses radar velocity and spatial information to segment dynamic objects and recover the monocular depth scale for Gaussian initialization, then jointly trains a velocity-guided point-tracking module under scene-flow supervision to maintain temporally consistent dynamic representations [34]. These downstream protocols are therefore not assigned to the original VoD and OmniHD-Scenes rows in Table V.

Existing benchmark protocols remain task-specific. Detection datasets report class-wise precision-recall and average precision, whereas semantic segmentation and occupancy use IoU-based metrics. Scene flow is evaluated by point-wise motion error, while localization and SLAM use trajectory and pose errors. These scores describe different outputs and cannot be compared as a single ranking of datasets or models.

Cross-dataset transfer is further complicated by front-end processing, point-cloud generation, sensor configuration, coordinate range, and annotation policy. Figure 9 visually illustrates differences in data representations and annotation strategies. As shown in Table V, many road-driving datasets provide object annotations and postprocessed point clouds, whereas few combine measurement-level radar with dense spatial and motion labels. No single road-driving resource in the table jointly covers raw measurements, processed tensors, point clouds, object labels, dense occupancy, motion, full calibration metadata, and diverse radar hardware. This fragmentation makes it difficult to separate transferable radar representations from models fitted to a particular sensor and dataset protocol.

X. Challenges and Future Directions

First, measurement-level automotive radar data remain scarce because many road-driving datasets release processed point clouds rather than measurement-level I/Q data. Future

TABLE V
OFFICIAL DATA FORMS AND TASK COVERAGE OF REPRESENTATIVE 4D RADAR DATASETS

Dataset	Radar form and synchronized modalities or labels	Object-level perception	Motion, localization, and mapping	Dense spatial perception	Dynamic reconstruction
VoD [36]	4D radar point clouds; LiDAR and stereo cameras	3D detection	–	–	–
K-Radar [37]	4D radar tensors; LiDAR, stereo cameras, and RTK-GPS	3D detection	–	–	–
TJ4DRadSet [38]	4D radar point clouds; synchronized camera	3D detection	Object tracking	–	–
Dual Radar [77]	Dual 4D radar point clouds; LiDAR and cameras	3D detection	Object tracking	–	–
ZJUSSet [31]	4D radar and LiDAR point clouds; cameras and point-wise semantic labels	–	–	Point-wise semantic segmentation	–
OmniHD-Scenes [41]	Six radars, six cameras, and LiDAR; dense occupancy ground truth	3D detection	–	Semantic segmentation and occupancy	–
DIDLM [39]	Radar/LiDAR, RGB/infrared/depth, and IMU/GPS	–	Localization, odometry, SLAM, mapping, and loop closure	–	–
NTU4DRadLM [76]	Radar/LiDAR, RGB/thermal, and IMU/RTK	–	Localization, odometry, mapping, and loop closure	–	–
V2X-Radar [78]	Vehicle-side/roadside radars, LiDAR, and multi-view cameras	Cooperative, single-vehicle, and roadside 3D detection	–	–	–
DSERT-RoLL [40]	Radar, event/RGB/thermal cameras, dual LiDAR, and odometry ground truth	2D/3D detection	Tracking and ego odometry	–	–
I/Q-1M† [3]	Complex range-Doppler-azimuth-elevation cubes; LiDAR and camera supervision	–	Ego-motion estimation	BEV/3D occupancy and semantic segmentation	–

Note: † denotes a cross-domain raw-radar resource rather than a conventional road-vehicle dataset.

datasets should jointly release raw I/Q, radar tensors, point clouds, BEV products, and synchronized auxiliary sensors together with the calibration, synchronization, waveform, array, CFAR, and point-cloud-generation parameters needed to reproduce front-end processing.

Second, cross-modal supervision should not be limited to LiDAR-like geometric completion. Although it can improve geometric completeness and dense prediction, treating unmatched radar returns indiscriminately as noise may suppress Doppler, RCS, multipath, and material-dependent cues. Future models should incorporate radar measurement equations, reflectivity models, Doppler consistency, multipath uncertainty, and sensor-noise statistics so that enhanced outputs remain compatible with the originating measurements.

Third, Doppler information remains underused. Existing studies exploit it in temporal detection, moving-object tracking, scene-flow estimation, and radar odometry. However, many methods still encode radial velocity only as an input feature, and current occupancy models generally do not explicitly estimate a velocity state for each voxel. Future research should make fuller use of Doppler information across detection, occupancy, motion, and mapping.

Fourth, multimodal fusion must progress from nominal accuracy gains to explicit robustness mechanisms. Real systems experience latency, extrinsic drift, sensor contamination, glare, precipitation, fog, and radar multipath. Radar-LiDAR calibration studies show that synchronization and extrinsic errors directly disrupt sparse cross-modal correspondence [47]. Fog-aware radar-camera fusion has begun to address condition-dependent visual degradation [28]. Future fusion models should adaptively weight modalities according to scene conditions and estimated reliability, with radar range and velocity contributing more strongly when visual reliability decreases or motion cues are critical.

Fifth, benchmark evaluation must become measurement-aware and task-specific. Generative enhancement, radar-data synthesis, and occupancy completion may infer spatial structure not directly supported by measured returns. Evaluation should therefore distinguish measured, inferred, free, occupied, and unknown regions and quantify velocity consistency, RCS preservation, uncertainty calibration, and planning impact. Current detection datasets commonly stratify results by class or range and, in K-Radar, by weather; future protocols should additionally stratify results by target velocity. Current scene-flow studies report point-wise flow errors; future motion evaluation should additionally report trajectory continuity and velocity consistency. SLAM and reconstruction should report pose error, map consistency, and dynamic-object handling.

Sixth, models must generalize across radar hardware, configurations, and vehicle platforms. Radar domain shift is strongly coupled to hardware because frequency band, antenna array, waveform, resolution, CFAR, and vendor postprocessing all alter the input. Sensor-parameter-conditioned encoding and multi-device pretraining are promising routes for exploiting raw and multi-sensor data. Evaluation should use cross-radar, cross-platform, and cross-city splits.

Seventh, real-time deployment and safety validation are essential. The engineering value of 4D radar lies in cost, environmental robustness, and manufacturability. Because diffusion, 3D voxel processing, large attention networks, and dynamic reconstruction add iterative or high-dimensional computation, their deployment cost should be reported explicitly. Future work should jointly report accuracy, latency, energy consumption, model size, calibration burden, failure detection, uncertainty, and planning impact on representative vehicle hardware.

From a system perspective, evaluation should extend beyond point density and aggregate mAP toward planning-relevant outputs. A useful 4D-radar interface should jointly represent objects; free, occupied, and unknown space; dynamic velocity; and uncertainty, while enforcing consistency among detection, occupancy, and motion estimates.

XI. Conclusion

4D millimeter-wave radar perception is progressing from sparse object detection toward spatial and motion understanding. This review systematically organizes the field around radar point-cloud and tensor enhancement, object-level perception, motion and localization, local and dense spatial perception, dynamic scene reconstruction, datasets, and benchmark protocols. Object detection is relatively mature, occupancy prediction is developing rapidly, whereas point-wise semantic understanding and dynamic scene reconstruction remain exploratory. Radar-only learning, multimodal fusion, and cross-modal supervision are suited to different tasks and deployment conditions, yet radar-specific physical information—including elevation, Doppler, RCS, measurement geometry, and motion consistency—has not been fully or consistently exploited. Future research should expand measurement-level datasets, make fuller use of radar-specific physical information, and improve multimodal robustness, cross-device generalization, uncertainty evaluation, and efficient vehicle deployment.

References

[1] Zeyu Han, Jiahao Wang, Zikun Xu, et al., “4D Millimeter-Wave Radar in Autonomous Driving: A Survey,” arXiv preprint arXiv:2306.04242, 2023.

[2] Xiangyuan Peng, Miao Tang, Huawei Sun, et al., “4D mmWave Radar for Sensing Enhancement in Adverse Environments: Advances and Challenges,” in 2025 IEEE 28th International Conference on Intelligent Transportation Systems (ITSC), 2025, pp. 2884-2891. doi: 10.1109/itsc60802.2025.11423145.

[3] Tianshu Huang, Akarsh Prabhakara, Chuhan Chen, et al., “Towards Foundational Models for Single-Chip Radar,” in 2025 IEEE/CVF International Conference on Computer Vision (ICCV), 2025, pp. 24655-24665. doi: 10.1109/ICCV51701.2025.02286.

[4] Zeyu Han, Junkai Jiang, Xiaokang Ding, et al., “DenserRadar: A 4D Millimeter-Wave Radar Point Cloud Detector Based on Dense LiDAR Point Clouds,” in 2024 IEEE 27th International Conference on Intelligent Transportation Systems (ITSC), 2024, pp. 930-936. doi: 10.1109/itsc58415.2024.10919654.

[5] Fuyang Liu, Jilin Mei, Fangyuan Mao, et al., “CORENet: Cross-Modal 4D Radar Denoising Network with LiDAR Supervision for Autonomous Driving,” in 2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2025, pp. 3084-3091. doi: 10.1109/IROS60139.2025.11246433.

[6] Boyuan Zheng, Shouyi Lu, Renbo Huang, et al., “R2LDM: An Efficient 4D Radar Super-Resolution Framework Leveraging Diffusion Model,” in 2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2025, pp. 769-776. doi: 10.1109/IROS60139.2025.11246739.

[7] Ruixin Wu, Zihan Li, Jin Wang, et al., “Diffusion-Based mmWave Radar Point Cloud Enhancement Driven by Range Images,” IEEE Robotics and Automation Letters, vol. 11, no. 5, pp. 6082-6089, 2026. doi: 10.1109/lra.2026.3673977.

[8] Dong Wang, JunSheng Huang, Zhe Zhang, “Dual-View Radar Reconstruction for Enhanced 3D Object Detection With 4D Imaging Radar Point Clouds,” Electronics Letters, vol. 62, no. 1, Art. no. e70568, 2026. doi: 10.1049/ell2.70568.

[9] Woo-Jin Jung, Dong-Hee Paek, Seung-Hyun Kong, “L2RDaS: Synthesizing 4D Radar Tensors for Model Generalization via Dataset Expansion,” arXiv preprint arXiv:2503.03637, 2025.

[10] Ruihan Liu, Xiaoyi Wu, Xijun Chen, et al., “4D-ROLLS: 4D Radar Occupancy Learning via LiDAR Supervision,” in 2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2025, pp. 13602-13609. doi: 10.1109/IROS60139.2025.11246471.

[11] Liye Jia, Runwei Guan, Haocheng Zhao, et al., “RadarNeXt: Lightweight and Real-Time 3D Object Detector Based on 4D mmWave Imaging Radar,” EURASIP Journal on Advances in Signal Processing, vol. 2025, no. 1, Art. no. 65, 2025. doi: 10.1186/s13634-025-01271-2.

[12] Yichun Xiao, Runwei Guan, Jin Jin, et al., “HyperDet: 3D Object Detection with Hyper 4D Radar Point Clouds,” arXiv preprint arXiv:2602.11554, 2026. [Online]. Available: `https://arxiv.org/abs/2602.11554`

[13] Caien Weng, Xin Bi, Panpan Tong, et al., “4DRadDet: Cluster-Queried Enhanced 3D Object Detection with 4D Radar,” in 2025 IEEE International Conference on Robotics and Automation (ICRA), 2025, pp. 16984-16990. doi: 10.1109/icra55743.2025.11127889.

[14] Fernando Amodeo, Luis Merino, Fernando Caballero, “4D Radar Gaussian Modeling and Scan Matching with RCS,” arXiv preprint arXiv:2604.14868, 2026. [Online]. Available: `https://arxiv.org/abs/2604.14868`

[15] Yongjun Cai, Jie Bai, Hui-Liang Shen, et al., “Development of Low-Cost Single-Chip Automotive 4D Millimeter-Wave Radar,” Sensors, vol. 25, no. 15, pp. 4640, 2025. doi: 10.3390/s25154640.

[16] Hanzhi Zhong, Zhiyu Xiang, Ruoyu Xu, et al., “CVFusion: Cross-View Fusion of 4D Radar and Camera for 3D Object Detection,” in 2025 IEEE/CVF International Conference on Computer Vision (ICCV), 2025, pp. 28188-28197. doi: 10.1109/iccv51701.2025.02617.

[17] Zhongyu Xia, Yousen Tang, Yongtao Wang, et al., “R4Det: 4D Radar-Camera Fusion for High-Performance 3D Object Detection,” arXiv preprint arXiv:2603.11566, 2026. [Online]. Available: `https://arxiv.org/abs/2603.11566`

[18] Xin Qiu, Wenjie Liu, “RPGFusion: 4D Radar Prior-Guided Multi-Modal Fusion for 3D Detection,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2026, pp. 284-294.

[19] Lianqing Zheng, Jianan Liu, Runwei Guan, et al., “Doracamom: Joint 3D Detection and Occupancy Prediction With Multi-View 4D Radars and Cameras for Omnidirectional Perception,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 36, no. 6, pp. 7630-7645, 2026. doi: 10.1109/tcsvt.2026.3657111.

[20] Ruoyu Xu, Zhiyu Xiang, Chenwei Zhang, et al., “SCKD: Semi-Supervised Cross-Modality Knowledge Distillation for 4D Radar Object Detection,” in Proceedings of the AAAI Conference on Artificial Intelligence, 2025, pp. 8933-8941. doi: 10.1609/aaai.v39i9.32966.

[21] Zhijun Pan, Fangqiang Ding, Hantao Zhong, et al., “RaTrack: Moving Object Detection and Tracking with 4D Radar Point Cloud,” in 2024 IEEE International Conference on Robotics and Automation (ICRA), 2024, pp. 4480-4487. doi: 10.1109/icra57147.2024.10610368.

[22] Fernando Amodeo, Luis Merino, Fernando Caballero, “4D Radar-Inertial Odometry Based on Gaussian Modeling and Multi-Hypothesis Scan Matching,” IEEE Robotics and Automation Letters, vol. 11, no. 5, pp. 5773-5780, 2026. doi: 10.1109/lra.2026.3675514.

[23] Dongying Zhu, Guanci Yang, “RDynaSLAM: Fusing 4D Radar Point Clouds to Visual SLAM in Dynamic Environments,” Journal of Intelligent & Robotic Systems, vol. 111, no. 1, pp. 11, 2025. doi: 10.1007/s10846-024-02204-1.

[24] Mingliang Zhai, Bing-Kun Bao, Xuezhi Xiang, “DMRFlow: 4D Radar Scene Flow Estimation With Decoupled Matching and Refinement,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 35, no. 8, pp. 7395-7408, 2025. doi: 10.1109/tcsvt.2025.3546274.

[25] Yufei Liu, Xieyuanli Chen, Neng Wang, et al., “Self-Supervised Diffusion-Based Scene Flow Estimation and Motion Segmentation With 4D Radar,” IEEE Robotics and Automation Letters, vol. 10, no. 6, pp. 5895-5902, 2025. doi: 10.1109/lra.2025.3563829.

[26] Jingyun Fu, Zhiyu Xiang, Na Zhao, “RaLiFlow: Scene Flow Estimation with 4D Radar and LiDAR Point Clouds,” in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 40, no. 5, 2026, pp. 4012-4021. doi: 10.1609/aaai.v40i5.37404.

[27] Ruiqi Cheng, Huijun Di, Jian Li, et al., “RadarMP: Motion Perception for 4D mmWave Radar in Autonomous Driving,” in Proceedings of the AAAI Conference on Artificial Intelligence, 2026, pp. 3282-3290. doi: 10.1609/aaai.v40i5.37323.

[28] Zhengqing Li, Baljit Singh, “Robust BEV Perception via Dual 4D Radar–Camera Fusion Under Adverse Conditions with Fog-Aware Enhancement,” Electronics, vol. 15, no. 6, pp. 1284, 2026. doi: 10.3390/electronics15061284.

[29] Xiaokai Bai, Jiahao Cheng, Songkai Wang, et al., “SD4R: Sparse-to-Dense Learning for 3D Object Detection with 4D Radar,” in 2025 IEEE 28th International Conference on Intelligent Transportation Systems (ITSC), 2025, pp. 4362-4368. doi: 10.1109/ITSC60802.2025.11423715.

[30] Fuyi Zhang, Zhu Yu, Chunhao Li, et al., “Structure-Aware Radar-Camera Depth Estimation,” in 2025 IEEE International Conference on Robotics and Automation (ICRA), 2025, pp. 13028-13035. doi: 10.1109/icra55743.2025.11128760.

[31] Chenwei Zhang, Zhiyu Xiang, Ruoyu Xu, et al., “RaSS: 4D mm-Wave Radar Point Cloud Semantic Segmentation with Cross-Modal Knowledge Distillation,” Sensors, vol. 25, no. 17, pp. 5345, 2025. doi: 10.3390/s25175345.

[32] Long Yang, Lianqing Zheng, Wenjin Ai, et al., "MetaOcc: Spatio-Temporal Fusion of Surround-View 4D Radar and Camera for 3D Occupancy Prediction with Dual Training Strategies," arXiv preprint arXiv:2501.15384, 2025.

[33] David Ninfa, Andras Palffy, Holger Caesar, "4DRC-OCC: Robust Semantic Occupancy Prediction Through Fusion of 4D Radar and Camera," arXiv preprint arXiv:2603.07794, 2026.

[34] Xiao Tang, Guirong Zhuo, Cong Wang, et al., "4DRadar-GS: Self-Supervised Dynamic Driving Scene Reconstruction with 4D Radar," arXiv preprint arXiv:2509.12931, 2025.

[35] Jianzhu Huai, Yiwen Chen, Binliang Wang, "RAMBA: 4D Radar Mapping by Bundle Adjustment," arXiv preprint arXiv:2605.25041, 2026. [Online]. Available: https://arxiv.org/abs/2605.25041

[36] Andras Palffy, Ewoud Pool, Srimannarayana Baratam, et al., "Multi-Class Road User Detection With 3+1D Radar in the View-of-Delft Dataset," IEEE Robotics and Automation Letters, vol. 7, no. 2, pp. 4961-4968, 2022. doi: 10.1109/LRA.2022.3147324.

[37] Dong-Hee Paek, Seung-Hyun Kong, Kevin Tirta Wijaya, "K-Radar: 4D Radar Object Detection for Autonomous Driving in Various Weather Conditions," in Advances in Neural Information Processing Systems 35, 2022, pp. 3819-3829. doi: 10.52202/068431-0276.

[38] Lianqing Zheng, Zhixiong Ma, Xichan Zhu, et al., "TJ4DRadSet: A 4D Radar Dataset for Autonomous Driving," in 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC), 2022, pp. 493-498. doi: 10.1109/itsc55140.2022.9922539.

[39] Weisheng Gong, Chen He, Kaijie Su, et al., "DIDLM: A SLAM Dataset for Difficult Scenarios Featuring Infrared, Depth Cameras, LiDAR, 4D Radar, and Others Under Adverse Weather, Low Light Conditions, and Rough Roads," IEEE Transactions on Intelligent Transportation Systems, vol. 27, no. 2, pp. 1908-1920, 2026. doi: 10.1109/TITS.2025.3632411.

[40] Hoonhee Cho, J.J. Kang, Yuhwan Jeong, et al., "DSERT-RoLL: Robust Multi-Modal Perception for Diverse Driving Conditions with Stereo Event-RGB-Thermal Cameras, 4D Radar, and Dual-LiDAR," arXiv preprint arXiv:2604.03685, 2026. [Online]. Available: http://arxiv.org/abs/2604.03685

[41] Lianqing Zheng, Long Yang, Qunshu Lin, et al., "OmniHD-Scenes: A Next-Generation Multimodal Dataset for Autonomous Driving," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 48, no. 6, pp. 7032-7049, 2026. doi: 10.1109/TPAMI.2026.3663672.

[42] Yiduo Hao, Sohrab Madani, Junfeng Guan, et al., "Bootstrapping Autonomous Driving Radars with Self-Supervised Learning," in 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024, pp. 15012-15023. doi: 10.1109/cvpr52733.2024.01422.

[43] Prashant Kumar Rai, Nataliya Strokina, Reza Ghabcheloo, "4DEgo: ego-velocity estimation from high-resolution radar data," Frontiers in Signal Processing, vol. 3, pp. 1198205, 2023. doi: 10.3389/frsip.2023.1198205.

[44] Chao Guo, Bangguo Wei, Bin Lan, et al., "DGRO: Doppler Velocity and Gyroscope-Aided Radar Odometry," Sensors, vol. 24, no. 20, pp. 6559, 2024. doi: 10.3390/s24206559.

[45] Changsong Pang, Xieyuanli Chen, Yimin Liu, et al., "RadarMOSEVE: A Spatial-Temporal Transformer Network for Radar-Only Moving Object Segmentation and Ego-Velocity Estimation," in Proceedings of the AAAI Conference on Artificial Intelligence, 2024, pp. 4424-4432. doi: 10.1609/aaai.v38i5.28240.

[46] Fangqiang Ding, Xiangyu Wen, Yunzhou Zhu, et al., "RadarOcc: Robust 3D Occupancy Prediction with 4D Imaging Radar," in Advances in Neural Information Processing Systems 37, 2024, pp. 101589-101617. doi: 10.52202/079017-3222.

[47] Chia-Le Lee, Chun-Yu Hou, Chieh-Chih Wang, et al., "Extrinsic and Temporal Calibration of Automotive Radar and 3-D LiDAR in Factory and On-Road Calibration Settings," IEEE Open Journal of Intelligent Transportation Systems, vol. 4, pp. 708-719, 2023. doi: 10.1109/ojits.2023.3312660.

[48] Yunfei Xie, Xiaohui Li, Dingheng Wang, et al., "Robust 3D Multi-Object Tracking via 4D mmWave Radar-Camera Fusion and Disparity-Domain Depth Recovery," Sensors, vol. 26, no. 7, pp. 2096, 2026. doi: 10.3390/s26072096.

[49] Yuyan Wu, Hae Young Noh, "4DRadarRBD: 4D mmWave radar-based road boundary detection in autonomous driving," Frontiers in Signal Processing, vol. 5, pp. 1667789, 2025. doi: 10.3389/frsip.2025.1667789.

[50] Hong Gao, Xiangkai Xu, Tianqi Zhu, et al., "Radar-Mamba: 4D Millimeter-Wave Point Cloud Enhancement via State Space Models," in Proceedings of the 33rd ACM International Conference on Multimedia (MM '25), 2025, pp. 1774-1782. doi: 10.1145/3746027.3755431.

[51] Lili Fan, Junhao Wang, Yuanmeng Chang, et al., "4D mmWave Radar for Autonomous Driving Perception: A Comprehensive Survey," IEEE Transactions on Intelligent Vehicles, vol. 9, no. 4, pp. 4606-4620, 2024. doi: 10.1109/TIV.2024.3380244.

[52] Arthur Venon, Yohan Dupuis, Pascal Vasseur, et al., "Millimeter Wave FMCW RADARs for Perception, Recognition and Localization in Automotive Applications: A Survey," IEEE Transactions on Intelligent Vehicles, vol. 7, no. 3, pp. 533-555, 2022. doi: 10.1109/TIV.2022.3167733.

[53] Arvind Srivastav, Soumyajit Mandal, "Radars for Autonomous Driving: A Review of Deep Learning Methods and Challenges," IEEE Access, vol. 11, pp. 97147-97168, 2023. doi: 10.1109/ACCESS.2023.3312382.

[54] Ashish Pandharipande, Chih-Hong Cheng, Justin Dauwels, et al., "Sensing and Machine Learning for Automotive Perception: A Review," IEEE Sensors Journal, vol. 23, no. 11, pp. 11097-11115, 2023. doi: 10.1109/JSEN.2023.3262134.

[55] Shanliang Yao, Runwei Guan, Zitian Peng, et al., "Exploring Radar Data Representations in Autonomous Driving: A Comprehensive Review," IEEE Transactions on Intelligent Transportation Systems, vol. 26, no. 6, pp. 7401-7425, 2025. doi: 10.1109/TITS.2025.3554781.

[56] Yi Zhou, Lulu Liu, Haocheng Zhao, et al., "Towards Deep Radar Perception for Autonomous Driving: Datasets, Methods, and Challenges," Sensors, vol. 22, no. 11, Art. no. 4208, 2022. doi: 10.3390/s22114208.

[57] Zhiqing Wei, Fengkai Zhang, Shuo Chang, et al., "MmWave Radar and Vision Fusion for Object Detection in Autonomous Driving: A Review," Sensors, vol. 22, no. 7, Art. no. 2542, 2022. doi: 10.3390/s22072542.

[58] Li Wang, Xinyu Zhang, Jun Li, et al., "Multi-Modal and Multi-Scale Fusion 3D Object Detection of 4D Radar and LiDAR for Autonomous Driving," IEEE Transactions on Vehicular Technology, vol. 72, no. 5, pp. 5628-5641, 2023. doi: 10.1109/TVT.2022.3230265.

[59] Yan Qiao, Yihan Wang, "MVFAN: Multi-View Feature Assisted Network for 4D Radar Object Detection," in Lecture Notes in Computer Science, 2024, pp. 493-511. doi: 10.1007/978-981-99-8070-3_38.

[60] Hongsi Liu, Jun Liu, Guangfeng Jiang, et al., "MSSF: A 4D Radar and Camera Fusion Framework With Multi-Stage Sampling for 3D Object Detection in Autonomous Driving," IEEE Transactions on Intelligent Transportation Systems, vol. 26, no. 6, pp. 8641-8656, 2025. doi: 10.1109/TITS.2025.3554313.

[61] Jianan Liu, Qiuchi Zhao, Weiyi Xiong, et al., "SMURF: Spatial Multi-Representation Fusion for 3D Object Detection with 4D Imaging Radar," in 2024 IEEE Intelligent Vehicles Symposium (IV), 2024, p. 3141. doi: 10.1109/IV55156.2024.10588505.

[62] Weigang Shi, Ziming Zhu, Kezhi Zhang, et al., "SMIFormer: Learning Spatial Feature Representation for 3D Object Detection from 4D Imaging Radar via Multi-View Interactive Transformers," Sensors, vol. 23, no. 23, Art. no. 9429, 2023. doi: 10.3390/s23239429.

[63] Lili Fan, Changxian Zeng, Yunjie Li, et al., "GRC-Net: Fusing GAT-Based 4D Radar and Camera for 3D Object Detection," SAE International Journal of Advances and Current Practices in Mobility, vol. 6, no. 5, pp. 2690-2696, 2023. doi: 10.4271/2023-01-7088.

[64] Weiyi Xiong, Jianan Liu, Tao Huang, et al., "LXL: LiDAR Excluded Lean 3D Object Detection With 4D Imaging Radar and Camera Fusion," IEEE Transactions on Intelligent Vehicles, vol. 9, no. 1, pp. 79-92, 2024. doi: 10.1109/TIV.2023.3321240.

[65] Zhiwei Lin, Zhe Liu, Yongtao Wang, et al., "RCBEVDet++: Toward High-accuracy Radar-Camera Fusion 3D Perception Network," arXiv preprint arXiv:2409.04979, 2024.

[66] Xun Huang, Ziyu Xu, Hai Wu, et al., "L4DR: LiDAR-4DRadar Fusion for Weather-Robust 3D Object Detection," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 39, no. 4, 2025, pp. 3806-3814. doi: 10.1609/aaai.v39i4.32397.

[67] Tiezhen Jiang, Runjie Kang, Qingzhu Li, "BSM-NET: Multi-Bandwidth, Multi-Scale and Multi-Modal Fusion Network for 3D Object Detection of 4D Radar and LiDAR," Measurement Science and Technology, vol. 36, no. 3, Art. no. 036107, 2025. doi: 10.1088/1361-6501/adafcb.

[68] Sheng Yang, Tong Zhan, Shichen Qiao, et al., "ZFusion: An Effective Fuser of Camera and 4D Radar for 3D Object Perception in Autonomous Driving," in 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2025, pp. 1752-1761. doi: 10.1109/CVPRW67362.2025.00362.

[69] Zecheng Li, Yuying Song, Fuyuan Ai, et al., "Semantic-Guided Depth Completion From Monocular Images and 4D Radar Data," IEEE Transactions on Intelligent Vehicles, vol. 9, no. 9, pp. 5606-5617, 2024. doi: 10.1109/TIV.2024.3355125.

[70] Jen-Hao Cheng, Sheng-Yao Kuan, Hou-I Liu, et al., "CenterRadarNet: Joint 3D Object Detection and Tracking Framework Using 4D FMCW Radar," in 2024 IEEE International Conference on Image Processing (ICIP), 2024, pp. 998-1004. doi: 10.1109/ICIP51287.2024.10648077.

[71] Jianan Liu, Guanhua Ding, Yuxuan Xia, et al., "Which Framework Is Suitable for Online 3D Multi-Object Tracking for Autonomous Driving with Automotive 4D Imaging Radar?," in 2024 IEEE Intelligent Vehicles Symposium (IV), 2024, pp. 1258-1265. doi: 10.1109/IV55156.2024.10588837.

[72] Bin Tan, Zhixiong Ma, Xichan Zhu, et al., "Tracking of Multiple Static and Dynamic Targets for 4D Automotive Millimeter-Wave Radar Point Cloud in Urban Environments," Remote Sensing, vol. 15, no. 11, Art. no. 2923, 2023. doi: 10.3390/rs15112923.

[73] Matthias Zeller, Vardeep S. Sandhu, Benedikt Mersch, et al., "Radar Instance Transformer: Reliable Moving Instance Segmentation in Sparse Radar Point Clouds," IEEE Transactions on Robotics, vol. 40, pp. 2357-2372, 2024. doi: 10.1109/TRO.2023.3338972.

[74] Fangqiang Ding, Andras Palffy, Dariu M. Gavrila, et al., "Hidden Gems: 4D Radar Scene Flow Learning Using Cross-Modal Supervision," in 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023, pp. 9340-9349. doi: 10.1109/CVPR52729.2023.00901.

[75] Yuan Zhuang, Binliang Wang, Jianzhu Huai, et al., "4D iRIOM: 4D Imaging Radar Inertial Odometry and Mapping," IEEE Robotics and Automation Letters, vol. 8, no. 6, pp. 3246-3253, 2023. doi: 10.1109/LRA.2023.3266669.

[76] Jun Zhang, Huayang Zhuge, Yiyao Liu, et al., "NTU4DRadLM: 4D Radar-Centric Multi-Modal Dataset for Localization and Mapping," in 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC), 2023, pp. 4291-4296. doi: 10.1109/ITSC57777.2023.10422606.

[77] Xinyu Zhang, Li Wang, Jian Chen, et al., "Dual Radar: A Multi-Modal Dataset with Dual 4D Radar for Autonomous Driving," Scientific Data, vol. 12, no. 1, Art. no. 439, 2025. doi: 10.1038/s41597-025-04698-2.

[78] Lei Yang, Xinyu Zhang, Jun Li, et al., "V2X-Radar: A Multi-modal Dataset with 4D Radar for Cooperative Perception," in Advances in Neural Information Processing Systems, vol. 38, 2025, pp. 126058-126078. doi: 10.52202/085713-3789.

[79] Pak Hung Chan, Sepeedeh Shahbeigi Roudposhti, Xinyi Ye, et al., "A Noise Analysis of 4D RADAR: Robust Sensing for Automotive?," IEEE Sensors Journal, vol. 25, no. 10, pp. 18291-18301, 2025. doi: 10.1109/JSEN.2025.3556518.

[80] Ignacio Roldan, Andras Palffy, Julian F. P. Kooij, et al., "A Deep Automotive Radar Detector Using the RaDelft Dataset," IEEE Transactions on Radar Systems, vol. 2, pp. 1062-1075, 2024. doi: 10.1109/TRS.2024.3485578.

[81] Yunting Yang, Jun Liu, Hongsi Liu, et al., "MKFusion: Multi-Modal Knowledge Distillation for 4D Radar Point Cloud Segmentation in Autonomous Driving," Knowledge-Based Systems, vol. 339, Art. no. 115616, 2026. doi: 10.1016/j.knosys.2026.115616.

[82] Nicolas Baumann, Michael Baumgartner, Edoardo Ghignone, et al., "CR3DT: Camera-RADAR Fusion for 3D Detection and Tracking," in 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2024, pp. 4926-4933. doi: 10.1109/IROS58592.2024.10801848.

[83] Jie Bai, Lianqing Zheng, Sen Li, Bin Tan, Sihan Chen, and Libo Huang, "Radar Transformer: An Object Classification Network Based on 4D MMW Imaging Radar," Sensors, vol. 21, no. 11, Art. no. 3854, 2021. doi: 10.3390/s21113854.

[84] Fangqiang Ding, Zhijun Pan, Yimin Deng, Jianning Deng, and Chris Xiaoxuan Lu, "Self-Supervised Scene Flow Estimation With 4-D Automotive Radar," IEEE Robotics and Automation Letters, vol. 7, no. 3, pp. 8233-8240, 2022. doi: 10.1109/LRA.2022.3187248.

[85] Wooseong Yang, Hyesu Jang, Ayoung Kim, "Ground-Optimized 4D Radar-Inertial Odometry via Continuous Velocity Integration using Gaussian Process," in 2025 IEEE International Conference on Robotics and Automation (ICRA), 2025, pp. 3815-3821. doi: 10.1109/ICRA55743.2025.11127572.

[86] Leon Ruddat, Laurenz Reichardt, Nikolas Ebert, Oliver Wasenmüller, "Sparsity-Robust Feature Fusion for Vulnerable Road-User Detection with 4D Radar," Applied Sciences, vol. 14, no. 7, Art. no. 2781, 2024. doi: 10.3390/app14072781.

[87] Mengjie Jiang, Gang Xu, Hao Pei, Zeyun Feng, Shuai Ma, Hui Zhang, and Wei Hong, "4D High-Resolution Imagery of Point Clouds for Automotive mmWave Radar," IEEE Transactions on Intelligent Transportation Systems, vol. 25, no. 1, pp. 998-1012, 2024, first published online Mar. 22, 2023. doi: 10.1109/TITS.2023.3258688.

[88] Xiaokai Bai, Lianqing Zheng, Runwei Guan, et al., "4DR360: State Reasoning for Joint 3D Detection and Occupancy Prediction in 4D Radar-Camera Full-Scene Perception," arXiv preprint arXiv:2607.09629, 2026. doi: 10.48550/arXiv.2607.09629.

[89] Jiarui Zhang, Zhihao Li, Chong Wang, and Bihan Wen, "RF4D: Neural Radar Fields for Novel View Synthesis in Outdoor Dynamic Scenes," in 2026 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2026, pp. 15387-15397.

[90] Yuhan Xiao, Yufei Liu, Kai Luan, et al., "Deep LiDAR-Radar-Visual Fusion for Object Detection in Urban Environments," Remote Sensing, vol. 15, no. 18, Art. no. 4433, 2023. doi: 10.3390/rs15184433.