Physics-Aware Gaussian Radar Enhancement for All-weather Perception

Suhani Grover[1] and Chhavi Dhiman[2]

***Abstract*— Autonomous robots operating in adverse weather conditions face a critical perception challenge: while LiDAR provides dense, accurate point clouds in clear conditions, it fails catastrophically in fog, rain, and snow where radar remains functional. However, radar's inherent sparsity severely limits its utility for fine-grained scene understanding. Existing radar enhancement methods treat radar detections as isolated points, ignoring the fundamental physics of radar beam propagation and detection patterns. We propose Physics-Aware Gaussian Radar Enhancement, a novel approach that explicitly models radar beam geometry to generate physics-consistent pseudo ground truth and employs a dual-stream neural architecture for radar-aware enhancement. Our method models each radar detection as a 3D frustum representing the actual beam pattern, performs stratified sampling within this volume, and projects samples to LiDAR surfaces while preserving radar physics. A dual-stream network with cross-modal attention processes geometric and physics features separately before fusion, predicting enhanced point clouds with radar-aware Gaussian parameters. Evaluated on nuScenes, PAGRE achieves stable training convergence and demonstrates a physics-consistent approach to radar enhancement, enabling robust all-weather robot perception. Code will be made available at https://github.com/Suhani92/Physics-Aware-Radar-Enhancement.**

I. Introduction

Autonomous robots deployed in real-world environments must maintain reliable perception capabilities across diverse weather conditions. While LiDAR sensors provide high-resolution 3D point clouds that have become the backbone of modern robotic perception systems, they suffer from critical limitations in adverse weather. Rain droplets, fog, and snow particles cause severe attenuation and scattering of laser pulses, leading to dramatically reduced range, increased noise, and complete sensor failure in extreme conditions. This weather dependency represents a fundamental barrier to deploying autonomous systems in safety-critical applications such as autonomous driving, search and rescue operations, and outdoor service robotics.

Cameras complement LiDAR with rich appearance features, yet they don't provide depth information and fail under low light and are sensitive to adverse weather conditions and occlusions. Additionally, both modalities are unable to measure object velocity without relying on temporal information.

Automotive radar systems offer a compelling solution to weather-induced perception failures. Operating in the millimeter-wave spectrum (76-81 GHz), radar signals penetrate atmospheric particles that block optical sensors, maintaining functionality in fog, rain, and snow. Radars provide not only spatial information (range, azimuth, elevation) but also unique physics-based measurements including Doppler velocity and radar cross-section (RCS), enabling direct velocity estimation and material property inference [1].

These advantages have led to widespread use of radar in Advanced Driving Assistance System (ADAS) applications such as adaptive cruise control and collision avoidance. However, radar's primary limitation is sparsity. While LiDAR typically returns 30,000+ points, automotive radar produces only hundreds of detections because each radar detection represents the strongest reflector within a 3D beam volume rather than a precise point measurement. Traditional approaches to radar processing either ignore this fundamental difference or apply point-cloud methods designed for dense, precise LiDAR data. Together, these factors yield point clouds that are both less informative and noisier than LiDAR, making radar challenging to directly employ in high-precision tasks such as detection, mapping, and localization.

Therefore, it is essential to address these challenges at the point cloud representation level rather than only at downstream tasks. By enhancing the radar point cloud to be geometrically faithful and physically consistent, we enable radar to move beyond a supplementary role and serve as a robust sensor modality for perception in safety-critical applications. Our approach combines multi-sensor fusion with physics-aware modeling: we use LiDAR geometry to guide radar enhancement while explicitly preserving radar-specific measurements rather than simply converting radar to LiDAR-like points. This differs from pure generative methods by grounding predictions in physical sensor characteristics.

This work proposes a Physics-Aware Gaussian Radar Enhancement to explicitly model radar beam physics for enhancement tasks (illustrated in Fig. 1). Our key contributions are:

- Beam-Aware Pseudo Ground Truth Generation: We model each radar detection as a 3D frustum representing the actual beam pattern, perform stratified sampling within beam volumes, and project samples to LiDAR surfaces while preserving radar physics.
- Dual-Stream Architecture: We propose a novel neural architecture with separate geometric and physics processing streams, using cross-modal attention for fusion and predicting radar-aware Gaussian parameters (representing each enhanced point as a 3D ellipsoid rather than a discrete location, enabling continuous surface representation).
- Physics-Preserving Enhancement: Our method maintains radar cross-section(RCS) information throughout

[1]Suhani Grover is with Delhi Technological University, Delhi, India. suhani1077@gmail.com

[2]Chhavi Dhiman is with the Department of Electronics and Communication Engineering, Delhi Technological University, Delhi, India. chhavi.dhiman@dtu.ac.in

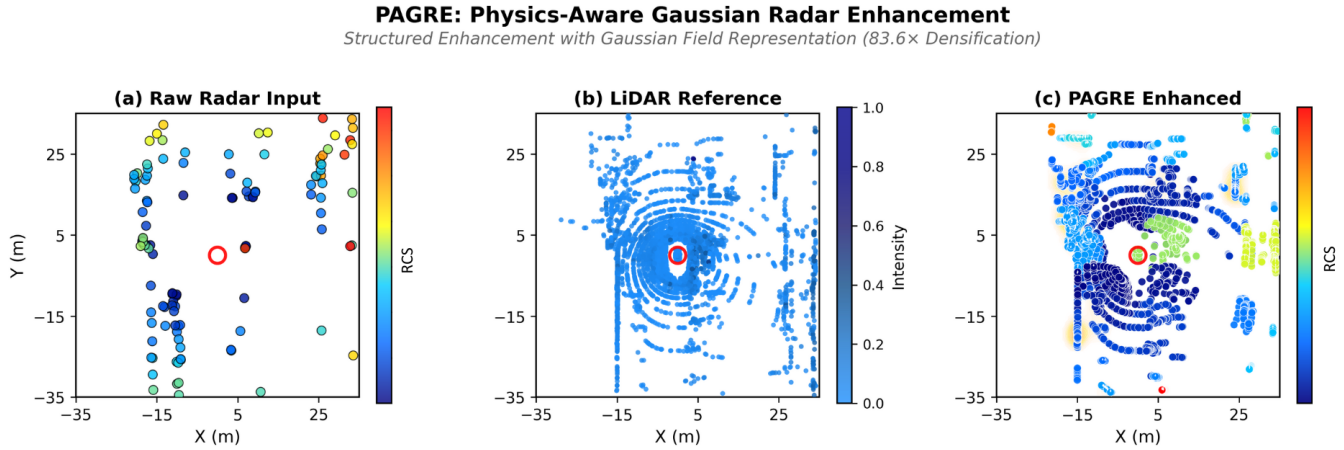

Fig. 1. Radar enhancement comparison: (a) raw sparse radar input, (b) LiDAR reference, and (c) PAGRE dense radar output with RCS preservation

the enhancement process, ensuring physical consistency in the output.

- Adaptive Loss Weighting: We introduce learnable uncertainty weighting to balance geometry, physics, and Gaussian losses during training.

II. Related Work

A. Classical Radar Signal Processing

Automotive mmWave radar traditionally relies on signal processing pipelines that form range Doppler angle tensors using FFTs and digital beamforming, followed by constant false-alarm rate (CFAR) detection and clustering to extract point targets [2], [3]. While efficient and interpretable, these pipelines assume ideal propagation and narrow beams; in practice, multipath, sidelobes, and specular reflections create clutter and ghost detections, while wide beam footprints yield sparse, ambiguous point clouds that lack surface detail.

B. Interpolation and Heuristic Densification

A straightforward way to increase density is to treat radar detections as isolated 3D points and apply point-wise interpolation or kernel upsampling (e.g., nearest-neighbor, KNN, or radial basis smoothing). Such techniques (often borrowed from LiDAR completion and depth upsampling) can fill gaps locally but ignore beam geometry and do not respect radar specific attributes such as radar cross-section (RCS). As a result, they tend to blur object boundaries and distort physical statistics, limiting their utility for safety-critical perception [4], [5].

C. Learning-Based Radar Enhancement

Learning-based enhancement directly maps sparse radar observations to denser geometric representations. Earlier used methods demonstrated that neural networks can super-resolve mmWave radar by training with cross-modal supervision from LiDAR. High-Resolution Point Clouds from mmWave Radar, RadarHD [6] learns to generate LiDAR-like point clouds from radar I/Q or low-resolution projections, establishing the viability of cross-modal training for radar densification. Closely related threads study robust radar imaging and enhancement under adverse weather [7] and learning to exploit radar for downstream tasks.

Recently, generative models, particularly diffusion, have set new performance levels for radar enhancement by injecting strong priors. Some approaches operate over range images or range-azimuth heatmaps and then lift to 3D, denoising and densifying with diffusion backbones [8], [9]. These methods achieve impressive geometry, but their reliance on image-like intermediate representations or heavy priors can obscure the physical meaning of outputs (e.g., RCS statistics) and risk hallucinations that are undesirable for safety-critical applications.

D. Radar Fusion for 3D Perception

Extensive research has explored multi-modal fusion approaches combining radar with cameras and LiDAR for autonomous perception tasks, including object detection, tracking, and semantic scene understanding. RadarNet [10] proposes a radar-centric pipeline for dynamic-object perception. CenterFusion [11] fuses radar with images via center-based detection, improving robustness to occlusion. While RCM-Fusion [12] performs multi-level radar–camera fusion to enhance 3D detection. Bi-LRFusion [13] explores bi-directional LiDAR–radar fusion that leverages complementary geometry and velocity cues. While these methods demonstrate the value of radar in a multi-modal stack, they predominantly focus on late or feature-level fusion for specific downstream tasks rather than improving the radar representation itself.

E. Point-Cloud Backbones and Attention

Point-based encoders such as PointNet and PointNet++ [14], [15] established permutation-invariant processing with local neighborhood aggregation for 3D perception. Attention mechanisms and variants (e.g., Attentional PointNet [16]) further improve feature selection under clutter. We leverage these principles via a dual-stream (geometry + physics) backbone fused with cross-modal attention, tailored to radar's measurement model and RCS.

III. Preliminaries

A. FMCW Radar Signal Model

Modern automotive radars operate using frequency-modulated continuous wave (FMCW) chirps. Each chirp sweeps linearly in frequency over a bandwidth B for a duration T_{chirp}. A radar frame typically consists of N_{chirp} chirps transmitted sequentially. With N_{TX} transmit and N_{RX} receive antennas, MIMO beamforming yields up to $N_{\text{TX}} \times N_{\text{RX}}$ virtual channels, improving angular resolution. The raw reflections sampled across time, antennas, and chirps form a 3D radar data cube, from which range, Doppler velocity, and angle of arrival can be estimated.

B. Range Estimation

The reflected signal is mixed with the transmitted chirp, producing an intermediate frequency (IF) beat signal whose frequency f_{beat} is proportional to the time delay τ between transmit and receive. The target range is given by:

$$R = \frac{c\, T_{\text{chirp}}\, f_{\text{beat}}}{2B} \tag{1}$$

where c is the speed of light. The maximum measurable range is $R_{\max} = \frac{cT_{\text{chirp}}}{2}$, and the range resolution is $\Delta R = \frac{c}{2B}$.

C. Velocity Estimation

Relative velocity is measured via the Doppler shift across successive chirps. The frequency shift f_d between chirps is related to the radial velocity:

$$v_r = \frac{\lambda f_d}{2} = \frac{c f_d}{2 f_c} \tag{2}$$

where $\lambda = c/f_c$ is the wavelength and f_c is the carrier frequency. A second FFT across the chirp dimension (Doppler FFT) is commonly applied to extract f_d.

D. Angle Estimation and Point Cloud Generation

Phase differences across the virtual antenna array enable angle-of-arrival (AoA) estimation. For azimuth angle θ and elevation ϕ, the 3D Cartesian coordinates of each detection are:

$$x = R \cdot \cos(\phi) \cdot \sin(\theta) \tag{3}$$

$$y = R \cdot \cos(\phi) \cdot \cos(\theta) \tag{4}$$

$$z = R \cdot \sin(\phi) \tag{5}$$

The angular resolution is approximately $\Delta\theta \approx \frac{\lambda}{N_{\text{TX}} N_{\text{RX}} d \cos\theta}$, where d is the antenna spacing. After range, Doppler, and angle FFTs, detections are extracted from the radar cube using target detectors such as CFAR, producing a sparse 3D point cloud augmented with velocity and reflectivity.

E. Radar Cross-Section (RCS)

Radar cross-section (RCS) quantifies the reflectivity of a target and provides material-specific information:

$$\sigma = 4\pi R^2 \frac{|E_s|^2}{|E_i|^2} \tag{6}$$

where E_s and E_i are scattered and incident field strengths, respectively. Unlike LiDAR intensity, RCS captures target size, shape, and material, making it valuable for classification and physics-aware modeling.

IV. Methodology

The pipeline is a structured process designed to transform sparse radar point clouds into dense, physics-aware representations. It begins with a novel pseudo-ground-truth generation process, followed by a dual-stream neural network that learns the enhancement mapping, all trained via a physics-preserving, adaptive multi-task loss function.

A. Physics-Aware Pseudo Ground Truth Generation

Unlike existing methods that treat radar detections as precise points, we model the underlying beam geometry that governs radar detection formation. This ensures that pseudo-ground-truth data reflects both radar's physics and LiDAR's geometric fidelity.

1) Radar Beam Modeling: Each radar detection originates from electromagnetic energy scattered within a 3D beam volume, as shown in Fig. 2. We model this as a truncated cone (frustum) defined by:

$$\mathcal{F}(r,\theta,\phi) = \left\{ \mathbf{p} \in \mathbb{R}^3 : \begin{array}{l} r_{\min} \le \|\mathbf{p}-\mathbf{o}\|_2 \le r_{\max}, \\ \angle(\mathbf{p}-\mathbf{o}, \mathbf{d}) \le \theta_{\text{beam}}/2 \end{array} \right\} \tag{7}$$

where $\mathbf{o} \in \mathbb{R}^3$ is the sensor origin, $\mathbf{d} \in \mathbb{R}^3$ is the detection direction, θ_{beam} is the beam width (3°), and $r_{\min}, r_{\max}$ define the range resolution bounds.

$$\mathcal{P}_{\text{enhanced}} = \{\mathbf{p}_i \mid \mathbf{p}_i \sim \text{Uniform}(\mathcal{F}_k), \text{inherit}(\text{RCS}_k)\} \tag{8}$$

with each $\mathbf{p}_i$ inheriting RCS from its parent radar detection.

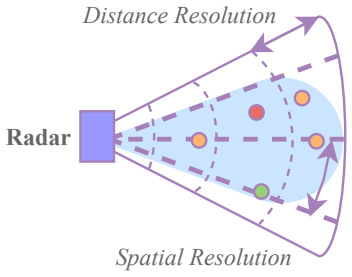

Fig. 2. Radar beam modeling with 3D frustum representation showing angular and range resolution constraints.

2) Surface Snapping: These randomly sampled points are then projected onto the true scene geometry. Each $\mathbf{p}_i$ is snapped to its nearest neighbor in the LiDAR cloud $\mathcal{L}$ if within a snapping threshold τ_{snap}.

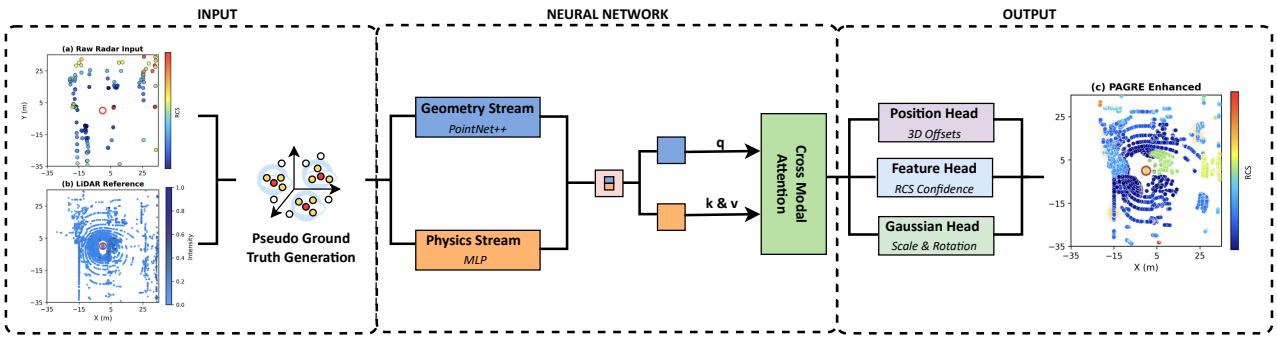

Fig. 3. PAGRE dual-stream architecture with geometric and physics processing streams, cross-modal attention fusion, and multi-head prediction outputs

3) Confidence Scoring: Not all generated points are equally reliable. Points generated far from the central axis of the original radar detection are geometrically less certain. This is modeled by assigning a Gaussian confidence score to each new point using a Gaussian function based on its Euclidean distance to the original radar point:

$$c_i = \exp\left(-\frac{\|\mathbf{p}_{\text{snap}} - \mathbf{r}_k\|_2^2}{2\sigma^2}\right), \tag{9}$$

where $\mathbf{r}_k$ is the original radar detection and σ controls rate of falloff. Points closer to the detection axis receive higher confidence, while points farther away are down-weighted during training (shown in Fig. 4). This score becomes a vital component for weighting the loss during network training, encouraging the model to prioritize predictions closer to the original measurements.

4) Physics Preservation: Every newly generated pseudo ground-truth point inherits radar-specific attributes (e.g., Radar cross-section (RCS)). This reflects radar's sensing mechanism more faithfully than purely geometric interpolation or CFAR-based thresholding.

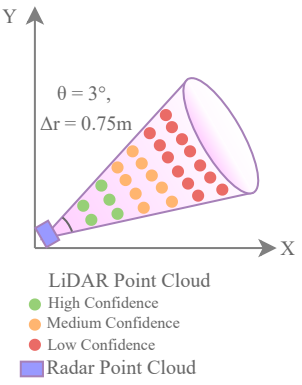

Fig. 4. LiDAR surface snapping with confidence-based weighting where colors indicate reliability levels.

B. Physics-Aware Network Architecture (PAGRENet)

The model is a dual-stream, attention-based architecture designed to first disentangle and then intelligently fuse geometric and physical information from the sparse radar input. The overall pipeline is shown in Fig. 3, where radar detections are processed by geometric and physics streams, fused via cross-modal attention and decoded into multi-head outputs.

a) Geometric Stream: The 3D coordinates (x,y,z) of the input radar points are processed with a hierarchical PointNet++ [15] backbone. This architecture captures local and global geometric features at multiple scales, allowing the network to learn and recognize spatial structures like planes, edges, and object clusters.

b) Physics Stream: The stream processes the physical attributes (RCS), using a simpler point-wise Multi-Layer Perceptron (MLP) to extract high-level feature embeddings from these physical measurements independently for each point.

c) Cross-Modal Attention Fusion: A simple concatenation of the two feature streams would fail to capture the nuanced, often non-linear relationship between an object's geometry and its physical properties. Instead, we employ cross-modal attention [17] where geometric features (from PointNet++ [15]) act as the Query (Q), while physics features (from MLP) serve as both the Key (K) and Value (V). This formulation allows the network to learn, for each point, which physical attributes from its local neighborhood are most relevant given its specific geometric context. For example, vertical surfaces with high curvature can selectively attend to high RCS values (indicating metal objects) while planar ground regions attend to lower RCS values (asphalt, vegetation). The attention mechanism explicitly models the geometry-physics relationship rather than assuming independence. A residual connection preserves the original geometric structure:

$$\mathbf{F}_{\text{fused}} = \text{Attn}(\mathbf{F}_{\text{geo}}, \mathbf{F}_{\text{phys}}) + \mathbf{F}_{\text{geo}} \tag{10}$$

where Attn denotes cross-modal attention with $\mathbf{F}_{\text{geo}}$ as queries and $\mathbf{F}_{\text{phys}}$ as keys/values, and the residual connection preserves geometric structure.

d) Output Heads: The fused features are then passed to three separate prediction heads:

- Position Head: An MLP that predicts 3D offsets for each input point, effectively learning to push the points towards their correct positions in the dense output cloud.

- Feature Head: An MLP that predicts the physical attributes.
- Gaussian Head: An MLP that predicts the parameters of a 3D Gaussian for each point, consisting of 3 values for scale(s) and 4 for a rotation quaternion(q). This provides a richer, continuous surface representation, moving beyond simple points.

$$\begin{aligned}\Delta\mathbf{p} &= \text{PositionHead}(\mathbf{F}_{\text{fused}}) \in \mathbb{R}^{N\times K\times 3} \\ \mathbf{s} &= \text{softplus}(\text{ScaleHead}(\mathbf{F}_{\text{fused}})) + \epsilon \in \mathbb{R}^{N\times K\times 3} \\ \mathbf{q} &= \text{normalize}(\text{RotationHead}(\mathbf{F}_{\text{fused}})) \in \mathbb{R}^{N\times K\times 4} \\ \phi_{\text{enhanced}} &= \text{FeatureHead}(\mathbf{F}_{\text{fused}}) \in \mathbb{R}^{N\times K\times 2}\end{aligned} \tag{11}$$

where $K = 5$ denotes the upsampling factor, $\Delta\mathbf{p}$ represents position offsets from the original radar detections, $\mathbf{s}$ are positive scale parameters ensuring valid Gaussian ellipsoids, $\mathbf{q}$ are normalized quaternions encoding 3D rotations, and ϕ_{enhanced} contains the enhanced radar features (RCS and confidence scores). The softplus activation ensures positive scales, while normalization constrains quaternions to the unit sphere for valid rotations. Final enhanced positions are computed as $\mathbf{p}_{\text{final}} = \mathbf{p}_{\text{radar}} + 0.1 \cdot \Delta\mathbf{p}$, where the scaling factor stabilizes training dynamics.

C. Adaptive Multi-Task Training

a) Physics-Preserving Augmentation: To ensure model robustness and generalization, random geometric augmentations (rotation, scaling) are applied during training. These transformations must be applied consistently to both spatial and kinematic data to preserve physical laws. Failing to apply consistent transformations would violate physical laws, corrupting training data and causing the model to learn incorrect geometry-motion correlations.

b) Adaptive Loss Function: Training PAGRENet involves optimizing three distinct objectives: geometric accuracy, physical feature fidelity, and Gaussian shape regularity.

Manually tuning the weights for these different loss components is difficult, time-consuming, and often leads to suboptimal performance. Therefore, an adaptive multi-task loss function based on uncertainty weighting is employed [18]. The total loss is formulated as:

$$\mathcal{L}_{\text{total}} = \frac{1}{\sigma_1^2}\mathcal{L}_{\text{geometry}} + \frac{1}{\sigma_2^2}\mathcal{L}_{\text{physics}} + \frac{1}{\sigma_3^2}\mathcal{L}_{\text{gaussian}} + \log(\sigma_1\sigma_2\sigma_3) \tag{12}$$

where $\sigma_1, \sigma_2, \sigma_3$ are learnable uncertainty parameters that automatically balance loss components during training.

The geometric loss employs confidence-weighted Chamfer distance which measures the similarity between predicted points and ground-truth points as:

$$\begin{aligned}\mathcal{L}_{\text{geometry}} = &\frac{1}{|\mathcal{P}|}\sum_{i=1}^{|\mathcal{P}|}\min_j \|\mathbf{p}_i - \mathbf{q}_j\|_2 \\ &+ \frac{1}{|\mathcal{Q}|}\sum_{j=1}^{|\mathcal{Q}|} c_j \min_i \|\mathbf{q}_j - \mathbf{p}_i\|_2\end{aligned} \tag{13}$$

where $\mathcal{P}$ and $\mathcal{Q}$ represent predicted and pseudo ground truth point sets, and c_j are confidence weights.

The physics consistency loss preserves RCS information:

$$\mathcal{L}_{\text{physics}} = \|\text{RCS}_{\text{pred}} - \text{RCS}_{\text{gt}}\|_2 + \text{BCE}(\text{conf}_{\text{pred}}, \text{conf}_{\text{gt}}) \tag{14}$$

where BCE denotes binary cross-entropy for confidence prediction.

The Gaussian regularization loss prevents degenerate ellipsoid parameters:

$$\mathcal{L}_{\text{gaussian}} = \|\mathbf{s} - \mathbf{s}_{\text{target}}\|_2^2 + \|\|\mathbf{q}\|_2 - 1\|_2^2 \tag{15}$$

where $\mathbf{s}_{\text{target}}$ encourages reasonable scale values and the second term enforces unit quaternion constraints.

This comprehensive training framework ensures that the model learns physics-consistent radar enhancement while automatically balancing competing objectives, providing a robust foundation for evaluation on real-world radar-LiDAR datasets.

V. Experimental Evaluation

A. Experimental Setup

1) Dataset: The method is evaluated on the large-scale nuScenes dataset [19], which provides synchronized data from a full automotive sensor suite, including radar and LiDAR. The dataset is processed by extracting paired radar and LiDAR scans for each keyframe. The proposed beam-aware pseudo ground truth generation is then run offline to create the dense pseudo-ground-truth point clouds and corresponding confidence maps required for training.

PAGRE is implemented in PyTorch and trained end-to-end using AdamW optimizer with learning rate 1×10^{-4}, batch size 4, and weight decay 1×10^{-4}. The learnable uncertainty parameters $\{\sigma_1, \sigma_2, \sigma_3\}$ are initialized to $\{0.0, -2.0, 0.0\}$ respectively. The adaptive loss weighting (Eq. 12) enables stable training convergence with all loss components decreasing monotonically over 50 epochs and reaching steady state performance by epoch 40. The learnable uncertainty parameters balance the three objectives without manual tuning.

2) Baselines: We compare our proposed model against several radar enhancement approaches. : (1) Random Upsampling: Simple replication with Gaussian noise, (2) KNN Interpolation: K-nearest neighbor interpolation from sparse radar, (3) Nearest Neighbor Copy: Direct copying of nearest radar measurements, and (4) RadarHD [6]: State-of-the-art deterministic radar super-resolution method to provide a fair comparison against existing work in the same problem category.

B. Evaluation Metrics

a) Evaluation Metrics: We evaluate enhancement quality by comparing the network's output against the original LiDAR point cloud, which serves as the geometric reference. The pseudo ground truth(generated via beam-aware sampling and LiDAR projection, as described in Section IV.A) is used only during training to supervise the network. At test time, we measure how well the enhanced radar point

TABLE I

COMPARISON OF PAGRE WITH BASELINES ON NUSCENES.

Method	Quantitative Metrics						Physics Metrics		
	Chamfer ↓	Hausdorff ↓	F-Score ↑	Earth Mover's Dist. ↓	Coverage ↑	FPS ↑	RCS Mean	RCS Std ↓	Temporal ↑
Random Upsampling	2.730	40.463	0.002	40.463	34.2	132.2	3.58	5.20	0.27
KNN Interpolation	2.707	40.208	0.006	40.208	32.2	9.4	3.38	5.10	0.45
NN Copy	2.735	40.687	0.004	40.687	36.0	**204.5**	3.42	5.09	0.27
RadarHD	2.754	40.037	**0.010**	40.037	24.5	7.4	3.37	5.10	0.65
PAGRE (Ours)	**1.652**	**27.664**	0.006	**27.664**	**46.1**	71.0	**5.95**	**4.17**	**0.75**

cloud matches the true LiDAR geometry while preserving radar-specific physics (RCS). We evaluate geometric accuracy using: Chamfer Distance (average closest point distance), Hausdorff Distance (maximum distance between point sets), F-Score (harmonic mean of precision/recall at 0.1m threshold), Earth Mover's Distance (optimal transport cost), and Coverage (percentage of LiDAR reference points within threshold). For physics consistency, we measure RCS preservation (mean and standard deviation of radar cross-section values) and Temporal consistency across consecutive frames. FPS measures computational efficiency.

C. Quantitative Results

Table I presents comprehensive quantitative evaluation results. PAGRE achieves substantial improvements over existing methods: 39% reduction in Chamfer distance compared to RadarHD (1.652 vs 2.754), 31% improvement in Hausdorff distance (27.664 vs 40.037), 31% reduction in Earth Mover's distance (27.664 vs 40.037), and 88% increase in coverage (46.1% vs 24.5%). These geometric improvements demonstrate the effectiveness of our dual-stream architecture and physics-aware pseudo ground truth generation. While RadarHD achieves the highest F-Score (0.010 vs 0.006), this metric can favor sparser predictions in low-density scenarios. PAGRE's superior performance across distance-based metrics (Chamfer, Hausdorff, EMD) indicates better overall geometric fidelity and point distribution matching. Crucially, PAGRE maintains superior physics consistency with 76% higher RCS preservation (5.95 vs 3.37 mean RCS) and 18% lower RCS variance (4.17 vs 5.10 standard deviation) compared to RadarHD. The temporal consistency metric shows 15% improvement (0.75 vs 0.65), indicating better frame-to-frame coherence essential for tracking applications. PAGRE achieves competitive computational efficiency (71.0 FPS) while significantly outperforming RadarHD (7.4 FPS) in accuracy, demonstrating the effectiveness of our streamlined dual-stream architecture for real-time deployment.

Robustness and Limitations: Our beam modeling uses fixed parameters ($\theta_{\text{beam}} = 3$, $\Delta r = 0.75$m) derived from the Continental ARS408 radar in nuScenes. While these radar parameters are sensor-specific, the frustum-based approach generalizes to other automotive radars by adjusting beam width and range resolution according to radar manufacturer specifications. The method's performance depends on LiDAR quality during training; in scenes with very sparse radar or heavy occlusion, enhancement quality degrades as fewer beam volumes contain reliable LiDAR surfaces for projection.

Weather Conditions: The nuScenes dataset contains 165 scenes with rain conditions where LiDAR performance typically degrades while radar maintains functionality. Our experiments include these adverse weather scenes in both training and evaluation, demonstrating that PAGRE maintains consistent performance across weather conditions. The physics-aware approach is particularly valuable in these scenarios where radar becomes the primary reliable sensor. Future work will include dedicated weather-stratified evaluation and additional adverse weather datasets to fully validate all-weather capabilities.

D. Ablation Studies

To better understand the contribution of each design choice in our framework, we perform an ablation study by systematically removing or altering key components of PAGRE. Specifically, we examine the effect of removing the physics stream, cross-attention, beam-aware pseudo ground truth, and reducing the model to a geometry-only baseline. Each variant is evaluated on both geometric fidelity (Chamfer distance, Hausdorff distance, F-Score, Coverage) and reflectivity consistency (RCS Error). Table II presents

TABLE II

ABLATION STUDY

Configuration	Chamfer	Hausdorff	F-Score	Coverage	RCS Error
w/o Physics Stream	1.810	29.900	0.005	44.8	5.994
w/o Cross-Attention	1.770	28.900	0.005	45.2	**3.961**
Geometry Only Baseline	1.840	30.200	0.005	43.5	5.789
w/o Beam-Aware Pseudo GT	1.900	31.100	0.006	44.0	5.803
PAGRE Full Model	**1.652**	**27.664**	**0.006**	**46.1**	5.727

a comprehensive ablation study analyzing the contribution of each core component in PAGRE. These results validate our core hypothesis: physics-consistent training data generation (beam-aware pseudo GT) is the most critical component, while cross-modal attention enables effective geometry-physics fusion. The geometry-only baseline confirms that RCS information provides valuable complementary signal beyond spatial coordinates alone. Removing the cross-modal attention mechanism shows the smallest geometric impact (1.770 vs 1.652 Chamfer distance), yet attention proves critical for physics preservation, achieving the best RCS error (3.961). The physics stream contributes substantially to spatial reconstruction (1.810 Chamfer when removed), val-

idating the importance of incorporating radar-specific measurements alongside geometric features. Most significantly, removing beam-aware pseudo ground truth generation causes the largest degradation across all metrics (1.900 Chamfer, 44.0 coverage), strongly validating our core hypothesis that physics-consistent training data generation is fundamental to effective radar enhancement.

Overall, these results demonstrate that the full PAGRE model offers the most favorable trade-off across accuracy, completeness, and radar consistency.

VI. Conclusion

In this paper, we introduced a physics-aware framework for radar point cloud super-resolution. The method contributes three core innovations: (i) a LiDAR-guided, beam-aware pseudo ground truth generation process for creating high-quality and physically consistent training data, (ii) a dual-stream attention network that jointly models geometric structure and radar-specific physical attributes, and (iii) an adaptive multi-task loss with uncertainty weighting for robust optimization across geometry, reflectivity, and Gaussian-based surface regularization. Extensive experiments on the nuScenes dataset [19] demonstrate that PAGRE achieves strong performance among deterministic radar super-resolution methods, producing dense point clouds that preserve both geometric fidelity and radar physics. By incorporating Gaussian representations, the approach is also compatible with neural rendering pipelines, offering a richer, continuous scene representation beyond discrete radar returns.

While this work focuses on representation enhancement at the point cloud level, the improved radar representations enable downstream perception tasks. The physics-preserving approach makes enhanced radar suitable for object detection, tracking, and mapping pipelines that rely on consistent sensor characteristics. Future work will extend PAGRE by: (i) evaluating cross-sensor generalization and performance under extreme weather conditions with dedicated weather-stratified benchmarks, (ii) leveraging temporal information to improve frame-to-frame consistency, (iii) incorporating velocity alongside RCS for full kinematic preservation in dynamic scenes, and (iv) integrating PAGRE with downstream tasks such as object detection, tracking, and SLAM. By unifying geometric and physical constraints, PAGRE paves the way toward reliable all-weather robotic perception and autonomy.

References

[1] K. Harlow, H. Jang, T. D. Barfoot, A. Kim, and C. Heckman, "A new wave in robotics: Survey on recent mmwave radar applications in robotics," *IEEE Transactions on Robotics*, vol. 40, pp. 4544–4560, 2024.

[2] M. I. Skolnik, *Radar Handbook*, 3rd ed. McGraw-Hill Education, 2008.

[3] M. A. Richards, J. A. Scheer, and W. A. Holm, *Principles of Modern Radar: Basic Principles*. SciTech Publishing, 2010.

[4] M. Jaritz, R. de Charette, E. Wirbel, X. Perrotton, and F. Nashashibi, "Sparse and dense data with cnns: Depth completion and semantic segmentation," 2018. [Online]. Available: https://arxiv.org/abs/1808.00769

[5] T. Roddick, A. Kendall, and R. Cipolla, "Orthographic feature transform for monocular 3d object detection," *British Machine Vision Conference*, 2019.

[6] A. Prabhakara, T. Jin, A. Das, G. Bhatt, L. Kumari, E. Soltanaghai, J. Bilmes, S. Kumar, and A. Rowe, "High resolution point clouds from mmwave radar," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 4135–4142.

[7] J. Guan, S. Madani, S. Jog, S. Gupta, and H. Hassanieh, "Through fog high-resolution imaging using millimeter wave radar," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 11 461–11 470.

[8] R. Wu, Z. Li, J. Wang, X. Xu, Z. Zheng, K. Huang, and G. Lu, "Diffusion-based mmwave radar point cloud enhancement driven by range images," 2025. [Online]. Available: https://arxiv.org/abs/2503.02300

[9] R. Zhang, D. Xue, Y. Wang, R. Geng, and F. Gao, "Towards dense and accurate radar perception via efficient cross-modal diffusion model," 2024. [Online]. Available: https://arxiv.org/abs/2403.08460

[10] B. Yang, R. Guo, M. Liang, S. Casas, and R. Urtasun, "Radarnet: Exploiting radar for robust perception of dynamic objects," 2020. [Online]. Available: https://arxiv.org/abs/2007.14366

[11] R. Nabati and H. Qi, "Centerfusion: Center-based radar and camera fusion for 3d object detection," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, January 2021, pp. 1527–1536.

[12] J. Kim, M. Seong, G. Bang, D. Kum, and J. W. Choi, "Rcm-fusion: Radar-camera multi-level fusion for 3d object detection," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 18 236–18 242.

[13] Y. Wang, J. Deng, Y. Li, J. Hu, C. Liu, Y. Zhang, J. Ji, W. Ouyang, and Y. Zhang, "Bi-lrfusion: Bi-directional lidar-radar fusion for 3d dynamic object detection," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 13 394–13 403.

[14] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," *arXiv preprint arXiv:1612.00593*, 2016.

[15] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," *arXiv preprint arXiv:1706.02413*, 2017.

[16] A. Paigwar, O. Erkent, C. Wolf, and C. Laugier, "Attentional pointnet for 3d-object detection in point clouds," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019, pp. 1297–1306.

[17] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2023. [Online]. Available: https://arxiv.org/abs/1706.03762

[18] R. Cipolla, Y. Gal, and A. Kendall, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7482–7491.

[19] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuscenes: A multimodal dataset for autonomous driving," *arXiv preprint arXiv:1903.11027*, 2019.