R²D-LPCC: Relevance-Ranking Guided Region-Adaptive Dynamic LiDAR Point Cloud Compression

Fangzhe Nan[1,2], Frederick W. B. Li[3], Gary K. L. Tam[4], Zhaoyi jiang[2], Bailin Yang[2*], Jingke Cui[1,2], Changshuo Wang[5]

[1]School of Statistics and Mathematics, Zhejiang Gongshang University, Hangzhou, 310000, Zhejiang, China
[2]School of Computer Science and Technology, Zhejiang Gongshang University, Hangzhou, 310000, Zhejiang, China
[3]Durham University, Durham, DH1 3LE, United Kingdom
[4]Swansea University, Swansea, SA2 8PP, United Kingdom
[5]Department of Computer Science, University College London, London, WC1E 6EA, United Kingdom
23010040030@pop.zjgsu.edu.cn, frederick.li@durham.ac.uk, K.L.Tam@Swansea.ac.uk, zyjiang@mail.zjgsu.edu.cn, ybl@zjgsu.edu.cn, 25010040035@pop.zjgsu.edu.cn, wangchangshuo1@gmail.com

Abstract

Dynamic LiDAR point cloud compression (LPCC) is crucial for the efficient transmission and storage of large-scale three-dimensional data in applications such as autonomous driving. However, many existing methods, which primarily focus on compressing geometric or motion information, face a fundamental limitation: they treat all points as equally important. This approach neglects the semantic priorities of a scene, resulting in inefficient bit allocation and particularly compromising the reconstruction quality of safety-critical regions, such as pedestrians and vehicles, which are vital to downstream perception tasks. To address these limitations, we propose R²D-LPCC, a relevance-ranking framework for region-adaptive LPCC that prioritizes fidelity in semantically important regions. Central to our approach is the Adaptive Relevance Learning (ARL) module, which integrates semantic context with uncertainty to evaluate regional significance and guide compression. We also introduce a Multi-scale Region-Adaptive Transform (MRAT) module to enhance semantic feature modeling and preserve fine-grained details in key areas. Additionally, we develop an adaptive multimodal motion estimation module to improve motion prediction in complex three-dimensional environments. Extensive experiments conducted on the SemanticKITTI benchmark demonstrate that R²D-LPCC significantly surpasses ten recent state-of-the-art methods, achieving a 45.48% BD-rate gain over the previous leading method, Unicorn, and a 98.58% gain over the GPCC standard, while ensuring superior reconstruction quality in semantically important regions. Project page with code: https://github.com/zj-nn/R2D-LPCC.

Introduction

LiDAR point cloud compression has witnessed substantial progress in response to the escalating need to efficiently manage massive and complex 3D data. Outdoor dynamic LiDAR captures complex scenes that include a wide variety of elements such as buildings, vegetation, ground surfaces, and moving vehicles, resulting in rich spatial distributions.

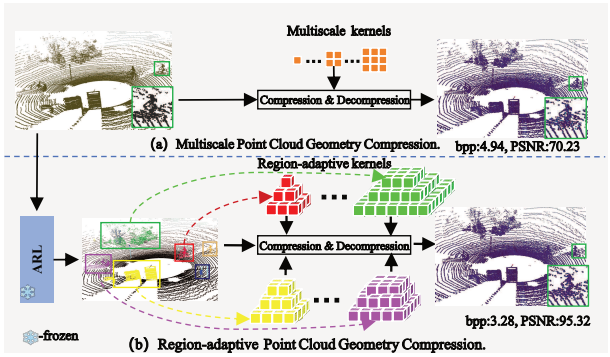

Figure 1: Paradigms of point cloud compression methods. (a) Traditional multi-scale compression assumes equal importance for all input points. (b) The proposed method generates region-adaptive kernels from different regions, enabling more effective compression.

These point clouds are indispensable for applications including autonomous driving, urban mapping, and robotic navigation. However, the raw data size is often enormous, reaching gigabytes or even terabytes, and successive frames possess significant spatiotemporal redundancy. This imposes serious challenges for real-time processing, transmission, and storage, making robust and efficient compression methods vital.

Traditional LiDAR point cloud compression (LPCC) approaches like VPCC and GPCC (Schwarz et al. 2018) mostly rely on 2D projections or native 3D decorrelation techniques for predictive coding. These conventional frameworks are grounded in handcrafted feature design and heuristic matching, which limits their flexibility and coding efficiency across diverse scenarios. Range-image–based methods (Sun et al. 2019; Wang et al. 2022) convert 3D points into 2D images to ease encoding but often sacrifice important geometric details by introducing distortions. Octree-based solutions (Huang et al. 2020; Que, Lu, and Xu 2021; Fu et al. 2022; Wang et al. 2023; Cui et al. 2023; Fan et al. 2023) hierarchically partition 3D space to exploit multiscale structure but suffer from the overhead of full octree construction

*Corresponding Author. Supplementary material is in the code.

This is the author accepted manuscript.

and traversal. Autoencoder-based models (Wiesmann et al. 2021; He et al. 2022; Huang et al. 2022; Wang et al. 2024) leverage learned latent spaces and refined loss functions, yet remain limited by their dependence on global statistics, which constrain their ability to capture local geometric details and semantic diversity within scenes.

Recently, semantic-guided compression has emerged by partitioning point clouds into categories such as foreground and background, applying class-specific or selective encoding (Sun et al. 2020; Varischio et al. 2021; Song et al. 2021; Sun and Luo 2023). While this marks a shift beyond purely geometric encoding, most existing methods rely on rigid category partitions and treat all points within each category as equally important. This simplistic assumption overlooks that within the so-called "important" foreground, objects vary significantly in relevance. Moreover, these approaches typically assume that the semantic or feature information used for guiding compression is accurate and complete, implicitly neglecting the uncertainty inherent in dynamic scenes. This kind of uncertainty arises because models have limited knowledge about dynamic objects and their changing appearances under diverse environmental conditions. Ignoring such uncertainty leads to suboptimal bit allocation and restricts compression efficiency.

Our review of existing methods reveals two primary challenges. First, semantic integration is still basic and fragmented, i.e., many approaches adopt coarse binary segmentations that fail to distinguish critical objects like vehicles and pedestrians, or rely on multi-model ensembles that increase training complexity and runtime overhead. Neither strategy enables unified fine-grained semantic prioritization for adaptive bit allocation. Furthermore, these frameworks implicitly assume that the input data and features are complete and reliable, thereby neglecting the uncertainty of model when dealing with dynamic or partially observed inputs. Second, existing feature extraction paradigms suffer from a lack of true receptive field adaptivity. While many networks employ multi-scale hierarchies to process information at various scales, the receptive field at each level is determined by handcrafted, static kernel sizes. This rigid structure prevents the dynamic adjustment of the information aggregation scope in response to real-time variations in scene content, thereby impairing the efficiency of spatiotemporal redundancy removal.Addressing these issues calls for novel compression frameworks that integrate fine-grained semantic awareness, uncertainty quantification, and adaptive context-sensitive feature extraction tailored for complex dynamic 3D scenes.

Motivated by these challenges, we introduce R²D-LPCC, a novel relevance-ranking guided, region-adaptive framework for dynamic LiDAR point cloud compression. Our main contributions include:

- We present R²D-LPCC, the first dynamic LiDAR point cloud compression framework that integrates relevance ranking with region-adaptive processing to efficiently prioritize semantically important and uncertain regions.
- We develop key adaptive modules, namely Adaptive Relevance Learning (ARL) for directing attentional resources to critical regions by learning their semantic context and uncertainty, Multi-Scale Region-Adaptive Transform (MRAT) for preserving fine details via relevance-guided feature aggregation, and Adaptive Multi-Modal Motion Estimation (AME) for robust motion alignment in dynamic scenes.
- Extensive evaluations on the SemanticKITTI benchmark demonstrate that R²D-LPCC consistently outperforms ten state-of-the-art methods, achieving a 45.48% BD-rate gain over the state-of-the-art Unicorn and enhanced reconstruction quality in critical semantic regions.

Related Work

Intra-LPCCs. Intra-LPCC aims to reduce spatial redundancy by encoding point clouds into compact forms. Range-image methods project 3D points into 2D planes (Sun et al. 2019; Wang et al. 2022; Sun et al. 2020; Zhao et al. 2022), but severely distort underlying geometry. More commonly, octree-based approaches perform hierarchical spatial partitioning and entropy coding, enhanced via contextual aggregation (Fu et al. 2022), hierarchical attention (Song et al. 2023a), multi-scale features (Fan et al. 2023), and sophisticated coding schemes (Jin et al. 2024). However, deep octree construction and traversal incur significant computation. Self-attention (Cui et al. 2023) and near/far partitioning (Güngördü and Tekalp 2024) offer further refinements. Alternatively, autoencoder-based techniques learn compact latent representations through variable-precision coding (Wiesmann et al. 2021), density preservation (He et al. 2022), block-wise schemes (Huang et al. 2022), multi-subset decoupling (Wang et al. 2024), and tiered designs (Xu, Zhang, and Wu 2024), including sparse tensor formulations (SparsePCGC (Wang et al. 2023), Unicorn (Wang et al. 2025)). Yet, these methods largely depend on global statistics or rigid partitioning and often neglect local geometric detail and semantic variance across regions. A key innovation of our method is Adaptive Relevance Learning (ARL), which combines semantic context with uncertainty. This enables dynamic prioritization of semantically critical and uncertain regions during compression, effectively addressing prior limitations.

Inter-LPCCs. Inter-LPCC methods fall into three categories: video-based, octree-based, and motion-based approaches. Video-based techniques perform frame prediction using uni- or bi-prediction to estimate subsequent point cloud frames and compress residuals (Tu et al. 2019; Zhao et al. 2022; Liu et al. 2023). However, these methods convert point clouds into 2D range images, which limits their ability to fully exploit 3D temporal redundancies. Octree-based methods enhance intra-frame coding by incorporating multi-frame context into the entropy model through sequential convolutions (Biswas et al. 2020), edge convolutions (Song et al. 2023b), or pose-based frame alignment (Que, Lu, and Xu 2021). Despite these advances, they predominantly process individual octree nodes independently, thus focusing more on spatial redundancy reduction than on temporal correlations. Motion-based methods encode explicit motion between frames. Traditional algorithms such as

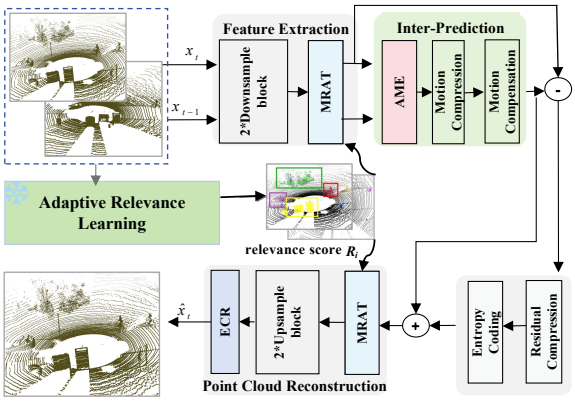

Figure 2: Architecture overview of R^{2}D-LPCC. Input frames (X_{t-1}, X_t), reconstruction $\hat{X}_t$. Components: Adaptive Relevance Learning (ARL), Multi-Scale Region Adaptive Transform (MRAT), Adaptive Motion Estimation (AME), Enhanced Coordinate Refinement (ECR).

ICP (Sun et al. 2021) or those leveraging sensor information (Feng, Liu, and Zhu 2020) often struggle with non-rigid or deformable objects. Recent deep learning approaches (Fan et al. 2022; Jiang et al. 2023) attempt to improve motion estimation but generally depend on fixed convolutional kernels, which lack adaptability for the irregular and sparse nature of point clouds. Our approach introduces an Adaptive Motion Estimation (AME) module that dynamically modulates motion prediction by integrating local self-attention and content-adaptive convolutional kernels, thus better capturing complex motion patterns in 3D scenes and overcoming the expressiveness limitations of prior methods.

Feature-Aware Compression. Recently, feature-aware compression has gained attention for improving point cloud compression by leveraging semantic and contextual information. Semantic-based methods (Sun et al. 2020; Wang et al. 2022; Zhao et al. 2022) typically perform semantic segmentation on range images to assign per-pixel labels, which are then used to predict the original range data. However, their performance heavily depends on the accuracy of the segmentation and prediction models, which can limit robustness under varying conditions. Selective-coding strategies (Varischio et al. 2021; Sun et al. 2022) enhance task-specific outcomes by removing less informative points, yet this pruning may disrupt feature continuity and harm downstream task. Class-specific coding schemes (Song et al. 2021; Wang and Liu 2022; Sun and Luo 2023) divide the point cloud into semantic classes and apply specialized encoders per category, improving compression for important classes but increasing model complexity significantly. Using semantic cues, these approaches mainly emphasize partitioning the data without evaluating the varying significance of different regions for compression prioritization. Based on ARL, we introduce the Multi-scale Region-Adaptive Transform (MRAT) module, which leverages a single network architecture to adaptively process diverse regions. This approach enables the network to enhance semantic feature modeling

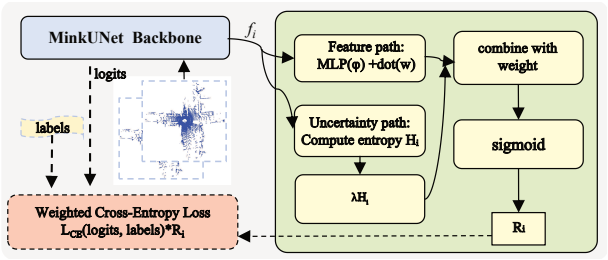

Figure 3: The **ARL module**. Computes the relevance score R_i for each point by jointly evaluating feature characteristics and predictive uncertainty. Semantic segmentation labels, shown as *labels*, are used solely for supervision during training to guide relevance learning. Logits are the raw prediction outputs from the backbone network.

while preserving fine-grained details in critical areas.

Building on these insights, our new framework redefines feature-aware compression through a principled relevance-ranking mechanism that dynamically prioritizes critical regions for optimized bit allocation. We also enhance semantic feature extraction by contextually modulating processing according to relevance, ensuring preservation of key details. By quantitatively assessing and leveraging regional importance beyond simple semantic cues, this approach effectively addresses prior limitations, delivering superior compression efficiency and reconstruction fidelity.

Methodology

Framework Overview

Our framework (Fig. 2) advances dynamic LiDAR point cloud compression by enabling context-aware prioritization of critical scene elements. Unlike traditional methods that treat all regions equally, it leverages semantic context and predictive uncertainty to focus compression resources on safety-critical objects like vehicles and pedestrians.

The approach includes Adaptive Relevance Learning (ARL), which ranks regions by semantic importance and uncertainty to guide bit allocation; the Multi-Scale Region-Adaptive Transform (MRAT), which uses relevance scores to enhance multi-scale feature extraction and preserve detail in key areas; Adaptive Multi-Modal Motion Estimation (AME), which integrates spatial-temporal and semantic cues from consecutive frames for robust motion prediction; and Enhanced Coordinate Refinement (ECR), which corrects coordinate drift to maintain geometric fidelity.

Given two consecutive point cloud frames, $X_{t-1} = \{C_{X_{t-1}}, F_{X_{t-1}}\}$ and $X_t = \{C_{X_t}, F_{X_t}\}$, where C and F represent coordinates and occupancy-based features, the R^{2}D-LPCC pipeline integrates these modules to deliver an efficient, high-fidelity, and semantically aware compression solution. We discuss each component below.

Adaptive Relevance Learning (ARL)

The Adaptive Relevance Learning (ARL) module quantifies the importance of each point in a dynamic LiDAR frame by combining semantic context and predictive uncertainty.

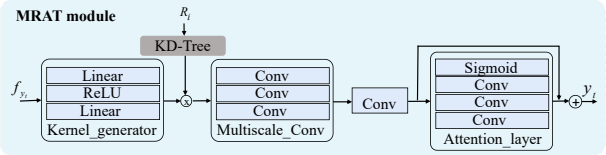

Figure 4: **MRAT module.** Multi-scale features are aggregated using convolution kernels dynamically generated and modulated by local features and region relevance scores.

Semantic context assigns high-level class labels (e.g., vehicles, pedestrians) to points, inferred from a semantic segmentation network, while predictive uncertainty measures the model's confidence. Uncertainty typically increases at object boundaries, ambiguous, or dynamic regions. ARL integrates these cues to compute continuous relevance scores, directing compression to focus on points that are both semantically important and uncertain.

Given a point cloud frame $X_t = \{C_{X_t}, F_{X_t}\}$, a sparse U-Net backbone extracts feature vectors $\mathbf{f}_i$ for each point. The relevance score for point i is

$$R_{t,i} = \frac{\exp\left(\mathbf{w}^\top \phi(\mathbf{f}_i) - \lambda H(\mathbf{f}_i)\right)}{\sum_j \exp\left(\mathbf{w}^\top \phi(\mathbf{f}_j) - \lambda H(\mathbf{f}_j)\right)}$$

where $\phi(\cdot)$ is a learnable semantic embedding function (MLP) encoding semantic context; $\mathbf{w}$ is a learnable attention vector identifying features correlated with critical classes; $H(\mathbf{f}_i) = -\sum_c p(c|\mathbf{x}_i) \log p(c|\mathbf{x}_i)$ is the predictive entropy quantifying per-point semantic uncertainty over classes c; and λ balances semantic context and uncertainty. The denominator normalizes scores so that $\sum_i R_{t,i} = 1$.

Since explicit ground truth for relevance is unavailable, ARL adopts a meta-learning paradigm (Ren et al. 2018) in which the computed relevance scores $R_{t,i}$ serve as adaptive weights in the loss function of a proxy task, semantic segmentation, which provides per-point class predictions and associated uncertainty estimates:

$$L_{\text{total},t} = \sum_{i=1}^{N} (N R_{t,i})\, L_{\text{task},i}$$

where N is the total number of points and $L_{\text{task},i}$ is the loss for point i. End-to-end training enables the model to learn scene-adaptive relevance scores that guide compression for efficient bit allocation and robust reconstruction, particularly in complex and dynamic scenes.

Multi-Scale Region Adaptive Transform (MRAT)

Standard convolutions process semantically diverse LiDAR point clouds uniformly, limiting focus on safety-critical regions like vehicles and pedestrians. This leads to suboptimal features in complex scenes with varied semantics and uncertainty. To address this, the Multi-Scale Region-Adaptive Transform (MRAT) dynamically adapts feature aggregation using pre-computed per-point relevance scores that capture semantic importance and uncertainty, combined with local geometry via a learnable kernel generation network.

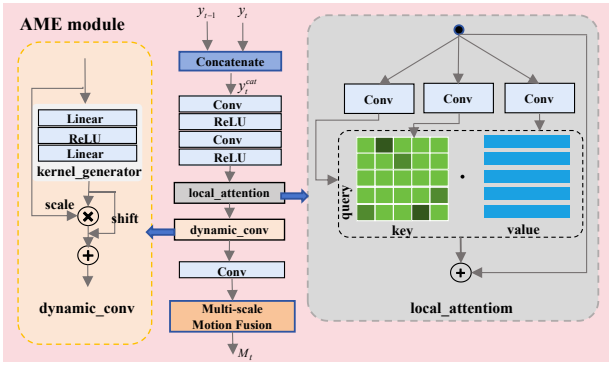

Figure 5: **AME module.** Each point utilizes its local features and region relevance score to dynamically generate kernels, enabling adaptive multi-scale aggregation and channel attention for context- and priority-aware feature extraction.

Each point $p_i \in C_{X_t}$ is assigned the relevance score $R_{t,i}$ of its nearest neighbor via a k-d tree, converting high-level semantic cues into point-wise modulation signals. A kernel generation sub-network (two fully connected layers) takes local features $\mathbf{f}_{y_t,i}$ and outputs normalized attention weights:

$$K_i' = \text{Softmax}\big(\Psi_{\text{kernel}}(\mathbf{f}_{y_t,i})\big)$$

These weights capture local geometric relations and are modulated by the relevance score and a global learnable parameter α:

$$K_i = \alpha \cdot R_{t,i} \cdot K_i'$$

This ensures feature aggregation is sensitive to both semantic relevance and local geometry.

To capture multi-scale context, MRAT uses three parallel Minkowski convolution branches with kernels of sizes 3, 5, and 7. Their outputs are fused by element-wise addition and a 1×1 convolution to form feature vector F_{fused}. A Channel Attention Module then recalibrates features:

$$y_t = F_{\text{fused}} \otimes \sigma\big(W_2\, \delta(W_1 F_{\text{fused}})\big)$$

where W_1, W_2 are 1×1 convolutions, δ is ReLU, σ is Sigmoid, and $\otimes$ denotes channel-wise multiplication. This final operation enhances informative channels while suppressing noise, yielding the refined output feature matrix y_t.

Adaptive Motion Estimation (AME)

The Adaptive Motion Estimation (AME) module enhances motion prediction for dynamic LiDAR point cloud compression by making the process content-aware and sensitive to the semantic relevance of different scene regions. Unlike traditional methods that apply uniform processing, AME adaptively aligns motion estimation with both geometric complexity and the importance of scene content. This enables more accurate and efficient motion prediction in heterogeneous and dynamic environments.

To achieve this, AME first concatenates feature representations from the current frame y_t and the previous frame y_{t-1} into joint spatio-temporal features y_t^{cat}, which are then fused through convolution. A local self-attention mechanism

Method	bpp1	PSNR	bpp2	PSNR	bpp3	PSNR	bpp4	PSNR	BD-Rate Gain ↓	Enc(s)	Dec(s)
GPCC TMC14	2.749	63.51	3.553	65.80	8.79	82.20	11.84	88.31	Anchor	2.53	0.66
RIDDLE	0.150	43.09	0.691	50.59	2.647	64.96	3.653	70.91	-19.99%	-	-
RPCC	1.831	60.95	2.186	65.00	2.664	69.88	3.425	76.14	-34.45%	0.098	0.098
OctAttention	0.394	58.95	0.968	65.00	2.101	70.99	3.852	77.02	-69.54%	0.4	120
Octformer	0.398	58.95	0.972	65.00	2.141	70.99	3.822	77.02	-69.34%	-	-
EHEM	0.64	60.37	1.00	63.16	2.15	69.80	3.69	76.46	-60.20%	-	-
MSLPCC	0.031	47.81	0.115	49.28	0.215	50.07	0.312	50.34	1.40%	-	-
D-DPCC	0.61	52.02	0.99	55.69	1.41	57.06	2.07	57.50	90.84%	1.5	0.17
Unicorn	0.71	46.96	1.299	63.18	2.216	69.46	3.785	77.06	-53.10%	-	-
RENO	0.71	58.85	1.65	65.24	2.72	69.89	5.00	77.13	-50.24%	**0.095**	**0.09**
SparsePCGC	0.395	58.95	0.939	65.00	2.016	70.99	3.694	77.02	-70.47%	1.47	1.07
Ours_w/o_ECR	0.398	67.03	1.01	69.55	2.04	74.71	3.65	78.58	-68.10%	2.96	0.17
Ours	**0.398**	**82.56**	**1.00**	**89.68**	**2.46**	**95.76**	**3.19**	**97.66**	**-98.58**%	2.96	0.17

Table 1: Comparison results of state-of-the-art representative LPCCs. The BD-Rate gain is calculated using GPCC as the anchor.

aggregates contextual relationships within each neighborhood $\mathcal{N}(i)$, producing context-enriched features $y_t'^{\text{cat}}$ that incorporate motion cues and environmental dependencies.

Adaptivity is further enhanced via a dynamic convolutional layer whose kernel generation sub-network Ψ_{kernel} predicts a unique channel-wise scaling vector $\mathbf{s}_i$ and bias $\mathbf{b}_i$ for each point based on its local features:

$$[\mathbf{s}_i, \mathbf{b}_i] = \Psi_{\text{kernel}}(y_t'^{\text{cat}})$$

These parameters define an affine transformation applied point-wise:

$$y_t''^{\text{cat}} = \mathbf{s}_i \odot y_t'^{\text{cat}} + \mathbf{b}_i,$$

allowing fine-grained adjustment of feature representations tailored to local motion characteristics and scene complexity, surpassing the rigidity of static, shared kernels.

Finally, the adaptively modulated features are decoded into the final motion vector field M_t through a multi-scale fusion head that integrates spatial, temporal, and contextual information at multiple scales. This comprehensive approach enables AME to deliver robust, expressive, and region-sensitive motion estimation suited for complex and dynamic 3D scenes.

Enhanced Coordinate Refinement (ECR)

Sparse convolutions in point cloud compression often lead to cumulative errors such as coordinate shifts from discretization and downsampling, degrading reconstruction accuracy. The Enhanced Coordinate Refinement (ECR) module addresses this by explicitly correcting these errors using a coordinate offset correction strategy. ECR employs a center offset loss to align the global geometric centers of original and reconstructed point clouds, a weighted Chamfer distance loss to preserve important local structures, and an offset regularization loss to prevent excessive corrections and ensure stable reconstruction. The combined training loss is:

$$\mathcal{L} = \lambda_1 + \lambda_2 \cdot \text{bpp} + \lambda_3 \cdot \mathcal{L}_{\text{center}} + \lambda_4 \cdot \mathcal{L}_{\text{Chamfer}} + \lambda_5 \cdot \mathcal{L}_{\text{offset}-\text{reg}},$$

coefficients $\lambda_3, \lambda_4, \lambda_5$ weight each refinement component.

By explicitly correcting coordinate drift across scales while enforcing stability, ECR significantly improves the spatial accuracy and fidelity of reconstructed point clouds, which is critical for applications such as autonomous driving and robotic perception.

Experiments and Results

Experimental Setup

Datasets: Our experimental evaluation leverages several large-scale autonomous driving benchmarks. The primary dataset is SemanticKITTI (Behley et al. 2019), which we use both for training and for assessing our model's compression performance. Following established conventions, we train on the ten official sequences (00–07, 09–10) and use the remaining sequences (08, 11–21) for testing, ensuring a strict separation between train and test data. To further demonstrate the generalization ability of our approach, we conduct additional evaluations on two widely recognized datasets, nuScenes (Caesar et al. 2020) and Ford (Pandey, McBride, and Eustice 2011). Additionally, to highlight the practical effectiveness of our compression algorithm in real-world perception scenarios, we measure its performance impact on the KITTI 3D object detection benchmark.

Training Strategy: Model training is carried out on an RTX4090 GPU using the Adam optimizer. We adopt $\beta_1 = 0.9$ and $\beta_2 = 0.999$ for the optimizer's momentum terms. Training is performed with a batch size of 1 and an initial learning rate set at 0.008.

Evaluation Metrics: We closely adhere to the Common Test Condition (CTC) as specified in the MPEG G-PCC and V-PCC standardization processes. For comprehensive assessment, we report compression performance using bits per point (bpp), and evaluate point cloud reconstruction quality with D1_PSNR and D2_PSNR scores (Fig.5 supppl).

Comparison Methods: We compare our approach with a broad selection of state-of-the-art (SOTA) methods, including GPCC TMC14 (Schwarz et al. 2018), RPCC (Wang et al. 2022), RIDDLE (Zhou et al. 2022), OctAttention (Fu et al. 2022), Octformer (Cui et al. 2023), EHEM (Song et al. 2023a), Sparse-PCGC (Wang et al. 2023), D-DPCC (Fan et al. 2022), MSLPCC (Wang et al. 2024), RENO (You et al. 2025), and Unicorn (Wang et al. 2025), thereby covering a variety of range-image, tree, point, and sparse tensor-based techniques. To ensure fairness in our comparisons, we standardize the experimental conditions and reconstruct key methods using official public code where available. Specifically, we acquired the code for GPCC TMC14 (Schwarz et al. 2018), RPCC (Wang et al. 2022), OctAttention (Fu

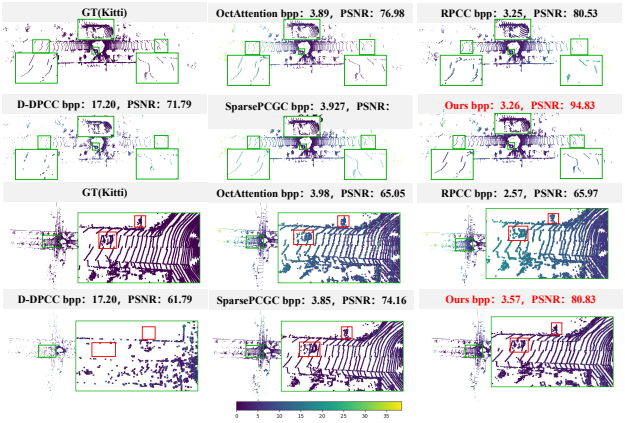

Figure 6: visualized compression results on SemanticKITTI, with points color-coded by error levels (lighter indicating greater discrepancy from ground truth).

et al. 2022), Sparse-PCGC (Wang et al. 2023), D-DPCC (Fan et al. 2022), and RENO (You et al. 2025), and retrained these models on SemanticKITTI using the same data splits and protocols as R²D-LPCC. In cases where official code could not be obtained, we instead reference the results as reported in the original papers. Recognizing that encoding and decoding runtimes are highly sensitive to hardware and software variability, we report codec runtimes only for those models whose results we could reliably reproduce.

Compression Performance

The results in Table 1 demonstrate that our model achieves SOTA reconstruction quality over the full bit-rate spectrum, yielding substantial BD-rate savings of 98.58% and 45.48% compared to GPCC and Unicorn, respectively. This performance is not simply a consequence of combining modules; rather, it results from directly addressing core bottlenecks of previous approaches. In contrast to methods that rely on distortion-prone projections (e.g., RPCC) or incur considerable processing time with octree-based structures (e.g., OctAttention), our model delivers a superior rate-distortion trade-off. This is primarily enabled by the MRAT and AME modules. MRAT, guided by relevance scores, generates fine-grained kernels for complex areas such as vehicles to enhance detail capture, while simpler areas like ground regions are efficiently processed using kernels with large receptive fields. This content-adaptive feature extraction ensures that bits are allocated preferentially to encode information critical for accurate reconstruction, resulting in lower distortion at the same bit-rate.

Fig. 6 provides a visual comparison illustrating the detailed reconstruction capability of our approach. Existing methods such as SparsePCGC and OctAttention, which employ uniform receptive fields, are unable to retain delicate structures during downsampling, as indicated in the green box. In contrast, our model maintains high fidelity in key regions, highlighted in the red box. This improvement stems from the combined operation of the ARL and MRAT modules: ARL first discerns high-importance regions, then

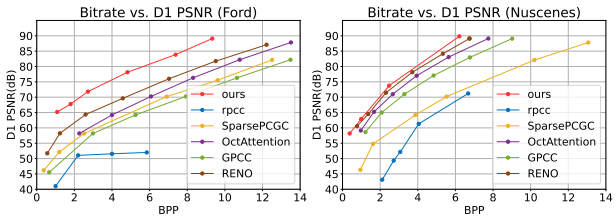

Figure 7: R-d performance on Ford and Nusecnes.

MRAT generates kernels tailored to local geometric characteristics. This cooperative mechanism allows our model to precisely preserve vehicle contour details that would otherwise be lost.

As depicted in Fig. 7, our method consistently achieves optimal Rate-distortion (R-d) trade-offs on both the Ford and nuScenes datasets. This strong cross-dataset generalization demonstrates that the model learns a fundamental, adaptive processing strategy rather than overfitting to any specific data distribution. The MRAT and AME modules together empower the network to dynamically adjust processing according to the underlying content, enabling our approach to robustly accommodate diverse sensor data and scene variations while maintaining SOTA performance.

Ablation Studies

To rigorously assess the contribution of each core module in our framework, we systematically compare the compression performance of different component combinations, as summarized in Table 2. The baseline (B) is D-DPCC. The analysis below clarifies how each component targets distinct technical challenges to improve compression efficacy.

Model Variant	bpp1	PSNR	bpp2	PSNR2	enc_time(s)
B	0.99	55.69	2.07	57.50	1.500
B+A	0.99	56.58	1.67	57.53	-
B+A+R	0.94	65.07	1.98	67.03	1.586
B+A+M	0.99	57.94	2.07	58.30	2.874
B+A+R+M	1.01	69.55	2.04	74.71	2.960
B+A+R+M+O	0.93	71.98	2.09	77.65	2.961
B+A+R+M+E	1.00	**89.68**	1.99	**94.20**	2.961

Table 2: Performance for different model variants.

- **ARL (A):** This module addresses the baseline's content-agnostic bottleneck by ranking relevance using semantic context and uncertainty. Semantic context identifies what matters (e.g., vehicles), while uncertainty pinpoints where compression is hardest (e.g., contours). This dual guidance enables smarter bit allocation, yielding an overall gain of +0.89 dB (bpp1). Fig. 8 exemplifies how gains come from prioritizing objects over ground points. Ground points are easy to reconstruct but contribute little to PSNR. Errors in sparse, meaningful regions like vehicles hurt PSNR more due to their shape and semantic importance. Our method preserves vehicle structure, while the baseline often misses these details (Fig. 3 suppl).

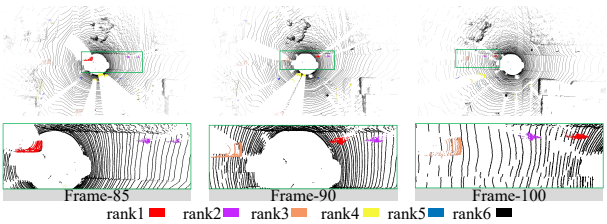

Figure 8: Relevance ranking visualization results.

- **MRAT (R):** MRAT mitigates the rigidity of fixed receptive fields by enabling "geometry-adaptive" feature extraction. With guidance from ARL, MRAT generates optimized convolutional kernels for different spatial regions, which greatly reduces redundancy in the feature space. This targeted adaptation leads to a notable PSNR gain of +8.49dB (bpp1) over the B+M model, indicating more effective capture of com-plex structures.
- **AME (M):** Extending adaptiveness into the temporal domain, AME introduces advanced motion estimation for sequences and lowers inter-frame prediction errors. Alone, AME offers a +1.36dB gain (B+M+A vs. B+M), but demonstrates synergy with MRAT, with a combined gain of +4.48dB(bpp1) (B+M+R+A vs. B+M+R), highlighting its benefit in handling temporal variability.
- **Boundary offset compensation (Offset (O)):** Adding only the coordinate offset mechanism from (Wang et al. 2023, 2025) without supplementing it with the specialized loss does not achieve optimal performance, suggesting that loss design is critical for extracting the benefits of this module.
- **ECR(E):** Serving as a dedicated residual refinement network, the ECR module performs final adjustments to compensate for subtle coordinate errors introduced by quantization. With its specialized loss, it resolves remaining alignment issues, which are often limiting factors for SOTA precision. This refinement yields a substantial additional gain of +20.13 dB (bpp1), reinforcing its essential role in the framework.

We further measure the computational cost of each module in Table 2 (with ARL's parameters frozen during inference and thus omitted). This analysis reveals how cost is distributed across modules. The AME module is the most resource-intensive, contributing an additional 1.374s to the encoding time compared to the baseline. The MRAT module remains highly efficient, adding only 0.086s, while the ECR module's overhead is nearly negligible at just 0.001s. This demonstrates that ours suits offline tasks like high-fidelity map generation as quality matters more than speed.

Downstream Task Evaluation

As highlighted in (You et al. 2025), LiDAR point-cloud compression serves two essential purposes: reducing storage and transmission costs, and preserving rich spatial detail necessary for real-time information exchange between devices or vehicles to facilitate collaborative perception and decision-making. For compressed point clouds to be useful in downstream applications, such as 3D object detection, it is crucial that they retain sufficient fidelity so detector performance closely matches that on uncompressed inputs.

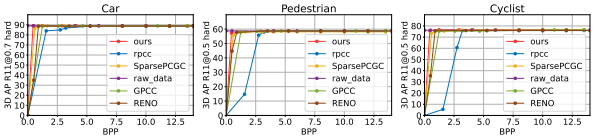

Figure 9: 3D object detection on the KITTI dataset.

To quantitatively evaluate this requirement, and following the protocol established in (You et al. 2025), we benchmark five representative encoders, namely GPCC TMC14, RPCC, SparsePCGC, D-DPCC, and Ours, by integrating their decoded point clouds into the OpenPCDet framework and assessing 3D detection performance using the pre-trained PV-RCNN (Bhattacharyya and Czarnecki 2020) model. Note that due to coordinate normalization in ActAttention decoding, an explicit inverse-normalization step is needed to enable downstream testing, which greatly increases the resulting bpp. For this reason, we exclude it from comparison.

As depicted in Fig. 9, the uncompressed raw point cloud delivers detection accuracies of about 89.07%, 58.73%, and 76.18% for cars, pedestrians, and bicycles, respectively. Our model consistently maintains high detection accuracy at very low bit-rates, outperforming both GPCC and RPCC. This benefit is directly enabled by targeted information retention in our compression framework: the ARL module first pinpoints small, sparse targets such as "pedestrians" and "cyclists," while MRAT adaptively generates specialized convolution kernels to preserve their fine-grained geometry, supporting robust detection. By contrast, competing approaches often experience diminished detection performance due to indiscriminate quantization or geometric distortion, which primarily affects these difficult instances. Our design avoids such failure modes, resulting in stronger downstream utility.

Conclusion

Existing LiDAR compression methods treat all regions as equally important, causing inefficient encoding and poor reconstruction of critical areas like dynamic objects. We propose R²D-LPCC, a framework for dynamic point cloud compression, to address this. Our approach uses a relevance ranking system (ARL module) leveraging semantic context and uncertainty to identify important regions. This ranking then guides an adaptive feature extraction module (MRAT) and an accurate motion estimation module (AME) to preferentially encode these critical areas. Extensive experiments validate our method's superior reconstruction quality by focusing on what matters. While the method effectively balances performance and efficiency, future work will optimize the framework for real-time applications, especially for inter-frame motion estimation.

Acknowledgments

This work was supported in part by the National Natural Science Foundation (Grant No. 62172366), the Zhejiang Province Natural Science Foundation (Grant No. LD24F020003), and the European Union's Horizon 2024 Research and Innovation Programme for the Marie Skłodowska-Curie Actions (Grant No. 101211118). Gary Tam is supported by the Royal Society grant IEC/NSFC/211159. For the purpose of Open Access, the author has applied a CC BY copyright licence to any Author Accepted Manuscript version arising from this submission.

References

Behley, J.; Garbade, M.; Milioto, A.; Quenzel, J.; Behnke, S.; Stachniss, C.; and Gall, J. 2019. Semantickitti: A dataset for semantic scene understanding of lidar sequences. In *Proceedings of the IEEE/CVF international conference on computer vision*, 9297–9307.

Bhattacharyya, P.; and Czarnecki, K. 2020. Deformable PV-RCNN: Improving 3D object detection with learned deformations. *arXiv preprint arXiv:2008.08766*.

Biswas, S.; Liu, J.; Wong, K.; Wang, S.; and Urtasun, R. 2020. Muscle: Multi sweep compression of lidar using deep entropy models. *Advances in Neural Information Processing Systems*, 33: 22170–22181.

Caesar, H.; Bankiti, V.; Lang, A. H.; Vora, S.; Liong, V. E.; Xu, Q.; Krishnan, A.; Pan, Y.; Baldan, G.; and Beijbom, O. 2020. nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11621–11631.

Cui, M.; Long, J.; Feng, M.; Li, B.; and Kai, H. 2023. OctFormer: Efficient Octree-Based Transformer for Point Cloud Compression with Local Enhancement. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(1): 470–478.

Fan, T.; Gao, L.; Xu, Y.; Li, Z.; and Wang, D. 2022. D-DPCC: Deep Dynamic Point Cloud Compression via 3D Motion Prediction. arXiv:2205.01135.

Fan, T.; Gao, L.; Xu, Y.; Wang, D.; and Li, Z. 2023. Multiscale latent-guided entropy model for lidar point cloud compression. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(12): 7857–7869.

Feng, Y.; Liu, S.; and Zhu, Y. 2020. Real-Time Spatio-Temporal LiDAR Point Cloud Compression. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 10766–10773. IEEE.

Fu, C.; Li, G.; Song, R.; Gao, W.; and Liu, S. 2022. OctAttention: Octree-Based Large-Scale Contexts Model for Point Cloud Compression. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36: 625–633.

Güngördü, O.; and Tekalp, A. M. 2024. Saliency-Aware End-to-End Learned Variable-Bitrate 360-Degree Image Compression. In *2024 IEEE International Conference on Image Processing (ICIP)*, 1795–1801. IEEE.

He, Y.; Ren, X.; Tang, D.; Zhang, Y.; Xue, X.; and Fu, Y. 2022. Density-preserving deep point cloud compression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2333–2342.

Huang, L.; Wang, S.; Wong, K.; Liu, J.; and Urtasun, R. 2020. Octsqueeze: Octree-structured entropy model for lidar compression. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1313–1323.

Huang, T.; Zhang, J.; Chen, J.; Ding, Z.; Tai, Y.; Zhang, Z.; Wang, C.; and Liu, Y. 2022. 3qnet: 3d point cloud geometry quantization compression network. *ACM Transactions on Graphics (TOG)*, 41(6): 1–13.

Jiang, Z.; Wang, G.; Tam, G. K.; Song, C.; Li, F. W.; and Yang, B. 2023. An end-to-end dynamic point cloud geometry compression in latent space. *Displays*, 80: 102528.

Jin, Y.; Zhu, Z.; Xu, T.; Lin, Y.; and Wang, Y. 2024. ECM-OPCC: Efficient Context Model for Octree-based Point Cloud Compression. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 7985–7989. IEEE.

Liu, C.-S.; Yeh, J.-F.; Hsu, H.; Su, H.-T.; Lee, M.-S.; and Hsu, W. H. 2023. BIRD-PCC: Bi-Directional Range Image-Based Deep Lidar Point Cloud Compression. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. IEEE.

Pandey, G.; McBride, J. R.; and Eustice, R. M. 2011. Ford campus vision and lidar data set. *The International Journal of Robotics Research*, 30(13): 1543–1552.

Que, Z.; Lu, G.; and Xu, D. 2021. Voxelcontext-net: An octree based framework for point cloud compression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6042–6051.

Ren, M.; Zeng, W.; Yang, B.; and Urtasun, R. 2018. Learning to reweight examples for robust deep learning. In *International conference on machine learning*, 4334–4343. PMLR.

Schwarz, S.; Preda, M.; Baroncini, V.; Budagavi, M.; Cesar, P.; Chou, P. A.; Cohen, R. A.; Krivokuća, M.; Lasserre, S.; Li, Z.; et al. 2018. Emerging MPEG standards for point cloud compression. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 9(1): 133–148.

Song, F.; Shao, Y.; Gao, W.; Wang, H.; and Li, T. 2021. Layer-wise geometry aggregation framework for lossless LiDAR point cloud compression. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(12): 4603–4616.

Song, R.; Fu, C.; Liu, S.; and Li, G. 2023a. Efficient hierarchical entropy model for learned point cloud compression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 14368–14377.

Song, R.; Fu, C.; Liu, S.; and Li, G. 2023b. Large-scale spatio-temporal attention based entropy model for point cloud compression. In *2023 IEEE International Conference on Multimedia and Expo (ICME)*, 2003–2008. IEEE.

Sun, X.; and Luo, Q. 2023. Density-Based Geometry Compression for LiDAR Point Clouds. In *EDBT*, 378–390.

Sun, X.; Ma, H.; Sun, Y.; and Liu, M. 2019. A novel point cloud compression algorithm based on clustering. *IEEE Robotics and Automation Letters*, 4(2): 2132–2139.

Sun, X.; Sun, Y.; Zuo, W.; Cheng, S. S.; and Liu, M. 2021. A novel coding scheme for large-scale point cloud sequences

based on clustering and registration. *IEEE Transactions on Automation Science and Engineering*, 19(3): 2384–2396.

Sun, X.; Wang, M.; Du, J.; Sun, Y.; Cheng, S. S.; and Xie, W. 2022. A task-driven scene-aware LiDAR point cloud coding framework for autonomous vehicles. *IEEE Transactions on Industrial Informatics*, 19(8): 8731–8742.

Sun, X.; Wang, S.; Wang, M.; Wang, Z.; and Liu, M. 2020. A novel coding architecture for lidar point cloud sequence. *IEEE Robotics and Automation Letters*, 5(4): 5637–5644.

Tu, C.; Takeuchi, E.; Carballo, A.; and Takeda, K. 2019. Real-time streaming point cloud compression for 3d lidar sensor using u-net. *IEEE Access*, 7: 113616–113625.

Varischio, A.; Mandruzzato, F.; Bullo, M.; Giordani, M.; Testolina, P.; and Zorzi, M. 2021. Hybrid point cloud semantic compression for automotive sensors: A performance evaluation. In *ICC 2021-IEEE International Conference on Communications*, 1–6. IEEE.

Wang, J.; Ding, D.; Li, Z.; Feng, X.; Cao, C.; and Ma, Z. 2023. Sparse Tensor-Based Multiscale Representation for Point Cloud Geometry Compression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(7): 9055–9071.

Wang, J.; Xue, R.; Li, J.; Ding, D.; Lin, Y.; and Ma, Z. 2025. A Versatile Point Cloud Compressor Using Universal Multiscale Conditional Coding – Part I: Geometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(1): 269–287.

Wang, M.; Huang, R.; Dong, H.; Lin, D.; Song, Y.; and Xie, W. 2024. msLPCC: A Multimodal-Driven Scalable Framework for Deep LiDAR Point Cloud Compression. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 5526–5534.

Wang, S.; Jiao, J.; Cai, P.; and Wang, L. 2022. R-PCC: A baseline for range image-based point cloud compression. In *2022 International Conference on Robotics and Automation (ICRA)*, 10055–10061. Philadelphia, PA: IEEE.

Wang, S.; and Liu, M. 2022. Point cloud compression with range image-based entropy model for autonomous driving. In *European Conference on Computer Vision*, 323–340. Springer.

Wiesmann, L.; Milioto, A.; Chen, X.; Stachniss, C.; and Behley, J. 2021. Deep compression for dense point cloud maps. *IEEE Robotics and Automation Letters*, 6(2): 2060–2067.

Xu, H.; Zhang, X.; and Wu, X. 2024. Fast Point Cloud Geometry Compression with Context-based Residual Coding and INR-based Refinement. In *European Conference on Computer Vision*, 270–288. Springer.

You, K.; Chen, T.; Ding, D.; Asif, M. S.; and Ma, Z. 2025. Reno: Real-time neural compression for 3d lidar point clouds. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 22172–22181.

Zhao, L.; Ma, K.-K.; Lin, X.; Wang, W.; and Chen, J. 2022. Real-time LiDAR point cloud compression using bi-directional prediction and range-adaptive floating-point coding. *IEEE transactions on broadcasting*, 68(3): 620–635.

Zhou, X.; Qi, C. R.; Zhou, Y.; and Anguelov, D. 2022. RIDDLE: Lidar Data Compression with Range Image Deep Delta Encoding. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 17191–17200. IEEE.