Received 12 July 2025; accepted 25 July 2025. Date of publication 31 July 2025; date of current version 19 August 2025.

Digital Object Identifier 10.1109/OJCOMS.2025.3594545

JSCC-Aided INR for High-Frequency Detail Preservation in LiDAR

AKIHIRO KUWABARA [1], SORACHI KATO [1], TOSHIAKI KOIKE-AKINO [2] (Senior Member, IEEE), AND TAKUYA FUJIHASHI [1] (Member, IEEE)

[1]Graduate School of Information Science and Technology, The University of Osaka, Suita 565-0871, Japan
[2]Mitsubishi Electric Research Laboratories, Cambridge, MA 02139, USA

CORRESPONDING AUTHOR: A. KUWABARA (e-mail: kuwabara.akihiro@ist.osaka-u.ac.jp)

This work was supported in part by JST·ASPIRE under Grant JPMJAP2432, and in part by JSPS KAKENHI under Grant JP22H03582.

ABSTRACT Light Detection and Ranging (LiDAR) sensors generate accurate 3D representations of real-world environments, which are essential for applications of 3D scene understanding. However, the substantial volume of LiDAR data poses significant challenges for efficient compression and transmission. Implicit neural representation (INR) has gained attention for its compact data representation, but its capacity to accurately represent high-frequency details is insufficient when using small models. In this paper, we propose a novel joint source-channel coding (JSCC) scheme that integrates INR with analog residual transmission for high-quality and efficient point cloud transmission. This scheme is designed to compensate for the limited high-frequency representation of INRs by transmitting the unmodeled details as residuals via pseudo-analog modulation. This integrated approach enables continuous reconstruction quality adaptation to varying wireless channel conditions and effectively mitigates the stair-case effect inherent in conventional digital schemes. Evaluations on the KITTI dataset demonstrate that the proposed scheme outperforms conventional and INR-based compression methods in terms of R-D performance and detection quality at low bitrates.

INDEX TERMS LiDAR, point clouds, joint source-channel coding, pseudo-analog transmission, implicit neural representation.

I. INTRODUCTION

LIGHT Detection and Ranging (LiDAR) sensors enable accurate three-dimensional (3D) mapping of the surrounding environment by emitting laser pulses and measuring reflected signals. The resulting 3D point cloud data play a critical role in various applications, such as digital archiving, remote spatial sharing, and the development of digital twins [1], [2], [3], [4]. However, with the advancement of LiDAR sensor resolution, the amount of data generated per scan has grown significantly, making efficient compression and transmission essential for practical deployment [5].

Conventional compression methods for LiDAR point clouds can be broadly categorized into geometry-based approaches, which voxelize or partition the 3D space hierarchically and assign bits to voxelized space [6], [7], [8], and projection-based approaches, which convert 3D point clouds into two-dimensional (2D) range image (RI) for image-based processing [9], [10]. The RI-based methods have gained attention as an effective way to reduce the structural complexity of 3D point clouds, thereby facilitating efficient compression and representation.

In view of compression, implicit neural representation (INR)-based compression [11] has emerged as a promising technology for compactly representing continuous spatial signals. INR [12], [13], [14], [15] represents a signal as a coordinate-to-value mapping using a small neural network, achieving high compression ratios with a limited number of parameters. Recent studies [16], [17] have used INR for LiDAR point cloud compression and have shown that it can reduce transmission traffic while achieving higher reconstruction accuracy than conventional geometry-based and RI-based methods. While INR-based methods require significant encoding time, they are efficient at the decoding stage. Fig. 1 shows the trade-off between decoding latency

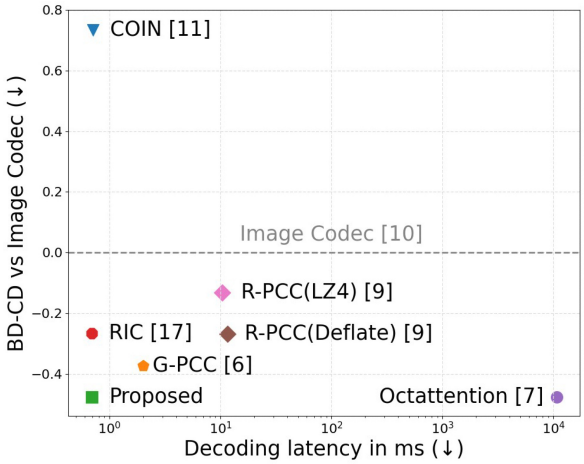


FIGURE 1. Rate-distortion performance (BD-CD) vs. decoding latency on KITTI dataset.

and rate-distortion (R-D) performance (BD-CD [18]). It suggests that INR-based approaches realize consistently low decoding latency despite variations in R-D performance.

However, one of the key issues in INR-based compression is its limited capacity to represent high-frequency components under a small model. This limitation often leads to a loss of fine structural details and degrades the performance of downstream tasks such as 3D object detection. Prior studies [13], [19] have proposed enhanced encoding schemes to mitigate this issue, but the expressive capacity of compact networks remains insufficient for capturing fine-grained detail.

To address this limitation without significantly increasing model complexity, we aim to integrate a power of communication with INR-based compression to compensate the high-frequency components that INR fails to model. Specifically, we introduce residual-aided transmission which is inspired by model-based compression [9], [20], [21]. These residuals represent high-frequency components not modeled by the INR, and are typically quantized, converted to binary, channel encoded, and modulated for wireless transmission to improve reconstruction fidelity. However, such digital schemes generally rely on fixed quantization levels and modulation formats, which do not adapt to time-varying wireless channel conditions. As a result, reconstruction quality improves only in discrete steps as channel conditions change, leading to the stair-case [22].

To solve the limitation of high-frequency components in a small INR and quality limitation of digital-based residual-aided transmission in time-varying wireless channels, we propose a novel scheme for efficient representation of LiDAR point clouds. It combines an INR-based digital LiDAR representation, RIC [17], with pseudo-analog residual transmission inspired by joint source–channel coding (JSCC). Specifically, the pseudo-analog modulation directly maps the residuals onto transmission symbols, so that the resulting reconstruction error scales smoothly with the instantaneous channel quality, i.e., gradual quality improvement under channel quality variation. In addition, the proposed scheme can control the ratio of digital and pseudo-analog symbols to maximize the R–D performance under the available bandwidth.

Evaluations on the KITTI dataset demonstrate that the proposed scheme enables gradual improvement in reconstruction quality under varying channel conditions, effectively mitigating the stair-case effect and preserving downstream task performance.

The major contributions of our study are three-fold:

- To the best of our knowledge, this is the first study to introduce a JSCC framework that incorporates INR-based representations for LiDAR point clouds, effectively addressing the fundamental limitation of modeling high-frequency components with compact networks.
- We design a residual communication scheme that enables smooth quality adaptation under varying channel conditions, mitigating the stair-case effect inherent in conventional digital approaches.
- We conduct extensive experiments on the KITTI dataset, covering both R-D performance and 3D object detection accuracy, to comprehensively evaluate the effectiveness of the proposed scheme.

II. RELATED WORK

A. POINT CLOUD COMPRESSION

LiDAR sensors capture 3D point clouds, where each point is defined by 3D coordinates, *i.e.*, (X, Y, Z). Compression methods for point clouds are categorized into 3D geometry-based approaches and 2D projection-based approaches using RIs.

Geometry-based compression approaches are typically divided into two types, known as graph-based and tree-based methods. The graph-based methods model point clouds as graph signals and apply the graph Fourier transform (GFT) to reduce redundancy in the spectral domain [23], [24], [25]. In addition, several studies have addressed graph signal reconstruction to reduce storage and transmission costs [26], [27]. In contrast, the tree-based compression methods structure point clouds by recursively subdividing the 3D space. A typical approach employs octree-based representations, such as point cloud library (PCL) and geometry-based point cloud compression (G-PCC) [6], [28]. Some recent studies have combined hierarchical tree structures with deep neural network (DNN) to further improve the efficiency of geometry compression [7], [29].

Projecting LiDAR measurements onto 2D RIs is a widely used technique to compactly represent spatial distance information. The RIs are typically generated either from raw LiDAR packets [30] or from 3D point clouds [9], [20], [21]. The obtained RIs are then compressed using intra-frame coding to reduce spatial redundancy [9], or inter-frame coding to exploit temporal coherence across frames [20], [21]. For

example, R-PCC [9] applies lossless compression, such as LZ4 and Deflate, to floating-point values.

Our study introduces INR-based compression for RIs to reduce storage and transmission costs for 3D point clouds. However, INRs often struggle to accurately represent fine-grained details when learning RIs. To address this problem, we define the residuals between the RI and the INR-based reconstruction. These residuals are then transmitted using pseudo-analog modulation, enabling the capture of high-fidelity details at low transmission costs.

B. IMPLICIT NEURAL COMPRESSION

INR is a technique that represents multi-dimensional signals, such as images and 3D point clouds, as continuous mappings from coordinates to signal values by overfitting them to a small neural network. A key limitation of INR is its insufficient accuracy in reconstructing high-frequency components. To address this issue, several methods have been proposed, including the use of sinusoidal activations in SIREN [13], positional encoding in NeRF [15], and its extension within the Neural Tangent Kernel framework [19]. Based on the results of the INR work, its application has been extended to image compression. In RI compression, a typical approach is to directly encode the entire RI using INR [11], [16], [31]. In contrast, RIC [17] improves coding efficiency by decomposing each RI into structurally distinct components and encoding them separately using dedicated INR networks. However, these methods suffer from insufficient representational capacity for high-frequency components. Furthermore, they remain susceptible to the stair-case effect caused by bit errors and irrecoverable quantization noise.

Our paper addresses these challenges by calculating residuals to compensate for high-frequency components. In addition, by directly mapping power-assigned residuals to transmission signals, we eliminate source and channel coding. As a result, the RI is reconstructed with high fidelity, adapting to instantaneous wireless channel conditions, and avoiding the stair-case effect.

C. JOINT SOURCE-CHANNEL CODING

Several JSCC schemes have been proposed to mitigate the stair-case effect caused by bit errors and to gradually improve the reconstruction quality of transmitted content according to the instantaneous wireless channel condition [22], [32], [33], [34]. These schemes eliminate quantization and entropy coding at the transmitter, and instead integrate decorrelation techniques, such as discrete cosine transform (DCT) [32] and discrete wavelet transform (DWT) [33], with pseudo-analog modulation to enable flexible adaptation to channel quality.

The signal processing-based and deep learning-based methods have been proposed to further improve the adaptability and compression efficiency of JSCC. Signal processing-based extensions include alternative decorrelation techniques using fixed or adaptive block divisions [35], [36], as well as error protection strategies tailored to channel conditions and downstream tasks [34], [37], [38]. In contrast, recent studies have introduced DNN-based architectures, leading to the development of deep JSCC [39], [40], [41], [42], [43], [44]. These approaches employ convolutional neural networks (CNNs) [40], transformer networks [41], [42], and graph neural networks (GNNs) [43], [44] to compress image and video signals into feature vectors, which are then directly mapped to pseudo-analog modulation formats for transmission.

However, image signals generally exhibit a wide dynamic range in pixel values, which causes a low reconstruction quality through pseudo-analog modulation in each channel condition. To address this issue, we compute and transmit the residuals using a pseudo-analog modulation format. Since the residual has a significantly smaller dynamic range compared to the RI, the proposed scheme can provide higher reconstruction quality and greater stability than conventional JSCC methods, even under the same channel conditions.

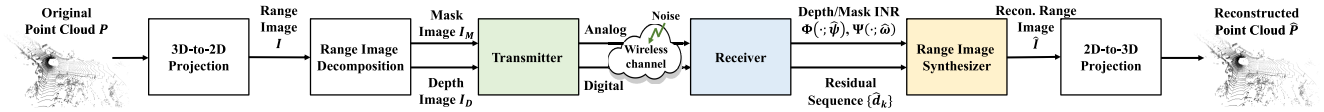
III. PROPOSED SCHEME

A. OVERVIEW

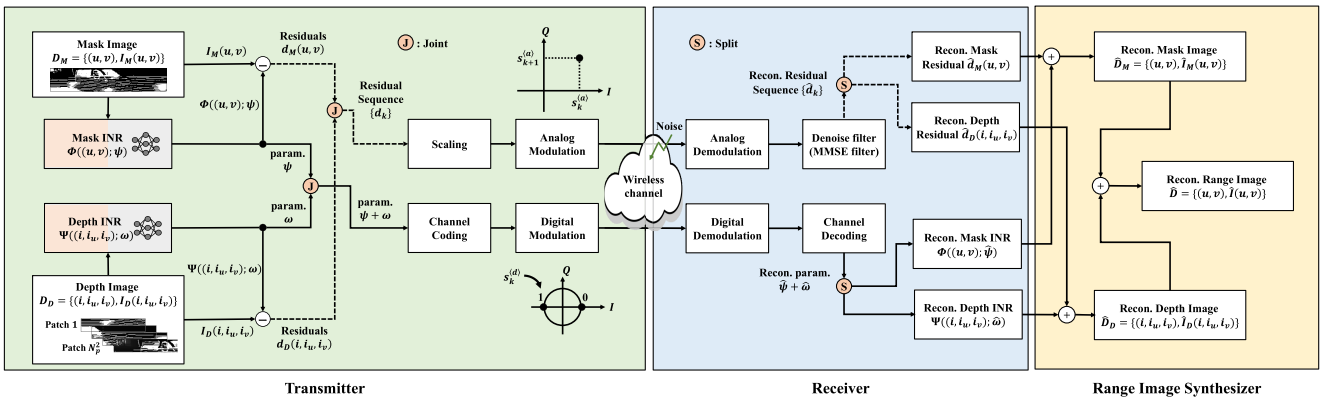
Fig. 2 shows an overview of the proposed scheme. Fig. 2 (a) specifically illustrates the end-to-end architecture of the proposed scheme. We consider that the LiDAR measurement to be compressed is a 3D point cloud consisting of N points, denoted as $\mathbf{P} = \{\mathbf{p}_i = [x_i, y_i, z_i] \mid i = 1, \ldots, N\}$, where $x_i, y_i, z_i \in \mathbb{R}$ represent the Cartesian coordinates of the i-th point. The proposed scheme first projects the input point cloud onto a 2D image plane via spherical coordinate transformation, resulting in a RI $I \in \mathbb{R}^{W \times H}$. The RI is then decomposed into a binary mask image $I_M \in \{0, 1\}^{W \times H}$, which indicates the presence or absence of valid measurements at each pixel, and a depth image $I_D \in \mathbb{R}^{W \times H}$ that stores the corresponding distance values. The depth image is further divided into rectangular patches.

Fig. 2 (b) shows the transceiver for the depth and mask images and the RI synthesizer on the receiver side. At the transmitter, the proposed scheme overfits two separate INR models, $\Phi(\cdot; \boldsymbol{\psi})$ and $\Psi(\cdot; \boldsymbol{\omega})$, to represent the mapping from coordinates to pixel values for the mask and depth images, respectively. The resulting parameters $\boldsymbol{\psi}$ and $\boldsymbol{\omega}$ are transmitted via digital modulation after channel coding. To capture the high-frequency components not modeled by the INRs, the proposed scheme computes residuals as the difference between the original images and the INR predictions. These residuals are scaled under a power constraint and transmitted using analog modulation. At the receiver, the INRs are reconstructed from the received parameters $\hat{\boldsymbol{\psi}}$ and $\hat{\boldsymbol{\omega}}$. The received residuals are then added to recover the mask and depth images, $\hat{I}_M$ and $\hat{I}_D$, which are combined to synthesize the final RI $\hat{I}$.

Finally, the LiDAR point cloud $\hat{\mathbf{P}}$ is reconstructed from the synthesized RI $\hat{I}$ by back-projecting it from the 2D image plane to 3D Cartesian coordinates via the spherical coordinate system.



(a) An end-to-end architecture of the proposed scheme



(b) The proposed transceiver and range image synthesizer.

FIGURE 2. Overview of the proposed scheme.

B. 3D-TO-2D PROJECTION

To reduce the computational complexity of point cloud compression, the proposed scheme first transforms all 3D points in the point cloud $\mathbf{P}$ into a 2D RI I via coordinate mapping. Specifically, the 3D-to-2D projection method consists of two steps: 1) mapping the 3D point cloud in their original 3D Cartesian coordinate system x-y-z to the spherical coordinate ρ-ϕ-θ, and 2) projecting these spherical coordinates onto an image coordinate system u-v.

Each point $\mathbf{p} \in \mathbf{P}$ in the 3D point cloud is initially represented in the Cartesian coordinate system as (x, y, z). This Cartesian point is then converted into its corresponding spherical coordinate $\mathbf{p}' = (\rho, \phi, \theta)$, where ρ denotes the length, ϕ the pitch, and θ the yaw of the coordinate system, as defined below:

$$\rho = \sqrt{x^2 + y^2 + z^2}, \ \phi = \arcsin\left(\frac{z}{\rho}\right), \ \theta = \arctan\left(\frac{y}{x}\right). \quad (1)$$

Subsequently, the spherical point is projected onto the 2D image coordinate (u, v) to generate the RI I using the following transformation:

$$u = \left\lfloor \frac{W}{2} \times \left(\frac{\theta}{\pi} + 1\right) \right\rfloor,$$
$$v = \left\lfloor H \times \left(1 - \frac{\phi + |\phi_{\text{down}}|}{\phi_{\text{up}} + |\phi_{\text{down}}|}\right) \right\rfloor. \quad (2)$$

Here, ϕ_{up} and ϕ_{down} denote the upper and lower bounds of the elevation angle ϕ observed in the dataset, $|\cdot|$ denotes the absolute value, and $\lfloor \cdot \rfloor$ denotes the floor function. Each pixel value $I(u, v)$ in the RI corresponds to the measured distance ρ computed in Eq. (1), expressed in an arbitrary physical unit. The parameters H and W in Eq. (2) represent the vertical and horizontal resolution of the RI, respectively, which are determined by the angular resolution of the LiDAR sensor in the elevation and azimuth directions. In our work, we set $H = 64$ and $W = 1024$.

Due to the sparsity of LiDAR measurements, not all pixels in the RI are necessarily assigned to a 3D point. Therefore, if a pixel (u', v') remains unassigned after the 3D-to-2D mapping of all points, we assign $I(u', v') = \rho_{\text{null}}$, where ρ_{null} is an arbitrary value indicating that no 3D point corresponds to that pixel. In practice, ρ_{null} is typically chosen to be greater than the maximum ρ value in the LiDAR data, or a negative number.

C. RANGE IMAGE DECOMPOSITION

Following the 3D-to-2D mapping, the RI is decomposed into a binary mask image $I_M \in \{0, 1\}^{W \times H}$ and a depth image $I_D \in \mathbb{R}^{W \times H}$.

The mask image I_M indicates whether a 3D point is assigned to each pixel in the RI, and is defined as follows:

$$I_M(u, v) = \begin{cases} 1 \text{ if } I(u, v) = \rho_{\text{null}}, \\ 0 \text{ otherwise}. \end{cases} \quad (3)$$

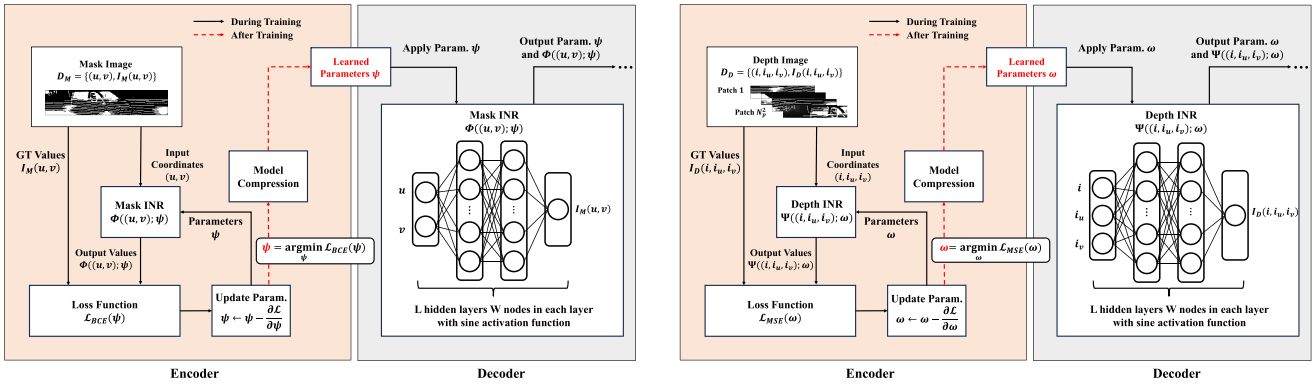
Based on the mask image, we construct a training dataset $\mathcal{D}_M$ for the mask INR $\Phi(\cdot; \boldsymbol{\psi})$, consisting of pixel coordinates and their corresponding binary values.:

$$\mathcal{D}_M = \{((u, v), I_M(u, v)) \mid u \in \{1, \ldots, W\}, v \in \{1, \ldots, H\}\}. \quad (4)$$

The depth image I_D is derived from the RI by masking out pixels with no assigned 3D point, which are treated as invalid and excluded from training:

$$I_D(u, v) = \begin{cases} \emptyset & \text{if } I(u, v) = \rho_{\text{null}}, \\ I(u, v) & \text{otherwise}. \end{cases} \quad (5)$$

To improve decoding performance and the quality of the reconstructed depth image, we divide the depth image into



(a) Mask INR

(b) Depth INR

FIGURE 3. The encoder and decoder architectures for the mask and depth INRs.

small rectangular patches, following [45]. Specifically, the RI is uniformly partitioned into N_p^2 patches $I'_D(i) \in \mathbb{R}^{\frac{W}{N_p} \times \frac{H}{N_p}}$, where N_p is a scaling factor and $i = 1, \ldots, N_p^2$. Each pixel in the patched RI is indexed as $I'_D(i, i_u, i_v)$, where i is the patch index and (i_u, i_v) are the in-patch coordinates with the origin at the top-left. Similar to the mask image, we construct a training dataset $\mathcal{D}_D$ for the depth INR $\Psi(\cdot; \boldsymbol{\omega})$, consisting of tuples of patch index, in-patch coordinates, and the corresponding depth value, excluding pixels with $\emptyset$:

$$\begin{aligned}\mathcal{D}_D = \{((i, i_u, i_v), I_D(i, i_u, i_v)) \mid\ & i \in \{1, \ldots, N_p^2\},\\ & i_u \in \{1, \ldots, W/N_p\},\\ & i_v \in \{1, \ldots, H/N_p\},\\ & I_D(i, i_u, i_v) \neq \emptyset\}. \end{aligned} \tag{6}$$

D. TRANSMITTER

1) DIGITAL TRANSMITTER

The digital transmitter consists of the mask and depth INRs, a channel encoder, and a digital modulator. It encodes the mask and depth images into compact representations by training the INRs, and then applying channel coding and digital modulation for wireless transmission. Fig. 3 shows the encoder and decoder architectures for the mask and depth INRs.

For the mask INR, we define a target function $\Phi_M : \mathbb{R}^2 \to \{0, 1\}$ that maps each pixel coordinate to a binary value indicating whether it is occupied by a projected LiDAR point. To approximate this target function, we train $\Phi(\cdot; \boldsymbol{\psi})$ using the dataset $\mathcal{D}_M$, by minimizing the binary cross-entropy (BCE) loss between the ground-truth values $I_M(u, v)$ and the predicted values $\Phi((u, v); \boldsymbol{\psi})$ as follows:

$$\begin{aligned}\mathcal{L}_{\mathrm{BCE}}(\boldsymbol{\psi}) = -\frac{1}{HW} \sum_{u}^{H} \sum_{v}^{W} \big[& I_M(u, v) \log(\Phi((u, v); \boldsymbol{\psi}))\\ & + (1 - I_M(u, v)) \log(1 - \Phi((u, v); \boldsymbol{\psi}))\big]. \end{aligned} \tag{7}$$

Similar to the mask INR, we define a target function $\Psi_D : \mathbb{R}^3 \to \mathbb{R}$ for the depth INR, which maps each input (i, i_u, i_v) to its corresponding depth value. To approximate this function, the depth INR $\Psi(\cdot; \boldsymbol{\omega})$ is trained using the dataset $\mathcal{D}_D$, by minimizing the mean squared error (MSE) between the ground-truth values $I_D(i, i_u, i_v)$ and the predicted values $\Psi((i, i_u, i_v); \boldsymbol{\omega})$, defined as:

$$\mathcal{L}_{\mathrm{MSE}}(\boldsymbol{\omega}) = \frac{1}{HW} \sum_{i}^{N_p^2} \sum_{i_u}^{W/N_p} \sum_{i_v}^{H/N_p} \|\Psi((i, i_u, i_v); \boldsymbol{\omega}) - I_D(i, i_u, i_v)\|^2. \tag{8}$$

After training both mask and depth INRs, their parameters, $\boldsymbol{\psi}$ and $\boldsymbol{\omega}$, become effective compressed representations of the depth and mask images. We then introduce model compression to these parameters to further reduce transmission and storage costs. As an initial step in our model compression procedure, the parameters are uniformly quantized to a bit depth of N_b. This quantization is layer-wise, meaning that given a parameter set corresponding to each layer in the depth and mask INRs as $\boldsymbol{\mu} \in \boldsymbol{\omega}$, a quantized parameter set $\boldsymbol{\mu}_q$ is obtained as follows:

$$\boldsymbol{\mu}_q = \mathrm{round}\left(\frac{\boldsymbol{\mu} - \boldsymbol{\mu}_{\min}}{2^{N_b}}\right) s + \boldsymbol{\mu}_{\min}, \quad s = \frac{\boldsymbol{\mu}_{\max} - \boldsymbol{\mu}_{\min}}{2^{N_b}}, \tag{9}$$

where $\mathrm{round}(\cdot)$ is a rounding function to the nearest integer and $\boldsymbol{\mu}_{\max}$ and $\boldsymbol{\mu}_{\min}$ are the maximum and minimum values in $\boldsymbol{\mu}$. To further minimize the bitrate, we then apply Huffman coding to the quantized tensor $\boldsymbol{\mu}_q$. This lossless entropy coding assigns variable-length codes based on the frequency of each parameter value, resulting in a more compact bitstream. The bitstream is processed by a channel encoder to provide robustness against transmission errors. We adopt a convolutional coding scheme with a rate of $1/2$, and the encoded bits are mapped to transmission symbols using digital modulation formats such as binary phase shift keying (BPSK), quadrature phase shift keying (QPSK), and quadrature amplitude modulation (QAM). In the case of BPSK, the k-th transmission symbol, denoted $s_k^{\langle \mathsf{d} \rangle}$, is defined as:

$$s_k^{\langle \mathsf{d} \rangle} = b_k, \quad b_k \in \mathbb{X} = \{\pm 1\}. \tag{10}$$

2) ANALOG TRANSMITTER

The analog transmitter consists of an analog residual encoder, a power scaler, and an analog modulator. This module improves the reconstruction quality by transmitting residuals that capture the high-frequency components not modeled by the INRs.

Specifically, the residuals for the mask and depth images are defined as the differences between the original images and the INR predictions at each pixel. For a pixel coordinate (u, v) in the mask image, the residual is

$$d_M(u, v) = I_M(u, v) - \Phi((u, v); \boldsymbol{\psi}), \tag{11}$$

and for a patch index i and in-patch coordinate (i_u, i_v) in the depth image, the residual is

$$d_D(i, i_u, i_v) = I_D(i, i_u, i_v) - \Psi((i, i_u, i_v); \boldsymbol{\omega}), \tag{12}$$

where $I_M(u, v)$ and $I_D(i, i_u, i_v)$ are the ground-truth values, and $\Phi(\cdot; \boldsymbol{\psi})$ and $\Psi(\cdot; \boldsymbol{\omega})$ are the predictions from the mask and depth INRs, respectively.

After calculating all residuals from the mask and depth images, they are jointly flattened into a single residual sequence $\{d_k\}$, where each d_k represents a residual value from either the mask image or the depth image. This sequence is then fed into the analog encoder for transmission.

In contrast, analog transmitters directly map residuals to transmission symbols, enabling the reconstruction quality to improve progressively as the wireless channel condition becomes better. To reduce the impact of channel noise, a scaling operation is applied prior to analog modulation. This operation, known as power allocation, aims to minimize the mean squared error (MSE) between the original and reconstructed residuals under a given transmission power constraint.

Let d_k denote the residual value at index k, and let g_k be the corresponding scaling factor. The analog-modulated transmission symbol $s_k^{\langle \mathsf{a} \rangle}$ is generated by scaling the residual as follows:

$$s_k^{\langle \mathsf{a} \rangle} = g_k d_k. \tag{13}$$

The goal is to determine the optimal set of scaling factors $\{g_k\}$ that minimize the mean squared error (MSE) between the original and reconstructed residuals, subject to an average transmission power constraint. This leads to the following optimization problem:

$$\min_{\{g_k\}} \quad \mathsf{MSE} = \frac{1}{N} \sum_{k=1}^{N} \frac{\sigma^2 \lambda_k}{g_k^2 \lambda_k + \sigma^2}, \tag{14}$$

$$\text{s.t.} \quad \frac{1}{N} \sum_{k=1}^{N} g_k^2 \lambda_k = P, \tag{15}$$

where $\lambda_k = |d_k|^2$ is the power of the k-th residual, σ^2 is the noise power of the wireless channel, and P is the transmission power budget.

The optimal scaling factor for each residual is obtained by solving the above problem, and is given by:

$$g_k = m \lambda_k^{-1/4}, \qquad m = \sqrt{\frac{P}{\sum_{k=1}^{N} \lambda_k^{1/2}}}. \tag{16}$$

The proposed scheme utilizes the scaling factor irrespective of the channel model, e.g., sub-optimal power allocation in Rayleigh fading channels.

Finally, for wireless transmission, every two scaled residuals $s_k^{\langle \mathsf{a} \rangle}$ and $s_{k+1}^{\langle \mathsf{a} \rangle}$ are jointly mapped onto the in-phase (I) and quadrature (Q) components of a complex-valued transmission symbol as follows:

$$x_k = s_k^{\langle \mathsf{a} \rangle} + \jmath s_{k+1}^{\langle \mathsf{a} \rangle}. \tag{17}$$

When the available bandwidth is B, i.e., the available number of transmission symbols/second, the proposed scheme can select and send up to $2B$ residuals/second, whose absolute value is large, because the proposed scheme assigns every two scaled residuals to the I and Q components of a complex-valued transmission symbol. This selection can reduce quality degradation even when the available bandwidth is insufficient to send all the residuals.

E. RECEIVER

1) DIGITAL RECEIVER

The digital receiver consists of the digital demodulator, the channel decoder, and the reconstructed mask and depth INRs. The receiver demodulates the digitally modulated symbols and decodes the channel-coded bitstreams to recover the parameter sets $\hat{\boldsymbol{\psi}}$ and $\hat{\boldsymbol{\omega}}$, which correspond to the mask and depth INRs, respectively. Using the recovered parameters, the receiver reconstructs the mask INR $\Phi(\cdot; \hat{\boldsymbol{\psi}})$ and the depth INR $\Psi(\cdot; \hat{\boldsymbol{\omega}})$.

2) ANALOG RECEIVER

The analog receiver consists of the analog demodulator, a denoising filter, and a decoder. The received symbols represent the analog-modulated residual values transmitted over the wireless channel. Specifically, each received symbol y_k can be modeled as:

$$y_k = h_k x_k + n_k, \tag{18}$$

where x_k is the transmitted analog symbol corresponding to the scaled residuals d_k and d_{k+1}, h_k is the channel gain, and n_k is additive noise with variance σ^2 accounting for channel distortion.

The goal of the receiver is to estimate the original residual d_k from the received symbol y_k. To this end, the transmitter provides the power of each residual component, defined as $\lambda_k = |d_k|^2$, as metadata. This enables the receiver to reconstruct the corresponding scaling factor g_k and apply the minimum mean squared error (MMSE) filter [32] as follows:

$$\hat{x}_k = \frac{h_k^* \mathbb{E}\left[|x_k|^2\right]}{|h_k|^2 \mathbb{E}\left[|x_k|^2\right] + \sigma^2} \cdot y_k, \tag{19}$$

where $\mathbb{E}[|x_k|^2] = g_k^2\lambda_k + g_{k+1}^2\lambda_{k+1}$. The estimates of the individual residuals $\hat{d}_k$ and $\hat{d}_{k+1}$ can be obtained by taking the real and imaginary parts of $\hat{x}_k$ and scaling them back as follows:

$$\hat{d}_k = \frac{\mathrm{Re}[\hat{x}_k]}{g_k},\ \hat{d}_{k+1} = \frac{\mathrm{Im}[\hat{x}_k]}{g_{k+1}}. \tag{20}$$

F. RANGE IMAGE SYNTHESIZER

The image synthesizer reconstructs the RI by applying the estimated residuals to the outputs of the mask and depth INRs. Specifically, the residuals are split into mask and depth components, which are added to the outputs of the mask and depth INRs, respectively. The reconstructed depth values are then selectively assigned to valid pixels as indicated by the mask image.

The mask image is reconstructed by feeding the coordinate set $\{(u, v) \mid u \in \{1, \ldots, W\},\ v \in \{1, \ldots, H\}\}$ to the mask INR $\Phi(\cdot; \hat{\boldsymbol{\psi}})$, and adding the residuals $\hat{d}_M(u, v)$ to the outputs. Formally, the binary mask is defined as:

$$\hat{I}_M(u, v) = \begin{cases} 1 \text{ if } \Phi\Big((u, v); \hat{\boldsymbol{\psi}}\Big) + \hat{d}_M(u, v) \geq 0.5, \\ 0 \text{ otherwise}. \end{cases} \tag{21}$$

The depth image is reconstructed in a two-stage manner. First, each patch is reconstructed by feeding the set of tuples $\{(i, i_u, i_v) \mid i \in \{1, \ldots, N_p^2\},\ i_u \in \{1, \ldots, W\},\ i_v \in \{1, \ldots, H\}\}$ to the depth INR $\Psi(\cdot; \hat{\boldsymbol{\omega}})$, and adding the corresponding residuals $\hat{d}_D(i, i_u, i_v)$. The reconstructed depth values are defined as:

$$\hat{I}_D(i, i_u, i_v) = \Psi\big((i, i_u, i_v); \hat{\boldsymbol{\omega}}\big) + \hat{d}_D(i, i_u, i_v). \tag{22}$$

The reconstructed patches are subsequently assembled to form the complete depth image $\hat{I}_D$.

The final RI $\hat{I}(u, v)$ is constructed by selectively assigning the reconstructed depth values to valid pixels based on the mask image. Let $\pi(u, v) = (i, i_u, i_v)$ denote the mapping from a pixel coordinate (u, v) to the corresponding patch index and in-patch coordinate in the depth domain. For each pixel (u, v), if the mask value $\hat{I}_M(u, v)$ is 0, the corresponding depth value $\hat{I}_D(\pi(u, v))$ is assigned. Otherwise, a null token ρ_{null} is used. Formally, the RI is defined as:

$$\hat{I}(u, v) = \begin{cases} \hat{I}_D(\pi(u, v)) & \text{if } \hat{I}_M(u, v) = 0, \\ \rho_{\text{null}} & \text{otherwise}. \end{cases} \tag{23}$$

G. 2D-TO-3D PROJECTION

The final stage of the decoding process reconstructs a 3D point cloud via a 2D-to-3D projection based on the RI $\hat{I}$. If $\hat{I}(u, v) \neq \rho_{\text{null}}$, the corresponding spherical coordinate $\hat{\mathbf{p}}' = (\hat{\rho}, \hat{\phi}, \hat{\theta})$ is computed as:

$$\begin{aligned} \hat{\rho} &= \hat{I}(u, v), \\ \hat{\phi} &= \Big(1 - \frac{v}{H}\Big)\big(\phi_{up} + |\phi_{\text{down}}|\big) - |\phi_{\text{down}}|, \\ \hat{\theta} &= -\Big(2\frac{u}{W} - 1\Big)\pi. \end{aligned} \tag{24}$$

Finally, the spherical coordinate is converted to the Cartesian coordinate $\hat{\mathbf{p}} = (\hat{x}, \hat{y}, \hat{z})$ as follows:

$$\hat{x} = \hat{\rho}\cos\hat{\phi}\cos\hat{\theta},\ \hat{y} = \hat{\rho}\cos\hat{\phi}\sin\hat{\theta}\ ,\ \hat{z} = \hat{\rho}\sin\hat{\phi}. \tag{25}$$

IV. EVALUATION

A. SETTINGS

Metric: We evaluate the decoded 3D point clouds using the Chamfer Distance (CD), a standard metric in the community:

$$\mathrm{CD} = \frac{1}{2}\left\{\frac{1}{|\mathbf{P}|}\sum_{\boldsymbol{p}\in\mathbf{P}}\min_{\widehat{\boldsymbol{p}}\in\widehat{\mathbf{P}}}\|\boldsymbol{p} - \widehat{\boldsymbol{p}}\|_2 + \frac{1}{|\widehat{\mathbf{P}}|}\sum_{\widehat{\boldsymbol{p}}\in\widehat{\mathbf{P}}}\min_{\boldsymbol{p}\in\mathbf{P}}\|\boldsymbol{p} - \widehat{\boldsymbol{p}}\|_2\right\}, \tag{26}$$

where $\mathbf{P}$ and $\widehat{\mathbf{P}}$ denote the original and decoded point sets, respectively. For the R-D performance assessment, we use the Bjøntegaard delta chamfer distance (BD-CD) [18] for calculating average chamfer distance (CD) improvement between R-D curves for the same bitrate. A higher BD-CD value indicates better reconstruction performance, with positive values representing improvements compared to the baselines.

Dataset: We use the KITTI dataset [46] as the source of 3D point cloud data. For R-D performance evaluation, we select five frames with frame indices 00, 25, 50, 75, and 100 from each sequence ranging from 00 to 06 in the KITTI Odometry dataset. To assess the effect of compression on downstream tasks, we perform 3D object detection using frame 000002 from the KITTI 3D Object Detection dataset.

Baselines: We evaluate the proposed scheme by comparing it with existing baselines in geometric 3D point cloud compression, 2D image compression, and INR-based compression.

1) As a baseline for 3D point cloud compression, we select **G-PCC** [6], a geometry-based method within the point cloud compression (PCC) family. We use the MPEG reference software TMC13-v14.0 for octree-based geometry compression. The compression level is adjusted by varying the positionQuantizationScale parameter from 0.05 to 0.95.
2) We also include **Draco** [8] as a baseline method for 3D point cloud compression within the PCC family. We use the official implementation of the Draco encoder, which applies KD-tree-based geometry compression [47]. The quantization parameter qp is varied from 5 to 13 to control the trade-off between bitrate and reconstruction quality.
3) In addition, we evaluate **OctAttention** [7], an octree-based autoencoder within the PCC family. This method improves the conventional octree structure by incorporating attention mechanisms for better context modeling. To evaluate its performance across different compression levels, we set the octree depth to values from 8 to 13.
4) As conventional image compression baselines, we select **Joint Photographic Experts Group 2000**

TABLE 1. The list of average BD-CD ↑ across the different LiDAR sequences for each SNR.

SNR	JPEG2000	HEIF	AVIF	R-PCC(Deflate)	R-PCC(LZ4)	G-PCC	Draco	Octattention	COIN	RIC
5 dB	0.591	0.366	0.232	0.033	0.139	0.062	0.086	-0.006	1.057	0.083
10 dB	0.812	0.502	0.270	0.065	0.183	0.110	0.145	0.026	1.157	0.115
15 dB	0.890	0.618	0.310	0.086	0.210	0.151	0.195	0.053	1.199	0.137
Average	0.764	0.495	0.271	0.061	0.177	0.108	0.142	0.024	1.138	0.112

(JPEG2000), **High-Efficiency Image File Format (HEIF)**, and **AV1 Image File Format (AVIF)**. To apply these codecs, we first convert the floating-point RI representations into 8-bit images. JPEG and JPEG2000 results are obtained using Pillow 8.4.0, while HEIF and AVIF results are obtained using pillow-heif 0.11.1 and pillow-avif-plugin 1.3.1, respectively.

5) **R-PCC** [9] is an RI based LiDAR compression baseline. It maps LiDAR point clouds to RIs and performs intra-coding using floating-point lossless coding methods. We adopt a uniform quantization framework and vary the accuracy parameter from 0 to 1 to control the degree of compression. For segmentation and modeling, we use Farthest Point Sampling (FPS) combined with plane fitting, setting the number of clusters to 100. Additionally, we employ Deflate and LZ4 compressors due to their better trade-off between compression efficiency and decompression speed.
6) **COmpression with Implicit Neural representations (COIN)** [11] is an INR–based image compression baseline. Its network is trained to directly map pixel coordinates to the corresponding pixel values of the RI. Since COIN does not apply depth/mask separation or JSCC scheme, we consider it a suitable reference for evaluating the effectiveness of both strategies in the proposed scheme.
7) We also include **RIC** [17] as a baseline method for an INR–based image compression. RIC enhances compression performance by decomposing the RI into separate depth and mask images and training individual INR models for each component. It further employs patch-wise learning for depth images and applies model compression techniques to reduce storage overhead. Since RIC does not incorporate JSCC, it provides a useful reference for evaluating the contribution of JSCC in the proposed scheme.

Network Architecture Details: The mask and depth INRs are implemented as multi-layer perceptrons (MLPs) with a fixed depth of $L = 6$ hidden layers and V nodes per layer, where sine activation functions are applied in all hidden layers. The mask INR takes a 2D coordinate (u, v) as input and outputs a scalar value with a sigmoid activation. To examine the effect of network complexity on compression performance, the number of nodes V is varied in $\{10, 20, 24, 28, 31, 34\}$. The depth INR takes a 3D input (i, i_u, i_v), where i is the patch index in a 16×16 grid and (i_u, i_v) are in-patch coordinates. It outputs a scalar value with an identity activation. To investigate the impact of model capacity on reconstruction quality, V is varied in $\{10, 20, 28, 31, 34, 37, 40, 42, 45\}$.

Hyperparameter Details: We use separate hyperparameter settings for mask and depth INRs. The general settings for both INRs include the Adam optimizer, an initial learning rate of 1×10^{-3}, 3,000 training epochs, and a batch size of 1. For depth INR, we adopt the cosine annealing scheduler with a warmup phase. The initial learning rate for the warmup phase is set to 1×10^{-4}, and the warmup period lasts for 300 epochs. During this period, the learning rate increases linearly to 1×10^{-3}. The learning rate then decreases according to a cosine curve for the remaining 2,700 epochs up to the minimum learning rate of 1×10^{-12}.

Wireless Channel Settings: The transmitted digital and analog symbols are impaired by the AWGN channels, which are modeled with a channel gain of $h_k = 1$, and the Rayleigh fading channels, where h_k follows a complex Gaussian distribution $h_k \sim \mathcal{CN}(0, 1)$. The wireless channel simulation is conducted using scikit-commpy 0.8.0. We adopt digital modulation schemes including BPSK, QPSK, and 16-QAM, all combined with a 1/2-rate convolutional code with a constraint length of 8.

Implementation Detail: All the evaluations exhibited in this paper are performed with CPUs of Intel Core i9-10850K and i9-13900KF and with GPUs of NVIDIA GeForce RTX 3080 and 4070. neural networks (NNs) for COIN and the proposed scheme are implemented, trained, and evaluated using PyTorch 2.2.0 with Python 3.10.

B. COMPARISON WITH BASELINES

1) RATE DISTORTION PERFORMANCE

We evaluate the R-D performance of the proposed scheme under three wireless channel conditions with SNR levels of 5 dB, 10 dB, and 15 dB. Table 1 lists the average BD-CD performance across the different LiDAR sequences for each SNR to evaluate the 3D reconstruction quality at certain SNRs, as well as the average BD-CD performance across SNRs. The results are averaged over 30 frames of sequences ranging from 00 to 06 in the KITTI Odometry dataset. At a low SNR of 5 dB, the proposed scheme outperforms AVIF (+0.232 BD-CD) and RIC (+0.083 BD-CD). It also achieves comparable performance compared to R-PCC (Deflate) (+0.033 BD-CD) and Octattention

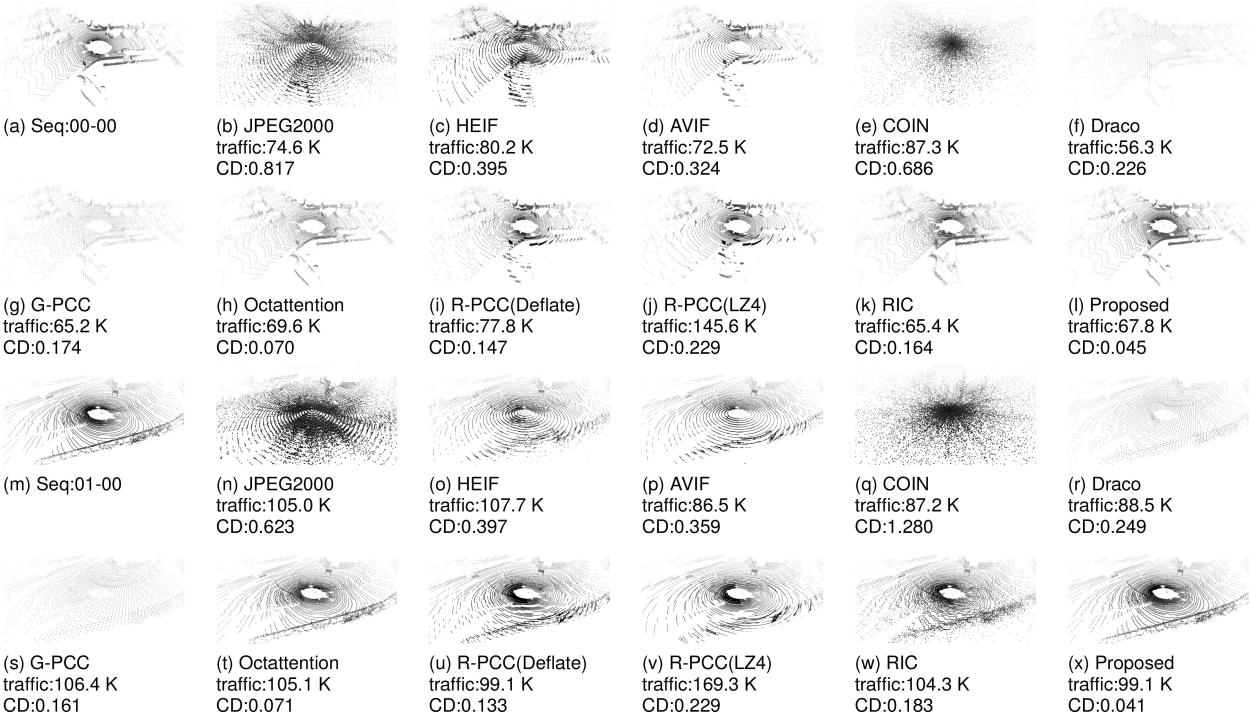


FIGURE 4. Snapshots of the reconstructed LiDAR point clouds obtained by the proposed and baseline methods under QPSK modulation and a 1/2-rate convolutional code, with a wireless channel SNR of 10 dB. The results in (a)–(l) and (m)–(x) correspond to sequences 00-00 and 01-00, respectively.

(−0.006 BD-CD). At a high SNR of 15 dB, the proposed scheme consistently surpasses all baseline methods, including Octattention (+0.053 BD-CD). This result highlights its superior reconstruction performance under favorable channel conditions.

Figs. 4 (a)–(x) show snapshots of the original and reconstructed LiDAR point clouds produced by each method at an SNR of 10 dB. Specifically, Figs. 4 (a)–(l) and (m)–(x) correspond to sequences 00-00 and 01-00, respectively. Image compression-based methods suffer from noise-induced structural degradation, which leads to visually noticeable distortions in the reconstructed point clouds. R-PCC methods exhibit persistent circular noise patterns. PCC methods such as G-PCC and Draco tend to produce sparse reconstructions with low point density, resulting in fragmented and incomplete geometric representations. The point clouds reconstructed by RIC accurately preserve the coarse object structure but exhibit low precision in reconstructing fine-grained geometric details. In contrast, the proposed scheme preserves both structural fidelity and point density, while reducing compression-induced distortions, thereby achieving structurally faithful point cloud reconstructions.

2) DOWNSTREAM TASK

This section evaluates the impact of point cloud compression methods on downstream task performance, using 3D object detection as a representative example. The downstream task performance of the proposed scheme is assessed under a wireless channel SNR of 10dB. We use PointPillar [48] as the 3D object detector. The input point clouds are compressed using AVIF, R-PCC (Deflate), G-PCC, Octattention, RIC, and the proposed scheme. The reflectance values of all point clouds are set to zero, and the reconstructed point clouds are used as input to PointPillar for inference. Each method is evaluated under three model sizes determined by the number of transmission symbols: approximately 50K (Low), 100K (Middle), and 150K (High). Table 2 shows the detection accuracy in terms of 2D IoU, Bird's Eye View (BEV) IoU, and 3D IoU for the object class "Car." The detection accuracy without compression (i.e., using the original point cloud) is 0.879 for 2D IoU, 0.866 for BEV IoU, and 0.780 for 3D IoU.

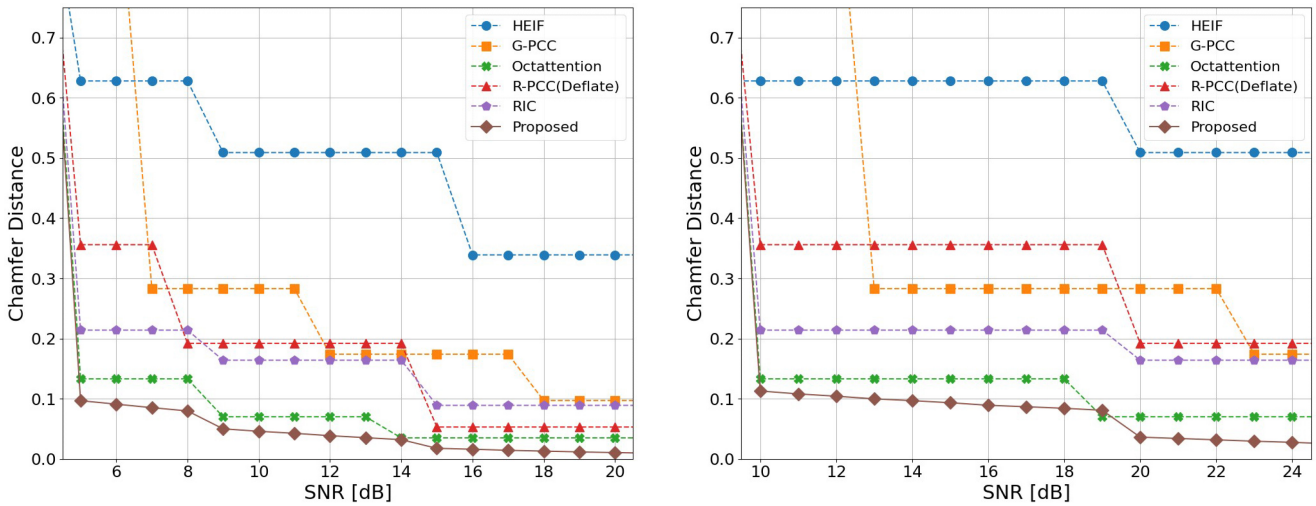
In the Low and Middle settings, the proposed scheme outperforms all baseline methods across all IoU metrics. In particular, under the Low setting, it demonstrates clear superiority even over high-performance approaches such as Octattention and RIC, showing that it can reliably preserve detection accuracy even under severe bitrate constraints. In the High setting, RIC achieves the highest detection accuracy.

3) EFFECT OF CHANNEL QUALITY FLUCTUATION

This section evaluates how the reconstruction quality of each method varies with wireless channel conditions, which often fluctuate due to environmental noise. Figs. 5 (a) and (b) show the reconstruction quality of the proposed and baseline schemes as a function of wireless channel quality under AWGN and Rayleigh fading channels, respectively. In both figures, the number of transmitted symbols is adjusted between 60K and 80K. All methods are considered adaptive modulation, where the optimal modulation scheme (BPSK,

TABLE 2. 3D object detection accuracy ↑ for different model sizes. The best and the second best results are denoted by pink and yellow.

Type	Model Size	AVIF	R-PCC (Deflate)	G-PCC	Oct-attention	RIC	Proposed	No compression
2D IoU	Low	0.190	0.662	0.727	0.761	0.477	0.835	0.879
	Mid	0.411	0.710	0.737	0.782	0.771	0.837	
	High	0.723	0.722	0.751	0.775	0.848	0.847	
BEV IoU	Low	–	0.602	0.747	0.753	0.183	0.846	0.866
	Mid	0.246	0.778	0.763	0.801	0.765	0.851	
	High	0.801	0.807	0.806	0.805	0.854	0.840	
3D IoU	Low	–	0.446	0.615	0.646	0.173	0.739	0.780
	Mid	0.221	0.610	0.633	0.680	0.653	0.736	
	High	0.659	0.637	0.664	0.681	0.742	0.735	



(a) AWGN channel.

(b) Rayleigh fading channel.

FIGURE 5. Reconstruction quality as a function of wireless channel quality.

TABLE 3. SNR thresholds (in dB) required for successful point cloud reconstruction under various modulation schemes and wireless channel models, for the proposed scheme and baseline methods.

Method	AWGN			Rayleigh fading		
	BPSK	QPSK	16QAM	BPSK	QPSK	16QAM
HEIF	5	9	16	9	20	30
G-PCC	7	12	18	13	23	35
Octattention	5	9	14	10	19	33
R-PCC (Deflate)	5	8	15	10	20	30
RIC	5	10	15	10	20	31
Proposed	5	10	15	10	20	31

QPSK, or 16QAM) is selected according to the SNR and combined with a 1/2-rate convolutional code of constraint length 8. Table 3 lists the SNR thresholds used to switch between modulation schemes.

In the baselines, the reconstruction quality exhibits the stair-case effect, with improvements occurring only at specific SNR thresholds. In contrast, the proposed scheme achieves smooth and continuous changes in reconstruction quality with respect to SNR, even under a fixed bitrate, and demonstrates graceful degradation as the SNR decreases. Moreover, it consistently outperforms the baselines across a wide SNR range even though the power allocation is suboptimal for Rayleigh fading channels, enabling robust and high-fidelity reconstruction under fluctuating channel conditions.

4) ENCODING LATENCY

Table 4 shows the average encoding latency of the proposed and baseline methods for LiDAR sequence 00-00. Here, the encoding latency for the RI-based schemes contains the conversion time from the point cloud to the RI.

It shows that the INR-based compression, including the proposed scheme, requires a significant encoding latency compared with the 3D point cloud compression and 2D image compression schemes. We note that the INR-based compression achieves extremely low decoding latency as shown in Fig. 1. It means the INR-based compression is effective for on-demand and quality-sensitive applications.

TABLE 4. Average encoding latency ↓.

Method	Latency per sequence
JPEG2000	10 ms
HEIF	55 ms
AVIF	94 ms
R-PCC (Deflate)	20 ms
R-PCC (LZ4)	60 ms
G-PCC	65 ms
Draco	10 ms
Octattention	130 ms
COIN	30 min
RIC	180 min
Proposed	90 min

TABLE 5. The list of BD-CD for KITTI dataset under different JSCC allocations to the depth and mask images of the RI. ✓ indicates that the corresponding component is transmitted with JSCC. Higher values indicate better reconstruction performance for the proposed scheme.

Depth	Mask	5dB	10dB	15dB	Average
		0.078	0.102	0.119	0.100
	✓	0.060	0.084	0.101	0.082
✓		0.066	0.032	0.020	0.039

C. ABLATION STUDY

1) EFFECT OF JSCC ALLOCATION TO DEPTH AND MASK IMAGES

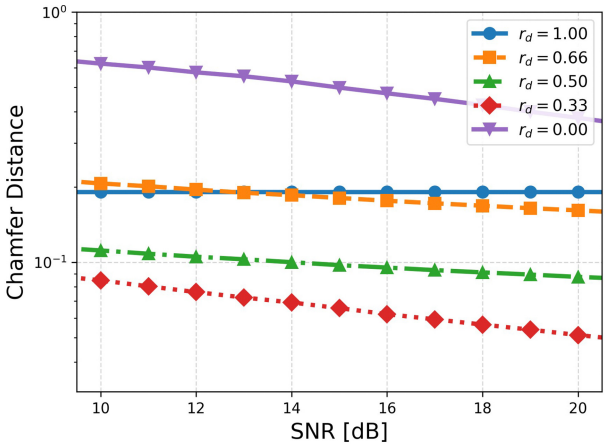
This section evaluates the impact of selectively applying JSCC to the depth and mask images of the RI, in order to clarify the individual contribution of JSCC to reconstruction quality. We compare four configurations: (i) JSCC is not applied to either image (i.e., RIC), (ii) JSCC is applied only to the mask image, (iii) JSCC is applied only to the depth image, and (iv) JSCC is applied to both images (i.e., the proposed scheme).

Table 5 shows the BD-CD for the KITTI dataset under each configuration. Larger values indicate that the compared configurations exhibit inferior performance compared to the proposed scheme. Among the tested configurations, the proposed scheme consistently achieves the best reconstruction performance across all SNR levels. Notably, applying JSCC only to the depth image yields an average BD-CD of 0.039, whereas applying it only to the mask image results in a higher BD-CD of 0.082. The configuration without JSCC applied to either image exhibits the worst performance, with an average BD-CD of 0.100. These results indicate that the depth image contributes more significantly to reconstruction quality when JSCC is applied to it, while the mask image also provides a moderate benefit.

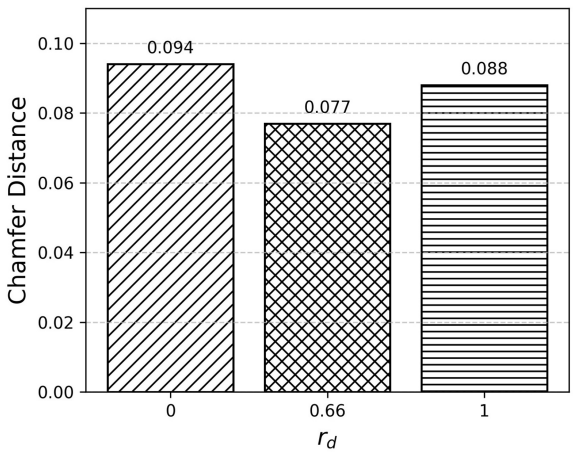
2) EFFECT OF DIGITAL-TO-ANALOG SYMBOL RATIO ON RECONSTRUCTION PERFORMANCE

Each of the depth and mask images is transmitted using both digitally encoded INR parameters and residuals transmitted in analog form. This section investigates how the ratio between digital and analog symbols affects reconstruction performance, by independently varying the ratio for the depth and mask images under a fixed total number of transmission symbols. We define a parameter $r_d \in [0, 1]$, which represents the ratio of transmission symbols allocated to digital symbols for INR. When $r_d = 1.0$, only the INR parameters are transmitted, and no residuals are sent. In contrast, $r_d = 0.0$ corresponds to a pure JSCC scheme, where the entire RI is transmitted using pseudo-analog modulation without relying on INR-based encoding.

Fig. 6 shows how the reconstruction quality varies as the ratio r_d is changed, while keeping the total number of transmission symbols fixed for either the depth or mask image. Fig. 6 (a) shows the reconstruction quality as a function of SNR when varying the ratio r_d for the depth image under an available bandwidth of 32 Ksymbols. To evaluate the effect of r_d on the depth image alone, the mask image is kept identical across all configurations.



(a) Depth image: Chamfer Distance as a function of SNR with varying r_d and an available bandwidth of 32 Ksymbols.



(b) Mask image: Chamfer Distance as a function of r_d under a fixed SNR of 10 dB and an available bandwidth of 5 Ksymbols.

FIGURE 6. Reconstruction quality across digital-to-analog symbol ratio r_d in the depth and mask images.

TABLE 6. The list of BD-CD results for the KITTI dataset when using different image compression schemes for RI reconstruction in the residual transmission framework. Higher values indicate better reconstruction performance for the proposed scheme.

Method	5dB	10dB	15dB	Average
JPEG2000	0.273	0.211	0.161	0.215
HEIF	0.211	0.177	0.150	0.179
AVIF	0.067	0.066	0.066	0.066
COIN	0.893	0.721	0.601	0.739

The best reconstruction quality is achieved when digital INR parameters and analog residuals are transmitted in combination. In particular, the configuration with $r_d = 0.33$, which corresponds to an digital-to-analog ratio of 1:2, consistently yields the lowest chamfer distance across all SNR levels. In contrast, the pure JSCC scheme ($r_d = 0.0$) shows limited improvement with increasing SNR and results in overall inferior reconstruction quality. Fig. 6 (b) shows the reconstruction quality as a function of r_d for the mask image under a wireless channel SNR of 10 dB and an available bandwidth of 5 Ksymbols. To evaluate the effect of r_d on the mask image alone, the depth image is kept unchanged across all configurations. The proposed hybrid scheme yields better 3D reconstruction quality compared with the pure digital and analog schemes under the limited bandwidth conditions.

3) EFFECT OF INTEGRATED COMPRESSION METHODS ON RESIDUAL TRANSMISSION

This section evaluates how different image compression methods affect reconstruction quality when integrated into the residual transmission framework. Specifically, we replace the RI compression module in the pipeline with JPEG2000, HEIF, AVIF, and COIN, respectively. In all configurations, the residual is computed from the reconstructed RI and transmitted using the same JSCC settings. Table 6 shows the BD-CD for the KITTI dataset when using different image compression schemes for RI reconstruction in the residual transmission framework. These results demonstrate that the proposed method outperforms all alternative approaches across all SNR conditions.

4) EFFECT OF NETWORK ARCHITECTURE

This section discusses the effect of the configurations for the depth INR architecture, specifically the patch size N_p and layer size L, on the quality of the reconstructed LiDAR point cloud. Here, a small patch size increases the complexity of intra-patch learning, while a large patch size increases the complexity of inter-patch learning.

Fig. 7 shows the 3D reconstruction quality of the proposed scheme under the different patch division sizes N_p. The evaluation results demonstrated that the patch size of $N_p = 16$ yields the best CD performance. However, either larger or smaller patch sizes degrade the 3D reconstruction quality under the same bitrate.

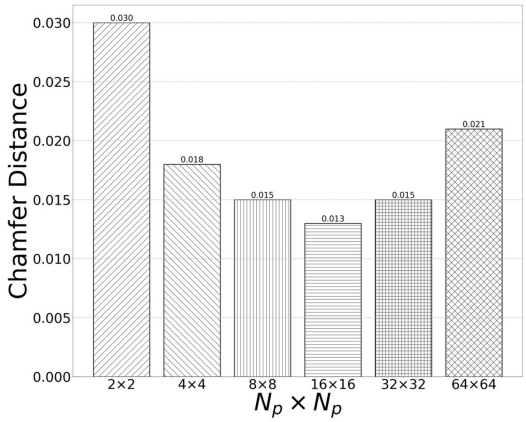


FIGURE 7. Chamfer distance under the different patch sizes.

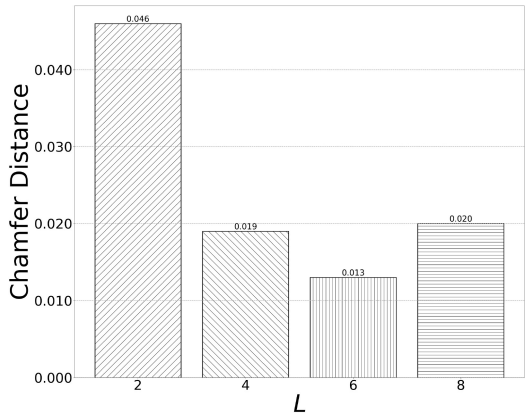


FIGURE 8. Chamfer distance under the different layer sizes.

Similarly, Fig. 8 shows the 3D reconstruction quality of the proposed scheme for different layer sizes L. The results indicate that a layer size of $L = 6$ is the most effective for CD performance.

5) EFFECT OF RESIDUAL SELECTION STRATEGY

This section evaluates the effect of the residual prioritization and available bandwidth on the 3D reconstruction quality. To discuss the effectiveness of residual transmission in the proposed scheme, we prepare two alternative strategies for sending residuals in band-limited conditions: (i) random selection and (ii) sequential selection. The random selection randomly chooses the transmission residuals, and the sequential selection sequentially chooses residuals from the top-left to the bottom-right of the RI to fit the available bandwidth. Here, we consider a retention ratio R, whose range is [0, 1]. For example, a retention ratio of 1 indicates that the available bandwidth is sufficient for sending all the residuals, whereas a retention ratio of 0.7 means that 30% of the residuals cannot be transmitted due to bandwidth limitations.

Fig. 9 shows the 3D reconstruction quality as a function of the retention ratio R of the residuals for three methods at the wireless channel SNR of 15 dB. It shows that the absolute value-based residual transmission in the proposed scheme

TABLE 7. The list of BD-CD ↑ for KITTI dataset at the wireless channel SNR of 5 dB.

Seq.	JPEG 2000	HEIF	AVIF	RPCC (Deflate)	RPCC (LZ4)	G-PCC	Draco	Oct-attention	COIN	RIC
00-00	0.486	0.330	0.235	0.051	0.130	0.064	0.081	-0.003	1.020	0.057
00-25	0.564	0.342	0.216	0.056	0.174	0.055	0.080	-0.006	0.990	0.067
00-50	0.674	0.372	0.232	0.061	0.153	0.053	0.058	-0.007	1.146	0.063
00-75	0.744	0.402	0.237	0.015	0.099	0.050	0.055	-0.006	0.989	0.058
00-100	0.500	0.325	0.219	0.015	0.107	0.061	0.075	-0.005	1.139	0.064
01-00	0.568	0.369	0.238	0.042	0.174	0.091	0.140	0.000	1.142	0.109
01-25	0.510	0.368	0.286	0.055	0.150	0.097	0.167	0.022	1.117	0.097
01-50	0.537	0.362	0.264	0.060	0.161	0.089	0.136	0.018	1.136	0.087
01-75	0.601	0.369	0.235	0.044	0.146	0.081	0.089	0.012	0.990	0.071
01-100	0.653	0.397	0.241	0.060	0.186	0.072	0.095	0.009	1.130	0.073
02-00	0.591	0.368	0.223	0.033	0.189	0.056	0.091	-0.020	1.026	0.098
02-25	0.476	0.329	0.239	0.052	0.139	0.069	0.119	0.003	1.024	0.085
02-50	0.493	0.337	0.250	0.045	0.129	0.085	0.126	0.001	1.048	0.091
02-75	0.506	0.345	0.248	0.030	0.128	0.067	0.117	0.002	0.999	0.077
02-100	0.470	0.315	0.235	0.036	0.131	0.074	0.099	0.008	1.150	0.061
03-00	0.602	0.342	0.210	0.057	0.168	0.049	0.060	-0.007	1.026	0.062
03-25	0.741	0.428	0.228	0.022	0.119	0.040	0.044	-0.016	0.968	0.080
03-50	0.583	0.363	0.231	0.011	0.115	0.054	0.069	-0.017	1.120	0.093
03-75	0.468	0.325	0.217	0.003	0.098	0.065	0.087	-0.013	0.964	0.095
03-100	0.472	0.329	0.224	0.009	0.112	0.068	0.105	-0.007	1.139	0.088
04-00	0.508	0.350	0.235	0.061	0.183	0.075	0.109	-0.010	1.125	0.104
04-25	0.661	0.396	0.230	0.022	0.130	0.060	0.081	-0.015	1.076	0.106
04-50	0.647	0.380	0.234	0.038	0.146	0.059	0.077	-0.001	1.116	0.071
04-75	0.577	0.348	0.225	0.021	0.133	0.051	0.064	-0.011	0.971	0.082
04-100	0.502	0.328	0.239	0.034	0.123	0.067	0.080	0.009	1.148	0.059
05-00	0.507	0.324	0.235	0.067	0.186	0.062	0.087	-0.008	1.049	0.076
05-25	0.520	0.327	0.228	0.016	0.105	0.064	0.080	-0.002	1.030	0.060
05-50	0.767	0.427	0.233	0.012	0.105	0.043	0.055	-0.017	1.021	0.077
05-75	0.813	0.466	0.240	0.011	0.114	0.041	0.048	-0.022	0.960	0.091
05-100	0.642	0.383	0.242	0.018	0.110	0.050	0.063	-0.009	1.132	0.068
06-00	0.575	0.351	0.209	0.046	0.184	0.054	0.077	-0.026	1.002	0.105
06-25	0.690	0.385	0.210	0.014	0.151	0.046	0.070	-0.022	1.008	0.109
06-50	0.672	0.391	0.211	0.012	0.133	0.047	0.065	-0.022	1.011	0.112
06-75	0.814	0.477	0.217	0.018	0.134	0.039	0.056	-0.026	0.957	0.109
06-100	0.536	0.349	0.224	0.007	0.107	0.069	0.103	-0.012	1.117	0.106
Average	0.591	0.366	0.232	0.033	0.139	0.062	0.086	-0.006	1.057	0.083

TABLE 8. The list of BD-CD ↑ for KITTI dataset at the wireless channel SNR of 10 dB.

Seq.	JPEG 2000	HEIF	AVIF	RPCC (Deflate)	RPCC (LZ4)	G-PCC	Draco	Oct-attention	COIN	RIC
00-00	0.673	0.424	0.275	0.082	0.172	0.109	0.134	0.027	1.122	0.089
00-25	0.768	0.463	0.256	0.082	0.211	0.096	0.136	0.022	1.067	0.095
00-50	0.964	0.536	0.266	0.085	0.197	0.084	0.099	0.014	1.291	0.087
00-75	1.061	0.586	0.271	0.039	0.128	0.080	0.094	0.014	1.064	0.079
00-100	0.688	0.418	0.257	0.044	0.143	0.100	0.123	0.022	1.279	0.094
01-00	0.757	0.496	0.288	0.080	0.297	0.160	0.232	0.049	1.234	0.151
01-25	0.670	0.454	0.319	0.079	0.189	0.156	0.256	0.063	1.206	0.133
01-50	0.688	0.464	0.298	0.082	0.185	0.139	0.201	0.052	1.217	0.115
01-75	0.806	0.518	0.266	0.075	0.170	0.123	0.136	0.046	1.065	0.098
01-100	0.907	0.574	0.267	0.099	0.234	0.119	0.157	0.041	1.277	0.097
02-00	0.811	0.517	0.266	0.068	0.253	0.117	0.168	0.021	1.110	0.137
02-25	0.624	0.416	0.273	0.078	0.177	0.114	0.180	0.036	1.108	0.117
02-50	0.619	0.421	0.285	0.072	0.156	0.137	0.189	0.034	1.114	0.120
02-75	0.688	0.439	0.287	0.064	0.165	0.112	0.185	0.035	1.085	0.105
02-100	0.626	0.400	0.264	0.063	0.163	0.116	0.152	0.036	1.276	0.084
03-00	0.870	0.485	0.243	0.080	0.216	0.083	0.109	0.018	1.124	0.089
03-25	1.030	0.613	0.272	0.054	0.162	0.077	0.090	0.014	1.052	0.113
03-50	0.807	0.490	0.269	0.046	0.164	0.101	0.138	0.017	1.266	0.128
03-75	0.626	0.416	0.259	0.039	0.141	0.124	0.148	0.024	1.048	0.135
03-100	0.612	0.421	0.267	0.043	0.150	0.122	0.175	0.029	1.276	0.126
04-00	0.665	0.451	0.279	0.096	0.243	0.135	0.172	0.031	1.214	0.144
04-25	0.897	0.558	0.273	0.058	0.178	0.115	0.143	0.022	1.167	0.146
04-50	0.908	0.554	0.265	0.069	0.185	0.096	0.128	0.027	1.211	0.096
04-75	0.808	0.491	0.258	0.055	0.181	0.097	0.122	0.020	1.052	0.113
04-100	0.699	0.421	0.264	0.063	0.156	0.104	0.130	0.034	1.289	0.081
05-00	0.688	0.417	0.271	0.096	0.237	0.112	0.141	0.022	1.127	0.108
05-25	0.704	0.424	0.265	0.045	0.136	0.109	0.128	0.026	1.113	0.087
05-50	1.144	0.659	0.268	0.045	0.154	0.082	0.108	0.011	1.113	0.107
05-75	1.167	0.700	0.276	0.043	0.166	0.080	0.096	0.007	1.044	0.123
05-100	0.939	0.561	0.272	0.049	0.157	0.088	0.115	0.018	1.283	0.093
06-00	0.778	0.494	0.260	0.083	0.234	0.118	0.141	0.017	1.088	0.147
06-25	0.977	0.566	0.254	0.055	0.197	0.106	0.144	0.016	1.107	0.152
06-50	0.930	0.569	0.255	0.052	0.172	0.106	0.124	0.017	1.097	0.154
06-75	1.112	0.698	0.268	0.053	0.174	0.089	0.116	0.009	1.041	0.147
06-100	0.724	0.466	0.270	0.045	0.150	0.135	0.169	0.028	1.271	0.152
Average	0.812	0.502	0.270	0.065	0.183	0.110	0.145	0.026	1.157	0.115

outperforms both random and sequential selection for any retention ratio. This result confirms that the absolute value-based strategy is an intuitive but highly effective and practical method for maintaining 3D reconstruction quality under the same available bandwidth. It also shows that degradation of 3D reconstruction quality in the proposed scheme is slight at a retention ratio of approximately 0.6 and becomes more significant at 0.5. It means that, to preserve 3D reconstruction quality, more than 60% of the residuals with larger absolute values should be transmitted using the proposed scheme. We

TABLE 9. The list of BD-CD ↑ for KITTI dataset at the wireless channel SNR of 15 dB.

Seq.	JPEG 2000	HEIF	AVIF	RPCC (Deflate)	RPCC (LZ4)	G-PCC	Draco	Oct-attention	COIN	RIC
00-00	0.735	0.515	0.307	0.096	0.198	0.151	0.172	0.050	1.162	0.108
00-25	0.811	0.563	0.288	0.095	0.229	0.129	0.172	0.042	1.092	0.110
00-50	1.045	0.688	0.303	0.097	0.219	0.110	0.133	0.032	1.335	0.101
00-75	1.176	0.724	0.303	0.054	0.149	0.102	0.128	0.030	1.098	0.093
00-100	0.748	0.509	0.288	0.062	0.164	0.135	0.159	0.043	1.327	0.112
01-00	0.849	0.625	0.337	0.110	0.340	0.207	0.327	0.088	1.287	0.181
01-25	0.741	0.552	0.354	0.102	0.408	0.195	0.344	0.096	1.254	0.160
01-50	0.803	0.601	0.328	0.111	0.221	0.204	0.281	0.086	1.286	0.142
01-75	0.860	0.638	0.297	0.091	0.119	0.162	0.169	0.071	1.093	0.113
01-100	0.981	0.709	0.299	0.121	0.257	0.155	0.200	0.067	1.322	0.112
02-00	0.835	0.620	0.315	0.086	0.265	0.166	0.217	0.051	1.132	0.157
02-25	0.667	0.494	0.312	0.094	0.191	0.153	0.233	0.065	1.143	0.138
02-50	0.664	0.486	0.319	0.087	0.175	0.178	0.240	0.060	1.146	0.139
02-75	0.757	0.547	0.318	0.087	0.185	0.154	0.243	0.064	1.126	0.125
02-100	0.684	0.482	0.292	0.082	0.186	0.154	0.196	0.060	1.327	0.099
03-00	0.937	0.620	0.278	0.092	0.261	0.115	0.147	0.038	1.158	0.105
03-25	1.140	0.742	0.317	0.075	0.205	0.111	0.127	0.037	1.092	0.134
03-50	0.877	0.611	0.318	0.068	0.197	0.145	0.195	0.048	1.317	0.151
03-75	0.672	0.499	0.302	0.061	0.163	0.168	0.208	0.054	1.082	0.158
03-100	0.661	0.501	0.307	0.065	0.172	0.163	0.230	0.059	1.325	0.151
04-00	0.726	0.551	0.323	0.117	0.272	0.178	0.225	0.065	1.259	0.172
04-25	1.000	0.701	0.326	0.086	0.311	0.174	0.198	0.056	1.215	0.178
04-50	0.989	0.684	0.301	0.089	0.199	0.133	0.166	0.049	1.248	0.112
04-75	0.881	0.608	0.295	0.077	0.209	0.137	0.170	0.044	1.088	0.133
04-100	0.758	0.521	0.285	0.078	0.175	0.132	0.169	0.052	1.334	0.094
05-00	0.758	0.521	0.310	0.114	0.268	0.162	0.191	0.050	1.167	0.131
05-25	0.755	0.517	0.296	0.062	0.153	0.144	0.166	0.048	1.147	0.103
05-50	1.225	0.759	0.306	0.063	0.176	0.114	0.148	0.032	1.143	0.125
05-75	1.289	0.823	0.313	0.067	0.204	0.117	0.143	0.032	1.084	0.146
05-100	1.019	0.676	0.311	0.068	0.185	0.121	0.155	0.040	1.330	0.108
06-00	0.860	0.621	0.323	0.113	0.264	0.170	0.193	0.053	1.132	0.177
06-25	1.092	0.721	0.313	0.086	0.180	0.162	0.210	0.053	1.162	0.184
06-50	1.016	0.694	0.310	0.079	0.199	0.155	0.171	0.050	1.139	0.183
06-75	1.338	0.919	0.316	0.092	0.060	0.147	0.184	0.041	1.106	0.185
06-100	0.785	0.578	0.323	0.069	0.175	0.190	0.220	0.062	1.324	0.182
Average	0.890	0.618	0.310	0.086	0.210	0.151	0.195	0.053	1.199	0.137

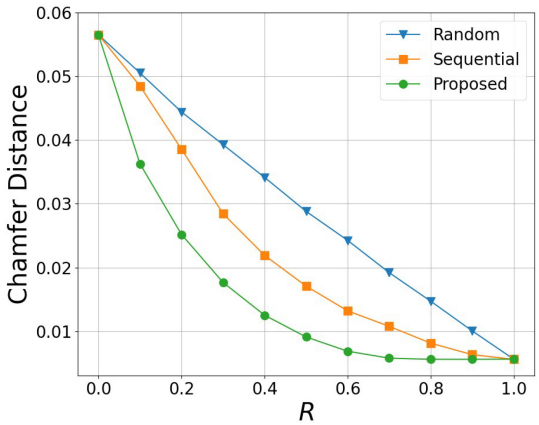


FIGURE 9. Reconstruction quality as a function of the retention ratio R of the residuals at the wireless channel SNR of 15 dB.

note that a similar trend was observed for different LiDAR sequences and various SNR regimes.

V. CONCLUSION

We proposed a novel scheme for LiDAR point cloud representation that combines an INR-based digital representation with pseudo-analog residual transmission. The proposed scheme is designed to efficiently represent high-frequency components in a small INR via residual transmission and to improve reconstruction quality under time-varying wireless channels by incorporating JSCC. Experiments on the KITTI dataset show that the proposed scheme outperforms existing methods for point cloud, image, RI, and INR-based compression in terms of R-D performance, achieving a BD-CD improvement of up to 1.199. In addition, it preserves 3D object detection accuracy even under severe bitrate constraints, demonstrating its effectiveness for downstream perception tasks.

In future work, we will develop a quantitative metric to measure the smoothness of quality adaptation across SNR regimes, i.e., gracefulness, of the baselines.

APPENDIX

This appendix provides further details for Table 1. Tables 7, 8, and 9 show the detailed BD-CD performance across the different LiDAR sequences for each SNR, respectively.

REFERENCES

[1] V. Croce, G. Caroti, L. De Luca, K. Jacquot, A. Piemonte, and P. Véron, "From the semantic point cloud to heritage-building information modeling: A semiautomatic approach exploiting machine learning," *Remote Sens.*, vol. 13, no. 3, p. 461, 2021.

[2] L. Jones and P. Hobbs, "The application of terrestrial LiDAR for geohazard mapping, monitoring and modelling in the British geological survey," *Remote Sens.*, vol. 13, no. 3, p. 395, 2021. [Online]. Available: https://www.mdpi.com/2072-4292/13/3/395

[3] S. Salamanca, P. Merchán, A. Espacio, E. Pérez, and M. J. Merchán, "Segmentation of 3D point clouds of heritage buildings using edge detection and supervoxel-based topology," *Sensors*, vol. 24, no. 13, p. 4390, 2024.

[4] E. Kim and G. Medioni, "Urban scene understanding from aerial and ground LIDAR data," *Mach. Vis. Appl.*, vol. 22, no. 4, pp. 691–703, Jul. 2011. [Online]. Available: https://doi.org/10.1007/s00138-010-0279-7

[5] C. Cao, M. Preda, and T. Zaharia, "3D point cloud compression: A survey," in *Proc. 24th Int. Conf. 3D Web Technol.*, 2019, pp. 1–9.

[6] D. Graziosi, O. Nakagami, S. Kuma, A. Zaghetto, T. Suzuki, and A. Tabatabai, "An overview of ongoing point cloud compression standardization activities: Video-based (V-PCC) and geometry-based (G-PCC)," *APSIPA Trans. Signal Inf. Process.*, vol. 9, p. e13, Apr. 2020.

[7] C. Fu, G. Li, R. Song, W. Gao, and S. Liu, "OctAttention: Octree-based large-scale contexts model for point cloud compression," in *Proc. AAAI*, vol. 36, no. 1, pp. 625–633, Jun. 2022.

[8] "Draco 3D data compression." 2022. [Online]. Available: https://google.github.io/draco/

[9] S. Wang, J. Jiao, P. Cai, and L. Wang, "R-PCC: A baseline for range image-based point cloud compression," in *Proc. ICRA*, 2022, pp. 10055–10061.

[10] *Information Technology High Efficiency Coding and Media Delivery in Heterogeneous Environments, Part 12: Image File Format*, ISO/IEC Standard 23008-12:2022, 2022.

[11] E. Dupont, A. Golinski, M. Alizadeh, Y. W. Teh, and A. Doucet, "COIN: Compression with implicit neural representations," in *Proc. ICLR Workshop Neural Compress.*, 2021, pp. 1–13.

[12] S. Saito, T. Simon, J. Saragih, and H. Joo, "Pifuhd: Multi-level pixel-aligned implicit function for high-resolution 3d human digitization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 84–93.

[13] V. Sitzmann, J. N. P. Martel, A. W. Bergman, D. B. Lindell, and G. Wetzstein, "Implicit neural representations with periodic activation functions," in *Proc. 34th Int. Conf. Neural Inf. Process. Syst.*, 2020, pp. 1–12.

[14] P. Wang, L. Liu, Y. Liu, C. Theobalt, T. Komura, and W. Wang, "NeuS: Learning neural implicit surfaces by volume rendering for multi-view reconstruction," in *Proc. 35th Conf. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 27171–27183.

[15] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "NeRF: Representing scenes as neural radiance fields for view synthesis," *Commun. ACM*, vol. 65, no. 1, pp. 99–106, 2021.

[16] R. Xue, J. Li, T. Chen, D. Ding, X. Cao, and Z. Ma, "NeRI: Implicit neural representation of LiDAR point cloud using range image sequence," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2024, pp. 8020–8024.

[17] A. Kuwabara, S. Kato, T. Fujihashi, T. Koike-Akino, and T. Watanabe, "Range image-based implicit neural compression for LiDAR point clouds," 2025, *arXiv:2504.17229*.

[18] G. Bjøntegaard, "Calculation of Average PSNR Differences between RD-curves," Int. Telecommun. Union, Geneva, Switzerland, document TU-T SG16/Q6 VCEG-M33, 2001.

[19] M. Tancik et al., "Fourier features let networks learn high frequency functions in low dimensional domains," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 7537–7547.

[20] C.-S. Liu, J.-F. Yeh, H. Hsu, H.-T. Su, M.-S. Lee, and W. H. Hsu, "BIRD-PCC: Bi-directional range image-based deep lidar point cloud compression," in *Proc. ICASSP*, 2023, pp. 1–5.

[21] L. Zhao, K.-K. Ma, Z. Liu, Q. Yin, and J. Chen, "Real-time scene-aware LiDAR point cloud compression using semantic prior representation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 8, pp. 5623–5637, Aug. 2022.

[22] T. Fujihashi, T. Koike-Akino, and T. Watanabe, "Soft delivery: Survey on a new paradigm for wireless and mobile multimedia streaming," *ACM Comput. Surv.*, vol. 56, no. 2, pp. 1–37, 2023.

[23] P. de Oliveira Rente, C. Brites, J. Ascenso, and F. Pereira, "Graph-based static 3D point clouds geometry coding," *IEEE Trans. Multimed.*, vol. 21, no. 2, pp. 284–299, Feb. 2019.

[24] T. Fujihashi, T. Koike-Akino, T. Watanabe, and P. V. Orlik, "HoloCast+: hybrid digital-analog transmission for graceful point cloud delivery with graph fourier transform," *IEEE Trans. Multimed.*, vol. 24, pp. 2179–2191, 2022.

[25] H. Kirihara, S. Ibuki, T. Fujihashi, T. Koike-Akino, and T. Watanabe, "Point cloud geometry compression using parameterized graph fourier transform," in *Proc. SIGCOMM Workshop Emerg. Multimedia Syst.*, 2024, pp. 52–57.

[26] S. Ueno, T. Fujihashi, T. Koike-Akino, and T. Watanabe, "Point cloud soft multicast for untethered XR users," *IEEE Trans. Multimed.*, vol. 25, pp. 7185–7195, 2023.

[27] T. Fujihashi, S. Kato, and T. Koike-Akino, "Implicit neural representation for low-overhead graph-based holographic-type communications," in *Proc. ICASSP*, 2024, pp. 2825–2829.

[28] J. Kammerl, N. Blodow, R. B. Rusu, S. Gedikli, M. Beetz, and E. Steinbach, "Real-time compression of point cloud streams," in *Proc. ICRA*, 2012, pp. 778–785.

[29] L. Huang, S. Wang, K. Wong, J. Liu, and R. Urtasun, "OctSqueeze: Octree-structured entropy model for LiDAR compression," in *Proc. CVPR*, 2020, pp. 1313–1323.

[30] X. Zhou, C. R. Qi, Y. Zhou, and D. Anguelov, "RIDDLE: Lidar data compression with range image deep delta encoding," in *Proc. CVPR*, 2022, pp. 17191–17200.

[31] E. Dupont, H. Loya, M. Alizadeh, A. Goliński, Y. W. Teh, and A. Doucet, "COIN++: Neural compression across modalities," in *Proc. TMLR*, 2022, pp. 1–26.

[32] S. Jakubczak and D. Katabi, "A cross-layer design for scalable mobile video," in *Proc. ACM Annu. Int. Conf. Mobile Comput. Netw.*, 2011, pp. 289–300.

[33] H. Cui, R. Xiong, C. Luo, Z. Song, and F. Wu, "Denoising and resource allocation in uncoded video transmission," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 1, pp. 102–112, Feb. 2015.

[34] J. Žádník, M. Kieffer, A. Trioux, M. Mäkitalo, and P. Jääskeläinen, "CV-Cast: Computer vision–oriented linear coding and transmission," *IEEE Trans. Mobile Comput.*, vol. 24, no. 2, pp. 1149–1162, Feb. 2025.

[35] X. Fan, F. Wu, D. Zhao, and O. C. Au, "Distributed wireless visual communication with power distortion optimization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 6, pp. 1040–1053, Jun. 2013.

[36] R. Xiong, F. Wu, X. Fan, C. Luo, S. Ma, and W. Gao, "Power-distortion optimization for wireless image/video SoftCast by transform coefficients energy modeling with adaptive chunk division," in *Proc. IEEE Int. Conf. Vis. Commun. Image Process.*, 2013, pp. 1–6.

[37] Y. Gui, L. Hancheng, F. Wu, and C. W. Chen, "LensCast: Robust wireless video transmission over mmWave MIMO with lens antenna array," *IEEE Trans. Multimedia*, vol. 24, pp. 33–48, 2020.

[38] X.-W. Tang, X.-L. Huang, and F. Hu, "QoE-driven UAV-enabled pseudo-analog wireless video broadcast: A joint optimization of power and trajectory," *IEEE Trans. Multimedia*, vol. 23, pp. 2398–2412, 2021.

[39] X. Luo, H.-H. Chen, and Q. Guo, "Semantic communications: Overview, open issues, and future research directions," *IEEE Wireless Commun.*, vol. 29, no. 1, pp. 210–219, Feb. 2022.

[40] E. Bourtsoulatze, D. B. Kurka, and D. Gündüz, "Deep joint source-channel coding for wireless image transmission," *IEEE Trans. Cogn. Commun. Netw.*, vol. 5, no. 3, pp. 567–579, Sep. 2019.

[41] H. Wu, Y. Shao, E. Ozfatura, K. Mikolajczyk, and D. Gündüz, "Transformer-aided wireless image transmission with channel feedback," *IEEE Trans. Wireless Commun.*, vol. 23, no. 9, pp. 11904–11919, Sep. 2024.

[42] L. X. Nguyen et al., "Swin transformer-based dynamic semantic communication for multi-user with different computing capacity," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 6, pp. 8957–8972, Jun. 2024.

[43] S. Ibuki, T. Okamoto, T. Fujihashi, T. Koike-Akino, and T. Watanabe, "Rateless deep joint source channel coding for 3d point cloud," *IEEE Access*, vol. 13, pp. 39585–39599, 2025.

[44] T. Fujihashi, T. K. Akino, S. Chen, and T. Watanabe, "Wireless 3D point cloud delivery using deep graph neural networks," in *Proc. IEEE Int. Conf. Commun.*, 2021, pp. 1–6.

[45] Y. Bai, C. Dong, C. Wang, and C. Yuan, "PS-NeRV: Patch-wise stylized neural representations for videos," in *Proc. ICIP*, 2023, pp. 41–45.

[46] A. Geiger, "Are we ready for autonomous driving? The kitti vision benchmark suite," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2012, pp. 3354–3361.

[47] O. Devillers and P.-M. Gandoin, "Geometric compression for interactive transmission," in *Proc. Inf. Vis. Conf.*, 2000, pp. 319–326.

[48] A. H. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang, and O. Beijbom, "PointPillars: Fast encoders for object detection from point clouds," in *Proc. CVPR*, 2019, pp. 12697–12705.

AKIHIRO KUWABARA received the B.S. degree from Osaka University, Osaka, Japan, in 2024, where he is currently pursuing the M.S. degree with the Graduate School of Information Science and Technology. His research interests include point cloud compression and delivery.

SORACHI KATO received the B.E. and M.E. degrees from Osaka University, Japan, in 2021 and 2023, respectively, where he is currently pursuing the Ph.D. degree with the Graduate School of Information Science and Technology. He is a Research Fellow (DC1) of Japan Society for the Promotion of Science in 2023. From 2023 to 2024, he was an Intern with Mitsubishi Electric Research Laboratories working with the Signal Processing Group. His research interests are in the areas of RF sensing, deep neural signal processing, and multimedia neural compression. He received the Outstanding Paper Award from the Information Processing Society of Japan in 2022.

TOSHIAKI KOIKE-AKINO (Senior Member, IEEE) received the B.S. degree in electrical and electronics engineering, the M.S. and Ph.D. degrees in communications and computer engineering from Kyoto University, Kyoto, Japan, in 2002, 2003, and 2005, respectively. From 2006 to 2010 he was a Postdoctoral Researcher with Harvard University. He is currently a Distinguished Research Scientist with Mitsubishi Electric Research Laboratories, Cambridge, MA, USA. He received the YRP Encouragement Award 2005, the 21st TELECOM System Technology Award, the 2008 Ericsson Young Scientist Award, the IEEE GLOBECOM'08 Best Paper Award in Wireless Communications Symposium, the 24th TELECOM System Technology Encouragement Award, and the IEEE GLOBECOM'09 Best Paper Award in Wireless Communications Symposium. He is a Fellow of Optica.

TAKUYA FUJIHASHI (Member, IEEE) received the B.E. and M.S. degrees from Shizuoka University, Japan, in 2012 and 2013, respectively, and the Ph.D. degree from the Graduate School of Information Science and Technology, Osaka University, Japan, in 2016, where he has been an Assistant Professor since April 2019. He was a Research Fellow (PD) of Japan Society for the Promotion of Science in 2016. From 2014 to 2016, he was a Research Fellow (DC1) of Japan Society for the Promotion of Science. From 2014 to 2015, he was an Intern with the Mitsubishi Electric Research Laboratories working with the Electronics and Communications Group. He was selected as one of the Best Paper candidates in IEEE International Conference on Multimedia and Expo in 2012. His research interests are in the area of video compression and communications, with a focus on multi-view video coding and streaming.