

Rateless Deep Joint Source Channel Coding for 3D Point Cloud

Fujihashi, Takuya; Koike-Akino, Toshiaki; Watanabe, Takashi

TR2025-069 June 04, 2025

Abstract

To enable untethered volumetric services, one of the primary challenges lies in reconstructing a high-quality 3D point cloud for each user, despite variations in channel conditions and bandwidth availability. Conventional digital point cloud coding approaches, such as tree-based or graph-based methods, reduce transmission data effectively; however, they are prone to quality degradation due to lossy compression and entropy coding, especially under fluctuating channels and limited bandwidth. To address these limitations, we propose a novel scheme inspired by deep joint source-channel coding (DJSCC). DJSCC compresses the 3D point cloud into coded symbols, which are then decoded via a graph auto-encoder (GAE) architecture. Additionally, the coded symbols are directly mapped to transmission symbols through analog modulation, allowing the point cloud quality to adapt dynamically to each user's channel conditions. Unlike existing DJSCC schemes, our proposed scheme includes a non-uniform dropout mechanism that provides a rateless feature, enabling the point cloud quality to enhance based on the available bandwidth progressively. Experimental results with a point cloud dataset demonstrate that the proposed scheme mitigates quality degradation due to variations in channel quality and bandwidth better than the existing DJSCC scheme. Moreover, the proposed scheme shows improved generalization performance, leading to superior point cloud quality in narrow-band scenarios.

IEEE Access 2025

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Rateless Deep Joint Source Channel Coding for 3D Point Cloud

SHOICHI IBUKI¹, (Non Member, IEEE), TSUBASA OKAMOTO¹, (Non Member, IEEE), TAKUYA FUJIHASHI¹, (Member, IEEE), TOSHIAKI KOIKE-AKINO², (Member, IEEE) TAKASHI WATANABE¹, (Member, IEEE),

¹Graduate School of Information Science and Technology, Osaka University, Osaka 565-0871, Japan

²Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA, USA

Corresponding author: Shoichi Ibuki (e-mail: ibuki.shoichi@ist.osaka-u.ac.jp).

S. Ibuki's work was supported by JSPS KAKENHI Grant Number and JP22H03582.

ABSTRACT To enable untethered volumetric services, one of the primary challenges lies in reconstructing a high-quality 3D point cloud for each user, despite variations in channel conditions and bandwidth availability. Conventional digital point cloud coding approaches, such as tree-based or graph-based methods, reduce transmission data effectively; however, they are prone to quality degradation due to lossy compression and entropy coding, especially under fluctuating channels and limited bandwidth. To address these limitations, we propose a novel scheme inspired by deep joint source-channel coding (DJSCC). DJSCC compresses the 3D point cloud into coded symbols, which are then decoded via a graph auto-encoder (GAE) architecture. Additionally, the coded symbols are directly mapped to transmission symbols through analog modulation, allowing the point cloud quality to adapt dynamically to each user's channel conditions. Unlike existing DJSCC schemes, our proposed scheme includes a non-uniform dropout mechanism that provides a rateless feature, enabling the point cloud quality to enhance based on the available bandwidth progressively. Experimental results with a point cloud dataset demonstrate that the proposed scheme mitigates quality degradation due to variations in channel quality and bandwidth better than the existing DJSCC scheme. Moreover, the proposed scheme shows improved generalization performance, leading to superior point cloud quality in narrow-band scenarios.

INDEX TERMS Point cloud, Deep Joint Source-Channel Coding

I. INTRODUCTION

The integration of virtual and physical worlds facilitates more precise simulations, supports informed decision-making, and enhances interactivity, propelling advancements in areas such as smart manufacturing and immersive experiences. To achieve this, volumetric content, including three-dimensional (3D) point clouds [1], plays a crucial role in replicating virtual or physical environments on remote devices. The 3D point cloud format is commonly used to represent 3D geometric information, comprising a collection of 3D points, each defined by coordinates (X, Y, Z). Transmitting 3D point clouds over wireless and mobile networks offers the capability to recreate these environments on multiple untethered devices. However, as these devices are mobile and operate in varied surroundings, they may encounter differing channel conditions and bandwidth availability. The difficulty in wire-

less point cloud transmission lies in ensuring the delivery of high-quality point clouds to multiple untethered users despite variations in channel quality and bandwidth constraints.

A common approach employs digital joint source-channel coding, where the point cloud undergoes sequential digital compression and transmission. In this process, digital compression encodes the 3D coordinates of the point cloud into a bitstream in a lossy manner. Specifically, tree-based coding like point cloud library (PCL) [2], [3] and Draco [4] are commonly employed to compress the 3D coordinates. In addition, the graph-based point cloud compression (PCC) treats a point cloud as a graph signal and uses graph signal processing (GSP) [5], [6] for compression [7]–[9]. In the transmission phase, redundancy is added to the bitstream through channel coding, followed by bit-to-symbol mapping to enhance transmission resilience. Here, the transmitter ad-

justs source and channel coding parameters based on the channel quality observed across multiple users, ensuring that the number of transmission symbols matches the lowest available bandwidth among users.

While the combination of digital compression and transmission is ideal for point-to-point communication [10], it leads to quality degradation for each user due to bandwidth variability and fluctuating channel conditions. Generally, the encoding parameters for 3D coordinates are set to match the transmission rate to the minimum available bandwidth across users, preventing playback stalls. This approach, however, restricts reconstruction quality since any extra bandwidth available to certain users does not contribute to quality improvements. Moreover, unstable channel quality may drop and even cap the reconstructed point cloud quality. Specifically, despite the use of channel coding, users do not decode the point cloud if the signal-to-noise ratio (SNR) or signal-to-interference-plus-noise ratio (SINR) falls below a required threshold. Additionally, while some users may experience better channel quality than others, it does not enhance the point cloud quality, as quantization errors remain unrecoverable on the user side.

Recently, DJSCC [11]–[14], also known as semantic communication, has emerged as a method for overcoming the limitations of digital-based schemes under fluctuating channel conditions. Inspired by analog joint source channel coding (JSCC) [15]–[19], these schemes aim to address the challenges posed by wireless channel instability. Existing DJSCC schemes use an auto-encoder (AE) architecture based on deep convolutional neural networks (DCNN) [20]–[22]. This architecture encodes each image into a fixed number of coded symbols, which are then mapped directly to in-phase (I) and quadrature-phase (Q) modulation for transmission. The AE architecture reconstructs the image, text, or video content from the received symbols. We introduced a graph auto-encoder (GAE) architecture [23] based on graph neural networks (GNN) [24], [25] specifically for DJSCC in 3D point cloud transmission, as illustrated in Fig. 1. These DJSCC schemes combine lossy compression through the AE architecture with direct symbol mapping, enabling adaptive reconstruction quality that aligns with real-time channel conditions and avoids decoding failures.

However, in conventional AE designs, all coded symbols are treated equally, which can result in quality degradation, especially in multi-user settings. When users have differing bandwidth availability, the number of coded symbols from the AE architecture is set to match the maximum bandwidth among the users. Users with limited bandwidth then receive fewer symbols, leading to significant quality degradation.

Existing schemes for wireless and mobile channels often lead to quality degradation and saturation due to unstable channel conditions and varying bandwidth among users. To address these issues, this paper introduces a novel DJSCC scheme for untethered point cloud transmission. To accommodate bandwidth diversity, the proposed scheme incorporates a rateless property within the GAE architecture, inspired

by the weighted dropout approach [26]. In the proposed GAE architecture, weighted dropout is applied during model training to adjust the relative importance of each coded symbol. This weighted dropout introduces unequal significance across the symbols, allowing the proposed scheme to transmit symbols progressively to users. Consequently, each user can gradually enhance the quality of the reconstructed point cloud, depending on the available bandwidth.

Evaluations on a point cloud dataset demonstrate that the rateless property in the proposed scheme enables a gradual improvement in the reconstructed point cloud quality, tailored to the bandwidth availability of each user. In contrast, the existing GAE-based DJSCC scheme shows a substantial quality loss for users who lack sufficient bandwidth to receive all coded symbols.

The contributions of this study are as follows:

- We present the first GAE architecture specifically designed to address channel quality fluctuations and bandwidth variations among users.
- Our scheme incorporates a weighted dropout mechanism based on the power function during the training phase, enabling the proposed GAE architecture to achieve a rateless property.
- We enhance the point cloud quality by adding a random feature [27] to each node in the GAE architecture, thereby improving quality without increasing transmission traffic.

In our preliminary work [28], we evaluated the proposed architecture using point clouds from a single category to highlight its foundational performance. However, if the architecture performs effectively only within the trained category, it would require a separate trained decoder for each point cloud category.

In contrast, this paper examines the generalization capabilities of the proposed GAE architecture. To this end, we train with point clouds from all categories, demonstrating that the rateless property reduces bandwidth requirements to achieve consistent 3D reconstruction quality across various categories. Moreover, the proposed scheme employs a unified decoder architecture to reconstruct multiple point cloud categories, effectively reducing communication overhead for point cloud reconstruction.

II. RELATED WORK

A. DEEP JOINT SOURCE CHANNEL CODING

Recent studies [11], [12] have developed DJSCC schemes for wireless multimedia delivery. The DCNN-based AE architecture presented in [11] represents a foundational work in DJSCC. This architecture encodes images into latent variables through a series of two-dimensional (2D) convolution layers with Leaky rectified linear unit (ReLU) activation. These latent variables are then power-normalized for wireless transmission. The received variables are processed through a series of 2D deconvolution layers to reconstruct pixel values, with the AE architecture trained using an mean squared

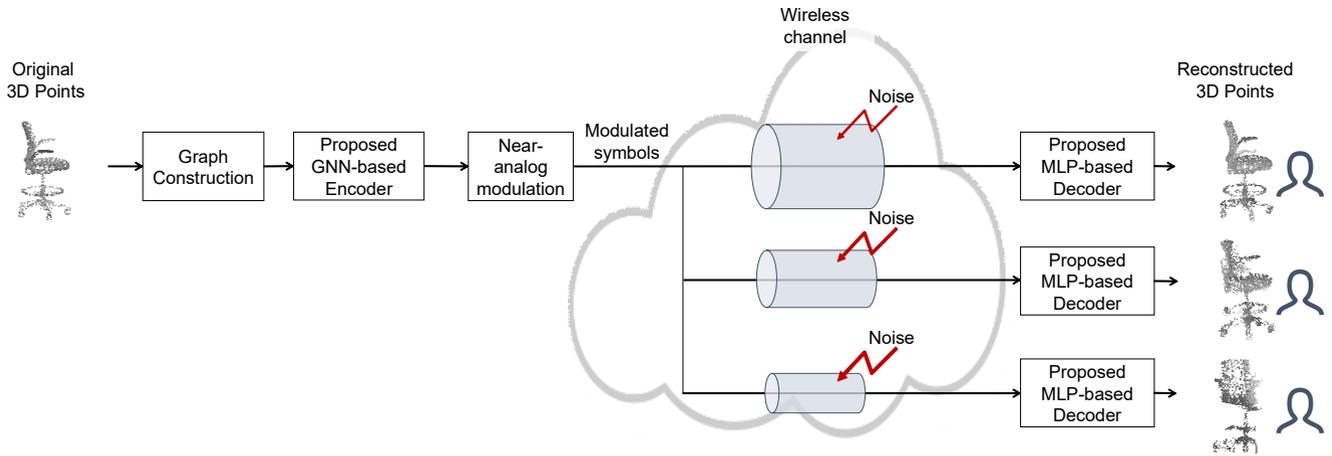


FIGURE 1. Schematics of the proposed graph-based DJSCC scheme.

error (MSE) loss function. Various studies have since experimented with alternative encoder-decoder architectures, including transformers [29], swin transformers [30], self-attention [31], [32], cross-attention [33], pre-trained generative adversarial networks (GAN) generator [34], feedback mechanisms [35], layered decoders [36], and channel-guided decoders [37]–[39] to enhance quality. Others have introduced semantic loss functions like multi-scale structural similarity index measure (MS-SSIM) [40] and learned perceptual image patch similarity (LPIPS) [41] to improve the perceptual quality of reconstructed images.

The proposed scheme contributes to the DJSCC framework by targeting 3D point clouds. While existing studies [23], [42]–[44] have developed deep graph convolutional networks (DGCN) and 2D/3D convolutional networks for point cloud encoding and decoding, our study uniquely integrates power function-based weighted dropout into the DJSCC scheme. This integration enables a rateless property in the coded symbols, enhancing generalization across diverse point cloud categories, particularly in bandwidth-constrained environments.

B. ANALOG JOINT SOURCE CHANNEL CODING

Analog JSCC schemes have been developed for wireless multimedia delivery to counteract quality degradation from fluctuating channel conditions [45]. A seminal work in this area is SoftCast [15], designed specifically for video transmission. SoftCast applies 3D-discrete cosine transform (DCT) to video signals and scales the coefficients before analog modulation, ensuring that the quality of the reconstructed video adapts proportionally to real-time channel conditions. Other studies [46]–[48] have refined these schemes with a focus on human visual perception, such as optimizing scaling operations in CV-Cast [48] to enhance accuracy for computer vision tasks rather than minimizing pixel distortion. Research [17], [49]–[52] has further extended analog JSCC to immersive and volumetric content. For example, HoloCast [50]–[52] employed graph Fourier transform (GFT)

for energy-efficient compression of 3D point clouds before analog modulation.

In line with these developments, our goal is to enable wireless 3D point cloud transmission that mitigates quality degradation and saturation due to channel variability. To achieve this, we propose a GAE architecture with a rateless property, which converts 3D point clouds into coded symbols. While both HoloCast and our scheme utilize GSP for JSCC, the proposed GAE architecture minimizes communication overhead during point cloud decoding by eliminating the need to share the graph basis matrix, a requirement that imposes significant overhead in GFT-based JSCC.

III. SYSTEM MODEL

In the system model, the transmitter sends a point cloud consisting of N 3D points, and the user reconstructs the same point cloud, maintaining N 3D points. The main objectives of this communication process are twofold: 1) to compress and decompress the point cloud using either traditional or neural networks (NN)-based encoding and decoding modules, and 2) to reconstruct the point cloud while addressing challenges posed by wireless channels, including fluctuating channel quality and varying bandwidth availability.

A. TRANSMITTER

The proposed system model is depicted in Figs. 2 (a) and (b). At the transmitter, the input is treated as a graph signal, where a 3D point cloud is represented as an unweighted, undirected graph $\mathcal{G} = (\mathbf{P}, \mathcal{E}, \mathbf{W})$. In this graph, \mathbf{P} and \mathcal{E} represent the sets of vertices and edges, respectively, with 3D points serving as vertices, and each point's 3D coordinates $\mathbf{p} = [x, y, z]^T \in \mathbb{R}^3$ as vertex attributes. To define connections among vertices, we employ a K -nearest neighbor graph. The adjacency matrix \mathbf{W} holds positive edge weights, with a binary structure indicating either 1 (connection) or 0 (no connection).

The attribute of 3D coordinates for N points is given by $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N]^T \in \mathbb{R}^{N \times 3}$, which is then mapped

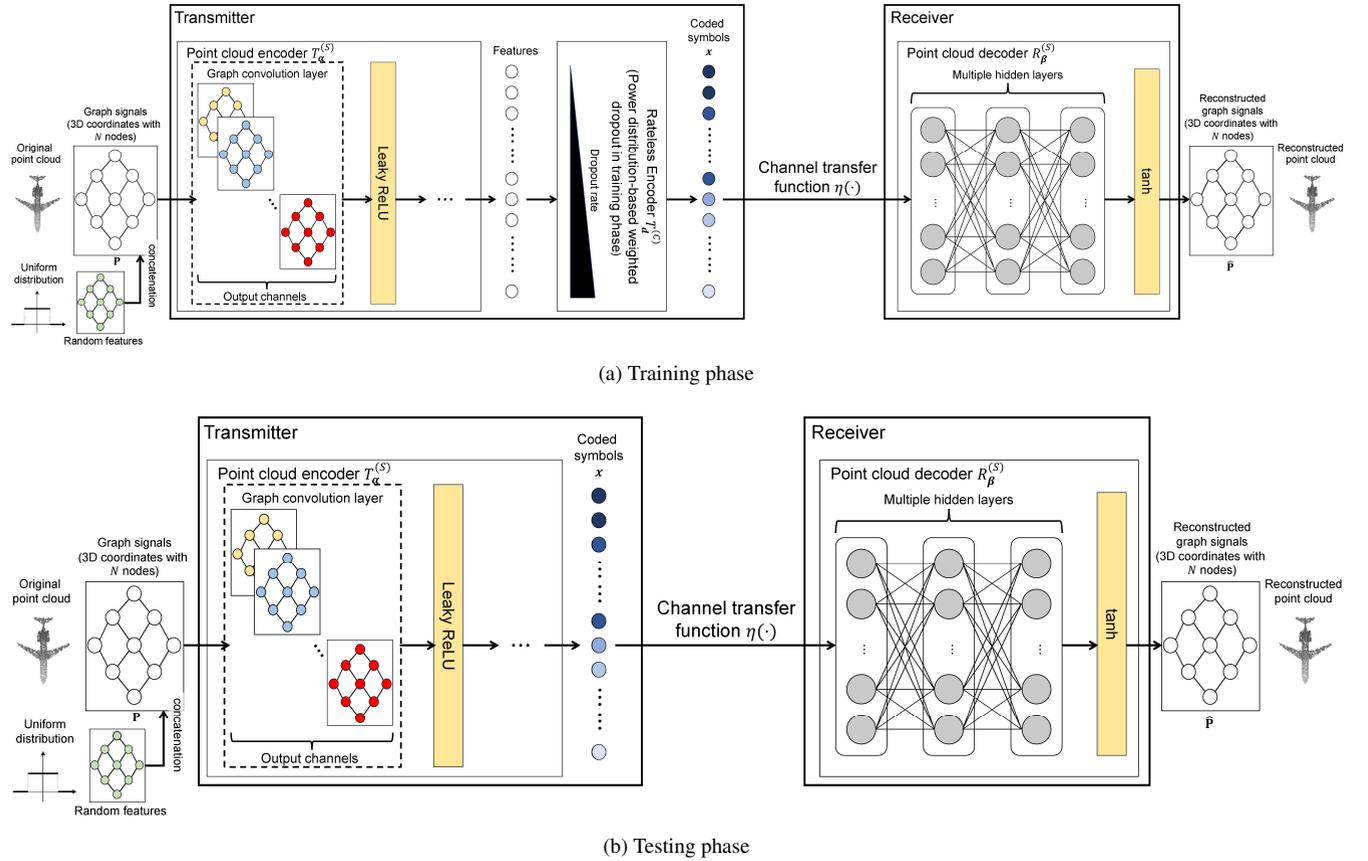


FIGURE 2. Proposed end-to-end transmitter and user in training and testing phases for wireless point cloud delivery.

into symbols, \mathbf{x} , for transmission over physical channels. The transmitter is composed of a point cloud encoder and a rateless encoder. The point cloud encoder uses a graph convolutional neural networks (GCNN), while the rateless encoder is based on weighted dropout, applied only during the training phase (Fig. 2 (a)). This setup provides a rateless property to the point cloud encoder’s coded symbols in the testing phase (Fig. 2 (b)). Let α represent the GCNN parameters for the point cloud encoder and d the weighted dropout distribution. The encoded symbol sequence, \mathbf{x} , in the training and testing phases can be defined as follows:

$$\mathbf{x} = \begin{cases} T_d^{(C)}(T_\alpha^{(S)}(\mathbf{P})), & \text{training phase} \\ T_\alpha^{(S)}(\mathbf{P}), & \text{testing phase} \end{cases} \quad (1)$$

where $T_\alpha^{(S)}(\cdot)$ and $T_d^{(C)}(\cdot)$ denote the point cloud encoder and rateless encoder with respect to parameters α and distribution d , respectively. Finally, the transmitted symbols $\mathbf{x} \in \mathbb{R}^k$ are normalized to satisfy the average transmit power constraint P , where k is the length of the symbols. Specifically, the normalization for the transmitted symbols $\hat{\mathbf{x}}$ is according to:

$$\hat{\mathbf{x}} = \mathbf{x} \cdot \sqrt{\frac{kP}{\sum_{x_i \in \mathbf{x}} x_i^2}}, \quad (2)$$

such that the power-normalized symbols $\hat{\mathbf{x}}$ satisfy the average transmit power constraint P . It ensures that the average power of the transmitted symbols remains within a constraint typically imposed by system design specifications or regulatory standards.

The power-normalized symbols, $\hat{\mathbf{x}}$, were transmitted directly over a physical channel, bypassing the traditional bit-to-symbol mapping by mapping the symbols directly to I and Q components for analog modulation. The wireless channel, represented by η , processes the input $\hat{\mathbf{x}}$ and produces the received signal \mathbf{y} . If \mathbf{h} is the channel coefficient, then the channel transfer function from the transmitter to the receiver can be modeled as:

$$\mathbf{y} = \eta(\hat{\mathbf{x}}) = \begin{cases} \mathbf{h}\hat{\mathbf{x}} + \mathbf{n} & \text{received} \\ 0 & \text{missed} \end{cases} \quad (3)$$

where $\mathbf{n} \sim \mathcal{CN}(0, \mathbf{I}\sigma^2)$ is a vector of additive white Gaussian noise (AWGN) with mean zero and variance σ^2 , \mathbf{I} is the identity matrix, \sim denotes “distributed as”, and $\mathcal{CN}(a, b)$ represents a complex Gaussian distribution with mean a and variance b . Symbols that are missed due to insufficient bandwidth are treated as zeros by the receiver.

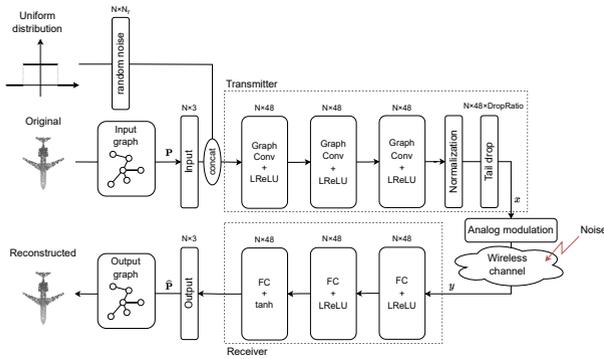


FIGURE 3. Proposed GAE structure.

B. RECEIVER

The receiver is equipped with a point cloud decoder that is responsible for reconstructing the point cloud from the received symbols. Let β represent the NN parameters of the point cloud decoder. The decoded point cloud, $\hat{\mathbf{P}}$, can be derived from the received signal, \mathbf{y} , as follows:

$$\hat{\mathbf{P}} = R_{\beta}^{(S)}(\mathbf{y}) \quad (4)$$

where $R_{\beta}^{(S)}(\cdot)$ denotes the point cloud decoder with respect to the parameters β .

The objective of the proposed scheme is to reconstruct the 3D coordinates of the point cloud as accurately as possible relative to the original coordinates. This is treated as a signal reconstruction task, aiming to minimize the errors between the original point cloud \mathbf{P} and the reconstructed point cloud $\hat{\mathbf{P}}$. The chamfer distance (CD) $\mathcal{L}_{CD}(\alpha, \beta)$ is employed as the loss function to measure the distance between \mathbf{P} and $\hat{\mathbf{P}}$, defined as:

$$\begin{aligned} \mathcal{L}_{CD}(\alpha, \beta) &= \frac{1}{2} \left\{ \frac{1}{|\mathbf{P}|} \sum_{\mathbf{p} \in \mathbf{P}} \min_{\hat{\mathbf{p}} \in \hat{\mathbf{P}}} \|\mathbf{p} - \hat{\mathbf{p}}\|_2 + \frac{1}{|\hat{\mathbf{P}}|} \sum_{\hat{\mathbf{p}} \in \hat{\mathbf{P}}} \min_{\mathbf{p} \in \mathbf{P}} \|\mathbf{p} - \hat{\mathbf{p}}\|_2 \right\}, \end{aligned} \quad (5)$$

The term $\min_{\hat{\mathbf{p}} \in \hat{\mathbf{P}}} \|\mathbf{p} - \hat{\mathbf{p}}\|_2$ ensures that each 3D coordinate \mathbf{p} in the original point cloud \mathbf{P} finds a close match in the reconstructed point cloud $\hat{\mathbf{P}}$, while the term $\min_{\mathbf{p} \in \mathbf{P}} \|\mathbf{p} - \hat{\mathbf{p}}\|_2$ enforces the same matching condition in the reverse direction.

IV. PROPOSED SCHEME

We designed a GAE architecture to address the challenges posed by varying channel quality and bandwidth availability among multiple users. Fig. 3 shows the detail of the proposed GAE architecture. The architecture includes a set of GCNN for the point cloud encoder, power function-based weighted dropout for the rateless encoder, and a multi layer perceptron (MLP) for the point cloud decoder. The weights of the encoder and decoder architectures in the proposed system are trained offline using a point cloud dataset and simulated

channels, such as AWGN. Once the proposed encoder and decoder are fully trained, the trained decoder is shared with the receiver to decode the received coded symbols transmitted from the transmitter.

A. POINT CLOUD ENCODER

The input to the proposed point cloud encoder, denoted as $\mathbf{P} \in \mathbb{R}^{N \times 3}$, consists of the 3D coordinates of a point cloud with N points. This point cloud is sampled from the dataset \mathcal{D} , which contains point clouds grouped into specific categories, such as airplanes, bags, and chairs.

Additionally, the proposed scheme introduces random features $\mathbf{r} \in \mathbb{R}^{N_r}$ with N_r features generated from a uniform distribution. Previous work [27] has shown that adding random features enhances the theoretical capacity of GNN architectures for various tasks. Following these findings, random features $\mathbf{R} = [\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_{N_r}]^T$ are concatenated with the input \mathbf{P} , forming $\text{CONCAT}([\mathbf{P}, \mathbf{R}]) \in \mathbb{R}^{N \times (N_r + 3)}$, where $\text{CONCAT}([\cdot, \cdot])$ represents concatenation along the feature dimension.

The concatenated features were processed through the GCNN-based point cloud encoder. This encoder comprises three consecutive graph convolution layers, each followed by a leaky ReLU (LReLU) activation function, with trainable parameters α , where each layer aggregates neighboring node features to sum up the features from the local neighborhood. Using the adjacency matrix \mathbf{W} , the graph convolution layers extract graph signal features, while the nonlinear activation function enables learning a nonlinear mapping from the source signals to the coded symbols. The encoder produces L channel features, $\mathbf{b} \in \mathbb{R}^{N \times L}$, from \mathbf{P} , where L depends on the configuration of the graph convolutions.

B. RATELESS ENCODER

The rateless encoder then transforms the features \mathbf{b} into coded symbols $\mathbf{x} \in \mathbb{R}^{NL}$ and normalizes these symbols to satisfy the power constraint $\|\mathbf{x}\|^2 = P$.

To achieve the rateless property, the rateless encoder applies non-uniform dropout to the coded symbols during training. This non-uniform dropout is applied across the width of the coded symbols, with increasing dropout rates assigned to each symbol. This approach prioritizes the upper coded symbols during testing, providing a progressive quality enhancement. The dropout rates are controlled by adjusting with a power function τ^γ , where τ represents the compression rate, and γ is an order parameter. Specifically, the rateless encoder first determines the dropout ratio as τ^γ , where τ follows a uniform distribution within the range $[0, 1]$, with a pre-determined order parameter γ . For example, a dropout ratio of 0.2 indicates that 80% of the last coded symbols are dropped, whereas a ratio of 0.8 means that only 20% of the last coded symbols are dropped during training. Based on the dropout ratio, the encoder generates a binary mask for NL coded symbols to determine which symbols are retained or dropped, applying the mask with a computational complexity of $O(NL)$. This complexity is equivalent to that

of a simple (i.e., uniform) dropout. A previous study [26] has shown that this power function performs effectively, and thus, we adopt it for dropout configuration in the subsequent evaluations.

The channel transfer function η takes the reshaped power-normalized symbols \hat{x} as input and produces the received signal \mathbf{y} at the user, as defined in Eq. (3). During the training phase, the proposed scheme synthetically simulates all potential distortions arising from channel coefficients and additive noise to learn optimized weights that minimize the CD over wireless channels.

C. POINT CLOUD DECODER

The received signal $\mathbf{y} \in \mathbb{R}^{N \times L}$ is processed by the point cloud decoder, transforming it into $\hat{\mathbf{b}} \in \mathbb{R}^{N \times L}$. The decoder consists of three consecutive fully connected (FC) layers: the first two layers are followed by leaky ReLU (LReLU) activation functions, while the final layer employs a hyperbolic tangent (tanh) activation function, and a trainable parameter set β .

The point cloud decoder maps the received symbols $\hat{\mathbf{b}}$ into an estimate of the 3D coordinates $\hat{\mathbf{P}} \in \mathbb{R}^{N \times 3}$. The CD loss function is computed at the user and back-propagated to the transmitter, allowing simultaneous updates to the trainable parameters of both the point cloud encoder and decoder.

V. PERFORMANCE EVALUATION

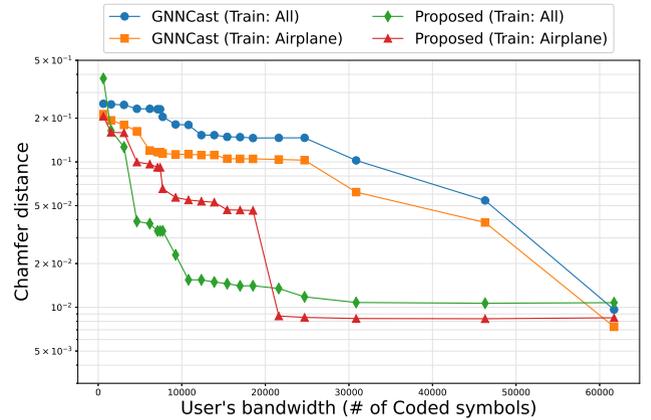
A. SIMULATION SETTINGS

Datasets: We use the ShapeNet benchmark dataset [54] for our experiments. ShapeNet includes approximately 17,000 3D point clouds across 16 categories: Airplane, Bag, Cap, Car, Chair, Earphone, Guitar, Knife, Lamp, Laptop, Motorbike, Mug, Pistol, Rocket, Skateboard, and Table. The dataset is split into training and test sets at a 9:1 ratio. Training data is used for network weight optimization, while test data assesses reconstruction quality.

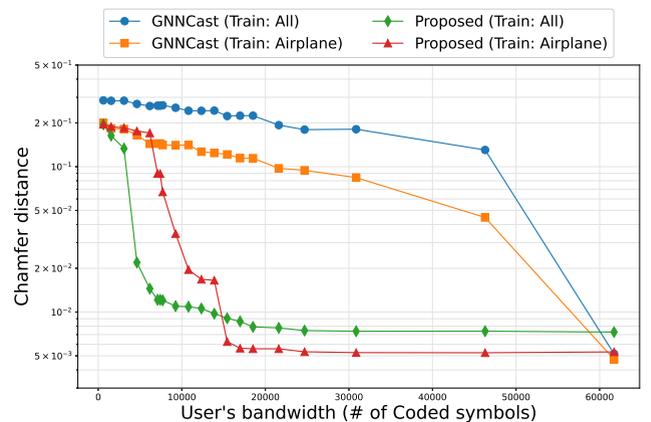
Quality Metric: The CD defined in Eq. 5 is used to evaluate the quality of the reconstructed 3D coordinates.

Wireless Environment: Unless otherwise mentioned, we simulate AWGN channels with noise power σ^2 ranging from 0 dB to -30 dB. The point cloud encoder and decoder are trained in deep graph joint source-channel coding schemes under wireless channel signal-to-noise ratios (SNR) of 10 dB or 20 dB. In Sec. VI-A, we consider Rayleigh fading channels. In such channels, the fading coefficient \mathbf{h} follows a zero-mean complex Gaussian distribution $\mathbf{h} \sim \mathcal{CN}(0, \mathbf{I}\sigma^2)$.

GAE Architecture: Our GAE architecture is implemented using PyTorch Geometric (PyG) [55]. Table 1 outlines the parameter configurations for the point cloud encoder, rateless encoder, and point cloud decoder in the deep graph joint source-channel coding schemes, including the proposed scheme and GNNCast [23]. The number of random features N_r is set to one. At the start of training, the trainable parameters α and β of the encoder and decoder are initialized by default setting of PyTorch and tuned to minimize Eq. (5). We utilize the adaptive momentum (ADAM) optimizer for training, with an initial learning rate of 0.0005, a batch size



(a) Trained SNR: 10dB



(b) Trained SNR: 20dB

FIGURE 4. CD under the different user's available bandwidth.

of 1, a momentum of 0.9, and a momentum2 of 0.999, and a training duration of 200 epochs.

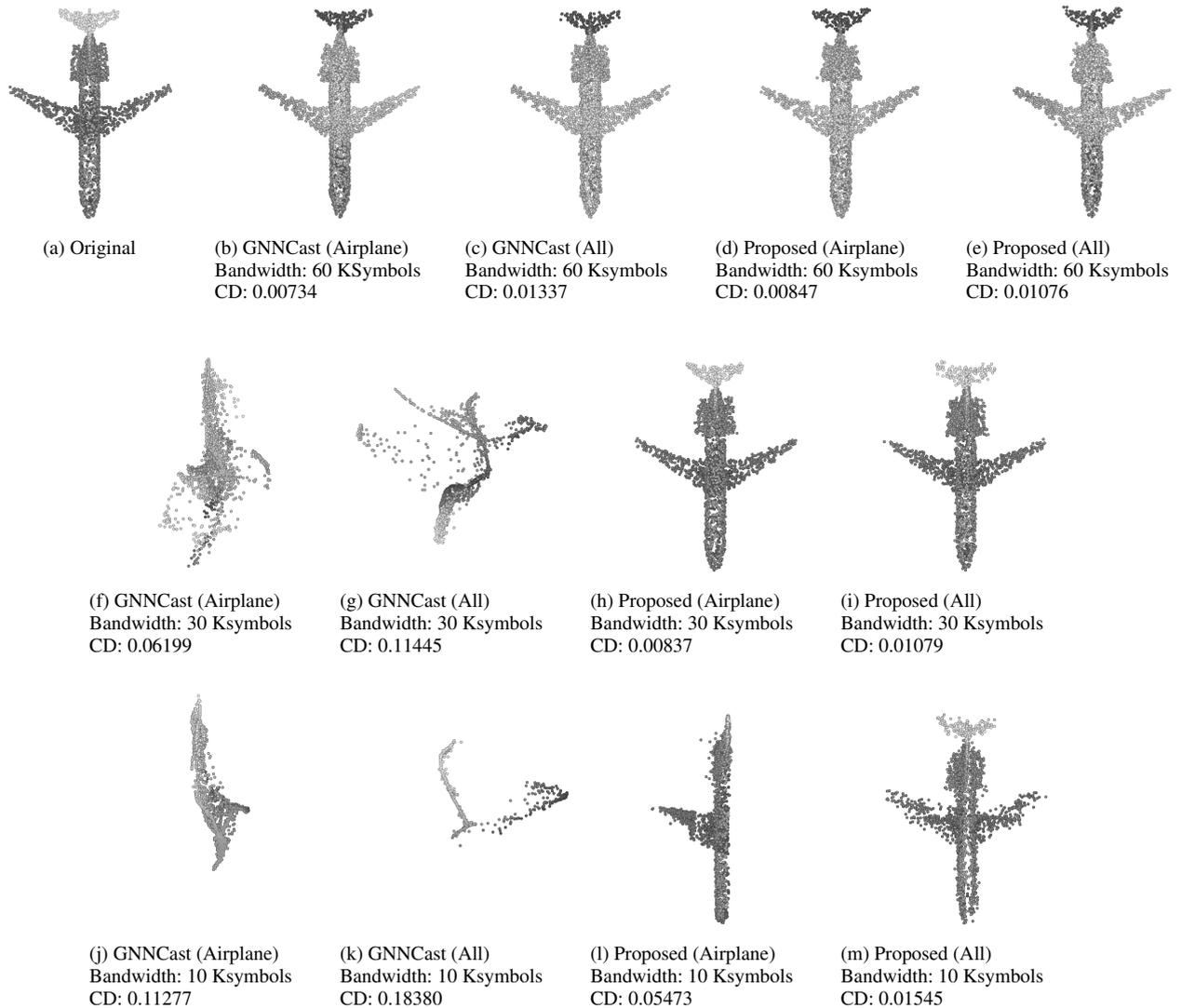
B. EFFECT OF BANDWIDTH AND CHANNEL QUALITY HETEROGENEITY

To evaluate the impact of the rateless property, we compare the reconstructed point cloud quality across users with varying available bandwidths. The baseline for comparison is the DJSCC scheme, GNNCast [23], which utilizes a GCNN-based point cloud encoder and an MLP-based point cloud decoder. The quality improvement in the proposed scheme primarily arises from the integration of the rateless encoder. We assess two variants for training both GNNCast and the proposed scheme: 1) training on point clouds from the ‘‘airplane’’ category only, and 2) training on point clouds from all 16 categories.

Figs. 4 (a) and (b) display the CD of the proposed and baseline schemes under different user bandwidth constraints. In these evaluations, the SNR of the wireless channel is fixed at 10 dB and 20 dB for training and testing in Fig. 4 (a) and (b), respectively. The results indicate that the proposed scheme, trained on point clouds from all categories, con-

TABLE 1. Parameter settings of the existing and proposed deep graph joint source-channel coding.

	Point cloud encoder			Rateless encoder	Point cloud decoder	
	graph convolutions	output channels in each layer	activation		output channels in each layer	activation
Proposed	GraphConv [53]	[48,48,48]	Leaky ReLU	1.0	[48,48,48]	Leaky ReLU
GNNCast	GraphConv [53]	[48,48,48]	Leaky ReLU	0.0	[48,48,48]	Leaky ReLU

**FIGURE 5.** Snapshots of reconstructed 3D point clouds under the different available bandwidths at wireless channel SNR of 10 dB.**TABLE 2.** BD-CD values across proposed and baseline schemes

Method	SNR (dB)	BD-CD
Proposed (Train: All) vs. GNNCast (Train: All)	10	-0.1560
	20	-0.0891
Proposed (Train: All) vs. GNNCast (Train: Airplane)	10	-0.0679
	20	-0.0530
Proposed (Train: All) vs. Proposed (Train: Airplane)	10	-0.0103
	20	-0.0087

sistently maintains high reconstruction quality, even with limited bandwidth, for both channel SNRs. In contrast, the

proposed scheme trained on “airplane” point clouds provides optimal performance when the available bandwidth is up to 34.7% and 25.5% of the maximum bandwidth requirement at channel SNRs of 10 dB and 20 dB, respectively. However, the reconstruction quality sharply declines when bandwidth availability drops further.

For GNNCast trained on the “airplane” category, the highest quality is achieved when the user receives all coded symbols. However, as bandwidth limitations increase, reconstruction quality diminishes due to the loss of essential coded symbols necessary for accurate reconstruction.

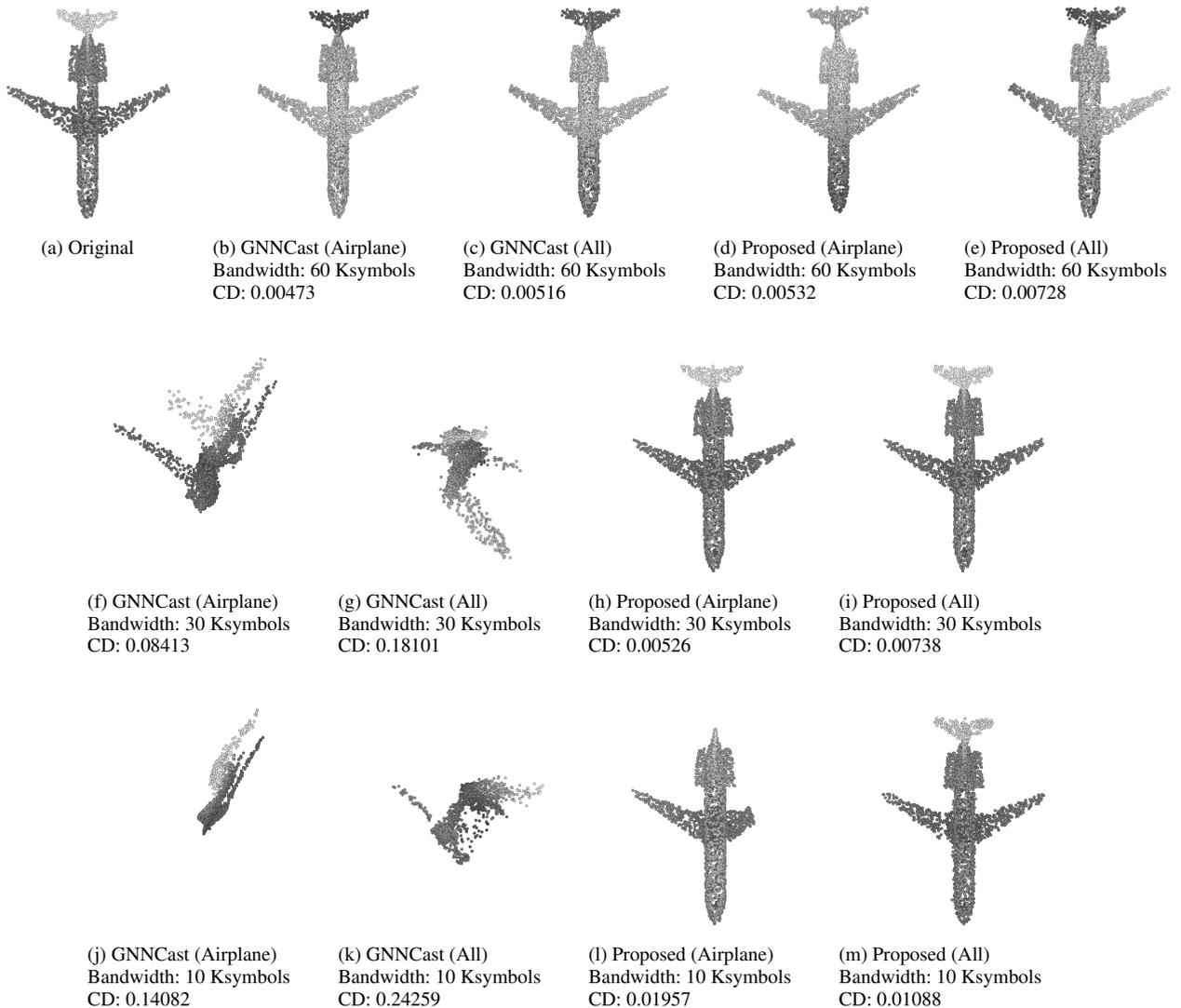


FIGURE 6. Snapshots of reconstructed 3D point clouds under the different available bandwidths at wireless channel SNR of 20 dB.

For the rate-distortion (R-D) performance assessment, we computed the Bjøntegaard delta Chamfer distance (BD-CD) [56] in Figs.4 (a) and (b) to quantify the average difference in CD between R-D curves over the same range of available bandwidth. Negative BD-CD values indicate quality improvements achieved by the proposed scheme over the baseline schemes within the evaluated bandwidth range. For the calculation, we used R-D curves within the bandwidth range of [618, 61,728] symbols.

Table 2 summarizes the BD-CD values for the proposed scheme, trained on all categories, compared to baseline schemes under channel SNRs of 10 and 20 dB. The results demonstrate that the proposed scheme achieves the highest 3D reconstruction quality across both narrowband and broadband environments, outperforming the baseline schemes.

Figs. 5 and 6 present snapshots of the original and reconstructed point clouds under varying user bandwidths at

channel SNRs of 10 dB and 20 dB, respectively. When user bandwidth is sufficient, GNNCast achieves the highest performance. However, the proposed schemes exhibit minimal visual quality degradation. When the bandwidth is reduced to 30 Ksymbols, GNNCast experiences severe degradation, resulting in reconstructed point clouds that no longer resemble the airplane shape, whereas the proposed schemes retain the shape at this bandwidth level. At a further reduced bandwidth of 10 Ksymbols, the proposed scheme trained on the airplane category suffers noticeable distortion, while the scheme trained on all categories preserves the visual quality, especially at a channel SNR of 20 dB. In the proposed scheme trained on the airplane category, the reconstructed 3D point clouds experience non-uniform missing points, particularly in the tail and left wing, as shown in Fig. 6 (l). Detailed evaluations reveal that the 3D points of the tail are the first to be lost at a bandwidth of 15 Ksymbols,

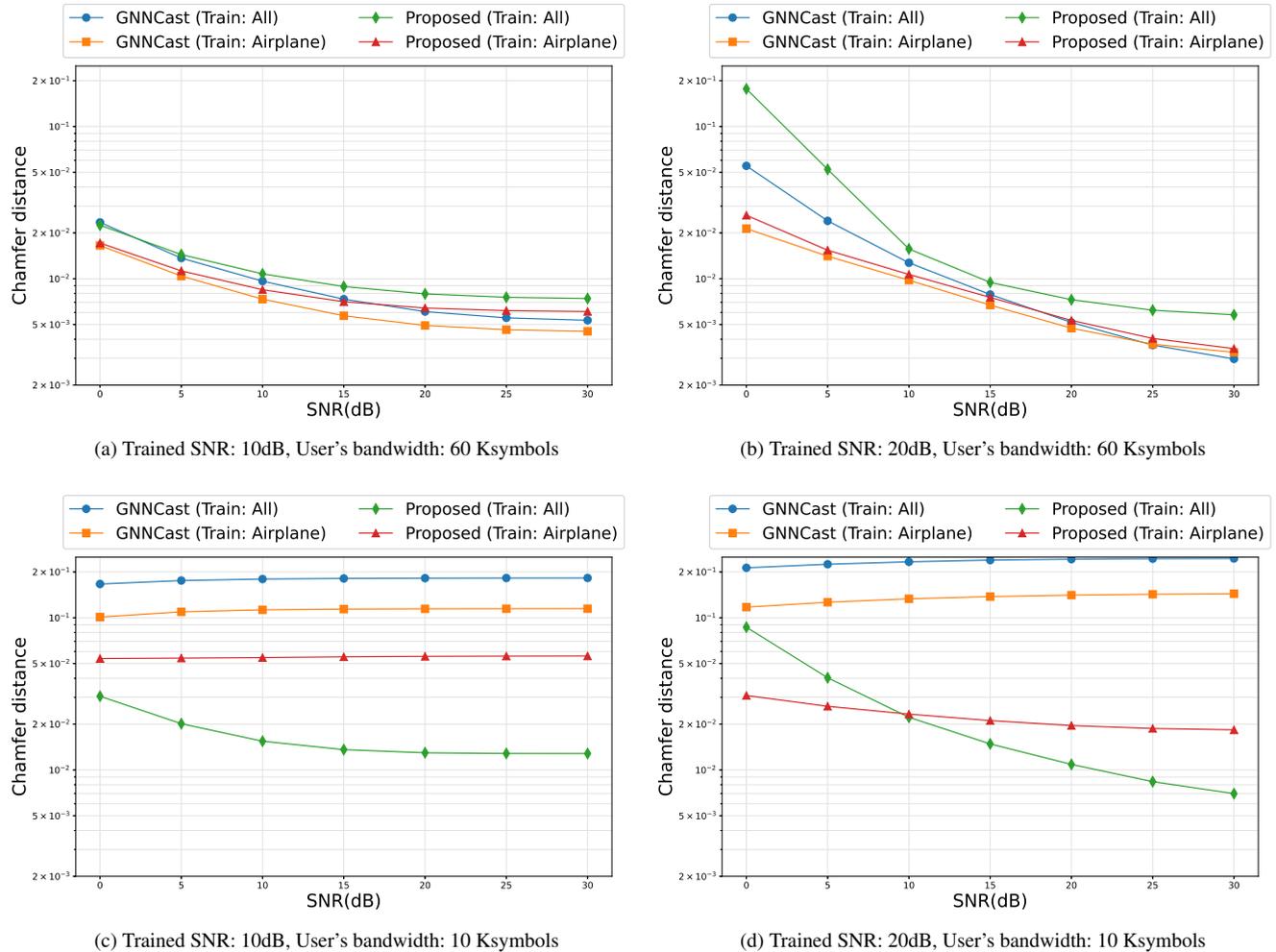


FIGURE 7. CD as a function of wireless channel SNRs under the different available bandwidths.

followed by additional losses in the left wing at a bandwidth of 13 Ksymbols.

In practical wireless settings, users encounter varying channel quality. We evaluate how channel diversity impacts reconstruction quality. Figs. 7 (a) through (d) show the CD as a function of channel SNR across different bandwidth conditions. Specifically, the available bandwidth is set to 60 Ksymbols in Figs. 7 (a) and (b), and 10 K symbols in Figs. 7 (c) and (d), ensuring a fair comparison across schemes. In Figs. 7 (a) and (c), each scheme is trained at a channel SNR of 10 dB, while Figs. 7 (b) and (d) use a training SNR of 20 dB. The key observations are:

- In broadband environments, GNNCast trained on “Airplane” point clouds performs best, indicating that the rateless encoder in the proposed scheme may slightly degrade quality for users who can receive all coded symbols.
- When trained at the channel SNR of 20 dB, reconstruction quality is superior in high-SNR environments but degrades at lower SNR regimes.
- Even in band-limited environments, the proposed

scheme gradually improves CD as channel quality increases, while GNNCast suffers due to discarded critical symbols.

- When trained at 10 dB, the proposed scheme trained on all categories consistently achieves the best quality across varying SNRs.

These results suggest that the proposed scheme trained on all categories demonstrates resilience to both bandwidth limitations and low-quality channels.

C. EFFECT OF CONTENT HETEROGENEITY

The previous evaluations assume that the user requests a point cloud from the “airplane” category. In this section, we analyze reconstruction quality across different categories of 3D point clouds, specifically “chair” and “bag.” For comparison, we include a baseline, “Proposed (Fine-tune)”, where the proposed scheme is initially trained on “airplane” point clouds and then fine-tuned using “bag” or “chair” categories to evaluate generalization performance.

Figs. 8 (a) and (b) show the CD results for the baseline and proposed schemes across varying bandwidths and point

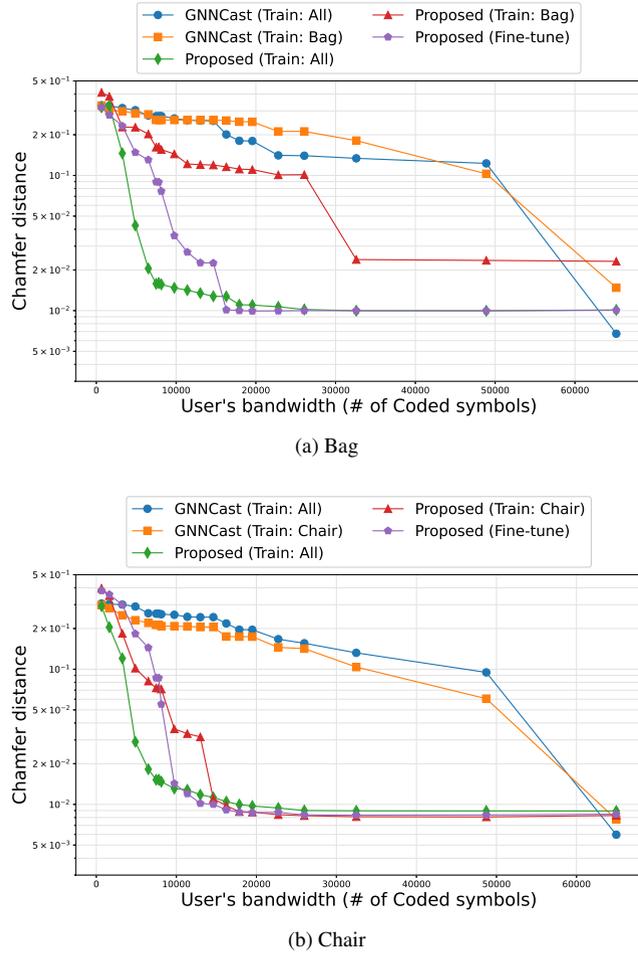


FIGURE 8. CD under different available bandwidths and point cloud categories: Bag and Chair. Here, the channel SNR in the training and testing phases is fixed at 20 dB.

cloud categories, with the channel SNR fixed at 20 dB for both training and testing. The key observations are:

- When the user’s requested point cloud belongs to a different category (e.g., “chair” or “bag”), the proposed scheme trained on all categories achieves the highest reconstruction quality for users with limited bandwidth.
- While the fine-tuned model provides comparable reconstruction quality to the proposed scheme in broadband environments, it suffers from degraded quality in narrow-band scenarios.

Finally, Figs. 9 and 10 present snapshots of the original and reconstructed “bag” and “chair” point clouds under varying bandwidth conditions at a channel SNR of 20 dB, respectively. As seen in Figs. 5 and 6, GNNCast exhibits substantial visual quality degradation for narrow-band users, while the proposed scheme trained on all categories maintains the shape of the reconstructed point cloud under the same bandwidth constraints. The visual quality of the fine-tuned model remains comparable at a bandwidth of 30 Ksymbols but displays clear visual distortion at 10 Ksymbols for both “bag” and “chair” point clouds.

D. ABLATION STUDY

Effect of Weighted/Unweighted Dropout Mechanism

Fig. 11 shows the performance of the proposed scheme with unweighted, i.e., uniform, and weighted dropout mechanisms under different available bandwidths. The proposed weighted dropout prioritizes the upper encoded symbols, preserving high 3D reconstruction quality in narrow-band environments.

Additionally, Fig. 12 discusses the effect of the order parameter γ on 3D reconstruction quality as a function of the available bandwidth. We consider γ values of 0.5, 1.0, and 2.0. The proposed scheme with $\gamma = 0.5$ achieves the highest 3D reconstruction quality when the number of coded symbols exceeds 30,000, i.e., in a broadband environment. However, the quality drops sharply when the number of coded symbols falls below 15,000 in a narrow-band environment. Conversely, with $\gamma = 2.0$, the 3D reconstruction quality remains stable even when the available bandwidth is reduced to 10,000 coded symbols. Nonetheless, when the number of coded symbols exceeds 30,000, the reconstruction quality at $\gamma = 2.0$ is lower than that of $\gamma = 0.5$.

VI. DISCUSSION

A. DISCUSSION ON FADING CHANNELS

In the above sections, the proposed scheme considered that the coded symbols are transmitted over AWGN wireless channels. However, the coded symbols are often transmitted via fading channels, such as Rayleigh fading channels.

To reduce the impact of fading, the proposed scheme can utilize two equalization techniques for the coded symbols: pre-equalization at the transmitter and post-equalization at the receiver. Let \hat{x}_i , h_i , and n_i be i -th power-normalized symbol \hat{x} and corresponding fading coefficient h and effective noise n . In the pre-equalization, the power-normalized symbol \hat{x}_i is pre-equalized as $\bar{x}_i = w_i \hat{x}_i$, where w_i is the pre-equalizer weight. Although there are many variants of pre-equalizer, we consider a simple coherent: $w_i = h_i^* / |h_i|$, where $[\cdot]^*$ denotes the conjugate operation. With this approach, the channel transfer function becomes $\eta(\hat{x}_i) = |h_i| \hat{x}_i + n_i$. In the post-equalization, the receiver takes zero-forcing post-equalization to counteract fading attenuation. Here, the received symbol y_i is equalized as $\hat{y}_i = y_i / h_i$ using the estimated fading coefficient h_i . In this case, the channel transfer function is $\eta(\hat{x}_i) = \hat{x}_i + n_i / h_i$.

In addition, a precoding method can be integrated with the equalization techniques. Specifically, it sorts the symbols \hat{x} according to the fading levels $|h_i|$ in ascending order. This sorting may facilitate optimizing the best-coded symbols to leverage diversity gain for quality enhancement.

Fig. 13 shows the effect of equalization and precoding methods on the performance of the proposed and GNNCast schemes in Rayleigh fading channels. Here, the channel SNR is set to 20 dB. We can see the following observations:

- Post-equalization achieves better CD compared with the pre-equalization.

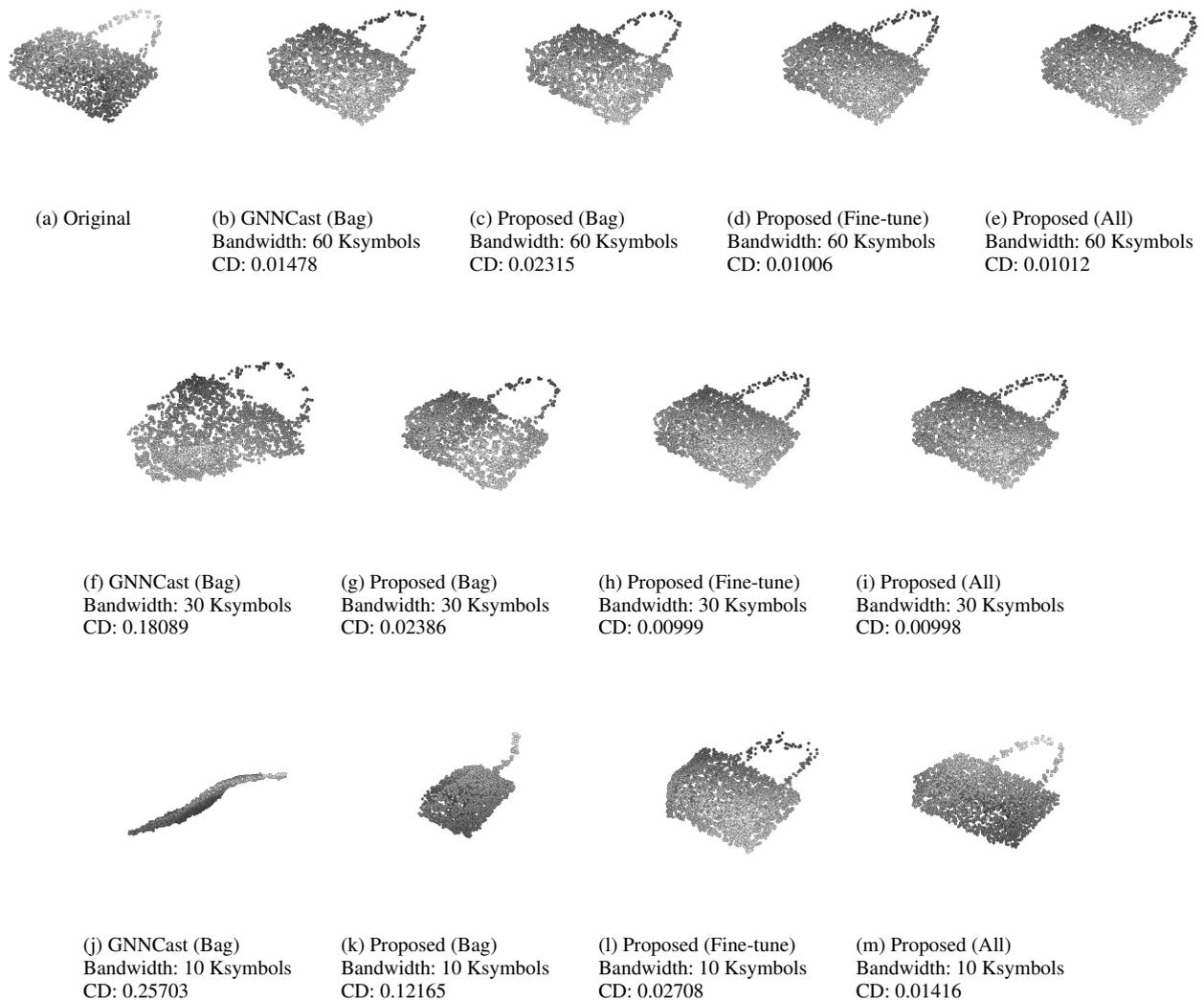


FIGURE 9. Snapshots of reconstructed 3D point clouds of “bag” category under the different available bandwidths at wireless channel SNR of 20 dB.

- Precoding in ascending order maintains superior 3D reconstruction quality, particularly in narrow-band environments.

These results highlight the effectiveness of post-equalization in improving overall performance and demonstrate the utilization of precoding for optimizing 3D reconstruction quality under challenging channel conditions.

B. DISCUSSION ON COMPUTATIONAL COSTS

The proposed scheme encodes and decodes the given point clouds with a certain number of 3D point clouds. This section evaluates the computational cost of the proposed scheme in terms of encoding and decoding time, comparing it with existing digital-based and recent DJSCC schemes for different scales of point clouds. For this purpose, we used point clouds from the ShapeNet (airplane category), KITTI, and longdress datasets, which contain an average of 2,572, approximately

51,000, and 858,000 3D points, respectively. We prepared PCL, Draco, GNNCast, and the point cloud-based semantic communication system (PCSC) [57] for comparison. PCL and Draco are digital-based schemes. PCL provides three compression profiles: “LOW”, “MED”, and “HIGH”. Draco determines the degree of compression for 3D coordinates using the “qp” parameter, where a larger qp value indicates higher 3D reconstruction quality. The default setting is qp 11. PCSC is a DJSCC scheme for point clouds based on 3D convolutional networks and voxelization. We set the voxelization of PCSC to $64 \times 64 \times 64$, following the default setting in [57]. For GNNCast and the proposed scheme, we measure the average encoding and decoding times when transmitting coded symbols ranging from 100% to 1% to the decoder.

Table 3 shows the average encoding and decoding times (in milliseconds) for different point clouds, measured as the mean of ten independent runs. We can see the following key

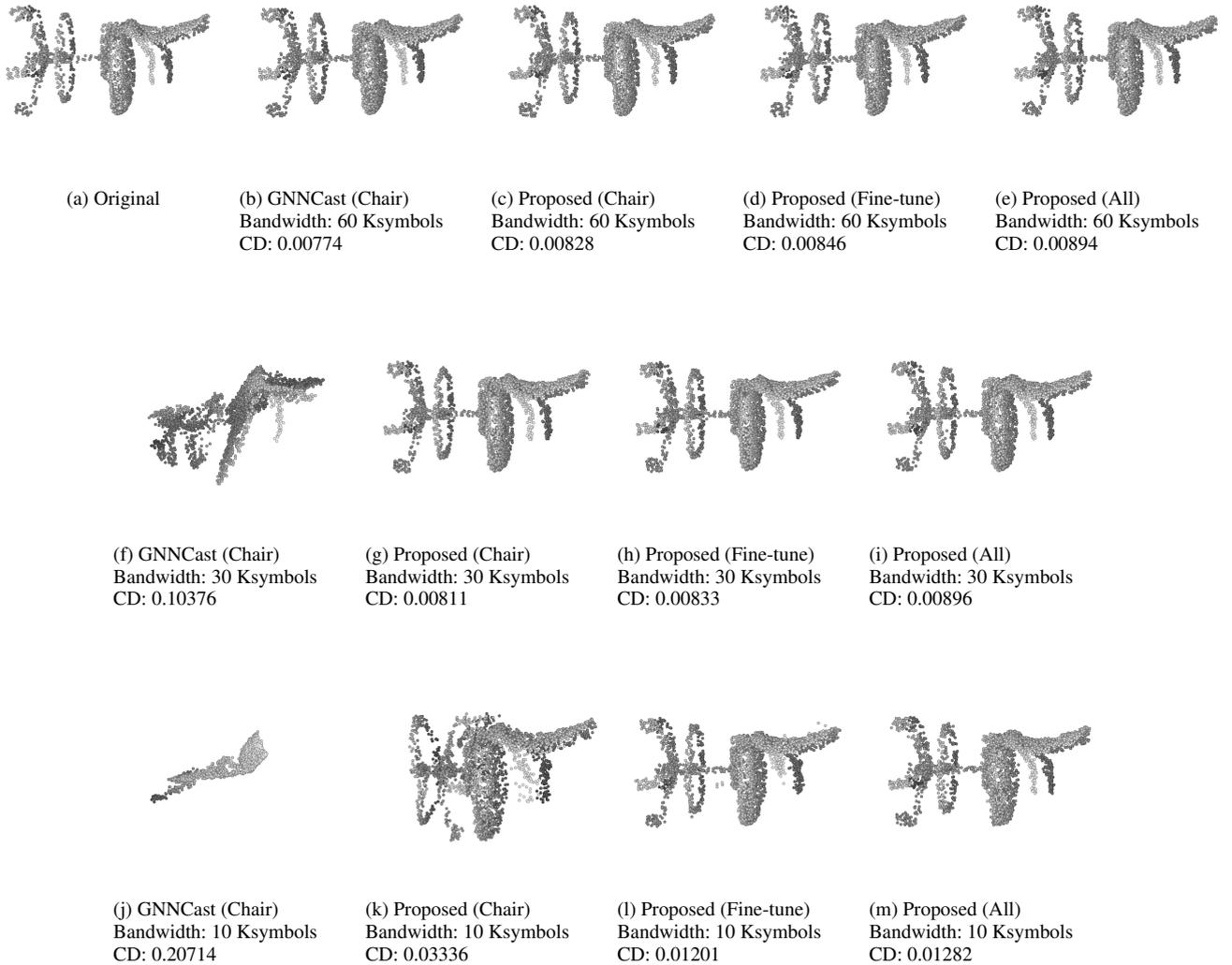


FIGURE 10. Snapshots of reconstructed 3D point clouds of “chair” category under the different available bandwidths at wireless channel SNR of 20 dB.

TABLE 3. Encoding and decoding times for different point clouds.

		Proposed	GNNCast	PCSC	PCL			Draco	
					LOW	MED	HIGH	qp:7	qp:11
ShapeNet Category: Airplane (Avg. 2573 points)	Enc. time (ms)	2.189×10^0	1.925×10^0	4.684×10^3	2.018×10^0	2.962×10^0	7.464×10^0	1.476×10^0	2.626×10^0
	Dec. time (ms)	1.259×10^1	1.906×10^1	3.779×10^5	8.839×10^{-1}	1.631×10^0	6.248×10^0	6.831×10^{-1}	1.307×10^0
KITTI Seq:00-00 (51240 points)	Enc. time (ms)	1.940×10^0	1.918×10^0	1.333×10^4	7.233×10^1	9.117×10^1	1.503×10^2	1.003×10^1	2.046×10^1
	Dec. time (ms)	3.828×10^2	3.719×10^2	6.925×10^5	5.764×10^1	6.960×10^1	1.466×10^2	7.407×10^0	1.669×10^1
KITTI Seq:01-00 (50807 points)	Enc. time (ms)	1.915×10^0	1.918×10^0	2.106×10^4	7.577×10^1	8.867×10^1	1.568×10^2	1.069×10^1	2.043×10^1
	Dec. time (ms)	3.895×10^2	3.538×10^2	1.229×10^6	6.026×10^1	7.430×10^1	1.539×10^2	8.963×10^0	1.808×10^1
KITTI Seq:02-00 (51249 points)	Enc. time (ms)	2.196×10^0	2.000×10^0	1.584×10^4	7.203×10^1	8.750×10^1	1.545×10^2	9.936×10^0	2.120×10^1
	Dec. time (ms)	2.466×10^2	2.894×10^2	8.677×10^5	5.666×10^1	6.925×10^1	1.442×10^2	9.026×10^0	1.627×10^1
SiVFB v2 longdress (857966 points)	Enc. time (ms)	3.333×10^0	3.259×10^0	8.063×10^4	1.762×10^3	1.944×10^3	3.140×10^3	1.834×10^2	3.611×10^2
	Dec. time (ms)	3.865×10^4	3.918×10^4	2.874×10^6	1.584×10^3	1.748×10^3	3.043×10^3	6.582×10^1	1.462×10^2

observations:

- When the 3D points were sampled from the ShapeNet dataset, the encoding and decoding times of the pro-

posed scheme were on a similar scale, i.e., several milliseconds, compared to the low-delay PCL and Draco baselines.

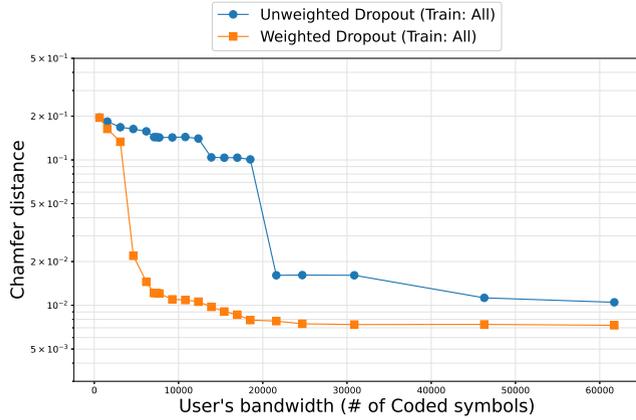


FIGURE 11. Effect of the unweighted and the proposed weighted dropout at channel SNR of 20 dB.

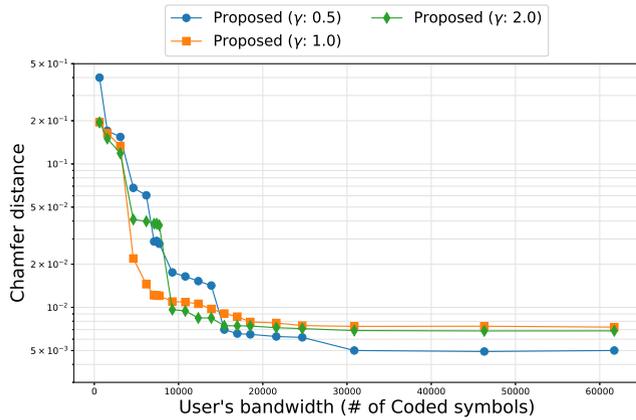


FIGURE 12. CD of the proposed scheme under different order parameter γ in weighted dropout at channel SNR of 20 dB.

- The computational cost of the proposed scheme and GNNCast is comparable since the computational complexity of the simple dropout and the proposed weighted dropout is equivalent.
- The encoding and decoding times of the PCSC are more than 10^3 through 10^4 times longer than those of the proposed scheme, respectively, when using the same ShapeNet point clouds.
- For large-scale point clouds from KITTI and 8iVFB v2, the encoding time of the proposed scheme is even shorter than that of the PCL and Draco baselines. However, the decoding delay is approximately ten times longer than the baselines, indicating that the proposed scheme requires sufficient decoding capability on the user's device.
- Although the gap between the proposed and PCSC schemes becomes smaller at a large-scale 3D point cloud, both the encoding and decoding times of the PCSC scheme remain significantly higher.

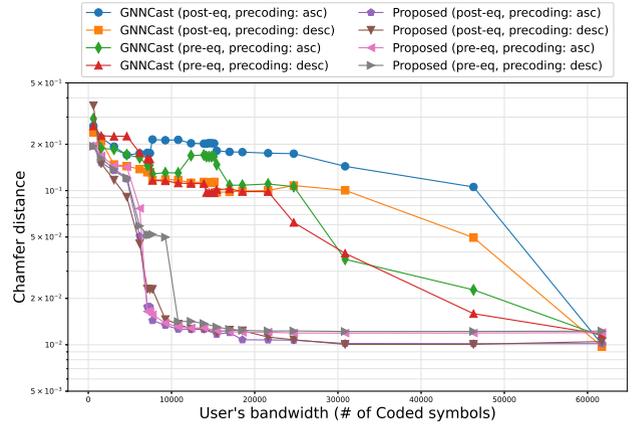


FIGURE 13. CD as a function of available bandwidths in Rayleigh fading channels. Here, channel SNR at 20 dB.

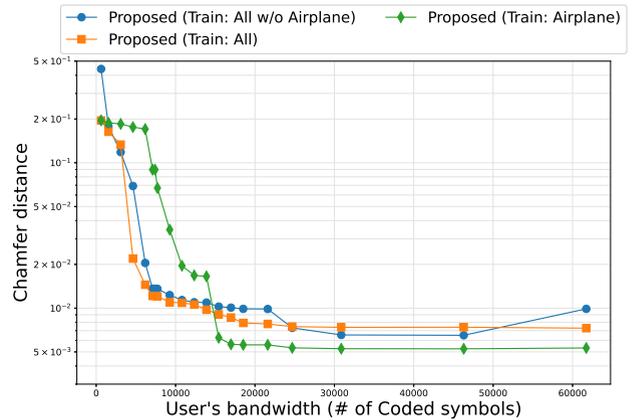


FIGURE 14. CD of proposed scheme variants when the "airplane" category is trained/untrained by proposed GAE architecture

C. DISCUSSION ON UNSEEN CATEGORY

This section examines the potential of the proposed scheme for untrained categories of point clouds. To this end, we introduced an additional baseline. Specifically, we trained the proposed GAE architecture using 15 categories, excluding the airplane point clouds, and then evaluated its performance using the trained GAE architecture.

Fig. 14 shows the performance of the proposed scheme variants when evaluated on airplane point clouds. When trained with all categories, the proposed scheme achieves superior 3D reconstruction quality in narrow-band environments. Interestingly, even when trained with only 15 categories, it still maintains acceptable 3D reconstruction quality for the airplane point clouds, despite them being an unseen category. These results demonstrate the robustness of the proposed scheme and its potential to achieve acceptable 3D reconstruction quality for previously unseen point cloud categories.

VII. CONCLUSION AND FUTURE WORK

We have introduced a novel graph-based DJSCC scheme designed for future 3D applications. This scheme combines a

GCNN-based point cloud encoder, a power function-based rateless encoder, an MLP-based point cloud decoder, and analog modulation to address challenges such as decoding failures, quality saturation from quantization, and quality limitations due to bandwidth diversity. Our evaluations demonstrate that the proposed scheme gradually enhances point cloud quality with improvements in channel conditions and available user bandwidth, enabled by the rateless property. Furthermore, this rateless capability supports generalization across diverse point cloud categories, lowering the bandwidth needed to achieve equivalent 3D reconstruction quality compared to existing DJSCC schemes.

In future work, we will evaluate and train the proposed scheme using alternative metrics, such as perceptual metrics. For this purpose, we consider Wasserstein distance and sliced Wasserstein distance [58], the latter being a low-cost approximation of the former. These metrics are widely recognized for their ability to quantify similarity between two point clouds and have recently been adopted as perceptual quality metrics [59]. Perception-aware metrics can effectively guide the training of the proposed scheme, potentially addressing the limitations of CD in penalizing global structure errors. Further exploration and optimization of such quality metrics to enhance 3D reconstruction quality remain as future work.

REFERENCES

- [1] A. Xiao, J. Huang, D. Guan, X. Zhang, S. Lu, and L. Shao, "Unsupervised point cloud representation learning with deep neural networks: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 9, pp. 11 321–11 339, 2023.
- [2] J. Kammerl, N. Blodow, R. B. Rusu, S. Gedikli, M. Beetz, and E. Steinbach, "Real-time compression of point cloud streams," in *IEEE International Conference on Robotics and Automation*, 2012, pp. 778–785.
- [3] 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 Transactions on Signal and Information Processing*, vol. 9, 2020.
- [4] "Draco 3D data compression," 2022. [Online]. Available: <https://google.github.io/draco/>
- [5] A. Ortega, P. Frossard, J. Kovacevic, J. M. F. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 808–828, 2018.
- [6] T. Fujihashi and T. Koike-Akino, "Graph-based EEG signal compression for human-machine interaction," *IEEE Access*, vol. 12, pp. 1163–1171, 2024.
- [7] P. de Oliveira Rente, C. Brites, J. Ascenso, and F. Pereira, "Graph-based static 3D point clouds geometry coding," *IEEE Transactions on Multimedia*, vol. 21, no. 2, pp. 284–299, 2019.
- [8] J. Zhang, Y. Chen, G. Liu, W. Gao, and G. Li, "Efficient point cloud attribute compression framework using attribute-guided graph fourier transform," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024, pp. 8426–8430.
- [9] N. Miyata, T. Fujihashi, T. Takahashi, S. Saruwatari, and T. Watanabe, "Point cloud geometry and attribute transmission over MIMO channels," in *IEEE Vehicular Technology Conference*, 2024, pp. 1–7.
- [10] C. E. Shannon, "Channels with side information at the transmitter," *IBM journal of Research and Development*, vol. 2, no. 4, pp. 289–293, 1958.
- [11] E. Boursoulatte, D. B. Kurka, and D. Gunduz, "Deep joint source-channel coding for wireless image transmission," *IEEE Transactions on Cognitive Communications and Networking*, vol. 5, no. 3, pp. 567–579, 2019.
- [12] H. Wu, A. Wang, J. Liang, S. Li, and P. Li, "DCSN-Cast: deep compressed sensing network for wireless video multicast," *Signal Processing: Image Communication*, vol. 76, pp. 56–67, 2019.
- [13] D. Gündüz, Z. Qin, I. E. Agerri, H. S. Dhillon, Z. Yang, A. Yener, K. K. Wong, and C.-B. Chae, "Beyond transmitting bits: Context, semantics, and task-oriented communications," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 5–41, 2022.
- [14] Z. Qin, X. Tao, J. Lu, W. Tong, and G. Y. Li, "Semantic communications: Principles and challenges," *arXiv preprint arXiv:2201.01389*, 2021.
- [15] S. Jakubczak and D. Katabi, "A cross-layer design for scalable mobile video," in *ACM Annual International Conference on Mobile Computing and Networking*, 2011, pp. 289–300.
- [16] J. Shen, L. Yu, L. Li, and H. Li, "Foveation based wireless soft image delivery," *IEEE Transactions on Multimedia*, vol. 20, no. 10, pp. 2788–2800, 2018.
- [17] J. Zhao, R. Xiong, and J. Xu, "Omnicast: Wireless pseudo-analog transmission for omnidirectional video," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 1, pp. 58–70, 2019.
- [18] T. Fujihashi, T. Koike-Akino, T. Watanabe, and P. V. Orlik, "High-quality soft video delivery with gmrf-based overhead reduction," *IEEE Transactions on Multimedia*, vol. 20, no. 2, pp. 473–483, 2018.
- [19] —, "FreeCast: Graceful free-viewpoint video delivery," *IEEE Transactions on Multimedia*, vol. 21, no. 4, pp. 1000–1010, 2019.
- [20] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.
- [21] Y. Chen and T. Pock, "Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1256–1272, 2017.
- [22] Y. Tai, J. Yang, X. Liu, and C. Xu, "Memnet: A persistent memory network for image restoration," in *IEEE International Conference on Computer Vision (ICCV)*, 2018, pp. 4549–4557.
- [23] T. Fujihashi, T. K. Akino, S. Chen, and T. Watanabe, "Wireless 3D point cloud delivery using deep graph neural networks," in *IEEE International Conference on Communications*, 2021, pp. 1–6.
- [24] C. T. Duong, T. D. Hoang, H. H. Dang, Q. V. H. Nguyen, and K. Aberer, "On node features for graph neural networks," *arXiv e-prints*, pp. 1–6, Nov. 2019.
- [25] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S. Y. Philip, "A comprehensive survey on graph neural networks," *IEEE transactions on neural networks and learning systems*, vol. 32, no. 1, pp. 4–24, 2020.
- [26] T. Koike-Akino and Y. Wang, "Stochastic bottleneck: Rateless auto-encoder for flexible dimensionality reduction," in *IEEE International Symposium on Information Theory (ISIT)*, 2020, pp. 2735–2740.
- [27] R. Sato, M. Yamada, and H. Kashima, "Random features strengthen graph neural networks," in *Proceedings of the SIAM International Conference on Data Mining*, 2021, pp. 333–341.
- [28] S. Ibuki, T. Okamoto, T. Fujihashi, T. Koike-Akino, and T. Watanabe, "Rateless deep graph joint source channel coding for holographic-type communication," in *IEEE Global Communications Conference*, 2023, pp. 3330–3335.
- [29] H. Wu, Y. Shao, C. Bian, K. Mikolajczyk, and D. Gündüz, "Deep joint source-channel coding for adaptive image transmission over mimo channels," *IEEE Transactions on Wireless Communications*, 2024.
- [30] K. Yang, S. Wang, J. Dai, X. Qin, K. Niu, and P. Zhang, "SwinJSCC: Taming swin transformer for deep joint source-channel coding," *IEEE Transactions on Cognitive Communications and Networking*, 2024.
- [31] T.-Y. Tung, D. B. Kurka, M. Jankowski, and D. Gündüz, "DeepJSCC-Q: Channel input constrained deep joint source-channel coding," in *IEEE International Conference on Communications*, 2022, pp. 3880–3885.
- [32] J. Xu, B. Ai, W. Chen, A. Yang, P. Sun, and M. Rodrigues, "Wireless image transmission using deep source channel coding with attention modules," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 4, pp. 2315–2328, 2021.
- [33] S. Wang, K. Yang, J. Dai, and K. Niu, "Distributed image transmission using deep joint source-channel coding," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2022, pp. 5208–5212.
- [34] E. Erdemir, T.-Y. Tung, P. L. Dragotti, and D. Gündüz, "Generative joint source-channel coding for semantic image transmission," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 8, pp. 2645–2657, 2023.
- [35] D. B. Kurka and D. Gündüz, "DeepJSCC-f: Deep joint source-channel coding of images with feedback," *IEEE Journal on Selected Areas in Information Theory*, vol. 1, no. 1, pp. 178–193, 2020.
- [36] —, "Bandwidth-agile image transmission with deep joint source-channel coding," *IEEE Transactions on Wireless Communications*, vol. 20, no. 12, pp. 8081–8095, 2021.

- [37] M. Yang, C. Bian, and H.-S. Kim, "Deep joint source channel coding for wireless image transmission with ofdm," in *IEEE International Conference on Communications*, 2021, pp. 1–6.
- [38] —, "OFDM-guided deep joint source channel coding for wireless multipath fading channels," *IEEE Transactions on Cognitive Communications and Networking*, vol. 8, no. 2, pp. 584–599, 2022.
- [39] S. Inokuma, Y. Sasaki, D. Hisano, Y. Nakayama, and K. Maruta, "Performance evaluation of mimo transmission in deep joint source-channel coding," in *IEEE Vehicular Technology Conference*, 2024, pp. 1–5.
- [40] J. Xu, T.-Y. Tung, B. Ai, W. Chen, Y. Sun, and D. Gündüz, "Deep joint source-channel coding for semantic communications," *IEEE Communications Magazine*, vol. 61, no. 11, pp. 42–48, 2023.
- [41] J. Wang, S. Wang, J. Dai, Z. Si, D. Zhou, and K. Niu, "Perceptual learned source-channel coding for high-fidelity image semantic transmission," in *IEEE Global Communications Conference*, 2022, pp. 3959–3964.
- [42] X. Liu, H. Liang, Z. Bao, C. Dong, and X. Xu, "A semantic communication system for point cloud," *IEEE Transactions on Vehicular Technology*, pp. 1–17, 2024.
- [43] Y. Zhu, Y. Huang, X. Qiao, Z. Tan, B. Bai, H. Ma, and S. Dustdar, "A semantic-aware transmission with adaptive control scheme for volumetric video service," *IEEE Transactions on Multimedia*, vol. 25, pp. 7160–7172, 2023.
- [44] Y. Huang, B. Bai, Y. Zhu, X. Qiao, X. Su, L. Yang, and P. Zhang, "ISCom: Interest-aware semantic communication scheme for point cloud video streaming on metaverse xr devices," *IEEE Journal on Selected Areas in Communications*, vol. 42, no. 4, pp. 1003–1021, 2024.
- [45] T. Fujihashi, T. Koike-Akino, and T. Watanabe, "Soft delivery: Survey on a new paradigm for wireless and mobile multimedia streaming," *ACM Computing Surveys*, vol. 56, no. 2, 2023.
- [46] J. Shen, L. Yu, L. Li, and H. Li, "Foveation-based wireless soft image delivery," *IEEE Transactions on Multimedia*, vol. 20, no. 10, pp. 2788–2800, 2018.
- [47] X.-W. Tang, X.-L. Huang, F. Hu, and Q. Shi, "Human-perception-oriented pseudo analog video transmissions with deep learning," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 9, pp. 9896–9909, 2020.
- [48] J. Žádník, M. Kieffer, A. Trioux, M. Mäkitalo, and P. Jääskeläinen, "CV-Cast: Computer vision-oriented linear coding and transmission," *IEEE Transactions on Mobile Computing*, pp. 1–14, 2024.
- [49] T. Fujihashi, T. Koike-Akino, T. Watanabe, and P. V. Orlik, "FreeCast: Graceful free-viewpoint video delivery," *IEEE Transactions on Multimedia*, vol. PP, no. 99, pp. 1–11, 2019.
- [50] T. Fujihashi, T. Koike-Akino, T. Watanabe, and P. Orlik, "HoloCast: Graph signal processing for graceful point cloud delivery," in *IEEE International Conference on Communications*, 2019, pp. 1–7.
- [51] 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 Transactions on Multimedia*, vol. 24, pp. 2179–2191, 2021.
- [52] S. Ueno, T. Fujihashi, T. Koike-Akino, and T. Watanabe, "Point cloud soft multicast for untethered XR users," *IEEE Transactions on Multimedia*, vol. 25, pp. 7185–7195, 2023.
- [53] C. Morris, M. Ritzert, M. Fey, W. L. Hamilton, J. E. Lenssen, G. Rattan, and M. Grohe, "Weisfeiler and leman go neural: Higher-order graph neural networks," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 4602–4609.
- [54] L. Yi, V. G. Kim, D. Ceylan, I.-C. Shen, M. Yan, H. Su, C. Lu, Q. Huang, A. Sheffer, and L. Guibas, "A scalable active framework for region annotation in 3d shape collections," *ACM Transactions on Graphics (ToG)*, vol. 35, no. 6, pp. 1–12, 2016.
- [55] M. Fey and J. E. Lenssen, "Fast graph representation learning with PyTorch geometric," *arXiv e-prints*, Mar. 2019.
- [56] G. Bjøntegaard, "Calculation of average PSNR differences between RD-curves," *ITU-T SG16/Q6 Input Document VCEG-M33*, 2001.
- [57] X. Liu, H. Liang, Z. Bao, C. Dong, and X. Xu, "A semantic communication system for point cloud," *IEEE Transactions on Vehicular Technology*, vol. 74, no. 1, pp. 894–910, 2025.
- [58] T. Nguyen, Q.-H. Pham, T. Le, T. Pham, N. Ho, and B.-S. Hua, "Point-set distances for learning representations of 3D point clouds," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 10 478–10 487.
- [59] P. L. Zeyu Yan, Fei Wen, "Optimally controllable perceptual lossy compression," in *International Conference on Machine Learning*, 2022.



SHOICHI IBUKI received the B.S. degree from Osaka University, Osaka, Japan, in 2023. He is now a M.S. student at the Graduate School of Information Science and Technology, Osaka University.



TSUBASA OKAMOTO received the B.S. and M.S. degrees from Osaka University, Osaka, Japan, in 2022 and 2024.



TAKUYA FUJIHASHI received his B.E. degree in 2012 and his M.S. degree in 2013 from Shizuoka University, Japan. In 2016, he received his Ph.D. degree from the Graduate School of Information Science and Technology, Osaka University, Japan. He is currently an Assistant Professor at the Graduate School of Information Science and Technology, Osaka University 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 at the Mitsubishi Electric Research Labs. (MERL) working with the Electronics and Communications group. He was selected as one of the Best Paper candidates in IEEE ICME (International Conference on Multimedia and Expo) 2012. His research interests are in the area of video compression and communications, with a focus on multi-view video coding and streaming.

PLACE
PHOTO
HERE

TOSHIAKI KOIKE-AKINO (M'05–SM'11) received the B.S. degree in electrical and electronics engineering, M.S. and Ph.D. degrees in communications and computer engineering from Kyoto University, Kyoto, Japan, in 2002, 2003, and 2005, respectively. During 2006–2010 he was a Postdoctoral Researcher at Harvard University, and is currently a Distinguished Research Scientist at Mitsubishi Electric Research Laboratories (MERL), 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.

he was a research fellow (DC1) of Japan Society for the Promotion of Science. From 2014 to 2015, he was an intern at the Mitsubishi Electric Research Labs. (MERL) working with the Electronics and Communications group. He was selected as one of the Best Paper candidates in IEEE ICME (International Conference on Multimedia and Expo) 2012. His research interests are in the area of video compression and communications, with a focus on multi-view video coding and streaming.



TAKASHI WATANABE (S'83–M'87) received his B.E., M.E., and Ph.D. degrees from Osaka University, Japan, in 1982, 1984, and 1987, respectively. He joined the Faculty of Engineering, Tokushima University, in 1987, and moved to the Faculty of Engineering, Shizuoka University, in 1990. He was a Visiting Researcher with the University of California, Irvine, from 1995 to 1996. He has been a Professor with the Graduate School of Information Science and Technology, Osaka

University, Japan, since 2013. His research interests include mobile networking, ad hoc sensor networks, Internet of Things/M2M networks, and intelligent transport systems, especially medium access control and routing. He is a member of IPSJ and IEICE. He has served on program committees of many networking conferences, such as the IEEE, ACM, IPSJ, and IEICE.

...