Towards Communication-efficient Digital Twin via AI-powered Transmission and Reconstruction

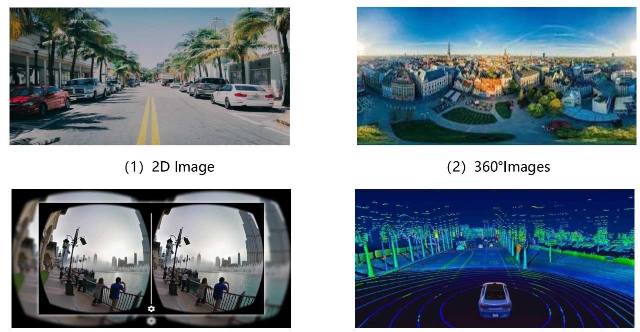
Cen Chen*, Xulei Yang, *Senior Member, IEEE*,
Yangfan Li*.

***Abstract*—Digital twin technology has recently gathered pace in engineering communities as it allows for the convergence of the real structure and its digital counterpart. 3D point cloud data is a more effective way to describe the real world and to reconstruct the digital counterpart than the conventional 2D images or 360-degree images. Large-scale, e.g., city-scale digital twins, typically collect point cloud data via internet-of-things (IoT) devices and transmit it over wireless networks. However, the existing wireless transmission technology can not carry real-time point cloud transmission for digital twin reconstruction due to mass data volume, high processing overheads, and low delay-tolerance. We propose a novel artificial intelligence (AI) powered end-to-end framework, termed AIRec, for efficient digital twin communication from point cloud compression, wireless channel coding, and digital twin reconstruction. AIRec adopts the encoder-decoder architecture. In the encoder, a novel importance-aware pooling scheme is designed to adaptively select important points with learnable thresholds to reduce the transmission volume. We also design a novel noise-aware joint source and channel coding is proposed to adaptively adjust the transmission strategy based on SNR and map the features to error-resilient channel symbols for wireless transmission to achieve a good tradeoff between the transmission rate and reconstruction quality. The decoder can accurately reconstruct the digital twins from the received symbols. Extensive experiments of typical datasets and comparison with baselines show that we achieve a good reconstruction quality under $24\times$ compression ratio.**

***Index Terms*—Communication-efficient, Digital twin, Deep neural networks, Point cloud.**

I. Introduction

DIGITAL twin technology has recently gathered pace in the engineering communities as it allows for the convergence of the real structure and its digital counterpart [1]. Recent advancements in the Internet of Things (IoT) [2] and remote sensing technologies provide large-scale, accurate and affordable measurement data, which greatly contribute to the burst of the scale of the digital twins, i.e, from object-scale and the infrastructure-scale to city-scale digital twins [3]. For example, the Singapore government launched the Virtual Singapore project, a digital representation of the country in 2014. The rich data environment provided by the large-scale digital twins facilitates a range of applications, such as Geographic Information Systems (GIS), smart transportation, resources and energy, disaster management and so on [4], [5]. However, because of the complicated shapes and dynamic status changes of objects in the physical world, together with the huge data amount and high update frequency, it is impossible to create and digitize objects solely by human hands. Thus, intelligent and automated techniques for real-time digital reconstruction are highly desired by both industry and academia.



(1) 2D Image (2) 360°Images (3) VR (4) Point Cloud

	2D Image	360° Images	VR	Point Cloud
Date Volume	~1Mbps	~100Mbps	~100Mbps	~1Gbps
Delay Tolerance	Medium	High	High	Very High
User Freedom	NA	2DoF	3DoF	6DoF

Fig. 1: Comparisons between point cloud streamings and other data streamings. The point cloud streaming requires an extreme bandwidth consumption and low delay [6].

3D point cloud data is an efficient representation of 3D objects and environmental data in Digital Twins. Techniques to reconstruct surface or volume representations from the the raw sparse, noisy, and non-uniform point clouds have been instrumental in creating digital twins [3], [7]–[9], like 3D city models, indoor and building models. Further, progress for intelligent 3D point cloud analysis tasks such as 3D shape classification, detection and tracking, segmentation [10] also facilitate a wide range of intelligent applications in city-scale digital twins, such as urban planning [11], 3D navigation [12] and so on.

Dynamically reconstructing a large-scale digital twin such as city-level digital twin for 3D navigation requires the transmission of intensive point cloud streams across the prevailing

Cen Chen is with the School of Future Technology, South China University of Technology, Guangzhou, 510641, China. Cen Chen is also with the Shenzhen Institute, Hunan University, China (Email: chencen@scut.edu.cn).
Xulei Yang is with the Institute for Infocomm Research (I2R), A*STAR, Singapore. (E-mail: yang_xulei@i2r.a-star.edu.sg).

Yangfan Li is with the School of Computer Science and Engineering, Central South University, Changsha, 410083, China. (E-mail: liyangfan37@csu.edu.cn).

*Corresponding authors.

network environment, including 5G. This task presents two significant challenges: (i) Shown in Fig. 1, the massive amount of raw point cloud frames collected by the IoT devices require the prohibitively high bandwidth and may be at a Gbps level even after conventional compression, surpassing the capability of the current 5G network [6]; (ii) Intelligent digital twin applications are sensitive to delays, and wireless communication is lossy in nature due to shadowing, channel fading and inter-symbol interference. Moreover, high decoding overhead for volumetric point cloud data also extremely slows down the transmission. The traditional network coding approaches encode the compressed features as bits for reliable transmission. These approaches suffer from the *cliff effect* and only achieve desired performance with accurate target channel quality estimation, making them unsuitable for low-latency digital twin applications [13].

To mitigate these communication bottlenecks, researchers have suggested traditional lossy and lossless point cloud compression schemes [14]–[16]. Unfortunately, lossy compression may lead to critical information loss under poor network conditions, while lossless compression techniques cannot achieve a satisfactory compression ratio. The complexity of codec software's encoding and decoding overhead hinders efficient transmission. Furthermore, these methods often overlook transmission errors induced by wireless networks, a key factor in wireless condition transmission delay.

Recently, the evolution of AI techniques in various fields has introduced new pathways for efficient wireless transmission. [13], [17] suggested that the deep networks can not only efficiently compress the data by encoding the key semantic information as dense features, but can also directly map the features to channel symbols, and outperform the state-of-the-art digital schemes for wireless image transmission. Nevertheless, these methods are limited for 2D domain, they fall short in achieving satisfactory reconstruction quality under high compression ratios for 3D digital twin communication. The 3D point cloud data is sparse, unordered and irregular in the European space and contains complex geometric information, and necessitates advanced compression techniques with lower latency transmission.

To address these issues, in this paper, we propose an AI-powered framework for efficient point cloud transmission and reconstruction in digital twins, termed as AIRec. This system trains an end-to-end encoder-decoder based deep neural network, facilitating digital twin communication from point cloud compression, wireless channel coding, to digital twin reconstruction. The encoder compresses the collected raw point cloud efficiently and maps the features to error-resilient channel symbols for wireless transmission, while the decoder accurately reconstructs the digital twins from the received symbols. AIRec greatly reduces the transmitted point cloud data and is robust against the noises in the wireless channels. Nevertheless, two technical challenges underpin the design and implementation.

The first challenge is enhancing the compression ratio of 3D point cloud while assuring reconstruction accuracy. We observe that not all points in the raw point cloud data are equal in importance in the complex physical world. For instance, flat areas like grounds or walls typically demand fewer points than complex surfaces like persons or vehicles for intelligent digital twins. With this motivation, we propose a novel importance-aware pooling strategy (IPAPool) in the encoder, implementing a learnable threshold method to adaptively remove unimportant points, thereby reducing transmission volume. Specifically, points below the importance threshold are progressively removed by a hierarchical encoder on the sender side. The importance threshold is automatically optimized by training, effectively eliminating random and manual factors introduced by hyperparameters. The information of the removed points is compressed into the dense features of the retained points through the hierarchical encoder, ensuring the accuracy of the data. These retained points, with compressed features, are then efficiently transmitted for digital counterpart reconstruction.

The second challenge is to ensure robust transmission and maintain reconstruction quality under the dynamic and bandwidth-limited wireless channel conditions. To tackle this, we propose a novel noise-aware joint channel-source coding scheme. Inspired by the recent advancements in joint source and channel coding (JSCC) [13], [18], we design a novel dynamic JSCC technique for point cloud transmission. We map the feature vectors directly into channel input symbols. The noisy channel output is used by the digital counterpart to reconstruct the physical world, without any explicit channel code involvement. Furthermore, the proposed scheme enables the JSCC to dynamically adjust the point transmission volume based on network conditions, thereby exploiting the tradeoff between transmission rate and signal quality.

In summary, our key contributions can be outlined as follows:

- We introduce AIRec, an end-to-end AI framework for efficient digital twin communication, which significantly reduces transmitted point cloud data and exhibits robustness against wireless channel noise.
- We propose an efficient IPAPool, which adaptively retains critical points in point cloud data via learnable importance, considerably reducing transmission by transferring only the remaining important point features.
- We devise a novel noise-aware joint channel-source coding scheme for point cloud transmission, capable of adaptively adjusting the transmission strategy to strike a balance between transmission rate and signal quality based on network conditions.
- Our evaluations on typical datasets, compared with existing baselines, demonstrate good reconstruction quality under a $24\times$ compression ratio.

We organize the remaining of our work as follows. Section II discusses the related work. Section III describes in detail our proposed framework AIRec. Section IV gives the experimental results and analysis, The work is summarized in Section V.

II. Related Work

A. Intelligent Digital Twin

Digital twin is a method to model and monitor intricacies, whose results of analysis and simulation can detect dangerous emergent behaviour and be helpful for mitigation unforeseen and adverse consequences to real-world systems and objects [19], [20]. In recent years, there has been a growing appeal for real-time semantically rich urban models such as digital twin cities (DTC), while the development of the Internet of Things (IoT) and remote sensing technologies has provided a large amount of precise and acceptable data from measurement. Data of point clouds can provide premium geometric measurements and restricted non-geometric semantic information, such as energy consumption in a DTC. Therefore, the output DTCs connect many fields, including building information modeling (BIM), geographic information systems (GIS), and autonomous vehicles [4], [21].

B. Point Cloud Compression and AI-based Point Cloud

The most original method of point cloud compression is to directly compress the point cloud to reduce the amount of data to transmit, including attribute compression, geometric compression, and motion-compensation compression [22], [23]. Plenty of these pay close attention to static Kdtree-based and octre-based solutions. [24] converts the original point cloud into voxels based on octree and supports an adaptive compression ratio. Draco realizes lossy compression by reducing quantization bits. [22] uses a B-spline wavelet to represent the geometry and attributes of the compressed point cloud. [25] points out that converting 3D point clouds into 2D maps and using conventional algorithms for compression will lead to the loss of key feature information.

Geometric compression [26] based on deep learning divides the original point cloud by 3D voxels, it will lead to a large amount of loss of the original data information due to the sparse Euclidean structure. Most of the current compression methods are based on spatial voxel characteristics, ignoring the transmission characteristics of the point cloud. [27] compresses the point cloud by extracting key features rather than the original point cloud, and can provide real-time delivery services within the existing network environment.

In the realm of AI-based point cloud feature processing, PointNet [28] stands as a trailblazer, introducing a direct approach for feature extraction from point clouds. This pioneering work has since inspired a range of point-based variants including PointNet++ [29], DGCNN [30], LDGCNN [31], F-PointNet [32], and DensePoint [33]. Notably, DGCNN models point clouds using a graph structure and leverages graph convolution to incorporate local neighborhood data, thereby enhancing feature representation.

C. Noise-aware Dynamic Joint Source and Channel Coding

Traditional wireless transmission is based on Shannon's separation theorem [34], and plenty of systems adopt a separate source coding and channel coding. This method has great limitations, and its effect is not as good as that of joint source channel coding (JSCC), such as plenty of schemes in index allocation [35] ,vector quantization and others. Due to its strong capacity to extract complex features, the JSCC system used for wireless image transmission has also begun to apply deep learning to solve problems. Plenty of deep JSCC methods only take into account fixed-rate training, taking into account the changes of the network condition, it is necessary to train plural models to achieve a multi-rate JSCC scheme, the concept of which was studied in [36], [37], but only a single SNR was considered for training in their models. In [38], it can realize a dynamical regulation for the rate to save channel bandwidth while transmitting enough source information content. In [39], [40], the image transmission scheme through additive white noise (AWGN) wireless channel is studied. A full convolution auto-encoder architecture is proposed in [39], which maps the input image directly to the channel symbol without any digital interface. It not only optimizes the channel coding and compression scheme of state-of-the-art, but also realizes the graceful degradation scheme of the channel signal-to-noise ratio (SNR). According to the results of [39] and [36], the auto-encoder architecture for image JSCC with channel output feedback on AWGN channel considers both noise and noise-free feedback, which not only learns how to map images to channel output, but also learns how to utilize channel output feedback. [17] proposed a JSCC scheme using channel output feedback, which presents simulated behavior and elegant degradation in the case of channel quality changes, and can suit forward or feedback channel changes.

III. DESIGN OF AIREC

A. System Overview

1) Overview of the AIRec: Our AIRec framework, depicted in Fig. 2, is designed to process data from the initial collection to the final reconstruction of digital twins.

Data Collection and Encoding at Intelligent 3D Sensing Devices. Intelligent devices equipped with 3D surface scanners, like Lidar sensors, capture raw 3D point cloud data from the physical world. These devices also have a small computational platform to preprocess the raw point clouds. The lightweight AIRec encoder, implemented in these devices, adaptively selects essential points with a hierarchical importance-aware feature extractor that uses a learnable threshold method. This reduces the transmission volume by removing non-essential points. The encoder then extracts the compressed features of the essential points and encodes these features as channel symbols for transmission.

Wireless Transmission through the Base Station. The intelligent 3D sensing devices are connected to a digital twin center server via a wireless network. Our system employs a novel noise-aware point cloud joint source and channel coding scheme that directly maps selected points and their compressed features to error-resilient channel symbols. The transmission strategy is adaptively adjusted based on the network condition, ensuring the resilience of the symbols during wireless transmission. These symbols are then transmitted in real-time to the base station.

Reconstruction and Task Execution at the Digital Twin Center Server. The AIRec Decoder, deployed at the digital twin center server, receives the compressed point features from the intelligent devices. Built on PUGAN [9], the decoder accurately reconstructs these features into a 3D scene in the

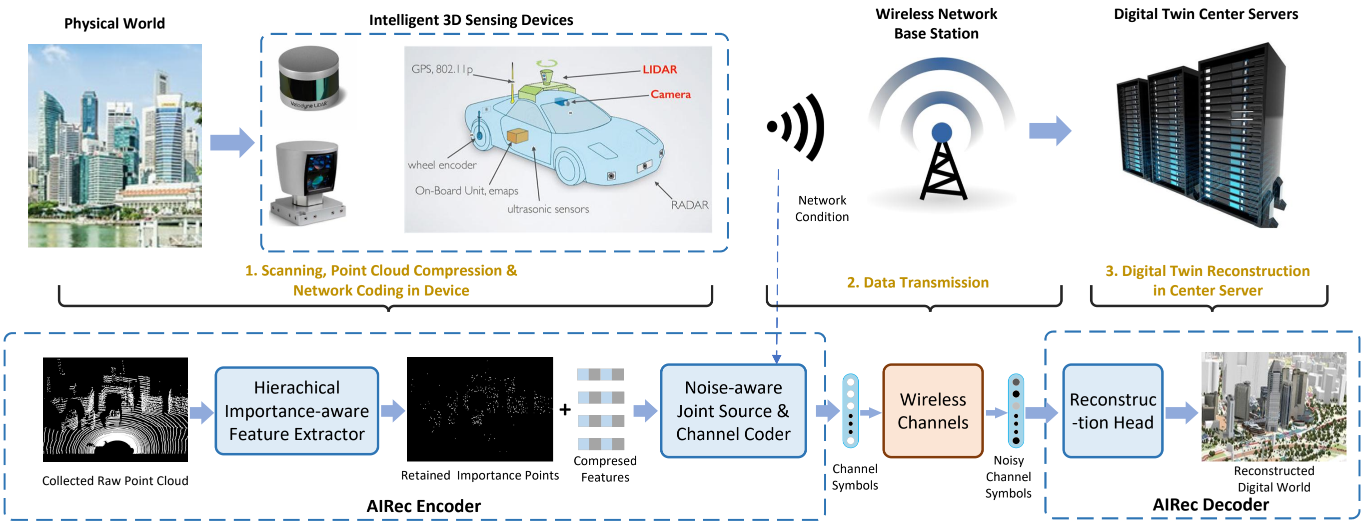


Fig. 2: System Overview of AIRec. The diagram shows the process from data collection by Intelligent 3D Sensing Devices, which utilize the encoder, to wireless transmission via the Base Station, and finally digital reconstruction at the Digital Twin Center Server using the decoder.

digital world. Additionally, the AIRec framework is versatile and supports multiple intelligent tasks in the digital twin. By replacing the reconstruction head with a 3D object detection head, for instance, the framework can efficiently transmit crucial points for 3D object detection.

2) Framework Deployment: The deployment of AIRec consists of two phases:

Offline Training Phase. This phase has two objectives. First, the AIRec encoder is trained to select important points, compress raw point cloud data, and convert the features into error-resilient wireless channel symbols. Network condition plays a critical role here, with the encoder learning to use more or fewer channel resources based on the SNR. Second, the AIRec decoder is trained for various intelligent tasks, such as reconstruction and 3D object detection.

Online Phase. Here, the trained lightweight AIRec encoder is deployed in intelligent 3D sensing devices, and the AIRec decoder is deployed in the digital twin center server. Raw point clouds are collected, important points are dynamically retained and transmitted in real-time, and the 3D scene is reconstructed upon receiving the compressed features at the center server. This ensures a balance between the transmission rate and task accuracy.

B. Hierarchical Encoder for Point Feature Extraction

Shown in Fig. 3, we designed a hierarchical encoder for point feature extraction based on PointNet++ [29], where the point cloud input can be classified and segmented directly. Due to the invariance of the arrangement of disordered point sets, we use multiple set abstraction layers for hierarchical feature learning to capture the local features of the original point cloud. Therefore, the output feature matrix from the point feature extraction encoder can represent the original point cloud abstractly with fewer points. Each set abstraction layer contains a sampling layer and a grouping layer. The sampling layer is designed to down-sample the points, which represents the center of the local area. The grouping layer finds a quantitative nearest neighbors around the center to construct the local region set. A pointnet network is further adopted to extract the features of the points from the above operations. Note that we adopt 4 layers of set abstraction in this hierarchical encoder. We apply D-FPS down sampling in the first two layers, which capture basic geometric details. In contrast, the last two layers, which hold richer and task-specific semantics, utilize our proposed importance-aware point pooling strategy (IPAPool).

C. Importance-aware point pooling strategy (IPAPool)

Down-sampling strategies are crucial for memory and computation efficient 3D point cloud processing. Researchers have proposed strategies such as random sampling [41] and farthest point sampling (FPS) sampling [29], [42] to reduce the amount of points to process while remains the essential information. However, these approaches are task-agnostic and neglect the fact that not all points are of equal importance for a specific task. For example, in the complex physical world, the flat areas such as grounds or walls usually require far fewer points than the complex surfaces such as people or vehicles for digital world reconstruction. Motivated by this, we propose a novel IPAPool scheme which explores a learnable threshold method to remove unimportant points. We also introduce a regularization strategy in our scheme to overcome the tendency of neural networks to retain more points for improved performance.

Shown in Fig. 4. IPAPool defines the importance score for every point in the dataset and then adaptively retains the significant points based on a new sampling method based on learnable thresholds. In contrast to other pooling methods, such as topK pooling, the amount of remained points are not fixed for different point cloud frame, making it possible to select more points in the complex frame while keeping less points in the simple scenarios, which further reduce the

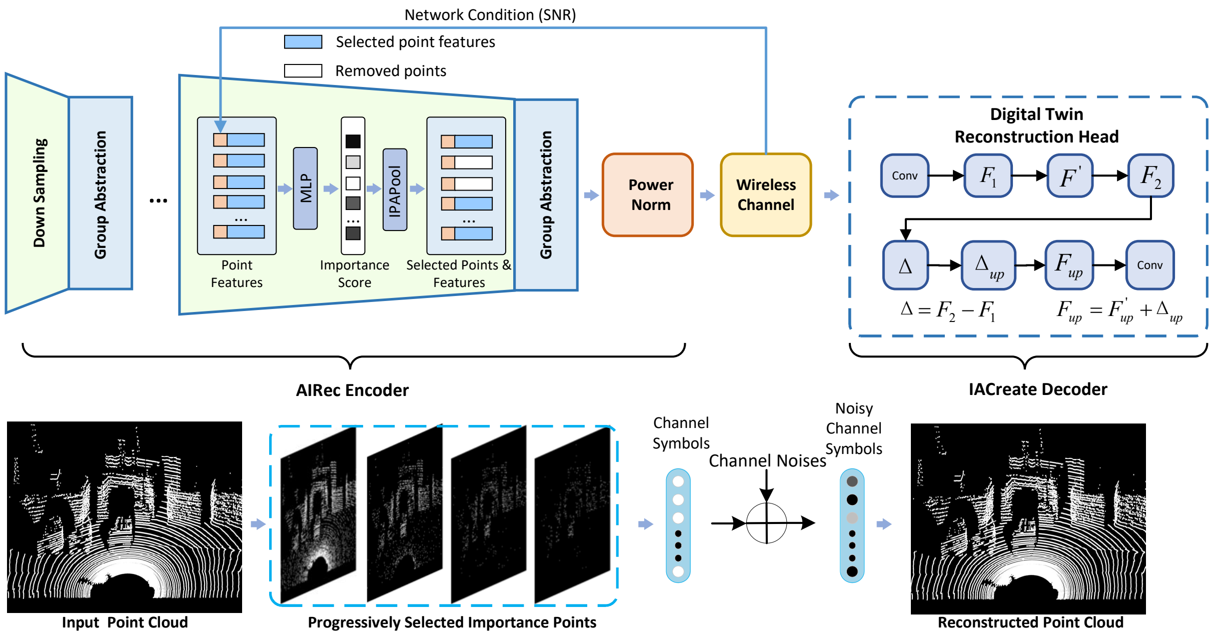


Fig. 3: Network Architecture of AIRec's Encoder and Decoder. This figure provides a detailed view of how the Encoder and Decoder work together in the AIRec framework to facilitate point cloud compression, wireless transmission, and digital twin reconstruction.

communication costs. Moreover, the thresholds are obtained through learning, which avoids tedious manual learning, and a nice tradeoff between the volume of transmission and reconstruction accuracy can be achieved.

We expect that significant points can more efficiently describe the overall shape of the expressed point cloud, such as points at boundary junctions or critical areas. By identifying significant points not only can we improve the model processing speed, but also reduce the noise introduced by non-critical points, thus improving the model accuracy, we employ a MLP to calculate the importance score based on the point features. Given the point i and its feature vector $F_i^{(l)}$, the importance score can be calculated as:

$$S_i^{(l)} = t_i \cdot F_i^{(l)} / \|t_i\|, \tag{1}$$

where $S_i^{(l)}$ is the importance of point i, and important points will have larger scores. t_i is the learnable weight parameter. With the importance scores for all the points, the points can be down-sampled by remaining only the points with high scores. A conventional approach is manually defining an threshold $\phi^{(l)}$ and remains important points whose $S_i^{(l)}$ is larger than $\phi^{(l)}$. Specifically, a binary mask matrix $M^{(l)}$ is calculated for all the points. If the importance score for the point $S_i^{(l)}$ is below the threshold, $M_i^{(l)} = 0$. Otherwise, $M_i^{(l)} = 1$. Formally, $M_i^{(l)}$ is defined as follows:

$$M_i^{(l)} = \begin{cases} 0 & S_i^{(l)} < \phi^{(l)}; \\ 1 & S_i^{(l)} \geq \phi^{(l)}. \end{cases} \tag{2}$$

Nevertheless, the hard pruning mask in non-differentiable. To address this prolem, we propose an adaptive learnable threshold scheme which facilitates the back-propagation to adaptively learn the threshold. We utilize the $\mathrm{sigmoid}(\cdot)$ function which transforms the hard non-differentiable mask $\tilde{M}^{(l)}$ to the differentiable soft mask:

$$\tilde{M}^{(l)} = \mathrm{sigmoid}\left(\frac{S_i^{(l)} - \tilde{\phi}^{(l)}}{R}\right), \tag{3}$$

where $\tilde{\phi}^{(l)}$ is the learnable importance threshold, R is the pooling influence factor. With the differentiable Eq. 3, $\tilde{\phi}^{(l)}$ can be optimized through the back-propagation of the whole network. Note that the importance threshold $\tilde{\phi}^{(l)}$ is initially based on the expected compression ratio of the network and $\tilde{\phi}^{(l)}$ is optimized base on the target loss function. The pooling influence factor R is to amplify the relationship between the importance score and threshold. With the amplifying, the value of points with importance scores $S_i^{(l)}$ larger than $\tilde{\phi}^{(l)}$ is approximately 1 in $\tilde{M}^{(l)}$, and 0 otherwise. After $\tilde{M}^{(l)}$ is obtained, we multiply the mask matrix $\tilde{M}^{(l)}$ by the feature matrix to remain importance points, i.e,

$$F^{(l+1)} = F^{(l)} \times \tilde{M}^{(l)}, \tag{4}$$

where $F^{(l)} \in \mathbb{R}^{n \times d}$ represents the input feature matrix of the layer l, $F^{(l+1)} \in \mathbb{R}^{n' \times d}$ $(n' < n)$ is the output feature, which acts as the input to the layer $l+1$.

D. Noise-aware Dynamic Joint Source and Channel Coding

Traditional network encoding methods convert extracted features into bits for digital wireless transmission. This approach, however, encounters the *cliff effect* phenomenon. This refers to the abrupt degradation in signal quality or even total signal loss when signal strength falls below a specific minimum level, which can lead to the loss of critical information during transmission. Furthermore, this approach relies heavily on precise channel quality estimation, making it unsuitable

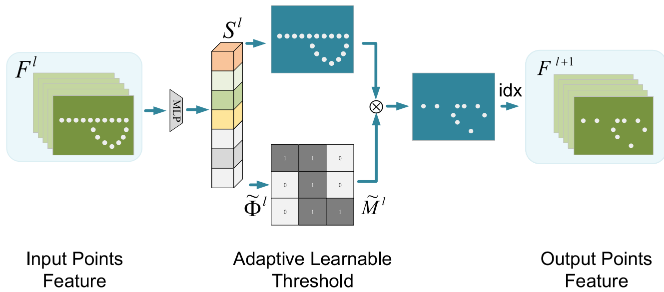


Fig. 4: Importance-aware point pooling strategy (IPAPool).

for low-latency digital twins applications [13]. Inspired by the recent advancements in joint source and channel coding [13], [18], we overcome this limitation by directly mapping features into channel input symbols. In real-world wireless network scenarios, network conditions are dynamic, and insufficiently transmitted points can lead to significant degradation in reconstruction performance under poor network conditions. To mitigate this, we propose a noise-aware pooling scheme that dynamically adjusts the number of transmitted points based on channel noise levels. In high noise scenarios, this scheme increases the transmission points to balance the adverse effects of noise. This approach ensures that the reconstruction accuracy remains robust without significantly compromising transmission speed, striking an effective balance between reconstruction quality and speed.

We assume the additive white gaussian noise (AWGN) channel for transmission following [13], [18]. The input of AWGN is composed of B complex channel input symbols x_i, and the channel input vector is represented by $\mathbf{x} \in \mathbb{C}^{\mathbf{B}}$. The output is given by **y=x+z**, and the channel output vector is expressed by $\mathbf{y} \in \mathbb{C}^{\mathbf{B}}$, where $z_i \sim \mathbb{C}N(0,\delta^2)$ is the independent identically distribution of the noise vector, and the noise vector is expressed by $\mathbf{z} \in \mathbb{C}^{\mathbf{B}}$, i= 1,2,...,B. The average power constraint is applied to the input vectors $\mathbf{x}$, such that $\frac{1}{B}\sum_{i=1}^{B}|x_i| \leq P = 1$, which is converted to the maximum accepted SNR in the static AWGN channel, which is $SNR = 10\log_{10}\left(\frac{1}{\sigma^2}\right)$ expressed in the dB scale

Our proposed noise-aware dynamic joint source and channel coding, depicted in Fig. 3, is designed to adaptively adjust to different transmission source requirements based on the network Signal-to-Noise Ratio (SNR) conditions. We introduce a novel noise-aware strategy within the IPAPool module to increase channel awareness during point retention, thus optimizing the joint encoding process. In Fig.5, the SNR condition is first transformed into a dense vector representation using an embedding layer. This dense vector is then replicated to match the total number of points and is concatenated with the point features for subsequent computation of each point's importance score. As the SNR condition varies, the system dynamically adjusts its behavior to transmit more or less information as needed.

Through the real-time input of the network condition (SNR), our noise-aware dynamic joint source and channel coding scheme decides to adaptively retain and transmit important points, providing a more efficient transmission strategy tailored to the current conditions. The retained points and their associated feature matrix are then processed through a power normalization module. This module converts the feature matrix into complex-valued transmission symbols with unit average power, using the first half of the features as the real part and the second half as the imaginary part. Following transmission across the noisy wireless channel, the output channel symbols are retrieved by the point Decoder. These symbols are then mapped back into the high-dimensional feature space, and the resulting feature vector is transmitted through the wireless network for digital twin reconstruction.

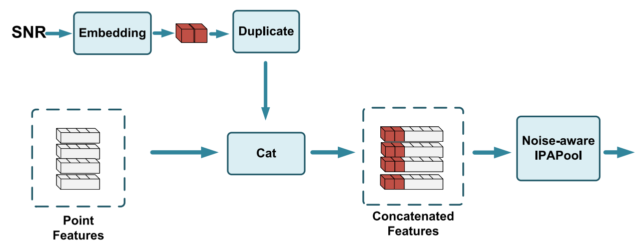


Fig. 5: Noise-aware dynamic joint source and channel coding.

E. Point Decoder for Intelligent Digital Twin Reconstruction

We design a point decoder for feature reconstruction based on PU-GAN [9]. The decoder is designed to accurately reconstruct the received features into a point cloud with more points so that the point cloud generated from a few important point features are closer to the original point cloud, which is convenient for Intelligent Digital Twin Reconstruction. Shown in Fig. 3, the generator of this network is composed of the feature extension part and the point set generation module.

The feature expansion module is implemented by an up-down-up unit. The feature $F_{up}{}^{'}$ of $rN * C$ is obtained through up-sampling from the input feature F_1 of $N * C$ through MLPs, and then the feature F_2 of $N * C$ is obtained through down-sampling feature operation. The offset vector Δ is calculated by the difference between F_2 and F_1, and the same up-sampling feature operation is performed on Δ to get Δ_{up}. Finally, the output F_{up} is obtained by adding the Δ_{up} to $F_{up}^{'}$. During the up-sampling, after r times of copying the input feature graph, the operator uses 2D network mechanism to generate unique 2D vector for each copy of the feature graph, which is equivalent to their ID card to push the new point away from the input point. At this time, the features of $rN * (C^{'} + 2)$ dimension are obtained.

After passing the self-attention module, the up-sampled features are output through a set of $MLPs$. In the self-attention module, the input features are respectively passed through three groups of $MLPs$ to obtain three groups of features, G, H and K. The output features are obtained by calculation of three groups of features, which can introduce remote context dependence after connection to enhance feature integration. The down-sampling operator is reshaped by simple reshaping features and then returned to the original features by a set of $MLPs$.

In the point set generation module, the point set generation module first regression the features in F_{up} into 3D coordinates through a set of $MLPs$, and only retains rN points with more distance as output points when sampling at the farthest point, which can further enhance the uniformity of point set distribution. The discriminator uses the basic network architecture of spot complete network, which can effectively combine local and global information and ensure lightweight network. After that, we design a self-attention module to improve feature learning, which is helpful to enhance feature integration and improve feature extraction ability. A set of $MLPs$ and maxpool were applied to generate global features.

F. Training Strategy

The training of our proposed AIRec is divided into 3 stages. (1) Train the reconstruction decoder with the patch-based scheme and the adversarial loss and the uniform loss proposed by PU-GAN [9]. (2) Load the pre-trained model of the decoder, fix the decoder parameters and train the encoder with reconstruction loss. Note that we assume the ideal channel and train the encoder without any threshold pruning scheme in this stage. The Earch Mover's distance [43] is adopted the reconstruction loss as follows:

$$\mathcal{L}_{rec} = \min_{\phi:Q\rightarrow\hat{Q}} \sum_{q_i\in Q} \|q_i - \phi(q_i)\|_2 \tag{5}$$

where $\phi : Q \rightarrow \hat{Q}$ is the bijection mapping.

(3) Train the threshold. Only with the reconstruction loss, the model is not able to learn a proper threshold because remaining more points always leads to a better performance, and the model will tend to remain maximum points. Therefore, we design a regularization term to mask matrix $\tilde{M}^{(l)}$ for punishment over more points of retention.

$$\mathcal{L}_{reg} = \frac{1}{N}\sum_{l=1}^{N} \|\tilde{M}^{(l)}\|_2, \tag{6}$$

where N is the number of IPAPool layers, and $\tilde{M}^{(l)}$ is the mask matrix of the l-th layer, which denotes the total number of points remained. If the learnable threshold $\tilde{\phi}^{(l)}$ is too small, more points will be remained in $\tilde{M}^{(l)}$, and $\mathcal{L}_{reg}$ gets larger in value. By minimizing this loss term, the threshold is expected to be larger and the total loss gets small, and more important points is expected to be remained near the threshold boundary.

Overall, the joint loss function is defined as follows:

$$\mathcal{L} = \mathcal{L}_{rec} + \alpha\mathcal{L}_{reg}^{(l)}, \tag{7}$$

where α is the loss value influence factor. Larger α denotes higher compression ratio. After the threshold has been obtained, the soft mask $\tilde{M}^{(l)}$ is binarized for finetuning the encoder parameters. Finally, in the inference phase, we fix the learned thresholds and all other parameters to prune the unimportant points for reconstruction.

IV. Evaluation

In this section, we conduct extensive experiments to validate the effectiveness of the proposed AIRec. We first present the experiment setup, and then we compare our AIRec with the 3 baselines demonstrating that It greatly reduces the transmitted point cloud data and is robust against the noises in the wireless channels for digital twin reconstruction.

A. Experiment Setup

Datasets. We construct experiments on 2 popular point cloud datasets to evaluate the proposed AIRec. (1) The **3D-Mesh** is a generic CAD point cloud dataset that contains 145 3D mesh models from the PU-GAN published dataset, where the number of points contained varies for each object. In our experiments, 40 simple, 40 medium and 40 complex models are used to train the network models and the remaining 25 models are used in the testing phase to evaluate the model performance. (2) The **KITTI** dataset is a real world point cloud dataset collected by LiDAR sensors with sparsity and non-uniformity of the inputs.

We adopt KITTI and datasets released by PU-GAN to evaluate the performance of our methods.

Baselines. We compare the proposed AIRec with three representative point cloud compression methods, including the conventional octree-based methods Octree [44], Draco [1] and a typical deep learning based method Geo-CNN [45].

- Octree [44] is a representative compression method for point sampled models based on an octree decomposition of space.
- Draco [1] is a popular library for compressing and decompressing 3D geometric meshes and point clouds.
- Geo-CNN [45] is a deep learning based geometry compression for point clouds based on deep convolution neural network with the uniform quantization.

Implementation Details. We implement AIRec using Pytorch and train it on a high-performance server equipped with eight Tesla A100 GPUs. The encoder of AIRec consists of 4 SA layers for extracting point-wise features. Recognizing that the semantic information in the shallow layers of the network is limited, we employ the D-FPS sampling strategy for the first 2 layers in the encoder. In contrast, the proposed IPAPool is solely used for the last 2 layers. The radius for the grouping operation is incremented to capture multi-scale features, specifically $([0.2, 0.8], [0.8, 1.6], [1.6, 4.8])$. We utilize the Adam optimizer with an initial learning rate of 0.001. In our experiments, the same pre-trained decoder is utilized across both datasets. However, due to the non-uniform nature of the KITTI dataset, the feature dimensions for each layer of the encoder are set to $[128, 256, 256, 512]$. For the 3D-Mesh dataset, which comprises uniformly generated point clouds, a reduced feature dimension of $[64, 128, 128, 256]$ is adequate. The initial value of the threshold is set at different values to attain various compression ratios. For each epoch, the SNR is randomly selected from a range of 0db to 5db.

[1] https://google.github.io/draco/

To ensure a fair comparison, we set the same compression ratio for the different models. For the octree-based methods, Octree and Draco, we adjust the value of parameter d to control the compression ratio. For Geo-CNN, the network is trained multiple times to derive models with different compression ratios, as the point cloud needs to be transformed into a voxel form with a fixed size for both the training set and the test set. We adopt AWGN as the channel model in our experiments and combine all the compared baselines with idealistic error-free transmission based on Shannon capacity. The bandwidth is set to 64 for all the baselines, corresponding to the transmission of 64 complex symbols through the channel.

Evaluation Metrics. We employ there commonly used metrics [9], [27]: (1) Chamfer distance (CD) [46], (2) Hausdorff distance (HD) [47] and (3) point-to-surface (P2F) [48] to evaluate the quality of the 3D point cloud reconstruction. The smaller the metric values are, the better the reconstruction results are.

B. Performance Results

We first demonstrate that the proposed AIRec is able to greatly reduce the transmission data while preserving the reconstruction accuracy in the ideal channel. Then, we further prove the AIRec is robust against the wireless network noises during the transmission.

Evaluation Result for Ideal Channels. Table I summarizes the evaluation results for the reconstruction of all the baselines. The origin 3D-Mesh dataset is denoted as the "Dense" setting. To further explore the performance over the sparse point cloud, we randomly removed some of the points from the point cloud and generate the "Moderate" setting (50% removed) and the "Sparse" setting (80% removed).

We observe that: (1) AIRec outperforms all the baselines in most cases in the "Dense" settings, and the conventional compression methods have a better performance in the "moderate" and "Sparse" setting. This is mainly because AIRec retains points based on semantic information, and reconstruction can be accomplished by extending these features and points. The performance of the deep learning-based approach degrades without enough semantics. (2) As the compression ratio increases, the performance on all the CD, CH, and P2F of all the baselines degrades. Nevertheless, our AIRec also significantly outperforms other baselines, especially when the compression ratio is high. This is mainly because when the compression ratio is high, AIRec is able to dynamically retain the most important points for the decoder to reconstruct with the proposed importance-aware hierarchical encoder. Different from the traditional approaches which compress the point cloud only based on the structure information, our proposed AIRec can not only retain the important points, but also can compress the information of the unimportant points into dense vectors, ensuring high compression rate while maintaining maximum retention of important information. With more points transferred, the differences between the reconstructed points and ground truth become smaller, AIRec is able to achieve higher reconstruction quality in the digital counterpart.

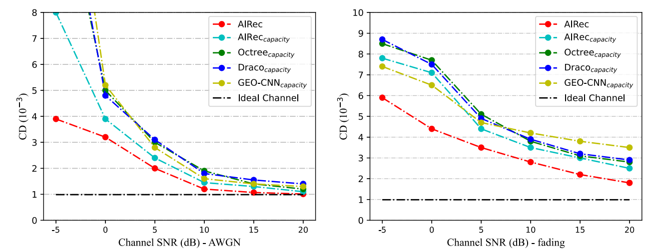


Fig. 6: Evaluation result in noisy AWGN and fading wireless channels. All metrics d in 10^{-3}.

Evaluation Result in Noisy Wireless Channels. Fig. 6 presents the evaluation result in noisy wireless channels with various SNR, including the AWGN and the fading channel for the CD evaluation metric. The experimental settings with the "capacity" in the brackets mean that the conventional digital approach with idealistic capacity-achieving channel codes proposed in [13] is adopted for transmission. For our methods, it means that the joint source and channel coding scheme is replaced with the digital approach.

The result shows that: (1) The noisy channel introduces additional perturbation to the channel symbols, and reduces the accuracy of all the proposed approaches. Nevertheless, the proposed AIRec outperforms all the compared baselines in all the scenarios. Comparing AIRec and AIRec (capacity-achieving), we observe that even under conditions of low channel noise (SNR=15dB), digital methods cannot achieve the accuracy of an ideal channel, while the proposed AIRec with joint source and channel coding can achieve the accuracy close to the ideal channel when SNR=10dB. (2) As the channel noise increases, the accuracy of the digital method deteriorates sharply, while the accuracy of the method proposed in this paper decreases more slowly. This is mainly because the digital approaches suffer from the *cliff effect*, which results in the sharp performance degradation, while our proposed joint source and channel coding scheme is able to learn the task-specific channel-noise resistant coding methods. The impact of channel noise on the reconstruction task is minimized by our proposed method. (3) In the settings of the digital approach, as the noise energy increases, our proposed method is not as effective as the baseline method, especially the traditional compression method. This may be due to the fact that the higher noise causes the change of features, which in turn greatly affects the semantic information in the features, and the feature expanding in our decoder feature further exacerbates the effects of noise and makes the reconstruction process less effective.

C. Ablation Study

In this section, some ablation studies are performed to examine the effectiveness of the proposed importance-aware down-sampling strategy, learnable threshold and the noise-aware strategy.

Effect of Importance-aware Down-sampling Strategy. To further demonstrate the effectiveness of the proposed

Input Points		Dense				Moderate				Sparse			
		Compression Ratio				Compression Ratio				Compression Ratio			
methods	metrics	3×	6×	12×	24×	3×	6×	12×	24×	3×	6×	12×	24×
Octree	CD	0.458	0.712	0.919	1.12	0.523	**0.741**	0.978	1.24	0.735	0.917	**0.118**	**1.61**
	HD	**4.37**	5.98	6.72	9.89	**5.51**	6.47	7.98	11.2	5.98	7.45	8.21	9.90
	P2F	3.46	4.21	4.93	5.68	3.61	4.37	4.88	6.12	**4.32**	4.98	5.18	7.54
Draco	CD	**0.420**	0.772	0.108	1.34	**0.489**	0.792	0.121	1.35	**0.698**	**0.912**	0.134	1.47
	HD	4.59	5.87	6.98	10.7	5.98	6.87	7.91	10.9	5.88	7.39	8.87	10.4
	P2F	3.46	3.68	4.21	6.23	**3.52**	**3.77**	4.91	6.81	4.45	**4.85**	**5.15**	6.92
Geo-CNN	CD	0.832	0.788	0.987	1.25	0.962	0.107	0.987	1.25	0.832	0.788	0.987	1.62
	HD	6.28	6.82	7.37	8.32	6.82	7.53	8.03	9.37	8.19	9.83	10.2	11.8
	P2F	4.09	4.36	4.72	5.38	4.62	4.82	4.98	5.76	5.43	5.89	6.53	7.16
(Our model)	CD	0.623	**0.692**	**0.746**	**0.987**	0.699	0.782	**0.886**	**1.11**	0.847	0.921	0.123	**1.52**
	HD	5.19	**5.72**	**6.23**	**7.10**	5.67	**6.27**	**6.93**	**8.02**	6.98	7.72	8.57	**9.02**
	P2F	**3.25**	**3.43**	**3.67**	**4.10**	3.65	3.79	**4.05**	**4.88**	4.54	4.98	5.46	**6.02**

TABLE I: Experimental results summary of all methods on 3D-Mesh datasets under ideal channel conditions (in CD, HD, and P2F metrics). Best results displayed in **boldface**. All metrics are measured in 10^{-3}.".

		Compression Ratio			
Methods	Metrics	3×	6 ×	12 ×	24×
D-FPS	CD	0.672	0.715	0.801	1.09
	HD	5.31	5.86	6.44	7.25
	P2F	3.51	3.75	3.92	4.39
Feat-FPS	CD	0.665	0.721	0.791	1.12
	HD	5.23	5.98	6.51	7.46
	P2F	3.35	3.62	3.87	4.42
IPAPool (ours)	CD	**0.623**	**0.692**	**0.746**	**0.987**
	HD	**5.19**	**5.72**	**6.23**	**7.10**
	P2F	**3.25**	**3.43**	**3.67**	**4.10**

TABLE II: Effect of importance-aware down-sampling strategy. All metrics are measure in 10^{-3}.

importance-aware down-sampling strategy, we replace the sampling module with the D-FPS [29] and Feat-FPS [42] and report the evaluation result in Table II. Note that the "Dense" setting with full points is adopted in this experiments. We observe that the proposed importance-aware down-sampling strategy is able to achieve higher performance under the same compression ratio. This is because both D-FPS and Feat-FPS are content-agnostic, they compress point cloud frames with different contents to the same number of points. For complex point cloud data, the number of points may not be enough to complete a satisfactory reconstruction, while for simple point cloud frames, they sample a large number of redundant points. Our method, on the other hand, is content-aware and can sample different numbers of points for different point cloud frames, and dynamically retain important points for various point cloud frames, greatly improving the overall compression ratio.

Effect of Learnable Threshold. To demonstrate the significance of our proposed learnable threshold pruning, we consider two variants: AIRec$_{Thres}$, which introduces a fixed pruning threshold instead of the learnable one, and $Topk$, which replaces learnable threshold pruning with top-K pruning. The 'training ratio' refers to the compression ratio used during training, while the 'testing ratio' denotes the compression ratio in the testing stage. For AIRec, the threshold is initialized to achieve the target ratio. Table III presents the experimental results. We observe that AIRec$_{Topk}$ achieves accuracy and efficiency comparable to AIRec$_{Thres}$. However, a manually set threshold may not offer the optimal setting and could impact the model performance. Our approach, on the other hand, strikes an optimal balance between processing efficiency and accuracy, delivering a higher speedup ratio with a marginal loss in accuracy, and vice versa.

Effect of Noise-aware pooling strategy. We showcase the efficacy of the proposed noise-aware joint source and channel coding strategy by contrasting AIRec with a variant AIRec$_{agno}$, which selects points for wireless transmission oblivious to the noise. In contrast, our method dynamically tailors the transmission strategy based on the network Signal-to-Noise Ratio (SNR) conditions. Fig. 7 illustrates the comparison of the remaining points and the reconstruction quality for the two methods under varied network conditions. As displayed in the left part of Fig. 7, with the escalating network noise, the performance of both AIRec and AIRec$_{agno}$ deteriorates, due to the interference of channel noise on feature symbols during the reconstruction process. However, AIRec is more resilient to noise than AIRec$_{agno}$. This is credited to our proposed noise-aware pooling scheme that mitigates the negative impact of noise by enhancing the number of transmission points in high noise scenarios, thereby ensuring the reconstruction accuracy does not substantially plummet, as depicted in the right part of Fig. 7. This strategy is particularly effective when transmitting point cloud data in a highly

		Metrics			
Variants	Training ratio	CD	HD	P2F	Testing ratio
$AIRec_{Thres}$	3×	0.643	5.42	3.54	3×
	6×	0.701	5.53	3.61	6×
	12×	0.748	6.33	3.82	12×
	24×	0.979	3.59	4.05	24×
$AIRec_{Topk}$	3×	0.647	5.39	3.62	3×
	6×	0.721	5.61	3.65	6×
	12×	0.762	6.31	3.85	12×
	24×	0.988	3.67	3.92	12×
AIRec	3×	0.623	5.19	3.25	2.91×
	6×	0.692	5.72	3.32	5.95×
	12×	0.746	6.23	3.67	11.8×
	24×	0.987	3.67	4.10	26.3×

TABLE III: Effect of learnable threshold. CD, HD and P2F are measured in 10^{-3}.

Method	SNR (dB)	Easy	Mod	Hard	Remained Point
3DSSD	0	70.27	62.95	59.98	512
	5	80.81	70.82	65.48	512
	10	84.32	75.54	71.98	512
	15	87.84	78.69	74.98	512
Ours	0	72.12	63.25	51.02	267
	5	81.09	71.73	65.98	125
	10	84.22	75.61	71.84	62
	15	87.76	78.75	75.21	44

TABLE IV: Experimental results on 3D object detection task compared with 3DSSD in AP for the "Car" class.

compressed format over noisy wireless networks, where noise on the compressed features can significantly affect the final outcome. Therefore, under favorable network conditions, our method can reconstruct the point cloud at a higher frame rate and speed, and adaptively lower the frame rate to maintain the reconstruction quality under challenging network conditions.

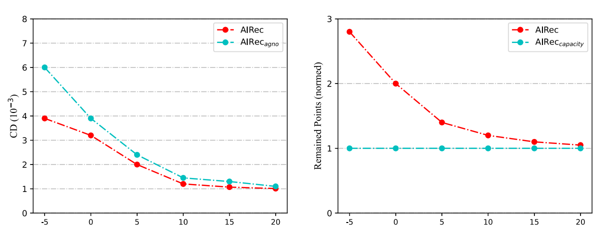


Fig. 7: Effectiveness of the noise-aware pooling.

D. Experiment on 3D Object Detection Task

To further evaluate the proposed schemes on a wider range of intelligent digital twin tasks other than reconstruction, we select 3D target detection tasks which can be widely applied to Intelligent digital twins, such as automatic driving, intelligent transportation and other tasks. In this task, we replaced our designed rebuild headers with 3DSSD [42]object detection headers to verify the performance of our network for intelligent tasks other than digital twin reconstruction. Note that the joint source and channel coding scheme is also adopted for 3DSSD transmission for fair comparison. As can be seen from the table IV, under different SNR conditions and difficulties, compared with 3DSSD, our method can greatly reduce the number of required transmission points, and the resulting accuracy is similar to that of 3DSSD, which ensures the accuracy and reduces the communication overhead at the same time. This is because AIRec can adaptively select the most important foreground points for 3D object detection, and reduce the noises introduced by the background information. Another observation is that the remained points increases in bad network condition, but are always less the fixed setting in 3DSSD.

E. Visualization on KITTI dataset

KITTI dataset is real world point cloud dataset collected by LiDAR sensors in urban scenarios. Reconstruction real LiDAR point cloud is a promising way toward the city-scale digital twin. However, the raw scanned point cloud is extremely sparse and ordered. Shown in Fig. 8, the nearby points are dense, while distant points are very sparse, while, due to the physical characteristics of the radar sensor, the scanned points are linear and almost unusable for surface reconstruction in digital reconstruction. We apply the proposed AIRec on the raw point cloud to reconstruct it and the result shows that we achieved a good reconstruction and we were able to reconstruct all points as a homogeneous surface regardless of the distance from the radar sensor. Note that in this case, the raw point cloud contains $71,825$ points, while we only select 384 important points with feature dimension 64 for transmission, achieving over $4\times$ in compression.

F. Discussion

In practical scenarios, our method presents potential solutions to various real-world challenges. Take smart manufacturing as an example. Plant operators often face difficulties due to inefficient data transmission, which can lead to inaccurate digital twins. The IPAPool technique refines the digital twin by selectively preserving essential points in point cloud data, thus providing a more accurate depiction of the production environment. In the context of autonomous driving, systems frequently struggle to maintain accurate and current digital twins due to noisy or unreliable networks. Our noise-aware coding adapts ensure accurate digital twins in variable network conditions, while IPAPool refines the digital twin's environment by preserving vital sensor data points. These improvements facilitate more efficient navigation and decision-making, thereby augmenting the safety and reliability of autonomous vehicles.

Furthermore, our method could pave the way for future research and applications in the field of digital twin technology. By optimizing the transmission and reconstruction of point cloud data, our method could potentially enable the development of more intricate and detailed digital twins. This

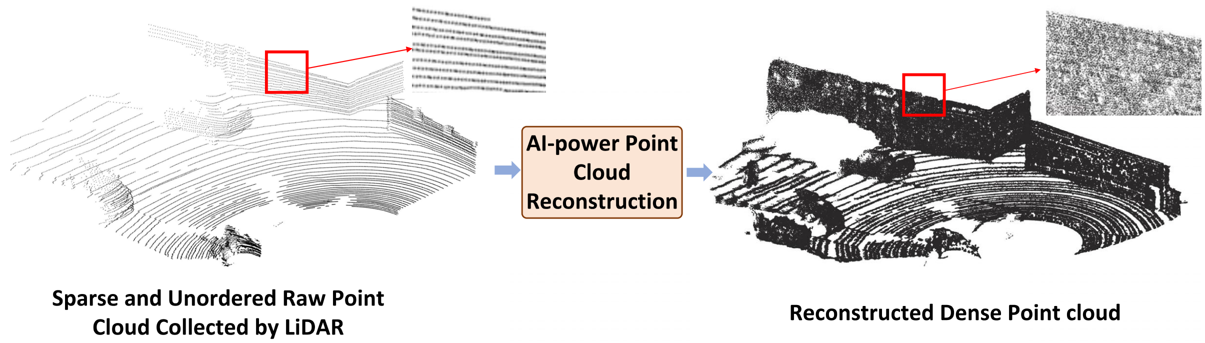


Fig. 8: Visualization on KITTI dataset.

advancement could open new research avenues. Researchers might leverage our methods to explore innovative strategies for optimizing digital twin-based systems, fostering fresh insights and groundbreaking innovations.

V. Conclusion

Existing wireless transmission technology struggles to support real-time point cloud transmission for digital twin reconstruction due to challenges such as massive data volumes, high processing overheads, and low delay-tolerance. To address these challenges, we have proposed a novel, AI-powered end-to-end framework, termed AIRec, for efficient communication within digital twins, encompassing point cloud compression, wireless channel coding, and digital twin reconstruction. The AIRec adopts an encoder-decoder architecture, incorporating a novel importance-aware pooling scheme in the encoder to selectively reduce transmission volume, and a unique noise-aware joint source and channel coding to adjust the transmission strategy based on SNR, enabling a beneficial tradeoff between the transmission rate and reconstruction quality. The decoder can accurately reconstruct the digital twins from the received symbols. Our extensive experiments and comparisons with baseline methods using typical datasets show that we achieve a high-quality reconstruction under a $24\times$ compression ratio. In the future, we plan to expand our AI-powered end-to-end framework to efficiently handle and reconstruct point cloud video streams in digital twins. This extension could broaden the applicability of our method, enabling real-time or near-real-time updating of digital twin models from point cloud video sources.

VI. Acknowledgment

This work was partially funded by the Cultivation of Shenzhen Excellent Technological and Innovative Talents (Ph.D. Basic Research Started) RCBS20200714114943014, and by the Basic research of Shenzhen Science and technology Plan under Grant JCYJ20210324123802006.

References

[1] H. V. Dang, M. Tatipamula, and H. X. Nguyen, "Cloud-based digital twinning for structural health monitoring using deep learning," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 6, pp. 3820–3830, 2021.

[2] S. Li, L. D. Xu, and S. Zhao, "The internet of things: a survey," *Information systems frontiers*, vol. 17, no. 2, pp. 243–259, 2015.

[3] F. Xue, W. Lu, Z. Chen, and C. J. Webster, "From lidar point cloud towards digital twin city: Clustering city objects based on gestalt principles," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 167, pp. 418–431, 2020.

[4] P. A. Burrough, R. A. McDonnell, and C. D. Lloyd, *Principles of geographical information systems*. Oxford university press, 2015.

[5] O. El Marai, T. Taleb, and J. Song, "Roads infrastructure digital twin: A step toward smarter cities realization," *IEEE Network*, vol. 35, no. 2, pp. 136–143, 2020.

[6] Z. Liu, Q. Li, X. Chen, C. Wu, S. Ishihara, J. Li, and Y. Ji, "Point cloud video streaming: Challenges and solutions," *IEEE Network*, vol. 35, no. 5, pp. 202–209, 2021.

[7] M. Chen, A. Feng, R. McAlinden, and L. Soibelman, "Photogrammetric point cloud segmentation and object information extraction for creating virtual environments and simulations," *Journal of Management in Engineering*, vol. 36, no. 2, p. 04019046, 2020.

[8] K. Mirzaei, M. Arashpour, E. Asadi, H. Masoumi, and H. Li, "Automatic generation of structural geometric digital twins from point clouds," *Scientific Reports*, vol. 12, no. 1, p. 22321, 2022.

[9] R. Li, X. Li, C.-W. Fu, D. Cohen-Or, and P.-A. Heng, "Pu-gan: a point cloud upsampling adversarial network," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 7203–7212.

[10] Y. Guo, H. Wang, Q. Hu, H. Liu, L. Liu, and M. Bennamoun, "Deep learning for 3d point clouds: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 12, pp. 4338–4364, 2020.

[11] G. Schrotter and C. Hürzeler, "The digital twin of the city of zurich for urban planning," *PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, vol. 88, no. 1, pp. 99–112, 2020.

[12] T. A. Pedersen, J. A. Glomsrud, E.-L. Ruud, A. Simonsen, J. Sandrib, and B.-O. H. Eriksen, "Towards simulation-based verification of autonomous navigation systems," *Safety Science*, vol. 129, p. 104799, 2020.

[13] M. Jankowski, D. Gündüz, and K. Mikolajczyk, "Wireless image retrieval at the edge," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 1, pp. 89–100, 2020.

[14] M. Krivokuća, P. A. Chou, and M. Koroteev, "A volumetric approach to point cloud compression–part ii: Geometry compression," *IEEE Transactions on Image Processing*, vol. 29, pp. 2217–2229, 2019.

[15] P. A. Chou, M. Koroteev, and M. Krivokuća, "A volumetric approach to point cloud compression—part i: Attribute compression," *IEEE Transactions on Image Processing*, vol. 29, pp. 2203–2216, 2019.

[16] S. Schwarz, M. Preda, V. Baroncini, M. Budagavi, P. Cesar, P. A. Chou, R. A. Cohen, M. Krivokuća, S. Lasserre, Z. Li *et al.*, "Emerging mpeg standards for point cloud compression," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 1, pp. 133–148, 2018.

[17] 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.

[18] M. Yang, C. Bian, and H.-S. Kim, "Deep joint source channel coding for wireless image transmission with ofdm," in *ICC 2021-IEEE International Conference on Communications*. IEEE, 2021, pp. 1–6.

[19] M. Grieves and J. Vickers, "Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems," in *Transdisciplinary perspectives on complex systems*. Springer, 2017, pp. 85–113.

[20] J. Du, Q. Zhu, Y. Shi, Q. Wang, Y. Lin, and D. Zhao, "Cognition digital twins for personalized information systems of smart cities: Proof of concept," *Journal of Management in Engineering*, vol. 36, no. 2, 2020.

[21] C. M. Eastman, C. Eastman, P. Teicholz, R. Sacks, and K. Liston, *BIM handbook: A guide to building information modeling for owners, managers, designers, engineers and contractors*. John Wiley & Sons, 2011.

[22] M. Krivokuca, P. A. Chou, and M. Koroteev, "A volumetric approach to point cloud compression, part ii: Geometry compression," *IEEE Transactions on Image Processing*, vol. 29, 2019.

[23] P. A. Chou, M. Koroteev, and M. Krivokuca, "A volumetric approach to point cloud compression, part i: Attribute compression," *IEEE Transactions on Image Processing*, pp. 1–1, 2019.

[24] R. B. Rusu and S. Cousins, "3d is here: Point cloud library (pcl)," in *2011 IEEE international conference on robotics and automation*. IEEE, 2011, pp. 1–4.

[25] A. Brock, T. Lim, J. M. Ritchie, and N. Weston, "Generative and discriminative voxel modeling with convolutional neural networks," *arXiv preprint arXiv:1608.04236*, 2016.

[26] T. Huang and Y. Liu, "3d point cloud geometry compression on deep learning," in *Proceedings of the 27th ACM international conference on multimedia*, 2019, pp. 890–898.

[27] Y. Huang, Y. Zhu, X. Qiao, Z. Tan, and B. Bai, "Aitransfer: Progressive ai-powered transmission for real-time point cloud video streaming," in *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, pp. 3989–3997.

[28] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 652–660.

[29] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," *Advances in neural information processing systems*, vol. 30, 2017.

[30] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic graph cnn for learning on point clouds," *Acm Transactions On Graphics (tog)*, vol. 38, no. 5, pp. 1–12, 2019.

[31] K. Zhang, M. Hao, J. Wang, C. W. de Silva, and C. Fu, "Linked dynamic graph cnn: Learning on point cloud via linking hierarchical features," *arXiv preprint arXiv:1904.10014*, 2019.

[32] C. R. Qi, W. Liu, C. Wu, H. Su, and L. J. Guibas, "Frustum pointnets for 3d object detection from rgb-d data," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 918–927.

[33] Y. Liu, B. Fan, G. Meng, J. Lu, S. Xiang, and C. Pan, "Densepoint: Learning densely contextual representation for efficient point cloud processing," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 5239–5248.

[34] C. E. Shannon, "A mathematical theory of communication," *The Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[35] A. Vosoughi, P. C. Cosman, and L. B. Milstein, "Joint source-channel coding and unequal error protection for video plus depth," *IEEE Signal Processing Letters*, vol. 22, no. 1, pp. 31–34, 2014.

[36] D. B. Kurka and D. Gündüz, "Successive refinement of images with deep joint source-channel coding," in *2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2019, pp. 1–5.

[37] ——, "Bandwidth-agile image transmission with deep joint source-channel coding," *IEEE Transactions on Wireless Communications*, vol. 20, no. 12, pp. 8081–8095, 2021.

[38] M. Yang and H.-S. Kim, "Deep joint source-channel coding for wireless image transmission with adaptive rate control," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 5193–5197.

[39] E. Bourtsoulatze, D. B. Kurka, and D. Gündüz, "Deep joint source-channel coding for wireless image transmission," *IEEE Transactions on Cognitive Communications and Networking*, vol. 5, no. 3, pp. 567–579, 2019.

[40] D. Burth Kurka and D. Gündüz, "Joint source-channel coding of images with (not very) deep learning," in *International Zurich Seminar on Information and Communication (IZS 2020). Proceedings*. ETH Zurich, 2020, pp. 90–94.

[41] Q. Hu, B. Yang, L. Xie, S. Rosa, Y. Guo, Z. Wang, N. Trigoni, and A. Markham, "Learning semantic segmentation of large-scale point clouds with random sampling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[42] Z. Yang, Y. Sun, S. Liu, and J. Jia, "3dssd: Point-based 3d single stage object detector," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11 040–11 048.

[43] H. Fan, H. Su, and L. J. Guibas, "A point set generation network for 3d object reconstruction from a single image," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 605–613.

[44] R. Schnabel and R. Klein, "Octree-based point-cloud compression." *PBG@ SIGGRAPH*, vol. 3, 2006.

[45] M. Quach, G. Valenzise, and F. Dufaux, "Learning convolutional transforms for lossy point cloud geometry compression," in *2019 IEEE international conference on image processing (ICIP)*. IEEE, 2019, pp. 4320–4324.

[46] M. A. Butt and P. Maragos, "Optimum design of chamfer distance transforms," *IEEE Transactions on Image Processing*, vol. 7, no. 10, pp. 1477–1484, 1998.

[47] M. Berger, J. A. Levine, L. G. Nonato, G. Taubin, and C. T. Silva, "A benchmark for surface reconstruction," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 2, pp. 1–17, 2013.

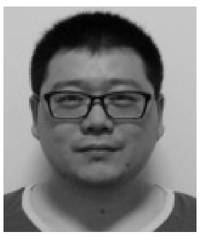
[48] M. Pauly, M. Gross, and L. P. Kobbelt, "Efficient simplification of point-sampled surfaces," in *IEEE Visualization, 2002. VIS 2002*. IEEE, 2002, pp. 163–170.

Mi Li was born in 1982, and received her bachelor degree from Wuhan University of Science and Technology in 2002, master degree from Shenzhen University in 2008, and doctor degree from Tongji University in 2012. Her main research interests include traffic scheduling, emergency evacuation, artificial intelligence, machine learning, big data and data mining. She also hosted in several provincial and national technological innovation funds, has published many journal articles and more than 20 sets of utility model patents and patents for innovation, got provincial innovative and entrepreneurial talent honor in 2022.

Cen Chen currently works as a professor at the school of Future Technology of South China University of Technology. He received the Ph.D. degree in computer science from Hunan University, Changsha, China, in 2019. He previously worked as a Scientist with Institute of Infocomm Research (I2R), Agency for Science, Technology and Research (A*STAR), Singapore. His research interest includes parallel and distributed computing, machine learning and deep learning. He has published more than 60 articles in international conferences and journals on machine learning algorithms and parallel computing, such as HPCA, DAC, IEEE TC, IEEE TPDS, AAAI, ICDM, ICPP, and ICDCS. He has served as a Guest Editor for Pattern Recognition and Neurocomputing.

Xulei Yang is a senior scientist from Institute for Infocomm Research (I2R), A*STAR, Singapore, with more than 10 years of research experiences in image analysis and machine learning. His current research interests focus on deep learning for computer vision and medical imaging. He has published more than 70 scientific papers and international patents.

Joey Tianyi Zhou is currently a senior scientist, Investigator and group manager with A*STAR Centre for Frontier AI Research (CFAR), Singapore. He is also holding an adjunct faculty position (adj. Assoc. Prof.) at National University of Singapore (NUS). Before working at CFAR, he was a senior research engineer with SONY US Research Center in San Jose, USA. Dr. Zhou received a Ph.D. degree in computer science from Nanyang Technological University (NTU), Singapore. His current interests mainly focus on improving the efficiency and robustness of machine learning algorithms. Dr. Zhou organized ICDCS annual workshop on Efficient AI meets Edge Computing, ACML'16 workshop on Learning on Big Data workshop and IJCAI'19 workshop on Multioutput Learning; is serving as an Associate Editor for IEEE Transactions on Emerging Topics in Computational Intelligence (TETCI) and IEEE Access, IET Image Processing, and TPC Chair in Mobimedia 2020; and received NeurIPS Best Reviewer Award in 2017.



Tao Zhang received his Ph.D. degree in the School of Computer Science and Engineering, Central South University, China. He is now an associate professor in Hunan Province Key Laboratory of Industrial Internet Technology and Security, Changsha University, China. His research interests include congestion control, load balancing, performance modeling, analysis, and data center networking.

Yangfan Li received his PhD degree in Computer Science, Hunan University, China in 2022. He is now a lecturer with the School of Computer Science and Engineering, Central South University, China. His research interest includes computer architecture, efficient machine learning and deep learning. He received the bachelor degree of engineering in the School of Automation, Huazhong University of Science and Technology in 2015.