Fast Inter-frame Motion Prediction for Compressed Dynamic Point Cloud Attribute Enhancement

Wang Liu[1], **Wei Gao**[1,2*], **Xingming Mu**[1]

[1] School of Electronic and Computer Engineering, Shenzhen Graduate School, Peking University
[2] Peng Cheng Laboratory
liuwang@stu.pku.edu.cn, {gaowei262, 1700012434}@pku.edu.cn

Abstract

Recent years have witnessed the success of deep learning methods in quality enhancement of compressed point cloud. However, existing methods focus on geometry and attribute enhancement of single-frame point cloud. This paper proposes a novel compressed quality enhancement method for dynamic point cloud (DAE-MP). Specifically, we propose a fast inter-frame motion prediction module (IFMP) to explicitly estimate motion displacement and achieve inter-frame feature alignment. To maintain motion continuity between consecutive frames, we propose a motion consistency loss for supervised learning. Furthermore, a frequency component separation and fusion module is designed to extract rich frequency features adaptively. To the best of our knowledge, the proposed method is the first deep learning-based work to enhance the quality for compressed dynamic point cloud. Experimental results show that the proposed method can greatly improve the quality of compressed dynamic point cloud and provide a fast and efficient motion prediction plug-in for large-scale point cloud. For dynamic point cloud attribute with severely compressed artifact, our proposed DAE-MP method achieves up to 0.52dB (PSNR) performance gain. Moreover, the proposed IFMP module has a certain real-time processing ability for calculating the motion offset between dynamic point cloud frame.

Introduction

In recent years, dynamic point cloud have been widely used in many scenarios, such as autonomous driving, urban modeling, and virtual/augmented reality. However, the existing methods of point cloud data acquisition and generation will cause storage and transmission burdens more than those of image or video. With the rise of point cloud applications, corresponding point cloud compression algorithms and standards have entered the field of vision of researchers. Therefore, the well-known Motion Picture Experts Group (Schwarz et al. 2018) (MPEG) proposed the geometry-based point cloud compression standard (GPCC) and the video-based point cloud compression standard (VPCC). Although the reduction of point cloud code stream in lossy mode is beneficial to the storage and transmission of large point cloud data, it will irreversibly produce information distortion.

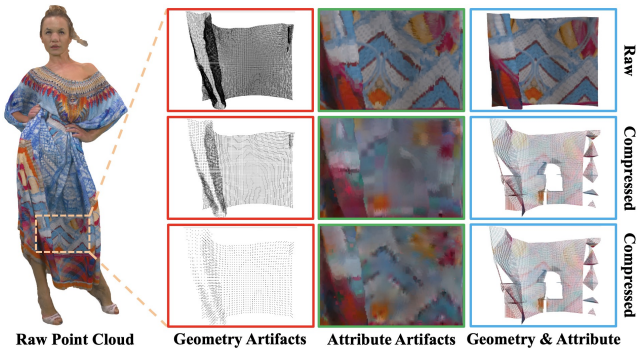

Figure 1: Illustration of point cloud compression artifacts. First column: raw geometry, geometry compressed by octree scheme under $Q_{step} = 2, 4$, respectively. Second column: raw attribute, attribute compressed under $Q_{step} = 51$ by RAHT scheme and predlift scheme, respectively. Last column: raw point cloud geometry and attribute, point cloud compressed by trisoup and RAHT scheme under $Q_{step} = 41$, compressed point cloud by trisoup and predlift scheme under $Q_{step} = 41$.

The geometry compression algorithms under the GPCC framework are divided into two categories: Octree-based (Schnabel and Klein 2006) and Trisoup-based (Anis, Chou, and Ortega 2016). Depending on the geometry compression process, attribute compression algorithms mainly include predlift-based method (Mammou et al. 2018) and RAHT-based method (De Queiroz and Chou 2016). There are significant gaps in point cloud distortion under different compression modes and compression algorithms. Figure 1 shows the subjective experience among them. The geometry distortion is mainly the reduction of the number of points and the position shift, while the attribute distortion shows the blur effect and block effect. Therefore, the goal of quality enhancement of compressed point cloud in a post-processing manner is not only to enhance the subjective experience or objective metric, but also to improve the performance in downstream tasks.

Compressed point cloud quality enhancement methods

*Corresponding author

based on deep learning have emerged currently and achieved remarkable results. A multi-scale sparse convolution framework in (Akhtar et al. 2020) is proposed to restore the compressed geometry information. In order to adaptively learn geometry information with different degrees of distortion, a multi-scale 3D convolution with cross-entropy loss is designed in (Fan et al. 2022b). For the attribute artifact removal task of compressed point cloud, Sheng et al. (Sheng et al. 2022) first propose a framework based on graph convolution and graph attention to organize point cloud and enhance attribute quality. However, existing methods focus on the geometry or attribute quality enhancement for the static compressed point cloud. At present, there is no end-to-end quality enhancement method for compressed dynamic point cloud. Due to the disorder and irregularity of geometry positions, it is difficult to organize point cloud data as efficiently as video data. The key is how to accurately establish the motion relationship between point cloud frames. Furthermore, quality enhancement will become more challenging when both geometry and attributes are lossy.

In this work, we propose a novel framework for quality enhancement of compressed attribute based on inter-frame motion prediction. In order to estimate the explicit inter-frame motion of large point cloud as accurately as possible, we attempt to align the geometry coordinate between frames in a supervised manner. Equipped with the Minkowski engine (Choy, Gwak, and Savarese 2019), we first propose a simple and fast single-frame attribute enhancement method (DAE-S). Combined with designed explicit inter-frame motion prediction, a novel multi-frame enhancement method (DAE-MP) is further developed through inter-frame feature alignment and fusion in our work. The main contributions of this paper are three-fold:

- We propose a novel inter-frame motion prediction module (IFMP) for dynamic point cloud to quickly and explicitly calculate the motion offset of adjacent frames. Through supervised training on geometry positions and a designed motion consistency loss function (MCL), it greatly improves the accuracy of inter-frame motion and ensures the effectiveness of inter-frame feature fusion.
- To the best of our knowledge, this is the first deep learning-based attribute quality enhancement framework for compressed dynamic point cloud. Specifically, we simultaneously present an end-to-end single-frame attribute enhancement model (DAE-S) and a two-stage multi-frame attribute enhancement model (DAE-MP).
- Considering the inconsistency of attribute distortion caused by the dynamic point cloud compression algorithm, we put forward a Frequency Component Perception (FCP) module to adaptively integrate deep feature with different frequencies.

Related Work

In this section, we briefly introduce the related literature for point cloud enhancement methods, compressed video enhancement methods and compressed point cloud enhancement methods.

Point Cloud Enhancement. Common point cloud enhancement tasks mainly include point cloud upsampling (Yu et al. 2018; Li et al. 2022), point cloud frame interpolation (Lu et al. 2021; Zeng et al. 2022), point cloud completion (Yuan et al. 2018; Zhang et al. 2022), point cloud denoising (Luo and Hu 2020) and point cloud compression quality enhancement, etc. PUNet (Yu et al. 2018) is the pioneer deep learning method for upsampling, which provides an infrastructural scheme for training and testing an upsampling model. PointINet (Zeng et al. 2022) extracts the relationship between point cloud frames through the 3D scene flow module designed by (Liu, Qi, and Guibas 2019). Without any structural assumption, PCNet (Yuan et al. 2018) proposes a deep learning-based completion framework that processes directly on raw point cloud. Luo et al. (Luo and Hu 2020) use latent surface reconstruction to remove noise under the framework of manifold learning. As one of the point cloud enhancement tasks, point cloud compression enhancement has borrowed from other tasks and achieved success.

Compressed Video Enhancement. Making full use of spatial-temporal information is the key factor for removing compressed video artifact. In order to achieve frame alignment and feature fusion in an end-to-end manner, Xue et al. (Xue et al. 2019) propose a task-oriented flow model to remove video blocking effects. Inter-prediction in video compression leads to severe quality fluctuations, which are ignored by previous methods. Guan et al. (Guan et al. 2019; Yang et al. 2018) design a multi-frame quality enhancement strategy to improve enhancement performance, which mainly use peak-quality frame (PQF) to enhance low-quality frame (LQF). However, PQF-based methods ignore the fact that some high-quality detail regions still exist in LQF. Xv et al (Xu et al. 2019) propose the NL-ConvLSTM model, which can capture hidden spatio-temporal information in the adjacent frames of the target frame for the reduction of compression artifacts. A large number of previous works on compressed video enhancement will provide a sufficient research paradigm for compressed point cloud enhancement.

Compressed Point Cloud Enhancement. The compressed point cloud enhancement task focuses on improving the quality of the compressed point cloud in a post-processing manner. According to the type of information that has been compressed, existing quality enhancement methods can be divided into geometry-based (Fan et al. 2022b; Borges, Garcia, and De Queiroz 2022; Akhtar et al. 2020), attribute-based (Sheng et al. 2022) and hybrid-based. Meanwhile, the above methods can be categorized as human perception-oriented and machine perception-oriented, which derive from the difference in subsequent processing tasks. The above methods are mainly applied to single-frame stationary point cloud. Therefore, the quality enhancement of compressed dynamic point cloud will be explored in this paper.

Proposed Method

Problem Formulation and Method Description

Given a point cloud sequence $\{P_{t-N}, ..., P_t, ..., P_{t+N}\}$ with compressed attribute, where $P_t = (p_t^c, p_t^f)$ represents geom-

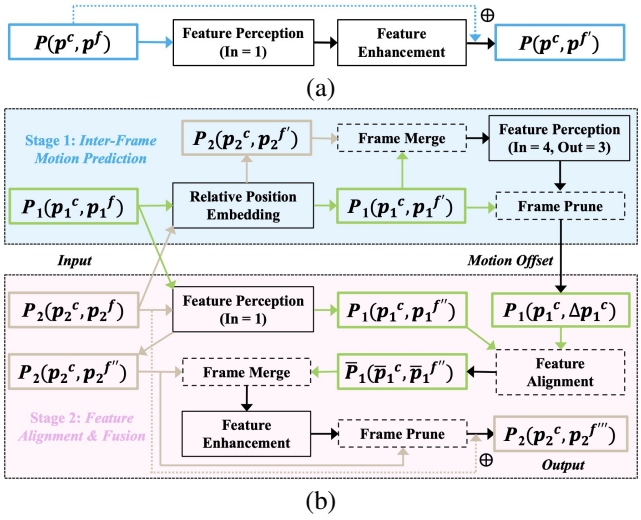

Figure 2: (a) The architecture of DAE-S. (b) The architecture of DAE-MP.

etry information p_t^c and attribute information p_t^f of the t-th frame. The goal of the attribute quality enhancement task for compressed dynamic point cloud is to generate a set of enhanced point cloud $\left\{\hat{P}_{t-N}, ..., \hat{P}_t, ..., \hat{P}_{t+N}\right\}$ with high-quality attributes, where $\hat{P}_t = (p_t^c, \hat{p}_t^f)$. To simplify the description, the enhanced frames of the intermediate frames are sequentially output with the reference of two adjacent frames, i.e., $N = 1$ in our method.

The overall framework is presented in Fig. 2. For the attribute quality enhancement task of compressing dynamic point cloud, we design a single-frame enhancement method (DAE-S) and a multi-frame enhancement method (DAE-MP) with motion prediction. Fig. 2(a) describes the structure of DAE-S, which contains a Feature Perception(FP) module and a Feature Enhancement(FE) module. The former is responsible for extracting the feature of a single frame point cloud, while the latter is used for the output of enhanced attribute. DAE-S is trained in an end-to-end manner in our experiments.

Fig. 2(b) presents the structure of DAE-MP, it mainly includes two stages: inter-frame motion prediction (IFMP) and inter-frame feature alignment. IFMP shares a similar network structure to FP module, except that the dimensions of the input and output are different. Given a target frame and a reference frame, the IFMP module is employed to explicitly compute the motion displacement from the reference frame to the target frame. The motion prediction process is supervised by lossless geometry coordinate information. At the same time, deep features of the target frame and reference frame are independently extracted by the FP module. After applying the calculated motion offset to the feature alignment process of the reference frame, the feature fusion process is performed by FE module. Therefore, the overall training process of DAE-MP is divided into two stages. The main function of the frame alignment module we propose is to propagate the attribute information of the reference

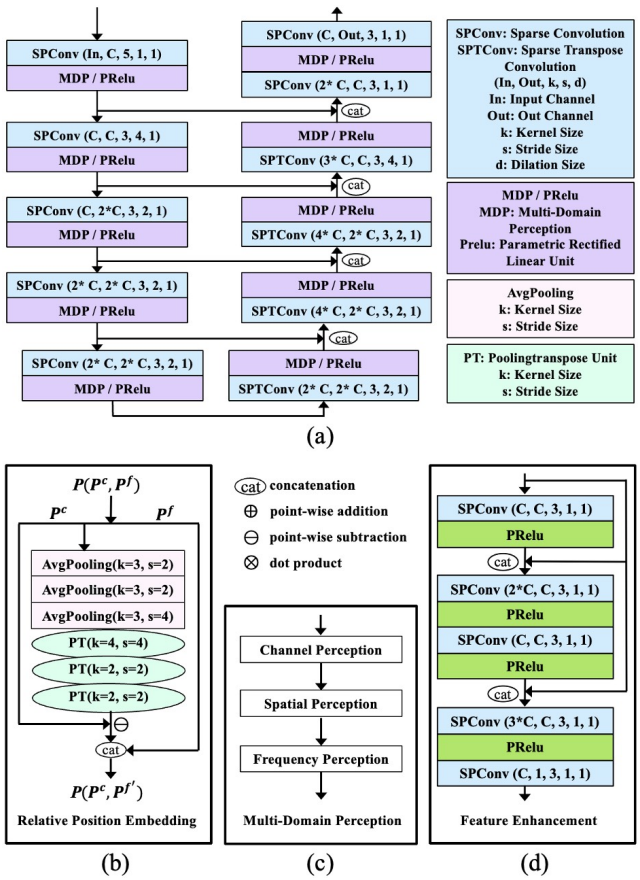

Figure 3: (a) Feature Perception. (b) Relative Position Embedding. (c) Multi-Domain Perception. (d) Feature Enhancement.

point cloud frame to the target frame through inter-frame motion prediction. Frame merge and frame prune operators for dynamic point cloud are described in detail in (Fan et al. 2022a).

Feature Perception. Inspired by point cloud geometry upsampling method (Akhtar et al. 2022), we design a multi-scale U-Net encoder-decoder structure (Ronneberger, Fischer, and Brox 2015) to extract deep features of point cloud. As shown in Fig. 3, the detailed structure of the feature perception module is drawn in Fig. 3(a). The encoding part consists of five sparse convolutional layers with different downsampling scales, while the decoding part contains six sparse transposeconvolutional layers with different upsampling scales. In each sparse convolutional layer, we insert a Multi-Damian Perception (MDP) module to obtain richer point cloud feature. In addition, features in the same layer of the encoding part and the decoding part are concatenated through skip connection. The sparse convolutional layer and sparse transposeconvolutional layer are marked as SPConv and SPTConv, respectively, where the parameters represent the input feature dimension, output feature dimension, convolution kernel size, stride size and dilation size in turn. We

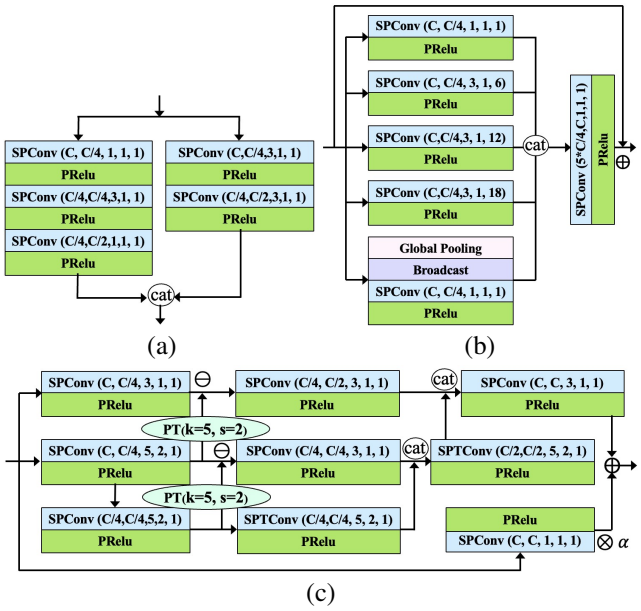

Figure 4: (a) Channel Perception. (b) Spatial Perception. (c) Frequency Component Perception. They are stacked sequentially to form an MDP module.

choose PRelu as the nonlinear activation layer, which shows better performance in the experiment.

The simple U-Net structure is difficult to fully extract rich and complex point cloud feature. Inspired by existing image and video feature extraction methods, we pay attention to the characteristics of the point cloud in the channel domain, spatial domain and frequency domain, respectively. For the purpose of extracting more discriminative features and speedind up the convergence of deep networks, Inception-Resnet block (Szegedy et al. 2017) is designed to adaptively learn the channel features of image. After that, Inception-Resnet block is widely applied to point cloud processing tasks (Wang et al. 2021; Akhtar et al. 2022). As a base plugin, it is also utilized in our approach and labeled as Channel Perception (CP) in Fig. 4(a). In addition to the channel domain, the way to expand the receptive field range by increasing the size of the convolution kernel has been proved in (He et al. 2015; Chen et al. 2017). In order to limit the increase in parameter size caused by expanding the convolution kernel and maintain the receptive field range, atrous convolution is proposed and considered to enhance enhance the spatial feature extraction ability of existing methods in (Chen et al. 2018). Similar to regular 2D atrous convolution implemented on image, sparse atrous convolution under the Minkowski engine are developed to extract deep feature of point cloud in (Akhtar et al. 2022). As a base unit, it is also adapted in our approach and labeled as Spatial Perception (SP) in Fig. 4(b).

The existing point cloud feature extraction methods mainly start from the channel domain and the spatial domain. This can be competent for the point cloud processing tasks that only contain geometry information. However, attribute enhancement of compressed dynamic point cloud involves explicit attribute feature extraction and fusion. However, point cloud compression algorithms inevitably cause blocking effects and noise on attribute features. These distortions show an inconsistent distribution of information loss at different frequency, which is difficult to be simultaneously enhanced by regular convolution operations. Inspired by the fact that frequency features are not uniformly distributed in image super-resolution (Li et al. 2021), we use adaptive downsampling and upsampling to separate and extract three categories of deep features with different frequency. Furthermore, learnable parameters α are introduced to adaptively adjust the weights of features with different frequency in the fusion process. The learning paradigm of adaptive separation and fusion of frequency features can be applied to other computer vision tasks.

Specifically, we design a simple point cloud frequency perception module with adaptive feature separation and fusion. As shown in Fig. 4(c), Frequency Component Perception (FCP) module mainly includes multi-scale downsampling convolution layer, upsampling convolution layer and point-wise feature operators. Given the current feature f, the output $\bar{f}$ is generated as follows:

$$
\begin{aligned}
f_1 &= SPConv1(f), \\
f_2 &= SPConv2(f), \\
f_3 &= SPConv3(f_2), \\
\hat{f}_1 &= SPConv4(f_1 - PT(f_2)), \\
\hat{f}_2 &= SPConv5(f_2 - PT(f_3)), \\
\tilde{f}_2 &= Cat(\hat{f}_2, SPTConv1(f_3)), \\
\tilde{f}_1 &= Cat(SPTConv2(\tilde{f}_2), \hat{f}_1), \\
\bar{f} &= SPConv6(\tilde{f}_1) + \alpha \cdot SPConv7(f),
\end{aligned}
\tag{1}
$$

where *SPConv2* and *SPConv3* represent sparse convolution with downsampling, *SPTConv1* and *SPTConv2* represent sparse transposeconvolution with upsampling, *PT* refers to the poolingtranspose operator, which is used to uniformly upsample features. After multi-scale feature sampling and subtraction, the FCP module can extract rich point cloud features with different frequency.

Motion Prediction. Inter-frame relationship establishment and feature interaction are missing in our proposed DAE-S model. In fact, inter-frame motion prediction in point cloud is crucial for improving the quality of the compressed attribute. On the one hand, dynamic point cloud data itself has significant temporal similarity and spatial similarity. On the other hand, the inter-frame prediction strategy in the attribute compression algorithm of dynamic point cloud explicitly establishes the spatio-temporal relationship. An implicit motion estimation method is proposed for deep dynamic point cloud compression (D-DPCC) in (Fan et al. 2022a). D-DPCC adopts the K-NearestNeighbor (KNN) search strategy for coarse geometry coordinate alignment, which is limited by the huge time and calculation overhead brought by its search process. In our proposed DAE-MP model, point-wise inter-frame motion is explicitly computed

via supervised learning.

Given the compressed target frame $P_t(p_t^c, p_t^f)$ and reference frame $P_r(p_r^c, p_r^f)$, an novel inter-frame motion prediction module(IFMP) is design to generate a new predicted frame $\bar{P}_t(\bar{p}_t^c, p_t^f)$. The module is trained by minimizing the Chamfer Distance(CD)(Fan, Su, and Guibas 2017) between the predicted point cloud and the target point cloud. In addition to the difference in input and output, IFMP adopts a network structure similar to that of the FP module. Different from the attribute information input of the DAE-S model, DAE-MP also considers the local similarity in the geometry distribution of adjacent frames. Inspired by the relative positional embedding used in the self-attention mechanism (Vaswani et al. 2017) and the irregularity of the point cloud distribution, we implement relative position information encoding by calculating local relative coordinates. As shown in Fig. 3(b), we perform multiple local average pooling and point-wise subtraction on the geometry coordinates of the input point cloud. Concatenating relative position information with point cloud attribute information, IFMP can more effectively predict the inter-frame motion offset of dynamic point cloud.

Training Scheme

Our proposed single-frame quality enhancement method(DAE-S) is trained in an end-to-end manner. Given a compressed point cloud frame $P(p^c, p^f)$ with lossless geometry p^c and lossy attribute p^f, the enhanced point cloud frame obtained by DAE-S is denoted as $\hat{P}(p^c, \hat{p}^f)$. Suppose the original frame with high attribute quality is labeled as $\tilde{P}(p^c, \tilde{p}^f)$, the Charbonnier (Charbonnier et al. 1994) loss function is set to optimize the model, which is defined as follows:

$$\mathcal{L} = \sqrt{(\tilde{p}^f - \hat{p}^f)^2 + \varepsilon}, \varepsilon = 1e^{-8}. \quad (2)$$

Compared with the DAE-S method, our proposed multi-frame enhancement model(DAE-MP) additionally includes an inter-frame motion prediction module(IFMP). In DAE-MP, IFMP is performed before inter-frame feature alignment and fusion. Given the compressed target frame $P_t(p_t^c, p_t^f)$ and reference frame $P_r(p_r^c, p_r^f)$, IFMP aim to find a geometry coordinate offset Δp_r^c and generate a new predicted frame $\bar{P}_t(\bar{p}_t^c, p_t^f)$ for the target frame. In order to make the predicted frame $\bar{P}_t$ close to the target frame P_t in terms of geometry position distribution, Chamfer Distance(CD)(Fan, Su, and Guibas 2017) is chosen to measure the geometry difference. The CD loss function is defined as follows:

$$\mathcal{L}_{CD} = \frac{1}{|\bar{p}_t|} \sum_{x \in \bar{p}_t} \min_{y \in p_t} \|x - y\|_2 + \frac{1}{|p_t|} \sum_{y \in p_t} \min_{x \in \bar{p}_t} \|y - x\|_2 . \quad (3)$$

Motion Consistency Loss. Taking into account the removal of noise values in motion offsets and motion continuity between adjacent frames, we design a motion consistency loss function to reduce abnormal offset values. Given an intermediate frame $P_2(p_2^c, p_2^f)$ and two adjacent reference frames $P_1(p_1^c, p_1^f), P_3(p_3^c, p_3^f)$, the forward offset $\Delta p_{1\to2}^c$ and backward offset $\Delta p_{3\to2}^c$ are calculated by IFMP module, respectively. Assuming that the movement trajectory of most points between consecutive frames is linear, the above two offsets are close in value and opposite in direction. Based on this assumption, the motion consistency loss is defined as follows:

$$\mathcal{L}_{MCL} = \sqrt{(\Delta p_{1\to2}^c + \Delta p_{3\to2}^c)^2 + \varepsilon}, \varepsilon = 1e^{-8}. \quad (4)$$

Then, the joint loss function for IFMP module is computed as:

$$\mathcal{L} = \mathcal{L}_{CD}(\bar{p}_1^c, p_2^c) + \mathcal{L}_{CD}(\bar{p}_3^c, p_2^c) + \lambda \cdot \mathcal{L}_{MCL}, \quad (5)$$

where $\bar{p}_1^c = p_1^c + \Delta p_{1\to2}^c$, $\bar{p}_3^c = p_3^c + \Delta p_{3\to2}^c$. The parameter λ is set to 1 in our experiments.

Experiments

Datasets

Following the previous work for deep dynamic point cloud compression(D-DPCC), we choose 8i Voxelized Full Bodies (8iVFB)(d'Eon et al. 2017) for model training and evulation. There are four dynamic sequences with published compression configuration file in 8iVFB, which includes *Longdress*, *Redandblcak*, *Solider* and *Loot*. Each sequence contains 300 frames and the precision of geometry coordinate is 10-bit. The first and third sequences are selected for model training, while the others for testing. All 600 frames of training data and 600 frames of test data are compressed in the G-PCC reference software TMC13v22. Our work targets compressed attribute quality enhancement for dynamic point cloud, so the Octree-based geometry lossless mode and the RAHT-based attribute lossy mode are adopted.

Implementation Settings

All training sequence and testing sequence are compressed at two Quantization Parameter (QPs), i.e., 51, 46. For compressed sequences with different quantization parameters, model training is performed separately. Since the chroma information (Y-channel) retains more attribute information in GPCC framework and the sensitivity of the human eye, we mainly pay attention to quality enhancement on Y-channel in YCrCb space. In addition, we choose increased Peak Signal-to-Noise Ratio ($\triangle$PSNR) (Wang et al. 2004) compared with compressed sequence as objective quality evaluation criteria to evaluate quality enhancement improvement.

Our method is implemented on the PyTorch platform with Minkowski Engine (Choy, Gwak, and Savarese 2019). Adam optimizer (Kingma and Ba 2014) is adopted to train our model with $\beta_1 = 0.9, \beta_2 = 0.999$ and $\varepsilon = 10^{-8}$. Learning rate is initially set to 5×10^{-4} and linearly decays to 2×10^{-4} after 200 epochs. Then, it linearly reduces to 1×10^{-4} after 200 epochs. Data augmentation skills are not involved in our experiments. The batch size is set to 16 and the model is trained on NVIDIA Tesla V100 GPU.

Comparison to Other Methods

The primary goal of our proposed DAE-S model and DAE-MP model is to improve the attribute quality of compressed dynamic point cloud. In addition to focusing on objective metric gains relative to compressed attribute, existing deep

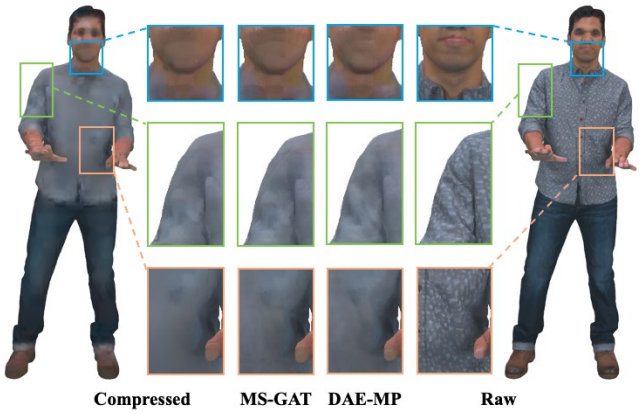

Figure 5: Comparison of subjective experience improvement.

QP	Sequence	MS-GAT	DAE-S	DAE-MP
51	*Loot*	0.32	0.39	0.52
	Redandblack	0.20	0.24	0.35
46	*Loot*	0.34	0.36	0.45
	Redandblack	0.26	0.31	0.39

Table 1: Overall performance comparison for △PSNR (dB) over standard compressed sequences at QP = 51, 46.

Sequence	Enc. & Dec.	MS-GAT	DAE-MP	Size(M)
Loot	10.4 + 11.4	6.7 + 12.2	3.77	1.18 / 2.63
Redandblack	9.7 + 11.3	6.1 + 10.5	3.29	

Table 2: Comparison for average inference time(s) and model parameter size over standard compressed sequences at QP = 51. The second column represents the inference time and data preprocessing time of MS-GAT. The last column records the parameter size: MS-GAT / DAE-MP.

learning-based single-frame attribute enhancement methods are considered in our experiments. Since there is no existing end-to-end compressed attribute quality enhancement work for dynamic point cloud, we mainly consider single-frame enhancement method (Sheng et al. 2022) for comparison. For fair comparision, we retain it on our datesets.

The overall performance comparison is shown in Table 1. The experimental results show that both our proposed single-frame enhancement model (DAE-S) and multi-frame enhancement model (DAE-MP) exceed MS-GAT in terms of the PSNR gain. In particular, the performance gain of the test sequence *Loot* under QP51 reaches 0.52dB after attribute enhancement. In addition to the comparison of objective evaluation metric, we also consider the subjective experience improvement after attribute enhancement. In Fig. 5, the blocking effect produced by the attribute compression algorithm RAHT is effectively removed in our proposed DAE-MP method. At the same time, the texture information of the point cloud is significantly enhanced.

Quality Fluctuation. Quality fluctuation is another commonly used objective criterion for dynamic point cloud quality metric, which reflects the range of quality variation between frames. Generally, small fluctuation means coherence and stability of the point cloud. We evaluate the quality fluctuation by calculating the standard deviation (SD) of PSNR value. As shown in Fig. 7, the PSNR values of a clip with 20 frames is plotted as a line chart. For the sequence *Loot*, the SD values of compressed attribute, MA-GAT and ours are 0.06, 0.07, and 0.06, respectively. Experimental results show that our method does not aggravate quality fluctuation with the overall quality improvement, which benefits from frame alignment and frame fusion.

Inference Speed. The time consumption of the point cloud compression process is another indicator to evaluate the performance of the compression algorithm. Similarly, the deep learning-based quality enhancement method of compressed point cloud also needs to consider the inference speed. The existing single-frame attribute enhancement method MS-GAT mainly uses the graph convolution framework to extract point cloud features. Although the point cloud processing method based on graph convolution has advantages in efficiently organizing point cloud data, it is limited by computational complexity and data preprocessing. In Table 2, we first measure the average time consumed by the dynamic point cloud sequence in the encoding part(Enc) and the decoding part(Dec), respectively. For MS-GAT, we split the test sequence into a series of non-overlapping spatial blocks of size 64 and enhance sequentially. In this case, data processing time and model inference time are calculated separately. Compared with MS-GAT, our proposed DAE-MP directly processes the entire point cloud, which prevents the semantic structure of the point cloud from being destroyed during the partitioning process. Although MS-GAT has a small parameter size, the inference speed of our method has been greatly improved.

Motion Prediction. To demonstrate the accuracy of the motion offset estimated by the proposed IFMP model, we compute the CD value of the aligned reference and target frame with respect to geometry coordinate. The overall results are shown in Table 5. In Table 3, we also analyze the running speed of the IFMP module for dynamic point cloud with different scales. For the test sequence *Redandblcak* with significant motion between frames, the designed IFMP module quickly reduce the geometry distance from 30.3 to 3.51. Next, we also visualize the explicit motion prediction computed by the IFMP module similar to a 2D optical flow map. As shown in Fig. 6, movement of human arm parts with large displacements are clearly marked in the red and blue dotted box. Furthermore, The CD value between the predicted frame of the reference frame and the target frame drops to 2.17. The above experiments show that our proposed IFMP module can be regarded as a lightweight explicit motion estimation plug-in and embedded in other dynamic point cloud processing tasks.

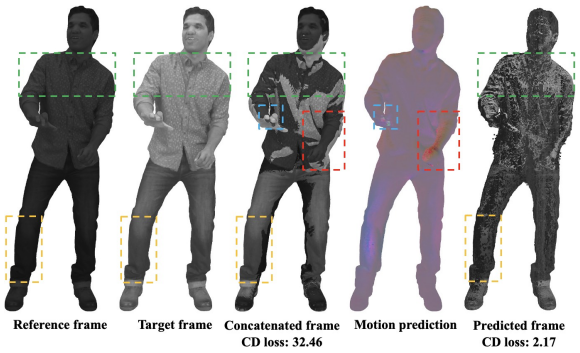

Figure 6: Visualization of explicit motion prediction.

Sequence	Number	Time(s)	CD
Loot	800K	0.53	15.65 / 2.87
Redandblcak	700K	0.42	30.30 / 3.51

Table 3: Comparison for point cloud scale, average speed of motion prediction and average CD value at QP = 51.

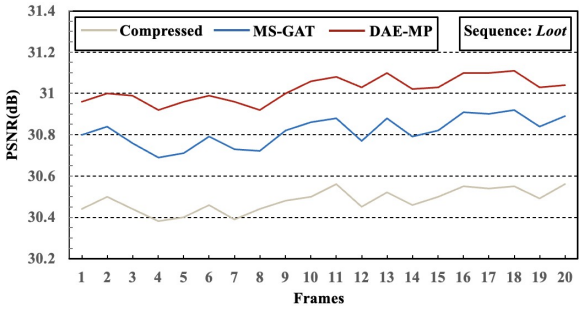

Figure 7: PSNR curves of the sequence *Loot* at QP = 51.

Sequence	*Loot*		*Redandblack*	
QP	51	46	51	46
w/o RPE	3.79	3.26	5.18	4.89
w/o MCL	4.48	5.01	8.63	7.26
IFMP	2.87	2.83	3.51	3.67

Table 5: Ablation experiments for CD value without RPE module and MCL module.

Ablation Study

Effectiveness of IFMP. To achieve inter-frame motion prediction and feature fusion, we propose a fast explicit motion offset estimation method(IFMP). In fact, the relationship between point cloud frames can be established implicitly through the network itself. Given a target frame and several adjacent reference frames, the frame concatenation operator is applied before participating in the DAE-S network. After that, the predicted value of the target frame is obtained by coordinate pruning. The scheme we proposed above is denoted as DAE-M, which is not involved in explicit motion prediction. As shown in Table 4, the DAE-M method without the IFMP module causes a significant performance drop. We speculate that the compressed attribute information with distortion limits the implicit inter-frame feature fusion in the DAE-M method. Therefore, supervised learning of coordinate offsets using lossless geometry information can guarantee the accuracy of motion estimation in IFMP.

Effectiveness of RPE. In this paper, we assume that dynamic point cloud are compressed in a geometry-lossless and attribute-lossy mode. Similar to self-attention mechanisms widely used in image processing tasks, we perform local relative positional embedding(RPE) on point cloud geometry coordinates. The emergence of RPE can help the network to use the similarity between point cloud frames to learn non-local correspondence, which will further improve the accuracy of motion prediction between frames. Therefore, we removed the RPE module in IFMP and observed its impact on motion prediction. Experiments presented in Table 5 show that the embedding of RPE improves the accuracy of motion prediction.

Sequence	*Loot*		*Redandblack*	
QP	51	46	51	46
DAE-M	0.41	0.37	0.27	0.35
w/o FCP	0.44	–	0.29	–
DAE-MP	0.52	0.45	0.35	0.39

Table 4: Ablation experiments for △PSNR (dB) without IFMP module and FCP module.

Effectiveness of FCP. A novel point cloud Frequency Component Perception approach (FCP) is designed and embedded into our feature perception module. As shown in Table 4, the intervention of FCP further improves the performance of the model.

Effectiveness of MCL. Although the aligned reference frame is globally close to the target frame, the CD loss function inevitably causes a few outliers in the motion offset. The above phenomenon motivates us to consider adopting an unsupervised strategy to impose consistency constraints on the inter-frame motion of adjacent frame. As shown in Table 5, removing the MCL constraint has a severe impact on the accuracy of motion prediction.

Conclusion

In this paper, we propose a novel attribute enhancement method for compressed dynamic point cloud. To achieve fast and accurate inter-frame information alignment and fusion, we design an inter-frame motion prediction module with relative position information encoding and motion consistency. Considering the inconsistency of attribute distortion, deep feature are adaptively integrated from channel domain, spatial domain and frequency domain in our method. Our proposed method outperform existing compression attribute enhancement methods in both objective metrics and subjective experience. We believe the proposed method can also extend to other dynamic point cloud attribute processing tasks.

Acknowledgments

This work was supported by Natural Science Foundation of China (62271013, 62031013), Shenzhen Fundamental Research Program (GXWD20201231165807007-20200806163656003), Shenzhen Science and Technology Program (JCYJ20230807120808017). This work was also sponsored by CAAI-MindSpore Open Fund, developed on OpenI Community (CAAIXSJLJJ-2023-MindSpore07).

References

Akhtar, A.; Gao, W.; Zhang, X.; Li, L.; Li, Z.; and Liu, S. 2020. Point cloud geometry prediction across spatial scale using deep learning. In *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*, 70–73. IEEE.

Akhtar, A.; Li, Z.; Van der Auwera, G.; Li, L.; and Chen, J. 2022. Pu-dense: Sparse tensor-based point cloud geometry upsampling. *IEEE Transactions on Image Processing*, 31: 4133–4148.

Anis, A.; Chou, P. A.; and Ortega, A. 2016. Compression of dynamic 3D point clouds using subdivisional meshes and graph wavelet transforms. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 6360–6364. IEEE.

Borges, T. M.; Garcia, D. C.; and De Queiroz, R. L. 2022. Fractional super-resolution of voxelized point clouds. *IEEE Transactions on Image Processing*, 31: 1380–1390.

Charbonnier, P.; Blanc-Feraud, L.; Aubert, G.; and Barlaud, M. 1994. Two deterministic half-quadratic regularization algorithms for computed imaging. In *Proceedings of 1st international conference on image processing*, volume 2, 168–172. IEEE.

Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; and Yuille, A. L. 2017. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4): 834–848.

Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; and Adam, H. 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, 801–818.

Choy, C.; Gwak, J.; and Savarese, S. 2019. 4d spatio-temporal convnets: Minkowski convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 3075–3084.

De Queiroz, R. L.; and Chou, P. A. 2016. Compression of 3D point clouds using a region-adaptive hierarchical transform. *IEEE Transactions on Image Processing*, 25(8): 3947–3956.

d'Eon, E.; Harrison, B.; Myers, T.; and Chou, P. 2017. ISO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG) Input Document WG11M40059/WG1M74006; 8i Voxelized Full Bodies-a Voxelized Point Cloud Dataset. *MPEG: Geneva, Switzerland*.

Fan, H.; Su, H.; and Guibas, L. J. 2017. 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*, 605–613.

Fan, T.; Gao, L.; Xu, Y.; Li, Z.; and Wang, D. 2022a. D-dpcc: Deep dynamic point cloud compression via 3d motion prediction. *arXiv preprint arXiv:2205.01135*.

Fan, X.; Li, G.; Li, D.; Ren, Y.; Gao, W.; and Li, T. H. 2022b. Deep Geometry Post-Processing for Decompressed Point Clouds. In *2022 IEEE International Conference on Multimedia and Expo (ICME)*, 1–6. IEEE.

Guan, Z.; Xing, Q.; Xu, M.; Yang, R.; Liu, T.; and Wang, Z. 2019. MFQE 2.0: A new approach for multi-frame quality enhancement on compressed video. *IEEE transactions on pattern analysis and machine intelligence*, 43(3): 949–963.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 37(9): 1904–1916.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Li, X.; Jin, X.; Yu, T.; Sun, S.; Pang, Y.; Zhang, Z.; and Chen, Z. 2021. Learning omni-frequency region-adaptive representations for real image super-resolution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 1975–1983.

Li, Z.; Li, G.; Li, T. H.; Liu, S.; and Gao, W. 2022. Semantic point cloud upsampling. *IEEE Transactions on Multimedia*.

Liu, X.; Qi, C. R.; and Guibas, L. J. 2019. Flownet3d: Learning scene flow in 3d point clouds. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 529–537.

Lu, F.; Chen, G.; Qu, S.; Li, Z.; Liu, Y.; and Knoll, A. 2021. Pointinet: Point cloud frame interpolation network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 2251–2259.

Luo, S.; and Hu, W. 2020. Differentiable manifold reconstruction for point cloud denoising. In *Proceedings of the 28th ACM international conference on multimedia*, 1330–1338.

Mammou, K.; Tourapis, A.; Kim, J.; Robinet, F.; Valentin, V.; and Su, Y. 2018. Lifting scheme for lossy attribute encoding in TMC1. *Document ISO/IEC JTC1/SC29/WG11 m42640, San Diego, CA, US*.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, 234–241. Springer.

Schnabel, R.; and Klein, R. 2006. Octree-based Point-Cloud Compression. *PBG@ SIGGRAPH*, 3.

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

Sheng, X.; Li, L.; Liu, D.; and Xiong, Z. 2022. Attribute artifacts removal for geometry-based point cloud compression. *IEEE Transactions on Image Processing*, 31: 3399–3413.

Szegedy, C.; Ioffe, S.; Vanhoucke, V.; and Alemi, A. 2017. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Wang, J.; Ding, D.; Li, Z.; and Ma, Z. 2021. Multiscale point cloud geometry compression. In *2021 Data Compression Conference (DCC)*, 73–82. IEEE.

Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4): 600–612.

Xu, Y.; Gao, L.; Tian, K.; Zhou, S.; and Sun, H. 2019. Non-local convlstm for video compression artifact reduction. In *Proceedings of the IEEE/CVF international conference on computer vision*, 7043–7052.

Xue, T.; Chen, B.; Wu, J.; Wei, D.; and Freeman, W. T. 2019. Video enhancement with task-oriented flow. *International Journal of Computer Vision*, 127: 1106–1125.

Yang, R.; Xu, M.; Wang, Z.; and Li, T. 2018. Multi-frame quality enhancement for compressed video. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 6664–6673.

Yu, L.; Li, X.; Fu, C.-W.; Cohen-Or, D.; and Heng, P.-A. 2018. Pu-net: Point cloud upsampling network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2790–2799.

Yuan, W.; Khot, T.; Held, D.; Mertz, C.; and Hebert, M. 2018. Pcn: Point completion network. In *2018 international conference on 3D vision (3DV)*, 728–737. IEEE.

Zeng, Y.; Qian, Y.; Zhang, Q.; Hou, J.; Yuan, Y.; and He, Y. 2022. Idea-net: Dynamic 3d point cloud interpolation via deep embedding alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6338–6347.

Zhang, R.; Gao, W.; Li, G.; and Li, T. H. 2022. QINet: Decision Surface Learning and Adversarial Enhancement for Quasi-Immune Completion of Diverse Corrupted Point Clouds. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–14.