

基于注意力融合多尺度特征的解压缩 点云质量增强方法

钟 芯¹, 唐春明^{2*}, 彭凌西³

(1. 广州大学计算机科学与网络工程学院, 广东广州 510000; 2. 广州大学数学与信息科学学院, 广东广州 510000;
3. 广州大学机械与电气工程学院, 广东广州 510000)

摘要: 基于几何的点云压缩算法(Geometry-based Point Cloud Compression, G-PCC)可以实现显著的点云压缩效率,但在低比特率场景下解压缩点云会产生严重的几何压缩伪影,并对整体视觉体验产生负面影响.为解决这一问题,本文提出了一种基于注意力融合多尺度特征的解压缩点云几何质量增强方法.具体地,该方法设计了多尺度输入模块对解压缩点云进行下采样操作,得到不同尺度的点云数据.接着,多尺度的点云被并行输入到离散卷积网络中提取从局部到全局的多尺度特征信息.最后,本文设计了跨尺度注意力特征融合模块来对多尺度特征进行融合,以增强特征的完整性和准确性.实验结果表明,本文所提出的方法在公开数据集上的平均峰值信噪比达到了 67.968 4 dB,相较于标准压缩算法 G-PCC 提高了 1.629 4 dB,主客观实验结果均表明本文方法能进一步提高解压缩点云的质量.

关键词: 点云压缩;深度学习;伪影去除;质量增强;多尺度特征;离散卷积

中图分类号: TP391

文献标识码: A

文章编号: 0372-2112(2025)08-2794-11

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20240914

A Method for Enhancing the Quality of Decompressed Point Clouds Based on Attention-Fused Multi-Scale Features

ZHONG Xin¹, TANG Chun-ming^{2*}, PENG Ling-xi³

(1. School of Computer Science and Cyber Engineering, Guangzhou University, Guangzhou, Guangdong 510000, China;
2. School of Mathematics and Information Sciences, Guangzhou University, Guangzhou, Guangdong 510000, China;
3. School of Mechanical and Electrical Engineering, Guangzhou University, Guangzhou, Guangdong 510000, China)

Abstract: Geometry-based point cloud compression (G-PCC) can achieve significant point cloud compression efficiency, but decompressing point clouds in low bit rate scenarios produces severe geometric compression artifacts and negatively affects the overall visual experience. To address this problem, this paper proposes a geometric quality enhancement method for decompressed point clouds based on attentional fusion of multiscale features. Specifically, the method designs a multi-scale input module to perform downsampling operations on the decompressed point cloud to obtain point cloud data at different scales. Then, the multi-scale point clouds are input in parallel into a discrete convolutional network to extract multi-scale feature information from local to global. Finally, a cross-scale attentional feature fusion module is designed in this paper to fuse the multi-scale features to enhance the completeness and accuracy of the features. The experimental results show that the proposed method achieves an average peak signal-to-noise ratio of 67.968 4 dB on the publicly available dataset, which is an improvement of 1.629 4 dB compared to the standard compression algorithm G-PCC, and the subjective and objective experimental results show that the method can further improve the quality of decompressed point clouds.

Key words: point cloud compression; deep learning; artifact removal; quality enhancement; multi-scale features; sparse convolution

1 引言

随着虚拟现实设备的进一步普及,人们越来越追

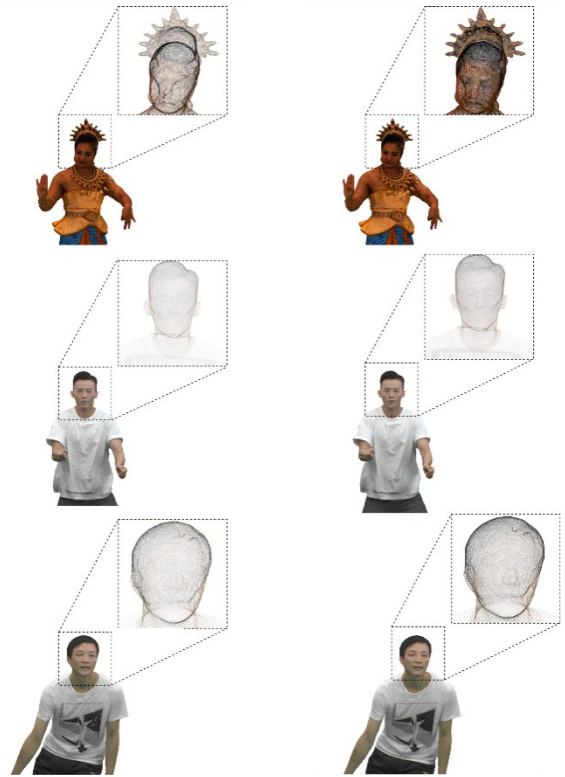
求更具沉浸感的视觉体验,点云作为三维场景的主要表示方式,被视为下一代沉浸式媒体中最为关键的内

容之一. 点云是在三维空间中稀疏且不规则分布的点的集合. 点云中的每个有效点都有其几何坐标, 如笛卡尔坐标系中的 (X, Y, Z) , 以及相关的属性信息, 如RGB颜色、法向量和反射率. 由于其灵活性和强大的表示能力, 点云越来越多地应用于身临其境的通信、机器人、地理信息系统和自动驾驶等领域^[1].

点云可根据其表示的内容划分为点云对象和点云场景. 点云场景通常通过三维扫描或激光雷达动态获取, 涵盖整个环境或场景^[2,3]. 而点云对象则表示特定的物体或目标, 并进一步分为静态和动态两种. 静态点云描述的是单个物体的几何和属性信息, 而动态点云除了描述物体的几何和属性信息外, 还包括其时间上的变化. 一帧点云的数据量可以从几百个点到几千万个点不等, 例如, 由8i实验室^[4]获取的典型点云序列中, 一帧静态点云可以达到几百万个点. 庞大的数据量不利于点云的储存和运输, 亟需高效的点云压缩方法.

为了解决上述问题, 国际标准化组织移动图像专家组(Moving Picture Experts Group, MPEG)自2017年以来一直在深入研究和推广高效点云压缩的潜在技术, 最终形成了针对动态点云的压缩标准, 即基于视频的点云压缩(Video-based Point Cloud Compression, V-PCC)^[5], 以及针对静态点云的压缩算法标准, 即基于几何的点云压缩(Geometry-based Point Cloud Compression, G-PCC)^[6]. 本文主要研究点云标准压缩算法G-PCC, 其主要针对静态点云对象. G-PCC其主要框架是首先对点云的几何信息进行编码, 然后再根据重建的几何信息对点云的属性信息进行编码. G-PCC拥有高效的压缩效率, 但是有损的压缩过程会导致解压缩点云中出现的伪影. 尤其在低比特率场景下, 压缩算法需要将更多的数据丢弃, 以便达到更小的文件大小, 从而产生更严重的伪影. 图1比较了G-PCC压缩后的点云和原始点云, 压缩后的点云会存在局部模糊或缺失和边缘欠平滑的现象, 这些压缩引起的几何伪影会直接导致解压缩点云的误差过大, 还可能导致数据丢失, 显著降低视觉质量, 严重影响压缩点云的后续应用.

在压缩伪影去除与质量增强的研究领域, 2D图像和视频压缩已成为一个广泛讨论的话题. 然而, 点云作为一种3D数据的表达形式, 其离散性和无序性的数据结构与图像和视频等传统数据有着本质的不同. 如果直接将2D图像或视频压缩的方法应用于点云处理, 将面临诸多挑战. 在点云处理的其他任务中, 一些研究尝试将3D点云转换为2D形式以进行特征提取^[7]. 这种方法的优势在于能够借鉴2D领域已有的成熟技术, 但也存在明显的缺点, 即3D数据向2D数据转换的过程中不可避免地引入额外误差. 因此, 直接在3D空间中处理点云已成为当前的主要研究方向. 离散卷积作为直接



注:左:使用G-PCC TMC13v14.0压缩后重建的点云;右:原始点云.

图1 点云

处理点云的一种方式, 正受到越来越多的关注, 例如谷歌提出的离散卷积方法和Minkowski Net^[8]. 同时, 随着几何点云压缩算法G-PCC和基于视频的压缩算法(V-PCC)的提出, 解压缩点云的质量增强问题也逐渐成为研究热点, 其中大部分工作基于深度学习方法展开. 例如, 对于V-PCC, Jia等人^[9]提出了一种两步法来去除几何伪影并提高V-PCC压缩效率. 第一步基于学习的伪运动补偿, 旨在去除伪影; 第二步利用几何图形中近场和远场深度场之间的强相关性, 进一步改善结果. Xing等人^[10]的工作为V-PCC引入了一种在二维几何图像上应用维纳滤波器的自适应去噪方法. Akhtar等人^[11]开发了一种新颖的环外点云几何伪影去除方案, 能在不增加带宽成本的情况下显著提高重建质量. Akhtar等人^[12]还通过使用稀疏卷积设计了PU-Dense, 采用渐进式缩放(Progressive Upscaling)和多尺度特征提取, 显著提高了重建点云质量, 但其主要适合密集点云的质量重建. Fan等人^[13]提出了DGPP, 使用生成多尺度占用概率, 并逐步累积, 采用由粗到精的方法提升点云恢复效果. 但该方法依赖于预测的占用概率进行点云重建, 如果概率预测不准确, 可能会影响最终的恢复质量, 且3D卷积网络计算量较大, 导致训练和推理时间较长, 对计算资源需求较高. Lee等人^[14]发现, 在V-PCC编码器将三维点云转换为二维斑块时, 尤其是在生成2D补丁

的精炼分割过程中,点云中的一些点可能因过度平滑而未被转换或被丢弃,因此他们提出了一种距离加权细化分割方法,减少遗漏点数量,从而提高视觉质量.对于G-PCC,Sheng等人^[15]提出了一种多尺度图注意力网络,用于消除G-PCC的属性压缩伪影.该方法通过切比雪夫图卷积提取点云属性特征,并采用多尺度方案捕捉当前点与其邻居之间的短程和长程相关性来去除压缩导致的伪影.王伟^[16]通过分析点云数据特性,利用压缩感知进行“压缩采样”与重建,并采用K-SVD算法提升信号的稀疏表示,从而减少了编码比特数,并更好地恢复原始数据.Ding等人^[17]为G-PCC标准开发了一种基于学习的自适应循环滤波器,以减少属性压缩伪影,他们的方法首先生成多个最可能的采样偏移作为潜在的压缩失真近似值,然后通过线性加权来减少伪影.

上述的研究方法中,对于G-PCC质量增强的研究更多的是集中在解压缩点云属性的质量增强.然而,点云的几何信息与属性信息密切相关,几何信息在点云中提供每个点的位置信息,属性信息的呈现依赖于这些位置信息.当解压缩后的点云中几何信息误差过大、局部几何结构缺失或模糊时,即相较于原始点云,部分点的位置发生偏差甚至直接丢失,这些点的属性信息必然也会受到影响,导致误差增加.因此,单纯去除属性伪影不足以显著提升点云的重建质量,几何伪影的去除同样至关重要.且上述研究方法中较少在3D空间内直接运用离散卷积来处理点云数据.因此,本研究将离散卷积运用在基于G-PCC的几何质量增强任务上.

此外,在针对点云的研究中,普遍存在局部上下文信息不足的问题,不能有效捕获点云数据中的全局信息和局部信息之间的关联性,限制了点云特征提取的有效性和准确性.在点云处理的其他任务中,运用多尺度技术对点云进行全局和局部信息的提取得到了很多关注^[18,19].多尺度技术在点云上应用是指对点云进行稀疏或者密集采样来获得不同尺度的点云数据.不同尺度的点云特征往往可以提供互补的信息,注重高尺度(密集点云)的细节信息和低尺度(稀疏点云)的上下文信息可以更全面地描述点云数据,从而提高重建点云的准确性.另外,注意力机制在点云任务上的研究也越来越成熟^[20,21].因此,本研究结合多尺度和注意力机制,对解压缩点云进行有效的特征提取和融合.

基于上述探讨,为了解决静态点云压缩标准算法G-PCC在压缩过程中带来的几何伪影问题,本文提出了一种基于Minkowski离散卷积的注意力融合多尺度特征的G-PCC解压缩点云几何质量增强方法.该方法的实现过程包括以下几个关键步骤.

首先,在多尺度采样阶段,以G-PCC解压缩点云作

为输入,对其进行下采样得到多个尺度点云.本文以G-PCC解压缩点云作为高尺度点云(密集点云)来提供几何细节信息,低尺度点云(稀疏点云)来提供全局上下文信息.这一多尺度表达方式注重从全局到局部不同层次特征的提取,使高尺度点云和低尺度点云能够共同提供丰富的特征信息,从而为后续神经网络对几何质量增强点云的重建奠定基础.

接着,为了直接在3D空间中提取解压缩点云的多尺度特征并重建几何质量增强点云,本文设计了一个基于Minkowski离散卷积的多尺度神经网络.该网络以多尺度点云为输入,通过并行结构分别对不同尺度的数据进行特征提取,最终输出相应尺度的重建几何质量增强点云.

最后,本文提出了一种基于点注意力的多尺度特征融合模块.该模块能够有效增强神经网络重建点云中的关键几何特征.通过融合不同尺度的重建点云,可以进一步恢复丢失或存在误差的几何信息,从而获得具有更高几何质量的增强点云.

本文的主要贡献如下:

(1)提出了一种基于Minkowski离散卷积的多尺度神经网络,该网络采用多分支特征提取结构,能够接收多尺度点云作为输入,并对其进行并行特征提取,提取从全局到局部的特征表达,这种设计有效支持了几何质量增强点云的重建.

(2)提出了一个基于点注意力的多尺度特征融合模块,该模块将神经网络输出的多尺度点云中的关键几何特征进行增强,并有效融合点云细节信息和上下文信息,恢复缺失或者存在误差的几何信息,获取几何质量增强的重建点云.

2 基于注意力融合多尺度特征的解压缩点云几何质量增强方法

本文基于深度学习技术,旨在去除经过G-PCC的解压缩点云在有损压缩过程中产生的几何伪影,提高解压缩点云重建质量.

首先,原始点云数据经过G-PCC得到解压缩点云数据.由于点云本身具有离散性和无序性,解压缩点云中的点和原始点云中的点并非一一对应,这会导致输入到神经网络进行质量增强时,对于解压缩点云无法直接获得与之对应的基准真实值以计算损失函数并进行反向传播.为了解决这一问题,通常采用的方法是在训练过程中,通过搜索算法在原始点云中找到与解压缩点云中每个点欧式距离最小的点,作为基准真实值进行损失计算.然而,这种方法存在显著缺点:因为搜索基准真实值的过程发生在神经网络的训练过程中,点云数据量巨大,搜索过程耗时较长,导致神经网络训

效率低下. 针对这一问题, 本文改进了基准真实值的生成方式, 在神经网络训练之前, 预先在原始点云数据中生成一组与解压缩点云点一一对应的基准真实值, 称为“与解压缩点云对齐的基准真实值”. 这一改进避免了训练过程中进行实时搜索的开销, 为后续神经网络进行梯度更新中计算解压缩点云与原始点云之间的误差提供参照. 接着, 通过块生成方法将解压缩点云和基准真实值分割成较小的块, 以减少计算量并适应计算资源. 然后, 将这些块输入到所提出的多尺度神经网络中进行几何特征学习, 从而提取点云的高级特征, 重建出尽可能接近原始点云的几何信息, 提高重建点云的几何质量. 最后, 将网络输出的几何质量增强点云块输入到所提出的多尺度特征融合模块中进行融合, 得到几何质量增强的重建点云.

2.1 问题建模

本文用 $P \in \mathbf{R}^{N \times 3}$ 表示原始点云, 其中, N 是点的数量, 3 对应于 3D 几何坐标 (X, Y, Z) . 经过 G-PCC 压缩并解码后, 得到的解压缩点云记为 $P' \in \mathbf{R}^{N \times 3}$. 通过将所提出的方法模型 Ψ 应用于 P' , 可以得到质量增强后的点云 $\hat{P} \in \mathbf{R}^{N \times 3}$. 本文的目标是最小化质量增强点云 \hat{P} 几何信息和原始点云 P 几何信息之间的差异. 其数学表达式如下:

$$\hat{P} = \Psi(P' | \Theta) \quad (1)$$

其中, Ψ 是本文所提出的方法, Θ 是它的可训练参数.

2.2 基于 Minkowski 离散卷积的多尺度神经网络

2.2.1 多尺度输入模块

由于点云数据规模不一, 从几千到数百万个点不等. 将整个点云直接输入神经网络进行学习会遇到很大的困难, 因此有必要将点云划分为较小的块以便进行后续处理. 本文采用了一种基于物理位置的块生成方法^[22], 将点云按空间位置分割成多个大小相等的正方体块, 并将每个正方体块内的点归为一块. 这样不仅避免了重复和遗漏点的取样问题, 同时也降低了计算复杂度.

使用该方法对原始点云和压缩后的点云分别进行分块, 使得二者在物理位置上的块相同, 即每个块内的点具有相似的几何特征. 这种一致性有助于模型更有效地学习压缩点云与原始点云之间的误差, 从而提高重建质量.

为了更全面地捕获点云的几何信息, 本文采用了最远点采样 (Farthest Point Sampling, FPS)^[23] 方法, 将点云下采样到多个尺度, 即 m 个采样级别.

FPS 是一种在点云处理中广泛应用且表现强大的算法, 能够在不同稀疏性条件下有效捕获重要的坐标特征. 通过这种多尺度处理, 高尺度上保留了更多细节信息, 低尺度则捕获了更广泛的上下文信息, 从而提供

更加丰富的特征表示. 这些不同尺度的点云被并行输入到网络中进行学习, 进一步提升了点云几何信息的整体表示能力.

2.2.2 离散卷积模块

所提议的网络结构基于 U-Net 架构, 通过 Minkowski Engine 库实现了离散卷积. 如图 2 所示, 网络由两部分组成: 一个编码器网络和一个解码器网络. 编码器通过一系列下采样卷积层和归一化层逐步提取输入数据的特征, 而解码器则通过反卷积层和归一化层逐步恢复空间分辨率, 并结合编码器提取的特征来重建输出.

在编码器部分, 首先接收多尺度输入模块生成的数据 I , 并进行多尺度特征的提取. 编码器的第一层是一个 $5 \times 5 \times 5$ 的离散卷积层:

$$E_0 = \text{CONV}(\text{MFIM}(I)) \quad (2)$$

其中, $\text{MFIM}()$ 表示多尺度输入模块, $\text{CONV}()$ 表示离散卷积. E_0 是从输入数据中提取的初始特征. 接着, 使用步长为 2 的 $3 \times 3 \times 3$ 离散卷积层, 进行下采样. 在下采样过程中, 特征维度增加一倍, 点的数量减少:

$$E_n = \text{D_CONV}(E_{n-1}) \quad (3)$$

其中, $\text{D_CONV}()$ 表示对前一层输出的稀疏卷积操作. E_n 是每个下采样层的中间输出. 在解码器部分, 采用步长为 2 的 $3 \times 3 \times 3$ 转置稀疏卷积, 进行上采样:

$$D_n = \text{Tr_CONV}(D_{n-1}) \quad (4)$$

其中, $\text{Tr_CONV}()$ 表示上一层输出的转置稀疏卷积操作, D_n 是每一层向上采样的中间输出. 在解码器的最后一层, 利用 $1 \times 1 \times 1$ 稀疏卷积来完善特征, 并生成与输入形状相同的最终输出. 网络转发过程中会存储中间编码器 (E_1, E_2, \dots, E_n) 和解码器 (D_1, D_2, \dots, D_n) 的中间输出. 为了获取上下文信息, 将编码器路径和解码器路径的信息结合起来.

最后利用跨尺度特征融合模块来融合所有尺度特征, 跨尺度特征融合模块在下一小节详细介绍. 所提出的多尺度神经网络是完全卷积网络, 因为它只包含卷积层, 没有任何密集层. 因此, 该框架具有能够处理不同大小的输入块的优势.

2.3 基于点注意力的跨尺度特征融合模块

在多尺度神经网络进行特征提取时, 多尺度输入模块将输入点云下采样到 m 个不同尺度, 并在离散卷积模块中提取各尺度的几何特征. 这些不同尺度的几何特征蕴含了不同层次的信息, 如局部细节和全局上下文信息. 为了增强各尺度涵盖的不同几何特征并进行融合, 重建出几何信息增强的点云, 受文献^[24]启发, 本文设计了一个基于点注意力的多尺度特征融合模块. 该模块结构如图 3 所示, 主要包括以下几个步

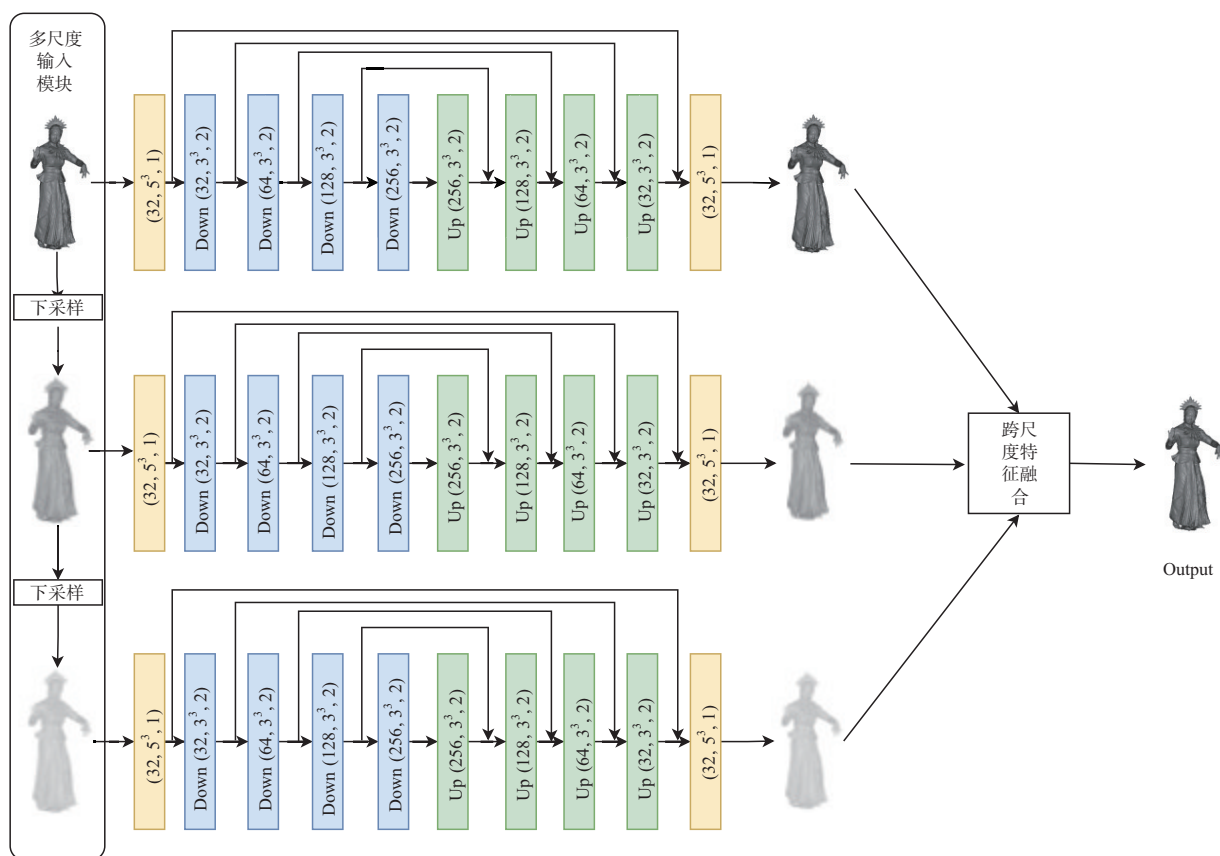


图2 基于Minkowski离散卷积的多尺度神经网络

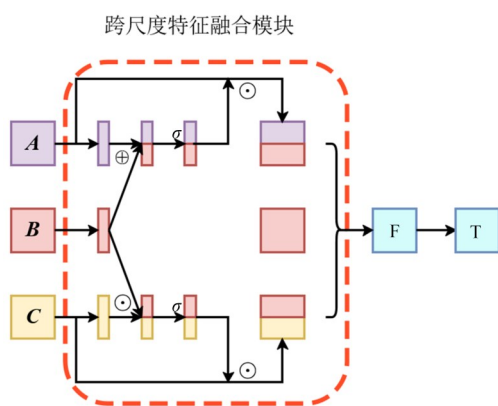


图3 基于点注意力的跨尺度特征融合模块

骤:融合前变换、跨尺度注意力特征学习、特征融合。

2.3.1 融合前变换

由于多尺度神经网络输出的特征具有不同的尺度,特征的大小和分布不尽相同,因此无法通过简单相加实现特征融合.为了统一特征的尺度,本研究采用了一种基于几何距离加权插值^[25]的方法.该方法将低尺度特征恢复到原始G-PCC解压缩点云的尺度.在将低尺度特征恢复到原始尺度的过程中,这一操作可以被视为一个从稀疏到密集的上采样过程.上采样过程需要生成新的点以确保稀疏特征能够完整地恢复为密集

特征,为确保低尺度特征在上采样过程的完整性,该方法利用几何距离的权重来决定上采样点的几何特征.具体来说,几何距离加权插值是一种常用于点云处理中点的插值方法.其基本原理是根据点云中已知点之间的几何距离来计算未知点的坐标.该方法能够充分利用点与点之间的几何关系,在上采样过程中保持点云的几何特征一致性.

假设低尺度特征的点数为 n_d 个点,上采样后原始尺度的点云包含 n 个点($n_d < n$).第 i^{th} 个上采样点的属性特征 h'_i 计算如下:

$$h'_i(\mathbf{x}_i) = \frac{\sum_{j=1}^K \psi_j(\mathbf{x}_i) h_j}{\sum_{j=1}^K \psi_j(\mathbf{x}_i)} \quad (5)$$

其中, $\psi_j(\mathbf{x}_i) = \frac{1}{d(\mathbf{x}_i, \mathbf{x}_j)^2}$, $i \in \{1, 2, \dots, n\}$ 是基于几何距离的权重函数,表示点 \mathbf{x}_i 与其第 j 个邻近点 \mathbf{x}_j 之间的距离倒数平方, $d(\mathbf{x}_i, \mathbf{x}_j)$ 是两点之间的欧几里得距离. K 表示每个上采样点在低分辨率点云 n_d 点中相邻点的数量.

使用该方法进行上采样的目的是确保在上采样过程中,新的点云能够保持与原始点云一致的几何特征.这种加权插值方法通过对近邻点的特征进行加权平

均,可以有效地平滑过渡,并捕捉到更准确的几何信息,从而为后续的特征融合模块提供更好的输入。

2.3.2 跨尺度注意力特征学习

在统一了原始尺度后,为了充分考虑特征中所包含的几何信息,本文设计了一个基于点注意力的特征融合模块。在本文中,多尺度神经网络将输入点云下采样到 $m=2$ 个不同尺度,并连同原始尺度点云并行进行特征提取。因此,本文定义原始尺度的点云为 A ,下采样一次的点云为 B ,下采样两次的点云为 C 。其中, A 来自高尺度点云,点数密集,代表细节信息; C 来自低尺度点云,点数稀疏,拥有上下文信息。相比于 B , A 拥有更多的细节信息,而 C 则包含更丰富的上下文信息。因此,本文提出在进行跨尺度注意力特征学习时,以 B 为基本尺度,在此基础上利用跨尺度注意力机制将点云 A 中的细节信息和点云 C 中的上下文信息作为输入利用跨尺度注意力机制进行有效的增强后并融合,重建出几何质量增强的点云。

具体来说,首先,将 B 和 A 进行逐元素加法操作,以计算细节注意力掩码 M_{res} ,同样地,将 C 和 B 进行逐元素乘法操作以计算语义注意力掩码 M_{sem} :

$$M_{\text{res}} = \sigma(z(A) + z(B)) \quad (6)$$

$$M_{\text{sem}} = \sigma(z(C) \cdot z(B)) \quad (7)$$

其中, σ 表示 sigmoid 激活函数, z 表示融合前变换操作。按照上述步骤,分别在 A 和 C 上分别应用对应的掩码 M_{res} 和 M_{sem} :

$$A' = A \cdot M_{\text{res}} \quad (8)$$

$$C' = C \cdot M_{\text{sem}} \quad (9)$$

其中,运算符“ \cdot ”表示逐元素(元素级别)的乘法运算。对于式(8), A 是原始尺度的点云特征, M_{res} 是细节注意力掩码。逐元素乘法操作将 A 中的每个元素与对应位置的 M_{res} 元素相乘,从而对 A 中的特征进行加权。这个加权过程是基于细节信息的注意力掩码,目的是突出原始尺度点云中的细节特征。对于式(9), C 是低尺度的点云特征, M_{sem} 是语义注意力掩码。逐元素乘法操作将 C 中的每个元素与对应位置的 M_{sem} 元素相乘,从而对 C 中的特征进行加权。这个加权过程是基于上下文信息的注意力掩码,目的是突出低尺度点云中的上下文特征。

通过跨尺度注意力特征学习,分别对包含不同几何信息的特征(包括细节和上下文)进行增强,提高了对多尺度神经网络的不同尺度特征的综合利用能力。同时,在通过跨尺度注意力特征学习之后,进行特征融合时,细节和上下文信息也得到了突出,增强了点云几何特征表示的准确性,构建出几何信息更准确的重建点云。

2.3.3 特征融合

通过跨尺度注意力特征学习获得了增强后的特征,将这些特征 A' 、 B 和 C' 进行堆叠操作(对应图3中的F),以便聚合来自不同尺度的信息。然后,使用多层感知器(Multi-Layer Perceptron, MLP, 见图3中的T)对堆叠后的特征进行平滑处理,从而得到融合后的多尺度特征。

2.4 损失函数

本文的损失函数由两部分组成,以衡量重建点云的几何质量与原始点云之间的差异。假设原始点云为 P ,通过本文方法重建的几何质量增强点云为 \hat{P} 。为了尽可能使 \hat{P} 接近 P ,本文定义了两部分损失函数,即 $P2\hat{P}$ 和 $\hat{P}2P$ 。

$P2\hat{P}$ 损失:以原始点云 P 为基准,对于 P 中的每个点 $p_i \in P$,计算其到重建点云 \hat{P} 的最近邻点 \hat{p}_j 的平方距离。其目的在于度量 P 和 \hat{P} 的相似性,从而引导重建点云 \hat{P} 更接近原始点云 P 。

$$L_{P \rightarrow \hat{P}} = \frac{1}{|\hat{P}|} \sum_{p_i \in P, \hat{p}_j \in \hat{P}} \min \|p_i - \hat{p}_j\|_2^2 \quad (10)$$

$\hat{P}2P$ 损失:以重建点云 \hat{P} 为基准,对于 \hat{P} 中的每个点 $\hat{p}_j \in \hat{P}$,计算其到原始点云 P 的最近邻点 p_i 的平方距离。 $\hat{P}2P$ 损失则用于直接反映 \hat{P} 和 P 之间的误差,使其最小化:

$$L_{\hat{P} \rightarrow P} = \frac{1}{|P|} \sum_{\hat{p}_j \in \hat{P}, p_i \in P} \min \|\hat{p}_j - p_i\|_2^2 \quad (11)$$

总损失函数为两部分的等权求和,即

$$L_{\text{total}} = \frac{1}{2} (L_{P \rightarrow \hat{P}} + L_{\hat{P} \rightarrow P}) \quad (12)$$

3 实验

3.1 实验数据

本实验在 8iVFBv2 数据集^[26]上对网络进行了训练,使用了其中的 Longdress、Loot 和 RedandBlack 三个序列来构建训练集。为了验证模型的泛化能力,在 8iVFBv2、8iVSLF^[4]和 Owl^[27]三个由 MPEG 提供的标准数据集上进行了测试。详细信息见表 1。训练集与测试集可视化如图 4 和图 5 所示。

3.2 评价指标

主观质量评估虽然能够提供有价值的视觉感知信息,但由于费时费力,且无法实时提供评估结果,在实际应用中并不实用。因此,自动化的客观质量评估方法显得尤为重要。在有损压缩的背景下,点对点的失真度量 D1 可作为几何失真评判标准,下面将介绍失真度量 D1 以及重建质量评估标准 PSNR (Peak Signal-to-Noise Ratio) 的计算过程。

表 1 实验数据集

数据集	点云序列	几何精度	帧数	几何信息	点数
训练集	Longdress	10	30	X,Y,Z	857 966
	RedandBlack	10	30	X,Y,Z	757 691
	Loot	10	30	X,Y,Z	805 285
测试集	Soldier	10	1	X,Y,Z	1 089 091
	Queen	10	1	X,Y,Z	1 000 993
	Basketball_Player	11	1	X,Y,Z	2 925 514
	Dancer	11	1	X,Y,Z	2 592 758
	Boxer	12	1	X,Y,Z	3 493 085
	Thaidancer	12	1	X,Y,Z	3 130 215
	Soldier	12	1	X,Y,Z	4 001 754

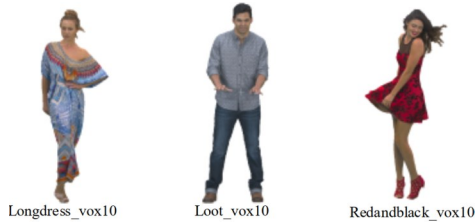


图 4 训练数据集



图 5 测试数据集

设 A 和 B 分别代表原始点云序列和重建点云,如图 6 所示, β_i 是 B 中的目标点, α_j 是 β_i 在点云 A 中找到的最近邻点,将重建点云 B 相对于点云 A 的失真度量表示为 $e_{B,A}$. A 与 B 之间的点到点失真 $e_{B,A}^{D1}(i)$ 的计算如下所示:

$$e_{B,A}^{D1}(i) = \frac{1}{N_B} \sum_{\forall \beta_i \in B} \left\| \mathbf{E}(\alpha_j, \beta_i) \right\|_2^2 \quad (13)$$

其中, $\mathbf{E}(\alpha_j, \beta_i)$ 表示点 α_j 和 β_i 之间的误差向量,误差向量的长度为点对点的欧式距离, $\|\cdot\|_2$ 表示 L_2 范数, N_B 为点云 B 中所包含的点的数量.

上述失真度 D1 对应的峰值信噪比 (PSNR) 计算公式如下:

$$\text{PSNR} = 10 \log_{10} \left(\frac{3p^2}{\max \{ e_{B,A}^{D1}(i), e_{A,B}^{D1}(i) \}} \right) \quad (14)$$

其中, p 由点云的体素化精度 d 决定,一般取为 2^d .

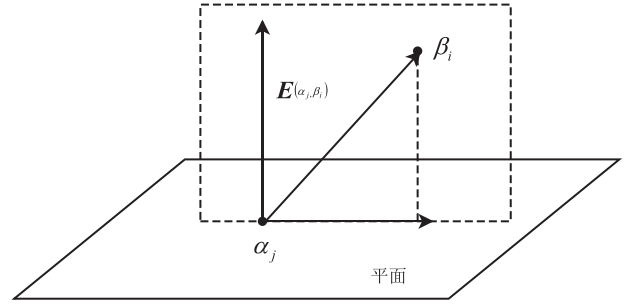


图 6 点对点的失真度量 D1 示意图

3.3 实现细节

本实验采用 MPEG-TMC13 v14(八叉树)^[28]来压缩点云,压缩后的点云和原始点云数据集分别按照本文 2.2 小节所提到的方法进行分块,基于点云密度分布特征和 GPU(Graphics Processing Unit)显存限制,分块大小设定为 $100 \times 100 \times 100$ 体素 (voxels) 的立方体单元. 点云下采样到 $m=2$ 个不同尺度;在上采样阶段, K 的值为 3. 为提升训练效率,剔除没有点云数据的空块,最终获得 8 530 个有效训练块.

本文的模型代码使用开源深度学习框架 PyTorch 和 Minkowski Engine 实现,在一块 NVIDIA GeForce RTX3080 GPU 上进行模型的训练与测试, CUDA 版本为 11.8. 模型在训练过程中,采用 Adam 优化器进行参数优化,其中初始学习率为 0.000 1,训练共进行了 5 000 次迭代,批次大小为 2.

3.4 对比实验

在对比实验中,选取了三种基线方法进行比较: MPEG-TMC13 (GPCC)^[28]、DGPP^[13] 和 PU-Dense^[12]. 其中 DGPP 利用 3D 卷积网络生成多尺度占用概率,通过逐步累加预测,实现从粗到细的点云几何细化. PU-Dense 通过渐进重新缩放和多尺度特征提取来分层重建上采样点云几何形状.

本实验采用均方误差点对点的 PSNR 作为衡量方法性能的客观指标,单位为分贝 (dB),精度为四位小数. 所有 PSNR 值均由 MPEG 标准误差工具计算^[29]得出. 为了验证所提出的点云质量增强方法的性能,本实验在相同数据集上进行了训练和测试,并将结果与几种代表性先进方法进行了定量和定性的对比分析,结果见表 2.

表 2 显示了本文方法与对比方法在测试集上的客观质量对比. 结果表明,本文所提的模型在总平均性能上优于其他现有的点云质量增强方法.

对于实体点云来说,实体点云通常具有较为明确的几何结构和较强的空间连贯性,本文方法在四个实体点云上的 PSNR 虽都有提升,但不如 DGPP 提升得高,这是由于 DGPP 利用 3D 卷积网络生成多尺度占用概率,通过逐步累加预测,实现从粗到细的点云几何细

化,从而提高 PSNR 的提升幅度. 本文方法虽然也在实体点云上有所改善,但由于未采用类似的逐步预测策略,无法像 DGPP 方法那样在细节恢复上提供更为精确的几何细化,因此提升幅度相对较小.

对于密集点云,本文方法比其他方法均获得了显著的 PSNR 提升. 具体来说,相比于 PU-Dense 平均 PSNR 提高了 1.724 6 dB. 这是因为 PU-Dense 只采用了坐标扩展类似方法来进行几何复原,在很大程度上忽略了压缩密集点云中的点位移问题. 相反,本文方法基

于点注意力机制,增强了多尺度点云中的关键几何特征,并有效融合了点云细节和上下文信息,从而实现了更精确的几何重建.

图7展示了本文方法与对比方法在主观可视化效果上的差异. 为了更清楚地对比不同方法的效果,点云被放大以展示细节. 放大后可以观察到,由于颈部线条丰富,压缩导致了大量信息的损失,使得重建变得较为困难. 然而,使用本文方法进行几何伪影去除后,点云在边界处更加平滑,保留了更多的局部信息,从而更加接近原始点云.

表 2 不同网络的点云质量增强 PSNR 结果对比

单位: dB

点云序列		方法			
		G-PCC	DGPP	PU-Dense	本文方法
实体点云	Soldier_vox10_	69.095 2	78.577 3	67.3884	72.631 5
	Queen_0 200_vox10	69.706 7	71.291 8	71.061 7	71.715 0
	Basketball Player_vox11	67.976 9	68.383 5	68.879 3	68.684 5
	Dancer_vox11	67.979 4	68.026 3	68.851 7	68.694 1
	Average	68.689 5	71.569 7	69.045 3	70.431 3
密集点云	Boxer_viewdep_vox12	63.202 0	61.584 0	60.985 1	65.042 9
	Thaidancer_viewdep_vox12	63.207 7	63.170 7	64.855 6	65.157 1
	Soldier_viewdep_vox12	63.205 2	63.336 5	62.229 7	63.854 0
	Average	63.205 0	62.697 1	62.690 1	64.684 7
Average		66.339 0	67.767 1	66.321 6	67.968 4

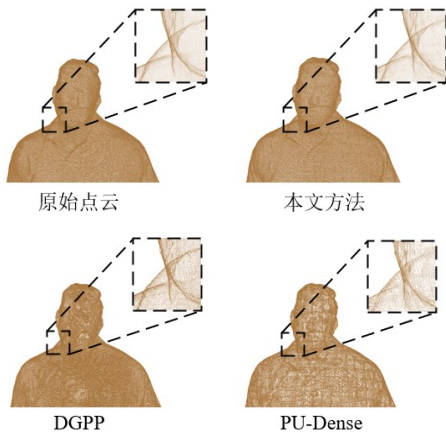


图7 不同点云质量增强网络的可视化结果

本文在表3收集比较方法和本文方法的模型大小和运行时间. 本文模型大小为 159 MB, 与 PU-Dense 模型的 154.6 MB 大小相近, 这是因为两者都使用了离散卷积来进行特征提取, 本文还增加了对多尺度特征的融合模块, 提高了重建点云的质量, 故增大了模型大小.

表3列出了不同方法的每个点云的平均运行时间. 这些方法运行在不同的平台上: G-PCC 是运行在 CPU (Central Processing Unit) 上的基于规则的算法, 而其他基于学习的方法主要运行在 GPU 上.

G-PCC 运行最快 (9.29 s), 因为它是传统方法, 不涉及深度学习. DGPP 运行时间高达 149.49 s. 这是因为

DGPP 使用密集的 3D 卷积来计算感受野内的所有体素, 而不管它们的占用状态, 这最终导致计算效率低下. 相比之下, 本文方法和 PU-Dense 采用稀疏计算, 其仅对紧密接近的有效点进行操作, 从而有效地降低了计算复杂度. 其中 PU-Dense 使用二进制体素占用分类损失, 该损失函数提高了模型对高分辨率点云的处理能力, 推理阶段的时间复杂度相对较低.

尽管本文模型在时间和内存占用上相较于 G-PCC 增加了开销, 但这一增加是为了实现更高的点云质量, 这种性能折中对于许多实际应用, 尤其是需要高质量点云的场景, 是完全可以接受的.

表 3 G-PCC、DGPP、PU-Dense 和本文模型大小、运行时间的比较

指标	G-PCC	PU-Dense	DGPP	本文方法
模型大小/MB	—	154.6	89.6	159.0
运行时间/s	9.29	24.09	149.49	36.78

3.5 消融实验

为验证本文多分支网络结构的有效性, 开展了消融实验, 结果如表4所示, 其中加粗数据为横向对比的最佳结果. 在测试集中, 三支网络 (即本文主体网络, 原始点云下采样两次, 对应图2所示的结构) 的 PSNR 值均高于单分支 (原始点云未经过下采样)、双分支 (原始点云经过一次下采样) 和四分支网络结构 (原始点云经过三次下采样). 这表明三支网络在捕捉多样化特

征并实现有效融合方面表现最佳. 相比之下, 单分支网络仅依赖原始点云特征, 未进行下采样, 因此缺乏足够的多尺度信息. 双分支网络由于特征源单一或特征融合不足, 导致对复杂几何结构的表达能力有限.

四分支网络尽管增加了特征的丰富性, 但同时引入了过多的冗余信息, 反而降低了信息的有效利用率. 因此, 三支网络能够在特征丰富性和冗余度之间取得平衡, 从而实现性能的提升.

表4 使用不同分支的点云质量增强 PSNR 结果对比

单位: dB

点云序列	方法			
	单分支	双分支	三支	四分支
soldier_vox10	72.974 3	73.065 0	72.631 5	71.418 7
Queen_0 200_vox10	72.640 7	72.494 2	71.715 0	71.403 0
Basketball Player_vox11	68.382 7	68.366 6	68.694 1	68.644 4
Dancer_vox11	68.417 5	68.350 7	68.684 5	68.599 3
Boxer_viewdep_vox12	63.895 0	64.163 3	65.042 9	63.876 3
Thaidancer_viewdep_vox12	64.039 5	64.381 2	65.157 1	64.041 1
Soldier_viewdep_vox12	63.677 5	64.043 3	63.854 0	64.296 2
Average	67.718 1	67.837 8	67.968 4	67.468 4

为了验证跨尺度特征融合模块的有效性, 在所有实验数据和参数保持不变的情况下, 实验对比了本文所提算法与去除跨尺度特征融合模块后的网络. 客观实验结果如表5所示, 其中加粗数据为横向对比的最佳结果. 当增加跨尺度特征融合模块时, 平均 PSNR 有所提升. 这一结果表明, 跨尺度特征融合模块能够有效增强特征融合能力, 从而提升网络的整体性能.

表5 在所提议算法上是否使用所提议特征融合模块的 PSNR 比较
单位: dB

点云序列	方法	
	本文方法 (无跨尺度特征融合 模块)	本文方法
soldier_vox10	70.549 0	72.631 5
Queen_0 200_vox10	72.434 4	71.715 0
Basketball Player_vox11	68.543 4	68.694 1
Dancer_vox11	68.527 4	68.684 5
Boxer_viewdep_vox12	64.391 0	65.042 9
Thaidancer_viewdep_vox12	64.526 5	65.157 1
Soldier_viewdep_vox12	64.060 6	63.854 0
Average	67.576 0	67.968 4

为了验证本文所提出的损失函数的有效性, 本文进行了损失函数消融实验, 比较了使用不同损失函数的模型表现. 本文的损失函数由两部分组成: $P2\hat{P}$ 和 $\hat{P}2P$. 假定通过本文方法重建的几何质量增强点云为 \hat{P} , 原始点云为 P , $\hat{P}2P$ 损失函数衡量增强后的点云 \hat{P} 与原始点云 P 之间的几何差异, 目的是尽量减少增强后的点云与原始点云的几何误差. $P2\hat{P}$ 损失函数目的在于度量 P 和 \hat{P} 的相似性, 从而引导重建点云 \hat{P} 更接近原始点云 P .

从表6(其中加粗数据为横向对比的最佳结果)中

可以看出, 单独使用 $P2\hat{P}$ 或 $\hat{P}2P$ 损失函数时, 不如两者结合使用时的提升效果显著. 这表明, 结合使用两种损失函数可以同时保持点云的点对点一致性和整体形态稳定性, 从而在点云质量增强任务中取得更好的效果.

表6 在所提议算法上使用不同损失函数的 PSNR 比较 单位: dB

点云序列	方法		
	Only $\hat{P}2P$	Only $P2\hat{P}$	$P2\hat{P} + \hat{P}2P$
soldier_vox10	71.418 7	71.866 0	72.631 5
Queen_0 200_vox10	70.565 2	70.162 7	71.715 0
Basketball Player_vox11	68.571 3	68.472 4	68.694 1
Dancer_vox11	68.599 3	68.499 0	68.684 5
Boxer_viewdep_vox12	63.876 3	63.638 2	65.042 9
Thaidancer_viewdep_vox12	64.041 1	63.825 3	65.157 1
Soldier_viewdep_vox12	63.884 9	64.322 4	63.854 0
Average	67.279 5	67.255 1	67.968 4

此外, 从图8中的主观实验结果可以观察到, 单独使用 $P2\hat{P}$ 或 $\hat{P}2P$ 损失函数时, 重建的点云出现了明显的块效应. 而当两者结合使用时, 块效应显著减小, 增强了重建点云的整体视觉质量. 这进一步证明了本文损失函数设计的有效性.

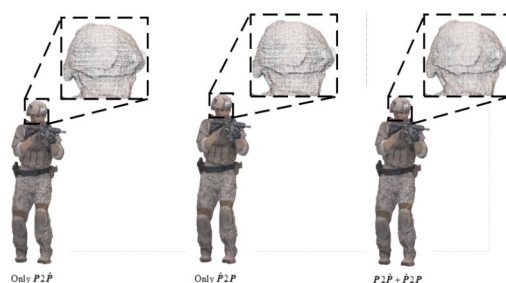


图8 本文方法使用不同损失函数的可视化结果

4 结论

G-PCC是目前静态点云的标准压缩方法,但研究发现使用G-PCC压缩后重新构建的点云会出现几何伪影.本文提出了一种基于多尺度特征提取的点云质量增强算法.该算法设计了多尺度输入模块对点云进行下采样得到不同尺度的点云数据.接着,多尺度的点云被并行输入到离散卷积网络中提取从局部到全局的多尺度特征信息.最后利用跨尺度注意力特征融合模块来对多尺度特征进行融合,以增强特征的完整性和准确性.主客观实验结果表明本文方法能进一步提高解压缩点云的质量.在未来的工作中,我们将重点提升网络在实体点云上的性能,并在此基础之上提高点云的颜色属性质量.

参考文献

- [1] WANG Q, KIM M K. Applications of 3D point cloud data in the construction industry: A fifteen-year review from 2004 to 2018[J]. *Advanced Engineering Informatics*, 2019, 39: 306-319.
- [2] GUO K W, XU F, YU T, et al. Real-time geometry, albedo and motion reconstruction using a single RGBD camera[J]. *ACM Transactions on Graphics*, 2017, 36(4): 1.
- [3] PERRA C MURGIA F, GIUSTO D. An analysis of 3D point cloud reconstruction from light field images[C]//2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA). Piscataway: IEEE, 2016: 1-6.
- [4] KRIVOKUCA M, CHOUP A, SAVIL P. 8i Voxelized surface light field (8iVSLF) dataset[EB/OL]. (2021-01-19)[2024-10-08]. <https://mpeg-pcc.org/index.php/pcc-content-database/8i-voxelized-surface-light-field-8ivslf-dataset/>.
- [5] V-PCC Codec Description. Document ISO/IEC JTC1/SC29/WG11 N19332[S/OL]. (2020-04-01) [2024-10-08]. <https://www.iso.org/committee/45316.html>.
- [6] G-PCC Codec Description. Document ISO/IEC JTC1/SC29/WG11 N19331[S/OL]. (2020-04-01) [2024-10-08]. <https://www.iso.org/committee/45316.html>.
- [7] ZHU W J, MA Z, XU Y L, et al. View-dependent dynamic point cloud compression[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021, 31(2): 765-781.
- [8] CHOY C, GWAK J, SAVARESE S. 4D spatio-temporal ConvNets: Minkowski convolutional neural networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2019: 3070-3079.
- [9] JIA W, LI L, LI Z, et al. Deep learning geometry compression artifacts removal for video-based point cloud compression[J]. *International Journal of Computer Vision*, 2021, 129(11): 2947-2964.
- [10] XING J R, YUAN H, CHEN C, et al. Wiener filter-based point cloud adaptive denoising for video-based point cloud compression[C]//Proceedings of the 1st International Workshop on Advances in Point Cloud Compression, Processing and Analysis. New York: ACM, 2022: 21-25.
- [11] AKHTAR A, GAO W, LI L, et al. Video-based point cloud compression artifact removal[J]. *IEEE Transactions on Multimedia*, 2021, 24: 2866-2876.
- [12] AKHTAR A, LI Z, AUWERA G V, et al. PU-dense: Sparse tensor-based point cloud geometry upsampling[J]. *IEEE Transactions on Image Processing*, 2022, 31: 4133-4148.
- [13] FAN X Q, LI G, LI D Q, et al. Deep geometry post-processing for decompressed point clouds[C]//2022 IEEE International Conference on Multimedia and Expo (ICME). Piscataway: IEEE, 2022: 1-6.
- [14] LEE M K, KIM Y H. Distance weighted refining segmentation method for visual quality improvement in V-PCC[C]//2023 IEEE International Conference on Visual Communications and Image Processing (VCIP). Piscataway: IEEE, 2023: 1-4.
- [15] SHENG X H, LI L, LIU D, et al. Attribute artifacts removal for geometry-based point cloud compression[J]. *IEEE Transactions on Image Processing*, 2022, 31: 3399-3413.
- [16] 王伟伟. 三维点云数据压缩与质量增强技术研究[D]. 济南: 山东大学, 2021.
WANG W W. Research on 3D Point Cloud Data Compression and Quality Enhancement Technology[D]. Jinan: Shandong University, 2021. (in Chinese)
- [17] DING D D, ZHANG J Z, WANG J Q, et al. CARNET: Compression artifact reduction for point cloud attribute[EB/OL]. (2022-09-17) [2025-05-12]. <https://arxiv.org/abs/2209.08276v1>.
- [18] LIU Q W, DING K, ZHANG C X, et al. A point cloud matching algorithm based on multiscale point pair features[C]//2023 IEEE International Conference on Real-time Computing and Robotics (RCAR). Piscataway: IEEE, 2023: 953-958.
- [19] ZHANG R J, XUE Y Q, WANG J, et al. MSFA-net: A multiscale feature aggregation network for semantic segmentation of historical building point clouds[J]. *Buildings*, 2024, 14(5): 1285.

- [20] YE T, LIU A, YAN X P, et al. An efficient 3D point cloud-based place recognition approach for underground tunnels using convolution and self-attention mechanism[J]. *Journal of Field Robotics*, 2025, 42(4): 1537-1549.
- [21] WU Y, HU X D, ZHANG Y, et al. SACF-net: Skip-attention based correspondence filtering network for point cloud registration[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023, 33(8): 3585-3595.
- [22] LI Z L, BAO J W, LIU Y, et al. Complement decoded point cloud with coordinate adjustment for video-based point cloud compression[J]. *Signal, Image and Video Processing*, 2024, 19(1): 48.
- [23] MOENNING C, DODGSON N A. *Fast Marching Farthest Point Sampling*[R]. Cambridge: University of Cambridge, Computer Laboratory, 2003.
- [24] NIE D, LAN R, WANG L, et al. Pyramid architecture for multi-scale processing in point cloud segmentation[C]// 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2022: 17263-17273.
- [25] QIC R, YIL, SUH, et al. PointNet++: Deep hierarchical feature learning on point sets in a metric space[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. California: NIPS, 2017: 5150-5114.
- [26] D'EONE, HARRISONB, MYERST, et al. 8i voxelized full bodies: A voxelized point cloud dataset, document WG11M40059/WG1M74006[EB/OL]. (2024-06-19) [2024-10-08]. <https://doi.org/10.48550/arXiv.2407.05677>.
- [27] XU Y, LU Y, WEN Z Y. OwlII dynamic human mesh sequence dataset[EB/OL]. (2021-01-05)[2024-10-08]. <https://mpeg-pcc.org/index.php/pcc-content-database/owlII-dynamic-human-textured-mesh-sequence-dataset/>.
- [28] CODINGM D G. G-PCC test model v14: ISO/IEC JTC1/SC29/WG7 output document N00094[S/OL]. (2021-01-05)[2024-10-08]. <https://www.iso.org/committee/45316.html>.
- [29] WATANABE S. Common test conditions for PCC: ISO/IEC JTC1/SC29 WG11 N19084[EB/OL]. (2020-01-01)[2024-10-08]. <https://datatracker.ietf.org/group/iso-iec-jtc1-sc29-wg11/about/>.

作者简介



钟 蕊 女, 1999 年出生于广西壮族自治区梧州市. 主要研究方向为机器学习、计算机视觉.
E-mail: 2112206163@gzhu.edu.cn



唐春明 男, 1972 年 9 月出生于湖南省怀化市. 教授, 博士生导师. 主要研究方向为密码学及其应用.
E-mail: ctang@gzhu.edu.cn



彭凌西 男, 1978 年出生. 教授, 博士生导师. 主要研究方向为人工智能技术及应用、网络安全.