

基于加权损失的点云占用图视频上采样

陈航, 李礼, 刘东, 李厚强

引用本文

陈航, 李礼, 刘东, 李厚强. 基于加权损失的点云占用图视频上采样[J]. 计算机科学, 2024, 51(1): 184-189.

CHEN Hang, LI Li, LIU Dong, LI Houqiang. [Weighted-loss-based Up-sampling for Point Cloud Occupancy Map Video](#) [J]. Computer Science, 2024, 51(1): 184-189.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于CodeBERT的设计模式语言模型](#)

CodeBERT-based Language Model for Design Patterns

计算机科学, 2023, 50(12): 75-81. <https://doi.org/10.11896/jsjcx.230100115>

[基于日志信息的不可重复构建原因分类](#)

Classification of Unreproducible Build Causes Based on Log Information

计算机科学, 2022, 49(12): 109-117. <https://doi.org/10.11896/jsjcx.220300227>

[端到端优化的图像压缩技术进展](#)

Advances in End-to-End Optimized Image Compression Technologies

计算机科学, 2021, 48(3): 1-8. <https://doi.org/10.11896/jsjcx.201100134>

[中国航空复杂网络的结构特征与应用分析](#)

Analysis of Characteristics and Applications of Chinese Aviation Complex Network Structure

计算机科学, 2019, 46(6A): 300-304.

[基于目标矩阵的CPU热点可持续冷却模型](#)

Sustainable Cooling Method of CPU Hot Spot Based on Target Matrix

计算机科学, 2019, 46(4): 329-333. <https://doi.org/10.11896/j.issn.1002-137X.2019.04.051>

基于加权损失的点云占用图视频上采样

陈航 李礼 刘东 李厚强

中国科学技术大学信息科学技术学院 合肥 230026

(ashlynchen@mail.ustc.edu.cn)

摘要 基于视频的点云压缩标准(Video-based Point Cloud Compression, V-PCC)中,3D点云会被分成数百个块并投影到2D平面中,形成记录点云纹理信息的纹理视频和记录点云空间信息的几何视频。同时,还需要生成一个占用图视频(Occupancy Map Video),以记录纹理视频和几何视频中每一个像素点是否对应重建点云中的某个点。因此,占用图视频质量与重建点云质量直接相关。为了节约编码比特数,占用图视频在编码端会先被下采样,然后在解码端通过简单的上采样恢复到原分辨率。文中的基本思路是引入深度学习来代替V-PCC中的简单上采样方法,使得上采样后的占用图视频质量更高,从而提高点云的重建质量。在网络训练阶段提出使用加权损失函数,使得在重建点云时能尽可能少地移除正常点并尽可能多地移除噪声点。实验结果证明,所提方法可以大幅提升V-PCC的主客观性能。

关键词: 点云压缩;基于视频的点云压缩标准;占用图视频;视频上采样;加权失真损失

中图分类号 TP391

Weighted-loss-based Up-sampling for Point Cloud Occupancy Map Video

CHEN Hang, LI Li, LIU Dong and LI Houqiang

School of Information and Technology, University of Science and Technology of China, Hefei 230026, China

Abstract In video-based point cloud compression(V-PCC), a 3D point cloud is divided into hundreds of patches and then mapped onto a 2D grid, generating a texture video that captures texture information and a geometry video that captures geometry information. Meanwhile, an occupancy map video is also generated to record whether each pixel in the former two videos corresponds to a point in the reconstructed point cloud. Therefore, the quality of the occupancy map video is directly linked to the quality of the reconstructed point cloud. To save bit cost, the occupancy map video is down-sampled at the encoder and up-sampled with a simplistic method at the decoder. This paper aims to use a deep learning-based up-sampling method to replace the simple up-sampling method in the original V-PCC to improve the quality of the up-sampled occupancy map videos as well as that of the reconstructed point cloud. A weighted distortion loss function in the network training process is introduced to remove the normal points as few as possible while removing the noisy points as many as possible when reconstructing a point cloud. Experimental results show that the proposed method can significantly improve the subjective and objective performances of the V-PCC.

Keywords Point cloud compression, Video-based point cloud compression standard, Occupancy map video, Video up-sampling, Weighted loss

1 引言

随着3D感知和采集技术逐渐成熟,以深度相机、激光雷达等设备为媒介,通过可视化数据对3D场景或物体进行更多细节的三维描述逐渐成为可能^[1],并在虚拟现实(Virtual Reality, VR)、增强现实(Augmented Reality, AR)、混合现实(Mixed Reality)等领域得到了广泛应用,给用户带来了更丰富的沉浸式与交互式体验^[2]。这一技术的实现依赖于对3D物体或场景进行扫描,从而获取具有丰富三维信息的海量点集,并进行加工处理。这一过程中采集到的点集数据被称作点云。

点云作为空间中一系列点的集合,包含丰富的空间信息,因此可以从三维角度对物体进行渲染。点云中的每一个独立的点均包含两种信息:空间信息(即点的3D坐标)和属性信息(包括颜色、反射度和法向量等)。根据物体和采集设备的状态,点云主要可被分为3类^[3]:1)静态点云,即被采集物体始终为静止的;2)动态点云,即采集时物体是运动的;3)动态获取点云,即采集设备是运动的。根据点与点之间距离的大小,点云又可被分为稀疏点云和紧致点云。稀疏点云为自动驾驶中的场景感知和目标检测带来了更智能化的应用^[4],而紧致点云则可以对目标和场景进行更细致的渲染。一帧动态

到稿日期:2023-06-20 返修日期:2023-10-27

基金项目:国家自然科学基金(62171429)

This work was supported by the National Natural Science Foundation of China(62171429).

通信作者:李礼(lil1@ustc.edu.cn)

紧致点云通常包含数十万甚至数百万空间中的点,其庞大的信息量不仅带来了更生动逼真的可视化效果,也为点云的传输和存储带来了巨大困难。一个30帧每秒、每帧100万个点的动态紧致点云,如果用30和24比特分别表示几何和属性信息,在未压缩的情况下传输占用的带宽可高达180 MB/S,这会给存储空间容量和网络传输带宽带来巨大的负担和挑战。因此,亟须研究动态紧致点云数据的高效压缩编码。

动态紧致点云编码的核心在于如何更有效地去除点云帧之间的时间相关性,去除时间相关性的核心技术是运动估计和运动补偿。根据运动估计和运动补偿所在的空间不同,动态紧致点云压缩方法可被简单分为3D空间法和2D空间法。顾名思义,3D空间法直接在3D空间进行运动估计和运动补偿^[5-6]。但是由于点云相邻帧点数不同,而且相邻帧的点不存在一一对应关系,这就使得3D运动估计和运动补偿成为一个难题,从而影响3D空间法的编码效率。近年来,一些实验结果表明,3D空间法更适合于编码稀疏点云,而非紧致点云^[7]。

不同于3D空间法,2D空间法的基本思路是把动态点云映射成为2D视频,使用成熟高效的2D视频编码框架,例如高性能视频编码(H.265/High Efficiency Video Coding, HEVC)^[8]等,来去除点云帧之间的时间相关性。许多2D空间法在多家研究机构响应动态图像专家组(Moving Picture Experts Group, MPEG)发布的动态紧致点云高效压缩提案征集的过程中被提出^[9-10]。其中,一种基于块的投影方法取得了最好的性能,并成为了MPEG基于视频的点云编码标准(V-PCC)^[3]的基础。V-PCC是目前为止压缩动态紧致点云最高效的方法。

在V-PCC中,输入的点云会被分解成大量的块,分解后的块会被分别投影到2D的网格中,形成记录空间信息的几何视频和记录属性信息的属性视频。由于几何视频并非每一个点都对应3D空间的一个点,还需要生成一个占用图视频来指示点云中的点与投影后形成的几何视频中像素的实际对应关系。在理想情况下,占用图视频应当与属性视频和几何视频保持相同的分辨率被编码,但这会造成较高的编码比特数消耗。为了达到比特数和性能的平衡,在V-PCC中,占用图视频通常在编码端压缩前被下采样为半分辨率或1/4分辨率,在解码端会对下采样后的占用图视频通过简单的上采样使其恢复到全分辨率,为点云的重建提供参考。

在占用图视频的上采样过程中,一些噪声点会被引入视频的边缘,在点云重建过程中被重建为3D点云中的噪声点,从而造成重建点云的质量严重下降。Jia^[11]首次提出使用基于深度学习的上采样方法来解决此问题,但是该方法没有考虑丢掉真实点和保留噪声点对点云质量的不同影响。此外,该方法需要额外引入几何视频作为输入,这会带来一定的复杂度提升。同样,为了提高重建点云质量,文献^[12]提出了基于深度学习的方法,对重建后低质量的点云进行质量修复,提高几何质量,但该方法并未考虑在压缩过程中直接去除压缩造成的影响,方法复杂度也较高。

本文的主要贡献如下:1)提出了加权失真损失函数,显式考虑丢掉真实点对重建点云质量的不良影响,使得在重建

点云时能尽可能少地移除正常点并尽可能多地移除噪声点,在保真的前提下尽可能提升点云质量;2)构建了一个基于残差块(Residual Block, Resblock)^[13]的多层卷积神经网络,在仅使用占用图视频作为网络输入的情况下提升点云重建质量;3)实验结果表明,本文方法可以为V-PCC带来显著的主观性能提升。

2 V-PCC 编码以及占用图视频上采样介绍

2.1 V-PCC 编码流程简介

V-PCC的核心思想在于首先将3D空间中的点云投影成一系列视频序列,如图1所示,包括几何视频、属性视频以及占用图视频。再通过现有成熟高效的视频编码技术对其进行压缩编码。由于视频编码的过程主要是调用现有视频编码标准,因此以下主要简要介绍V-PCC中生成视频序列的过程。

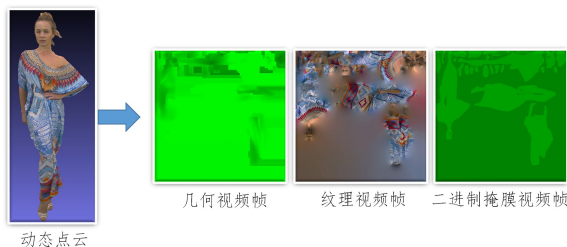


图1 一个典型点云序列投影产生的视频帧

Fig. 1 Generated video frames from a typical dynamic point cloud

V-PCC生成视频序列的过程可以大致分为3步。1)聚类:V-PCC通过计算每一个点的法向量,将每一个点与其邻近的法向量相近的点聚成一块,从而将3D的点云模型分割成几十到几百个大小不一的3D块。少数孤立点在此过程中无法形成3D块,在点云有损编码过程中被直接丢弃。2)投影:每一个聚类形成的3D块会被投影到2D平面中,分别形成记录点的空间信息的几何视频和记录点的属性信息的纹理视频。为了避免投影过程中存在多个点被投影到一个像素中的情况,每一个块会被投影到两层2D平面中,第一层记录多个点位于点云最外层表面点的信息,第二层记录位于点云最里层点的信息。3)填充:由于点云分割后的块投影到2D平面的形状的不规则性,平面中会存在大量未被投影的空白区域,这些空白区域会导致视频空间不连续,从而影响视频编码效率^[14]。因此,空白区域会根据周围像素进行填充,以提高视频的空间相关性以及后续的编码压缩效率。

除了几何视频和纹理视频之外,在V-PCC有损点云编码中还会生成一个占用图视频。占用图视频中的每一个像素用于指示纹理几何视频中的像素点是否实际对应3D点云中的一个点。具体来说,在占用图视频中,如果像素点值为1,则表示该点对应的几何视频中的像素点与点云中一个实际点相对应,反之,若值为0,则不存在对应的点。从以上描述可以看出,占用图视频直接关系到空间中的位置是否会存在相应的点,对点云的重建起着指导作用。因此,为了提高重建点云的质量,文献^[15]提出通过深度学习的方法挖掘占用图视频中的信息,对解码端的纹理视频帧做增强处理。但该方法仅考虑了纹理视频质量的修复,忽略了编解码中的采样过程对占用图视频质量带来的损伤,也未考虑占用图视频质量对

重建点云质量的直接影响。

2.2 占用图视频上采样

由于占用图视频中的像素点与几何视频中的像素点一一对应,理想情况下,占用图视频应当在与几何视频相同分辨率的情况下进行编码。但全分辨率的占用图视频编码会消耗大量的编码比特数。通常情况下,V-PCC会在编码端对其进行两倍或四倍下采样,再在解码过程中进行上采样,以恢复到全分辨率。

如图2所示,在编码端下采样过程中,全分辨率视频中每一 $P \times P$ (P 为2或者4)大小的块会被下采样为一个像素点。假设该块中有一个或多于一个像素点的像素值为1,则该块会被下采样为一个值为1的像素点;反之,若该块为零块,则会被下采样为值为0的像素点。在解码端上采样过程中,下采样后的占用图视频会被上采样回全分辨率,也就是说下采样视频中每一个像素点会被上采样回一个 $P \times P$ 的块。假设下采样视频中像素点的像素值为0,则会被上采样为一个像素值全为0的 $P \times P$ 块;反之,假设下采样视频中像素点像素值为1,则会被上采样为一个像素值全为1的 $P \times P$ 块。由图2中最左边和最右边的子图可以看出,这种较为粗糙的上下采样方式能保证不会有任何真实的点被去除,这也是V-PCC选择使用这种简单上下采样方式的原因。但是这种方式会引入大量噪声点,尤其在块的边缘,会导致重建点云主观质量的显著降低。在最极端的情况下,原始全分辨率视频中一个仅有一个像素且像素值为1的块,经历编码端四倍下

采样和解码端四倍上采样后会被采样为一个 4×4 的像素值全部为1的块,由此造成多达15个噪声点的引入。本文的主要出发点就是在上采样过程中尽可能少地引入噪声点。

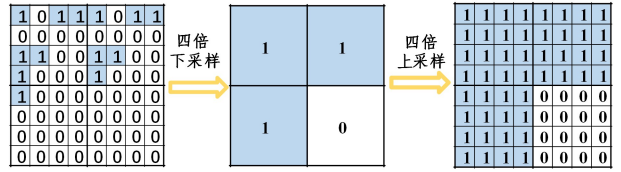


图2 V-PCC中的占用图视频下采样和上采样过程

Fig. 2 Down-sampling and up-sampling processes of an occupancy map video in VPCC

3 基于加权损失的点云占用图视频上采样方法

3.1 基本框架

图3给出了本文提出的基于加权损失的点云占用图视频上采样与V-PCC结合的框架图。本文方法不改变V-PCC对占用图视频进行的下采样操作,因此也不会改变占用图视频压缩所需要的码率。本文主要对使用V-PCC上采样之后的占用图视频使用本文提出的基于卷积神经网络的方法进行类似滤波的处理。需要注意的是,本文拟提出的网络的输入和输出均为原分辨率占用图视频。然后将处理后的原分辨率占用图视频和辅助信息以及几何视频信息结合起来重建几何信息,并进一步间接影响属性信息的重建以及点云的整体重建质量。

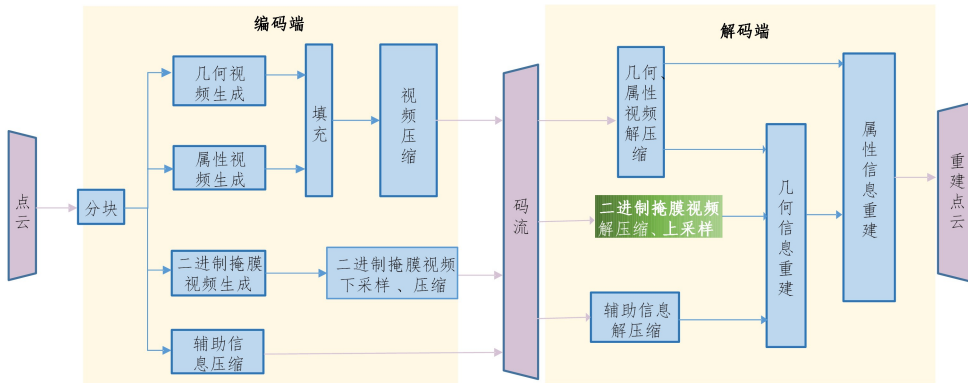


图3 基于加权损失的点云占用图视频上采样框架

Fig. 3 Weighted-loss-based up-sampling framework for occupancy map videos

3.2 网络结构

图4给出了本文提出的基于残差块的卷积神经网络框图。由于二进制上采样是一个较为简单的任务,本文仅使用上采样后的占用图视频作为输入,没有加入其他辅助信息。输入的上采样占用图视频帧首先经过一层卷积神经网络,然后经过 N 个残差块,最后再经过两层卷积层得到最终滤波后的上采样掩膜视频帧。本文使用的网络的所有卷积核的大小均为 3×3 ,每一层特征图的数量均为32。在实验中选择 N 的大小为10,以在性能和复杂度之间取得较好的均衡。由于批正则化层对像素级的恢复任务通常效果不好,本文使用的残差块也去掉了批正则化层,主要由两层卷积层和一层非线性层构成。本文提出的网络结构主要借鉴了增强深度超分辨率网络(Enhanced Deep Residual Networks,EDSR)^[16]中的结构。

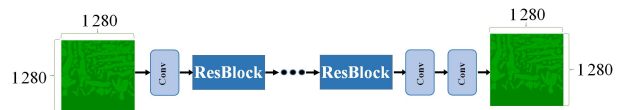


图4 本文提出的基于加权损失的卷积神经网络框图

Fig. 4 Weighted-loss-based convolutional neural network

3.3 加权失真损失函数

占用图视频编码端下采样,然后在解码端上采样到原分辨率,相比直接使用原分辨率可以节省大量码率,但同时因为少传了大量信息必然会引入两种失真:一种是丢掉原本存在的正常点;另一种是引入不该存在的噪声点。V-PCC现有编码框架从保真的角度出发处于其中一个极端,即完全不丢掉原本存在的正常点,引入大量噪声点。需要指出的是,V-PCC现有方法有一个优点:如前文所述,在V-PCC投影生成视频

的过程中会丢掉少量点,而 V-PCC 中引入的噪声点存在刚好与投影过程丢掉点距离较近的可能,从而提升重建点云的客观质量和保真程度。Jia 等^[11]提出的方法把占用图视频上采样问题看成一个分割问题,旨在为每一个像素找到一个合适的二分类,在损失函数中把丢掉正常点和引入噪声点看成是同等重要的任务,违背了动态紧致点云压缩尽可能保真的原则,从保真的角度会带来较多的损失。

不同于 Jia 等^[11]使用的二分类损失函数,本文认为占用图视频恢复任务是一个典型的像素重建任务,故而首先提出均方误差 (Mean Square Error, MSE) 作为损失函数。

$$L_m = \frac{1}{N} \sum_{i=0}^N (y_i - x_i)^2 \quad (1)$$

其中, x_i 为第 i 个预测像素值, y_i 为对应的标签, N 为总的像素个数。同 Jia 等^[11]使用的二分类损失函数一样,简单的均方误差也把丢掉噪声点和引入噪声点看成是同等重要的任务,从保真的角度会引入较大损失。本文拟在尽可能保真的前提下,去除大量的噪声点,因而拟对标签为 1 预测为 0 的情况施加较大的惩罚,而对标签为 0 预测为 1 的情况施加较小的惩罚,修改后的基于加权失真的损失函数如式(2)所示:

$$L_w = \frac{1}{N} \sum_{i=0}^N [\omega (y_i - x_i)^2 \text{sgn}(y_i) + (1-\omega)(y_i - x_i)^2 \text{sgn}(1-y_i)] \quad (2)$$

其中,当输入为 1 时 $\text{sgn}()$ 函数的值为 1,当输入为 0 时其值为 0,以区分当前的错误是把 1 变成 0 还是把 0 变成 1; ω 和 $1-\omega$ 分别表示两类错误的权重。如果 ω 大于 0.5,则会对标签为 1 预测为 0 的情况施加更大的惩罚,从而达到尽可能保真的目的,例如当 ω 取 0.8 时,不丢失真实点的权重是去除噪声点的权重的 4 倍。需要注意的是,如果 ω 取 0.5,本文所提的基于加权失真的损失函数就退化为 MSE。

4 实验结果

4.1 数据集与训练

目前面向静态点云编码的数据集^[17]较多,但面向动态点云编码的数据集较为缺乏,主要有 8i 公司利用传感器以 30 帧每秒采集目标的几何的颜色信息形成数据集^[18]。本实验选用 8i 提供的 Loot, Longdress, Redandblack 以及 Soldier 一共 4 个动态点云作为数据集,每个序列包含 300 帧。其中 Soldier 序列的前 256 帧作为训练集,后 44 帧作为验证集,剩余 3 个序列的前 32 帧作为测试集,训练集、验证集和测试集之间没有重复的动态紧致点云。

在大部分码率范围,V-PCC 使用 4 倍下采样可以带来最好的码率和点云重建质量之间的平衡,因此本文主要面向 4 倍上采样进行优化,本文方法也可以直接扩展到 2 倍下采样。在具体产生训练数据时,本文使用 V-PCC 参考软件 VPCC-8.0^[19]对 Soldier 序列的全部 300 帧进行编码,通过调整占用图视频的下采样倍数产生训练数据。对于 Soldier 序列,不进行下采样的占用图视频分辨率为 1280×1280 ,4 倍下采样之后的占用图视频分辨率为 320×320 ,进行简单上采样之后的视频分辨率也为 1280×1280 。然后将每一帧的亮度分量裁剪为 64×64 大小的块,共获得 102400 对 64×64

大小的训练块和 17600 对 64×64 大小的验证块。在训练过程中,采用自适应矩估计算法 (Adaptive Moment Estimation, Adam)^[20]作为梯度下降优化器,并使用其默认参数,网络的批大小被设置为 128,训练步数为 800,学习率为 0.0001。

4.2 实验设置

为了验证本文方法的性能,本文将提出的算法集成在 V-PCC 参考软件 VPCC-8.0 中进行测试,以便和 V-PCC 中简单的点云上采样方法进行对比。本实验测试了使用 4 倍下采样的从低码率点 r1 到高码率点 r4 的 4 个码率点。测试序列为 Loot, Longdress 和 Redandblack 这 3 个序列的前 32 帧,本文认为前 32 帧已经能代表整个序列的性质。本文分别使用点到点的误差 (Point-to-point error, D1) 以及点到平面的误差 (Point-to-plane error, D2) 来衡量点云几何信息重建质量^[21]; 本文使用亮度和色度分量的均方误差来衡量点云颜色信息重建质量。由于本文方法和 V-PCC 中简单的点云上采样方法的码率不完全一致,文本使用 BD-rate^[22]来衡量本文算法的性能。在没有特殊说明的情况下,本文的主要结果均基于权重 ω 为 0.8。

4.3 实验结果和分析

表 1 列出了本文提出的基于加权损失的上采样方法和 V-PCC 简单上采样方法的客观性能对比。表 1 中 Geom. BD-TotalRate 和 Attr. BD-TotalRate 分别展示了几何信息以及纹理信息上的性能表现, Luma 表示亮度信息, Cb 和 Cr 表示色度信息。表中的负值表示性能提升,正值表示性能损失。从表中可以看出,本文提出的方法在 D1 性能指标上损失 4.2% 的情况下在 D2 性能指标上有 13.0% 的提升。

表 1 本文提出的基于加权损失的上采样和 V-PCC 简单上采样的客观性能对比

Table 1 Objective performance comparison of the weighted-loss-based up-sampling and V-PCC simple up-sampling (%)

| 测试序列 | Geom. BD-TotalRate | | Attr. BD-TotalRate | | |
|-------------|--------------------|-------|--------------------|------|------|
| | D1 | D2 | Luma | Cb | Cr |
| Loot | 1.3 | -12.2 | 0.2 | 0.6 | 2.0 |
| Longdress | 6.6 | -13.1 | 0.1 | 0.0 | 0.3 |
| Redandblack | 4.6 | -13.7 | 0.4 | -0.5 | -0.1 |
| 平均 | 4.2 | -13.0 | 0.2 | 0.0 | 0.7 |

此外,本文方法不会对其他信息的编码造成较大的影响。特别地,对于 Loot 序列,本文方法只会引入 1.3% 的 D1 指标上的性能损失,但在 D2 指标上能带来 12.2% 的性能提升。总的来说,本文方法对 D2 指标性能的提升远大于 D1 指标性能的损失。

本文认为本文方法会导致部分 D1 性能损失的原因主要有两方面。一是点数的减少,相比 V-PCC 会引入更多的噪声点,本文方法会消除 V-PCC 中的噪声点,导致总输出点数的减少,这会影响到依赖于点与点之间的距离计算的 D1 性能指标,导致在 D1 指标上出现损失。其次,由于在投影过程中存在点的丢失,因此本文使用的标签并不对应原始点云所有的点,也存在添加一些标签以为的噪声点能提升 D1 指标的可能,这一点是本文方法的固有缺点,未来计划结合几何信息使用 3D 标签来解决该问题。

表 2 列出了本文方法与 Jia 等^[11]提出的上采样方法的对比结果。由表 2 可见,相比 Jia 等^[11]的方法在 D1 和 D2 上的性能两极化,即在 D1 上损失严重但在 D2 上增益明显,甚至在 Longdress 序列上 D1 的损失超过了 D2 的增益,本文方法既平衡了 D1 的损失,又提升了 D2 的性能。

表 2 本文提出的基于加权损失的上采样和 Jia 等的上采样的客观性能对比

Table 2 Objective performance comparison of the weighted-loss-based up-sampling and V-PCC simple up-sampling (%)

| 测试序列 | (本文方法) | | (Jia 等的方法) | |
|-------------|--------------------|-------|--------------------|-------|
| | Geom. BD-TotalRate | | Geom. BD-TotalRate | |
| | D1 | D2 | D1 | D2 |
| Loot | 1.3 | -12.2 | 6.3 | -16.9 |
| Longdress | 6.6 | -13.1 | 22.0 | -18.5 |
| Redandblack | 4.6 | -13.7 | 16.2 | -20.4 |
| 平均 | 4.2 | -13.0 | 14.8 | -18.6 |

除了客观性能外,本文方法的另一个优势是更好的主观性能。图 5 给出了本文提出的基于加权损失的上采样和 V-PCC 简单上采样在相同码率下的主观质量对比。图 5 中,从左至右分别是 V-PCC 简单上采样的重建点云、本文方法的重建点云以及未压缩的原始点云。图 5 给出了 Longdress 序列测试码率为 r_2 、Redandblack 序列和 Loot 序列测试码率为 r_1 的第 0 帧的重建结果。从图中可以看出,本文方法由于删除部分噪声点可以带来明显的主观质量提升,从图中也可以看出本文方法会删掉部分真实点。总的来说,本文方法可以提升重建点云的主观质量。

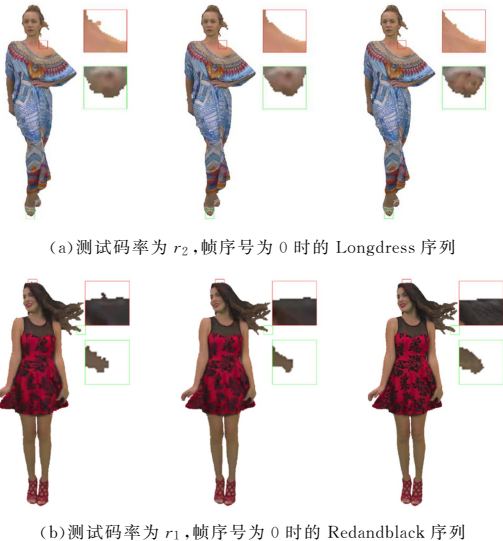


图 5 本文提出的基于加权损失的上采样和 V-PCC 简单上采样的主观性能对比

Fig. 5 Subjective performance comparison of the weighted-loss-based up-sampling and V-PCC simple up-sampling

4.4 消融实验

4.4.1 不同加权权重对编码性能的影响

为了更好地探究加权权重对编码性能的影响,表 3 列出了不同权重对 D1、D2 以及 D1+D2 的影响。表 3 列出了 3 个测试序列 Loot、Longdress 和 Redandblack 的平均结果。从结

果可以看出,权重越小,D1 指标下的性能损失越大,D2 指标下的性能提升越明显。如果认为 D1 和 D2 的客观指标同等重要,则可以看到在权重为 0.8 时性能达到了最优 8.8%,这也是在前面所述的所有实验中选择 w 为 0.8 的原因。出现 D1 指标下性能递减和 D2 指标下性能递增的主要原因是 D1 指标主要由去掉的真实点数量决定,真实点丢失越多,性能越差。而 D2 主要由去掉的噪声点数量决定,噪声点去掉越多,性能越好。

表 3 不同加权权重对编码性能的影响

Table 3 Influences of the weights on compression performance (%)

| 权值 | BD-TotalRate (D1) | BD-TotalRate (D2) | BD-TotalRate(D1)+BD-TotalRate(D2) |
|-----|-------------------|-------------------|-----------------------------------|
| 0.6 | 9.6 | -15.4 | -5.8 |
| 0.7 | 6.2 | -14.1 | -7.9 |
| 0.8 | 4.2 | -13.0 | -8.8 |
| 0.9 | 2.0 | -10.6 | -8.6 |

4.4.2 不同加权权重对重建点云点数的影响

为了更好地解释不同加权权重对编码性能的影响,表 4 以 Loot 序列为例,列出了不同加权权重对重建点云点数的影响。由表 4 可知,重建点云的点数随着码率的增加逐渐增加,这主要是由于随着码率的增加,几何视频编码使用的量化参数越来越小,因此不同点被量化为同一个点的概率越来越小,从而导致重建点云点数的增加。此外,随着权值的增加,重建点云的点数单调递增,这主要受到两个方面的影响:一方面,随着权重的增加,保留的真实点越来越多;另一方面,随着权重的增加,去掉的噪声点越来越少。这两方面均导致点数的增加,这也反映了本文提出的基于加权损失的上采样方法的确达到了预期的效果。

表 4 不同加权权重对重建点云点数的影响

Table 4 Influences of the weights on the number of points in reconstructed point clouds

| 权值 | 重建点云点数(Loot) | | | |
|-----|--------------|--------|--------|--------|
| | r_1 | r_2 | r_3 | r_4 |
| 0.6 | 640425 | 710367 | 728993 | 757329 |
| 0.7 | 644892 | 714920 | 733367 | 761581 |
| 0.8 | 648653 | 718586 | 736776 | 764818 |
| 0.9 | 654286 | 724043 | 741872 | 769700 |

4.4.3 不同残差块数量对上采样后的占用图视频质量的影响

为了比较不同残差块数量对性能的影响,表 5 列出了 3 个测试序列在不同残差块数量下占用图视频前 32 帧的平均峰值信噪比(Peak Signal to Noise Ratio, PSNR)。

表 5 不同残差块数量对上采样后的占用图视频质量的影响

Table 5 Influences of the number of Resblocks on the qualities of up-sampled occupancy map videos

| 序列 | 前三十二帧的平均峰值信噪比 (dB) | | |
|-------------|--------------------|---------|---------|
| | 前 32 帧的平均峰值信噪比 | | |
| | N=5 | N=10 | N=15 |
| Longdress | 20.1096 | 20.1525 | 20.1733 |
| Loot | 20.8027 | 20.9579 | 20.9736 |
| Redandblack | 19.7329 | 19.8784 | 19.8954 |

从表 5 可以看出,当残差块的数量增加时,占用图视频的 PSNR 也随之提高,其中从 5 层到 10 层提高较为明显,然后

趋于饱和。因此,基于网络复杂度和性能的综合考量,本实验中 Resblock 的数量设为 10。此外,需要注意的是,表 5 列出的 PSNR 看似整体偏低,主要原因是本文处理的掩膜视频为二进制视频,像素值非 0 即 1,因此一个像素值的错误会对整体的 PSNR 造成较大的影响。

结束语 本文提出了一种基于加权损失的可学习的点云占用图视频上采样方法,构建了一个基于残差块的多层卷积神经网络,在仅使用占用图视频作为网络输入的情况下提升点云重建质量,进一步提出了加权损失函数在尽可能保留真实点的情况下删除大量的噪声点。本文将所提算法集成在基于视频的点云视频编码标准(V-PCC)参考软件中,实验结果表明本文方法可明显改善 V-PCC 重建点云的主客观质量。

参 考 文 献

- [1] SUI L C,ZHANG B Y. Principle and Trend of Airborne Laser Scanning Remote Sensing [J]. Journal of Zhengzhou Institute of Surveying and Mapping, 2006, 23(2): 127-129.
- [2] BRUDER G,STEINICKE F,NÜCHTER A. Poster: Immersive point cloud virtual environments[C]//IEEE Symposium on 3D User Interfaces (3DUI). Minneapolis, MN, USA. IEEE, 2014: 161-162.
- [3] SCHWARZ S,PRED A M,BARONCINI V,et al. Emerging MPEG standards for point cloud compression[J]. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 2019, 9(1): 133-148.
- [4] YANG J H,ZHAO X,FANG Y Y,et al. 3D Laser Point Cloud Based Vehicle Target Recognition Algorithm[C]//International Conference on Mechanical and Electronics Engineering(IC-MEE). Xi'an, China, 2022: 286-291.
- [5] DE QUEIROZ R L,CHOU P A. Motion-compensated compression of dynamic voxelized point clouds[J]. IEEE Transactions on Image Processing, 2017, 26(8): 3886-3895.
- [6] LI L,LI Z,LIU S,et al. Motion estimation and coding structure for inter-prediction of LiDAR point cloud geometry[J]. IEEE Transactions on Multimedia, 2022, 24: 4504-4513.
- [7] PERRY S,CONG H P,DA SILVA CRUZ L A,et al. Quality evaluation of static point clouds encoded using MPEG codecs [C]//IEEE International Conference on Image Processing. Abu Dhabi, United Arab Emirates. IEEE, 2020: 3428-3432.
- [8] SULLIVAN G J,OHM J R,HAN W J,et al. Overview of the high efficiency video coding(HEVC) standard[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2012, 22(12): 1649-1668.
- [9] SCHWARZ S,SHEIKHIPOUR N,FAKOUR SEVOM V,et al. Video coding of dynamic 3D point cloud data [J]. APSIPA Transactions on Signal and Information Processing, 2019, 8: e31.
- [10] HE L Y,ZHU W J,XU Y L. Best-effort projection based attribute compression for 3D point cloud [C] // 23rd Asia-Pacific Conference on Communications(APCC). Perth, WA, Australia. IEEE, 2017: 1-6.
- [11] JIA W,LI L,AKHTAR A,et al. Convolutional neural network-based occupancy map accuracy improvement for video-based point cloud compression[J]. IEEE Transactions on Multimedia, 2022, 24: 2352-2365.
- [12] AKHTAR A,GAO W,LI L,et al. Video-Based Point Cloud Compression Artifact Removal[J]. IEEE Transactions on Multimedia, 2022, 24: 2866-2876.
- [13] HE K M,ZHANG X Y,REN S Q,et al. Deep residual learning for image recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA, IEEE, 2016: 770-778.
- [14] LI L,LI Z,LIU S,et al. Efficient projected frame padding for video-based point cloud compression[J]. IEEE Transactions on Multimedia, 2021, 23: 2806-2819.
- [15] CHEN P L,WANG S Q,LI Z. Occupancy Map Guided Attributes Deblocking for Video-based Point Cloud Compression [C]//Data Compression Conference (DCC). Snowbird, UT, USA, 2023: 332-332.
- [16] LIM B,SON S,KIM H,et al. Enhanced deep residual networks for single image super-resolution [C] // IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu, HI, USA. IEEE, 2017: 1132-1140.
- [17] CHANG A X,FUNKHOUSER T,GUIBAS L,et al. Shapenet: An information-rich 3d model repository [J]. arXiv: 1512. 03012, 2015.
- [18] D'EON E, HARRISON B, MYERS T, et al. 8i voxelized full bodies-a voxelized point cloud dataset. ISO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG) input document WG11M40059/WG11M74006[S]. 2017, 7: 8.
- [19] Point cloud compression category 2 reference software, TMC2-4. 0 [OL]. <http://mpegx.intevry.fr/software/MPEG/PCC/TM/mpeg-pcc-tmc2.git>, accessed: 2021.
- [20] KINGMA D P,BA J. Adam: A method for stochastic optimization[J]. arXiv: 1412. 6980, 2014.
- [21] SCHWARZ, MARTIN-COCHER G, FLYNN D, et al. Budagavi, "Common Test Conditions for Point Cloud Compression," Document ISO/IEC JTC1/SC29/WG11 w17766[S]. Ljubljana, Slovenia, 2018.
- [22] BJØNTEGAARD G. Calculation of Average PSNR Differences Between RD Curves", ITU-T SG16/Q6, 13th VCEG Meeting [S]. Austin, Texas, USA, April 2001, Doc. VCEG-M33.



CHEN Hang, born in 2000, postgraduate, is a student member of CCF(No. P3178G). Her main research interests include point cloud compression and video sampling.



LI Li, born in 1990, Ph.D. His main research interests include image/video coding and processing.