arXiv:2303.04027v2 [cs.MM] 9 Mar 2023

BIRD-PCC：基于双向距离图像的深度激光雷达点云压缩

Chia-Sheng Li u[1], *Jia-Fong Yeh*[1], *Hao Hsu*[1],*Hung-Ting Su*[1],*Ming-Sui Lee*[1],*Winston H. Hsu*[1] ,[2]

[1] 国立台湾大学 [2] Mobile Drive Technology

摘要

激光雷达传感器采集的大量数据带来了激光雷达点云压缩（PCC）问题。以往的激光雷达点云压缩研究采用距离图像表示，并遵循预测编码范式，构建了编码框架的基本原型。然而，由于忽略了距离图像中的无效像素，并且未考虑时间步中的未来帧，其预测方法得出的结果并不准确。此外，人工设计的残差编码方法无法充分利用空间冗余。为解决这些问题，我们提出了 BIRD-PCC 编码框架。该框架的预测模块能够感知距离图像中无效像素的坐标，并采用双向方案。此外，我们引入了深度学习残差编码模块，以进一步利用残差帧中的空间冗余。在 SemanticKITTI 和 KITTI-360 数据集上进行的实验表明，在大多数比特率条件下，BIRD-PCC 的表现优于其他方法，并且能够很好地泛化至未见环境。

索引词— 压缩、深度学习、激光雷达、点云、距离图像。

1. 引言

近年来，激光雷达点云压缩（PCC）在多媒体压缩领域日益受到关注。激光雷达点云能够提供三维信息，因此被用于自动驾驶场景中感知周围环境。然而，存储或传输激光雷达数据的成本很高。因此，本文深入探讨激光雷达点云压缩这一具有挑战性的问题。

激光雷达点云压缩主要有两种主流方法：一种基于树，另一种基于距离图像。基于树的方法内存效率高，但存在分辨率退化问题，即重建点云中的点数减少，正如 [1]中所观察到的那样。基于距离图像的方法利用激光雷达的工作机制，通过球面投影将点云表示为距离图像。图 1 概述了近期遵循预测编码范式的基于距离图像的方法。Sun 等人 [2]提出了一种基于实例的聚类帧内编码方法。该方

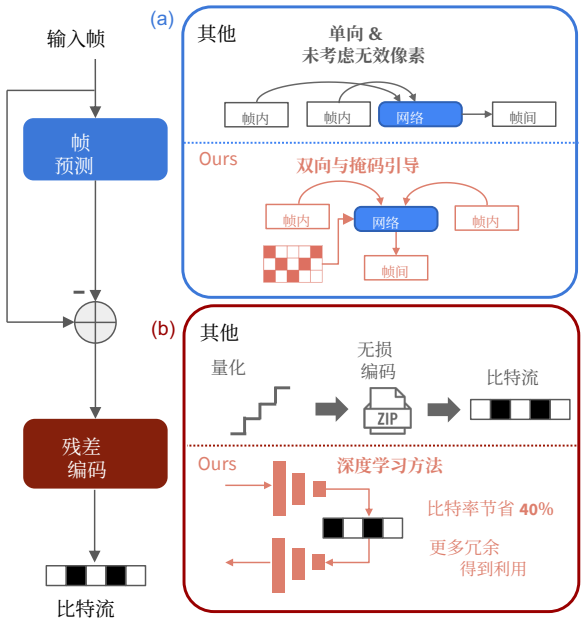

图1。我们的方法与其他方法的比较。(a) 先前的方法省略了距离图像中无效像素的坐标，导致预测不准确（如图5所示）。此外，该方法仅将前面的帧作为参考帧。(b) 先前的手工设计方法难以捕捉复杂的依赖关系，因此压缩率较低。

他们方法的出发点是，属于同一物体的区域之间存在空间依赖关系。他们的聚类算法将距离图像分割为若干不相交区域。随后，计算每个区域的平均深度值，并用其替换该区域内的所有像素值。最后，对真实值与平均深度值之间的残差值进行编码。Wang 等人沿用他们的工作 [2] 并提出了 R-PCC [1]。为了提高分割的运行速度，他们采用最远点采样，提出了一种基于区域的聚类方法。上述工作 [1, 2] 只考虑了单帧的空间冗余，却没有利用时间冗余。随后，Sun 等人 [3] 受到传统视频编码标准的启发，设计了一种同时考虑空间冗余和时间冗余的编码架构。给定一个帧序列，他们的方法首先将每一帧分类为帧内帧或帧间帧。然后，将[2] 中的方法用于帧内编码。至于帧间编

码，一个由 convLSTM 单元组成的神经网络以先前时间步的帧作为输入，并进行帧级预测。最后，将真实值与预测值之间的差异量化，并使用现成的无损编码器（如 bzip2、LZ4 或 Deflate）进行编码。

关于基于距离图像的方法，已有研究提供了很好的启发，但仍有若干方面可以改进。首先，距离图像的特性与彩色图像不同。因此，直接将彩色图像预测算法应用于距离图像并非良策。如何利用距离图像的属性来开发更优的压缩算法这一问题仍未得到充分探索。在本研究中，我们的预测算法会考虑距离图像中无效像素的坐标。其次，仅考虑前一时间步帧的单向预测并非最优。我们证明，双向预测能够实现更高的准确度。第三，残差帧中仍存在空间冗余。先进行量化再进行无损编码，会导致较差的率失真性能。作为替代方案，我们采用深度学习框架处理残差编码。通过利用更多冗余，我们的方法能够实现更好的率失真性能。总而言之，我们的贡献有三点：

- 我们提出了 BIRD-PCC，这是一种激光雷达点云压缩框架，利用距离图像的特性，并采用深度学习的残差编码方法。
- BIRD-PCC 通过考虑无效像素的坐标并采用双向方案，实现了准确的预测。此外，其残差编码模块更擅长探索残差帧内部的依赖关系。
- BIRD-PCC 在大多数比特率条件下均实现了更优的率失真性能，并且能够很好地泛化到未见过的更大数据集。在相同失真水平下，与基线的基于距离图像的方法相比，其比特率几乎降低了 20%。

2. 相关工作

基于树的激光雷达点云压缩。树形数据结构是一种节省内存的稀疏点云表示方法。Draco [4] 和 MPEG G-PCC [5] 分别采用 kd 树和八叉树数据结构，是广为人知的基于树的方法，但二者都未采用深度学习方法。近年来，涌现出多种基于深度学习的八叉树压缩技术。Huang 等人 [6] 提出使用多层 MLP 提取树节点特征，并根据祖先节点的特征预测节点符号的分布。为考虑时间依赖关系，Biswas 等人 [7] 提出使用前一个时间步的节点特征，使熵模型能够更准确地预测分布。为进一步利用树节点之间的依赖关系，Fu 等人 [8] 提出了一种熵模型，不仅考虑祖先节点，还考虑兄弟节点。此外，他们还提出了一种注意力机制，用于计算不同节点的权重。

近年来，基于树的方法得到了广泛研究，并被视为激光雷达点云压缩的解决方案。然而，无法保证解码点云中的点数与原始点云相同。这可能导致真实值与解码点云之间的分辨率不一致，而基于距离图像的方法则不存在这一问题。

3. 所提出的方法

3.1. 概述

我们的压缩流程如图 2 所示。该设置与先前的工作一致 [1, 2, 3], ，仅关注三维几何坐标，而非点属性。给定一系列激光雷达点云，每个点云都会按照文献 [1]第 III 节中的公式转换为距离图像表示$\boldsymbol{x} = \{x_1, x_2, \cdots, x_T\}$。随后，距离图像序列被划分为若干编码单元。对于每个编码单元，时间步中的首帧和末帧被视为帧内帧，中间 k 帧则被视为帧间帧。请注意， k 是一个超参数，本工作中取 $k = 1$ 。假设在时间步 t，编码单元记为 $\{x_{t-1}, x_t, x_{t+1}\}$。采用 R-PCC [1] 进行帧内编码。对于帧间编码，我们的预测模块以解码后的帧内帧作为参考帧，并生成预测 $\bar{x}$ （第 3.2 节）。残差帧 r_t 定义为真实值帧 x_t 与预测帧 $\bar{x}_t$之差。随后，残差帧通过残差编码模块（第 3.3 节）编码为比特流。

3.2. 帧间预测

一些研究 [12, 13, 14] 表明，带有跳跃连接的 U-Net 架构能够进行帧插值或外推。帧预测的一种简单方法是，将参考帧和预测帧分别直接作为 U-Net 的输入和输出。然而，这会导致输出的距离图像质量较低。为解决这一问题，深度学习视频压缩领域通常会利用光流等运动信息来实现更准确的预测 [15, 16]。但是，由于彩色图像与距离图像之间存在内在差异，这种策略并不适用于距离图像。

可以注意到，距离图像中的许多坐标都是无效像素，其像素值为零。这些无效像素是由于没有激光雷达激光信号返回造成的。在我们的预测设计中，我们利用了这一特性

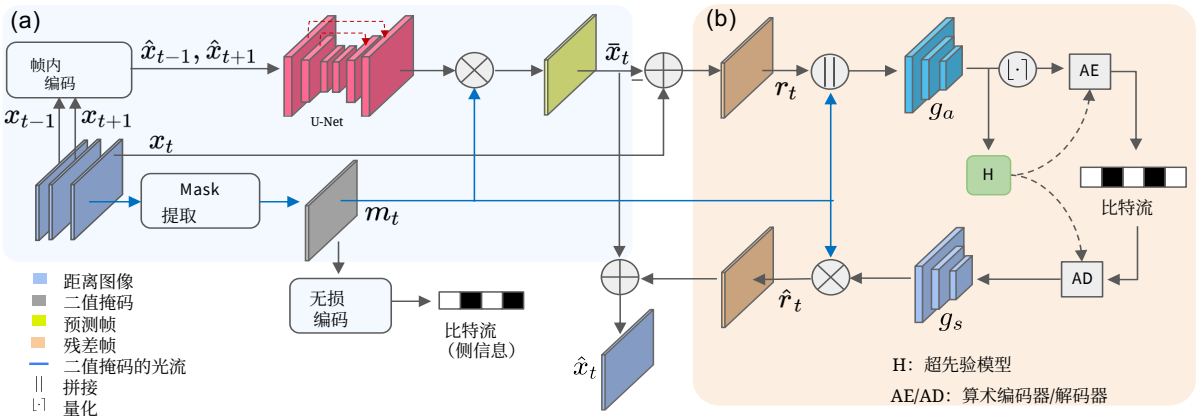

图2。我们的帧间压缩流程。(a) 预测网络以解码帧内帧 $\hat{x}_{t-1}$、$\hat{x}_{t+1}$ 和掩码 m_t 作为输入，并生成预测结果 $\bar{x}_t$。(b) 残差编码网络 g_a 和 g_s 由卷积、残差块 [9] 和注意力块 [10]堆叠而成。有关超先验模型的更多详细信息，请参阅 [11]。请注意，编码后的掩码和编码后的残差帧都应被视为需要存储或传输的比特流。

并证明这些无效像素的坐标是实现准确预测的关键信息。具体而言，我们按照以下公式从帧间x_t 中提取二值掩码 m_t：

$$m_t(i,j) = \begin{cases} 0 & \text{if } x_t(i,j) = 0 \\ 1 & \text{if } x_t(i,j) \neq 0 \end{cases}, \tag{1}$$

其中（i、 j）是图像坐标。该二值掩码作为滤波器应用于 U-Net 的最后一层，以细化预测结果。请注意，二值掩码被视为侧信息 [17]。因此，我们使用 bzip2 将其无损编码为比特流。对于网络训练，目标函数是真实值与预测帧之间的 L1 损失。

3.3. 残差帧编码

残差帧编码的任务是将残差帧编码为比特流。已有工作 [1, 2, 3]的方法首先量化残差帧，再采用 bzip2 等无损编码器进行残差编码。然而，前一阶段的预测可能不够准确，因此残差帧中存在空间冗余。以往的手工设计方法无法充分利用这些冗余，导致压缩率较低。

我们基于学习的方法受到近期学习式彩色图像压缩研究 [9, 10, 11, 18]的启发。我们采用变分自编码器（VAE）实现变换编码。我们证明，该方法同样适用于激光雷达残差信号。具体而言，在编码阶段，编码器网络 g_a 执行非线性变换，将输入从像素空间映射到潜在空间。为了辅助 g_a进行特征提取，我们将二值掩码 m_t 与输入残差帧 r_t拼接。随后，对所得潜在编码进行量化，并通过算术编码器进行无损编码为比特流 [19]。为估计量化潜在编码的分布，已有多种熵模型被提出 [10, 11, 18]。鉴于其简洁高效，我们在设计中选择了超先验模型 [11]。在解码阶段，解码器网络 g_s 将量化后的潜在编码进行逆变换，得到重建的残差帧 $\hat{r}_t$。二值掩码 m_t 被用作 g_s最后一层的滤波器，从而优化解码残差帧。拉格朗日训练目标由码率和失真（均方误差）构成，与 [11]相同。

4. 实验结果

4.1. 设置

数据集。预测网络和残差编码网络分别使用来自 SemanticKITTI[20] 数据集的 18.5k 帧和 34k 帧进行训练。与基线方法的评估则在规模更大的未见过的 KITTI-360 [21]数据集上进行。

评估指标。 我们采用每点比特数（bpp）作为比特率指标。对于失真指标，我们使用倒角距离，该指标也在 [6, 22] 中采用并定义。

基线方法。实验中选取了两种基于树的基线方法和一种基于距离图像的基线方法。Draco [4] 是 Google 发布的一种基于 kd 树的开源压缩算法。G-PCC [5] 是 MPEG 提出的一种基于八叉树的压缩标准。R-PCC [1] 是一种仅执行帧内编码的基于距离图像的方法。

4.2. 率失真曲线的定量结果

为展示 BIRD-PCC 的泛化能力，我们在未见过的 KITTI-360 数据集上，将其率失真性能与选定的基线方法进行比较。图4(a)展示了定量结果，表明我们的方法在低比特率下优于 R-PCC，在高比特率下则优于基于树的方法。

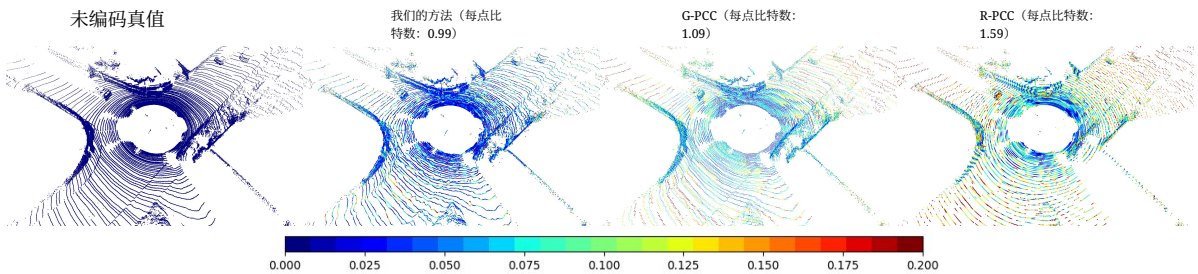

图 3。各种压缩方法的真实值与解码点云的可视化结果。从左到右依次为：未编码真值、我们的方法、G-PCC 和 R-PCC。误差色条显示在底部。建议在彩色显示并放大后查看。

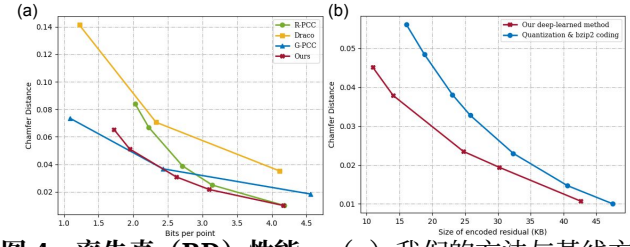

图 4。率失真（RD）性能。（a）我们的方法与基线方法在 KITTI-360 上的率失真曲线。（b）SemanticKITTI 上采用各种残差编码方法的率失真曲线。

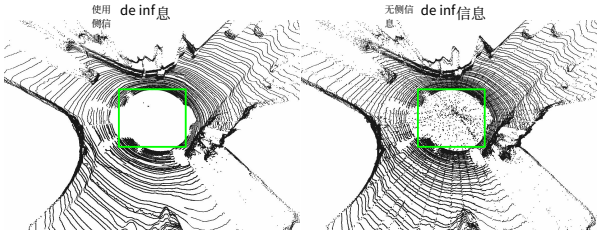

图 5。应用掩码（左）和未应用掩码（右）时预测帧的对比。图中展示了反投影至点云的预测帧。显然，在未应用掩码的情况下，由绿色区域界定的中心区域包含更多噪声点。

4.3. 定性结果

图3展示了我们的方法、R-PCC 和 G-PCC 的解码点云。我们的方法能够实现更低的失真，同时比特率仍低于其他方法。此外，我们的方法能够保持点的分布模式，这一点在中央的 O 形区域尤为明显，该区域的点密度理应较高。然而，基于八叉树的 G-PCC 因体素化而无法保持这种模式。

表1. 各种设计的预测结果。

方法	L1 ↓	均方根误差[b]↓	准确率@0.1[b]↑
帧间平均距离	2.250	0.322	49.2%
不使用侧信息的双向方法[a]	1.354	0.393	52.5%
使用侧信息的单向方法[a]	0.338	0.333	59.7%
使用侧信息的双向方法[a]	**0.288**	**0.302**	**63.6%**

[a] 侧信息：辅助信息（掩码）[b] 均方根误差和准确率指标定义见 [3] 和 [22]。

4.4. 消融研究

预测模块的选择。 我们比较了不同框架选择下的预测质量。在表1中，我们展示了二值掩码对帧预测的影响。此外，我们还比较了单向方法 [3] 和双向方法。为确保公平，我们使用前两帧作为参考帧进行预测；在单向方法实验中，网络骨干仍采用 U-Net。实验表明，我们的设计在所有指标上都取得了最佳预测质量。为展示二值掩码的优势，我们在图5中可视化了预测结果。

残差编码模块的选择。 我们比较了不同的残差帧编码方法，包括我们提出的深度学习方法和先前的手工设计方法。在不同比特率条件下，使用各种方法对相同的残差帧进行编码。结果如图4(b)所示。值得一提的是，在失真程度相同的情况下，我们的方法比先前的方法可实现低 40% 的比特率。

5. 结论

随着激光雷达传感器采集的数据量不断增长，探索如何有效压缩这类数据变得越来越有价值。本文提出 BIRD-PCC，旨在为激光雷达点云寻找一种帧间编码方案。我们采用双向预测框架，并考虑无效像素的坐标，以实现更准确的预测。此外，我们采用基于学习的方法，更有效地利用残差帧中的冗余。KITTI-360 数据集上的实验结果表明，BIRD-PCC 的率失真性能优于先前基于距离图像的方法，并且能够很好地泛化到未见环境。

6. 致谢

本研究部分由台湾国家科学及技术委员会（NSTC）资助，项目编号为 NSTC 111-2634-F-002-022；另获高通通过台湾大学研究合作项目，以及联咏科技奖学金的支持。

7. 参考文献

[1] Sukai Wang、Jianhao Jiao、Peide Cai 和 Lujia Wang，〈R-pcc：基于距离图像的点云压缩基准〉，载于国际机器人与自动化会议（*ICRA*）。IEEE，2022年。

[2] Xuebin Sun、Han Ma、Yuxiang Sun 和 Ming Liu，“一种基于聚类的新型点云压缩算法”，*IEEE*机器人与自动化快报（*RA-L*），2019年。

[3] Xuebin Sun、Sukai Wang、MiaohuiWang、ZhengWang 和 Ming Liu，“一种用于激光雷达点云序列的新型编码架构”，*IEEE*机器人与自动化快报（*RA-L*），2020年。

[4]Google，“Draco”，https://github.com/google/draco。

[5]D Graziosi、O Nakagami、S Kuma、A Zaghetto、T Suzuki 和 A Tabatabai，“点云压缩标准化工作的概述：基于视频的（V-PCC）和基于几何的（G-PCC）”，*AP-SIPA*信号与信息处理汇刊，2020年。

[6] Lila Huang、Shenlong Wang、Kelvin Wong、Jerry Liu 和 Raquel Urtasun，“OctSqueeze：用于激光雷达压缩的八叉树结构熵模型”，载于*IEEE/CVF*计算机视觉与模式识别会议（*CVPR*）论文集，2020年。

[7]Sourav Biswas、Jerry Liu、Kelvin Wong、Shenlong Wang 和 Raquel Urtasun，《MUSCLE：使用深度熵模型进行激光雷达多扫描压缩》，神经信息处理系统进展，2020年。

[8] Chunyang Fu、Ge Li、Rui Song、Wei Gao 和 Shan Liu，《OctAttention：基于八叉树的点云压缩大规模上下文模型》，载于*AAAI* 人工智能会议论文集（*AAAI*），2022年。

[9] Dailan He、Ziming Yang、Weikun Peng、Rui Ma、Hong-wei Qin 和 Yan Wang，《ELIC：采用不均匀分组的空间-通道上下文自适应编码的高效学习式图像压缩》，载于*IEEE/CVF* 计算机视觉与模式识别会议论文集（*CVPR*），2022年。

[10] Zhengxue Cheng、Heming Sun、Masaru Takeuchi 和 Jiro Katto，《采用离散高斯混合似然和注意力模块的学习式图像压缩》，载于*IEEE/CVF* 计算机视觉与模式识别会议论文集（*CVPR*），2020年。

[11] JohannesBall´、David Minnen、Saurabh Singh、Sung Jin Hwang 和 Nick Johnston，“基于尺度超先验的变分图像压缩”，载于国际学习表征会议（*ICLR*），2018年。

[12] Huaizu Jiang、Deqing Sun、Varun Jampani、Ming-Hsuan Yang、Erik Learned-Miller 和 Jan Kautz，“Super slomo：用于视频插帧的高质量多中间帧估计”，载于*IEEE/CVF*计算机视觉与模式识别会议论文集（*CVPR*），2018年。[13] Lili Zhao、Xuhu Lin、Wenyi Wang、Kai-Kuang Ma 和 Jianwen Chen，“Rangeinet：快速激光雷达点云时间插值”，载于国际声学、语音与信号处理会议（*ICASSP*），2022年。[14] B. Mersch、X. Chen、J. Behley 和 C. Stachniss，“使用三维时空卷积网络进行自监督点云预测”，载于机器人学习会议论文集（*CoRL*），2021年。

[15] Guo Lu、Wanli Ouyang、Dong Xu、Xiaoyun Zhang、Chunlei Cai 和 Zhiyong Gao，“Dvc：端到端深度视频压缩框架”，载于*IEEE/CVF* 计算机视觉与模式识别会议论文集（*CVPR*），2019年。

[16] Eirikur Agustsson、David Minnen、Nick Johnston、Jo- hannes Balle、Sung Jin Hwang 和 George Toderici，“用于端到端优化视频压缩的尺度空间光流”，载于*IEEE/CVF* 计算机视觉与模式识别会议论文集（*CVPR*），2020年。

[17] Thomas MCover，信息论基础，John Wiley & Sons，1999年。

[18] David Minnen、Johannes Ball´e 和 George D Toderici，“用于学习式图像压缩的联合自回归先验与层次先验”，载于神经信息处理系统进展，2018年。

[19] Jorma Rissanen 和 Glen G Langdon，“算术编码”，*IBM*研究与开发杂志，1979年。

[20]J. Behley、M. Garbade、A. Milioto、J. Quenzel、S. Behnke、C. Stachniss 和 J. Gall，“SemanticKITTI：用于激光雷达序列语义场景理解的数据集”，发表于*IEEE/CVF*国际计算机视觉会议（*ICCV*）论文集，2019年。

[21] Yiyi Liao、Jun Xie 和 Andreas Geiger，“KITTI-360：用于二维和三维城市场景理解的新型数据集与基准”，*IEEE*模式分析与机器智能汇刊，2022年。

[22] Xuanyu Zhou、Charles R Qi、Yin Zhou 和 Dragomir Anguelov，“RIDDLE：基于距离图像深度增量编码的激光雷达数据压缩”，发表于*IEEE/CVF*计算机视觉与模式识别会议（*CVPR*）论文集，2022年。