点云质量评估：一种采用基于 Transformer 架构的全新无参考方法

Marouane Tliba[1], *Aladine Chetouani*[1],*Giuseppe Valenzise*[2] 和 *Fr´ederic Dufaux*[2]

[1]PRISME 实验室，奥尔良大学，奥尔良，法国[2]巴黎萨克雷大学、法国国家科学研究中心、中央理工-高等电力学院，信号与系统实验室

arXiv:2303.08634v1 [cs.CV] 15 Mar 2023

摘要

随着人们对沉浸式体验的兴趣日益浓厚，点云应运而生，并被广泛采用，成为表示三维媒体的首选方式。除了从采集到渲染过程中可能影响三维内容的各种失真之外，通过传统通信系统传输此类体积内容，往往需要以牺牲最终感知质量为代价。为估计此类质量下降的程度，采用质量指标已成为不可避免的解决方案。在本研究中，我们提出一种新颖的深度学习无参考质量指标，可直接处理完整点云，无需大量预处理，从而能够对传输和渲染环节进行实时评估。为此，我们采用一种新颖的模型设计，主要由交叉注意力层和自注意力层组成，旨在学习最佳的局部语义亲和性，同时在从基础特征提取到深层表示建模的多个层次上，保留几何信息与颜色信息的最佳组合。

索引词— 三维点云、图像质量评估、图神经网络、深度学习。

1. 引言

近年来，点云（PC）日益重要，因为它能够捕捉现实世界物体和环境的细致、准确表征。使用点云的主要优势之一是能够捕捉高水平的细节，包括细微特征和不规则形态，而这些内容可能难以通过其他三维数据格式捕捉。然而，三维点云只有包含大量信息（可达数千甚至数百万个点），才能准确表示三维场景。因此，有损压缩方案经常被采用，也成为不可避免的应用。为了优化沉浸式三维体验的发展，并在给定比特率下实现最佳视觉质量，需要准确的质量指标。因此，近年来，点云质量评估（PCQA）领域受到了研究人员的广泛关注。

点云感知质量可以通过主观或客观方法进行评估。为此，已有多种有效的全参考客观方法。这些方法可分为三大类：基于点、基于特征和基于投影的指标。基于点的指标，例如点到点（Po2Po） [1], 点到平面（Po2Pl）[2], 平面到平面（Pl2Pl） [3] 和点到网格（Po2Mesh）[4], , 通过计算参考点云与其失真版本之间逐点的几何距离和/或特征距离来预测质量。Po2Po 通过测量点对之间的相对距离来估计最终质量；Po2Pl 在 Po2Po 的基础上，将误差向量沿局部法线方向投影；Pl2Pl 则通过测量与参考内容及退化内容中各点相关联的表面之间的角度相似性来量化质量。Po2Mesh 根据参考样本创建多边形网格，然后计算每个失真点与相应表面之间的距离。目前，动态图像专家组（MPEG）正将 Po2Po 均方误差和 Po2Pl 均方误差及其对应的峰值信噪比（PSNR）作为标准点云几何质量指标。基于特征的点云质量指标以全局或局部方式提取逐点几何及其相关属性。在这些指标中，包括 PC-MSDM [5]，它通过考虑局部曲率统计，将二维结构相似性（SSIM）指标 [6] 扩展至点云；利用局部二值模式（LBP） [8] 描述符的 Geotex [7] 指标；以及结合几何和颜色特征的 PCQM [9]。在基于投影的点云质量指标中，三维点或其相关特征会被投影到二维规则网格 [10]中。

也有人提出了一些无参考方法，尤其是基于深度学习的方法。然而，这些方法仍存在明显的局限性和缺点。例如，为了适应不规则的散布式三维点云结构，往往需要进行繁重的预处理，如二维图像投影或转换（体素化）。因此，额外的预处理也使无参考质量评估耗时增加，并且在实践中难以实现。近期，已有一些高效的深度学习指标问世，例如 PointNet-SSNR [11]和 PointNet-DCCFR[12] 。它们利用了点云数据的内在特征。但这些方法的缺点在于，它们关注点云的不同局部区域，并假设整个点云的感知质量一致。

我们提出的无参考点云质量评估方法解决了以往方法中的一个重大局限，即这些方法往往缺乏有效考虑点云区域内部复杂结构关系的能力。

通过采用多层自注意力和交叉注意力，我们能够捕捉点的局部语义亲和性，为评估其视觉质量提供更全面的表征。总而言之，本文的主要贡献总结如下：

1. 我们提出了一种新颖、高效的端到端深度学习方法，用于点云质量评估（PCQA）。该方法直接处理完整点云，无需任何投影或其他变换，从而能够以更具实证性和全面的方式评估点云质量。

2. 我们的方法旨在捕捉点表示中的局部语义亲和性，并利用自注意力在相距较远的特征之间建立连接。这有助于更好地从局部区域提取特征。

3. 我们提出了一种新系统，采用并行双流架构处理点云中的几何特征和颜色特征。双流网络独立运行，同时通过交叉注意力网络设计，在多个层级上动态地相互作用。

2. 所提方法

如图 1 所示，我们方法的整体流程包含三个主要步骤：首先进行预处理，然后利用自注意力和交叉注意力机制提取特征并进行表示映射，最后进行质量估计。

2.1. 预处理

预处理是我们方法中的关键阶段，旨在优化点云（PC）处理流程，包括将点云划分为多个垂直点切片（分区）。这种划分有两个主要目的：（i）实现点的并行处理；（ii）满足图形处理器处理所需的内存要求，因为某些点云可能规模很大（例如包含数百万个点）。为平衡计算负载，分区数量会根据原始点云的大小在 8 到 24 之间变化。随后，每个分区会进一步划分为局部补丁。为此，我们首先选取足够数量的质心，然后应用 k 近邻聚类方法。值得注意的是，所选取的彼此间距较大的质心可确保生成一组覆盖整个分区内所有点的补丁。随后，每个补丁都会输入我们的模型，以进行特征提取和表示建模。通过应用这种预处理方法，我们确保为模型提供一个完整的整体，其中包含所有显著特征，从而增强模型从连贯区域中学习连贯局部信息的能力。

2.2. 提出的 Transformer 模型

2.2.1. 概述

我们的模型设计借鉴了两种现有架构：PointNet [13] 和 Transformer [14]。具体而言，我们采用了 PointNet 的一项核心特性：使用置换不变函数从点集中提取信息。为此，我们使用了两个并行流网络，分别处理颜色和几何信息。该网络通过特征嵌入 层，将输入点的几何信息和颜色信息映射到更高维空间。为了捕获更丰富的局部结构信息，我们在几何流的特征嵌入 基础上引入多头自注意力，以提取相邻点之间的语义亲和性。

为了结合几何和颜色表示，我们采用了多头交叉注意力。这不仅为相邻点表示之间增添了局部连通信息，还补充了相应的颜色特征。随着点嵌入的更新，点表示的注意力连通性会在网络的每个层级动态更新，从而捕捉不同层级的语义亲和性。因此，我们的方法结合了 PointNet 和 Transformer 的优势，同时引入了新特征以提升性能。

总体而言，我们提出的方法与原始 Transformer [14] 几乎秉持相同的理念。PointNet 特征嵌入层取代了输入嵌入和后续的前馈网络。然而，我们的方法并非捕捉有序序列中的上下文，而是旨在寻找点表示之间的语义亲和性，无需施加顺序，因为输入包含明确的几何坐标信息。

2.2.2. 双流网络

我们将几何和颜色视为彼此独立的信息，分别对应两种不同的处理流。根据所考虑的处理流不同，输入点$\mathbf{x}_i$ 所包含的信息也不同。对于几何流，$\mathbf{x}_i = (x_i,\ y_i,\ z_i)$包含点的三维坐标；对于颜色流，点则表示RGB属性信息。值得注意的是，也可以在其他处理流中加入额外特征。

颜色流和几何流均由三个特征嵌入层组成，每个层都包含一系列由非线性函数穿插连接的一维卷积。在几何流中的每个特征嵌入层之后，都会应用一个多头自注意力层，以提取相邻点之间的语义关联。随后，我们采用交叉多头注意力，将学习到的颜色信息与几何表示对齐。

形式上，对于给定的输入点云分区，该分区由 M个图块 $\{P^0, P^1, P^2, .., P^M\}$组成，每个图块 P^i表示为一个大小为 $\mathbb{R}^{N\times F}$的矩阵，其中 N 是图块中的点数，F是每点特征数。在

每一层中，都会应用一个标准的特征嵌入层 $f_{\mathbf{\Theta}}$： $\mathbb{R}^F \to \mathbb{R}^{F'}$，以生成所提供的 X^i_{xyz}的新表示；对于双流中的每一条流，X^i_{rgb} 分别指几何和颜色原始输入或后续生成的表示。随后，将一个多头自注意力层 $MHSA_{\mathbf{\Theta}}$： $\mathbb{R}^{F'} \to \mathbb{R}^{F'}$应用于几何流输出 f（X_{xyz}），生成更新后的表示 X'_{xyz}。从数学上讲，所提出的$MHSA_{\mathbf{\Theta}}$ 可表示如下：

$$\mathbf{q} = X_{xyz}W_q, \mathbf{k} = X_{xyz}W_k, \mathbf{v} = X_{xyz}W_v \tag{1}$$

$$\mathbf{A} = \frac{softmax(\mathbf{q}\mathbf{k}^T)}{\sqrt{\frac{C}{h}}} \tag{2}$$

$$MHSA_{\mathbf{\Theta}}(f(X_{xyz})) = \mathbf{A}\mathbf{v} \tag{3}$$

其中，$W_q, W_k, W_v \in R^{C(C/h)}$ 为可学习参数，C 和h分别为嵌入维度和头数。

X'_{xyz} 表示可视为由图传播层诱导的嵌入。点之间的注意力分数形成一种软连接，用于模拟邻接矩阵。更准确地说，通过设置注意力分数阈值，可以获得点之间的实际连接。因此，为了加快收敛，我们在双流上都采用 GraphNorm 操作 [15]。尽管颜色表示 X'_{rgb} 不能被视为源自类似图的函数，但我们发现，在双流上应用相同类型的归一化有助于保持特征尺度固定。GraphNorm 操作是针对图归一化调整的 InstanceNorm [16], 变体，其中包含一个可学习参数 α，用于决定在平移操作中保留多少通道平均值。该操作表示如下：

$$\mathbf{x}' = \frac{\mathbf{x}' - \alpha \cdot \mathbb{E}[\mathbf{x}']}{\sqrt{\text{Var}[\mathbf{x}' - \alpha \cdot \mathbb{E}[\mathbf{x}']] + \epsilon}} \cdot \gamma + \beta \tag{4}$$

其中，γ 和 β 是可学习的仿射参数，与其他归一化技术中使用的参数类似。

随后，两个流产生的表示 X'_{rgb} 和 X'_{xyz} 最终通过多头交叉注意力层 $MHSCA_{\mathbf{\Theta}}$ 进行融合： $\mathbb{R}^{F'} \to \mathbb{R}^{F'}$。我们注意到，交叉注意力只是自注意力的一种变体，也是一种直观的信息融合方法，其中一个分布的注意力用于突出另一个分布中提取的特征。此处的融合通过测量查询 $\mathbf{q}$（X'_{rgb} 的线性投影）与键 $\mathbf{k}$（X'_{xyz} 的线性投影）之间的相似度，并利用该相似度调整值向量 $\mathbf{v}$（X'_{xyz} 的线性投影）来实现。因此，与 方程 (3) 相比，$MHCA_{\mathbf{\Theta}}$ 函数接收两个输入 f（X'_{xyz}）和 f（X'_{rgb}）。所得向量被视为 X'_{xyz} 的更新版本，并用于后续的网络嵌入层。关于交叉注意力计算的更多细节，请读者参阅 [14, 17]。总而言之，在双流网络中，每一层都会结合几何与颜色流，以利用相应的颜色特征信息更新几何表示。

经过三个连续的（特征嵌入、*MHSA Graph-Norm*、*MHCA*）层块后，会对每个图块独立应用**点嵌入聚合**。此处对特征的通道维度进行最大池化和平均池化操作。随后，将两种池化操作的结果拼接成一个向量，因此，每个包含 M个图块的点云分区都会转换为由 $2F'$维向量组成的 M 序列。

2.2.3. 特征聚合与质量估计

为了聚合从每个点云分区获得的表示并捕捉它们之间的亲和关系，我们采用了**图块聚合**方法。该方法包含一个多头自注意力层 [14]，[18]，随后通过最大池化生成一个表示各分区的向量。最后，使用一个浅层多层感知机估计质量分数，并通过计算分区分数序列的平均值获得整体分数。

3. 实验

3.1. 训练与实现细节

我们使用均方误差（MSE）作为损失函数，以端到端的方式训练模型。该网络旨在建立输入点云与质量平均意见分数（MOS）之间的映射函数。损失函数定义如下：

$$\mathcal{L} = \text{MSE}\left(\text{mean}\left(\sum_i Out_i\right), \mathcal{Y}\right), \tag{5}$$

其中，Out_i 表示每个预测的分区分数，$\mathcal{Y}$表示 MOS。

该模型使用 PyTorch 实现，并采用 Adam 优化器 [19] 进行训练，初始学习率为 0.0001，批次大小为 1。训练轮数因训练折而异，在 ICIP 数据集中为 80 至 200 轮。

3.2. 评估协议与结果分析

为评估我们模型的有效性，我们在两个公开可用的基准上进行了实验。这两个基准均采用主观评分，并使用不同的新兴压缩方案：ICIP20 [20] 和 PointXR [21]。ICIP20 包含 6 个参考点云，每个点云均以 5 个级别进行压缩，通过三种压缩方式生成了 90 个失真版本。PointXR 包含 5 个点云，每个点云均采用 G-PCC 进行压缩，其中几何压缩使用八叉树编码，颜色压缩使用提升和 RAHT，共生成 45 个失真版本。

我们采用 6 折交叉验证协议，在 ICIP20 上训练和测试模型；每次迭代使用 5 个参考点云进行

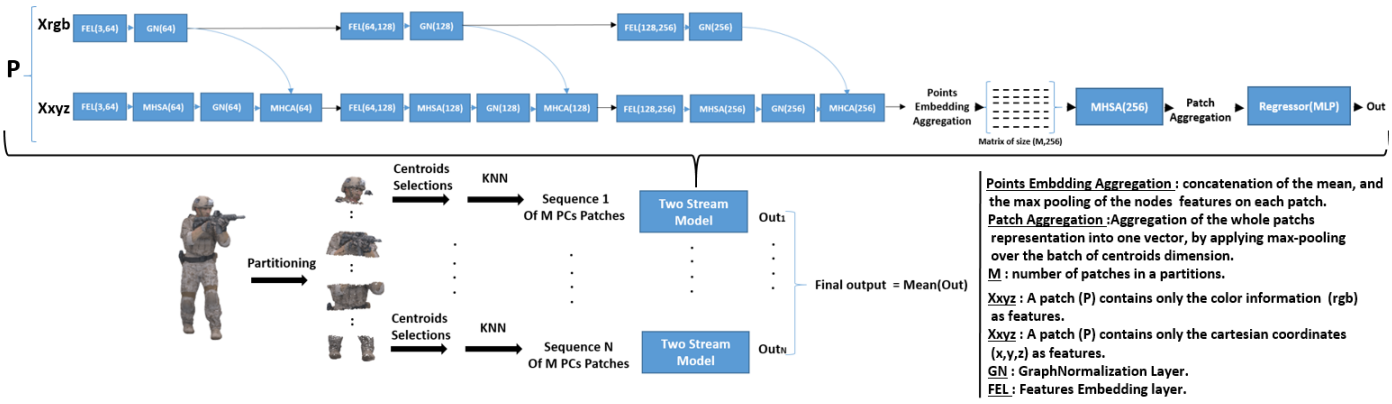

图1。我们提出方法的总体流程

表1。在 ICIP20 上训练模型并在 PointXR 上测试所得的结果

模型	PLCC ↑	SROCC ↑
点对点均方误差	0.887	0.978
点对平面均方误差	0.855	0.942
点对点均方误差峰值信噪比	**0.983**	**0.978**
点对平面均方误差峰值信噪比	0.972	0.950
PointNet-DCCFR	0.981	0.964
PointGraph	0.967	**0.988**
Our	**0.969**	**0.990**

训练，并使用 1 个点云进行测试。为评估我们的方法在未知点云上预测质量的泛化能力，我们还在 ICIP20 上训练，并在 PointXR 上测试，使用皮尔逊相关系数和斯皮尔曼相关系数进行评估。结果见表2和表1，相关系数均值根据所有折计算得出。我们使用相同的协议，将本方法的性能与当前最先进的方法进行了比较。

表2展示了我们的方法在 ICIP20 数据集上的结果，并将其与最先进方法进行了比较。结果表明，我们的方法与主观真实值具有很强的相关性；与大多数现有方法相比，在 PLCC 和 SROCC 指标上均呈现明显优势。我们提出的方法优于所有其他方法，取得了最高的 SROCC 分数，相关系数达到 **0.976**；同时，与 po2planeMSe 方法并列取得最高的 PLCC 分数，相关系数达到 **0.959**。值得注意的是，我们的方法甚至优于大多数全参考方法。特别是，表中列出的所有方法均为全参考方法，只有 PointNet-SSNR 和 PointNet-Graph 属于无参考方法。

跨数据集评估结果如表1所示，表明我们的方法实现了较高的相关性，并优于部分对比方法。这些结果体现了我们的指标在预测未知点云质量方面的泛化能力，说明所提出的方法在不同验证集上表现稳定。随着未来出现更多带标注的点云质量数据集，

表2. 使用六折交叉验证在 ICIP20 数据集上获得的结果

模型	PLCC ↑	SROCC ↑
点到点均方误差	**0.946**	0.934
点到平面均方误差	**0.959**	0.951
点到点均方误差峰值信噪比	0.868	0.855
点到点 HAU 峰值信噪比	0.548	0.456
点到平面 HAU 峰值信噪比	0.580	0.547
颜色 Y 通道均方误差	0.876	0.892
颜色 Cb 通道均方误差	0.683	0.694
颜色 Cr 通道均方误差	0.594	0.616
亮度 Y 峰值信噪比	0.887	0.892
色度 Cb 峰值信噪比	0.693	0.694
色度 Cr 峰值信噪比	0.626	0.617
点到平面 平均值	0.922	0.910
点到平面 均方误差	0.925	0.912
PCQM	0.796	0.832
GraphSim	0.931	0.893
PointNet-SSNR	0.908	0.955
PointNet-DCCFR	**0.947**	**0.973**
PointGraph	**0.946**	**0.973**
Our	**0.959**	**0.976**

我们的方法还可以得到进一步验证。

4. 结论

本文提出一种采用基于学习的方法、用于点云的全新无参考质量指标。我们的网络采用多头自注意力和交叉注意力机制，以捕获局部语义亲和性，并建立点与点之间的概率连接。由于感知质量可能在几何和颜色层面均发生劣化，我们提出一种双流架构，并行处理颜色失真和几何失真，同时在网络的多个层级动态交互。我们的方法在两个点云质量评估基准上取得了最先进的相关性。

5. 参考文献

[1] C. Tulvan R. Mekuria、Z. Li 和 P. Chou，“PCC（点云压缩）的评估准则”，载于*ISO/IEC MPEG* 文档。N16332，2016 年，第 II 卷，第 803–806 页。[2] D. Tian 等人，“点云压缩的几何失真度量”，载于*IEEE ICIP*，2017 年。[3] E. Alexiou 和 T. Ebrahimi，“基于角度相似性的点云质量评估指标”，载于*IEEE ICME-W*，2018 年。[4] C. Rochinni P. Cignoni 和 R. Scopigno，“Metro：简化曲面误差的测量”，载于*Computer GraphicsForum*，1998 年，第 17 卷，第 167–174 页。[5] J. Digne G. Meynet 和 G. Lavou´e，“PC-MSDM：一种三维点云质量指标”，2019 年。[6] Zhou Wang、A.C. Bovik、H.R. Sheikh 和 E.P. Simoncelli，“图像质量评估：从误差可见性到结构相似性”，*IEEE Transactions on Image Processing*，第 13 卷，第 4 期，第 600–612 页，2004 年。[7] Rafael Diniz、Pedro Garcia Freitas 和 Myl`ene C. Q. Farias，“基于局部二值模式的点云质量评估模型探索”，载于*QoMEX*，2020 年，第 1–6 页。[8]Matti Pietik¨ainen 和 Guoying Zhao，“局部二值模式的二十年：综述”，*CoRR*，第 abs/1612.06795 卷，2016 年。[9] J. Digne G. Meynet、Y. Nehm´e 和 G. Lavou´e，“PCQM：一种彩色三维点云全参考质量指标”，2020 年。

[10] Aladine Chetouani、Maurice Quach、Giuseppe Valenzise 和 Frederic Dufaux，“用于有参考点云质量评估的卷积神经网络”，载于*2021*年*IEEE*第*23*届多媒体信号处理国际研讨会（*MMSP*），2021年，第1–6页。

[11] Marouane Tliba、Aladine Chetouani、Giuseppe Valenzise 和 Fr´ederic Dufaux，“无参考点云质量评估的表示学习优化”，载于*2022*年*IEEE*图像处理国际会议（*ICIP*），2022年，第3702–3706页。

[12] Marouane Tliba、Aladine Chetouani、Giuseppe Valenzise 和 Fr´ed´eric Dufaux，“利用深层特征交叉相关进行点云质量评估”，载于第*2*届视觉多媒体应用体验质量研讨会论文集，2022年，第63–68页。

[13] C. Qi、L. Yi、Hao Su 和 Leonidas J.Guibas，《Pointnet ++：度量空间中点集的深度层次特征学习》，2017。

[14]Ashish Vaswani、Noam Shazeer、Niki Parmar、Jakob Uszkoreit、Llion Jones、Aidan N. Gomez、Łukasz Kaiser 和 Illia Polosukhin，《注意力就是你所需的一切》，载于第 *31* 届神经信息处理系统国际会议论文集，美国纽约州雷德胡克，2017，NIPS’17，第 6000–6010 页，Curran Associates Inc.[15] Tianle Cai、Shengjie Luo、Keyulu Xu、Di He、Tie yan Liu 和 Liwei Wang，《Graphnorm：一种加速图神经网络训练的原则性方法》，2020。[16] Dmitry Ulyanov、Andrea Vedaldi 和 Victor S. Lempitsky，《实例归一化：快速风格化缺失的关键》，*arXiv*，第 abs/1607.08022 卷，2016。[17]Chun-Fu Richard Chen、Quanfu Fan 和 Rameswar Panda，《Crossvit：用于图像分类的交叉注意力多尺度视觉 Transformer》，载于*2021* 年 *IEEE/CVF* 国际计算机视觉会议（*ICCV*），2021，第 347–356 页。[18] M Tliba 等，《Satsal：一种基于多层自注意力的视觉显著性预测架构》，*IEEE Access*，第 10 卷，第 20701–20713 页，2022。[19] Diederik P. Kingma 和 Jimmy Ba，《Adam：一种随机优化方法》，*CoRR*，第 abs/1412.6980 卷，2015。[20] S. Perry 等，《使用 MPEG 编解码器编码的静态点云质量评估》，第 3428–3432 页，2020。[21] N. Yang、E. Alexiou 和 T Ebrahimi，《PointXR：用于虚拟现实中点云可视化和主观评估的工具箱》，2020。