No-Reference Point Cloud Quality Assessment via Graph Convolutional Network

Wu Chen, Qiuping Jiang, Wei Zhou, Feng Shao, Guangtao Zhai, Weisi Lin

***Abstract*—Three-dimensional (3D) point cloud, as an emerging visual media format, is increasingly favored by consumers as it can provide more realistic visual information than two-dimensional (2D) data. Similar to 2D plane images and videos, point clouds inevitably suffer from quality degradation and information loss through multimedia communication systems. Therefore, automatic point cloud quality assessment (PCQA) is of critical importance. In this work, we propose a novel no-reference PCQA method by using a graph convolutional network (GCN) to characterize the mutual dependencies of multi-view 2D projected image contents. The proposed GCN-based PCQA (GC-PCQA) method contains three modules, i.e., multi-view projection, graph construction, and GCN-based quality prediction. First, multi-view projection is performed on the test point cloud to obtain a set of horizontally and vertically projected images. Then, a perception-consistent graph is constructed based on the spatial relations among different projected images. Finally, reasoning on the constructed graph is performed by GCN to characterize the mutual dependencies and interactions between different projected images, and aggregate feature information of multi-view projected images for final quality prediction. Experimental results on two publicly available benchmark databases show that our proposed GC-PCQA can achieve superior performance than state-of-the-art quality assessment metrics. The code will be made available soon.**

***Index Terms*—Point cloud, multiple views, projection, graph convolution, no-reference, quality assessment.**

I. Introduction

In recent years, the development of three-dimensional (3D) visual information acquisition technology makes point clouds easier to obtain and gradually becomes a popular type of visual data. A 3D Point cloud is mainly used to describe a complete 3D scene or object, including geometric attributes (position of each point in 3D space), color attributes (RGB attributes of each point), and others (normal vector, opacity, reflectivity, time, etc.) [1]. Point clouds have been widely studied and used in a wide range of application scenarios such as 3D reconstruction [2], [3], classification and segmentation [4], [5], facial expression representation [6], autonomous driving [7], [8], and virtual reality [9], etc. Although point cloud can realistically record 3D objects through a huge point set, it also consumes a lot of memory, and it is difficult to achieve data transmission under limited network bandwidth [10], [11]. This new and effective data representation put forward a challenge to the current hardware storage and network transmission. Therefore, in order to achieve efficient storage and transmission, compression of point clouds is necessary [12]–[15]. However, point cloud compression may introduce artifacts, resulting in the degradation of point cloud visual quality. Point cloud visual quality is an important way to compare the performance of various point cloud processing algorithms. Effective point cloud quality assessment (PCQA) methods can not only help people evaluate the distortion degree of point clouds and the performance of compression algorithms but also be beneficial to optimize the visual quality of distorted point clouds. Thus, how to accurately assess the perceptual quality of point clouds has become a critical issue.

Similar to image quality assessment (IQA), PCQA can also be divided into subjective and objective methods. The subjective method is mainly based on the perception of the human visual system (HVS). It is difficult to be widely applied because this kind of assessment requires a large number of participants to ensure the rationality and accuracy of the assessment results in a statistical sense. Currently, the results obtained from subjective assessment experiments are generally served as the ground-truth data for benchmarking different objective methods [16]. According to the participation of original point clouds, objective PCQA methods can have three categories: full reference (FR), reduced reference (RR), and no reference (NR). Since the original point clouds are not always available, NR-PCQA methods that do not rely on any original information as a reference are more suitable in practical applications.

The traditional NR-PCQA methods [17], [18] generally predict the quality score by extracting quality-aware features based on the analysis of point cloud attributes such as geometry and color. Recently, the great success of deep learning in the field of NR-IQA has promoted the development of deep learning-based NR-PCQA metrics [19]–[24]. The common practice of these deep NR-PCQA metrics is to directly apply the ordinary convolution operation on the point cloud for automatic feature learning in a data-driven manner. Nonetheless, point cloud is a typical kind of non-Euclidean data which is sparsely distributed over the 3D space, and a large number of useless pixels are also involved with pixel-by-pixel convolution, thus resulting in a huge waste of resources and inefficient data processing. In order to solve this problem, some related works try to represent non-Euclidean

W. Chen, Q. Jiang, and F. Shao are with the School of Information Science and Engineering, Ningbo University, Ningbo 315211, China (e-mail: jiangqiuping@nbu.edu.cn).

W. Zhou is with the School of Computer Science and Informatics, Cardiff University, Cardiff CF24 4AG, United Kingdom (e-mail: zhouw26@cardiff.ac.uk)

G. Zhai is with the Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai 200240, China (zhaiguangtao@sjtu.edu.cn)

W. Lin is with the School of Computer Science and Engineering, Nanyang Technological University, Singapore (wslin@ntu.edu.sg)

Corresponding author: Qiuping Jiang

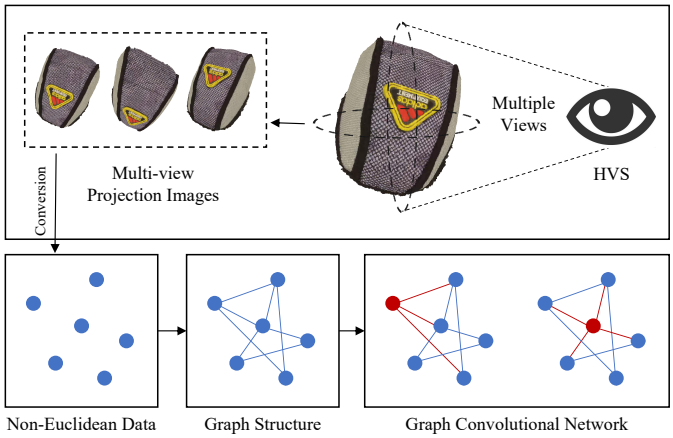

Fig. 1. We simulate the perceptual process of HVS to perform multi-view projection on the 3D point cloud and build the graph based on the projected images. The red node in graph convolution represents the central node of the current convolution process, and the red line represents the adjacency relationship. The central node will constantly exchange information with neighboring nodes to aggregate feature information from neighbors.

data with the graph which includes node information and complex adjacency relations between nodes. With the graph-based non-Euclidean data as input, the current works then introduce to use graph convolutional network (GCN) rather than traditional convolutional neural network (CNN) for more effective feature representation learning [25]. For instance, Thomas et al. [26] proposed to convert non-Euclidean data into a graph based on which the GCN is used to realize graph feature extraction. As a typical kind of non-Euclidean data, GCN has also been applied to many point cloud-based vision tasks, such as point cloud classification [27], [28], point cloud segmentation [29], point cloud data analysis [30], action recognition [31], etc. Moreover, it has also been applied to infer the perceptual quality of various multimedia data, e.g., traditional 2D images [32], [33], 360-degree images [34], [35], and meshes [36], [37].

Due to the strong capability of GCN in handling non-Euclidean data including 3D point cloud, this paper presents a novel GCN-based NR-PCQA method (GC-PCQA). One of the most critical issues is to effectively create a graph of the point cloud so that the GCN can be applied for feature learning. Since the goal of PCQA is to predict the quality of the test point cloud consistent with human perception, how to construct a highly perception-consistent graph of point clouds is the key to its success. It is known that the HVS reconstructs 3D objects in their mind based on multiple two-dimensional (2D) plane images observed from different viewpoints. In order to imitate the process of the HVS to perceive 3D objects, it is natural to perform multi-view projection on the point cloud to obtain a set of projected images with each corresponding to a specific viewpoint. Although these projected images are independent individuals, there is a certain extent of correlation between each other. Therefore, we regard all projected images as a set of non-Euclidean data and then establish a graph according to the dependencies between each individual projected image. Finally, GCN is applied to realize feature extraction from the constructed graph for quality prediction. The entire process is simply illustrated in Fig. 1. Experimental results demonstrate that our proposed GC-PCQA method outperforms state-of-the-art reference and non-reference PCQA methods on two public PCQA databases. Overall, the main contributions of this paper are as follows:

1) We perform multi-view projection on the point cloud to obtain a set of projected images based on which a highly perception-consistent graph is constructed to model the mutual dependencies of multi-view projected images. The graph nodes are defined with the projected images and connected by spatial relations between each other.
2) We perform GCN on the proposed graph to characterize the interactions between different projected images and aggregate the feature information of multi-view projected images for final quality prediction. The ablation study validates the effectiveness of the GCN architecture and the source code is available for public research usage.
3) We fuse the horizontally and vertically projected image features extracted by two GCNs that do not share weights to boost the performance. Experimental results show that the proposed GC-PCQA can predict subjective scores more accurately than the existing state-of-the-art PCQA metrics.

The rest of this paper is organized as follows. In Section II, we introduce the related works. In Section III, we illustrate the proposed GC-PCQA with technical details. We conduct experiments and analyze the results in Section IV, and finally draw conclusions in Section V.

II. Related Work

PCQA metrics have developed rapidly and can be mainly divided into PC-based metrics and projection-based metrics. PC-based metrics evaluate the quality score through the characteristic information of each point in the point cloud. While projection-based metrics use the projected images of the point cloud instead of the point cloud itself.

A. PC-based Metrics

As one of the important evaluation methods, the FR method has been widely investigated in PC-based metrics. The initial methods calculate quality scores based on geometric information of point clouds, such as $PSNR_{MSE,p2po}$ and $PSNR_{HF,p2po}$ [38], $PSNR_{MSE,p2pl}$ and $PSNR_{HF,p2pl}$ [39]. Among them, the point-to-point methods (p2point) compute the L_2 norm of the nearest point pair as the distortion measure of the point, while point-to-plane methods (p2plane) increase the normal vector of the plane. Besides, Alexiou et al. [40] captured the distortion point cloud degradation through the angular similarity between the corresponding points. Javaheri et al. [41] used the generalized Hausdorff distance for PCQA. In addition to geometric properties, the color properties of point clouds can also be used as one of the important features for quality assessment. $PSNR_Y$ [42] evaluates texture distortion of colored point clouds based on point-to-point color components. Viola et al. [43] use global color statistics, such as color histograms and correlograms, to

evaluate the degree of distortion of a point cloud. On the basis of PC-MSDM [44], Meynet et al. [45] proposed a linear model PCQM based on curvature and color attributes to predict 3D point cloud visual quality. Inspired by the idea of similarity, Alexiou et al. [46] used the structural similarity index based on geometric and color features for evaluation. Diniz et al. [47]–[49] extract statistical information of point clouds based on local binary pattern descriptors and local luminance pattern descriptors, which are assessed by distance metrics. Yang et al. [50] proposed to construct the local graph representation of the reference point cloud and the distorted point cloud respectively with the key point as the center and calculate the similarity feature by extracting three color gradient moments between the central key point and all other points, so as to estimate the quality score of the distorted point cloud. They also believed that point clouds have potential energy, and used multiscale potential energy discrepancy (MPED) [51] to quantify point cloud distortion. Furthermore, Viola et al. [52] extracted part of the geometric, color, and normal features from the point clouds, and then evaluated the distorted point cloud by finding the best combination of features through a linear optimization algorithm. Liu et al. [53] proposed an analytical model with only three parameters to accurately predict the MOS of V-PCC compressed point clouds from geometric and color features. All of these methods use reference point clouds, and despite the advanced performance achieved, these methods may not be useful in practical applications. Therefore, it is meaningful to research NR methods to overcome the problem of missing reference point clouds. Zhang et al. [18] used 3D natural scene statistics (3D-NSS) and entropy to extract geometric and color features related to quality, and then predicted quality scores through a support vector regression (SVR) model. With structure-guided resampling, Zhou et al. [54] estimated point cloud quality based on geometry density, color naturalness, and angular consistency. The success of deep learning in various research fields has prompted researchers to introduce it into PCQA. Chetouani et al. [19] used deep neural networks (DNNs) to learn the mapping of low-level features such as geometric distance, local curvature, and luminance values to quality scores. Liu et al. [20] constructed a large-scale PCQA dataset named LS-PCQA, which contains more than 22,000 distortion samples, and then proposed an NR metric based on sparse CNN.

B. Projection-based Metrics

In addition to the above methods that directly use point clouds for quality evaluation, projection-based metrics also play an important role in PCQA. The projection-based PCQA metrics project the point cloud from 3D space to 2D plane, so as to transform PCQA into IQA which has been relatively mature. Therefore, the existing IQA methods can be directly used to evaluate the quality of 2D projection images, such as PSNR [55], SSIM [56], MS-SSIM [57], IW-SSIM [58], VIFP [59], etc. Freitas et al. [60] used a multi-scale rotation invariant texture descriptor called Dominant Rotated Local Binary Pattern (DRLBP) to extract statistical features from these texture maps and calculate texture similarity. Finally, texture features and similarity features were fused to predict the visual quality of point clouds. Hua et al. [17] proposed a blind quality evaluator of colored point cloud based on visual perception, which reduces the influence of visual masking effect by projecting the point cloud onto a plane to extract geometric, color and joint features. Tao et al. [21] projected the color point cloud in 3D space into a 2D color and geometric projection map, and then weighted the quality scores of local blocks in the map based on a multi-scale feature fusion network. Liu et al. [22] projected the six planes of the 3D point cloud, and then extracted multi-view features through a DNN to classify the distortion types of the point cloud. Finally, the final quality score was obtained by multiplying the probability vector and the quality vector. Tu et al. [23] designed a two-stream CNN to extract the features of texture projection maps and geometric projection maps. Yang et al. [24] used natural images as the source domain and point clouds as the target domain, and predicted point cloud quality through unsupervised adversarial domain adaptation.

Based on the above statements, PC-based metrics and projection-based metrics have achieved certain results. However, most of the existing projection-based PCQA metrics are evaluated based on six projection planes, which do not take into account the quality perception of HVS for 3D point clouds from multiple views, and do not make full use of the correlation between different projected images for modeling. Thus, we propose a novel non-reference PCQA method by using GCN to characterize the mutual dependencies of multi-view 2D projected image contents.

III. Proposed Method

We first give a brief overview of the proposed GC-PCQA method. Then, the details of each module in our GC-PCQA method will be illustrated. Finally, we describe how the network is trained.

A. Overview

The framework of our proposed GC-PCQA method is shown in Fig. 2. It is mainly composed of three parts: multi-view projection, graph construction, and GCN-based quality prediction. Firstly, considering the behavior of the HVS when observing 3D point clouds, multi-view projection is performed on the point cloud to obtain a set of horizontally and vertically projected 2D images. The horizontally (vertically) projected 2D image set covers the visual contents that can be perceived within the horizontal (vertical) visual field by an observer. All these projected images are fed into a pre-trained backbone and an attention block for attentive feature extraction. Secondly, a multi-level fusion of attentive feature maps is carried out through the multi-level conversion module, and graph construction is performed according to spatial relations among different projected images. Thirdly, reasoning on the constructed graph is performed by GCN to model the mutual dependencies between nodes and generate more effective feature representations. Finally, multi-level feature fusion is carried out on the feature representations obtained by two GCNs (corresponding to the horizontal visual field and the

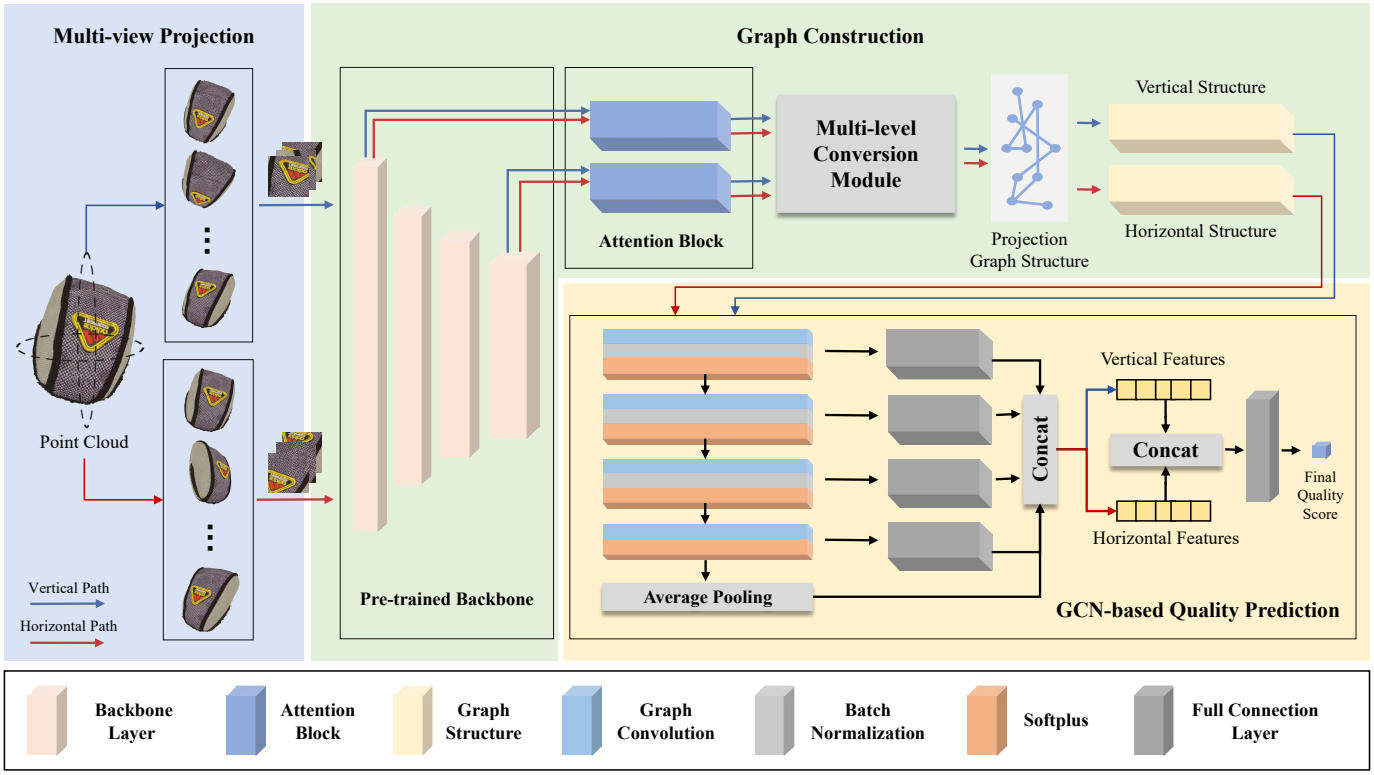

Fig. 2. Framework of the proposed method. It is mainly composed of three parts: multi-view projection, graph construction, and GCN-based quality prediction. Firstly, multi-view projection is performed on the point cloud to obtain a set of horizontally and vertically projected 2D images. All these projected images are fed into a pre-trained backbone and an attention block for attentive feature extraction. Secondly, a multi-level fusion of attentive feature maps is carried out through the multi-level conversion module, and graph construction is performed according to spatial relations among different projected images. Thirdly, reasoning on the constructed graph is performed by GCN to model the mutual dependencies between nodes and generate more effective feature representations. Finally, multi-level feature fusion is carried out on the feature representations obtained by two GCNs to predict the final quality score.

vertical visual field, respectively) to predict the final quality score.

B. Multi-view Projection

Point cloud consists of huge point sets, which is mainly used to describe 3D objects in detail, resulting in a large volume of point cloud and it is difficult to directly input point cloud data into the network. In order to reduce the cost of processing large-scale point clouds, many point cloud resampling strategies [61]–[63] and point cloud projection methods [21]–[24] have been proposed to simplify point cloud data. Since deep learning is very effective in the field of image processing, we choose to use the multi-view projection method to convert the point cloud into an image, so as to take advantage of deep learning for PCQA.

The visual field of the human eye is divided into horizontal visual field and vertical visual field. The range of view in different directions is limited, i.e., the horizontal view limit is approximately 190 degrees while the vertical view limit is 135 degrees. When human eyes observe 3D objects such as 3D point clouds, it is difficult to directly observe all contents within 360 degrees [34]. In general, human reconstructs 3D objects in the brain by observing from different viewpoints, so as to have a clearer perception of 3D objects. In order to imitate the process of HVS to perceive 3D objects, we perform a multi-view projection operation on the point cloud. Just as the human eye perceives 3D objects, we rotate and project the point cloud onto the 2D space in both horizontal and vertical directions, with a rotation stride (RS) of $360/N$ degrees. Finally, for each direction, we obtain a multi-view projected image group $\mathbf{P}$ containing N projected images:

$$\mathbf{P} = [\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3, \cdots, \mathbf{Y}_N] \in \mathbb{R}^{N\times 3\times H\times W}, \quad (1)$$

where $\mathbf{Y}_i$ represents the i-th projected image, and N is the total number of projected images. H and W indicate the height and width of each individual projected image, respectively. We use $\mathbf{P}_H$ to denote the horizontally projected image group and $\mathbf{P}_V$ to denote the vertically projected image group.

C. Graph Construction

The graph construction module aims to construct a perception-consistent graph based on the features extracted from multi-view projected images. It is mainly composed of a pre-trained backbone network, an attention block, and a multi-level conversion module.

1) Feature Extraction: We adopt the pre-trained ResNet101 [64] as the backbone for feature extraction. The input image size of the backbone network is 224×224. Since the point cloud only exists in the middle of the projected image and the white background may have a negative effect, an additional image pre-processing is performed as follows. First, the projected image is cropped to best remove the useless white

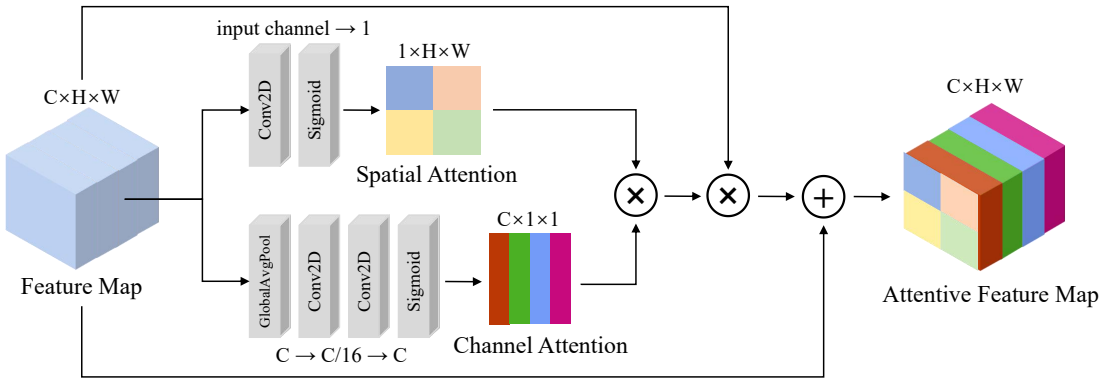

Fig. 3. The attention block is used on the feature map extracted by the pre-trained backbone to obtain the attentive feature map. The upper part extracts spatial attention for the input feature map, and the lower part extracts channel attention. H: image height, W: image width, C: image channel.

background regions. Then, an informative 224×224 image patch is obtained by resizing as the network input.

2) Attention Block: The attention mechanism comes from the research on human vision and draws lessons from the attention thinking of the human vision, which can make the feature extractor focus more on those significant areas of the target while suppressing the most unimportant information to improve the performance of DNNs. At present, a variety of attention modules have been proposed such as SE attention [65], CBAM attention [66], scSE attention [67], etc. Therefore, in order to further improve the feature representation capability of the network, we devise an attention block to impose appropriate attention weights to the feature maps obtained by the pre-trained backbone. The structure of our devised attention block is shown in Fig. 3.

We first introduce the upper flow, i.e., the spatial attention. The input feature maps from the pre-trained backbone can be represented as $\mathbf{F}^{(i)} \in \mathbb{R}^{C\times H\times W}$, where i represents the output of the i^{th} layer of the backbone. Firstly, 2D convolution with a convolution kernel size of 1 is used to reduce the number of channels to 1. Then, a sigmoid activation function is applied to map the range of feature values into [0,1]. Finally, the spatial attention map $\mathbf{F}_S \in \mathbb{R}^{H\times W}$ is expressed as follows:

$$\mathbf{F}_S = \delta(\phi(\mathbf{F})) = \begin{bmatrix} \delta(S_{11}) & \delta(S_{12}) & \cdots & \delta(S_{1W}) \\ \delta(S_{21}) & \delta(S_{22}) & \cdots & \delta(S_{2W}) \\ \vdots & \vdots & \vdots & \vdots \\ \delta(S_{H1}) & \delta(S_{H2}) & \cdots & \delta(S_{HW}) \end{bmatrix}, \tag{2}$$

where ϕ represents the 2D convolution with a convolution kernel size of 1, $\delta(\cdot)$ denotes the sigmoid function, and $S_{ij}(i \in \{1,2,\ldots,H\}, j \in \{1,2,\ldots,W\})$ indicates the relative importance of the eigenvalues at position (i,j).

For the lower flow, global average pooling is first applied on the input feature maps $\mathbf{F} \in \mathbb{R}^{C\times H\times W}$. Then, the channel dimension of the feature maps is reduced and then increased by two 2D convolution layers with a convolution kernel size of 1. The final channel attention map $\mathbf{F}_C \in \mathbb{R}^{C\times 1\times 1}$ is generated by attaching a sigmoid activation function $\delta(\cdot)$ in the end, which can be expressed as follows:

$$\mathbf{F}_C = \delta(\phi(avg(\mathbf{F}))) = [\delta(U_1), \delta(U_1), \cdots, \delta(U_C)], \tag{3}$$

where $U_i(i \in \{1,2,\ldots,C\})$ denotes the relative importance of the i^{th} channel among all channels, avg denotes the global average pooling. After obtaining two attentions, the spatial attention map and the channel attention map are multiplied to obtain a mixed attention map $\mathbf{F}_{SC} \in \mathbb{R}^{C\times H\times W}$, which makes the information important in both space and channel dimensions more prominent and will encourage the network to learn more meaningful features. Then, the skip connection is used to multiply the original feature map and the mixed attention map pixel-by-pixel to complete information calibration. Finally, the residual connection is used to alleviate the gradient disappearance problem caused by increasing depth in the DNN. Mathematically, the final attentive feature map $\hat{\mathbf{F}}$ is generated as follows:

$$\hat{\mathbf{F}} = (\mathbf{F}_S \times \mathbf{F}_C) \odot \mathbf{F} + \mathbf{F}, \tag{4}$$

where $\odot$ denotes pixel-by-pixel multiplication.

3) Multi-level Conversion Module: Feature fusion [68]–[70] is an important way to make full use of the information from each individual feature input. Generally, the low-level features in shallow layers have higher resolution and contain more detailed information, while the high-level features in deep layers have lower resolution and stronger semantic representation ability. In addition, it has been demonstrated that the HVS tends to perform multi-level feature fusion in perceiving image quality [57]. In our method, we design a multi-level conversion module to fuse the low-level and high-level features from the 1st layer and the 4th layer of the backbone network to exploit the complementarity between them.

The multi-view projected images are obtained from the point cloud under different viewpoints. Therefore, there is a certain inter-dependency between each other. To capture and exploit such a kind of dependency, we build a graph based on the correlation between those multi-view projected images by taking each projected image feature representation as a node in the graph. As a consequence, the obtained multi-level attentive feature maps need to be converted into the feature vector $\mathbf{h}_{v_i} \in \mathbb{R}^D$ of the graph node v_i $(i \in \{1,2,\ldots,N\})$ by the multi-level conversion module, where D is the feature dimension. Formally, the conversion process of a feature vector

can be expressed as

$$\mathbf{h}_v = avg(\hat{\mathbf{F}}^{(1)}) \oplus avg(\hat{\mathbf{F}}^{(4)}), \tag{5}$$

where $\hat{\mathbf{F}}^{(1)}$ and $\hat{\mathbf{F}^{(4)}}$ represent the attentive feature maps output by the attention block, and the input of the attention block comes from the 1st and 4th layers of the pre-trained backbone, respectively, $\oplus$ represents the concatenation of the channels of the feature maps. Finally, we create a set of nodes $\mathbf{V} = [\mathbf{h}_{v_1}, \mathbf{h}_{v_2}, \cdots, \mathbf{h}_{v_N}]^{\mathrm{T}}$ based on all the feature vectors, Meanwhile, the adjacency relation between any two nodes v_i, v_j can be expressed as

$$\mathbf{A}(v_i, v_j) = \begin{cases} 1, & \text{if } AugularDist(v_i, v_j) \leq \theta \\ 0, & \text{otherwise} \end{cases}, \tag{6}$$

where $\mathbf{A} \in \mathbb{R}^{N \times N}$ is the adjacency matrix representing the correlation between nodes of the graph, $AugularDist(\cdot)$ computes the angle between the projected images corresponding to two nodes, and the angular distance threshold θ is set to $36°$ with experiments. Specifically, if the angle between the center points of two projected images is less than θ, the two projected images are considered to be connected, otherwise they are not adjacent. Then, by multiplying both sides of the adjacency matrix $\mathbf{A}$ by the square root of the degree matrix for normalization [26], [71], the graph nodes with many neighbor nodes are avoided to have too much influence. The normalization formula is as follows

$$\widehat{\mathbf{A}} = \mathbf{D}^{-\frac{1}{2}}(\mathbf{A} + \mathbf{I})\mathbf{D}^{\frac{1}{2}}, \tag{7}$$

where $\widehat{\mathbf{A}}$ is the normalized adjacency matrix. $\mathbf{D}$ is the degree matrix, which takes the degree of the corresponding node as the value only on the diagonal and is 0 in the rest of the positions. Concretely, $\mathbf{D}_{ii} = \sum_{j=0}^{N} \mathbf{A}(v_i, v_j)$ $(i \in \{1, 2, \cdots, N\})$. $\mathbf{I}$ is the identity matrix and adds self-join to the adjacency matrix. Finally, we construct the graph $\mathbf{G} = (\mathbf{V}, \widehat{\mathbf{A}})$, so that the correlation between nodes v_i and v_j can be represented by the corresponding value $\mathbf{A}(v_i, v_j)$ in the adjacency matrix.

D. GCN-based Quality Prediction

After graph construction, we use GCN to model the interaction between the contents of different projected 2D images according to the graph $\mathbf{G}$, thus completing the quality prediction.

1) Graph Convolutional Network: We take the two graphs corresponding to the horizontal and vertical projection image feature groups into two GCNs without weight sharing, respectively, and update the new node representations by constantly exchanging neighborhood information based on the $\widehat{\mathbf{A}}$. The GCN consists of four graph convolutional blocks, and the number of output channels of these blocks are [512, 128, 32, 1]. The graph convolutional block include a graph convolutional layer, asoftplus activation function, and a batch normalization layer. The process of GCN can be described as

$$(\mathbf{H}^{(1)}, \mathbf{H}^{(2)}, \mathbf{H}^{(3)}, \mathbf{H}^{(4)}) = M_G(\mathbf{G}, \widehat{\mathbf{A}}; \mathbf{w}_G), \tag{8}$$

where $\mathbf{H}^{(l)}$ is the feature matrix after activation of the l^{th} layer of GCN, $\mathbf{H}^{(0)} = \mathbf{V}$, M_G denotes the GCN, $\mathbf{w}_G$ is the network parameter that is constantly updated during training. The layer-wise propagation rule of GCN is defined as follows

$$\mathbf{H}^{(l+1)} = \sigma(BN(\widehat{\mathbf{A}}\mathbf{H}^{(l)}\mathbf{w}_G^{(l)})), \tag{9}$$

where σ represents the softplus activation function, BN represents batch normalization, $\mathbf{w}_G^{(l)}$ is the trainable weight matrix of the l^{th} layer, and the size of the matrix is related to the number of input and output channels. Based on the above propagation rules, the GCN continuously learns the dependencies between nodes and uses the information in the adjacency matrix to aggregate the features of itself and its neighbors to extract richer features.

2) Quality Prediction: Here, we fuse the multi-level feature matrices $\mathbf{H}^{(l)}(l \in \{1, 2, 3, 4\})$ output by GCN. Above all, the first dimension of the first three feature matrices is averaged pooling, so as to imitate the HVS to aggregate the feature information of different projection images. Then, the 1-dimensional perceptual feature matrix is obtained through the dimension reduction of the fully connected layer. Since the number of channels of the feature matrix output by the last layer is already 1, the average pooling layer and the fully connected layer are used for the feature information of the first dimension of the matrix, so as to enhance the diversity of features. The final obtained multi-level fusion feature matrix $\bar{H}$ is described as follows

$$\begin{aligned} \bar{\mathbf{H}} =& L(\alpha(\mathbf{H}^{(1)})) \oplus L(\alpha(\mathbf{H}^{(2)})) \oplus \\ & L(\alpha(\mathbf{H}^{(3)})) \oplus L(\mathbf{H}^{(4)}) \oplus \alpha(\mathbf{H}^{(4)}), \end{aligned} \tag{10}$$

in which $\alpha(\cdot)$ represents average pooling, $L(\cdot)$ represents the fully connected layer. From this, we extract the multi-level fusion feature matrix of horizontal and vertical projection image groups respectively, which are denoted as $\bar{\mathbf{H}}_H$ and $\bar{\mathbf{H}}_V$. The two groups of features from different projection directions have different detail information and can complement each other. Finally, after fusing the two feature matrices, we use the fully connected layer to automatically assign weights to $\bar{\mathbf{H}}_H$ and $\bar{\mathbf{H}}_V$ to predict the perceptual quality score of the point cloud.

E. Network Training

For the whole network, we simultaneously input 20 images from the two projected image groups to jointly optimize the two branches. The loss function used to optimize the model is l_1, which can be defined as

$$l_1 = \frac{1}{n} \sum_{i=0}^{n} |y_i - \bar{y}_i|, \tag{11}$$

$$\bar{y}_i = Q(\mathbf{P}, \widehat{\mathbf{A}}; \mathbf{w}_F), \tag{12}$$

where n indicates batch size, y_i and $\bar{y}_i$ indicates the i^{th} subjective quality score and objective prediction score in the batch respectively. $\bar{y}_i$ is extracted through the GC-PCQA network Q. $\mathbf{w}_F$ is the trainable parameters of the network, which are updated by minimizing l_1.

TABLE I
PERFORMANCE COMPARISON RESULTS ON SJTU-PCQA AND WPC DATABASES. ALL INDICATORS ADOPT ABSOLUTE VALUES FOR PERFORMANCE COMPARISON FOR BETTER VISIBILITY. THE FIRST, SECOND, AND THIRD OF THE FOUR INDICATORS ARE MARKED IN RED, BLUE AND GREEN, RESPECTIVELY.

Ref	Type	Metric	SJTU-PCQA				WPC			
			SRCC↑	PLCC↑	KRCC↑	RMSE↓	SRCC↑	PLCC↑	KRCC↑	RMSE↓
FR	PC-Based	$PSNR_{MSE,p2po}$	0.6002	0.7622	0.4917	1.4382	0.1607	0.2673	0.1147	20.6947
		$PSNR_{MSE,p2pl}$	0.5505	0.7381	0.4375	1.5357	0.1182	0.2879	0.0851	21.1898
		$PSNR_{HF,p2po}$	0.6744	0.7737	0.5217	1.4481	0.0557	0.3555	0.0384	20.8197
		$PSNR_{HF,p2pl}$	0.6208	0.7286	0.4701	1.6000	0.0989	0.3263	0.0681	21.11
		AS_{Mean}	0.5317	0.5297	0.3723	2.7129	0.2484	0.3397	0.1801	21.5013
		AS_{RMS}	0.5653	0.7156	0.4144	1.6550	0.2479	0.3347	0.1802	21.5325
		AS_{MSE}	0.5472	0.5115	0.3865	2.6431	0.2484	0.3397	0.1801	21.5013
		$PSNR_Y$	0.7871	0.8124	0.6116	1.3222	0.5823	0.6166	0.4164	17.9001
		PCQM	0.7748	0.8301	0.6152	1.2978	0.5504	0.6162	0.4409	17.9027
		PointSSIM	0.7051	0.7422	0.5321	1.5601	0.4639	0.5225	0.3394	19.3863
		GraphSIM	0.8853	0.9158	0.7063	0.9462	0.6217	0.6833	0.4562	16.5107
	Projection-Based	SSIM	0.8667	0.8868	0.6988	1.0454	0.6483	0.6690	0.4685	16.8841
		MS-SSIM	0.8738	0.8930	0.7069	1.0091	0.7179	0.7349	0.5385	15.3341
		IW-SSIM	0.8638	0.8932	0.6934	1.0268	0.7608	0.7688	0.5707	14.5453
		VIFP	0.8624	0.8977	0.6934	1.0173	0.7426	0.7508	0.5575	15.0328
RR	PC-Based	PCMRR	0.5622	0.6699	0.4091	1.7589	0.3605	0.3926	0.2543	20.9203
NR	PC-Based	3D-NSS	0.7819	0.7813	0.6023	1.7740	0.6309	0.6284	0.4573	18.1706
		ResSCNN	0.8328	0.8865	0.6514	1.0728	0.4362	0.4531	0.2987	20.2591
	Projection-Based	PQANet	0.7593	0.7998	0.5796	1.3773	0.6368	0.6671	0.4684	16.6758
		IT-PCQA	0.8286	0.8605	0.6453	1.1686	0.4329	0.4870	0.3006	19.896
		Ours	0.9108	0.9301	0.7546	0.8691	0.8054	0.8091	0.6246	13.3405

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we first introduce the adopted subject-rated databases and performance measures. Then, the implementation details are provided. Finally, we conduct extensive experiments and analyze the results to verify our proposed method, including both performance comparison and ablation test.

A. Databases and Performance Measures

1) Databases: We perform experiments on two publicly available 3D point cloud databases which consist of SJTU-PCQA [72] and WPC [73].

The SJTU-PCQA database includes nine pristine and 378 distorted point clouds generated from seven distortion types. Each distortion type corresponds to six distortion levels. The subjective scores are in the form of MOS values ranging from 1 to 10.

The WPC database has 20 original point clouds. For each reference point cloud, 37 distorted point clouds are created by simulating five distortion types (i.e., Downsample, Gaussian white noise, G-PCC(T), V-PCC, G-PCC(O)), leading to 740 distorted point clouds in total. Each distorted point cloud also relates to a MOS value. The range of MOS is [0, 100].

2) Performance Measures: We apply four measures to evaluate and compare different PCQA methods, including Spearman's Rank Correlation Coefficient (SRCC), Pearson's linear correlation coefficient (PLCC), Kendall Rank Correlation Coefficient (KRCC), and root mean squared error (RMSE). The SRCC and KRCC are used to measure the monotonicity, while PLCC and RMSE are used to evaluate the accuracy. Higher correlation coefficients and lower RMSE represent better performance. It should be noted that before calculating PLCC and RMSE, we utilize a five-parameter logistic function [74] which can be formulated as

$$y = \beta_1 (\frac{1}{2} - \frac{1}{1 + exp(\beta_2 (x - \beta_3))}) + \beta_4 x + \beta_5, \qquad (13)$$

where x is the raw predicted result of the PCQA metric. y indicates the mapped objective quality score through the five-parameter logistic function, and $\beta_i (i \in \{1, 2, \cdots, 5\})$ are the fitting parameters.

B. Implementation Details

In our experiments, we employ PyTorch as the deep learning framework and the computer operating system is Ubuntu18.04. Moreover, the GPU is used to accelerate the training and testing procedures. The adaptive moment estimation optimizer (Adam) [75] is used for model training. We set batch size and initial learning rate as 32 and 1e-3, respectively. Additionally, the learning rate is reduced to 0.5 times of the original one every 10 epochs until the final convergence.

The network F is trained for 50 epochs and the training terminates early when there is no further optimized w_F for 20 epochs. Meanwhile, we exploit data augmentation methods such as horizontal and vertical flipping to enhance the generalization ability of the proposed network.

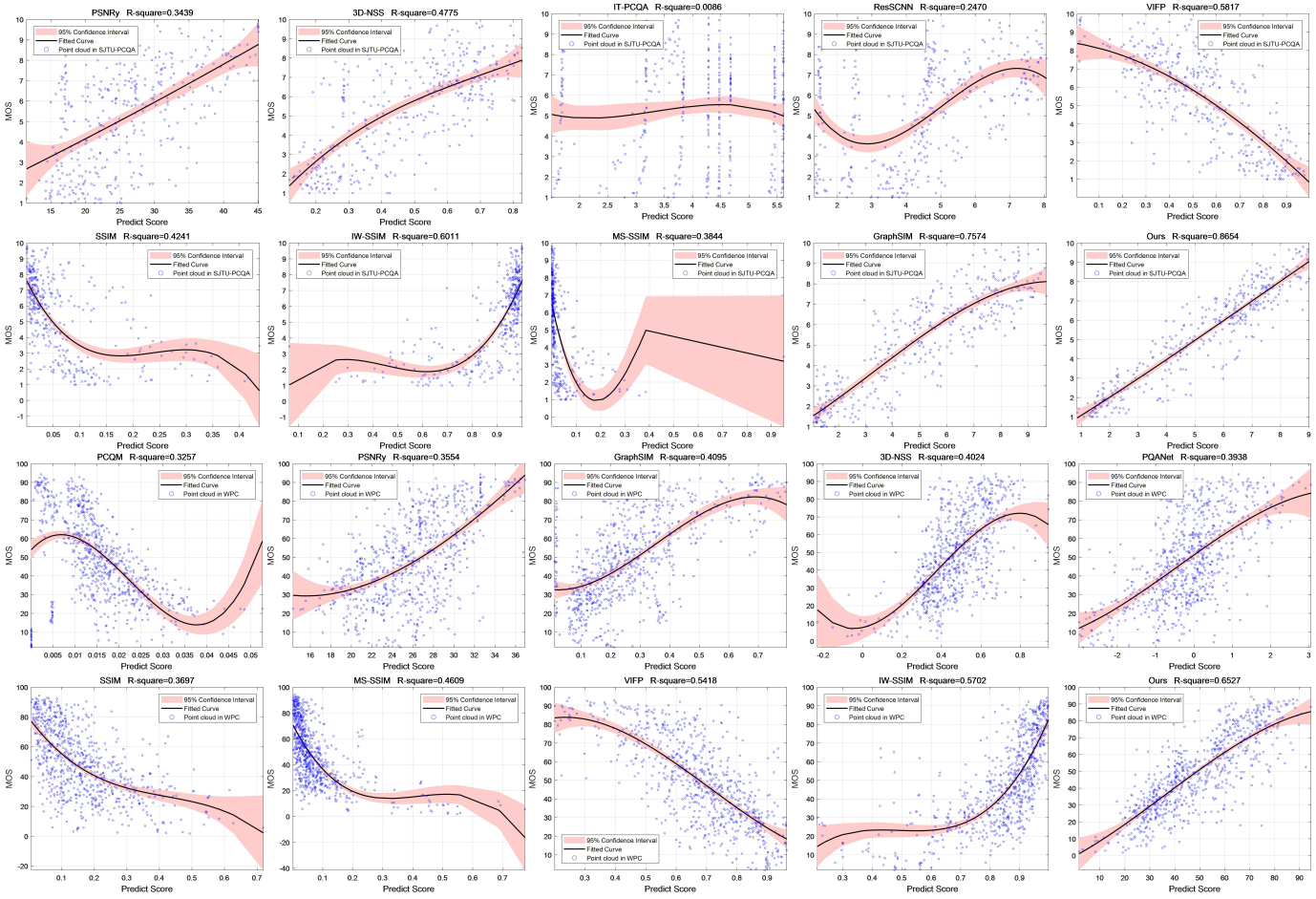

Fig. 4. Scatter plot between objective prediction scores and MOS for the top ten PCQA metrics in the experiment. The X-axis is the objective prediction score of the PCQA metric, and the Y-axis is the corresponding MOS. The first ten figures are the results on SJTU-PCQA database, from top left to bottom right are $PSNR_Y$ [42], 3D-NSS [18], IT-PCQA [24], ResSCNN [20], VIFP [59], SSIM [56], IW-SSIM [58], MS-SSIM [57], GraphSIM [49] and our proposed method. The last ten figures are the results on the WPC database, from top left to bottom right are PCQM [45], $PSNR_Y$ [42], GraphSIM [49], 3D-NSS [18], PQANet [22], SSIM [56], MS-SSIM [57], VIFP [59], IW-SSIM [58] and our proposed method. R-Square score, 95% confidence interval, and fitted curve are calculated for each scatter plot.

In addition, the k-fold cross-validation strategy is used for the performance test. For each point cloud quality database, $\frac{K-1}{K}$ distorted samples are randomly selected from the database as the train sets, and the rest point clouds are used as the test sets. Specifically, we choose K equalling to 9 and 5 for the SJTU-PCQA and WPC databases, respectively. The final results can be obtained by averaging the performance values from K times.

C. Performance Comparison

We compare our proposed GC-PCQA with 20 state-of-the-art quality assessment methods. As mentioned in Section II, existing PCQA metrics can be divided into two types: PC-based metrics and projection-based metrics. PC-based metrics are directly evaluated from 3D point clouds, including $PSNR_{MSE,p2po}$ [38], $PSNR_{HF,p2po}$ [38], $PSNR_{MSE,p2pl}$ [39], $PSNR_{HF,p2pl}$ [39], AS_{Mean} [40], AS_{RMS} [40], AS_{MSE} [40], $PSNR_Y$ [42], PCQM [45], PointSSIM [46], GraphSIM [49], PCMRR [52], 3D-NSS [18], and ResSCNN [20]. The projection-based metrics operate on the projected 2D images of point clouds, including SSIM [56], MS-SSIM [57], IW-SSIM [58], VIFP [59], PQANet [22], and IT-PCQA [24]. It is worth mentioning that the performance results of SSIM, MS-SSIM, IW-SSIM, and VIFP are the average values from six perpendicular projections [72], [76].

The performance comparison results on SJTU-PCQA and WPC databases are shown in Table I. From the table, we can draw several conclusions: (1) Compared to existing quality assessment methods, the proposed GC-PCQA achieves the best performance on both databases, which demonstrates the effectiveness of the proposed method. To be specific, the SRCC score of GC-PCQA is 0.0255 higher than that of the second-place GraphSIM on the SJTU-PCQA database, and 0.0446 higher than that of the second-place IW-SSIM on the WPC database. (2) Since the WPC database has more point cloud data and more complex distortions, the performance of PCQA metrics on the WPC database shows a significant degradation compared to that on the SJTU-PCQA database. For example, ResSCNN performs well on the SJTU-PCQA database, but its SRCC decreases by 0.397 on the WPC database. Compared with other comparison methods, our proposed metric achieves promising results on both databases without excessive

TABLE II

SRCC PERFORMANCE EVALUATION OF EXISTING PCQA METRICS BASED ON POINT CLOUD CONTENT AND DISTORTION TYPE IS PERFORMED ON THE SJTU-PCQA DATABASE. ABSOLUTE SRCC IS USED FOR COMPARISON TO OBTAIN BETTER VISIBILITY. THE LETTERS A-R IN THE TABLE STAND FOR $PSNR_{MSE,p2po}$, $PSNR_{MSE,p2pl}$, $PSNR_{HF,p2po}$, $PSNR_{HF,p2pl}$, AS_{RMS}, $PSNR_Y$, PCQM, POINTSSIM, GRAPHSIM, SSIM, MS-SSIM, IW-SSIM, VIFP, PCMRR, 3D-NSS, RESSCNN, PQANET, IT-PCQA AND OUR PROPOSED METHOD IN TURN. THE FIRST, SECOND, AND THIRD PLACES IN THE SPCC INDICATOR ARE MARKED IN RED, BLUE AND GREEN, RESPECTIVELY.

	Subset	FR													RR	NR				
		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	Ours
Content	Redandblack	0.6196	0.5943	0.7421	0.6819	0.5799	0.7478	0.8024	0.6670	0.8702	0.8603	0.8718	0.8911	0.8885	0.6506	0.8647	0.8003	0.8603	0.8557	0.9057
	Romanoillamp	0.4247	0.3617	0.7457	0.6032	0.6022	0.4278	0.5145	0.5150	0.8525	0.7509	0.7869	0.7939	0.7882	0.6044	0.6885	0.6193	0.7509	0.7248	0.9041
	Loot	0.6738	0.6405	0.7447	0.6391	0.4817	0.7875	0.8426	0.7299	0.8868	0.8693	0.8809	0.8846	0.8619	0.6770	0.8890	0.8780	0.8693	0.8778	0.9481
	Soldier	0.6781	0.6478	0.7493	0.6329	0.5404	0.8336	0.8684	0.7718	0.9118	0.8917	0.8843	0.8843	0.8744	0.5809	0.8731	0.9123	0.8917	0.8050	0.9253
	ULB Unicorn	0.7085	0.6082	0.8500	0.8081	0.4773	0.8687	0.7496	0.5715	0.8597	0.9084	0.8981	0.8548	0.8514	0.5148	0.4101	0.8364	0.9084	0.9129	0.9109
	Longdress	0.6640	0.6437	0.7885	0.7096	0.5704	0.9326	0.8896	0.8608	0.9499	0.9245	0.9191	0.8710	0.8976	0.6474	0.9005	0.8650	0.9245	0.8243	0.9441
	Statue	0.5678	0.5362	0.5883	0.5652	0.6291	0.8241	0.7483	0.7391	0.8744	0.8578	0.8663	0.8428	0.8637	0.4181	0.8520	0.9002	0.8578	0.8757	0.8633
	Shiva	0.4129	0.4074	0.1168	0.2689	0.7057	0.8375	0.8060	0.7896	0.8595	0.8968	0.8914	0.8744	0.8903	0.4884	0.8198	0.8599	0.8968	0.8243	0.8866
	Hhi	0.6526	0.5150	0.7443	0.6785	0.5012	0.8242	0.7524	0.7010	0.9028	0.8409	0.8658	0.8773	0.8462	0.4785	0.7394	0.8240	0.8409	0.7577	0.9089
Distortion	OT	0.4407	0.4407	0.3788	0.3524	0.5210	0.3068	0.6495	0.7108	0.7049	0.2198	0.2712	0.3382	0.3743	0.1800	0.4068	0.1683	0.0883	0.0189	0.8892
	CN	NaN	NaN	NaN	NaN	NaN	0.5588	0.6070	0.7660	0.7779	0.6283	0.6453	0.7531	0.7429	0.7157	0.1480	0.2265	0.5507	0.0655	0.9021
	DS	0.4495	0.4489	0.6847	0.3286	0.3653	0.4697	0.6990	0.8500	0.8654	0.3246	0.4718	0.4535	0.4546	0.1489	0.5051	0.4292	0.2958	0.0556	0.8918
	D+C	0.5735	0.5979	0.7619	0.7499	0.4025	0.7397	0.8014	0.7449	0.8846	0.5062	0.6281	0.6661	0.6932	0.6120	0.5895	0.5158	0.4899	0.0468	0.9500
	D+G	0.6779	0.7058	0.7423	0.7196	0.8915	0.5413	0.7476	0.9288	0.8833	0.6920	0.7589	0.8222	0.7989	0.7439	0.7442	0.5263	0.5033	0.0411	0.9664
	GGN	0.7008	0.7144	0.7453	0.7328	0.9376	0.5727	0.7143	0.9027	0.9064	0.7436	0.7783	0.8324	0.8436	0.7813	0.8435	0.4497	0.3771	0.0798	0.9546
	C+G	0.7577	0.7758	0.8205	0.8025	0.9241	0.6692	0.7078	0.7991	0.9334	0.7307	0.7948	0.8406	0.8463	0.8329	0.8645	0.5523	0.6137	0.1044	0.9681

TABLE III

SRCC PERFORMANCE EVALUATION OF EXISTING PCQA METRICS BASED ON POINT CLOUD CONTENT AND DISTORTION TYPE IS PERFORMED ON THE WPC DATABASE. ABSOLUTE SRCC IS USED FOR COMPARISON TO OBTAIN BETTER VISIBILITY. THE LETTERS REPRESENT THE SAME PCQA METRICS AS THE TABLE ABOVE. THE FIRST, SECOND, AND THIRD PLACES IN THE SPCC INDICATOR ARE MARKED IN RED, BLUE AND GREEN, RESPECTIVELY.

	Subset	FR													RR	NR				
		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	Ours
Content	Bag	0.6669	0.5751	0.4363	0.4365	0.4325	0.8051	0.5955	0.4829	0.7164	0.7300	0.7584	0.7309	0.7093	0.6069	0.7731	0.1603	0.3504	0.6174	0.7587
	Banana	0.6471	0.5691	0.1933	0.2033	0.3147	0.6211	0.4649	0.2202	0.5045	0.8011	0.7677	0.7790	0.7771	0.5287	0.6524	0.2475	0.6949	0.2485	0.5503
	Biscuits	0.5252	0.4160	0.3085	0.3368	0.3505	0.7764	0.6245	0.5816	0.7198	0.9173	0.9500	0.7992	0.7416	0.4310	0.6645	0.4765	0.6147	0.3570	0.7468
	Cake	0.3074	0.1798	0.1724	0.1796	0.0609	0.5180	0.4566	0.3177	0.4251	0.7390	0.7691	0.6534	0.6477	0.3070	0.4547	0.4467	0.5835	0.7300	0.8532
	Cauliflower	0.3501	0.2058	0.0918	0.1653	0.1781	0.5927	0.4903	0.4237	0.5529	0.8004	0.8608	0.8182	0.7008	0.4187	0.5517	0.5095	0.6238	0.0593	0.9239
	Flowerpot	0.6509	0.5298	0.4348	0.4515	0.3629	0.6385	0.5875	0.3784	0.6609	0.8303	0.9066	0.9047	0.8954	0.0477	0.6958	0.4900	0.2357	0.8127	0.7591
	GlassesCase	0.5845	0.4390	0.2020	0.3238	0.4288	0.7826	0.5861	0.5258	0.6546	0.7617	0.7577	0.7304	0.7459	0.3883	0.4790	0.2003	0.7674	0.7750	0.8826
	HoneydewMelon	0.4890	0.3299	0.2768	0.2300	0.3228	0.6740	0.4500	0.5609	0.7248	0.8549	0.8917	0.9180	0.8279	0.5742	0.7229	0.4026	0.7418	0.7352	0.7387
	House	0.5866	0.4483	0.3429	0.3434	0.4522	0.7798	0.5880	0.5590	0.7373	0.7788	0.7793	0.7357	0.7200	0.4905	0.7646	0.4780	0.8668	0.4201	0.9343
	Litchi	0.5109	0.4291	0.3478	0.3204	0.3554	0.7027	0.5965	0.6422	0.6958	0.7748	0.8623	0.7496	0.7018	0.4839	0.8113	0.1994	0.7207	0.0868	0.8879
	Mushroom	0.6396	0.5156	0.3486	0.3105	0.2911	0.6550	0.5725	0.5443	0.6802	0.7821	0.8781	0.8160	0.7897	0.2556	0.8153	0.0754	0.5835	0.3570	0.8608
	PenContainer	0.7720	0.6688	0.2159	0.3635	0.5465	0.7328	0.6394	0.5948	0.8250	0.8954	0.8758	0.8485	0.8397	0.6830	0.7809	0.5676	0.6470	0.7859	0.8928
	Pineapple	0.3777	0.2785	0.1376	0.1831	0.2155	0.7217	0.6427	0.5386	0.6401	0.7307	0.7805	0.5856	0.6441	0.4011	0.6074	0.5275	0.6318	0.5913	0.8862
	PingpongBat	0.5924	0.4984	0.4958	0.4357	0.4521	0.5428	0.5783	0.6051	0.7697	0.8054	0.8812	0.7570	0.7539	0.5092	0.6935	0.3518	0.6358	0.4737	0.8760
	PuerTea	0.6069	0.4746	0.1173	0.0384	0.4734	0.7639	0.5685	0.4139	0.7999	0.8917	0.8668	0.8359	0.7866	0.4308	0.4763	0.1456	0.7359	0.5467	0.6584
	Pumpkin	0.4947	0.3423	0.3092	0.3068	0.3220	0.6901	0.5934	0.5699	0.6517	0.9111	0.8156	0.9042	0.8976	0.3241	0.5768	0.4052	0.7857	0.5536	0.8435
	Ship	0.7464	0.6267	0.3404	0.5158	0.4943	0.7786	0.5434	0.4488	0.7558	0.8973	0.8578	0.8340	0.8013	0.4400	0.6935	0.6612	0.5349	0.3777	0.8054
	Statue	0.8040	0.6707	0.2450	0.4487	0.4900	0.7001	0.5714	0.5085	0.7390	0.8985	0.9372	0.9099	0.8950	0.1811	0.6368	0.5782	0.3762	0.4976	0.8743
	Stone	0.6219	0.5129	0.3551	0.3424	0.3649	0.7115	0.6475	0.6126	0.1920	0.8426	0.8881	0.8587	0.8196	0.3632	0.6968	0.2122	0.8234	0.1790	0.8222
	ToolBox	0.3937	0.2969	0.1972	0.1884	0.2984	0.8706	0.6304	0.4927	0.7935	0.7821	0.8255	0.8056	0.7411	0.5239	0.5806	0.5026	0.8653	0.4694	0.9336
Distortion	Downsampling	0.4815	0.3251	0.5356	0.4879	0.2465	0.5542	0.4537	0.8319	0.7903	0.8234	0.8834	0.8822	0.8828	0.7407	0.7508	0.2899	0.7234	0.3327	0.8148
	Gaussian noise	0.6155	0.6194	0.6149	0.6150	0.6844	0.7644	0.8775	0.5844	0.7469	0.6264	0.7118	0.8560	0.8847	0.7762	0.7460	0.5459	0.7938	0.1718	0.8533
	G-PCC (T)	0.3451	0.3568	0.2811	0.3085	0.1342	0.5916	0.7775	0.6745	0.7457	0.4669	0.6042	0.6742	0.6304	0.2702	0.5947	0.2531	0.4710	0.1987	0.8038
	V-PCC	0.1602	0.1992	0.2051	0.2370	0.3877	0.3203	0.5534	0.3546	0.5989	0.5141	0.5812	0.7063	0.7410	0.2966	0.3927	0.1028	0.0045	0.0090	0.6830
	G-PCC (O)	NaN	NaN	NaN	NaN	0.0350	0.8072	0.8944	0.7917	0.8258	0.5290	0.7214	0.7128	0.7116	0.6468	0.2891	0.0247	0.4204	0.1180	0.8874

performance degradation. Therefore, GC-PCQA has stronger learning ability and can maintain better performance in more complex distorted point cloud databases. (3) Four and five of the top-five PCQA metrics belong to projection-based metrics for the SJTU-PCQA and WPC databases, respectively. This proves the effectiveness of converting point clouds into multiple projected 2D images for quality assessment. That is, the projection can help evaluate the visual quality of 3D point clouds. To get a more intuitive understanding of the performance for different PCQA metrics, we draw a scatter plot regarding the predicted scores and MOS, as shown in Fig. 4. In this figure, we adopt the fitted curve to describe the relationship between objective predictions and MOS values, and the closer to the diagonal line, the better. Besides, R-Square is used to evaluate the goodness of the model fit, and the value range is [0, 1]. The closer the value is to 1, the better the model fitted the data. It can be clearly observed that the proposed GC-PCQA achieves the best R-Square scores on both SJTU-PCQA and WPC databases, indicating that the objective predictions of the model fit the MOS well.

Since point cloud quality databases involve various contents and distortion types, it is interesting to test the performance of existing PCQA metrics regarding each individual point cloud content and distortion type. We report the experimental

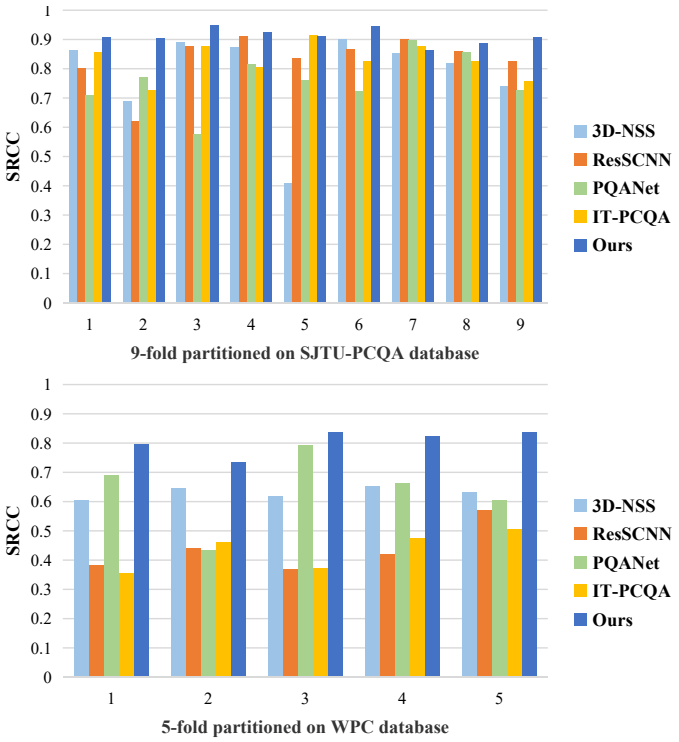

Fig. 5. Performance comparison results of NR method in different partition test databases of SJTU-PCQA and WPC.

TABLE IV
PERFORMANCE CONTRIBUTION RESULTS OF PROJECTED IMAGES AND GCN ON SJTU-PCQA AND WPC DATABASES. THE BEST PERFORMANCE IS INDICATED IN **BOLD.**

Model	SJTU-PCQA		WPC	
	SRCC↑	PLCC↑	SRCC↑	PLCC↑
FC Layer with P_H	0.8667	0.8941	0.6819	0.6989
FC Layer with P_V	0.8705	0.8889	0.6934	0.7000
FC Layer with P_{HV}	0.8849	0.8944	0.7075	0.7126
GCN with P_{HV}	**0.9108**	**0.9301**	**0.8054**	**0.8091**

TABLE V
PERFORMANCE COMPARISON RESULTS OF DIFFERENT BACKBONES ON SJTU-PCQA AND WPC DATABASES. THE BEST PERFORMANCE IS INDICATED IN **BOLD.**

Backbone	SJTU-PCQA		WPC	
	SRCC↑	PLCC↑	SRCC↑	PLCC↑
VGG16	0.8876	0.9148	0.7322	0.7444
ResNet50	0.8930	0.9138	0.7210	0.7288
ResNet101	**0.9108**	**0.9301**	**0.8054**	**0.8091**
InceptionV3	0.8605	0.9008	0.6606	0.6688
DenseNet121	0.8764	0.9124	0.6962	0.7063
MobileNetV2	0.8829	0.9155	0.7382	0.7282
EfficientNet-B0	0.8877	0.9094	0.7499	0.7528

results in Tables II and III. From the two tables, we can observe that: (1) On the smaller database subset, compared with the NR method, the FR method achieves more top-three performances due to the existence of the reference point cloud. (2) Our proposed method achieves the top-three SRCC performance for 14 times among the 16 subset experiments on the SJTU-PCQA database, including the best performance 12 times. Meanwhile, in the 25 subset experiments on the WPC database, it achieves the top-three SRCC performance 15 times, including the best performance 8 times. Thus, our model can deliver superior performance, which further demonstrates the effectiveness of our proposed method. (3) The projection-based metric achieves the best performance 15 and 23 times for the SJTU-PCQA and WPC databases. This also proves the advantages of the projection method in PCQA.

Considering that NR methods do not rely on original reference point clouds, they are more suitable in practical application scenarios. Therefore, here we perform k-fold cross-validation for NR methods and visualize all the results on the SJTU-PCQA and WPC databases, as shown in Fig. 5. The abscissa represents the test database with different partitions, and the ordinate represents the value of SRCC. From this figure, it can be observed that the SRCC results of our proposed method are better than the other NR methods on seven folds out of a total number of nine folds on the SJTU-PCQA database, and the results of the remaining two folds are not far from the best performance. Moreover, it consistently outperforms other NR methods over all folds on the more complex WPC database. This shows that our proposed NR method can predict the perceptual quality of point clouds more accurately and maintain a certain performance in complex visual environments.

D. Ablation Study

1) Contributions of Multi-view Projection and GCN: Our proposed method uses multi-view projection and GCN to improve the network performance. To measure the contributions of these operations, we conduct ablation experiments while keeping the default experimental settings unchanged. The experimental results are shown in Table IV. Here, P_H and P_V represent the horizontally and vertically projected image groups, respectively. P_{HV} means using both horizontally and vertically projected image groups. On the one hand, both the horizontal and vertical projection image groups contribute to the model performance, and the combination of the two can achieve better results. On the other hand, GCN can help the model better aggregate the feature information of multi-view projections, and obtain larger performance improvement on the more complex WPC database.

2) Backbone Comparison: In the network, 2D-CNN plays an important role as a feature extractor of images. To compare the performance of different backbones, we conduct experiments by replacing the backbones while keeping the other modules unchanged. Specifically, the backbones used for comparison include VGG16 [77], ResNet50 [64], ResNet101 [64], InceptionV3 [78], DenseNet [79], MobileNetV2 [80], and EfficientNet [81]. The experimental results are presented in Table V. It can be seen that the networks with other backbones for feature extraction are still competitive with state-of-the-art PCQA metrics, and there will be no significant fluctuations. Among them, ResNet101 has the best performance when used as the backbone.

TABLE VI
PERFORMANCE COMPARISON OF DIFFERENT RSS ON THE SJTU-PCQA AND WPC DATABASES. THE BEST PERFORMANCE IS INDICATED IN **BOLD.**

Rotation Stride (RS)	SJTU-PCQA		WPC	
	SRCC↑	PLCC↑	SRCC↑	PLCC↑
24°	0.8963	0.9150	0.7702	0.7625
36°	**0.9108**	**0.9301**	**0.8054**	**0.8091**
48°	0.8936	0.9159	0.7474	0.7447
60°	0.8923	0.9178	0.7533	0.7430

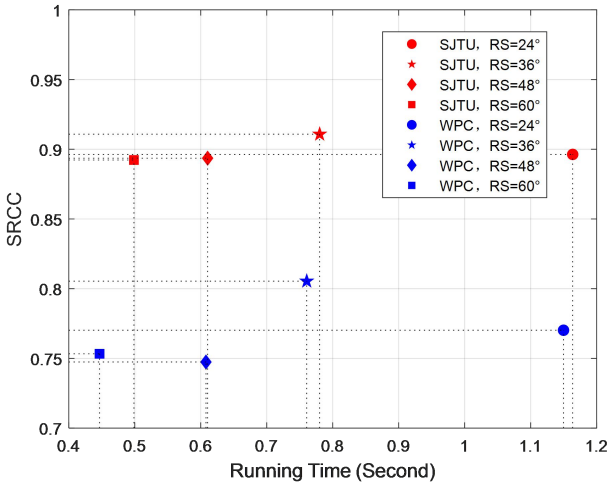

Fig. 6. Running time vs. SRCC performance of GC-PCQA under four different rotation strides on SJTU-PCQA and WPC databases. The red and blue points represent the results of the PCQA metric on the SJTU-PCQA and WPC databases, respectively. RS represents the rotation stride.

3) RS Comparison: We also conduct experiments for choosing an appropriate RS for projection to obtain a more effective multi-view projection image group P. In the experiments, we choose four RSs, i.e., 24°, 36°, 48°, 60° for verification. The performance comparison results are shown in Table VI. The image numbers obtained by projection through different strides are also different. For example, based on the RS of 36°, the final number of multi-view projection images is $360/36 = 10$. According to the experimental results, the performance gap between different RSs is not large, and the best performance is achieved when the RS is 36°.

To compare the inference efficiency of our NR-PCQA metric under four different RSs, we report the average running time for processing a point cloud. The details are shown in Fig. 6. It can be observed that although larger strides can produce fewer images and speed up network training, more projected images can bring more feature information and get better results. When the RS is reduced to 36°, the network performance reaches the limit. Further reducing the stride will negatively affect the performance. This is because the projected image at this time has been able to cover most contents of the point cloud, and more images will lead to a large number of redundancy feature information. In addition, when the RS is 60°, the network still achieves good performance while requiring only half of the inference time of the best configuration.

V. CONCLUSION

In this paper, we propose a new GCN-based NR-PCQA method. Our main inspiration comes from the fact that the HVS depends on a set of projected images from multiple viewpoints when perceiving 3D objects and the mutual dependencies among different projected images can be well modeled by GCN. Therefore, we first project the point cloud in the horizontal and vertical directions to obtain the projected image group under multiple views. Then, graph construction is performed on the projected image group and GCN is used to model the mutual dependencies between different projected 2D image contents to aggregate the feature information of images under different viewpoints, so as to imitate the viewing behavior of HVS and better measure the quality of point cloud. The experimental results on two benchmark datasets show that our proposed method has better performance than other state-of-the-art methods.

REFERENCES

[1] Q. Liu, H. Yuan, J. Hou, R. Hamzaoui, and H. Su, "Model-based joint bit allocation between geometry and color for video-based 3D point cloud compression," *IEEE Transactions on Multimedia*, vol. 23, pp. 3278–3291, 2021.

[2] Z. Huang, Y. Yu, J. Xu, F. Ni, and X. Le, "PF-Net: Point fractal network for 3D point cloud completion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 7659–7667.

[3] R. Chen, S. Han, J. Xu, and H. Su, "Visibility-aware point-based multi-view stereo network," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 10, pp. 3695–3708, 2021.

[4] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3D classification and segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 652–660.

[5] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep hierarchical feature learning on point sets in a metric space," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[6] G. G. Demisse, D. Aouada, and B. Ottersten, "Deformation-based 3D facial expression representation," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 14, no. 1s, pp. 1–22, 2018.

[7] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, "Point cloud rendering after coding: Impacts on subjective and objective quality," *IEEE Transactions on Multimedia*, vol. 23, pp. 4049–4064, 2020.

[8] Y. Cui, R. Chen, W. Chu, L. Chen, D. Tian, Y. Li, and D. Cao, "Deep learning for image and point cloud fusion in autonomous driving: A review," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 2, pp. 722–739, 2022.

[9] E. Alexiou, N. Yang, and T. Ebrahimi, "Pointxr: A toolbox for visualization and subjective evaluation of point clouds in virtual reality," in *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2020, pp. 1–6.

[10] S. Schwarz, M. Preda, V. Baroncini, M. Budagavi, P. Cesar, P. A. Chou, R. A. Cohen, M. Krivokuća, S. Lasserre, Z. Li *et al.*, "Emerging mpeg standards for point cloud compression," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 1, pp. 133–148, 2019.

[11] S. Gu, J. Hou, H. Zeng, and H. Yuan, "3D point cloud attribute compression via graph prediction," *IEEE Signal Processing Letters*, vol. 27, pp. 176–180, 2020.

[12] D. Group and et al., "Text of iso/iec cd 23090-5: Video-based point cloud compression," *ISO/IEC JTC1/SC29/WG11 Doc. N18030*, 2018.

[13] D. Group *et al.*, "Text of iso/iec cd 23090-9 geometry-based point cloud compression," *ISO/IEC JTC1/SC29/WG11 Doc. N18478*, 2019.

[14] Q. Liu, H. Yuan, R. Hamzaoui, and H. Su, "Coarse to fine rate control for region-based 3D point cloud compression," in *2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE, 2020, pp. 1–6.

[15] J. Xu, Z. Fang, Y. Gao, S. Ma, Y. Jin, H. Zhou, and A. Wang, "Point AE-DCGAN: A deep learning model for 3D point cloud lossy geometry compression," in *2021 Data Compression Conference (DCC)*. IEEE, 2021, pp. 379–379.

[16] S. Perry, H. P. Cong, L. A. da Silva Cruz, J. Prazeres, M. Pereira, A. Pinheiro, E. Dumic, E. Alexiou, and T. Ebrahimi, "Quality evaluation of static point clouds encoded using mpeg codecs," in *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2020, pp. 3428–3432.

[17] L. Hua, G. Jiang, M. Yu, and Z. He, "BQE-CVP: Blind quality evaluator for colored point cloud based on visual perception," in *2021 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*. IEEE, 2021, pp. 1–6.

[18] Z. Zhang, W. Sun, X. Min, T. Wang, W. Lu, and G. Zhai, "No-reference quality assessment for 3D colored point cloud and mesh models," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 11, pp. 7618–7631, 2022.

[19] A. Chetouani, M. Quach, G. Valenzise, and F. Dufaux, "Deep learning-based quality assessment of 3D point clouds without reference," in *2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE, 2021, pp. 1–6.

[20] Y. Liu, Q. Yang, Y. Xu, and L. Yang, "Point cloud quality assessment: Dataset construction and learning-based no-reference metric," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 19, no. 2s, pp. 1–26, 2023.

[21] W.-x. Tao, G.-y. Jiang, Z.-d. Jiang, and M. Yu, "Point cloud projection and multi-scale feature fusion network based blind quality assessment for colored point clouds," in *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, pp. 5266–5272.

[22] Q. Liu, H. Yuan, H. Su, H. Liu, Y. Wang, H. Yang, and J. Hou, "PQA-Net: Deep no reference point cloud quality assessment via multi-view projection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 12, pp. 4645–4660, 2021.

[23] R. Tu, G. Jiang, M. Yu, T. Luo, Z. Peng, and F. Chen, "V-PCC projection based blind point cloud quality assessment for compression distortion," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 7, no. 2, pp. 462–473, 2023.

[24] Q. Yang, Y. Liu, S. Chen, Y. Xu, and J. Sun, "No-reference point cloud quality assessment via domain adaptation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2022, pp. 21 179–21 188.

[25] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S. Y. Philip, "A comprehensive survey on graph neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 4–24, 2021.

[26] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[27] Y. Zhang and M. Rabbat, "A graph-CNN for 3D point cloud classification," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 6279–6283.

[28] Y. Li and Y. Tanaka, "Structural features in feature space for structure-aware graph convolution," in *2021 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2021, pp. 3158–3162.

[29] M. Xu, W. Dai, Y. Shen, and H. Xiong, "MSGCNN: Multi-scale graph convolutional neural network for point cloud segmentation," in *2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM)*. IEEE, 2019, pp. 118–127.

[30] L. Hansen, J. Diesel, and M. P. Heinrich, "Multi-kernel diffusion CNNs for graph-based learning on point clouds," in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, September 2018.

[31] C. Si, W. Chen, W. Wang, L. Wang, and T. Tan, "An attention enhanced graph convolutional LSTM network for skeleton-based action recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2019, pp. 1227–1236.

[32] S. Sun, T. Yu, J. Xu, W. Zhou, and Z. Chen, "GraphIQA: Learning distortion graph representations for blind image quality assessment," *IEEE Transactions on Multimedia*, 2022.

[33] Y. Huang, L. Li, Y. Yang, Y. Li, and Y. Guo, "Explainable and generalizable blind image quality assessment via semantic attribute reasoning," *IEEE Transactions on Multimedia*, 2022.

[34] J. Xu, W. Zhou, and Z. Chen, "Blind omnidirectional image quality assessment with viewport oriented graph convolutional networks," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 5, pp. 1724–1737, 2020.

[35] J. Fu, C. Hou, W. Zhou, J. Xu, and Z. Chen, "Adaptive hypergraph convolutional network for no-reference 360-degree image quality assessment," in *Proceedings of the 30th ACM International Conference on Multimedia*, 2022, pp. 961–969.

[36] I. Abouelaziz, A. Chetouani, M. El Hassouni, H. Cherifi, and L. J. Latecki, "Learning graph convolutional network for blind mesh visual quality assessment," *IEEE Access*, vol. 9, pp. 108 200–108 211, 2021.

[37] M. El Hassouni and H. Cherifi, "Learning graph features for colored mesh visual quality assessment," in *2022 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2022, pp. 3381–3385.

[38] R. Mekuria, Z. Li, C. Tulvan, and P. Chou, "Evaluation criteria for point cloud compression," *ISO/IEC MPEG*, no. 16332, 2016.

[39] D. Tian, H. Ochimizu, C. Feng, R. Cohen, and A. Vetro, "Geometric distortion metrics for point cloud compression," in *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2017, pp. 3460–3464.

[40] E. Alexiou and T. Ebrahimi, "Point cloud quality assessment metric based on angular similarity," in *2018 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2018, pp. 1–6.

[41] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, "A generalized hausdorff distance based quality metric for point cloud geometry," in *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2020, pp. 1–6.

[42] R. Mekuria, S. Laserre, and C. Tulvan, "Performance assessment of point cloud compression," in *2017 IEEE Visual Communications and Image Processing (VCIP)*. IEEE, 2017, pp. 1–4.

[43] I. Viola, S. Subramanyam, and P. Cesar, "A color-based objective quality metric for point cloud contents," in *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2020, pp. 1–6.

[44] G. Meynet, J. Digne, and G. Lavoué, "PC-MSDM: A quality metric for 3D point clouds," in *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2019, pp. 1–3.

[45] G. Meynet, Y. Nehmé, J. Digne, and G. Lavoué, "PCQM: A full-reference quality metric for colored 3D point clouds," in *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2020, pp. 1–6.

[46] E. Alexiou and T. Ebrahimi, "Towards a point cloud structural similarity metric," in *2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE, 2020, pp. 1–6.

[47] R. Diniz, P. G. Freitas, and M. C. Farias, "Towards a point cloud quality assessment model using local binary patterns," in *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2020, pp. 1–6.

[48] R. Diniz, P. G. Freitas, and et al., "Multi-distance point cloud quality assessment," in *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2020, pp. 3443–3447.

[49] R. Diniz, P. G. Freitas, and M. C. Farias, "Local luminance patterns for point cloud quality assessment," in *2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP)*. IEEE, 2020, pp. 1–6.

[50] Q. Yang, Z. Ma, Y. Xu, Z. Li, and J. Sun, "Inferring point cloud quality via graph similarity," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 6, pp. 3015–3029, 2022.

[51] Q. Yang, Y. Zhang, S. Chen, Y. Xu, J. Sun, and Z. Ma, "MPED: Quantifying point cloud distortion based on multiscale potential energy discrepancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 5, pp. 6037–6054, 2023.

[52] I. Viola and P. Cesar, "A reduced reference metric for visual quality evaluation of point cloud contents," *IEEE Signal Processing Letters*, vol. 27, pp. 1660–1664, 2020.

[53] Q. Liu, H. Yuan, R. Hamzaoui, H. Su, J. Hou, and H. Yang, "Reduced reference perceptual quality model with application to rate control for video-based point cloud compression," *IEEE Transactions on Image Processing*, vol. 30, pp. 6623–6636, 2021.

[54] W. Zhou, Q. Yang, Q. Jiang, G. Zhai, and W. Lin, "Blind quality assessment of 3D dense point clouds with structure guided resampling," *arXiv preprint arXiv:2208.14603*, 2022.

[55] S. Wolf and M. Pinson, "Reference algorithm for computing peak signal to noise ratio (psnr) of a video sequence with a constant delay," *ITU-T Contribution COM9-C6-E*, 2009.

[56] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[57] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, vol. 2. Ieee, 2003, pp. 1398–1402.

[58] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1185–1198, 2011.

[59] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.

[60] X. G. Freitas, R. Diniz, and M. C. Farias, "Point cloud quality assessment: unifying projection, geometry, and texture similarity," *The Visual Computer*, pp. 1–8, 2022.

[61] S. Chen, D. Tian, C. Feng, A. Vetro, and J. Kovačević, "Contour-enhanced resampling of 3D point clouds via graphs," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 2941–2945.

[62] S. Chen, D. Tian, C. Feng, A. Vetro, and et al., "Fast resampling of three-dimensional point clouds via graphs," *IEEE Transactions on Signal Processing*, vol. 66, no. 3, pp. 666–681, 2018.

[63] J. Qi, W. Hu, and Z. Guo, "Feature preserving and uniformity-controllable point cloud simplification on graph," in *2019 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2019, pp. 284–289.

[64] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 2016, pp. 770–778.

[65] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 2018, pp. 7132–7141.

[66] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018, pp. 3–19.

[67] A. G. Roy, N. Navab, and C. Wachinger, "Concurrent spatial and channel 'squeeze & excitation'in fully convolutional networks," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I*. Springer, 2018, pp. 421–429.

[68] S. Chaib, H. Liu, Y. Gu, and H. Yao, "Deep feature fusion for vhr remote sensing scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 8, pp. 4775–4784, 2017.

[69] T. Akilan, Q. J. Wu, A. Safaei, and W. Jiang, "A late fusion approach for harnessing multi-cnn model high-level features," in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2017, pp. 566–571.

[70] J. Yan, Y. Fang, L. Huang, X. Min, Y. Yao, and G. Zhai, "Blind stereoscopic image quality assessment by deep neural network of multi-level feature fusion," in *2020 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2020, pp. 1–6.

[71] C. Wu, X.-J. Wu, and J. Kittler, "Spatial residual layer and dense connection block enhanced spatial temporal graph convolutional network for skeleton-based action recognition," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, Oct 2019.

[72] Q. Yang, H. Chen, Z. Ma, Y. Xu, R. Tang, and J. Sun, "Predicting the perceptual quality of point cloud: A 3D-to-2D projection-based exploration," *IEEE Transactions on Multimedia*, vol. 23, pp. 3877–3891, 2021.

[73] H. Su, Z. Duanmu, W. Liu, Q. Liu, and Z. Wang, "Perceptual quality assessment of 3D point clouds," in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 3182–3186.

[74] V. Q. E. Group *et al.*, "Final report from the video quality experts group on the validation of objective models of video quality assessment, phase ii," *2003 VQEG*, 2003.

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

[76] W. Zhou, G. Yue, R. Zhang, Y. Qin, and H. Liu, "Reduced-reference quality assessment of point clouds via content-oriented saliency projection," *arXiv preprint arXiv:2301.07681*, 2023.

[77] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[78] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 2016, pp. 2818–2826.

[79] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, July 2017, pp. 4700–4708.

[80] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 2018, pp. 4510–4520.

[81] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning*, vol. 97. PMLR, 09–15 Jun 2019, pp. 6105–6114.