IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 16 January 2023, accepted 13 February 2023, date of publication 22 February 2023, date of current version 21 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3247191

RESEARCH ARTICLE

No-Reference 3D Point Cloud Quality Assessment Using Multi-View Projection and Deep Convolutional Neural Network

SALIMA BOURBIA[1,2], AYOUB KARINE[1], ALADINE CHETOUANI[3], (Senior Member, IEEE), MOHAMMED EL HASSOUNI[2], (Member, IEEE), AND MAHER JRIDI[1]

[1]L@bISEN, Vision-AD, ISEN Yncréa Ouest, 44470 Carquefou, France
[2]FLSH, FSR, Mohammed V University in Rabat, Morocco
[3]Laboratoire PRISME, Université d'Orléans, 45100 Orléans, France

Corresponding author: Salima Bourbia (salima.bourbia@isen-ouest.yncrea.fr)

This work was supported in part by the National Center for Scientific and Technical Research (CNRST), Morocco; and in part by the French Institute (FI), Morocco.

ABSTRACT Digital representation of 3D content in the form of 3D point clouds (PC) has gained increasing interest and has emerged in various computer vision applications. However, various degradation may appear on the PC during acquisition, transmission, or treatment steps in the 3D processing pipeline. Therefore, several Full-Reference, Reduced-Reference, and No-Reference metrics have been proposed to estimate the visual quality of PC. However, Full-Reference and Reduced-Reference metrics require reference information, which is not accessible in real-world applications, and No-Reference metrics still lack precision in evaluating the PC quality. In this context, we propose a novel deep learning-based method for No-Reference Point Cloud Quality Assessment (NR-PCQA) that aims to automatically predict the perceived visual quality of the PC without using the reference content. More specifically, in order to imitate the human visual system during the PC quality evaluation that captures the geometric and color degradation, we render the PC into different 2D views using a perspective projection. Then, the projected 2D views are divided into patches that are fed to a Convolutional Neural Network (CNN) to learn sophisticated and discriminative visual quality features for evaluating the local quality of each patch. Finally, the overall quality score of the PC is obtained by pooling the quality score patches. We conduct extensive experiments on three benchmark databases: ICIP2020, SJTU, and WPC, and we compare the proposed model to the existing Full-Reference, Reduced-Reference, and No-Reference state-of-the-art methods. Based on the experimental results, our proposed model achieves high correlations with the subjective quality scores and outperforms the state-of-the-art methods.

INDEX TERMS Point cloud, quality assessment, point cloud rendering, convolutional neural network (CNN).

I. INTRODUCTION

In the past years, the digital representation of 3D models has gained increased interest and has been used in prevalent 3D computer vision applications such as virtual and augmented reality, immersive communications, and cultural heritage [1], [2], [3], [4], [5], [6], [7], [8]. PC is considered one of the most widely used data for digital representation to model 3D realistic content. A PC is a set of unstructured points with geometric coordinates that represent point position and optional associated attributes related to the point appearance including color, curvatures, and opacity [9], [10].

As for images and videos, the PC objects may be affected by several factors from the point cloud processing pipeline (acquisition, representation, compression, and rendering) that could degrade their perceived visual quality. Therefore, it is

The associate editor coordinating the review of this manuscript and approving it for publication was Amin Zehtabian.

essential to develop effective methods that accurately assess the quality of PC and preserve the quality of the user experience. Two families of methods are usually adopted to evaluate PC degradation [11], [12], [13]: subjective and objective metrics. The subjective ones are based on human judgment for the evaluation, which makes it cumbersome and expansive in practical real-world situations. However, this type of method is used to construct a PC annotated database. In this case, the annotation refers to the subjective quality scores (ground truth) of the PC, often called Mean Opinion Scores (*MOS*). On the other hand, the objective methods automatically predict the perceived quality score that should be highly correlated with the ground truth.

As for 2D images, we can classify the objective point cloud quality assessment (*PCQA*) methods into three branches: Full-Reference (*FR*), Reduced-Reference (*RR*) and No-Reference (*NR*). In NR-PCQA methods, the perceived quality of the visual stimuli (distorted point clouds) is assessed without the need for the reference PC, while in the FR-PCQA and RR-PCQA methods, the quality is estimated through a full or partial (PC features) information from the reference PC, respectively. However, in practical situations, the reference PC is not often available. Consequently, the NR-PCQA methods can be considered a fruitful solution.

In this context, due to the great success of deep learning in several computer vision applications, CNNs are widely adopted in the NR image quality assessment task [14], [15], [16], [17], [18], [19], [20]. However, contrary to the 2D images that have regular grids and spatial structures, PCs are unordered and unstructured. Based on these considerations, PCs cannot be directly processed with CNNs, which uses discrete convolutions. To circumvent this limitation and to mimic the human visual system during the subjective evaluation, we propose to generate 2D projections from each PC object on multiple views. The projected images are then split into vertical overlapping patches to discard the useless information in the background. After that, we feed the extracted patches into a CNN that automatically and hierarchically learns discriminant visual features in order to predict the visual quality score. Finally, the quality score of the PC is computed by aggregating the score of each patch.

The main contributions of our paper are as follows:

1) We propose a novel multi-view deep learning-based method that aims to automatically evaluate the perceived visual quality of the PC without relying on the reference content.
2) We conduct a large study on the impact of different model parameters, such as the number of rendering views, the type of the CNN feature extractor, and the spatial pooling technique used to aggregate the predicted quality scores of the views. This study is not taken into consideration in the projection-based NR state-of-the-art metrics.
3) We compare the proposed method with various Full-Reference, Reduced-Reference, and No-Reference state-of-the-art methods on three PCQA benchmark databases: ICIP2020, SJTU, and WPC. Extensive experiments show that the proposed method outperforms all the NR and RR methods and is competitive or even better than the FR methods. Moreover, our method shows better performance over the 3 databases, and on mixed and individual types of distortion.

The remainder of this paper is organized as follows. We present in Section II the state-of-the-art of 3D point cloud quality assessment. After that, the proposed approach is described in Section III. Finally, Section IV presents the experimental results that are followed by conclusions and future works.

II. RELATED WORK

In the literature, most of the existing PCQA methods are Full-Reference (FR). They can be broadly classified into point-based, feature-based, and projected-based metrics.

Point-based metrics establish a correspondence between the reference PC and its degraded version. After that, a distance is adopted to quantify the visual quality. Concomitantly, point-based approaches can be divided into two categories: geometry-based and joint-geometry-and-color-based. The sub-categories of geometry-based metrics are Point-to-Point [21], Point-to-Plane [22], Plane-to-Plane [23], and Point-to-Distribution [24]. In [21], Mekuria et al. proposed a Point-to-Point metric where the distance between the corresponding points of the reference and the distorted PC is measured using the Mean Squared Error (MSE) or Hausdorff distance to evaluate the geometric PC quality. In the same vein, Tian et al. [22] proposed a Point-to-Plane metric that projects the point-to-point distance along the normal vector of the reference PC. These metrics predict the geometric distortions accurately, however, they fall short when dealing with structure loss. To solve this issue, Alexiou et al. [23] proposed a Plane-to-Plane metric based on measuring the angular similarity between the tangent planes of the distorted PC and its reference. However, the performance of this latter depends upon the used method to estimate the normal, which is error-prone. As a Point-to-Distribution method, Javaheri et al. [24] computed the Mahalanobis distance between the distributions of reference and degraded PCs. The joint-geometry-and-color-based methods assess the quality based on a combination of color and geometry information. Javaheri *et al.* [25] proposed to fuse the geometry and the color distortion after calculating them independently.

For the feature-based metrics, the quality score is calculated through the distance between the attributes and/or geometry features of original and distorted PC objects. Meynet et al. [26] adapted the Structural Similarity (SSIM) metric [27] to evaluate the PC quality by capturing the local curvature statistics changes between the reference and distorted PC. The same authors proposed the so-called Point Cloud Quality Metric (PCQM) [28], which aggregates a set of geometry-based and color-based features through logistic regression. Similarly, in [29], the authors linearly combined

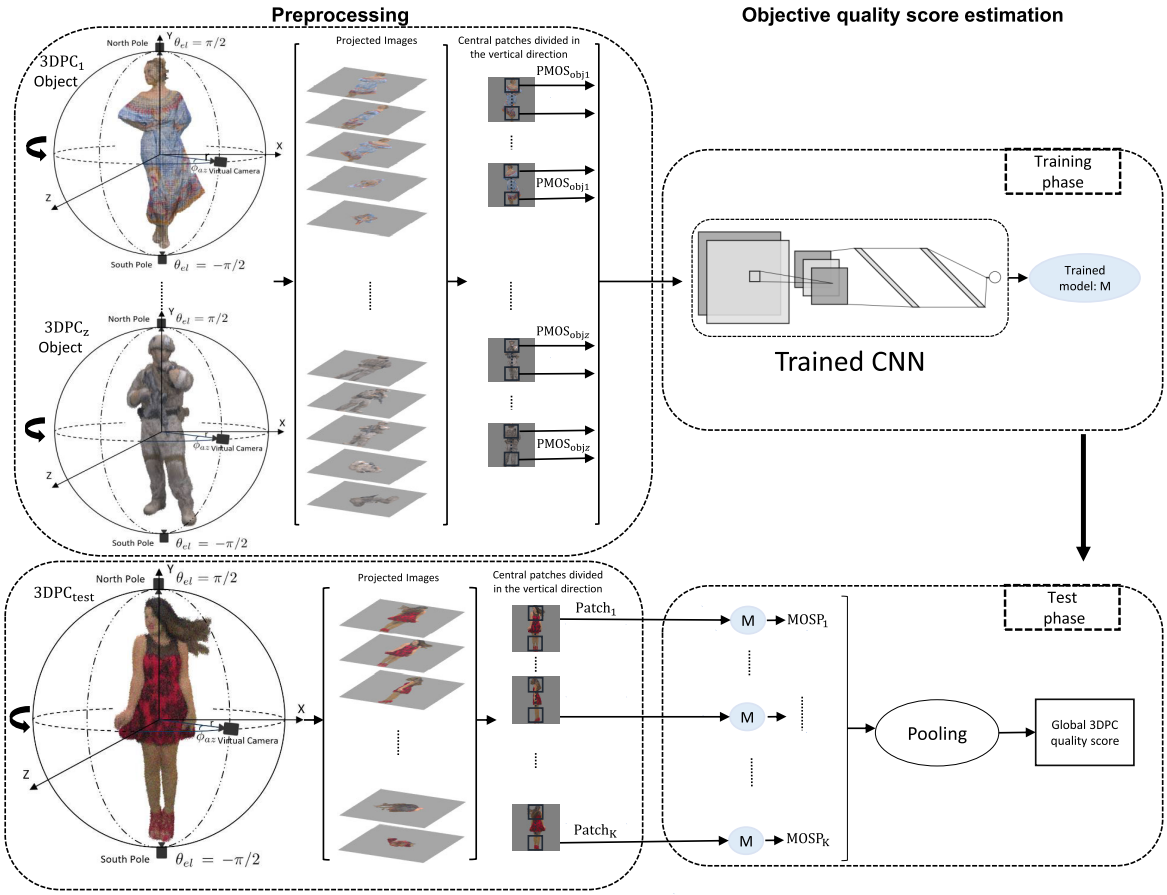


FIGURE 1. Flowchart of the proposed NR-PCQA method. It is composed of two main steps: Preprocessing and objective quality score estimation of the PC. In the first step, we represent the PC with different 2D viewpoints. From this later, we extract central overlapping patches using a sliding step in the vertical direction to discard the useless information in the background. In the second step, we feed the extracted patches into a deep CNN network to learn discriminative and meaningful visual features in the training phase. Afterward, in the test phase, we predict the local quality score of each extracted degraded patch using the learned CNN model (M) in the training phase. Finally, the global PC quality score is obtained by pooling all the patches local quality scores.

the color statistic and Point-to-Plane metric extracted from the reference and altered PC to estimate the overall quality score. Alexiou et al. [30] extracted a set of statistical dispersion features including geometry, normal, curvature, and color, to analyze the local changes between the reference and the degraded PC. Diniz et al. [31], [32] proposed two frameworks where geometry and texture features are extracted, and the distance between their statistics is used to compare the reference and the degraded object. In [31] the geometry and the texture features are computed independently while in [32] the texture information is extracted with respect to their geometry information. In [33], Yang et al. proposed a GraphSIM approach that construct graphs in the reference and degraded PC in order to calculate the similarity index for PC quality evaluation. Zhang et *al.* [34] approached the problem of the PCQA from the perspective of transformational complexity to avoid the complicated process of feature selection. The PC quality is estimated by calculating the complexity of transforming the distorted PC back to its reference.

For the projected-based metrics, the PC is projected into multiple 2D planes and evaluated by comparing the corresponding images from the reference and the distorted PC using classical 2D image quality assessment metrics. Torlig et al. [35] projected the voxelized PC into six orthographic viewpoints and employed 2D objective quality methods, including Peak Signal-to-Noise Ratio (PSNR) [36], Information Fidelity in Pixel domain (VIFP) [37], Structural Similarity Index (SSIM) [27] and Multi Scale Structural Similarity Index (MS-SSIM) [38], to predict the perceptual quality of the stimuli. Correspondingly, Yang et al. [39] performed for each PC six perpendicular color texture and their corresponding depth images. The quality score is then obtained by combining the local and the global image-based features of all the projected planes. In the same vein, Alexiou et al. [40] studied the impact of the view number on the performance of the algorithm used to assess the perceived quality of the content. Additionally, they weighted the projected views depending on the user interaction in subjective evaluation experiments. Javahri et al. [41] address the problem

of misalignment between the reference and the degraded projected images when geometric degradation exists in the PC content. To avoid this problem, the authors proposed to assign the same geometry condition to both the reference and distorted PC through a recoloring step before applying the projection step. Diniz et al. [42] proposed to calculate the visual textures similarities from the 2D projections of the reference and the degraded PC and combining them with the geometrical similarities to assess the PC quality.

Limited works are introduced in Reduced-Reference (RR) point cloud quality assessment. Viola et al. [43] proposed to extract a set of features from the reference and the evaluated PC that are transmitted through the processing channel and used on the receiver side to evaluate the visual quality of the PC. Liu et al. [44] proposed to evaluate the degradation of the V-PCC compressed PC using the geometry and color quantization parameters.

Recently, several No-Reference (NR) point cloud quality assessment methods have been proposed in the literature. Tao et al. [45] proposed to project the PC into 2D projections that are fed to a multi-scale feature fusion network to evaluate the visual PC quality blindly. Yang et al. [46] proposed to represent the PC with six texture and depth images, and then aggregate the features extracted from these maps as point cloud quality index. Liu et al. [47] proposed point cloud quality assessment network (PQA-Net) framework that consists of multi-view projection feature extraction, followed by the distortion type classification, and perceptual quality prediction sub-tasks to assess PCs affected with only an individual degradation. In [48], the authors proposed the use of transfer learning in order to leverage the rich subjective scores of 2D images in 3D quality score assessment through domain adaptation. To achieve this, Generative Adversarial Networks (GANs) are used to extract effective latent features and minimize the domain discrepancy between 2D and 3D data, then a quality regression network is utilized to find the final MOS. Liu et al. [49] proposed a method based on sparse convolutional layers and residual blocks to extract the hierarchical features of the PC, which are then pooled and sent to a regression model to predict PC quality score. Other interesting approaches have been recently proposed in the literature that more focus on the perceptual impact of compression schemes or have been applied to 3D meshes.

III. PROPOSED METHOD

The overall objective of this work is to predict automatically the visual quality score of the PC without relying on the reference content. To achieve this goal, we compose our model into two major steps, which are the preprocessing step and the objective quality score estimation step, as depicted in the flowchart 1. In the preprocessing step, we project 2D views from the degraded PC using the perspective projection. Then, we extract central overlapping patches in the vertical direction. In the quality score estimation step, the feature learning is performed in the training phase using a CNN model, and the visual quality score prediction of the PC is performed in the test phase after pooling all the estimated scores of the extracted patches from the PC.

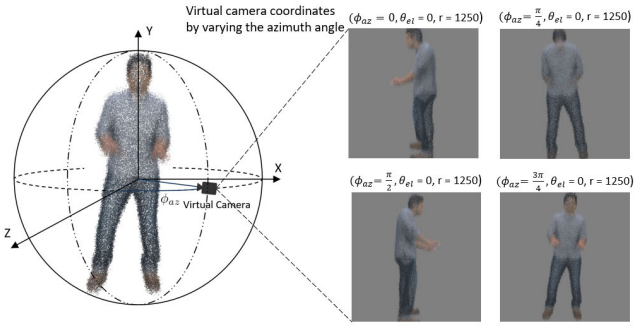


FIGURE 2. An example of the projected views from a PC with Downscaling and Geometry Gaussian noise degradation from the SJTU database. The virtual camera coordinates are r, θ_{el} and ϕ_{az}. ϕ_{az} is the azimuth angle, θ_{el} is the elevation angle and r is the radius.

A. PREPROCESSING

1) FROM PC TO 2D PROJECTION VIEWS

The first step of the proposed method consists of projecting each distorted PC object into different 2D viewpoints. For doing so, we exploit a perspective projection to mimic the perception of the human visual system when evaluating the quality of the PC. This projection captures information about the geometric and color distortions, as well as the stimuli depth information. To be specific, we fix virtual cameras at different angles to surround the PC. The centroid of the 3D object is set to the origin of a spherical coordinate system $(r, \theta_{el}, \phi_{az})$ where r is the radius that represents the distance between the virtual camera and the origin, $\theta_{el} \in [0, 2\pi]$ and $\phi_{az} \in [0, 2\pi]$ are the elevation and the azimuth angles, respectively [50]. We note that the virtual camera coordinates are obtained by varying the azimuths angle with $\frac{\pi}{24}$ and setting the elevation angle to zero. Intuitively, the distance r is changing according to the size of each PC in order to cover it entirely and clearly. In Fig. 2, we give an example of 2D projected images from a degraded PC when varying the azimuth angle with $\frac{\pi}{4}$. Additionally, we capture two other projections from the north and the south poles. In our work, the size of each projection is 512×512 pixels.

2) PATCH EXTRACTION AND NORMALIZATION

Since the 3D object is concentrated in the middle of each 2D projected image, we extract central overlapping patches of size 224×224 pixels using a sliding step (stride) in the vertical direction to discard the useless information in the background. We give an example of patch extraction with a stride equal to 40 to obtain 4 patches in Fig. 3. This splitting allows us to augment the data and evaluate locally the distortion that appears in the PC.

After that, we normalize each extracted patch as demonstrated in Equation 1. The use of normalization not only remedies the saturation problem, but also provides a decorrelation effect and makes the neural network more resilient to

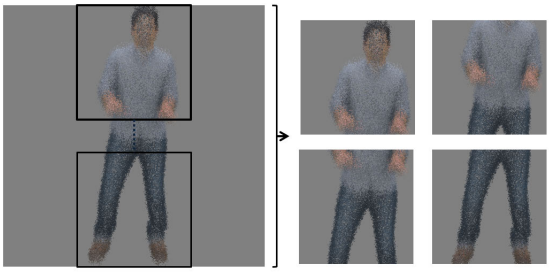


FIGURE 3. An example of patch extraction of a distorted projected image from a PC with Downscaling and Geometry Gaussian noise degradation from the SJTU database.

FIGURE 4. The reference samples from ICIP2020 (1) and SJTU (2) WPC (3) databases.

brightness and contrast changes [14].

$$\hat{I}(i,j) = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + C} \tag{1}$$

where $\hat{I}(i,j)$ is the normalized intensity value of the $I(i,j)$ pixel at the (i,j) location, C is a constant value that is set at 1 to prevent division by zero, $\mu(i,j)$ represents the mean and $\sigma(i,j)$ refers to the variance.

B. OBJECTIVE QUALITY SCORE ESTIMATION

We recall that the ultimate goal of our work is to predict the objective quality score of a given PC. For doing so, we develop a novel NR-PCQA method based on a deep learning approach. More precisely, we exploit the high potential of the deep CNN that has demonstrated excellent performance in various computer vision tasks to learn discriminative and meaningful visual features without relying on handcrafted ones [51]. Such a method requires two phases: training phase to learn the model and test one to evaluate the performance of the model.

1) TRAINING PHASE

In the training phase, we first construct an annotated database. Since the degraded PC samples from all state-of-the-art PC databases have homogeneous distortions, we affect the normalized Mean Opinion Score *MOS* (subjective quality score) as an annotation to all extracted patches from the same PC. After that, using the built database, we train the CNN model using three different networks for comparison: AlexNet [52], VGG [53] and ResNet [54]. This comparison allows studying the effect of the architecture as well as the impact of depth on the performance of the proposed method.

- AlexNet [52]: is a CNN classification model that won the Imagenet Large-scale Visual Recognition Challenge (ILSVRC) in 2012. It is composed of 5 convolutional layers with max-pooling, followed by 3 fully connected layers. The authors introduce the use of overlapping pooling, and the use of Rectified Linear Unit (*ReLu*) in addition to the dropout to prevent overfitting and improve learning.
- VGG [53]: is proposed by the Oxford Visual Geometry Group. It won the ILSVRC in 2014. Different versions are provided depending on the number of convolutional layers. In this study, we compare VGG16 and VGG19 which consisted of 16 and 19 layers, respectively. This network is characterized by its deep structure and small convolution kernels that reduce the computational complexity of the model while improving its generalization ability.
- ResNet [54]: is won the ILSVRC in 2015. The authors introduced the Residual blocks in order to reduce the training time and to improve the accuracy. In this work, we use ResNet18 and ResNet50.

We underline that we adjust all the baselines to make them suitable for the NR-PCQA task by modifying the size of their last three fully-connected layers into 512, 512, 1 neurons, respectively. Each fully-connected layer is followed by a dropout set to 0.5 and a *ReLu* function, except for the output layer that has a *Sigmoid* function. Then, using the pre-trained CNN models, mentioned above, on ImageNet dataset [52], we fine-tune the networks on our dataset to adapt their weights to the quality prediction task by minimizing the following L_1 Norm loss function:

$$\mathcal{L}oss = \frac{1}{N}\sum_{i=1}^{N} |PMOS_i - MOS_i| \tag{2}$$

where MOS_i and $PMOS_i$ represent the subjective and estimated quality score of the patches, respectively. N is the total number of patches. To select the parameters of the trained model, we use a validation set that contains PC objects different from the training and the test sets. To optimize the model parameters, we adopt stochastic gradient descent (*SGD*) with momentum. In addition, we initialize the learning rate with 10^{-3} that is descended every 70 epoch with a decay rate of 10^{-1}.

2) TEST PHASE

In the test phase, we estimate the quality score of each patch using the model learned in the training phase (M). After that, we aggregate the quality scores of all patches to derive the final visual quality score of the overall PC object. For this purpose, we use five spatial pooling methodologies that are defined by the following Equations:

- Minimum pooling: is the minimum score in K predicted scores of the PC object patches, and it is calculated by the following Equation:

$$MOS_{global} = \min_{1,\ldots,N} PMOS_i^N, i = 1, \ldots, N \quad (3)$$

- Maximum pooling: is the maximum score of all predicted quality scores of the patches extracted from the PC object, and it is computed by the following Equation:

$$MOS_{global} = \max_{1,\ldots,N} PMOS_i^N, i = 1, \ldots, N \quad (4)$$

- Median pooling: is given by ordering the set of the predicted quality scores $PMOS_i$ of the PC and then taking the middle value of the predicted scores using the following Equation:

$$MOS_{global} = \begin{cases} PMOS\left[\frac{K+1}{2}\right] & \text{if } K \text{ is odd} \\ \frac{\left(PMOS\left[\frac{K}{2}\right] + PMOS\left[\frac{K}{2}+1\right]\right)}{2} & \text{if } K \text{ is even} \end{cases} \quad (5)$$

- Weighted average pooling: is calculated by applying to each $PMOS_i$ estimated quality score a corresponding w_i weight. The predicted scores are divided into 5 categories and the frequency of each category is used as w_i. We represent the weights mean pooling by the following Equation:

$$MOS_{global} = \frac{1}{\sum_{i=1}^{K} w_i} \sum_{i=1}^{K} w_i PMOS_i \quad (6)$$

- Average pooling: is calculated by averaging all the predicted quality score patches of the PC, as shown in the following Equation:

$$MOS_{global} = \frac{1}{K} \sum_{i=1}^{K} PMOS_i \quad (7)$$

IV. EXPERIMENTAL RESULTS

In this section, we first provide a description of the experimental setting including the used PCQA databases, the validation protocol, and the evaluation metrics. Subsequently, we make an ablation study, a comparison with the state-of-the-art, and a cross-dataset evaluation.

A. POINT CLOUD DATABASES

We evaluate our proposed PCQA method on two databases:

- **ICIP2020** [55]: is composed of 6 reference PC objects that represent human body models, including four watertight/full-coverage objects (Soldier, LongDress, Loot and RedandBlack) and two semi-coverage PC (Ricardo10 and Sarah9). Each original PC is degraded with 3 types of compression distortion caused by G-PCC Octree, G-PCC Trisoup, and V-PCC with a 5-level rating scale, 1 indicating the lowest quality and 5 the highest quality. Consequently, the number of degraded objects is 90.
- **SJTU** [39]: contains 9 reference PC and 378 distorted ones, which include both human body models and inanimate objects. Each original PC is degraded according to 7 types of distortion with 6 different levels caused by 4 independent distortions: Octree-based compression (OT), Color noise (CN), Geometry Gaussian noise (GGN) and Downsampling (DS), and 3 superimposed distortions: Downscaling and Color noise (DS+CN or D+C), Downscaling and Geometry Gaussian noise (DS+GGN or D+G), Color noise and Geometry Gaussian noise (CN+GGN or C+G). These types of distortions are used to characterize the distortions that may be present in the PC during the processing.
- **WPC** [46]: is composed of 20 reference inanimate objects, and it is corresponding 740 distorted PC degraded by 3 compression types including G-PCC Trisoup with 12 different levels, G-PCC Octree with 4 different levels and VPCC with 9 different levels, Gaussian noise (GN) with 9 different levels and Downsampling (DS) with 3 different levels.

B. VALIDATION PROTOCOL AND EVALUATION METRICS

To evaluate the performance and the effectiveness of our proposed method for the point cloud quality assessment, we use the 4-fold cross-validation protocol for ICIP2020, 7-fold for SJTU and 20-fold for WPC database. For all datasets, the different folds are organized as follows: one fold as a test set, another fold as the validation set, and the remaining folds as the training set. In this way, we assure that we do not have an overlapping between the three. Besides, we adopt the following metrics that are commonly used in the field of quality assessment:

1) Spearman Rank Order Coefficient (SROCC): it measures the monotonicity of the model estimation. The

SROCC metric can be defined as follows:

$$SROCC = 1 - \frac{\sum_{i=1}^{J}(rank(MS_i) - rank(MP_i))^2}{J(J^2-1)} \quad (8)$$

where MP_i and MS_i are the estimated and the ground truth quality score while J indicates the total number of the objects.

2) Pearson Linear Correlation Coefficient (PLCC): it computes the prediction accuracy between the predicted and the subjective score, and is calculated with the following Equation:

$$PLCC = \frac{\sum_{i=1}^{J}(MP_i - \overline{MP_m})(MS_i - \overline{MS_m})}{\sqrt{\sum_{i=1}^{J}(MP_i - \overline{MP_m})^2}\sqrt{(MS_i - \overline{MS_m})^2}} \quad (9)$$

where $\overline{MS_m}$ and $\overline{MP_m}$ are the mean values of MP and MS.

The highest absolute values of the PLCC and SROCC criteria (close to 1) indicate the best quality prediction performance of the model.

C. ABLATION STUDY

The proposed method has three degrees of freedom: the number of rendering views, the number of patches, the architecture of the fine-tuned networks and the spatial pooling methodology. We study in the following the influence of these parameters on the ICIP2020 database.

1) IMPACT OF THE RENDERING VIEWS NUMBER

In our work, we represent the PC with different perspectives (2D images) which are obtained by fixing virtual cameras at different angles, as illustrated in Fig. 1. The number of views is related to the used rotation angles. In other words, using small angles increases the number of projected views and vice versa. Therefore, in Table 1, we investigate the effect of this parameter (i.e. the number of views) while setting the other model parameters unchanged (i.e. we fix the CNN model to VGG, the number of the patches to 4 and the pooling methodology to average pooling) to determine how many projected images are needed to represent all the information of the 3D object. According to this table, the best performance is given when we use the $\pi/24$ that provides 50 views (48 views + the top and bottom views). When we use a smaller angle ($\pi/48$), the performance of the model decreases, this is due to the redundancy of the information from the captured images. Similarly, the use of larger angles ($\pi/2$, $\pi/4$, $\pi/6$) causes a loss of information that prevents the PC object from being represented accurately. It is noteworthy that when we increase the number of views from 10 ($pi/4$) to 14 ($pi/6$), we notice a small decrease in the correlation. However, this difference is not statistically significant (P-value $>$ 0.5) according to the One-way Analysis of Variance (ANOVA) test. In addition, we examine the performance of the proposed model with and without the top and bottom views, as presented in Table 2. We observe a slight gap in the performance of the proposed model when the top and bottom views are added. In the rest of the paper, the experiments are performed using 50 views.

2) IMPACT OF THE NUMBER OF THE PATCHES

As mentioned above, we extract patches from each projected 2D image. After that, we aggregate the local quality of all sampled patches to measure the quality score of the PC. Since we use a fixed patch size (224×224), the number of extracted patches is related to the stride size. We vary this parameter in Table 3 while keeping the rest of the model unchanged. As we can observe, the number of patches does not significantly affect the quality performance. In the rest of the experiments, we use 4 patches.

3) IMPACT OF THE FINE-TUNED CNN NETWORKS

The extracted patches are used to fine-tune the pre-trained CNN architecture. To study the influence of this later on the quality performance, we conduct experiments with three different CNN architectures pre-trained on ImageNet dataset: AlexNet, VGG, and ResNet. As recorded in Table 4, the performance of the proposed model varies from a pre-trained model to another. The use of residual networks (ResNet18 and ResNet50) and shallow network (AlexNet) decrease the performance since the difference is statistically significant between them and the VGG based models, P-value $<$ 0.5. These results might be explained by the fact that AlexNet is a shallow model and thus cannot extract enough relevant features. For ResNet based models, it seems that the skip connections introduce a kind of redundancy that is not relevant for the quality task [56], [57]. However, it is giving competitive results compared to the state-of-the-art methods on ICIP2020 database. Moreover, we test the impact of the depth for VGG and ResNet based models. As we can observe, the depth did not affect the performance results since the p-value $>$ 0.5, which indicates that the difference is not statistically significant. From the previous remarks, we conclude that the network architecture has more influence on the model performance compared to the effect of the depth. Based on all the obtained results, we use VGG16 as our pre-trained CNN model to compare the performance of our method with the state-of-the-art metrics.

4) IMPACT OF THE SPATIAL POOLING METHODOLOGY

The final step of our proposed method is to aggregate all the predicted quality scores of the extracted patches to obtain the final PC quality score. To study the influence of the spatial pooling methodology on the performance of the proposed model. We conduct a comparison with five pooling methods: Minimum, maximum, median, weighted average, and average pooling. Based on the obtained results in Table 5, we observe that the best results on both metrics SROCC and PLCC are obtained when using the spatial average pooling. We use this later in our experiment.

5) QUALITATIVE RESULTS

In Fig. 5, we present a qualitative study to test the ability of the proposed model in evaluating the PC quality

TABLE 1. Correlation coefficients SROCC and PLCC of the proposed model using different numbers of rendering views on ICIP2020 database.

Number of views	6 views ($\pi/2$)	10 views ($\pi/4$)	14 views ($\pi/6$)	26 ($\pi/12$) views	50 views ($\pi/24$)	98 views ($\pi/48$)
SROCC	0.963	0.971	0.969	0.971	**0.982**	0.980
PLCC	0.943	0.959	0.947	0.956	**0.958**	0.949

TABLE 2. Performance of the proposed model with and without top and bottom views on ICIP2020 database.

	SROCC	PLCC
Without top and bottom views	0.981	0.941
With top and bottom views	**0.982**	**0.958**

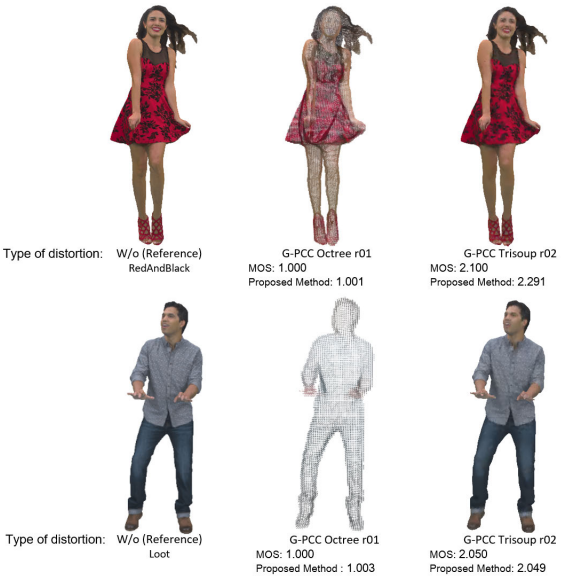


FIGURE 5. Distorted point clouds with G-PCC Octree r01 and Trisoup r02 on ICIP2020 database.

by comparing the predicted and the corresponding ground truth quality scores on different distorted samples. More precisely, we compare the quality scores of 2 different samples: RedAndBlack and Loot that are degraded with GPCC octree and G-PCC trisoup distortions. The PC with high quality is noted by 5, and the PCwith the lowest quality is scored by 1. As we can observe from Fig. 5, our model is able to predict quality scores that are close to the ground truth *MOS*, which proves the effectiveness of our proposed model.

D. COMPARISON STUDY WITH THE STATE-OF-THE-ART

In this section, we conduct a comparison study of our model against different Full-Reference (FR), Reduced-Reference (RR), and No-Reference (NR) state-of-the-art methods. The best results are highlighted in bold.

1) COMPARISON WITH FR AND RR METHODS

In Table 6, 7 and 8, we list the correlation coefficient results on ICIP2020, SJTU, and WPC databases of our method in comparison with FR and RR methods.

In ICIP2020 database (Table 6), considering all the PC objects distortions and regarding the FR methods, the Point-to-Point MSE (i.e. P2P MSE) [21], Point-to-Point Hausdorff (i.e. P2P Hausdorff) [21], Point-to-Plane MSE (i.e. P2Pl MSE) [22], Point-to-Plane Hausdorff (i.e. P2Pl Hausdorff), Plane-to-Plane (i.e. Pl2Pl) [22] and Point-to-Distribution-Geometry MMD (i.e. P2D-G MMD) [24] metrics are based on the geometry structure and are given the lower performance. One reason is that these methods do not consider the color information and are based only on a simple geometric distance to compute the quality of the PC. This is proved by the better performance in the Point-to-Distribution-Geometry-Joint-Geometry-and-Color MMD (i.e. P2D-JGC MMD) [24], PCQM [28], BitDance [31], and GraphSIM [33] metrics that include the color information for the PC evaluation. We denote that the FR projection-based methods including, JGC-ProjQM DISTS, JGC-ProjQM LPIPS methods [41], achieve good correlations on each degradation type and on all the database. However, the quality evaluation of these methods depends on the reference information that is not available in the majority of practical applications. For the RR methods, the PCM metric [43], which is based on the comparison between the original and degraded PC in the receiver side to evaluate the transmitted contents, archives a PLCC and a SROCC of 0.882 and 0.627, respectively. By studying the different types of distortions individually, we remark that the proposed method outperforms the state-of-the-art methods on the three compression distortion types (VPCC, G-PCC Trisoup, and G-PCC Octree).

In SJTU database (Table 7), the proposed method provides the best correlation coefficient results on all the distortions and outperforms the state-of-the-art metrics in SROCC and PLCC correlation coefficients. It is worth noting that the correlation values of our method and all the state-of-the-art methods are less than the correlation values of ICIP2020 dataset. This could be justified by the type of distortions in the two databases. ICIP2020 database consists only of objects with compression types, while SJTU database has more challenging types of degradation such as acquisition noise and resampling, that can be applied individually (Octree-based compression (OT), Color noise (CN), Geometry Gaussian noise (GGN) and Downsampling (DS)) or superimposed (Downscaling and Color noise (D+C), Downscaling and Geometry Gaussian noise (D+G), Color noise and Geometry Gaussian noise (C+G)). In addition, we compare the SROCC and PLCC values for each of the seven degradation types. We denote that the correlations of the metrics marked with '−' are undefined because their variance is equal to zero (since they are divided by zero). As shown in Table 7, our

TABLE 3. Correlation coefficient of the proposed model on different patches number on ICIP2020 database.

Number of patches	2 patches (Stride = 224)		4 patches (Stride = 90)		8 patches (Stride = 40)	
	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
Correlation score	0.980	0.955	**0.982**	**0.958**	0.982	0.955

TABLE 4. Performance of the proposed model using different pre-trained models on ICIP2020 database.

AlexNet		VGG-16		VGG-19		ResNet-18		ResNet-50	
SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
0.971	0.919	**0.982**	**0.958**	0.968	0.967	0.965	0.951	0.980	0.952

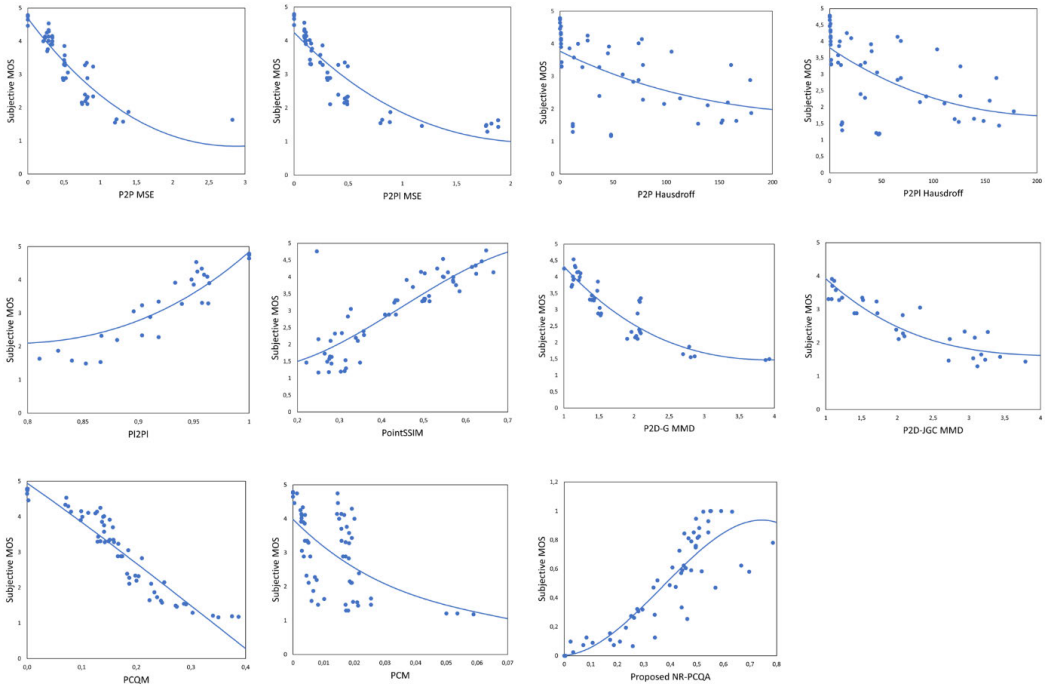


FIGURE 6. Scatter distributions of the objective scores versus the *MOS* scores on the database ICIP2020. X-axis is the objective quality values predicted by the metrics, Y-axis is the subjective quality scores and the curve presents the estimated no-linear logistic function.

TABLE 5. Performance of the proposed model using different spatial pooling strategies on ICIP2020 database.

	SROCC	PLCC
Minimum	0.823	0.795
Maximum	0. 775	0.820
Median	0.844	0.814
Weighted average	0. 796	0.827
Average	**0. 982**	**0.958**

model exhibits robust performance for each type of degradation and shows good correlations with the subjective quality scores. The geometric-based methods, P2P MSE [21], P2P Hausdorff [21], P2Pl MSE [22], P2Pl Hausdorff [22], Pl2Pl [23] and P2D-G MMD [24] metrics, fail to reflect the color degradation in the PC and perform the worst SROCC and PLCC. In the same vein, all the compared metrics present inferior performance in terms of PLCC for superimposed distortions, expect PCQM [28], PointSSIM [30], JGC-ProjQM DISTS, JGC-ProjQM LPIPS [41], Yang et al. method [39] and our proposed method. We note that the PCM metric [43] provides poor results on all types of degradation, which could be explained by the large set of features used in their method that prevents its generalization across all types of degradation.

In WPC database (Table 8), our approach outperforms all the compared methods for each distortion type and for the whole database. We note that most of the compared methods achieve lower performance on the WPC database compared to the other two databases. This can be explained by the fact that some WPC reference object samples are less sensitive to distortions in perceived visual quality. For example, Banana, Honeydew-melon, and Litchi (See Fig. 4), etc, samples show less complexity in geometry structure and in texture color, making the distortion difference less obvious and harder to measure. We highlight our model is able to keep nearly the same performance on the larger and more difficult database, which proves its high robustness.

TABLE 6. Performance comparison of the proposed method against the FR and RR state-of-the-art methods on ICIP2020 database.

		Distortion type							
		VPCC		G-PCC Octree		G-PCC Trisoup		All	
Method type	Methods	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
FR	P2P MSE [21]	0.954	0.615	0.963	0.817	0.944	0.864	0.947	0.673
	P2P Hausdorff [21]	0.682	0.615	0.944	0.692	0.975	0.615	0.656	0.673
	P2Pl MSE [22]	0.971	0.618	0.932	0.848	0.971	0.618	0.975	0.670
	P2Pl Hausdorff [22]	0.735	0.491	0.932	0.838	0.735	0.491	0.704	0.521
	Pl2Pl [23]	0.902	0.626	0.628	0.603	0.447	0.292	0.902	0.626
	P2D-G MMD [24]	0.960	0.784	0.831	0.871	0.906	0.906	0.954	0.806
	P2D-JGC MMD [25]	0.965	0.881	0.906	0.880	0.906	0.880	0.962	0.889
	PCQM [28]	0.977	0.942	0.966	0.978	0.977	0.955	0.977	0.942
	PointSSIM [30]	0.546	0.246	0.628	0.603	0.447	0.292	0.795	0.717
	TCDM [34]	0.822	–	0.885	–	0.970	–	0.935	0.942
	GraphSIM [33]	0.855	–	0.939	–	0.770	–	0.890	0.872
	BitDance [31]	–	–	–	–	–	–	0.936	0.932
	JGC-ProjQM DISTS [41]	0.968	0.983	0.969	0.954	0.960	0.961	0.974	0.937
	JGC-ProjQM LPIPS [41]	0.968	0.983	0.969	0.954	0.960	0.961	0.978	**0.960**
RR	PCM [43]	0.882	0.627	0.830	0.749	0.510	0.407	0.882	0.627
NR	**The proposed NR-PCQA**	**0.982**	**0.958**	**1.000**	**0.987**	**1.000**	**0.957**	**0.982**	0.958

TABLE 7. Performance comparison of the proposed method against the FR and RR state-of-the-art methods on SJTU database.

		Distortion type															
		OT		CN		GGN		DS		D+C		D+G		C+G		All	
Method type	Methods	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
FR	P2P MSE [21]	0.349	0.481	-	-	0.801	0.385	0.646	0.499	0.661	0.165	0.837	0.385	0.757	0.416	0.803	0.606
	P2P Hausdorff [21]	0.286	0.496	0.286	0.496	0.858	0.395	0.451	0.260	0.383	0.167	0.761	0.386	0.818	0.417	0.687	0.606
	P2Pl MSE [22]	0.345	0.470	-	-	0.846	0.369	0.757	0.448	0.746	0.165	0.837	0.385	0.809	0.405	0.715	0.568
	P2Pl Hausdorff [22]	0.377	0.492	-	-	0.858	0.395	0.451	0.351	0.383	0.167	0.801	0.467	0.828	0.438	0.683	0.562
	Pl2Pl [23]	0.359	0.290	0.059	0.058	0.579	0.515	0.361	0.210	0.241	0.164	0.723	0.665	0.572	0.503	0.772	0.615
	P2D-G MMD [24]	0.269	0.075	0.067	0.190	0.768	0.462	0.548	0.188	0.747	0.166	0.742	0.478	0.754	0.501	0.604	0.628
	P2D-JGC MMD [25]	0.260	0.039	0.778	0.699	0.792	0.494	0.639	0.315	0.762	0.174	0.772	0.509	0.794	0.594	0.755	0.667
	PCQM [28]	0.741	0.786	0.812	0.801	0.903	0.771	0.864	0.787	0.937	0.857	0.883	0.712	0.920	0.813	0.855	0.813
	PointSSIM [30]	0.806	**0.831**	0.742	0.765	0.936	0.964	0.866	0.902	0.733	0.741	0.951	0.955	0.809	0.811	0.733	0.715
	TCDM [34]	0.793	–	0.819	–	0.921	–	0.876	–	0.934	–	0.944	–	0.951	–	0.910	0.930
	GraphSIM [33]	0.693	–	0.778	–	0.916	–	0.872	–	0.886	–	0.888	–	0.941	–	0.841	0.856
	BitDance [31]	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.730	0.714
	Yang *et al.* method [39]	0.760	0.790	0.790	**0.850**	0.730	0.760	0.580	0.730	0.880	0.890	0.680	0.690	0.930	0.950	0.602	0.607
	JGC-ProjQM DISTS [41]	**0.811**	0.822	0.823	0.838	0.747	0.876	0.896	0.927	0.883	0.872	0.806	0.900	0.786	0.869	0.671	0.600
	JGC-ProjQM LPIPS [41]	0.738	0.765	0.780	0.813	0.747	0.836	0.862	0.862	0.921	0.894	0.819	0.867	0.814	0.866	0.690	0.665
RR	PCM [43]	0.279	0.271	0.029	0.014	0.175	0.187	0.428	0.398	0.006	0.093	0.430	0.509	0.132	0.265	0.219	0.263
NR	**The proposed NR-PCQA**	0.641	0.816	**0.853**	0.830	**0.976**	**0.975**	**0.927**	**0.978**	**0.967**	**0.966**	**0.959**	**0.980**	**0.992**	**0.986**	**0.915**	**0.943**

TABLE 8. Performance comparison of the proposed method against the FR and RR state-of-the-art methods on WPC database.

		Distortion type											
		DS		GN		G-PCC Trisoup		G-PCC Octree		VPCC		All	
Method type	Methods	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
FR	P2P MSE [21]	0.900	0.779	0.728	0.686	0.464	0.534	-	-	0.697	0.684	0.566	0.399
	P2P Hausdorff [21]	0.904	0.755	0.688	0.662	0.293	0.243	-	-	0.445	0.254	0.258	0.166
	P2Pl MSE [22]	0.849	0.724	0.737	0.677	0.462	0.521	-	-	0.705	0.702	0.446	0.395
	P2Pl Hausdorff [22]	0.861	0.614	0.692	0.664	0.355	0.299	-	-	0.558	0.377	0.313	0.226
	Pl2Pl [23]	0.643	0.618	0.683	0.677	0.141	0.132	0.035	0.025	0.382	0.392	0.321	0.315
	P2D-G MMD [24]	0.786	0.688	0.863	0.826	0.502	0.509	0.062	0.032	0.790	0.734	0.411	0.420
	P2D-JGC MMD [25]	0.786	0.688	0.863	0.826	0.502	0.509	0.786	0.709	0.790	0.734	0.431	0.413
	PCQM [28]	0.875	–	0.886	–	0.821	–	0.894	–	0.643	–	0.743	0.751
	PointSSIM [30]	0.835	0.872	0.586	0.670	0.681	0.657	0.791	0.783	0.365	0.379	0.450	0.460
	TCDM [34]	0.882	–	0.857	–	0.832	–	0.795	–	0.640	–	0.807	0.804
	GraphSIM [33]	0.898	–	0.840	–	0.816	–	0.855	–	0.612	–	0.841	0.856
	Projection PSNR [35]	0.5399	0.6783	0.6538	0.8292	0.1968	0.3291	0.7809	0.7730	0.1998	0.2903	0.4601	0.4989
	Projection SSIM [35]	0.8039	0.8529	0.7509	0.8213	0.6144	0.6065	0.8391	0.8258	0.3195	0.3299	0.6138	0.6013
	Projection MS-SSIM [35]	0.8876	0.9375	0.7493	0.8372	0.6572	0.6545	0.8770	0.8774	0.4213	0.4397	0.6656	0.6701
	Projection VIFP [35]	0.9212	0.9700	0.8067	0.8467	0.8153	0.8105	0.8976	0.8950	0.7484	0.7448	0.7689	0.7670
RR	PCM [43]	0.737	0.661	0.780	0.788	0.243	0.304	0.672	0.662	0.282	0.251	0.345	0.367
NR	**The proposed NR-PCQA**	**0.939**	**0.971**	**1.000**	**0.999**	**0.914**	**0.936**	**1.000**	**0.995**	**0.956**	**0.966**	**0.930**	**0.925**

TABLE 9. Performance comparison of the proposed method against the NR state-of-the-art methods on SJTU [39] and WPC databases.

Methods	SJTU		WPC	
	SROCC	PLCC	SROCC	PLCC
PQA-Net [47]	0.23	0.28	0.69	0.70
IT-PCQA [48]	0.63	0.58	0.54	0.55
ResSCNN [49]	0.81	0.86	0.75	0.72
The proposed NR-PCQA	**0.83**	**0.87**	**0.86**	**0.85**

TABLE 10. Cross-database evaluation: The method is trained on SJTU and is tested on ICIP2020 and WPC databases.

	ICIP2020		WPC	
	PLCC	SROCC	SROCC	PLCC
Proposed method	0.765	0.725	0.234	0.419

2) COMPARISON WITH NR METHODS

To make a fair comparison of our method with the NR ones, we follow the protocol proposed in [49] and [48] on SJTU and WPC databases. We take 75% of the database for training and the remaining for testing. As shown in Table 9, the proposed method achieves the best SROCC and PLCC results on both databases, largely outperforming the compared NR methods, except for ResSCNN which shows competitive results on SJTU database. This can be explained by the projection-based NR metrics, including PQA-Net and IT-PCQA methods, not taking into consideration the impact of the number of views and utilizing much unnecessary background information that could drop the model accuracy. In our proposed model, we address this issue by conducting a patch extraction in the central axis that contributes to removing the majority of the background information and focusing on evaluating the local degradation in the PC.

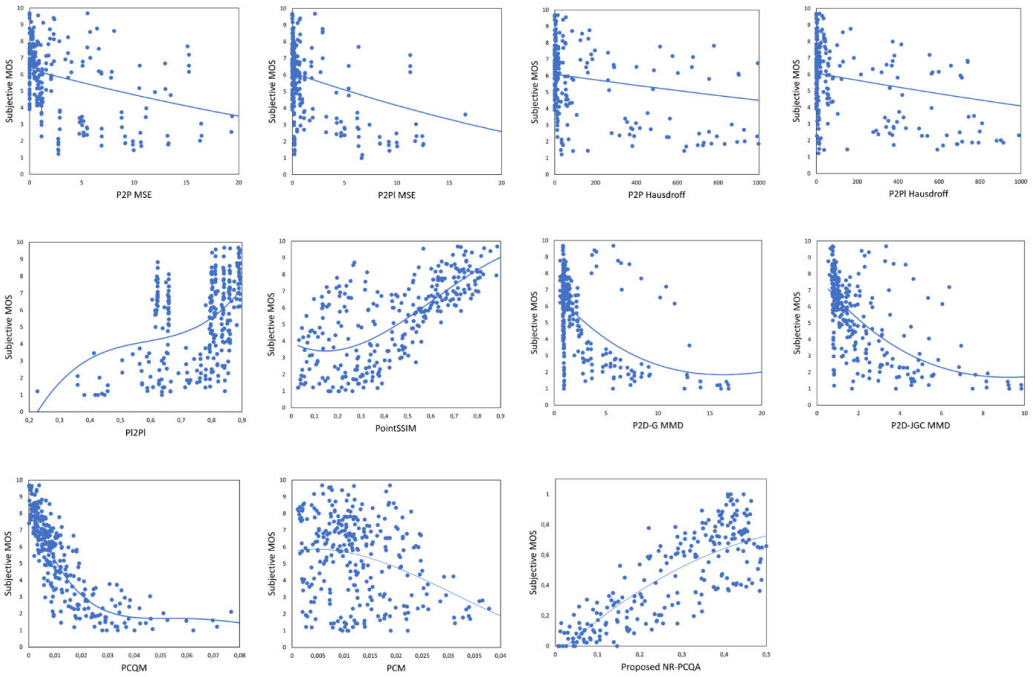


FIGURE 7. Scatter distributions of the objective scores versus the *MOS* scores on the database SJTU. X-axis is the objective quality values predicted by the metrics, Y-axis is the subjective quality scores and the curve presents the estimated no-linear logistic function.

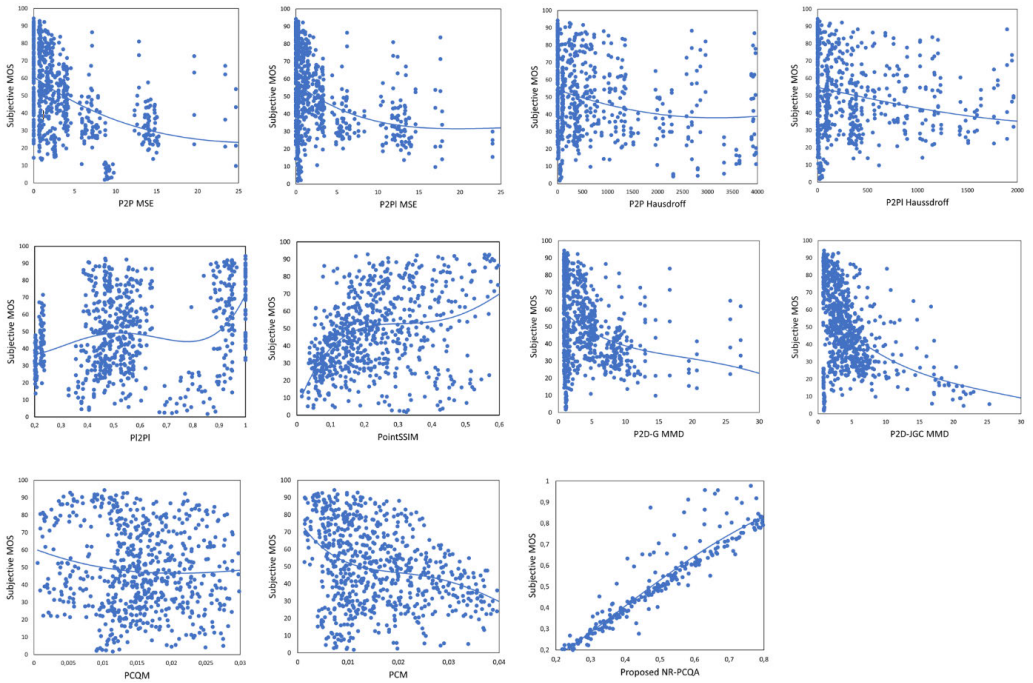


FIGURE 8. Scatter distributions of the objective scores versus the *MOS* scores on the database WPC. X-axis is the objective quality values predicted by the metrics, Y-axis is the subjective quality scores and the curve presents the estimated no-linear logistic function.

To better illustrate the accuracy of the compared state-of-the-art metrics, we provide the scatter plots of the predicted quality values and the subjective scores of the 3 benchmark databases: ICIP2020, SJTU and WPC, in Fig. 6, 7 and 8. Note that the blue line presented in each sub-plot is obtained by exploiting a nonlinear curve fitting process according to the

suggestion of the Video Quality Experts Group (VQEG) [58] that is defined in Equation 10,

$$MS_p = a_0 + a_1 * (MP) + a_2 * (MP)^2 + a_3 * (MP)^3 \quad (10)$$

where MS_p is the predicted subjective score of the proposed model, MP is the objective quality score and a_0, a_1, a_2 and a_3 are parameters to be estimated. The closer the points are to the blue line, the better the model correlation and vice versa. As expected, our proposed model shows a good fit with the nonlinear regression curve on the 3 databases while some of the methods based on the simple geometric distance and the PCM metric are away from the fitted curve specially on the WPC database.

E. CROSS-DATASET EVALUATION

In this section, we investigate the generalization capability and the robustness of the proposed NR-PCQA model by evaluating the quality scores using a cross-database test. To do so, we first train the proposed model on SJTU database then we test it on the ICIP2020 and WPC databases. As presented in Table 10, our model provides decent correlation results (SROCC = 0.765, PLCC = 0.725) on ICIP2020 database. We note that only a single distortion type (G-PCC Octree) is shared between the training and the test database while the other distortions are different. When tested on WPC, our model shows decreased results, which can be explained by the large difference between the distribution of the objects and the distortions in the training and the test database, being more challenging for PCQA models.

V. CONCLUSION

In this paper, we proposed a novel no-reference point cloud quality assessment method to predict the visual quality of distorted 3D point clouds. We fine-tuned a pre-trained CNN model to learn the visual quality features of the 2D patches extracted from the rendered views of the 3D point cloud. The final quality score of the evaluated object is obtained by averaging the predicted quality scores of all patches. Interestingly, our proposed method does not require any reference information for evaluation, which is a promising solution for real-world applications. We conducted several ablation experiments to select the best parameters for our method. Moreover, the comparison with the Full-Reference, Reduced-Reference and No-Reference state-of-the-art methods proved that our proposed method provides superior correlations with the subjective quality scores on the three benchmark datasets and across different degradation types. Finally, the proposed method has demonstrated the generalization ability through cross-dataset experiments. In future work, we will project to generate our own pre-trained model. This will be done by training a generative adversarial network to learn to reconstruct the patches, and only use its encoder in the point cloud quality assessment task.

REFERENCES

[1] G. Bruder, F. Steinicke, and A. Nuchter, "Poster: Immersive point cloud virtual environments," in *Proc. IEEE Symp. 3D User Interfaces (3DUI)*, Mar. 2014, pp. 161–162, doi: 10.1109/3DUI.2014.6798870.
[2] A. H. Hoppe, K. Westerkamp, S. Maier, F. van de Camp, and R. Stiefelhagen, "Multi-user collaboration on complex data in virtual and augmented reality," in *HCI International 2018—Posters' Extended Abstracts*, C. Stephanidis, Ed. Cham, Switzerland: Springer, 2018, pp. 258–265.
[3] L. Han, T. Zheng, Y. Zhu, L. Xu, and L. Fang, "Live semantic 3D perception for immersive augmented reality," *IEEE Trans. Vis. Comput. Graphics*, vol. 26, no. 5, pp. 2012–2022, May 2020, doi: 10.1109/TVCG.2020.2973477.
[4] M. Liu, "Robotic online path planning on point cloud," *IEEE Trans. Cybern.*, vol. 46, no. 5, pp. 1217–1228, May 2016, doi: 10.1109/TCYB.2015.2430526.
[5] L. Zhang, P. Van Oosterom, and H. Liu, "Visualization of point cloud models in mobile augmented reality using continuous level of detail method," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 44, pp. 167–170, Jun. 2020.
[6] Q. Wang and M.-K. Kim, "Applications of 3D point cloud data in the construction industry: A fifteen-year review from 2004 to 2018," *Adv. Eng. Informat.*, vol. 39, pp. 306–319, Jan. 2019.
[7] D. Fernandes, A. Silva, R. Névoa, C. Simões, D. Gonzalez, M. Guevara, P. Novais, J. Monteiro, and P. Melo-Pinto, "Point-cloud based 3D object detection and classification methods for self-driving applications: A survey and taxonomy," *Inf. Fusion*, vol. 68, pp. 161–191, Apr. 2021.
[8] T. Xu, D. An, Y. Jia, and Y. Yue, "A review: Point cloud-based 3D human joints estimation," *Sensors*, vol. 21, no. 5, p. 1684, Mar. 2021.
[9] C. Cao, M. Preda, and T. Zaharia, "3D point cloud compression," in *Proc. 24th Int. Conf. 3D Web Technol.*, Los Angeles, CA, USA, Jul. 2019, pp. 1–9, doi: 10.1145/3329714.3338130.
[10] L. Cui, R. Mekuria, M. Preda, and E. S. Jang, "Point-cloud compression: Moving picture experts group's new standard in 2020," *IEEE Consum. Electron. Mag.*, vol. 8, no. 4, pp. 17–21, Jul. 2019, doi: 10.1109/MCE.2019.2905483.
[11] E. Alexiou and T. Ebrahimi, "On subjective and objective quality evaluation of point cloud geometry," in *Proc. 9th Int. Conf. Quality Multimedia Exper. (QoMEX)*, May 2017, pp. 1–3, doi: 10.1109/QoMEX.2017.7965681.
[12] H. Fretes, M. Gomez-Redondo, E. Paiva, J. Rodas, and R. Gregor, "A review of existing evaluation methods for point clouds quality," in *Proc. Workshop Res., Educ. Develop. Unmanned Aerial Syst. (RED UAS)*, Nov. 2019, pp. 247–252, doi: 10.1109/REDUAS47371.2019.8999725.
[13] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, "Subjective and objective quality evaluation of 3D point cloud denoising algorithms," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jul. 2017, pp. 1–6.
[14] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1733–1740.
[15] S. Bourbia, A. Karine, A. Chetouani, and M. E. Hassoun, "A multi-task convolutional neural network for blind stereoscopic image quality assessment using naturalness analysis," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Anchorage-Alaska, France, Sep. 2021, pp. 1434–1438. [Online]. Available: https://hal.archives-ouvertes.fr/hal-03258262
[16] X. Yang, F. Li, and H. Liu, "A survey of DNN methods for blind image quality assessment," *IEEE Access*, vol. 7, pp. 123788–123806, 2019.
[17] G. Zhai and X. Min, "Perceptual image quality assessment: A survey," *Sci. China Inf. Sci.*, vol. 63, no. 11, Nov. 2020.
[18] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 206–219, Jan. 2018, doi: 10.1109/TIP.2017.2760518.
[19] S. Bianco, L. Celona, P. Napoletano, and R. Schettini, "On the use of deep learning for blind image quality assessment," *Signal, Image Video Process.*, vol. 12, no. 2, pp. 355–362, Feb. 2018.
[20] S. Bosse, D. Maniry, T. Wiegand, and W. Samek, "A deep neural network for image quality assessment," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 3773–3777.
[21] R. Mekuria, Z. Li, C. Tulvan, and P. Chou, *Evaluation Criteria for PCC (Point Cloud Compression)*, Standard ISO/IEC JTC1/SC29/WG11 MPEG, N16332, w16332, Jun. 2016.

[22] D. Tian, H. Ochimizu, C. Feng, R. Cohen, and A. Vetro, "Geometric distortion metrics for point cloud compression," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 3460–3464, doi: 10.1109/ICIP.2017.8296925.

[23] E. Alexiou and T. Ebrahimi, "Point cloud quality assessment metric based on angular similarity," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2018, pp. 1–6, doi: 10.1109/ICME.2018.8486512.

[24] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, "Mahalanobis based point to distribution metric for point cloud geometry quality evaluation," *IEEE Signal Process. Lett.*, vol. 27, pp. 1350–1354, 2020, doi: 10.1109/LSP.2020.3010128.

[25] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, "A point-to-distribution joint geometry and color metric for point cloud quality assessment," 2021, *arXiv:2108.00054*.

[26] G. Meynet, J. Digne, and G. Lavoue, "PC-MSDM: A quality metric for 3D point clouds," in *Proc. 11th Int. Conf. Quality Multimedia Exper. (QoMEX)*, Jun. 2019, pp. 1–3.

[27] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004, doi: 10.1109/TIP.2003.819861.

[28] G. Meynet, Y. Nehme, J. Digne, and G. Lavoue, "PCQM: A full-reference quality metric for colored 3D point clouds," in *Proc. 12th Int. Conf. Quality Multimedia Exper. (QoMEX)*, May 2020, pp. 1–6, doi: 10.1109/QoMEX48832.2020.9123147.

[29] I. Viola, S. Subramanyam, and P. Cesar, "A color-based objective quality metric for point cloud contents," in *Proc. 12th Int. Conf. Quality Multimedia Exper. (QoMEX)*, May 2020, pp. 1–6.

[30] E. Alexiou and T. Ebrahimi, "Towards a point cloud structural similarity metric," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jul. 2020, pp. 1–6, doi: 10.1109/ICMEW46912.2020.9106005.

[31] R. Diniz, P. Garcia Freitas, and M. Farias, "Color and geometry texture descriptors for point-cloud quality assessment," *IEEE Signal Process. Lett.*, vol. 28, pp. 1150–1154, 2021.

[32] R. Diniz, P. Garcia Freitas, and M. C. Q. Farias, "Point cloud quality assessment based on geometry-aware texture descriptors," *Comput. Graph.*, vol. 103, pp. 31–44, Apr. 2022.

[33] Q. Yang, Z. Ma, Y. Xu, Z. Li, and J. Sun, "Inferring point cloud quality via graph similarity," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 6, pp. 3015–3029, Jun. 2022, doi: 10.1109/TPAMI.2020.3047083.

[34] Y. Zhang, Q. Yang, Y. Zhou, X. Xu, L. Yang, and Y. Xu, "Evaluating point cloud quality via transformational complexity," 2022, *arXiv:2210.04671*.

[35] E. M. Torlig, E. Alexiou, T. A. Fonseca, R. L. de Queiroz, and T. Ebrahimi, "A novel methodology for quality assessment of voxelized point clouds," *Proc. SPIE*, vol. 10752, Sep. 2018, Art. no. 107520I.

[36] S. Wolf, and M. H. Pinson, *Reference Algorithm for Computing Peak Signal to Noise Ratio (PSNR) of a Video Sequence With a Constant Delay*, document J.340 (06/10), ITU-T Contribution COM9-C6-E, 2009. [Online]. Available: https://www.itu.int/rec/T-REC-J.340-201006-I/

[37] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006, doi: 10.1109/TIP.2005.859378.

[38] Z. Wang, E. Simoncelli, and A. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, vol. 2, 2003, pp. 1398–1402, doi: 10.1109/ACSSC.2003.1292216.

[39] 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 Trans. Multimedia*, vol. 23, pp. 3877–3891, 2021, doi: 10.1109/TMM.2020.3033117.

[40] E. Alexiou and T. Ebrahimi, "Exploiting user interactivity in quality assessment of point cloud imaging," in *Proc. 11th Int. Conf. Quality Multimedia Exper. (QoMEX)*, Jun. 2019, pp. 1–6, doi: 10.1109/QoMEX.2019.8743277.

[41] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, "Joint geometry and color projection-based point cloud quality metric," *IEEE Access*, vol. 10, pp. 90481–90497, 2022, doi: 10.1109/ACCESS.2022.3198995.

[42] X. P. G. Freitas, R. Diniz, and M. C. Q. Farias, "Point cloud quality assessment: Unifying projection, geometry, and texture similarity," *Vis. Comput.*, vol. 38, no. 3, pp. 1–8, Mar. 2022.

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

[44] 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 Trans. Image Process.*, vol. 30, pp. 6623–6636, 2021, doi: 10.1109/TIP.2021.3096060.

[45] 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 *Proc. 29th ACM Int. Conf. Multimedia*, New York, NY, USA, Oct. 2021, pp. 5266–5272, doi: 10.1145/3474085.3475645.

[46] Q. Liu, H. Su, Z. Duanmu, W. Liu, and Z. Wang, "Perceptual quality assessment of colored 3D point clouds," *IEEE Trans. Vis. Comput. Graphics*, early access, Apr. 13, 2022, doi: 10.1109/TVCG.2022.3167151.

[47] 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 Trans. Circuits Syst. Video Technol.*, vol. 31, no. 12, pp. 4645–4660, Dec. 2021, doi: 10.1109/TCSVT.2021.3100282.

[48] Q. Yang, Y. Liu, S. Chen, Y. Xu, and J. Sun, "No-reference point cloud quality assessment via domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 21179–21188.

[49] Y. Liu, Q. Yang, Y. Xu, and L. Yang, "Point cloud quality assessment: Dataset construction and learning-based no-reference metric," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 19, no. 2s, pp. 1–26, Jun. 2023.

[50] S. Bai, X. Bai, Z. Zhou, Z. Zhang, Q. Tian, and L. J. Latecki, "GIFT: Towards scalable 3D shape retrieval," *IEEE Trans. Multimedia*, vol. 19, no. 6, pp. 1257–1271, Jun. 2017, doi: 10.1109/TMM.2017.2652071.

[51] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[52] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012.

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

[54] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015, *arXiv:1512.03385*.

[55] 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 *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2020, pp. 3428–3432, doi: 10.1109/ICIP40778.2020.9191308.

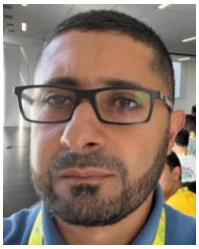
[56] A. Chetouani, "Image quality assessment without reference by mixing deep learning-based features," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2020, pp. 1–6.

[57] A. Chetouani, M. Quach, G. Valenzise, and F. Dufaux, "Deep learning-based quality assessment of 3d point clouds without reference," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jul. 2021, pp. 1–6.

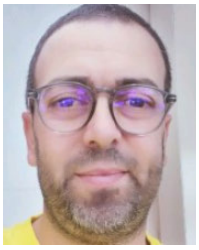
[58] A. Rohaly, P. Corriveau, J. Libert, A. Webster, V. Baroncini, J. Beerends, J.-L. Blin, L. Contin, T. Hamada, D. Harrison, A. Hekstra, J. Lubin, Y. Nishida, R. Nishihara, J. Pearson, A. Pessoa, N. Pickford, A. Schertz, M. Visca, and S. Winkler, "Video quality experts group: Current results and future directions," *Proc. SPIE*, vol. 4067, pp. 742–753, May 2000, doi: 10.1117/12.386632.

SALIMA BOURBIA received the master's degree in computer science and telecommunications from the Computer Science and Telecommunications Laboratory Research "Laboratoire de Recherche en Informatique et Telecommunications (LRIT)," in 2020. She is currently pursuing the Ph.D. degree with Mohammed V University, Rabat, Morocco, where she started her research in December 2020. She was a Visiting Researcher with the Vision-AD Laboratory, ISEN Yncréa Ouest, Nantes, France, and with the PRISME Laboratory, University of Orléans, Orléans, France. Her research interests include 3-D computer vision, machine/deep learning, and, mainly, 3-D point cloud quality assessment.

AYOUB KARINE received the M.Sc. degree from Mohammed V University, Rabat, Morocco, in 2014, and the joint Ph.D. degree from Mohammed V University and the Université de Bretagne Occidentale, France, in 2018. From 2017 to 2019, he was a temporary Assistant and Researcher with the Université de Haute-Alsace, France. He is currently an Associate Professor with ISEN Yncréa Ouest, Nantes, France. His publications are about machine/deep learning and computer vision. He served as a TPC member and a reviewer for many leading international conferences and journals.



ALADINE CHETOUANI (Senior Member, IEEE) received the master's degree in computer science from Pierre and Marie Curie University, France, in 2005, the Ph.D. degree in image processing from the University of Paris 13, France, in 2010, and the Habilitation degree from Université d'Orléans, in 2020. His Habilitation degree was titled "On the use of visual attention and deep learning for blind quality assessment of multimedia contents." From 2010 to 2011, he was a Postdoctoral Researcher with the L2TI Laboratory, University of Paris 13. In 2020, he was benefited from a CNRS delegation year with the L2S Laboratory, Centrale Supélec, Université Paris-Saclay, France. He is currently an Associate Professor with the PRISME Laboratory, Orléans, France. He led and participated in several research projects. He has supervised more than 20 students (Ph.D. and master's degrees). He is the coauthor of more than 120 research publications in international refereed journals and conference proceedings. His current interests include image quality, perceptual analysis, visual attention, and for cultural heritage using deep learning models for different multimedia contents (image, stereo, and 3-D). He serves as a reviewer for major conferences and journals in the field of image analysis and pattern recognition. He has co-edited different special issues in international journals (such as IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, *MTAP*, and *JEI*) and organized different special sessions in international conferences (IEEE ICIP and IEEE ICME). He has served for several program committees and as the Area Chair/Meta-Reviewer for conferences, such as IEEE ICIP, IEEE ICASSP, IEEE ICME, and QOMEX. He is also a member of the MMSP and IVMSP Technical Committees of the IEEE Signal Processing Society, in 2023. He is one of the chairs of the International Conference on Communication and Signal Processing & Information Technology (CSP) of the IEEE International Multi-Conference on Systems, Signals and Devices (SSD-22). He is also the General Chair of the Content-Based Multimedia Indexing 2023 Conference. He is also an Associate Editor of IEEE TRANSACTIONS ON MULTIMEDIA.



MOHAMMED EL HASSOUNI (Member, IEEE) received the Ph.D. degree in computer science from the University of Burgundy, in 2005, and the Habilitation degree from Mohammed V University, Rabat, in 2012. He was an invited professor at several universities, such as Bordeaux, Orléans, Dijon, and Konstanz. He is currently a Full Professor in computer science with Mohammed V University, where he is affiliated with the Faculty of Letters and Human Sciences and the LRIT Laboratory. His Ph.D. dissertation concerned the proposition of a new method using higher order statistics for video denoising in a severely corrupted environment. He has supervised more than 15 Ph.D. and master's students in the area of image and video processing. He led and participated in several national and international research projects. He has authored or coauthored numerous original journal articles and conference papers. His current research interests include the modeling and analysis of ND images, visual quality assessment, complex networks, and deep learning. He is also a member of the IEEE Signal Processing Society, INNS Moroccan Chapter, and several conference program committees. Since 2014, he has been the Co-Founder and the Co-Chair of the Quality of Multimedia Services QUAMUS Workshop. He was the Co-Chair of the "Multimedia, Computer Vision, and Image Processing" Track of ACS/IEEE International Conference on Computer Systems and Applications (AICCSA) 2015 and 2016 and the third International Conference on Advanced Technologies in Signal and Image Processing (ATSIP 2017). He was the Co-Chair of the COMBI 2017, ACMMIPH 2018, COMBI 2019, and CIVIA 2020 tracks of ACM-SAC. He co-edited different special issues, namely, the Special Issue on Perceptually Driven Visual Information Analysis of *JEI* (SPIE) and the Special Issue on Advances in Computational Intelligence for Multimodal Biomedical Imaging of *MTAP* (Springer). He serves as a reviewer for major conferences and journals in the field of image analysis and pattern recognition.

MAHER JRIDI received the engineering degree from the Graduate Telecommunication School, Sup'Com, Tunisia, in 2003, and the Ph.D. degree in electronics from the University of Bordeaux I, France, in 2007. Currently, he is a Professor with ISEN Yncrea Ouest. He is also the Head of the Vision and Data Analysis Group and a Research Referent with L@bISEN, Nantes. He has coauthored more than 80 journal articles or conference papers and edited many special issues. His research interests include digital electronics design for data processing algorithms, IoT physical layer communication, CNN embedding in reconfigurable architectures, and, more particularly, algorithm architecture co-design for resource-constrained systems. He received the professorial qualification from Université de Bretagne Occidentale, in 2018.