

Point cloud quality assessment using cross-correlation of deep features

Marouane Tliba, Aladine Chetouani, Giuseppe Valenzise, Frédéric Dufaux

▶ **To cite this version:**

Marouane Tliba, Aladine Chetouani, Giuseppe Valenzise, Frédéric Dufaux. Point cloud quality assessment using cross-correlation of deep features. 2nd Workshop on Quality of Experience in Visual Multimedia Applications (ACMMM QoEVMA 2022), ACM Multimedia, Oct 2022, Lisbon, Portugal. 10.1145/3552469.3555710 . hal-03801014

HAL Id: hal-03801014
https://hal.science/hal-03801014

Submitted on 6 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Point Cloud Quality Assessment Using Cross-correlation of Deep Features

Marouane Tliba
University of Orleans, PRISME
Orleans, France
marouane.tliba@univ-orleans.fr

Aladine Chetouani
University of Orleans, PRISME
Orleans, France
aladine.chetouani@univ-orleans.fr

Giuseppe Valenzise
University of Paris-Saclay, CNRS, CentraleSupélec
Paris, France
giuseppe.valenzise@l2s.centralesupelec.fr

Frederic Dufaux
University of Paris-Saclay, CNRS, CentraleSupélec
Paris, France
frederic.dufaux@l2s.centralesupelec.fr

ABSTRACT

3D point clouds have emerged as a preferred format for recent immersive communication systems, due to the six degrees of freedom they offer. The huge data size of point clouds, which consists of both geometry and color information, has motivated the development of efficient compression schemes recently. To support the optimization of these algorithms, adequate and efficient perceptual quality metrics are needed. In this paper we propose a novel end-to-end deep full-reference framework for 3D point cloud quality assessment, considering both the geometry and color information. We use two identical neural networks, based on a residual permutation-invariant architecture, for extracting local features from a sparse set of patches extracted from the point cloud. Afterwards, we measure the cross-correlation between the embedding of pristine and distorted point clouds to quantify the global shift in the features due to visual distortion. The proposed scheme achieves comparable results to state-of-the-art metrics even when a small number of centroids are used, reducing the computational complexity.

CCS CONCEPTS

• **Computing methodologies** → *Machine learning*.

KEYWORDS

Point Cloud, Quality Assessment, Cross Correlation, Deep Learning

ACM Reference Format:
Marouane Tliba, Aladine Chetouani, Giuseppe Valenzise, and Frederic Dufaux. 2022. Point Cloud Quality Assessment Using Cross-correlation of Deep Features. In *Proceedings of the 2nd Workshop on Quality of Experience in Visual Multimedia Applications (QoEVMA '22), October 14, 2022, Lisboa, Portugal.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3552469.3555710

1 INTRODUCTION

Recently, with the abundant availability of low-cost 3D acquisition devices, and the increased interest in immersive technologies, multiple formats have been proposed to represent three-dimensional content, offering sufficient flexibility to interact with the geometrical representation from any point of view [10][13]. In this context, 3D point clouds have become a successful representation, thanks to their ability to provide six degrees of freedom interaction in applications such as mixed reality, autonomous driving, cultural heritage, etc. A point cloud is as a set of unordered points, determined by their Cartesian's 3D coordinates and associated attributes such as color, reflectance, or normal vectors.

A drawback of point clouds is the sheer amount of data they require to be transmitted and stored. This has motivated in recent years the development of point cloud coding (PCC) tools, notably in standardization committees [3]. However, lossy compression methods may introduce significant visual distortion, which calls for effective methods to quantify the quality of experience of compressed point clouds. A number of point cloud quality metrics have been thus developed in the past few years. They can be classified into three main groups: Point-based, Feature-based and Projection-based metrics.

Point-based metrics such as Point-to-Point (Po2Po)[24], Point-to-Plane (Po2Pl)[11], Plane-to-Plane (Pl2Pl)[1] and Point-to-Mesh (Po2Mesh)[21], predict the quality through point-wise geometric and/or features distance between the reference PC and its distorted version. Po2Po measures the relative distance between point pairs to estimate the final quality, Po2Pl extends Po2Po by projecting the error vector along the local normal, and Pl2Pl quantify the quality through measuring the angular similarity between surfaces associated to the points from the reference and degraded contents. Po2Mesh creates a polygonal mesh from the reference sample and then compute the distance between each distorted point and the corresponding surface. Currently, MPEG is adopting Po2Po MSE and Po2Pl MSE with the associated PSNRs as the standard point cloud geometry quality metrics.

Feature-based PC quality metrics extract the geometry with the associated attributes from point-wise level in a global or local way. Among those metrics, we can cite PC-MSDM [17] that extends the 2D SSIM metric [25] to PC by considering local curvature statistics, the Geotex [8] metric that exploits the Local Binary Pattern (LBP)

[22] descriptors, and PCQM [18] that combines the geometry and color features.

In projection-based PC Quality Metrics, the 3D points or their associated features are projected into 2D regular grids, and later, 2D methods are applied on these views [2], or 2D grid of features [6][5].

All the aforementioned point based metrics or projection based metrics take long time for pre-processing. In particular, point based metrics do the computing with each point independently the nearest neighbor search is applied as many times as the number of points in the PC, in order to find the corresponding couples of points from the reference and distorted images. Thus, the computational complexity of metric computation is a critical aspect to improve. In this paper, we propose a novel deep quality metric that operates directly on a set of points. Unlike the point-wise distance mechanism adopted in previous methods, our approach compares the embedding of a sparse set of points (centroids) and their corresponding neighborhoods. This results in an important gain in terms of computational cost and efficiency. Also, to the best of our knowledge, for full-reference point cloud quality assessment, we are the first who consider operating on point cloud directly without any prior projection or voxelization.

The main contribution of this paper are detailed in what follow:

- We propose an efficient end-to-end full-reference method for point cloud quality assessment. The proposed method considers both local and global features on the native point cloud, without the need for voxelization or 2D projections. As an advantage regarding the state of the art, our method operates only on a small number of non-overlapped patches, which leads to important gains in computational complexity.
- To quantify the global feature shift due to visual distortion, we computed the cross-correlation on the embedding of pairs of 3D point patches. We interpret the cross-correlation as a measure of the relative representation displacement between a pair of reference-distorted point cloud regions.

Two publicly available datasets are used to show the effectiveness of the proposed method compared to state-of-the-art methods.

2 PROPOSED FRAMEWORK

The proposed quality assessment method consists in extracting relevant local features from a pair of compressed (distorted) and pristine point clouds, in such a way that the shift between the features enables the accurate prediction of the PC visual impairment. A general scheme of the proposed quality assessment method is shown in Figure 1. We extract features from a sparse set of corresponding point regions (from the original and distorted point clouds), using a shallow permutation invariant network, similar to the PointNet architecture [23]. The number of these regions is identified by local *centroids*, i.e., sampled points in the original point cloud. We consider both the geometry and color information as input to the embedding network. Next, we compute the cross-correlation on the embedding and measure their distance from the identity matrix, which would correspond to the case where the distorted points are the same as the original, i.e., no distortion and perfect quality. In other terms, the distance of the cross-correlation matrix from the target identity matrix is a proxy to the perceptual distortion of the compressed point cloud. The distance is computed to cover the global shift of 3D point sets, determined by their centroids, extracted from the original and compressed point clouds. In practice, to map the distance vector into target global mean opinion scores, the vector of distance is fed to a shallow regressor to estimate the final quality score.

2.1 Pre-Processing

The pristine and distorted PCs are firstly divided into set of local 3D patches by considering the (x, y, z) coordinates and the RGB color information. To do so, we employ the farthest point sampling [23] and K nearest neighbor algorithms as follows:

- From a given distorted PC X_d, we first apply the farthest point sampling algorithm to select M centroids (i.e. $C = \{P_{d1}, P_{d2}, P_{d3}, .., P_{dm}\}$) from the distorted image.
- In order to form a patch around each centroid, we then employ the K nearest neighboring clustering method. The sub-set of points of the i_{th} centroid, denoted as $\{P^1_{di}, P^2_{di}, P^3_{di}, .., P^K_{di}\}$, represents the i_{th} patch. K was fixed here to 512.
- We use the same set of centroids C for the pristine PC X_r as well, to form corresponding patches in the reference PC, we opted to select the set of centroids from the distorted version since it usually contains less number of points, so for each point from X_d there is a corresponding or a close point from X_r.
- Finally, we compute the position and color differences between each sub-set of points and its corresponding centroid from both of X_d and X_r. We obtain the final sub-set of points, denoted here S_{di} and S_{ri} respectively, that are fed as inputs to the parallel identical network model.

It is worth noting that the use of patches allows for decreasing the computational time and increases the model efficiency, where the information is extracted to sufficiently cover the intrinsic features from the local level in a parallelized way; and simultaneously allows the model to effectively compute the empirical cross-correlation to assess feature embedding shift globally.

2.2 Model Architecture

Figure 2, represents the architecture of the model encoder, which is based on PointNet [23]. Essentially, we extend the features extractor by a series of residual 1D convolution and remove the symmetry transformation network.

To sum up, the architecture of our model is composed of a feature embedding network including a residual symmetric invariant features extractor (i.e. series of 1D convolutions followed by a max pooling operation over point representation). In order to capture the global shift between the patches of the reference and distorted images, we compute the cross-correlation between the representation extracted from the set of patches S_{di} and S_{ri} respectively. We measure the relative entropy as distance between the empirical cross-correlation matrix and the identity matrix, which represents here the perfect correlation (i.e., zero distortion). The measured distance vector is lastly fed to a shallow regressor to interpolate the final quality score.

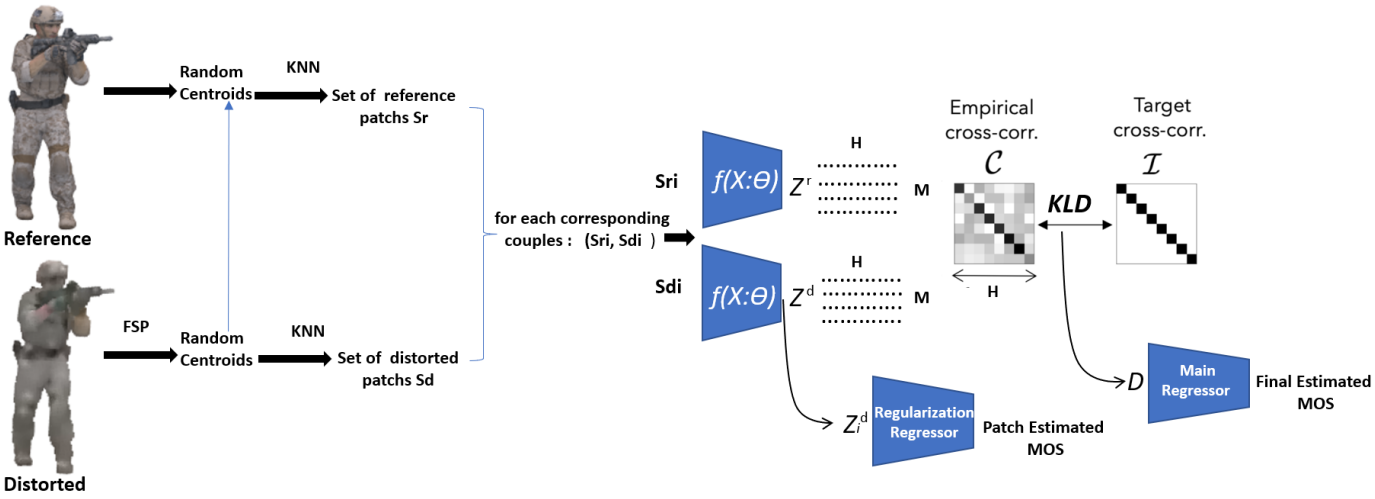


Figure 1: Overview of the proposed framework. A set of M random centroids are extracted from the original and distorted point cloud. An encoder neural network with point-wise convolutions is used to extract local features from the 3D patches around the selected centroids. The features extracted from the original and the distorted PC are then cross-correlated, yielding the cross-correlation matrix C. The distance between C and the identity matrix is then used to predict the final quality score. The regularization regressor forces the learned embedding to carry information about the target quality scores. M is the number of centroids. H is the feature dimension which is equal to 1024.

2.3 Empirical Cross Correlation

The main motivation of our model is inspired from recent advances in joint embedding methods [26][4][14]. Our method operates to measure the representation shift or relative displacement between the embedding of two identical networks fed with a set of corresponding patches from reference and distorted point clouds.

More specifically, the set of corresponding patches from the reference and distorted PCs, S_d and S_r respectively, are fed to f_θ, a deep neural network with learnable parameters θ. This yields a set of embedding Z^r and Z^d respectively. Z^r and Z^d are matrices of size $M \times H$, where $H = 1024$ in this paper is the size of the feature vectors, and M is the number of centroids. Without loss of generality, we remove the mean from each line of the matrices, such that each unit Z_i^r or Z_i^d has mean output 0 over the set of patches. We compute then the cross-correlation matrix C between the features of the distorted and original patches:

$$C_{jk} \triangleq \frac{\sum_i z^r_{i,j} z^d_{i,k}}{\sqrt{\sum_i (z^r_{i,j})^2}\sqrt{\sum_i (z^d_{i,k})^2}} \tag{1}$$

where i indexes patches samples, and j and k are indexes over the feature vector components. C is a square matrix of size $H \times H$, values varying in $[-1, 1]$, representing the anti-correlation, and the perfect correlation, respectively.

A cross-correlation matrix equal is equal to the identity if: a) the feature vectors of the original and distorted patches are the same or they are perfectly correlated; b) features elements in different positions of the feature vectors are uncorrelated. Condition a) is satisfied when the input distorted point cloud is the same to the reference, i.e., the distortion is zero. On the other hand, condition b) enforces the features to contain as much information as possible, and has been shown to avoid feature collapse and learn better representations [26]. We show in the ablation studies that computing the cross-correlation matrix and comparing it to identity leads in general to better results than simply optimizing condition a) alone, e.g., by computing the distance between the embedding of the original and compressed PCs.

2.4 Regularization network

Our architecture is completed with a shallow regularization regressor network whose target is the groun-truth MOS. The goal of this additional module is keeping the produced embedded vector Z_i^d aligned with the downstream task of PC quality assessment.

But also, the additional task oriented feedback from this small regressor help to decorrelate the different vector components of the embedding, which in turn, reduces the redundancy between output units.

2.5 Quality Estimation

To efficiently estimate the perceptual quality, we computed the distance between the empirical cross-correlation matrix C and the identity matrix I, denoted here as the target-correlation. To this end, we consider each row of C as a probability mass function, and we compute the Kullback-Leibler divergence with respect to the corresponding row of the identity matrix. This yields a vector of H divergences:

$$\mathcal{D}_j = \mathrm{KLD}(C_j || I_j) \tag{2}$$

where C_j and I_j denote the j-th row of the cross-correlation and identity matrix, respectively.

As last step, $\mathcal{D}$ is fed to a shallow regressor composed of three layer of 1D convolution, in order to produce one value that estimate the final quality score of the input PC.

2.6 Training Protocol

We trained our model in an end-to-end manner, using the mean square error as a loss function. The main regressor network (denoted in the equations as **Main**) is trained to create a mapping function between $\mathcal{D}$ and the mean opinion quality score Y, where the regularisation regressor (denoted in the equations as **Regularization**) is trained to create a mapping function between each Z_i^d and the mean opinion quality score Y.

Thus, the global loss function is defined as the sum of the two fallowing losses:

$$\mathcal{L}_1 = \text{MSE}\left(\textbf{Main}(\mathcal{D}), \mathcal{Y}\right), \tag{3}$$

$$\mathcal{L}_2 = \text{mean}\left(\sum_i \text{MSE}(\textbf{Regularization}(Z_i^d), \mathcal{Y})\right). \tag{4}$$

2.7 Implementation details

We implemented our model in PyTorch, initializing the parameters from scratch using Xavier initialization [19]. Since the initial PC is divided into patches we set the batch size to one. The model is trained using Adam optimizer [20] with initial learning rate of 0.000001, the number of epoch is varied based on the training folds.

3 EXPERIMENTS

We evaluate the effectiveness of our model on two publicly available benchmark datasets with subjective scores, which adopt different compression schemes with multiple encoding levels: **ICIP20** [15] and **PointXR** [9]. First, we provide a short description about the used datasets for evaluation, then we discuss the performance achieved by our model comparing to state of the art quality metrics, using different validation protocols.

ICIP20 is composed of 6 reference PCs from which 90 degraded versions were derived through three types of compression: V-PCC, G-PCC with triangle soup coding and G-PCC with octree coding [16][7][12]. Each reference PC was compressed using five different levels.

PointXR is composed of 5 PCs from which 45 degraded versions were derived through G-PCC with octree coding for geometry compression and, Lifting and RAHT for color compression.

Lastly, to farther emphasize the effectiveness and demonstrate the generalization ability of our network we carried out a cross dataset validation.

3.1 k-fold Cross Validation

To fairly study the performance and for the sake of providing a solid validation baseline protocol to future works, we randomly select 128 centroids for validation as we do during the training. We adopt a 6-fold cross validation protocol on ICIP20 dataset through splitting the dataset into training validation 6 times, where 6 refers to the number of reference point cloud samples. More precisely, at each iteration 5 reference point cloud samples and their compressed versions are used for training, and one reference point cloud sample and its compressed versions are used for testing. Lastly, to emphasize the effectiveness and validate the generalization ability of our model on unseen compression, we apply a cross-dataset validation protocol by training our model on ICIP20 and testing it on PointXR. Pearson Correlation Coefficient (PCC) and Spearman Rank-Order Coefficient Correlation (SROCC) are computed to evaluate the quality prediction ability of our method.

The outcomes of these measures varies in the range $[0, 1]$ as absolute values, where one corresponds to the best correlation, and zero means the total absence of correlation. The reported correlations are computed regarding each fold independently during the validation, and finally averaging the results over the different folds.

Table 1. shows the performance of our method on ICIP20 dataset compared to a well representative set of state-of-the-art methods. As can be seen, our model is quite competitive with state-of-the-art methods, especially in terms of ranking (i.e. SROCC) of the compression degradation effects. It is also worth to note that unlike our method, all of listed state-of-the-art methods require heavy pre-processing which result in expensive computation.

Table 1: Results obtained on ICIP20 dataset

Model	PLCC ↑	SROCC ↑
po2point MSE	0.946	0.934
po2plane MSE	**0.959**	0.951
PSNR po2point MSE	0.868	0.855
PSNR po2point HAU	0.548	0.456
PSNR po2plane HAU	0.580	0.547
color Y MSE	0.876	0.892
color Cb MSE	0.683	0.694
color Cr MSE	0.594	0.616
color Y PSNR	0.887	0.892
color Cb PSNR	0.693	0.694
color Cr PSNR	0.626	0.617
pl2plane AVG	0.922	0.910
pl2plane MSE	0.925	0.912
PCQM	0.796	0.832
Our	**0.947**	**0.973**

3.2 Cross-dataset evaluation

Table 2. shows the results obtained for the cross-dataset evaluation. As can be seen, high correlations are reached by our method, outperforming some of the compared ones. This results shows the generalization ability of our method to predict the quality of unseen PCs. It is remarkable that our method is the most consistent one comparing to the performance achieved on ICIP20.

Table 2: Results obtained by training the model on ICIP20 and testing it on PointXR

Model	PLCC ↑	SROCC ↑
po2pointMSE	0.887	0.978
po2planeMSE	0.855	0.942
PSNRpo2pointMSE	0.983	0.978
PSNRpo2planeMSE	0.972	0.950
Our	**0.981**	**0.964**

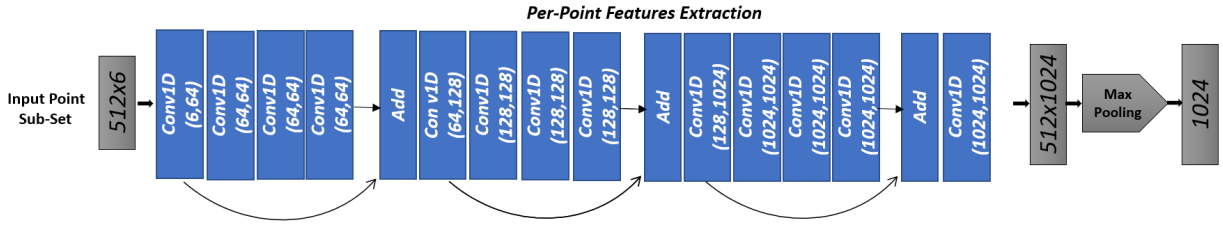


Figure 2: Model Encoder, $f(X:\theta)$

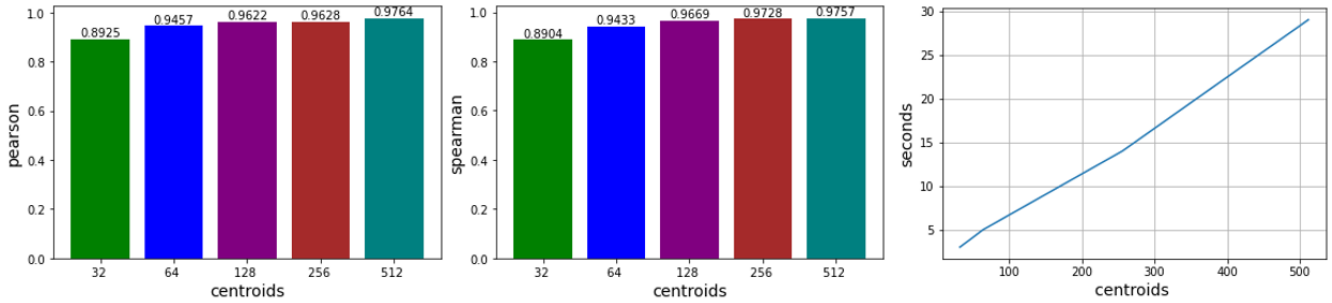


Figure 3: Effect of the number of centroids on the quality prediction performance and the computation time. Left: PLCC. Middle: SROCC. Right: computation time (seconds).

3.3 Ablation study

In this section, we study the impact of the number of centroids on the performance and the computational complexity. We also compare the KLD-based quality estimation with component-wise MSE to verify the importance of using the whole cross-correlation matrix to assess distances.

3.3.1 ***Impact of the number of centroids.*** The number of centroids is one of the key points of our method. Using a sparse set of centroids considerably reduces the complexity of our method since only certain parts of the PCs are considered. However, it can be expected that by decreasing the number of centroids also the precision of quality estimation decreases. In order to study the impact of the number of centroids on performance, we varied it from 32 to 512 and calculated the correlations reached for each of them. We also show the corresponding computation time.

Fig. 3 shows the Pearson and Spearman correlations obtained on one fold of the ICIP20 dataset. As can be seen, the correlations increase with the number of selected centroids. Even using only 64 centroids, we can achieve correlations which are competitive with state-of-the-art point-based methods, but using a fraction of their complexity. As can be seen, setting the number of centroids to 128 leads to a calibrated balance between correlations and the required time for processing. It is worth noting that the performance can further increase by adding more centroids.

3.3.2 ***Impact of using KLD to measure cross-correlation shift.*** In this experiment, we aim to prove the effectiveness of using the full cross-correlation matrix as a proxy to the perceptual quality, and in particular the use of KLD as distance with respect to the identity matrix. To this end, we compared the correlations obtained by computing the proposed KLD-based quality estimation with those obtained by computing simply the Mean Square Error (MSE) distance between feature vectors of distorted and original patches.

Table 3: Comparison of KLD and MSE on feature vectors (considering the whole and only the diagonal of cross-correlation matrix), as a proxy to predicted quality scores.

Distance	PLCC ↑	SROCC ↑
MSE Cross-Correlation Diagonal Shift	0.859	0.929
MSE Whole Cross-Correlation Shift	0.865	0.887
KLD Cross-Correlation Diagonal Shift	0.829	0.884
KLD Whole Cross-Correlation Shift	**0.947**	**0.973**

This experiment was carried out of ICIP20 dataset, over all folds using k-fold cross validation. We retrained the network to use MSE to measure the shift between embedding of distorted and original patches, in two options considering the whole cross-correlation matrix C and only the diagonal elements of C, instead of KLD.

Moreover, to prove the importance of attending to the whole cross-correlation matrix C elements in estimating the shift between the embedding using the KLD as a distribution displacement metric, we retrained the network from scratch on the whole folds using the KLD only on diagonal elements. Table 3. shows the correlations achieved for the four strategies. As can be seen, the proposed KLD-based quality estimation approach on the whole cross-correlation matrix C outperformed all the rest of options by a significant gain in terms of performance, demonstrating the usefulness of considering the relation between feature elements, and confirming the finding in [26] that enforcing independence among feature components can help learning more discriminative representations.

4 CONCLUSION

In this paper we were set out to build, implement and test a novel end-to-end deep learning based method for perceptual quality assessment of 3D point cloud data. Unlike previous deep learning-based methods, ours works directly on the native point cloud representation without requiring any projection or transformation step as voxelization. A main element of our method is the computation of feature cross-correlation and its distance with respect to the identity cross-correlation matrix as a way to capture the global shift in the feature domain induced by compression affects. To the best of our knowledge, our method is the first deep full-reference point cloud quality metric that operates directly on point cloud data and considers geometry and color information. In addition, a highlight of our method is the reduced computational complexity compared to point-based methods, which require computing k-nearest neighbors for each point. Instead, we only compute nearest neighbors for a set of sparse centroids, without significant losses in performances.

REFERENCES

[1] E. Alexiou and T. Ebrahimi. 2018. Point Cloud Quality Assessment Metric based on Angular Similarity. In *IEEE ICME-W*.

[2] Salima Bourbia, Ayoub Karine, Aladine Chetouani, and Mohammed El Hassouni. 2022. Blind Projection-based 3D Point Cloud Quality Assessment Method using a Convolutional Neural Network, Giovanni Maria Farinella, Petia Radeva, and Kadi Bouatouch (Eds.). 518–525. https://doi.org/10.5220/0010872700003124

[3] Chao Cao, Marius Preda, Vladyslav Zakharchenko, Euee S. Jang, and Titus Zaharia. 2021. Compression of Sparse and Dense Dynamic Point Clouds—Methods and Standards. *Proc. IEEE* 109, 9 (2021), 1537–1558. https://doi.org/10.1109/JPROC.2021.3085957

[4] Xinlei Chen and Kaiming He. 2021. Exploring Simple Siamese Representation Learning. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2021), 15745–15753.

[5] Aladine Chetouani, Maurice Quach, Giuseppe Valenzise, and Frederic Dufaux. 2021. Convolutional Neural Network for 3D Point Cloud Quality Assessment with Reference. In *2021 IEEE 23rd International Workshop on Multimedia Signal Processing (MMSP)*. 1–6. https://doi.org/10.1109/MMSP53017.2021.9733565

[6] Aladine Chetouani, Maurice Quach, Giuseppe Valenzise, and Frederic Dufaux. 2021. Deep Learning-Based Quality Assessment Of 3d Point Clouds Without Reference. In *2021 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*. 1–6. https://doi.org/10.1109/ICMEW53276.2021.9455967

[7] Ricardo L. de Queiroz and Philip A. Chou. 2016. Compression of 3D Point Clouds Using a Region-Adaptive Hierarchical Transform. *IEEE Transactions on Image Processing* 25 (2016), 3947–3956.

[8] Rafael Diniz, Pedro Garcia Freitas, and Mylène C. Q. Farias. 2020. Towards a Point Cloud Quality Assessment Model using Local Binary Patterns. In *QoMEX*. 1–6. https://doi.org/10.1109/QoMEX48832.2020.9123076

[9] N. Yang E. Alexiou and T Ebrahimi. 2020. PointXR: A Toolbox for Visualization and Subjective Evaluation of Point Clouds in Virtual Reality. (2020).

[10] A. Chetouani et al. 2020. Classification of engraved pottery sherds mixing deep-learning features by compact bilinear pooling. *Pattern Recognition Letters* 131 (2020), 1–7. https://doi.org/10.1016/j.patrec.2019.12.009

[11] D. Tian et al. 2017. Geometric distortion metrics for point cloud compression. In *IEEE ICIP*.

[12] E. Pavez et al. 2018. Dynamic polygon clouds: representa- tion and compression for vr/ar. 7 (2018).

[13] M. Tliba et al. 2021. 2D-Based Saliency Prediction Framework for Omnidirectional–360° Video. In *11th International Conference of Pattern Recognition Systems (ICPRS 2021)*, Vol. 2021. 31–37. https://doi.org/10.1049/icp.2021.1459

[14] M. Tliba et al. 2022. Self Supervised Scanpath Prediction Framework for Painting Images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. 1539–1548.

[15] S. Perry et al. 2020. Quality Evaluation Of Static Point Clouds Encoded Using MPEG Codecs. (2020), 3428–3432.

[16] S. Schwarz et al. 2019. Emerging MPEG Standards for Point Cloud Compression. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* 9, 1 (2019), 133–148.

[17] J. Digne G. Meynet and G. Lavoué. 2019. PC-MSDM: A quality metric for 3D point clouds. (2019).

[18] J. Digne G. Meynet, Y. Nehmé and G. Lavoué. 2020. PCQM: A Full-Reference Quality Metric for Colored 3D Point Clouds. (2020).

[19] Xavier Glorot and Yoshua Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. In *AISTATS*.

[20] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. *CoRR* abs/1412.6980 (2015).

[21] C. Rochinni P. Cignoni and R. Scopigno. 1998. Metro: measuring errors on simplified surfaces,. In *Computer Graphics Forum*, Vol. 17. 167–174.

[22] Matti Pietikäinen and Guoying Zhao. 2016. Two decades of local binary patterns: A survey. *CoRR* abs/1612.06795 (2016). arXiv:1612.06795 http://arxiv.org/abs/1612.06795

[23] C. Qi, L. Yi, Hao Su, and Leonidas J. Guibas. 2017. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. (2017).

[24] C. Tulvan R. Mekuria, Z. Li and P. Chou. 2016.. Evaluation criteria for PCC (Point Cloud Compression). In *in ISO/IEC MPEG Doc*, Vol. II. N16332, 803–806.

[25] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* 13, 4 (2004), 600–612. https://doi.org/10.1109/TIP.2003.819861

[26] Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. 2021. Barlow Twins: Self-Supervised Learning via Redundancy Reduction. In *ICML*.