



A No Reference Deep Quality Assessment Index for 3D Colored Meshes

Zaineb Ibork, Anass Nouri, Olivier Lézoray, Christophe Charrier, Raja Touahni

► To cite this version:

Zaineb Ibork, Anass Nouri, Olivier Lézoray, Christophe Charrier, Raja Touahni. A No Reference Deep Quality Assessment Index for 3D Colored Meshes. IEEE International Conference on Systems, Man, and Cybernetics (SMC2024), IEEE, Oct 2024, Sarawak, Malaysia. hal-04775149

HAL Id: hal-04775149

<https://hal.science/hal-04775149v1>

Submitted on 9 Nov 2024

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.

A No Reference Deep Quality Assessment Index for 3D Colored Meshes

Zaineb Ibork^{1,2}, Anass Nouri^{1,2}, Olivier Lézoray², Christophe Charrier², Raja Touahni¹

Abstract—The advent of 3D data has revolutionized various industries, from architecture and engineering to healthcare and entertainment, enabling more precise simulations and realistic visualizations. However, 3D data is susceptible to noise and loss during generation and transmission, making quality assessment crucial for ensuring accuracy and usability. While existing literature addresses quality assessment for 3D point clouds and meshes separately, a gap exists in assessing the quality of 3D colored meshes due to the lack of reference datasets. This paper proposes an approach for No Reference 3D Colored Mesh Visual Quality Assessment (CMVQA), based on previous work related to quality assessment of 3D non colored meshes quality assessment. Our approach combines geometric and color features with spatial domain features extracted from mesh projections. Through extensive experiments and comparisons with full-reference metrics, including image quality metrics, our proposed approach demonstrates superior performance.

Index Terms—3D Colored Mesh, Mesh Visual Quality Assessment, Convolutional Neural Network, Deep learning, No reference quality assessment, Brisque.

I. INTRODUCTION

The emergence of 3D data has transformed industries ranging from architecture and engineering to healthcare and entertainment, enabling more accurate simulations and realistic visualizations. This shift has opened new avenues for innovation and creativity, driving advancements in fields such as virtual reality, 3D printing, and computer-aided design. However, 3D data is prone to geometry/color noise and compression/simplification loss during generation and transmission procedures. Therefore, the quality assessment of 3D data has recently become crucial to ensure accuracy and fidelity, thereby guaranteeing effectiveness and usability. 3D data can take two distinct type of forms: 3D point clouds and 3D meshes. Both representations have been considered in the literature for quality assessment, see e.g., [1] and [2].

However, 3D point clouds and 3D meshes are two very different representations. 3D point clouds are made of a set of 3D points whereas 3D meshes also provide a graph representing a triangulation of the 3D surface of the object. The consequence is that 3D meshes naturally have a better visual quality than 3D point clouds, as no holes are obtained during their rendering process, making them more suitable for practical applications. Figure 1 shows this difference. In

*This work received funding from PHC TOUBKAL TBK/22/142-CAMPUS N°47259YH.

¹SETIME Laboratory, Information Processing and A.I Team, Faculty of Sciences, Ibn Tofail University, Kenitra, Morocco {zaineb.ibork, anass.nouri, touahni.raja}@uit.ac.ma

²Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, 14000 Caen, France {olivier.lezoray, christophe.charrier}@unicaen.fr



Fig. 1: Visual difference between a 3D colored point cloud and a 3D colored mesh, the latter has a better visual quality, as no holes do appear.

this paper we aim at designing an approach for the estimation of the quality of 3D colored meshes without any reference.

In the literature, many approaches have been proposed to estimate the quality of 3D meshes without reference, see [2]–[8]. All these approaches consider uncolored meshes and the distortions used to build a database for evaluation are solely based on independent geometric perturbations related to noise, smoothing and compression. In contrast, some of the no reference quality assessment approaches for 3D point cloud consider the color information in addition to geometry, see [1], [9], [10]. As a consequence, there is a gap in the literature for the estimation of the quality of 3D colored meshes. The reason for this is mainly due to the lack of reference datasets. Fortunately, Nehme *et al.* have recently introduced in [11] the first subject rated 3D colored mesh dataset, named the CMDM database. They have also proposed an approach for full-reference quality assessment of 3D colored meshes.

In this paper, we propose to adapt our recent work [8] for no-reference 3D mesh quality assessment to the case of 3D Colored Mesh Visual Quality Assessment (CMVQA). In particular, we extend our previous work by considering the learned combination of several quality indices extracted at both the view and patch levels. As some no-reference 3D Point Cloud Quality Assessment (PCQA) approaches have also considered the CMDM database for evaluation, we will compare our approach to these works [12].

In the sequel we introduce the approach, present the obtained results and conclude.

II. NO-REFERENCE 3D COLORED MESH QUALITY ESTIMATION

A. Flowchart of the approach

To assess the quality of colored meshes, our proposed method, namely Colored Mesh Visual Quality Assessment (CMVQA), follows a series of steps. Initially, multiple 2D projections, referred to as views, are generated by varying the viewpoint around the mesh. These views may contain irrelevant white background, so they are cropped based on the mesh bounding box. Subsequently, each processed view is divided into four overlapping patches. These 2D images, both views and patches, are then input into a pre-trained convolutional network (VGG16 [13] in our case), to extract deep features representing the image content. The resulting feature vectors are utilized to predict the quality of each view or patch using a Multi-Layer Perceptron (MLP) regressor, which takes the Mean Opinion Score (MOS) associated with each mesh M_i as reference. As each mesh comprises multiple views and patches, a vector of quality scores is obtained. To enhance this vector, we propose to augment it with the scores estimated by the BRISQUE no-reference image quality index [14] on views and patches. This augmented vector of quality scores is then utilized to compute the Predictive Mean Opinion Score (PMOS) for the 3D mesh. The estimation can be achieved through either averaging, as in our previous work [8], or via non-linear regression. Figure 3 illustrates the outlined approach, and subsequent sections will delve into each step of our proposed Mesh Visual Quality Assessment (MVQA) approach. This approach extends our previous work on uncolored 3D meshes, where we introduced a no-reference mesh quality assessment index based on deep convolutional features named DCFQI (Deep Convolutional Features Quality Index) [8].

B. 2D projections construction

Initially, we aim to generate diverse 2D renderings of each 3D mesh M_i from a database containing N meshes. To ensure consistent positioning, the centroids of the meshes are shifted to the origin of the coordinate system. Subsequently, each mesh is rendered from 11 different viewpoints by systematically adjusting azimuth (θ_a) and elevation (θ_e) angles in increments of $\pi/3$ (60 degrees) as illustrated in Figure 2. This technique ensures varied perspectives and detailed views of the mesh. Following rendering, the obtained 2D views, denoted as V_i^j , undergo cropping and resizing to eliminate white background and standardize their size to 512×512 pixels. Additionally, to capture intricate details, we extract four overlapping patches of size 288×288 pixels from each 2D view. This strategy enhances information extraction compared to previous methods that utilized smaller patches [2]. We will denote the extracted patches from the 2D views V_i^j of a mesh M_i by $P_i^{j,k}$ with $j \in [1, 11]$ and $k \in [1, 4]$. The extracted views and patches undergo global normalization, optimizing the quality assessment process. This pre-processing prepares the data for subsequent analysis, ensuring comprehensive evaluation of visual quality.

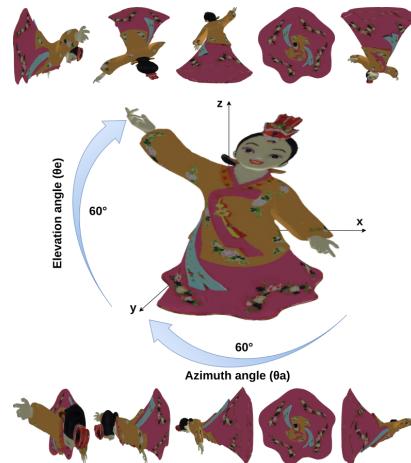


Fig. 2: Illustrating viewing angles in the rendering process of a coloured 3D mesh: Azimuth angle (θ_a) in the horizontal plane with $z = 0$ and Elevation angle (θ_e) from the xz plane with $y = 0$.

C. Quality estimation of a 2D projection

Through the pre-processing phase, we have derived two distinct datasets from the original database \mathcal{B} comprising N meshes: 1) \mathcal{B}_V , focused solely on views (2D images), and 2) \mathcal{B}_P , dedicated to patches. The $\mathcal{B}_V = V_i^j$ dataset contains $N \times 11$ images, with i ranging from 1 to N and j ranging from 1 to 11. Meanwhile, the $\mathcal{B}_P = P_i^{j,k}$ dataset encompasses $N \times 11 \times 4$ images, where i spans from 1 to N , j spans from 1 to 11, and k spans from 1 to 4. For simplicity in notation, we will represent an image from the set \mathcal{B}_S as I_i^j , where S can be either V or P , i ranges from 1 to N , j ranges from 1 to N_S , and $N_V = 11$ while $N_P = 44$.

Our aim is to assess the quality of 2D projection images derived from 3D meshes, utilizing the mean opinion score (MOS) of the corresponding mesh as a reference. To achieve this, we employ a pre-trained VGG16 convolutional network [13] to extract deep features from each image I_i^j , which is resized to 224×224 before processing. Denoting the feature extractor as ϕ , it transforms the input image into a flattened vector of size 25088. Each image I_i^j in dataset \mathcal{B}_S is thus represented by the feature vector $\phi(I_i^j)$, serving as input to a shallow multi-layer perceptron (MLP) regressor. The objective is to predict a quality score close to the MOS of the associated mesh M_i . With separate datasets for views (\mathcal{B}_V) and patches (\mathcal{B}_P), two distinct MLP regressors, denoted as MLPR_S , are trained to estimate quality scores. The predicted image mean opinion score (PIMOS) for each image I_i^j is obtained as $\text{PIMOS}_S(I_i^j) = \text{MLPR}_S(\phi(I_i^j))$, where S indicates whether I_i^j is a view (V) or a patch (P), and i, j iterate over the respective datasets.

We have just demonstrated how to train models to estimate the quality of 2D projections from datasets of views or patches. However, since 2D projections are essentially images, another avenue to explore is the No-Reference Im-

age Quality Assessment (IQA) metrics. One such metric, BRISQUE (Blind/Referenceless Image Spatial Quality Evaluator), stands out in the literature [14]. BRISQUE evaluates perceived image quality without requiring a reference image for comparison. It achieves this by analyzing spatial domain features within the image, such as local mean and standard deviation. A machine learning model is then trained on a dataset of natural images with known quality scores to predict the image quality score. Higher BRISQUE scores typically indicate lower image quality, while lower scores suggest higher image quality. We propose to incorporate BRISQUE as a no-reference IQA method to estimate the quality of 2D projections. BRISQUE offers the advantage of being referenceless and pre-trained. We denote this estimation as $\text{PIMOS}_S^B(I_i^j) = \text{BRISQUE}_S(I_i^j)$, where $S \in \{V, P\}$, $i \in [1, N]$, and $j \in [1, N_S]$.

D. CMVQA from 2d projection quality scores

Once a regressor MLPR_S has been trained on a dataset \mathcal{B}_S , a vector of N_S image-based quality scores is obtained for each mesh $M_i : \mathbf{PQ}_S(M_i) = [\text{PIMOS}_S^B(I_i^j) : j \in [1, N_S]]^T$. A similar vector of scores can be obtained with the BRISQUE IQA index: $\mathbf{PQ}_{S,B}(M_i) = [\text{PIMOS}_S^B(I_i^j) : j \in [1, N_S]]^T$. Finally, as we have two datasets \mathcal{B}_S for views and patches, we obtain four different vectors depending whether the 2D quality evaluators are performed on views or patches:

- $\mathbf{PQ}_V(M_i)$ for quality scores with MLPR_V on views,
- $\mathbf{PQ}_P(M_i)$ for quality scores with MLPR_P on views' patches,
- $\mathbf{PQ}_{V,B}(M_i)$ for quality scores with BRISQUE on views,
- $\mathbf{PQ}_{P,B}(M_i)$ for quality scores with BRISQUE on views' patches.

To estimate the quality of a mesh M_i , a fusion strategy is needed to aggregate all the $\mathbf{PQ}_{S,*}(M_i)$ scores into a single one that provides the global mesh quality. We quote this final quality score of a mesh as the Predicted Mesh Mean Opinion Square, denoted by PMMOS. We consider two aggregation strategies: 1) an averaging and 2) a MLP non-linear regression. Both will be performed on a vector of quality scores \mathbf{PQ} .

The first aggregation scheme, the averaging, can be expressed as:

$$\overline{\text{PMMOS}}(M_i) = \frac{1}{|\mathbf{PQ}|} \sum_{j=1}^{|\mathbf{PQ}|} \mathbf{PQ}^j(M_i) \quad (1)$$

with \mathbf{PQ}^j the j -th element of the vector \mathbf{PQ} , and $|\mathbf{PQ}|$ the cardinality of \mathbf{PQ} .

The second aggregation scheme, the MLP non-linear regressor, can be expressed as:

$$\text{PMMOS}(M_i) = \text{MLPR}(\mathbf{PQ}(M_i)) \quad (2)$$

So far we do not have mentioned what are the elements that do constitute the vector \mathbf{PQ} . We could solely use the quality scores on views and patches separately, but we could also make the most of them by using both. Therefore, we will investigate the following configurations:

- $\mathbf{PQ}(M_i) = \mathbf{PQ}_V(M_i)$: the vector of 11 quality scores estimated from views with MLPR_V
- $\mathbf{PQ}(M_i) = \mathbf{PQ}_P(M_i)$: the vector of 44 quality scores estimated from views' patches with MLPR_P
- $\mathbf{PQ}(M_i) = \mathbf{PQ}_V(M_i) \cup \mathbf{PQ}_P(M_i)$: the vector of 55 quality scores estimated from views and views' patches with MLPR_V and MLPR_P
- $\mathbf{PQ}(M_i) = \mathbf{PQ}_{V,B}(M_i)$: the vector of 11 quality scores estimated from views with BRISQUE
- $\mathbf{PQ}(M_i) = \mathbf{PQ}_{P,B}(M_i)$: the vector of 44 quality scores estimated from views' patches with BRISQUE.
- $\mathbf{PQ}(M_i) = \mathbf{PQ}_{V,B}(M_i) \cup \mathbf{PQ}_{P,B}(M_i)$: the vector of 55 quality scores estimated from views and views' patches with BRISQUE.
- $\mathbf{PQ}(M_i) = \mathbf{PQ}_V(M_i) \cup \mathbf{PQ}_P(M_i) \cup \mathbf{PQ}_{V,B}(M_i) \cup \mathbf{PQ}_{P,B}(M_i)$: the vector of 110 quality scores estimated from views and views' patches with MLPR_V , MLPR_P , and BRISQUE.

All these configurations will give rise to different CMVQA methods that we will compare in the next section.

III. EXPERIMENTAL RESULTS

A. Database

We conduct our experimental evaluation on the CDM database, the sole one with MOS ground truth currently existing in the state of the art that concerns 3D colored meshes [11]. This dataset has five source models (see Figure 4) subjected to geometry and color distortions. The distortions include uniform geometric quantization, uniform LAB color quantization, color-ignorant simplification, and color-aware simplification, each with four different strength levels. In total, the CDM database contains 80 distorted models, each associated with five subjective quality scores based on various viewpoints and animation types. To obtain a single quality score per mesh, we compute the average of these subjective scores. Figure 5 showcases a sample of the four distorted colored meshes from the CDM Dataset along with their source model.

B. Metrics

To evaluate the performance of our Predicted Mean Opinion Scores (PMOS) against the ground truth provided in the database, we employ two criteria. Firstly, we use the Pearson Linear Correlation Coefficient (PLCC) to gauge the prediction accuracy. Secondly, we utilize the Spearman Rank-Order Correlation Coefficient (SROCC) to measure the monotonicity of the predictions. Higher PLCC and SROCC values indicate better correlation with human visual perception. It's worth noting that SROCC operates solely on the rank of the data and does not consider the relative distance between data points.

C. Training and evaluation protocol

1) *MLP Regressors training* : For 2D projections quality estimation, the MLP regressor architecture (MLPR_S) consists of a single hidden layer with 512 neurons, with a Rectified Linear Unit (ReLU) activation function. A dropout

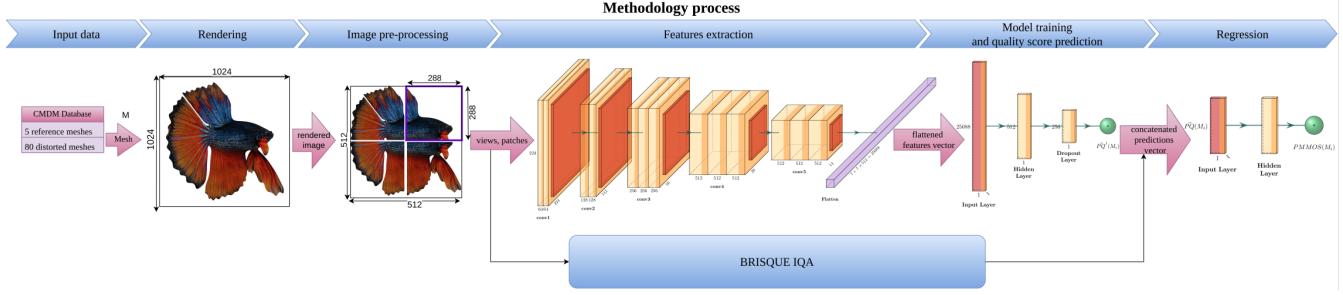


Fig. 3: Illustration of the proposed Mesh Visual Quality Assessment approach CMVQA.



Fig. 4: The 3D colored meshes from the CMDM dataset. From top left to bottom right: Samourai, Fish, Ari, Chameleon, Aix.

regularization with a rate of 0.5 is applied to prevent overfitting, and weight initialization employs the Glorot uniform method. Training employs Mean Average Error loss and RMSprop optimizer with a fixed learning rate of 0.001. A batch size equal to one-third of the training set size is found to yield the best correlation scores. Early Stopping technique is used for regularization and model generalization enhancement.

For the aggregation of image quality scores, another MLP regressor (MLPR) is employed, also consisting of a single hidden layer with ReLU activation and a sigmoid output layer. The number of neurons in the hidden layer equals the number of inputs (i.e., $|\mathbf{PQ}|$). The training procedure for all regressors is referred to as the "Base Model (BM)".

2) Leave-One-Model-Out Evaluation Protocol: To assess the accuracy of the MLP regressors MLPR_S , we use a Leave-One-Model-Out Cross-Validation (LOMO-CV) procedure. During training, all meshes except one source model and its distorted versions are considered. This means that when training a MLP regressor MLPR_S , the images associated with a specific mesh are excluded. The trained neural network is then tested on these excluded images to evaluate its performance on unseen data. This LOMO-CV process is repeated for each model in the dataset, ensuring an objective assessment of the MLPR model as it is evaluated on strictly independent data that it hasn't been trained on.

3) Evaluated configurations: In the approach description, we have discussed various setups based on how the vector \mathbf{PQ} is formed. Its size can range from 11 (only views), 44 (only patches), 55 (both views and patches), to 110 (both views and patches along with BRISQUE scores). Additionally, we can opt for either a base model or a cumulative model depending on the chosen training method (will be detailed thereafter). To distinguish and compare these setups, we will use a specific naming convention. The different setups used to design CMDQA will be indicated by terms separated by dashes:

CMDQA -	BM	-	A	-	VPB
	RCM		R		VP

CMDQA -	BM	-	A	-	VPB
	RCM		R		VP
			P		P
			V		V

where the first part refers to the training protocol (BM refers to the Base Model, RCM to the Retrained Cumulative Model), the second part refers to the aggregation scheme (A for Average Aggregation i.e., Eq.1, R for MLP Regression i.e., Eq.2), and the third part specifies how the aggregated quality scores are obtained: V refers to quality scores from views, P to quality scores from patches, VP to quality scores from both views and patches, and VPB to quality scores from both views and patches with the addition of BRISQUE scores. When exclusively using BRISQUE, the configurations are denoted as follows:

BRISQUE -	R	-	VP
			P
			V

For BRISQUE, there are no base or cumulative models, and we will exclusively use MLP Regression for aggregating scores. These setups will help evaluate the performance of BRISQUE IQA solely for CMVQA. The various configurations are summarized in Table III and depicted in Figure 3 (for MLP Regression aggregation only).

D. Results and evaluation

1) Our proposed metric performance versus the state-of-the-art: In this section, we compare our proposed Base Model (CMDQA-BM) that considers deep geometric and color features (extracted with VGG16) along with the spatial

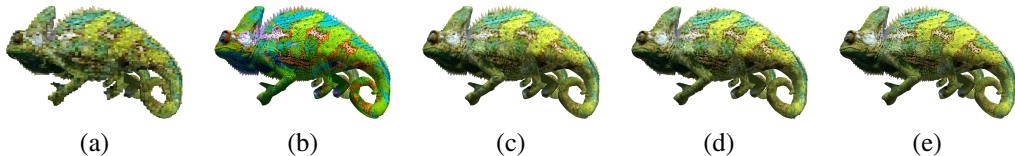


Fig. 5: Visual difference between the different noise types applied to the Chameleon mesh. (a) Uniform geometric quantization. (b) Uniform LAB color quantization. (c) Color-ignorant simplification. (d) Color-aware simplification. (e) Original Chameleon mesh.

domain features within the mesh projections (computed with BRISQUE). We will see in the sequel that our model achieves a better performance than actual state-of-the-art approaches. To compare the performances of our proposed metric versus the existing No-Reference (NR) and Full-Reference (FR) metrics, we compare the SROOC and PLCC correlation coefficients on the CDM database. With the LOMO-CV training protocol, we obtain scores per mesh that can be either averaged or computed for the meshes altogether. We will consider both scores for comparison purposes as in the state-of-the-art these two scores are considered.

To ease the reading of the tables, we have put: best rates bolded in each column for each category of approach, best SOTA approach is shown with color ■, the best BRISQUE protocol is shown with color ■, our best Base Model is shown with color ■.

In Table I, a comparison between different state-of-the-art (SOTA) NR metrics and our proposed metrics is provided to analyze their performances in predicting the quality scores of the CDM database.

Among the SOTA NR metrics, NCMQE [15] and LGF-CMVQA [16] stand out with the highest SROOC and PLCC values, indicating strong correlations with human perception. In particular NCMQE is similar to 3D-NSS [17] but trained specifically on colored meshes whereas 3D-NSS is trained on 3D colored point clouds. This shows that it is better to directly consider 3D colored meshes for learning the quality. On the other hand, metrics like NR-SVR [18], NR-CNN [4], NIQUE [19], and BRISQUE [14] show moderate correlations, while NR-GRNN [3] and 3D-NSS exhibit stronger correlations.

Looking now to our proposed metrics, we can see that learned combination of BRISQUE scores from views and patches enables to enhance its SOTA performance (which was computed using a logistic regression from the average score of the 2D views [12]). However, it is still far from the best SOTA approach. In contrast, our proposed metric CMDQA-BM achieves very competitive or better scores. One can see that the quality is always better when estimated from views than from views' patches, but the combination of all the quality scores (VPB) from both views and patches enables to obtain the better performance, overpassing the best SOTA scores. This shows the complementary effect of these quality scores obtained from views and patches.

Unfortunately, the detailed LOMO-CV results on the CDM dataset for the previously mentioned NR metrics are

Comparison of SROOC and PLCC values with NR metrics			
Type	Metric	Average SROOC	PLCC
No-Reference	NR-SVR [18]	0,449	0,608
	NR-GRNN [3]	0,695	0,660
	NR-CNN [4]	0,502	0,520
	3D-NSS [17]	0,875	0,863
	NIQUE [19]	0,477	0,406
	BRISQUE [14]	0,488	0,579
	GMS-3DQA [12]	0,839	0,876
	NCMQE [15]	0,879	0,889
	LGF-CMVQA [16]	0,899	0,901
Ours	Brisque-R-V	0,515	0,603
	Brisque-R-P	0,605	0,598
	Brisque-R-VP	0,579	0,613
	CMDQA-BM-A-V	0,953	0,962
	CMDQA-BM-A-P	0,872	0,925
	CMDQA-BM-R-V	0,963	0,965
Ours	CMDQA-BM-R-P	0,821	0,897
	CMDQA-BM-R-VP	0,774	0,848
	CMDQA-BM-R-VPB	0,979	0,970

TABLE I: Comparison of SROOC and PLCC values of our proposed methods with No Reference state of the art approaches

not available from the corresponding papers. Therefore, we also compare the performance of our metrics for each mesh of the dataset against a set of FR metrics. Specifically, we compare our approach with the three metrics proposed by Nehmé *et al.* [20] depending on the kind of features they are using: CDM_Geo (geometry features only), CDM_Col (color features only), and CDM (composed of four features: curvature contrast, lightness contrast, structure, and chroma comparison). Additionally, we compare our results with three additional FR Image Quality Metrics (FR-IQMs) Nehmé *et al.* also used: a classical color distance D LAB, SSIM [21], HDR-VDP2 [22], and iCID [23]. Table II summarizes the correlation coefficients SROOC and PLCC of these metrics on the CDM database.

In the provided table, each column represents a correlation score evaluated with LOMOC-CV on each original mesh and its distorted versions (Aix, Ari, Chameleon, Fish, Samurai). For "All Models," the correlations are computed across the whole set of 85 meshes of the dataset. The "Average" column is the average of the individual LOMO-CV scores.

The results highlight CDM's strong correlations across all models, while CDM_Geo and CDM_Col exhibit weaker performance due to their narrow focus on specific features. Metrics such as SSIM, HDR-VDP2, and iCID also

show robust correlations. Our proposed base models are all competitive with the trial FR SOTA approaches and can overpass them when the combination of all the quality scores (VPB) from both views and patches is considered, as it was the case when compared to NR SOTA approaches. For specific meshes, in particular Chameleon, the best scores are obtained with patches only whereas for other meshes it is preferable to consider views. Obviously, this depends on the different details that do appear on the mesh. This advocates towards our proposal in combining quality scores estimated from both views and patches, as none stands out for every configuration.

Overall, these results indicate that our proposed methods outperforms the cited FR metrics from the SOTA.

2) Ablation study on the performance of the input features: We conducted an ablation study on various configurations, either by considering the complete VPB setup (as previously presented) or by removing one or more components to analyze their individual contributions. In this investigation, we assessed the performance of the CMDQA index with different feature configurations: views (V), patches (P), or a combination of views and patches (VP). Additionally, we conducted an ablation study on the BRISQUE scores vector, either by considering the VP combination setup or by separating its components. The findings of this study are summarized in Table III with best rates in gray. One can see that the BRISQUE IQA solely is clearly not sufficient to predict the quality of a 3D colored mesh. The use of deep features extracted from views or patches enables to obtain much better results. However, there is no configuration where using only views or views' patches is always better, this depends on the mesh under consideration. Even combining both views and patches does not enable to go beyond the results we obtained with the full configuration CMDQA-BM-R-VPB (shown in Table II) that combines all the features (deep and brisque) from views and views' patches. This underlines the interest of our proposed approach.

3) Evaluation with a cumulative model: When using LOMO-CV training for the MLP regressors and initializing each one with Glorot uniform initialization, we obtain individual trained MLP regressors for each fold, enabling measurement of their generalization abilities. However, these trained networks cannot be directly used to evaluate the quality of new unseen 3D meshes. To address this limitation, we propose a different learning strategy termed "Cumulative". This approach aims to produce a single neural network capable of assessing the visual quality of 3D meshes, surpassing the performance of the base models derived from LOMO-CV.

The cumulative training strategy, introduced in our previous work [8], involves training the MLP model with Glorot initialization on the first fold using early stopping. Subsequently, the same MLP is trained on the next fold, and so forth. This process results in a final cumulative MLP that can be deployed for future predictions on unseen data, thereby enhancing its accuracy. We trained two cumulative models, one on views (CMDQA-CM-V) and the other on

patches (CMDQA-CM-P). However, evaluating the performances of these final models can lead to overestimated results, as they were gradually trained over the entire dataset.

To mitigate this potential overfitting effect and better evaluate the performances of the Cumulative Model (CM), we re-train it using LOMO-CV to obtain a final Retrained Cumulative Model (coined as RCM). In this process, on each fold, a new MLP is initialized with the weights of the CM and trained for an optimal number of epochs using early stopping. Results are presented in Table IV with best rates in gray and demonstrate that the RCM model outperforms our Base model, and refitting it provides a more accurate evaluation of its generalization abilities.

Table IV also provides a comparison of our approach with the full reference approach of Nehmé *et al.* [20]. Indeed, they have proposed a similar approach for combining the models obtained from several folds. Using their metric based on geometric features (CMDM_Geo), or colored features (CMDM_Col), they created a third metric averaging the regression weights obtained for each fold of their LOMO-CV tests. This enables them to obtain more representative weights for their metric. Both their metrics and ours are tested on the entire dataset, and the results are summarized in Table IV. As it can be seen, our cumulative (no reference) model provides better results than the cumulative (full reference) model of Nehmé *et al.* [20]. However, the difference is small, and a larger database might be needed to differentiate them.

IV. CONCLUSION

In this study, we introduced a novel approach for assessing the visual quality of 3D colored meshes without reference, addressing a significant gap in the literature. Our proposed approach, CMDQA, leverages deep features extracted from 3D meshes' 2D projections views and patches extracted from, combined with the BRISQUE no-reference image quality index. Through extensive experiments on the CMDM database, we have shown the effectiveness of our approach in predicting overall quality, and demonstrated its superior performance compared to existing (full or no reference) metrics of the state-of-the-art.

REFERENCES

- [1] Y. Liu, Q. Yang, Y. Xu, and L. Yang, "Point cloud quality assessment: Dataset construction and learning-based no-reference metric," *ACM Trans. Multim. Comput. Commun. Appl.*, vol. 19, no. 2s, pp. 80:1–80:26, 2023.
- [2] I. Abouelaziz, A. Chetouani, M. E. Hassouni, L. J. Latecki, and H. Cherifi, "No-reference mesh visual quality assessment via ensemble of convolutional neural networks and compact multi-linear pooling," *Pattern Recognit.*, vol. 100, p. 107174, 2020.
- [3] I. Abouelaziz, M. E. Hassouni, and H. Cherifi, "A curvature based method for blind mesh visual quality assessment using a general regression neural network," in *SITIS*, 2016, pp. 793–797.
- [4] ——, "A convolutional neural network framework for blind mesh visual quality assessment," in *ICIP*, 2017, pp. 755–759.
- [5] A. Nouri, C. Charrier, and O. Lézoray, "3d blind mesh quality assessment index," in *3DIPM*, 2017, pp. 9–16.
- [6] I. Abouelaziz, A. Chetouani, M. E. Hassouni, L. J. Latecki, and H. Cherifi, "3d visual saliency and convolutional neural network for blind mesh quality assessment," *Neural Comput. Appl.*, vol. 32, no. 21, pp. 16 589–16 603, 2020.

Correlation coefficients SROOC () and PLCC () of different objective metrics on the CMDM database															
Type	Metric	Aix	Ari	Chameleon	Fish	Samurai	All meshes	Average							
		SROOC	PLCC												
Full Reference	CMDM	0.956	0.958	0.91	0.96	0.83	0.83	0.914	0.93	0.944	0.933	-	-	0.91	0.92
	CMDM_Geo	0.621	0.53	0.468	0.68	0.4774	0.457	0.554	0.554	0.407	0.462	0.437	0.501	0.51	0.54
	CMDM_Col	0.791	0.778	0.553	0.491	0.788	0.764	0.903	0.941	0.779	0.76	0.746	0.745	0.76	0.75
	D_LAB	0.826	0.791	0.497	0.282	0.747	0.776	0.779	0.734	0.659	0.546	0.603	0.55	0.70	0.63
	SSIM	0.909	0.896	0.932	0.973	0.868	0.823	0.2929	0.959	0.915	0.957	0.799	0.797	0.78	0.92
	HDR-VDP2	0.853	0.893	0.947	0.976	0.818	0.849	0.897	0.895	0.962	0.978	0.84	0.853	0.90	0.92
Ours	iCID	0.932	0.958	0.929	0.953	0.921	0.924	0.935	0.954	0.968	0.966	0.83	0.825	0.94	0.95
	Brisque-R-V	0.022	0.291	0.458	0.592	0.772	0.648	0.664	0.738	0.659	0.746	0.497	0.566	0.515	0.603
	Brisque-R-P	0.287	0.231	0.108	0.086	0.853	0.868	0.980	0.987	0.797	0.818	0.473	0.456	0.605	0.598
	Brisque-R-VP	0.105	0.126	0.174	0.215	0.863	0.905	0.951	0.986	0.801	0.836	0.473	0.492	0.579	0.613
	CMDQA-BM-A-V	0.973	0.990	0.836	0.849	0.993	0.991	0.985	0.992	0.980	0.987	0.699	0.682	0.953	0.962
	CMDQA-BM-A-P	0.975	0.998	0.441	0.647	0.998	0.999	0.990	0.995	0.956	0.988	0.662	0.670	0.872	0.925
	CMDQA-BM-R-V	0.949	0.971	0.924	0.891	0.995	0.993	0.985	0.982	0.961	0.988	0.813	0.859	0.963	0.965
	CMDQA-BM-R-P	0.843	0.959	0.417	0.613	0.958	0.990	0.988	0.992	0.900	0.933	0.711	0.732	0.821	0.897
	CMDQA-BM-R-VP	0.831	0.949	0.520	0.626	0.919	0.957	0.850	0.899	0.748	0.811	0.565	0.597	0.774	0.848
	CMDQA-BM-R-VPB	0.985	0.986	0.998	1.000	0.946	0.871	0.998	1.000	0.971	0.992	0.876	0.887	0.979	0.970

TABLE II: Comparison of SROOC and PLCC values of our proposed methods with Full Reference state of the art approaches.

Correlation coefficients SROOC and PLCC of different protocols on the CMDM database												
Model's input	Protocol	Aggregation	Configuration	Aix out	Chameleon out	Ari out	Fish out	Samurai out	All meshes	Average		
Brisque	views		MLP	SROOC	0.022	0.458	0.772	0.664	0.659	0.497	0.515	
	patches			PLCC	0.291	0.592	0.648	0.738	0.746	0.566	0.603	
	views + patches			SROOC	0.287	0.108	0.853	0.980	0.797	0.473	0.605	
	views		MLP	PLCC	0.231	0.086	0.868	0.987	0.818	0.456	0.598	
	views + patches			SROOC	0.105	0.174	0.863	0.951	0.801	0.473	0.579	
	Deep features		Average	SROOC	0.973	0.836	0.993	0.985	0.980	0.699	0.953	
		views		PLCC	0.990	0.849	0.991	0.992	0.987	0.682	0.962	
		patches		SROOC	0.949	0.924	0.995	0.985	0.961	0.813	0.963	
		Average		PLCC	0.971	0.891	0.993	0.982	0.988	0.859	0.965	
		MLP		SROOC	0.843	0.417	0.958	0.988	0.900	0.711	0.821	
		views + patches		PLCC	0.959	0.613	0.990	0.992	0.933	0.732	0.897	

TABLE III: Ablation study on the input features applied to the CMDM Database with SROOC and PLCC values of the different configurations.

metrics	SROOC	PLCC
CMDM_Geo	0.437	0.501
CMDM_Col	0.746	0.745
CMDMrec	0.9	0.913
CMDQA-BM-R-V	0.813	0.859
CMDQA-BM-R-P	0.711	0.732
CMDQA-BM-R-VPB	0.876	0.887
CMDQA-RCM-R-V	0.996	0.998
CMDQA-RCM-R-P	0.995	0.99
CMDQA-RCM-R-VPB	0.912	0.944

TABLE IV: Comparison of SROOC and PLCC values of the Cumulative Models of Nehmé *et al.* [20] and ours.

- [7] I. Abouelaziz, A. Chetouani, M. E. Hassouni, H. Cherifi, and L. J. Latecki, "Learning graph convolutional network for blind mesh visual quality assessment," *IEEE Access*, vol. 9, pp. 108 200–108 211, 2021.
- [8] Z. Ibork, A. Nouri, O. Lézoray, C. Charrier, and R. Touahni, "No reference 3d mesh quality assessment using deep convolutional features," in *International Symposium on Image and Signal Processing and Analysis (ISPA - IEEE)*, 2023, pp. 1–6.
- [9] S. Bourbia, A. Karine, A. Chetouani, M. E. Hassouni, and M. Jridi, "No-reference 3d point cloud quality assessment using multi-view projection and deep convolutional neural network," *IEEE Access*, vol. 11, pp. 26 759–26 772, 2023.
- [10] Z. Wang, Y. Zhang, Q. Yang, Y. Xu, J. Sun, and S. Liu, "Point cloud quality assessment using 3d saliency maps," in *IEEE International Conference on Visual Communications and Image Processing (VCIP)*, 2023, pp. 1–5.
- [11] Y. Nehmé, F. Dupont, J. Farrugia, P. L. Callet, and G. Lavoué, "Visual quality of 3d meshes with diffuse colors in virtual reality: Subjective and objective evaluation," *IEEE Trans. Vis. Comput. Graph.*, vol. 27, no. 3, pp. 2202–2219, 2021.
- [12] Z. Zhang, W. Sun, H. Wu, Y. Zhou, C. Li, X. Min, G. Zhai, and W. Lin, "GMS-3DQA: projection-based grid mini-patch sampling for 3d model quality assessment," *CoRR*, vol. abs/2306.05658, 2023.
- [13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *ICLR*, 2015.
- [14] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, 2012.
- [15] Z. Zhang, W. Sun, X. Min, T. Wang, W. Lu, W. Zhu, and G. Zhai, "A no-reference visual quality metric for 3d color meshes," in *IEEE International Conference on Multimedia & Expo Workshops*, 2021, pp. 1–6.
- [16] M. E. Hassouni and H. Cherifi, "Learning graph features for colored mesh visual quality assessment," in *IEEE International Conference on Image Processing, ICIP*, 2022, pp. 3381–3385.
- [17] Z. Zhang, W. Sun, X. Min, T. Wang, W. Lu, and G. Zhai, "No-reference quality assessment for 3d colored point cloud and mesh models," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 11, pp. 7618–7631, 2022.
- [18] I. Abouelaziz, M. E. Hassouni, and H. Cherifi, "No-reference 3d mesh quality assessment based on dihedral angles model and support vector regression," in *ICISP*, vol. LNCS 9680, 2016, pp. 369–377.
- [19] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, 2013.
- [20] Y. Nehmé, J. Delanoy, F. Dupont, J. Farrugia, P. L. Callet, and G. Lavoué, "Textured mesh quality assessment: Large-scale dataset and deep learning-based quality metric," *ACM Trans. Graph.*, vol. 42, no. 3, pp. 31:1–31:20, 2023.
- [21] 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, 2004.
- [22] I. Lissner and P. Urban, "Toward a unified color space for perception-based image processing," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 1153–1168, 2012.
- [23] Y. Liu, J. Wang, S. Cho, A. Finkelstein, and S. Rusinkiewicz, "A no-reference metric for evaluating the quality of motion deblurring," *ACM Trans. Graph.*, vol. 32, no. 6, pp. 175:1–175:12, 2013.