IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 15 May 2024, accepted 17 July 2024, date of publication 29 July 2024, date of current version 12 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3435377

RESEARCH ARTICLE

No Reference 3D Mesh Quality Assessment Learned From Quality Scores on 2D Projections

ZAINEB IBORK[1,2], ANASS NOURI[1,2], OLIVIER LÉZORAY[2], (Senior Member, IEEE), CHRISTOPHE CHARRIER[2], (Member, IEEE), AND RAJA TOUAHNI[1]

[1]Information Processing and AI Team, SETIME Laboratory, Faculty of Sciences, Ibn Tofail University, Kénitra 14000, Morocco
[2]Normandie Univ, UNICAEN, ENSICAEN, GREYC, CNRS, 14000 Caen, France

Corresponding author: Olivier Lézoray (olivier.lezoray@unicaen.fr)

This work was supported by Partenariat Hubert Curien (PHC) TOUBKAL TBK/22/142-Campus under Grant 47259YH.

ABSTRACT With the widespread availability and utilization of 3D meshes across various applications, the need for accurately assessing their visual quality has become increasingly important. Despite the significance of this task, the literature offers few No-Reference (NR) approaches for evaluating the visual quality of 3D meshes. In response to this gap, this paper proposes a novel NR approach tailored specifically to score the quality of 3D meshes. After rendering a 3D mesh into 2D views and patches, a pre-trained convolutional neural network automatically extracts deep features from. These features are then employed in a Multi-Layer Perceptron regressor to predict the quality score of the rendered images. The obtained scores are combined with their corresponding BRISQUE scores, and an additional MLP regressor is used to predict the final score. We present experimental results demonstrating the effectiveness and robustness of our approach across a diverse range of 3D mesh datasets. Comparative analyses with existing NR methods underscore the superior performance and versatility of the proposed approach. Overall, this paper contributes to the advancement of NR techniques for assessing 3D mesh quality, offering a valuable tool for researchers, practitioners, and developers working with 3D models across various domains.

INDEX TERMS 3D mesh, mesh visual quality assessment, convolutional neural network, deep learning, no reference quality assessment, BRISQUE.

I. INTRODUCTION

Three dimensional meshes are widely used in many fields and applications such as computer graphics and games. 3D acquisition techniques have recently undergone a number of advances and developments and many different techniques can be now considered for 3D acquisition. We can quote laser scanners, structured light cameras and photogrammetry. Using the latter, it is now easily possible to generate 3D models by simply taking pictures around an object or a person. Once a 3D mesh model has been acquired, it can be used in many various contexts such as computer graphics, games, digital animation, virtual to augmented reality, digital twins, heritage or forensics, 3D printing, industrial design and many others. Depending on the application context, the required precision of the 3D mesh is not always the same. For instance, for conservation purposes in digital heritage, one would like to have very detailed meshes with millions of vertices. But if the mesh is to be distributed via the Internet for viewing by the general public, the size of the mesh has to be greatly reduced to enable a fast and interactive user experience. As a consequence, raw acquired 3D meshes always undergo a lot of pre-processing steps [1], from denoising [2] to simplification [3], [4], watermarking [5], [6] and compression [7], [8]. Inevitably, all these processing steps can introduce visual distortions and affect the 3D mesh quality that can hinder a proper visualisation for end-users.

The direct consequence is that the development of 3D acquisition techniques has put forward the need for 3D Mesh Visual Quality Assessment (MVQA) methods. 3D meshes are now subject to the same challenges as 2D images, with their distribution via the Internet and the need for

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

methods to assess their quality. If there exists a lot of state-of-the-art methods for Image Quality Assessment (IQA) (see [9] for a recent review) that are now used in practical applications, this is far from being the case for 3D meshes. For 2D IQA, we can categorize objective methods regarding the availability of a reference. A metric is termed ''No-Reference'' (NR) or ''no-reference'' when the evaluation is conducted without access to the reference data. Conversely, a metric is referred to as ''Reduced-Reference'' (RR) or ''Full Reference'' (FR) when the evaluation is performed using partial or complete reference data, respectively. For 3D MVQA, the same categories do apply. For NR methods, the Mean Opinion Score (MOS) associated to the meshes can be used as a ground truth for training a prediction model. Historically, the majority of established objective methods for 3D mesh quality assessment have predominantly fallen into the categories of FR or RR. However, no-reference methods (i.e. NR) hold particular significance as they operate without the need for a reference object to assess mesh quality. This characteristic makes them valuable for real-time applications where speed is crucial, and the luxury of having data references is not always feasible. Hence, this serves as the motivation for undertaking this work and our objective is to propose a NR MVQA approach. As the focus of the paper is on NR methods, we will not describe in depth FR or NR MVQA approaches. We refer the reader to [10] and [11] that provide a good review of. These approaches usually compute local geometrical features that are combined into a quality score, such as the FR method MSDM2 [12], and the RR method FMPD [13].

The organization of the paper unfolds as follows: in section II, we provide a brief overview of the state-of-the-art of NR MVQA. Following that, section III describes the details of our proposed approach. Next, section IV delves into the experimental setup with the considered databases, the training and evaluation validation protocols, and a comparison with the state-of-the-art (SOTA). Section V concludes the paper and presents our perspectives on the topic.

II. NO-REFERENCE 3D MESH VISUAL QUALITY ASSESSMENT SOTA

In this section we provide a brief overview of the main methods of the SOTA for NR MVQA.

Several NR metrics have been developed for MVQ assessment. Abouelaziz et al. [14] proposed a no-reference method that uses mean curvature features and a General Regression Neural Network (GRNN). Nouri et al. [15] employed visual saliency and Support Vector Regression (SVR) for 3D No-Reference Mesh Quality Assessment Index (BMQI). In [16], Lin et al. proposed BMQA-GSES (No-Reference Mesh Quality Assessment with Graph Spectral Entropy ans Spatial features). Using Graph Spectral Entropy and Spatial features, the 3D mesh is considered as a graph signal. In the graph spectral domain, Gaussian curvature signal undergoes conversion with Graph Fourier transform, extracting smoothness and information entropy for distortion evaluation. In the spatial domain, four key spatial features describe concave, convex, and structural information. The random forest regression fuses all features to predict the objective quality score of the 3D mesh. In order to take advantage of Convolutional Neural Networks (CNNs), Abouelaziz et al. proposed various approaches feeding them with 2D projection rendered images to predict visual quality [17], [18], [19]. In their primary work, they introduced a patch-selection strategy based on mesh saliency to emphasize attractive regions [20]. They have finally proposed an approach called CNNs-CMP [10] that combines deep features extracted from several pre-trained CNNs (VGG16/AlexNet/ResNet) and combine them using Compact Multi-linear Pooling. This approach is actually the best approach for NR MVQA.

In a previous work [21], we have presented DCFQI (Deep Convolutional Features Quality Index). Our work is inline with the approaches of Abouelaziz et al. that consider 2D rendered projections of 3D meshes and extract deep features from. By rendering a 3D mesh into 2D views that can be further decomposed into patches, a pre-trained CNN firstly automatically extracts deep features. These features are then used in a Multi Layer Perceptron (MLP) regressor to the quality of each rendered image. The quality scores of the 2D views are then averaged to estimate the final 3D mesh quality. In this paper we will extend our previous work at different levels:

- We use a MLP regressor instead of a simple average to estimate the final 3D mesh quality score from rendered images,
- We combine quality scores from both 2D views and patches, and also those obtained from the BRISQUE NR IQA index [22] on the rendered images,
- We perform a thorough evaluation of the different configurations that can be obtained with our approach on four different 3D mesh quality databases,
- We demonstrate the effectiveness and robustness of our approach versus the SOTA.

Details about the proposed approach are provided in the next section.

III. NO-REFERENCE 3D MESH QUALITY ASSESSMENT LEARNED FROM QUALITY SCORES ON 2D PROJECTIONS

A. FLOWCHART OF THE APPROACH

In order to assess the visual quality of a given 3D mesh, our proposed method starts with the rendering of multiple 2D projection (that will be called views), by varying the point of view around the mesh. These views can potentially contain a lot of white background that is not informative for quality estimation. Therefore, they are cropped according to the mesh bounding box in the view. The processed views are then divided into four overlapping patches. Each 2D view or patch issued from a view are 2D images and are fed to a pre-trained convolutional network (VGG16 [23] in our case), to extract deep features describing the image. Afterwards, this obtained

feature vector serves to predict the quality of a view or a patch, using a MLP regressor from the Mean Opinion Score (MOS) associated to each mesh M_i. As a mesh is described by several views and patches, a vector of quality scores is obtained. We also propose to enlarge the latter by appending to it the scores estimated by the BRISQUE no-reference image quality metric on views and patches. The final vector of quality scores is used to estimate the Predictive Mean Opinion Score (PMOS) for the 3D mesh. This estimation can be done by two different fusion schemes: averaging (as in our previous work [21]) or by a non-linear regression. Figure 1 illustrates the pipeline of the proposed approach. In the sequel we will describe each step of the proposed MVQA approach.

B. 2D PROJECTIONS PRE-PROCESSING

Given a database containing N 3D meshes, our objective is to perform several 2D renderings of each 3D mesh M_i, where $i \in [1, N]$. To ensure that the different meshes are positioned in a similar way, the meshes centroids' are shifted to the origin of the coordinate system. This ensures a consistent positioning for all meshes, and the distance from the rendering camera is adjusted for each mesh. Both are crucial factors for achieving reliable quality assessments. Once well positioned, a 3D mesh M_i is rendered from 11 different viewpoints by systematically changing the azimuth (θ_a) and elevation (θ_e) angles by $\frac{\pi}{3}$ (60 degrees) for each viewpoint. This technique is illustrated in Figure 2.

To capture diverse views, the elevation angle is maintained at 0 degrees, while the Azimuth angle is systematically varied (and vice versa). The camera position and the distance to the object are set manually. Indeed, the bounding box of a mesh can strongly vary from one mesh to another, and it is essential to adjust the rendering to each. This ensures that the object is localized close to the camera, maximizing the visibility of intricate details. The adjustment of these parameters facilitates the creation of an extensive dataset of 2D views. This dataset encompasses various perspectives and highlights crucial object details. Figure 2 presents the obtained rendered 2D views of the Jessy 3D mesh from the UWB Compression database [24].

After this rendering step, 11 rendered views are obtained for each mesh M_i. We will denote them by V_i^j with $j \in [1, 11]$. The size of the rendered views is 1024×1024. This fixed size was chosen to meticulously capture crucial distinguishing details vital for visual quality assessment. Nevertheless, these images come with a considerable expanse of white background. To mitigate the influence of this non-informative background on quality assessment, we conduct a cropping and resizing process. The objective is to retain solely the mesh bounding box, effectively eliminating the majority of the surrounding white backdrop. Given that the cropping operation is contingent on the size of the mesh bounding box, the resultant images may exhibit varying sizes. To ensure uniformity, we resize all images to a standardized 512×512.

With the goal of capturing intricate details from 3D meshes through 2D views, the acquisition of only $11 \times N$ views might fall short in a supervised learning setting. To address this limitation, we adopt a strategy of extracting four overlapping patches from each 2D view. Therefore, we extract four patches of size 288×288: the patches are larger than 256×256 as we enforce 12.5% of overlap. This way of splitting a view into several patches is different from the one proposed in [18]. Indeed, they extract tiny patches of size 32×32 and this has several drawbacks. First, small patches may lack sufficient information for comprehensive quality assessment. Second, small patches can predominantly consist of background, necessitating specific strategies for background elimination: in [10], they have proposed to eliminate background patches with the help of saliency detection. Third, as the number of eliminated background patches varies from one 2D view to another, this produces a different number of retained extracted patches per view and produces an unbalanced dataset. With our approach of larger patch extraction, we cope with all these problems. 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 $k \in [1, 4]$.

To conclude this pre-processing, the extracted images V_i^j and $P_i^{j,k}$ of a mesh M_i undergo normalization, scaling its values between 0 and 1. Unlike the local contrast normalization employed in [10], our proposed method opts for global normalization performed on the L^* lightness channel within the CIELAB color space. This choice is motivated by the nonlinear nature of perceived lightness (L^*), mirroring human perception, and aims to enhance the quality assessment process.

C. QUALITY ESTIMATION OF A 2D PROJECTION

With the pre-processing step, from an initial database $\mathcal{B}$ of N meshes, two distinct datasets have been generated: 1) $\mathcal{B}_V$ considering only 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 $i \in [1, N]$, and $j \in [1, 11]$. The $\mathcal{B}_P = \{P_i^{j,k}\}$ dataset contains $N \times 11 \times 4$ images with $i \in [1, N], j \in [1, 11]$, and $k \in [1, 4]$. To simplify the notations we will denote as I_i^j an image of a set $\mathcal{B}_S$ with $S \in \{V, P\}$, $i \in [1, N], j \in [1, N_S]$, $N_V = 11$ and $N_P = 44$.

1) LEARNED 2D PROJECTION QUALITY ESTIMATION

From these two datasets, our objective is to learn to score the quality of a 2D projection image from its corresponding 3D mesh MOS as reference value, whether it is a view or a patch issued from a view. To do so, we first extract a deep feature vector from an image I_i^j using a pre-trained VGG16 convolutional network [23] (we do not train it from scratch but re-use the pre-trained weights obtained on ImageNet). Each image is resized to 224×224 before being fed to VGG16. The latter was chosen for its superior feature extraction performance compared to alternative models like AlexNet and ResNet [10]. The resize operation to

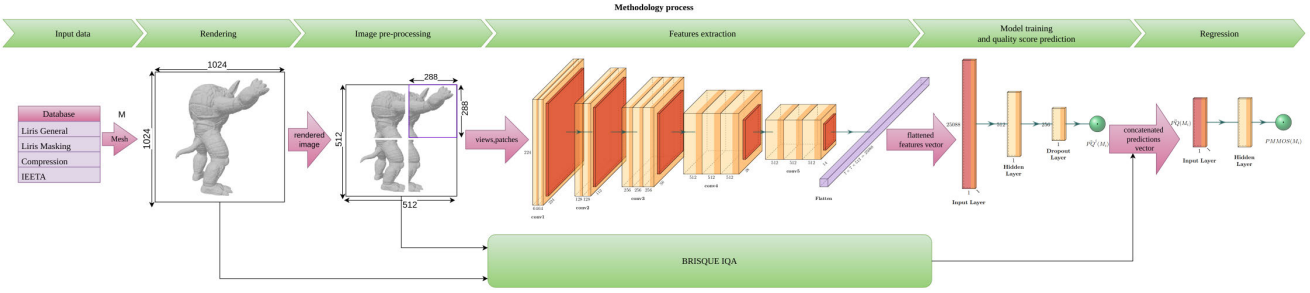


FIGURE 1. Illustration of the proposed Mesh Visual Quality Assessment approach.

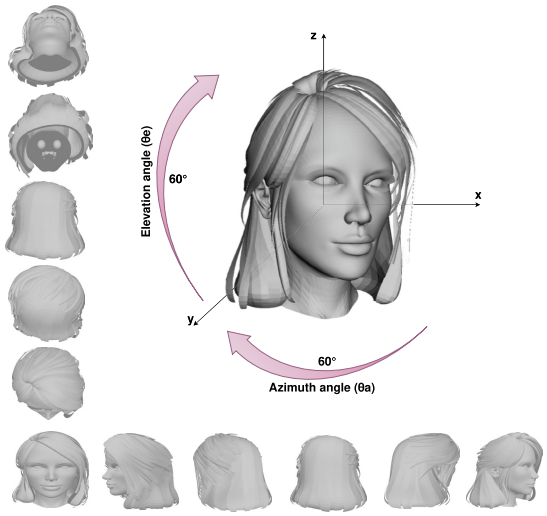


FIGURE 2. Illustrating viewing angles in the rendering process of a 3D mesh: Azimuth angle (θ_a) in the horizontal plane with $z = 0$ and Elevation angle (θ_e) from the xz plane with $y = 0$.

224×224 is needed as we consider a pre-trained VGG16 the input size of which is fixed. For 2D views, that have a size of 512×512, this resize operation has a strong reduction factor and this can alter the extracted features. This is the reason why we also consider patches extracted from views. As the size of the extracted patches is 288×288, this time the reduction factor is negligible. An alternative would be to re-train the VGG16 feature extractor for views and patches, we will consider this possibility in future works. Anyway, as it will be demonstrated in Section IV, the experimental results show that the features extracted from the resized views are sufficient for an accurate quality prediction, but they can be enhanced by combining them with predictions made on patches.

Formally we denote as $\boldsymbol{\phi}$ the VGG16 deep feature extractor (i.e., obtained with the VGG16 layer before its final Fully Connected layers). It takes as input an image of size 224×224 and outputs a flattened vector of size $7 \times 7 \times 512 = 25088$. It is defined as $\boldsymbol{\phi} : \mathbb{R}^{224 \times 224} \rightarrow \mathbb{R}^{25088}$.

Given a dataset $\mathcal{B}_S$, each image $I_i^j \in \mathcal{B}_S$ is now described by a feature vector $\boldsymbol{\phi}(I_i^j)$. The latter is used as input to a shallow Multi-Layer Perceptron (MLP) the objective of which is to predict a quality score close to the MOS of the associated mesh M_i. Therefore we train a MLP non-linear regressor with each input vector being $\boldsymbol{\phi}(I_i^j)$ and the associated value to predict is MOS(M_i). As we have two distinct datasets of views and patches $\mathcal{B}_V$ and $\mathcal{B}_P$, we finally obtain two distinct MLP non-linear regressors denoted by MLPR_S that can estimate the quality of an image (being either a view or a patch, as given by S). We quote this as the Predicted Image Mean Opinion Square and denote it by $\text{PIMOS}_S(I_i^j) = \text{MLPR}_S(\boldsymbol{\phi}(I_i^j))$ with $S \in \{V, P\}$, $i \in [1, N]$, and $j \in [1, N_S]$.

2) BRISQUE IQA FOR 2D PROJECTION QUALITY ESTIMATION

We just have shown how to learn to estimate the quality of 2D projections from reference datasets of views or patches. However, as 2D projections are images, we could also consider NR Image Quality Assessment (IQA) metrics to do so. The BRISQUE (No-Reference/Referenceless Image Spatial Quality Evaluator) IQA stands out from the literature as a no-reference image quality assessment algorithm [22]. It measures the perceived quality of images without the need for a reference image for comparison. This algorithm leverages spatial domain features within the image, extracting statistical measures like local mean and standard deviation. Then, a machine learning model is trained to predict the quality score, with the model being trained on a dataset of natural images with known quality scores. High Brisque scores typically indicate lower image quality, while lower scores suggest higher image quality. We propose to also consider such a no-reference IQA to estimate the quality of 2D projections. The interest of using BRISQUE is twofold: it is referenceless and is already trained. We quote this estimation by $\text{PIMOS}_S^B(I_i^j) = \text{BRISQUE}_S(I_i^j)$ with $S \in \{V, P\}$, $i \in [1, N]$, and $j \in [1, N_S]$. As with VGG16, the BRISQUE model we consider is pre-trained: it has been trained beforehand on the LIVE IQA dataset [22]. BRISQUE extracts coefficients from Mean Substracted Contrast Normalized (MSCN) images and a Generalized Gaussian Distribution (GGD) is fit onto. Features are extracted from the fitted GGD and used to train a Support Vector Regressor model that predicts the image quality. To be able to use this pre-trained BRISQUE model, we have

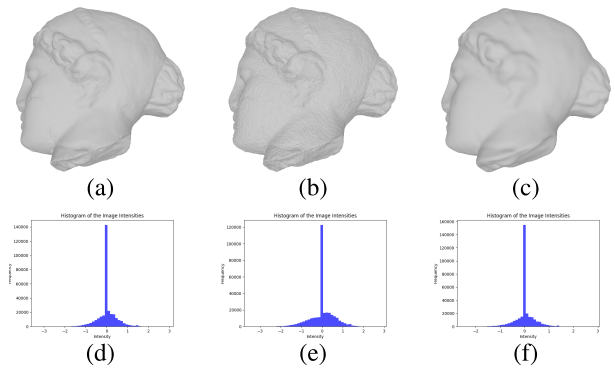


FIGURE 3. Illustration of the Gaussianity of 2D projections. Figure (a)-(c) present a 2D projection of an original mesh and two of its degraded versions (with added noise in (b) and simplification in (c)). The histograms of the normalized luminance coefficients follow a nearby Gaussian distribution.

to ensure that the 2D projections also do follow a GGD. Figure 3 presents a 2D projection for a reference mesh and two degraded versions, obtained by adding noise or simplifying the mesh. As it can be seen the MSCN coefficient nearby follow a Gaussian distribution, and this validates our assumption that BRISQUE can be used on 2D projections. Here we have considered a pre-trained BRISQUE model, but we could have re-trained it on a dedicated dataset of 2D projections. This will be considered in future investigations.

D. MVQA 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: $\mathbf{PQ_S}(M_i) = [\text{PIMOS}_S(I_i^j) : j \in [1, N_S]]^T$. 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 evalxuators 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:

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

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

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

$$\text{PMMOS}(\text{M}_i) = \text{MLPR}(\mathbf{PQ}(M_i)) \qquad (2)$$

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 MVQA methods that we will compare in the next section.

IV. EXPERIMENTS

Now that we have introduced our approach to MVQA, we investigate its properties by experiments with several configurations on different databases. Let us first present the databases we have considered.

A. DATABASES

We give a short description of the databases we used. For each database, a histogram of the meshes' MOS is shown to illustrate the database MOS diversity (or lack thereof).

1) LIRIS/EPFL GENERAL PURPOSE DATABASE

The LIRIS/EPFL General-Purpose Database [5] contains 88 models ranging from 40K to 50K vertices. These models are derived from 4 reference meshes (Armadillo, Dynosaur, Venus, and RockerArm). Distortions, such as noise addition and smoothing, are applied to the 3D meshes with varying strengths at four locations: uniformly (on the entire object), or on smooth, rough, and intermediate areas. A total of 12 observers participated in the subjective evaluation and their subjective scores were averaged to obtain a MOS per mesh. Figure 4 presents the 3D meshes of this database. Figure 5 presents the MOS histogram for this database (few meshes of low quality are present).

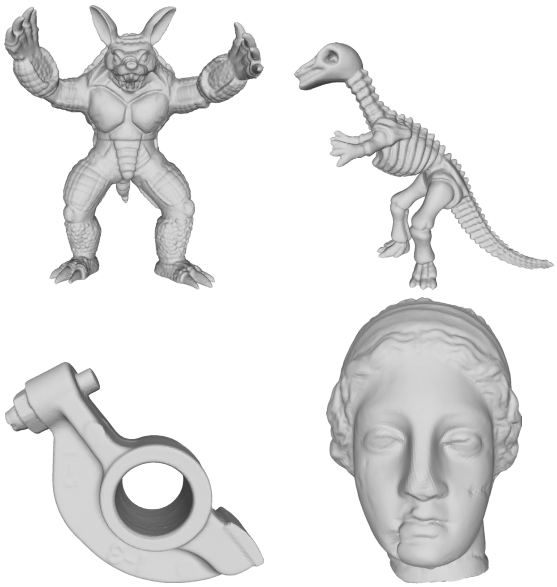


FIGURE 4. The meshes of the Liris/EPFL General Purpose Database. From top-left to bottom-right: Armadillo, Dyno, RockerArm, Venus.

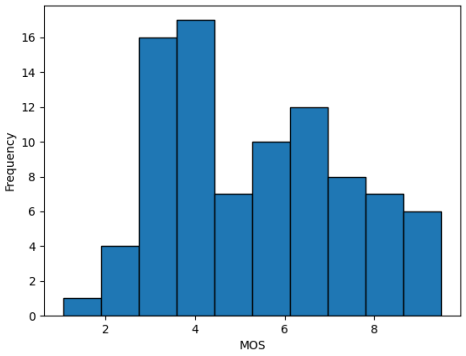


FIGURE 5. Liris/EPFL General Purpose Database histogram of MOS.

FIGURE 6. The meshes of the LIRIS Masking Database. From top-left to bottom-right: Armadillo, Dyno, Bimba, Lion.

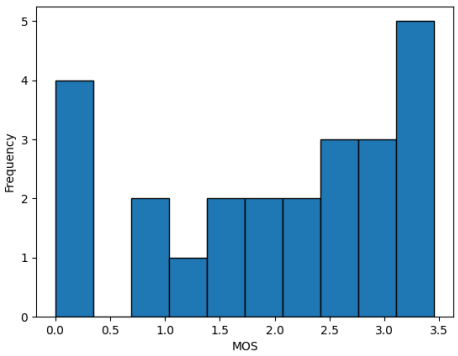


FIGURE 7. LIRIS Masking Database histogram of MOS.

2) LIRIS MASKING DATABASE

The LIRIS Masking Database [25] contains 26 models ranging from 9K to 40K vertices, generated from 4 reference meshes (Armadillo, Bimba, Dynosaur, and Lion). These objects were selected for their distinct smooth and rough areas. The sole applied distortion is noise addition, with three different strengths, exclusively on either smooth or rough regions. The primary objective of this database is to assess metric behavior concerning the visual masking effect. As noise is less visible on rough regions, metrics should align with this perceptual mechanism. A total of 11 observers participated in the subjective evaluation and their subjective scores were averaged to obtain a MOS per mesh. Figure 6 presents the 3D meshes of this database. Figure 7 presents the MOS histogram for this database.

3) IEETA SIMPLIFICATION DATABASE

The IEETA Simplification Database [26] contains 30 models generated from 5 reference meshes (Bunny, Foot, Head, Lung, and Strange) ranging from 2K to 25K vertices. Reference models have undergone simplification using three different methods at two levels (20% and 50% of the original number of faces). A total of 65 observers contributed to the subjective evaluation and their subjective scores were averaged to obtain a MOS per mesh. Figure 8 presents the 3D meshes of this database. Figure 9 presents the MOS histogram for this database, that is very unbalanced.

4) UWB COMPRESSION DATABASE

The UWB Compression Database [24], contains a total of 78 altered models generated from 6 reference meshes. Each original model undergoes 13 distinct compression types implemented by various algorithms. The subjective evaluation study involves the participation of 69 observer and their subjective scores were averaged to obtain a MOS per mesh. Figure 10 presents the 3D meshes of this database. Figure 11 presents the MOS histogram for this database.

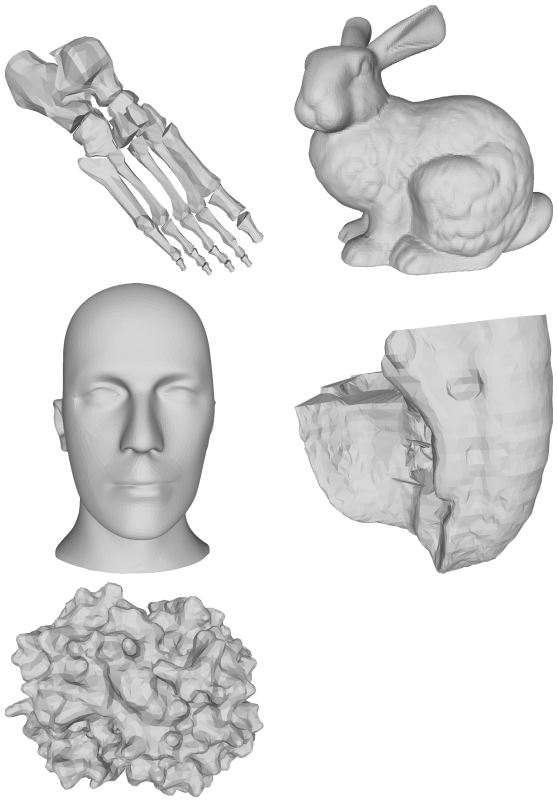


FIGURE 8. The meshes of the IEETA Simplification Database. From top-left to bottom-right: Bones, Bunny, Head, Lung, Strange.

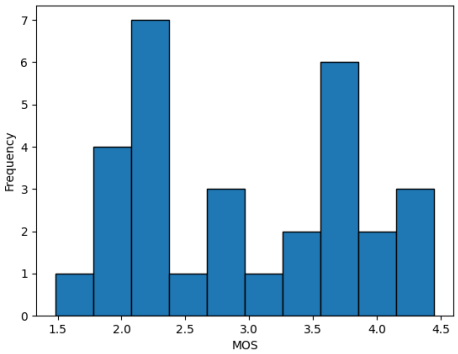


FIGURE 9. IEETA Simplification Database histogram of MOS.

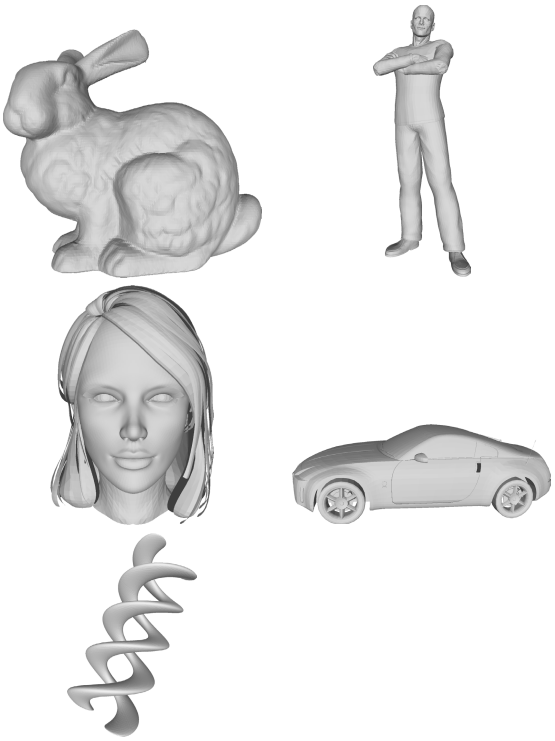


FIGURE 10. The meshes of the UWB Compression Database. From top-left to bottom-right: Bunny, James, Jessy, Nissan, Helix.

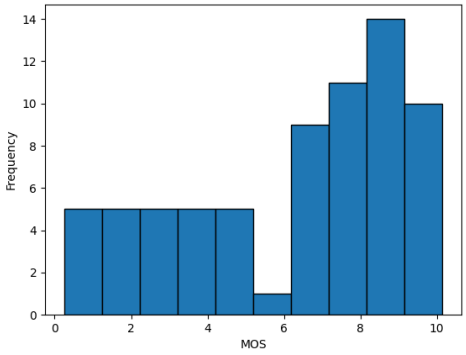


FIGURE 11. UWB Compression Database histogram of MOS.

B. EVALUATION METRICS

To evaluate the quality assessment performance of our approach, we will employ two standard metrics: the Spearman Rank-Order Correlation Coefficient *SROOC* and the Pearson Linear Correlation Coefficient *PLCC*. Widely used in visual quality assessment, these metrics measure the agreement and similarity between predicted scores and ground truth values. They form the basis for comparing the performance of our proposed approach with existing no-reference mesh quality assessment metrics. By quantifying the relationship between predicted and actual scores or rankings, these coefficients offer valuable insights into the model's performance.

The *SROOC* coefficient (r_s) is a statistical measure used to assess the strength and direction of the monotonic relationship between the predicted quality scores PMMOS(M_i) and the reference mean opinion scores MOS(M_i). On the other hand, the *PLCC* metric (r_p) assesses the linear relationship or correlation between the predicted scores and the ground truth values.

C. TRAINING AND EVALUATION PROTOCOL

We provide details on the MLP regressors' MLPR$_S$ architecture used to perform quality estimation from 2D projections

TABLE 1. The different configurations of our approach that we tested in the experiments. To each configuration an acronym is given (see text for details).

Configuration name	Configuration
Base models	
BRISQUE-R-V	PMMOS learned from $PQ_{V,B}(M_i)$
BRISQUE-R-P	PMMOS learned from $PQ_{P,B}(M_i)$
BRISQUE-R-VP	PMMOS learned from $PQ_{V,B}(M_i) \cup PQ_{P,B}(M_i)$
DCFQI-BM-A-V	$\overline{\text{PMMOS}}$ averaged from $PQ_V(M_i)$
DCFQI-BM-A-P	$\overline{\text{PMMOS}}$ averaged from $PQ_P(M_i)$
DCFQI-BM-R-V	PMMOS learned from $PQ_V(M_i)$
DCFQI-BM-R-P	PMMOS learned from $PQ_P(M_i)$
DCFQI-BM-R-VP	PMMOS learned from $PQ_V(M_i) \cup PQ_P(M_i)$
DCFQI-BM-R-VPB	PMMOS learned from $PQ_V(M_i) \cup PQ_P(M_i) \cup PQ_{V,B}(M_i) \cup PQ_{P,B}(M_i)$
Cumulative models	
DCFQI-CM-A-V	Cumulative $\overline{\text{PMMOS}}$ averaged from $PQ_V(M_i)$
DCFQI-CM-A-P	Cumulative $\overline{\text{PMMOS}}$ averaged from $PQ_P(M_i)$
DCFQI-CM-R-V	Cumulative PMMOS learned from $PQ_V(M_i)$
DCFQI-CM-R-P	Cumulative PMMOS learned from $PQ_P(M_i)$
DCFQI-CM-R-VP	Cumulative PMMOS learned from $PQ_V(M_i) \cup PQ_P(M_i)$
DCFQI-CM-R-VPB	Cumulative PMMOS learned from $PQ_V(M_i) \cup PQ_P(M_i) \cup PQ_{V,B}(M_i) \cup PQ_{P,B}(M_i)$

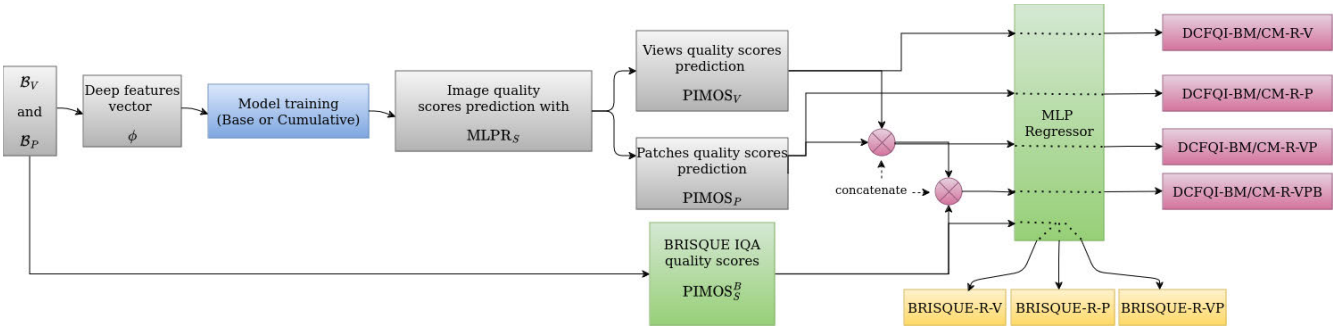


FIGURE 12. Illustration of the different configurations that can be considered with our proposed approach (see text for details).

as well on how they are trained (two procedures are considered) and evaluated.

1) MLP REGRESSORS TRAINING

For 2D projections quality estimation, the considered MLP regressor MLPR$_S$ architecture is composed by a single hidden layer of 512 neurons using a Rectified Linear Unit (ReLU) activation function, followed by a single-neuron output layer using a sigmoid activation function. A dropout layer with a rate of 0.5 is appended after the dense layer to counteract overfitting. Weight initialization for the dense layer employs the Glorot uniform initialization method [27]. The training is performed with the Mean Average Error loss and the RMSprop optimizer with a learning rate fixed to 0.001. Based on extensive tests, we have found that the best correlation scores are obtained with a batch size of one-third of the training set size. This parameter will be fixed this way for all the trials. In order to enhance the results, we employ the Early Stopping technique which is a form of regularization that helps prevent overfitting and improves generalization of the trained model.

For the aggregation of image quality scores, a third MLP regressor MLPR is used. It is composed of a a single hidden layer of neurons using a Rectified Linear Unit (ReLU) activation function, followed by a single-neuron output layer using a sigmoid activation function. The number of neurons of the hidden layer is the same than its number of inputs (i.e., $|\mathbf{PQ}|$). The other parameters are the same than for the two other regressors (activation function, optimizer, etc.).

We will refer to this training procedure of the three regressors as the "Base Model (BM)".

2) LEAVE-ONE-MESH-OUT EVALUATION PROTOCOL

In order to evaluate the accuracy of the MLP regressors MLPR$_S$, we will perform a Leave-One-Mesh-Out Cross-Validation procedure (LOMO-CV). At training, all the meshes are considered except one mesh and its distorted versions. In such a protocol, while training a MLP regressor MLPR$_S$, the images associated to a specific mesh are all excluded from the training process. The trained neural network is then tested on these excluded images to evaluate its performance as they represent unseen data. This LOMO-CV process is repeated for each mesh (a reference mesh and its degraded versions) of a dataset. By employing cross-validation, we ensure an objective assessment of the MLPR model as it is evaluated on strictly independent data that it has not been trained on.

When training the third regressor MLPR, the same procedure is applied and the quality scores of the excluded mesh and its degraded version are excluded.

3) CUMULATIVE TRAINING OF THE MLP REGRESSORS

By using a LOMO-CV training of the MLP regressors, we are able to measure their generalization abilities. However, we are not able to use a resulting network to evaluate the quality of new unseen 3D meshes. Indeed, we obtain a trained MLP regressor for each fold and not a single trained model. To cope with this problem, we propose to use a different learning strategy that we call ''Cumulative''. It allows to obtain a single neural network that can be used to assess the visual quality of 3D meshes and that can go beyond the performances of the base models obtained by LOMO-CV. To do so, we consider a cumulative training the principle of which was presented in our previous work [21]. This learning strategy begins by training and testing the MLPR model with Glorot initialization on the first fold for 1000 epochs. Then, the same MLPR is used for training on the next fold and so on. As a consequence, the final cumulative MLPR can be used to perform future predictions on unseen data and this also helps in improving its accuracy. However, if we measure the performances of this final model, this obviously leads to overestimated results as this cumulative training was gradually trained over the entire dataset. To mitigate this effect and better evaluate the performances of the Cumulative Model (CM), we re-train it using LOMO-CV in order to obtain a final Retrained Cumulative Model. To do so, on each fold, a new MLPR is initialized with the weights of the CM and is trained for a fixed number of epochs. The latter is determined by finding the global minimum of the average of the losses of all folds during the cumulative training. We will see in the results that the CM model performs better than the Base model and refitting it enables to have a better evaluation of its generalization abilities.

4) EVALUATED CONFIGURATIONS

As we have seen in the description of the approach, we can consider different configurations depending on the way the vector **PQ** is built. Its size can vary from 11 (only views are considered), 44 (only patches are considered), 55 (both views and patches are considered), to 110 (both views and patches are considered with the addition of BRISQUE scores on both views and patches). In addition, we can consider either a base model or a cumulative model depending on the chosen training procedure. To differentiate all these configurations and compare them, we will use a specific notation. The approach will be denoted by DCFQI (Deep Convolutional Features Quality Index) and the different configurations will be specified by the different terms between dashes:

DCFQI -	BM / CM	-	A / R	-	V / P / VP / VPB

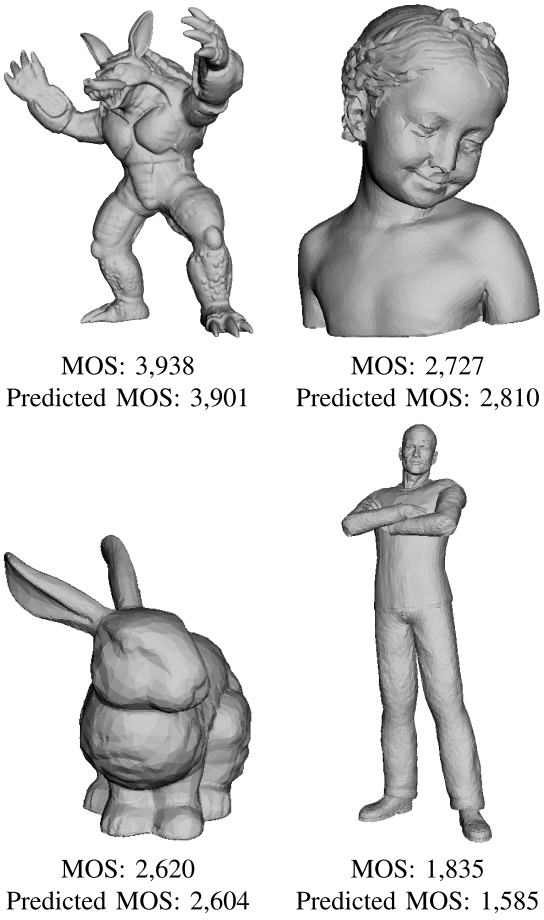


FIGURE 13. From top-left to bottom right: Reference MOS and predicted MOS with our model DCFQI-CM-R-VPB on meshes from Liris General purpose, Liris masking, IEETA Simplification, and UWB Compression databases.

where BM stands for Base Model, CM for Cumulative Model, A for Average Aggregation (Equation 1), R for MLP Regression (Equation 2), V for quality scores from views, P for quality scores from patches, VP for quality scores from both views and patches, and VPB for quality scores from both views and patches with the addition of BRISQUE scores on both views and patches. When using only BRISQUE, the configuration are specified by:

BRISQUE -	R	-	V / P / VP

There is no base or cumulative model for BRISQUE and we will consider only the MLP Regression for aggregation of the scores. These configurations will be used to assess the performance of the BRISQUE IQA taken solely for the MVQA. The different configurations are summed up in Table 1 and illustrated in Figure 12 (for only the MLP Regression aggregation).

D. RESULTS AND COMPARISON WITH SOTA

We have compared the different configurations of our proposed approach (the ones resumed in Table 1) on the

TABLE 2. Comparison of SROOC and PLCC values of our proposed methods with the state of the art applied to the Liris general purpose database.

Correlation coefficients SROOC and PLCC of different objective metrics on the LIRIS/EPFL general-purpose database													
Type	Metric	Armadillo		Dyno		Venus		Rocker		All models		Average	
		SROOC	PLCC	SROOC	PLCC	SROOC	PLCC	SROOC	PLCC	SROOC	PLCC	SROOC	PLCC
Full Reference	HD [29]	69,5	30,2	30,9	22,6	1,6	0,8	18,1	5,5	13,8	1,3	30,03	14,78
	RMS [30]	62,7	32,2	0,3	0	90,1	77,3	7,3	3	26,8	7,9	40,10	28,13
	MSDM2 [12]	81,6	72,8	85,9	73,5	89,3	76,5	89,6	76,1	80,4	66,2	86,60	74,73
	TPDM [31]	**84,5**	78,8	**92,2**	89	90,6	91	**92,2**	91,4	**89,6**	86,2	**89,88**	87,55
	Yildiz *et al.* [32]	-	86	-	79	-	89	-	88	-	-	-	85,50
	TPDMPW [33]	-	-	-	-	-	-	-	-	87,2	87,7	-	-
	Chetouani [34]	75,7	**86,1**	90,6	**90**	**94,9**	**95,5**	91,4	**92,1**	88,1	**90,9**	88,15	**90,93**
Reduced Reference	3DWPM1 [35]	65,8	35,7	62,7	35,7	71,6	46,6	87,5	53,2	69,3	38,4	71,90	42,80
	3DWPM2 [35]	74,1	43,1	52,4	19,9	34,8	16,4	37,8	29,9	49	24,6	49,78	27,33
	FMPD [13]	**75,4**	**83,3**	89,6	**88,9**	87,5	**83,9**	**88,8**	**84,7**	**81,9**	**83,5**	**85,33**	**85,20**
	DAME [36]	60,3	76,3	**92,8**	**88,9**	**91**	**83,9**	85	80,1	76,6	75,2	82,28	82,30
No-Reference	NR-SVR [19]	76,8	91,5	78,6	84,1	85,7	88,6	86,2	86,6	81,5	87,8	81,83	87,70
	NR-GRNN [14]	87,1	**97,3**	91,2	**94,1**	86,3	85	78,6	74,8	86,2	88,7	85,80	87,80
	NR-CNN1 [17]	87,2	84,3	86,4	86,2	92,2	85,6	91,3	85,2	83,6	82,7	89,28	85,33
	NR-CNN2 [18]	93,4	95,6	86,2	84,3	**94,1**	90,3	80,4	82,2	81,8	82,5	88,53	88,10
	BMQI [15]	20,1	-	83,5	-	88,9	-	92,7	-	78,1	-	71,30	-
	CNNs-CMP [10]	**95,8**	95,6	**93,6**	92,9	93,4	**91,3**	**94,5**	**95,2**	**94,4**	**94,8**	**94,33**	**93,75**
Our Approach - Base Model	Brisque-R-V	20,84	41,61	75,72	83,48	52,23	69,6	68,38	71,61	38,81	44,91	54,29	66,58
	Brisque-R-P	27,61	52,24	84,64	90,5	80,01	80,6	85,66	83,07	63,72	70,34	69,48	76,60
	Brisque-R-VP	22,98	38,87	89,84	91,96	91,53	95,69	93,56	93,2	54,36	61,18	74,48	79,93
	DCFQI-BM-A-V [21]	96,3	98,2	78,9	89	98,4	99,5	96,2	95,7	90,4	89,9	92,45	95,60
	DCFQI-BM-A-P [21]	98,9	98	81,4	98,1	99,5	99,7	98,1	97	92,4	92,1	94,48	98,20
	DCFQI-BM-R-V	95,1	98,4	95,1	95,7	99,3	99,6	97	96,4	96,4	96,6	96,63	97,53
	DCFQI-BM-R-P	95,7	98,6	91,8	93,1	99,4	99,6	98,8	98,6	95,9	96,7	96,43	97,48
	DCFQI-BM-R-VP	95,7	**98,7**	97	97,6	**99,9**	**99,8**	98,8	**99**	97	97,4	97,85	**98,78**
	DCFQI-BM-R-VPB	**96,3**	**98,7**	**98,6**	**98**	99,8	99,6	**99,1**	98,7	**97,9**	**98,3**	**98,45**	98,75
Our Approach - Cumulative Model	DCFQI-CM-A-V	99,2	99,8	99,5	99,8	99,7	99,8	98,6	99,2	96,5	96,6	99,25	99,65
	DCFQI-CM-A-P	99,7	99,9	99,8	99,8	**99,9**	99,9	99,1	99,6	99,1	99,4	99,63	99,80
	DCFQI-CM-R-V	98,8	99,8	99,8	99,9	99,3	99,8	98,8	99,4	98,7	99	99,18	99,73
	DCFQI-CM-R-P	99,7	99,9	99,7	99,8	99,8	99,8	**99,4**	**99,8**	99	99,2	99,65	99,83
	DCFQI-CM-R-VP	**99,8**	**100**	**99,9**	99,9	**99,9**	99,9	99,3	99,7	99,1	99,4	**99,73**	**99,88**
	DCFQI-CM-R-VPB	98,9	99,7	**99,9**	**100**	**99,9**	**100**	99,3	99,7	**99,4**	**99,7**	99,50	99,85

TABLE 3. Comparison of SROOC and PLCC values of our proposed methods with the state of the art applied to the Liris masking database.

Correlation coefficients SROOC and PLCC of different objective metrics on the LIRIS masking database													
Type	Metric	Armadillo		Lion		Bimba		Dyno		All models		Average	
		SROOC	PLCC	SROOC	PLCC	SROOC	PLCC	SROOC	PLCC	SROOC	PLCC	SROOC	PLCC
Full Reference	HD [29]	48,6	37,7	71,4	25,1	25,7	7,5	48,6	31,1	26,6	4,1	48,58	25,35
	RMS [30]	65,7	44,6	71,4	23,8	71,4	21,8	71,4	50,3	48,8	17	69,98	35,13
	MSDM2 [12]	88,6	65,8	**94,3**	87,5	**100**	93,7	**100**	91,7	89,6	76,2	95,73	84,68
	TPDM [31]	88,6	91,4	82,9	88,4	**100**	97,2	**100**	97,1	90	88,6	92,88	93,53
	TPDMPW [33]	-	-	-	-	-	-	-	-	**94,2**	91,7	-	-
	Chetouani [34]	**99**	**99**	83	**94**	99	**99**	93	**98**	93,9	**97,8**	**93,50**	**97,50**
Reduced Reference	3DWPM1 [35]	58	41,8	20	9,7	20	8,4	66,7	45,3	29,4	10,2	41,18	26,30
	3DWPM2 [35]	48,6	37,9	38,3	22	37,1	14,4	71,4	50,1	37,4	18,2	48,85	31,10
	FMPD [13]	94,2	88,6	93,5	94,3	**98,9**	**100**	**96,9**	**94,3**	**80,8**	**80,2**	**95,88**	**94,30**
	DAME [36]	**94,3**	**96**	**100**	**99,5**	97,7	88	82,9	89,4	68,1	58,6	93,73	93,23
No-Reference	NR-SVR [19]	89,5	84,7	**100**	96,3	94,2	93,6	94,4	89,7	90,4	91,2	94,53	91,08
	NR-GRNN [14]	82,3	80,5	94,1	**97**	90,2	94,3	78,2	82,3	90,2	82,4	86,20	88,53
	NR-CNN1 [17]	95,2	**97,6**	89,4	91,6	93,4	**98,7**	**96,3**	89,9	88,2	85,4	93,58	**94,45**
	BMQI [15]	94,3	-	**94,3**	-	**100**	-	83	-	78,1	-	92,90	-
	CNNs-CMP [10]	**96,2**	95,5	93,1	92,4	92,5	92,8	94,2	**94**	**95,8**	**95,5**	**94,00**	93,68
Our Approach - Base Model	Brisque-R-V	-20	-44,9	42,86	33,6	82,86	88,04	81,17	83,15	41,79	41,75	46,72	39,97
	Brisque-R-P	-20	-13,59	94,29	96,33	100	99,58	81,17	85,88	74,07	78,69	63,87	67,05
	Brisque-R-VP	-8,57	-17,42	54,29	70,63	94,29	98,75	92,76	90,76	69,93	72,44	58,19	60,68
	DCFQI-BM-A-V [21]	82,9	92,2	**100**	92,5	94,3	92,6	**100**	92,3	85,4	85,3	94,30	92,40
	DCFQI-BM-A-P [21]	**100**	**98,1**	**100**	96,6	82,9	93,4	**100**	96,8	92,5	88,1	95,73	96,23
	DCFQI-BM-R-V	71,4	90,9	**100**	99	**100**	97,8	81,2	94,1	**93,8**	95,6	88,15	95,45
	DCFQI-BM-R-P	82,9	72,2	94,3	89,1	**100**	96,1	98,6	95,6	89,2	82,7	93,95	88,25
	DCFQI-BM-R-VP	92,8	93,2	98,6	91,6	98,6	92,1	98,6	80,5	93,4	87,3	**97,15**	89,35
	DCFQI-BM-R-VPB	82,9	88,5	**100**	**99,3**	**100**	**99,9**	98,6	**99,4**	92,6	**94,1**	95,38	**96,78**
Our Approach - Cumulative Model	DCFQI-CM-A-V	98,6	86,6	98,6	96,8	92,8	94,5	94,1	92,6	89,1	89,6	96,03	92,63
	DCFQI-CM-A-P	**100**	96,8	**100**	94,2	**100**	94,8	**100**	90,8	94,1	89,7	**100,00**	94,15
	DCFQI-CM-R-V	92,8	78,1	98,6	93	98,6	91,2	98,6	92,4	91,2	86,3	97,15	88,68
	DCFQI-CM-R-P	98,6	78,3	98,6	92	98,6	91,5	81,2	88,2	87	82,4	94,25	87,50
	DCFQI-CM-R-VP	98,6	80,5	92,8	93,2	98,6	92,1	98,6	91,6	93,4	87,3	97,15	89,35
	DCFQI-CM-R-VPB	94,3	85,6	**100**	98,9	**100**	**99,9**	98,6	**99,3**	**97**	**96**	98,23	**95,93**

four databases introduced at the beginning of this section (Liris/EPFL General Purpose Database, Liris Masking Database, IEETA Simplification Database, UWB Compression Database). The results are repectively shown in Tables 2, 3, 4, 5. To ease the reading of the tables, we have: best rates bolded in each column for each category of approach, best SOTA approach is shown with color , our best Base Model is shown with color , best Cumulative Model is shown with color . The three overall best scores are shown in each column with colors , , and .

If the scores are presented for all the meshes of the databases, we will draw our conclusions only from the global scores: "All models" and "Average", as they are more representative. The Average column is the average of the SROOC and PLCC scores of all the meshes. The All models column is the computation of the SROOC and PLCC directly on all the models quality scores. This last measure is the most important one to analyze.

Whatever the database, using only the BRISQUE IQA scores for MVQA does not produce good results. This shows that classical IQA algorithms are not solely sufficient to learn 3D Mesh quality scores from views or patches. If we compare the BRISQUE-R-* scores with those of the DFCQI Base Models, using a simple averaging of the quality scores estimated from views or patches (DCFQI-BM-A-*), there is a systematic gain of performance for all the models. This shows the interest of learning a dedicated image quality score on projections instead of using BRISQUE. Now, if we have a look to all of the DFCQI Base Models, we can see that it is always beneficial to replace the averaging score quality aggregator by a MLP non-linear Regressor. Therefore, the relationship between the estimated 2D image quality scores

TABLE 4. Comparison of SROOC and PLCC values of our proposed methods with the state of the art applied to the IEETA database.

Correlation coefficients SROOC and PLCC of different objective metrics on the IEETA simplification database															
Type	Metric	Bones		Bunny		Head		Lung		Strange		All models		Average	
		SROOC	PLCC	SROOC	PLCC	SROOC	PLCC	SROOC	PLCC	SROOC	PLCC	SROOC	PLCC	SROOC	PLCC
Full Reference	HD [29]	94,3	84,8	39,5	14,3	88,6	53	88,6	64,9	37,1	27,4	49,4	25,5	69,62	48,88
	RMS [30]	94,3	71,1	77,1	79,2	42,9	23,1	94,3	71,3	94,3	92,4	64,3	34,4	80,58	67,42
	MSDM2 [12]	77,1	**96,7**	94,3	**96,3**	**88,6**	**79**	65,7	85,3	**100**	**98,1**	86,7	79,6	85,14	**91,08**
	TPDM [31]	**99**	94,3	**98**	94,3	63,1	65,7	**98,6**	94,3	98,7	**94,3**	**86,9**	**88,2**	**91,48**	88,58
Reduced Reference	FMPD [13]	**88,6**	**96**	**94,3**	**98**	**65,7**	**70,4**	**88,6**	**95,5**	**65,7**	**96**	**87,2**	**89,3**	**80,58**	**91,18**
No-Reference	CNNs-CMP [10]	**91,3**	**88,9**	**91,1**	**92,4**	**91,8**	**91,5**	**95,3**	**89,4**	**91,1**	**88,9**	**91**	**91,1**	**92,12**	**90,22**
Our Approach - Base Model	Brisque-R-V	65,71	74,61	54,29	51,21	64,71	92,91	77,14	92,34	89,86	81,67	59,96	62,47	70,34	78,55
	Brisque-R-P	88,57	95,95	-60	-66,59	100	99,02	94,29	99,72	81,17	91,02	51,99	58,4	60,81	63,82
	Brisque-R-VP	88,57	95,44	-25,71	-30,06	100	97,91	77,14	99,58	81,17	92	64,33	66,2	64,23	70,97
	DCFQI-BM-A-V [21]	60	57,55	71,43	94,47	82,35	90,85	82,86	79,23	95,49	95,05	60,34	51,07	78,43	83,43
	DCFQI-BM-A-P [21]	71,43	80,53	82,86	**98,64**	**100**	99,21	77,14	96,7	95,49	97,59	83,7	74,5	85,38	94,53
	DCFQI-BM-R-V	94,29	82,91	71,43	85,7	95,49	93,93	88,57	99,74	77,01	58,53	86,48	77,81	85,36	84,16
	DCFQI-BM-R-P	88,57	97,99	**88,57**	96,84	82,35	98,49	94,29	99,81	89,33	68,24	78,84	73,14	88,62	92,27
	DCFQI-BM-R-VP	**100**	**99,83**	71,43	76,94	82,35	99,02	94,29	99,11	95,49	**98,78**	79,1	78,66	88,71	94,74
	DCFQI-BM-R-VPB	88,57	99,25	60	86,01	**100**	**99,75**	**100**	**99,94**	**98,56**	98,54	**89,56**	**89,76**	**89,43**	**96,70**
Our Approach - Cumulative Model	DCFQI-CM-A-V	71,43	96,16	94,29	**99,79**	**100**	99,43	94,11	87,42	95,49	87,86	88,59	81,55	91,06	94,13
	DCFQI-CM-A-P	**94,29**	95,9	82,86	98,81	**100**	98,32	**100**	84,4	**100**	96,16	84,09	74,59	**95,43**	94,72
	DCFQI-CM-R-V	88,57	98,79	100	87,26	100	98,34	94,11	98,49	89,33	88,8	90,15	82,45	94,40	94,34
	DCFQI-CM-R-P	77,14	83,52	65,71	75,91	100	99,27	92,76	99,64	89,33	91,69	79,31	72,42	84,99	90,01
	DCFQI-CM-R-VP	88,57	**99,22**	82,86	92,85	100	99,19	88,04	99,37	98,56	95,31	86,19	86,53	91,61	97,19
	DCFQI-CM-R-VPB	77,14	97,83	**94,29**	98,85	**100**	**99,82**	94,29	**99,95**	92,76	**98,88**	**92,92**	**89,52**	91,70	**99,07**

TABLE 5. Comparison of SROOC and PLCC values of our proposed Indexes with the state of the art applied to the compression database.

Correlation coefficients SROOC and PLCC of different objective metrics on the UWB compression database															
Type	Metric	Bunny		James		Jessy		Nissan		Helix		All models		Average	
		SROOC	PLCC	SROOC	PLCC	SROOC	PLCC	SROOC	PLCC	SROOC	PLCC	SROOC	PLCC	SROOC	PLCC
Full Reference	HD [29]	34,1	52,2	-16,8	6,8	-23,6	12,5	14,4	23,6	45,1	46,4	10,6	28,3	10,64	28,30
	RMS [30]	34,2	20,9	14	10,8	0	14,8	17,8	29,7	46,9	44,6	22	24,1	22,58	24,16
	MSDM2 [12]	**97,4**	90,1	82,6	69,2	84,3	63,1	**84,4**	73,1	98,1	94,7	89,3	78	89,36	78,04
	TPDM [31]	95,1	**96,5**	**90,8**	**73,6**	**85,8**	**75,8**	82,7	**73,4**	**98,7**	**95**	**91,5**	82,9	**90,62**	**82,86**
	TPDMPW [33]	-	-	-	-	-	-	-	-	-	-	91,3	**96,4**	-	-
Reduced Reference	3DWPM1 [35]	94,7	93,4	77,3	72,3	87,2	**89,5**	63,6	59,3	98	**95,2**	84,1	81,9	84,16	81,94
	3DWPM2 [35]	96	91,2	76,9	65,3	**86,9**	85,9	56,3	67,6	95,5	94,3	82,3	80,9	82,32	80,86
	FMPD [13]	94,2	89,6	95,3	91,2	63,3	60	92,4	**77,5**	**98,4**	90,8	88,8	81,8	88,72	81,82
	DAME [36]	**96,8**	**93,4**	**95,7**	**93,4**	84,4	70,5	**93,9**	75,3	96,6	**95,2**	**93,5**	**85,6**	**93,48**	**85,56**
No-Reference	CNNs-CMP [10]	**95,6**	**94,8**	**92,5**	**90,6**	**92,5**	**87,1**	**88,7**	**89**	**90,7**	**90,4**	**92,7**	**93,8**	**92,00**	**90,38**
Our Approach - Base Model	Brisque-R-V	41,54	31,65	17,6	20,63	42,9	41,82	92,09	82,29	4,18	-1,17	38,6	35,72	39,66	35,04
	Brisque-R-P	96,48	95,18	79,43	81,65	97,69	99,04	94,29	97,38	57,36	57,43	55,93	60,09	85,05	86,14
	Brisque-R-VP	92,97	97,65	89,55	85,09	98,13	99,38	94,29	97,4	70,99	64,66	63,59	65,01	89,19	88,84
	DCFQI-BM-A-V [21]	**99,12**	**98,3**	81,1	93,7	92,09	97,82	92,51	98,3	**99,12**	99,3	75,39	65,41	92,79	97,48
	DCFQI-BM-A-P [21]	95,59	93,59	83,3	83,3	90,77	96,14	88,11	98,02	97,36	99,16	67,67	67,92	91,03	94,04
	DCFQI-BM-R-V	96,81	97,45	70,11	83,85	80,22	96,13	97,25	95,04	90,33	97,43	85,27	73,73	86,94	93,98
	DCFQI-BM-R-P	86,03	92,31	94,73	93	91,65	98,29	95,49	99	91,21	98,32	87,41	**90,63**	91,82	96,18
	DCFQI-BM-R-VP	89,55	93,33	96,48	94,65	95,6	98,57	95,93	99,53	92,09	97,85	82	79,17	93,93	96,79
	DCFQI-BM-R-VPB	93,29	94,95	**96,92**	**98,42**	**100**	**99,9**	**99,89**	**99,89**	96,92	**99,75**	**86,88**	88,49	**97,40**	**98,58**
Our Approach - Cumulative Model	DCFQI-CM-A-V	95,15	96,25	99,12	95,75	97,25	97,84	**99,56**	98,45	**100**	**98,78**	87,2	86,6	98,22	97,41
	DCFQI-CM-A-P	**99,56**	95,88	**99,56**	97,13	98,24	97,93	96,78	97,53	99,56	96,56	97,09	96,57	**98,74**	97,01
	DCFQI-CM-R-V	94,7	95,68	95,05	96,17	96,15	97,83	84,93	93,65	91,21	97,2	88,19	87,69	92,41	96,11
	DCFQI-CM-R-P	99,45	95,87	**99,56**	97,08	96,48	97,42	97,22	97,96	95,14	96,5	93,33	94,11	97,57	96,97
	DCFQI-CM-R-VP	95,93	96,66	99,45	97,26	97,69	96,76	97,69	96,76	96,92	97,22	95,62	94,97	97,54	96,93
	DCFQI-CM-R-VPB	98,13	**98,22**	98,68	**98,57**	**99,56**	**99,45**	98,57	**99,7**	95,6	98,4	**97,94**	**97,79**	98,11	**98,87**

is clearly not linear. This enhances the results of our previous works that used an averaging aggregator [21].

If we compare the best DFCQI Base Models to the best state-of-the-art approach CNNs-CMP [10], we are very competitive (on Liris Masking, IEETA databases) or better (on Liris General Purpose,Compression databases) when using a combination of views and patches (DCFQI-BM-R-VP). In addition, CNNs-CMP [10] uses a complex approach that relies on the combination of features from three pre-trained models (VGG/AlexNet/ResNet) combined with saliency patch-based selection (with small patches of size 32×32). Our approach shows that, even with the use of views, we can obtain better results and enhance them by combining predictions from both views and patches.

Lastly if we consider a cumulative model, we surpass the state-of-the-art for all meshes, except for the IEETA Database where this depends on the considered evaluation metrics but we are always better with SROOC versus CNNs-CMP. As shown in [10], the deep features extracted from ResNet [28] enable to obtain better results on this database, and this could be considered to further enhance our results. A cumulative model using both view and patches (DCFQI-CM-R-VP) works well on all the databases, but the results can be further enhanced by incorporating the BRISQUE scores (DCFQI-BM-R-VPB). So, if the BRISQUE IQA scores solely are not of interest, they can bring some supplementary information to our learned 2D projection quality scores.

All these results show the benefit of our approach at different levels and how it can be tuned to overpass the state-of-the-art. To end these experiments, we provide some qualitative results. Figure 13 presents distorted meshes from each of the considered databases along with the reference MOS and the PMMOS predicted by our model DCFQI-CM-R-VPB. One can see that the predicted values are very close to the reference ones.

V. CONCLUSION

In this paper, we have introduced an innovative no-reference approach for assessing mesh quality. The method involves rendering the mesh into 2D views, which are subsequently divided into patches. Deep features are then extracted from these images using the pre-trained VGG16 Convolutional Neural Network (CNN) and fed into a Multilayer Perceptron (MLP) to predict the quality of the 2D projections based on the reference Mean Opinion Score (MOS) of the 3D mesh. Our base model demonstrates competitiveness with the state-of-the-art, particularly when incorporating scores from both views and patches. Additionally, we proposed a

cumulative training approach to obtain a final unified model for prediction, which surpasses the current state-of-the-art performance across all considered databases. Future research directions will explore the application of our metrics for estimating the quality of colored meshes [11] and the use of alternative deep representation such as Vision Transformers.

REFERENCES

[1] A. S. Lalos, E. Vlachos, G. Arvanitis, K. Moustakas, and K. Berberidis, "Signal processing on static and dynamic 3D meshes: Sparse representations and applications," *IEEE Access*, vol. 7, pp. 15779–15803, 2019, doi: 10.1109/ACCESS.2019.2894533.

[2] H. Chen, Z. Li, M. Wei, and J. Wang, "Geometric and learning-based mesh denoising: A comprehensive survey," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 20, no. 3, pp. 1–28, Nov. 2023.

[3] H. Hoppe, "Progressive meshes," in *Proc. 23rd Annu. Conf. Comput. Graph. Interact. Techn.*. New York, NY, USA: Association for Computing Machinery, Aug. 1996, pp. 99–108.

[4] J. Li, D. Chen, F. Hu, Y. Wang, P. Li, and J. Peethambaran, "Shape-preserving mesh decimation for 3D building modeling," *Int. J. Appl. Earth Observ. Geoinformation*, vol. 126, Feb. 2024, Art. no. 103623.

[5] G. Lavoué, E. D. Gelasca, F. Dupont, A. Baskurt, and T. Ebrahimi, "Perceptually driven 3D distance metrics with application to watermarking," in *Proc. SPIE*, vol. 6312, 2006, p. 63120.

[6] H. Zhou, K. Chen, Z. Ma, F. Wang, and W. Zhang, *Triangle Mesh Watermarking and Steganography*. Springer, 2023.

[7] A. Maglo, G. Lavoué, F. Dupont, and C. Hudelot, "3D mesh compression: Survey, comparisons, and emerging trends," *ACM Comput. Surveys*, vol. 47, no. 3, pp. 1–41, Feb. 2015.

[8] G. Luo, X. Zhao, Q. Chen, Z. Zhu, and C. Xian, "Dynamic data reshaping for 3D mesh animation compression," *Multimedia Tools Appl.*, vol. 81, no. 1, pp. 55–72, Jan. 2022.

[9] J. Yang, M. Lyu, Z. Qi, and Y. Shi, "Deep learning based image quality assessment: A survey," *Proc. Comput. Sci.*, vol. 221, pp. 1000–1005, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050923008384

[10] I. Abouelaziz, A. Chetouani, M. El 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, Apr. 2020, Art. no. 107174, doi: 10.1016/j.patcog.2019.107174.

[11] 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, Nov. 2022, doi: 10.1109/TCSVT.2022.3186894. https://doi.org/10.1109/TCSVT.2022.3186894

[12] G. Lavoué, "A multiscale metric for 3D mesh visual quality assessment," *Comput. Graph. Forum*, vol. 30, no. 5, pp. 1427–1437, Aug. 2011, doi: 10.1111/j.1467-8659.2011.02017.x.

[13] K. Wang, F. Torkhani, and A. Montanvert, "A fast roughness-based approach to the assessment of 3D mesh visual quality," *Comput. Graph.*, vol. 36, no. 7, pp. 808–818, Nov. 2012, doi: 10.1016/j.cag.2012.06.004.

[14] I. Abouelaziz, M. El Hassouni, and H. Cherifi, "A curvature based method for blind mesh visual quality assessment using a general regression neural network," in *Proc. 12th Int. Conf. Signal-Image Technol. Internet-Based Syst. (SITIS)*, Nov. 2016, pp. 793–797, doi: 10.1109/SITIS.2016.130.

[15] A. Nouri, C. Charrier, and O. Lézoray, "3D blind mesh quality assessment index," in *Proc. 3DIPM*, 2017, pp. 9–16, doi: 10.2352/ISSN.2470-1173.2017.20.3DIPM-002.

[16] Y. Lin, M. Yu, K. Chen, G. Jiang, F. Chen, and Z. Peng, "Blind mesh assessment based on graph spectral entropy and spatial features," *Entropy*, vol. 22, no. 2, p. 190, Feb. 2020, doi: 10.3390/e22020190.

[17] I. Abouelaziz, M. E. Hassouni, and H. Cherifi, "A convolutional neural network framework for blind mesh visual quality assessment," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 755–759, doi: 10.1109/ICIP.2017.8296382.

[18] I. Abouelaziz, A. Chetouani, M. E. Hassouni, and H. Cherifi, "A blind mesh visual quality assessment method based on convolutional neural network," in *Proc. 3DIPM*, 2018, pp. 1–5, doi: 10.2352/ISSN.2470-1173.2018.18.3DIPM-423.

[19] I. Abouelaziz, M. E. Hassimosouni, and H. Cherifi, "No-reference 3D mesh quality assessment based on dihedral angles model and support vector regression," in *Proc. ICISP*, vol. 9680, 2016, pp. 369–377, doi: 10.1007/978-3-319-33618-3_37.

[20] I. Abouelaziz, A. Chetouani, M. E. Hassouni, L. J. Latecki, and H. Cherifi, "Convolutional neural network for blind mesh visual quality assessment using 3D visual saliency," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Athens, Greece, 2018, pp. 3533–3537, doi: 10.1109/ICIP.2018.8451763.

[21] Z. Ibork, A. Nouri, O. Lézoray, C. Charrier, and R. Touahni, "No reference 3D mesh quality assessment using deep convolutional features," in *Proc. Int. Symp. Image Signal Process. Anal. (ISPA)*, Rome, Italy, Sep. 2023, pp. 1–6, doi: 10.1109/ispa58351.2023.10278663.

[22] 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, Dec. 2012, doi: 10.1109/TIP.2012.2214050. https://doi.org/10.1109/TIP.2012.2214050

[23] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. ICLR*, 2015, pp. 1–14.

[24] L. Váša and J. Rus, "Dihedral angle mesh error: A fast perception correlated distortion measure for fixed connectivity triangle meshes," *Comput. Graph. Forum*, vol. 31, no. 5, pp. 1715–1724, Aug. 2012, doi: 10.1111/j.1467-8659.2012.03176.x.

[25] G. Lavoué, "A local roughness measure for 3D meshes and its application to visual masking," *ACM Trans. Appl. Perception*, vol. 5, no. 4, pp. 1–23, Jan. 2009, doi: 10.1145/1462048.1462052.

[26] S. Silva, B. Santos, C. Ferreira, and J. Madeira, "A perceptual data repository for polygonal meshes," in *Proc. Int. Conf. Visualisation*, Aug. 2009, pp. 207–212.

[27] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. AISTATS*, vol. 9, 2010, pp. 249–256.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[29] N. Aspert, D. S. Cruz, and T. Ebrahimi, "MESH: Measuring errors between surfaces using the Hausdorff distance," in *Proc. ICME*, 2002, pp. 705–708, doi: 10.1109/ICME.2002.1035879.

[30] P. Cignoni, C. Rocchini, and R. Scopigno, "Metro: Measuring error on simplified surfaces," *Comput. Graph. Forum*, vol. 17, no. 2, pp. 167–174, Jun. 1998, doi: 10.1111/1467-8659.00236.

[31] F. Torkhani, K. Wang, and J. Chassery, "A curvature tensor distance for mesh visual quality assessment," in *Proc. Comput. Vis. Graph. Int. Conf. (ICCVG)*, in Lecture Notes in Computer Science, vol. 7594, Warsaw, Poland, L. Bolc, R. Tadeusiewicz, L. J. Chmielewski, and K. W. Wojciechowski, Eds., Germany: Springer, 2012, pp. 253–263, doi: 10.1007/978-3-642-33564-8_31.

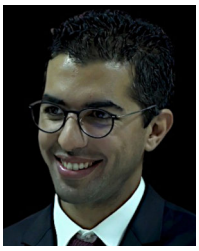
[32] Z. C. Yildiz and T. Capin, "A perceptual quality metric for dynamic triangle meshes," *EURASIP J. Image Video Process.*, vol. 2017, no. 1, p. 12, Dec. 2017, doi: 10.1186/s13640-016-0157-y.

[33] X. Feng, W. Wan, R. Yi Da Xu, S. Perry, P. Li, and S. Zhu, "A novel spatial pooling method for 3D mesh quality assessment based on percentile weighting strategy," *Comput. Graph.*, vol. 74, pp. 12–22, Aug. 2018, doi: 10.1016/j.cag.2018.04.005.

[34] A. Chetouani, "Three-dimensional mesh quality metric with reference based on a support vector regression model," *J. Electron. Imag.*, vol. 27, no. 4, p. 1, Aug. 2018, doi: 10.1117/1.jei.27.4.043048.

[35] M. Corsini, E. D. Gelasca, T. Ebrahimi, and M. Barni, "Watermarked 3-D mesh quality assessment," *IEEE Trans. Multimedia*, vol. 9, no. 2, pp. 247–256, Feb. 2007, doi: 10.1109/TMM.2006.886261.

ZAINEB IBORK received the master's degree in embedded electronics from Ibn Tofail University, Kénitra, Morocco, in 2020. She is currently pursuing the Ph.D. degree in computer science with Ibn Tofail University and the University of Caen Normandy, France. She performs the Ph.D. research with the SETIME Laboratory, Ibn Tofail University; and GREYC, UMR CNRS 6072 Laboratory, University of Caen Normandy. Her research interests include 3-D computer vision, machine/deep learning, and 3-D meshes quality assessment.



ANASS NOURI received the M.Sc. degree in intelligent systems and image analysis from Ibn Tofail University and the University of Poitiers, France, in 2013, and the Ph.D. degree in computer science from the University of Caen Normandy, France, in 2016. From 2016 to 2017, he was an Assistant Professor with the Computer Science Department, National School of Engineering of Caen (ENSICAEN), affiliated with the GREYC Laboratory (UMR 6072), focusing his research on 3-D mesh analysis. From 2017 to 2018, he was a Postdoctoral Fellow with the Institut du Thorax Laboratory (UMR 1087) and RMES Laboratory (UMR 1229), University of Nantes, concentrating on the development of automatic detection and characterization techniques for 3-D bifurcations to predict intracranial aneurysms. From 2018 to 2019, he held the position of an Associate Professor of computer science with the School of Engineering, Institut Suprieur de lElectronique et du Numrique (ISEN), Brest, France. He was part of the Vision Research Team, L@bIsen. Currently, he is an Associate Professor of computer science with Ibn Tofail University, Kénitra, Morocco. He has been awarded two international patents and has contributed to over 16 papers at refereed international conferences, five in international journals and one book chapter. He teaches courses in computer programming, image analysis, and machine learning. His current research focuses on 2-D and 3-D image/volume processing, visual saliency, objective quality assessment, medical imaging, and machine learning.

OLIVIER LÉZORAY (Senior Member, IEEE) received the M.Sc., Ph.D., and Habilitation degrees in computer science from the University of Caen Normandy, France, in 1996, 2000, and 2007, respectively.

From 1999 to 2000, he was an Assistant Professor with the Computer Science Department, University of Caen. In 2000, he joined the West Normandy Institute of Technology, as an Associate Professor of computer science with the Communication Networks and Services Department; and the Director of Graduate Studies, from 2000 to 2009. In 2010, he became a Full Professor with the Multimedia and Internet (MMI) Department, West Normandy Institute of Technology. He was the Head of the MMI Department, from 2015 to 2021. Since 2022, he has been the Deputy Director of the GREYC, CNRS Research Laboratory. He is the author of two books, 11 book chapters, 50 papers in international journals, and more than 120 in international conferences. His research interests include graph-based signal processing, adaptive and multidimensional mathematical morphology, and machine learning. Applications of his works include computational photography, computer-aided diagnosis, and computer vision. He is a member of SPIE, IAPR, and EURASIP. He was an Associate Editor of IEEE TRANSACTIONS ON SIGNAL PROCESSING and *IET Image Processing*. He has guest-edited the following journals: *Signal Processing: Image Communication*, *Journal of Electronic Imaging*, *Signal Image and Video Processing*, *Computerized Medical Imaging and Graphics*, *Signal Processing*, and *EURASIP Journal on Advances in Signal Processing*.

CHRISTOPHE CHARRIER (Member, IEEE) received the Ph.D. degree in computer science from the University Jean Monnet, Saint-Etienne, in 1998. From 1998 to 2001, he was a Research Assistant with the Laboratory of Radio Communications and Signal Processing, Laval University, Quebec City, QC, Canada. In 2001, he joined the West Normandy Institute of Technology, Saint-Lo, as an Assistant Professor. In 2008, he joined the GREYC Laboratory and the Security, Architecture, Forensics and biomEtrics (SAFE) Research Unit. In 2008, he was a Visiting Scholar with the Laboratory for Image and Video Engineering, The University of Texas at Austin. From 2009 to 2011, he was an Invited Professor with the Computer Department, University of Sherbrooke, Sherbrooke, QC. Since 2016, he has been the Head of the SAFE Research Group; and the Head of the Multimedia and Internet (MMI) Department, West Normandy Institute of Technology, since 2021. He is the author of 13 book chapters, 24 articles in international journals, and more than 90 papers in international conferences. His current research interests include digital image and video forensics (deepfake video detection), image and video coding, processing, quality assessment, computational vision, and biometrics (evaluation of biometric systems and fingerprint quality assessment). He has been an Associate Editor of *IET Biometrics*, since 2020.

RAJA TOUAHNI has been a Professor of higher education with the Physics Department, Faculty of Sciences, since 1987. She is a member of the Systmes Electroniques, Traitement de lInformation, Mcanique (SETIME) Laboratory. She is the Head of the Data Processing and Artificial Intelligence Research Team. She has co-authored more than 90 scientific papers publishing in refereed international conferences and journals and three book chapters. She has supervised several national thesis, responsible, and a member of several projects (PHC Volubilis, Erasmus, and Tempus Plus-Scolamar). Her current research interests include image processing, data analysis, and e-learning. She is a Founding Member of the Moroccan Association of Classification (SMC) and the Moroccan Association of the Advanced Materials (A2MA).