3D visual saliency and convolutional neural network for blind mesh quality assessment

Ilyass Abouelaziz, Aladine Chetouani, Mohammed El Hassouni, Lj Latecki, Hocine Cherifi

Neural Computing and Applications
https://doi.org/10.1007/s00521-019-04521-1

IAPR-MEDPRAI

3D visual saliency and convolutional neural network for blind mesh quality assessment

Ilyass Abouelaziz[1] · Aladine Chetouani[2] · Mohammed El Hassouni[1,3] · Longin Jan Latecki[4] · Hocine Cherifi[5]

Received: 26 October 2018 / Accepted: 5 October 2019

Abstract
A number of full reference and reduced reference methods have been proposed in order to estimate the perceived visual quality of 3D meshes. However, in most practical situations, there is a limited access to the information related to the reference and the distortion type. For these reasons, the development of a no-reference mesh visual quality (MVQ) approach is a critical issue, and more emphasis needs to be devoted to blind methods. In this work, we propose a no-reference convolutional neural network (CNN) framework to estimate the perceived visual quality of 3D meshes. The method is called SCNN-BMQA (3D visual saliency and CNN for blind mesh quality assessment). The main contribution is the usage of a CNN and 3D visual saliency to estimate the perceived visual quality of distorted meshes. To do so, the CNN architecture is fed by small patches selected carefully according to their level of saliency. First, the visual saliency of the 3D mesh is computed. Afterward, we render 2D projections from the 3D mesh and its corresponding 3D saliency map. Then the obtained views are split into 2D small patches that pass through a saliency filter in order to select the most relevant patches. Finally, a CNN is used for the feature learning and the quality score estimation. Extensive experiments are conducted on four prominent MVQ assessment databases, including several tests to study the effect of the CNN parameters, the effect of visual saliency and comparison with existing methods. Results show that the trained CNN achieves good rates in terms of correlation with human judgment and outperforms the most effective state-of-the-art methods.

Keywords Mesh visual quality assessment · Mean opinion score · Mesh visual saliency · Convolutional neural network

1 Introduction

Technological advances in computer graphics, telecommunication, and computer-aided design over the two past decades have contributed to the development of three-dimensional (3D) data. Thanks to the rapid development of hardware and software for both professionals (3D modeling tools) and users (graphic cards, smart-phones capable of viewing 3D models), 3D data is widely used nowadays. Indeed, several application fields are directly concerned by this type of data such as architecture, cultural heritage (3D

✉ Ilyass Abouelaziz
ilyass.abouelaziz@um5s.net.ma

Aladine Chetouani
aladine.chetouani@univ-orleans.fr

Mohammed El Hassouni
mohamed.elhassouni@gmail.com

Longin Jan Latecki
latecki@temple.edu

Hocine Cherifi
hocine.cherifi@u-bourgogne.fr

[1] LRIT, URAC No 29, Faculty of Sciences, Mohammed V University in Rabat, B.P. 1014 RP, Rabat, Morocco

[2] University of Orleans PRISME Laboratory, Orléans, France

[3] FLSHR, Mohammed V University in Rabat, Rabat, Morocco

[4] Department of Computer and Information Sciences, Temple University, Philadelphia, USA

[5] LE2I, UMR 6306 CNRS, University of Burgundy, Dijon, France

Published online: 19 October 2019

digitizing of the ancient statues), digital recreation (video games, 3D movies), scientific visualization and so forth [1].

3D data (describing a human, animal or an object) can be represented in different ways; however, in most applications, it is represented by polygonal meshes which model the surface of objects by a set of vertices and faces. This representation, particularly triangular meshes, is widely used rather than other surface models such as implicit surfaces or parametric surfaces.

In practical situations, several geometric operations can be applied to 3D meshes such as compression [2, 3] to reduce the size of large 3D data, watermarking [4, 5] to protect the intellectual property of 3D content, simplification [6, 7] to decrease the number of vertices, noise perturbation during the transmission process and so forth.

In all such scenarios, it is crucial to identify how much the original model has been modified and assess the perceived visual quality of distorted models.

Two types of evaluation can be conducted. The most reliable evaluation process is the subjective visual quality assessment. Each distorted mesh is given a quality score by human observers in a controlled environment. Although this evaluation is very reliable, it is a time-consuming solution that is often too expensive to be adopted. The second evaluation process is called objective visual quality assessment. It relies on the computation of a quality metric that tries to mimic an ideal human observer [8]. It is a good solution to automatically assess the perceived visual quality of a distorted mesh. However, it must correlate well with the subjective assessment process.

Many objective methods have been proposed in the literature in order to estimate the perceived visual quality of distorted meshes. The well-known root-mean-squared error (RMS) [9] and the Hausdorff distance [10] are two methods that use simple geometric distances to assess the difference between two meshes. This type of methods is based only on pure geometric distances, and it does not take into consideration the perceptual information that describes the main operations of the HVS. Consequently, the predicted visual quality is not well reflected as proven by the moderate correlation with human perception [11, 12]. To overcome these drawbacks, many researchers have recently developed perceptually driven quality methods for 3D meshes [13, 14].

The remainder of this paper is organized as follows: We present in Sect. 2 the related work on 3D mesh quality assessment and the motivation behind our proposition. A detailed description of the proposed method SCNN-BMQA is given In Sect. 3. Section 4 is dedicated to the experimental setup and the obtained results. Finally, we present some concluding remarks and perspectives in Sect. 5.

2 Related work

Mesh visual quality (MVQ) objective methods can be generally classified depending on the availability of the reference: full-reference (FR), reduced-reference (RR) and no-reference (NR).

FR-MVQ assessment methods require the original model in order to compare it to the distorted version. Several methods have been proposed in this context. Karni and Gostman [15] proposed the Geometric Laplacian method (GL) to estimate the visual quality of compressed meshes. This method is based on a distance between the original model and its distorted version as a measure of smoothness of the vertices. For the quality estimation, this method computes a weighted sum of the vertex Laplacian coordinate error and the vertex root-mean-squared error. It is proven that the Laplacian coordinates are suitable to MVQ assessment since they are strongly related to the surface normal. Following this argument, Sorkine et al. [16] improved later this method by giving more importance to the Laplacian values. In the improved method, a greater weight is assigned to the Laplacian coordinate error rather than the vertex root-mean-squared error. Pan et al. [17] proposed another FR method for MVQ assessment. In their method, they experimentally studied the influence of many parameters as like the geometric resolution and the resolution of texture. Bian et al. [18] proposed a strain energy field-based measure (SEF). It is based on the energy introduced by a specific mesh deformation. The mesh is considered as an elastic object, and the visual deformation is related to the level of energy which causes the deformation. The perceptual distance is computed as the level of strain energy of the normalized triangular faces. Based on the well-known image quality metric, the SSIM (structural similarity) index [19], Lavoué et al. [20] proposed a metric called the Mesh Structural Distortion Measure (MSDM). In their method, they extend the SSIM to 3D meshes by using the mesh mean curvature as alternative for the pixel intensities in the SSIM index. The limitation of this method is that it assumes that the reference mesh and its distorted version share the same connectivity, which is not always valid. To overcome this issue, the author improves this method and proposes MSDM2 [21]. Compared to MSDM, the improved version MSDM2 can compare triangle meshes with different connectivities, by a vertex correspondence processing step. Another improvement is that a multiscale approach is used to evaluate the visual difference which leads to a considerable amelioration in predicting the perceived visual quality. Torkhani et al. [22] proposed a quality method called Tensor-based Perceptual Distance Measure (TPDM). This method computes a distance

between curvature tensors of the reference mesh and its distorted version. The curvature amplitudes and the principal curvature directions, which are obtained from the tensor eigenvalues and eigenvectors respectively, are used to compute a perceptually oriented tensor distance. FR-MVQ assessment is mostly employed for guiding mesh compression and watermarking. Although FR methods can correctly estimate the perceived visual quality, they cannot be used in most practical applications since the original models are not always available.

To overcome the limitations of FR-MVQ assessment, RR-MVQ is presented. This type of metrics requires only partial information about the reference mesh (i.e., extracted features and attributes). It is mostly used in many application fields such as real-time visual information communications.

In order to evaluate watermarking algorithms, Corsini et al. [23] proposed two RR quality methods. In their work, the authors suppose that the roughness of the 3D model is related to the visual quality. The first method relies on statistical parameters extracted from the dihedral angles to measure the global roughness of the 3D surface. The second method relies on a smoothing algorithm applied to a reference mesh to estimate the difference between the original model and the distorted version. The dihedral angles are also used by Vasa and Rus [24]. This method computes the difference between the original and the distorted mesh using oriented dihedral angles extracted from both meshes. In [25], Wang et al. proposed a RR method called fast mesh perceptual distance (FMPD). This method is based on the Gaussian curvature to extract a mesh local roughness measure. The perceptual distance between the reference mesh and the distorted version is defined as the difference between the normalized surface integrals of the local roughness measure. However, we still face the same issue that with FR methods. Indeed, when using RR methods, we need to extract features from the original models. This is not suitable for a lot of real-world applications. Another issue is that the extracted RR features need to be transmitted or embedded in the distorted images. This introduces an additional burden for quality assessment.

NR-MVQ seems to be a good solution to overcome the limitation of FR and RR approaches. This type of metrics relies only on the distorted mesh, and the quality estimation can be performed without knowing the reference. For this reason, NR-MVQ assessment methods are more appealing in practical situations, and the development of such methods becomes crucial to remedy this limitation. Abouelaziz et al. [26, 27] proposed two NR methods based on extracted features from the distorted meshes used to feed machine learning techniques. The first method considers the dihedral angles as a structural information descriptor. The feature distribution is fitted using statistical models in order to learn the support vector regression (SVR) for the quality prediction. The second method is based on the mean curvature features and uses the general regression neural network (GRNN) for the feature learning step. Nouri et al. [28] proposed a NR method called 3D blind mesh quality assessment index (BMQI). This method uses the SVR to learn the visual saliency and roughness features extracted from distorted meshes. Unlike for MVQ, several blind methods have been proposed to evaluate the visual quality of images, and they successfully estimate the perceived quality in terms of the correlation with subjective scores [29–31].

Convolutional neural networks (CNN) have recently attracted the attention of many researchers. They have been successfully employed in various computer vision applications allowing to reach high performances [32]. One of their main advantage over classical neural networks is that they adequately consider the spatial structure of the input data. Moreover, CNN allows the important property of weights sharing between the convolutional layers which restrict the number of parameters to learn. Their use in blind image quality assessment (BIQA) has shown notable improvement in terms of the correlation with the human judgment [33]. However, it has not yet been exploited for MVQ assessment. Indeed, studies in mesh quality assessment tend more to adopt FR and RR approaches, since they usually perform better than blind methods.

Building on these works, we propose a novel NR-MVQ assessment method called SCNN-BMQA (3D visual saliency and CNN for blind mesh quality assessment). It relies on the assumption that the human visual system (HVS) is more sensitive to distortions in salient regions, whereas in non-salient regions their influence on the overall judgment can be neglected. In this context, mesh visual saliency is used to indicate the most relevant regions of the 3D mesh. These regions are presented in the form of 2D small patches which are used to feed a CNN to learn an effective representation and estimate the perceived visual quality. SCNN-BMQA exploits the sensitivity of the HVS to mesh degradation together with the efficiency of the CNN learning approach. The proposed method is the first one that is able to outperform FR and RR approaches. Additionally, it proves to be significantly better than the alternative blind methods.

3 Proposed method

In this section, we describe the proposed MVQ assessment method in detail. SCNN-BMQA relies on the hypothesis that the HVS is more sensitive to degradation in salient

regions and that more importance must be given to these regions in the overall perceived visual quality. Consequently, mesh saliency is used to determine the most relevant patches from specific views of the 3D mesh. Instead of using handcrafted features, our method aims to acquire an effective mesh representation from raw images representing 2D projections of the 3D shape. SCNN-BMQA consists of three modules: mesh rendering, saliency-based patch selection, and feature learning to estimate a single objective quality score. Mesh rendering allows to obtain 2D projections in order to represent the 3D mesh from multiple views. The views are then split into small patches, and a saliency-based technique is further used to select the most relevant patches. Feature learning is implemented by a convolutional neural network, and a quality score is finally obtained using a regression method.

3.1 Flowchart

The flowchart of SCNN-BMQA is depicted in Fig. 1. Given a distorted 3D mesh, we first compute the level of saliency for each vertex to obtain a 3D saliency map. Then, we render 2D projections (views) from the 3D object and its corresponding 3D saliency map. The obtained projections are split into small patches of size 32×32. The saliency of each patch is used as a selection criterion as follows: All the patches with a level of saliency superior to a fixed threshold are considered as relevant patches, whereas the other patches are neglected. The selected patches are then used as input for the network after a normalization. We use a CNN with a regression method to estimate the objective score for each selected patch. The final quality score for the 3D object is obtained by averaging the scores.

3.2 Mesh views rendering

The first step of SCNN-BMQA consists of rendering 2D projections to represent the 3D mesh from multiple views. To do so, virtual cameras are fixed at different angles around the 3D mesh according to the axes X and Y. As illustrated in Fig. 2, the centroid of the 3D object is placed at the origin of the coordinate system. The coordinates (x_i, y_i) of the virtual cameras are obtained by varying the angles $x \in [0, 2\pi]$ and $y \in [0, 2\pi]$ by $\frac{\pi}{6}$ (30 degrees). Twelve angles are obtained for each axis. Hence, for each combination of x_i and y_i a virtual camera is placed and a 2D projection is obtained. In total, 144 projections are obtained from each 3D mesh.

3.3 Saliency-based patch selection

Our attention is generally attracted by salient visual stimuli. It is important for complex biological systems to quickly detect the most relevant regions in a given field of view. Visual saliency is a subjective phenomena that makes a region remarkable compared to others and immediately attracts our attention. The HVS has evolved to automatically detect salient regions. Visual saliency has been used previously for MVQ assessment. Anass et al. [28] proposed a blind mesh quality assessment index (BMQI) based on the estimation of visual saliency and roughness. In their work, they suppose that the quality of a 3D mesh is more affected in salient regions. The same authors proposed a full reference method [34] using a multiscale visual saliency map to compare the structural information between an original mesh and a distorted version. The relationship between visual saliency and distortion perception has been studied in [35]. It is claimed that the annoyance of the distortions depends strongly on the saliency of the regions that they appear in.

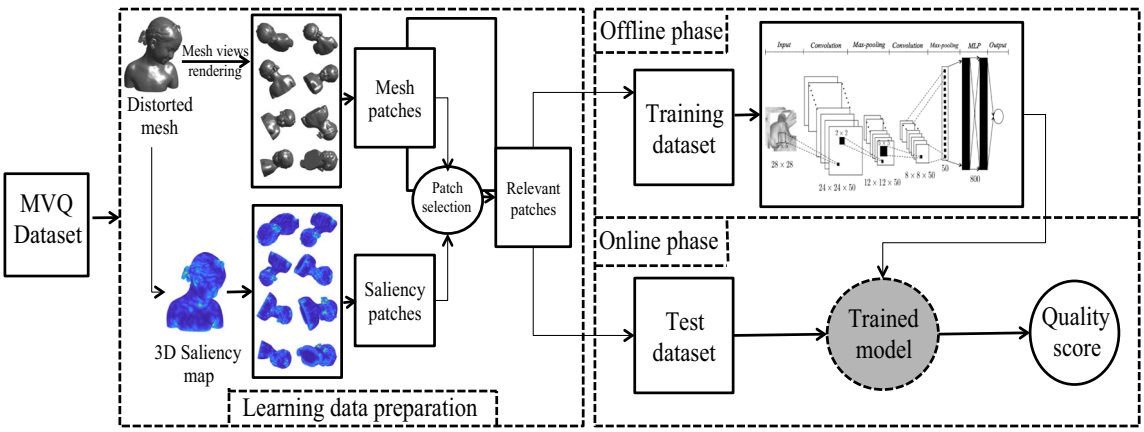

Fig. 1 Flowchart of the proposed saliency and CNN-based blind mesh quality assessment method : SCNN-BMQA

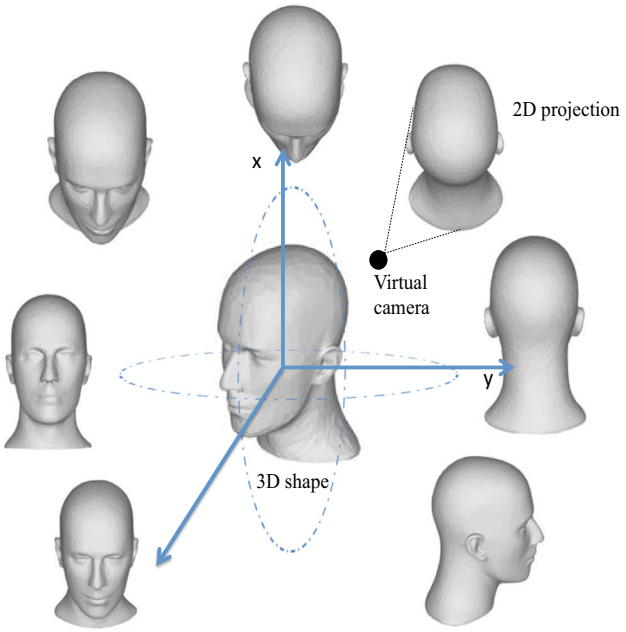

Fig. 2 Mesh views rendering, a virtual camera is placed for each combination of x and y. One hundred forty-four projections are obtained for each 3D mesh

Following this principle, we make the assumption that the subjective evaluation of the visual quality of a distorted mesh is strongly related to the distortion applied to salient regions. In other words, the human visual perception is impacted by the modification in salient regions since the visual attention is attracted automatically to these locations. In our work, we use mesh saliency to detect the salient patches from specific views of the 3D mesh. In order to compute the saliency map of 3D meshes and thus to detect perceptually important regions on mesh surfaces, we use the method proposed in [36]. This method is inspired by low-level HVS operations, and it is based on the center-surround mechanism adopted by the well-known Itti et al.'s method [37]. The process of computing mesh saliency is as follows: First, the mean curvature is computed at mesh vertices. The mean curvature is then filtered with a fine and coarse Gaussian. After that, the saliency is computed as the difference between the filtered mean curvatures within different scales by varying the Gaussian's standard deviation. The final saliency map is obtained by a nonlinear normalization sum of all the multiscale saliency maps. Figure 3 shows some examples of 3D meshes and their saliency maps. We can notice that some regions in the 3D shape are considerably more distinct and hence judged as salient. It is remarkable that regions with high curvature levels such as ears, nose, eyes, and paws attract more attention compared to smooth regions where the level of curvature is low.

It is noteworthy that in our work we do not propose a saliency method; we are decided to use an existing one. This method was evaluated by the authors, and the obtained results validate its good performance in capturing the salient regions. Our purpose in using visual saliency is to demonstrate its importance and usefulness in quality assessment and show how the salient regions are more susceptible to degradations that are easily detected by the human eye compared to the non-salient regions. Once the saliency map is obtained, we render 2D views following the same procedure described in the last section. The views obtained from the saliency map are used to select the salient locations in the corresponding 3D mesh as follows:

- First, we sample non-overlapping patches of size 32×32 from the 2D projections of the 3D mesh and its corresponding saliency map.
- For each patch of the saliency map, we compute the local level of saliency (LoS) which corresponds here to its average saliency value. The level of saliency LoS is used afterward to select the most relevant patches with a saliency threshold S_t set experimentally to 0.4 (more details can be found in Sect. 4.4). All the patches with $\text{LoS} \geq S_t$ are considered as relevant regions, whereas patches with $\text{LoS} < S_t$ are neglected. We note that the LoS is computed using only the pixels that contain the saliency information, the background pixels at object boundary are not considered. Thus, informative patches (with high saliency) at object boundary are not ignored.
- After that, we perform a local normalization on the retrieved patches which correspond to the salient regions in the 3D mesh.
- Finally, the selected patches are then used as input to our CNN model.

3.4 Feature Learning and quality score estimation

After the relevant patches are selected, the next step is to use machine learning to estimate the quality score. For that, a CNN model is used.

3.4.1 Input normalization

Before the training process, a simple local contrast normalization is applied on the input patches. The normalized value $\hat{I}(i,j)$ of a pixel $I(i, j)$ at location (i, j) is computed as follows:

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

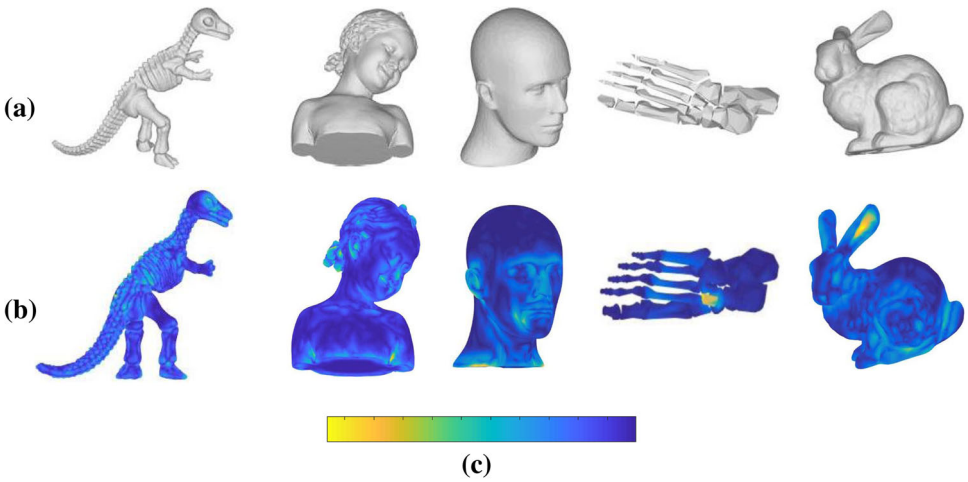

Fig. 3 Examples of mesh saliency: **a** 3D models, **b** their corresponding saliency maps and **c** is the colormap. The yellow color presents the most salient regions and the blue color presents the no salient regions

$$\mu(i,j) = \frac{1}{(2M+1)\times(2N+1)} \sum_{m=-M}^{m=M} \sum_{n=-N}^{n=N} I(i+m, j+n) \tag{2}$$

$$\sigma(i,j) = \sqrt{\sum_{m=-M}^{m=M} \sum_{n=-N}^{n=N} (I(i+m, j+n) - \mu(m,n))^2} \tag{3}$$

where c is a constant that prevents instabilities from dividing by zero. M and N are the normalization window sizes. The used normalization is very important to decrease the influence of the saturation problem. It is a type of distortion where the image is limited to some maximum values [38]. In addition, the normalization makes the network robust to illumination and contrast variation.

3.4.2 Network architecture

The next step consists in using a CNN in order to estimate the perceived visual quality. The CNN is fed by the normalized patches of size 32×32. Figure 4 shows the different layers of the used CNN.

Note that several network configurations have been tested in order to choose the best architecture for our method. The elaborated architecture in this section is the one that led to the best results (more details in Sect. 4.2).

The first layer is a convolutional layer; it filters the input patch with 32 kernels of size (5×5). The convolution process is defined as follows:

$$Y_i = W_i * X + b_i, \quad i = 1, 2, \ldots, N \tag{4}$$

where X is the input patch of the CNN and $*$ is the convolution operation. $\{W_i\}_{i=1}^{N}$ denotes the convolutional kernels and $\{b_i\}_{i=1}^{N}$ are the biases values. Thirty-two feature maps (28×28) are generated by this layer.

The second layer in our network is a max-pooling layer. It applies the max-pooling operation on the feature maps generated by the previous layer in order to reduce the dimension of the filter responses. The max-pooling operation is defined as follows:

$$M_{x,y}^{n} = \max_{(x,y)\in\Omega} (Y_{x,y}^{n}) \tag{5}$$

where $M_{x,y}^{n}$ denotes the output of the max-pooling layer (maximum values). $n = 1, 2, \ldots, N$ where N is the number of filters. Ω is a local window used in the pooling operation. In this layer, we use a local window of size 2×2 with a stride equals to 2. This provides 32 feature maps (14×14).

Another convolutional layer is introduced with 32 kernels (5×5). This layer provides 32 maps of size 10×10. It filters the feature maps obtained from the previous layer using the same process in Eq. 4.

After the second convolutional layer, we introduce another max-pooling layer of size 10×10. The size of the local window, in this case, is the same size of the maps produced by the previous layer; in other words, each feature map will be reduced to a single value and its output is thus a vector of size 1×32.

In order to estimate the perceived quality scores of a given distorted 3D model, the obtained feature vectors are then used to feed two fully connected layers with 500 neurons each. In our work, we adopt the nonlinear activation function ReLU (Rectified Linear Units) [39].

For the training process, we use the objective function adopted in [38] defined as follows:

$$\begin{aligned} L &= \frac{1}{N} \sum_{n=1}^{N} ||S(p_n; \omega) - \mathrm{MOS}_n||_{l1} \\ \hat{\omega} &= \min_{\omega} L \end{aligned} \tag{6}$$

where MOS_n is the subjective score assigned to a given input patch p_n and $S(p_n; \omega)$ is the estimated objective score of p_n with network weights ω. The Stochastic Gradient

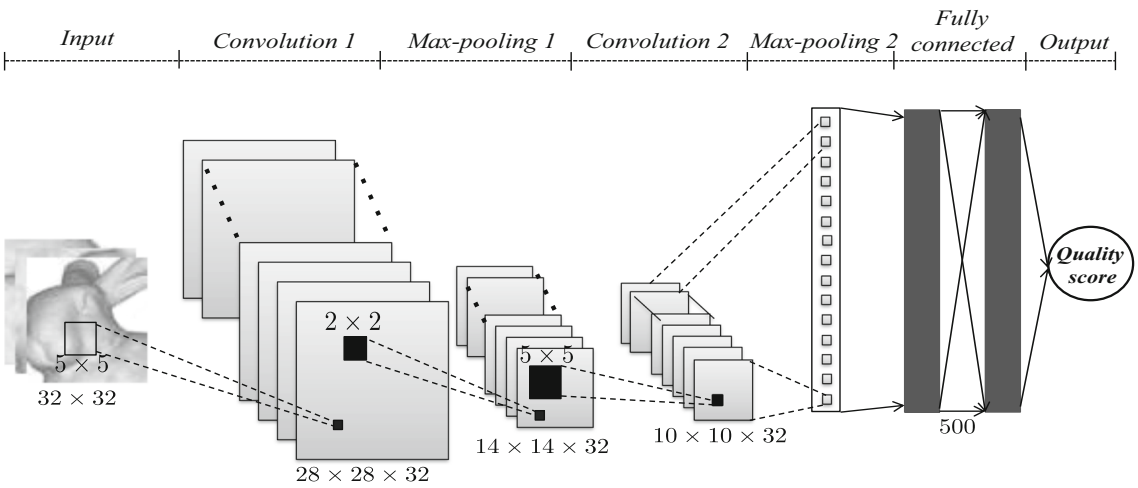

Fig. 4 Convolutional neural network configuration of the proposed method

Descent (SGD) and back propagation are used to learn the parameters of the CNN by minimizing the objective function defined in Eq. 6. We note that we perform SGD for 40 epochs in our experiments.

The leave-one-out cross-validation is used for the training process as follows: First, we build a training model as an offline phase using the patches from all the existing distorted objects in the repository except one group of distorted meshes (one object and its distorted versions). Then, the patches from the excluded object are used for the test (online phase) using the constructed model.

In the training, each patch is labeled by a quality score the same as the ground truth score of the source mesh as used in image quality assessment in [40]. Although some distortions are non-uniform in the tested databases, we can consider the same scores on all the salient patches since they tend to have the global MOS according to the assumption that the HVS is more sensitive to distortions in salient regions.

In order to study the effect of the CNN layers and parameters, several configurations are tested (see results in the next section).

4 Experiments

In this section, we evaluate the performance of SCNN-BMQA and the effectiveness of the CNN architecture for mesh visual quality assessment. We begin by describing the validation protocol and the used databases, i.e., the LIRIS/EPFL general-purpose database, the LIRIS masking database, the UWB compression database, and the IEETA simplification database. Then, we examine how the CNN parameters affect the performance of the network. After that, we investigate the importance of including the visual saliency-based patch selection technique in our method and how the performance of the CNN is affected. Finally, we present the experimental results and comparative analysis on mesh visual quality assessment state of the art.

4.1 Protocol

Objective MVQ methods goal is to provide quality scores that correlate well with human judgments. The performance of SCNN-BMQA is tested using four databases specially designed for MVQ assessment evaluation. Each database contains reference models and a variety of distorted versions. Depending on the level of distortion, a MOS value is assigned for each distorted mesh by a careful subjective evaluation. Figure 5 shows examples of the reference meshes from the four used databases that are defined as follows:

- LIRIS/EPFL general-purpose database [20]: Comprises 88 models including four references and 84 distorted versions (21 distortions for each reference model obtained by applying smoothing or noise addition either locally or globally). Twelve human observers participated in the subjective evaluation by giving a quality score between 0 (good quality) and 10 (bad quality). The MOS is obtained for each distorted model by averaging the twelve scores.
- LIRIS Masking database [41]: Comprises 28 models including four reference and 24 distorted versions (six distortions for each reference model obtained by applying only the local noise addition with different levels). This database is designed to capture the visual masking effect. Eleven human observers participated in the subjective evaluation and provide a quality score between 0 (bad quality) and 4 (good quality) for each distortion.

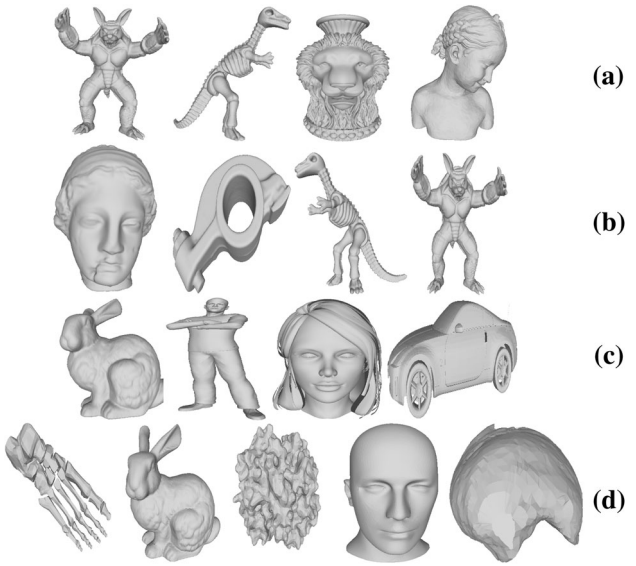

Fig. 5 The reference models from: the LIRIS masking database (**a**), the general-purpose database (**b**), the UWB compression database (**c**) and the IEETA simplification database (**d**)

- UWB compression database [24]: Contains 68 models (five references and 63 distorted versions). The distortions are obtained by applying 13 levels of compression distortions. A quality score between 0 (bad quality) and 1 (good quality) is provided for each distortion.
- The IEETA simplification database [42]: This database contains five reference models and 30 simplified versions (six distorted versions for each reference). The simplified models were obtained using three simplification algorithms with two different vertex reduction ratios.

It is noteworthy that the available databases for MVQ contain a limited amount of data; the largest database (the general-purpose database) includes only 88 models while training a CNN requires a huge amount of data. However, the decomposition of meshes as views and then patches allows us to obtained a sufficient set to train our model. For example, to evaluate the visual quality of the distorted versions of Dinosaur model on the general-purpose database, we use Armadillo, Venus and Rocker models , and their 21 deformed versions to build the training model in the offline phase. Then, the group of Dinosaur model and its 21 deformed versions are used for the test.

During the training, each distorted model provides around 13,787 training samples (patches). Talking about the last example, 3×21 models are used for the training, in total $3 \times 21 \times 13{,}787$ samples which represent 75% of the database versus $1 \times 21 \times 13{,}787$ for the test (25%).

To test the performance of quality metrics, it is common to compute the correlation between the subjective scores provided in the database and the objective scores obtained by the MVQ method. To do so, two correlation coefficients are used:

- The Pearson linear correlation coefficient (r_p) defined as follows:

$$r_p = \frac{\sum_{i=1}^{n}(Qs_i - \bar{Qs})(MOS_i - \bar{MOS})}{\sqrt{\sum_{i=1}^{n}(Qs_i - \bar{Qs})^2}\sqrt{\sum_{i=1}^{n}(MOS_i - \bar{MOS})^2}} \tag{7}$$

where Qs_i denotes the objective scores obtained by SCNN-BMQA. MOS_i is the subjective scores, and n is the number of distortions in the database.

- The Spearman rank-order correlation coefficient (r_s) defined as follows:

$$r_s = 1 - \frac{\sum_{i=1}^{n}(\mathrm{rank}(MOS_i) - \mathrm{rank}(Qs_i))^2}{n(n^2-1)} \tag{8}$$

The Pearson correlation coefficient r_p measures the linear dependence between the objective and subjective scores and is generally considered a more effective, and more important index than the Spearman coefficient r_s. It compares the actual score values by measuring the linear dependence between the objective and subjective scores. r_s measures how well the relationship between the objective and subjective scores can be described by a monotonic function [43]: As the value of the subjective scores increases, so does the value of the objective scores. Only the ranks of the scores are used in the computation of the r_s, not the actual score values.

4.2 CNN configuration

The CNN involves several parameters and provides an important degree of freedom to design an effective architecture for a specific application. In this work, several network architectures have been tested in order to investigate how the performance is affected by these parameters and choose the best configuration for our method. To do so, we first fix the patch size (32×32) and the kernel size (5×5) while testing the network with a different number of convolutional kernels. After that, we adjust the kernel size while fixing the number of kernels and the patch size. Finally, we examine the performance of the network by varying the patch size while fixing the number and the size of kernels to the best configuration obtained.

4.2.1 Effect of the number of filters in the convolutional layers

As the convolutional layer is the core building of a CNN. The number of convolutional adopted for this layer could also have an influence on the performance of the network. To demonstrate the influence of this parameter, we test the ability of our network in predicting the visual quality by using a variety of convolution kernels while fixing the other parameters. Table 1 presents the performance of the network regarding the correlation coefficients with respect to the size of convolution kernels.

It is shown from Table 1 that the number of kernels significantly affects the performance of the network. Using 32 kernels instead of 10 leads to an important improvement; however, using more kernels than 32 decreases the performance of the network in predicting the visual quality.

4.2.2 Effect of the size of filters

Another parameter tested in our experiments is the size of the convolution kernels. To do so, we fix the input patch size and the number of convolution kernels while testing different sizes of the kernels. Table 2 presents the performance of the network regarding the correlation coefficients with respect to the number of convolution kernels.

The kernel size also affects the performance of the network as shown in Table 2. Using a greater window size than 7×7 leads to lower correlations; however, the network is not strongly sensitive to the kernel size when using 3×3, 5×5 and 7×7 especially regarding the SROCC correlation.

4.2.3 Effect of the size of input data (patches)

As mentioned earlier, the proposed CNN is fed by small patches. Since these latter are sampled in a non-overlapping way, the size affects the number of patches obtained per views (i.e., smaller size leads to a larger number of patches). In this experiment, we examine how the input size affects the performance of our CNN in predicting the perceived visual quality. Table 3 presents the performance of the network regarding the correlation coefficients with respect to the input patch size variation.

As we can see from Table 3, the performance of the network is sensitive to the size of the input patch. The best correlations are provided when using input patches with size 32×32. Otherwise, using patches with size 128×128 provides the lowest results since the number of patches decreases strongly and thus the learning dataset becomes smaller.

According to these experiments, we adopt the CNN configuration that leads to the best correlation scores ($r_s = 93.3\%$ and $r_p = 92.2\%$):

- Input: (32×32)
- Conv1–32 (5×5)
- Max-pool1 (2×2)
- Conv2–32 (5×5)
- Max-pool2 (10×10)
- FC-500

4.3 Effect of the number of views

The 3D mesh is represented by different views obtained by fixing virtual cameras at different angles. The number of views is inversely proportional to the rotation angle of the virtual camera, i.e., smaller angle provides more views. In this experiment, we test how the number of input views affects the performance of our method. Table 4 presents the performance of our method on the general-purpose database using a different number of views.

It is shown in Table 4 that the best performance is obtained using the angle $\frac{\pi}{6}$. Smaller angle (i.e., $\frac{\pi}{12}$) provides 576 views; although this number seems representative, many views may have the same information. Greater angle (i.e., $\frac{\pi}{4}$) provides 64 views, which is not enough to represent the 3D shape since a lot of information is missed.

4.4 Effect of the saliency-based patch selection

As mentioned earlier, SCNN-BMQA relies on the assumption that the subjective evaluation of the visual quality of a distorted mesh is strongly related to the

Table 1 Performance of the network with respect to the number of kernels

Network configuration	Input: (32×32) Conv1–10 (5×5) Max-pool1 (2×2) Conv2–10 (5×5) Max-pool2 (10×10) FC-500	Input: (32×32) Conv1–32 (5×5) Max-pool1 (2×2) Conv2–32 (5×5) Max-pool2 (10×10) FC-500	Input: (32×32) Conv1–50 (5×5) Max-pool1 (2×2) Conv2–50 (5×5) Max-pool2 (10×10) FC-500
r_s	88.6	**93.3**	92.4
r_p	89.7	**92.2**	91.2

The best correlations are highlighted in bold

Table 2 Performance of the network with respect to the size of convolutional kernels

Network configuration	Input: (32 × 32) Conv1–32 (3 × 3) Max-pool1 (2 × 2) Conv2–32 (3 × 3) Max-pool2 (13 × 13) FC-500	Input: (32 × 32) Conv1–32 (5 × 5) Max-pool1 (2 × 2) Conv2–32 (5 × 5) Max-pool2 (10 × 10) FC-500	Input: (32 × 32) Conv1–32 (7 × 7) Max-pool1 (2 × 2) Conv2–32 (7 × 7) Max-pool2 (7 × 7) FC-500	Input: (32 × 32) Conv1–32 (9 × 9) Max-pool1 (2 × 2) Conv2–32 (9 × 9) Max-pool2 (4 × 4) FC-500
r_s	93.0	**93.3**	93.2	89.4
r_p	**92.4**	92.2	90.4	88.9

The best correlations are highlighted in bold

Table 3 Performance of the network with respect to the input patch size variation

Network configuration	Input: (16 × 16) Conv1–32 (5 × 5) Max-pool1 (2 × 2) Conv2–32 (5 × 5) Max-pool2 (2 × 2) FC-500	Input: (32 × 32) Conv1–32 (5 × 5) Max-pool1 (2 × 2) Conv2–32 (5 × 5) Max-pool2 (10 × 10) FC-500	Input: (64 × 64) Conv1–32 (5 × 5) Max-pool1 (2 × 2) Conv2–32 (5 × 5) Max-pool2 (26 × 26) FC-500	Input: (128 × 128) Conv1–32 (5 × 5) Max-pool1 (2 × 2) Conv2–32 (5 × 5) Max-pool2 (58 × 58) FC-500
r_s	92.7	**93.3**	92.7	89.3
r_p	90.3	**92.2**	91.8	86.8

The best correlations are highlighted in bold

Table 4 Correlation coefficients r_s (%) and r_p (%) of SCNN-BMQA using different number of views on the general-purpose database

Number of views (Rotation angle)	576 ($\frac{\pi}{12}$)		144 ($\frac{\pi}{6}$)		64 ($\frac{\pi}{4}$)	
	r_s	r_p	r_s	r_p	r_s	r_p
Correlation score	90.6	90.1	93.3	92.4	88.6	87.3

distortion applied to salient regions. To demonstrate this, we use a patch selection approach based on mesh visual saliency and only the salient patches are used to feed our CNN. The salient patches are selected by fixing a saliency threshold S_t. To do so, several thresholds are tested on the general-purpose database. We choose to conduct the experiments in this database because it contains the greater number of deformations per object. As shown in Table 5, the performance of SCNN-BMQA is sensitive to the saliency threshold. Starting from $S_t = 0.1$, the use of a greater threshold leads to a better performance until $S_t = 0.4$ that provides the best performance. However, the performance decreases when the threshold value exceeds 0.4. We note that this value is fixed in our experiments as constant for all the other databases.

We also compare the performance of SCNN-BMQA with and without using the patch selection approach. Table 6 presents the correlation coefficients $r_s(\%)$ and $r_p(\%)$ of SCNN-BMQA with and without the patch selection on the four used databases. We note that the given correlation scores are for the whole repository.

Table 5 Correlation coefficients r_s (%) and r_p using different saliency threshold S_t values on the LIRIS/EPFL general-purpose database

Threshold S_t	0.1	0.2	0.4	0.6	0.8	0.9
r_s	89.9	90.8	**93.3**	92.8	88.5	80.5
r_s	91.5	92.1	**92.2**	90.6	86.3	80.2

The best correlations are highlighted in bold

It is shown in Table 6 that the patch selection process improves significantly the correlations scores. Especially on the masking database where the Spearman coefficient increases by 4.1% and the Pearson coefficients increases by 3.5%, and on the general-purpose database where r_s and r_p coefficients increase by 3.3% and 0.9%, respectively. On the UWB compression and simplification databases, the performance is slightly improved except for the Pearson correlation on the compression database which is decreased by 0.1%. Therefore, the used patch selection strategy based on visual saliency is very effective, especially on the LIRIS masking and the general-purpose database.

Table 6 Correlation coefficients r_s (%) and r_p (%) of SCNN-BMQA with and without the patch selection strategy on the four tested databases

	Masking database		General-purpose database		Compression database		Simplification database	
	r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p
Without patch selection	91.4	90.8	90.0	92.0	90.1	88.3	90.4	90.2
With patch selection	95.5	94.3	93.3	92.4	90.4	88.2	90.4	90.5

4.5 Evaluation and comparison with the state of the art

In order to evaluate the performance of our method, a comparative analysis is conducted. SCNN-BMQA is compared to the state of the art including FR, RR and NR methods:

- Full reference methods: HD [10], RMS [9], MSDM2 [25], TPDM [26].
- Reduced reference methods: 3DWPM1 [27], 3DWPM2 [27], FMPD [25], DAME [28].
- No reference methods: NR-SVR [30], NR-GRNN [31], BMQI [36].

The correlation coefficients values r_s and r_p on the LIRIS masking, LIRIS/EPFL general-purpose, UWB compression and the IEETA simplification databases are listed, respectively, in Tables 7, 8, 9 and 10.

As shown in Tables 7, 8, 9 and 10, the geometric measures HD and RMS perform the worst. One reason is that these methods do not include the main operations of the HVS and the visual quality is computed by a simple geometric distance. For the other FR measures, MSDM2 and TPDM incorporate the perceptual information, represented in the mesh curvature. As such, the perceptual information is included and better prediction is achieved compared to the geometric measures as proven by the high correlation coefficients.

The RR method FMPD also provide good correlations compared to MSDM2 and TPDM. This method (FMPD) includes a roughness measure which is an important feature in mesh processing.

SCNN-BMQA shows excellent performance on all the available subjectively-rated MVQ databases, as proven by its high scores on the individual models as well as on the whole repositories. On the LIRIS masking database, SCNN-BMQA provides the highest Spearman and Pearson correlation coefficients on the whole corpus ($r_s = 95.5\%$ and $r_p = 94.3\%$) and overcome the NR methods (BMQI, NR-SVR, and NR-GRNN) as well as the most effective FR and RR methods.

The general-purpose database is the largest MVQ database; so far, it comprises the highest number of distorted meshes among all the other databases. That is, 84 distorted meshes and a variety of distortion types. On this database, SCNN-BMQA shows a good performance and provides the highest correlation coefficients ($r_s = 93.3\%$ and $r_p = 92.4\%$) that contend all the compared methods. The high correlation scores obtained by SCNN-BMQA in the

Table 7 Correlation coefficients r_s (%) and r_p (%) of different objective metrics on the LIRIS masking database

Type	Metric	Armadillo		Lion		Bimba		Dyno		All models	
		r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p
Full reference	HD [10]	48.6	37.7	71.4	25.1	25.7	7.5	48.6	31.1	26.6	4.1
	RMS [9]	65.6	44.6	71.4	23.8	71.4	21.8	71.4	50.3	48.8	17.0
	MSDM2 [25]	81.1	88.6	93.5	94.3	96.8	100	95.6	100	87.3	89.6
	TPDM [26]	91.4	88.6	88.4	82.9	97.1	100	71.1	100	**88.6**	**90.0**
Reduced reference	3DWPM1 [27]	58.0	41.8	20.0	9.7	20.0	8.4	66.7	45.3	29.4	10.2
	3DWPM2 [27]	48.6	37.9	38.3	22.0	37.1	14.4	71.4	50.1	37.4	18.2
	FMPD [25]	94.2	88.6	93.5	94.3	98.9	100	96.9	94.3	**80.8**	**80.2**
	DAME [28]	96.0	94.3	99.5	100	88.0	97.7	89.4	82.9	58.6	68.1
No-reference	NR-SVR [30]	89.5	84.7	100	96.3	94.2	93.6	94.4	89.7	90.4	91.2
	NR-GRNN [31]	82.3	80.5	94.1	97.0	90.2	94.3	78.2	82.3	90.2	82.4
	BMQI [36]	94.3	NA	94.3	NA	100	NA	83.0	NA	92.9	NA
	SCNN-BMQA	92.4	91.7	92.2	93.1	97.9	97.3	93.4	92.6	**95.5**	**94.3**

The best correlations for each evaluation type (no-reference, reduced reference and no reference) are highlighted in bold

Table 8 Correlation coefficients r_s (%) and r_p (%) of different objective metrics on the LIRIS/EPFL general-purpose database

Type	Metric	Armadillo		Dyno		Venus		Rocker		All models	
		r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p
Full reference	HD [10]	69.5	30.2	30.9	22.6	1.6	0.8	18.1	5.5	13.8	1.3
	RMS [9]	62.7	32.3	0.3	0.0	90.1	77.3	7.3	3.0	26.8	7.9
	MSDM2 [25]	81.6	85.3	85.9	85.7	89.3	87.5	89.6	87.2	80.4	81.4
	TPDM [26]	84.5	78.8	92.2	89.0	90.6	91.0	92.2	91.4	**89.6**	**89.2**
Reduced reference	3DWPM1 [27]	65.8	35.7	62.7	35.7	71.6	46.6	87.5	53.2	69.3	38.4
	3DWPM2 [27]	74.1	43.1	52.4	19.9	34.8	16.4	37.8	29.9	49.0	24.6
	FMPD [25]	75.4	83.2	89.6	88.9	87.5	83.9	88.8	84.7	**81.9**	**83.5**
	DAME [28]	60.3	76.3	92.8	88.9	91.0	83.9	85.0	80.1	76.6	75.2
No-reference	NR-SVR [30]	76.8	91.5	78.6	84.1	85.7	88.6	86.2	86.6	81.5	87.8
	NR-GRNN [31]	87.1	97.3	91.2	94.1	86.3	85.0	78.6	74.8	86.2	88.7
	BMQI [36]	20.1	NA	83.5	NA	88.9	NA	92.7	NA	78.1	NA
	SCNN-BMQA	89.8	91.4	91.6	92.2	94.6	93.8	91.9	93.9	**93.3**	**92.4**

The best correlations for each evaluation type (no-reference, reduced reference and no reference) are highlighted in bold

Table 9 Correlation coefficients r_s (%) and r_p (%) of different objective metrics on the UWB compression database

Type	Metric	Bunny		James		Jessy		Nissan		Helix		All models	
		r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p
Full reference	HD [10]	34.1	52.2	− 16.8	6.8	− 23.6	12.5	14.4	23.6	45.1	46.4	10.6	28.3
	RMS [9]	34.2	20.9	14.0	10.8	0.0	14.8	17.8	29.7	46.9	44.6	22.0	24.1
	MSDM2 [25]	97.4	90.1	82.6	69.2	84.3	63.1	84.4	73.1	98.1	94.7	89.3	78.0
	TPDM [26]	95.1	96.5	90.8	73.6	85.8	75.8	82.7	73.4	98.7	95.0	**91.5**	**82.9**
Reduced reference	3DWPM1 [27]	94.7	93.4	77.3	72.3	87.2	89.5	63.6	59.3	98.0	95.2	84.1	81.9
	3DWPM2 [27]	96.0	91.2	76.9	65.3	86.9	85.9	56.3	67.6	95.5	94.3	82.3	80.9
	FMPD [25]	94.2	89.6	95.3	91.2	63.3	60.0	92.4	77.5	98.4	90.8	88.8	81.8
	DAME [28]	96.8	93.4	95.7	93.4	84.4	70.5	93.9	75.3	96.6	95.2	**93.5**	**85.6**
No-reference	SCNN-BMQA	95.8	91.7	96.2	95.6	92.3	90.5	88.7	84.7	96.7	94.6	**90.4**	**88.2**

The best correlations for each evaluation type (no-reference, reduced reference and no reference) are highlighted in bold

general-purpose database prove its strength in MVQ assessment task.

On the UWB compression database, SCNN-BMQA performs the best in terms of PLCC score ($r_p = 88.2\%$) overcoming the most effective methods. In addition, it provides competitive SROCC scores on the whole repository ($r_s = 90.4\%$) against $r_s = 91.5\%$ for TPDM and $r_s = 93.5\%$ for the RR method DAME. We note that the methods NR-SVR, NR-GRNN and BMQI are not evaluated on this database.

On the IEETA simplification database, SCNN-BMQA provides the highest correlation coefficients ($r_s = 90.4\%$ and $r_p = 90.5\%$). The perceptual methods MSDM2, TPDM and FMPD also perform well in this database. The results of 3DWPM1, 3DWPM2 and DAME are missing because these metrics have mesh connectivity constraint and they cannot be applied to compare two meshes with different connectivities. The results of NR-SVR, NR-GRNN and BMQI are also missing since these methods are not evaluated on the IEETA simplification database.

4.6 Psychometric curve fitting

Since the objective scores obtained by an MVQ method and the corresponding subjective scores are nonlinear, it is important to introduce a psychometric fitting function to partially remove the nonlinearity and make the correlation values interpretable by users. In our work, we use the cumulative Gaussian psychometric function [44] adopted by [22, 25] defined as follows:

$$p(a,b,X) = \frac{1}{\sqrt{2\pi}} \int_{a+bX}^{\infty} \exp - \left(\frac{t^2}{2}\right) \mathrm{d}t \tag{9}$$

Table 10 Correlation coefficients r_s (%) and r_p (%) of different objective metrics on the IEETA simplification database

Type	Metric	Bones		Bunny		Head		Lung		Strange		All models	
		r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p
Full reference	HD [10]	92.0	94.3	37.8	39.5	72.8	88.6	80.6	88.6	52.3	37.1	50.5	49.4
	RMS [9]	86.4	94.3	94.5	77.1	49.6	42.9	89.0	100	90.4	88.6	59.6	70.2
	MSDM2 [25]	98.3	94.3	98.1	77.1	88.9	88.6	92.3	60.0	99.0	94.3	**89.2**	86.7
	TPDM [26]	99.0	94.3	98.0	94.3	63.1	65.7	98.6	94.3	98.7	94.3	86.9	**88.2**
Reduced reference	FMPD [25]	96.0	88.6	98.0	94.3	70.4	65.7	95.5	88.6	96.0	65.7	89.3	87.2
No-reference	SCNN-BMQA	96.8	93.7	96.7	90.9	96.4	93.6	94.3	89.6	95.4	93.5	**90.4**	**90.5**

The best correlations for each evaluation type (no-reference, reduced reference and no reference) are highlighted in bold

where X is the quality score obtained by the objective method, a and b are two parameters to be determined. The values of a and b are retrieved using the MOS values and the objective values obtained by SCNN-BMQA for each database. Figure 6 shows the scatter plots of the predicted scores obtained by SCNN-BMQA and the subjective MOSs. As illustrated by this figure, the subjective vs objective scores point cloud is close enough to the psychometric curve with regard to the four tested databases. The good fitting of these plots is another indicator of the good performance of SCNN-BMQA.

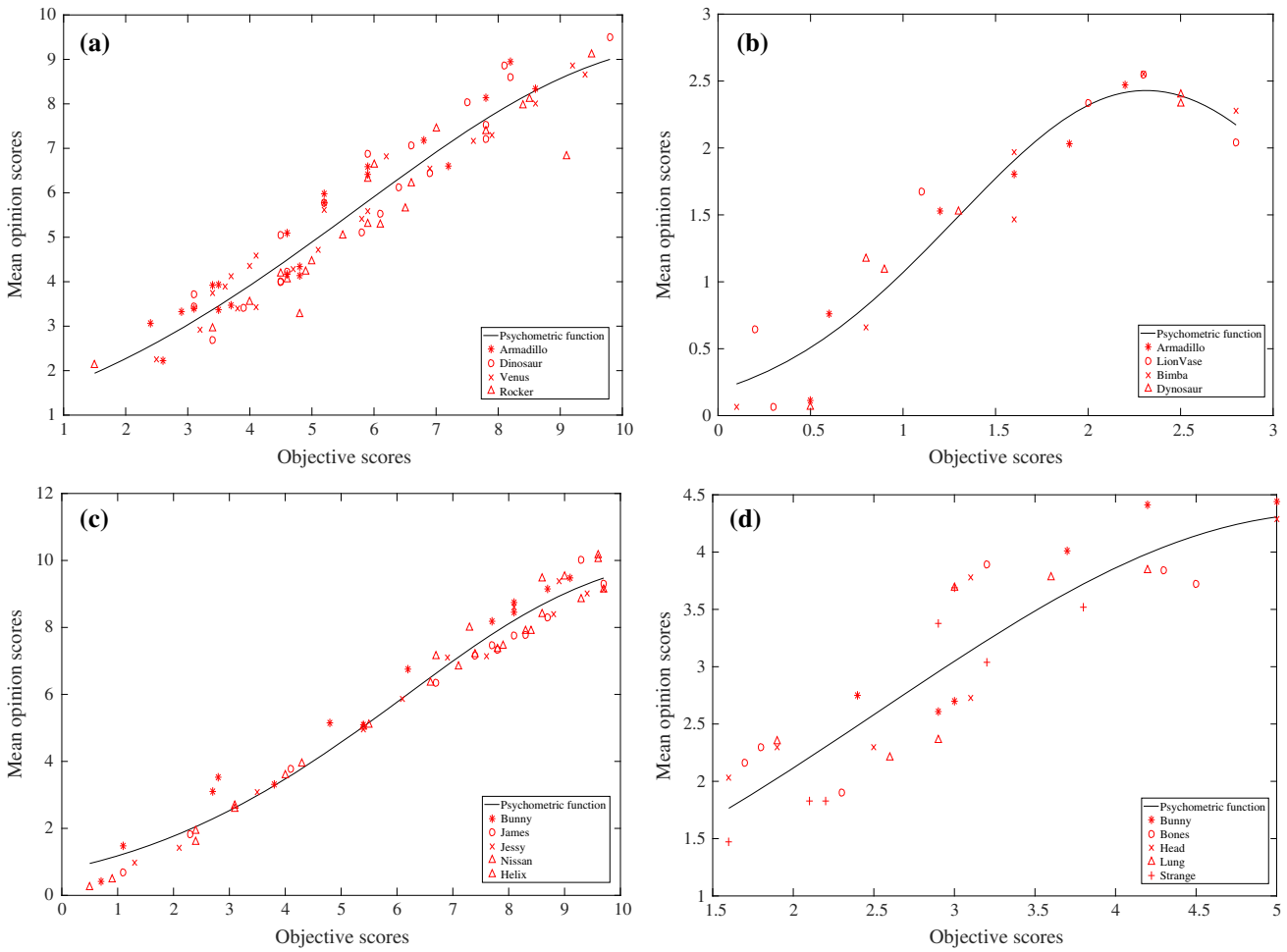

Fig. 6 Scatter plots of the mean opinion scores (MOS) versus the objective scores obtained from SCNN-BMQA. **a** LIRIS/EPFL general-purpose database. **b** LIRIS masking database. **c** The UWB compression database. **d** IEETA simplification database

Table 11 Correlation coefficients r_s (%) and r_p (%) obtained by training on the general-purpose and testing on LIRIS masking, UWB compression and IEETA simplification

Database	Object	Scores		Database	Object	Scores		Database	Object	Scores	
		r_s	r_p			r_s	r_p			r_s	r_p
LIRIS masking	Armadillo	88.6	86.5	UWB compression	Bunny	89.6	88.8	IEETA simplification	Bones	82.6	82.9
	Lion	89.5	88.8		James	86.4	84.7		Bunny	91.1	90.8
	Bimba	94.8	94.3		Jessy	82.9	82.5		Head	80.6	79.6
	Dyno	92.9	91.2		Nissan	90.4	88.7		Lung	82.4	83.8
					Helix	84.1	83.0		Strange	84.2	83.1
	All models	91.3	90.5		All models	83.7	82.8		All models	81.8	81.3

4.7 Cross dataset evaluation

In this section, we investigate the generalization ability of SCNN-BMQA. To do so, we perform a cross database evaluation by training our network on the general-purpose database and using the other databases for the test. We choose this database for the training process because it contains the highest number of distorted models and rich variety of distortion types. Table 11 shows the results of the cross dataset evaluation. This table presents the correlation coefficients of each 3D object in the three tested databases (i.e., LIRIS masking, UWB compression and IEETA simplification) as well as the scores for the whole repositories.

As we can see, our network successfully estimate the perceived visual quality using the cross dataset evaluation as proven by the high correlation coefficients obtained. These results ensure the generalization ability of SCNN-BMQA.

5 Conclusion

In this paper, we propose a no-reference MVQ assessment method to accurately estimate the perceived visual quality of distorted meshes. A CNN architecture is used to learn sets of 2D patches rendered from the 3D mesh. The visual saliency is adopted to select the most relevant regions with high saliency level. SCNN-BMQA successfully predicts the visual quality of distorted meshes as proven by the high correlations with human judgment. In addition, it can be useful in practical situations since it does not require any information about the reference unlike the full reference and reduced reference methods. We have tested many network configurations (the number and the size of kernels, and the size of the input data). It is demonstrated from the experiments that the CNN parameters significantly affect the performance of the network. It is proven also that the used patch selection strategy based on visual saliency is very effective; hence, we can conclude that the distortions in salient regions are more important than in the normal regions, and thus, the saliency information impacts more the overall subjective score.

The current stage of development for the proposed method is focused on using the CNN fed by 2D patches. In order to process directly the 3D mesh, a possible direction of future work would be using a network especially conceived for the MVQ assessment task.

References

1. Botsch M, Kobbelt L, Pauly M, Alliez P, Lévy B (2010) Polygon mesh processing. CRC Press, Boca Raton
2. Alliez P, Gotsman C (2005) Recent advances in compression of 3d meshes. In: Dodgson N, Floater M, Sabin M (eds) Advances in multiresolution for geometric modelling. Springer, Berlin, pp 3–26
3. Lee H, Dikici Ç, Lavoué G, Dupont F (2011) Joint reversible watermarking and progressive compression of 3d meshes. Vis Comput 27(6):781–792
4. Wang K, Lavoué G, Denis F, Baskurt A (2008) A comprehensive survey on three-dimensional mesh watermarking. IEEE Trans Multimed 10(8):1513–1527
5. Wang Y-P, Shi-Min H (2009) A new watermarking method for 3d models based on integral invariants. IEEE Trans Vis Comput Graph 15(2):285–294
6. Garland M, Heckbert PS (1997) Surface simplification using quadric error metrics. In: Proceedings of the 24th annual conference on Computer graphics and interactive techniques. ACM Press/Addison-Wesley Publishing Co., pp 209–216
7. Luebke DP (2001) A developer's survey of polygonal simplification algorithms. IEEE Comput Graph Appl 21(3):24–35
8. Corsini M, Larabi M-C, Lavoué G, Petřík O, Váša L, Wang K (2013) Perceptual metrics for static and dynamic triangle meshes. In: Computer graphics forum, vol 32, issue 1. Blackwell, Oxford, UK, pp 101–125
9. Cignoni P, Rocchini C, Scopigno R (1998) Metro: measuring error on simplified surfaces. In: Computer graphics forum, vol 17, issue 2. Blackwell, Oxford, UK, pp 167–174

10. Aspert N, Santa-Cruz D, Ebrahimi T (2002) Mesh: measuring errors between surfaces using the Hausdorff distance. In: Proceedings. 2002 IEEE international conference on multimedia and expo, 2002. ICME'02, vol 1. IEEE, pp 705–708
11. Lavoué G, Corsini M (2010) A comparison of perceptually-based metrics for objective evaluation of geometry processing. IEEE Trans Multimed 12(7):636–649
12. Bulbul A, Capin T, Lavoué G, Preda M (2011) Assessing visual quality of 3-d polygonal models. IEEE Signal Process Mag 28(6):80–90
13. Lin W, Jay Kuo C-C (2011) Perceptual visual quality metrics: a survey. J Vis Commun Image Represent 22(4):297–312
14. Lin W, Ebrahimi T, Loizou PC, Moller S, Reibman AR (2012) Introduction to the special issue on new subjective and objective methodologies for audio and visual signal processing. IEEE J Sel Top Signal Process 6(6):614–615
15. Karni Z, Gotsman C (2000) Spectral compression of mesh geometry. In: Proceedings of the 27th annual conference on computer graphics and interactive techniques. ACM Press/Addison-Wesley Publishing Co., pp 279–286
16. Sorkine O, Cohen-Or D, Toledo S (2003) High-pass quantization for mesh encoding. In: Symposium on geometry processing, vol 42
17. Pan Y, Cheng I, Basu A (2005) Quality metric for approximating subjective evaluation of 3-d objects. IEEE Trans Multimed 7(2):269–279
18. Bian Z, Shi-Min H, Martin RR (2009) Evaluation for small visual difference between conforming meshes on strain field. J Comput Sci Technol 24(1):65–75
19. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612
20. Lavoué G, Gelasca ED, Dupont F, Baskurt A, Ebrahimi T (2006) Perceptually driven 3d distance metrics with application to watermarking. In: SPIE optics + photonics. International Society for Optics and Photonics, pp 63120L–63120L
21. Lavoué G (2011) A multiscale metric for 3d mesh visual quality assessment. In: Computer graphics forum, vol 30, issue 5. Blackwell, Oxford, UK, pp 1427–1437
22. Torkhani F, Wang K, Chassery J-M (2012) A curvature tensor distance for mesh visual quality assessment. In: Computer vision and graphics, pp 253–263
23. Corsini M, Gelasca ED, Ebrahimi T, Barni M (2007) Watermarked 3-d mesh quality assessment. IEEE Trans Multimed 9(2):247–256
24. Váša L, Rus J (2012) Dihedral angle mesh error: a fast perception correlated distortion measure for fixed connectivity triangle meshes. In: Computer graphics forum, vol 31. Wiley Online Library, pp 1715–1724
25. Wang K, Torkhani F, Montanvert A (2012) A fast roughness-based approach to the assessment of 3d mesh visual quality. Comput Graph 36(7):808–818
26. Abouelaziz I, El Hassouni M, Cherifi H (2016) No-reference 3d mesh quality assessment based on dihedral angles model and support vector regression. In: International conference on image and signal processing. Springer, pp 369–377
27. Abouelaziz I, El Hassouni M, Cherifi H (2016) A curvature based method for blind mesh visual quality assessment using a general regression neural network. In: 2016 12th international conference on signal-image technology & internet-based systems (SITIS). IEEE, pp 793–797
28. Nouri A, Charrier C, Lézoray O (2017) 3d blind mesh quality assessment index. Electron Imaging 2017(20):9–26
29. Moorthy AK, Bovik AC (2010) A two-step framework for constructing blind image quality indices. IEEE Signal Process Lett 17(5):513–516
30. Saad MA, Bovik AC, Charrier C (2010) A dct statistics-based blind image quality index. IEEE Signal Process Lett 17(6):583–586
31. Li C, Bovik AC, Wu X (2011) Blind image quality assessment using a general regression neural network. IEEE Transa Neural Netw 22(5):793–799
32. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436
33. Zhang W, Chenfei Q, Ma L, Guan J, Huang R (2016) Learning structure of stereoscopic image for no-reference quality assessment with convolutional neural network. Pattern Recognit 59:176–187
34. Nouri A, Charrier C, Lézoray O (2016) Full-reference saliency-based 3d mesh quality assessment index. In: 2016 IEEE international conference on image processing (ICIP). IEEE, pp 1007–1011
35. Engelke U, Pepion R, Le Callet P, Zepernick H-J (2010) Linking distortion perception and visual saliency in h. 264/AVC coded video containing packet loss. In: Visual communications and image processing 2010. International Society for optics and photonics, vol 7744, p 774406
36. Lee CH, Varshney A, Jacobs DW (2005) Mesh saliency. ACM Trans Graph: TOG 24:659–666
37. Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Pattern Anal Mach Intell 20(11):1254–1259
38. Mittal A, Moorthy AK, Bovik AC (2012) No-reference image quality assessment in the spatial domain. IEEE Trans Image Process 21(12):4695–4708
39. Nair V, Hinton GE (2010) Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th international conference on machine learning (ICML-10), pp 807–814
40. Kang L, Ye P, Li Y, Doermann D (2014) Convolutional neural networks for no-reference image quality assessment. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1733–1740
41. Lavoué G, Larabi MC, Váša L (2016) On the efficiency of image metrics for evaluating the visual quality of 3d models. IEEE Trans Vis Comput Graph 22(8):1987–1999
42. Silva S, Santos BS, Ferreira C, Madeira J (2009) A perceptual data repository for polygonal meshes. In: Second international conference in visualisation, 2009. VIZ'09. IEEE, pp 207–212
43. Wang Z, Bovik AC (2006) Modern image quality assessment. Synth Lect Image Video Multimed Process 2(1):1–156
44. Engeldrum PG (2001) Psychometric scaling: avoiding the pitfalls and hazards. In: PICS, pp 101–107