

A REDUCED-REFERENCE QUALITY ASSESSMENT METRIC FOR TEXTURED MESH DIGITAL HUMANS

Zicheng Zhang, Yingjie Zhou, Chunyi Li, Kang Fu, Wei Sun,
Xiaohong Liu, Xiongkuo Min[†], and Guangtao Zhai[†]

Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University, China

ABSTRACT

In an era where 3D Digital Humans (DHs) are becoming increasingly prevalent in fields like gaming, automotive, and the metaverse, the demand for high DH visual quality is rising. This paper presents the first-ever reduced-reference (RR) quality assessment metric tailored specifically for textured mesh DHs, aiming to optimize transmission systems and improve Quality of Experience (QoE) for viewers in resource-constrained environments. Four critical geometric curvature-related attributes and two texture-related indicators are computed, which are then statistically analyzed and utilized in a Support Vector Regression (SVR) model for robust and efficient quality prediction. Experimental results confirm that our method outperforms existing full-reference (FR) metrics, making it an invaluable tool for the future of 3D DHs in various applications.

Index Terms— Quality assessment, 3D digital humans, textured mesh, reduced-reference

1. INTRODUCTION

Digital humans (DHs) are computer-simulated people, which are widely used in gaming, automotive, and the metaverse. However, the high storage and bandwidth requirements necessitate compression/simplification, often at the cost of visual quality [1]. Therefore, in this paper, we introduce the first reduced-reference (RR) quality assessment metric specifically designed for textured mesh DHs, which can provide useful guidelines for the compression/simplification algorithms and improve Quality of Experience (QoE) for the viewers as shown in Fig. 1.

The 3D digital human quality assessment (DHQA) is intrinsically linked to the 3D model quality assessment (3DQA). However, the content of DHs is more complicated than common 3D models, and people tend to be more sensitive to perceptual distortions. Previous 3DQA methods extract features either from the projections of the 3D models (projection-based) [2, 3, 4, 5, 6] or directly from the 3D models themselves (model-based) [7, 8, 9, 10, 11]. Projection-based methods transfer the problem to an image

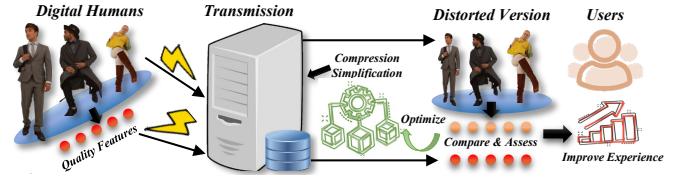


Fig. 1. Application of the proposed RR DHQA method. The reference quality features are extracted and transmitted along with the DHs. Then the distorted quality features are compared with the reference quality features for evaluation, which can help optimize the compression/simplification settings and improve the experience of the users.

quality assessment (IQA) task but are constrained by rendering configurations and viewpoint selection [2]. On the other hand, extracting features directly from the 3D model is more stable but also increases computational complexity [10]. Given the limited prevalence of rendering modules in transmission systems, extracting features directly from 3D models is more universally applicable. While full-reference (FR) methods provide a thorough comparison of distorted and reference models, transmitting these reference models in real-time systems leads to significant overhead. Additionally, deep neural networks are expensive for high-quality DHs with vast faces and vertices. Thus it is more cost-effective to prioritize the abstraction and transmission of quality features from the reference model [12] with handcrafted descriptors. Hence, our model-based RR DHQA metric aims to strike a balance between efficiency and perceptual loss capture.

Specifically, our method rigorously assesses the quality of DHs by extracting quality-aware features from both the 3D geometric mesh and the 2D texture map. In terms of geometry, the proposed method calculates four key curvature-related feature attributes to capture the geometry quality patterns. For the texture aspect, quality indicators are extracted from the gray-scale illumination map and the gradient magnitude map. These features are quantitatively evaluated using a variety of statistical parameters and subsequently employed in a Support Vector Regression (SVR) model for quality prediction. The experimental results show that our method outperforms all the compared FR metrics on the two selected DHQA databases.

2. PROPOSED METHOD

Since the 3D DHs are represented in textured meshes, we propose to extract the quality-aware features from the 3D geometry mesh and 2D texture separately. Given a textured mesh DH, we define the 3D mesh \mathcal{M} as:

$$\mathcal{M} = \{\mathcal{V}, \mathcal{E}, \mathcal{F}\}, \quad (1)$$

where \mathcal{V} , \mathcal{E} , and \mathcal{F} represent the set of vertices, edges, and faces respectively. Then the 2D texture is simply defined as \mathcal{T} , which is a common RGB image. The framework of the proposed method is shown in Fig. 2.

2.1. Mesh Feature Extraction

The mesh representations serve as the foundational geometric structures for DHs, encompassing facial features, limbs, and the torso. Unlike conventional 3D objects encountered in everyday life, the 3D representation of a digital human entails a higher level of complexity [1]. In particular, curvature attributes often serve as critical indicators of geometric complexity. Therefore, we propose mapping the mesh geometry into the curvature domains for feature extraction. To ensure diversity and comprehensiveness, we initially focus on two primary curvature-related attributes: **Vertex Defect** and **Dihedral Angle**. Subsequently, we construct a local spherical subspace within the mesh and compute the **Discrete Gaussian Curvature** and **Discrete Mean Curvature** within this subspace [13], which can provide an enhanced representation of local curvature patterns and furnish a more complete picture of the geometric features intrinsic to DHs.

1) **Vertex Defect**: The vertex defect $\phi^D(v)$ can be computed as 2π minus the sum of the angles incident to the corresponding vertex v :

$$\phi^D(v) = 2\pi - \sum \theta_v, \quad (2)$$

where $\sum \theta_v$ denotes the sum of the angles incident to vertex v . Each vertex in the mesh has a defect value and we can get the set of the vertex defects as $\tilde{\phi}_V^D$.

2) **Dihedral Angle**: The dihedral angle $\phi^A(e)$ between two adjacent faces f_1 and f_2 sharing an edge e can be computed using their respective normals \mathbf{N}_1 and \mathbf{N}_2 :

$$\phi^A(e) = \arccos \left(\frac{\mathbf{N}_1 \cdot \mathbf{N}_2}{\|\mathbf{N}_1\| \|\mathbf{N}_2\|} \right), \quad (3)$$

where each edge in the mesh has a dihedral angle value and we can obtain the set of the dihedral angles as $\tilde{\phi}_E^A$.

3) **Discrete Gaussian Curvature**: The discrete Gaussian curvature $\phi^G(v)$ at the vertex v can be approximated as the sum of the neighboring vertex defects:

$$\phi^G(v) = \sum_{\tilde{v} \in \mathcal{S} \cap \mathcal{V}} \phi^D(\tilde{v}), \quad (4)$$

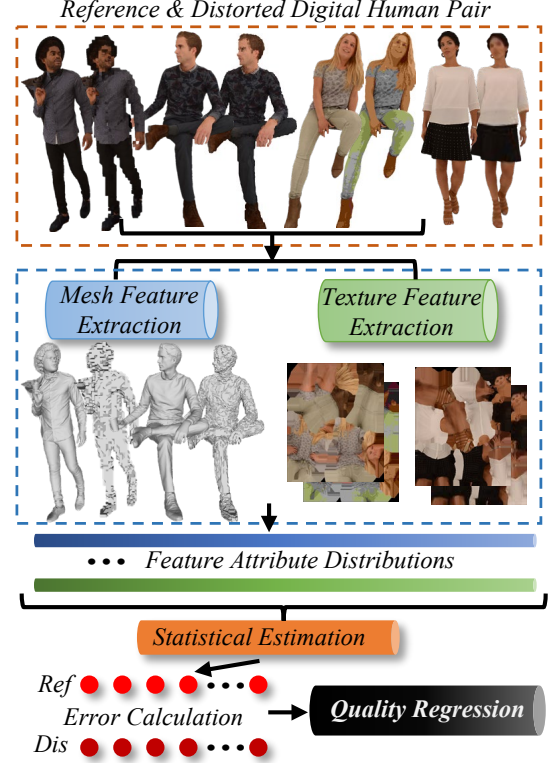


Fig. 2. The framework of the proposed RR DHQA method. The geometry and the texture attribute distributions are extracted from the mesh and texture map respectively. Afterward, the statistical parameters are estimated from the distributions. The computed error between the reference and distorted parameters is then fed into the quality regression module.

where \mathcal{S} is a spherical region centered at the vertex v with a radius of $r = c_1 * BB$ (c_1 is a constant and BB is the longest side of the bounding box), \tilde{v} stands for the neighboring vertices that belongs to the intersection of \mathcal{S} and \mathcal{V} . Therefore, each vertex has a discrete Gaussian curvature value and we can get the set of the discrete Gaussian curvature as $\tilde{\phi}_V^G$.

4) **Discrete Mean Curvature**: The discrete mean curvature $\phi^M(v)$ at the vertex v can be obtained as a weighted sum of the neighboring dihedral angles:

$$\phi^M(v) = \sum_{e \in \mathcal{E}} |e \cap \mathcal{S}| \cdot \text{sgn}(\phi^A(e)), \quad (5)$$

where $|e \cap \mathcal{S}|$ indicates the length of intersection between the edge e and the spherical region \mathcal{S} as described above (if e has no intersection with \mathcal{S} , the length is zero), the sign $\text{sgn}(\cdot)$ of the neighboring $\phi^A(e)$ is positive if e is convex and negative if it is concave. Thus each vertex has a discrete mean curvature value and we can get the set of the discrete mean curvature as $\tilde{\phi}_V^M$. As shown in Fig. 3, it is clear that distortions significantly impact the patterns of curvature.

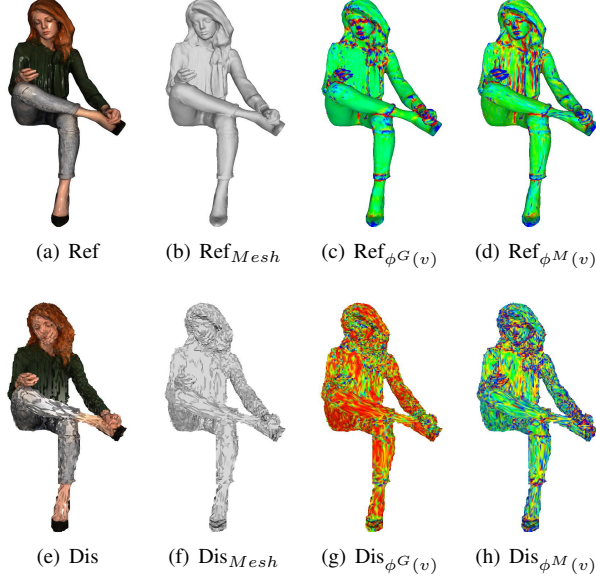


Fig. 3. Visualization results of discrete Gaussian/mean curvature. (a), (b), (e), and (f) represent the reference/compression-distorted textured mesh DH examples and the corresponding geometry meshes. (c), (d), (g), and (h) denote the reference discrete Gaussian/mean curvature and the distorted discrete Gaussian/mean curvature projections respectively, where the red/green colors signify relatively greater/smaller curvature values. It is obvious that compression notably alters the DH geometry patterns and the corresponding visual quality.

2.2. Texture Feature Extraction

Textures can take many forms, from simple color patterns to complex, high-resolution images. They are used for various purposes, including simulating complex surface properties like fur, grass, opacity, or even subsurface scattering. In practical applications, the DH textures can be distorted with sub-sampling, compression, and noise. The texture feature extraction can be recognized as a similar task to image quality feature extraction.

Therefore, we simply use the gray-scale illumination map and gradient magnitude map of the texture for feature extraction. Given the texture \mathcal{T} , the gray-scale illumination map $\Phi_{\mathcal{T}}^L$ can be derived as the maximum of RGB channels:

$$\Phi_{\mathcal{T}}^L = \max_{c \in \{r, g, b\}} \mathcal{T}(i, j), \quad (6)$$

where i and j are pixel indexes of the texture. The gradient information has been proven to be very effective in many IQA methods [14]. A Sobel gradient operator is applied to calculate the gradient maps. Then the gradient magnitude map $\Phi_{\mathcal{T}}^G$ can be computed as follow:

$$\Phi_{\mathcal{T}}^G = \sqrt{(\Phi_{\mathcal{T}}^L \otimes S_x)^2 + (\Phi_{\mathcal{T}}^L \otimes S_y)^2}, \quad (7)$$

where the symbol \otimes denotes the convolution operation, and S_x and S_y are the horizontal and vertical Sobel filters.

Table 1. Overview of the employed feature domains.

Group	Symbol	Description
Mesh	$\tilde{\phi}_V^D$	Vertex Defect
	$\tilde{\phi}_E^A$	Dihedral Angle
	$\tilde{\phi}_V^G$	Discrete Gaussian Curvature
	$\tilde{\phi}_V^M$	Discrete Mean Curvature
Texture	$\Phi_{\mathcal{T}}^L$	Gray-scale Illumination Map
	$\Phi_{\mathcal{T}}^G$	Gradient Magnitude Map

2.3. Statistical Estimation

It has been proved in many works that different types of distortions can corrupt the appearance of geometry and texture attribute distributions [19, 10]. Therefore, we employ several basic statistical parameters (mean, variance, and entropy) and classical distributions (generalized Gaussian distribution (GGD) and the general asymmetric generalized Gaussian distribution (AGGD) [20]) to quantify the perceptual loss from the feature domains described above. A summary of the feature domains is illustrated in Table 1. The estimation process can be described as:

$$\begin{aligned} \mathcal{D} &\sim Basic(\mu, \sigma^2, E), \\ \hat{\mathcal{D}} &\sim GGD(\alpha_1, \beta_1), \\ \hat{\mathcal{D}} &\sim AGGD(\alpha_2, \beta_2, \sigma_l^2, \sigma_r^2), \\ \mathcal{D} &\in \{\tilde{\phi}_V^D, \tilde{\phi}_E^A, \tilde{\phi}_V^G, \tilde{\phi}_V^M, \Phi_{\mathcal{T}}^L, \Phi_{\mathcal{T}}^G\}, \end{aligned} \quad (8)$$

where \mathcal{D} and $\hat{\mathcal{D}}$ represent the original and normalized feature distributions, (μ, σ^2, E) parameters in the $Basic(\cdot)$ function stand for (mean, variance, entropy), (α_1, β_1) indicate the GGD estimated features, and $(\alpha_2, \beta_2, \sigma_l^2, \sigma_r^2)$ stand for the AGGD estimated features.

2.4. Quality Regression

Finally, we can get a quality-aware feature vector for a single textured mesh DH, which consists of $54=6 \times 9$ features in total. The mapped error between the features extracted from the reference and distorted DHs are taken as the input features F_{in} :

$$F_{in} = \frac{1}{1 + e^{-(F_R - F_D)}}, \quad (9)$$

where F_R and F_D indicate the features extracted from the reference and distorted DHs respectively, and the sigmoid function is used to bind the values of the feature error. Furthermore, we utilize the support vector regression (SVR) as the regression model. The SVR is calibrated with the extracted features and quality labels and then validated for performance exhibition.

Table 2. Benchmark Performance on the SJTU-H3D and DHHQA databases. Best in **RED** and second in **BLUE**.

Ref	Type	Method	SJTU-H3D				DHHQA			
			SRCC \uparrow	PLCC \uparrow	KRCC \uparrow	RMSE \downarrow	SRCC \uparrow	PLCC \uparrow	KRCC \uparrow	RMSE \downarrow
FR	Projection-based	PSNR	0.5139	0.4974	0.3170	0.9441	0.8347	0.8371	0.6405	11.5822
		SSIM [15]	0.7336	0.6888	0.5416	0.8626	0.7355	0.7253	0.5388	14.5221
		MS-SSIM [16]	0.2417	0.2776	0.1822	1.0150	0.8557	0.8396	0.6653	11.4953
		GMSD [17]	0.2574	0.3538	0.1855	0.9833	0.8411	0.8350	0.6534	11.6441
		G-LPIPS [3]	0.6930	0.6112	0.5343	0.7966	0.8389	0.8069	0.6446	12.0051
	Model-based	PSNR $_{p2po}$ [7]	0.2636	0.2680	0.2154	1.0134	0.2891	0.2916	0.2359	21.0813
		PSNR $_{p2pl}$ [8]	0.2101	0.2114	0.1686	1.0244	0.2698	0.2961	0.2250	21.0520
		PSNR $_{yuv}$ [18]	0.5247	0.5638	0.4141	0.9199	0.1761	0.2272	0.1369	21.4299
RR	Model-based	Proposed	0.8420	0.8486	0.6612	0.5422	0.8796	0.8862	0.6969	9.9143

Table 3. Ablation study results on the SJTU-H3D and DHHQA databases.

Feature Group	SJTU-H3D		DHHQA	
	SRCC \uparrow	PLCC \uparrow	SRCC \uparrow	PLCC \uparrow
G1	0.7984	0.8084	0.8485	0.8470
G2	0.8312	0.8355	0.6548	0.6719
G3	0.6598	0.6901	0.6546	0.7480
G4	0.7973	0.8128	0.8274	0.8429
G5	0.8257	0.8326	0.8405	0.8326
All	0.8420	0.8486	0.8796	0.8862

3. EXPERIMENT

3.1. Databases & Experimental Setup

The publicly available full-body DH quality assessment (SJTU-H3D) database [21] and the DH head quality assessment (DHHQA) database [22] are selected for validation. The SJTU-H3D database contains 1,120 distorted full-body textured mesh DHs while the DHHQA database contains 1,540 distorted textured mesh DH heads. For both databases, we chose a k value of 5 for conducting k -fold cross-validation, aiming for a balanced assessment over various subsets. It is important to note that the training and testing folds do not share any overlapping content.

The Support Vector Regression (SVR) model with RBF kernel is implemented with the Python scikit-learn package. The popular Spearman Rank Correlation Coefficient (SRCC), Pearson Linear Correlation Coefficient (PLCC), Kendall’s Rank Order Correlation Coefficient (KRCC), and Root Mean Squared Error (RMSE) are utilized as evaluation criteria.

3.2. Competitors

The competitors’ selection is conducted to ensure high diversity, which includes both projection-based and model-based methods. The former includes: PSNR, SSIM [15], MS-SSIM [16], GMSD [17], and G-LPIPS [3]. These methods are applied to the six cube-like projections of the DHs, and the averaged scores are recorded. The latter includes: PSNR $_{p2po}$ [7], PSNR $_{p2pl}$ [8], and PSNR $_{yuv}$ [18]. These methods are developed by the MPEG group and we convert the textured meshes into point clouds for validation.

3.3. Performance Discussion

The overall experimental performance is exhibited in Table 2, from which we can make several useful conclusions. 1) The proposed RR DHQA method achieves the best performance among all the FR competitors, which confirms the effectiveness of the proposed method for predicting the visual quality of DHs. 2) Benefiting from the mature and advanced development of IQA, the projection-based competitors seem to achieve relatively better performance than the model-based competitors. However, the projection-based FR methods except SSIM all experience significant performance drops from the DHHQA to SJTU-H3D database, which indicates the projection-based methods are less robust.

3.4. Ablation Study

To better understand the impact of different feature sets, we conduct a comprehensive ablation study in this section. We define the groups of features as follows: G1 \leftrightarrow w/o vertex defect & dihedral angle features, G2 \leftrightarrow w/o discrete Gaussian & mean curvature features, G3 \leftrightarrow w/o texture features, G4 \leftrightarrow w/o basic statistical parameters, and G5 \leftrightarrow w/o GGD & AGGD statistical parameters. Our observations indicate that the removal of any feature group results in a decrease in performance, underscoring the importance of each feature set in the overall result. Upon closer examination, it becomes apparent that G3 achieves the lowest performance in terms of SRCC, which suggests that texture features play a particularly significant role in contributing to the model’s performance.

4. CONCLUSION

This paper introduces a pioneering reduced-reference (RR) quality assessment metric specifically crafted for 3D Digital Humans. Through the careful extraction and statistical analysis of critical geometric and texture-related features, the proposed metric offers an efficient yet robust means for quality prediction. Not only does it outperform existing full-reference methods, but it also holds the potential to significantly enhance the Quality of Experience (QoE) for viewers. This makes it an indispensable asset for advancing the visual quality of 3D Digital Humans across diverse applications.

5. REFERENCES

- [1] Wenmin Zhu, Xiumin Fan, and Yanxin Zhang, “Applications and research trends of digital human models in the manufacturing industry,” *Elsevier VRIH*, 2019.
- [2] Qi Yang, Hao Chen, Zhan Ma, Yiling Xu, Rongjun Tang, and Jun Sun, “Predicting the perceptual quality of point cloud: A 3d-to-2d projection-based exploration,” *IEEE TMM*, 2020.
- [3] Yana Nehmé, Johanna Delanoy, Florent Dupont, Jean-Philippe Farrugia, Patrick Le Callet, and Guillaume Lavoué, “Textured mesh quality assessment: Large-scale dataset and deep learning-based quality metric,” *ACM TOG*, 2022.
- [4] Qi Liu, Hui Yuan, Honglei Su, Hao Liu, Yu Wang, Huan Yang, and Junhui Hou, “Pqa-net: Deep no reference point cloud quality assessment via multi-view projection,” *IEEE TCSVT*, 2021.
- [5] Qi Liu, Honglei Su, Zhengfang Duanmu, Wentao Liu, and Zhou Wang, “Perceptual quality assessment of colored 3d point clouds,” *IEEE TVCG*, 2022.
- [6] Wei Zhou, Guanghui Yue, Ruizeng Zhang, Yipeng Qin, and Hantao Liu, “Reduced-reference quality assessment of point clouds via content-oriented saliency projection,” *IEEE SPL*, vol. 30, pp. 354–358, 2023.
- [7] Dong Tian, Hideaki Ochimizu, Chen Feng, R Cohen, and Anthony Vetro, “Evaluation metrics for point cloud compression,” *ISO/IEC JTC m74008, Geneva, Switzerland*, vol. 1, no. 3, 2017.
- [8] Dong Tian, Hideaki Ochimizu, Chen Feng, Robert Cohen, and Anthony Vetro, “Updates and integration of evaluation metric software for pcc,” *ISO/IEC JTC1/SC29/WG11 input document MPEG2017 M*, vol. 40522, pp. 26, 2017.
- [9] Qi Yang, Zhan Ma, Yiling Xu, Zhu Li, and Jun Sun, “Inferring point cloud quality via graph similarity,” *IEEE TPAMI*, 2020.
- [10] Zicheng Zhang, Wei Sun, Xiongkuo Min, Tao Wang, Wei Lu, and Guangtao Zhai, “No-reference quality assessment for 3d colored point cloud and mesh models,” *IEEE TCSVT*, 2022.
- [11] Wei Zhou, Qi Yang, Qiuping Jiang, Guangtao Zhai, and Weisi Lin, “Blind quality assessment of 3d dense point clouds with structure guided resampling,” *arXiv preprint arXiv:2208.14603*, 2022.
- [12] Qi Liu, Hui Yuan, Raouf Hamzaoui, Honglei Su, Junhui Hou, and Huan Yang, “Reduced reference perceptual quality model with application to rate control for video-based point cloud compression,” *IEEE TIP*, 2021.
- [13] David Cohen-Steiner and Jean-Marie Morvan, “Restricted delaunay triangulations and normal cycle,” in *SoCG*, 2003.
- [14] Zicheng Zhang, Wei Sun, Xiongkuo Min, Wenhan Zhu, Tao Wang, Wei Lu, and Guangtao Zhai, “A no-reference evaluation metric for low-light image enhancement,” in *IEEE ICME*, 2021.
- [15] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE TIP*, vol. 13, no. 4, pp. 600–612, 2004.
- [16] Zhou Wang, Eero P Simoncelli, and Alan C Bovik, “Multiscale structural similarity for image quality assessment,” in *ACSSC*, 2003.
- [17] Wufeng Xue, Lei Zhang, Xuanqin Mou, and Alan C Bovik, “Gradient magnitude similarity deviation: A highly efficient perceptual image quality index,” *IEEE TIP*, 2013.
- [18] Rafael Mekuria, Kees Blom, and Pablo Cesar, “Design, implementation, and evaluation of a point cloud codec for tele-immersive video,” *IEEE TCSVT*, 2016.
- [19] Ilyass Abouelaziz, Mohammed El Hassouni, and Hocine Cherifi, “No-reference 3d mesh quality assessment based on dihedral angles model and support vector regression,” in *Springer Image and Signal Processing*, 2016, pp. 369–377.
- [20] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik, “No-reference image quality assessment in the spatial domain,” *IEEE TIP*, 2012.
- [21] Zicheng Zhang, Wei Sun, Yingjie Zhou, Haoning Wu, Chunyi Li, Xiongkuo Min, Xiaohong Liu, Guangtao Zhai, and Weisi Lin, “Advancing zero-shot digital human quality assessment through text-prompted evaluation,” *arXiv preprint arXiv:2307.02808*, 2023.
- [22] Zicheng Zhang, Yingjie Zhou, Wei Sun, Xiongkuo Min, Yuzhe Wu, and Guangtao Zhai, “Perceptual quality assessment for digital human heads,” in *IEEE ICASSP*, 2023.