Spectrum AUC Difference (SAUCD): Human-aligned 3D Shape Evaluation

Tianyu Luan[1,2*] Zhong Li[2] Lele Chen[2] Xuan Gong[3,2,1]
Lichang Chen[2,4] Yi Xu[2] Junsong Yuan[1]

[1]State University of New York at Buffalo [2]OPPO US Research Center
[3] Harvard Medical School [4]University of Maryland, College Park

Abstract

Existing 3D mesh shape evaluation metrics mainly focus on the overall shape but are usually less sensitive to local details. This makes them inconsistent with human evaluation, as human perception cares about both overall and detailed shape. In this paper, we propose an analytic metric named Spectrum Area Under the Curve Difference (SAUCD) that demonstrates better consistency with human evaluation. To compare the difference between two shapes, we first transform the 3D mesh to the spectrum domain using the discrete Laplace-Beltrami operator and Fourier transform. Then, we calculate the Area Under the Curve (AUC) difference between the two spectrums, so that each frequency band that captures either the overall or detailed shape is equitably considered. Taking human sensitivity across frequency bands into account, we further extend our metric by learning suitable weights for each frequency band which better aligns with human perception. To measure the performance of SAUCD, we build a 3D mesh evaluation dataset called Shape Grading, along with manual annotations from more than 800 subjects. By measuring the correlation between our metric and human evaluation, we demonstrate that SAUCD is well aligned with human evaluation, and outperforms previous 3D mesh metrics. Our project page: `https://bit.ly/saucd`.

1. Introduction

With the recent progress of 3D reconstruction and processing techniques, 3D mesh shapes have increasing applications in fields such as video games, industrial design, 3D printing, etc. In these applications, assessing the visual quality of the 3D mesh shape is a crucial task. To meet the requirements of various applications, a promising evaluation metric should not only reflect the geometry measurement but also align with human visual perception. Considering that human beings perceive 3D meshes in both overall shape and local details, it is a challenging task to find an evaluation metric that can align well with humans.

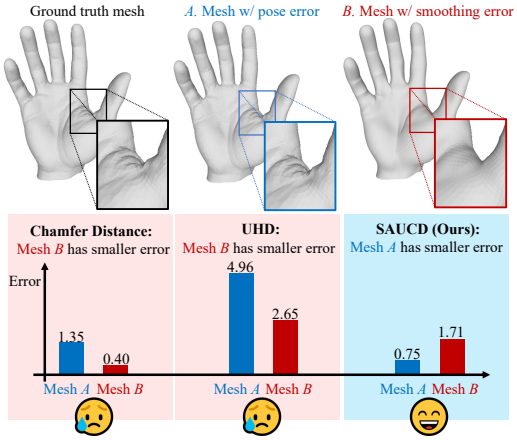

Figure 1. An example of how previous spatial domain 3D shape metrics (Chamfer Distance [6] and UHD [47]) deviate from human evaluation. we create Mesh A by adding a small pose error to the ground truth mesh, and by applying a large smoothing kernel to ground truth, we create Mesh B. Contrary to human perception, previous spatial domain metrics evaluate Mesh B better than Mesh A. This indicates that while they are sensitive to general shape differences, they tend to overlook high-frequency details. Note that different metrics use different units of measurement.

Previous metrics have the following disadvantages in this scenario. Traditional spatial domain measurements such as Chamfer Distance [6] which calculates the mean distance between a vertex on one mesh and its nearest vertex on the other mesh, can accurately measure the spatial distance. However, it does not guarantee capturing all shape details. In fact, such measurements in the spatial domain often overlook finer shape details, as the details tend to get overwhelmed by the overall shape. Fig. 1 illustrates the discrepancy between spatial measurements and human evaluation as mesh details change. Specifically, When we remove the wrinkles from the ground truth mesh (resulting in Mesh B), the errors detected by previous metrics are not as significant as when we slightly

*Work done while Tianyu was an intern at OPPO US Research Center.

change the pose of the hand (Mesh *A*). However, humans tend to sense a significant difference between ground truth and Mesh *B*, but barely recognize the difference between ground truth and Mesh *A*. To mitigate this problem, previous works propose learning-based approaches, such as Single Shape Fréchet Inception Distance (SSFID) [46] based on learnable features from 3D shape. They compare the difference between the test mesh and the ground truth mesh in the latent feature space, and the design is expected to better align human perception. However, such learning-based methods would require a large amount of data to train the network. Their accuracy and generalizability are limited by the size of the dataset, data distribution, and annotation quality, not to mention the potential bias in collecting human perception feedback, which could mislead the learned metrics. An analytic metric that can better explain the shape difference is thus preferred.

To address the above limitations, we design an analytic-based 3D shape evaluation metric named Spectrum Area Under the Curve Difference (SAUCD). Our metric measures mesh shape differences with a balanced consideration of both overall and detailed shape, making it better aligned with human evaluation. To allow our metric to capture detail variations, we leverage the 3D shape spectrum to decompose different levels of shape details from the overall shape, with details corresponding to higher-frequency components. The advantage of transforming the shape signal into the spectrum domain is that the high-frequency details are explicitly separated from the low-frequency overall shape. Therefore, it provides appropriate consideration to the information in different frequency bands, not just the low-frequency information of the overall shape in the dominant place. Thus, the details that human perception cares about will be better represented. Besides, the frequency analysis method allows the metric to be mostly analytic and better explained.

We design SAUCD following the above inspiration. To begin with, both the test mesh and the ground truth mesh are transformed from the spatial to the spectrum domain using the discrete Laplace-Beltrami operator (DLBO), which encodes the mesh geometry information into a semidefinite Laplacian matrix. Once in the spectrum domain, we compare the regions under the two spectrums. Our Spectrum Area Under the Curve Difference metric is defined as the area of the non-overlapping region under the two spectrums – a larger area indicates a greater difference. Moreover, to better align with human evaluation, we further extend our design by learning a spectrum weight for SAUCD. However, different from previous learning-based approaches that use deep networks, large datasets, and extensive learning processes, our learning-based method requires the training of a weight vector. This vector measures the sensitivity of human perception across frequency bands, making the learned metric better aligned with human perception. We then evaluated the effectiveness of SAUCD on our provided user study benchmark dataset named *Shape Grading*. Using *Shape Grading*, we compare our metrics with previous metrics by calculating the correlation between each metric and human scoring. In summary, our contributions are listed as follows.

- We design an analytic-based 3D mesh shape metric named Spectrum AUC Difference (SAUCD), which evaluates the difference between a 3D mesh and its ground truth mesh. Our metric considers both the overall shape and intricate details, to align more closely with human perception.
- We further extend our design to a learnable metric. The extended metric explores the human perception sensitivity in different frequency bands, which further improves this metric.
- We build a user study benchmark dataset named *Shape Grading* which is annotated by more than 800 subjects. The provided dataset verifies that both versions of our metrics are consistent with human evaluation and outperform previous methods. This dataset can also facilitate 3D mesh metric evaluation in future research.

Our experiments show that both SAUCD and its extended version outperform previous methods with good generalizability to different types of objects.

2. Related Works

Metrics in 3D mesh reconstruction. Chamfer Distance [6] is a popular metric used in 3D mesh reconstruction tasks such as those in [21, 24, 32, 35, 45, 48–50]. Other spatial domain metrics, such as 3D Intersection over Union (IoU) in [10, 16, 17, 27, 33, 39]. F-score in [4, 15, 40, 43], and Unidirectional Hausdorff distance (UHD) in [47] are commonly focused on the geometry accuracy of mesh shapes. These metrics can provide accurate geometry measurements, but they are not designed to align with human evaluation. Deep-learning-based methods such as Single Shape Fréchet Inception Distance [46] are also used in 3D reconstruction. While these metrics have the capacity to adapt from human evaluation, they are more like black boxes, with performances subject to dataset size and annotation bias. Moreover, most previous works miss out on user study validation to verify if their metrics align with human evaluation.

3D shape generation metrics. Multiple metrics have been used in 3D shape generation, such as Minimal Matching Distance (MMD) [3], Jensen-Shannon Divergence (JSD) [22], Total Mutual Difference (TMD) [47], Fréchet Pointcloud Distance (FPD) [36], *etc*.. These metrics are designed to measure the differences between the generated distributions, while our task is to build a metric to compare the shape of two meshes.

3D mesh compression and watermarking metrics. Previous works [8, 12, 23, 41] focused on evaluating mesh errors in mesh compression and watermarking. Since compression and watermarking pursue mesh errors that cannot

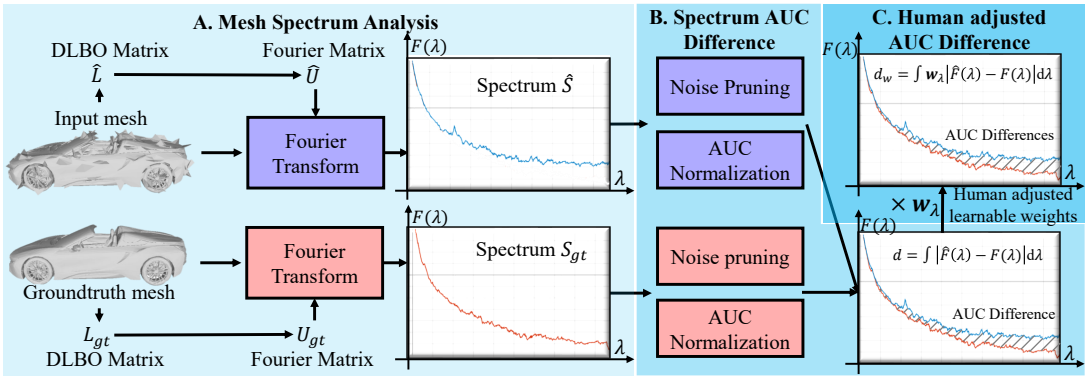

Figure 2. Our SAUCD metric is designed as follows: *A.* We use mesh Fourier Transform to analyze the spectrums of test and ground truth mesh. *B.* We compare the difference between two spectrum curves by calculating the Area Under the Curve (AUC) difference. *C.* We further extend our metric by multiplying the AUC difference with a learnable weight to capture human sensitivity in each frequency band.

be detected by humans, they mainly focus on barely noticed errors. However, our task is to build a metric that can handle generally occurring errors that happen in 3D reconstruction tasks and applications.

3. Proposed Method

Our task is to design a metric aligned with human evaluation to measure the shape difference between a test triangle mesh and its corresponding ground truth triangle mesh. Specifically, given a test mesh $\hat{M}$ and its ground truth mesh M_{gt}, Spectrum AUC Difference (SAUCD) can be abstracted as

$$d = D(\hat{M}, M_{gt}). \tag{1}$$

d is the distance between the test and the ground truth mesh. In this section, we will elaborate on how the distance function $D(\cdot)$ is designed.

3.1. Overview

As shown in Fig. 2, our metric is calculated via the following steps: First, we use mesh Fourier transform to analyze the spectrums of the test and ground truth mesh (in Sec. 3.2). Then we leverage each frequency band by calculating the Area Under the Curve (AUC) difference of the spectrum curves (in Sec. 3.3). Moreover, we further extended our metric by multiplying the AUC difference with a learnable weight to capture the human sensitivity on each frequency band (in Sec. 3.4). We will discuss each step in detail.

3.2. Mesh spectrum analysis

In order to capture the overall shape as well as shape details, we choose to decompose the mesh signal into a spectrum. Considering the mesh as a function on a discretized manifold space, we can calculate the spectrum using the manifold space Fourier transform. In Hilbert space, the Fourier operator is defined as the eigenfunctions of the Laplacian operator [14]. The same definition and similar theories are extended to continuous and discrete manifold space by [5, 9]. The Laplacian operator on discrete manifold spaces, *i.e.* mesh space in our task, is named the Discrete Laplace-Beltrami operator (DLBO). Similar to the Laplacian operator in image space that encodes the pixel information by capturing the local pixel differences [20, 28, 30, 34, 38, 44], DLBO encodes the mesh shape information by capturing the local shape fluctuation. The "Cotan formula" defined in [25] is the most widely used discretization, which can be represented in matrix form as

$$L_{ij} = \begin{cases} \sum_{j \in N(i)} \frac{1}{2A_i}(\cot \alpha_{ij} + \cot \beta_{ij}), & i = j \\ -\frac{1}{2A_i}(\cot \alpha_{ij} + \cot \beta_{ij}), & i \neq j \wedge j \in N(i) \\ 0, & i \neq j \wedge j \notin N(i), \end{cases} \tag{2}$$

where $L \in \mathbb{R}^{N \times N}$ is the DLBO matrix, with N the vertex number of the mesh. L_{ij} indicates its entry in ith row and jth column, which represents the edge weight between vertex v_i and v_j. A_i is the mixed Voronoi area of vertex v_i on the mesh. As shown in Fig. 3, the v_i's mixed Voronoi area is defined as the area of the polygon in which the vertices are the circumcenters of v_i's surrounding faces. $N(i)$ is the index set of v_i's adjacent vertices. If v_i and v_j are adjacent, α_{ij} and β_{ij} are the opposite angles of edge (v_i, v_j) in each of the edge's two neighbor triangle faces, respectively (shown in Fig. 3). If not, α_{ij} and β_{ij} are not defined and L_{ij} is 0. As shown in Fig. 2, the DLBO matrix is used for mesh Fourier transform to get mesh spectrum. We calculate the Fourier operator $U^\top$, which is the eigenfunction of L as

$$L = U \Lambda U^\top, \tag{3}$$

where Λ is a diagonal matrix whose diagonal elements are Fourier mesh frequencies.

To ensure the mesh frequencies are non-negative, we need the DLBO matrix L to be positive semidefinite. Our experiment in Fig. 7 gives an example of the counterintuitive

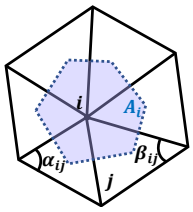

Figure 3. Variables defined in our discrete Laplace-Beltrami operator design.

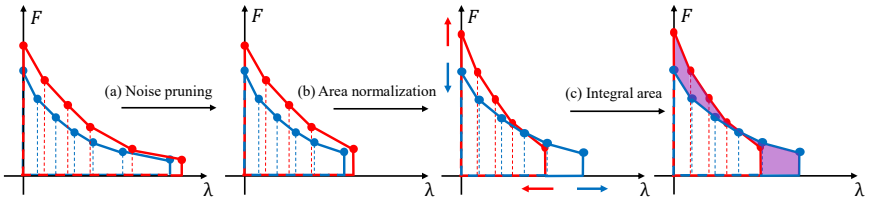

Figure 4. Spectrum Area Under the Curve Difference. We design our metric using the AUC difference of the spectrums. The blue curve and red curve are the test and ground truth mesh spectrum, respectively. The purple area in the last graph is the Spectrum AUC Difference. Please find details in Sec. 3.3.

results when there are negative frequencies. However, the Cotan formula in Eq. (2) does not guarantee to be positive semidefinite. We provide a simple example in Supplementary Materials Sec. 3 in which L is not positive semidefinite when the mesh is not Delaunay triangulated and the mixed Voronoi areas are not all equal to each other. In our metric design, we made two small changes to the original Cotan formula to make it positive semidefinite. a) Inspired by the symmetric normalization of the topology Laplacian matrix in [11], we make L symmetric by changing the normalization parameter A_i into a symmetric normalized manner $A_i^{\frac{1}{2}} A_j^{\frac{1}{2}}$. b) We replace $\cot \alpha_{ij} + \cot \beta_{ij}$ with $|\cot \alpha_{ij} + \cot \beta_{ij}|$. This ensures all the edge weights in the Laplacian matrix to be non-negative. Thus, our revision of DLBO is defined as

$$L_{ij} = \begin{cases} \frac{1}{2} \sum_{j \in N(i)} A_i^{-\frac{1}{2}} A_j^{-\frac{1}{2}} |\cot \alpha_{ij} + \cot \beta_{ij}|, & i = j \\ -\frac{1}{2} A_i^{-\frac{1}{2}} A_j^{-\frac{1}{2}} |\cot \alpha_{ij} + \cot \beta_{ij}|, & i \neq j \wedge j \in N(i) \\ 0, & i \neq j \wedge j \notin N(i). \end{cases} \tag{4}$$

We prove that our revision of the Cotan formula is positive semidefinite in Supplementary Materials Sec. 2. In Tab. 5, our experiments show that our DLBO matrix design outperforms the origin Cotan formula in [25], and the topology Laplacian matrix defined in [11].

Finally, we obtain the mesh spectrum by acting Fourier operator on the mesh vertices

$$F_i = \sqrt{G_{i,x}^2 + G_{i,y}^2 + G_{i,z}^2}, G = U^\top v, \tag{5}$$

where $v \in \mathbb{R}^{N \times 3}$ indicates the 3D coordinates of N mesh vertices. The result spectrum $F \in \mathbb{R}^N$. Fig. 5 shows an example of the mesh spectrum (left side) and how the meshes look in different frequency bands (right side). This provides an illustration of the information contained in different frequency bands of the mesh spectrum.

3.3. Spectrum AUC Difference

To reduce the noise and normalize the mesh scale, we also design noise pruning and AUC normalization procedures before calculating the Spectrum AUC Difference.

Noise pruning. As shown in Fig. 4 process *(a)*, we prune a small portion of the highest frequency information to reduce the interference of noise. From the observation of the first two meshes (A and B) in Fig. 5, we can see that humans can barely notice the shape differences when the highest frequency parts are removed. Thus, if we try to evaluate the mesh shape that aligns with human perception, it is reasonable to remove high-frequency noise without losing much information that humans care about. Empirically, we choose to prune the highest 0.1% frequency information as noise. Our experiments in Tab. 4 show that this portion is more consistent with human perception while preserving good mesh quality.

AUC normalization. Given a spectrum $F(\lambda)$, its Area Under the Curve (AUC) can be defined as $\int_\infty F(\lambda) d\lambda$. AUC normalization means using spectrum AUC to normalize the mesh scale. If the mesh scale increases by s times in length, the mixed Voronoi area, *i.e.* A_i in Eq. (4), will decrease by s^2 times. Thus, each element in the DLBO matrix L will also decrease by s^2 times. Then, according to Eq. (3), the frequency λ will decrease s^2 times to λ / s^2, and according to Eq. (5), the spectrum amplitude F will change to sF because v is increased by s time and $U^\top$ is still orthonormal. Then the area under the spectrum curve (the area boxed with red or blue lines in Fig. 4) changes as $A' = \int sF(\lambda) d\lambda / s^2 = \frac{1}{s} A$.

In our approach, we normalize the area under the spectrum curve A to 1 to resolve the scale difference, which means $s = A$, λ decreases by A^2 times, and spectrum amplitude F increases by A times (Fig. 4 process *(b)*). AUC normalization fairly normalizes the scale of objects in different shapes by only changing the scale, not shape details. It normalizes the spectrum AUC of all mesh to 1, making the mesh spectrums differ only in distributions. Our experiments in Tab. 5 demonstrate this design can bring a fairer comparison of the spectrums and improve the human consistency of the metric. The experiment also demonstrates that this design outperforms the spatial domain scale normalization.

Spectrum AUC Difference. In order to capture the difference between two mesh shapes in the spectrum domain, we design Spectrum AUC Difference (SAUCD) on the spectrum analysis results after noise pruning and AUC normalization:

$$d = D(\hat{M}, M_{gt}) = \int_\lambda |\hat{F}(\lambda) - F_{gt}(\lambda)| d\lambda, \tag{6}$$

dataset	Raw	w/ IQR removal
number of valid scores	24304	23775
Scoring range	$[0, 6]$	$[0, 6]$
95% confidence interval	0.318	0.303
Relative 95% confidence interval	5.33%	**5.04%**

Table 1. Dataset statistics and error analysis.

where $\hat{F}$ and F_{gt} are the test and groundtruth mesh spectrum. As shown in Fig. 4 process *(c)*, our metric is defined as the AUC difference of the two spectrum curves (the purple area). In Tab. 5, we compare our design with an alternative design which changes the amplitude difference $|\hat{F}(\lambda) - F_{gt}(\lambda)|$ to energy difference $|\hat{F}(\lambda)^2 - F_{gt}(\lambda)^2|$. The result shows our design is more consistent with human evaluation. Besides, experiments in Tab. 2 show our SAUCD metric aligns well with human evaluation, and outperforms SOTA metrics under multiple evaluation methods. Experiments in Fig. 8 show SAUCD has the capability to improve mesh detail qualities in 3D reconstruction when adapted into training loss.

3.4. Human-adjusted Spectrum AUC Difference

We also provide an extended metric version, in which we design a learnable weight parameter along the frequency bands. The weight parameter indicates the adjustment of human sensitivity to each frequency band. Specifically, we design the extended metric as

$$d_w = D_w(\hat{M}, M_{gt}) = \int_{\lambda} w(\lambda)|\hat{F}(\lambda) - F_{gt}(\lambda)|d\lambda. \quad (7)$$

$w(\lambda)$ is weight parameters indicating human sensitivity along frequency bands. Our training loss is defined as

$$\mathcal{L} = \lambda_p \mathcal{L}_{plcc} + \lambda_{sr} \mathcal{L}_{srocc} + \lambda_r \mathcal{L}_{regu}, \quad (8)$$

where $\mathcal{L}_{plcc}$ and $\mathcal{L}_{srocc}$ are Pearson correlation loss and Spearman rank order loss. They are defined the same as Pearson's linear correlation [29] and Spearman's rank order correlation [37]. $\mathcal{L}_{regu} = 1/N \sum_i (w_i - 1)^2$ is the regularization loss, which regularizes weight w_i close to 1. λ_p, λ_{sr}, and λ_r are the loss weights of $\mathcal{L}_{plcc}$, $\mathcal{L}_{srocc}$, and $\mathcal{L}_{regu}$. More details of the loss functions can be found in Supplementary Materials Sec. 1. Our experiments in Sec. 4.3 show that after adjustment, the consistency between our metric output and human-annotated ground truth is improved.

4. Experiments

4.1. Dataset

We build a user study benchmark dataset *Shape Grading* to evaluate whether our metric is aligned with human evaluation. The dataset contains the human evaluation scores for a variety of distorted meshes. Using this dataset, we can calculate the correlation between metric outputs and human evaluation scores to see how aligned the test metrics are to human evaluation.

Dataset design. We choose 12 objects as ground truth 3D triangle mesh from public object/scene/human mesh datasets such as [18, 26, 42] and commercial datasets such as [1, 2]. These objects are picked from different categories including humans, animals, buildings, plants, *etc*.. For each object, we synthesize 7 different types of distortions which commonly occur in 3D reconstruction. For each distortion type we synthesize 4 distortion levels, which gives us $7 \times 4 = 28$ distorted objects for every ground truth object. We rotate and render each distorted object into 3 videos using different materials for the mesh. In total, we generate $12 \times 28 \times 3 = 1,008$ distorted mesh videos. Supplementary Materials Sec. 4 shows the meshes and distortion types we use in our dataset.

Human scoring procedure. We use a pairwise comparison scoring process similar to [31]. Each subject will evaluate all 28 distorted objects of one ground truth object with a certain material. The scoring follows a Swiss system tournament principle used in [31], in which each subject takes 6 rounds of pairwise comparison to score the distorted meshes. After 6 rounds of scoring, the meshes are scored from 0 to 6. 0 means the object loses in every round and 6 means it wins in every round. This process will largely reduce the biases among subjects, since the subjects are compelled to distribute an equal amount of points to the 28 distorted objects. The process will take about 15 minutes for each subject, avoiding the fatigue problem in [7]. For every object rendered with every material, we have 24 to 25 subjects scoring it. In total, we have 868 subjects (536 males, 316 females, and 16 others) who give us $868 \times 28 = 24304$ scores. More details of the scoring procedure can be found in Supplementary Materials Sec. 5.

Outlier detection. We use the interquartile range (IQR) method [13] which is widely used in statistics to detect and remove outliers. For each distorted mesh, we first find the 25 percentile and the 75 percentile of the scores. The score range in between is called the IQR range. We remove the scores that are 1.5IQR smaller than the 25 percentile or 1.5IQR larger than the 75 percentile. Our dataset error analysis in Tab. 1 shows, that by removing 2.2% of the scores using IQR, we can decrease the uncertainty of the final scoring result by nearly 6%.

Dataset error analysis. We analysis the average 95% confidence interval of our dataset scores in Tab. 1. The confidence interval of score x can be calculated as $\sigma_{\bar{x}} = z_{0.95} \times \sigma/\sqrt{N}$ where σ is the standard derivation of x, N is the number of valid scores, and $z_{0.95} \approx 1.96$. We report the average 95% confidence interval and the relative 95% confidence interval (which is the confidence interval divided by the scoring range). The result shows that dataset scoring

Metrics \ Object No.	1	2	3	4	5	6	7	8	9	10	11	12	Overall
Chamfer Distance [6]	0.54	0.15	-0.10	0.57	-0.06	-0.12	-0.20	0.07	0.04	0.30	-0.20	0.17	0.097
Point-to-Surface	0.45	0.19	-0.04	0.66	-0.08	-0.25	-0.32	-0.20	0.01	0.13	-0.21	-0.12	0.017
Normal Difference	0.46	0.11	0.06	0.28	0.11	0.21	0.29	0.47	0.27	0.39	0.11	0.27	0.253
IoU [16]	0.60	0.63	0.01	0.51	0.30	0.02	-0.07	0.20	0.14	0.47	-0.09	-0.01	0.225
F-score [43]	0.58	0.09	0.05	0.33	0.03	0.06	0.16	0.34	0.27	0.25	0.01	0.34	0.208
SSFID [46]	0.71	**0.74**	-0.04	**0.74**	**0.39**	0.24	0.13	0.32	0.25	0.64	0.25	-0.02	0.363
UHD [47]	0.29	0.22	0.11	0.15	-0.04	0.18	0.41	0.55	0.13	0.18	0.25	0.33	0.231
SAUCD (Ours)	0.73	0.21	**0.60**	0.63	0.31	0.51	**0.83**	0.65	**0.77**	**0.80**	**0.69**	0.08	0.567
Adjusted SAUCD (Ours)	**0.79**	0.19	0.56	0.64	0.36	**0.54**	0.79	**0.76**	0.75	0.77	0.67	**0.36**	**0.598**

a. Pearson's linear correlation coefficient.

Metrics \ Object No.	1	2	3	4	5	6	7	8	9	10	11	12	Overall
Chamfer Distance [6]	0.33	0.14	-0.09	0.43	-0.08	-0.06	-0.15	0.17	-0.04	0.24	-0.16	0.22	0.079
Point-to-Surface	0.42	0.39	0.14	0.59	0.11	0.05	-0.10	0.20	0.18	0.40	-0.11	0.18	0.205
Normal Difference	0.44	0.22	0.33	0.42	0.19	0.29	0.33	0.56	0.33	0.32	0.21	0.34	0.331
IoU [16]	0.57	0.61	0.28	0.50	0.36	0.21	0.12	0.31	0.262	0.56	0.03	0.30	0.342
F-score [43]	0.47	0.25	0.20	0.52	0.21	0.11	0.07	0.36	0.30	0.42	-0.01	0.35	0.27
SSFID [46]	0.63	**0.81**	0.28	**0.70**	0.33	0.23	0.10	0.33	0.32	0.65	0.16	0.34	0.407
UHD [47]	0.38	0.20	0.11	0.32	0.13	0.35	0.41	0.60	0.06	0.27	0.37	0.35	0.296
SAUCD (Ours)	0.79	0.25	**0.57**	0.59	0.36	0.56	**0.83**	0.79	0.69	**0.69**	**0.83**	0.24	0.598
Adjusted SAUCD (Ours)	**0.83**	0.21	0.55	0.59	**0.38**	**0.60**	0.82	**0.80**	**0.69**	0.68	0.75	**0.42**	**0.611**

b. Spearman's rank order correlation coefficient.

Metrics \ Object No.	1	2	3	4	5	6	7	8	9	10	11	12	Overall
Chamfer Distance [6]	0.25	0.14	-0.08	0.31	-0.04	-0.02	-0.09	0.15	0.013	0.19	-0.07	0.22	0.080
Point-to-Surface	0.33	0.30	0.07	0.45	0.10	0.08	-0.03	0.17	0.13	0.30	-0.01	0.16	0.171
Normal Difference	0.34	0.16	0.17	0.31	0.18	0.22	0.26	0.44	0.25	0.23	0.16	0.27	0.250
IoU [16]	0.42	0.44	0.24	0.37	0.28	0.22	0.14	0.26	0.20	0.41	0.10	0.23	0.275
F-score [43]	0.37	0.17	0.14	0.42	0.15	0.11	0.09	0.28	0.23	0.34	0.01	**0.30**	0.216
SSFID [46]	0.48	**0.62**	0.24	**0.51**	0.25	0.24	0.12	0.29	0.26	**0.48**	0.17	0.23	0.322
UHD [47]	0.27	0.13	0.07	0.22	0.09	0.26	0.29	0.42	0.048	0.19	0.28	0.24	0.209
SAUCD (Ours)	0.60	0.16	**0.42**	0.41	0.27	0.45	**0.65**	0.57	**0.55**	0.47	**0.60**	0.19	0.445
Adjusted SAUCD (Ours)	**0.64**	0.14	0.40	0.41	**0.29**	**0.48**	0.63	**0.59**	0.55	0.45	0.57	0.29	**0.453**

c. Kendall's rank order correlation coefficient.

Table 2. Correlations between different metrics and human annotation. "SAUCD" is our basic version metric. "Adjusted SAUCD" is the human-adjusted version of our metric. The ranges of all three correlation coefficients are $[-1, 1]$, and the higher the better.

is accurate with a 5% error range with IQR outlier removal.

Evaluation methods. We use 3 different evaluation methods to evaluate the correlation between our metrics and the human scoring (ground truth) on our *Shape Grading* benchmark dataset. Pearson's linear correlation coefficient (**PLCC**) [29] is used to evaluate the linear alignment between our metric and human perception. We also used Spearman's rank order correlation coefficient (**SROCC**) [37] and Kendall's rank order correlation coefficient (**KROCC**) [19] to evaluate the ranking order correlation between our metric and human perception. The possible ranges of 3 metrics are all $[-1, 1]$. Higher numbers mean stronger correlations. More details of the three evaluation methods can be found in Supplementary Materials Sec. 1.

4.2. Implementation details

We implement our basic version metric following Eq. (6). $\hat{F}(\lambda)$ and $F_{gt}(\lambda)$ in Eq. (6) are both piece-wise functions, so we implement the integration by simply adding every piece area together. We implement our human-adjusted version following Eq. (7). We use a 20-dimensional weight $w(\lambda)$ to avoid overfitting. We interpolate w to all frequencies of the ground truth and test meshes and element-wisely multiply them to the spectrums. In spectrum weight training, SROCC and PLCC are used as part of the loss function as Eq. (8). KROCC is not used in training but only for testing. We use a k-fold strategy for training the human-adjusted weight. Each time we choose 1 object for testing and the rest 11 objects for training, which means $k = 12$. More implementation details can be found in Supplementary Materials Sec. 1.

4.3. Quantitive and qualitative results

SOTA comparison. Tab. 2 shows our results compared to previous 3D mesh shape metrics. We evaluated the correlation between each metric and the human scoring via three different evaluation methods. We observe that **a)** without any learning-based design, our metric outperforms the SOTA learning-based (SSFID) and non-learning-based met-

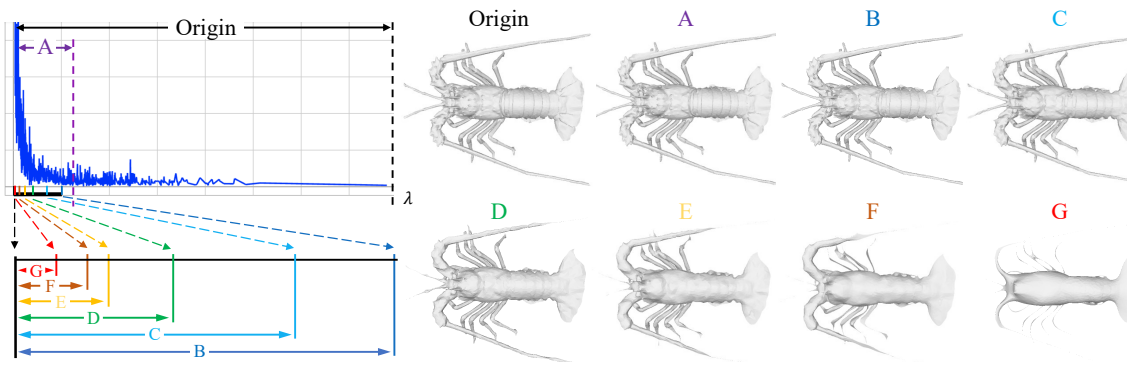

Figure 5. An example of mesh spectrum curve: We do mesh Fourier transform on the "Origin" mesh and show the spectrum in the left graph. The λ-axis is the eigenvalues of the DLBO matrix, the larger the higher frequency. We also show how mesh changes when gradually removing high-frequency information (mesh A to G). The frequency bands of the meshes are shown as the colored arrows in the left graph.

rics (Chamfer Distance, IoU, F-score, and UHD), **b)** our extended version metric with learned weights has better linearity and slightly better ranking order correction with human evaluation, and **c)** our results on different objects show that our metrics have good generalizability.

Spectrum example. We first show an example of mesh spectrum in Fig. 5. We decompose the "origin" mesh using the Fourier Transform and get the resulting spectrum (top-left graph). The meshes on the right (from mesh A to G) are generated by gradually removing high-frequency information. The frequency bands of the meshes are shown as colored arrows in and under the graph. As we see, the details gradually disappear as we remove high-frequency information.

Frequency band separation. We explored the consistency between human perception and the information obtained from every frequency band. Specifically, we separate the frequency band exponentially and build metrics only using information from that frequency band. The results are shown in Tab. 3, we find the frequency bands $[0, 0.001]$ and $[0.01, 0.03]$ have the best consistency with human perception. Moreover, it shows that if we put all frequencies together, they can achieve better results.

Trained weight. We show our trained weights in Human-adjusted metric in Fig. 6. Different lines represent different folds, and the bold **purple** line is the average weight. We can see the weights trained on each fold have similar patterns. We also observe that the weight curves have a small peak in the range A and two much larger peaks between A and B, which means our extended metric relies more on the information between A and B. We show an example of mesh shapes in the range A, B, and **C** at the bottom of Fig. 6. Mesh A obviously has fewer details than Mesh B, and the weight curve shows that this difference is what the learning process tries to emphasize.

Negative frequencies. In Fig. 7 we illustrate how our revised Cotan formula DLBO in Eq. (4) improves frequency analysis compared to the original Cotan formula in Eq. (2) [25]. The first and second rows are the results of the original Cotan formula DLBO and our revised Cotan formula DLBO, respectively. The original Cotan formula can yield negative frequencies due to its lack of positive semidefiniteness, whereas our revision ensures all frequencies are non-negative. For both objects in the figure, we remove different portions of high-frequency information and show the remaining low-frequency parts (resulting in "Filtered mesh 1" and "Filtered mesh 2"). For the left object, notice the counterintuitive sharp shapes in the red circle when using the Cotan formula. The right object is a much more severe case. Sharp shapes in low-frequency parts show improper decomposition and high-frequency aliasing with low-frequency shapes, making the Cotan formula unsuitable for spectral

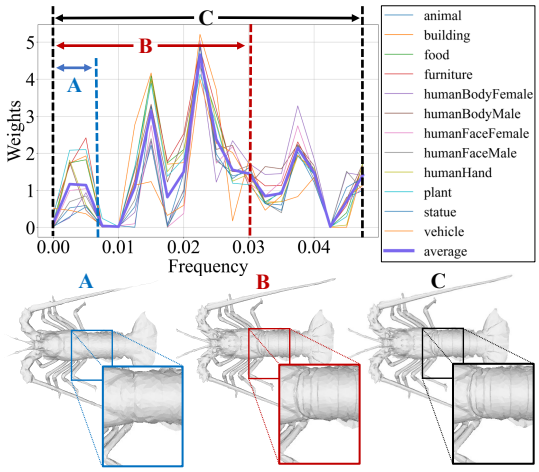

Figure 6. Learned spectrum weights on all 12 folds. The name of colorful thin lines means the test object name of that fold. The bold **purple** line is the average weights of all folds. We also show some examples of mesh shape information in different frequency bands. Frequency band A is $[0, 0.0075)$, B is $[0, 0.03)$, and C is $[0, 0.05)$.

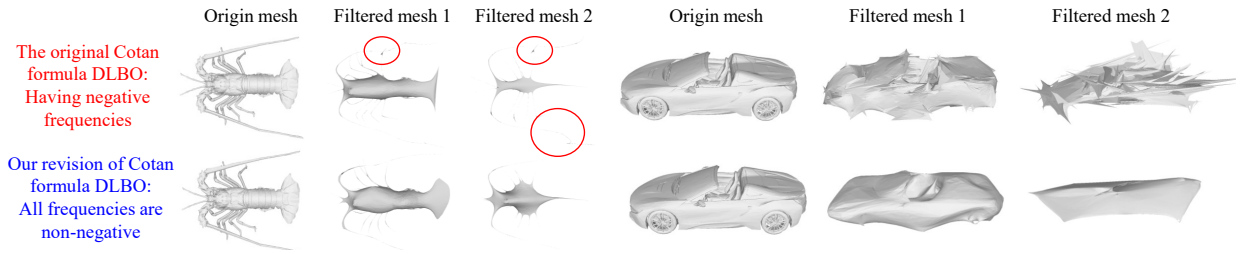

Figure 7. Counterintuitive low-frequencies information if some of the mesh frequencies are negative. We can see if we remove the high-frequency part of the mesh (resulting in "Filtered mesh 1" and "Filtered mesh 2") using the original Cotan formula, the mesh's low-frequency parts show artifacts (sharp shapes). The red circles show the artifacts in the left object. The right object shows a case when these artifacts occur much more often. These artifacts do not occur using our revised Cotan formula DLBO.

Frequency band	PLCC ↑	SROCC ↑	KROCC ↑
$[0, 0.001)$	0.434	0.515	0.376
$[0.001, 0.003)$	0.240	0.409	0.281
$[0.003, 0.01)$	0.255	0.455	0.340
$[0.01, 0.03)$	0.421	0.528	0.391
$[0.03, 0.1)$	0.287	0.351	0.250
$[0.1, \infty)$	0.318	0.192	0.155
$[0, \infty)$	**0.567**	**0.598**	**0.445**

Table 3. Results when building metrics using each frequency band separately. The bottom row is our proposed metric.

Pruning Portion	PLCC ↑	SROCC ↑	KROCC ↑
0%	0.513	0.549	0.393
0.1%	**0.567**	0.598	**0.445**
1%	0.554	**0.602**	0.462
10%	0.517	0.581	0.442
20%	0.503	0.587	0.445

Table 4. Results with different pruning portions. The metric achieves better results with pruning portion to be 0.1% or 1%. We use pruning portion as 0.1% in our design.

Modules	PLCC ↑	SROCC ↑	KROCC ↑
Topology Laplacian [11]	0.298	0.327	0.235
Cotan formula [25]	0.417	0.470	0.340
Energy difference	0.268	0.315	0.215
w/o normalization	0.257	0.507	0.353
Spatial normalization	0.269	0.542	0.392
Ours	**0.567**	**0.598**	**0.445**

Table 5. Module replacement. We replace each module of our metric with alternative designs to verify the design of each module.

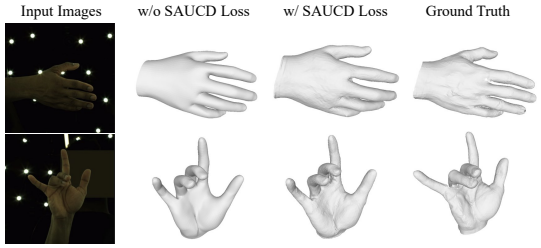

Figure 8. We Adapt SAUCD into a loss function and use it in monocular-image-based 3D hand reconstruction. From left to right: input images, reconstruction result w/o SAUCD loss, reconstruction result w/ SAUCD loss, and ground truth mesh. We can see that the enhancement of SAUCD loss in mesh details is clearly noticeable.

mesh comparison. In contrast, our revised formula yields smooth low-frequency components without these artifacts.

Noise pruning portion. Tab. 4 shows our SAUCD metric performance by changing the noise pruning portion (Sec. 3.3). The metric achieves better results when the pruning portion is 0.1% or 1%. In our proposed metric, we choose the pruning portion to be 0.1% to best avoid possible high-frequency information loss.

Module replacement. Tab. 5 shows our SAUCD metric performance by replacing some modules with alternative designs. First, we replace our revision of the discrete Laplace-Beltrami operator in Eq. (4) with topology Laplacian matrix in [11] and "Cotan formula" in [25]. Second, we change the AUC difference defined in Eq. (6) into the energy difference, which means changing $|\hat{F}(\lambda) - F_{gt}(\lambda)|$ in Eq. (6) into $|\hat{F}(\lambda)^2 - F_{gt}(\lambda)^2|$. In the third experiment, we replace AUC normalization (in Sec. 3.3) with spatial normalization, where we normalize the meshes by their maximum range along all 3 spatial axes. We also removed the AUC normalization module for another comparison. Our experiments show SAUCD has better performance than alternative designs.

Adapting SAUCD to loss function. We adapted our metric into a loss function to enhance the visual quality of 3D mesh reconstructions, as evident from the hand reconstruction results in Fig. 8. Details on the experiment's implementation are available in Supplementary Materials Sec. 7. From this experiment, we can see that the enhancement of SAUCD loss in mesh details is clearly noticeable.

Visualized examples. We visualize examples in our dataset and their evaluation result using different metrics in Supplementary Materials Sec. 6.

5. Conclusions

In order to propose a 3D shape evaluation that better aligns with human perception, we design an analytic metric named Spectrum AUC Difference (SAUCD). Our proposed SAUCD leverages mesh spectrum analysis to evaluate 3D shape that aligns with human evaluation, and its extended version Human-adjusted SAUCD further explores the sensitivity of human perception of each frequency band. To evaluate our new metrics, we build a user study dataset to compare our metrics with existing metrics. The results validate that both our new metrics are well aligned with human perceptions and outperform previous methods.

References

[1] Gobotree - photos, cut-outs, 3d people. https://www.gobotree.com/.

[2] Sketchfab - the best 3d viewer on the web. https://sketchfab.com/.

[3] Panos Achlioptas, Olga Diamanti, Ioannis Mitliagkas, and Leonidas Guibas. Learning representations and generative models for 3d point clouds. In *ICML*, pages 40–49, 2018.

[4] Jan Bechtold, Maxim Tatarchenko, Volker Fischer, and Thomas Brox. Fostering generalization in single-view 3d reconstruction by learning a hierarchy of local and global shape priors. In *CVPR*, pages 15880–15889, 2021.

[5] Alexander I Bobenko, John M Sullivan, Peter Schröder, and G Ziegler. *Discrete differential geometry*. Springer, 2008.

[6] Gunilla Borgefors. Distance transformations in arbitrary dimensions. *Computer vision, graphics, and image processing*, pages 321–345, 1984.

[7] RECOMMENDATION ITU-R BT. Methodology for the subjective assessment of the quality of television pictures. *International Telecommunication Union*, 2002.

[8] Abdullah Bulbul, Tolga Capin, Guillaume Lavoué, and Marius Preda. Assessing visual quality of 3-d polygonal models. *IEEE Signal Processing Magazine*, pages 80–90, 2011.

[9] William L Burke, William L Burke, and William L Burke. *Applied differential geometry*. Cambridge University Press, 1985.

[10] Zhiqin Chen, Vladimir G Kim, Matthew Fisher, Noam Aigerman, Hao Zhang, and Siddhartha Chaudhuri. Decor-gan: 3d shape detailization by conditional refinement. In *CVPR*, pages 15740–15749, 2021.

[11] Fan RK Chung. *Spectral graph theory*. American Mathematical Soc., 1997.

[12] Massimiliano Corsini, Mohamed-Chaker Larabi, Guillaume Lavoué, Oldřich Petřík, Libor Váša, and Kai Wang. Perceptual metrics for static and dynamic triangle meshes. In *Comput. Graph. Forum*, pages 101–125, 2013.

[13] Frederik Michel Dekking, Cornelis Kraaikamp, Hendrik Paul Lopuhaä, and Ludolf Erwin Meester. *A Modern Introduction to Probability and Statistics: Understanding why and how*. Springer, 2005.

[14] Javier Duoandikoetxea and Javier Duoandikoetxea Zuazo. *Fourier analysis*. American Mathematical Soc., 2001.

[15] Kyle Genova, Forrester Cole, Avneesh Sud, Aaron Sarna, and Thomas Funkhouser. Local deep implicit functions for 3d shape. In *CVPR*, pages 4857–4866, 2020.

[16] Paul Henderson and Vittorio Ferrari. Learning to generate and reconstruct 3d meshes with only 2d supervision. *arXiv preprint arXiv:1807.09259*, 2018.

[17] Tao Hu, Liwei Wang, Xiaogang Xu, Shu Liu, and Jiaya Jia. Self-supervised 3d mesh reconstruction from single images. In *CVPR*, pages 6002–6011, 2021.

[18] Rasmus Jensen, Anders Dahl, George Vogiatzis, Engin Tola, and Henrik Aanæs. Large scale multi-view stereopsis evaluation. In *CVPR*, pages 406–413, 2014.

[19] Maurice George Kendall et al. The advanced theory of statistics. *The advanced theory of statistics*, 1946.

[20] Dilip Krishnan and Rob Fergus. Fast image deconvolution using hyper-laplacian priors. *NeurIPS*, 22, 2009.

[21] Audrius Kulikajevas, Rytis Maskeliunas, Robertas Damasevicius, and Tomas Krilavicius. Auto-refining 3d mesh reconstruction algorithm from limited angle depth data. *IEEE Access*, pages 87083–87098, 2022.

[22] Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, pages 79–86, 1951.

[23] Guillaume Lavoué. A local roughness measure for 3d meshes and its application to visual masking. *ACM Transactions on Applied Perception*, pages 1–23, 2009.

[24] Peizhen Lin, Hongliang Zhong, Lei Wang, and Jun Cheng. 3d mesh reconstruction of indoor scenes from a single image in-the-wild. In *International Conference on Graphics and Image Processing*, pages 457–465, 2022.

[25] Mark Meyer, Mathieu Desbrun, Peter Schröder, and Alan H Barr. Discrete differential-geometry operators for triangulated 2-manifolds. In *Visualization and mathematics III*, pages 35–57. Springer, 2003.

[26] Gyeongsik Moon, Takaaki Shiratori, and Kyoung Mu Lee. Deephandmesh: A weakly-supervised deep encoder-decoder framework for high-fidelity hand mesh modeling. In *ECCV*, 2020.

[27] Yinyu Nie, Xiaoguang Han, Shihui Guo, Yujian Zheng, Jian Chang, and Jian Jun Zhang. Total3dunderstanding: Joint layout, object pose and mesh reconstruction for indoor scenes from a single image. In *CVPR*, pages 55–64, 2020.

[28] Sylvain Paris, Samuel W Hasinoff, and Jan Kautz. Local laplacian filters: edge-aware image processing with a laplacian pyramid. *ACM TOG*, page 68, 2011.

[29] Karl Pearson. Notes on the history of correlation. *Biometrika*, pages 25–45, 1920.

[30] Patrick Pérez, Michel Gangnet, and Andrew Blake. Poisson image editing. In *SIGGRAPH*, pages 313–318, 2003.

[31] Nikolay Ponomarenko, Vladimir Lukin, Alexander Zelensky, Karen Egiazarian, Marco Carli, and Federica Battisti. Tid2008-a database for evaluation of full-reference visual quality assessment metrics. *Advances of Modern Radioelectronics*, pages 30–45, 2009.

[32] Marie-Julie Rakotosaona, Paul Guerrero, Noam Aigerman, Niloy J Mitra, and Maks Ovsjanikov. Learning delaunay surface elements for mesh reconstruction. In *CVPR*, pages 22–31, 2021.

[33] Hari Santhanam, Nehal Doiphode, and Jianbo Shi. Automated line labelling: Dataset for contour detection and 3d reconstruction. In *WACV*, pages 3136–3145, 2023.

[34] Jianbing Shen, Xiaogang Jin, Chuan Zhou, and Charlie CL Wang. Gradient based image completion by solving the poisson equation. *Computers & Graphics*, pages 119–126, 2007.

[35] Rakesh Shrestha, Zhiwen Fan, Qingkun Su, Zuozhuo Dai, Siyu Zhu, and Ping Tan. Meshmvs: Multi-view stereo guided mesh reconstruction. In *3DV*, pages 1290–1300. IEEE, 2021.

[36] Dong Wook Shu, Sung Woo Park, and Junseok Kwon. 3d point cloud generative adversarial network based on tree structured graph convolutions. In *ICCV*, pages 3859–3868, 2019.

[37] Charles Spearman. Correlation calculated from faulty data. *British journal of psychology*, page 271, 1910.

[38] Shen-Chuan Tai and Shih-Ming Yang. A fast method for image noise estimation using laplacian operator and adaptive edge detection. In *International Symposium on Communications, Control and Signal Processing*, pages 1077–1081, 2008.

[39] Jiaxiang Tang, Xiaokang Chen, Jingbo Wang, and Gang Zeng. Point scene understanding via disentangled instance mesh reconstruction. In *ECCV*, pages 684–701, 2022.

[40] Maxim Tatarchenko, Stephan R Richter, René Ranftl, Zhuwen Li, Vladlen Koltun, and Thomas Brox. What do single-view 3d reconstruction networks learn? In *CVPR*, pages 3405–3414, 2019.

[41] Kai Wang, Guillaume Lavoué, Florence Denis, Atilla Baskurt, and Xiyan He. A benchmark for 3d mesh watermarking. In *Shape Modeling International Conference*, pages 231–235. IEEE, 2010.

[42] Lizhen Wang, Zhiyuan Chen, Tao Yu, Chenguang Ma, Liang Li, and Yebin Liu. Faceverse: a fine-grained and detail-controllable 3d face morphable model from a hybrid dataset. In *CVPR*, pages 20333–20342, 2022.

[43] Nanyang Wang, Yinda Zhang, Zhuwen Li, Yanwei Fu, Wei Liu, and Yu-Gang Jiang. Pixel2mesh: Generating 3d mesh models from single rgb images. In *ECCV*, pages 52–67, 2018.

[44] Xin Wang. Laplacian operator-based edge detectors. *IEEE TPAMI*, pages 886–890, 2007.

[45] Xingkui Wei, Zhengqing Chen, Yanwei Fu, Zhaopeng Cui, and Yinda Zhang. Deep hybrid self-prior for full 3d mesh generation. In *ICCV*, pages 5805–5814, 2021.

[46] Rundi Wu and Changxi Zheng. Learning to generate 3d shapes from a single example. *arXiv preprint arXiv:2208.02946*, 2022.

[47] Rundi Wu, Xuelin Chen, Yixin Zhuang, and Baoquan Chen. Multimodal shape completion via conditional generative adversarial networks. In *ECCV*, pages 281–296, 2020.

[48] Rongfei Zeng, Mai Su, and Xingwei Wang. Cd2: Fine-grained 3d mesh reconstruction with twice chamfer distance. *arXiv preprint arXiv:2206.00447*, 2022.

[49] Zhihao Zhang, Xinyang Ren, and Xianqiang Yang. Parametric chamfer alignment based on mesh deformation. *Measurement and Control*, pages 192–201, 2023.

[50] Xinxin Zuo, Sen Wang, Minglun Gong, and Li Cheng. Unsupervised 3d human mesh recovery from noisy point clouds. *arXiv preprint arXiv:2107.07539*, 2021.