Pacific Graphics 2022
N. Umetani, E. Vouga, and C. Wojtan
(Guest Editors)

*Volume 40 (2022), Number 7*

# Learning 3D Shape Aesthetics Globally and Locally

Minchan Chen and Manfred Lau

City University of Hong Kong

**Abstract**

*There exist previous works in computing the visual aesthetics of 3D shapes "globally", where the term global means that shape aesthetics data are collected for whole 3D shapes and then used to compute the aesthetics of whole 3D shapes. In this paper, we introduce a novel method that takes such "global" shape aesthetics data, and learn both a "global" shape aesthetics measure that computes aesthetics scores for whole 3D shapes, and a "local" shape aesthetics measure that computes to what extent a local region on the 3D shape surface contributes to the whole shape's aesthetics. These aesthetics measures are learned, and hence do not consider existing handcrafted notions of what makes a 3D shape aesthetic. We take a dataset of global pairwise shape aesthetics, where humans compares between pairs of shapes and say which shape from each pair is more aesthetic. Our solution proposes a point-based neural network that takes a 3D shape represented by surface patches as input and jointly outputs its global aesthetics score and a local aesthetics map. To build connections between global and local aesthetics, we embed the global and local features into the same latent space and then output scores with the weights-shared aesthetics predictors. Furthermore, we designed three loss functions to supervise the training jointly. We demonstrate the shape aesthetics results globally and locally to show that our framework can make good global aesthetics predictions while the predicted aesthetics maps are consistent with human perception. In addition, we present several applications enabled by our local aesthetics metric.*

**CCS Concepts**

• ***Computing methodologies*** → *Shape analysis; Perception;*

## 1. Introduction

The computational measurement of aesthetics is a significant research problem due to its wide potential applications in areas where visual experience is involved, including product designs, artificial intelligence, and human-computer interactions. There have been previous works [DLT17, RSL*17, ZZL*21] that explore the automatic assessment of image aesthetics with deep learning and datasets with thousands of annotated images. For 3D shapes, early works [Pha99,PZ03,Séq05,BR13,MR14] focused on building relationships between aesthetic properties suggested by art and philosophy with handcrafted geometric features (e.g., curvature, symmetry, proportion) or mathematical criteria (e.g., bending energy, minimum variation surface). Recently, Dev and Lau [DL22] first proposed a learning-based shape aesthetics metric based on a human shape aesthetics dataset. Given a 3D shape represented by multi-view images, they learned a neural network to compute the shape's aesthetic score. Although Dev and Lau [DL22] can rate the shapes' aesthetics automatically, their work cannot automatically figure out which elements on the shape surface contribute to the shape's overall beauty. To the best of our knowledge, there is no existing work that predicts shape aesthetics maps with a learning-based method and not using manually defined aesthetics features.

The key contribution of this paper is in simultaneously predicting the global aesthetics scores and local aesthetics maps for 3D shapes. By the term "global", we mean that the shape aesthetics data are collected for whole 3D shapes and then used to compute the aesthetics of whole 3D shapes. We introduce a novel method in

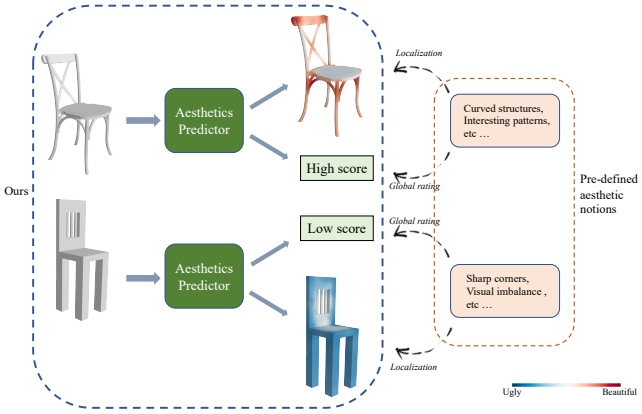


**Figure 1:** *Instead of indirectly formulating global and local shape aesthetics attributes, we propose a learning-based framework to directly output the global shape aesthetics score and local shape aesthetics map simultaneously.*

this paper that takes such "global" shape aesthetics data, and simultaneously learn a "global" shape aesthetics measure that computes aesthetics scores for whole 3D shapes, and a "local" shape aesthetics measure that computes to what extent a local region on the 3D shape surface contributes to the whole shape's aesthetics. The "local" shape aesthetics measure can then be used to compute shape aesthetics maps on the surfaces of 3D shapes to indicate each local region's influence towards the whole shape's aesthetics.

In this work, we do not consider any manually defined shape aesthetics features. Instead, we learn the global shape aesthetics score and the local shape aesthetics map directly from a dataset of global pairwise shape aesthetics comparisons (see Figure 1). This data comes from asking humans to compare between pairs of whole 3D shapes and to say which shape from each pair is more aesthetic.

We propose a learning-based method to solve the following two problems simultaneously: (i) whether a 3D shape looks aesthetic overall, and (ii) which surface region(s) or element(s) contribute to a shape's beauty. We introduce a point-based neural network (see Figure 3) with several deliberately designed structures and loss functions to encourage the learning of meaningful local shape aesthetics information from the global shape aesthetics data. Specifically, to capture the local attributes of the 3D shapes, we chose a patch-based representation [HLK*17] which focuses on localizing the style-defined elements over the shape surface. But unlike this previous work [HLK*17] which clusters the patches and then iteratively selects the discriminating elements, our method feeds each shape with its patches to the neural network, and then simultaneously outputs the global aesthetics score for the shape and the local aesthetic scores for the surface patches. As the global aesthetics comparisons provide weak supervision for the local predictions, in our networks, we embed the global features and local features into the same latent space and then output scores with the weights-shared aesthetic predictors. In addition, to further build connections between the global aesthetics and local aesthetics maps, we make the assumption that "if shape A is more aesthetic than shape B, A is likely to have more beautiful patches than B" as an additional constraint. Details of our network architecture and loss functions can be found in Section 3.

We evaluate the performance of our method globally and locally. For the global aesthetics prediction, we quantitatively compare our results with the results of Dev and Lau [DL22] and show that our performance is comparable or even better. For the local aesthetics maps, we conduct user studies to quantitatively and qualitatively demonstrate that our predictions are consistent with users' perceptions. Furthermore, to demonstrate the effectiveness of our method, we develop several applications such as aesthetics-based patch galleries, aesthetics-driven subparts extraction, and aesthetics-guided shape editing that can benefit from the types of shape aesthetics maps that we compute.

In summary, our work makes the following contributions:

- We propose a novel learning-based framework to simultaneously predict global aesthetics scores and local aesthetics maps for 3D shapes.
- Our framework is the first method to learn the local aesthetics attributes from "global" shape aesthetics data consisting of shape pairwise comparisons, without considering existing handcrafted shape aesthetics features.
- We present several applications that are made possible by our shape aesthetics maps: aesthetics-based patch gallery, aesthetics-driven subparts extraction, and aesthetics-guided shape editing.

## 2. Related Work

Our work simultaneously predicts the perceptual attribute of shape aesthetics "globally" and "locally" via point-based neural networks. In this section, we discuss previous works related to: metrics of aesthetics, and the global and local perceptual analysis of 3D shapes.

### 2.1. Metrics of Aesthetics

The concept of aesthetics has been analyzed in many studies of philosophy, psychology and art. In computer graphics and vision, while many works have proposed aesthetic metrics for 2D images based on commonly established photographic rules and learning-based methods [DLT17,ZZL*21], there have been fewer works that explore the aesthetics of 3D shapes.

Early work [Pha99] tried to propose a systematic approach for exploring the interactions of aesthetics principals and term-based design variables in 3D designs by integrating knowledge from fields such as philosophy, psychology and arts. Then the following work [PZ03] used parametric geons to derive the membership functions for linguistic geometric descriptions and aesthetic characteristics by a series of user studies. Furthermore, mathematical geometric-based criteria [BR13] were defined by extending the existing criteria for image aesthetic assessment, and mathematical models [MR14] were proposed to formulate aesthetic curves and surfaces directly. Instead of using specific handcrafted features to formulate the aesthetic metric, recent work [DL22] proposed a cross-category shape aesthetics metric via multi-view image-based neural networks. While their work only predicts a global aesthetics score for an input shape, our patch-based network also captures the local surface information and outputs a local shape aesthetics map simultaneously.

### 2.2. Global Perceptual Analysis of 3D Shapes

From the "global" perspective of 3D shapes as a whole, existing works explored metrics to rate a single shape's perceptual attribute or rate the pairwise attribute comparison between two shapes.

Some works aimed to learn intra-category perceptual attributes. Xu et al. [XLZ*10] explored the style between shapes in the same functional class by anisotropic structural ratios. Rather than focus on a single specific attribute, Yumer et al. [YCHK15] conducted user studies to explore the most relevant attributes of shapes in the same category and then formulate the score function for each attribute fitted by the crowdsourcing rates. Other works aimed to work with heterogeneous 3D shape collections and explore inter-category instincts. Machine learning with handcrafted geometric descriptors [LHLF15, LKS15] or deep learning with image-based representations [LGK16, WYA*20, DL22] were proposed to explore style compatibility or aesthetics measuring for 3D shapes

crossing categories. Our work chooses the locally-defined geometric descriptors as the input of a point-based neural network and trains an aesthetics predictor to capture both global and local information.

### 2.3. Local Perceptual Analysis of 3D shapes

From the "local" perspective of local regions on 3D shape surfaces, existing works tried to measure the contributions of local regions towards the perceptual attributes.

Lee et al. [LVJ05] proposed handcrafted mesh saliency via local geometric cues to measure the regional visual importance of 3D meshes, and Song et al. [SLMR14] took global information also into consideration. Without the connectivity information of the mesh surface, Shtrom et al. [SLT13] manually define hierarchical saliency for dense point clouds via multi-scale distinctness and association between point-based features. Instead of directly defining the saliency via handcrafted geometric descriptors, some works also used data-based and learning methods. Regression models [CSPF12] or neural networks [SXX*18] were proposed to predict a probability distribution for points of saliency or interest from pointwise marked data in a large-scale user study. Besides visual perception, Lau et al. [LDS*16] learned a tactile saliency metric via crowdsourcing pairwise comparison data.

As local-based perceptual datasets are sometimes difficult to obtain, style co-analysis supervised by global style labels [HLK*17] or semi-supervisions like global triplet comparisons [YZX*18] were proposed to extract the cross-category style-defining local elements. Moreover, Remil et al. [RXCW19] unsupervisedly learned the content-revealing patches and style-revealing patches from a set of shapes in the same category. Our work also learns the local perceptual information from the global dataset. However, in contrast to the above previous works, which identify the attribute-defining elements over the input shapes via machine learning, we design neural networks to predict end-to-end pointwise aesthetics maps directly.

## 3. Methods

### 3.1. Shape Representation

3D shapes are composed of a series of surface patches. We represent each 3D shape with the patch centers and the patch-based features.

After all the shapes have been aligned and normalized, we first sample points on the shape surface uniformly using Poisson disk sampling with a fixed grid size $d = 0.004$. Then we sample $N = 1024$ points from each dense point cloud by farthest point sampling as the patch centers. We generate surface patches by growing from each center until the geodesic radius reaches a threshold $r = 0.07$ of the shape's bounding box diagonal. Here the radius $r$ is decided based on experience from early patch-based works [HLK*17, RXCW19], which should be local but large enough to provide meaningful perceptual information. Motivated by the style-learning works [LKS15, HLK*17], we use a feature collection that contains both low-level geometric information and high-level perceptual information. As shown in Figure 2, to represent the surface patch, we use geometric histograms such as point-feature histogram ($3 \times 11$ bins) [RMBB08], spin image (64 bins) [JH99], and pairwise distance distribution (32 bins) to encode low-level geometric information. Moreover, we compute each point's principal curvatures and saliency [SLT13] for the dense point cloud, and then we use histograms of points' curvature distribution ($4 \times 16$ bins) and saliency distribution ($3 \times 16$ bins) within the patch to encode high-level perceptual information. In this way, each 3D shape is represented by the patch centers and the concatenation of a series of patch-based histograms.

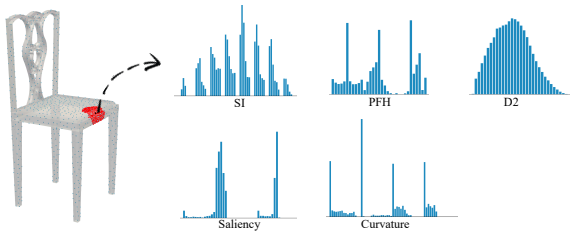


**Figure 2:** *We use a series of patches to represent a 3D shape where each patch is encoded with its patch center and several patch-based histograms. Here the blue points over the surface are patch centers.*

### 3.2. Network Architecture

Given a 3D shape, we design a network to output the global aesthetic score of the 3D shape and the local aesthetic scores of the patches. The core idea of the network design is to enhance the connection between the local information and global information. In the feature embedding, we use several network structures to aggregate each patch to its neighbors and the whole shape; and in the aesthetic predicting, we use the weights-shared predictors to facilitate a local-global related aesthetics metric.

Figure 3 shows the architecture of our network, and it is inspired by the framework of Point Cloud Transformer [GCL*21]. The network's input is the sparse patch centers with the patches' encoded features. First, we use two EdgeConv layers [WSL*19] to aggregate the local attributes with the neighbors further. By concatenating the output (a 64D feature) of each EdgeConv layer, we obtain the new embedded patch feature. Then the feature is fed into four Offset-Attention (OA) layers [GCL*21] and each layer yields a 128D feature. A fully-connected layer [QSMG17] further fuses the concatenated OA features and obtains the local features $f_l \in \mathbb{R}^{N\times 512}$. And the global feature $f_g \in \mathbb{R}^{512}$ is obtained after the max-pooling operation. We expect our learned aesthetic features are general and regular enough, so we follow the idea of variant auto-encoder (VAE) [KW14] and map the global aesthetic feature to a latent space. More concretely, we assume the latent variable follows the Gaussian distribution, and we use two fully-connected layers to predict the mean $\mu \in \mathbb{R}^{512}$ and the standard deviation $\sigma \in \mathbb{R}^{512}$ respectively. During training, we sample the latent code $z = \mu + \sigma \times n$, where $n \in \mathbb{R}^{512} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is the random noise. Then the sampled feature will be fed into the aesthetics prediction network (MLPs) to output a global aesthetic score $s_g$. To predict the local patches' scores $s_l$, we embed the local features into the same latent space and then feed them to the same aesthetics prediction network. The advantage of using weights-shared networks instead

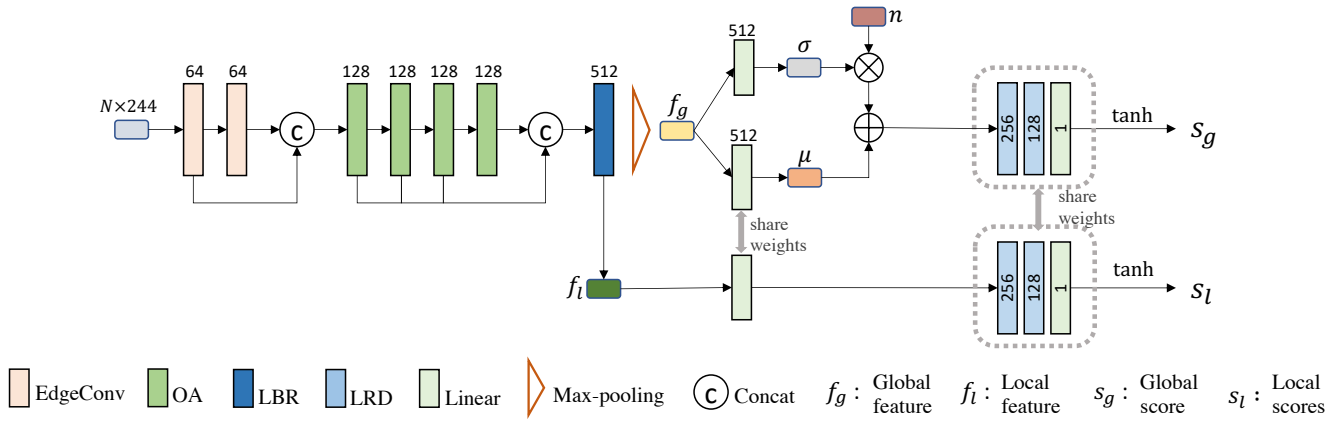


**Figure 3:** *Overview of our network architecture for aesthetics predictions. Given the patch-based represented input, our framework embeds the global and local features into the same latent space and outputs the scores with weights-sharing aesthetics predictors. Here LBR combines Linear, BatchNorm and ReLU layers; LRD combines Linear, BatchNorm and Dropout layers.*

of an additional aesthetic predictor will be demonstrated in Section 4.4.

### 3.3. Loss Function

We propose the following loss functions according to our training dataset which is in the form of pairwise global comparisons: each data sample consists of a sorted pair of 3D shapes, where the former shape is labeled more aesthetic than the latter shape by a user.

#### 3.3.1. Comparison Loss from Global Predictions

Similar to the method from Dev and Lau [DL22], we also consider the margin loss to constrain the pairwise comparison. However, different from their constant margin $m$, we propose a dynamic margin weighted by "user's agreement".

Supposing that there are $n_1$ $(A,B)$ pairs and $n_2$ $(B,A)$ pairs in the dataset $\mathcal{S}$, we define users' agreement on the unsorted pair $\{A,B\}$:

$$a_{\{A,B\}} = \frac{|n_1 - n_2|}{n_1 + n_2} \tag{1}$$

As humans have various perceptions of aesthetics and some shape pairs do not have a high difference in aesthetics, the users' agreement of the pairwise comparisons varies among the pairs. Considering that a high agreement indicates that $\{A,B\}$ is likely to have a high aesthetics difference, and a low agreement indicates that $\{A,B\}$ is likely to have a low aesthetics difference, we propose the dynamic margin loss as follows:

$$\mathcal{L}_g = \frac{1}{|\mathcal{S}|} \sum_{(A,B)\in\mathcal{S}} \max(0, a_{\{A,B\}} \cdot m - (s_{gA} - s_{gB})) \tag{2}$$

#### 3.3.2. Comparison Loss from Local Predictions

We aim to have our network be able to make good "local" aesthetics predictions, even though the dataset has "global" shape aesthetics comparisons. To further build a bridge between the global and local aesthetics, we make an assumption that if a shape $A$ is more aesthetic than $B$, then $A$ is likely to have more attractive patches than $B$. That is, the average local aesthetic score $\overline{s_{lA}}$ is likely to be greater than $\overline{s_{lB}}$. Therefore, we also define the comparison loss from local predictions:

$$\mathcal{L}_l = \frac{1}{|\mathcal{S}|} \sum_{(A,B)\in\mathcal{S}} \max(0, a_{\{A,B\}} \cdot m - (\overline{s_{lA}} - \overline{s_{lB}})) \tag{3}$$

Please note that it is not easy to use other functions such as "variance" instead of "the average" to bridge the global and local aesthetics, as it is hard to assume that if a shape A is more aesthetic than B, then A is likely to have higher (or lower) local aesthetic variance than B. While "the average" looks naive, the evaluations in Section 4.3 show it is enough to make a good performance.

#### 3.3.3. KL-divergence Loss

As proposed in the vanilla VAE model [KW14], the latent feature $z = \mu + \sigma \times n \in \mathbb{R}^{512}$ follows a standard Gaussian distribution with the assumption of independence of each dimension. Therefore, we add a Kullback-Leibler divergence loss as the constraint:

$$\mathcal{L}_{KL-div} = \frac{1}{2} \sum_{i=1}^{512} (\mu_i^2 + \sigma_i^2 - 1 - \log(\sigma_i^2)) \tag{4}$$

#### 3.3.4. Network Training

The total loss is a weighted sum of the loss functions described above:

$$\mathcal{L} = \mathcal{L}_g + \lambda_1 \cdot \mathcal{L}_l + \lambda_2 \cdot \mathcal{L}_{KL-div} \tag{5}$$

We set $\lambda_1 = 0.01$, $\lambda_2 = 4e-4$, $m = 0.5$. We use the Adam optimizer to train the network for 30 epochs, setting the batch size to 48 and the learning rate to follow a cosine annealing decay schedule from $2e-4$ to $1e-6$. All the hyperparameters were jointly decided according to the training performance.

## 4. Results and Analysis

### 4.1. Dataset

We used the dataset provided by Dev and Lau [DL22], which contains 277 chairs, 40 tables, 75 mugs, and 88 lamps with 5100, 2875, 825, and 2500 pairwise aesthetics comparisons respectively. We split the data for each shape category by 8:2 for training and testing (denoted as $\mathcal{I}_t$), and we trained with all the categories together.

Moreover, we conducted an additional user study to augment the "more reliable" subset. Specifically, we randomly chose 30 shapes from each category and invited 30 participants to select the top 5 most beautiful shapes and the top 5 most ugly shapes from the 30 shapes per category. Each user selects top 5 shapes (denoted as $X$), bottom 5 (i.e. most ugly) shapes (denoted as $Y$), and then the remaining 20 shapes are denoted as $Z$. For each category, 225 aesthetics comparison pairs can be generated from $(X,Y)$, $(X,Z)$, and $(Z,Y)$. For each unsorted pair $\{A,B\}$, if it is proposed by at least 15 users, and the agreement is greater than $\frac{1}{3}$, its corresponding sorted pair is recognized as a general and reliable comparison. In this way, we have another 223 chair pairs, 202 table pairs, 199 mug pairs, and 220 lamp pairs for evaluation (denoted as $\mathcal{I}_r$).

### 4.2. Global Aesthetics

We use the accuracy of pairwise comparison on $\mathcal{I}_t$ and $\mathcal{I}_r$ to evaluate the performance of our global aesthetics predictor. We denote the complete network architecture with all the loss functions introduced in Section 3 as the full version, and the network without VAE embedding (so the corresponding loss $\mathcal{L}_{KL\_div}$ is also excluded) and local aesthetics loss $\mathcal{L}_l$ as the naive version. Table 1 shows the quantitative comparison of our method with the method from Dev and Lau [DL22]. Our full version global aesthetics predictor outperforms their work in three object categories on both the original "messy" test set $\mathcal{I}_t$ and the "more reliable" aggregated set $\mathcal{I}_r$. Besides, excluding the VAE structure and the local-related loss does not affect the performance much.

**Table 1:** *Quantitative evaluation of the global aesthetics prediction performance. Our method performs better than the method of Dev and Lau [DL22] for most categories in both the original test set $\mathcal{I}_t$ and the "more reliable" augmentation set $\mathcal{I}_r$.*

| | Method | Chair | Table | Mug | Lamp |
|---|---|---|---|---|---|
| $\mathcal{I}_t$ | Dev and Lau [DL22] | 64.4% | 71.7% | 68.0% | **72.8%** |
| | Ours (full) | **65.5%** | 72.5% | **75.2%** | 66.8% |
| | Ours (naive) | 65.1% | **72.9%** | 74.5% | 66.4% |
| $\mathcal{I}_r$ | Dev and Lau [DL22] | 80.7% | **93.9%** | 79.5% | 79.5% |
| | Ours (full) | **90.6%** | 90.4% | **86.3%** | **90.4%** |
| | Ours (naive) | 89.6% | 90.4% | 85.1% | 89.2% |

### 4.3. Local Aesthetics Maps

To the best of our knowledge, since there are no existing labeled datasets for assessing the local aesthetics maps, we invited 20 users and conducted two user studies to evaluate the quality of our local aesthetics maps.

#### 4.3.1. Pairwise Local Comparisons

Similar to the data collection in the Tactile Saliency work [LDS*16], we asked users to compare pairs of vertices of a 3D shape and rate which vertex (with its local patch) contributes more to the shape's aesthetics. More specifically, we sampled ten shapes from each shape category, for a total of 40 shapes. For each shape, we sampled ten pairs of surface patches. Each user rated one pair from each shape (so each user rated 40 pairs from different shapes in total), and each pair was rated by two users. Hence we collected 800 labeled pairs as ground truth, denoted by $\mathcal{S}_{gt}$.

We compute the accuracy of our predicted local aesthetics scores in two ways. First, we conduct an evaluation directly on the unfiltered 800 pairs. As shown in the second column of Table 2, in the full version method, we achieved 71.8% accuracy. While aesthetics ratings are personal and sometimes users have different opinions about the same pair, there are potential conflicts in the original ground truth. Therefore, we also filter out the pairs which received different answers from their two users. We find there are 244 (i.e. about 60%) non-repeated pairs left, and we denoted them as $\mathcal{S}'_{gt}$. As shown in the third column of Table 2, we achieved a higher accuracy of 85.6% on $\mathcal{S}'_{gt}$.

**Table 2:** *Quantitative evaluation of the local aesthetics prediction performance in different settings. The full version achieves the highest accuracy in the unfiltered pairwise comparison $\mathcal{S}_{gt}$ and the filtered subset $\mathcal{S}'_{gt}$. When we remove some loss function or network structure, the accuracy decreases.*

| Ablation | $\mathcal{S}_{gt}$ | $\mathcal{S}'_{gt}$ |
|---|---|---|
| Full | **71.8%** | **85.6%** |
| w/o $\mathcal{L}_l$ | 70.2% | 83.2% |
| w/o WS | 66.0% | 76.2% |
| w/o VAE | 68.2% | 80.0% |

#### 4.3.2. Global Rating

Since the pairwise comparisons do not cover the shape patches and their relative aesthetics densely, we also invited each user to rate our aesthetics maps directly. Users were shown the aesthetics maps and told that: "red" represents the local surface region contributing to the beauty of the shape, "blue" represents the local surface region contributing to the ugliness of the shape, and the color intensity represents the degree of contribution. Then we asked each user to rate whether they agree with the aesthetics maps on a 7-point scale, from -3 (strongly disagree) to 0 (neutral) to 3 (strongly agree). After each user rated all 40 shapes, we selected two to three shapes that received the lowest scores from that user and conducted a semi-interview to learn more about why the user disagreed with the aesthetics maps.

We visualize the aesthetics maps and the results of the users' ratings in Figures 4 and 5. Qualitatively, as the aesthetic maps in Figure 4 show, local regions with high curvatures can be beautiful (e.g., complex patterns), neutral (e.g., cylinder holders) or ugly (e.g., sharp corners). Our learned aesthetic maps distinguish aesthetics better than the naive curvatures, and the feature collection

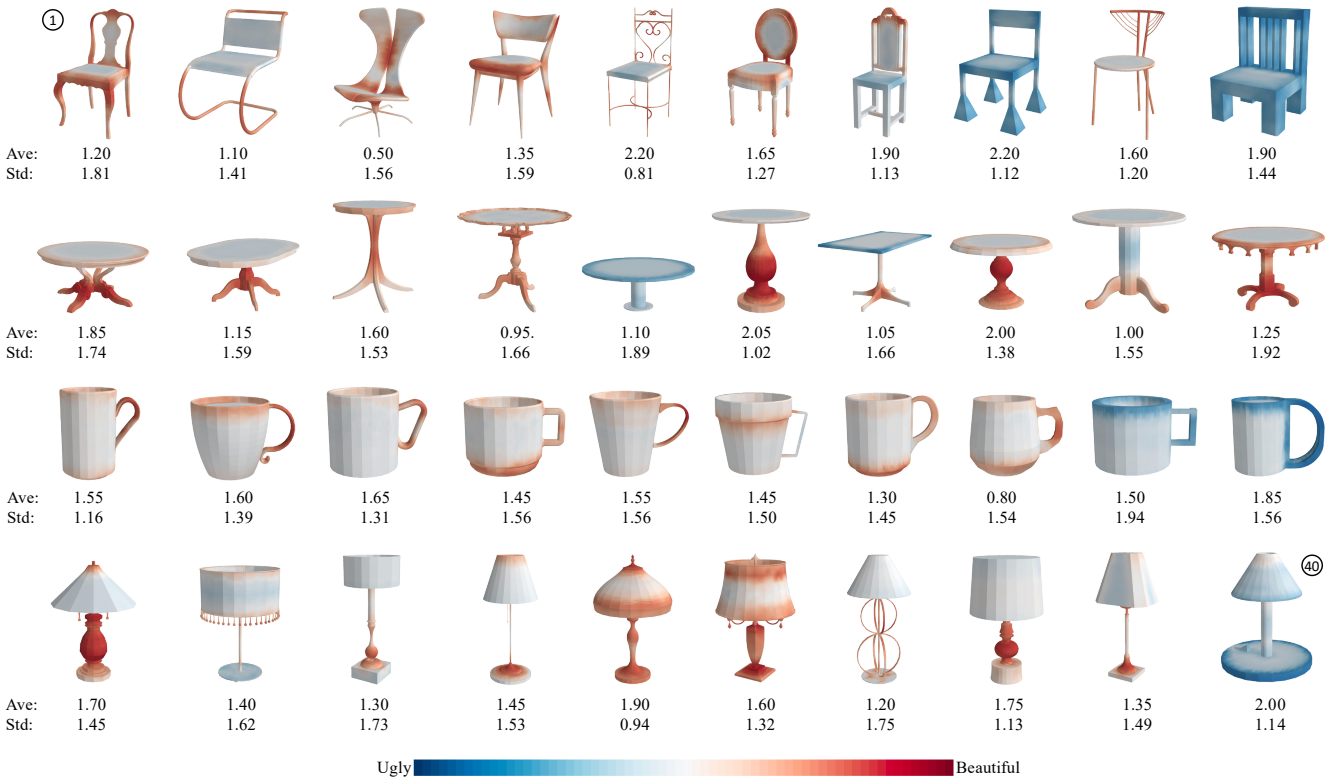


**Figure 4:** *Results of local shape aesthetics maps generated by our method. The shapes are labeled from ID 1 to ID 40, in order from top to bottom, and then left to right. For each shape, we compute the average and standard deviation of the users' ratings (from -3 to 3).*

we mentioned in Section 3.1 jointly helps to make proper predictions. For quantitative evaluation, we compute the average users' ratings and the standard deviation for each shape, based on all the 20 users. As Figure 5 shows, for most shapes, the majority of the users agree with our aesthetics maps. The overall average rating is 1.499, and the standard deviation is 1.531. Also, we can see that, for most shapes, there are always some users proposing negative ratings. This is natural because the perception of aesthetics can vary among different users.

In addition to the quantitative evaluation from the user ratings, we perform a qualitative evaluation from our semi-interviews. We discuss below the reasons for the users' negative ratings. First, each user has his/her personal preferences. One user (U16) mentioned: *"For shape 11, I don't like the curved and complex legs, so I think the legs should be blue ... Also, I don't like the elements like sphere decorations. Instead, I prefer simple designs like that cup (shape 29), and it looks beautiful"*. Second, some users would like to consider global aesthetics criteria when assessing the local aesthetics maps. For example, *"For shape 6, the chair looks rounded and soft overall, so the edge of the back should be blue because it is not soft enough"* (U3) and *"For shape 1, I think the center part of the back should be red as it is the component of the decoration like the edges"* (U9). Third, although we asked the users to rate the maps from the perspective of aesthetics, some users subliminally considered other factors such as functionality. For example, user 5 said *"For shape 37, although the circle decorations around the pillar look okay, I think they will affect my usage of the lamp, so I prefer that area to be blue."*

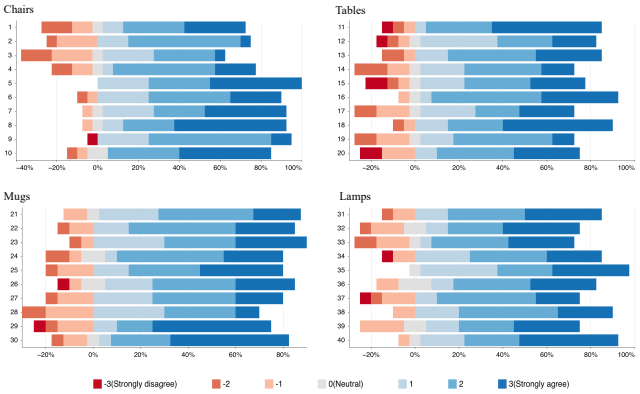


**Figure 5:** *The users' rating distribution for each aesthetics map. The ID corresponds to the shapes in Figure 4. The further the bars extend to the right (relative to zero), the more users agree with our aesthetics maps.*

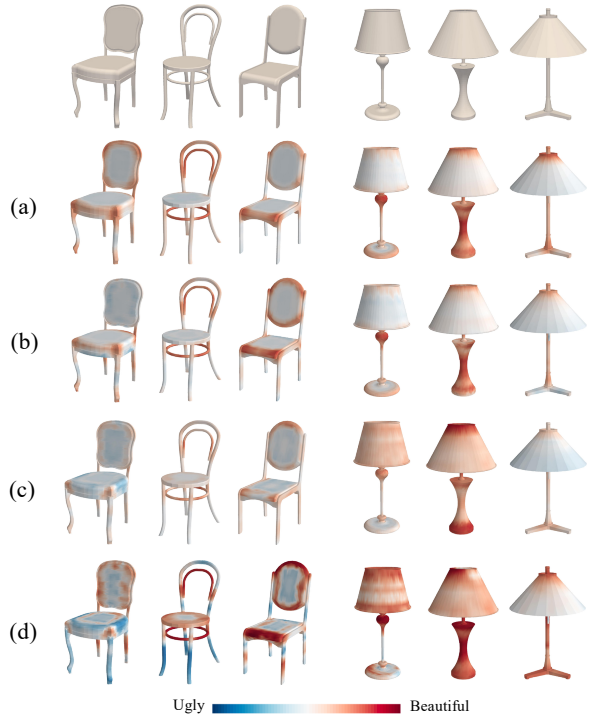


**Figure 6:** *Comparison of the aesthetics maps in different cases. (a) Our full version method; (b) Without local comparison loss; (c) Without VAE; (d) Using independent aesthetics predictor. Overall, the full version achieves the best qualitative performance.*

### 4.4. Ablation Study

Our key idea is to simultaneously learn the global aesthetics metric and the local aesthetics maps via a point-based neural network. In this section, we analyze the effects of our network architecture and loss functions with a series of ablation studies. As we mentioned that the naive version of our method does not affect the global performance much, the following discussion will focus on the local evaluations.

#### 4.4.1. Role of Local Comparison Loss

According to our network architecture design, the global aesthetics feature $f_g$ comes from a dimensional-wise max-pooling operation of the local patches' features $f_l$. Then they are further embedded and used to predict scores with the same network. Ideally, if our network indeed learns the intrinsic criteria of 3D shape aesthetics with the global comparison loss $\mathcal{L}_g$, the aesthetics feature space will be good enough for the local predictions. We conducted experiments without the local comparison loss function $\mathcal{L}_l$ (denoted as w/o $\mathcal{L}_l$), and Table 2 shows the corresponding evaluation accuracy on the local pairwise dataset. From the table, we see that our networks still achieve acceptable accuracy. However, with the local comparison loss as a further constraint, the performance improves. Besides, from the visualization of some aesthetics maps in Figures 6 and 7, we find that some aesthetics-contributing regions are missed.

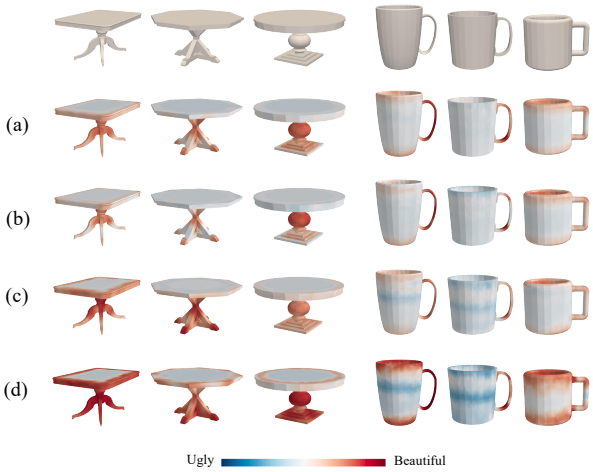


**Figure 7:** *Comparison of the aesthetics maps in different cases. (a) Our full version method; (b) Without local comparison loss; (c) Without VAE; (d) Using independent aesthetics predictor. Overall, the full version achieves the best qualitative performance.*

#### 4.4.2. Weights Sharing

As shown in Figure 3, our global aesthetics predictor shares weights with the local aesthetics predictor. We propose this strategy because of the weak local constraints from the global-based dataset. Our training data are the pairwise global comparisons. To build a connection between local aesthetics and global aesthetics, we assume that if shape A is more aesthetic than B, A will have more attractive patches than B, and propose the corresponding loss $\mathcal{L}_l$. However, this loss is still a weak constraint for each patch, and the assumption itself also fails in some cases. If we use independent networks (denoted as w/o WS) to predict the local aesthetics, overfitting will occur. Quantitatively, as Table 2 shows, the accuracies decrease drastically. Qualitatively, Figure 6 (d) and Figure 7 (d) show that the aesthetics maps will have greater differences between different regions and the smoothness decreases.

#### 4.4.3. Role of the VAE Architecture

Due to the lack of strong local constraints from the global-based data, we proposed to let the local and global predicting networks share weights. This requires the embedded feature space to be general and regular. Therefore, we follow the idea of VAE to map the global and local aesthetics features into the same latent space and assume the embedded latent code follows a Gaussian distribution. We conducted experiments with VAE excluded (denoted as w/o VAE) to evaluate its role. As shown in Table 2, the accuracy is lower than the full version. Moreover, through the visualizations in Figures 6 (c) and Figure 7 (c), we find that the aesthetics maps will become less stable.

## 5. Applications

In this section, we describe several applications enabled by our local shape aesthetics measure and the resulting shape aesthetics maps.

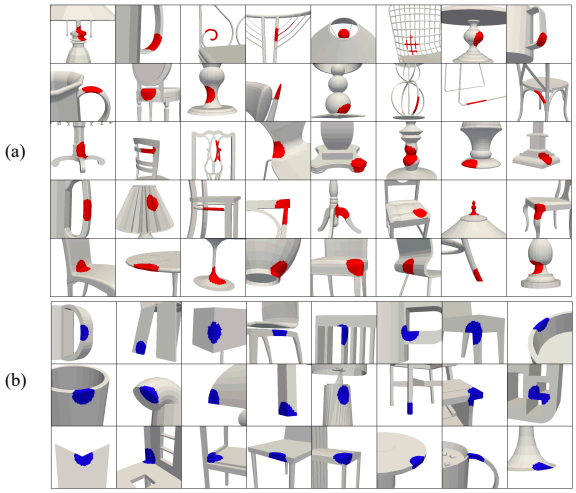


**Figure 8:** *We extract the aesthetic contributing elements (a) and aesthetic suppressing elements (b) with our aesthetics maps. The aesthetic contributing elements are more varied and can come from different parts of a shape, while the aesthetic suppressing elements are more similar.*

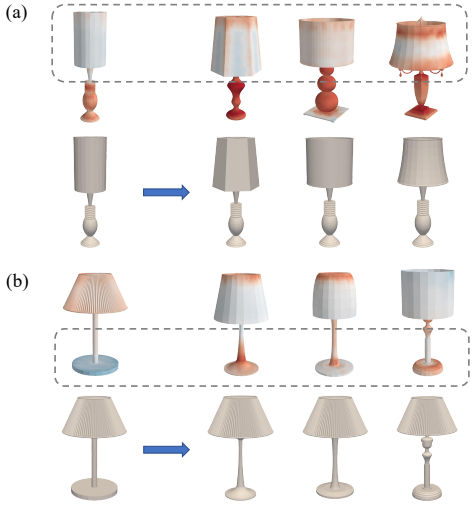


**Figure 9:** *Two examples of our aesthetics-guided shape editing. In case (a), the system detects that the lamp shader is less aesthetic; in case (b), the lamp holder is detected to be less aesthetic. For both cases, similar but more aesthetic corresponding parts are searched from the dataset. Finally, a series of more aesthetic shapes are obtained via substitution.*

### 5.1. Aesthetics-based Patch Galleries

Inspired by related works in shape style [HLK*17, YZX*18] that extract the style-revealing patches over the surface, we show the application of aesthetics-revealing patch galleries. For each shape, we extract the patches whose scores are higher than 0.4 or lower than -0.4 and then do clustering [ZMP04] with the patch-based representation introduced in Section 3.1 to obtain no more than ten representative patches, respectively. While aesthetics can be creative and those mentioned patches can make up a large gallery, we do clustering further for all the categories to find the more general aesthetic-related elements. As shown in Figure 8, we obtain the aesthetic contributing elements such as curves and waves, and the aesthetic suppressing elements such as hard corners. our aesthetics-revealing patch gallery can act as the inspiration and reference for novice designers. Novice designers can learn the typical aesthetics-contributing elements in the existing 3D shapes, and then they can incorporate those elements as they create and edit new shapes. Moreover, designers can learn from the aesthetics suppressing elements and use them less in their own designs.

### 5.2. Aesthetics-driven Subparts Extraction

Combining our aesthetics maps with 3D shape segmentation or cuboid abstraction algorithms, we can extract the aesthetic subparts and build a series of aesthetic-driven sub-datasets. More specifically, we use the existing segmenting method [YC21] to segment each shape into subparts respectively. Then we propose a part-based aesthetic score with our learned aesthetic map for each subpart to extract the top-scoring subset.

We define the part-based aesthetics from the local aesthetics maps with the following considerations. First, If the subpart looks aesthetic overall, it will generally not have many low score patches, which means the average aesthetics over the subpart are generally not low. Second, it is usually a small number of aesthetic patches that make a big global difference. Therefore, to emphasize more about the contributions of the aesthetic patches, we also add the average scores of the top 10% patches into the computation. Finally, we define the part-based aesthetics as the sum of the average of all scores and the average of the top 10% scores in the subpart. Figure 10 shows the aesthetic chair backs, chair seats, and lamp holders extracted automatically from our dataset with the part-based aesthetics.

### 5.3. Aesthetics-guided Shape Editing

Our aesthetics maps are helpful in automatically recognizing the regions that cause a 3D shape to be ugly. This is an important ability if we aim to make a shape more beautiful by making as few changes as possible.

We propose a system based on the local aesthetics maps and global aesthetics metric to facilitate editing shapes to make them more aesthetic (illustrated in Figure 9). Given an input shape, the system detects the relatively less aesthetic part in the whole shape. Then the system suggests more aesthetic parts searched from the dataset according to a weighted consideration of similarity differences [CTSO03] and aesthetic increases. Here the weight is a user-controllable parameter to balance the editing changeability and aesthetics improvement. However, sometimes part substitutions will cause a shape to be globally incompatible. We use our global aesthetics measure to check for these cases and filter them automatically. With our proposed system, users can easily obtain a series of more aesthetic shapes by choosing from the substitution suggestions.

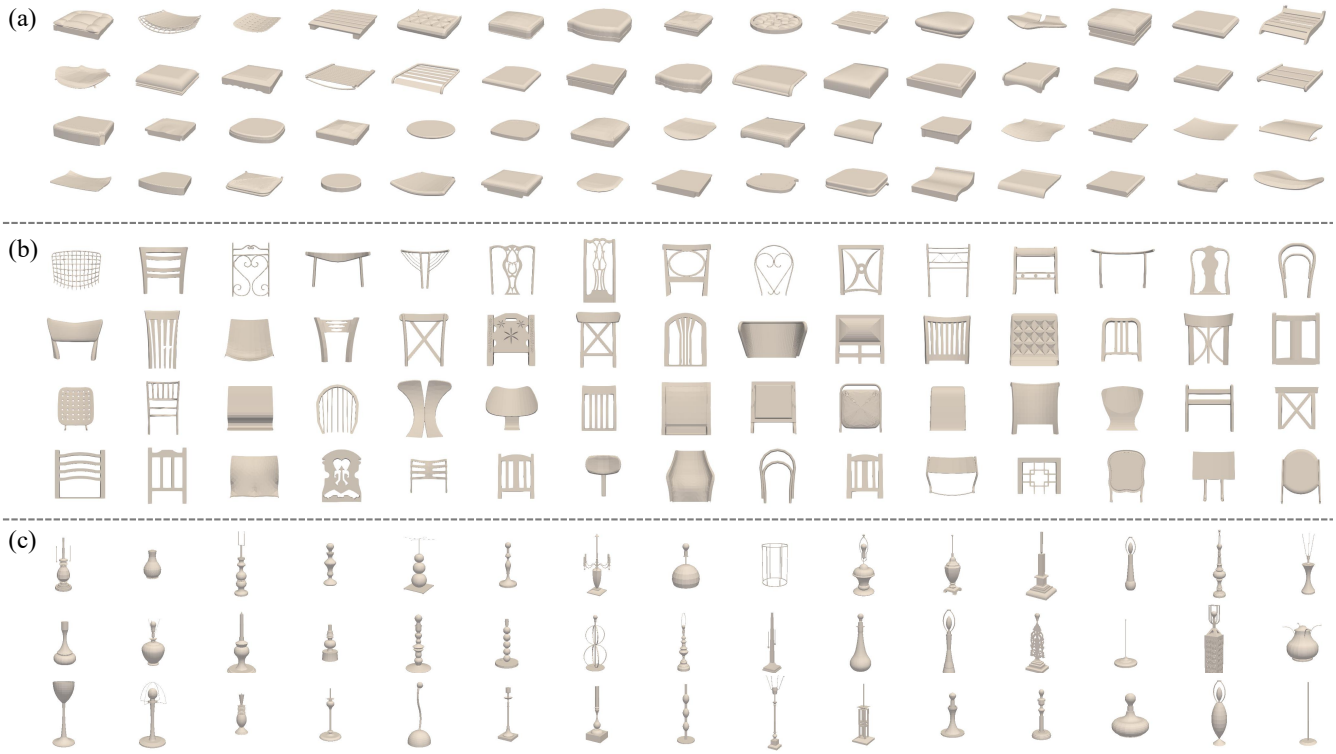


**Figure 10:** *Examples of aesthetic subparts: (a) chair seats, (b) chair backs and (c) lamp holders extracted from the dataset with our aesthetics maps.*

## 6. Limitations and Future Work

We designed a framework to predict the global shape aesthetics score and local shape aesthetics maps via a point-based network architecture, based on user data from global pairwise shape comparisons. While our work was able to achieve a good global aesthetics measure and compute local aesthetics maps on 3D shape surfaces, there are some limitations.

First, we design the network from the perspective of small patches, so our aesthetics map may not reflect the overall shape style or structural disharmony. Sometimes, a shape looks ugly because it violates the visual balance, and the components that are inconsistent with other parts in the overall shape proportion or style negatively contribute to the shape's aesthetic. Figure 11 shows some failure cases where our aesthetics maps fail to highlight the incompatible part. For future research, we can learn the aesthetics metrics hierarchically. For example, we can consider disentangling the aesthetics into two independent branches: the coarse structural level and the fine patch-based level; and then combine them to output an aesthetics map that can capture both the local patterns and global balance.

Furthermore, our method learns the generic global and local aesthetics, and it is not personalized. As discussed in Section 4.3.2, we find that the users' aesthetics preferences sometimes vary by a lot: some users prefer the soft, curved or elaborate designs, while other users prefer the rigid, linear or simple designs. A good intelligent perceptual tool should be personalized, and personalized image aesthetics assessment [RSL*17, ZZL*21] has received increasing attention in the 2D field in recent years. For 3D shapes, we can learn from the algorithms proposed for the images, such as learning a prior generic model first and then fine-tuning with a small number of individually annotated dataset. Compared to the images, the more complex representations and less available annotated datasets make personalized shape aesthetics assessment an interesting but challenging topic.

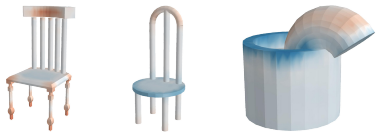

**Figure 11:** *Our aesthetics maps work well in showing the local patch-based aesthetics contributions, but it fails to detect the style incompatibility, proportional imbalance or other structural-based ugliness.*

Additionally, we learn the aesthetics metrics from the 3D geometry represented by surface patches, and we use the patch center with patch-based histograms to encode each patch. While information loss inevitably occurs during the handcrafted encoding, we can perform learning from the dense point cloud directly [BLZ*20].

Finally, our work only considers the geometry of the 3D shapes. However, in real applications, the aesthetics of a 3D shape is also closely affected by other factors such as texture and material. For example, a flat plane can look rugged with some specific normal map. For future work, more data collection and studies are needed to build a more complete aesthetics metric for 3D objects.


## Acknowledgements

We thank the anonymous reviewers for their comments. This work was partially supported by grants from the Hong Kong Research Grants Council (General Research Fund numbers 11206319 and 11205420).


## References


[BLZ*20] BAI X., LUO Z., ZHOU L., FU H., QUAN L., TAI C.-L.: D3feat: Joint learning of dense detection and description of 3d local features. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020), pp. 6359–6367. 9

[BR13] BERGEN S., ROSS B. J.: Aesthetic 3d model evolution. *Genetic Programming and Evolvable Machines 14*, 3 (2013), 339–367. 1, 2

[CSPF12] CHEN X., SAPAROV A., PANG B., FUNKHOUSER T.: Schelling points on 3d surface meshes. *ACM Transactions on Graphics (TOG) 31*, 4 (2012), 1–12. 3

[CTSO03] CHEN D.-Y., TIAN X.-P., SHEN Y.-T., OUHYOUNG M.: On visual similarity based 3d model retrieval. vol. 22, Wiley Online Library, pp. 223–232. 8

[DL22] DEV K., LAU M.: Learning perceptual aesthetics of 3D shapes from multiple views. *IEEE Computer Graphics and Applications 42*, 1 (2022), 20–31. 1, 2, 4, 5

[DLT17] DENG Y., LOY C. C., TANG X.: Image aesthetic assessment: An experimental survey. *IEEE Signal Processing Magazine 34*, 4 (2017), 80–106. 1, 2

[GCL*21] GUO M.-H., CAI J.-X., LIU Z.-N., MU T.-J., MARTIN R. R., HU S.-M.: Pct: Point cloud transformer. *Computational Visual Media 7*, 2 (2021), 187–199. 3

[HLK*17] HU R., LI W., KAICK O. V., HUANG H., AVERKIOU M., COHEN-OR D., ZHANG H.: Co-locating style-defining elements on 3d shapes. *ACM Transactions on Graphics (TOG) 36*, 3 (2017), 1–15. 2, 3, 8

[JH99] JOHNSON A. E., HEBERT M.: Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE Transactions on pattern analysis and machine intelligence 21*, 5 (1999), 433–449. 3

[KW14] KINGMA D. P., WELLING M.: Auto-Encoding Variational Bayes. In *International Conference on Learning Representations (ICLR)* (2014). 3, 4

[LDS*16] LAU M., DEV K., SHI W., DORSEY J., RUSHMEIER H.: Tactile mesh saliency. *ACM Transactions on Graphics (TOG) 35*, 4 (2016), 1–11. 3, 5

[LGK16] LIM I., GEHRE A., KOBBELT L.: Identifying style of 3d shapes using deep metric learning. In *Computer Graphics Forum* (2016), vol. 35, Wiley Online Library, pp. 207–215. 2

[LHLF15] LIU T., HERTZMANN A., LI W., FUNKHOUSER T.: Style compatibility for 3d furniture models. *ACM Transactions on Graphics (TOG) 34*, 4 (2015), 1–9. 2

[LKS15] LUN Z., KALOGERAKIS E., SHEFFER A.: Elements of style: learning perceptual shape style similarity. *ACM Transactions on graphics (TOG) 34*, 4 (2015), 1–14. 2, 3

[LVJ05] LEE C. H., VARSHNEY A., JACOBS D. W.: Mesh saliency. In *ACM SIGGRAPH 2005 Papers*. 2005, pp. 659–666. 3

[MR14] MIURA K. T., RU G.: Aesthetic curves and surfaces in computer aided geometric design. *International Journal of Automation Technology 8*, 3 (2014), 304–316. 1, 2

[Pha99] PHAM B.: Design for aesthetics: interactions of design variables and aesthetic properties. In *Human Vision and Electronic Imaging IV* (1999), vol. 3644, SPIE, pp. 364–371. 1, 2

[PZ03] PHAM B., ZHANG J.: A fuzzy shape specification system to support design for aesthetics. In *Soft Computing in Measurement and Information Acquisition*. Springer, 2003, pp. 39–50. 1, 2

[QSMG17] QI C. R., SU H., MO K., GUIBAS L. J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), pp. 652–660. 3

[RMBB08] RUSU R. B., MARTON Z. C., BLODOW N., BEETZ M.: Learning informative point classes for the acquisition of object model maps. In *2008 10th International Conference on Control, Automation, Robotics and Vision* (2008), IEEE, pp. 643–650. 3

[RSL*17] REN J., SHEN X., LIN Z., MECH R., FORAN D. J.: Personalized image aesthetics. In *Proceedings of the IEEE international conference on computer vision* (2017), pp. 638–647. 1, 9

[RXCW19] REMIL O., XIE Q., CHEN H., WANG J.: 3d shape synthesis via content–style revealing priors. *Computer-Aided Design 115* (2019), 87–97. 3

[Séq05] SÉQUIN C. H.: Cad tools for aesthetic engineering. *Computer-aided design 37*, 7 (2005), 737–750. 1

[SLMR14] SONG R., LIU Y., MARTIN R. R., ROSIN P. L.: Mesh saliency via spectral processing. *ACM Transactions On Graphics (TOG) 33*, 1 (2014), 1–17. 3

[SLT13] SHTROM E., LEIFMAN G., TAL A.: Saliency detection in large point sets. In *Proceedings of the IEEE International Conference on Computer Vision* (2013), pp. 3591–3598. 3

[SXX*18] SHU Z., XIN S., XU X., LIU L., KAVAN L.: Detecting 3d points of interest using multiple features and stacked auto-encoder. *IEEE transactions on visualization and computer graphics 25*, 8 (2018), 2583–2596. 3

[WSL*19] WANG Y., SUN Y., LIU Z., SARMA S. E., BRONSTEIN M. M., SOLOMON J. M.: Dynamic graph cnn for learning on point clouds. *Acm Transactions On Graphics (tog) 38*, 5 (2019), 1–12. 3

[WYA*20] WEISS T., YILDIZ I., AGARWAL N., ATAER-CANSIZOGLU E., CHOI J.-W.: Image-driven furniture style for interactive 3d scene modeling. In *Computer Graphics Forum* (2020), vol. 39, Wiley Online Library, pp. 57–68. 2

[XLZ*10] XU K., LI H., ZHANG H., COHEN-OR D., XIONG Y., CHENG Z.-Q.: Style-content separation by anisotropic part scales. In *ACM SIGGRAPH Asia 2010 papers*. 2010, pp. 1–10. 2

[YC21] YANG K., CHEN X.: Unsupervised learning for cuboid shape abstraction via joint segmentation from point clouds. *ACM Transactions on Graphics (TOG) 40*, 4 (2021), 1–11. 8

[YCHK15] YUMER M. E., CHAUDHURI S., HODGINS J. K., KARA L. B.: Semantic shape editing using deformation handles. *ACM Transactions on Graphics (TOG) 34*, 4 (2015), 1–12. 2

[YZX*18] YU F., ZHANG* Y., XU K., MAHDAVI-AMIRI A., ZHANG H.: Semi-supervised co-analysis of 3d shape styles from projected lines. *ACM Transactions on Graphics (TOG) 37*, 2 (2018), 1–17. 3, 8

[ZMP04] ZELNIK-MANOR L., PERONA P.: Self-tuning spectral clustering. *Advances in neural information processing systems 17* (2004). 8

[ZZL*21] ZHU H., ZHOU Y., LI L., LI Y., GUO Y.: Learning personalized image aesthetics from subjective and objective attributes. *IEEE Transactions on Multimedia* (2021). 1, 2, 9