Enhancing the Aesthetics of 3D Shapes via Reference-based Editing

MINCHAN CHEN, City University of Hong Kong, China
MANFRED LAU, City University of Hong Kong, China

Fig. 1. Given an input shape, our framework automatically produces a set of aesthetically enhanced shapes via editing the input according to some reference shapes that are aesthetic. There are four sets of examples here. The inputs are in blue, the reference shapes are in orange, and the beautified shapes are in yellow.

While there have been previous works that explored methods to enhance the aesthetics of images, the automated beautification of 3D shapes has been limited to specific shapes such as 3D face models. In this paper, we introduce a framework to automatically enhance the aesthetics of general 3D shapes. Our approach employs a reference-based beautification strategy. We first performed data collection to gather the aesthetics ratings of various 3D shapes to create a 3D shape aesthetics dataset. Then we perform reference-based editing to edit the input shape and beautify it by making it look more like some reference shape that is aesthetic. Specifically, we propose a reference-guided global deformation framework to coherently deform the input shape such that its structural proportions will be closer to those of the reference shape. We then optionally transplant some local aesthetic parts from the reference to the input to obtain the beautified output shapes. Comparisons show that our reference-guided 3D deformation algorithm outperforms existing techniques. Furthermore, quantitative and qualitative evaluations demonstrate that the performance of our aesthetics enhancement framework is consistent with both human perception and existing 3D shape aesthetics assessment.

CCS Concepts: • **Computing methodologies** → *Perception*; *Shape modeling*.

Additional Key Words and Phrases: perception, shape aesthetics, 3D shape beautification, reference-guided editing

ACM Reference Format:
Minchan Chen and Manfred Lau. 2024. Enhancing the Aesthetics of 3D Shapes via Reference-based Editing. *ACM Trans. Graph.* 43, 6, Article 279 (December 2024), 15 pages. https://doi.org/10.1145/3687954

Authors' addresses: Minchan Chen, City University of Hong Kong, Hong Kong, China, mcchen7-c@my.cityu.edu.hk; Manfred Lau, City University of Hong Kong, Hong Kong, China, manfred.lau@gmail.com.

0730-0301/2024/12-ART279 $15.00
https://doi.org/10.1145/3687954

1 INTRODUCTION

The concept of visual aesthetics is intricately intertwined with daily human experience, and objects of aesthetic appeal have the power to catch human gazes effortlessly. Since analyzing and enhancing the aesthetics of existing objects is a time-consuming task that may require specific knowledge from artists or designers, being able to automatically beautify 3D shapes is desirable. In the past, researchers have achieved much success in the automatic beautification of two-dimensional images [Deng et al. 2018; Islam et al. 2017; Leyvand et al. 2008; Yan et al. 2016]. There has also been work in the establishment of quantifiable aesthetics metrics [Datta et al. 2006; Deng et al. 2017; Nishiyama et al. 2011; Yang et al. 2022; Zhu et al. 2021]. However, fewer works have been proposed to enhance the aesthetic appearance of three-dimensional models. While automatic 3D aesthetics assessment is achieved via rudimentary handcrafted aesthetic-related features [Bergen and Ross 2013] or state-of-the-art data-driven deep learning techniques [Chen and Lau 2022; Dev and Lau 2022], the automated beautification of existing 3D shapes is restricted to specific domains such as 3D faces [Arsalane et al. 2022; Liao et al. 2012; O'Toole et al. 1999], 3D layouts [Xu et al. 2019] and 3D sketch strokes [Machuca et al. 2018]. To the best of our knowledge, no existing work automatically modifies the shape of general 3D models with the goal of enhancing their aesthetics. Our approach is data-driven as we collect 3D shape aesthetics data. Hence the high-level meaning of "beautification" from our view is that our output shapes will generally be similar to shapes that have been identified as aesthetic by users, and each output shape will have a higher aesthetic score compared to the corresponding input shape.

It is meaningful to take existing ugly 3D shapes and beautify them, since novice modellers are likely to produce unpleasing results (for example, with poor global proportions or lacking local details) if they were to attempt to beautify the shapes themselves. In addition, modellers usually wish to make the shapes more aesthetic while keeping some of the original shapes' attributes. Therefore, in this

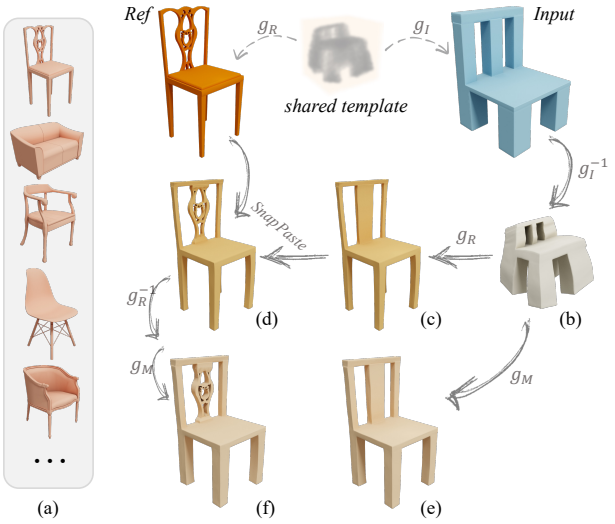

Fig. 2. High-level Overview of our Framework. We build a 3D shape aesthetics dataset (a) and then perform reference-based editing to generate beautification results. We first co-consider the input shape and a reference shape that is aesthetic, and disentangle them into a joint neural template and their own deformation flows g_I and g_R. The input shape is then deformed with its inverse deformation flow to obtain a template (b). We then apply g_R to obtain a globally adjusted shape (c). Furthermore, we may optionally perform local substitution (d) via blending some aesthetic regions of the reference shape with the globally adjusted shape. If the beautified shape looks far from the input, we may also perform shape morphing with the blended flow g_M to produce additional results (e) and (f).

paper, we propose a framework to automatically beautify general 3D shapes. The key idea is to follow a reference-based beautification strategy, as shown in Figure 1. More specifically, we take some 3D models from ShapeNet [Chang et al. 2015], and collect 3D shape aesthetics data to build a shape aesthetics sub-dataset. We then use this sub-dataset to guide the editing of an input shape, and make it look more like some target reference shapes in the aesthetics sub-dataset. Since the reference shape has been judged to be aesthetic by users, and the output shape will have global and local features of the reference, the output is expected to be aesthetic.

Our idea is inspired by the success of image beautification with rich well-labeled aesthetics-related datasets [Kong et al. 2016; Sun et al. 2017]. In contrast with image aesthetics, however, the advancement of 3D shape aesthetics assessment and enhancement has been hindered by the limited availability of 3D aesthetics-related datasets. Existing large 3D datasets [Chang et al. 2015; Mo et al. 2019] are annotated with low-level semantic information, such as categorization and segmentation. Furthermore, high-level semantic labels, such as style [Hu et al. 2017], tactile saliency [Lau et al. 2016], and other descriptive attributes [Yumer et al. 2015], are typically collected for only a small number of 3D models. Recently, Dev and Lau [2022] introduced a learning-based aesthetics metric, trained on a crowd-sourced shape aesthetics dataset, for general 3D shapes. However, their dataset presents several limitations in both aesthetic quantity and structural variation. For example, the dataset only includes data for 277 "dining chairs" (which typically feature one back, one seat, and four legs) from ShapeNet, while chairs with more varying structures are not taken into consideration. Therefore, there remains a significant demand for well-labeled aesthetics-related 3D datasets. To address this gap, we performed data collection to collect aesthetics ratings for some 3D shapes to construct an aesthetics sub-dataset. We gathered aesthetic ratings for 3D shapes of lamps, airplanes, cars, chairs, and tables, and we obtained hundreds of aesthetic shapes for each category. Figure 2a and Figure 3 show several example aesthetic shapes with variant structures.

Given an input 3D model, our approach automatically applies global deformations and local substitutions to enhance its aesthetics by leveraging our aesthetics sub-dataset, as illustrated in Figure 2. The approach is novel in that it co-considers an input 3D shape and an aesthetic reference 3D shape, computes a joint neural template for them, and applies sequences of forward, inverse, and blended deformation flows. This allows us to start with an input, convert between the template space and shape space, and generate new output shapes that are aesthetic and resemble the input.

Our approach involves identifying suitable reference shapes in the aesthetics sub-dataset, and editing the input shape to make it look more like these aesthetic references while preserving its inherent attributes. Inspired by the concept of Neural-Template[Hui et al. 2022], a topology-shape-disentangled framework designed for 3D shape reconstruction, we use a topology-aware searching method and introduce a reference-guided editing framework. Specifically, Neural-Template disentangles each input shape into a topology-aware template and a topology-preserving deformation flow which will deform the template to reconstruct the input. To identify suitable reference shapes, we take our aesthetics sub-dataset and search in Neural-Template's latent topology space to find the aesthetic shapes whose templates are closest to the input. In other words, the found references have similar topology structures to the input. We then edit the input shape globally and locally to make it look more like the reference shape.

To perform global editing, for each (input, reference) pair, we first re-disentangle the input and reference into a joint neural template and their own deformation flows (Figure 2). Then the input is deformed with its inverse deformation flow to obtain an explicit template (Figure 2b), which is then deformed by the reference's deformation flow. Through this process, the input is globally and coherently deformed such that its global proportions will approximate the reference (Figure 2c). See Section 3 for more details about the joint neural template and deformation flows.

While an ugly shape can be aesthetically enhanced globally, its local geometries may also improve its aesthetics. Hence, we further optionally perform local substitution and transplant some local aesthetic regions from the reference shape to the input shape. These local regions can be pre-labelled manually when we collect the aesthetic shapes. Also, they can be extracted automatically with previous works such as aesthetic maps [Chen and Lau 2022]. As the input and reference shapes are well-aligned after global deformation, we can achieve these transplants via Snappaste [Sharf et al. 2006] (Figure 2d). In addition, if we wish the output shape to still resemble the input shape even as it tries to imitate the reference shape, we may optionally perform shape morphing (Figure 2e,f) for the edited

shapes, via blending the deformation flows of the input and reference shapes. See Section 3.6 for more details of this shape morphing step.

We evaluate the performance of our 3D shape aesthetics enhancement framework through both user studies and state-of-the-art 3D shape aesthetics assessment [Chen and Lau 2022]. We also compare our beautification results with text-guided geometry manipulation works [Gao et al. 2023; Ma et al. 2023; Michel et al. 2022] using the prompt "beautiful" to demonstrate that current large language models based methods cannot deal with high-level abstract descriptions. In addition, we compare our reference-guided 3D deformation method with the original Neural-Template method [Hui et al. 2022] and other target-driven deformation techniques [Liu et al. 2021; Yifan et al. 2020] to demonstrate that our approach yields superior results from the perspective of reference-guided deformation.

In summary, our work makes these contributions:

- We collect 3D shape aesthetics data and build an aesthetics sub-dataset. The sub-dataset can be used to augment existing datasets for 3D shape analysis such as our shape beautification problem.
- We introduce a novel reference-based editing framework to enhance the aesthetics of existing 3D shapes. To the best of our knowledge, our framework is the first to beautify general 3D shapes automatically. We demonstrate the effectiveness of our framework with evaluations and comparisons.
- Our approach is novel in that it co-considers the input and reference 3D shapes, computes a joint neural template, and applies sequences of forward, inverse, and blended deformation flows.

2 RELATED WORK

Our work enhances the aesthetics of 3D shapes by editing them based on aesthetic reference shapes. In this section, we discuss previous works in: aesthetics assessment and beautification, reference-guided 3D deformation, and semantic-driven shape editing.

2.1 Aesthetics Assessment and Beautification

While aesthetics is a highly abstract and theoretical concept in the domains of philosophy and art, researchers in computer graphics and vision endeavour to investigate computational metrics for the explicit evaluation of aesthetics.

For 2D images, early works [Datta et al. 2006; Li et al. 2010; Nishiyama et al. 2011] extracted handcrafted features related to visual and composition attributes to explicitly model the commonly established photographic rules. More recent works [Liu et al. 2020; Lu et al. 2015; Ren et al. 2017; Yang et al. 2022; Zhu et al. 2021] focused on end-to-end learning of aesthetics, facilitated by the accelerated progress of neural networks and the substantial expansion of labeled image datasets. Correspondingly, with the supervision of the established aesthetics metrics or aesthetics-revealing features, general photo beautification is enabled through color enhancement [Aydın et al. 2014; Deng et al. 2018; Yan et al. 2016] and composition adjustment [Guo et al. 2018; Islam et al. 2017; Liu et al. 2010]. Moreover, geometric manipulation and conditional generative models are also applied to the beautification of portrait photos [Diamant et al. 2019; Leyvand et al. 2008; Li et al. 2015; Xiao et al. 2020] and sketch drawings [Fišer et al. 2016; Shen 2021; Zitnick 2013].

In the 3D domain, early works [Pham 1999; Pham and Zhang 2003] proposed exploring the interactions between aesthetic characteristics and linguistic or parametric design variables in 3D designs via empirical investigations. Furthermore, mathematical models were proposed for 3D aesthetic designs by extending the existing aesthetics criteria for images to quantifiable 3D geometric features [Bergen and Ross 2013] or formulating aesthetic curves and surfaces with well-defined splines [Miura and RU 2014]. More recently, Dev and Lau [2022] built a shape aesthetics dataset and proposed a cross-category end-to-end 3D geometry aesthetics assessment via multi-view image-based neural networks instead of using handcrafted features. Based on that shape aesthetics dataset, Chen and Lau [2022] proposed a patch-based learning framework to further predict global aesthetics scores and local aesthetics maps for 3D shapes simultaneously. Taking advantage of 3D geometries and 2D textures, Beauty3DFaceNet [Xiao et al. 2021] enabled 3D facial attractiveness prediction via training deep convolutional neural networks. Although these works focused on 3D aesthetics assessment, enhancing the aesthetic appeal of 3D shapes has received relatively less attention. Existing works explored aesthetic enhancement for specific cases, such as face models [Arsalane et al. 2022; Liao et al. 2012; O'Toole et al. 1999], 3D layouts [Xu et al. 2019], or 3D sketch strokes [Machuca et al. 2018], with the supervision of predefined geometric and spatial constraints related to symmetry, angles, proportions, alignment, and primitives. To the best of our knowledge, no existing work proposed automatic aesthetics enhancement for the geometry of general 3D shapes, and our reference-based beautification framework fills this gap. We employ a reference-based strategy rather than predetermined constraints such as normalization used in facial enhancement [O'Toole et al. 1999] because the aesthetics of general objects has more diversity in global topology and detailed geometry. Moreover, although no existing 3D aesthetics metrics were taken into consideration in our framework, the evaluations demonstrate that our results are consistent with both human perception and the metrics introduced in previous works.

2.2 Reference-guided 3D Deformation

The problem of reference-guided 3D deformation involves the process of warping a 3D source shape to a target 3D shape while maintaining certain attributes of the source shape. It typically requires "as rigid as possible" preservation of local identity with the aim of pose transfer [Gao et al. 2018; Song et al. 2021, 2023], or local geometric modification is allowed for better shape/style transfer. Our focus will be on the latter, and we discuss the works from the perspectives of generation-based and displacement-based approaches.

The generation-based strategies follow the idea of embedding the source and target data into a shared latent feature space, conducting manipulation in the latent space, and subsequently reconstructing the output through the decoder networks. In terms of network architecture, the shape representation formats are typically voxels [Wu et al. 2019], unordered point clouds [Segu et al. 2020; Wang et al. 2022b], or implicit representations [Chen and Zhang 2019; Deng et al. 2021; Park et al. 2019]. While these works may be adequate for

tasks such as style transfer or topology interpolation, we believe that they are unsuitable for our purpose of enhancing the aesthetics of existing 3D shapes. This is due to the inherent imprecision associated with the "encoding" operation. For example, the complex detailed geometry and the delicate surface patterns could be missing in the decoded results.

The displacement-based strategies seek to displace the source vertices under the consideration of satisfying the desired target. The displacements of the vertices can be accomplished either directly [Groueix et al. 2019; Wang et al. 2019] or indirectly through the use of grid-based [Hanocka et al. 2018], cage-based [Yifan et al. 2020], primitive-based [Xu et al. 2010; Yin et al. 2021], and on-surface [Liu et al. 2021] controllers. A challenge in this context is to strike a balance between the degrees of freedom (DoF) associated with the parameters that need to be resolved. A high DoF [Wang et al. 2019] can lead to shape or structural anomalies, while a low DoF [Yifan et al. 2020] may not be able to cope with flexible deformations. Our reference-guided deformation framework is inspired by Neural-Template [Hui et al. 2022] which decomposes each shape into its own template and deformation flow. We recompose the input and reference shapes into a joint neural template, and then use their deformation flows to guide the displacements of the vertices. We demonstrate that our framework outperforms Neural-Template [Hui et al. 2022] and other leading works [Liu et al. 2021; Yifan et al. 2020].

2.3 Semantic-driven Shape Editing

Semantic-driven shape editing refers to the idea of editing a given input shape according to some high-level semantic information without any reference shapes.

Early works proposed to establish their own metrics to guide the corresponding semantic-driven editing. Yumer et al. [2015] formulated score functions for descriptive attributes such as "comfortable" and "durable", and then global deformation was performed with the supervision of attribute scores. Hu et al. [2017] proposed to learn the style-defining local elements of 3D shapes and then style-driven modeling (e.g. transferring a Japanese chair into a Ming chair) was achieved with the substitution of local elements. Our work does not establish specific aesthetics metrics to guide the editing, as aesthetics is a highly diverse and abstract concept when compared to these attributes or styles. Instead, we perform aesthetics-driven editing indirectly via reference shapes.

With the rapid development of large language models (LLMs), recent works are interested in performing shape editing with the guidance of texts. Most of them [Gao et al. 2023; Hong et al. 2022; Ma et al. 2023] rely on the pretrained vision-language models such as CLIP [Radford et al. 2021]. Typically, these works require a concrete description as prompt and abstract conceptions such as aesthetics are more difficult for them. Our experiments show that we outperform these text-guided works in addressing the 3D shape beautification problem.

3 3D SHAPE BEAUTIFICATION

3.1 Shape Aesthetics sub-Dataset

We construct a shape aesthetics sub-dataset by collecting aesthetics perceptual data from users. We take shapes from ShapeNet [Chang et al. 2015] in the categories of chairs, lamps, airplanes, cars, and tables. Some 3D shapes exhibit flaws such as missing faces or are misclassified. Therefore, we pre-reviewed all the shapes and manually filtered the unqualified models for each category. As a result, 6600 chairs, 2100 lamps, 3900 airplanes, 3450 cars and 7600 tables are retained. Then the remaining models in each category were randomly divided into groups with 50 shapes for the data collection.

Fig. 3. Examples of 3D shapes in our aesthetics sub-collection.

To carry out the user studies for the data collection, we invited participants on our campus to rate the visual aesthetics of the 3D models. Specifically, Each user was presented with a group of 50 models one by one. For each 3D model displayed on the screen, participants were encouraged to manipulate the model by rotating and scaling it to gain a comprehensive observation. Then, participants were requested to choose the extent to which they agreed that the 3D model is aesthetic on a 5-point Likert scale: strongly disagree, disagree, neutral, agree, and strongly agree.

While the users' aesthetics preferences can vary significantly, we aim to construct a dataset that captures the common opinion of the users. To achieve this, we employed a strategy whereby each participant only rated a group of models and each group of models was evaluated by three participants. Subsequently, we gathered the models that were positively rated (i.e., agree and strongly agree) by all three participants into the aesthetics sub-collection. Conversely, shapes that received at least two negative ratings (i.e., disagree and strongly disagree) were designated as test shapes. After performing this step, there are 415 chairs, 253 lamps, 352 cars, 518 airplanes and 537 tables in the aesthetics sub-collection. Figure 2a and Figure 3 show some examples of the aesthetic shapes.

3.2 Preliminary

The searching of reference shapes and the reference-guided deformation algorithms are developed based on Neural-Template [Hui et al. 2022]. We briefly overview it for a better comprehension of the following sections.

Neural-Template is a framework proposed for the reconstruction of high-quality 3D meshes from voxel-based or image-based inputs. It disentangles the reconstruction problem into two steps: learning a topology-aware neural template and performing topology-preserving 3D deformation. Specifically, Neural-Template first encodes the input into a topology code $\boldsymbol{t} \in \mathbb{R}^{128}$ and a shape code

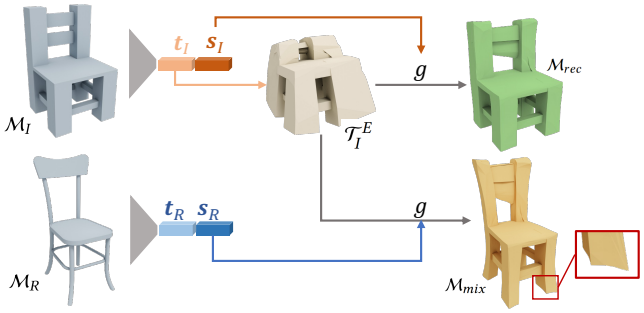

Fig. 4. Neural-Template's application of shape mixing. Although the mixed shape imitates the reference's proportions, it is far from appealing: the back area is incorrectly reconstructed, the legs are still thick and the bottom of the legs cannot touch the ground stably because of distortion.

$\boldsymbol{s} \in \mathbb{R}^{128}$. Then the topology code $\boldsymbol{t}$ is fed into the topology formulation module f to induce an implicit template $\mathcal{T} = f(\cdot, \boldsymbol{t})$, which should be close to 0 when the query point is inside the template and close to 1 otherwise. And the shape code $\boldsymbol{s}$ is fed into a deformation module g to predict a deformation flow $g(\cdot, \boldsymbol{s})$, which computes the deformed position of a query point. Finally, the reconstruction $\mathcal{M}$ can be implicitly obtained by applying the deformation flow g to the template $\mathcal{T}$, i.e. $\mathcal{M} = g(\mathcal{T}, \boldsymbol{s})$.

One of the important designs of Neural-Template is that the topology formulation module f is implemented with BSP-Net [Chen et al. 2020], so the implicit neural template $\mathcal{T}$ can be easily transformed into a collection of convex polyhedra, resulting in an explicit representation with vertices and faces, i.e., $\mathcal{T}^E = (V^{\mathcal{T}}, F^{\mathcal{T}})$. The other important design is that the deformation module g is implemented with the neural ordinary differential equation (NODE) [Gupta and Chandraker 2020; Jiang et al. 2020; Yang et al. 2019]. In this way, the deformation flow $g(\cdot, \boldsymbol{s})$ is a diffeomorphism between the template space and the reconstructed shape space, leading to the effective embedding of topology information in the topology code $\boldsymbol{t}$. In our work, we utilize topology codes to identify appropriate reference shapes, and leverage the deformation flows to edit an input shape such that it looks more like each reference shape.

3.3 References Acquisition

We enhance the aesthetics of a given 3D model by referencing the pre-built shape aesthetics sub-dataset. To achieve this goal, we first need to identity suitable shapes from the sub-dataset that can be easily imitated by the input mesh. If the reference and input shapes have large topological differences, the globally deformed shape may look poor. Hence, our method finds a reference shape for each input. It entails that the input mesh and a reference shape should possess comparable topologies, thereby enabling natural and continuous deformations from the input shape to the target shape without compromising the overall structure of the input.

According to this analysis, we utilize the topology code $\boldsymbol{t}$ in Neural-Template [Hui et al. 2022] as the reference-searching feature. Given an input model $\mathcal{M}_I$ represented as vertices V_I and faces F_I, we first voxelize it and feed it into the pre-trained Neural-Template's encoder to obtain its topology code $\boldsymbol{t}_I$ and shape code $\boldsymbol{s}_I$. The 3D models in our aesthetic sub-dataset were also preprocessed this way. Then the K aesthetic shapes whose topology codes are the top-K nearest to $\boldsymbol{t}_I$ in l_2 distance are found as the reference shapes.

3.4 Reference-guided Global Deformation

After finding the suitable reference shapes, we perform reference-guided deformation for each pair of input and reference shapes. Assuming $\mathcal{M}_R$ is one of the reference shapes, with the precomputed topology code $\boldsymbol{t}_R$ and shape code $\boldsymbol{s}_R$, we introduce a deformation framework to make $\mathcal{M}_I$ look more like $\mathcal{M}_R$ while preserving $\mathcal{M}_I$'s detailed attributes.

The high-level idea is to apply the reference's deformation flow to the input's template. Figure 4 shows Neural-Template's application of shape mixing. As mentioned in Section 3.2, the input's topology code $\boldsymbol{t}$ can be decoded into an explicit template $\mathcal{T}_I^E = (V_I^{\mathcal{T}}, F_I^{\mathcal{T}})$, which is further deformed by its deformation flow $g(\cdot, \boldsymbol{s}_I)$ to produce the reconstruction $\mathcal{M}_{rec}$. As shown in Figure 4, deforming $\mathcal{T}_I^E$ with the reference's deformation flow $g(\cdot, \boldsymbol{s}_R)$ can generate a shape-remixed object $\mathcal{M}_{mix} = (g(V_I^{\mathcal{T}}, \boldsymbol{s}_R), F_I^{\mathcal{T}})$ which imitates the reference's global shape. However, this simplistic utilization does not satisfy our beautification standards. First, the reconstructed explicit template $\mathcal{T}_I^E$ comprises no more than 32 convexes, which leads to reconstruction errors or omissions of details in the final shape. For example, the back areas of the chairs $\mathcal{M}_{rec}$ and $\mathcal{M}_{mix}$ have inaccurate geometry (Figure 4). Second, even though the topology codes $\boldsymbol{t}_I$ and $\boldsymbol{t}_R$ are close, the deformation flow of the reference shape may not be compatible with the input's template. Consequently, the desired outcome cannot be achieved (e.g., the legs of the chair $\mathcal{M}_{mix}$ are still quite thick in Figure 4), potentially resulting in distortions (e.g., the leg corners distort and the bottom parts of the legs do not touch the ground stably).

Given the above issues with applying a naive version of Neural-Template, we propose a novel framework to achieve our reference-guided global deformation, as illustrated in Figure 5. Overall, we re-disentangle the input and reference shapes into one joint template, and subsequently incorporate the reference's deformation flow into the template, while taking into account the geometric constraints. Figure 8 shows some results of our reference-guided beautification, in particular for the global deformation. The inputs are smoothly deformed to obtain a better global proportion. We describe the reference-guided deformation framework in detail in the following sections.

3.4.1 Template Consistency. As the reference's deformation flow is associated with the reference's template, naively incorporating the reference's flow to the input's template can potentially generate unsatisfactory results, particularly if the templates are not well-aligned. Therefore, we introduce a technique to re-disentangle the two shapes to make the input and the reference share the same joint template. More specifically, the input $\mathcal{M}_I$ and the reference $\mathcal{M}_R$ are disentangled into a sharing topology code $\tilde{\boldsymbol{t}}$, with their own shape codes $\widetilde{\boldsymbol{s}_I}$ and $\widetilde{\boldsymbol{s}_R}$. In this way, the input and the reference share a joint neural template $\mathcal{T} = f(\cdot, \tilde{\boldsymbol{t}})$ (an occupancy field), and then deformation flows $g_I = g(\cdot, \widetilde{\boldsymbol{s}_I})$ and $g_R = g(\cdot, \widetilde{\boldsymbol{s}_R})$ map the sharing template into the occupancy fields of $\mathcal{M}_I$ and $\mathcal{M}_R$ respectively.

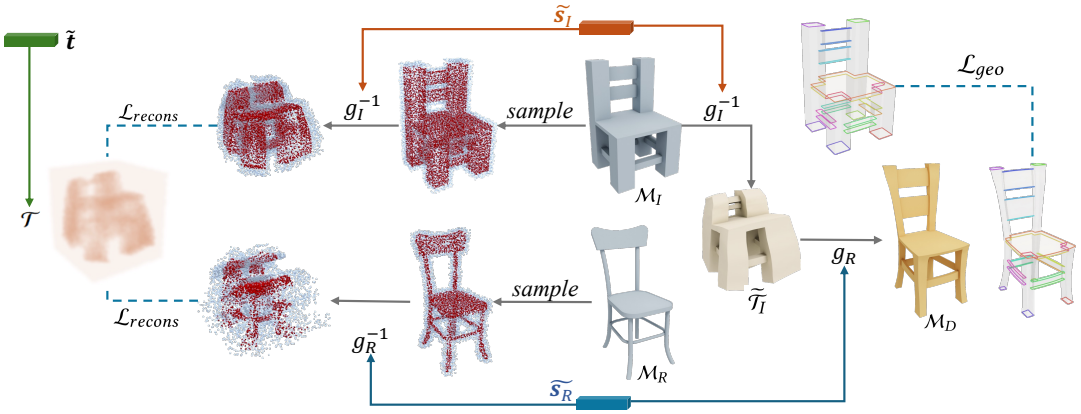

Fig. 5. Our Reference-Guided Deformation Framework. We disentangle the input $\mathcal{M}_I$ and the reference shapes $\mathcal{M}_R$ into a sharing topology code $\tilde{t}$ (which induces a neural template $\mathcal{T}$) and their own shape codes $\widetilde{s_I}$ and $\widetilde{s_R}$ (that guides their deformation flows respectively). During deformation, $\mathcal{M}_I$ is first inversely deformed with $g_I^{-1} = g^{-1}(\cdot, \widetilde{s_I})$ to obtain an explicit template $\widetilde{\mathcal{T}_I}$. Then $\widetilde{\mathcal{T}_I}$ is deformed with $g_R = g(\cdot, \widetilde{s_R})$ to obtain the output $\mathcal{M}_D$. To supervise the disentanglement and the deformation, reconstruction loss $\mathcal{L}_{recons}$ is used to penalize the difference between the joint template and the ground-truth templates. In addition, we apply a curve-based geometric loss $\mathcal{L}_{geo}$ to penalize geometric changes between $\mathcal{M}_I$ and $\mathcal{M}_D$.

To supervise the disentanglement, we penalize the reconstruction loss between the original shapes and the new reconstructions. Specifically, we follow the sampling method from IM-Net [Chen and Zhang 2019] to sample $N = 16384$ points in the shape space and obtain a group of occupancy ground-truth $\{(p_I^i, o_I^i)\}_{i=1}^N$ and $\{(p_R^j, o_R^j)\}_{j=1}^N$ for the input and the reference, which represents if a point p is inside the 3D mesh (with $o = 1$) or not (with $o = 0$). As g is an diffeomorphism between the template space and the original shape space, to predict if a point p is inside the 3D mesh is to predict if the inversely mapped point $g^{-1}(p)$ is inside the template. Therefore, the occupancy prediction of p for the input is $O(p, \mathcal{M}_I) := f(g^{-1}(p, \widetilde{s_I}), \tilde{t})$, and $O(p, \mathcal{M}_R) := f(g^{-1}(p, \widetilde{s_R}), \tilde{t})$ for the reference. Then the reconstruction loss is defined as:

$$\begin{aligned}\mathcal{L}_{recons} &= \mathcal{L}_{occ\_input} + \mathcal{L}_{occ\_ref} \\ &= \frac{1}{N}\Big(\sum_{i=1}^{N}[o_I^i * O(p_I^i, \mathcal{M}_I) + (1 - o_I^i) * (1 - O(p_I^i, \mathcal{M}_I))] + \\ &\sum_{j=1}^{N}[o_R^j * O(p_R^j, \mathcal{M}_R) + (1 - o_R^j) * (1 - O(p_R^j, \mathcal{M}_R))]\Big)\end{aligned} \tag{1}$$

3.4.2 Global Deformations. To incorporate the reference's deformation flow with the input's template, we need to obtain the input's template first. As the shared joint template is for both the input and reference, it ignores the detailed and inherent attributes of the input shape. Therefore, instead of decoding the explicit template from $\tilde{t}$ as Neural-Template [Hui et al. 2022] did, we inversely deform the input $\mathcal{M}_I = (V_I, F_I)$ into the template space via $g^{-1}(\cdot, \widetilde{s_I})$ to obtain the input's new template, i.e:

$$\widetilde{\mathcal{T}_I} = (g^{-1}(V_I, \widetilde{s_I}), F_I) \tag{2}$$

Then, we apply the reference's deformation flow $g(\cdot, \widetilde{s_R})$ to $\widetilde{\mathcal{T}_I}$ to obtain the globally deformed shape:

$$\mathcal{M}_D = g(\widetilde{\mathcal{T}_I}, \widetilde{s_R}) = (g(g^{-1}(V_I, \widetilde{s_I}), \widetilde{s_R}), F_I) \tag{3}$$

As we introduced in Section 3.2, the deformation flow g is implemented with the neural ordinary differential equation (NODE) [Gupta and Chandraker 2020], which has a diffeomorphic nature. Therefore, $\mathcal{M}_I$ is diffeomorphic to $\widetilde{\mathcal{T}_I}$, and $\widetilde{\mathcal{T}_I}$ is diffeomorphic to $\mathcal{M}_D$. Consequently, $\mathcal{M}_I$ is diffeomorphic to $\mathcal{M}_D$. Hence, $\mathcal{M}_I$ is continuously and smoothly deformed into $\mathcal{M}_D$. Therefore, the deformed shape $\mathcal{M}_D$ will preserve the topology of the input shape $\mathcal{M}_I$ regardless of its structural or detailed complexity. For example, the complex topology of the input lamp and rifle shapes in Figure 6 are preserved. Moreover, because the deformation is smooth, the local geometry of the inputs can be preserved (for example, the details in the chair legs and car tires in Figure 6).

3.4.3 Geometric Constraints. The deformed shape $\mathcal{M}_D$ should maintain some attributes of $\mathcal{M}_I$ while it imitates $\mathcal{M}_R$. Considering the fact that normal or Laplacian regulations used in previous works[Liu et al. 2021; Yifan et al. 2020] can be naturally changed during our beautification procedure, we also use curve-based geometric constraints that are inspired by IWIRES [Gal et al. 2009]. We extract locally planar feature loops following the method in IWIRES [Gal et al. 2009]. We first extract sharp edges, chain the edges into a group of loops (i.e., the colorful wireframes in Figure 5), and then define the following geometric constraints for each loop $l = (v_0, v_1, \cdots, v_n)$ from a local level and a global level.

First, we expect that the locally planar loop l should still be locally planar after deformation. Therefore, we penalize the changes of triple product values of continuous edges:

$$\mathcal{L}_{tri} = \frac{1}{n}\sum_{i=0}^{n-1} |e_i \cdot (e_{i+1} \times e_{i+2}) - \widetilde{e_i} \cdot (\widetilde{e_{i+1}} \times \widetilde{e_{i+2}})| \tag{4}$$

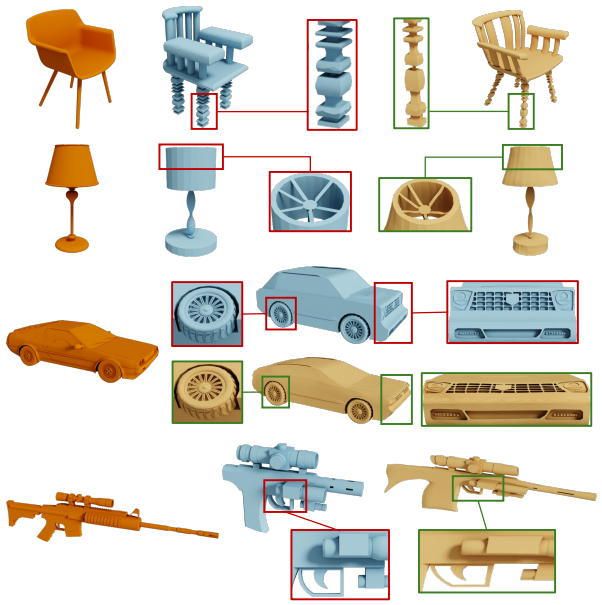

Fig. 6. Examples of results with more complex 3D shapes. These examples show that our reference-guided global deformation can preserve the global topology and the local details. The reference shapes are in orange, the input shapes are in blue, and the deformed shapes are in yellow. Details are shown in the rectangular boxes.

where e_i is the normalized edge $\overrightarrow{v_i v_{i+1}}$ and $\widetilde{e_i}$ is the corresponding edge after deformation. The subscripts are cyclical to n.

Furthermore, we expect that the general contours of the loops will remain relatively unchanged. For example, we hope that circular shapes will be deformed into circular shapes, while rectangular shapes will be deformed into rectangular shapes. So we also penalize the changes of angles:

$$\mathcal{L}_{ang} = \frac{1}{n}\sum_{i=0}^{n-1} |e_i \cdot e_{i+1} - \widetilde{e_i} \cdot \widetilde{e_{i+1}}| \tag{5}$$

While the formulas $\mathcal{L}_{tri}$ and $\mathcal{L}_{ang}$ are defined from a local level, we also define similar losses from a more global level to avoid the accumulation of local errors:

$$\mathcal{L}'_{tri} = \frac{1}{n}\sum_{i=0}^{n-1} |e'_i \cdot (e'_{i+k} \times e'_{i+2k}) - \widetilde{e'_i} \cdot (\widetilde{e'_{i+k}} \times \widetilde{e'_{i+2k}})| \tag{6}$$

$$\mathcal{L}'_{ang} = \frac{1}{n}\sum_{i=0}^{n-1} |e'_i \cdot e'_{i+k} - \widetilde{e'_i} \cdot \widetilde{e'_{i+k}}| \tag{7}$$

where e'_i is the normalized edge $\overrightarrow{v_i v_{i+k}}$ and $k = \lfloor n/6 \rfloor$. And finally, the geometry loss is:

$$L_{geo} = \mathcal{L}_{tri} + \mathcal{L}'_{tri} + \mathcal{L}_{ang} + \mathcal{L}'_{ang} \tag{8}$$

3.4.4 Implementation and Training. In our implementation, we use the same network architecture as Neural-Template [Hui et al. 2022] to implement the template module f and deformation module g.

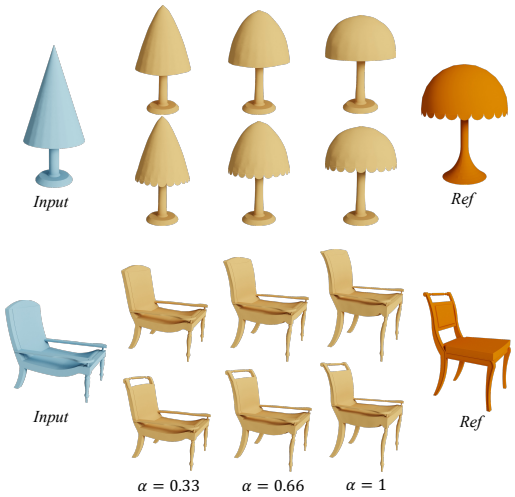

Fig. 7. Examples of Shape Morphing: For each pair of input and reference shapes, the first row shows the morphing with the globally deformed shape, and the second row shows the morphing with further local substitutions.

We solve for the desired network's parameters and latent codes via minimizing the following loss function:

$$\arg\min_{\theta,\tilde{t},\widetilde{s_I},\widetilde{s_R}} \mathcal{L} = \mathcal{L}_{recons} + \lambda \mathcal{L}_{geo} \tag{9}$$

where θ represents the neural parameters of g and f, which is initialized from the pretrained Neural-Template [Hui et al. 2022] model. Besides, we initialize $\tilde{t} = (t_I + t_R)/2$, $\widetilde{s_I} = s_I$, and $\widetilde{s_R} = s_R$. The learning rate for θ is 0.0005 and the learning rate for $\tilde{t}, \widetilde{s_I}, \widetilde{s_R}$ is 0.0025. We set $\lambda = 0.7$ for cars and $\lambda = 0.2$ for the other categories. We run 300 epochs with the Adam optimizer to obtain the deformed shape, and it takes about 8 minutes on a Tesla V100 GPU. Then the globally deformed shape $\mathcal{M}_D$ is obtained according to Equation 3.

3.5 Reference-guided Local Substitution

The aesthetics of shapes can be analyzed from both a broad, global perspective and a more specific, local perspective. In the last section, we perform global deformations to the input such that it looks more like the reference shape in terms of its global structural proportions. As the deformation flow cannot convey delicate local geometry, we may optionally perform local substitutions to transplant some aesthetic patches from the reference shape to the input. Here the aesthetic patches were pre-labelled with the establishment of the shape aesthetics dataset. Additionally, these areas can be automatically identified using techniques such as aesthetic maps [Chen and Lau 2022] or be specified by users.

Assuming Ω is one of the aesthetic local region of the reference shape $\mathcal{M}_R$, we transplant Ω into $\mathcal{M}_D$ automatically via SnapPaste[Sharf et al. 2006]. More specifically, SnapPaste is an interactive tool for mesh composition. If two mesh parts are placed approximately close and they have overlapping snapping regions, SnapPaste can snap the two parts via its soft iterative closest point (Soft-ICP) algorithm. To achieve our local patch transplantation, we need to identify the snapping regions on both the input and reference shapes.

Fig. 8. Results of our aesthetics enhancement. These examples demonstrate the global deformations in particular. The input shapes are in blue, the reference shapes are in orange, and the globally deformed shapes are in yellow.

The snapping region of the reference can be pre-labelled, and then its boundary loops can be detected with orientation. Here the orientation is defined so that the snapping region always lies on the right of the boundary's orientation. As we obtain a good alignment between $\mathcal{M}_D$ and $\mathcal{M}_R$, if the snapping region Ω_s of Ω is predefined with oriented boundary loops, the corresponding snapping region and the corresponding replaced region on $\mathcal{M}_D$ can be obtained via projecting the oriented boundary loops of Ω_s onto $\mathcal{M}_D$'s surface. After snapping regions are decided, Soft-ICP [Sharf et al. 2006] is performed to overlap the snapping regions, and finally the local substitution is achieved.

3.6 Results Morphing

Following the global deformation and local substitution operations, the edited shape will adhere to the reference's overall structural proportions, while also capturing some of its local details. These results are referred to as "full-imitation". As editing becomes more extensive, the input will increasingly resemble the reference, but may lose some of its inherent structural attributes. To meet the user's potential desire to preserve more of the input's original attributes, we also allow for some "semi-imitation" outputs.

As defined in Formula 3, we obtained the globally deformed shape $\mathcal{M}_D$ via deforming input's template $\widetilde{\mathcal{T}_I}$ with reference's deformation flow $g(\cdot, \widetilde{s_R})$. To preserve input's global attribute, we use a blended deformation flow $g(\cdot, s')$, where s' is an interpolation of the input's shape code $\widetilde{s_I}$ and the reference's shape code $\widetilde{s_R}$, i.e., $s' = \alpha \cdot \widetilde{s_R} + (1-\alpha) \cdot \widetilde{s_I}$. More specifically, if no local substitution is applied to $\mathcal{M}_D$, we can directly apply $g(\cdot, s')$ to $\widetilde{\mathcal{T}_I}$ to obtain the morphing shape. Moreover, if $\mathcal{M}_D$ is further edited into $\mathcal{M}_F$ via local substitution, we first inversely deform $\mathcal{M}_F$ via the inverse flow $g^{-1}(\cdot, \widetilde{s_R})$ to obtain it's template, and then apply $g(\cdot, s')$, i.e., the morphing shape is $g(g^{-1}(\mathcal{M}_F, \widetilde{s_R}), s')$. Figure 7 shows some examples of "semi-imitation" morphing. As α gets closer to 0, more of the input's global attributes are preserved.

4 RESULTS AND EVALUATIONS

4.1 Quality of Beautification

To evaluate the performance of our reference-based 3D beautification framework, a total of 90 chairs, 50 lamps, 50 cars, 60 airplanes and 80 tables were randomly selected from the"ugly models" as our testing dataset. The"ugly models" refer to the shapes which have previously received at least two low ratings, which means at least two users disagreed or strongly disagreed with the aesthetic quality of the shapes in our data collection process described in Section 3.1. For each shape, we generated five globally deformed results with the guidance of its top-five reference shapes. Then local substitutions were performed if the found reference was pre-labelled with some aesthetic regions. Table 1 shows the number of shapes we had. For example, we have 450 globally-deformed outputs and 132 further locally-blended outputs for the category of chairs.

Table 1. Number of inputs to be beautified and the generated outputs through our global and local editing techniques.

	Chair	Lamp	Car	Airplane	Table
input	90	50	50	60	80
global	450	250	250	300	400
global & local	132	158	68	89	107

4.1.1 Qualitative Evaluation. Figures 8 and 9 show some of our shape beautification results. As shown in Figure 8, some inputs are not aesthetically pleasing due to their dullness or inappropriate global proportions. Therefore, by making them look more like the reference at a global scale, we are able to enhance their visual appeal. In Figure 9, we further conduct local substitutions for some

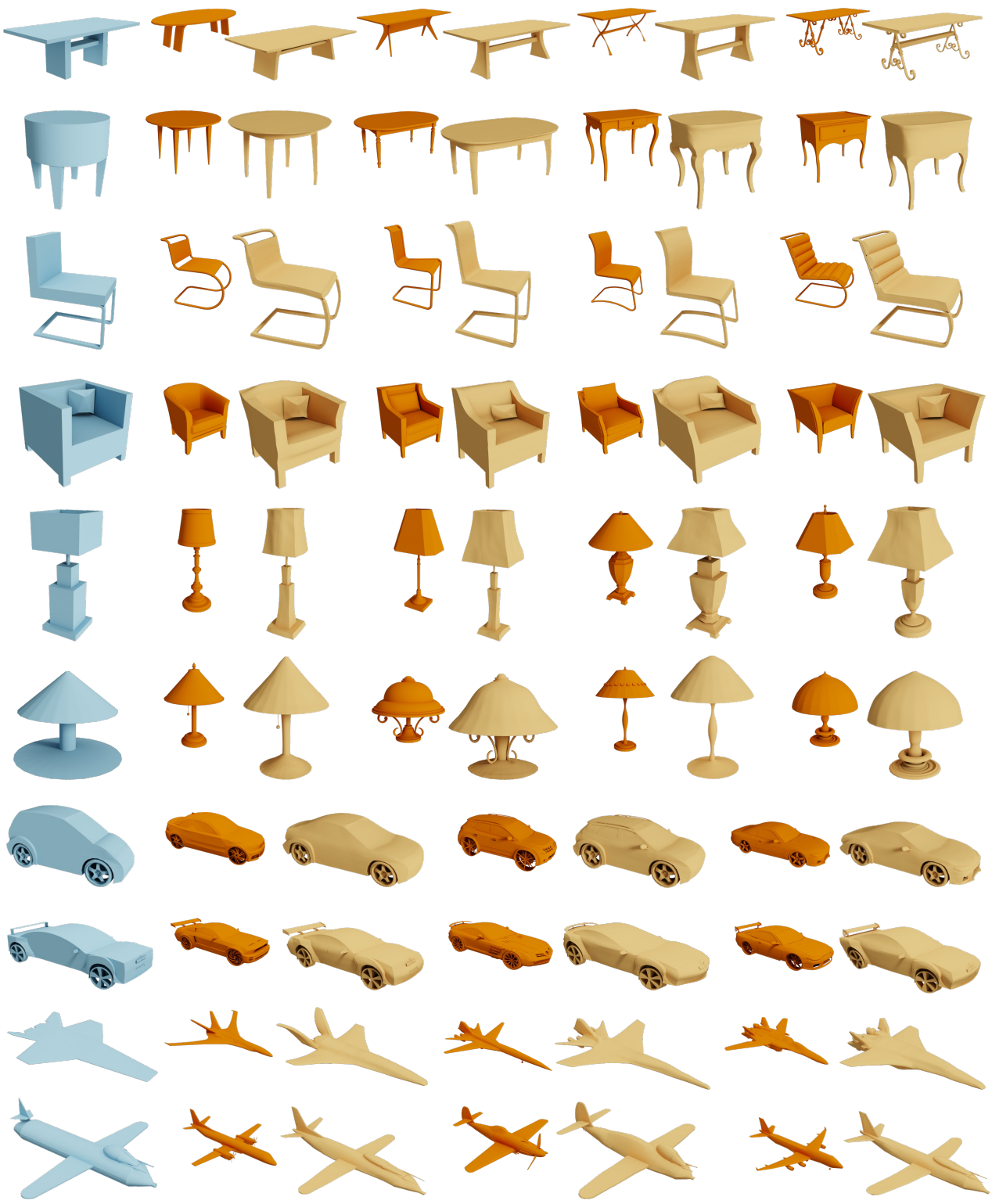

Fig. 9. Results of our reference-based beautification algorithm. Given the input shapes (shown in blue), several aesthetic references (shown in orange) are found. Then global deformations and optional local substitutions are performed to edit the input shapes according to their respective reference shapes. The output shapes are are shown in yellow.

shapes after global deformations, which make the edited shapes more delicate. Overall, the reference-based edited results exhibit a higher level of visual appeal compared to the original inputs.

In addition, we compare the beautification quality of our method with the leading text-guided geometric editing works with the prompt "a beautiful *X*". As depicted in Figure 10, Text2Mesh [Michel et al. 2022] and X-Mesh [Ma et al. 2023] only perform near-surface displacement, which limits their ability to achieve significant global deformations and improve the dull or inappropriate global proportions. Although TextDeformer [Gao et al. 2023] can conduct large-scale deformations, the deformation is not controllable with only "beautiful" as supervision. While our reference-based beautification strategy can allow for both large global editing (such as the chair and lamp in Figure 10) or slight global editing (such as the airplane in Figure 10), TextDeformer [Gao et al. 2023]'s deformation results are not visually pleasing. It is worth noting that current text-guided geometric editing works struggle with abstract tasks, such as beautification, due to limitations in vision-language models like CLIP [Radford et al. 2021] that cannot parse abstract concepts.

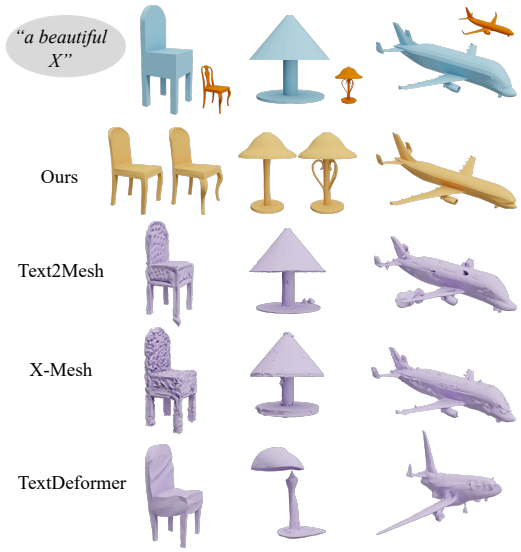

Fig. 10. Comparison of beautification quality to the text-guided geometric editing techniques: Text2Mesh [Michel et al. 2022], X-Mesh [Ma et al. 2023] and TextDeformer [Gao et al. 2023]

4.1.2 Quantitative Evaluation. We quantitatively evaluate the quality of our beautification work via a previous learning-based aesthetic metric [Chen and Lau 2022] and an additional user study.

Chen and Lau [2022]'s aesthetic metric was trained on four categories (chairs, lamps, cups, and tables) with a limited number of shapes and structures. Therefore, we only used their metric to evaluate the aesthetics scores for our chairs, lamps and tables. After applying global deformations (Table 2), 71.8% of the deformed chairs, 78.4% of the deformed lamps and 72.5% of the deformed tables have higher aesthetic scores than their input shapes. After further local substitutions, 87.9% of the edited chairs, 90.5% of the edited lamps and 89.7% of the edited tables have higher aesthetic scores than their deformed shapes.

Table 2. Percentage of 3D shapes whose aesthetics scores [Chen and Lau 2022] increase after editing.

	Chair	Lamp	table
global	71.8%	78.4	72.5%
global & local	87.9%	90.5	89.7%

In addition, we conducted a user study to evaluate if our algorithm effectively improves the appearance of the input shapes. Specifically, we randomly selected 12 shapes from each category, resulting in a total of 48 shapes. For each input shape, we randomly chose three results based on three different references, leading to three (input, output) pairs. We then invited 24 users to select which shape they found more beautiful from each pair, with each user rating only one pair from each shape. Thus, each participant rated 48 pairs of different shapes. In total, we obtained 144 pairs, and each pair was rated by 8 participants. Our statistical analysis, as presented in Figure 11, indicates that, across all categories, the majority of shape pairs achieved an apparent enhancement, with at least 6 out of 8 participants perceiving the output shape as more aesthetically pleasing.

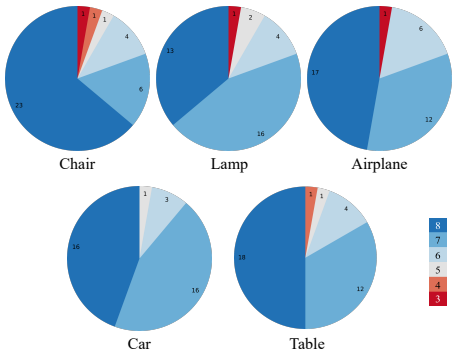

Fig. 11. Shape pairs' ratings distributions. The legend represents "for each pair, how many participants out of eight think the edited shape is more beautiful", from 8 (dark blue) to 3 (dark red). The pie charts count the rating distribution of pairs. For example, for the Chair category, there are 23 pairs for which all of the eight users think that the output shape is more beautiful.

4.2 Quality of Target-driven Global Deformation

To evaluate the quality of the beautification, we also conduct evaluation of our target-driven global deformation framework. We compare our results to Neural-Cage [Yifan et al. 2020] which proposed cage-based global deformation and DeepMetaHandles [Liu et al. 2021] which achieved global deformation via the combinational deformations of surface key points. We also compare with the shape-mixing results of Neural-Template [Hui et al. 2022]. As Neural-Cage [Yifan et al. 2020] and DeepMetaHandles [Liu et al. 2021] focused on "chair", "table" and "car", we perform comparisons on these three categories.

Fig. 12. Qualitative comparison of our method with other methods: (DM) DeepMetaHandles [Liu et al. 2021], (NC) Neural-Cages [Yifan et al. 2020] and (NT) Neural-Template [Hui et al. 2022]. Our approach yields superior results: DeepMetaHandles's and Neural-Cage's deformed shapes still remain much different from the target shapes at a overall level. Besides, Neural-Template's convex-compositional results display much roughness and it cannot convey delicate details such as the bottom areas of the shapes in column 5, the leg areas of the shapes in column 7, and the side-curves of the shapes in columns 8 and 9.

In our evaluation, we primarily care about the source-target pairs that have comparable topological structures. In addition to the 450 chair pairs, 250 car pairs, and 400 table pairs we used in Section 4.1, we further randomly sampled 200 chair pairs, 250 car pairs, and 250 table pairs. Those pairs have shapes whose topology distances (i.e. the l_2 distance of the pre-computed topology latent codes) are within the nearest 50 shapes to each other.

Figure 12 shows the qualitative results. Our method achieves better visual outcomes compared to prior deformation works [Liu et al. 2021; Yifan et al. 2020] according to the following aspects. First, Neural-Cages [Yifan et al. 2020] and DeepMetaHandles [Liu et al. 2021] are restricted by the degrees of freedom for deformation. Their results look more close to the "axis-aligned" scaling, and the deformed shapes still remain much different from the target shapes at a global level. Second, their outputs exhibit a lower level of "man-made" quality compared to ours. Specifically, the seating area and tabletop area are no longer flat, the wheels of the car are no longer circular, and the areas that are supposed to touch the ground no longer touch the ground. As for Neural-Template [Hui et al. 2022], although its outputs match the targets on a global scale, the convex-compositional shapes exhibit greater roughness when compared to the original shapes. It also cannot deal with delicate shape transferring. For example, the bottom areas of the table in Figure 12 column 5 and the leg areas of the table in Figure 12 column 7 are not transferred well. In addition, the side curves of the cars in Figure 12 columns 8-9 are not effectively conveyed.

Inspired by Neural-Cages [Yifan et al. 2020] and DeepMetaHandles [Liu et al. 2021], for quantitative evaluation, we utilized Chamfer distance (CD, computed over 100,000 uniformly sampled points) between the deformed shape and the target shape to measure the alignment error, and we used the differences in cotangent Laplacians (CotL) to measure distortion. Please note that "CotL" and "CD" cannot directly stand for the quality of beautification. "CD" measures the difference of the deformed shape compared to the reference, and smaller "CD" values correspond to smaller differences. "CotL" measures the distortion of the deformed shape compared to the original input, and smaller "CotL" values correspond to smaller distortions. Table 3 shows the quantitative results. Compared to other methods [Hui et al. 2022; Liu et al. 2021; Yifan et al. 2020], our approach achieves the lowest alignment errors. Besides, our "CotL" error is higher than Neural-Cages [Yifan et al. 2020] and DeepMetaHandles [Liu et al. 2021]. However, we illustrate that it is normal because our results achieve better alignment and larger distortions are needed to deform the input into the reference. Moreover, the higher "CotL" errors do not mean that our approach produces unnatural or incoherent deformations. "CotL" error only measures the distortion of the deformed shape compared to the original input,

and it cannot measure if the deformed shape has visually unpleasing distortions. As Figure 12 shows, our inputs are deformed in a natural and smooth manner to look more like and approximate the target shapes, leading to a greater difference from the original source shapes and consequently yielding a higher "CotL" error.

Table 3. Chamfer distance (CD) and cotangent Laplacians error (CotL) comparison between different methods: (DM) DeepMetaHandles [Liu et al. 2021], (NC) Neural-Cages [Yifan et al. 2020] and (NT) Neural-Template [Hui et al. 2022]. The units of CD and CotL are both 10^{-3}.

	Chair		Table		Car	
	CD	CotL	CD	CotL	CD	CotL
DM	3.407	0.344	16.851	0.331	1.115	0.208
NC	4.290	0.127	10.964	0.279	1.412	0.474
NT	1.292	–	1.321	–	0.630	–
Ours	**0.867**	0.697	**0.723**	0.632	**0.624**	4.677

4.3 Ablation Study

To perform reference-guided global deformation, our key idea is to disentangle the input shape and the reference shape into a joint neural template and their own deformation flows, with the supervision of the reconstruction loss and geometric constraints. In this section, we analyze the effects of our framework designs with the ablation studies. First, we compare the deformation results with the basic Neural-Template flows. That is, we map the vertices of the input shape to the template domain using its inverse flow g_I^{-1} and apply the reference's deformation flow g_R. Here the flows are deduced from the original Neural-Template networks without any optimization. As Figure 13a shows, the deformed results cannot transfer the references' global shapes precisely, and unpleasing distortions occur. In other words, it has the shortcomings of both "only without using joint neural template" (Figure 13b) and "only without geometric constraints" (Figure 13c). Next, we describe these ablation studies in more detail.

4.3.1 Joint Neural Template. In order to explore the role of utilizing a joint neutral template between the input and the reference, we conducted experiments (denoted as "w/o ST") in which we disentangled the input and the reference into two independent templates (referred to as two independent topology codes). The network was then trained with reconstruction loss and geometric loss as we described in Section 3.4.4.

Qualitatively, as Figure 13b shows, the deformation cannot precisely deform the input's global shape towards the reference. For example, the areas of the car trunk, back legs, table bottom, and lampshade are not well deformed into the references' shapes. This is because the reference's deformation flow cannot be applied effectively with the input's template if the input's template is not close to the reference's template. Also, we computed "CD" and "CotL" errors as we did in Section 4.2. Table 4 shows the statistical results. All of the "CD" errors increase when using independent templates, which quantitatively demonstrates that the ability of the target-driven shape transferring will decline without using a joint neural template.

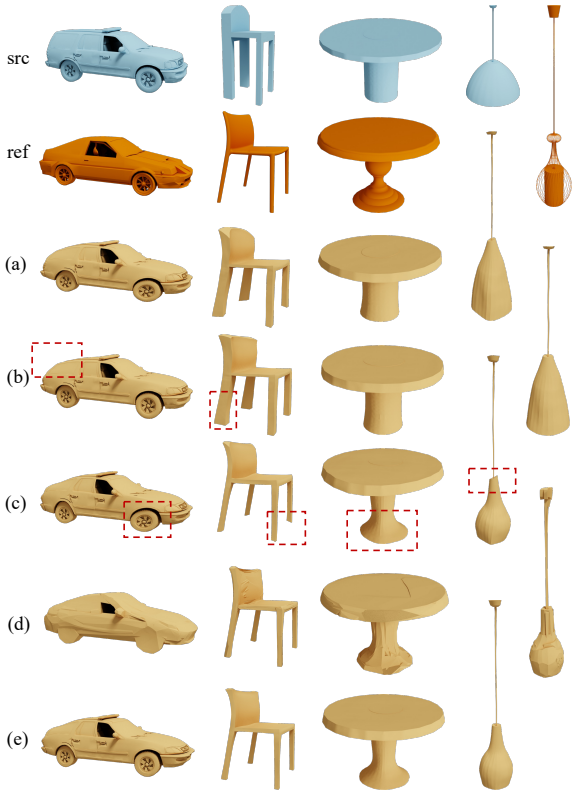

Fig. 13. Comparison of reference-guided global deformation results in different cases: (a) basic Neural-Template's flows; (b) without shared template; (c) without geometric constraint; (d) without inverse deformation; (e) full version. Overall, the full version achieves the best qualitative performance.

Table 4. Chamfer distance (CD) and cotangent Laplacians error (CotL) comparison between different ablation studies. The units of CD and CotL are both 10^{-3}.

	w/o ST		w/o $\mathcal{L}_{geo}$		Full	
	CD	CotL	CD	CotL	CD	CotL
Chair	1.319	0.618	0.862	0.721	0.867	0.697
Table	1.411	0.504	0.712	0.682	0.723	0.632
Car	0.716	3.310	0.629	5.14	0.624	4.677
Lamp	1.193	0.581	1.125	0.632	1.113	0.607
Airplane	0.702	1.278	0.588	2.018	0.590	1.851

4.3.2 Geometric Constraints. In order to explore the role of the geometric constraints, we conducted experiments to train the network without the supervision of geometric loss (denoted as "w/o $\mathcal{L}_{geo}$"), i.e. we set $\lambda = 0$ in Formula 9.

Qualitatively, as Figure 13c shows, there are distortions in the wheels of the cars or the bottom areas of the chairs and tables. Without the geometric constraints, the circular shapes are likely to lose their circularity and the flat planes are likely to lose their flatness. Quantitatively, as Table 4 shows, all of the "CotL" errors increase without the supervision of the geometric loss. While the

Chamfer distances do not change much without $\mathcal{L}_{geo}$, larger "CotL" errors indicate a higher probability of the undesirable distortions.

4.3.3 Inverse Deformation. In our global deformation framework, after learning the shared topology code $\tilde{t}$ and the deformation flows, we obtain the input's template via inversely deforming the input mesh with its learned deformation flow, instead of reconstructing a shared explicit template from $\tilde{t}$. This strategy helps to keep the detailed geometries of the input shape, because the shared template may ignore details to accommodate both the input and the reference. As Figure 13d shows, without using inverse deformations, detailed geometries such as the car wheels or the patterns on the car's body will appear uneven or be incorrect.

4.4 Generality of our Framework

Although we built the aesthetics dataset and conducted evaluations on five shape categories, our reference-based 3D shape aesthetics enhancement framework can be adapted to more categories. More specifically, given an input shape, if a structurally-similar aesthetic reference shape is provided, we can perform beautification with our algorithm. Figure 14 shows some examples.

Fig. 14. Generality of our Framework: Examples of reference-based beautification for a cabinet, rifle, display and ship. The inputs are in blue, the references are in orange and the outputs are in yellow.

5 DISCUSSION, LIMITATIONS, AND FUTURE WORK

While our reference-based 3D shape aesthetics enhancement framework enables the automatic beautification of general 3D shapes via making an input shape look more like or imitate some beautiful reference shapes, it has the following limitations.

First, we acknowledge that our framework's ability to enhance the aesthetics of 3D models is constrained by the inherent limitations of the aesthetics sub-dataset, which is a common challenge encountered by data-driven or reference-based techniques. If an input shape is ugly and has unusual global structure (such as in Figure 15a) and no suitable references can be found in our aesthetic dataset, the reference-guided global beautification will fail. Furthermore, the detail-preserving global deformation will also preserve some irregular parts (such as the highlighted cone-shaped part in Figure 15b). In addition, the beautification may not work well on some sub-categories of shapes. For example, there is no truck in the set of aesthetic cars. As Figure 15c shows, the beautification will fail if we attempt to make the truck follow the SUV's shape (and the result in the case is a mixture of the truck and SUV). Therefore, larger

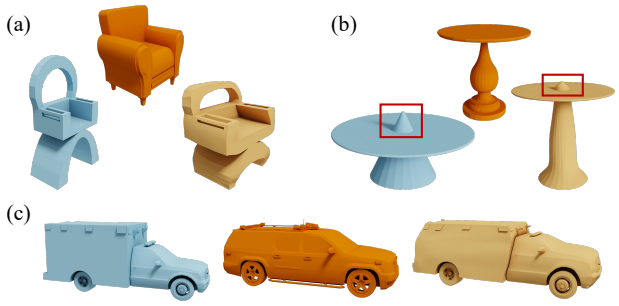

Fig. 15. Failure cases of our beautification framework. The inputs are in blue, the references are in orange and the outputs are in yellow.

aesthetic-related datasets and semantic-aware aesthetic supervision are needed. In the future, we will collect additional aesthetics-related ratings for a broader range of 3D models to augment our shape aesthetics sub-dataset. Additionally, we aim to develop more comprehensive 3D shape aesthetics metrics through the expanded dataset, and then metric-guided end-to-end beautification can be explored.

Second, our method needs 8 minutes to beautify each input shape. However, our method is intended for offline beautification of an input 3D shape, and it works well for this purpose. Eight minutes is acceptable compared to the time a user would spend manually to edit an ugly model to make it more beautiful.

Third, our framework will not be as effective if the reference shapes that are found are not more aesthetically pleasing than the input shapes. Consequently, it is unable to handle inputs that are already visually appealing. By allowing the input to imitate multiple references simultaneously, rather than just one at a time, this issue may be resolved. However, this presents a challenging problem as it necessitates a more comprehensive analysis of 3D aesthetics and 3D blending. Additionally, beyond imitating existing shapes, learning aesthetics-conditional generative neural networks for 3D shapes is also a challenging but potential problem. Future work may study these unexplored issues.

Finally, the current work is exclusively devoted to beautifying a 3D model by means of modifications to its geometric shape, with no regard for its textural information. In practical applications, texture is an inseparable part of a complete model. Current leading works [Hong et al. 2022; Ma et al. 2023; Wang et al. 2022a] generate textured 3D shapes with the rapid development of differentiable rendering techniques or representations [Bangaru et al. 2022; Chen and Zhang 2019; Mildenhall et al. 2021; Wang et al. 2021]. Theoretically, the manipulation of the texture of an existing model will affect its aesthetics qualities. Therefore, in future works of aesthetics enhancement, we expect that both texture and geometry will be simultaneously considered and modified.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their comments. This work was supported in part by the Research Grants Council of the Hong Kong SAR, China (CityU 11206319 and 11205420), and in part by the Chow Sang Sang Group Research Fund/Donation.

REFERENCES

Zarghili Arsalane, Majda Aicha, Oufkir Ayat Allah, et al. 2022. 3D facial attractiveness enhancement using free form deformation. *Journal of King Saud University-Computer and Information Sciences* 34, 6 (2022), 3497–3505.

Tunç Ozan Aydın, Aljoscha Smolic, and Markus Gross. 2014. Automated aesthetic analysis of photographic images. *IEEE transactions on visualization and computer graphics* 21, 1 (2014), 31–42.

Sai Praveen Bangaru, Michaël Gharbi, Fujun Luan, Tzu-Mao Li, Kalyan Sunkavalli, Milos Hasan, Sai Bi, Zexiang Xu, Gilbert Bernstein, and Fredo Durand. 2022. Differentiable rendering of neural sdfs through reparameterization. In *SIGGRAPH Asia 2022 Conference Papers*. 1–9.

Steve Bergen and Brian J Ross. 2013. Aesthetic 3D model evolution. *Genetic Programming and Evolvable Machines* 14 (2013), 339–367.

Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. 2015. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012* (2015).

Minchan Chen and Manfred Lau. 2022. Learning 3D Shape Aesthetics Globally and Locally. In *Computer Graphics Forum*, Vol. 41. Wiley Online Library, 579–588.

Zhiqin Chen, Andrea Tagliasacchi, and Hao Zhang. 2020. Bsp-net: Generating compact meshes via binary space partitioning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 45–54.

Zhiqin Chen and Hao Zhang. 2019. Learning implicit fields for generative shape modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5939–5948.

Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z Wang. 2006. Studying aesthetics in photographic images using a computational approach. In *Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006, Proceedings, Part III 9*. Springer, 288–301.

Yubin Deng, Chen Change Loy, and Xiaoou Tang. 2017. Image aesthetic assessment: An experimental survey. *IEEE Signal Processing Magazine* 34, 4 (2017), 80–106.

Yubin Deng, Chen Change Loy, and Xiaoou Tang. 2018. Aesthetic-driven image enhancement by adversarial learning. In *Proceedings of the 26th ACM international conference on Multimedia*. 870–878.

Yu Deng, Jiaolong Yang, and Xin Tong. 2021. Deformed implicit field: Modeling 3d shapes with learned dense correspondence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10286–10296.

Kapil Dev and Manfred Lau. 2022. Learning perceptual aesthetics of 3D shapes from multiple views. *IEEE Computer Graphics and Applications* 42, 1 (2022), 20–31.

Nir Diamant, Dean Zadok, Chaim Baskin, Eli Schwartz, and Alex M Bronstein. 2019. Beholder-gan: Generation and beautification of facial images with conditioning on their beauty level. In *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 739–743.

Jakub Fišer, Paul Asente, Stephen Schiller, and Daniel Sỳkora. 2016. Advanced drawing beautification with shipshape. *Computers & Graphics* 56 (2016), 46–58.

Ran Gal, Olga Sorkine, Niloy J Mitra, and Daniel Cohen-Or. 2009. iWIRES: An analyze-and-edit approach to shape manipulation. In *ACM SIGGRAPH 2009 papers*. 1–10.

Lin Gao, Jie Yang, Yi-Ling Qiao, Yu-Kun Lai, Paul L Rosin, Weiwei Xu, and Shihong Xia. 2018. Automatic unpaired shape deformation transfer. *ACM Transactions on Graphics (TOG)* 37, 6 (2018), 1–15.

William Gao, Noam Aigerman, Groueix Thibault, Vladimir Kim, and Rana Hanocka. 2023. TextDeformer: Geometry Manipulation using Text Guidance. In *ACM Transactions on Graphics (SIGGRAPH)*.

Thibault Groueix, Matthew Fisher, Vladimir G Kim, Bryan C Russell, and Mathieu Aubry. 2019. Unsupervised cycle-consistent deformation for shape matching. In *Computer Graphics Forum*, Vol. 38. Wiley Online Library, 123–133.

Guanjun Guo, Hanzi Wang, Chunhua Shen, Yan Yan, and Hong-Yuan Mark Liao. 2018. Automatic image cropping for visual aesthetic enhancement using deep neural networks and cascaded regression. *IEEE Transactions on Multimedia* 20, 8 (2018), 2073–2085.

Kunal Gupta and Manmohan Chandraker. 2020. Neural Mesh Flow: 3D Manifold Mesh Generation via Diffeomorphic Flows. *Advances in Neural Information Processing Systems* 33 (2020), 1747–1758.

Rana Hanocka, Noa Fish, Zhenhua Wang, Raja Giryes, Shachar Fleishman, and Daniel Cohen-Or. 2018. Alignet: Partial-shape agnostic alignment via unsupervised learning. *ACM Transactions on Graphics (TOG)* 38, 1 (2018), 1–14.

Fangzhou Hong, Mingyuan Zhang, Liang Pan, Zhongang Cai, Lei Yang, and Ziwei Liu. 2022. AvatarCLIP: Zero-Shot Text-Driven Generation and Animation of 3D Avatars. *ACM Transactions on Graphics (TOG)* 41, 4 (2022), 1–19.

Ruizhen Hu, Wenchao Li, Oliver Van Kaick, Hui Huang, Melinos Averkiou, Daniel Cohen-Or, and Hao Zhang. 2017. Co-locating style-defining elements on 3d shapes. *ACM Transactions on Graphics (TOG)* 36, 3 (2017), 1–15.

Ka-Hei Hui, Ruihui Li, Jingyu Hu, and Chi-Wing Fu. 2022. Neural template: Topology-aware reconstruction and disentangled generation of 3d meshes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 18572–18582.

Md Baharul Islam, Wong Lai-Kuan, and Wong Chee-Onn. 2017. A survey of aesthetics-driven image recomposition. *Multimedia Tools and Applications* 76 (2017), 9517–9542.

Chiyu Jiang, Jingwei Huang, Andrea Tagliasacchi, and Leonidas J Guibas. 2020. Shapeflow: Learnable deformation flows among 3d shapes. *Advances in Neural Information Processing Systems* 33 (2020), 9745–9757.

Shu Kong, Xiaohui Shen, Zhe Lin, Radomir Mech, and Charless Fowlkes. 2016. Photo aesthetics ranking network with attributes and content adaptation. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer, 662–679.

Manfred Lau, Kapil Dev, Weiqi Shi, Julie Dorsey, and Holly Rushmeier. 2016. Tactile mesh saliency. *ACM Transactions on Graphics (TOG)* 35, 4 (2016), 1–11.

Tommer Leyvand, Daniel Cohen-Or, Gideon Dror, and Dani Lischinski. 2008. Data-driven enhancement of facial attractiveness. In *ACM SIGGRAPH 2008 papers*. 1–9.

Congcong Li, Alexander C Loui, and Tsuhan Chen. 2010. Towards aesthetics: A photo quality assessment and photo selection system. In *Proceedings of the 18th ACM international conference on Multimedia*. 827–830.

Jianshu Li, Chao Xiong, Luoqi Liu, Xiangbo Shu, and Shuicheng Yan. 2015. Deep face beautification. In *Proceedings of the 23rd ACM international conference on Multimedia*. 793–794.

Qiqi Liao, Xiaogang Jin, and Wenting Zeng. 2012. Enhancing the symmetry and proportion of 3D face geometry. *IEEE transactions on visualization and computer graphics* 18, 10 (2012), 1704–1716.

Dong Liu, Rohit Puri, Nagendra Kamath, and Subhabrata Bhattacharya. 2020. Composition-aware image aesthetics assessment. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 3569–3578.

Ligang Liu, Renjie Chen, Lior Wolf, and Daniel Cohen-Or. 2010. Optimizing photo composition. In *Computer graphics forum*, Vol. 29. Wiley Online Library, 469–478.

Minghua Liu, Minhyuk Sung, Radomir Mech, and Hao Su. 2021. Deepmetahandles: Learning deformation meta-handles of 3d meshes with biharmonic coordinates. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 12–21.

Xin Lu, Zhe Lin, Hailin Jin, Jianchao Yang, and James Z Wang. 2015. Rating image aesthetics using deep learning. *IEEE Transactions on Multimedia* 17, 11 (2015), 2021–2034.

Yiwei Ma, Xiaoqing Zhang, Xiaoshuai Sun, Jiayi Ji, Haowei Wang, Guannan Jiang, Weilin Zhuang, and Rongrong Ji. 2023. X-Mesh: Towards Fast and Accurate Text-driven 3D Stylization via Dynamic Textual Guidance. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2749–2760.

Mayra D Barrera Machuca, Paul Asente, Wolfgang Stuerzlinger, Jingwan Lu, and Byungmoon Kim. 2018. Multiplanes: Assisted freehand vr sketching. In *Proceedings of the 2018 ACM Symposium on Spatial User Interaction*. 36–47.

Oscar Michel, Roi Bar-On, Richard Liu, Sagie Benaim, and Rana Hanocka. 2022. Text2mesh: Text-driven neural stylization for meshes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 13492–13502.

Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Commun. ACM* 65, 1 (2021), 99–106.

Kenjiro T Miura and Gobithaasan RU. 2014. Aesthetic curves and surfaces in computer aided geometric design. *International Journal of Automation Technology* 8, 3 (2014), 304–316.

Kaichun Mo, Shilin Zhu, Angel X Chang, Li Yi, Subarna Tripathi, Leonidas J Guibas, and Hao Su. 2019. Partnet: A large-scale benchmark for fine-grained and hierarchical part-level 3d object understanding. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 909–918.

Masashi Nishiyama, Takahiro Okabe, Imari Sato, and Yoichi Sato. 2011. Aesthetic quality classification of photographs based on color harmony. In *CVPR 2011*. IEEE, 33–40.

Alice J O'Toole, Theodore Price, Thomas Vetter, James C Bartlett, and Volker Blanz. 1999. 3D shape and 2D surface textures of human faces: The role of "averages" in attractiveness and age. *Image and Vision Computing* 18, 1 (1999), 9–19.

Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. 2019. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 165–174.

Binh Pham. 1999. Design for aesthetics: interactions of design variables and aesthetic properties. In *Human Vision and Electronic Imaging IV*, Vol. 3644. SPIE, 364–371.

Binh Pham and Jinglan Zhang. 2003. A fuzzy shape specification system to support design for aesthetics. *Soft Computing in Measurement and Information Acquisition* (2003), 39–50.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*. PMLR, 8748–8763.

Jian Ren, Xiaohui Shen, Zhe Lin, Radomir Mech, and David J Foran. 2017. Personalized image aesthetics. In *Proceedings of the IEEE international conference on computer vision*. 638–647.

Mattia Segu, Margarita Grinvald, Roland Siegwart, and Federico Tombari. 2020. 3dsnet: Unsupervised shape-to-shape 3d style transfer. *arXiv preprint arXiv:2011.13388*

(2020).

Andrei Sharf, Marina Blumenkrants, Ariel Shamir, and Daniel Cohen-Or. 2006. Snappaste: an interactive technique for easy mesh composition. *The Visual Computer* 22 (2006), 835–844.

I-Chao Shen. 2021. Data-Driven Sketch Beautification With Neural Feature Representation. *IEEE Computer Graphics and Applications* 42, 4 (2021), 72–79.

Chaoyue Song, Jiacheng Wei, Ruibo Li, Fayao Liu, and Guosheng Lin. 2021. 3D pose transfer with correspondence learning and mesh refinement. *Advances in Neural Information Processing Systems* 34 (2021), 3108–3120.

Chaoyue Song, Jiacheng Wei, Ruibo Li, Fayao Liu, and Guosheng Lin. 2023. Unsupervised 3d pose transfer with cross consistency and dual reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2023).

Wei-Tse Sun, Ting-Hsuan Chao, Yin-Hsi Kuo, and Winston H Hsu. 2017. Photo filter recommendation by category-aware aesthetic learning. *IEEE Transactions on Multimedia* 19, 8 (2017), 1870–1880.

Can Wang, Menglei Chai, Mingming He, Dongdong Chen, and Jing Liao. 2022a. Clip-nerf: Text-and-image driven manipulation of neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3835–3844.

Haoran Wang, Jiaxin Li, Alexandru Telea, Jiří Kosinka, and Zizhao Wu. 2022b. USTNet: Unsupervised Shape-to-Shape Translation via Disentangled Representations. In *Computer Graphics Forum*, Vol. 41. Wiley Online Library, 141–152.

Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. 2021. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *arXiv preprint arXiv:2106.10689* (2021).

Weiyue Wang, Duygu Ceylan, Radomir Mech, and Ulrich Neumann. 2019. 3dn: 3d deformation network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1038–1046.

Zhijie Wu, Xiang Wang, Di Lin, Dani Lischinski, Daniel Cohen-Or, and Hui Huang. 2019. Sagnet: Structure-aware generative network for 3d-shape modeling. *ACM Transactions on Graphics (TOG)* 38, 4 (2019), 1–14.

Qinjie Xiao, Xiangjun Tang, You Wu, Leyang Jin, Yong-Liang Yang, and Xiaogang Jin. 2020. Deep shapely portraits. In *Proceedings of the 28th ACM International Conference on Multimedia*. 1800–1808.

Qinjie Xiao, You Wu, Dinghong Wang, Yong-Liang Yang, and Xiaogang Jin. 2021. Beauty3DFaceNet: deep geometry and texture fusion for 3D facial attractiveness prediction. *Computers & Graphics* 98 (2021), 11–18.

Kai Xu, Honghua Li, Hao Zhang, Daniel Cohen-Or, Yueshan Xiong, and Zhi-Quan Cheng. 2010. Style-content separation by anisotropic part scales. In *ACM SIGGRAPH Asia 2010 papers*. 1–10.

Pengfei Xu, Guohang Yan, Hongbo Fu, Takeo Igarashi, Chiew-Lan Tai, and Hui Huang. 2019. Global Beautification of 2D and 3D Layouts With Interactive Ambiguity Resolution. *IEEE transactions on visualization and computer graphics* 27, 4 (2019), 2355–2368.

Zhicheng Yan, Hao Zhang, Baoyuan Wang, Sylvain Paris, and Yizhou Yu. 2016. Automatic photo adjustment using deep neural networks. *ACM Transactions on Graphics (TOG)* 35, 2 (2016), 1–15.

Guandao Yang, Xun Huang, Zekun Hao, Ming-Yu Liu, Serge Belongie, and Bharath Hariharan. 2019. Pointflow: 3d point cloud generation with continuous normalizing flows. In *Proceedings of the IEEE/CVF international conference on computer vision*. 4541–4550.

Yuzhe Yang, Liwu Xu, Leida Li, Nan Qie, Yaqian Li, Peng Zhang, and Yandong Guo. 2022. Personalized image aesthetics assessment with rich attributes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 19861–19869.

Wang Yifan, Noam Aigerman, Vladimir G Kim, Siddhartha Chaudhuri, and Olga Sorkine-Hornung. 2020. Neural cages for detail-preserving 3d deformations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 75–83.

Kangxue Yin, Jun Gao, Maria Shugrina, Sameh Khamis, and Sanja Fidler. 2021. 3dstylenet: Creating 3d shapes with geometric and texture style variations. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 12456–12465.

Mehmet Ersin Yumer, Siddhartha Chaudhuri, Jessica K Hodgins, and Levent Burak Kara. 2015. Semantic shape editing using deformation handles. *ACM Transactions on Graphics (TOG)* 34, 4 (2015), 1–12.

Hancheng Zhu, Yong Zhou, Leida Li, Yaqian Li, and Yandong Guo. 2021. Learning personalized image aesthetics from subjective and objective attributes. *IEEE Transactions on Multimedia* (2021).

C Lawrence Zitnick. 2013. Handwriting beautification using token means. *ACM Transactions on Graphics (TOG)* 32, 4 (2013), 1–8.