UDFStudio: A Unified Framework of Datasets, Benchmarks and Generative Models for Unsigned Distance Functions

Junsheng Zhou*, Weiqi Zhang*, Baorui Ma, Kanle Shi, Yu-Shen Liu, Zhizhong Han

Fig. 1: Diverse shapes with and without open surfaces generated by our UDiFF model. **Top-Left:** Conditional generation of clothes with prompts 'A short-sleeved dress in spiderman style', 'A Batman upper with long sleeves', 'A superman pant', 'A camouflage slip dress'. **Around:** A shape gallery generated by UDiFF conditionally and unconditionally.

***Abstract*—Unsigned distance functions (UDFs) have emerged as powerful representation for modeling and reconstructing geometries with open surfaces. However, the development of 3D generative models for UDFs remains largely unexplored, limiting current methods from generating diverse open-surface 3D content. Moreover, mainstream 3D datasets predominantly consist of watertight meshes, revealing a critical challenge: the absence of standardized datasets and benchmarks specifically tailored for open-surface generation and reconstruction. In this paper, we begin by introducing UDiFF, a novel diffusion-based 3D generative model specifically designed for UDFs. UDiFF supports both conditional and unconditional generation of textured 3D shapes with open surfaces. At its core, UDiFF generates UDFs in the spatial-frequency domain using a learnable wavelet transform. Instead of relying on manually selected wavelet transforms, which are labor-intensive and prone to information loss, we introduce a data-driven approach that learns the optimal wavelet transformation from UDFs datasets. Beyond UDiFF, we present the UWings dataset, comprising 1,509 high-quality 3D open-surface models of winged creatures. Using UWings, we establish comprehensive benchmarks for evaluating both generative and reconstruction methods based on UDFs.**

***Index Terms*—Unsigned distance field, generative modeling, dataset, benchmark, diffusion model.**

Junsheng Zhou, Weiqi Zhang and Baorui Ma are with the School of Software, Tsinghua University, Beijing, China (E-mail: {zhou-js24, zwq23}@mails.tsinghua.edu.cn, mabaorui2014@gmail.com).

Kanle Shi is with Kuaishou Technology, Beijing, China (E-mail: shikanle@kuaishou.com).

Yu-Shen Liu is with the School of Software, Tsinghua University, Beijing, China (E-mail: liuyushen@tsinghua.edu.cn).

Zhizhong Han is with the Department of Computer Science, Wayne State University, USA (E-mail: h312h@wayne.edu).

Junsheng Zhou and Weiqi Zhang contribute equally to this work. Yu-Shen Liu is the corresponding author. This work was supported by National Key R&D Program of China (2022YFC3800600), and the National Natural Science Foundation of China (62272263, 62072268), and in part by Tsinghua-Kuaishou Institute of Future Media Data. Project page: https://weiqi-zhang.github.io/UDiFF. Data link: https://drive.google.com/drive/folders/10A5ROEU0fLYRr8ykTbmoznm_DUzA4riF?usp=sharing.

I. Introduction

PROBABILISTIC diffusion models [1], [2] have largely revolutionized 2D content generation. Recent advancements, such as DALL-E 2 [3] and Stable Diffusion [4], have been widely used in text-to-image generation, image inpainting, etc. A series of stuides [5], [6] try to replicate these success in 3D content generation by developing diffusion models for point clouds or voxels, but fail to produce high

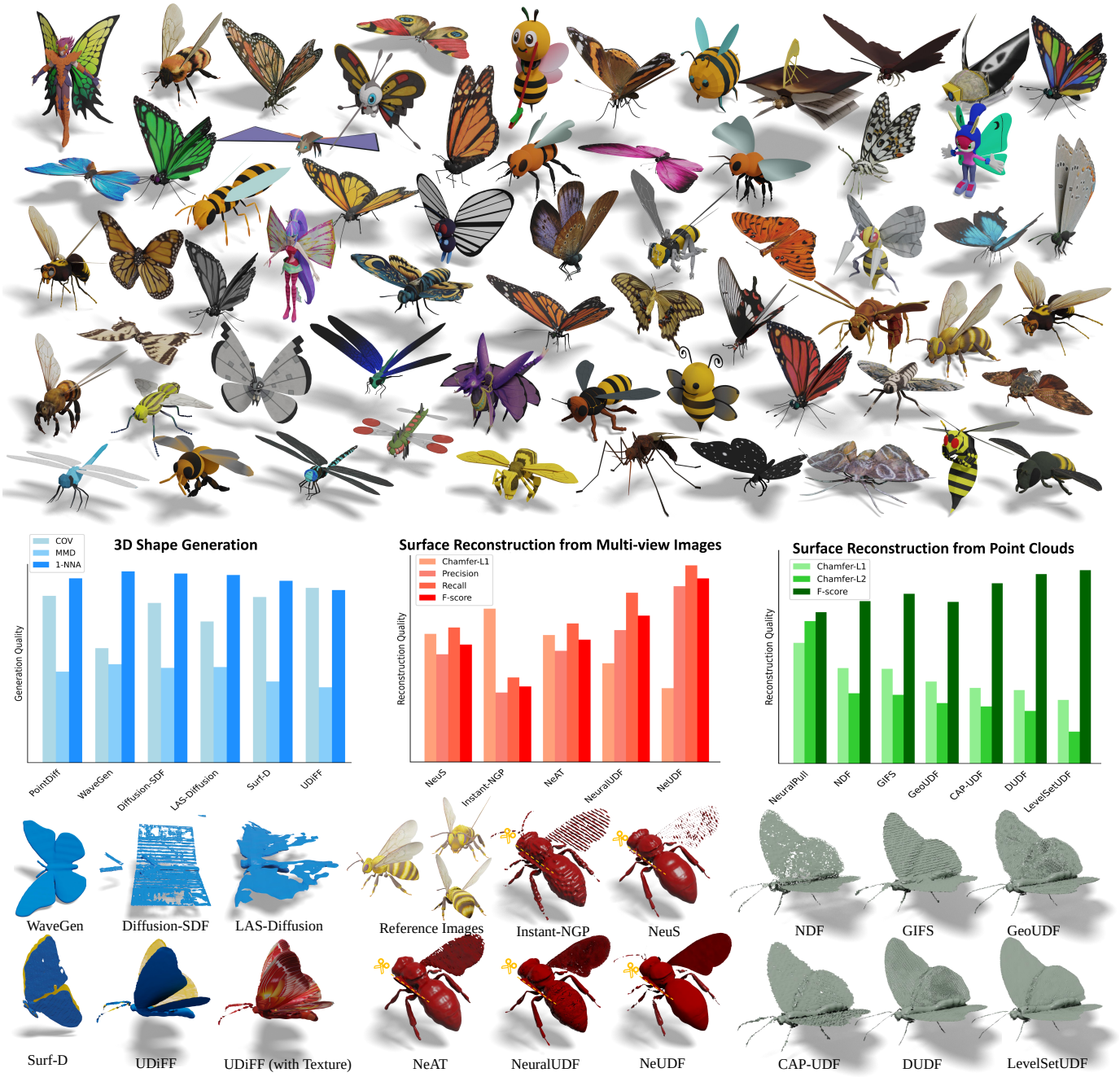

Fig. 2: The illustration of UWings dataset and the benchmarks under UWings dataset.

fidelity results due to the limited resolution in voxels and the discreteness of points. Recent approaches [7]–[9] explore diffusion models to generate 3D shapes represented by neural implicit functions, e.g. signed distance function (SDF) [10], [11] and occupancy function (Occ) [12]. However, these methods are constrained to generating closed shapes, as both SDF and Occ rely on modeling the internal and external relationships of 3D locations to represent 3D shape. This makes previous 3D implicit diffusion models not capable of generating diverse 3D real world contents with open surfaces.

Another challenge in diffusion-based 3D generative models is how to define an appropriate compression transform scheme for achieving compact implicit representations which can be learned by diffusion models efficiently. Some approaches train a variational auto-encoder (VAE) [13] for converting shapes into triplane [14], [15] or single latents [16] for latent diffusion. However, the relatively limited 3D data makes it difficult to train a stable VAE. Instead, another series of approaches (e.g. WaveGen [7]) seek to leverage explicit transform (e.g. wavelet transform [17]) for direct compression. Nevertheless, they need to select an appropriate wavelet type, which often requires extensive manual efforts and can still result in significant information loss during the inverse wavelet transformation.

To address these issues, we propose UDiFF, a 3D diffusion model for unsigned distance fields (UDFs) [18], [19] which

is capable of generating textured 3D shapes with open surfaces. Compared to commonly-used SDF or Occ, UDF has proven to be an advanced representation that supports arbitrary typologies and remains strong generalization. Going beyond unconditioned models, we can also incorporate conditions achieved from CLIP [20] models into UDiFF by introducing conditional cross-attentions. This enables us to control 3D generation using the text and image signals. Previous studies merely focus on generating geometries which lead to a lack of appearance and prevent them from creating diverse and visual-appealing 3D models. In contrast, we get inspiration from Text2Tex [21] to simultaneously generate textures for universal 3D content creation.

Directly applying existing SDF-based diffusion models to UDF does not work well. The difficulty arises from the significantly greater complexity of UDF than SDF, particularly in the context of the non-differential zero-level set. To solve this issue, we introduce UDiFF, a diffusion model in the spatial-frequency domain via an optimal wavelet transformation, which produces a compact representation space for UDF generation. Instead of engaging in selecting a suitable wavelet transformation, which is tedious and often results in significant information loss, we employ a data-driven approach to obtain an optimal wavelet filter for representing UDFs. We minimize the unsigned distance errors during a self-reconstruction through the wavelet transformation, especially near the zero-level set of UDFs. This preserves the geometry details during wavelet transformation, which leads to the high-fidelity generation of 3D geometries.

Our method is initially reported in CVPR 2024 [22]. In this paper, we extend our CVPR work by collecting a large-scale dataset for UDFs, performing more comprehensive evaluations and benchmarking methods for implicit generation and reconstruction with UDFs under the new dataset.

Previous approaches that learn signed distance fields or occupancy for 3D reconstruction and generation are limited to modeling closed topologies, with evaluations conducted on datasets of closed shapes such as ShapeNet [23], DTU [24], ABO [25] and Thingi10K [26]. With recent advances in UDF-based methods for reconstructing and generating 3D shapes with arbitrary topologies, there is an urgent need for a large 3D dataset of open-surface shapes to evaluate their performance. In this paper, we propose the UWings dataset, a high-quality 3D dataset containing 1,509 shapes with open surfaces. Specifically, UWings is a diverse dataset of creatures with wings featuring complex open-surface geometries and thin structures. The dataset is curated from the large Objaverse-XL dataset [27] through both automated and manual filtering. The diversity of the UWings dataset is highlighted in two key aspects: (1) Diverse categories. UWings includes 10 categories of creatures with wings such as bees, butterflies, dragonflies, beetles, mosquitoes, flies, grasshoppers, ladybugs, crickets, and moths. (2) Diverse motions. The 3D models in UWings dataset feature various motions, such as different wing angles and postures.

To facilitate the development of 3D generation, reconstruction and perception with unsigned distance fields, we further establish comprehensive benchmarks based on the UWings dataset, including: (1) 3D shape generation, (2) surface reconstruction from multi-view images, and (3) surface reconstruction from point clouds. We will release all textured meshes, multi-view renderings, and sampled point clouds as part of our dataset. We believe that the new UWings benchmarks provide a large-scale, unbiased platform for comparing both existing and future methods.

We evaluate our proposed method UDiFF for generating 3D shapes with both open and closed surfaces, either conditionally or unconditionally, on the DeepFashion3D [28], ShapeNet [23] and UWings datasets. The experimental results demonstrate that UDiFF achieves promising generation performance compared to the existing state-of-the-art approaches, in both qualitative and quantitative evaluations. Our main contributions can be summarized as follows.

- We propose UDiFF, a 3D diffusion model for unsigned distance fields which is capable of generating real world textured 3D shapes with open surfaces unconditionally or from text conditions.
- We introduce an optimal wavelet transformation for UDFs through data-driven optimization, and justify that the spatial-frequency domain learned through this transformation is a compact domain suitable for UDF generation.
- We introduce the UWings dataset, a high-quality 3D dataset containing 1,509 shapes with complex open-surface geometries and thin structures. UWings includes diverse categories of winged creatures, capturing a range of forms and motions.
- On the proposed UWings dataset, we establish comprehensive benchmarks, including: (1) 3D shape generation, (2) surface reconstruction from multi-view images, and (3) surface reconstruction from point clouds.
- We evaluate UDiFF for generating 3D shapes with both open and closed surfaces, and show our superiority over the state-of-the-art methods.

II. Related Work

With the rapid development of deep learning, the neural networks have shown great potential in 3D applications [29]–[42]. We mainly focus on learning generative Neural Implicit Functions with networks for generating 3D shapes.

A. Neural Implicit Representations

Recently, Neural Implicit Functions (NIFs) have shown promising results in surface reconstruction [10], [12], [43], novel view synthesis [44], [45], image super-resolution [46], [47], etc. The NIFs approaches train a neural network to represent shapes and scenes with signed distance functions (SDFs) [10], [48], [49] or binary occupancy [12], [50], where the marching cubes algorithm [51] is then used to extract surfaces from the learned NIFs. OccNet and DeepSDF [10], [12] are the pioneers of NIFs which learn global latent codes for representing 3D shapes with MLP-based decoder to achieve occupancies or signed distances. The subsequent approaches [50], [52] leverage more latent codes to represent detailed local geometries. PCP [53] and OnSurf [54] introduce

predictive context priors and on-surface prior to enhance the representation ability of NIFs.

Occupancy and SDFs are mainly suitable to represent closed shapes. Recent studies explore the neural unsigned distances (UDFs) [18], [19], [55]–[59] to represent shapes and scenes with open surfaces. NDF [18] designs a hierarchical neural network to learn UDFs with ground truth distance supervisions. GIFS [60] learns UDFs and represents shapes with query relationships. CAP-UDF [19] and LevelSetUDF [59] develop consistency-aware constraints and level set projections to stabilize the optimization of UDFs and produce more accurate geometries.

B. Diffusion-based 3D Generative Models

Generating 3D contents plays the key role in augmented/virtual reality and has been widely explored in the past few years. Earlier studies transfer the success of GAN [61], VAE [13] and the flow-based model [62] in image generation to the 3D domain for generating 3D shapes represented as point clouds [63]–[67] and voxels [6], [68]. For example, PointDiff [5] introduces the powerful diffusion models for point cloud generation. Some advanced methods [69], [70] combining the voxel with point representations were proposed for more robust 3D generation with diffusion models.

More recently, some approaches [7], [14], [71]–[73] try to combine the diffusion models with neural implicit representations for generating high-quality 3D shapes. These methods generate signed distance fields [7], [8], [15], [72], [74] or occupancy fields [9] with diffusion models and extract the meshes from the fields with the marching cubes [51]. For the efficient training of diffusion models, methods like Diffusion-SDF [8] and 3D-LDM [16] train a VAE for converting shapes into latent codes for latent diffusion. But the relative small number of 3D samples for training makes it difficult to train a stable VAE. WaveGen [7] was proposed to explicitly compress SDFs in frequency domain with wavelet transform, but it is limited to the information loss during the wavelet recovery.

The advances in NIFs-based 3D generative models have shown significant improvements in the generation qualities, however, they are limited to generate closed surfaces. This prevents them from generating diverse 3D contents in real world. In this work, we focus on generating UDFs for open surfaces with textures using a 3D diffusion model.

C. 3D Datasets and Benchmarks for Implicit Functions

Collecting large-scale 3D datasets is both costly and challenging. To train and evaluate the representations of implicit functions, researchers usually leverage the existing 3D datasets collected for 3D analysis. For 3D generation tasks, commonly used datasets include ShapeNet [23], ABO [25] and 3D-FUTURE [75]. For the task of surface reconstruction from multi-view images and point clouds, widely-used datasets include DTU [24], FAMOUS [76], Thingi10K [26] and OmniObject3D [77]. However, these datasets are limited to closed shapes, making them inadequate for evaluating the performance of implicit functions in modeling shapes with arbitrary typologies.

Recent advances in unsigned distance functions (UDFs) have demonstrated their promising capability in generating and reconstructing open surfaces with arbitrary topologies. Current studies conduct benchmarks on DeepFashion3D [28], a 3D garment dataset containing 563 instances of open garments. However, the DeepFashion3D dataset is of relatively low quality and lacks diversity. The DeepFashion3D dataset is collected using scanning sensors, inevitably producing self-occlusions and invisible regions. Additionally, the garment geometries often lack detail and do not present significant challenges for learning to model. Therefore, there is an urgent need to develop a diverse and large-scale 3D dataset of open-surface shapes with high-quality geometric details, which is used to properly evaluate UDF-based methods. To achieve this purpose, we collect and develop the UWings dataset for evaluating UDF methods, as will be introduced in detail below.

III. The UWings Dataset

A. Data Collection, Filtering and Processing

Data Collection. We collect shapes for UWings dataset from the large Objaverse [78] and Objaverse-XL [27] datasets. Objaverse-XL is the largest available 3D asset dataset, compiled from diverse internet sources such as GitHub, Sketchfab, Thingiverse and Polycam. We aim to collect a high-quality set of winged creature models from Objaverse-XL dataset to create a dedicated dataset for open-surface shapes. However, the Objaverse-XL dataset is highly noisy, containing lots of low-quality models which significantly affect the performance of training generative models. To this end, we choose to collect data for UWings dataset from the Objaverse dataset and the curated subset of Objaverse-XL dataset. Since Objaverse-XL dataset involves all the models in the Objaverse dataset, we will refer the used data source as Objaverse-XL.

Data Filtering. To collect shapes with wings from the Objaverse-XL dataset, we first utilized the Cap3D [79] dataset and its extended version, i.e. DiffuBank [80], to automatically filter insect models with thin wings. Cap3D captions over 1 million 3D models in the Objaverse-XL dataset by leveraging BLIP2 [81] to generate captions for renderings, where GPT-4 Vision [82] is used for refining the final captions. Specifically, we filtered out the target shapes from Objaverse-XL by querying relevant text prompts from the Cap3D captions. We set the query prompts to 10 insect categories with thin wings, including bees, butterflies, dragonflies, beetles, mosquitoes, flies, grasshoppers, ladybugs, crickets and moths. The models that do not contain these category-specific texts in their Cap3D captions were excluded, and the remaining models were collected as the automatically filtered set.

Manual filtering is also necessary for improving the quality of our dataset, since the Objaverse-XL dataset is highly noisy, which contains a significant proportion of corrupted or mismatched models, even within the curated subset. We manually reviewed all the 3D models in the automatically filtered set and removed defective or mismatched models. To enhance the diversity of motion patterns, we also captured frames from animated objects to include 3D models in various

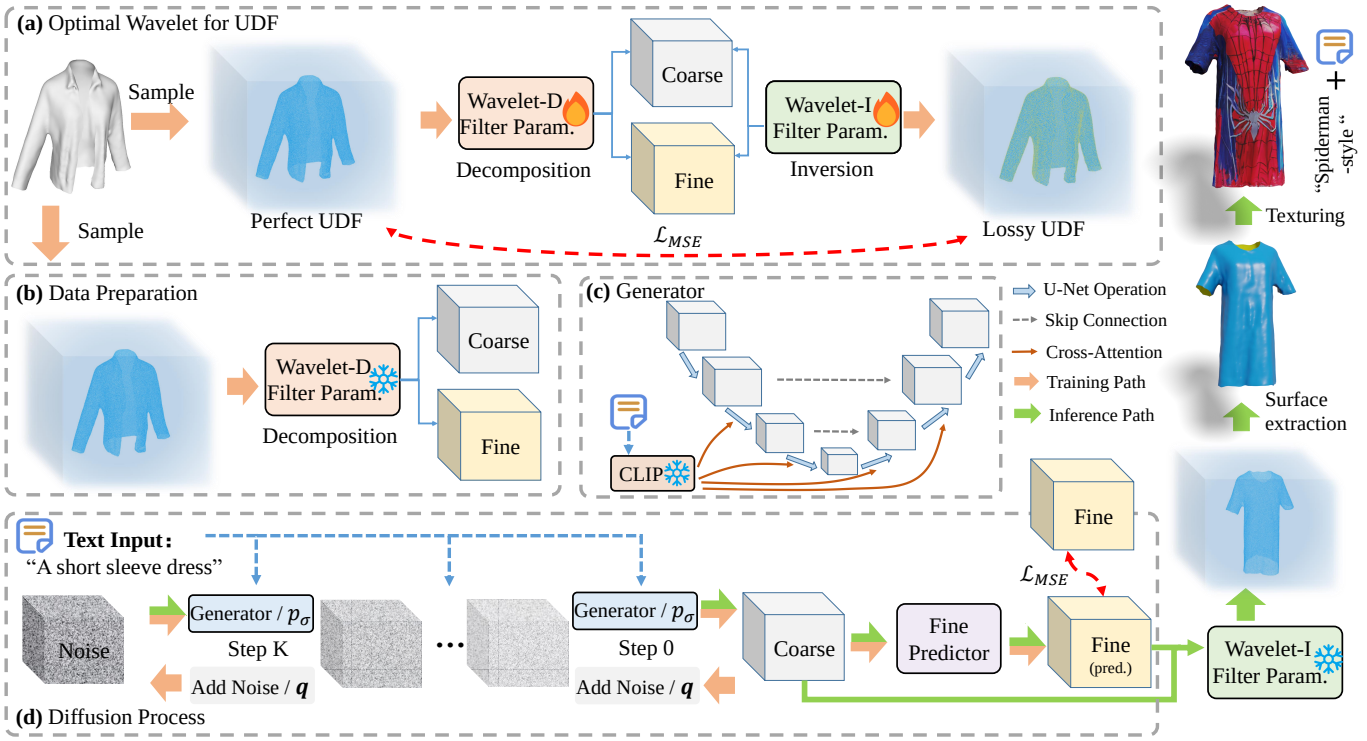

Fig. 3: **Overview of UDiFF. (a)** We propose a data-driven approach to attain the optimal wavelet transformation for UDF generation. We optimize wavelet filter parameters through the decomposition and inversion by minimizing errors in UDF self-reconstruction. **(b)** We fix the learned decomposition wavelet parameters and leverage it to prepare the data as a compact representation of UDFs including pairs of coarse and fine coefficient volumes. **(c)** is the architecture of the generator in diffusion models, where text conditions are introduced with cross-attentions. **(d)** The diffusion process of UDiFF. We train the generator to produce coarse coefficient volumes from random noises guided by input texts and train the fine predictor to predict fine coefficient volumes from the coarse ones. Follow the **green** arrows for inference, we start from a random noise and an input text condition to leverage the trained generator to produce a coarse coefficient volume. The trained fine predictor then predicts the fine coefficient volume. Together with the coarse one, we recover the UDFs with the fixed pre-optimized inversion wavelet filter parameters. Finally, we extract surfaces from UDFs and further texture them with the guiding text.

motions and poses. In total, we obtained 1,509 clean, high-quality 3D models of winged creatures from the 4,073 models in the automatically filtered set.

Data Processing. We begin by normalizing all models to fit within a unit sphere before sampling 3D/2D data for downstream tasks. These normalized objects are ready to use in training 3D generative models. For surface reconstruction from point clouds, we uniformly sample 10,000 points from each model to create a pair consisting of a point cloud and its corresponding ground truth mesh. For multi-view reconstruction, we render 93 views with uniformly distributed camera poses around each model.

B. Benchmarking UWings

We conduct comprehensive benchmarks on the developed UWings dataset to evaluate the performance of various unsigned distance field-based methods on 3D generation and reconstruction. Here, we mainly focus on three tasks: (1) 3D shape generation, (2) surface reconstruction from multi-view images, and (3) surface reconstruction from point clouds. In Sec. VI, we report the benchmarks and evaluations on the state-of-the-art methods.

IV. The UDiFF Model

Overview. The overview of UDiFF is shown in Fig. 3. UDiFF is a 3D generative model which takes texts as conditions and generates general textured 3D shapes with either open or closed surfaces. We will start by introducing the novel approach to obtain an optimal wavelet transform for a compact UDF representation and the data preparation process for training diffusion models in Sec. IV-A. We then present the designed conditional diffusion framework for UDF generation and the generator network in Sec. IV-B. Finally, we extract surfaces from the generated UDF and further add textures on the mesh with the guiding text in Sec. IV-C.

A. Optimal Wavelet Transformation for UDFs

One main challenge in diffusion-based 3D generative models is to search for a compact representation space for diffusion model to learn efficiently. WaveGen [7] adopts an explicit wavelet transform on the SDF volumes (256^3) to decompose them into coarse coefficient volumes and fine coefficient volumes with much lower resolutions. The naive wavelet transform leads to large information loss since the manually

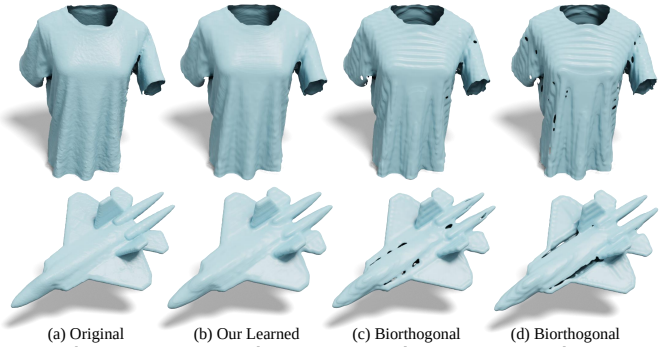

Fig. 4: **Comparisons of reconstructions with different wavelet filters.** **(a)** The input shapes from DeepFashion3D [28] and ShapeNet [23], from which we sample UDFs to produce compact wavelet representations. **(b)** The surfaces extracted from the recovered UDF with decomposition and inversion by our learned wavelet filter. **(c,d)** The surfaces extracted from the recovered UDF with manual chosen wavelet filters.

selected wavelet is not capable of representing various shapes as accurate distance functions.

To represent UDFs in a compact way, we follow WaveGen to adopt multi-scale wavelet transform [17], [83] as the compressing schema, keeping only the coefficients at a relative small scale of $\mathcal{J} = 3$ for efficient shape learning. However, the UDF is significantly more complex and unstable than SDF, particularly in the area of non-differential zero-level sets, where the geometry details that the wavelet compressing does not preserve will severely affect the generation of UDFs. Thus, a suitable wavelet filter with much less information loss but remains compact and efficient for UDFs is vital.

To this end, instead of manually searching for the appropriate wavelet filter which demands costly efforts and is still hard to reduce the information loss, we propose a data-driven approach to learn the optimal wavelet filter parameters for UDFs through learning-based optimization as shown in Fig. 3(a). Specifically, we define a learnable biorthogonal wavelet filter which consists of a decomposition filter ϕ_θ^D and an inversion filter ϕ_δ^I with learnable filter parameters θ and δ. Given a set of shapes $\{S_i\}_{i=1}^N$, we first sample the UDF volume U_i for each shape at a resolution of 256^3 and truncate the distance values in U_i to [0, 0.1], and then compress it into a coarse coefficient volume and a fine coefficient volume with the learnable decomposition filter ϕ_θ^D as:

$$\{C_i, F_i\} = \phi_\theta^D(U_i). \tag{1}$$

We then predict the lossy UDF $\hat{U}$ from C_i and F_i with the learnable inversion filter ϕ_δ^I as:

$$\{\hat{U}_i\} = \phi_\delta^I(C_i, F_i). \tag{2}$$

The target is to optimize the filter parameters θ and δ by minimizing the information loss during wavelet decomposition and inversion, formulated as:

$$\min_{\theta,\delta} \sum_{i=1}^{N} \mathcal{L}_{\text{MSE}}(w_i^\gamma \hat{U}_i, w_i^\gamma U_i), \tag{3}$$

where w_i^γ is the weights for enforcing the optimization to focus on the space near the zero-level set of UDF. w_i^γ has the same size as U_i for weighting each grid in the UDF volume, where we define w_i^γ according to a threshold γ to mask the grids with distances larger than γ.

After data-driven optimization of the wavelet filters ϕ_θ^D and ϕ_δ^I, we learn the optimal wavelet transform with much less information loss and can faithfully reconstruct the original UDF while remains compact. We show the comparison on the wavelet filters in Fig. 4, where the surfaces reconstructed from UDF with our learned wavelet filter in Fig. 4(b) are much smoother and more accurate than the reconstructions with common filters like Biorthogonal wavelet 3-3 in Fig. 4 (d). Specifically, Biorthogonal wavelet 6-8 in Fig. 4 (c) is the carefully chosen filter by WaveGen from a series of wavelet filters, where our learned filter significantly outperforms the manually selected filters in compressing and recovering UDF. The reason is that the filters learned by data-driven optimizing from UDF datasets are much more suitable to specific characters of UDFs, which preserves more geometry details.

With the learned optimal wavelet filter, we then leverage it to represent UDFs as a compact representation for training diffusion models. As shown in Fig. 3 (b), we fix the parameters for ϕ_θ^D and produce the paired coarse efficient $\{C_i\}_{i=1}^M$ volumes and fine efficient volumes $\{F_i\}_{i=1}^M$ by decomposing U_i with Eq. (1).

B. Conditional UDF Diffusion

Generator Architecture. We first introduce the network details of diffusion generators for 3D volumes, as shown in Fig. 3(c). The generator shares a similar U-Net architecture as Stable-Diffusion [4], [84], where the 2D convolutions are replaced with 3D ones for handling 3D volumes. Each U-Net operation in Fig. 3(c) contains $3\times3\times3$ residual blocks, pooling layers and down/up-sampling layers. For introducing text conditions to diffusion models, we first encode the input texts with frozen CLIP [20] models to produce text embeddings and then fuse them into the volume features with cross-attention layers.

Learning Diffusion Models. We develop our 3D generative model UDiFF based on diffusion probabilistic models [1], [2]. The diffusion process is to generate coarse coefficient volumes which represents the general geometry of 3D shapes from random volume noises, as shown in Fig. 3 (d). We define $\{\mathcal{C}_0, \mathcal{C}_1, ..., \mathcal{C}_T\}$ as the forward process $q(\mathcal{C}_{0:T})$ which gradually transforms a real data $\mathcal{C}_0$ into Gaussian noise $(\mathcal{C}_T)$ by adding noises, where $\mathcal{C}_0$ is a sample from the coarse coefficient data $\{C_i\}_{i=1}^M$. The diffusion backward process $p_\sigma(\mathcal{C}_{0:T})$ leverages the generator with parameter σ to denoise $\mathcal{C}_T$ into a real data sample. The learning schema is to train the generator to maximize the generation probability of the target,

i.e. $p_\sigma(\mathcal{C}_0)$. We follow DDPM [2] to simplify the optimization target to predict noises ϵ_σ with the generator, formulated as:

$$\min_{\sigma} \mathbb{E}_{\mathcal{C}_0,t,\epsilon\sim\mathcal{N}(0,1)} \left[\|\epsilon - \epsilon_\sigma\left(\mathcal{C}_t, t\right)\|^2 \right], \tag{4}$$

where t is a time step and ϵ is a noise volume sampled from the unit Gaussian distribution $\mathcal{N}$.

Condition-Guided 3D Diffusion. Up to this point, we have covered the generative diffusion process without conditions. For a controllable generation of unsigned distance fields, we further introduce a conditioning mechanism [4] into the diffusion process by cross-attention. Specifically, given an input text y, we first leverage a frozen CLIP text encoder τ to project y into the condition embedding $\tau(y)$. The embedding is then fused into the U-Net layers of generator with cross attention modules implemented as $\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)\cdot V$, where $Q = W_Q^{(i)}\cdot\varphi_i(\mathcal{C}_t)$, $K = W_K^{(i)}\cdot\tau(y)$ and $V = W_V^{(i)}\cdot\tau(y)$. Here, $\varphi_i(\mathcal{C}_t)$ is the output of an intermediate layer of the U-Net and $W_V^{(i)}$, $W_Q^{(i)}$ & $W_K^{(i)}$ are learnable matrices.

The cross-attention mechanism learns a mapping from the input text condition to the coefficient volumes which represent the geometric generations. The optimizing target in Eq. (4) is then modified as:

$$\min_{\sigma} \mathbb{E}_{\mathcal{C}_0,y,t,\epsilon\sim\mathcal{N}(0,1)} \left[\|\epsilon - \epsilon_\sigma\left(\mathcal{C}_t, \tau(y), t\right)\|^2 \right]. \tag{5}$$

Fine Predictor. The last module for learning to generate UDFs is the fine predictor f which predicts fine coefficient volumes from the generated coarse ones. We follow WaveGen [7] to implement f with the similar U-Net architecture as the generator. We train f with pairs of coarse and fine coefficient volumes $\{C_i, F_i\}$ with MSE loss to minimize the differences between F_i and the prediction $f(C_i)$.

C. Generating Novel 3D Shapes

Generating UDFs at Inference. With the learned optimal wavelet filters and the trained conditional diffusion models, we can now generate novel 3D shapes as shown in the green arrows in Fig. 3. Starting from a random volume noise and an input text y, we leverage the generator to produce a coarse coefficient C' volume by removing noises iteratively with the guidance of y. The fine predictor then predicts the fine coefficient volume F', together with C' to generate the UDF U' by the wavelet inversion with the learned filter ϕ_δ^I as Eq. (2).

Surface Extraction and Texturing. After generating a novel UDF U', we extract the zero-level set of U' as a surface. The recent works [19], [85] leverage the gradients at UDF as the signals to mesh UDFs, however, the approximated gradients of generated UDF may not be stable enough at the zero-level set, which leads to errors and holes. We therefore adopt DCUDF [86] with double covering to mesh the generated UDF of UDiFF. Finally, to create visual-appealing 3D models, we drew inspiration from Text2Tex [21]. This helps to generate textures for the extracted mesh while leveraging the text guidance within a progressive rendering-based texturing framework.

TABLE I: **Quantitative comparison of generated shapes on DeepFashion3D dataset.** MMD-CD scores and MMD-EMD scores are scaled by 10^3 and 10^2, respectively.

Method	COV ↑		MMD ↓		1-NNA ↓	
	CD	EMD	CD	EMD	CD	EMD
PointDiff [5]	68.67	64.56	**11.01**	15.53	83.21	87.69
WaveGen [7]	62.34	51.89	15.56	17.03	92.93	94.83
Diffusion-SDF [8]	67.09	62.03	14.79	16.63	88.98	92.63
LAS-Diffusion [87]	67.40	56.01	14.59	16.53	88.61	91.41
Ours	**69.62**	**67.72**	11.60	**14.01**	**81.83**	**82.14**

Fig. 5: Visual comparison of generated shapes with state-of-the-arts trained under DeepFashion3D dataset. The front and back faces are rendered in different colors for a visualization of open surfaces.

V. Experiment on UDiFF

In this section, we evaluate our proposed UDiFF in shape generation. We first demonstrate the performance of UDiFF in generating novel shapes with open surfaces in Sec. V-A. Next, we conduct experiments on generating shape with closed surfaces in Sec. V-B. The ablation studies are shown in Sec. V-C. The implementation details are reported in Sec. V-D.

A. Open-Surface Shape Generation

Dataset. For evaluations in generating shapes with open surfaces, we conduct experiments on DeepFashion3D dataset

Fig. 6: Conditional generations produced by UDiFF and Shap·E. The front and back faces are rendered in different colors for a clear visualization of open surfaces.

Fig. 7: Image conditioned generation with UDiFF. (a) Open-surface geometries generated with image guidance. (b) An example of generating textured shapes with image guidance.

[28]. The DeepFashin3D dataset is a real-captured 3D dataset of open-surface clothes, containing 1,798 models reconstructed from real garments. It covers 10 categories and 563 garment instances. The dataset is randomly split into training and testing sets by the ratio 80% and 20 %. To get the text conditions for training UDiFF, we first render each model from the front facing view to obtain the image representing the model. We then leverage BLIP2 [81] for captioning the images and keep the caption description of the rendered image for each model as the text condition for the model. We further mix the category description of each model provided in the dataset into the text condition as a supplementary.

Metrics. For a fair comparison with various methods, we conduct the quantitative evaluations on the unconditional shape generation. We randomly generate 1,000 shapes with the trained model and uniformly sample 2,048 points on each generated shape. We follow previous works [5], [7] to evaluate the generation quality using Minimum matching distance (MMD), Coverage (COV) and 1-NN classifier accuracy (1-NNA). MMD measures the geometry accuracy of the generated shapes. COV indicates the ability of the generated shapes to cover the shapes in the test set. 1-NNA is designed to measure how well a classifier differentiates the generated shapes from the given shapes in the testing set. Lower is better for MMD, higher is better for COV and the closer to 50 % the better for 1-NNA.

Baselines. We compare UDiFF with the state-of-the-art methods in terms of the shape generation quality. The methods includes PointDiff [5], WaveGen [7], Diffusion-SDF [8] and LAS-Diffusion [87]. PointDiff uses point cloud data for training, where we sample 2,048 points on each model and leverage the official code for training. All the previous implicit-based shape generation methods represent shapes as SDF or Occ, where the watertight meshes are required to generate the SDF/Occ data for training. Therefore, we leverage the commonly-used manifold method [88] for preprocessing the open-surfaces in DeepFashion3D. After that, we follow the official codes of these methods for training unconditional models with the watertight meshes.

Comparison. The quantitative comparison is shown in Tab. I. where UDiFF achieves the best performance compared to the previous state-of-the-art methods. The main reason is that all the previous implicit-based methods fail to handle the open-surfaces, where the needed manifold preprocessing leads to large bias on the original open-surface shapes. While the proposed UDiFF represents shapes as unsigned distance fields and is able to handle general shapes with or without open surfaces, leading to superior performance compared to other methods.

The visual comparison is shown in Fig. 5, where the proposed UDiFF significantly outperforms the previous works in generating visual-appealing clothes with open surfaces. We render the inside and outside surfaces in different colors for a clear difference on open surfaces. The PointDiff generates the point cloud to represente shapes, which do not require the manifold preprocess. However, it struggles to produce high-fidelity generations due to the discreteness of points.

Text-conditional Generation. For evaluations in generation with conditions, we further train a conditional model and generate shapes with the guidance from provided text prompts. We visually compare the generations with those produced by Shap·E under the same texts as shown in Fig. 6. The results demonstrate that UDiFF generates more accurate and high-fidelity predictions from the texts. UDiFF also produces more realistic textures thanks to the powerful Text2Tex [21]. On the contrary, Shap·E struggles to generate correct geometries and textures.

Image-conditional Generation. We further justify that UDiFF can receive diverse signals except texts (e.g. images) for conditional generation. This is achieved by leveraging the pre-aligned text and image representations of the CLIP model, where we adopt the frozen CLIP image encoder to achieve the image embeddings to guide UDiFF generation by cross-attention, without requiring extra training on images. We show the image-conditional generations of UDiFF in Fig. 7. The

TABLE II: **Quantitative comparison of generated shapes on ShapeNet dataset.** MMD-CD scores and MMD-EMD scores are scaled by 10^3 and 10^2, respectively.

Method	Chair						Airplane					
	COV ↑		MMD ↓		1-NNA ↓		COV ↑		MMD ↓		1-NNA ↓	
	CD	EMD	CD	EMD	CD	EMD	CD	EMD	CD	EMD	CD	EMD
IM-GAN [89]	**56.49**	54.50	11.79	14.52	61.98	63.45	61.55	62.79	3.320	8.371	76.21	76.08
Voxel-GAN [90]	43.95	39.45	15.18	17.32	80.27	81.16	38.44	39.18	5.937	11.69	93.14	92.77
PointDiff [5]	51.47	55.97	12.79	16.12	61.76	63.72	60.19	62.30	3.543	9.519	74.60	72.31
SPAGHETTI [91]	49.48	50.22	14.7	15.85	72.34	69.46	56.86	58.83	4.260	8.930	79.36	78.86
SALAD (Global) [74]	49.71	48.75	11.71	14.12	62.72	61.25	54.88	59.33	3.877	8.958	82.20	80.35
SALAD [74]	56.42	55.16	11.69	14.29	**57.82**	**58.41**	63.16	**65.39**	3.636	8.238	**73.92**	**71.08**
WaveGen [7]	49.63	50.15	12.12	14.25	65.04	62.87	60.94	59.09	3.528	7.964	75.77	72.93
Ours	52.58	**55.99**	**11.67**	**14.04**	65.96	63.42	**64.77**	63.78	**3.151**	**7.798**	74.48	78.99

Fig. 8: Visual comparison of the generated shapes with state-of-the-arts on ShapeNet dataset.

textures on the right of Fig. 7 is achieved with Text2Tex [21] on the text prompt predicted from the image with BLIP2 [81], i.e., 'A white floral shirt with a long sleeves' .

Category-conditional Generation. We provide more shape generations obtained by the UDiFF model trained on DeepFashion3D [28] dataset with the cloth categories as the conditions. Specifically, we generate 8 categories of cloth shapes, including *"long sleeve dress"*, *"long sleeve upper"*, *"pants"*, *"no sleeve dress"*, *"no sleeve upper"*, *"dress"*, *"shot sleeve dress"* and *"shot sleeve upper"*. The visualizations are shown in Fig. 14, where UDiFF generates diverse and novel shapes correctly corresponds to the text conditions.

B. Closed Shape Generation

Dataset and metrics. For the closed shape generation, we follow the common setting of previous methods [7], [74] to conduct generation experiments under the airplane and chair classes of the ShapeNet [23] dataset. We randomly generate 2,000 shapes with the trained model and uniformly sample 2,048 points on each generated shape. We follow previous works [5], [7] to evaluate the generation quality using MMD, COV and 1-NNA. We compare our method with all the baselines using their officially provided pretrained models and codes.

Comparison. We compare UDiFF with the state-of-the-art methods including IM-GAN [89], Voxel-GAN [90], PointDiff [5], SPAGHETTI [91], WaveGen [7] and SALAD [74]. We show the quantitative comparison in Tab. II, where the results are directly borrowed from WaveGen and SALAD for a fair comparison.

The comparison demonstrates UDiFF also has the capability to generate high-fidelity watertight geometries with only closed surfaces. We justify that UDiFF is a general shape generator to produce general shapes with either open surfaces or closed surfaces. We achieve the comparable performance with the state-of-the-art method SALAD [74], and also significantly outperform the baseline WaveGen [7] which also leverages wavelet transformation as the compact representation. The reason is that our proposed approach for learning optimal wavelet filter largely reduces the information loss during transformation, which leads to more accurate and diverse generations. We further show the visual comparison of some generated shapes of different methods in Fig. 8. We can see that the shapes generated by our method are more faithful than IM-GAN and SPAGHETTI by producing finer details and cleaner surfaces, and have less bumpy geometries than WaveGen thanks to the optimal wavelet filter to significantly reduce information loss.

Conditional Generation. We further train a text-conditional model under the 'Chair' category of the ShapeNet dataset. We visually compare the generations produced by AutoSDF [92] and our propose UDiFF with the same text conditions as shown in Fig. 9. The results demonstrate that UDiFF generates more accurate and high-fidelity predictions from the texts compared to AutoSDF.

More Visualizations. We further provide more unconditional shape generation results achieved by the UDiFF model trained

Fig. 9: Text-conditioned generation with UDiFF and AutoSDF on ShapeNet dataset.

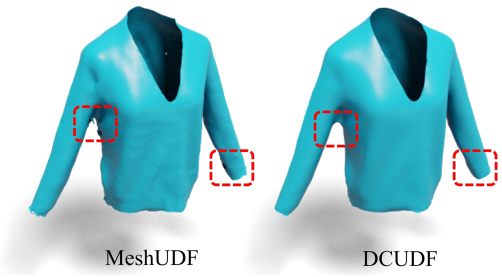

Fig. 10: Mesh extraction comparisons between MeshUDF and DCUDF.

TABLE III: **Ablation studies on the framework design.** MMD-CD scores and MMD-EMD scores are scaled by 10^3 and 10^2.

Method	COV ↑		MMD ↓		1-NNA ↓	
	CD	EMD	CD	EMD	CD	EMD
W/o learned wavelet	64.52	65.02	13.24	15.26	85.06	86.22
W/o fine predictor	66.36	65.18	12.37	14.48	83.62	84.17
Full	**69.62**	**67.72**	**11.60**	**14.01**	**81.83**	**82.14**

under single categories of ShapeNet [23] dataset. We generate shapes of the *"chair"* and *"airpane"* categories. The visualizations are shown in Fig. 15, where UDiFF generates visual-appealing shapes.

C. Ablation Studies

Framework Design. To evaluate the major components in our methods, we conduct ablation studies under the DeepFashion3D dataset [28] and report the performance in Tab. III. We first justify the effectiveness of the proposed optimal wavelet transformation by replacing our learned wavelet filter with the previous carefully chosen wavelet filter by WaveGen [7], i.e., Biorthogonal 6-8. The result is shown as 'W/o learned wavelet'. We then remove the fine predictor of UDiFF to recover the 3D shapes with only the generated coarse coefficients as shown in 'W/o fine predictor'. The ablation study results demonstrate that effect of designs in UDiFF by significantly improving the generation performance.

TABLE IV: **Ablation studies on the effect of wavelet optimization.** We report the L2 Chamfer Distance scaled by 10^5.

Method	Haar	Biorthogonal3-3	Biorthogonal6-8
CD	264.8	46.04	42.92

Method	Learnable ϕ_θ^D	Learnable ϕ_δ^I	Both
CD	36.12	32.15	**28.51**

The Effect of Wavelet Optimization. We further evaluate the effect of our proposed wavelet optimization to achieve optimal wavelet filter. The result is shown in Tab. IV, where we conduct evaluations under the test set of DeepFashion3D [28] and report the L2 Chamfer Distance between the ground truth meshes and the recovered meshes with wavelet filters Haar, Biorthogonal3-3, Biorthogonal6-8 and ours. We show the performance of only optimizing decomposition filter parameters ϕ_θ^D with fixed inversion filter parameters ϕ_δ^I as 'Learnable ϕ_θ^D', and only optimizingϕ_δ^I with fixed ϕ_θ^D as 'Learnable ϕ_δ^I'. The best performance is achieved with optimizing both ϕ_δ^I and ϕ_θ^D as 'Both'.

The Meshing Approach. We further conduct ablation studies on the meshing approaches for extracting geometries from the generated UDFs. We show the visual comparison of meshing the generated UDFs with MeshUDF [85] and DCUDF [86] in Fig. 10.

D. Implementation Details

Meshing. Different from SDFs, UDFs fail to extract surfaces by the marching cubes [51] since UDFs cannot perform inside/outside tests on 3D grids. Recent works [19], [85] leverage the gradients at UDF grids as the signals to mesh UDFs. However, for the generated UDFs, the approximated gradients may not be stable enough at the zero-level set, which results in errors and holes. The approximated gradient at a grid point q is defined as the direction from q to the neighbour grid q_n where the UDF from q to q_n increases rapidly the most. We adopt DCUDF [86] with double covering to mesh the generated UDF of UDiFF, which results in more continuous surfaces. We make an adaption to DCUDF on the double covering operation to replace the time-consuming optimizations with an explicit vertices refinement strategy. We move each vertices against the surface normals with a stride of unsigned distances to reach the zero-level sets, and then leverage the min-cut algorithm to achieve the final model.

Texturing. We leverage Text2Tex [21] to generate textures for the extracted meshes. This is achieved with a progressive texture generation process and a texture refinement process. Specifically, we first render the texture-less initial mesh from the preset viewpoint and generate the appearance according to the text prompt with the depth-guided stable-diffusion [4]. We then adjust to the next preset viewpoint and repeat the appearance generation process until the last preset viewpoint where the whole mesh is textured. Finally, we optimize the textures with automatically selected viewpoints for refinement.

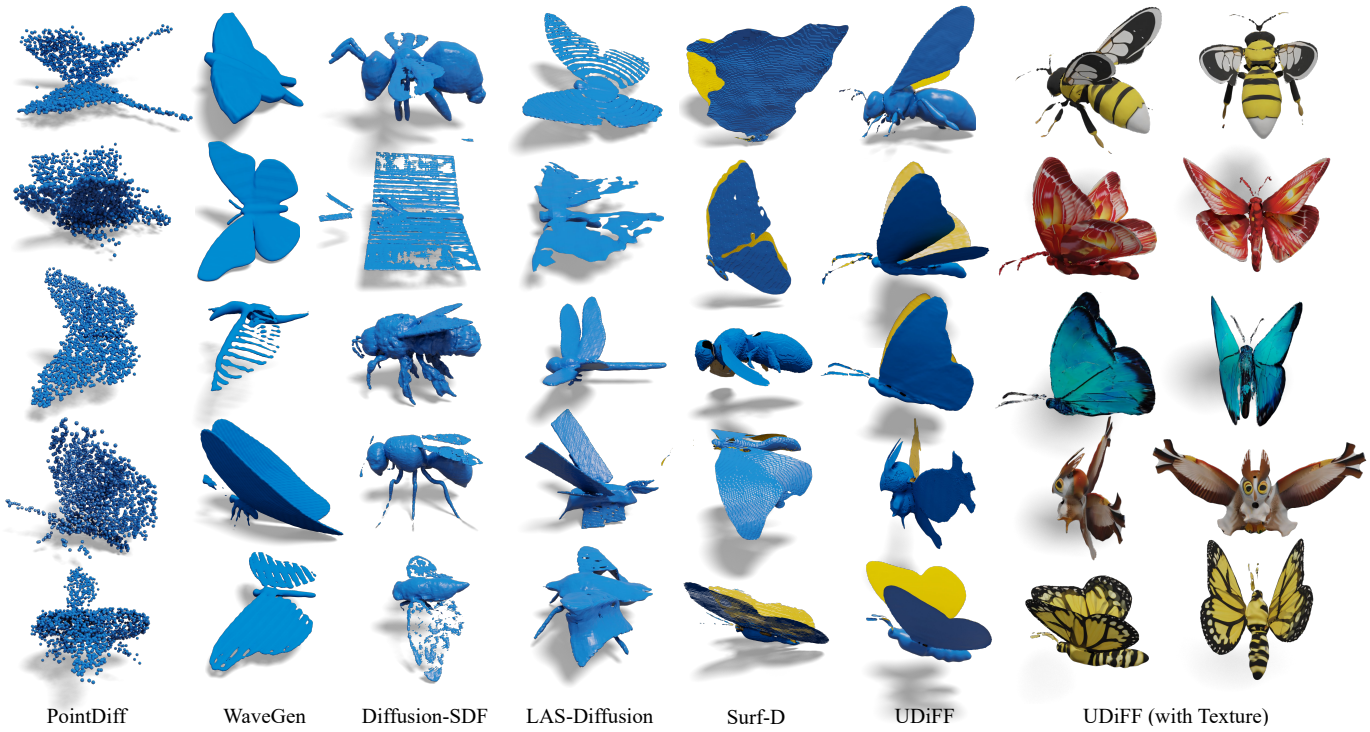

Fig. 11: Visual comparisons of 3D shape generations under the UWings dataset.

TABLE V: **Quantitative comparison of generated shapes on UWings dataset. MMD-CD scores and MMD-EMD scores are scaled by 10^3 and 10^2, respectively.**

Method	COV ↑		MMD ↓		1-NNA ↓	
	CD	EMD	CD	EMD	CD	EMD
PointDiff [5]	71.56	70.21	23.82	16.75	78.97	82.67
WaveGen [7]	64.10	63.00	30.83	19.40	83.51	87.44
Diffusion-SDF [8]	69.60	67.77	27.16	17.54	81.93	83.66
LAS-Diffusion [87]	66.30	67.40	27.88	17.88	80.99	86.33
Surf-D [93]	71.20	72.89	22.67	16.87	77.66	83.51
UDiff	**74.35**	**75.45**	**16.63**	**13.78**	**73.53**	**80.67**

VI. Benchmarks on UWings

A. 3D Shape Generation

Data Settings and Metrics. We conduct experiments under the UWings dataset for evaluating the 3D generation quality and diversity. For a comparison among various methods, we report the quantitative performance in unconditional shape generation. We randomly generate 1,000 shapes with the trained models and uniformly sample 2,048 points on each generated shape. We adopt the same settings as experiments in Sec.V-A to evaluate the generation quality using Minimum matching distance (MMD), Coverage (COV) and 1-NN classifier accuracy (1-NNA).

Baselines. Using shapes with open surfaces in the UWings dataset, we evaluate the generation performance of the state-of-the-art methods, including PointDiff [5], WaveGen [7], Diffusion-SDF [8], LAS-Diffusion [87], Surf-D [93], and UDiFF [22]. PointDiff is a point cloud-based 3D generation method, where we sample 2,048 points from each model and use the official code for training. For the state-of-the-art SDF-based 3D generation methods (e.g. WaveGen, Diffusion-SDF and LAS-Diffusion), we follow the same setup as experiments in Sec.V-A by first preprocessing the open surfaces in the UWings dataset using manifold methods, and then generating SDF data for training.

UDiFF is the first method to explore UDF-based 3D generation for creating open surfaces. For a comprehensive comparison of UDF-based 3D generation methods, we also include Surf-D, a recently released 3D generative model for UDF. Both UDiFF and Surf-D [93] can be directly trained on shapes with open surfaces in the UWings dataset without the need for manifold operations.

Comparisons. The quantitative comparison is shown in Tab. V. UDiFF proposed in Sec. IV achieves the best performance compared to all the state-of-the-art methods. The previous SDF-based implicit generation methods fail to handle open surfaces, resulting in errors and fat double-layer typologies on the thin structures. UDiFF also significantly outperforms the recent work Surf-D [93] which focuses on the same task to generate unsigned distance fields. We also present the visual comparison of the 3D generations in Fig. 11, where UDiFF also achieves the best performance in generating visual-appealing 3D shapes with open and thin structures. We observe that LAS-Diffusion performs best in the SDF-based methods by producing smooth surfaces, while the point-based PointDiff struggles in generating clean shapes. Surf-D can generate open surfaces but struggle in producing high-fidelity details.

The benchmark conducted on the high-quality UWings dataset for 3D generation of shapes with open surfaces significantly contributes to the field of shape generation with arbitrary typologies. The benchmark provides a fair and com-

TABLE VI: Quantitative comparisons under the UWings dataset.

ScanID	S0	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	Mean
NeuS [94]	4.35	3.95	4.23	3.30	3.11	6.68	4.88	2.60	3.92	5.34	4.75	4.55	3.66	4.34	4.48	4.37	4.82	5.96	4.41
Instant-NGP [45]	5.20	4.91	4.39	3.33	3.43	9.37	5.48	2.65	4.41	4.32	5.23	5.27	3.76	15.18	4.74	3.74	3.50	5.02	5.28
NeAT [95]	4.30	3.65	4.54	3.55	3.37	5.39	4.79	2.89	4.55	3.26	5.84	4.70	4.31	4.21	3.80	5.11	4.87	5.52	4.37
NeuralUDF [57]	2.51	3.46	4.56	2.62	3.41	2.48	4.15	1.68	2.13	2.11	3.86	3.50	2.28	4.98	4.97	3.91	4.76	3.80	3.40
NeUDF [58]	3.42	2.86	3.28	1.89	2.15	2.93	2.41	1.57	2.07	1.64	2.12	4.16	2.31	2.46	2.40	2.11	2.57	3.35	2.54

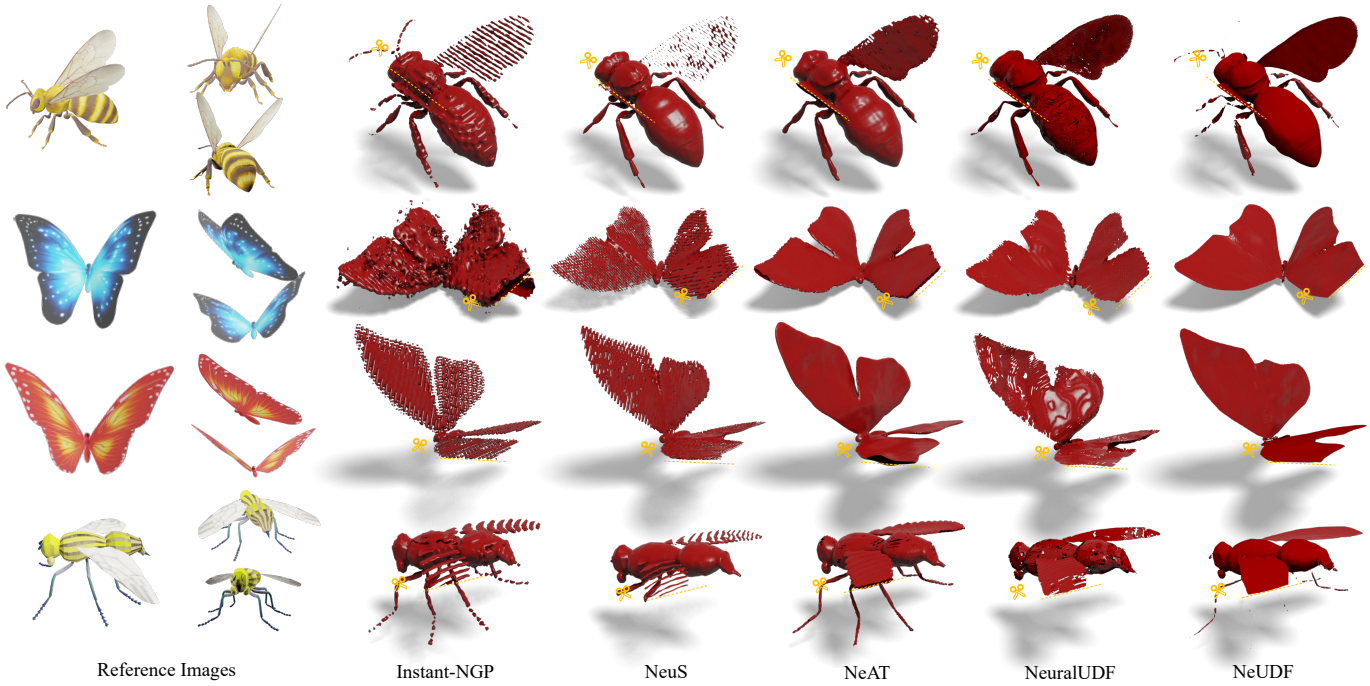

Fig. 12: Visual comparisons of surface reconstructions from multi-view images in UWings dataset.

prehensive comparison among the explicit and implicit-based 3D generation methods.

B. Surface Reconstruction from Multi-view Images

Neural surface reconstruction from multi-view images has been shown to be powerful for recovering 3D surfaces via image-based neural rendering. Most of the previous methods focus on learning signed distance fields [94] or occupancy fields [96] for modeling 3D geometries, but they are limited to reconstructing closed shapes. A series of recent methods optimize unsigned distance fields which are capable of modeling open structures. However, their performance under open-surface shapes are only evaluated limitedly under some simple garment models from DeepFashion3D [28] or MGN [97] dataset.

Data Settings and Metrics. For a comprehensive comparison among the implicit-based multi-view reconstruction methods on reconstructing open-surface 3D shapes with complex details, we conduct a benchmark on the proposed UWings dataset. We evaluate a wide range of baselines under 18 shapes from the UWings dataset. The shapes are chosen to have complex geometries and contain diverse categories and motions. We adopt the widely-used L1 Chamfer distance as the metric to evaluate the errors of the randomly sampled points on the reconstructed surfaces compare to the ones sampled on the ground truth meshes.

Baselines. We evaluate the multi-view reconstruction performance of the state-of-the-art methods, including NeuS [94], Instant-NGP [45], NeAT [95], NeuralUDF [57] and NeUDF [58]. NeuS and Instant-NGP are SDF-based and NeRF-based approaches, which can not handle shapes with open surfaces well. NeAT trains an SDF network with a validity probability function to mask out the extra surfaces extracted, and finally produces open surfaces. Neural-UDF and NeUDF are the pioneers in learning unsigned distance fields from multi-view images for open surface reconstruction.

Comparisons. Tab. 12 shows the quantitative comparison. The UDF-based methods (i.e., NeuralUDF and NeUDF) significantly outperform the NeuS and Instant-NGP. Visual comparisons are also provided in Fig. 12. The results demonstrate that unsigned distance fields provide a convincing solution for modeling open surfaces, and the UWings dataset provides a comprehensive benchmark for the multi-view surface reconstruction methods.

C. Surface Reconstruction from Point Clouds

Surface reconstruction from point clouds [11], [99], [101] plays an important role in computer graphics. Similar to the

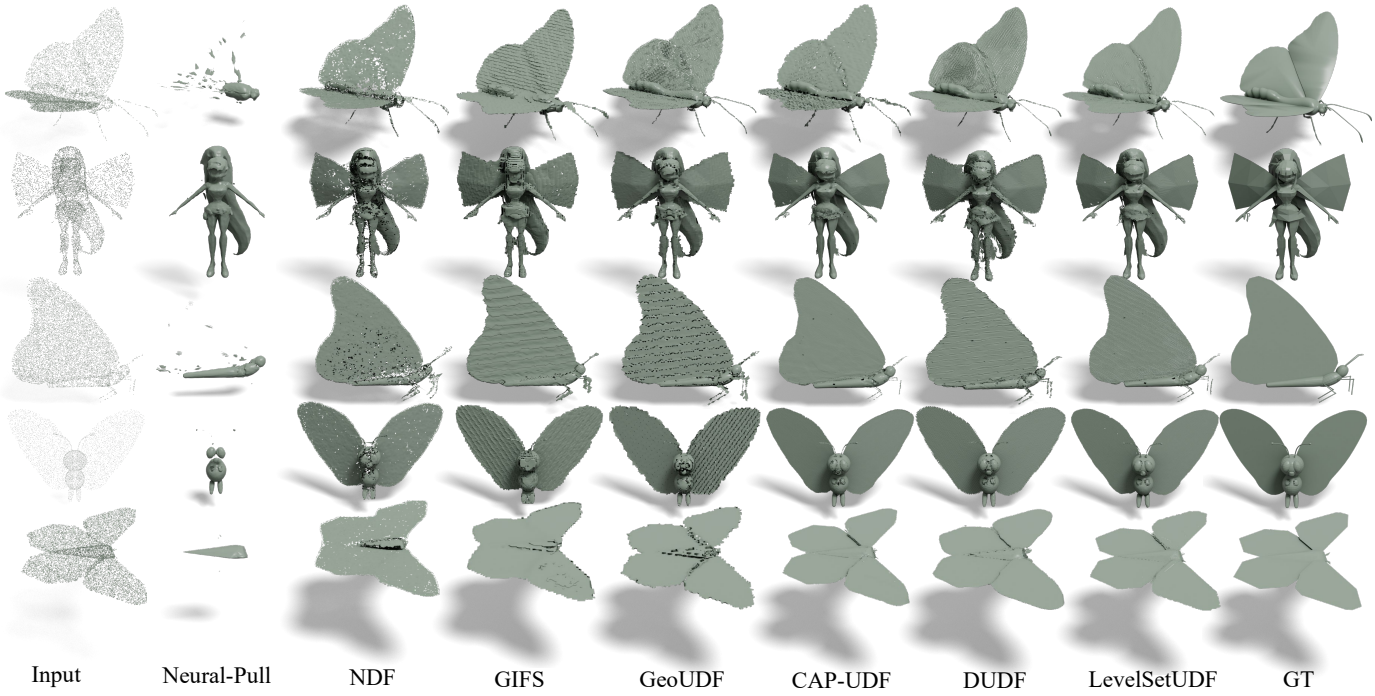

Fig. 13: Visual comparisons of surface reconstructions from point clouds in the UWings dataset.

TABLE VII: Surface reconstruction evaluation using Chamfer distances and F-scores. Chamfer-L1 and Chamfer-L2 values are reported, along with F-Score at thresholds 0.005 and 0.01.

Method	Chamfer-L1		Chamfer-L2		F-Score	
	Mean	Median	Mean	Median	$F1^{0.005}$	$F1^{0.01}$
NeuralPull [11]	2.229	0.885	41.798	3.242	62.44	68.85
NDF [18]	0.227	0.217	0.080	0.071	96.06	99.40
GIFS [60]	0.224	0.222	0.076	0.069	96.38	99.42
GeoUDF [98]	0.191	0.186	0.061	0.058	95.89	99.46
CAP-UDF [19], [99]	0.183	0.178	0.057	0.046	97.26	99.82
DUDF [100]	0.181	0.178	0.053	0.052	97.39	99.83
LevelSetUDF [59]	**0.175**	**0.170**	**0.045**	**0.039**	**97.75**	**99.90**

recent advances in 3D generation and multi-view reconstruction, the latest works learn unsigned distance fields [18], [19] to model open surfaces instead of the signed distance fields [10] and occupancy fields [12] adopt in previous methods.

Data Settings and Metrics. We aim to conduct a comprehensive benchmark on the Uwings dataset for evaluating the performance of existing approaches and for the subsequent works to make a fair and convenient comparison on reconstructing complex and open-surface geometries. We evaluate a wide range of baselines under the full set of the UWings dataset. For evaluating the performances, we follow the common setting [19], [60] to sample 100K points from the reconstructed surfaces and leverage Chamfer Distance, Normal Consistency [12] and F-score with a threshold of 0.005 / 0.01 as the evaluation metrics.

Baselines. We evaluate the state-of-the-art methods under the task of surface reconstruction from point cloud, including Neural-Pull [11], NDF [102], GIFS [60], CAP-UDF [19], [99], GeoUDF [98], DUDF [100] and LevelSetUDF [59]. Neural-Pull is the SDF-based method and GIFS learns a query relationship for modeling open surfaces. The rest methods are UDF-based methods, which are capable of representing arbitrary typologies.

Comparisons. We show the quantitative and qualitative comparison in Tab. 13 and Fig. 13. The SDF-based method Neural-Pull struggles in reconstructing thin geometries with open structures. The comprehensive comparisons among UDF-based methods demonstrate that LevelSetUDF achieves the best performance with high-quality and smooth reconstructions.

VII. Conclusion

In this work, we present UDiFF, a 3D diffusion model for generating textured 3D shapes with open and closed surfaces, either conditionally or unconditionally. We leverage a diffusion model to learn distributions of UDFs in a spatial-frequency space established through an optimal wavelet transformation for UDFs, which is learned by data-driven based self-reconstruction. The evaluations on widely used benchmarks show our superior performance over the latest methods in generating shapes with either open and closed surfaces. Additionally, we introduce the UWings dataset, which contains 1,509 high-quality 3D models of winged creatures which contains open surfaces, for shape modeling with UDFs. We establish comprehensive benchmarks on UWings dataset to provide a large-scale unbiased platform for evaluating the UDF-based methods.

References

[1] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics," in *International conference on machine learning*. PMLR, 2015, pp. 2256–2265.

Fig. 14: Category conditional generations under DeepFashion3D dataset.

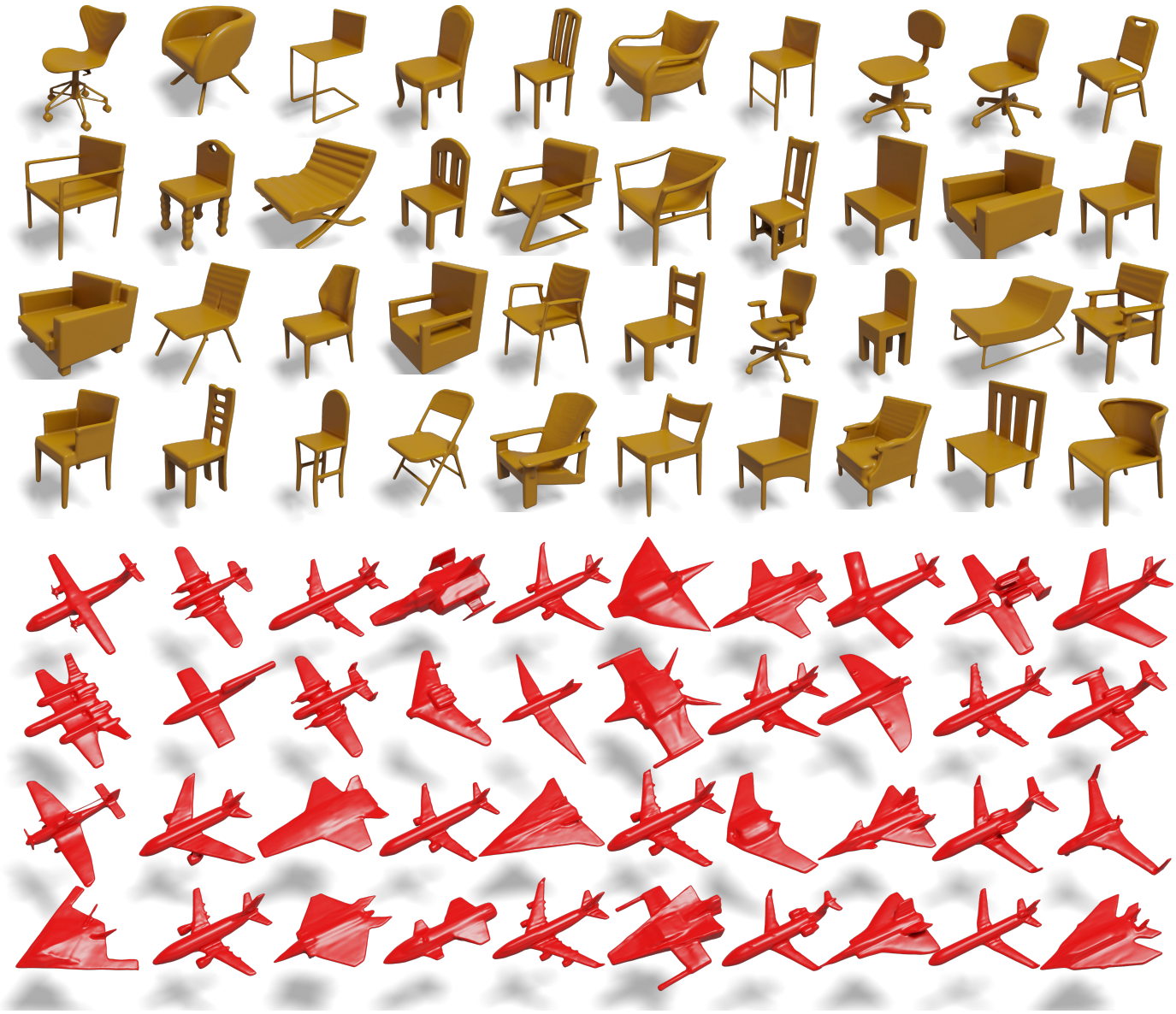

Fig. 15: Unconditional generations under the *"chair"* and *"airplane"* categories of the ShapeNet dataset.

[2] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.

[3] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen, "Hierarchical text-conditional image generation with clip latents," 2022.

[4] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 10 684–10 695.

[5] S. Luo and W. Hu, "Diffusion probabilistic models for 3d point cloud generation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 2837–2845.

[6] E. J. Smith and D. Meger, "Improved adversarial systems for 3d object generation and reconstruction," in *Conference on Robot Learning*. PMLR, 2017, pp. 87–96.

[7] K.-H. Hui, R. Li, J. Hu, and C.-W. Fu, "Neural wavelet-domain diffusion for 3d shape generation," in *SIGGRAPH Asia 2022 Conference Papers*, 2022, pp. 1–9.

[8] G. Chou, Y. Bahat, and F. Heide, "Diffusion-sdf: Conditional generative modeling of signed distance functions," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 2262–2272.

[9] B. Zhang, J. Tang, M. Niessner, and P. Wonka, "3dshape2vecset: A 3d shape representation for neural fields and generative diffusion models," *arXiv preprint arXiv:2301.11445*, 2023.

[10] J. J. Park, P. Florence, J. Straub, R. Newcombe, and S. Lovegrove, "DeepSDF: Learning continuous signed distance functions for shape representation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 165–174.

[11] B. Ma, Z. Han, Y.-S. Liu, and M. Zwicker, "Neural-Pull: Learning signed distance function from point clouds by learning to pull space onto surface," in *International Conference on Machine Learning*. PMLR, 2021, pp. 7246–7257.

[12] L. Mescheder, M. Oechsle, M. Niemeyer, S. Nowozin, and A. Geiger, "Occupancy networks: Learning 3D reconstruction in function space," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4460–4470.

[13] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[14] J. R. Shue, E. R. Chan, R. Po, Z. Ankner, J. Wu, and G. Wetzstein, "3d neural field generation using triplane diffusion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 20 875–20 886.

[15] A. Gupta, W. Xiong, Y. Nie, I. Jones, and B. Oğuz, "3DGen: Triplane latent diffusion for textured mesh generation," *arXiv preprint arXiv:2303.05371*, 2023.

[16] G. Nam, M. Khlifi, A. Rodriguez, A. Tono, L. Zhou, and P. Guerrero, "3D-LDM: Neural implicit 3d shape generation with latent diffusion models," *arXiv preprint arXiv:2212.00842*, 2022.

[17] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE transactions on information theory*, vol. 36, no. 5, pp. 961–1005, 1990.

[18] J. Chibane, G. Pons-Moll *et al.*, "Neural unsigned distance fields for implicit function learning," *Advances in Neural Information Processing Systems*, vol. 33, pp. 21 638–21 652, 2020.

[19] J. Zhou, B. Ma, Y.-S. Liu, Y. Fang, and Z. Han, "Learning consistency-aware unsigned distance functions progressively from raw point clouds," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

[20] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark *et al.*, "Learning transferable visual models from natural language supervision," in *International conference on machine learning*. PMLR, 2021, pp. 8748–8763.

[21] D. Z. Chen, Y. Siddiqui, H.-Y. Lee, S. Tulyakov, and M. Nießner, "Text2tex: Text-driven texture synthesis via diffusion models," *arXiv preprint arXiv:2303.11396*, 2023.

[22] J. Zhou, W. Zhang, B. Ma, K. Shi, Y.-S. Liu, and Z. Han, "UDiFF: Generating conditional unsigned distance fields with optimal wavelet diffusion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 21 496–21 506.

[23] A. X. Chang, T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su *et al.*, "Shapenet: An information-rich 3D model repository," *arXiv preprint arXiv:1512.03012*, 2015.

[24] R. Jensen, A. Dahl, G. Vogiatzis, E. Tola, and H. Aanæs, "Large scale multi-view stereopsis evaluation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 406–413.

[25] J. Collins, S. Goel, K. Deng, A. Luthra, L. Xu, E. Gundogdu, X. Zhang, T. F. Y. Vicente, T. Dideriksen, H. Arora *et al.*, "ABO: Dataset and benchmarks for real-world 3d object understanding," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 21 126–21 136.

[26] Q. Zhou and A. Jacobson, "Thingi10k: A dataset of 10,000 3d-printing models," *arXiv preprint arXiv:1605.04797*, 2016.

[27] M. Deitke, R. Liu, M. Wallingford, H. Ngo, O. Michel, A. Kusupati, A. Fan, C. Laforte, V. Voleti, S. Y. Gadre *et al.*, "Objaverse-xl: A universe of 10m+ 3d objects," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[28] H. Zhu, Y. Cao, H. Jin, W. Chen, D. Du, Z. Wang, S. Cui, and X. Han, "Deep fashion3d: A dataset and benchmark for 3d garment reconstruction from single images," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16*. Springer, 2020, pp. 512–530.

[29] P. Xiang, X. Wen, Y.-S. Liu, Y.-P. Cao, P. Wan, W. Zheng, and Z. Han, "Snowflake point deconvolution for point cloud completion and generation with skip-transformer," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 5, pp. 6320–6338, 2023.

[30] B. Ma, J. Zhou, Y.-S. Liu, and Z. Han, "Towards better gradient consistency for neural signed distance functions via level set alignment," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 17 724–17 734.

[31] X. Wen, J. Zhou, Y.-S. Liu, H. Su, Z. Dong, and Z. Han, "3D shape reconstruction from 2D images with disentangled attribute flow," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 3803–3813.

[32] W. Zhang, R. Xing, Y. Zeng, Y.-S. Liu, K. Shi, and Z. Han, "Fast learning radiance fields by shooting much fewer rays," *IEEE Transactions on Image Processing*, 2023.

[33] J. Zhou, B. Ma, W. Zhang, Y. Fang, Y.-S. Liu, and Z. Han, "Differentiable registration of images and lidar point clouds with voxelpoint-to-pixel matching," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

[34] C. Jin, T. Wu, and J. Zhou, "Multi-grid representation with field regularization for self-supervised surface reconstruction from point clouds," *Computers & Graphics*, 2023.

[35] J. Zhou, X. Wen, B. Ma, Y.-S. Liu, Y. Gao, Y. Fang, and Z. Han, "3D-OAE: Occlusion auto-encoders for self-supervised learning on point clouds," *IEEE International Conference on Robotics and Automation (ICRA)*, 2024.

[36] H. Huang, Y. Wu, J. Zhou, G. Gao, M. Gu, and Y.-S. Liu, "NeuSurf: On-surface priors for neural surface reconstruction from sparse input views," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024.

[37] S. Li, J. Zhou, B. Ma, Y.-S. Liu, and Z. Han, "Learning continuous implicit field with local distance indicator for arbitrary-scale point cloud upsampling," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024.

[38] B. Ma, H. Deng, J. Zhou, Y.-S. Liu, T. Huang, and X. Wang, "Geodream: Disentangling 2d and geometric priors for high-fidelity and consistent 3d generation," *arXiv preprint arXiv:2311.17971*, 2023.

[39] X. Wen, P. Xiang, Z. Han, Y.-P. Cao, P. Wan, W. Zheng, and Y.-S. Liu, "PMP-Net++: Point cloud completion by transformer-enhanced multi-step point moving paths," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, pp. 852–867, 2023.

[40] J. Zhou, J. Wang, B. Ma, Y.-S. Liu, T. Huang, and X. Wang, "Uni3D: Exploring unified 3d representation at scale," *International Conference on Learning Representations*, 2024.

[41] J. Zhou, X. Wen, Y.-S. Liu, Y. Fang, and Z. Han, "Self-supervised point cloud representation learning with occlusion auto-encoder," *arXiv e-prints*, pp. arXiv–2203, 2022.

[42] S. Li, J. Zhou, B. Ma, Y.-S. Liu, and Z. Han, "NeAF: Learning neural angle fields for point normal estimation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023.

[43] B. Ma, Y.-S. Liu, and Z. Han, "Learning signed distance functions from noisy 3d point clouds via noise to noise mapping," in *International Conference on Machine Learning (ICML)*, 2023.

[44] B. Mildenhall, P. Srinivasan, M. Tancik, J. Barron, R. Ramamoorthi, and R. Ng, "NeRF: Representing scenes as neural radiance fields for view synthesis," in *European Conference on Computer Vision*, 2020.

[45] T. Müller, A. Evans, C. Schied, and A. Keller, "Instant neural graphics primitives with a multiresolution hash encoding," *ACM Transactions on Graphics (ToG)*, vol. 41, no. 4, pp. 1–15, 2022.

[46] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein, "Implicit neural representations with periodic activation functions," *Advances in Neural Information Processing Systems*, vol. 33, pp. 7462–7473, 2020.

[47] T. Brooks, A. Holynski, and A. A. Efros, "Instructpix2pix: Learning to follow image editing instructions," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 18 392–18 402.

[48] G. Chou, I. Chugunov, and F. Heide, "Gensdf: Two-stage learning of generalizable signed distance functions," in *Advances in Neural Information Processing Systems*.

[49] J. Zhou, B. Ma, and Y.-S. Liu, "Fast learning of signed distance functions from noisy point clouds via noise to noise mapping," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[50] S. Peng, M. Niemeyer, L. Mescheder, M. Pollefeys, and A. Geiger, "Convolutional occupancy networks," in *European Conference on Computer Vision*. Springer, 2020, pp. 523–540.

[51] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3D surface construction algorithm," *ACM Siggraph Computer Graphics*, vol. 21, no. 4, pp. 163–169, 1987.

[52] C. Jiang, A. Sud, A. Makadia, J. Huang, M. Nießner, T. Funkhouser *et al.*, "Local implicit grid representations for 3D scenes," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 6001–6010.

[53] B. Ma, Y.-S. Liu, M. Zwicker, and Z. Han, "Surface reconstruction from point clouds by learning predictive context priors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

[54] B. Ma, Y.-S. Liu, and Z. Han, "Reconstructing surfaces for sparse point clouds with on-surface priors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

[55] W. Chen, C. Lin, W. Li, and B. Yang, "3PSDF: Three-pole signed distance function for learning surfaces with arbitrary topologies," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 18 522–18 531.

[56] L. Wang, J. Yang, W. Chen, X. Meng, B. Yang, J. Li, and L. Gao, "HSDF: Hybrid sign and distance field for modeling surfaces with arbitrary topologies," in *Advances in Neural Information Processing Systems*.

[57] X. Long, C. Lin, L. Liu, Y. Liu, P. Wang, C. Theobalt, T. Komura, and W. Wang, "NeuralUDF: Learning unsigned distance fields for multi-view reconstruction of surfaces with arbitrary topologies," *arXiv preprint arXiv:2211.14173*, 2022.

[58] Y.-T. Liu, L. Wang, J. Yang, W. Chen, X. Meng, B. Yang, and L. Gao, "NeUDF: Leaning neural unsigned distance fields with volume

rendering," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 237–247.
[59] J. Zhou, B. Ma, S. Li, Y.-S. Liu, and Z. Han, "Learning a more continuous zero level set in unsigned distance fields through level set projection," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2023.
[60] J. Ye, Y. Chen, N. Wang, and X. Wang, "GIFS: Neural implicit function for general shape representation," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.
[61] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.
[62] D. P. Kingma and P. Dhariwal, "GLOW: Generative flow with invertible 1x1 convolutions," *Advances in neural information processing systems*, vol. 31, 2018.
[63] L. Hui, R. Xu, J. Xie, J. Qian, and J. Yang, "Progressive point cloud deconvolution generation network," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XV 16*. Springer, 2020, pp. 397–413.
[64] R. Li, X. Li, K.-H. Hui, and C.-W. Fu, "SP-GAN: Sphere-guided 3D shape generation and manipulation," *ACM Transactions on Graphics (TOG)*, vol. 40, no. 4, pp. 1–12, 2021.
[65] R. Cai, G. Yang, H. Averbuch-Elor, Z. Hao, S. Belongie, N. Snavely, and B. Hariharan, "Learning gradient fields for shape generation," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*. Springer, 2020, pp. 364–381.
[66] G. Yang, X. Huang, Z. Hao, M.-Y. Liu, S. Belongie, and B. Hariharan, "Pointflow: 3d point cloud generation with continuous normalizing flows," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 4541–4550.
[67] A. Nichol, H. Jun, P. Dhariwal, P. Mishkin, and M. Chen, "Point-e: A system for generating 3d point clouds from complex prompts," *arXiv preprint arXiv:2212.08751*, 2022.
[68] J. Wu, C. Zhang, T. Xue, B. Freeman, and J. Tenenbaum, "Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling," *Advances in neural information processing systems*, vol. 29, 2016.
[69] L. Zhou, Y. Du, and J. Wu, "3d shape generation and completion through point-voxel diffusion," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 5826–5835.
[70] A. Vahdat, F. Williams, Z. Gojcic, O. Litany, S. Fidler, K. Kreis *et al.*, "Lion: Latent point diffusion models for 3d shape generation," *Advances in Neural Information Processing Systems*, vol. 35, pp. 10 021–10 039, 2022.
[71] J. Gao, T. Shen, Z. Wang, W. Chen, K. Yin, D. Li, O. Litany, Z. Gojcic, and S. Fidler, "Get3d: A generative model of high quality 3d textured shapes learned from images," *Advances In Neural Information Processing Systems*, vol. 35, pp. 31 841–31 854, 2022.
[72] M. Li, Y. Duan, J. Zhou, and J. Lu, "Diffusion-sdf: Text-to-shape via voxelized diffusion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 12 642–12 651.
[73] Y.-C. Cheng, H.-Y. Lee, S. Tulyakov, A. G. Schwing, and L.-Y. Gui, "Sdfusion: Multimodal 3d shape completion, reconstruction, and generation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 4456–4465.
[74] J. Koo, S. Yoo, M. H. Nguyen, and M. Sung, "SALAD: Part-level latent diffusion for 3d shape generation and manipulation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 14 441–14 451.
[75] H. Fu, R. Jia, L. Gao, M. Gong, B. Zhao, S. Maybank, and D. Tao, "3D-Future: 3D furniture shape with texture," *International Journal of Computer Vision*, vol. 129, pp. 3313–3337, 2021.
[76] P. Erler, P. Guerrero, S. Ohrhallinger, N. J. Mitra, and M. Wimmer, "Points2surf learning implicit surfaces from point clouds," in *European Conference on Computer Vision*. Springer, 2020, pp. 108–124.
[77] T. Wu, J. Zhang, X. Fu, Y. Wang, J. Ren, L. Pan, W. Wu, L. Yang, J. Wang, C. Qian *et al.*, "Omniobject3d: Large-vocabulary 3d object dataset for realistic perception, reconstruction and generation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 803–814.
[78] M. Deitke, D. Schwenk, J. Salvador, L. Weihs, O. Michel, E. VanderBilt, L. Schmidt, K. Ehsani, A. Kembhavi, and A. Farhadi, "Objaverse: A universe of annotated 3d objects," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 13 142–13 153.
[79] T. Luo, C. Rockwell, H. Lee, and J. Johnson, "Scalable 3D captioning with pretrained models," *Advances in Neural Information Processing Systems*, vol. 36, 2024.
[80] T. Luo, J. Johnson, and H. Lee, "View selection for 3D captioning via diffusion ranking," in *European Conference on Computer Vision*. Springer, 2024, pp. 180–197.
[81] J. Li, D. Li, S. Savarese, and S. Hoi, "BLIP-2: Bootstrapping language-image pre-training with frozen image encoders and large language models," in *International conference on machine learning*. PMLR, 2023, pp. 19 730–19 742.
[82] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat *et al.*, "GPT-4 technical report," *arXiv preprint arXiv:2303.08774*, 2023.
[83] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.
[84] A. Q. Nichol and P. Dhariwal, "Improved denoising diffusion probabilistic models," in *International Conference on Machine Learning*. PMLR, 2021, pp. 8162–8171.
[85] B. Guillard, F. Stella, and P. Fua, "MeshUDF: Fast and differentiable meshing of unsigned distance field networks," *European Conference on Computer Vision*, 2022.
[86] F. Hou, X. Chen, W. Wang, H. Qin, and Y. He, "Robust zero level-set extraction from unsigned distance fields based on double covering," *ACM Transactions on Graphics (TOG)*, vol. 42, no. 6, pp. 1–15, 2023.
[87] X.-Y. Zheng, H. Pan, P.-S. Wang, X. Tong, Y. Liu, and H.-Y. Shum, "Locally attentional SDF diffusion for controllable 3D shape generation," *ACM Transactions on Graphics (ToG)*, vol. 42, no. 4, pp. 1–13, 2023.
[88] J. Huang, H. Su, and L. Guibas, "Robust watertight manifold surface generation method for shapenet models," *arXiv preprint arXiv:1802.01698*, 2018.
[89] Z. Chen and H. Zhang, "Learning implicit fields for generative shape modeling," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5939–5948.
[90] M. Kleineberg, M. Fey, and F. Weichert, "Adversarial generation of continuous implicit shape representations," *arXiv preprint arXiv:2002.00349*, 2020.
[91] A. Hertz, O. Perel, R. Giryes, O. Sorkine-Hornung, and D. Cohen-Or, "SPAGHETTI: Editing implicit shapes through part aware generation," *ACM Transactions on Graphics (TOG)*, vol. 41, no. 4, pp. 1–20, 2022.
[92] P. Mittal, Y.-C. Cheng, M. Singh, and S. Tulsiani, "Autosdf: Shape priors for 3d completion, reconstruction and generation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 306–315.
[93] Z. Yu, Z. Dou, X. Long, C. Lin, Z. Li, Y. Liu, N. Müller, T. Komura, M. Habermann, C. Theobalt *et al.*, "Surf-D: High-quality surface generation for arbitrary topologies using diffusion models," in *European Conference on Computer Vision*. Springer, 2024.
[94] P. Wang, L. Liu, Y. Liu, C. Theobalt, T. Komura, and W. Wang, "NeuS: Learning neural implicit surfaces by volume rendering for multi-view reconstruction," *Advances in Neural Information Processing Systems*, vol. 34, pp. 27 171–27 183, 2021.
[95] X. Meng, W. Chen, and B. Yang, "Neat: Learning neural implicit surfaces with arbitrary topologies from multi-view images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 248–258.
[96] M. Oechsle, S. Peng, and A. Geiger, "Unisurf: Unifying neural implicit surfaces and radiance fields for multi-view reconstruction," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 5589–5599.
[97] B. L. Bhatnagar, G. Tiwari, C. Theobalt, and G. Pons-Moll, "Multi-garment net: Learning to dress 3D people from images," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 5420–5430.
[98] S. Ren, J. Hou, X. Chen, Y. He, and W. Wang, "GeoUDF: Surface reconstruction from 3D point clouds via geometry-guided distance representation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 14 214–14 224.
[99] J. Zhou, B. Ma, S. Li, Y.-S. Liu, Y. Fang, and Z. Han, "CAP-UDF: Learning unsigned distance functions progressively from raw point clouds with consistency-aware field optimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.
[100] M. Fainstein, V. Siless, and E. Iarussi, "DUDF: Differentiable unsigned distance fields with hyperbolic scaling," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 4484–4493.

[101] S. Peng, C. Jiang, Y. Liao, M. Niemeyer, M. Pollefeys, and A. Geiger, "Shape as points: A differentiable poisson solver," *Advances in Neural Information Processing Systems*, vol. 34, 2021.
[102] J. Chibane, T. Alldieck, and G. Pons-Moll, "Implicit functions in feature space for 3D shape reconstruction and completion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 6970–6981.

Junsheng Zhou received the B.S. degree in software engineering from Xiamen University, China, in 2021, and the M.S. degree from the School of Software, Tsinghua University, Beijing, China, in 2024. He is currently working toward the Ph.D. degree in the School of Software, Tsinghua University, Beijing, China. His research interests include deep learning, 3D shape analysis and 3D reconstruction.

Weiqi Zhang received the B.S. degree in computer science from Southwest Jiaotong University, China, in 2023. He is currently working toward the M.S. degree in the School of Software, Tsinghua University, Beijing, China. His research interests include deep learning, 3D computer vision and 3D reconstruction.

Baorui Ma received the B.S. degree in computer science and technology from Jilin University, China, in 2018, and the PhD degree from the School of Software, Tsinghua University, Beijing, China, in 2023. He is currently a researcher with Beijing Academy of Artificial Intelligence, BAAI. His research interests include deep learning and 3D reconstruction.

Kanle Shi received his B.S. degree in 2007, and Ph.D. degree in 2012, from Tsinghua University. From 2012 to 2013, he was a Post-Doctoral Researcher in INRIA, on computational geometry. From 2014 to 2019, he was an assistant/associate researcher at the School of Software, Tsinghua University. He is currently a researcher in Kuaishou Technology. His research interests cover computer graphics, geometry and machine learning.

Yu-Shen Liu (M'18) received the B.S. degree in mathematics from Jilin University, China, in 2000, and the Ph.D. degree from the Department of Computer Science and Technology, Tsinghua University, Beijing, China, in 2006. From 2006 to 2009, he was a Post-Doctoral Researcher with Purdue University. He is currently an Associate Professor with the School of Software, Tsinghua University. His research interests include shape analysis, pattern recognition, machine learning, and semantic search.

Zhizhong Han received the Ph.D. degree from Northwestern Polytechnical University, China, 2017. He was a Post-Doctoral Researcher with the Department of Computer Science, at the University of Maryland, College Park, USA. Currently, he is an Assistant Professor of Computer Science at Wayne State University, USA. His research interests include 3D computer vision, digital geometry processing and artificial intelligence.