From Air to Wear: Personalized 3D Digital Fashion with AR/VR Immersive 3D Sketching

Ying Zang, Yuanqi Hu, Xinyu Chen, Yuxia Xu, Suhui Wang, Chunan Yu, Lanyun Zhu, Deyi Ji, Xin Xu, Tianrun Chen*

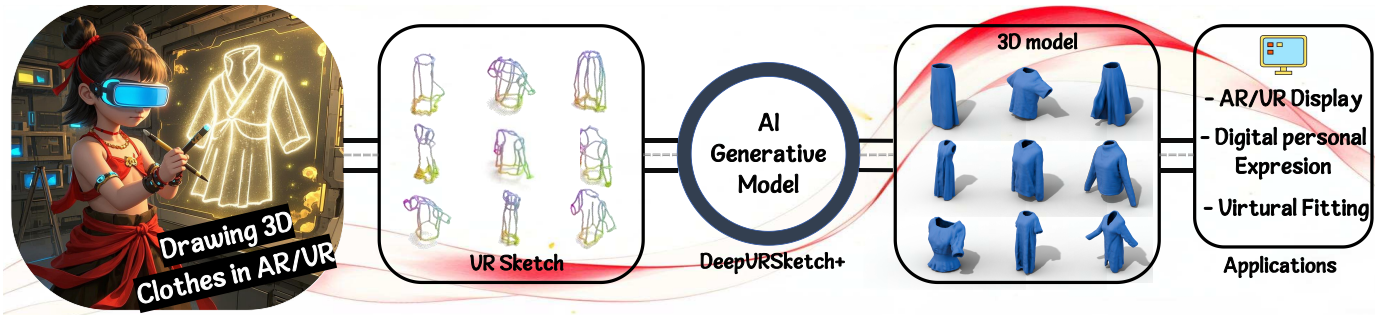

Fig. 1. In this work, we propose a novel method that allows everyday users to create personalized 3D garments by simply sketching in immersive AR/VR environments (in 3D). A carefully designed generative AI model transforms these freehand 3D sketches into realistic, detailed garment models, which can be used for personalized expression in the metaverse, AR/VR visualization, and virtual try-on applications.

***Abstract*—In the era of immersive consumer electronics, such as AR/VR headsets and smart devices, people increasingly seek ways to express their identity through virtual fashion. However, existing 3D garment design tools remain inaccessible to everyday users due to steep technical barriers and limited data. In this work, we introduce a 3D sketch-driven 3D garment generation framework that empowers ordinary users — even those without design experience — to create high-quality digital clothing through simple 3D sketches in AR/VR environments. By combining a conditional diffusion model, a sketch encoder trained in a shared latent space, and an adaptive curriculum learning strategy, our system interprets imprecise, free-hand input and produces realistic, personalized garments. To address the scarcity of training data, we also introduce KO3DClothes, a new dataset of paired 3D garments and user-created sketches. Extensive experiments and user studies confirm that our method significantly outperforms existing baselines in both fidelity and usability, demonstrating its promise for democratized fashion design on next-generation consumer platforms.**

***Index Terms*—AR/VR, 3D Sketch, Shape-from-X, Content Creation, Metaverse.**

I. Introduction

CLOTHING is one of the most personal and powerful forms of self-expression [1]–[3]. People don't just wear clothes — they choose them to reflect their identity, taste, and mood. In the digital age, clothing is no longer limited to the physical world. As AR/VR technologies become part of everyday consumer electronics — from headsets and smart mirrors to mobile AR apps — people are spending more and more time in immersive virtual spaces. In these spaces, clothing still matters. Just like in real life, people want to express their identity through what they wear in the virtual world. But unlike the real world, where options are limited by manufacturing and logistics, digital garments can be limitless — as long as we give people the tools to create them.

However, for a long time, fashion design was a privilege reserved for the elite — professional designers with years of training and expensive tools [4], [5]. Designing even a single 3D garment required mastering complex modeling software and navigating tedious workflows. Ordinary users still don't have accessible tools to design clothing in 3D. That's the gap we aim to fill. Just as AR headsets and VR devices are becoming more affordable and user-friendly — turning once high-end tech into everyday consumer products — we believe 3D fashion creation should follow the same path. Therefore, in this work, we develop a fashion design tool that is just as democratized - intuitive, creative, and open to all, with the growing accessibility of AR/VR platforms.

To make 3D clothing creation accessible to everyone, the first challenge we need to solve is how to simplify the design process. Traditional 3D modeling tools are complex and time-consuming, often requiring years of training and professional software — far beyond the reach of everyday users [6]. But with the rise of consumer-grade AR/VR devices, the way people interact with 3D space is rapidly changing. Here, we innovatively propose to use AR/VR devices as the creation tool.

AR/VR headsets not only let users view 3D objects in immersive environments, but also allow them to create directly in 3D space. Past researches have shown that with motion controllers or hand tracking, anyone can sketch in the air — drawing curves and shapes as if they were sculpting with invisible tools [6]–[10]. We build on this capability by pairing

This work is supported by the Public Welfare Research Program of Huzhou Science and Technology Bureau (2022GZ01) and ZJU Kunpeng & Ascend Center of Excellence. (*Corresponding author: Tianrun Chen*)

Ying Zang, Yuanqi Hu, Xinyu Chen, Yuxia Xu, Suhui Wang and Chunan Yu are with the School of Information Engineering, Huzhou University. (e-mail: 02750@zjhu.edu.cn)

Lanyun Zhu is with the Information Systems Technology and Design Pillar, Singapore University of Technology and Design.

Deyi Ji is with the School of Information Science and Technology, University of Science and Technology of China

Xin Xu is with KOKONI, Moxin (Huzhou) Technology Co., LTD.

Tianrun Chen is with the College of Computer Science and Technology, Zhejiang University and KOKONI, Moxin (Huzhou) Technology Co., LTD. (email: tianrun.chen@zju.edu.cn)

*Corresponding Author

these 3D sketches with our carefully designed generative AI model, which instantly transforms rough strokes into high-quality 3D garments that match the user's vision (see main figure). With AI generative model, we removes the need for accurate drawing or manual modeling. Instead, users can focus on expressing ideas, while the AI fills in the structure and detail. We believe sketching is one of the most natural and powerful ways for people to communicate creative ideas. It's quick, intuitive, and nearly everyone can do it [5].

The real challenge, though, is making sure the system understands the sketch — especially when it's loose or imprecise. For beginners, drawing precise lines is often difficult, especially when the clothing has a freely shaped surface, making it challenging to accurately depict. But clothing is complex, often with soft, flowing shapes that are hard to capture in rough lines. While past methods rely on supervised learning or direct regression to match sketches to models [11], [12], they often struggle when the sketch is unclear. Our goal is to go further: to create a system that's not only easy to use, but also forgiving — one that helps all users, regardless of skill, generate high-quality, faithful 3D designs.

To address the challenge, we adopt a generative approach instead of determinstic regression to obtain 3D design. Rather than directly regressing a 3D model from a user's sketch — a process that often requires precise input — our network treats the sketch as a loose condition. This allows the model to "imagine" a complete and realistic 3D garment that aligns with the user's intent, even when the input is rough or incomplete. The generative model learns to interpret abstract lines and transform them into high-fidelity 3D shapes with natural structure and flow, lowering the technical barrier for users.

However, generative models typically rely on large datasets to perform well — and in the 3D fashion domain, data scarcity remains a pressing issue. The most widely used dataset, provided by Zhu et al. [11], contains only 1,212 3D garments and lacks paired human-drawn sketches. This makes it difficult to train reliable, generalizable models. As a result, finding ways to fully leverage this limited data becomes a key challenge in making sketch-based generative 3D clothing design truly practical and accessible.

To tackle the limitations posed by scarce training data, we introduce a three-stage strategy that leverages a shared 3D point cloud-based latent space, curriculum learning, and a newly collected dataset. First, we pre-train a conditional diffusion model on a large-scale, diversified dataset of clothes 3D shapes and point clouds. This diffusion model serves as a strong shape generator, producing high-quality garments from abstract latent features. In the second stage, we freeze the pre-trained diffusion model and train a sketch encoder that maps a hand-drawn 3D sketch into the same latent space. These encoded features are injected into the intermediate layers of the diffusion model to condition the generation process, allowing it to adaptively synthesize garments that reflect the user's input sketch while preserving shape realism and plausibility. Finally, we jointly fine-tune the entire system — the sketch encoder and the diffusion generator — to further enhance the alignment between user sketches and generated garments. During training, we observed that the model struggled to generalize across diverse sketching styles and garment geometries, especially under limited supervision. To mitigate this, we adopt an adaptive curriculum learning strategy that gradually increases the complexity of training samples, helping the model learn from simpler shapes before progressing to more intricate examples.

To further alleviate data scarcity, we contribute a new dataset — KO3DClothes — which includes paired 3D garments and human-drawn 3D VR sketches. We select 3D garment models from the DeepFashion3D dataset [11] and invite 10 non-professional participants to create corresponding sketches using custom VR sketching tools. The resulting dataset contains 969 high-quality paired samples, enriching the available resources for research on sketch-based 3D garment generation and facilitating future work in this direction.

Through comprehensive experimental validation, our method outperforms existing benchmarks in both model quality and realism, even generating satisfactory results for previously unseen data drawn by novice users. Additionally, we demonstrate that, even with intricate and complex structures, the system can successfully generate accurate 3D models based on detailed 3D drawings from users. In user studies, participants expressed high satisfaction with the generated 3D models, further validating the practicality and effectiveness of our approach. We believe that our work takes a step toward democratizing digital fashion by making 3D garment design intuitive, expressive, and widely accessible for next-generation consumer platforms.

II. Related Works

A. 3D Garment Design with Deep Learning Algorithm.

Given the significance of 3D garment design, many methods have been proposed [13]–[18] to reconstruct or digitize 3D garments from images. Some approaches [11], [19]–[21] employ neural implicit representations, such as occupancy fields and Signed Distance Functions (SDF), to construct 3D models of clothed humans from single-view or sparse multi-view images using supervised learning. However, these methods face limitations when it comes to independently modeling garments separate from the body. ReEF [15] addresses this issue by employing a technique that independently models garments through the learning of explicit boundary curves and segmentation fields. In contrast, xCloth [16] offers a more efficient representation with the added advantage of generating texture maps. Nonetheless, these approaches still rely on high-quality real-world datasets of clothed humans, which often lack diversity in style and appearance, mainly due to the high cost of acquiring large-scale datasets. Furthermore, the generated garments are usually tightly coupled with the underlying body pose, often resulting in suboptimal surface quality.

Another research direction, inspired by the actual garment creation process, has proposed both analytical [22] and neural network-based [23] methods for procedurally generating unposed 3D garments ready for production. However, these methods rely on complex sewing patterns, which are not

intuitive for designers. More recent approaches [24], [25] have bypassed panel-based generation by using parametric human body templates to generate garment models. However, these methods typically focus on modeling tight-fitting garments and need substantial expertise to create the 3D garment.

In contrast to previous approaches, our method exclusively uses **VR sketches** as the input modality, allowing novice users to create the 3D garment without worrying about the issue of spatial perception in 3D in 2D and drawing skills. There are some existing sketch-based 3D garment design methods [12], [26]–[28], but they are limited by the 2D input. The user needs to "imagine" 3D before drawing, which is hard. Despite the input differences, we also adopt a more complex but effective network configuration and training scheme compared to these previous works to tackle the data scarcity and low output quality issue. We design a multi-stage generative network that, even with limited training data, can generate high-quality 3D shapes while accurately capturing the user's design intentions.

B. 3D Model Generation with Generative Models

In recent years, significant progress has been made in 3D shape generation. Many studies have explored various generative models, including Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), autoregressive models, normalizing flows, and more recently, diffusion models. In this study, we adopt a diffusion model-based network architecture, which has achieved state-of-the-art results in the field of 3D shape generation [29]–[31], capable of producing high-quality and highly detailed 3D shapes.

Additionally, 3D model generation methods based on images and text have also made significant strides in recent years [32]–[37]. However, in contrast to traditional approaches, we introduce an innovative input modality—3D VR sketches, to the 3D garment creation. We argue that using 3D VR sketches as input in AI generative models offers unique advantages over other methods. **Image input** struggles with freely creating 3D models from scratch, while **text input** is much less intuitive and precise in conveying spatial and geometric information compared to hand-drawn sketches.

C. 2D Sketches or 3D Sketches

As sketches are natural form of computer-human interaction, some methods for 3D model generation from **2D sketches** have been proposed [6], [38]–[49]. However, researches have found that 2D sketches are inherently ambiguous and abstract, leading to issues with occlusion and information loss when the viewpoint is limited. Creating view-consistent 2D sketches by hand is also nearly impossible. In contrast, with the growing popularity in AR/VR devices as consumer electronics, **3D VR sketches** can convey more comprehensive information, such as accurately describing internal features of complex objects, and also easier for users to understand, which have been confirmed by previous research [50]. In this research, we are the first to our knowledge to expand the 3D VR sketches in the field of 3D garment creation.

III. KO3DClothes Dataset

Given the limited availability of datasets for VR sketch-to-3D garment generation, we first create a novel dataset, KO3DClothes. To ensure realism and capture the nuances of human-created sketches, we opted for manual annotation by human volunteers rather than relying on synthetic data. Sketches, unlike simple edge detections, inherently reflect human intention and imprecision, and only through manual annotation can we accurately capture the errors and subtle variations characteristic of hand-drawn strokes.

We invited 10 participants to create sketches on existing 3D models from the DeepFashion3D dataset [11]. During data collection, participants wore VR headsets and used controllers to draw sketches in a virtual environment, while employing dedicated software for the creation process. The system tracked the controllers' movements in real-time via VR. The data collection protocol was adapted from previous studies [50], [51]. Specifically, participants were asked to draw 3D sketches around manually crafted 3D models within a predefined boundary box. Each participant's strokes were recorded as a series of 3D coordinates, forming point cloud data that represents the 3D structure of the sketch. In each sample, the point cloud data was uniformly sampled with N=4096 points. Furthermore, the sketches were manually aligned to ensure consistent X, Y, and Z directions and positions, providing a standardized reference frame for accurate comparison and analysis. As a result, we obtained aligned and paired 3D sketches and 3D models. The 3D models used were selected from the high-quality models of the DeepFashion3D dataset [11], consisting of 969 unique 3D shapes. Each 3D sketch-shape pair went through a quality control step by another participant. Fig. 2 presents representative examples of the user-created 3D sketches of 3D clothes.

Fig. 2. The visualization of hand-drawn 3D sketch samples from KO3DClothes dataset.

IV. Method

A. Preliminary: Conditional 3D Diffusion Model

1) Principle: Our method employs a conditional diffusion model to generate 3D shapes, which have shown excellent performance in generating diverse and high-quality 3D models. We train the diffusion model by reversing the noise diffusion process to sample from the target distribution. Given a sample z, we gradually add Gaussian noise z_t to the sample based

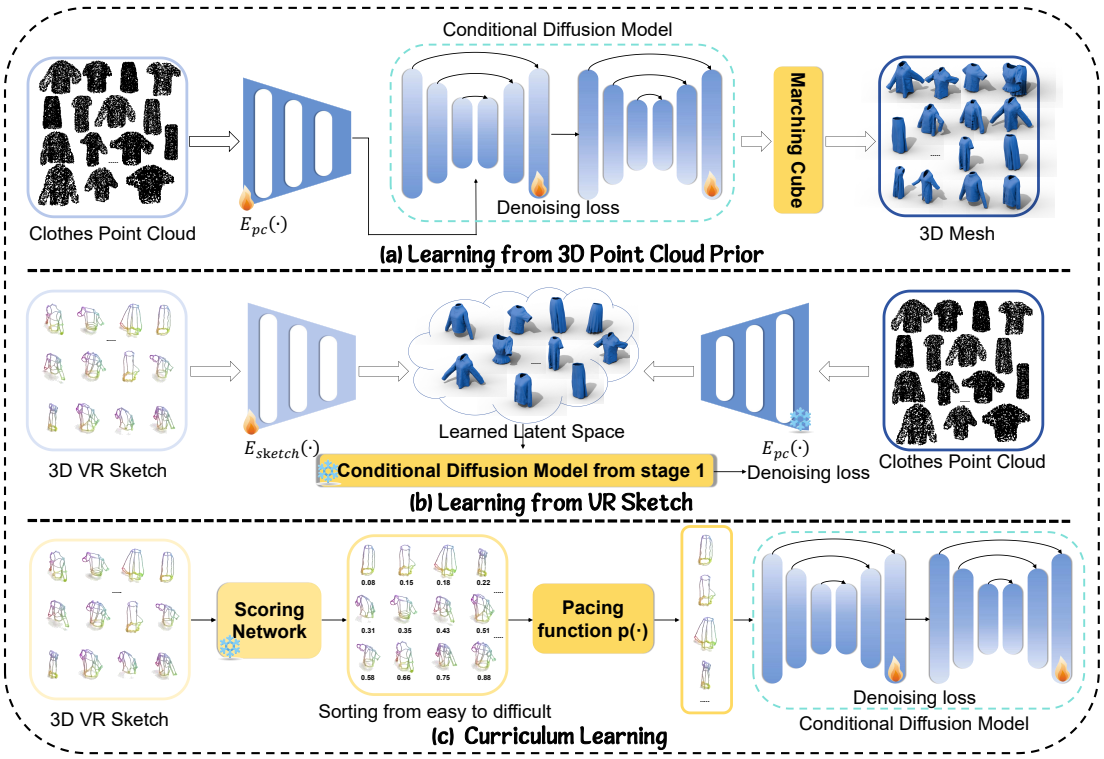

Fig. 3. The Overview of Deep3DVRSketch+. (a) Pre-training a conditional diffusion model by sampling ground truth (GT) point clouds. (b) Fine-tuning the sketch encoder to project sketches onto the diffusion manifold. (c) Curriculum learning leverages a limited set of sketch-shape pairs.

on a pre-defined variance schedule, with t ranging from 1 to T, producing the corresponding z_t. Next, we use a time-conditioned 3D UNet (denoted as ϵ_θ) to denoise the noise. Finally, the UNet generates a new 3D shape from the denoised Gaussian noise sample.

2) Coarse-to-Fine Diffusion Network: To achieve high-fidelity 3D shape representations, we use high-resolution discrete signed distance fields (SDFs) to accurately model shape details. However, directly generating a dense SDF grid can incur substantial computational and memory costs due to its cubic complexity. To mitigate this computational burden while ensuring high-quality model output, we reference the method proposed in [52] and design a two-stage diffusion framework, employing a self-conditioned continuous diffusion model. Specifically, the first stage uses a 5-layer UNet to generate a low-resolution 3D occupancy volume $C \in R^{n\times n\times n}$, providing a coarse representation of the 3D shape. Next, the second stage constructs a high-resolution sparse volume $F \in R^{N\times N\times N}$ using a 4-layer UNet, where an octree convolutional neural network is employed to handle the sparse voxel format of the SDF data. Both UNets are trained with a denoising loss [53], which is formulated as:

$$L(\theta) = \mathbb{E}_{z,\epsilon\sim N(0,1),t}\left[\left\|\epsilon - \epsilon_\theta\left(z_t, t, \{\mathbf{c}_i\}_{i=1}^N\right)\right\|^2\right] \quad (1)$$

where $\mathcal{N}(0,1)$ represents the Gaussian distribution, and $\{\mathbf{c}_i\}_{i=1}^N$ denotes the conditions required for the generative process. In the implementation, n is set to 32, and N is set to 128.

3) Learning the Conditional Distribution: We use the conditional diffusion model as the core structure to generate 3D content, where the user-provided condition $\{\mathbf{c}_i\}_{i=1}^N$ is injected into the model to generate results based on the user's needs. Conditional features are extracted through a dedicated encoder, which converts condition data, such as point clouds, into 1024-dimensional feature embeddings l. These feature embeddings are then integrated into the UNet architecture using a multi-head cross-attention mechanism. Due to the limitations of the conditional input, the diffusion model may restrict the diversity of generated samples. We use classifier-free guidance [54] to mitigate this issue.

B. The Proposed Multi-stage Training Steategy

In this section, we introduce our multi-stage training strategy. We first train a 3D diffusion generator that can generate high-quality shapes based on processed point cloud data, which is provided by the point cloud encoder $E_{pc}(\cdot)$. This process generates a feature encoding l, which guides the diffusion model for fine-tuning and generates realistic 3D shapes. Then, in the 3D sketch mapping stage, we keep the diffusion generator unchanged and train the sketch encoder $E_{sketch}(\cdot)$ to map the input 3D sketch S_{vr} into the latent space Z from the first stage. It is important to note that while the output at this stage is close to the latent space, it still deviates from perfect alignment with the pre-trained input. Therefore, we introduce a joint fine-tuning stage, where both the encoder $E(\cdot)$ and the diffusion generator are fine-tuned simultaneously to optimize the alignment and improve the quality of alignment between $E_{pc}(\cdot)$ and $E_{sketch}(\cdot)$. This multi-stage training approach helps us fully leverage the knowledge of the pre-trained model and has been proven to be crucial for improving the final output quality.

1) Stage 1: Pre-Training the 3D Priors: In this stage, our primary goal is to train the conditional diffusion model to

generate 3D shapes in the latent space, providing important prior information for the subsequent 3D sketch-to-3D model conversion. To achieve this, we use 3D point cloud data as conditional input and employ a supervised diffusion model for generation. Specifically, we use a pre-trained point cloud encoder, Uni3D [55], which plays a key role in encoding 3D point cloud data. This encoder $E_{pc}(\cdot)$ converts the 3D point cloud into a latent encoding $l_{pc}^{+} = E_{pc}(S_{pc}) \in \mathbb{R}^{1024}$ in the diffusion model. This approach allows the model to accurately capture the latent morphological features present in different 3D garment shape datasets and generate high-quality 3D shapes.

2) Stage 2: 3D Sketch Mapping: In this stage, we input hand-drawn 3D sketches into the system and map them to the latent space Z defined in Stage 1, ensuring alignment between the sketches and the point cloud data. During this stage, we keep the diffusion model pre-trained in the first stage unchanged. Since 3D sketches are represented through point clouds, we design a Transformer-based point cloud encoder $E_{sketch}(\cdot)$ to map the sketch S_{vr} into a latent encoding $l_{vr}^{+} = E_{sketch}(S_{vr}) \in \mathbb{R}^{1024}$, with dimensions consistent with the Uni3D features from the previous stage.

3) Stage 3: Joint Fine-Tuning: We found that simply adjusting the 3D sketch encoder $E_{sketch}(\cdot)$ does not ensure optimal alignment between the generated shapes and the sketch input. Inspired by certain approaches in image diffusion generation methods [56], we decided to simultaneously fine-tune both the 3D sketch encoder $E_{sketch}(\cdot)$ and the diffusion model to enhance spatial-semantic alignment. This strategy helps to fully leverage the knowledge of the pre-trained model and is crucial for improving the quality of the generated results.

Experimental results show that, although the fine-tuned model generates results of high-quality, there is still room for improvement, especially in model details, as shown in IV. To address this challenge, we propose an adaptive curriculum learning strategy that focuses on improving the model's details.

C. Adaptive Curriculum Learning

Data scarcity and the complexity of 3D sketches are two major challenges we face. In our framework, we observed that when the training data consists of a limited number of hand-drawn 3D sketches and their corresponding 3D models, the network typically struggles to effectively generalize across a broader range of sketch styles and geometries when mapping these abstract sketches to the latent space and conditioning them. Inspired by curriculum learning, we address these issues by simulating the way humans learn during the sketching process. Just as beginners in sketching typically start to learn with with simple and flexible shapes and gradually progress to more complex forms, we aim to apply this progressive learning strategy to our framework.

1) Sample Difficulty Score: In curriculum learning, carefully selecting and ordering samples from simple to complex is crucial for effective progressive skill development. The selection of samples is based on their difficulty scores. Inspired by the curriculum DeepSDF [57], we treat points with estimation errors as difficult samples, points with correct estimations as easy samples, and points with values between 0 and the true value as semi-difficult samples. We use the following difficulty scoring formula:

$$s = 1 + \alpha \,\mathrm{sgn}(y)\,\mathrm{sgn}(\bar{y} - y) \tag{2}$$

where y is the SDF value corresponding to the hand-drawn sketch, $\bar{y}$ is the predicted SDF value, and α controls the coefficients for difficult and semi-difficult samples. The function $\mathrm{sgn}(v)$ is defined as:

$$\mathrm{sgn}(v) = \begin{cases} 1 & \text{if } v \geq 0, \\ -1 & \text{if } v < 0. \end{cases}$$

2) Adaptive Curriculum: Unlike traditional manually designed curriculum learning, we adopt an adaptive curriculum learning strategy [58]. Specifically, we first use a pre-trained network to obtain the initial difficulty scores and, based on these scores, sort the initial dataset Λ in ascending order to form a sample pool $\Lambda^{'}$. Then, we divide $\Lambda^{'}$ into different mini-batches $B = [B_1, ..., B_m]$ and sequentially input them into the target network for training. Next, we design a pacing function $p(\cdot)$, which is a monotonically increasing function that determines the rate at which we learn from simpler to more complex samples. Finally, at the end of the forward propagation, we update the difficulty scores and calculate the new sample pool $\Lambda^{''}$. The difficulty score at the $(k+1)$-th position can be represented as:

$$s_{k+1} = (1 - \beta)s_k + \beta s \tag{3}$$

where $k = \lfloor B_m / inv \rfloor$, inv controls the frequency of difficulty score updates, and β controls the speed at which the difficulty scores are updated.

3) Pacing Function: To manage the pace at which the network learns from the samples, we need a monotonically increasing pacing function $p(\cdot)$ to control the rate at which data is fed into the target network. This function can be expressed as:

$$p(i) = n \times \min(1, p_0 \times q^{\lfloor i / r_0 \rfloor}) \tag{4}$$

where n represents the number of samples, p_0 is the sample ratio in the initial step, q controls the speed of sample ratio growth, r_0 controls the frequency of sample ratio growth, and i is the current step. In actual training, we set p_0 to 0.2, q to 1.9, and r_0 to 1.

V. Experiments

A. Implementation Details

In this study, we first pre-train the generative diffusion model on 3D shape dataset (DeepFashion3D) to enable high-quality synthesis (Stage 1). Subsequently, the model undergoes fine-tuning on a paired dataset of hand-drawn 3D sketches and 3D shapes with our KO3DClothes+ dataset (Stages 2-3). We split the KO3DClothes dataset into training and test sets in an 8:2 ratio. In the initial pre-training stage, we train the first UNet using the Adam optimizer [59], with a learning rate of 2e-4 for 800 epochs. For training the second UNet, we adopt the AdamW optimizer [60], adjusting the learning rate to 1e-4 and training for 500 epochs. During the 3D sketch mapping stage, we train the sketch encoder using the Adam

optimizer with a learning rate of 2e-4 for 300 epochs. Finally, in the joint fine-tuning stage, we simultaneously train both the diffusion model and the 3D sketch encoder using the Adam optimizer for 300 epochs, maintaining a learning rate of 2e-4. The training process is conducted on Ascend 910b GPUs under Mindspeed framework. Following previous study [50], we use widely-used voxel IoU (Intersection over Union) and Chamfer Distance (CD) to evaluate the model performance.

B. Qualitative and Quantitative Assessment

To evaluate the performance of our method, we conducted comparison on several recent 3D reconstruction and generation methods based on VR sketches. 3DSketch2Shape [61] is an early attempt of 3D sketch to 3D shape based on normalizing flow, our diffusion model exhibits better shape production capability. Deep3DVRSketch [50] also use diffusion model, but this approach is initially designed for object 3D content creation, while our approach is specifically designed for garment creation. We use point-cloud 3D prior instead of image priors as in [50] to better fit the complex nature of the clothes, which demonstrate better performance in the experiment.

We train each baseline model with the same KO3DClothes dataset. As shown in Table I, our method performs well in shape accuracy. To view the shape quality, Fig. 4 presents visualization of results. Our approach is capable of generating plausible and high-quality 3D garments. It's important to note that the primary focus of our experiment is capturing the **overall shape** of the garment (w/ or w/o sleeves, the length of the sleeves, etc.). Fine details like wrinkles and folds are not well-suited for, nor necessary to obtain through, 3D sketching. Leveraging the capabilities of numerous commercially available clothing simulators, we can readily generate dynamic garment with realistic clothing behavior with natural wrinkles across various poses from clothes simulations once the overall shape is established.

TABLE I
QUANTITATIVE EVALUATION OF THE REAL-WORLD KO3DCLOTHES DATASET

Methods	IoU↑	CD↓
3DSketch2Shape	0.3188	0.0820
Deep3DVRSketch	**0.3252**	0.0606
Ours	0.3190	**0.0597**

C. User Study

1) 3D Model Quality and Fidelity: We evaluate the quality of generated results from user study. Following previous research (Chen et al., 2023a), we used the widely adopted 5-point Mean Opinion Score (MOS) metric for evaluation. In this study, participants were asked to rate the generated 3D models based on two aspects, with scores from 1 to 5: Q1) The degree of fidelity between the generated 3D model and the input sketch; Q2) Participants' overall evaluation of the quality of the generated 3D model. We invited 15 designers to participate and presented them with 12 results generated by our algorithm for evaluation. Before the experiment began, we provided detailed explanations of the "fidelity" and "quality" rating criteria to ensure that participants had a consistent understanding of the evaluation standards. The average ratings of the experimental results are shown in Tab. II. Compared to existing methods, our approach performed obtains higher user ratings, which further validating the advantages and effectiveness of our method in generating high-quality 3D garments.

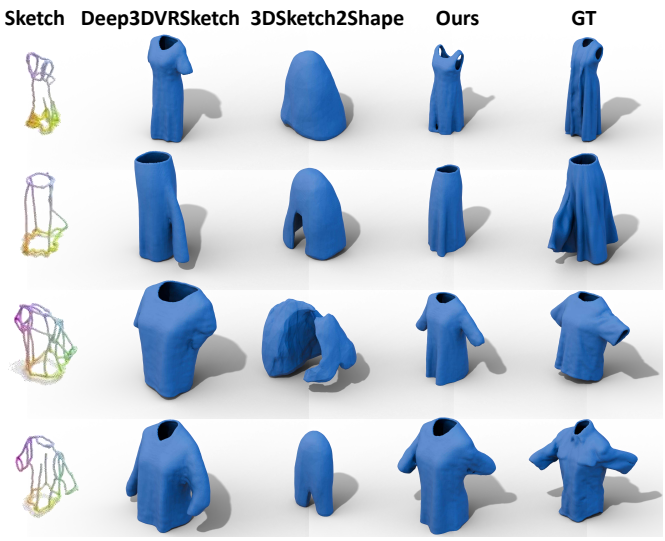

Fig. 4. Comparison with the existing state-of-the-art methods.

TABLE II
MEAN OPINION SCORES (1-5) FROM USER STUDY

Methods	(Q1): Fidelity	(Q2): Quality
3DSketch2Shape	1.2175 ± 0.2813	1.3480 ± 0.6109
Deep3DVRSketch	3.2824 ± 0.6563	3.3921 ± 0.5014
Ours	**4.5925 ± 0.4821**	**4.6225 ± 0.3428**

2) Comparison with 2D Sketches: To verify the advantages of our method in terms of controllability, additional user experiments were conducted by recruiting a group of 15 3D designers from a 3D printing company. We asked the designers to draw their desired shape contours on a 2D plane and use baseline models Sketch2model [38]. The designers also described and sketched their desired 3D models and generated textured 3D shapes using our approach. Participants were asked to evaluate the controllability and usefulness of each method, which are crucial factors in the assessment of user interface usability and user experience [62]–[64]. Based on the settings in [63], [64], we employed a 7-point Likert scale ranging from strongly disagree to strongly agree. The results are shown in Table III. Compared to existing 2D-to-3D methods, participants gave our 3D-to-3D approach higher ratings in controllability and usefulness.

TABLE III
MEAN OPINION SCORES (1-7) FROM USER STUDY

Methods	(Q1): Controllability	(Q2): Usefulness
Sketch2model	2.5233 ± 1.0717	2.4883 ± 1.3165
Ours	**6.4533 ± 0.6258**	**6.6217 ± 0.3684**

D. Ablation Study

In the ablation study, we isolate the effectiveness of using shape priors and curriculum learning. We found that removing pretraining with point cloud and directly training a conditional diffusion model on sketches (using the same network architecture) leads to a significant drop in performance. As shown in Fig. 5 and Tab. IV, models without pretraining fail to effectively learn to generate reasonable shapes. It is a effective

way to overcome the challenge of limited annotated data in the sketch domain.

TABLE IV
QUANTITATIVE EVALUATION OF ABLATION STUDY

	IoU↑	CD↓
w/o Point Cloud Prior	0.3178	0.0749
w/o Curriculum Learning	0.2680	0.0620
Ours	**0.3190**	**0.0597**

The curriculum learning method guides the model through a step-by-step process from simple to complex learning tasks, which significantly improves the model's ability to adapt to diverse sketch styles as shown in Tab. IV.

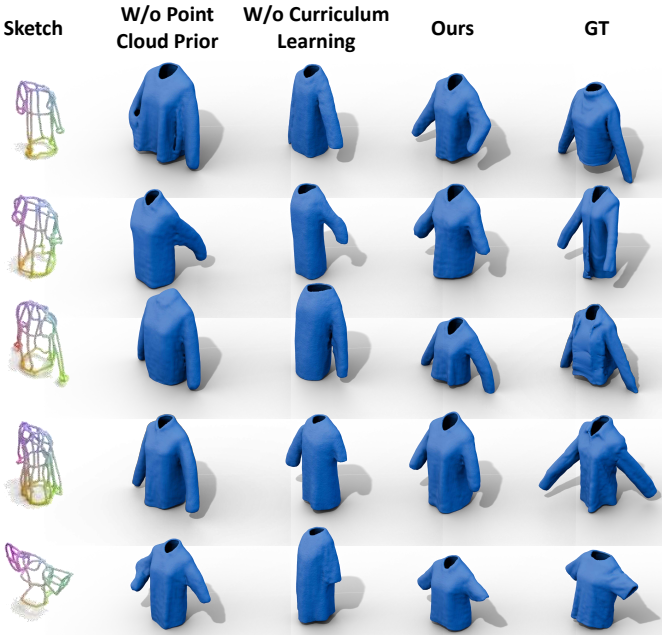

Fig. 5. Qualitative Evaluation for Ablation Studies.

VI. CONCLUSION

As immersive consumer electronics continue to redefine how people interact with 3D content, our work takes a step toward democratizing digital fashion by making 3D garment design intuitive, expressive, and widely accessible. By harnessing the natural interaction affordances of AR/VR devices and the power of generative AI, we enable users to turn simple 3D sketches into detailed, wearable virtual garments without the need for professional tools or training. Our novel architecture and data collection efforts not only push the frontier of sketch-based modeling but also pave the way for broader adoption in consumer-facing applications, such as personalized avatars, virtual try-ons, and digital self-expression. Looking forward, we envision expanding this framework to real-time, interactive design experiences on everyday AR/VR devices — bringing the future of fashion creation to the fingertips of every user.

REFERENCES

[1] H. Zhang, X. Mu, G. Li, Z. Xu, X. Yu, and J. Ma, "Mannequin2real: a two-stage generation framework for transforming mannequin images into photorealistic model images for clothing display," *IEEE Transactions on Consumer Electronics*, 2024.

[2] X. Hu, C. Fang, K. Yang, J. Liang, R. Luo, and T. Peng, "Towards high-fidelity 3d virtual try-on via global collaborative modeling," *IEEE Transactions on Consumer Electronics*, 2024.

[3] K. Plażyk, "The democratization of luxury–a new form of luxury," *Studia University of Economics in Katowice*, pp. 158–165, 2015.

[4] J. Hopkins, *Fashion design: The complete guide*. Bloomsbury Publishing, 2021.

[5] R. Arora, R. H. Kazi, F. Anderson, T. Grossman, K. Singh, and G. W. Fitzmaurice, "Experimental evaluation of sketching on surfaces in vr." in *CHI 2017*, vol. 17, 2017, pp. 5643–5654.

[6] C. Gao, Q. Yu, L. Sheng, Y.-Z. Song, and D. Xu, "Sketchsampler: Sketch-based 3d reconstruction via view-dependent depth sampling," in *European Conference on Computer Vision*. Springer, 2022, pp. 464–479.

[7] M. F. Deering, "Holosketch: a virtual reality sketching/animation tool," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 2, no. 3, pp. 220–238, 1995.

[8] D. F. Keefe, D. A. Feliz, T. Moscovich, D. H. Laidlaw, and J. J. LaViola Jr, "Cavepainting: A fully immersive 3d artistic medium and interactive experience," in *Proceedings of the 2001 symposium on Interactive 3D graphics*, 2001, pp. 85–93.

[9] K. C. Kwan and H. Fu, "Mobi3dsketch: 3d sketching in mobile ar," in *CHI 2019*, 2019, pp. 1–11.

[10] P. Xu, H. Fu, Y. Zheng, K. Singh, H. Huang, and C.-L. Tai, "Model-guided 3d sketching," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 10, pp. 2927–2939, 2018.

[11] Z. Zheng, T. Yu, Y. Liu, and Q. Dai, "Pamir: Parametric model-conditioned implicit representation for image-based human reconstruction," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 6, pp. 3170–3184, 2021.

[12] Y. He, H. Xie, and K. Miyata, "Sketch2cloth: Sketch-based 3d garment generation with unsigned distance fields," in *2023 Nicograph International (NicoInt)*. IEEE, 2023, pp. 38–45.

[13] F. Zhao, S. Liao, J. Huo, Z. Huo, W. Wang, J. Han, and C. Shan, "Weakly supervised joint transfer and regression of textures for 3d human reconstruction," *IEEE Transactions on Consumer Electronics*, 2024.

[14] B. Jiang, J. Zhang, Y. Hong, J. Luo, L. Liu, and H. Bao, "Bcnet: Learning body and cloth shape from a single image," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XX 16*. Springer, 2020, pp. 18–35.

[15] H. Zhu, L. Qiu, Y. Qiu, and X. Han, "Registering explicit to implicit: Towards high-fidelity garment mesh reconstruction from single images," https://arxiv.org/abs/2203.15007, 2022. [Online]. Available: https://arxiv.org/abs/2203.15007

[16] A. Srivastava, C. Pokhariya, S. S. Jinka, and A. Sharma, "xcloth: Extracting template-free textured 3d clothes from a monocular image," in *Proceedings of the 30th ACM International Conference on Multimedia*, 2022, pp. 2504–2512.

[17] 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.

[18] 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," https://arxiv.org/abs/2003.12753, 2020. [Online]. Available: https://arxiv.org/abs/2003.12753

[19] S. Saito, Z. Huang, R. Natsume, S. Morishima, A. Kanazawa, and H. Li, "Pifu: Pixel-aligned implicit function for high-resolution clothed human digitization," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 2304–2314.

[20] S. Saito, T. Simon, J. Saragih, and H. Joo, "Pifuhd: Multi-level pixel-aligned implicit function for high-resolution 3d human digitization," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 84–93.

[21] Y. Xiu, J. Yang, X. Cao, D. Tzionas, and M. J. Black, "Econ: Explicit clothed humans optimized via normal integration," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 512–523.

[22] N. Pietroni, C. Dumery, R. Falque, M. Liu, T. A. Vidal-Calleja, and O. Sorkine-Hornung, "Computational pattern making from 3d garment models." *ACM Trans. Graph.*, vol. 41, no. 4, pp. 157–1, 2022.

[23] M. Korosteleva and S.-H. Lee, "Generating datasets of 3d garments with sewing patterns," *arXiv preprint arXiv:2109.05633*, 2021.

[24] E. Corona, A. Pumarola, G. Alenya, G. Pons-Moll, and F. Moreno-Noguer, "Smplicit: Topology-aware generative model for clothed peo-

ple," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 11 875–11 885.

[25] Z. Su, T. Yu, Y. Wang, and Y. Liu, "Deepcloth: Neural garment representation for shape and style editing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 2, pp. 1581–1593, 2022.

[26] 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," *arXiv preprint arXiv:2311.17050*, 2023.

[27] P. N. Chowdhury, T. Wang, D. Ceylan, Y.-Z. Song, and Y. Gryaditskaya, "Garment ideation: Iterative view-aware sketch-based garment modeling," in *2022 International Conference on 3D Vision (3DV)*. IEEE, 2022, pp. 22–31.

[28] H. Bandyopadhyay, S. Koley, A. Das, A. Sain, P. N. Chowdhury, T. Xiang, A. K. Bhunia, and Y.-Z. Song, "Doodle your 3d: From abstract freehand sketches to precise 3d shapes," *arXiv preprint arXiv:2312.04043*, 2023.

[29] 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.

[30] D. Kong, Q. Wang, and Y. Qi, "A diffusion-refinement model for sketch-to-point modeling," in *Proceedings of the Asian Conference on Computer Vision*, 2022, pp. 1522–1538.

[31] 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.

[32] A. Sanghi, H. Chu, J. G. Lambourne, Y. Wang, C.-Y. Cheng, M. Fumero, and K. R. Malekshan, "Clip-forge: Towards zero-shot text-to-shape generation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 18 603–18 613.

[33] C. Li, C. Zhang, A. Waghwase, L.-H. Lee, F. Rameau, Y. Yang, S.-H. Bae, and C. S. Hong, "Generative ai meets 3d: A survey on text-to-3d in aigc era," *arXiv preprint arXiv:2305.06131*, 2023.

[34] R. Fu, X. Zhan, Y. Chen, D. Ritchie, and S. Sridhar, "Shapecrafter: A recursive text-conditioned 3d shape generation model," *Advances in Neural Information Processing Systems*, vol. 35, pp. 8882–8895, 2022.

[35] X. Tian, Y.-L. Yang, and Q. Wu, "Shapescaffolder: Structure-aware 3d shape generation from text," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 2715–2724.

[36] M. Liu, C. Xu, H. Jin, L. Chen, Z. Xu, H. Su *et al.*, "One-2-3-45: Any single image to 3d mesh in 45 seconds without per-shape optimization," *arXiv preprint arXiv:2306.16928*, 2023.

[37] Y. Shi, P. Wang, J. Ye, M. Long, K. Li, and X. Yang, "Mvdream: Multi-view diffusion for 3d generation," *arXiv preprint arXiv:2308.16512*, 2023.

[38] S.-H. Zhang, Y.-C. Guo, and Q.-W. Gu, "Sketch2model: View-aware 3d modeling from single free-hand sketches," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 6012–6021.

[39] B. Guillard, E. Remelli, P. Yvernay, and P. Fua, "Sketch2mesh: Reconstructing and editing 3d shapes from sketches," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 13 023–13 032.

[40] T. Chen, C. Fu, L. Zhu, P. Mao, J. Zhang, Y. Zang, and L. Sun, "Deep3dsketch: 3d modeling from free-hand sketches with view-and structural-aware adversarial training," *arXiv preprint arXiv:2312.04435*, 2023.

[41] L. Olsen, F. F. Samavati, M. C. Sousa, and J. A. Jorge, "Sketch-based modeling: A survey," *Computers & Graphics*, vol. 33, no. 1, pp. 85–103, 2009.

[42] A. Bonnici, A. Akman, G. Calleja, K. P. Camilleri, P. Fehling, A. Ferreira, F. Hermuth, J. H. Israel, T. Landwehr, J. Liu *et al.*, "Sketch-based interaction and modeling: where do we stand?" *AI EDAM*, vol. 33, no. 4, pp. 370–388, 2019.

[43] Z. Lun, M. Gadelha, E. Kalogerakis, S. Maji, and R. Wang, "3d shape reconstruction from sketches via multi-view convolutional networks," in *2017 International Conference on 3D Vision (3DV)*. IEEE, 2017, pp. 67–77.

[44] C. Li, H. Pan, Y. Liu, X. Tong, A. Sheffer, and W. Wang, "Robust flow-guided neural prediction for sketch-based freeform surface modeling," *ACM Transactions on Graphics (TOG)*, vol. 37, no. 6, pp. 1–12, 2018.

[45] J. Wang, J. Lin, Q. Yu, R. Liu, Y. Chen, and S. X. Yu, "3d shape reconstruction from free-hand sketches," in *European Conference on Computer Vision*. Springer, 2022, pp. 184–202.

[46] Y. Zhong, Y. Gryaditskaya, H. Zhang, and Y.-Z. Song, "Deep sketch-based modeling: Tips and tricks," in *2020 International Conference on 3D Vision (3DV)*. IEEE, 2020, pp. 543–552.

[47] T. Chen, C. Fu, Y. Zang, L. Zhu, J. Zhang, P. Mao, and L. Sun, "Deep3dsketch+: Rapid 3d modeling from single free-hand sketches," in *International Conference on Multimedia Modeling*. Springer, 2023, pp. 16–28.

[48] Y. Zang, C. Fu, T. Chen, Y. Hu, Q. Liu, and W. Hu, "Deep3dsketch+: Obtaining customized 3d model by single free-hand sketch through deep learning," *arXiv preprint arXiv:2310.18609*, 2023.

[49] Y. Zang, C. Ding, T. Chen, P. Mao, and W. Hu, "Deep3dsketch+\+: High-fidelity 3d modeling from single free-hand sketches," *arXiv preprint arXiv:2310.18178*, 2023.

[50] T. Chen, C. Ding, S. Zhang, C. Yu, Y. Zang, Z. Li, S. Peng, and L. Sun, "Rapid 3d model generation with intuitive 3d input," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 12 554–12 564.

[51] L. Luo, Y. Gryaditskaya, Y. Yang, T. Xiang, and Y.-Z. Song, "Fine-grained vr sketching: Dataset and insights," in *2021 International Conference on 3D Vision (3DV)*. IEEE, 2021, pp. 1003–1013.

[52] 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," *arXiv preprint arXiv:2305.04461*, 2023.

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

[54] J. Ho and T. Salimans, "Classifier-free diffusion guidance," *arXiv preprint arXiv:2207.12598*, 2022.

[55] J. Zhou, J. Wang, B. Ma, Y.-S. Liu, T. Huang, and X. Wang, "Uni3d: Exploring unified 3d representation at scale," *arXiv preprint arXiv:2310.06773*, 2023.

[56] T. Wang, T. Zhang, B. Zhang, H. Ouyang, D. Chen, Q. Chen, and F. Wen, "Pretraining is all you need for image-to-image translation," *arXiv preprint arXiv:2205.12952*, 2022.

[57] Y. Duan, H. Zhu, H. Wang, L. Yi, R. Nevatia, and L. J. Guibas, "Curriculum deepsdf," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VIII 16*. Springer, 2020, pp. 51–67.

[58] Y. Kong, L. Liu, J. Wang, and D. Tao, "Adaptive curriculum learning," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 5067–5076.

[59] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[60] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017.

[61] L. Luo, P. N. Chowdhury, T. Xiang, Y.-Z. Song, and Y. Gryaditskaya, "3d vr sketch guided 3d shape prototyping and exploration," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 9267–9276.

[62] C. Oh, J. Song, J. Choi, S. Kim, S. Lee, and B. Suh, "I lead, you help but only with enough details: Understanding user experience of co-creation with artificial intelligence," in *CHI 2018*, 2018, pp. 1–13.

[63] B. Albert and T. Tullis, *Measuring the User Experience: Collecting, Analyzing, and Presenting UX Metrics*. Morgan Kaufmann, 2022.

[64] Y. Zang, Y. Han, C. Ding, J. Zhang, and T. Chen, "Magic3dsketch: Create colorful 3d models from sketch-based 3d modeling guided by text and language-image pre-training," *arXiv preprint arXiv:2407.19225*, 2024.