

# DreamReward: Text-to-3D Generation with Human Preference

JunLiang Ye<sup>1,2†</sup>, Fangfu Liu<sup>1†</sup>, Qixiu Li<sup>1</sup>, Zhengyi Wang<sup>1,2</sup>, Yikai Wang<sup>1</sup>,  
Xinzhou Wang<sup>1,2</sup>, Yueqi Duan<sup>1✉</sup>, and Jun Zhu<sup>1,2✉</sup>

<sup>1</sup> Tsinghua University  
<sup>2</sup> ShengShu, Beijing, China

**Abstract.** 3D content creation from text prompts has shown remarkable success recently. However, current text-to-3D methods often generate 3D results that do not align well with human preferences. In this paper, we present a comprehensive framework, coined **DreamReward**, to learn and improve text-to-3D models from human preference feedback. To begin with, we collect 25k expert comparisons based on a systematic annotation pipeline including rating and ranking. Then, we build Reward3D—the first general-purpose text-to-3D human preference reward model to effectively encode human preferences. Building upon the 3D reward model, we finally perform theoretical analysis and present the Reward3D Feedback Learning (DreamFL), a direct tuning algorithm to optimize the multi-view diffusion models with a redefined scorer. Grounded by theoretical proof and extensive experiment comparisons, our DreamReward successfully generates high-fidelity and 3D consistent results with significant boosts in prompt alignment with human intention. Our results demonstrate the great potential for learning from human feedback to improve text-to-3D models. Project Page: <https://jamesyj1.github.io/DreamReward/>.

**Keywords:** 3D Generation · RLHF · Human Preference

## 1 Introduction

3D content generation has wide applications in various fields (*e.g.*, films, animation, game design, architectural design, and virtual reality). In recent years, significant advancements in diffusion models have greatly propelled the development of automated 3D generation. 3D creation can be classified into two principal categories [43]: inference-only 3D native methods [11, 16] and optimization-based 2D lifting methods [4, 5, 14, 17, 21, 33, 42, 49]. Given a text or an image, these models are capable of generating highly intricate 3D content, some even overcoming multi-face issues to produce high-quality and viewpoint-consistent 3D models [23, 24, 39]. Despite rapid advancements, some researchers [39] indicate that the 3D content generated by existing generative models struggles to align with human preferences. Typically, this inconsistency includes but is not limited to text-3D alignment, overall quality, and multi-view consistency.

---

† Equal contribution.

✉ Corresponding authors.

Recently, some works have applied reinforcement learning from human feedback (RLHF) to natural language processing (NLP) [32, 41] and text-to-image generation [2, 45, 52, 56]. These algorithms typically begin by constructing and annotating datasets based on human feedback, and then training reward models. Finally, they finetune large models (such as large language models or diffusion models) using reinforcement learning techniques. This allows the fine-tuning models to better align with human preferences.

Inspired by the aforementioned works, we recognize the effectiveness of RLHF in improving the performance of generative models. In this work, we propose **DreamReward**, which greatly boosts high-text alignment and high-quality text-to-3D generation through human preference feedback. We propose the first general-purpose human preference reward model for text-to-3D generation, Reward3D. First, we use a clustering algorithm to extract 5k of the most representative prompts from Cap3D [26] and generate a corresponding 3D dataset. Subsequently, we produce 10 3D contents for each prompt and filtered them based on quality, resulting in 2530 prompt sets, where each prompt corresponds to 4~10 3D contents. After collecting 25k pairs of expert comparisons, we annotate the comparison and trained the Reward3D model based on it.

After constructing the annotated 3D Dataset, we train a 3D-aware scoring model for text-to-3D generation on the constructed 3D dataset. Given the most representative 110 prompts generated by GPTEval3D [51] and compared to the 2D scoring models ImageReward [52] and CLIP [34], which lack 3D-aware capabilities, our Reward3D can consistently align with human preference ranking and exhibit higher distinguishability among different 3D models. With its high alignment in human preference observed from experiments, we suggest that Reward3D could serve as a promising automatic text-to-3D evaluation metric.

Building upon this, we further explore an optimization approach to improve 3D generation results—Reward3D Feedback Learning (DreamFL), which is a direct tuning algorithm designed to optimize multi-view diffusion models using a redefined scorer. Based on Reward3D, we carefully design the LossReward and incorporate it into the SDS pipeline for 3D generation. Grounded by our mathematical derivation, we find that the LossReward effectively drives the optimization of 3D models towards higher quality and alignment. Extensive experimental results demonstrate that 3D assets generated by DreamFL not only achieve impressive visualization but also outperform other text-to-3D generation methods in terms of quantitative metrics such as GPTEval3D [51], CLIP [34], ImageReward [52], and our Reward3D.

To summarise, we make the following contributions:

- **Labeled-3D dataset**: We are among the first to construct and annotate a diverse 3D dataset suitable for training and testing models aligned with human preferences.
- **Reward3D**: We train the Reward3D scoring model with 3D-aware capabilities, enabling it to effectively evaluate the quality of generated 3D content.
- **DreamFL**: We propose the Reward3D Feedback Learning (DreamFL) algorithm to enhance the human preference alignment in our 3D results.

## 2 Related Work

### 2.1 Text-to-image Generation

Diffusion models [7, 13, 40] combining with large-scale language encoders [34, 37], have become the leading approach in text-to-image generation. Typically, they involve a two-step process. Initially, noise is progressively added to the original data until it aligns with a prior distribution, such as the Gaussian distribution. Subsequently, a neural network is employed to predict the previously added noise, allowing the initialized samples from the prior distribution to undergo a step-by-step reverse denoising process. Leveraging this technique, recent advancements [29, 36, 38] have demonstrated the ability to synthesize images of exceptional quality.

### 2.2 Text-to-3D Generation

With the high development of text-to-image diffusion models, there has been a surge of studies in text-to-3D generation recently. Due to limited diverse 3D datasets [3] compared to 2D, DreamFusion [33] and SJC [47] have shifted towards exploring the route of distilling score from 2D diffusion priors to optimize a 3D representation such as NeRF [28], and show very promising results. Such distillation-based methods [4, 21, 33, 39, 48–50, 55, 57] have undergone rapid improvements in recent years. However, there still exists a significant gap in the generation quality between 3D and 2D generation. 3D generation still lacks in terms of generation speed, diversity of themes, and alignment with human preferences. To improve the efficiency of 3D generation, DreamGaussian [5] transformed the 3D representation from NeRF to gaussian-splatting [17], resulting in high-quality generation effects. On the other hand, Shap-E [16] and Point-E [30] achieve 3D generation in a matter of minutes through pretraining on massive undisclosed 3D datasets. In this work, we will continue to narrow the gap between 3D and 2D generation through the DreamReward framework, which guides 3D models towards high-quality and highly aligned generation.

### 2.3 Text-to-3D Generation Evaluation Metrics.

Evaluating text-to-3D generation models is a highly challenging task, requiring both 3D awareness and understanding of textual semantics. The existing text-to-3D evaluation methods mainly include approaches that utilize multimodal embeddings, such as CLIP [15, 34] and BLIP [18, 19], as well as methods, such as GPTEval3D [51] and T3batch [12] that employ large-scale multimodal language models GPT-4V [1]. To obtain a fair and reliable collection of text prompts, GPTEval3D [51] created a text-prompt generator using language instruction. This enables us to better assess the strengths and weaknesses of different 3D models. Our proposed Reward3D in this work will serve as a novel evaluation model for text-to-3D, assisting users in evaluating 3D results effectively without relying on large language models, which may incur costs. It will also provide promising scores and rankings aligned with human preferences.

## 2.4 Learning from human feedback

The alignment of large language models (LLMs) [31, 44] with human preferences is an issue that has garnered considerable attention. Reinforcement Learning from Human Feedback (RLHF) [32, 41, 58] uses a strategy that leverages human feedback with reinforcement learning policies to address this challenge. Recent literature [2, 9, 46, 52, 53, 53] has demonstrated that incorporating human feedback enhances the performance of text-to-image models as well. ImageReward [52] introduces a reward model based on human preferences specifically for text-to-image tasks and puts forward a novel approach termed Reward Feedback Learning, which is designed to refine diffusion models. Meanwhile, DiffusionDPO [46] presents a technique that aligns diffusion models with human preferences by directly optimizing with human comparative data, an adaptation of Direct Preference Optimization (DPO) [35]. Further, DPOK [9] amalgamates policy optimization with KL regularization within the framework of text-to-image diffusion models. Despite the proven efficacy of these approaches, learning from human feedback in the domain of text-to-3D generation still requires investigation.

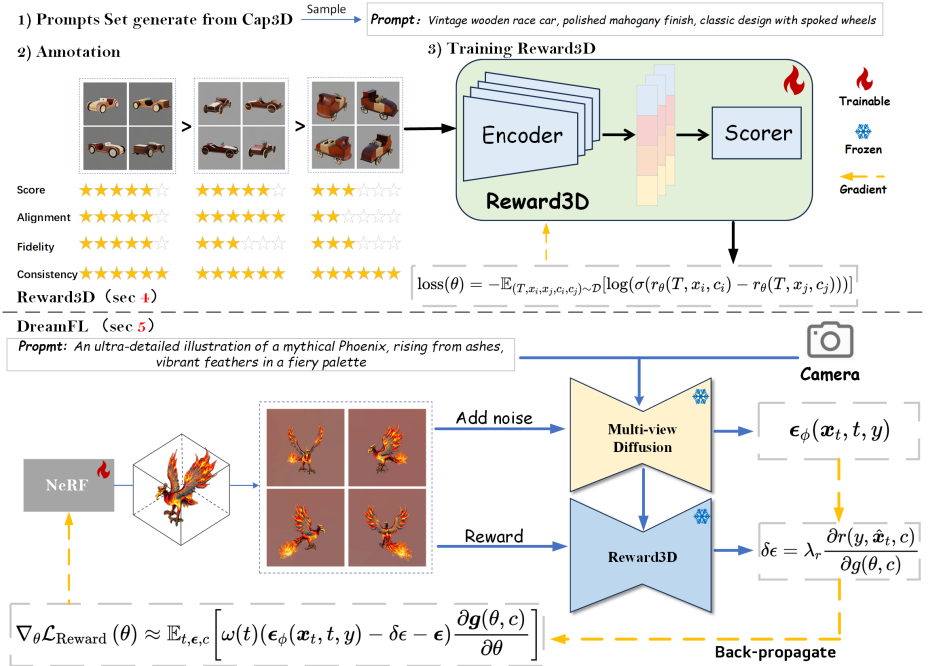
## 3 Overall Framework

We hereby present the overall framework of our DreamReward, a novel text-to-3D framework to achieve human preference alignment. The complete pipeline is depicted in Figure 1. Initially, in Sec. 4, we introduce the Reward3D architecture, which encompasses the construction of a 3D dataset (Sec. 4.1), the development of a data annotation pipeline (Sec. 4.1), and the training of the Reward Model (RM) (Sec. 4.2). After training the 3D-aware Reward3D model, we proceed to delineate the core of our DreamReward framework—the Reward3D Feedback Learning (DreamFL) in Sec. 5. Specifically, In Sec. 5.2, we first identify that the distribution obtained by existing diffusion models for distilling 3D assets diverges from the desired distribution at two distinct levels. Then in Sec. 5.3, we demonstrate the efficacy of our Reward3D in bridging the gap between these two distributions through both detailed mathematical derivation and demonstration. More detailed implementation specifics of our algorithm can be found in our supplementary materials.

## 4 Reward3D

### 4.1 Annotation Pipeline Design

**Prompt Selection and 3D Collection.** Our proposed new dataset utilizes a diverse selection of prompts from cap3D [26], which is a re-annotation of the large-scale 3D dataset Objaverse [6], with better alignment compared to the original prompts in Objaverse [6]. To ensure diversity in selected prompts, we employ a graph-based algorithm that leverages language model-based prompt similarity. This selection yields 5000 candidate prompts, each accompanied by 4-10 sampled 3D assets generated from *ashawkey/mvdream-sd2.1-diffusers* [39]



**Fig. 1: The overall framework of our DreamReward.** (Top) Reward3D involves data collection, annotation, and preference learning. (Bottom) DreamFL utilizes feedback from Reward3D to compute RewardLoss and incorporate it into the SDS loss for simultaneous optimization of NeRF.

**Dataset Filtering.** From our empirical observations, we notice that results we generate are prone to encountering mode collapses. This is attributed to the complexity of the selected prompts [51], leading to significant collapses of the corresponding 3D assets under the same prompt. Therefore, prior to annotation, we conduct a filtering process on the generated 3D dataset with a selection of 2530 prompts, each from 4-10 assets. Then we obtain 25,304 candidate pairs for labeling.

**Human Annotation Design.** In the annotation process, annotators rate images on a scale of 1-6 based on text-3D alignment, overall quality, and multi-view consistency, and rank them according to the average scores. To avoid potential conflicts in rankings, we maintain a real-time data structure for each prompt set. When conflicts arise, conflicting pairs are flagged, followed by secondary verification and correction. We recruit some annotators from universities for labeling, and additionally sought assistance from data institutions, with annotation documents showcased in the appendix.

## 4.2 Reward3D Training

Similar to RM training for language model of previous works [32, 41, 58], we formulate the preference annotations as rankings. We have 9 3D-model ranked for the same prompt  $T$ , and get at most  $C_9^2$  comparison pairs if there are no ties between two 3D-model. For each comparison, if  $x_i$  is better and  $x_j$  is worse, the loss function can be formulated as:

$$\text{loss}(\theta) = -\mathbb{E}_{(T, x_i, x_j, c_i, c_j) \sim \mathcal{D}} [\log(\sigma(r_\theta(T, x_i, c_i) - r_\theta(T, x_j, c_j)))], \quad (1)$$

where  $r$  represents the Reward3D model,  $c_i$  and  $c_j$  represent cameras.

**Training Detail.** We use ImageReward [52] as the backbone of our Reward3D. We extract image and text features, combine them with cross-attention, and use an MLP to generate a scalar for preference comparison. During the training stage, we observe rapid convergence and consequent overfitting, which harms its performance. To address this, we freeze some backbone transformer layers’ parameters, finding that a proper number of fixed layers can improve the model’s performance. We train Reward3D on a single 4090 GPU (24GB) with a batch size set to 8. We utilize the AdamW [25] optimizer with a learning rate of 1e-5 and a fixed rate set to 80%.

## 5 DreamFL

After training the 3D reward model, we now present DreamFL algorithm. Our pipeline is depicted in Figure 1. Before delving into the specifics of our approach, let’s start by revisiting the preliminaries of Score Distillation Sampling theory.

### 5.1 Preliminaries

**Score Distillation Sampling (SDS).** Score Distillation Sampling (SDS) [33], an optimization method that distills 3D knowledge from pretrained 2D diffusion models, has significantly advanced the rapid development of 3D generation [21, 33, 48, 49, 55] in recent years. Given a differentiable rendering mapping function  $g(\theta, c)$ , a pretrained 2D diffusion model  $\phi(x_t|y)$  and its corresponding noise prediction network  $\epsilon_\phi(x_t, t, y)$ , SDS optimizes the parameter  $\theta$  by solving:

$$\nabla_\theta \mathcal{L}_{\text{SDS}}(\theta) \approx \mathbb{E}_{t, \epsilon, c} \left[ \omega(t) (\epsilon_\phi(x_t, t, y) - \epsilon) \frac{\partial g(\theta, c)}{\partial \theta} \right], \quad (2)$$

where  $\theta$  is the parameter of 3D representation and  $c$  is desired camera. To elaborate further, we denote  $q_t^\theta(x_t|c)$  as the distribution at time  $t$  of the forward diffusion process, initiated from the rendered image. The SDS optimization algorithm described above can be regarded as

$$\min_{\theta \in \Theta} \mathcal{L}_{\text{SDS}}(\theta) := \mathbb{E}_{t, c} [(\sigma_t / \alpha_t) \omega(t) D_{\text{KL}}(q_t^\theta(x_t|c) \| p_t(x_t|y))]. \quad (3)$$

## 5.2 DreamFL

**Why do SDS-generated 3D assets lack alignment?** Two major challenges arise when seeking to align 3D generation models with human preferences. (1) the conditional distribution  $p_t(\mathbf{x}_t|y)$  obtained by pre-trained diffusion models diverges from human preference and actual user-prompt distributions [52, 53]. Consequently, 3D assets distilled from this deviant distribution inherently fail to align with human preferences, often to an even more pronounced degree. (2) The capability to maintain multi-view consistency acquired through score distillation from diffusion models is profoundly limited [8, 20, 22].

In recent years, many related works [2, 46] have emerged in the field of text-to-image generation to address the aforementioned problem (1). However, due to the multi-step nature of the process, fine-tuning a diffusion model suitable for the text-to-3D domain is challenging. The main difference lies in the timestamp, where fine-tuning a diffusion model for generation typically requires fine-tuning the last 20%-30% of denoising steps [52]. In contrast, text-to-image tasks often only require the last 10 steps for fine-tuning, as they use around 40 steps. Through empirical studies, it is found that 3D generation often requires over 1000 denoising steps [10, 49], meaning over 20 times the computational workload. On the other hand, the inherent multi-view inconsistency of the SDS algorithm will exacerbate the perceived quality of the generated 3D results among humans. Suppose there exists a diffusion model  $\phi$  that aligns well with human preferences. However, the presence of problem (2). causes it to lack awareness of 3D, resulting in the generated 3D assets still not meeting our desired criteria. Based on the above analysis, we have learned that the distribution  $p_t(\mathbf{x}_t|y)$ , followed by our current diffusion model, deviates significantly from the ideal distribution  $p_t^r(\mathbf{x}_t|y)$ , where  $p_t^r(\mathbf{x}_t|y)$  represents the target distribution aligned with human preferences and possessing 3D awareness.

To address these challenges, our approach aims to leverage our existing distribution  $p_t(\mathbf{x}_t|y)$  to approximate the challenging distribution  $p_t^r(\mathbf{x}_t|y)$ . Inspired by ProlificDreamer, which used a LoRA [54] to approximate the distribution of NeRF, we found that approximating the predicted noise of a distribution is sufficient to approximate the distribution itself. Therefore, we denote the noise generated from distributions  $p_t(\mathbf{x}_t|y)$  and  $p_t^r(\mathbf{x}_t|y)$  as  $\epsilon_\phi(x_t, t, y)$  and  $\epsilon_\phi^r(x_t, t, y)$ , respectively. Our goal is to approximate  $\epsilon_\phi^r(x_t, t, y)$  using  $\epsilon_\phi(x_t, t, y)$ , meaning to obtain  $\delta\epsilon = \epsilon_\phi(x_t, t, y) - \epsilon_\phi^r(x_t, t, y)$ , which will effectively address the problem. With this understanding, we proceed to the derivation below.

**Approximate  $\delta\epsilon$  using Reward3D.** First, let us rewrite Eq. 3 by replacing  $p_t(\mathbf{x}_t|y)$  with  $p_t^r(\mathbf{x}_t|y)$ .

$$\min_{\theta \in \Theta} \mathcal{L}_{\text{Reward}}(\theta) := \mathbb{E}_{t,c} [(\sigma_t/\alpha_t) \omega(t) D_{\text{KL}}(q_t^\theta(\mathbf{x}_t|c) \| p_t^r(\mathbf{x}_t|y))]. \quad (4)$$

Its gradient is approximated by

$$\nabla_{\theta} \mathcal{L}_{\text{Reward}}(\theta) \approx \mathbb{E}_{t,c} \left[ \omega(t) (\epsilon_\phi^r(\mathbf{x}_t, t, y) - \epsilon) \frac{\partial g(\theta, c)}{\partial \theta} \right]. \quad (5)$$

However, as mentioned above, obtaining the distribution  $p_t^r(\mathbf{x}_t|y)$  and  $\epsilon_\phi^r(x_t, t, y)$  is a very challenging task. To overcome this challenge, we approximate  $\epsilon_\phi^r(x_t, t, y) = \epsilon_\phi(x_t, t, y) - \delta\epsilon$  using the pretrained  $\epsilon_\phi(x_t, t, y)$  prediction network and Reward3D. Let  $\mathbf{g}(\theta, c) = x \in \mathbb{R}^{4 \times H \times W \times C}$  denote the multiview images with continuous angles spanning 360 degrees,  $\mathbf{x}_t = \alpha_t \mathbf{x} + \sigma_t \epsilon$  and  $\hat{\mathbf{x}}_t = \frac{1}{\alpha_t} [\mathbf{x}_t - \sigma_t \epsilon_\phi(x_t, t, y)]$  denote the prediction. The difference  $\delta\epsilon$  between the  $\epsilon_\phi^r(x_t, t, y)$  and  $\epsilon_\phi(x_t, t, y)$  can be calculated by:

$$\delta\epsilon = -\frac{\partial r(y, \hat{\mathbf{x}}_t, c)}{\partial \epsilon_\phi(x_t, t, y)} = \lambda_r \frac{\partial r(y, \hat{\mathbf{x}}_t, c)}{\partial g(\theta, c)}, \quad (6)$$

Where  $r$  stands for Reward3D, and  $c$  stands for camera. Therefore, Eq. 5 can be reorganized as:

$$\nabla_\theta \mathcal{L}_{\text{Reward}}(\theta) \approx \mathbb{E}_{t, \epsilon, c} [\omega(t) (\underbrace{\epsilon_\phi(x_t, t, y) - \lambda_r \frac{\partial r(y, \hat{\mathbf{x}}_t, c)}{\partial g(\theta, c)}}_{\epsilon_\phi^r(x_t, t, y)} - \epsilon) \frac{\partial \mathbf{g}(\theta, c)}{\partial \theta}]. \quad (7)$$

Eq. 4 can be reorganized as:

$$\mathcal{L}_{\text{Reward}}(\theta) \approx \mathcal{L}_{\text{SDS}}(\theta) - \lambda_r r(y, \hat{\mathbf{x}}_t, c). \quad (8)$$

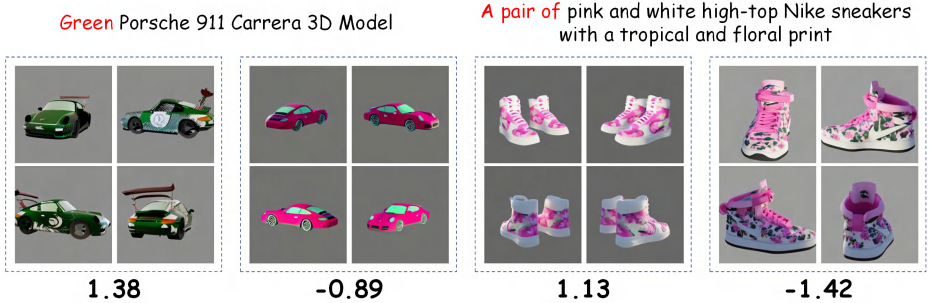
### 5.3 Implementation Details

We use MVDream [39] as our backbone, which is capable of generating multi-view consistent 3D assets. MVDream can align well with our multi-view optimization pipeline. For weighting factors, we define  $\lambda_r$  as  $t(\mathcal{L}_{\text{SDS}}, r(y, \hat{\mathbf{x}}_t, c))$  multiplied by  $\mu$ , where  $t$  is a weighting function used to ensure consistency in the magnitudes of  $\mathcal{L}_{\text{SDS}}$  and  $r(y, \hat{\mathbf{x}}_t, c)$ .  $\mu$  increases from 0 to 0.25 throughout the training process until reaching 0.6. Upon completion of training, we deactivate the  $\mathcal{L}_{\text{SDS}}$  and solely fine-tune the 3D results for 200 steps only using  $r(y, \hat{\mathbf{x}}_t, c)$  with a large  $\mu$ . Experimental results demonstrate that this approach enhances the aesthetics and stability of the training process. Please refer to our supplementary materials for more implementation details.

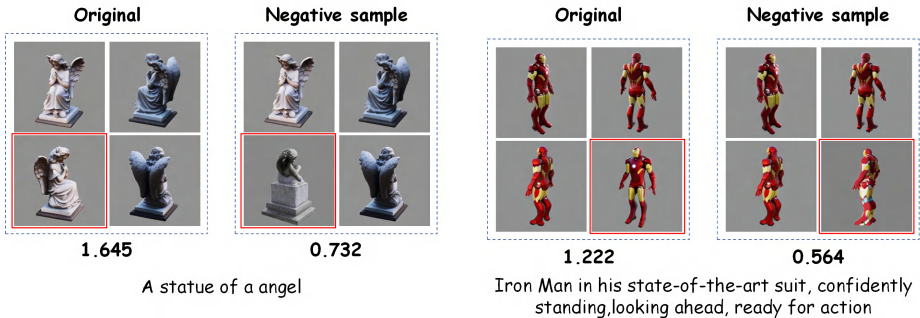
## 6 Experiments

In this section, we conduct extensive experiments to evaluate our text-to-3D method DreamFL and text-to-3D evaluation model Reward3D. In Sec. 6.2, We first present qualitative results compared with five baselines. Then we report the quantitative results with four evaluation metrics and user studies. All of this indicates that our DreamReward model beats the other five models and conforms to human preference. In Sec. 6.3, the experimental results demonstrated the evaluating capacity of our Reward3D model in assessing the quality of 3D assets as well as its capability in evaluating text-to-3D models, showing that its assessment aligns with human preferences. Consequently, it can serve as a substitute for human evaluation. Both qualitative evaluations and illustrative examples of representative cases have been presented in this part. Please refer to the supplementary for ablation studies.





**Fig. 2:** Representative examples from our constructed 3D dataset, along with the scores assigned by Reward3D. Reward3D gives lower scores to 3D assets deviating from the prompt description.



**Fig. 3:** The utilization of Reward3D in scoring both positive examples and negative examples (*left*: inconsistency, *right*: multi-face issue) reveals that the model can effectively distinguish negative examples.

## 6.1 Experiment Setup

Our experiments can be divided into two parts: (1) Comparative experiments on DreamReward Sec. 6.2, and (2) Comparative experiments on Reward3D Sec. 6.3.

In Sec. 6.2, we compare our proposed DreamReward with five baseline 3D models: DreamFusion [33], ProlificDreamer [49], Latent-NeRF [27], MVDream [39], and Fantasia3D [4]. To align with the test results provided in GPTEval3D [51], we utilize the official implementations of each method when available. Alternatively, we turn to threestudio’s [10] implementation. The comparative results of the visualization are presented. Simultaneously, to obtain convincing results, four evaluation metrics are used here: CLIP [36], GPTEval3D [51], ImageReward [52], and our Reward3D. For the test dataset, we use 110 prompts generated by GPTEval3D, consisting of prompts with varying levels of creativity and complexity. Additionally, we conduct a user study to further demonstrate the alignment of our method with human preferences.

**Table 1: Quantitative comparisons** on 110 prompts generated by GPTEval3D [51]. We compared our DreamReward with DreamFusion [33], ProlificDreamer [49], Latent-NeRF [27], MVDream [39], and Fantasia3D [4]. We calculate CLIP $\uparrow$  [36], ImageReward $\uparrow$  [52], GPTEval3D  $\uparrow$  [51] and Reward3D $\uparrow$ . Meanwhile, to demonstrate the alignment of our trained Reward3D model with human preferences, we annotated all generated results by researchers. The ranking results in the table below indicate a high alignment between our Reward3D model and the annotated results.

Dataset & Model	110 Prompts from GPTEval3D [51]									
	Human Eval		GPTEval3D		Reward3D		ImageReward		CLIP	
	Rank	Win	Rank	Score	Rank	Score	Rank	Score	Rank	Score
DreamFusion [33]	6	97	6	1000	6	-1.597	5	-1.489	5	0.224
Fantasia3D [4]	5	167	5	1006	5	-1.582	6	-1.521	6	0.222
ProlificDreamer [49]	4	246	4	1152	4	-0.195	4	-0.639	3	0.252
Latent-NeRF [27]	3	287	3	1173	3	-0.012	2	-0.350	2	0.257
MVDream [39]	2	375	2	1224	2	0.246	3	-0.541	4	0.243
Spearman $\rho$ to Human Eval.	-		<b>1.00</b>		<b>1.00</b>		0.80		0.60	
<b>DreamReward(Ours)</b>	<b>1</b>	<b>478</b>	<b>1</b>	<b>1480</b>	<b>1</b>	<b>2.594</b>	<b>1</b>	<b>1.833</b>	<b>1</b>	<b>0.274</b>

In Sec. 6.3, we conduct detailed comparative experiments and user studies on the aforementioned prompt set and 3D baselines. In the course of assessing these models, a large amount of results indicate that our Reward3D better aligns with human preferences compared to existing methods.

## 6.2 DreamFL

**Qualitative Comparison.** Figure 4 and Figure 5 show the 3D assets generated by four baselines for multiple prompts, allowing for intuitive visual comparison. We observe that the generated results using DreamFusion [33], ProlificDreamer [49] and Magic3D [21] deviate from the text content and also suffer from multi-face problems. While MVDream [39] can generate high-quality 3D assets with visual consistency, its results still deviate from the given prompt content (as indicated by the red-highlighted text). In comparison, our model can generate 3D assets that align closely with the given prompt while maintaining visual consistency and meeting human aesthetic preferences.

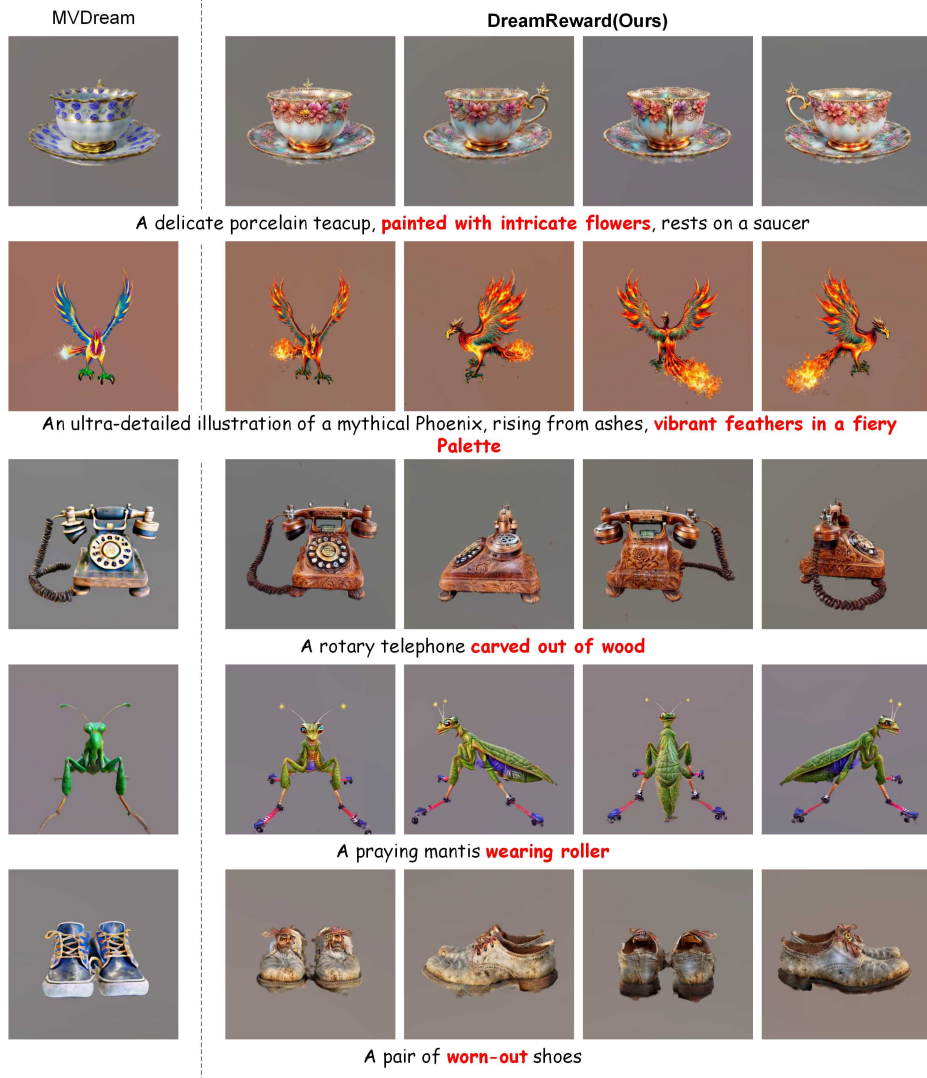
**Quantitative Comparison.** In Table 1, we compare our DreamReward with five baselines. It indicates that our results consistently outperform other baseline models across multiple evaluation criteria. In Figure 6, we demonstrate that using GPTEval3D [51], our results consistently outperform other baseline models across all 6 criteria: Text-asset alignment, 3D plausibility, Texture details, Geometry details, Texture-geometry coherency, and overall. For the user study, we extract 120 images from the 60-degree rotating videos rendered by threestudio [10]. Each annotator randomly receives 5 sets of multi-view images generated by random methods and is asked to rate them on three aspects: multi-view consistency, consistency between text and model, and personal preference. In



**Fig. 4:** Comparison with four baselines. The results indicate that existing 3D generation models do not align well with human preferences (as highlighted in red). Conversely, our DreamReward results conform more closely to human preferences.

addition, they should also choose the favorite one. Finally, we collect the results from 30 participants on 20 text prompts, as shown in Table 2 and Figure 6. In Table 2 we observe that most users consider our results with the highest alignment, generation quality and second consistency. In Figure 6 we observe that our DreamReward is preferable (65%) by the raters on average.

**Ablation Study.** We conduct an ablation study based on the backbone. To verify the strength of our method, we implement our DreamFL algorithm on the basis of the DreamFusion architecture. For a fair comparison, we chose the same 2D diffusion model *stabilityai/stable-diffusion-2-1-base* [38]. The results of Figure 7 indicate that even for backbones with average performance, incorpo-

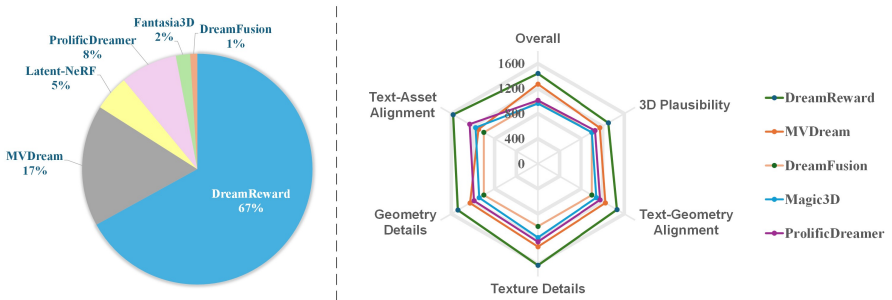


**Fig. 5:** More generated results using our DreamReward. Our work can generate 3D assets of higher alignment, while maintaining consistency across multiple perspectives.

rating our DreamFL algorithm can achieve better generation quality and text alignment

### 6.3 Evaluation of Reward3D

**Quantitative Comparison.** We evaluate the capability of our Reward3D model to determine whether it possesses judgment abilities that align with



**Fig. 6: Left:** User study of the rate from volunteers’ preference for each method in the inset pie chart, **Right:** Holistic evaluation using GPTEval3D. The Radar charts report the Elo rating for each of the 6 criteria. The results indicate that our results consistently rank first across all metrics.

**Table 2:** Quantitative comparisons on the alignment, quality and multi-view consistency score in a user study, rated on a scale of 1-6, with higher scores indicating better performance.

	Alignment $\uparrow$	Quality $\uparrow$	Consistency $\uparrow$	Average $\uparrow$
DreamFusion [33]	2.65	1.95	2.95	2.52
Fantasia3D [4]	2.88	3.23	2.50	2.87
ProlificDreamer [49]	3.90	3.78	3.13	3.60
Latent-NeRF [27]	3.45	3.10	3.18	3.24
MVDream [39]	3.88	4.40	<b>5.38</b>	4.55
<b>DreamReward(Ours)</b>	<b>4.88</b>	<b>5.03</b>	5.30	<b>5.07</b>

human aesthetics. This serves to substantiate the feasibility of employing the Reward3D model instead of human evaluation for assessing the quality of 3D assets. The results presented in Table 1 indicate that the Reward3D model’s assessments of 3D assets generated by various models are consistent with human aesthetics, whereas ImageReward [52] and CLIP [34] are not. It is also noteworthy that the results in Table 1 reveal that GPTEval3D’s [51] (a non-feature model using GPT-4V [1]) evaluating competency is also nearly identical to human assessments. However, in comparison to GPT-4V, our Reward3D model is exceptionally lightweight and offers far faster inference speeds, while still retaining strong evaluation abilities for 3D assets. This suggests a significant advantage of utilizing our Reward3D model for evaluating text-to-3D models as well as for the DreamFL method.

**Representative Cases.** We further present more visualization results with representative cases. In Figure 3, We select two 3D results and replace one of the viewpoint images with incorrect content, thus creating two negative examples (left: inconsistency, right: multi-face issue). To ease identification, we mark the altered viewpoint images with a red border. We observe that our Reward3D gives



**Fig. 7: Ablation study.** we change the backbone of our DreamFL to DreamFusion [33] and select *stabilityai/stable-diffusion-2-1-base* [38] for 2D diffusion model. We can observe that our method still generates 3D content with higher quality and text alignment.

lower scores to negative examples. These results show that Reward3D accurately assesses 3D content based on human preference and strong 3D awareness.

## 7 Conclusion

In this paper, we propose a novel text-to-3D framework called DreamReward for human preference alignment. To the best of our knowledge, we are among the first to utilize RLHF for 3D generation. Specifically, we first construct a new 3D dataset annotated with human preferences consisting of 2530 prompts. Then we train a 3D evaluation model Reward3D on this dataset to better align with human preferences. Powered by the Reward3D model, we further introduce our DreamFL algorithm, which achieves high-fidelity, multi-view consistent, and faithfully human preference-aligned text-to-3D generation. Extensive quantitative and qualitative experiments verify that our DreamReward framework can generate 3D assets with strong human preference alignment.

**Limitations and future work.** Although our proposed DreamReward can generate high-quality and preference-aligned 3D results, there remain requirements for further improvement in the diversity limited by the size of annotated 3D dataset. In our future work, we will continue to optimize our Reward3D on larger datasets and attempt to incorporate more cameras and orientation information into the Reward3D architecture for better performance.



## References

1. Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F.L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., et al.: Gpt-4 technical report. arXiv preprint arXiv:2303.08774 (2023) [3](#), [13](#), [20](#)
2. Black, K., Janner, M., Du, Y., Kostrikov, I., Levine, S.: Training diffusion models with reinforcement learning (2024) [2](#), [4](#), [7](#)
3. Chang, A.X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., Xiao, J., Yi, L., Yu, F.: Shapenet: An information-rich 3d model repository (2015) [3](#)
4. Chen, R., Chen, Y., Jiao, N., Jia, K.: Fantasia3d: Disentangling geometry and appearance for high-quality text-to-3d content creation (2023) [1](#), [3](#), [9](#), [10](#), [13](#), [20](#)
5. Chen, Y., Chen, Z., Zhang, C., Wang, F., Yang, X., Wang, Y., Cai, Z., Yang, L., Liu, H., Lin, G.: Gaussianeditor: Swift and controllable 3d editing with gaussian splatting (2023) [1](#), [3](#)
6. Deitke, M., Schwenk, D., Salvador, J., Weihs, L., Michel, O., VanderBilt, E., Schmidt, L., Ehsani, K., Kembhavi, A., Farhadi, A.: Objaverse: A universe of annotated 3d objects (2022) [4](#)
7. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis (2021) [3](#)
8. Ding, L., Dong, S., Huang, Z., Wang, Z., Zhang, Y., Gong, K., Xu, D., Xue, T.: Text-to-3d generation with bidirectional diffusion using both 2d and 3d priors (2023) [7](#)
9. Fan, Y., Watkins, O., Du, Y., Liu, H., Ryu, M., Boutilier, C., Abbeel, P., Ghavamzadeh, M., Lee, K., Lee, K.: Dpok: Reinforcement learning for fine-tuning text-to-image diffusion models (2023) [4](#)
10. Guo, Y.C., Liu, Y.T., Shao, R., Laforte, C., Voleti, V., Luo, G., Chen, C.H., Zou, Z.X., Wang, C., Cao, Y.P., Zhang, S.H.: threestudio: A unified framework for 3d content generation. <https://github.com/threestudio-project/threestudio> (2023) [7](#), [9](#), [10](#)
11. Gupta, A., Xiong, W., Nie, Y., Jones, I., Oğuz, B.: 3dgen: Triplane latent diffusion for textured mesh generation (2023) [1](#)
12. He, Y., Bai, Y., Lin, M., Zhao, W., Hu, Y., Sheng, J., Yi, R., Li, J., Liu, Y.J.: T<sup>3</sup>bench: Benchmarking current progress in text-to-3d generation (2023) [3](#)
13. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models (2020) [3](#)
14. Hong, Y., Zhang, K., Gu, J., Bi, S., Zhou, Y., Liu, D., Liu, F., Sunkavalli, K., Bui, T., Tan, H.: Lrm: Large reconstruction model for single image to 3d (2023) [1](#)
15. Jain, A., Mildenhall, B., Barron, J.T., Abbeel, P., Poole, B.: Zero-shot text-guided object generation with dream fields (2022) [3](#)
16. Jun, H., Nichol, A.: Shap-e: Generating conditional 3d implicit functions (2023) [1](#), [3](#)
17. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering (2023) [1](#), [3](#)
18. Li, J., Li, D., Savarese, S., Hoi, S.: Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models (2023) [3](#)
19. Li, J., Li, D., Xiong, C., Hoi, S.: Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation (2022) [3](#)
20. Li, W., Chen, R., Chen, X., Tan, P.: Sweetdreamer: Aligning geometric priors in 2d diffusion for consistent text-to-3d (2023) [7](#)
21. Lin, C.H., Gao, J., Tang, L., Takikawa, T., Zeng, X., Huang, X., Kreis, K., Fidler, S., Liu, M.Y., Lin, T.Y.: Magic3d: High-resolution text-to-3d content creation (2023) [1](#), [3](#), [6](#), [10](#), [19](#)

22. Liu, F., Wu, D., Wei, Y., Rao, Y., Duan, Y.: Sherpa3d: Boosting high-fidelity text-to-3d generation via coarse 3d prior (2023) [7](#)
23. Liu, M., Shi, R., Chen, L., Zhang, Z., Xu, C., Wei, X., Chen, H., Zeng, C., Gu, J., Su, H.: One-2-3-45++: Fast single image to 3d objects with consistent multi-view generation and 3d diffusion (2023) [1](#)
24. Liu, R., Wu, R., Hoorick, B.V., Tokmakov, P., Zakharov, S., Vondrick, C.: Zero-1-to-3: Zero-shot one image to 3d object (2023) [1](#)
25. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization (2019) [6](#)
26. Luo, T., Rockwell, C., Lee, H., Johnson, J.: Scalable 3d captioning with pretrained models (2023) [2, 4](#)
27. Metzger, G., Richardson, E., Patashnik, O., Giryes, R., Cohen-Or, D.: Latent-nerf for shape-guided generation of 3d shapes and textures (2022) [9, 10, 13, 20](#)
28. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis (2020) [3](#)
29. Nichol, A., Dhariwal, P., Ramesh, A., Shyam, P., Mishkin, P., McGrew, B., Sutskever, I., Chen, M.: Glide: Towards photorealistic image generation and editing with text-guided diffusion models (2022) [3](#)
30. Nichol, A., Jun, H., Dhariwal, P., Mishkin, P., Chen, M.: Point-e: A system for generating 3d point clouds from complex prompts (2022) [3](#)
31. OpenAI, :, Achiam, J., Adler, S., Agarwal, S., et al.: Gpt-4 technical report (2023) [4](#)
32. Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C.L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., Lowe, R.: Training language models to follow instructions with human feedback (2022) [2, 4, 6](#)
33. Poole, B., Jain, A., Barron, J.T., Mildenhall, B.: Dreamfusion: Text-to-3d using 2d diffusion (2022) [1, 3, 6, 9, 10, 13, 14, 19, 20](#)
34. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I.: Learning transferable visual models from natural language supervision (2021) [2, 3, 13](#)
35. Rafailov, R., Sharma, A., Mitchell, E., Ermon, S., Manning, C.D., Finn, C.: Direct preference optimization: Your language model is secretly a reward model (2023) [4](#)
36. Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., Chen, M.: Hierarchical text-conditional image generation with clip latents (2022) [3, 9, 10](#)
37. Roberts, A., Raffel, C., Lee, K., Matena, M., Shazeer, N., Liu, P.J., Narang, S., Li, W., Zhou, Y.: Exploring the limits of transfer learning with a unified text-to-text transformer. Tech. rep., Google (2019) [3](#)
38. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models (2022) [3, 11, 14](#)
39. Shi, Y., Wang, P., Ye, J., Long, M., Li, K., Yang, X.: Mvdream: Multi-view diffusion for 3d generation (2023) [1, 3, 4, 8, 9, 10, 13, 18, 19, 20, 21](#)
40. Sohl-Dickstein, J., Weiss, E.A., Maheswaranathan, N., Ganguli, S.: Deep unsupervised learning using nonequilibrium thermodynamics (2015) [3](#)
41. Stiennon, N., Ouyang, L., Wu, J., Ziegler, D.M., Lowe, R., Voss, C., Radford, A., Amodei, D., Christiano, P.: Learning to summarize from human feedback (2022) [2, 4, 6](#)
42. Tang, J., Chen, Z., Chen, X., Wang, T., Zeng, G., Liu, Z.: Lgm: Large multi-view gaussian model for high-resolution 3d content creation (2024) [1](#)
43. Tang, J., Ren, J., Zhou, H., Liu, Z., Zeng, G.: Dreamgaussian: Generative gaussian splatting for efficient 3d content creation (2023) [1](#)



44. Team, G., Anil, R., Borgeaud, S., et al.: Gemini: A family of highly capable multimodal models (2023) [4](#)
45. Wada, Y., Kaneda, K., Saito, D., Sugiura, K.: Polos: Multimodal metric learning from human feedback for image captioning (2024) [2](#)
46. Wallace, B., Dang, M., Rafailov, R., Zhou, L., Lou, A., Purushwalkam, S., Ermon, S., Xiong, C., Joty, S., Naik, N.: Diffusion model alignment using direct preference optimization (2023) [4](#), [7](#)
47. Wang, H., Du, X., Li, J., Yeh, R.A., Shakhnarovich, G.: Score jacobian chaining: Lifting pretrained 2d diffusion models for 3d generation (2022) [3](#)
48. Wang, Z., Li, M., Chen, C.: Luciddreaming: Controllable object-centric 3d generation (2023) [3](#), [6](#)
49. Wang, Z., Lu, C., Wang, Y., Bao, F., Li, C., Su, H., Zhu, J.: Prolificdreamer: High-fidelity and diverse text-to-3d generation with variational score distillation (2023) [1](#), [3](#), [6](#), [7](#), [9](#), [10](#), [13](#), [19](#), [20](#)
50. Wei, M., Zhou, J., Sun, J., Zhang, X.: Adversarial score distillation: When score distillation meets gan (2023) [3](#)
51. Wu, T., Yang, G., Li, Z., Zhang, K., Liu, Z., Guibas, L., Lin, D., Wetzstein, G.: Gpt-4v(ision) is a human-aligned evaluator for text-to-3d generation (2024) [2](#), [3](#), [5](#), [9](#), [10](#), [13](#), [19](#), [20](#), [21](#), [24](#)
52. Xu, J., Liu, X., Wu, Y., Tong, Y., Li, Q., Ding, M., Tang, J., Dong, Y.: Imagereward: Learning and evaluating human preferences for text-to-image generation (2023) [2](#), [4](#), [6](#), [7](#), [9](#), [10](#), [13](#)
53. Yang, K., Tao, J., Lyu, J., Ge, C., Chen, J., Li, Q., Shen, W., Zhu, X., Li, X.: Using human feedback to fine-tune diffusion models without any reward model (2023) [4](#), [7](#)
54. Zhang, L., Rao, A., Agrawala, M.: Adding conditional control to text-to-image diffusion models (2023) [7](#)
55. Zhu, J., Zhuang, P.: Hifa: High-fidelity text-to-3d generation with advanced diffusion guidance (2023) [3](#), [6](#)
56. Zhu, Z., Zhao, H., He, H., Zhong, Y., Zhang, S., Guo, H., Chen, T., Zhang, W.: Diffusion models for reinforcement learning: A survey (2024) [2](#)
57. Zhuang, J., Wang, C., Liu, L., Lin, L., Li, G.: Dreameditor: Text-driven 3d scene editing with neural fields (2023) [3](#)
58. Ziegler, D.M., Stiennon, N., Wu, J., Brown, T.B., Radford, A., Amodei, D., Christiano, P., Irving, G.: Fine-tuning language models from human preferences (2020) [4](#), [6](#)

## A The derivation of DreamReward

In this section, we provide a more detailed derivation of **DreamReward**. First, let me begin by revisiting the core assumption of this paper:

$$\nabla_{\theta} \mathcal{L}_{\text{Reward}}(\theta) \approx \mathbb{E}_{t, \epsilon, c} \left[ \omega(t) \left( \epsilon_{\phi}^r(\mathbf{x}_t, t, y) - \epsilon \right) \frac{\partial \mathbf{g}(\theta, c)}{\partial \theta} \right] \quad (9)$$

Where  $\epsilon_{\phi}^r(\mathbf{x}_t, t, y)$  denotes an ideal noise prediction network aligned with human preference. We approximate it using Reward3D and  $\epsilon_{\phi}(\mathbf{x}_t, t, y)$ , as follows:

$$\epsilon_{\phi}^r(\mathbf{x}_t, t, y) = \epsilon_{\phi}(\mathbf{x}_t, t, y) - \delta\epsilon \quad (10)$$

The difference  $\delta\epsilon$  between the  $\epsilon_{\phi}^r(\mathbf{x}_t, t, y)$  and  $\epsilon_{\phi}(\mathbf{x}_t, t, y)$  can be calculated by:

$$\delta\epsilon = -\frac{\partial r(y, \hat{\mathbf{x}}_t, c)}{\partial \hat{\mathbf{x}}_t} \frac{\partial \hat{\mathbf{x}}_t}{\partial \epsilon_{\phi}} = \lambda \frac{\partial r(y, \hat{\mathbf{x}}_t, c)}{\partial \hat{\mathbf{x}}_t} \frac{\partial \hat{\mathbf{x}}_t}{\partial \mathbf{x}} = \lambda \frac{\partial r(y, \hat{\mathbf{x}}_t, c)}{\partial \mathbf{x}} \quad (11)$$

Where  $\hat{\mathbf{x}}_t$  represents the predicted  $\mathbf{x}_0$  using  $x_t$ . In Eq. 11,  $\lambda$  encompasses all constant terms to ensure the equation holds. In practical implementation, we replace  $\lambda$  with  $\lambda_r$ , defined as follows:

$$\lambda_r = t(\mathcal{L}_{\text{SDS}}, r(y, \hat{\mathbf{x}}_t, c)) \times \mu \quad (12)$$

Where  $\mu$  denotes a constant hyperparameter, and  $t$  represents the weighting function. Therefore, Eq. 9 can be reorganized as:

$$\nabla_{\theta} \mathcal{L}_{\text{Reward}}(\theta) \approx \mathbb{E}_{t, \epsilon, c} \left[ \underbrace{\omega(t) (\epsilon_{\phi}(\mathbf{x}_t, t, y) - \lambda_r \frac{\partial r(y, \hat{\mathbf{x}}_t, c)}{\partial \mathbf{g}(\theta, c)})}_{\epsilon_{\phi}^r(\mathbf{x}_t, t, y)} - \epsilon \right] \frac{\partial \mathbf{g}(\theta, c)}{\partial \theta}. \quad (13)$$

SDS can be reorganized as:

$$\mathcal{L}_{\text{Reward}}(\theta) \approx \mathcal{L}_{\text{SDS}}(\theta) - \lambda_r r(y, \hat{\mathbf{x}}_t, c). \quad (14)$$

## B Additional Implementation Details

### B.1 Pseudo-code for DreamFL

A more detailed pseudo-code for DreamFL is presented in Algorithm 1.

### B.2 Training Details

The time consumption for our DreamReward is 40 minutes (which is similar to MVDream [39]). It is worth mentioning that our DreamFL can also be applied to traditional RLHF. Specifically, by utilizing the DreamFL algorithm, a pre-trained 3D asset can be fine-tuned in just 2 minutes. The resulting 3D asset will exhibit finer texture details, more appealing colors, and content that better aligns with textual descriptions.

---

**Algorithm 1** Pseudo-code for DreamFL

---

**Input:** Large pretrained text-to-image diffusion model  $\epsilon_\phi$ . Reward3D model  $r$ . Learning rate  $\eta_1$  for 3D structures parameters. A prompt  $y$ . The weight of reward loss  $\lambda_r$ . Evaluating threshold of noising time  $t_{\text{threshold}}$ .

- 1: **initialize** A 3D structure presenting with NeRF  $\theta$ .
- 2: **while** not converged **do**
- 3: Randomly sample a camera pose  $c$ , 2D noise  $\epsilon \sim \mathcal{N}(\mathbf{0}, I)$ , and timestep  $t \sim \text{Uniform}(\{1, \dots, T\})$ .
- 4: Render at pose  $c$  to get a multiview image  $\mathbf{x}_0 = \mathbf{g}(\theta, c) \in \mathbb{R}^{4 \times H \times W \times C}$ .
- 5: **if**  $t < t_{\text{threshold}}$  **then**
- 6: Add noise  $\epsilon$  to  $\mathbf{x}_0$  and get  $\mathbf{x}_t$ .
- 7: Denoise with the predicting noise  $\hat{\mathbf{x}}_t = (\mathbf{x}_t - \sqrt{1 - \alpha_t} \epsilon_\phi(\mathbf{x}_t, t, y)) / \sqrt{\alpha_t}$ .
- 8:  $\mathcal{L}_{\text{Reward}} = \mathcal{L}_{\text{SDS}} - \lambda_r r(y, \hat{\mathbf{x}}_t, c)$ .
- 9:  $\delta\epsilon \leftarrow \lambda_r \nabla_{\mathbf{g}(\theta, c)} r$
- 10: **else**
- 11:  $\mathcal{L}_{\text{Reward}} = \mathcal{L}_{\text{SDS}}$
- 12:  $\delta\epsilon \leftarrow 0$ .
- 13: **end if**
- 14:  $\nabla_\theta \mathcal{L}_{\text{Reward}}(\theta) \approx \mathbb{E}_{t, \epsilon, c} \left[ \omega(t) (\epsilon_\phi(\mathbf{x}_t, t, y) - \delta\epsilon - \epsilon) \frac{\partial \mathbf{g}(\theta, c)}{\partial \theta} \right]$
- 15:  $\theta \leftarrow \theta - \eta \nabla_\theta \mathcal{L}$ .
- 16: **end while**
- 17: **return**

---

## C Additional Experiments

In this section, we provide more visual examples to demonstrate that our DreamReward is more aligned with human preferences.

### C.1 More Visual Results

Figure 9, 10 presents additional comparison results with MVDream [39]. We adhere to the definition used in GPTEval3D [51], dividing prompts into the subject part and the properties part (e.g., in "a sleeping cat," "cat" is the subject while "sleeping" represents the properties). Based on experiments [39], it is observed that MVDream [39] can effectively generate the subject part but encounters difficulties with certain more complex properties. For instance, in the "A dancing elephant" depicted in Figure 9, MVDream [39] can depict the elephant accurately but struggles to convey its dancing attribute. On the other hand, our DreamReward can effectively overcome this issue, thus generating high-quality 3D assets that align with the properties described.

### C.2 More Qualitative Comparisons

Figure 12 presents additional comparison results with DreamFusion [33], ProflificDreamer [49], Magic3D [21] and MVDream [39]. These examples indicate that the properties of the 3D assets generated by baseline models fail to align

with the provided prompts. In contrast, our DreamReward not only ensures consistency between text and 3D output but also provides results that are more in line with human aesthetics.

### C.3 More Quantitative Comparison

When evaluating the 3D assets generated by our DreamReward using GPTEval3D [51], we employ the latest official open-source code and run it with default configuration settings to maintain fairness. To ensure comprehensive assessment, we conduct evaluations through two different comparison methods: (1) competing against 13 baseline models and (2) competing solely against MVDream. We conduct 220 and 110 inquiries/competitions respectively. The Radar charts presented in the main text are derived from the results of the first method, while the specific ELo scores extracted from these charts are detailed in Table 3. Following the completion of evaluation (1), we observe that although we achieve the best results across various metrics, there are instances of structural collapse for certain prompts. Later, we choose a new hyperparameter  $\lambda_r$  and uniformly regenerate 110 3D assets using this updated value. We conduct evaluation (2) based on the new 3D assets, with the ELO scores presented in Table 4.

### C.4 GPT-4V Comparison Examples

To better illustrate the higher quality of our generated results, we choose GPTEval3D [51] for additional pairwise comparison. As shown in Figure 11, we present four pairs of comparison results from the tournament [51]. GPT4V [1] scores each pair from six aspects: Text-asset alignment, 3D plausibility, Texture details, Geometry details, Texture-geometry coherence, and overall.

**Table 3:** The first evaluation results using GPTEval3D [51] correspond to the Radar charts in the main text. We bold the best result and underline the second-best result.

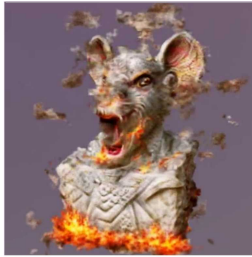
	Text-Asset Alignment	3D-Plausibility	Text-Geometry Alignment	Geometry-Details	Texture-Details	Overall
DreamFusion [33]	1000	1000	1000	1000	1000	1000
Fantasia3D [4]	1068	892	1006	1028	1109	934
ProlificDreamer [49]	<u>1262</u>	1059	1152	1181	1246	1013
Latent-NeRF [27]	1222	1145	1157	1161	1180	1179
MVDream [39]	1098	<u>1148</u>	<u>1251</u>	<u>1256</u>	<u>1325</u>	<u>1271</u>
<b>Ours</b>	<b>1567</b>	<b>1305</b>	<b>1465</b>	<b>1479</b>	<b>1622</b>	<b>1442</b>

**Table 4:** The second evaluation results using GPTEval3D [51]. We bold the best result.

	Text-Asset Alignment	3D- Plausibility	Text- Geometry Alignment	Geometry- Details	Texture- Details	Overall
MVDream [39]	779	815	810	784	780	784
<b>Ours</b>	<b>1221</b>	<b>1185</b>	<b>1190</b>	<b>1216</b>	<b>1220</b>	<b>1216</b>



$$\lambda_r = t(\mathcal{L}_{\text{SDS}}, r(\mathbf{y}, \hat{\mathbf{x}}_t, \mathbf{c}))\mu$$



$$\lambda_r = 10000$$



$$\lambda_r = 0$$

**Fig. 8:** More ablation study about  $\lambda_r$ , the prompt is: "A marble bust of a mouse with its mouth open, spewing fire."

### C.5 More Ablation Study

We conduct ablation study on the weight parameter  $\lambda_r$  of the reward. Figure 8 indicate that when  $\lambda_r$  is set to 0 (i.e., degenerating into a standard SDS), the generated results only consist of subjects without properties. When  $\lambda_r$  is set to a constant 10000, the form of the generated results tends to collapse. However, with the application of our designed weight function  $t$ , the generated results maintain both high quality and alignment simultaneously.

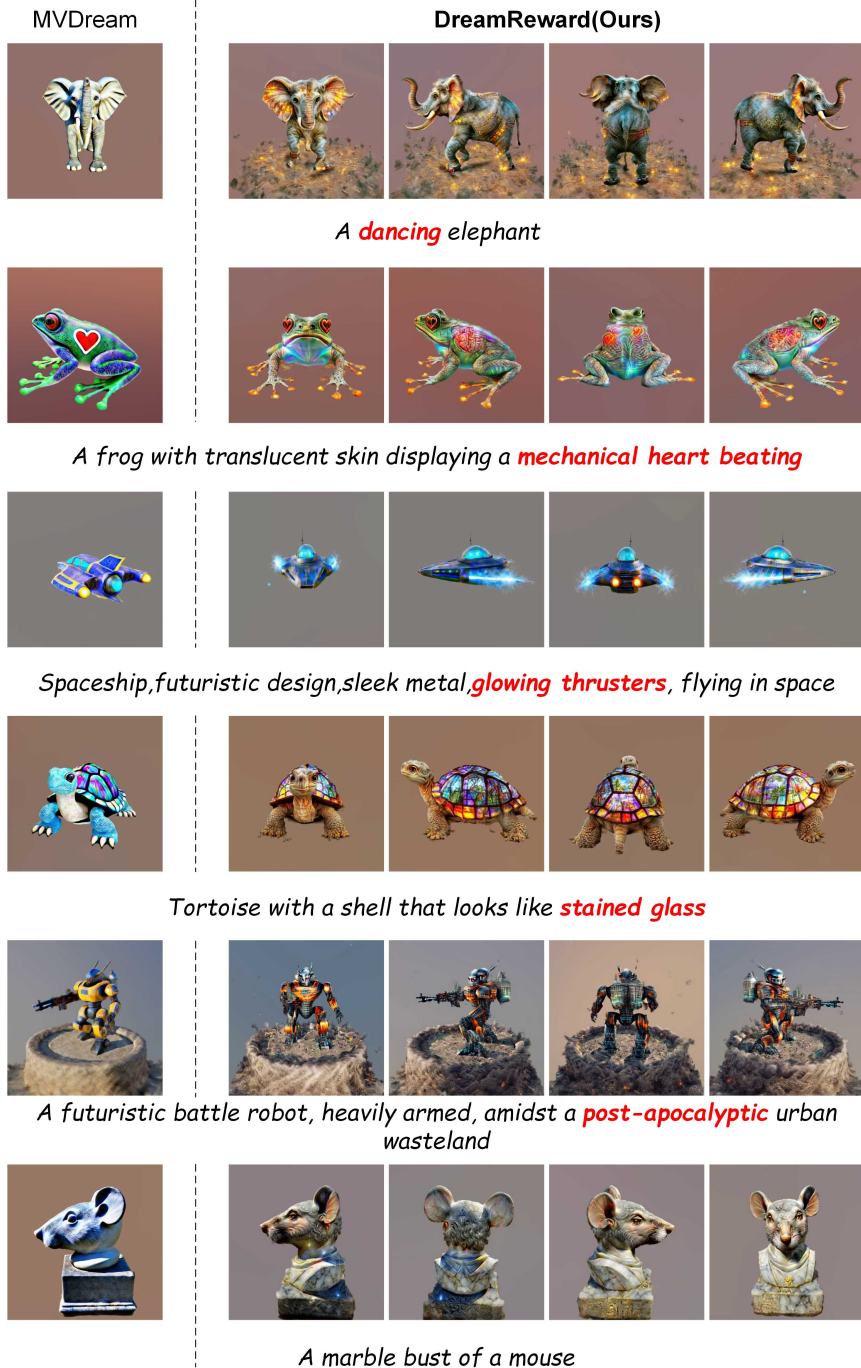


Fig. 9: More generated results using our DreamReward

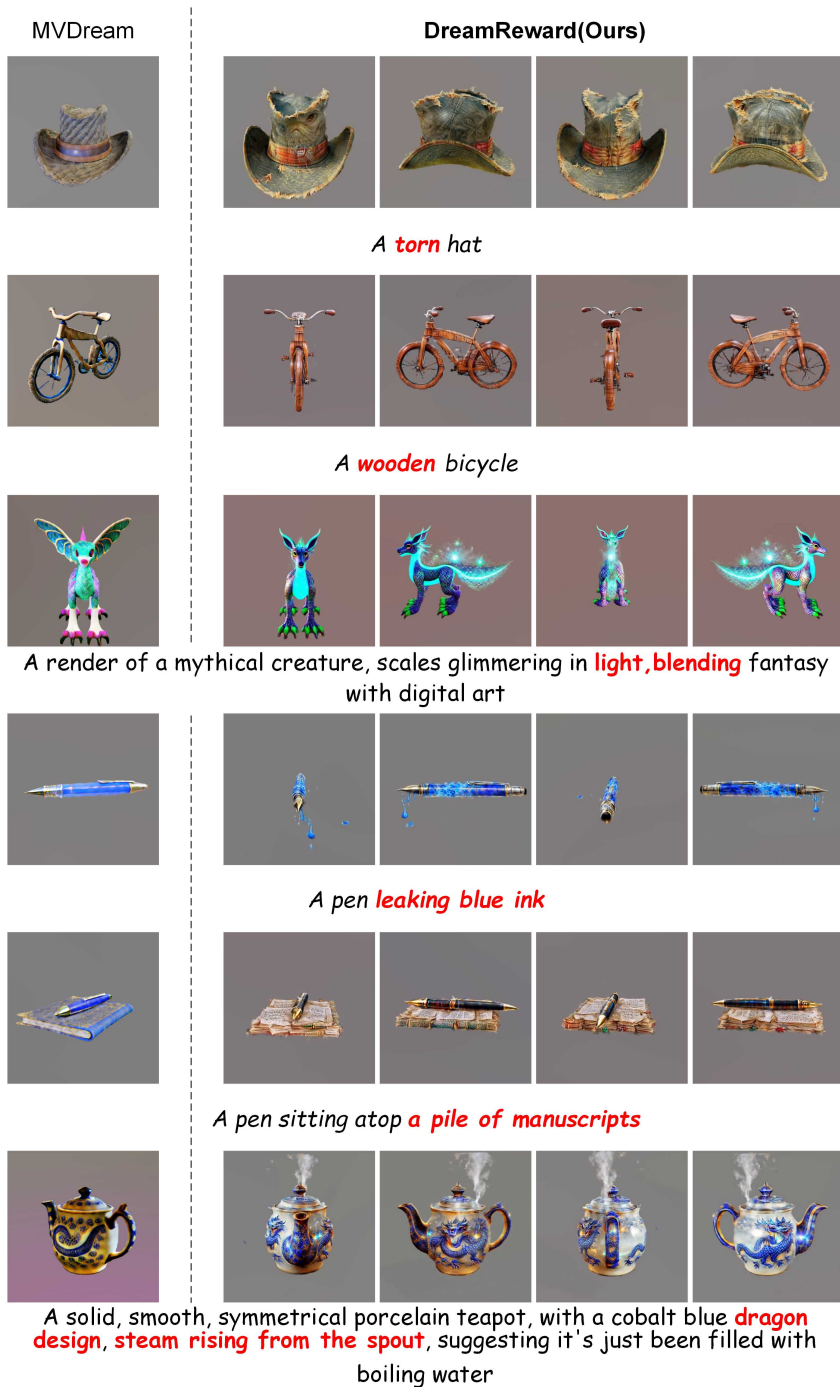


Fig. 10: More generated results using our DreamReward



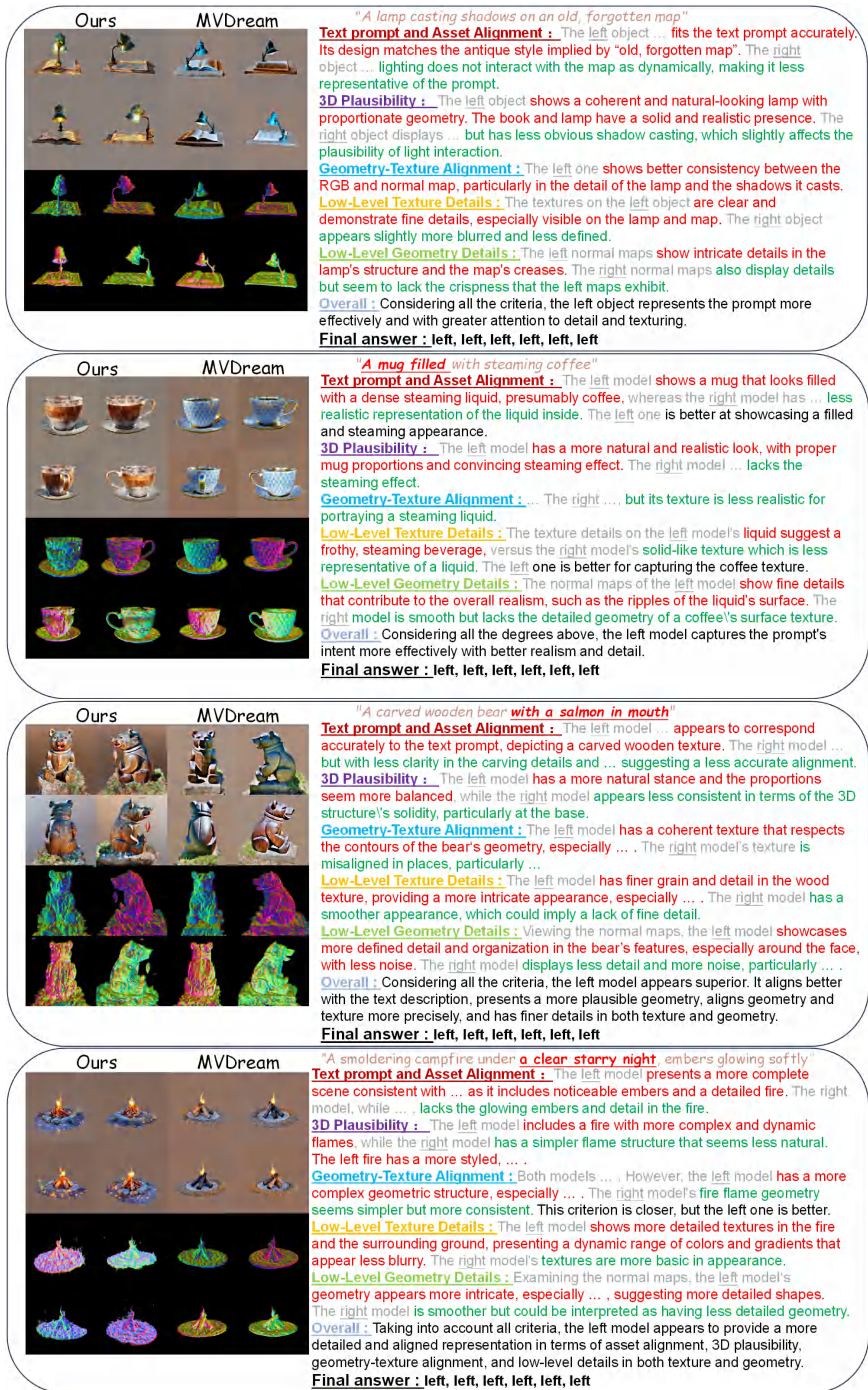


Fig. 11: Examples of the analysis by GPTEval3D [51]

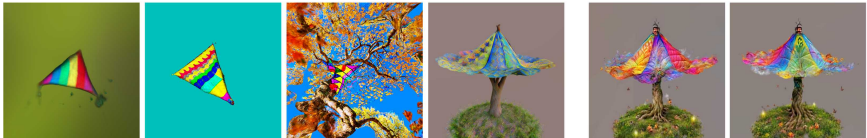




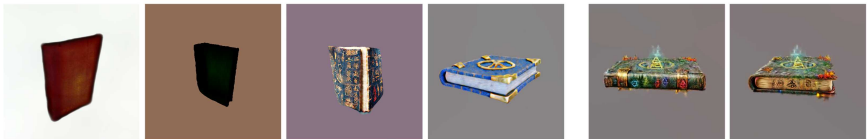
*A mesmerizing dance performed by a **kaleidoscope** of butterflies*



*Sand hourglass, sand **glitters like stars***



*A **colorful** kite tangled **in the branches of an oak tree**, fabric fluttering in the wind.*



*A book **bound in mysterious symbols***



*An ancient, weathered statue, now covered in a blanket of moss*

**Fig. 12: More comparisons with four baselines**