
Machine Unlearning in 3D Generation: A Perspective-Coherent Acceleration Framework

Anonymous Author(s)

Affiliation

Address

email

1 In this document, we present supplementary materials that couldn't be accommodated within the
2 main manuscript due to page limitations. Specifically, we offer additional details on the complete
3 algorithmic design of our framework, the threshold selection strategy for dynamic acceleration,
4 detailed setups for various unlearning tasks, qualitative results and hyperparameter configurations,
5 3D reconstruction outcomes, and a discussion of current limitations.

6 1 Supplementary Methods

7 1.1 Framework Algorithms

8 We propose a novel two-stage framework that integrates dynamic timestep skipping with directional
9 unlearning, enabling efficient and precise removal of targeted concepts from a diffusion-based gener-
10 ative model. This section provides supplementary algorithms for our proposed framework, including
11 Algorithm 1 and Algorithm 2.

Algorithm 1 Dynamic Skipping via Interpolation

Require: Total timesteps T , base angles θ_b , all the sample angles θ , each sample angle θ_s , weight factor w_f , noise perturbation $\epsilon \sim \mathcal{N}(0, 1)$, angle threshold θ_{th} , interpolation upper time step t_{upper} , interpolation lower time step t_{lower} , and noise weight factor ϵ_w , $x_t(\theta)$ represents the denoised result at timestep t during the diffusion process, conditioned on the angle θ . At timestep $t = 0$, $x_0(\theta)$ is the final denoised image, and at timestep $t = T$, $x_T(\theta)$ is the noisy image or latent representation.

1: **Examples of base angles:**
2: • 3 base angles: $\theta_b = \{0^\circ, 120^\circ, -120^\circ\}$
3: • 4 base angles: $\theta_b = \{0^\circ, 90^\circ, -90^\circ, 180^\circ\}$
4: • 8 base angles: $\theta_b = \{0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, -135^\circ, -90^\circ, -45^\circ\}$
5: **for** each sample angle θ_s **do**
6: Compute CLIP similarity $S(\theta_s, \theta_b)$ with key angles
7: Select the two most similar key angles as θ_1, θ_2
8: Determine skip steps based on threshold and similarity
9: Interpolate $x_t(\theta_s)$ from $x_t(\theta_1)$ and $x_t(\theta_2)$ using:
10: $x_t(\theta_s) = \lambda x_t(\theta_1) + (1 - \lambda)x_t(\theta_2)$
11: $\lambda = \frac{S(\theta_s, \theta_1)}{S(\theta_s, \theta_1) + S(\theta_s, \theta_2)}$
12: **if** $|\theta_s - \theta_b| < \theta_{th}$ **then**
13: Use interpolated x_t at $t = T - t_{upper}$
14: **else**
15: Use interpolated x_t at $t = T - t_{lower}$

Description of Algorithm 1: This algorithm accelerates the diffusion process by dynamically skipping denoising steps through interpolation. For each sample-conditioned angle θ_s , it identifies the two most similar base angles θ_1 and θ_2 using a similarity metric (e.g., CLIP similarity). Then, it interpolates the intermediate denoised result $x_t(\theta_s)$ from the known results at θ_1 and θ_2 via a weighted average governed by their similarity scores:

$$x_t(\theta_s) = \lambda x_t(\theta_1) + (1 - \lambda)x_t(\theta_2), \quad \lambda = \frac{S(\theta_s, \theta_1)}{S(\theta_s, \theta_1) + S(\theta_s, \theta_2)}.$$

Depending on whether the sample angle is sufficiently close to a base angle (determined by a threshold θ_{th}), the algorithm either uses the interpolated result at a higher or lower timestep (i.e., fewer or more skipped steps). This allows the system to trade off between fidelity and speed while maintaining semantic consistency.

Algorithm 2 Unlearning via Dynamic Acceleration with Remain and Forget Losses

Require: Pre-trained score network S_t , unlearned model for the target object f_u , fake score network S_f , remain set D_r , unlearn set D_f , override set D_o , all the sample angles θ , each sample angle θ_s , batch size B

- 1: Initialize S_f and f_u from pre-trained model
 - 2: **for** each epoch **do**
 - 3: Sample batch $X_r \sim D_r, X_f \sim D_f, X_o \sim D_o$
 - 4: Call **Algorithm 1** with θ_s ▷ Interpolation Acceleration
 - 5: **for** each sample angle θ_s in θ **do** ▷ Train Fake Score Network
 - 6: Compute $\mathcal{L}_{fn \text{ remain}}(\theta_s)$
 - 7: Compute $\mathcal{L}_{fn \text{ forget}}(\theta_s)$
 - 8: Update S_f using gradient descent ▷ Train Generator
 - 9: Compute $\mathcal{L}_g \text{ remain}(\theta_s)$
 - 10: Compute $\mathcal{L}_g \text{ forget}(\theta_s)$
 - 11: Update f_u using gradient descent
-

Description of Algorithm 2: This algorithm presents a training framework for concept unlearning by alternately optimizing the generator and a fake score network using supervision from the remain, forget, and override datasets. A key innovation of this framework lies in the use of dynamic skipping (realized by Algorithm 1) to accelerate the diffusion process for arbitrary sample angles, enabling efficient training while preserving semantic consistency in generated outputs.

At the beginning of each epoch, Algorithm 1 is invoked to perform interpolation sampling across all base angles. This preprocessing step prepares the interpolated denoising results, allowing for fast inference at any sample angle θ_s by reusing the precomputed intermediate states.

Subsequently, the algorithm iterates through all sample angles θ_s defined in the training setup. For each θ_s , it alternates between training the fake score network and the generator. The score network is updated using remain and forget losses to reflect the desired unlearning behavior, while the generator is optimized using the same objectives to remove target concepts while preserving unrelated features.

By integrating dynamic acceleration and angle-wise alternating optimization, this framework achieves fine-grained control over the forgetting process in diffusion models, while significantly reducing the computational burden of full denoising for every training step.

1.2 Dynamic Acceleration Threshold Selection Basis

Recall that in the main paper (see Eq. (10)), we empirically set the angular threshold $\tau = 20^\circ$ to guide the dynamic adjustment of the denoising timestep t_{jump} . Below, we provide supplementary justification for this choice.

Specifically, we precompute and cache intermediate denoising results for a discrete set of reference viewpoints across all sampling steps. For each non-reference training view, we identify its nearest reference angle via cosine similarity and interpolate the cached features at matched time steps to

approximate the denoising trajectory. This enables a skip-sampling mechanism in which certain sampling steps are bypassed by reusing spatially coherent representations.

Motivated by the inherent geometric consistency among nearby viewpoints, we hypothesize that smaller angular distances to reference views indicate higher structural similarity and, consequently, greater tolerance for step skipping. Based on this observation, we design a dynamic skipping scheme where the number of skipped steps is conditioned on the angular proximity to the nearest reference angle. In later experiments, we quantitatively assess the trade-off between generation quality and sampling efficiency under this dynamic scheme using SSIM, LPIPS, and Δ PSNR, as well as overall training speedup.

1.2.1 Marginal Benefit Analysis

We introduce the concept of marginal benefit as a key indicator for dynamic acceleration threshold selection.

Combining SSIM decrease and LPIPS increase into a single quality loss metric:

$$\text{Marginal Benefit} = \frac{\Delta Q_{\text{total}}}{\Delta Q_{\text{total}}} = \frac{\alpha \cdot \Delta Q_{\text{SSIM}} + \beta \cdot \Delta L_{\text{LPIPS}}}{\alpha \cdot (Q_{\text{previous}} - Q_{\text{current}}) + \beta \cdot (L_{\text{current}} - L_{\text{previous}})} \quad (1)$$

- Objective: Find the threshold range that **maximizes** marginal benefit, i.e., achieve the greatest acceleration improvement with the minimal quality degradation.

- Weight coefficients α and β need to be adjusted based on business requirements (defaulting to 0.5 each).

- Physical meaning:

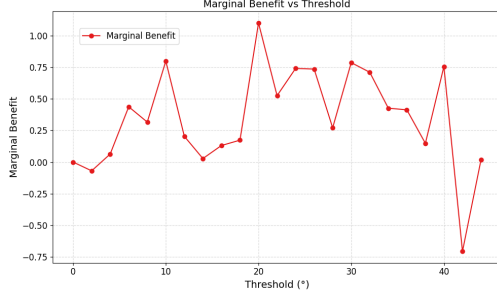
- $\Delta Q_{\text{SSIM}} = Q_{\text{previous}} - Q_{\text{current}}$ (SSIM decrease, larger value means more quality loss).
- $\Delta L_{\text{LPIPS}} = L_{\text{current}} - L_{\text{previous}}$ (LPIPS increase, larger value means more perceptual difference).
- $\Delta S = S_{\text{current}} - S_{\text{previous}}$ (Speed-up Ratio increase, larger value means faster reasoning).

1.2.2 Experimental Setup and Threshold Determination

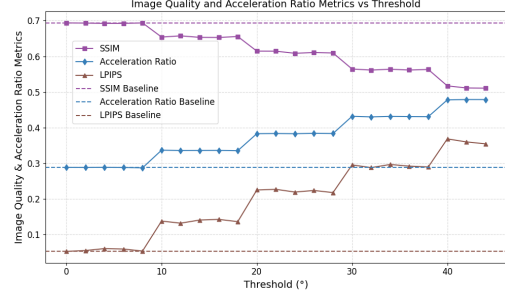
To determine the optimal dynamic acceleration threshold, we sampled 36 viewpoints within the $[0^\circ, 45^\circ]$ range from the base view at 2° intervals, using 4 reference views. Experiments were conducted on the Yellow Car unlearning task. For each candidate threshold, we computed SSIM, LPIPS, and acceleration ratio under the dynamic 4-view, 12-step sampling configuration, and subsequently calculated the marginal benefit. The angle yielding the highest marginal benefit was selected as the optimal dynamic threshold. As shown in Figure 1a, the marginal benefit peaks at a threshold of 20° .

To further validate the effectiveness of the proposed dynamic strategy, we compared it with a static configuration using 4 reference views and 12 uniform steps, without threshold adaptation. This comparison demonstrates that our strategy can achieve acceleration while maintaining high generation quality. The results of this comparative experiment are presented in Figure 1b.

From the figure, we observed that as the angular threshold increases from 0° to 45° , the SSIM decreases from 0.69 to 0.51, while LPIPS increases from 0.05 to 0.36, indicating a consistent trade-off between fidelity and efficiency. Meanwhile, the acceleration ratio improves from the static baseline of 0.29 up to 0.48. Notably, the 20° threshold yields a balanced performance—achieving 0.62 SSIM, 0.22 LPIPS, and 0.38 acceleration ratio—and represents the optimal marginal gain point. Beyond 20° , marginal returns diminish: from 20° to 30° , acceleration increases only 0.06, while LPIPS worsens 0.07. After 30° , visual degradation accelerates, with LPIPS exceeding 0.3 and SSIM dropping below 0.6. These results confirm that moderate thresholds (approximately 20°) achieve the best trade-off, while aggressive skipping leads to diminishing quality returns.



(a) Marginal benefit vs. threshold angle.



(b) Comparison between dynamic and static strategies.

Figure 1: Analysis of threshold selection and strategy comparison.

2 Supplementary Experimental Results

2.1 Experimental Settings

Implementation Details

To cover the full 360° horizontal field of view, we define the angular range as $(-180^\circ, 180^\circ]$. Given the desired number of reference directions $N_{\text{reference}}$, we generate a set of reference angles starting from 0° with a fixed interval of $360^\circ/N_{\text{reference}}$. As these angles are initially defined in the $[0^\circ, 360^\circ)$ range, we map them into $(-180^\circ, 180^\circ]$ to align with the defined coordinate system.

During the training stage of the unlearning task, we additionally sample one angle every 10° over the range from -180° to 180° , excluding the reference angles. This ensures dense and uniform coverage across the entire horizontal span. Such a setup helps the model generalize to diverse viewpoints while maintaining consistency between the training and evaluation angular distributions.

Hyper-parameter Settings

In all experiments, we employ the Adam optimizer, where β_1 and β_2 denote the exponential decay rates for the first and second moment estimates, respectively. The parameters λ and μ represent the regularization coefficients used in the objective function. The term ϵ_t denotes the standard Gaussian noise added during the diffusion process at time step t , with $\epsilon_t \sim \mathcal{N}(0, 1)$.

The *Number of References* represents the number of pre-cached reference angles used for subsequent interpolation to estimate noise; the *Skip Steps* indicates the initial steps skipped during the sampling process.

Table 1: Hyperparameters for Fake Score and Generator

Parameter	Fake Score	Generator
λ	1.0	1.0
μ	0.01	0.01
Optimizer	Adam	Adam
Learning Rate	4×10^{-6}	6×10^{-6}
β_1	0.0	0.0
β_2	0.999	0.999
ϵ_t	10^{-8}	10^{-8}

Table 2: Experimental Settings for Different Reference Angles and Unlearn Effects

Parameter	Reference Angles Experiment	Target Forget Images Experiment
GPU	NVIDIA A100 80GB	NVIDIA A6000 48GB
Batch Size	8	2
Sample Steps	32	32
Training Epochs	5	-
Number of References	-	4
Skip Steps	-	12

2.2 Multi-angle presentation of the results from the main experiment

We provide additional experimental results to supplement the main paper. The following provides concrete examples of the unlearning implementation for the retargeting, stylization, and partial tasks in our experiments.

Table 3: Representative application cases categorized by type.

Category	Case Examples
Style Transfer	Yellow Car Transformation, Metal Style Ice-cream Transformation, Bronze Statue Transformation
Whole Object Retarget	Cherry to Banana, Barrier to Fire Hydrant, Football to Phone
Partial Edit Replacement	Barrel Add Black Lid, Doraemon with Hat, Minion with Backpack, Stool with Pot

Unlearning task 1: Style Transfer

- **Yellow Car Transformation:** In this experimental setup, a frontal image of a *silver car* is designated as the *forget image*, representing the category to be unlearned, while a frontal image of a *yellow car* which is generated by adjusting the color tone of the original image, changing the car body color to yellow while keeping other visual content unchanged, serves as the *override image*, representing the target category. During training, the *forget image* combined with a given *sample angle* is replaced by the *override image* with the same corresponding *sample angle*. This configuration aims to evaluate the model’s ability to forget and override when the object’s appearance attributes, such as color, change.
- **Metal Style Ice-cream Transformation:** The forget image is a Green ice cream cone, while the override image is generated by changing the color of the ice cream to a metallic sheen. Similar to the Yellow Car Transformation case, both the forget angle and the override angle are aligned with the sample angle.
- **Bronze Statue Transformation:** The forget image is a white marble sculpture, while the override image is generated by changing the color of the sculpture to bronze. Again, both the forget angle and the override angle are aligned with the sample angle.

Unlearning task 2: Whole Object Retarget

- **Cherry to Banana:** In this experimental setup, a frontal image of a *cherry* is designated as the *forget image*, representing the category to be unlearned, while a frontal image of a *banana* serves as the *override image*, representing the target category. During training, the *forget image* combined with a given *sample angle* is replaced by the *override image* with the same corresponding *sample angle*. This configuration is designed to evaluate the model’s capability in performing semantic transformation between different object categories.
- **Barrier to Fire Hydrant:** The forget image is a barrier, while the override image is a fire hydrant. Similar to the Cherry to Banana case, both the forget angle and override angle are aligned with the sample angle.
- **Football to Phone:** The forget image is a football, while the override image is a phone. Again, both the forget angle and override angle are aligned with the sample angle.

136 Unlearning task 3: Partial Edit Replacement

137 • **Minion With Backpack** This setting aims to evaluate the model’s response to viewpoint
 138 variations and additional attribute modifications. The *forget image* is a frontal view of
 139 a minion. When the *sample angle* lies within the range $[-90^\circ, 90^\circ]$, the *override image*
 140 is the same as the *forget image*, and both the *forget angle* and *override angle* match the
 141 *sample angle*. However, when the *sample angle* falls outside this range (i.e., side or rear
 142 views), the *override image* is replaced by a rear view of the minion wearing a red backpack,
 143 and the *override angle* is defined as the *sample angle* plus 180° (i.e., the opposite viewing
 144 direction). This setting simulates the unlearning and rewriting behavior when the target
 145 object undergoes structural or appearance changes under different viewpoints. The specific
 146 angle relationships are as follows:

- 147 – If the original *forget angle* is within $[-90^\circ, 90^\circ]$, the guidance condition uses the orig-
 148 inal *forget image* and *forget angle*.
- 149 – If the *forget angle* lies in $[-180^\circ, -90^\circ]$, the guidance condition replaces the image
 150 with the *override image* and adjusts the angle to *forget angle* plus 180° .
- 151 – If the *forget angle* lies in $(90^\circ, 180^\circ]$, the guidance condition replaces the image with
 152 the *override image* and adjusts the angle to *forget angle* minus 180° .

153 • **Barrel Add Black Lid:** The forget image is a wooden barrel, while the override image is
 154 generated by adding a big black lid to the original barrel. Both the forget angle and the
 155 override angle are aligned with the sample angle.

156 • **Doraemon With Hat:** The forget image is a Doraemon, while the override image is gen-
 157 erated by putting a red cap on Doraemon’s head.. Both the forget angle and the override
 158 angle are aligned with the sample angle.

159 • **Stool With Pot:** The forget image is a wooden stool, while the override image is generated
 160 by placing a small plant in a pot on the stool. Again, both the forget angle and override
 161 angle are aligned with the sample angle.

162 Figure 2 presents multi-view visualizations of the unlearning outcomes for various target objects
 163 across multiple categories, demonstrating the consistency and robustness of the unlearning effect
 164 under different viewing angles.

165 In each row, the left-most pair shows the original source object (left) and the desired unlearned target
 166 (right). The right panel visualizes the unlearned results rendered from multiple canonical perspec-
 167 tives (front, side, back, etc.). It can be observed that across diverse object types—including vehicles,
 168 statues, characters, and everyday items—the model consistently applies unlearning effects to gen-
 169 erate novel outputs aligned with the desired target identity or semantics. For instance, the “car” is
 170 reliably altered to resemble a yellow sports model across all views, while the “Doraemon” character
 171 is consistently altered to wear a red hat across all viewpoints, suggesting strong disentanglement and
 172 generalization capacity in the forgetting process.

173 These results validate that our method does not overfit to a single viewpoint, but achieves semanti-
 174 cally coherent forgetting across multiple 3D-consistent renderings, highlighting the model’s capacity
 175 for multi-perspective semantic consistency in unlearning tasks.

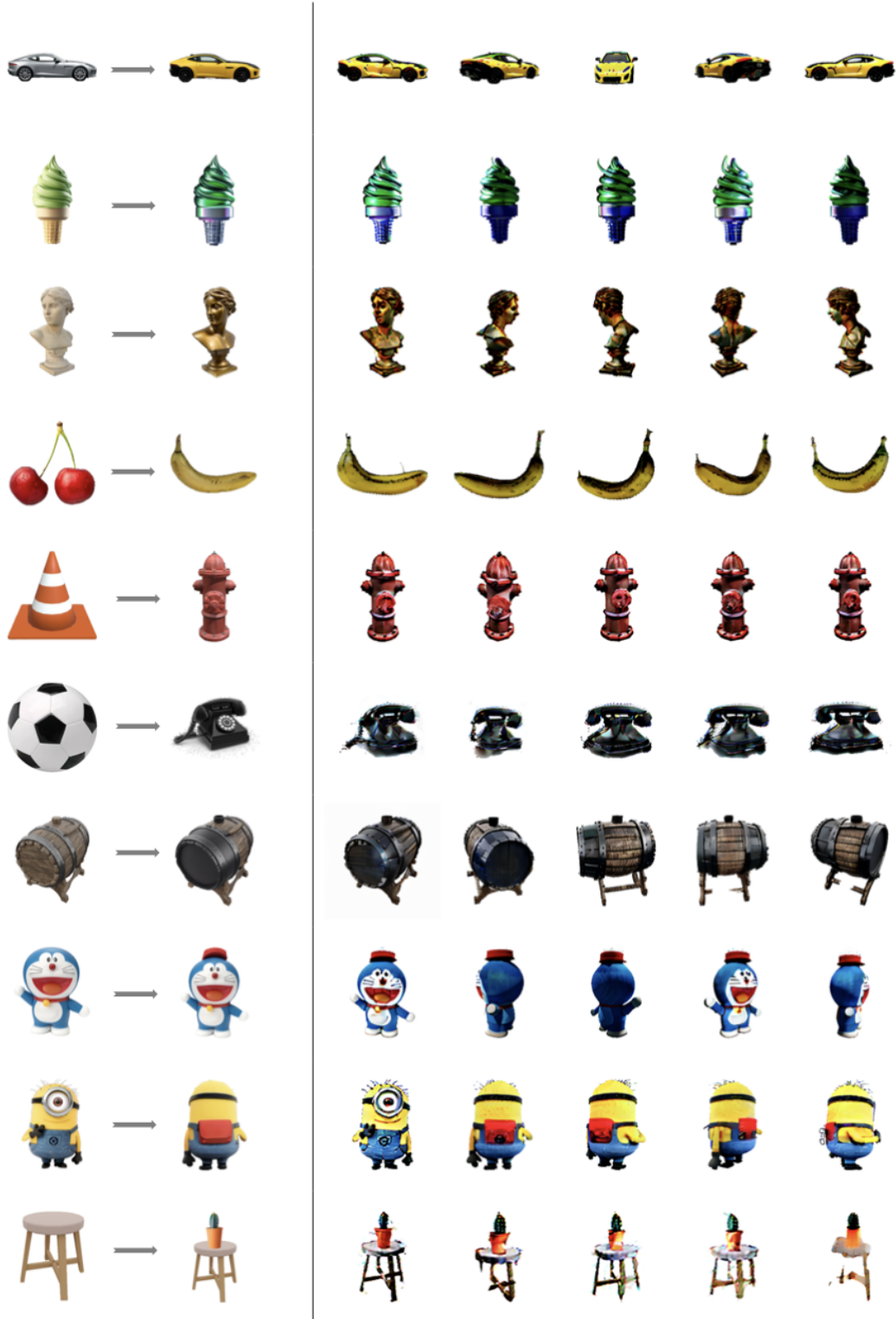


Figure 2: Demonstration of multi-perspective effects on the forget set for different unlearning tasks.

2.3 3D Reconstruction Demonstrations

We showcase 3D reconstruction results for several tasks, presenting pairs of rendered images and depth maps. These results demonstrate that our unlearning strategy not only performs effectively in multi-view consistency settings but also extends to full 3D geometry. Specifically, we observe consistent suppression of undesired concepts across different viewpoints and depth cues, indicating that unlearning has been successfully integrated into the volumetric representation. This highlights the generalizability and spatial coherence of our method beyond view-based supervision, ensuring that undesired features are removed holistically rather than superficially.



Figure 3: Qualitative 3D reconstruction results across different tasks after unlearning.

2.4 Effect of View-consistent Acceleration without Unlearning

To isolate the effect of acceleration from unlearning, we conduct an ablation study on the Zero123 baseline by applying our multi-view consistency-guided acceleration without any unlearning objective. Specifically, we introduce skip-step sampling with different reference view counts (3/4/8 views) to observe how generation quality changes purely due to acceleration.

Table 4 reports the Δ FID scores, computed as the difference between the FID of accelerated models and the baseline Zero-1-to-3 model. Positive values indicate improved fidelity relative to the baseline, while negative values indicate a degradation.

Table 4: Δ FID Comparison between Accelerated Models and Baseline (Zero-1-to-3).

Method	Steps Skipped	Delta FID
3 View	8	+0.9779
	12	-6.7059
	16	-15.8247
4 View	8	+0.3379
	12	-7.1140
	16	-34.8952
8 View	8	+2.8421
	12	-10.5640
	16	-74.9905

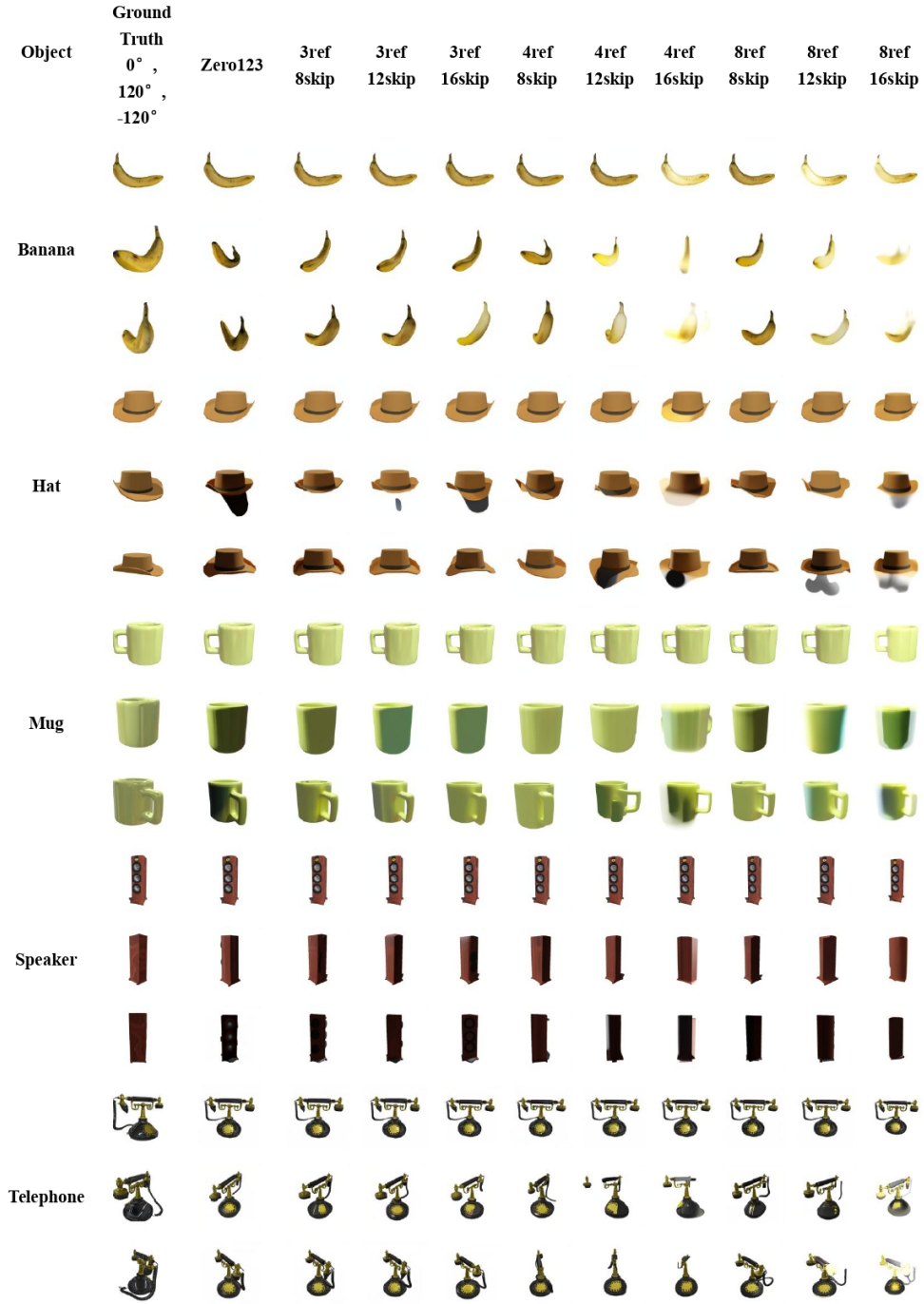


Figure 4: Qualitative results demonstrating the effect of view-consistent acceleration without un-learning. (Corresponds to Table 4 data)

192 We analyze the results from two perspectives: the number of reference (baseline) views used for
 193 interpolation, and the number of diffusion steps skipped during inference.

Effect of Reference View Number: When fixing the number of skipped steps, increasing the number of reference views generally leads to a more accurate initialization for the diffusion process due to finer angular coverage and closer interpolation points. This advantage is reflected in the 8-step skip setting, where the model with 8 reference views achieves the highest positive ΔFID (+2.8421), compared to 3 and 4 views (+0.9779 and +0.3379 respectively). This suggests that a denser set of baseline views provides a better starting point, facilitating high-fidelity multi-view synthesis.

Effect of Skipped Diffusion Steps: Across all reference view counts, increasing the number of skipped diffusion steps significantly degrades performance. For example, under 8 reference views, the Delta FID drops from +2.8421 at 8 skipped steps to -10.5640 at 12 skipped steps and further to -74.9905 at 16 skipped steps. This trend indicates that while skipping steps can accelerate inference, excessive step skipping undermines the model’s ability to refine the initial interpolated latent, leading to poorer image quality.

Interaction between Reference Views and Skipped Steps: Interestingly, the degradation caused by skipping more steps is more pronounced as the number of reference views increases. This is likely because the interpolation between two nearby reference views produces a finer but potentially more complex latent initialization that requires sufficient diffusion steps to properly refine. When too many steps are skipped, the model lacks the capacity to adequately recover details and enforce multi-view consistency, resulting in a sharper performance drop.

Qualitative results illustrating these effects are shown in Figure 4.

3 Limitation and Social Impact

While our method introduces a novel framework for machine unlearning in 3D generation, several limitations remain. We categorize these into technical limitations and broader societal concerns, and outline promising future directions to address them.

Technical Limitations.

- **Model Generalization.** Our framework is currently validated on Zero123 and Zero123XL. Its applicability to other 3D generation paradigms (e.g., NeRFs, mesh-based models, point-based representations) remains untested and may require architectural adaptations.
- **View Similarity Estimation.** The dynamic skipping mechanism leverages CLIP-based similarity to approximate view-level correspondence. While practical, this may be suboptimal for objects with subtle geometric or structural variations that CLIP embeddings cannot fully capture.
- **Manual Target Selection.** The forget/remain/retarget sets are manually specified. Real-world deployment would benefit from automatic identification of privacy-sensitive or biased content, requiring new detection or attribution tools.
- **Hyperparameter Sensitivity.** Our method depends on empirically chosen parameters, such as the angular threshold τ and skip-step schedule. These may require retuning on new datasets or under different acceleration regimes.
- **Lack of Robustness Evaluation.** We do not assess the robustness of the unlearned model against adversarial attacks such as model inversion, concept re-injection, or prompt-based data recovery.

Future Work.

- **Broader Model Applicability.** We aim to adapt our framework to a wider range of 3D generation backbones, including volumetric NeRFs, implicit surfaces, and real-time rendering architectures.
- **Privacy-Aware Target Detection.** Future work will explore integrating privacy or attribution detectors to automatically identify sensitive content for targeted unlearning without human intervention.
- **Unlearning Without Retargeting.** While our method currently aligns forgotten content with a retargeted distribution, we plan to investigate pure erasure techniques without replacement, suitable for content removal rather than transformation.

244 • **Online and Continual Unlearning.** Extending our method to dynamic settings—such as
 245 continual learning or post-deployment unlearning requests—is an important direction for
 246 practical applications.

247 • **Trustworthy Unlearning Evaluation.** We plan to develop formal verification protocols
 248 and benchmarks to quantify the effectiveness and irreversibility of unlearning across di-
 249 verse tasks and threat models.

250 **Social Impact Considerations.** Our framework raises potential concerns regarding privacy leak-
 251 age and model bias, especially in the context of modular or pre-trained model reuse.

252 • **Privacy Risk.** By reusing pretrained parameters, the unlearned model may unintentionally
 253 retain latent traces of upstream data. Adversaries could potentially reconstruct sensitive
 254 content through model inversion or prompt tuning. One mitigation strategy is to increase
 255 the diversity and number of pretrained models used, ensuring that no single model contains
 256 sufficient information to recover sensitive content.

257 • **Model Bias.** Biases present in the original training data or source models may propagate
 258 through the unlearning process. To mitigate this, we propose diversifying the source model
 259 pool and introducing diversity-promoting regularization during training. This helps prevent
 260 over-reliance on any single biased component and encourages fairer predictions.

261 We consider these directions essential for improving the robustness, fairness, and ethical deploy-
 262 ment of 3D unlearning systems, and plan to extend our study to address these limitations in future
 263 iterations of this research.