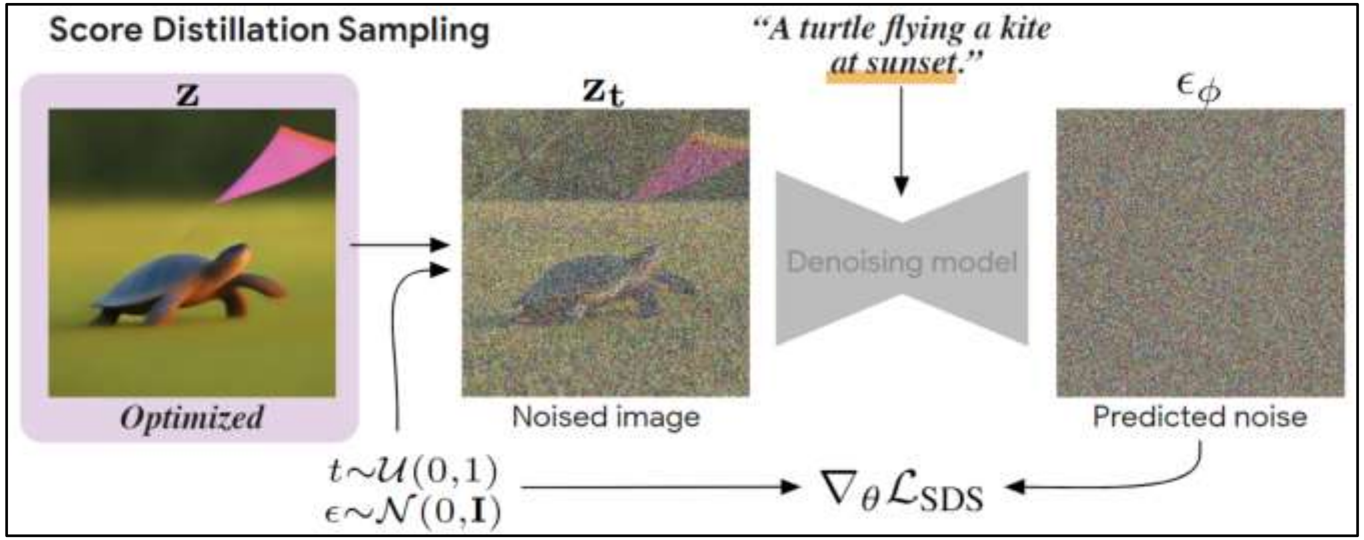
DreamTime: An Improved Optimization Strategy for Diffusion-Guided 3D Generation

Yukun Huang[1,2], Jianan Wang[1], Yukai Shi[1], Boshi Tang[1], Xianbiao Qi[1], Lei Zhang[1]

[1] International Digital Economy Academy (IDEA)
[2] The University of Hong Kong (HKU)

Diffusion-Guided 3D Generation

With score distillation sampling (SDS) techniques, we could use pre-trained 2D diffusion model for 3D asset generation.



Text-to-3D [1]

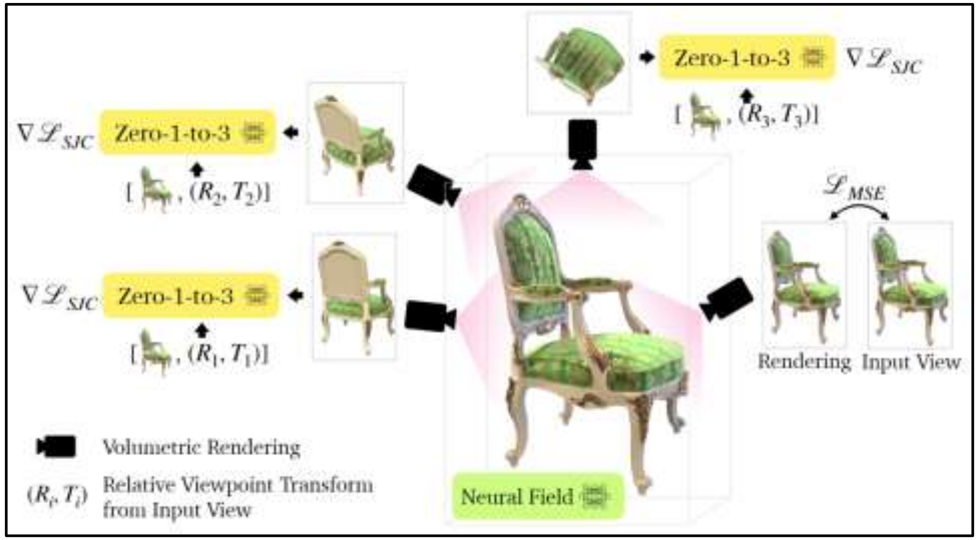


Image-to-3D [2]

[1] "DreamFusion: Text-to-3D using 2D Diffusion." ICLR 2023.
[2] "Zero-1-to-3: Zero-shot One Image to 3D Object." ICCV 2023.

Drawbacks of SDS for 3D Generation

Random t-sampling in SDS: $t \sim U(1, T)$

Slow Convergence

- DreamFusion: 6 TPU hours
- Magic3D: 5.3 A100 hours
- Fantasia3D: 6 RTX3090 hours
- ProlificDreamer: several A100 hours

Misaligned Supervision

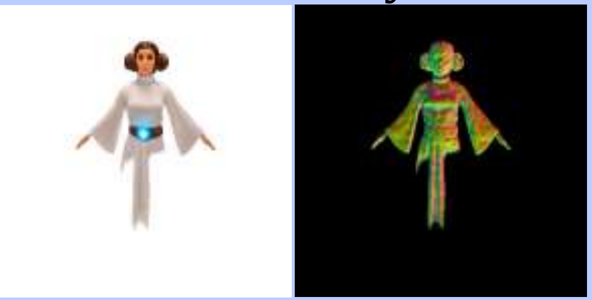
conflicts with coarse-to-fine generation nature

Out of Distribution

cannot handle low-frequency bias of early 3D renderings

Quality Concerns

Geometry



incomplete geometry

Texture

blurriness — color distortion

Semantics

"a peacock on a surfboard" — *"a chimpanzee dressed like Henry VIII king of England"*

Mode Collapse

Observation 1: Mathematical Formulation

We contrast SDS loss:

$$\mathcal{L}_{\mathrm{SDS}}(\phi, \mathbf{x}_t) = \mathbb{E}_{t \sim \mathcal{U}(1,T)} \left[w(t) \| \boldsymbol{\epsilon}_\phi(\mathbf{x}_t; y, t) - \boldsymbol{\epsilon} \|_2^2 \right]$$

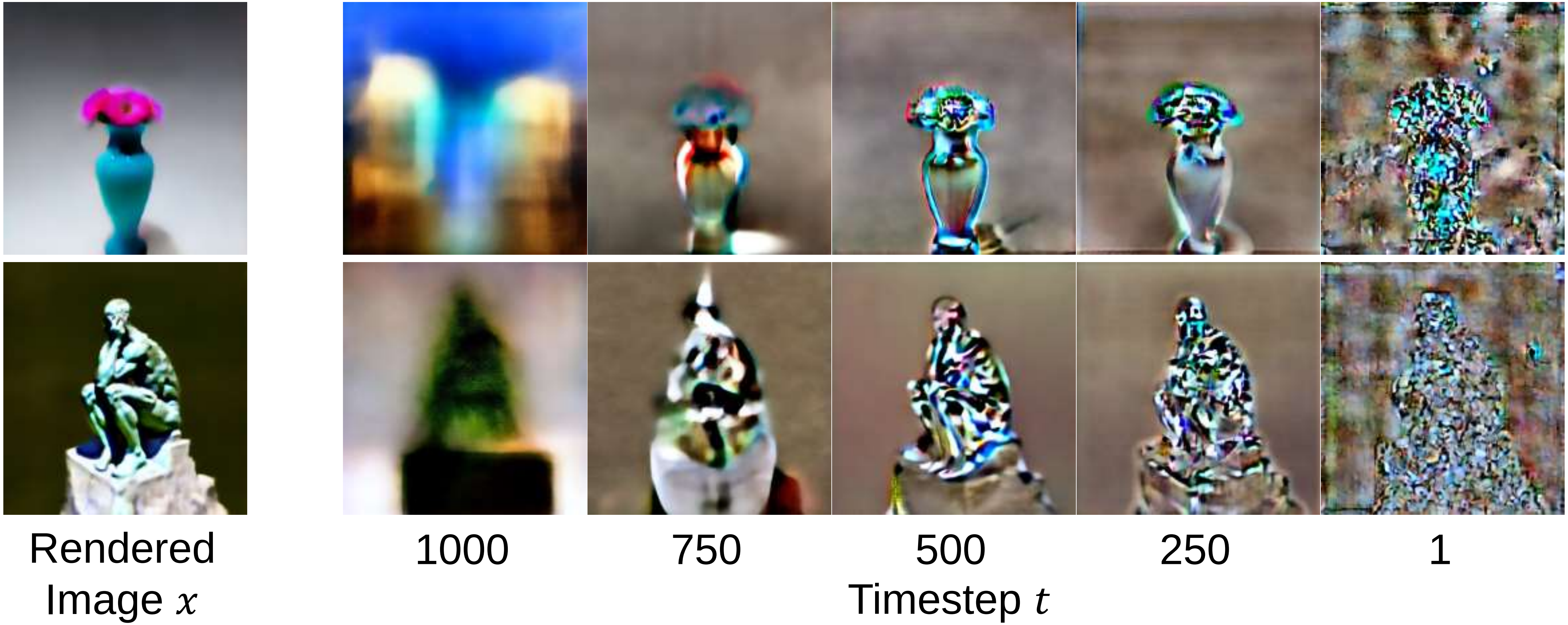
with DDPM sampling process, i.e., for $t = T \to 1$:

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_\phi(\mathbf{x}_t; y, t) \right) + \sigma_t \boldsymbol{\epsilon},$$

The randomly uniform t-sampling in SDS for 3D Generation is unaligned with the non-increasing t-sampling in DDPM for 2D Generation.

Observation 2: Supervision Misalignment

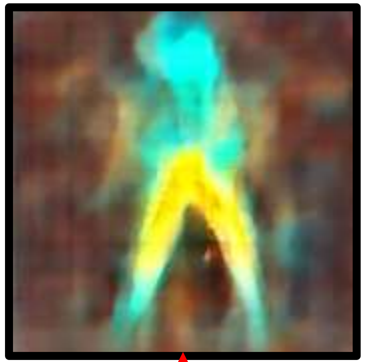
For diffusion models, score prediction provides different granularity of supervision at different timestep t: from coarse structure to fine details as t decreases.



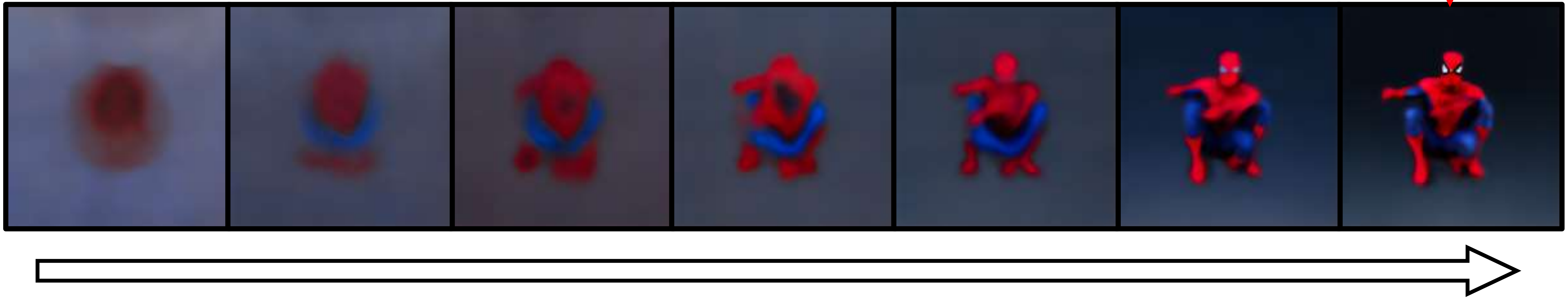
Observation 2: Supervision Misalignment

As SDS optimization proceeds, the trained 3D representation (e.g., NeRF) presents a coarse-to-fine process, in which different stages prefer different granularity of supervision.

However, randomly uniform timestep sampling in the vanilla SDS makes such requirements difficult to guarantee.



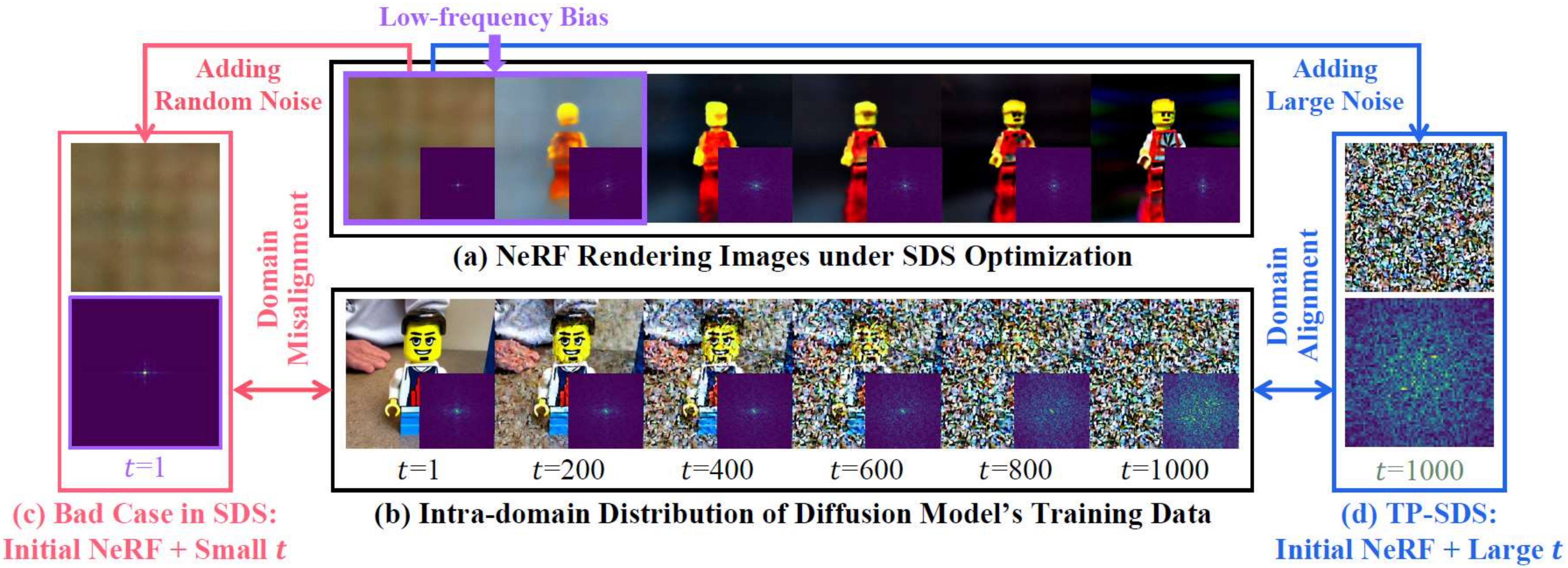
Misalignment !



3D Rendered Images with SDS Optimization in Progress

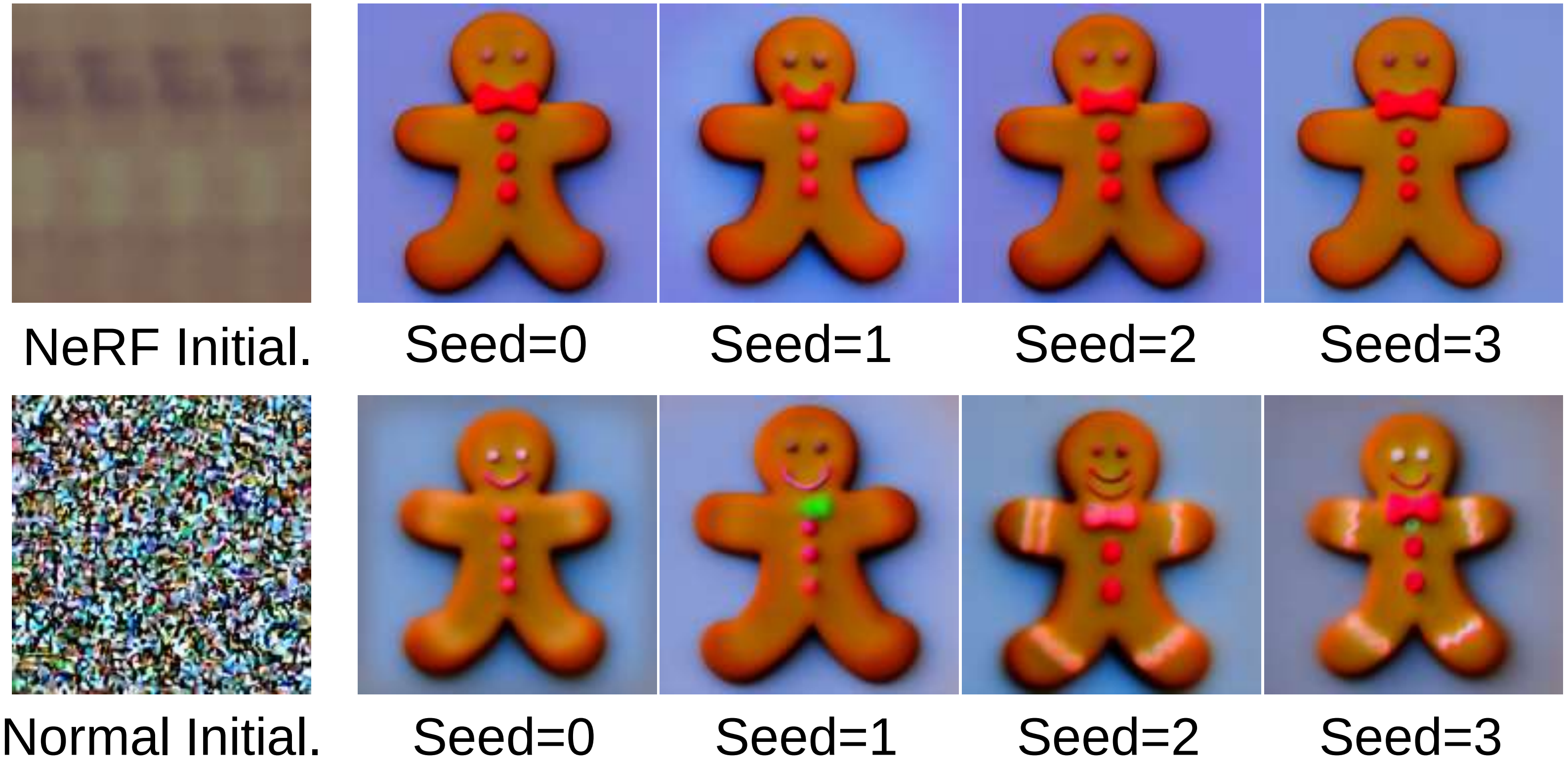
Observation 3: Out-of-Distribution Inputs

The OOD issue is significant when using rendered images from the early training stage (low-frequency bias) as diffusion inputs and timestep t is small.



Observation 3: Out-of-Distribution Inputs

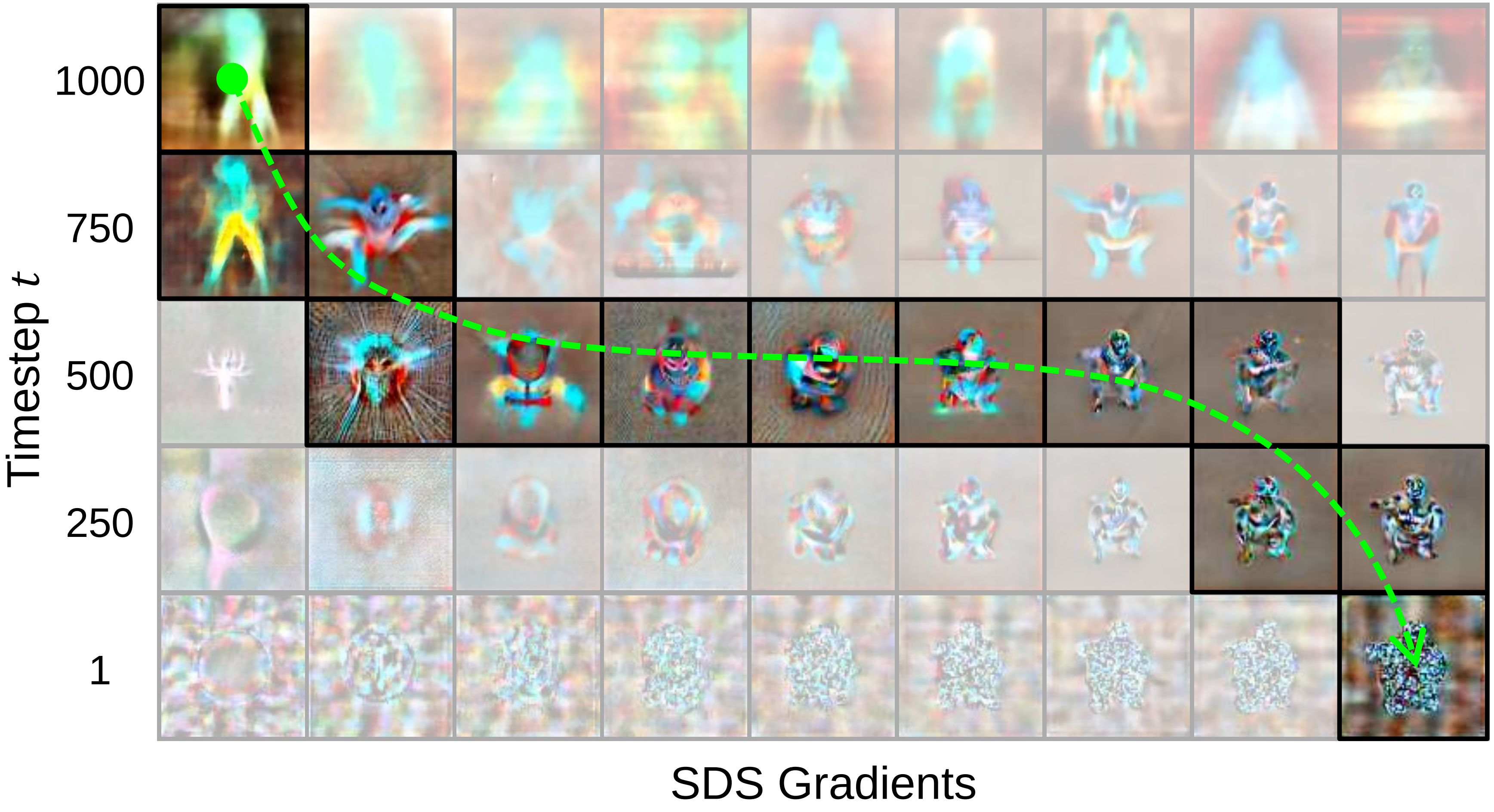
We provide an 2D generated example to demonstrate that low-frequency bias in the initial input image (common in NeRF) could lead to low-diversity generation.



Text Prompt: “gingerbread man”

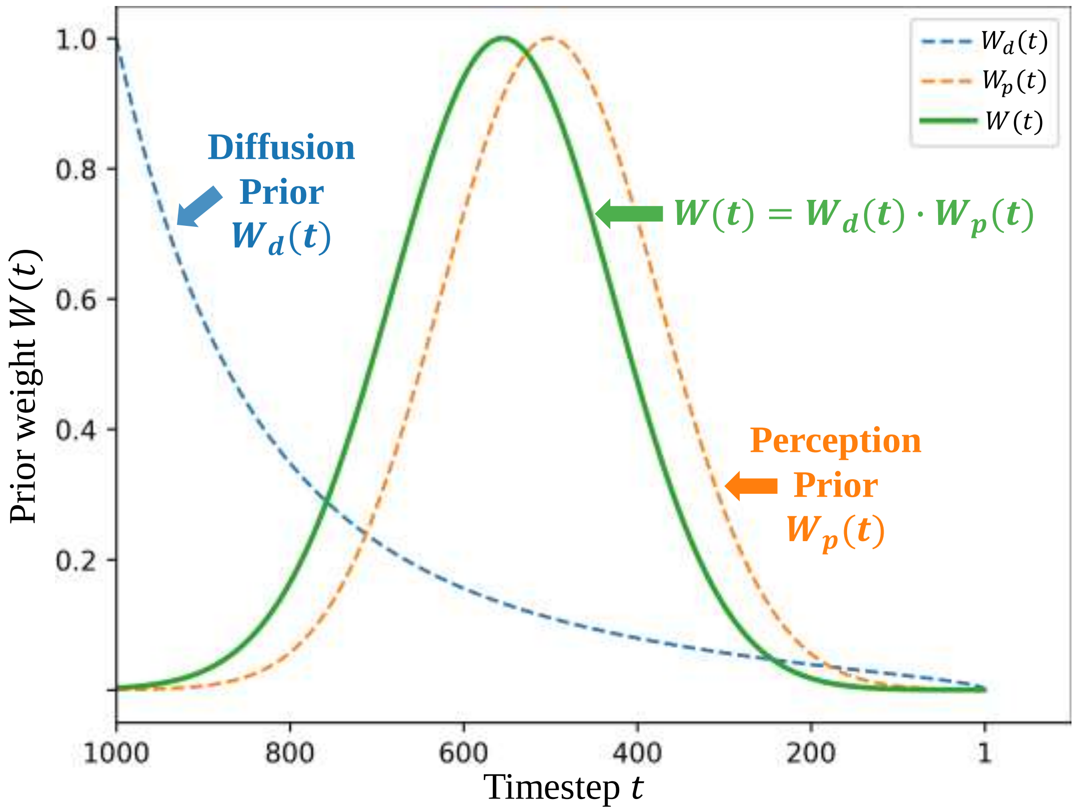
Method: Time Prioritized Score Distillation

We argue that non-increasing t-sampling (indicated by - - -➝) is more effective for diffusion-guided 3D optimization compared to randomly uniform t-sampling.



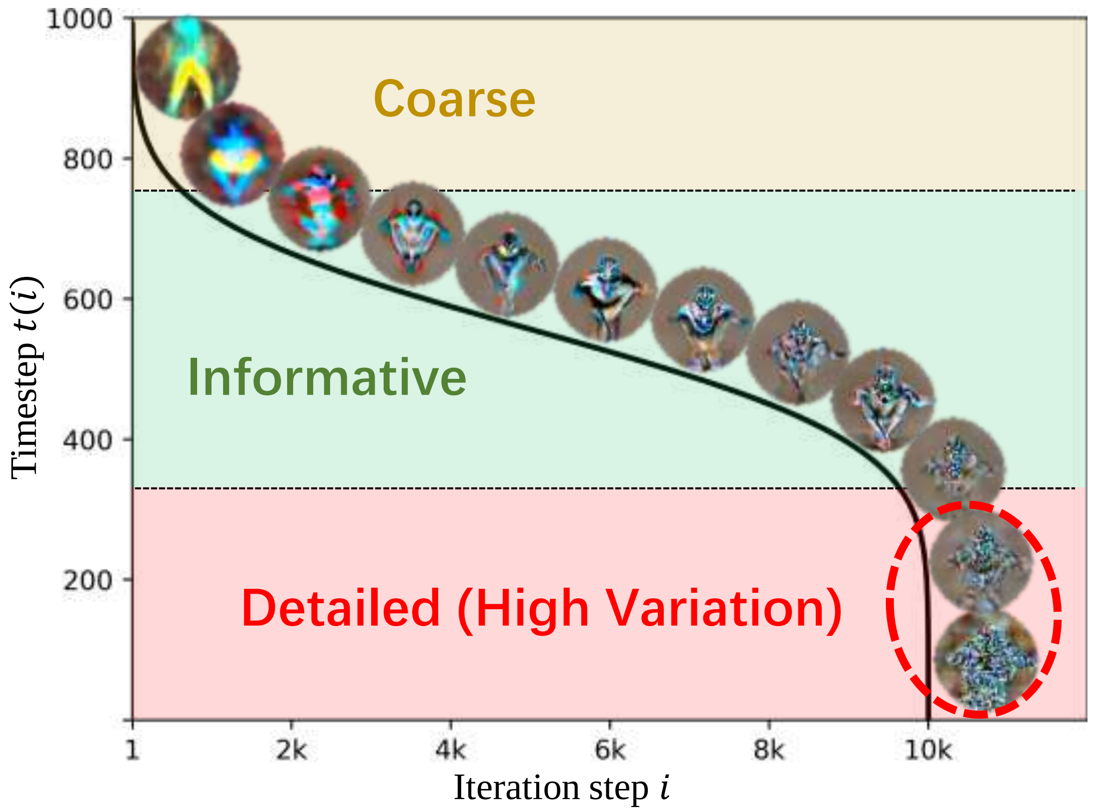
Method: Time Prioritized Score Distillation

Based on the characteristics of diffusion training and 3D generation, we carefully design a **weight function** $W(t)$ to modulate the timestep descent process.



(a) Prior weight function $W(t)$.

modulate



(b) Weighted non-increasing t-sampling function $t(i)$.

Results: Faster Convergence

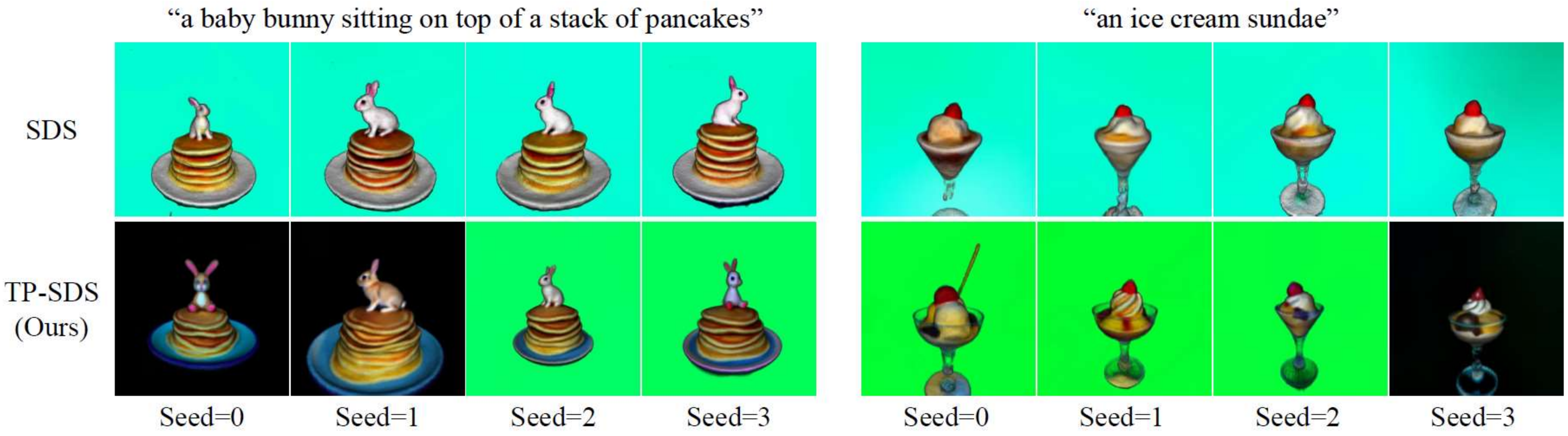
The proposed Time Prioritized Score Distillation Sampling (TP-SDS) leads to faster 3D content generation than the SDS baseline.

Results: Better Quality

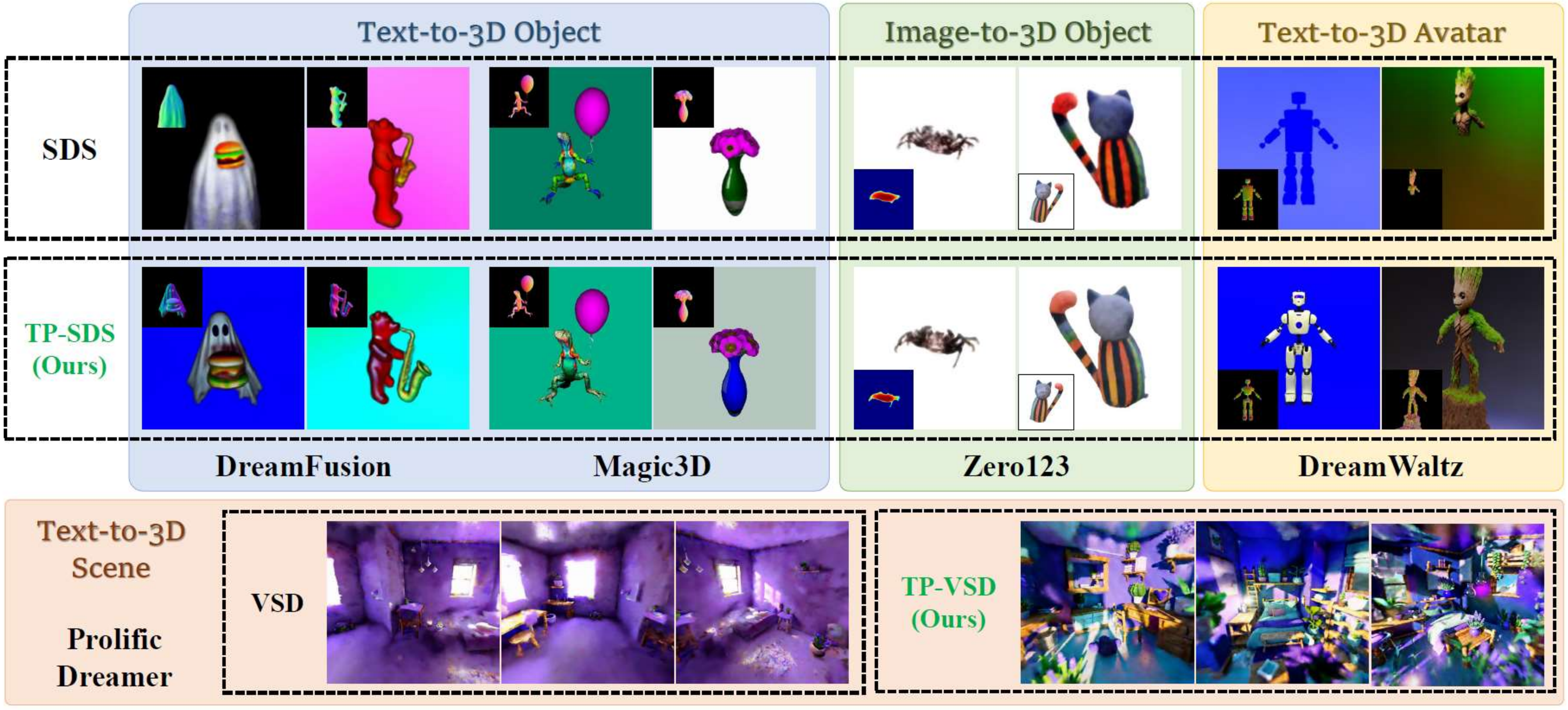
Our method can alleviate some common quality problems in SDS optimization, such as attribute missing, unsatisfactory geometry, and compromised details, as highlighted by the colored circles.

Results: Higher Diversity

Given different random seeds, our TP-SDS is able to generate visually distinct 3D objects, while the results produced by SDS baseline all look alike.



Results: Versatility



Thank you!

Please feel free to contact us if you have any questions:

yukun@hku.hk