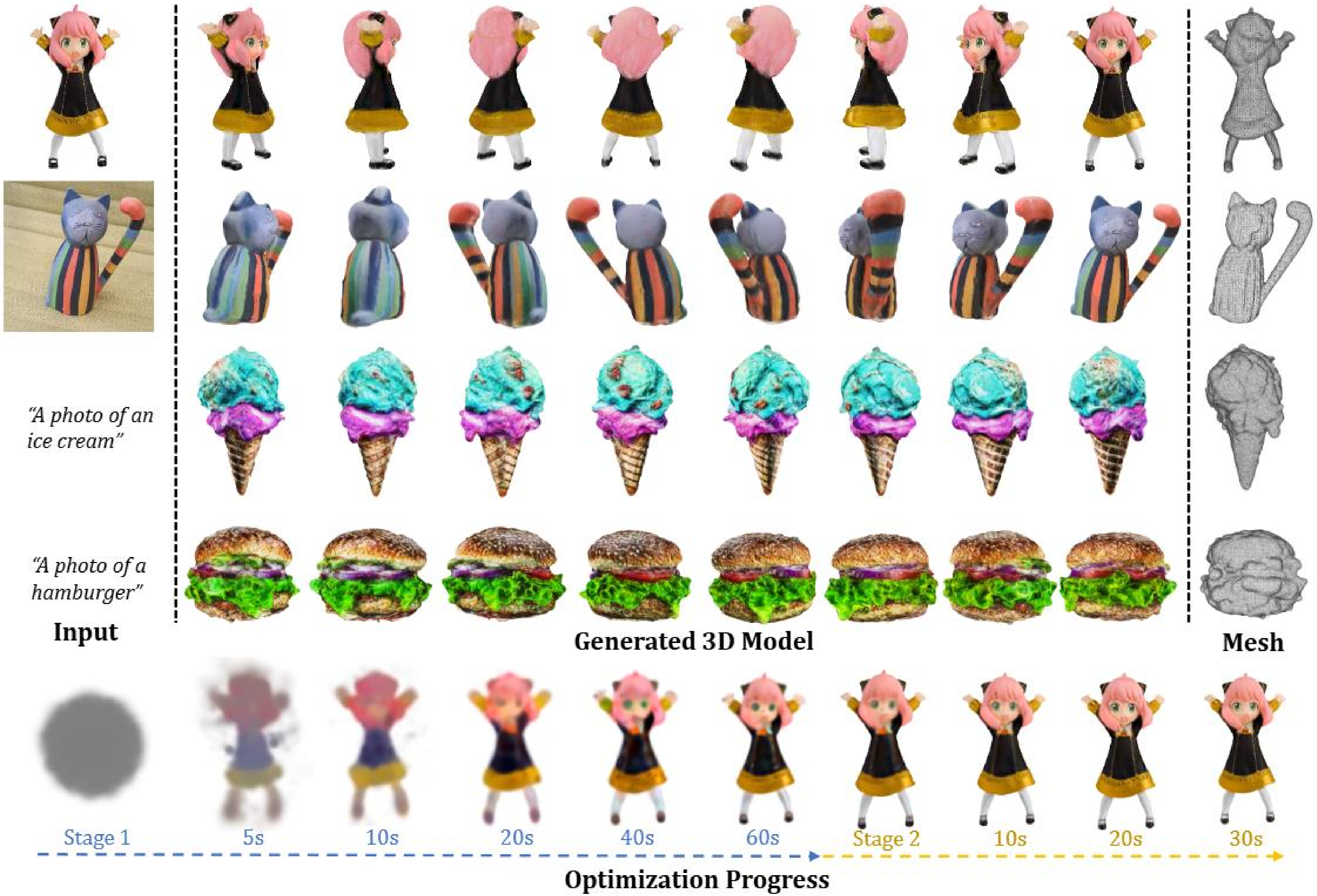
DreamGaussian: Generative Gaussian Splatting for Efficient 3D Content Creation

ICLR 2024 (Oral)

Jiaxiang Tang[1], Jiawei Ren[2], Hang Zhou[3], Ziwei Liu[2], Gang Zeng[1]

[1] Peking University [2] S-Lab, Nanyang Technological University [3] Baidu

整体目的


"A photo of an ice cream"
"A photo of a hamburger"
Input
Generated 3D Model
Mesh
Stage 1
5s
10s
20s
40s
60s
Stage 2
10s
20s
30s
Optimization Progress

Gaussian splatting

$$G(x) = exp\left(-\frac{1}{2}x^T \sum^{-1} x\right)$$

center position $\mu \in \mathbb{R}^3$

opacity $\alpha \in \mathbb{R}$

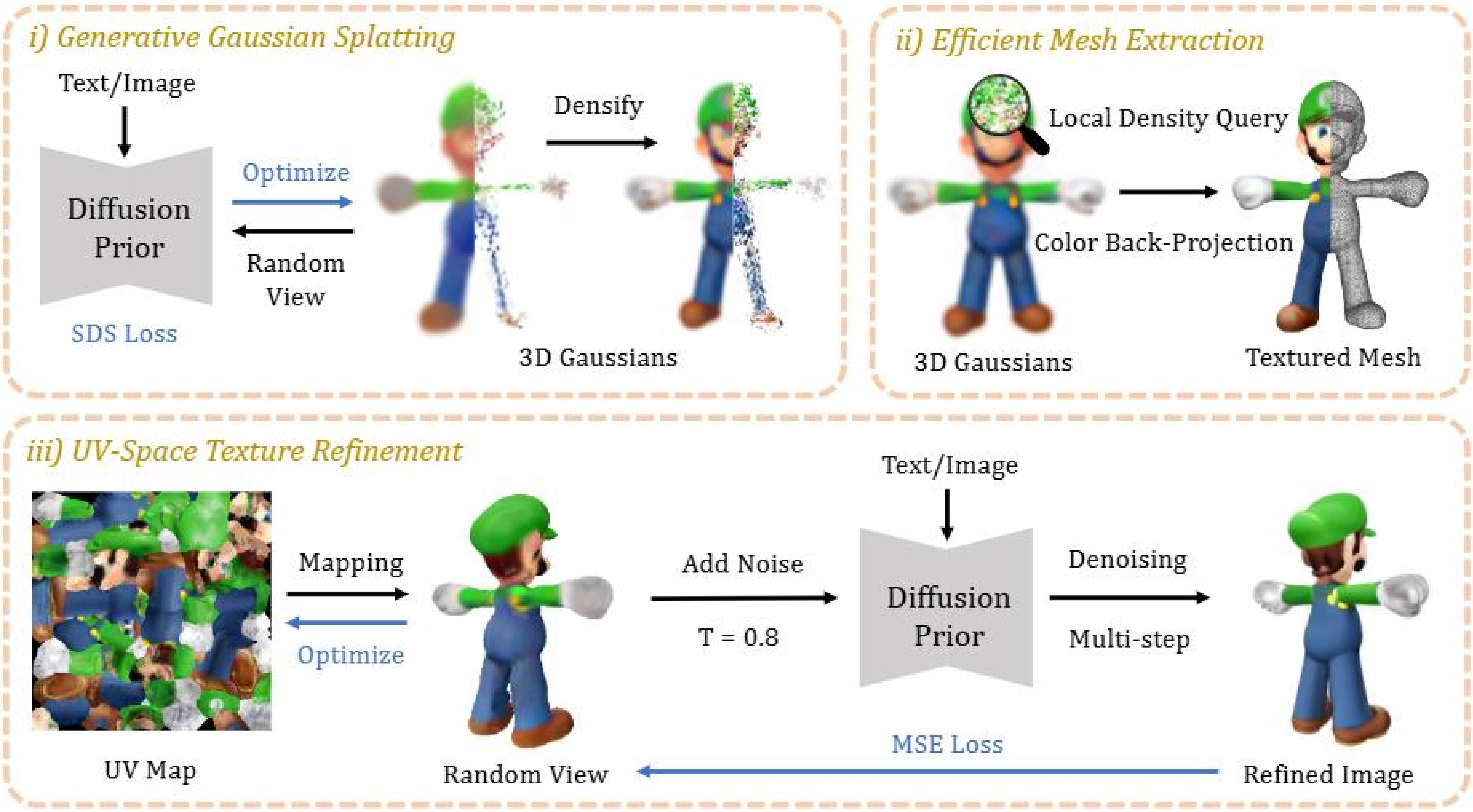
color $c \in \mathbb{R}^3$

scale $S \in \mathbb{R}^3$

rotation $R \in \mathbb{R}^{3\times 3}$

covariance matrix $\sum = RSS^T R^T$

框架



i) Generative Gaussian Splatting
Text/Image
Diffusion Prior
SDS Loss
Optimize
Random View
Densify
3D Gaussians
ii) Efficient Mesh Extraction
Local Density Query
Color Back-Projection
3D Gaussians
Textured Mesh
iii) UV-Space Texture Refinement
UV Map
Mapping
Optimize
Random View
Add Noise
T = 0.8
Text/Image
Diffusion Prior
Denoising
Multi-step
Refined Image
MSE Loss

Diffusion prior

Zero-1-to-3: Zero-shot One Image to 3D Object

Ruoshi Liu
Columbia University

Rundi Wu
Columbia University

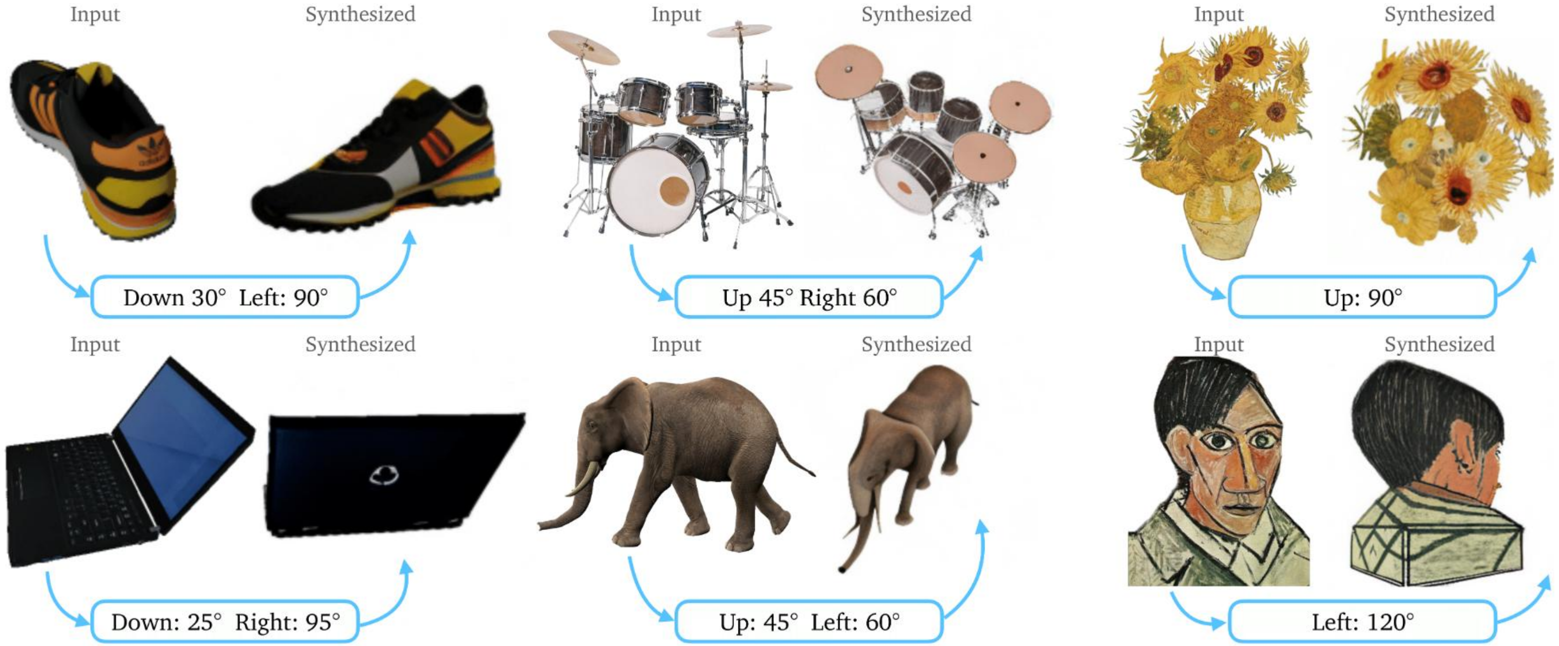
Basile Van Hoorick
Columbia University

Pavel Tokmakov
Toyota Research Institute

Sergey Zakharov
Toyota Research Institute

Carl Vondrick
Columbia University

TL;DR: We learn to control the camera perspective in large-scale diffusion models, enabling zero-shot novel view synthesis and 3D reconstruction from a single image.



$$\hat{x}_{R,T} = f(x, R, T)$$

$$\min_{\theta} \mathbb{E}_{z\sim\mathcal{E}(x),t,\epsilon\sim\mathcal{N}(0,1)}||\epsilon - \epsilon_\theta(z_t, t, c(x, R, T))||_2^2$$

SDS损失函数

Image-3D

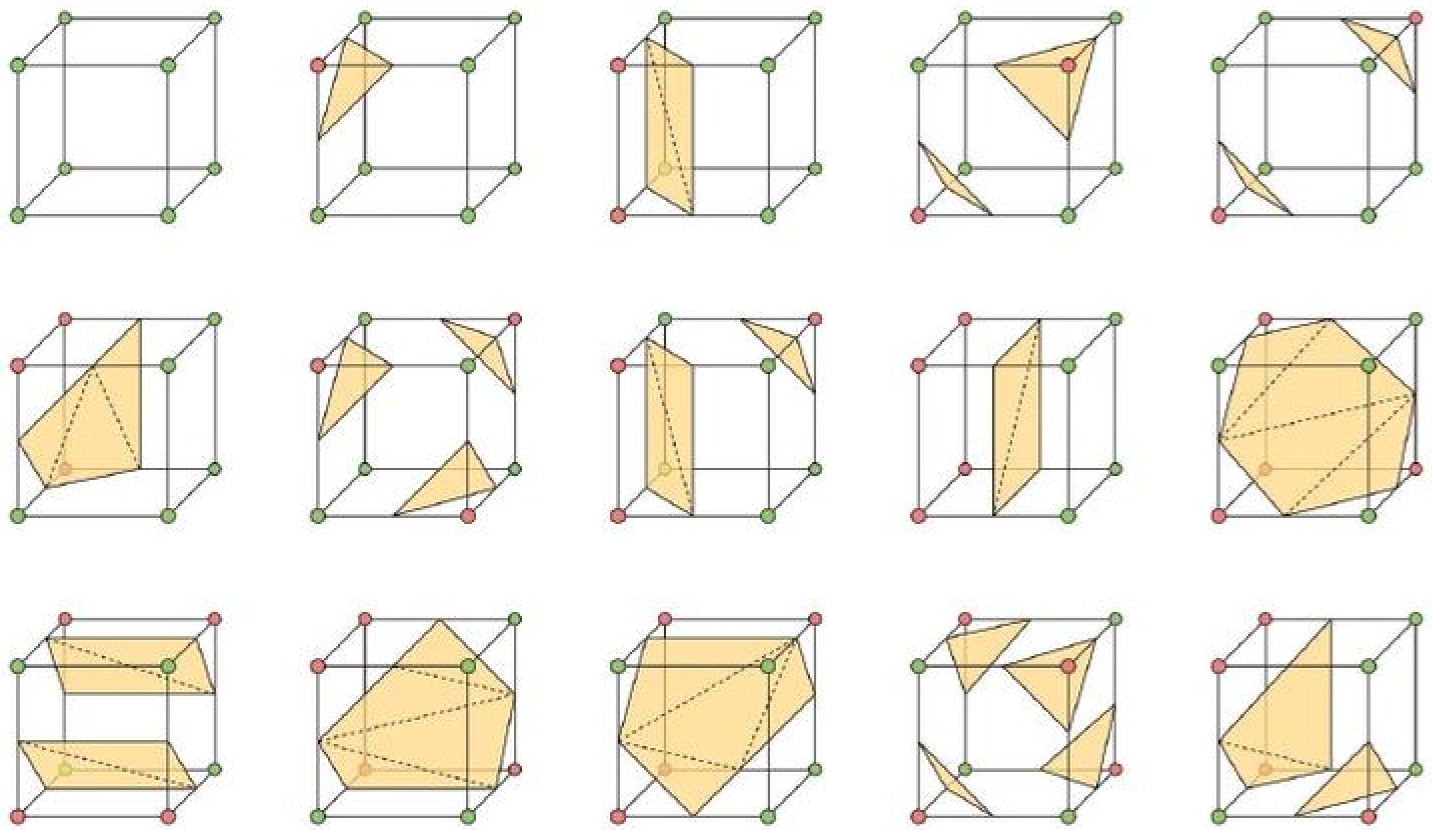
$$\nabla_{\Theta}\mathcal{L}_{\text{SDS}} = \mathbb{E}_{t,p,\epsilon}\left[(\epsilon_{\phi}(I^{p}_{\text{RGB}}; t, \tilde{I}^{r}_{\text{RGB}}, \Delta p) - \epsilon)\frac{\partial I^{p}_{\text{RGB}}}{\partial\Theta}\right]$$

$$\mathcal{L}_{\text{Ref}} = \lambda_{\text{RGB}}||I^{r}_{\text{RGB}} - \tilde{I}^{r}_{\text{RGB}}||^{2}_{2} + \lambda_{\text{A}}||I^{r}_{\text{A}} - \tilde{I}^{r}_{\text{A}}||^{2}_{2}$$

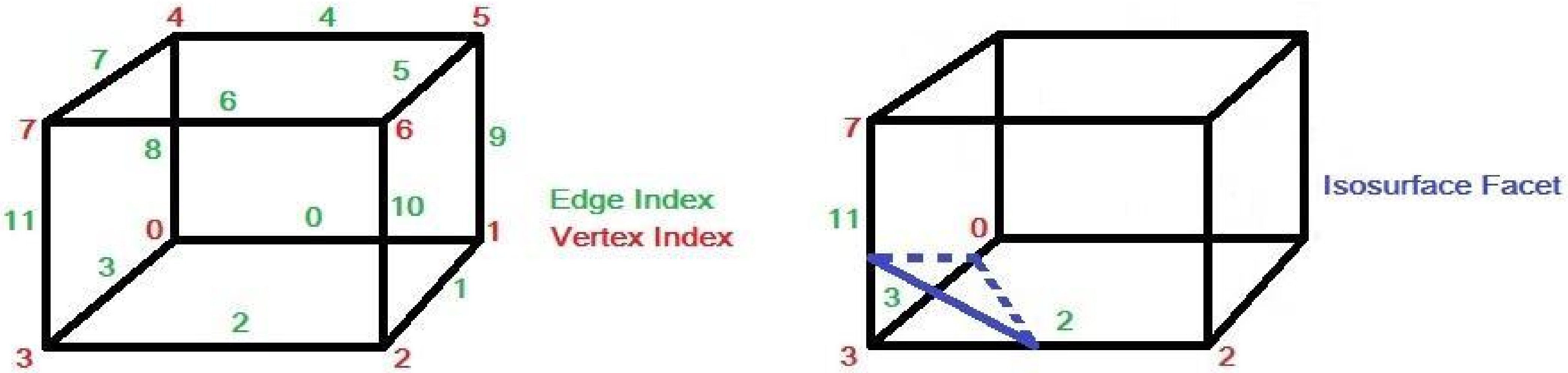
Text-3D

$$\nabla_{\Theta}\mathcal{L}_{\text{SDS}} = \mathbb{E}_{t,p,\epsilon}\left[(\epsilon_{\phi}(I^{p}_{\text{RGB}}; t, e) - \epsilon)\frac{\partial I^{p}_{\text{RGB}}}{\partial\Theta}\right]$$

提取mesh面

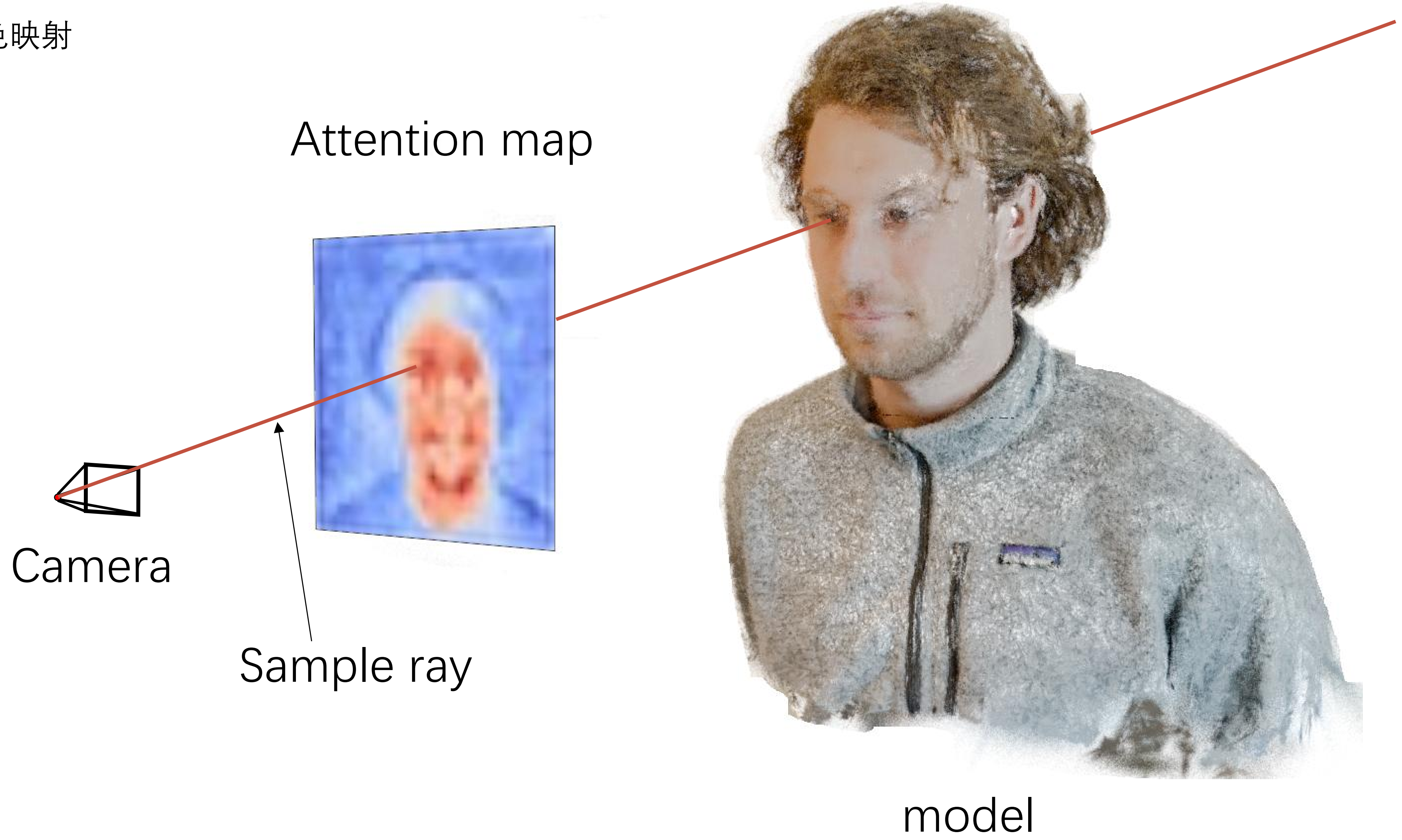


Grid构建策略

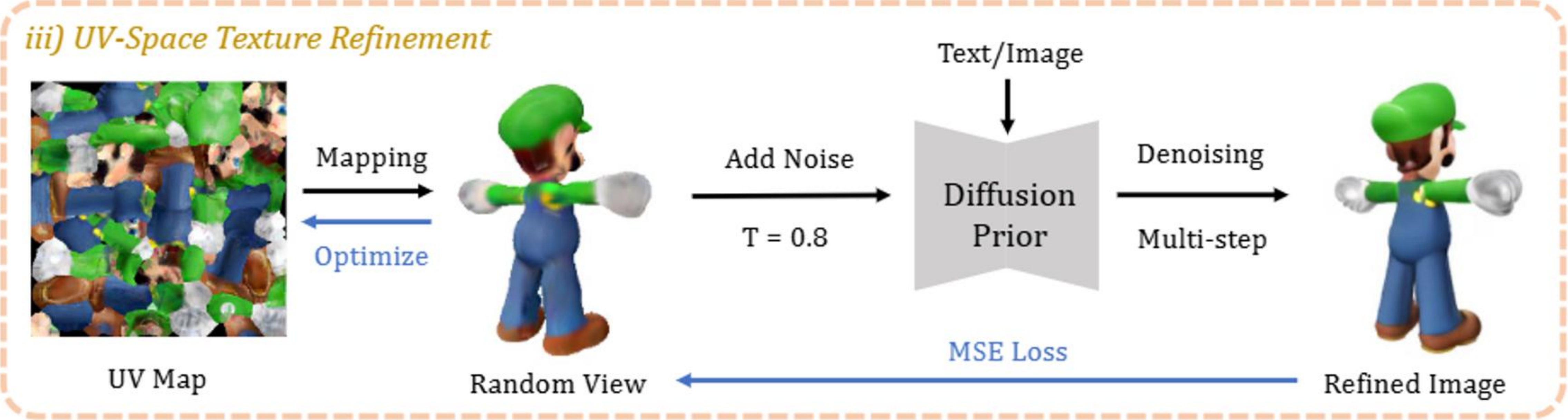


$$d(\mathbf{x}) = \sum_i \alpha_i \exp(-\frac{1}{2}(\mathbf{x}-\mathbf{x_i})^T \Sigma_i^{-1} (\mathbf{x}-\mathbf{x_i}))$$

颜色映射



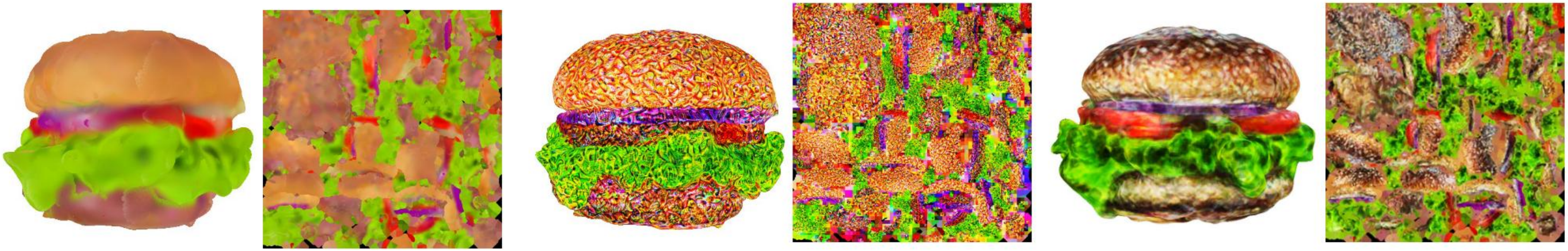
纹理优化



$$I^p_{\text{fine}} = f_\phi(I^p_{\text{coarse}} + \epsilon(t_{\text{start}}); t_{\text{start}}, c)$$

$$\mathcal{L}_{\text{MSE}} = ||I^p_{\text{fine}} - I^p_{\text{coarse}}||^2_2$$

纹理优化结果

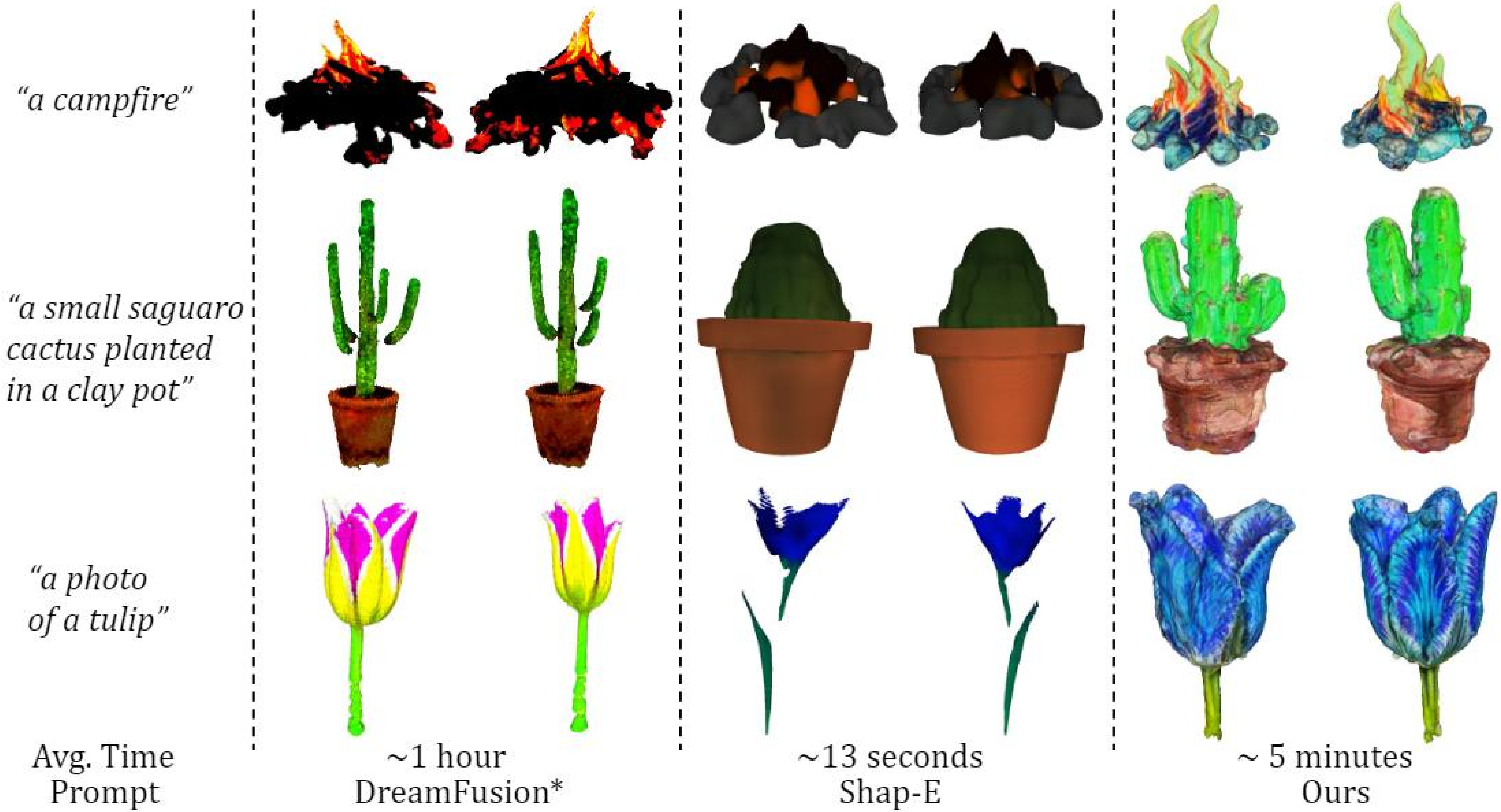


Stage 1　　Stage 2 (SDS)　　Stage 2 (MSE)

图片引导定性对比


Avg. Time
Input
~20 minutes
Zero-1-to-3
~45 seconds
One-2-3-45
~27 seconds
Shap-E
~2 minutes
Ours

文本引导定性对比


"a campfire"
"a small saguaro cactus planted in a clay pot"
"a photo of a tulip"
Avg. Time
Prompt
~1 hour
DreamFusion*
~13 seconds
Shap-E
~ 5 minutes
Ours

定量分析对比

	Type	CLIP-Similarity ↑	Generation Time ↓
One-2-3-45 (Liu et al., 2023a)	Inference-only	0.594	45 seconds
Point-E (Nichol et al., 2022)	Inference-only	0.587	78 seconds
Shap-E (Jun & Nichol, 2023)	Inference-only	0.591	27 seconds
Zero-1-to-3 (Liu et al., 2023b)	Optimization-based	0.647	20 minutes
Zero-1-to-3* (Liu et al., 2023b)	Optimization-based	0.778	30 minutes
Ours (Stage 1 Only)	Optimization-based	0.678	1 minute
Ours	Optimization-based	0.738	2 minutes

谢谢!