上海人工智能实验室
Shanghai Artificial Intelligence Laboratory

MATLABER: Material-Aware Text-to-3D via LAtent BRDF auto-EncodeR

Xudong Xu

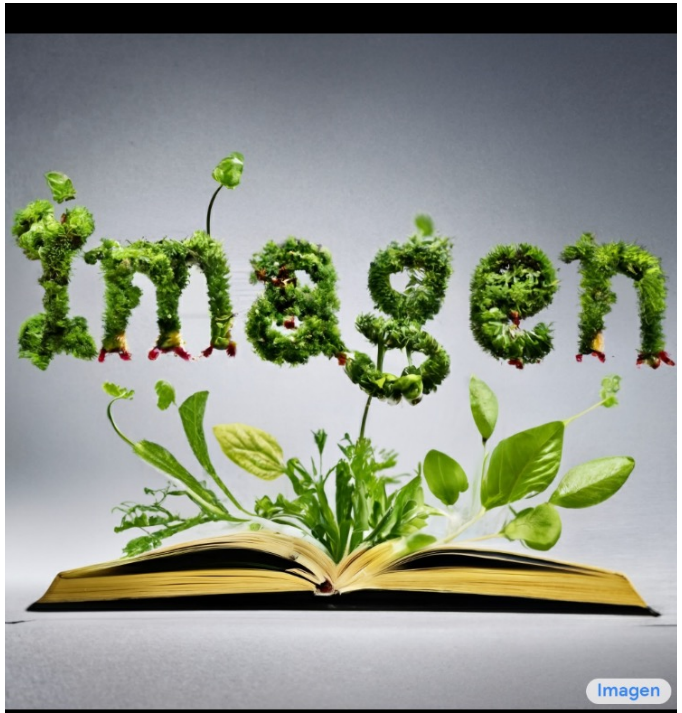
Shanghai AI Lab

上海人工智能实验室
Shanghai Artificial Intelligence Laboratory

MATLABER

Background: Text-to-image synthesis

- Thanks to powerful diffusion model and massive text-image pair data, we have witnessed great progress in text-to-image synthesis.

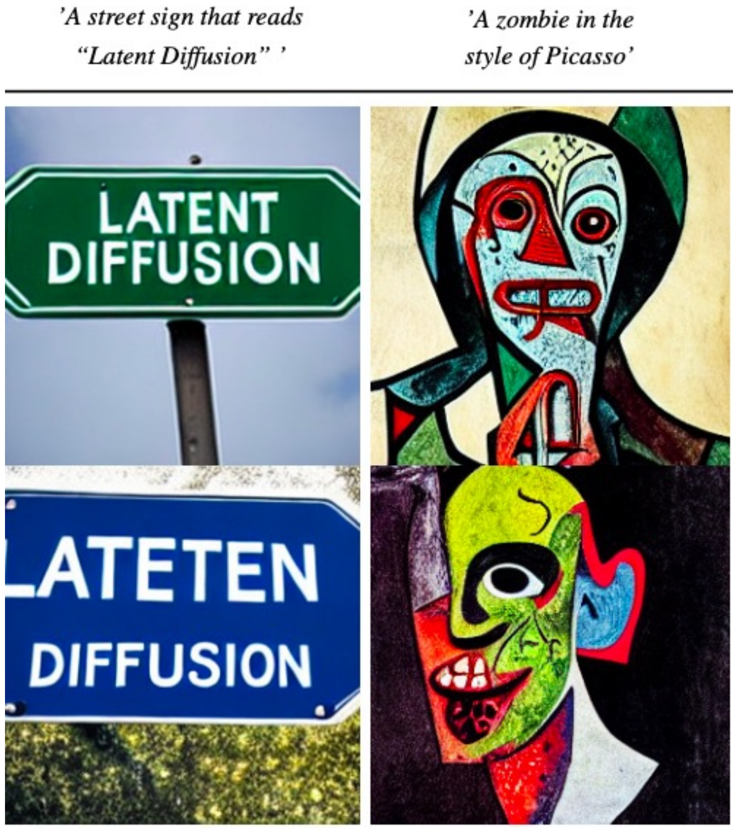


Sprouts in the shape of text 'Imagen' coming out of a fairytale book.

Imagen

"an illustration of albert einstein wearing a superhero costume"

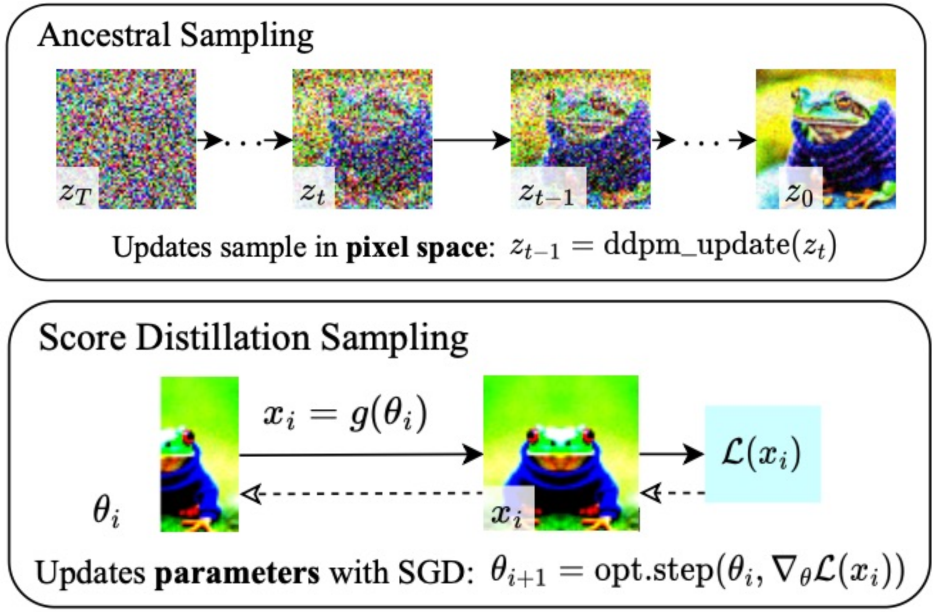
Glide



Stable Diffusion

Background: Text-to-3D generation

- Text-to-3D generation: AIGC exploration from 2D to 3D domain.
- Relying on promising Score Distillation Sampling (**SDS**), DreamFusion successfully takes the first step in text-to-3D generation.

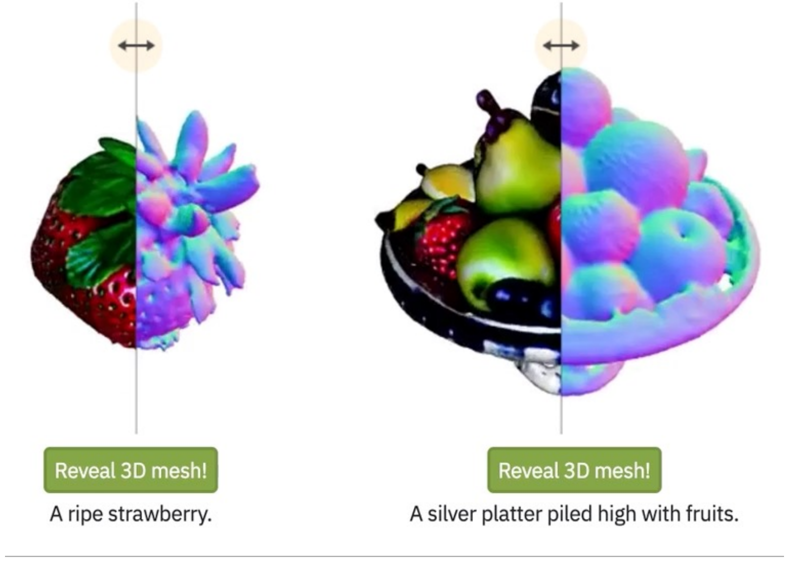


$$\nabla_\theta \mathcal{L}_{\text{SDS}}(\phi, \mathbf{x} = g(\theta)) \triangleq \mathbb{E}_{t,\epsilon}\left[w(t)\left(\hat{\epsilon}_\phi(\mathbf{z}_t; y, t) - \epsilon\right)\frac{\partial \mathbf{x}}{\partial \theta}\right]$$

Background: Text-to-3D generation

Representative follow-ups:

- Magic3D: increase resolution from 64 to 512
- Fantasia3D: better geometry and realistic appearance
- ProlificDreamer: high-fidelity and **diverse** generation



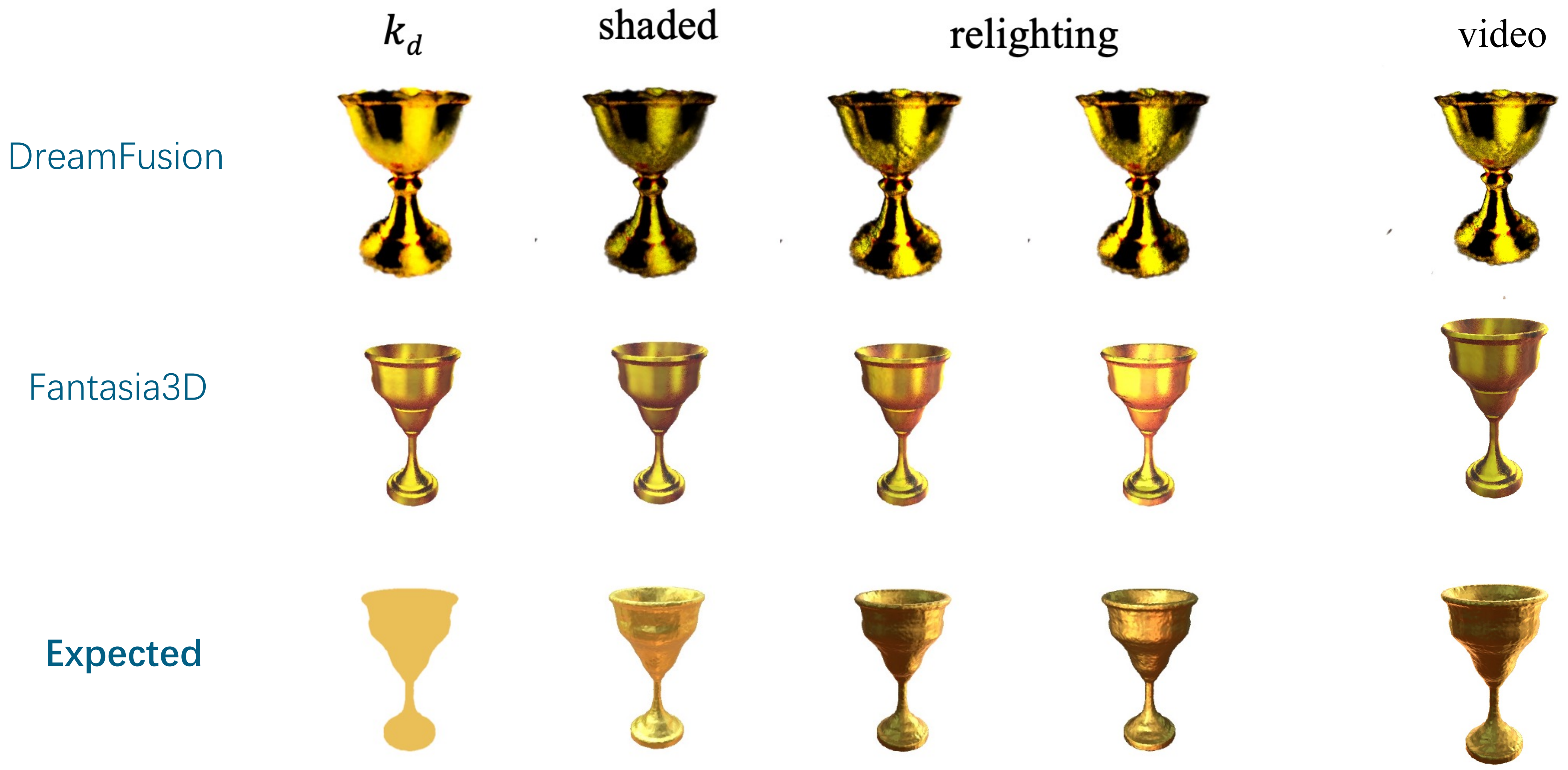
Magic3D

Fantasia3D

ProlificDreamer

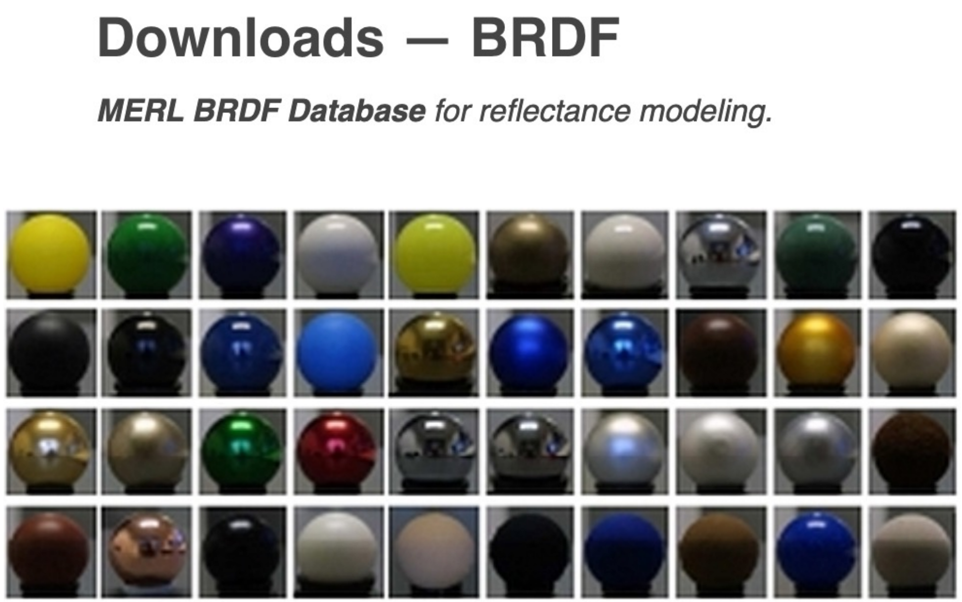
Background: No one cares materials

- DreamFusion: only consider Lambertian reflectance
- Fantasia3D: BRDF materials entangled with environment lights

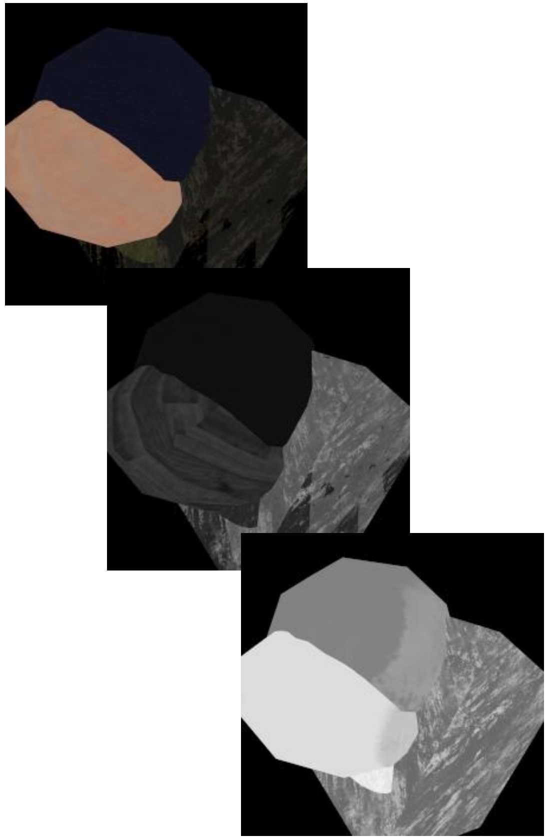


Background: Text-material data?

- Unfortunately, there does not exist text-material paired dataset.
- However, there are several BRDF material datasets.



Adobe Substance 3D Assets



two-shot-brdf

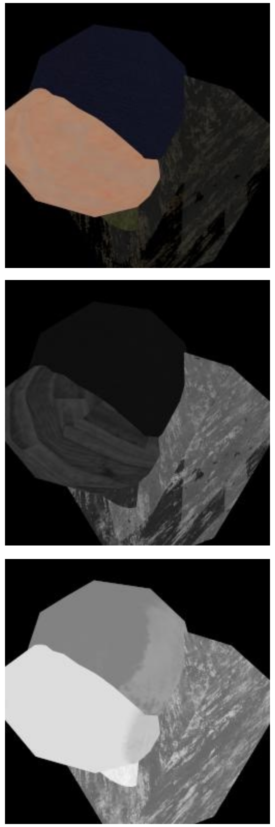
Method: BRDF auto-encoder

- We train a latent BRDF auto-encoder, **which acts as a material prior**.
- Smoothness and KL losses are imposed to latent codes for a smooth latent space.

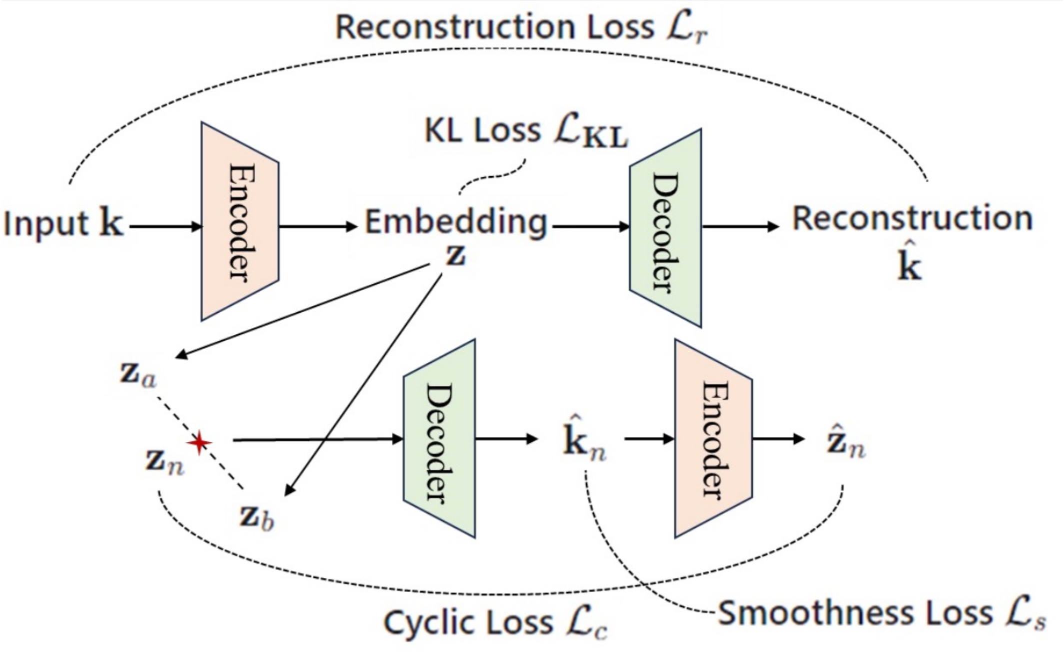
diffuse

specular

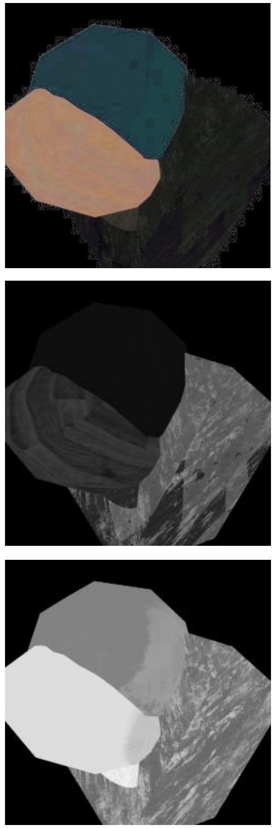
roughness



7-dim BRDF



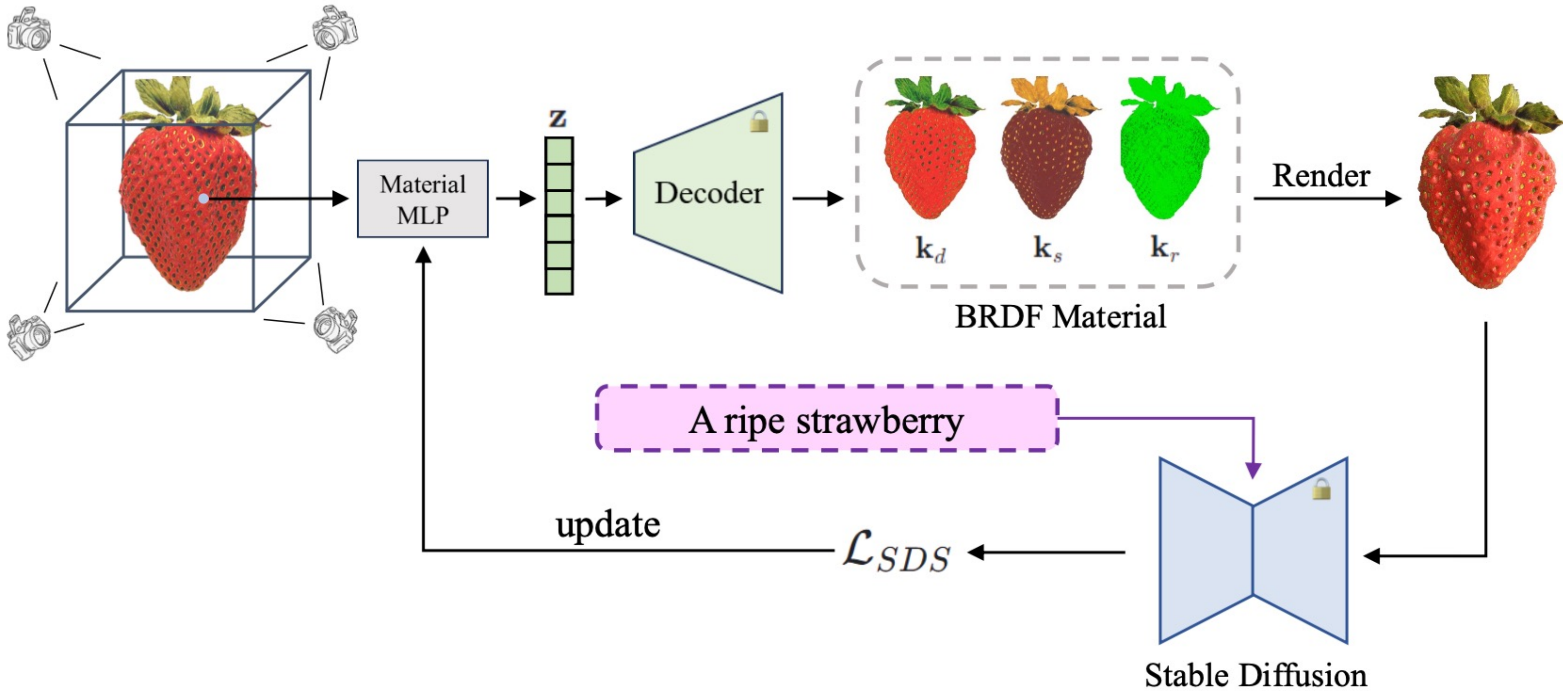
$$\mathcal{L} = \mathcal{L}_r + \lambda_{\mathrm{KL}}\mathcal{L}_{\mathrm{KL}} + \lambda_s\mathcal{L}_s + \lambda_c\mathcal{L}_c$$



Reconstructed BRDF

Method: Text-to-3D generation pipeline

- For geometry modeling, we follow the method proposed in Fantasia3D.
- Material MLP predicts the BRDF latent code z, rather than the BRDF material.
- The obtained latent code is then decoded to 7-dim BRDF via our decoder.
- SDS loss can be applied to rendered images, thus enabling the network training.



Method: Rendering equations

- For a surface point x, we first apply positional encoding and then leverage a material MLP to predict BRDF latent code z, which is then transferred to BRDF k.

$$\mathbf{z_x} = \Gamma(\beta(\mathbf{x}); \gamma), \quad \mathbf{k_x} = \mathcal{D}(\mathbf{z_x}).$$

- Similar to prior works, we also leverage the split-sum method for rendering.

$$L(\mathbf{x}, \boldsymbol{\omega}_o) = L_d(\mathbf{x}) + L_s(\mathbf{x}, \boldsymbol{\omega}_o),$$

$$L_d(\mathbf{x}) = \mathbf{k}_d(1-m)\int_\Omega L_i(\mathbf{x}, \boldsymbol{\omega}_i)(\boldsymbol{\omega}_i \cdot \mathbf{n})\mathrm{d}\boldsymbol{\omega}_i,$$

$$L_s(\mathbf{x}, \boldsymbol{\omega}_o) = \int_\Omega \frac{DFG}{4(\boldsymbol{\omega}_o \cdot \mathbf{n})(\boldsymbol{\omega}_i \cdot \mathbf{n})} L_i(\mathbf{x}, \boldsymbol{\omega}_i)(\boldsymbol{\omega}_i \cdot \mathbf{n})\mathrm{d}\boldsymbol{\omega}_i,$$

- Specifically, for the specular term:

$$F(\boldsymbol{\omega}_o, \mathbf{h}, k_r) = F_0 + \big(\max(1-k_r, F_0) - F_0\big)\big(1 - (\boldsymbol{\omega}_o \cdot \mathbf{h})\big)^5,$$

$$L_s(\mathbf{x}, \boldsymbol{\omega}_o) = \big(F(\boldsymbol{\omega}_o, \mathbf{h}, k_r)B_0(\boldsymbol{\omega}_o \cdot \mathbf{n}, k_r) + B_1(\boldsymbol{\omega}_o \cdot \mathbf{n}, k_r)\big)\int_\Omega D(\boldsymbol{\omega}_i, \boldsymbol{\omega}_o, \mathbf{n}, k_r)L_i(\mathbf{x}, \boldsymbol{\omega}_i)\mathrm{d}\boldsymbol{\omega}_i,$$

Method: Training

- We leverage multiple HDRs and keep rotating them to encourage the predicted BRDF materials to disentangle from environment lights.

- The SDS loss w.r.t. the parameters of material network becomes:

$$\nabla_{\gamma}\mathcal{L}_{\text{SDS}}(\phi, x) = \mathbb{E}_{t,\epsilon}\left[w(t)(\epsilon_{\phi}(z_t; y, t) - \epsilon)\frac{\partial z}{\partial x}\frac{\partial x}{\partial \mathbf{k}}\frac{\partial \mathbf{k}}{\partial \gamma}\right]$$

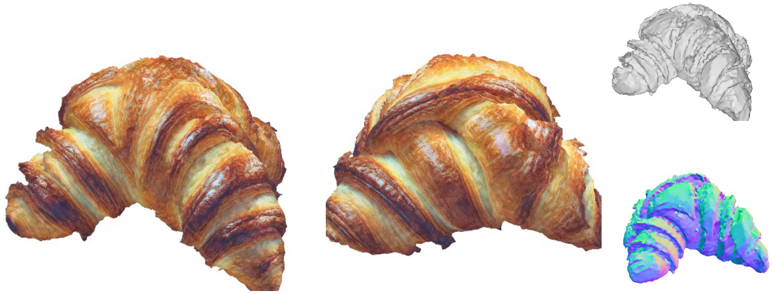
- A material smoothness regularizer is used for enforcing smooth diffuse materials.

$$\mathcal{L}_{\text{mat}} = \sum_{\mathbf{x}\in\mathcal{S}} |\mathbf{k}_d(\mathbf{x}) - \mathbf{k}_d(\mathbf{x} + \boldsymbol{\epsilon})|$$

Results: Gallery of generated 3D assets

A sliced loaf of fresh bread

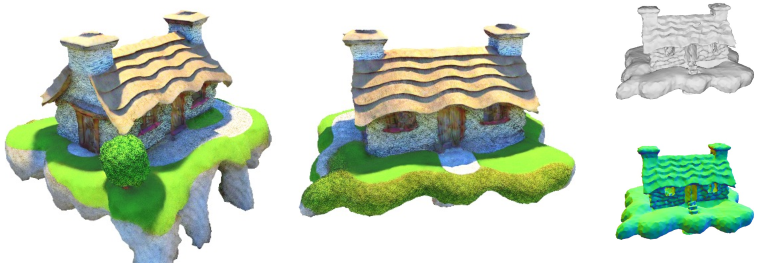
A DSLR photo of a hamburger



A delicious croissant

A pineapple

A blue tulip



A 3D model of an adorable cottage with a thatched roof

A plate piled high with chocolate chip cookies

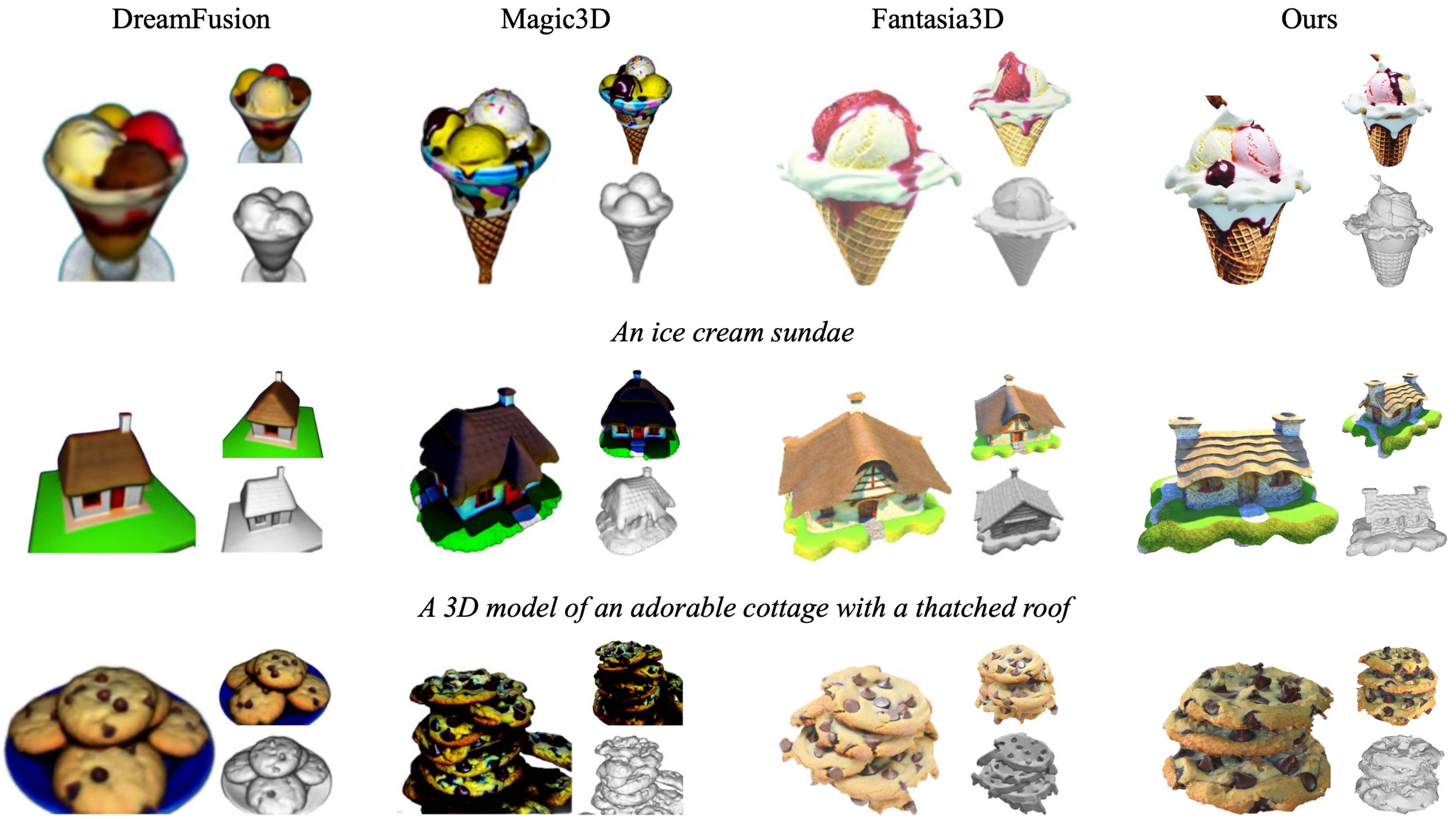
A rabbit, animated movie character, high detail 3d model

A car made out of sushi

Results: Qualitative comparison

上海人工智能实验室
Shanghai Artificial Intelligence Laboratory

- Compared to baselines, our results have more natural textures and richer details.



Results: Quantitative results

- Our method MATLABER outperforms baselines on realism, details and disentanglement.

Table 1: Mean opinion scores in range $1 \sim 5$, where 1 means the lowest score and 5 is the highest score.

Method	Alignment	Realism	Details	Disentanglement
DreamFusion [4]	3.97 (± 0.66)	3.56 (± 0.43)	3.23 (± 0.61)	3.48 (± 0.59)
Magic3D [8]	**4.01** (± 0.59)	3.84 (± 0.72)	3.70 (± 0.66)	3.14 (± 0.89)
Fantasia3D [7]	3.76 (± 0.82)	4.17 (± 0.65)	4.27 (± 0.75)	2.93 (± 0.95)
Ours	3.81 (± 0.75)	**4.35** (± 0.60)	**4.31** (± 0.70)	**3.89** (± 0.65)

- For the **alignment**, I want to talk about the deficiency of stable diffusion (CLIP).

a frog wearing a sweater*

DreamFusion

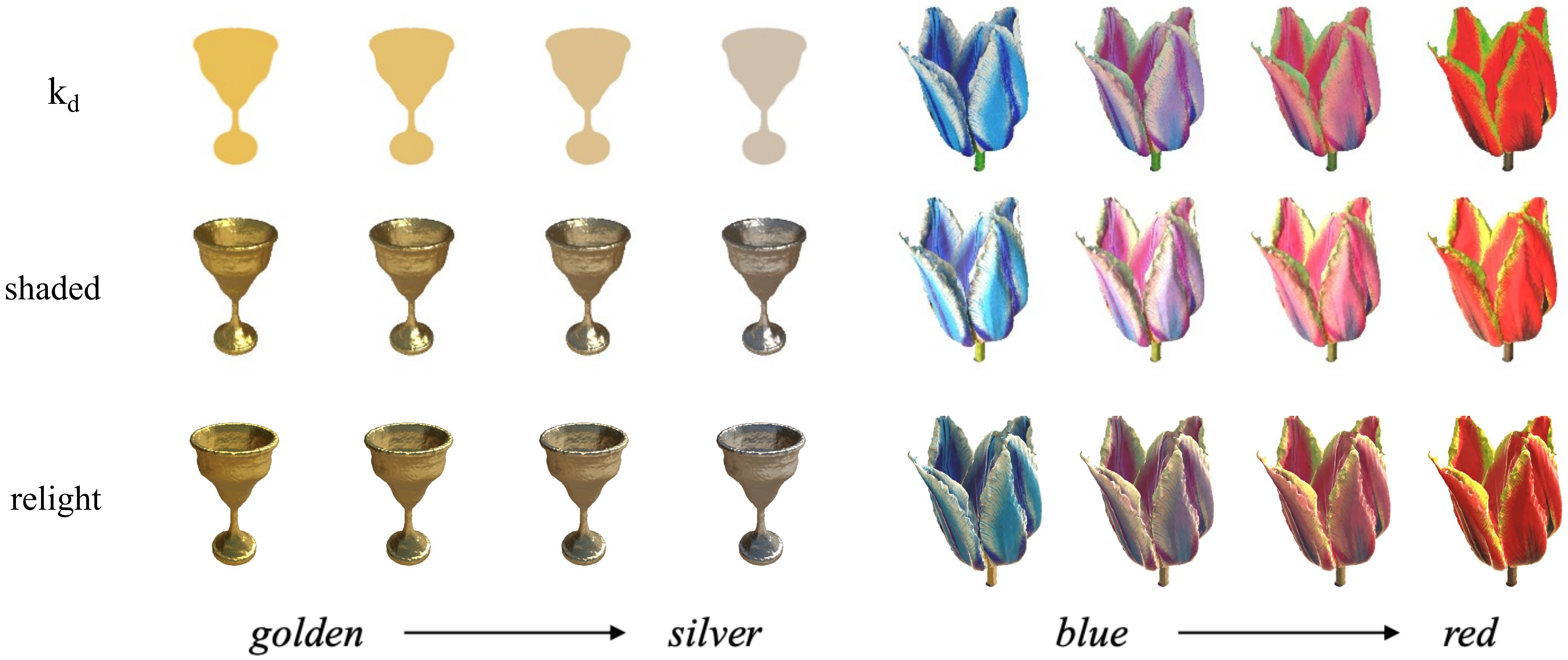
Fantasia3D/Ours

Results: Relighting results

- Our generated realistic and coherent materials naturally allows relighting.
- We show our 3D assets relit under a rotating environment light.

Results: Material interpolation

- Thanks to the smooth latent space of our BRDF auto-encoder, we can conduct a linear interpolation on the BRDF embeddings to achieve material interpolation.



Results: Failure cases

- Owing to imperfect geometry, our generated 3D objects will present clear artifacts under some novel illuminations.

Future work

- Refine geometry for shape-appearance alignment
- Larger BRDF dataset for better materials
- Better disentanglement
- Diversity problem
- From 3D objects to others
- …

Project page:

Thanks for listening!