Text-to-3D using 2D Diffusion

SUMMER OF MACHINE LEARNING AT SKOLTECH

Kirill Struminsky

Research Fellow @ HSE Centre for Deep Learning and Bayesian Methods

ЦИФРОВАЯ ЭКОНОМИКА НАЦИОНАЛЬНЫЕ ПРОЕКТЫ РОССИИ

p(B|A)yesgroup.ru

HSE UNIVERSITY

Skoltech Applied AI Center

Motivation

Image taken from “The Next Leap: How A.I. will change the 3D industry” talk by Andrew Price

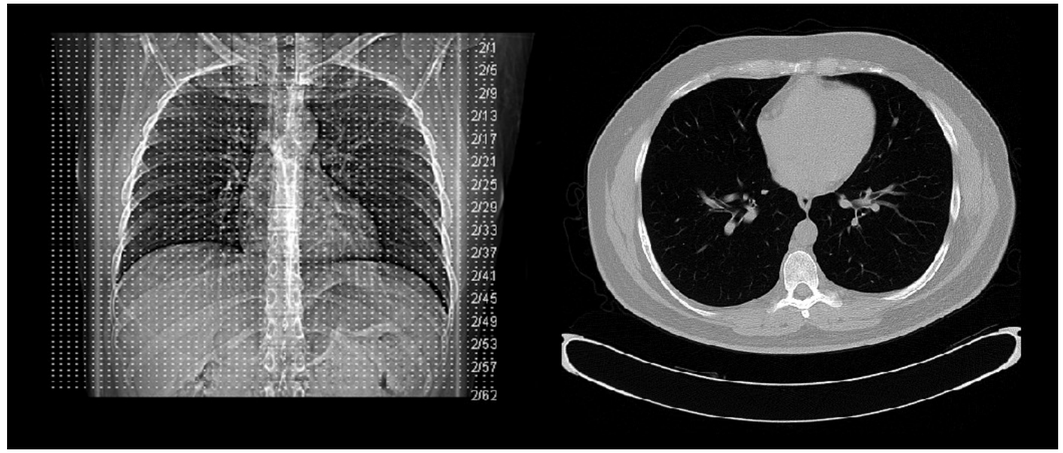
- 3D assets are expensive
- Demand goes beyond game industry

SUMMER OF MACHINE LEARNING AT SKOLTECH

Motivation

SUMMER OF MACHINE LEARNING AT SKOLTECH

- Inverse problems
  - Observe
  - Find
- Generative model
  - Find
  - Such that are plausible samples
- Can we use as an objective?



Setup

- Pre-trained text-to-2D diffusion model
- Input: text prompt

  A DSLR photo of a squirrel wearing a kimono reading a book
- Output: 3D model

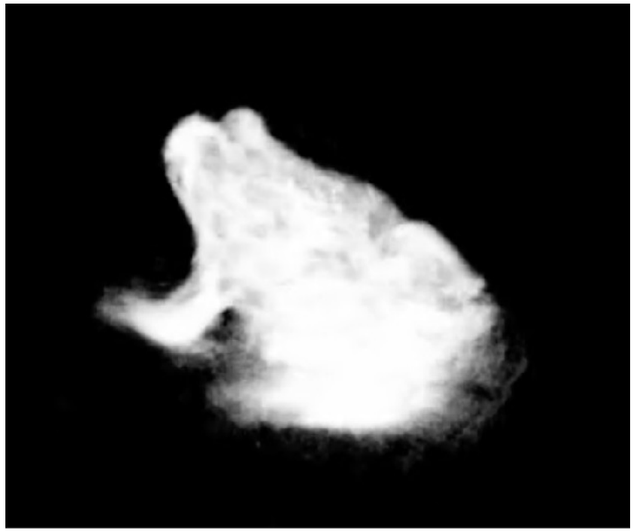
Video taken from “DreamFusion: Text-to-3D using 2D Diffusion” project page

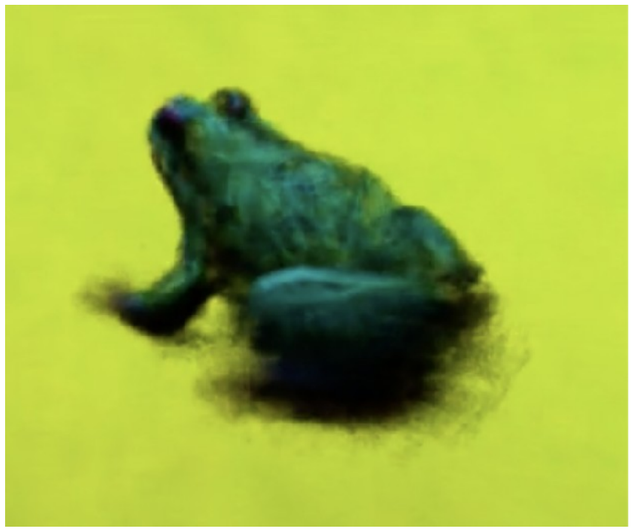
Parameterising & Rendering 3D Models

SUMMER OF MACHINE LEARNING AT SKOLTECH

- Neural Radiance Fields
- Volumetric density
- Radiance field

- Often, and are neural networks





Parameterising & Rendering 3D Models

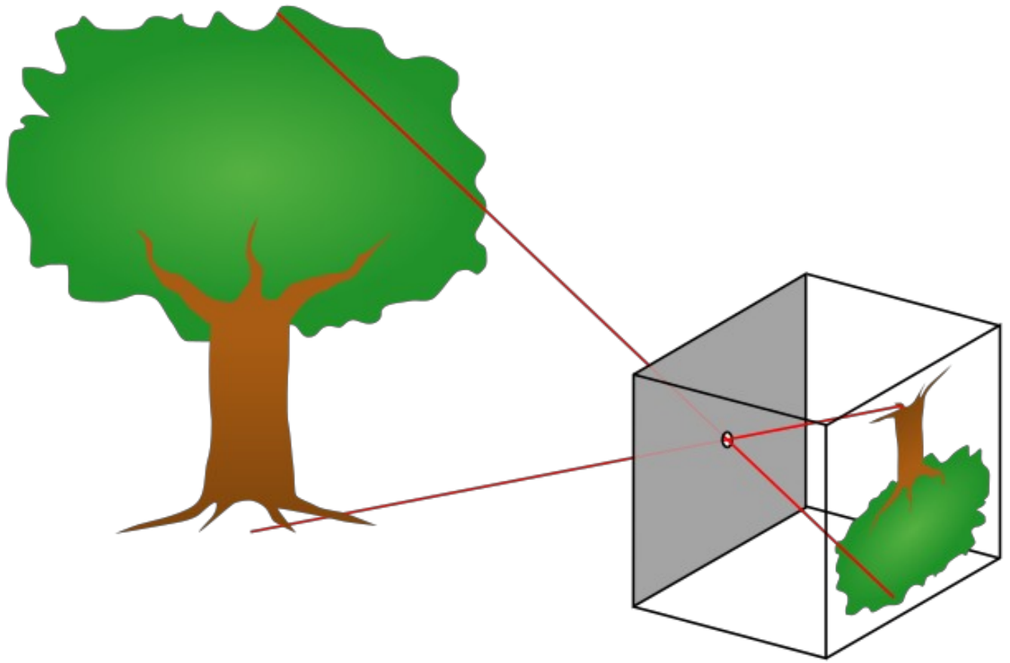
SUMMER OF MACHINE LEARNING AT SKOLTECH

- Fields and parameterise the scene
- Render image pixel by pixel

$$r = (r_o, r_d)\, t_i$$

$$\sigma_i = \sigma(r_o + t_i r_d) \qquad C_i = c(r_o + t_i r_d)$$

- Pixel color is



https://github.com/sunset1995/DirectVoxGO

Diffusion Models Recap

SUMMER OF MACHINE LEARNING AT SKOLTECH

Mapping data to noise

$$x_t = \alpha_t x_0 + \sigma_t \varepsilon$$

Trained reverse map

$$\hat{x}_\phi(x_t, t) = \frac{\left(x_t - \sigma_t \varepsilon_\phi(x_t, t)\right)}{\alpha_t}$$

$$\nabla_{x_t} \log p_t(x_t) = s(x_t) \approx -\frac{\varepsilon_\phi(x_t, t)}{\sigma_t}$$

https://arxiv.org/abs/2011.13456

DreamFusion: Text-to-3D using 2D Diffusion

SUMMER OF MACHINE LEARNING AT SKOLTECH

- Consider rendered image
  - Scene parameterisation
  - View angle, camera parameters, etc.
- How do we find ?
- Optimize ?
  - Training optimises
  - Can we optimise for
- Minimize

Score Distillation Sampling

SUMMER OF MACHINE LEARNING AT SKOLTECH

- Loss
- Image and
- Compute gradient

- Omitting score Jacobian helps

Score Distillation Sampling Loss

SUMMER OF MACHINE LEARNING AT SKOLTECH

- Let
- Can we interpret ?

Score Distillation Sampling Loss

SUMMER OF MACHINE LEARNING AT SKOLTECH

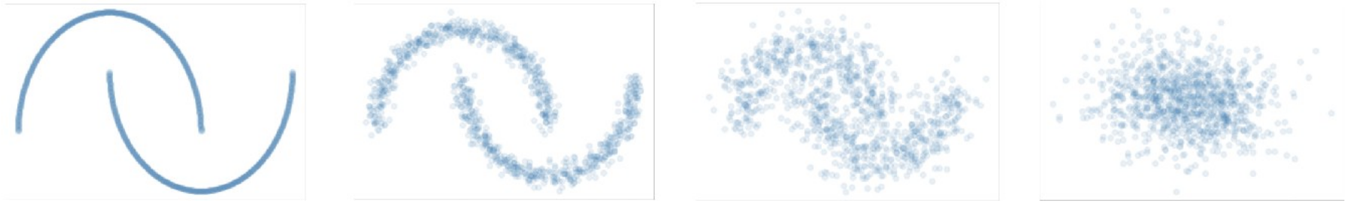
- Let
- Can we interpret ?
- Recall and rewrite

- Finally

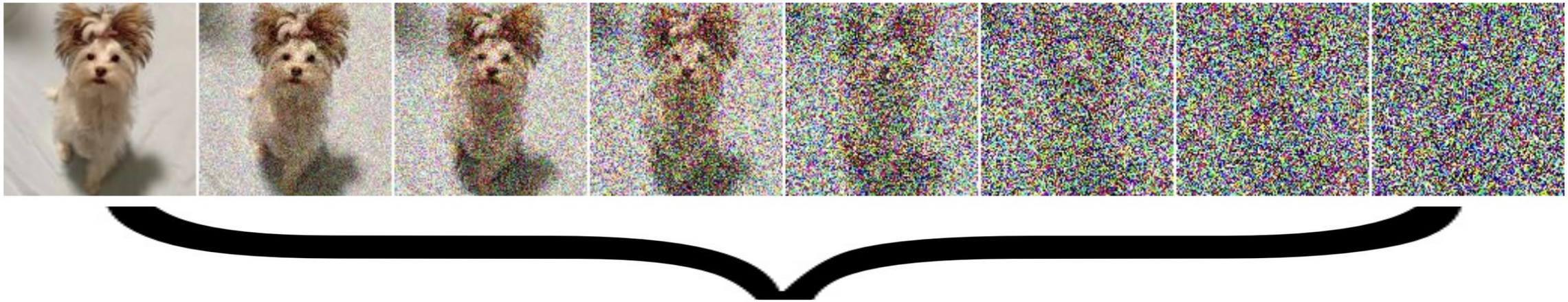
Score Distillation Sampling Loss

SUMMER OF MACHINE LEARNING AT SKOLTECH

- In sum,



L_{Diff}

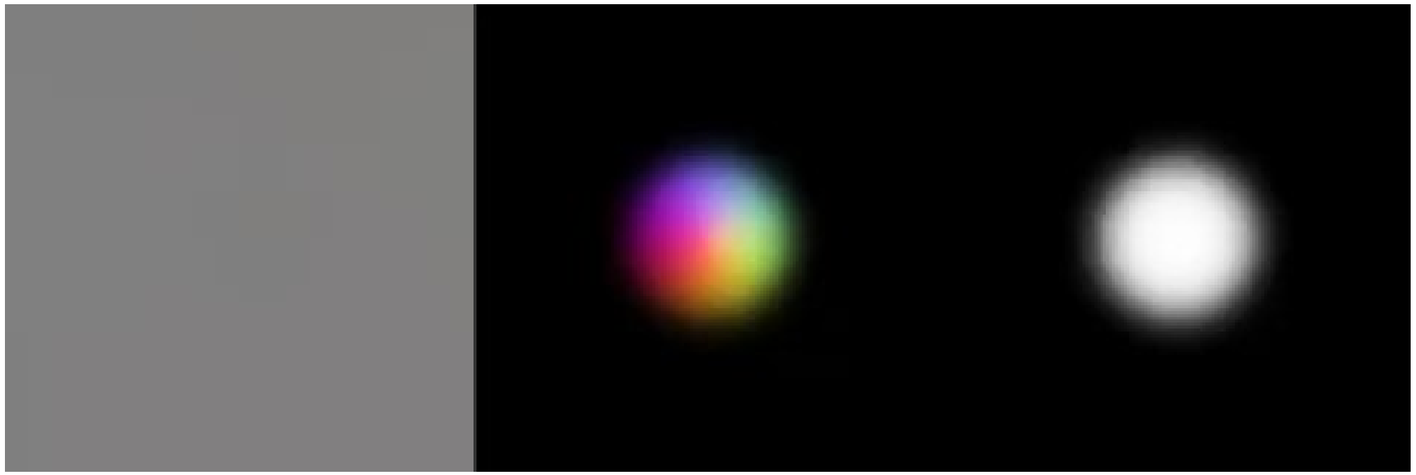


L_{SDS}

Training Dreamfusion

SUMMER OF MACHINE LEARNING AT SKOLTECH

- Choose prompt (omitted for clarity)
- Diffusion model (Imagen)
- Training loop:
  - Sample camera position
  - Render image
  - Sample and
  - Update with



Additional Heuristics (I)

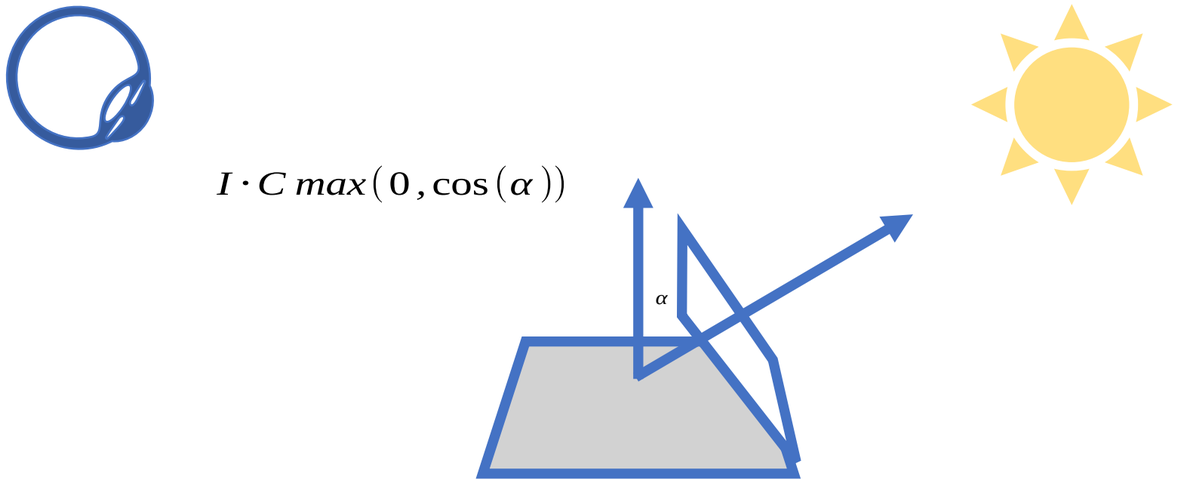
SUMMER OF MACHINE LEARNING AT SKOLTECH

- Janus problem
- Remedy: specify camera angle in prompts

Additional Heuristics (II)

SUMMER OF MACHINE LEARNING AT SKOLTECH

- Poor geometry
- Remedy (1/2): simulate shadows



- Remedy (2/2): process colourless images

Problems to Fix

SUMMER OF MACHINE LEARNING AT SKOLTECH

- Poor model quality
- Seemingly ad-hoc objective
- Low sample diversity
- Janus problem

Adapting to Latent Diffusion

SUMMER OF MACHINE LEARNING AT SKOLTECH

- Imagen is not public
- Naive adaptation works fine
  - Render image
  - Encode
  - Compute SDS loss in latent space

a blue poison-dart frog sitting on a water lily

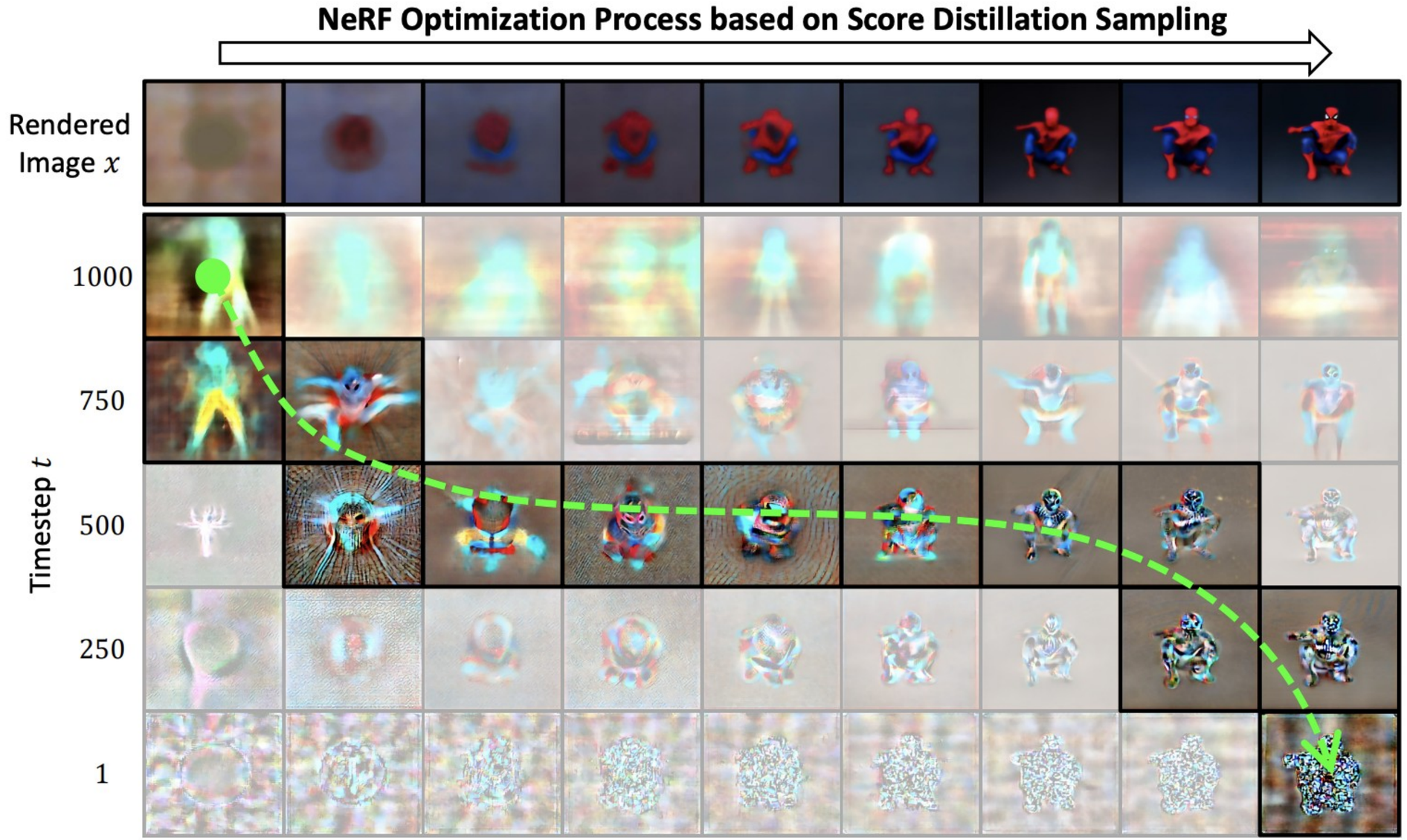
neuschwanstein castle, aerial view

(1) 64 rendering | (2) + 512 rendering | (3) + Annealed t | (4) + VSD

https://research.nvidia.com/labs/dir/magic3d/
https://ml.cs.tsinghua.edu.cn/prolificdreamer/

Time Annealing

SUMMER OF MACHINE LEARNING AT SKOLTECH



https://arxiv.org/abs/2306.12422

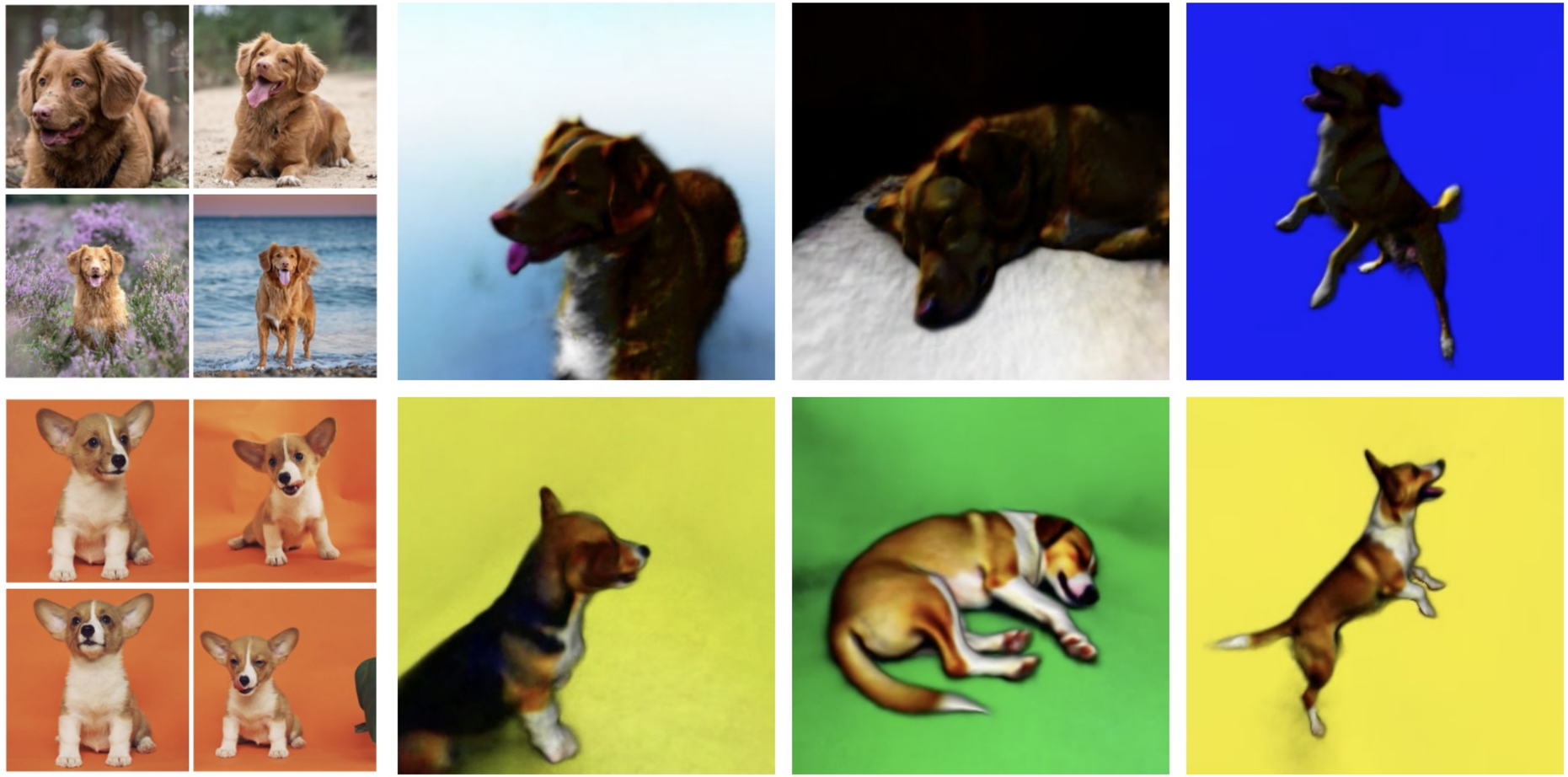
Tackling Low Sample Diversity

SUMMER OF MACHINE LEARNING AT SKOLTECH

https://dreambooth.github.io/

Dreambooth 3D

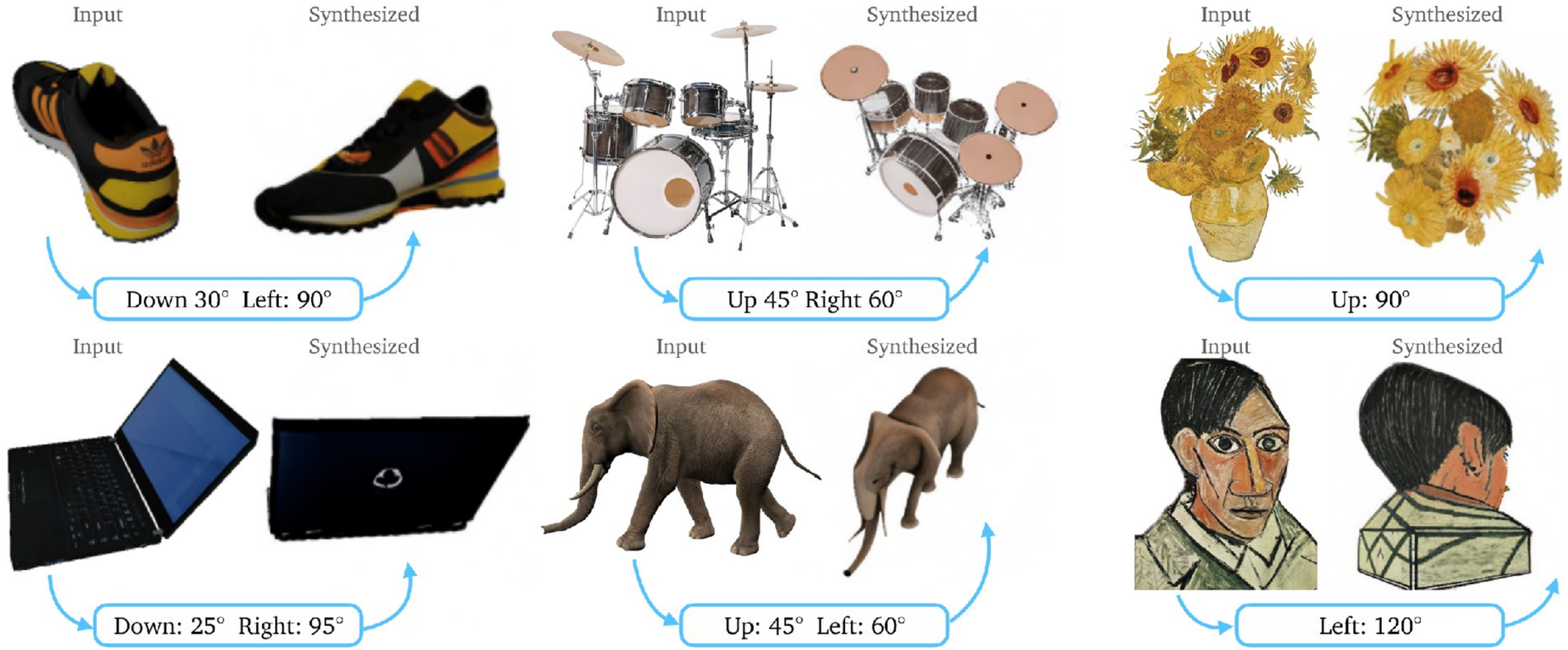
SUMMER OF MACHINE LEARNING AT SKOLTECH



https://dreambooth3d.github.io/

Zero-123: viewangle control

SUMMER OF MACHINE LEARNING AT SKOLTECH



https://zero123.cs.columbia.edu/

Zero-123: examples

https://zero123.cs.columbia.edu/

Magic123

SUMMER OF MACHINE LEARNING AT SKOLTECH

https://guochengqian.github.io/project/magic123/

Key takeaways

SUMMER OF MACHINE LEARNING AT SKOLTECH

- Text-to-2D models can solve to text-to-3D
- Generation with SDS loss
- Model fine-tuning can address some of SDS issues