Team 2
Midterm Project Presentation

Sync3D: Single Image Reconstruction via Diffusion Syncing in 3D Space

Asiman Ziyaddinov, Jinhyuk Jang, Prin Phunyaphibarn

SiNeRF

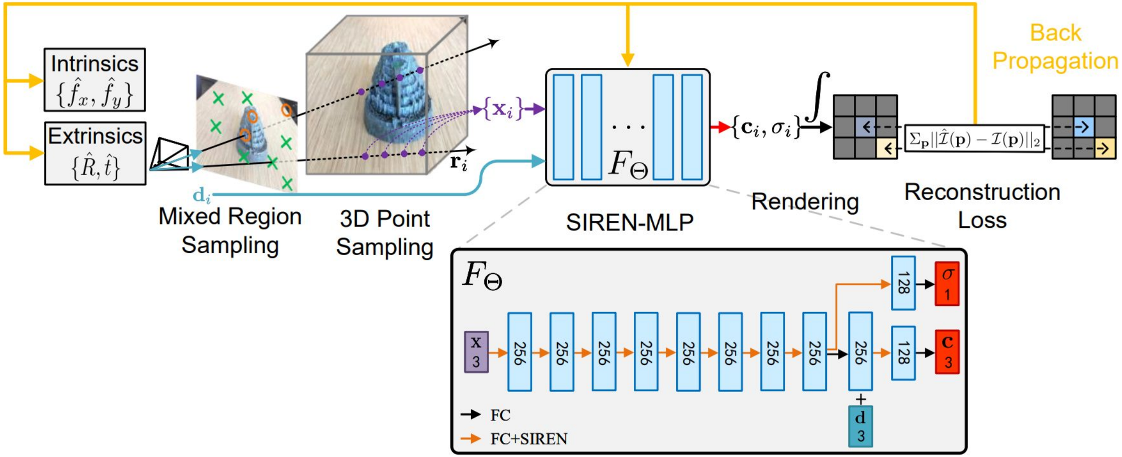
- Sinusoidal Activation functions

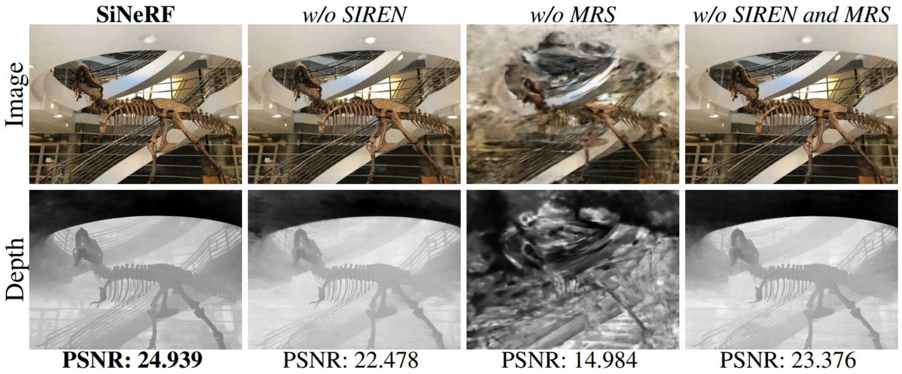
Pros:

- Increased Accuracy
- Reduced Artifacts
- Effective for Periodic Structures

Cons:

- Increased Computation
- Implementation Complexity
- Not Universally optimal





Zero 1-to-3

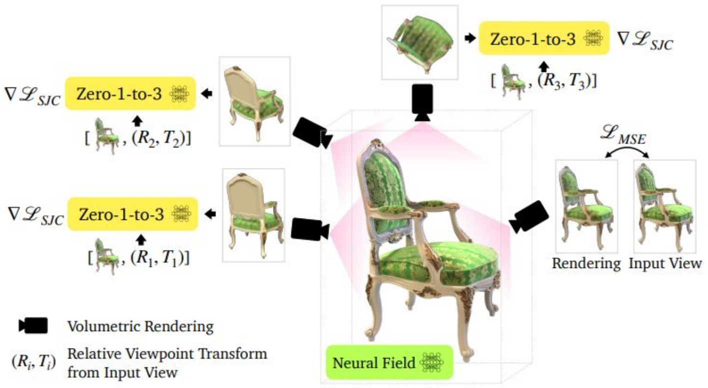
- single-image 3D scene generation

Pros:

- Data Efficiency
- Versatile Applications
- High-Quality Output

Cons:

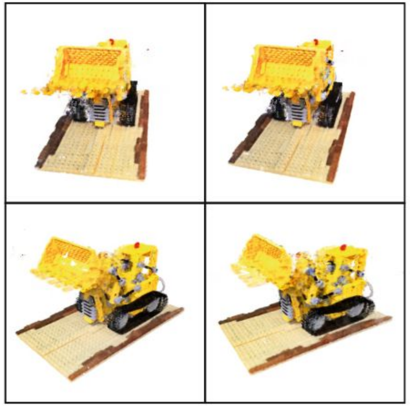
- Limited Control
- Inconsistent Detail
- Dependence on Pre-trained Models



Combining Zero-1-to-3 and SinNeRF for Iterative 3D Scene Synthesis

SinNeRF

- \+ 3D Consistent
- \- Low Quality (Blurry)



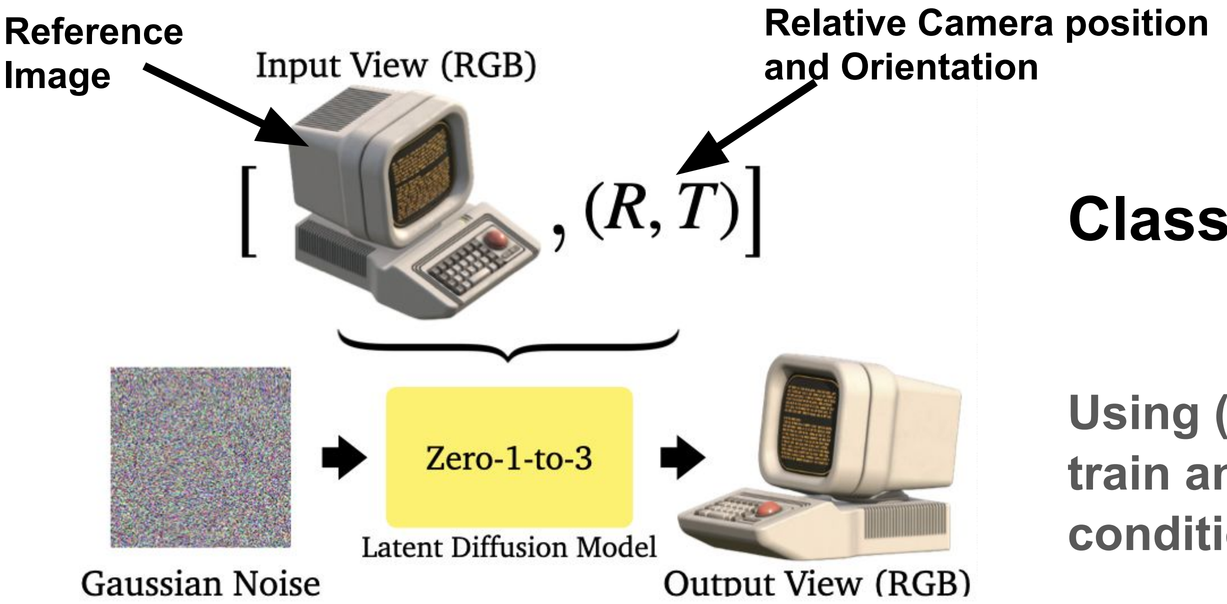
Zero 1-to-3

- \+ High Quality
- \- 3D Inconsistent

Guided Diffusion Refinement

- \+ High Quality
- \+ 3D Consistent

Zero 1-to-3 Predicted Images



Novel View Synthesis

Classifier Free Guidance

Using (image, condition) pair to train and generate image in right condition

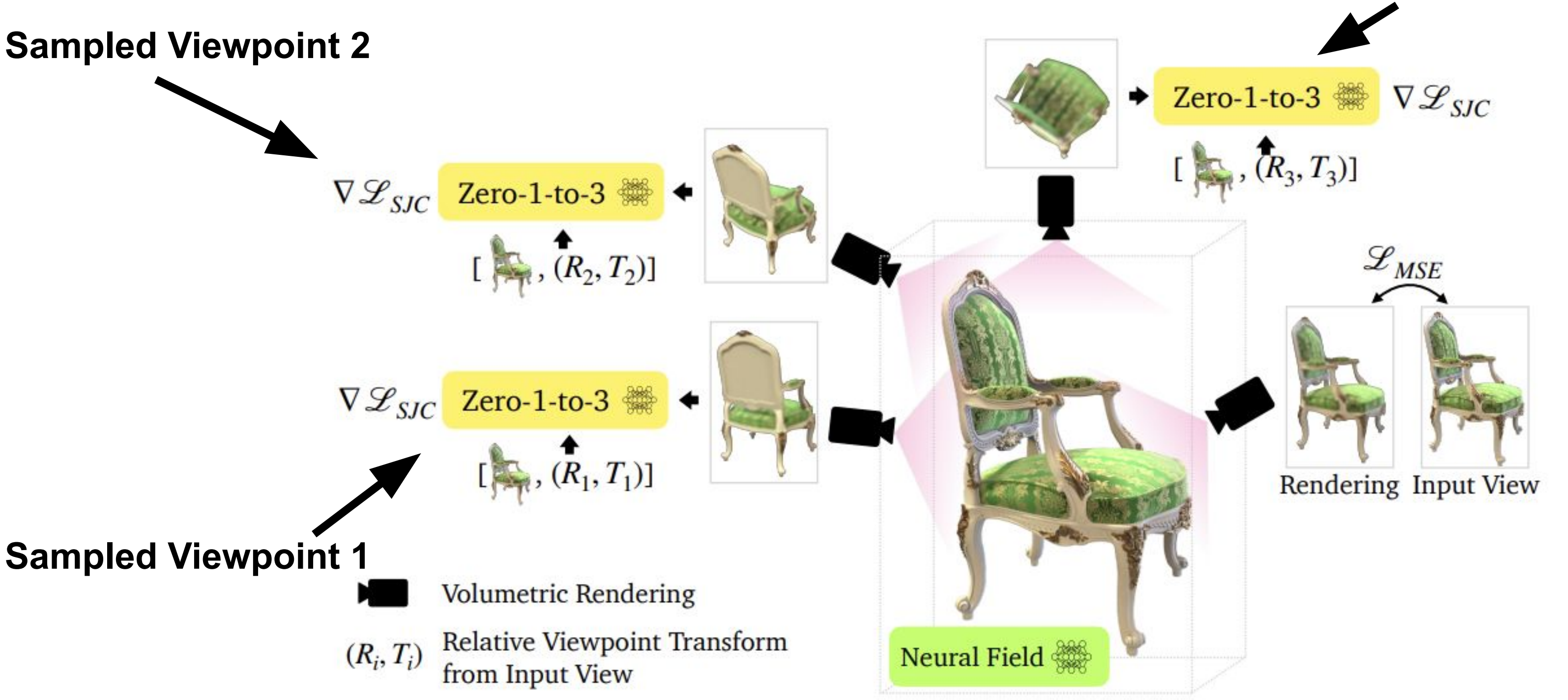
Zero 1-to-3 Predicted Images

Input

Classifier free guidance

Zero 1-to-3

NeRF/3DGS Pseudo GT



Sample different viewpoints → Volumetric Rendering

NeRF/3DGS Pseudo GT

Zero 1-to-3

NeRF / 3D Gaussian Splatting

Pseudo Ground Truth

Combining Zero-One-to-Three and SineRF for Iterative 3D Scene Synthesis

Step 1. ***Zero-One-to-Three for Initial 2D Views***

Generate multiple 2D images with angles from a single input image.
Challenge: High quality but inconsistent views due to stochasticity of diffusion process.

Step 2. NeRF ***for 3D Scene Reconstruction***

Use generated 2D views as input to NeRF to create a preliminary 3D scene.
Result: A consistent but blurry and inaccurate 3D representation.

Step 3. ***Guided Diffusion Refinement***

Use the initial 3D scene as guidance for a diffusion model.
Process: Iteratively refine to improve 3D scene accuracy.

Goal: *Refine a 3D scene from a single image through iteration*

Semantic Guidance: Guide Zero 1-to-3 via Pseudo-GT

Universal Guidance of Diffusion Process

The Diffusion Process can be guided using the gradient of a loss function

Noisy Images

“Predicted”
Clean Images
(Tweedies)

Inject Guidance

Bansal, Arpit, et al. "Universal guidance for diffusion models." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.

Universal Guidance of Diffusion Process

The Diffusion Process can be guided using the gradient of a loss function

$$\underbrace{\hat{\epsilon}_\theta(z_t, t)}_{\text{Updated Noise}} = \epsilon_\theta(z_t, t) + s(t) \cdot \underbrace{\nabla_{z_t} \ell(c, f(\overbrace{\hat{z}_0}^{\text{Tweedies}}))}_{\text{Gradient Guidance}}$$

Bansal, Arpit, et al. "Universal guidance for diffusion models." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.

LPIPS: Capturing Low-level Structure

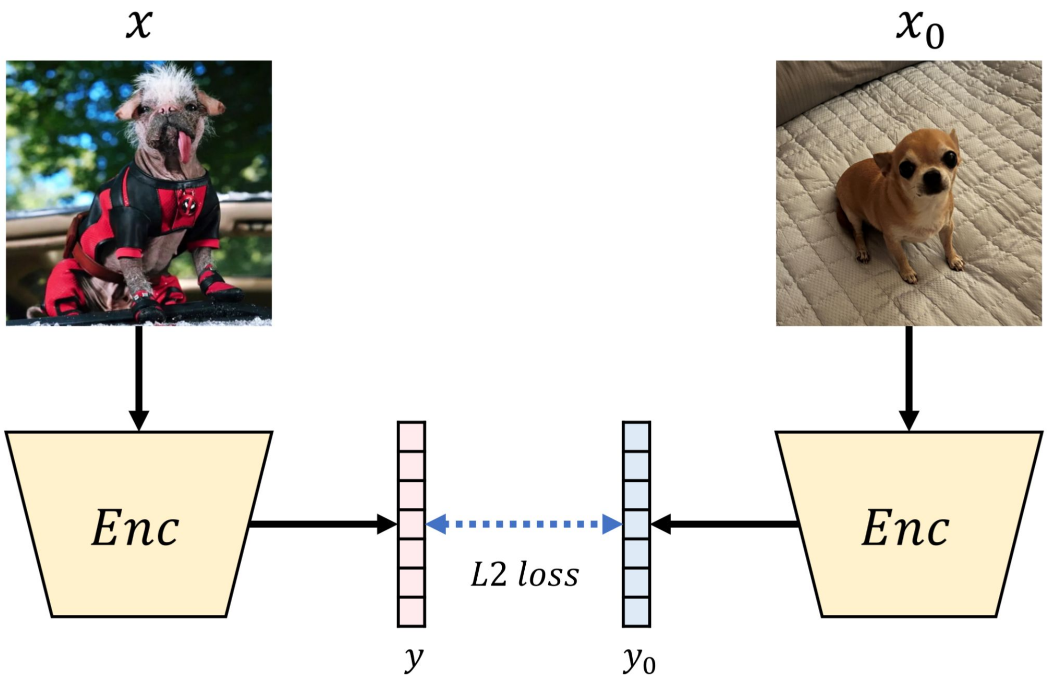


Figure from https://ethanswinery.tistory.com/26

LPIPS: Capturing Low-level Structure

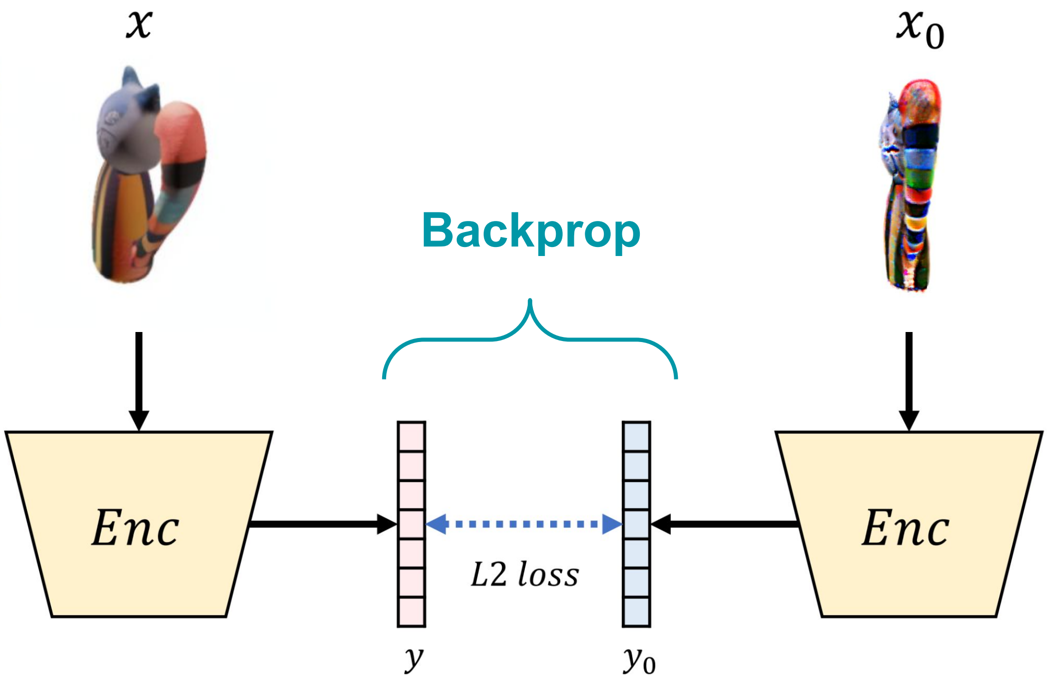


Figure from https://ethanswinery.tistory.com/26

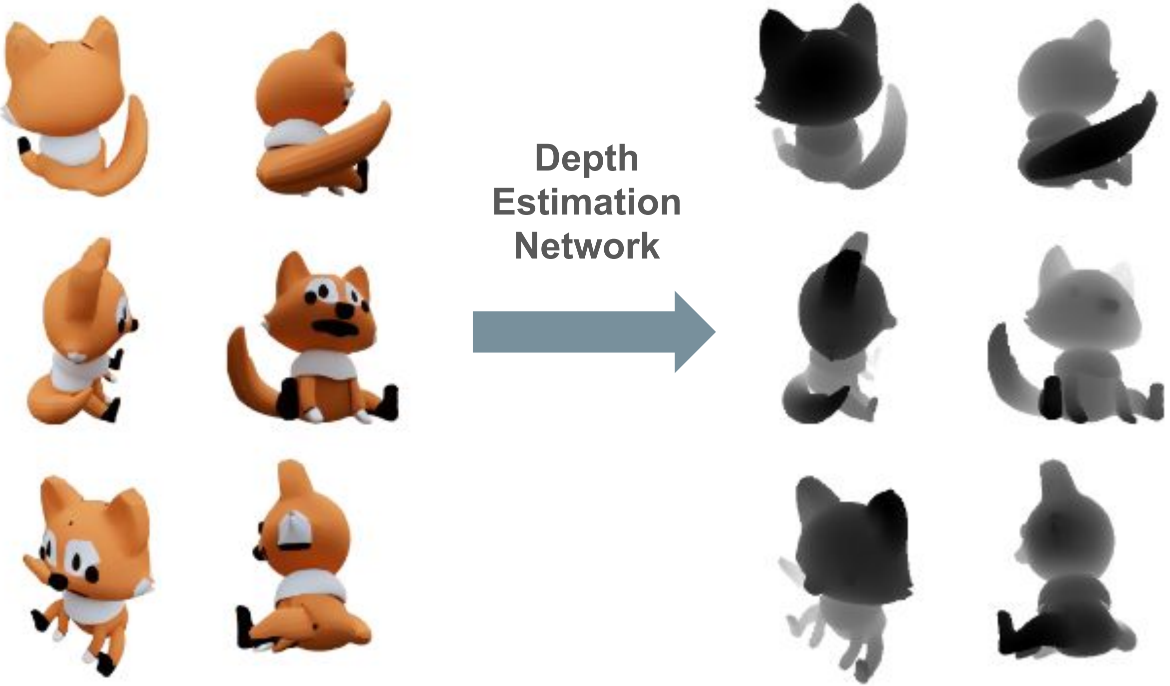
Geometric Guidance: Depth-map warping

Generate Pseudo Depth Maps by warping

Shi, Ruoxi, et al. "Zero123++: a single image to consistent multi-view diffusion base model." arXiv preprint arXiv:2310.15110 (2023).

Geometric Guidance: Depth-map warping

Zero 1-to-3 Renderings

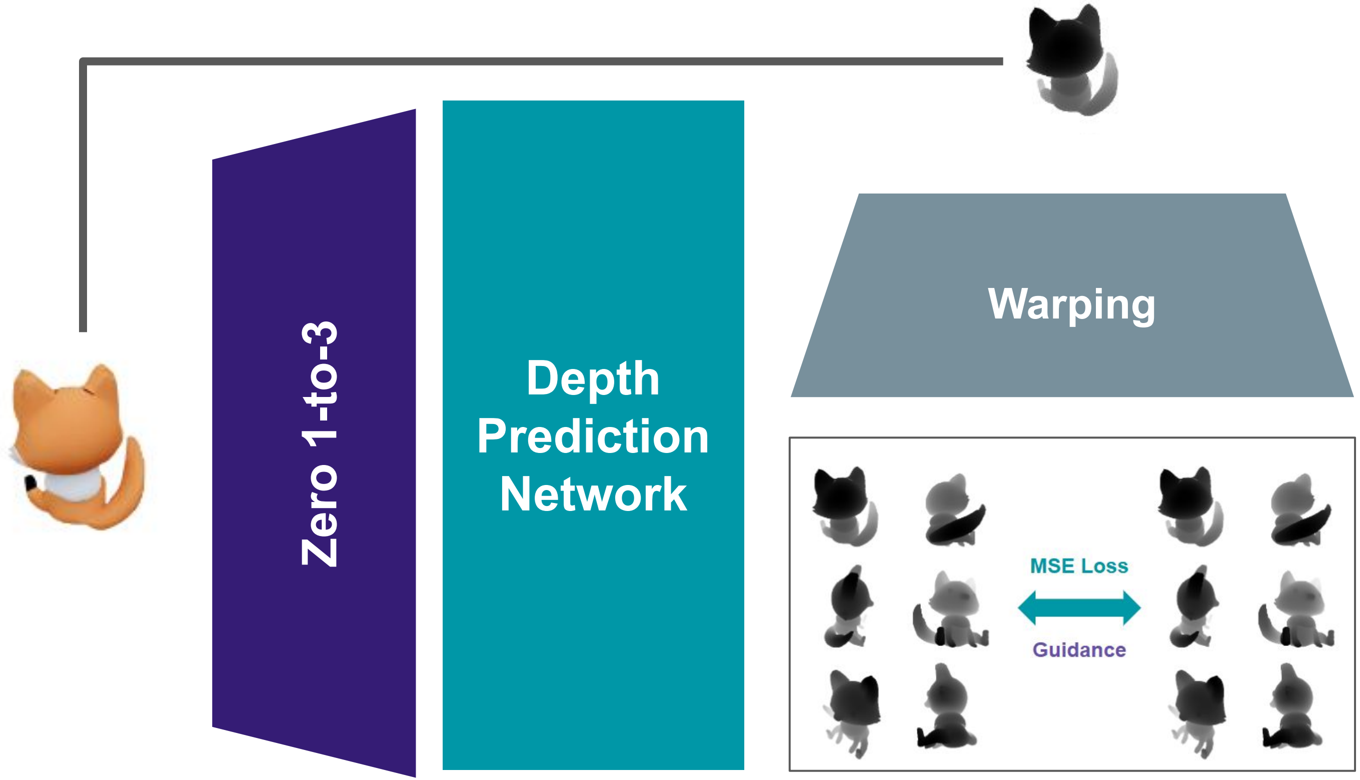


Shi, Ruoxi, et al. "Zero123++: a single image to consistent multi-view diffusion base model." arXiv preprint arXiv:2310.15110 (2023).

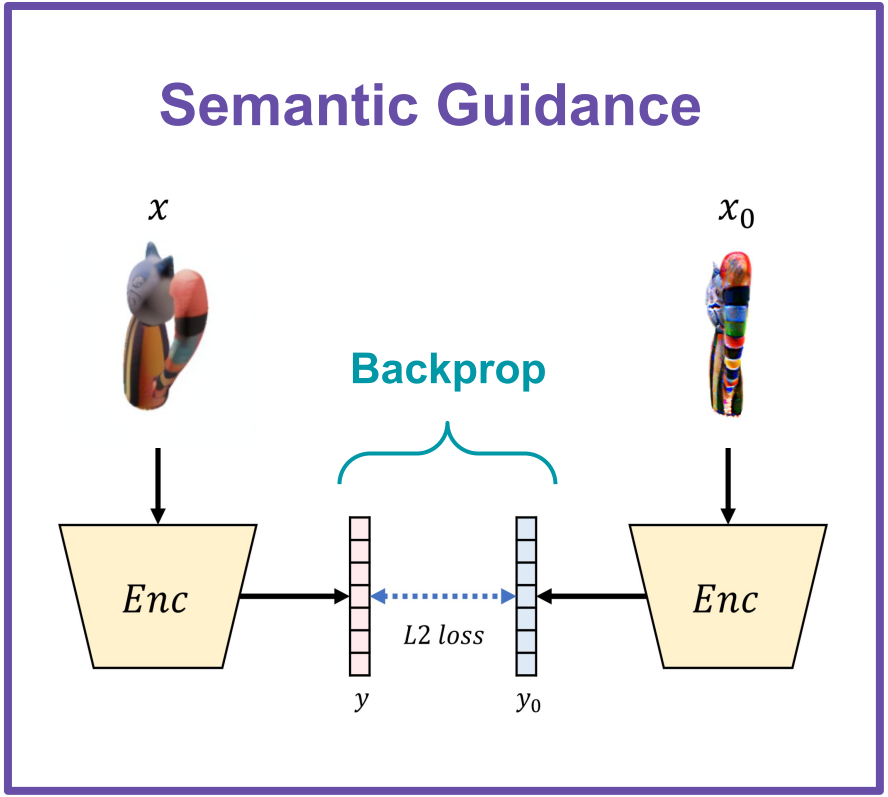
Geometric Guidance: Depth-map warping

Shi, Ruoxi, et al. "Zero123++: a single image to consistent multi-view diffusion base model." arXiv preprint arXiv:2310.15110 (2023).

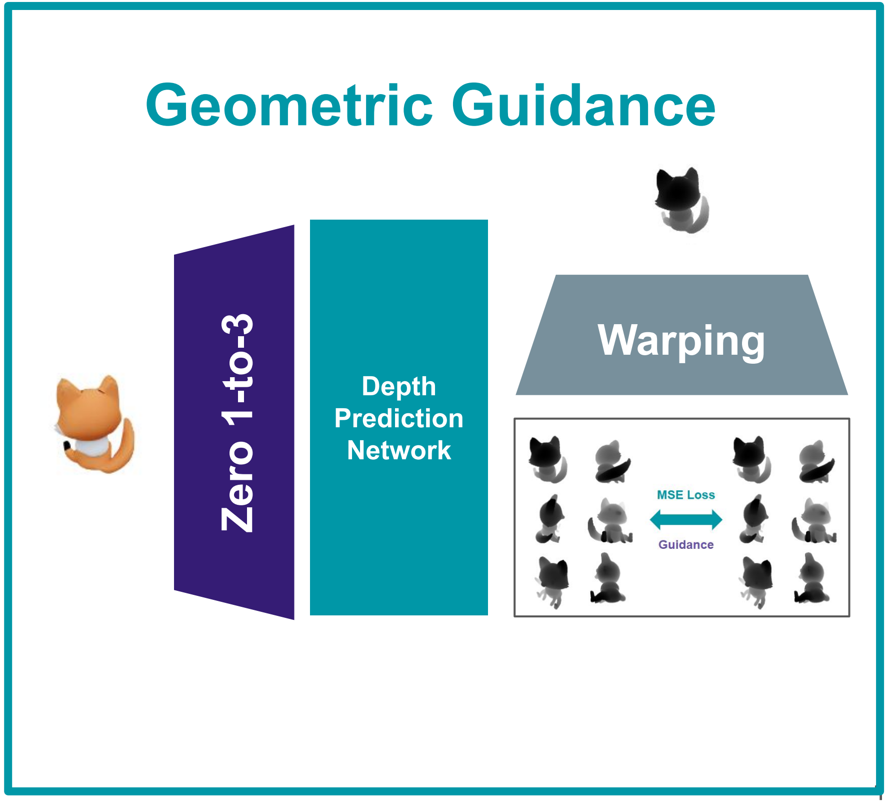
Geometric Guidance: Depth-map warping



Conclusion



Semantic Guidance
x
x_0
Backprop
Enc
Enc
L2 loss
y
y_0



Geometric Guidance
Zero 1-to-3
Depth Prediction Network
Warping
MSE Loss
Guidance