Generative AI for VFX & Games

Workshop - AI4CC - AI for Content Creation

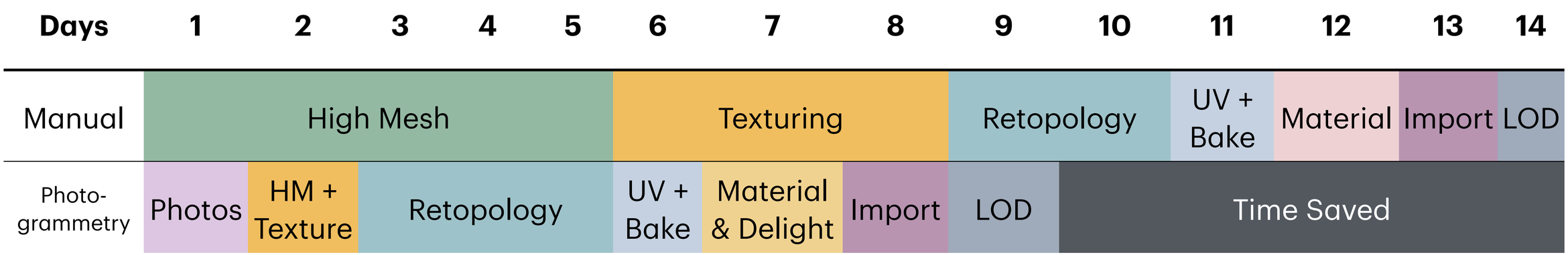
Mark Boss - Researcher @ Stability AI

Why 3D assets?

- Games and movies rely on 3D objects for rendering
- Artists are used to work with meshes
    - 3D asset generation can help teams scale

3D is time-consuming

- Creating assets is highly time-consuming



Estimated time for a single asset

- Pipeline per asset is long and involved
- Speeding up or replacing any tasks with AI allows to scale visual media tremendously

Table from: Create Photorealistic Game Assets - E-Book - Unity Technologies - 2017

Challenges

- 3D assets of artists are highly detailed and reach photorealism
  - No current 3D generation solution is capable of reaching that
- 3D datasets are small
  - Even more if we require more niche properties (Animation, additional material data)

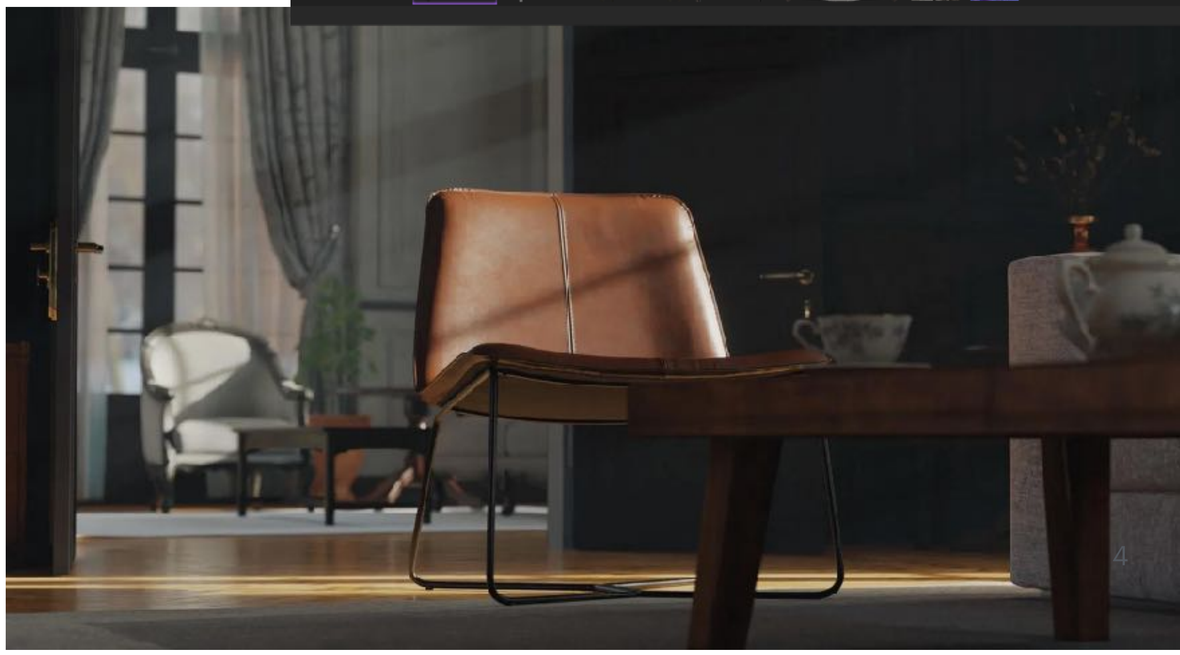


Image Datasets
5 - 10 billion samples

3D Datasets
0.25-10 million samples

Quality levels and goals

- Several tasks have different goals and requirements

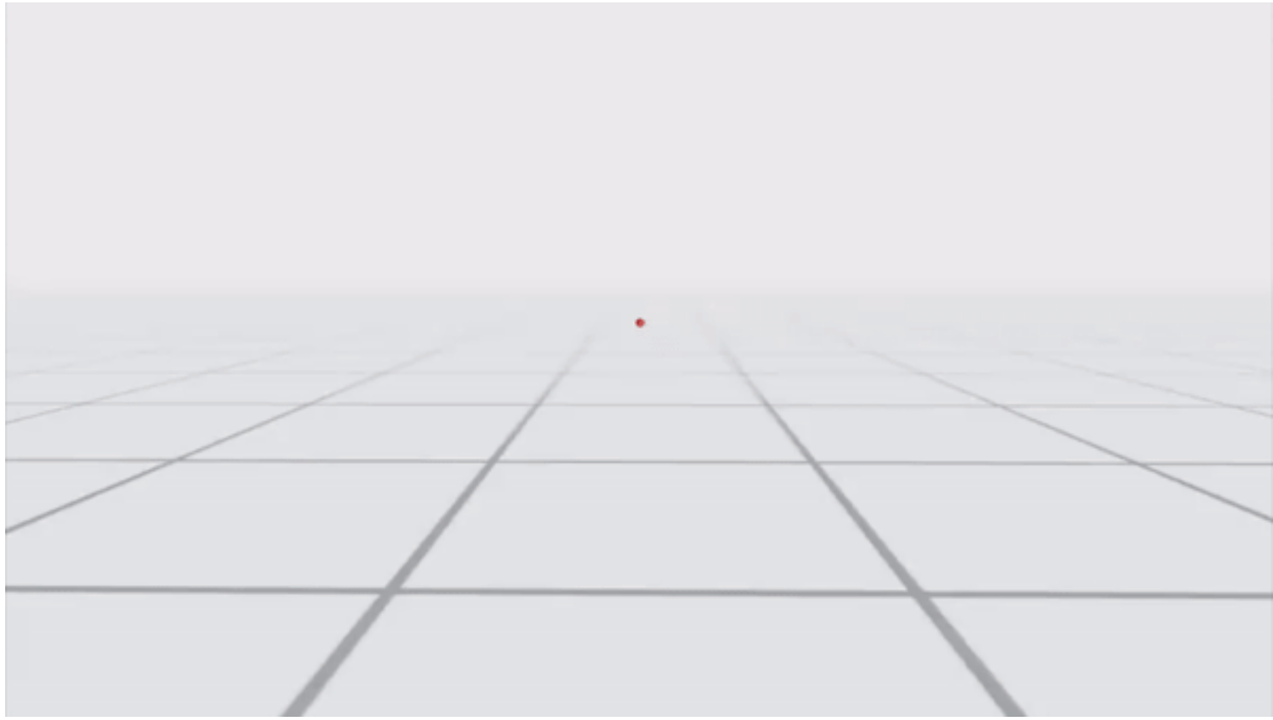
Pre-production
- Fast iteration
- Lower fidelity

Production
- Slower iteration speed
- Artistic intent
- Medium fidelity

Post-production
- Only polish
- Precision
- Iteration speed not a concern

Pre-production

- Rapid iteration essential
- Getting a feel for the scene
- Large benefits in the velocity of overall production
- Pre-production can include 3D asset generation or fast image-based explorations



Example from Evan Jones (Sped up to in-house SF3D speed)

3D Generation

Project Page

SF3D: Stable Fast 3D Mesh Reconstruction with UV-Unwrapping and Illumination Disentanglement

Mark Boss, Zixuan Huang, Aaryaman Vasishta, Varun Jampani

Saturday Afternoon Poster Session - Poster #37

Image to Relightable Object

SF3D

Image to Relightable Object

Improvements of SF3D

- Baked-in illumination

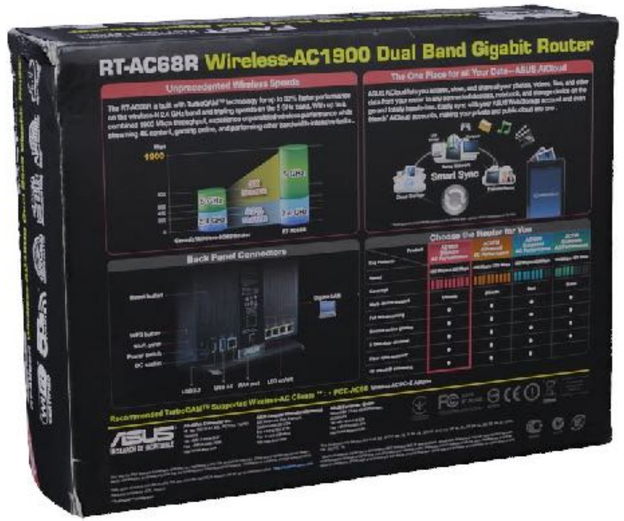
Ground Truth

TripoSR

Ours (SF3D)

Improvements of SF3D

- Baked-in illumination
- Vertex Color do not produce sharp textures



Ground Truth

TripoSR

Ours (SF3D)

Improvements of SF3D

- Baked-in illumination
- Vertex Color do not produce sharp textures
- Marching Cubes can generate obvious shading artifacts

Ground Truth

TripoSR

Ours (SF3D)

Improvements of SF3D

- Baked-in illumination
- Vertex Color do not produce sharp textures
- Marching Cubes can generate obvious shading artifacts
- No material parameters

Ground Truth

TripoSR

Ground Truth

TripoSR

Improvements of SF3D

- Baked-in illumination
- Vertex Color do not produce sharp textures
- Marching Cubes can generate obvious shading artifacts
- No material parameters

Ground Truth

Ours

Ground Truth

Ours

Overview

Overview

Overview

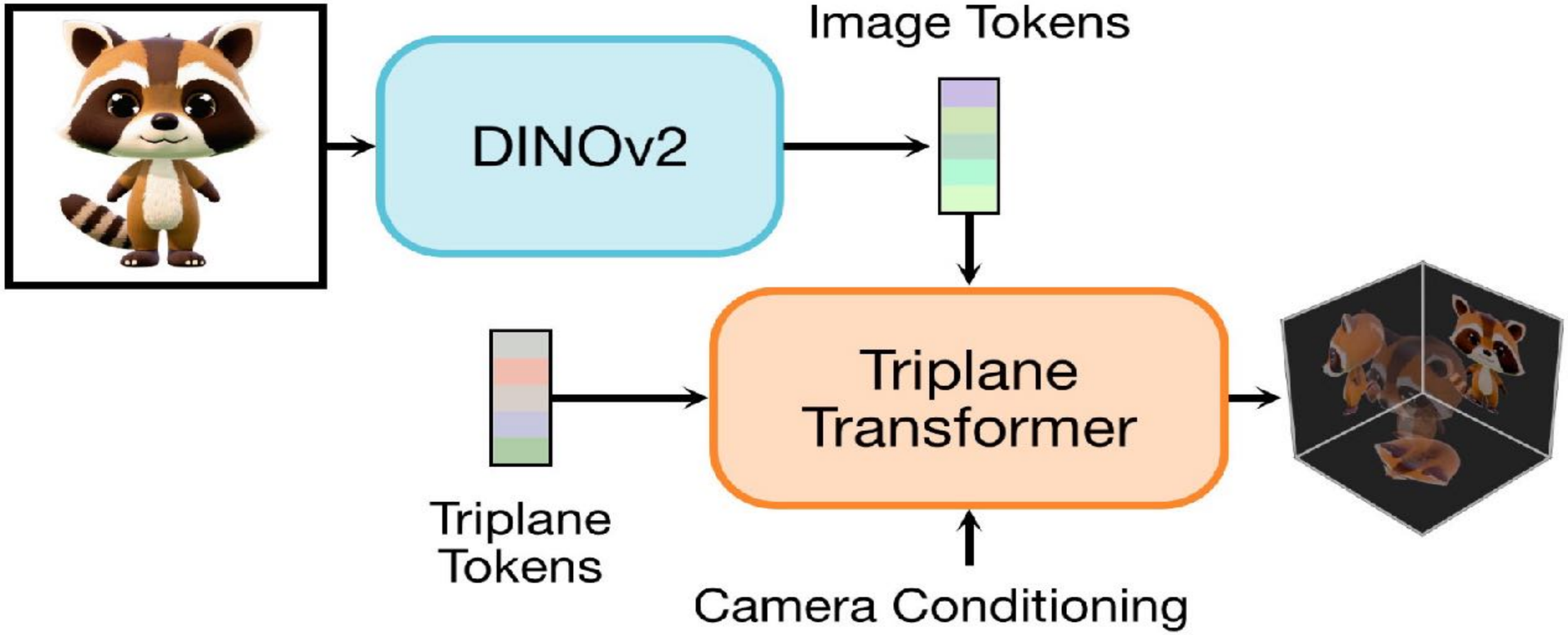
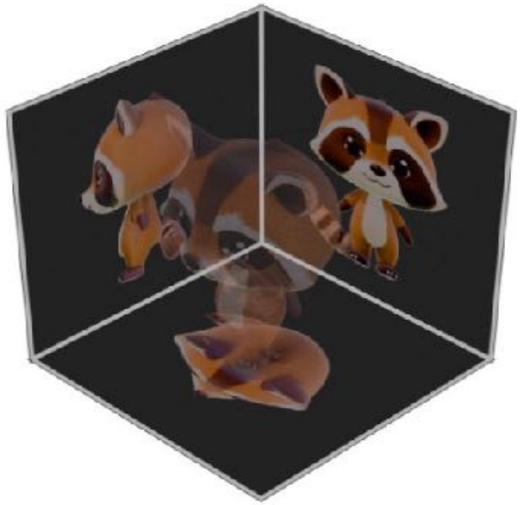
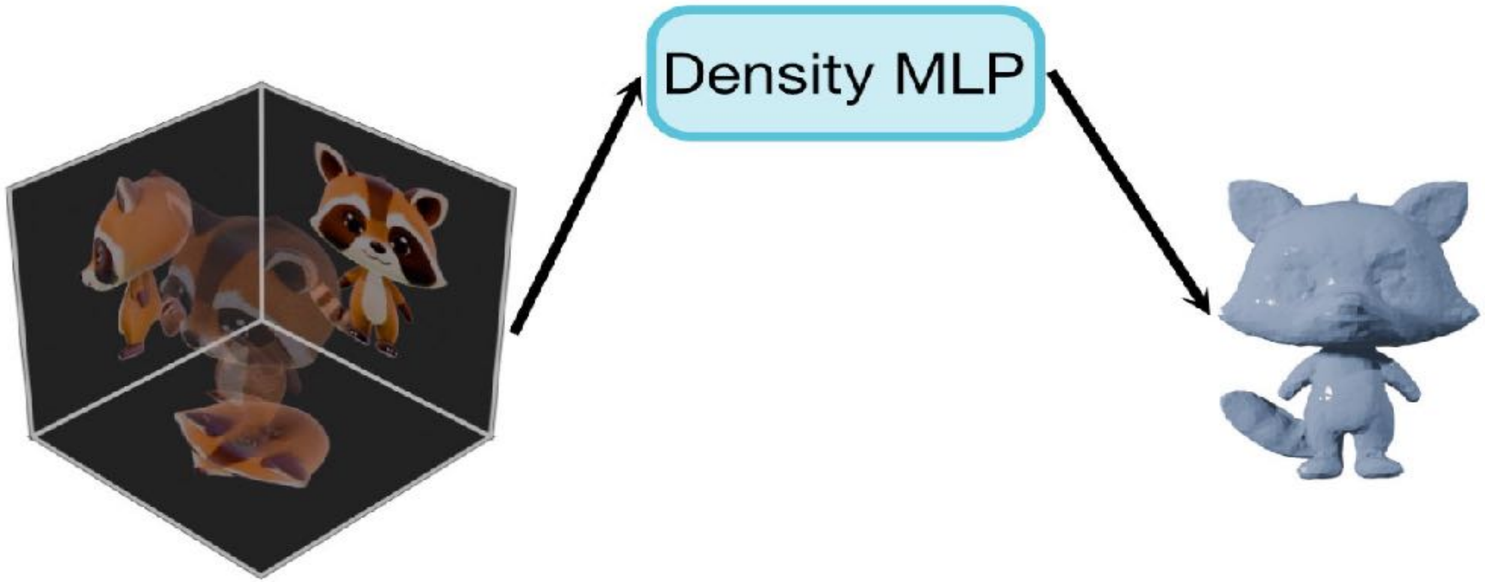


Image Tokens
DINOv2
Triplane
Transformer
Triplane
Tokens
Camera Conditioning

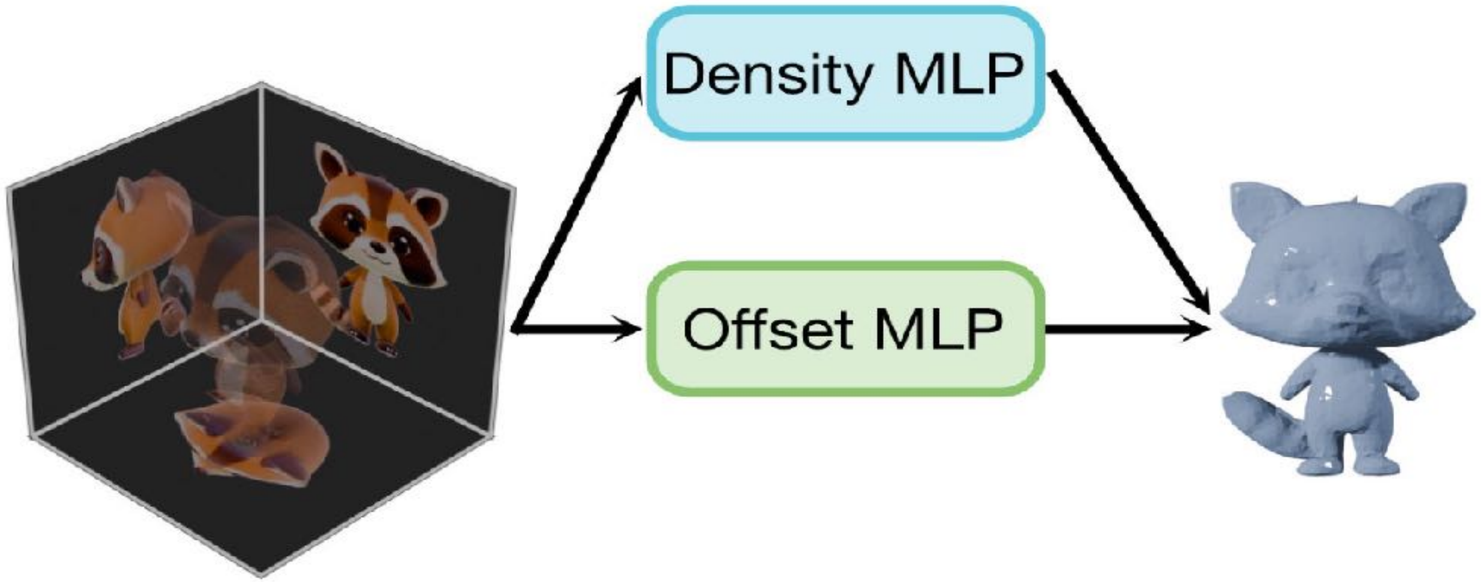
Overview



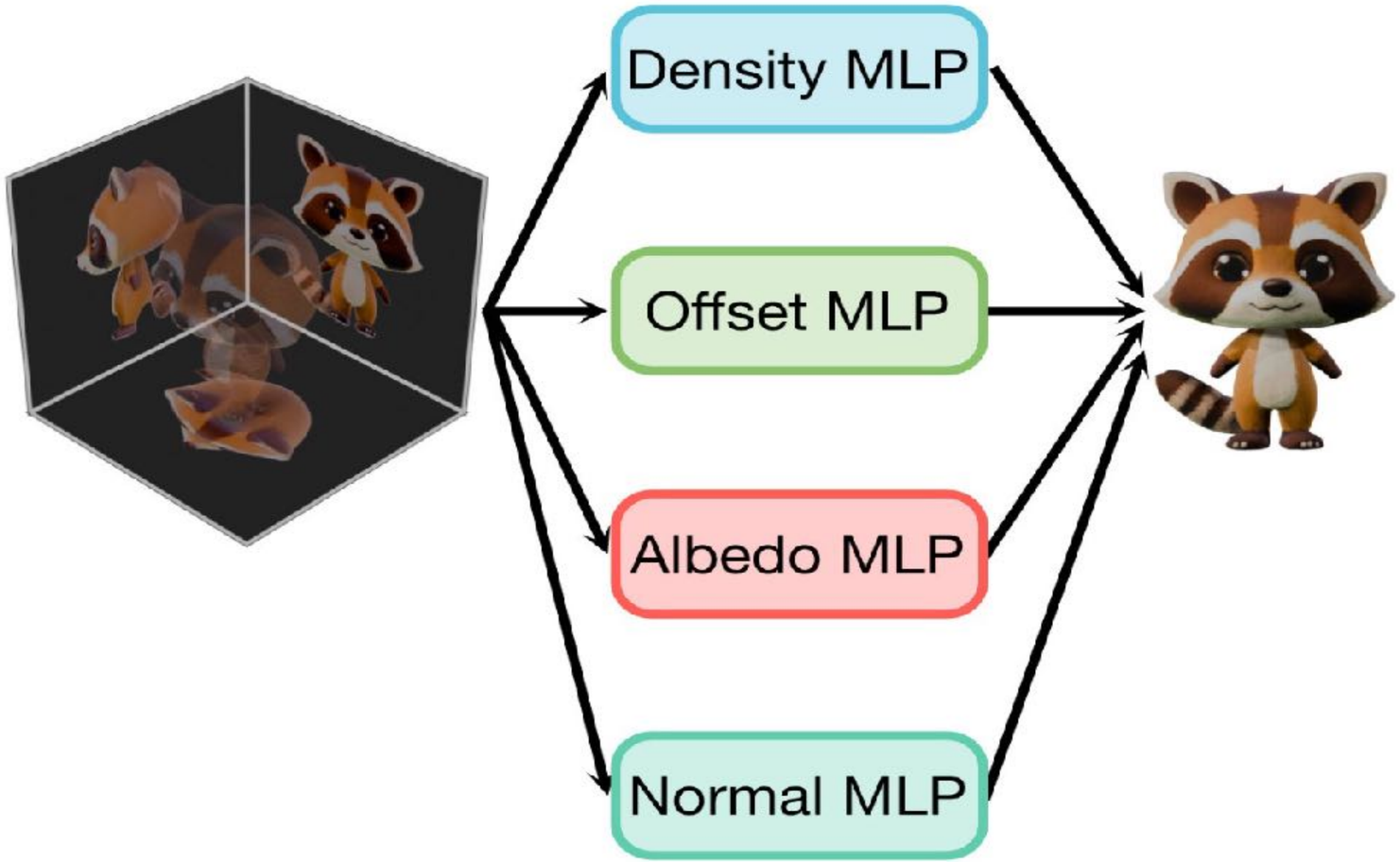
Overview



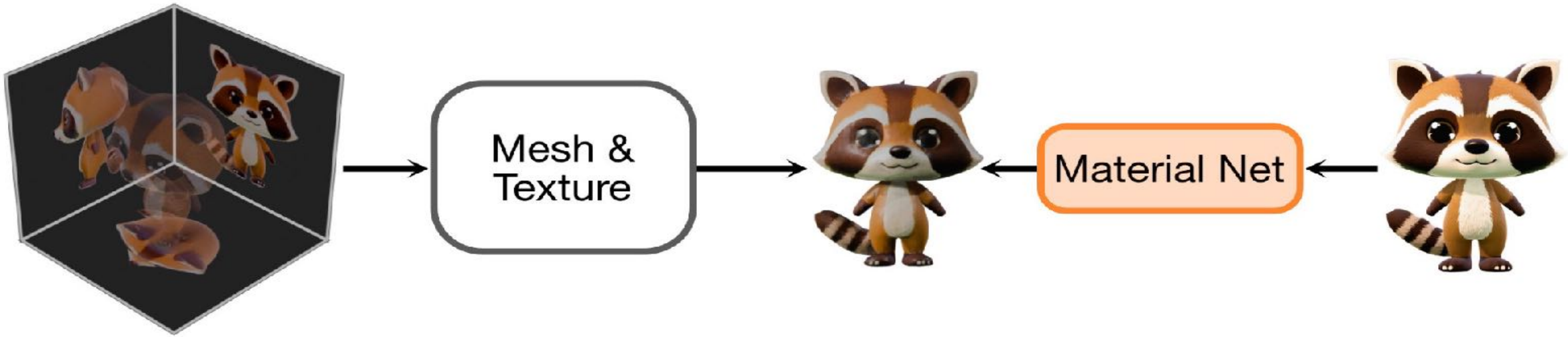
Overview



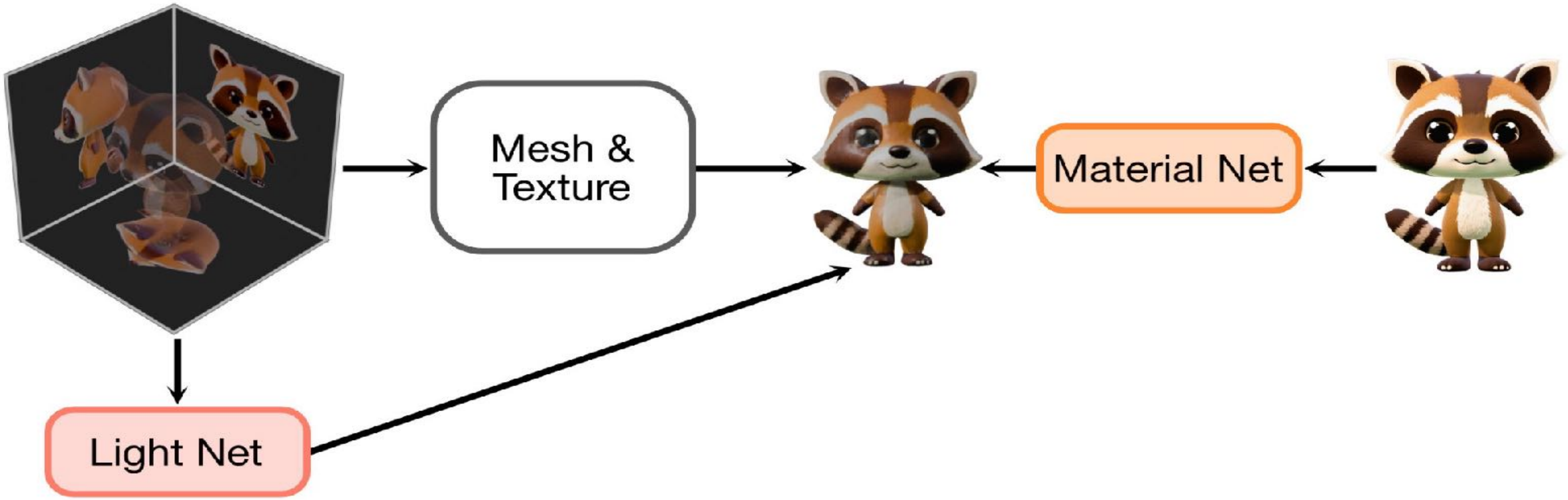
Overview



Overview



Overview



Aliasing Issues

- Previous methods used a low resolution triplane (64 x 64)
- We found that this result in grid artifacts
- These are aliasing issues (higher frequency)
- We predict high resolution 384 x 384 triplanes instead

Ground Truth

Low Resolution

Ours (High Resolution)

Comparison

TripoSR

InstantMesh

CRM

Ours (SF3D)

Comparison

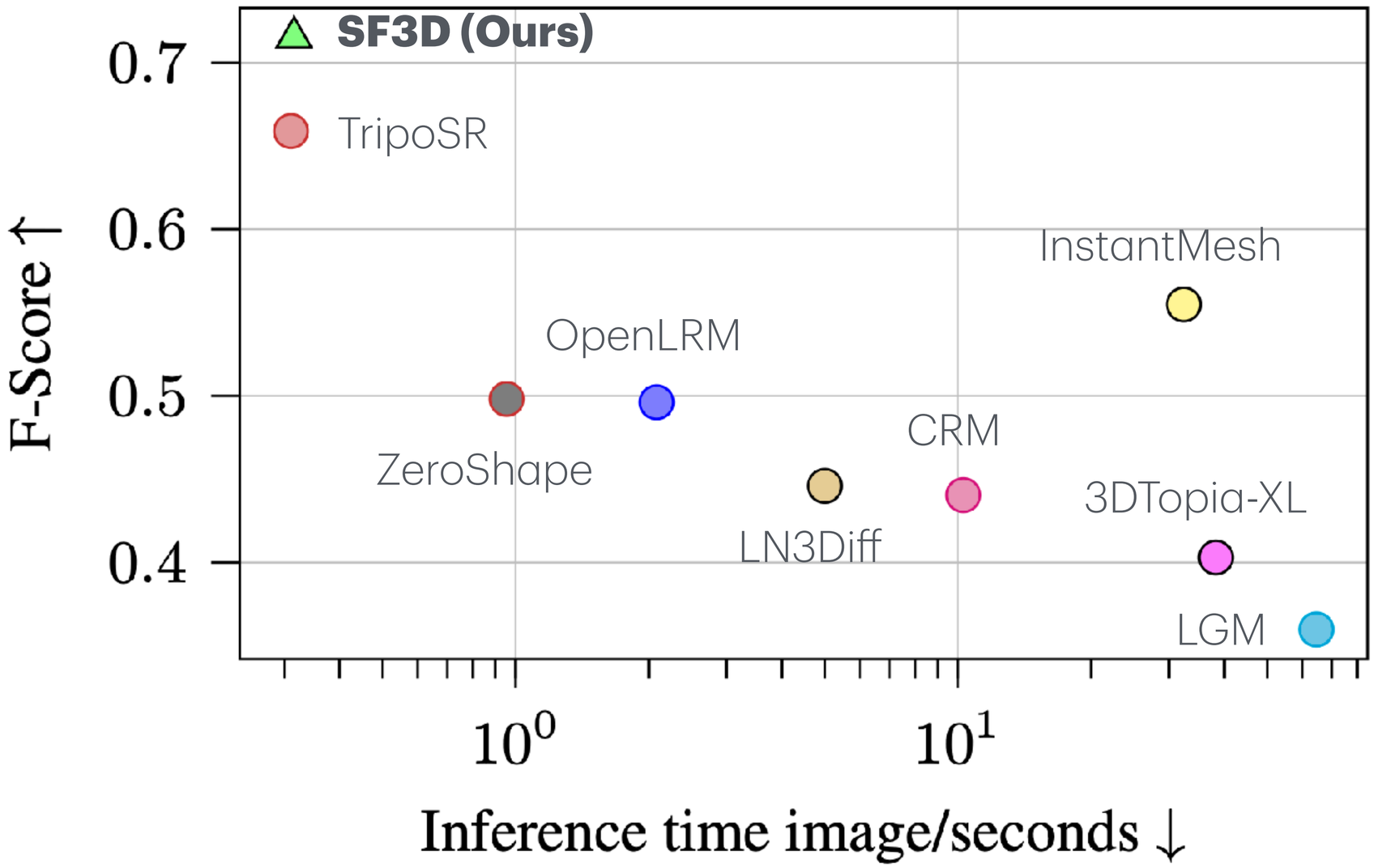
TripoSR

InstantMesh

CRM

Ours (SF3D)

Quantitative Comparison



Conclusion

- Highly efficient method in generating objects from a single image (~0.3s)
  - Fast UV unwrapping proposed in method
- Enhance triplane resolution helps in texture reproduction
- Explicit mesh extraction training helps in assets quality

Project Page

SPAR3D: Stable Point-Aware Reconstruction of 3D Objects from Single Images

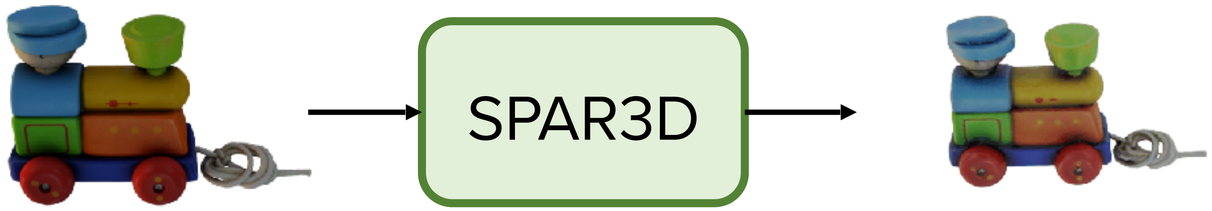
Zixuan Huang, Mark Boss, Aaryaman Vasishta, James Matthew Rehg, Varun Jampani

Saturday Afternoon Poster Session - Poster #99

Single Image To 3D

- Fast Image to 3D ~0.7s
- Based on Stable Fast 3D
- Several new key contributions

Regression-based and Generative Modelling



Regression-based methods:

+ Fast and align well to input

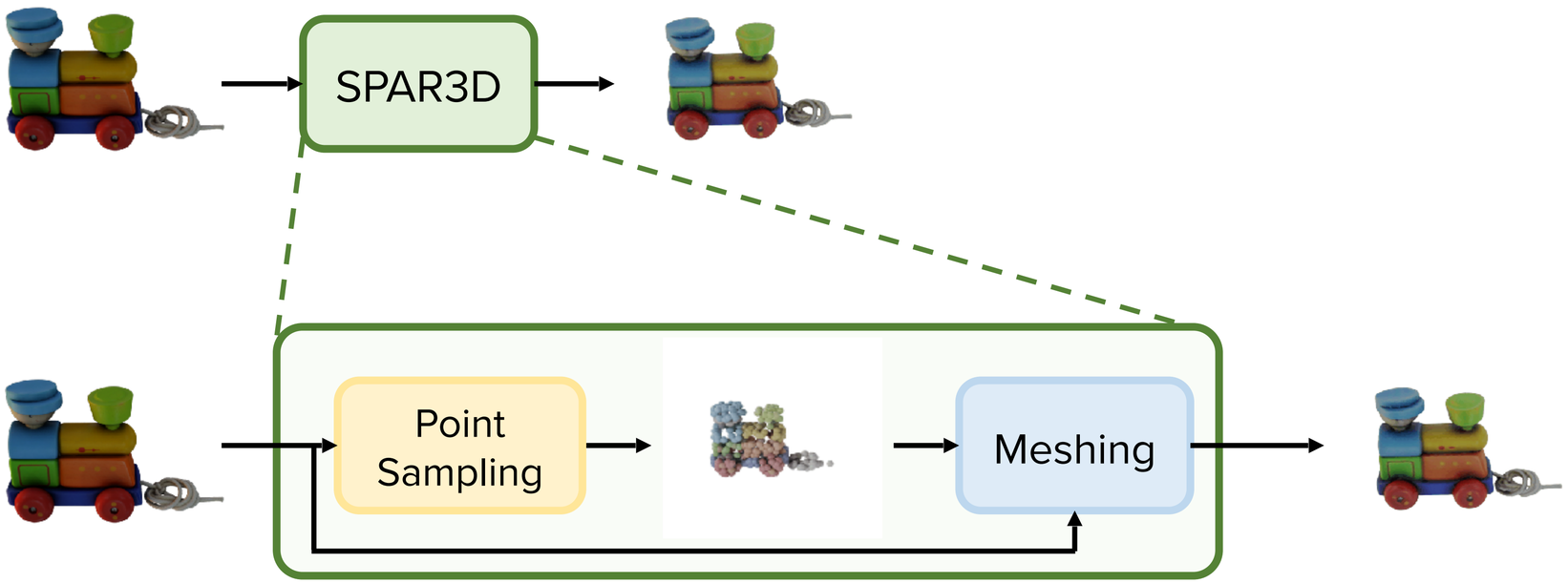
- Oversmoothed back surface

Diffusion-based methods:

+ Better back surface

- Slow and low output fidelity

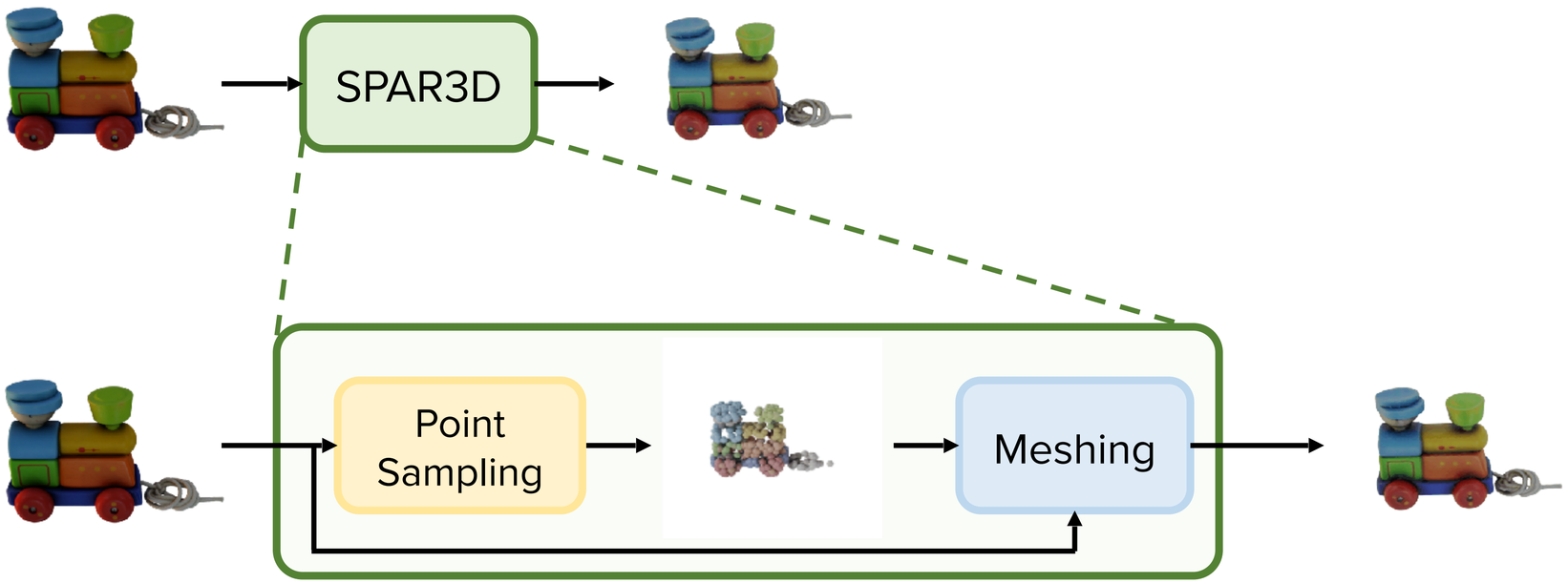
Sparse Point Clouds as Bridge



SPAR3D aims to take the best of regression and diffusion-based modeling:

- Point sampling generates a sparse point cloud via point diffusion
- Meshing creates a detailed mesh from the point cloud and the image

Sparse Point Clouds as Bridge

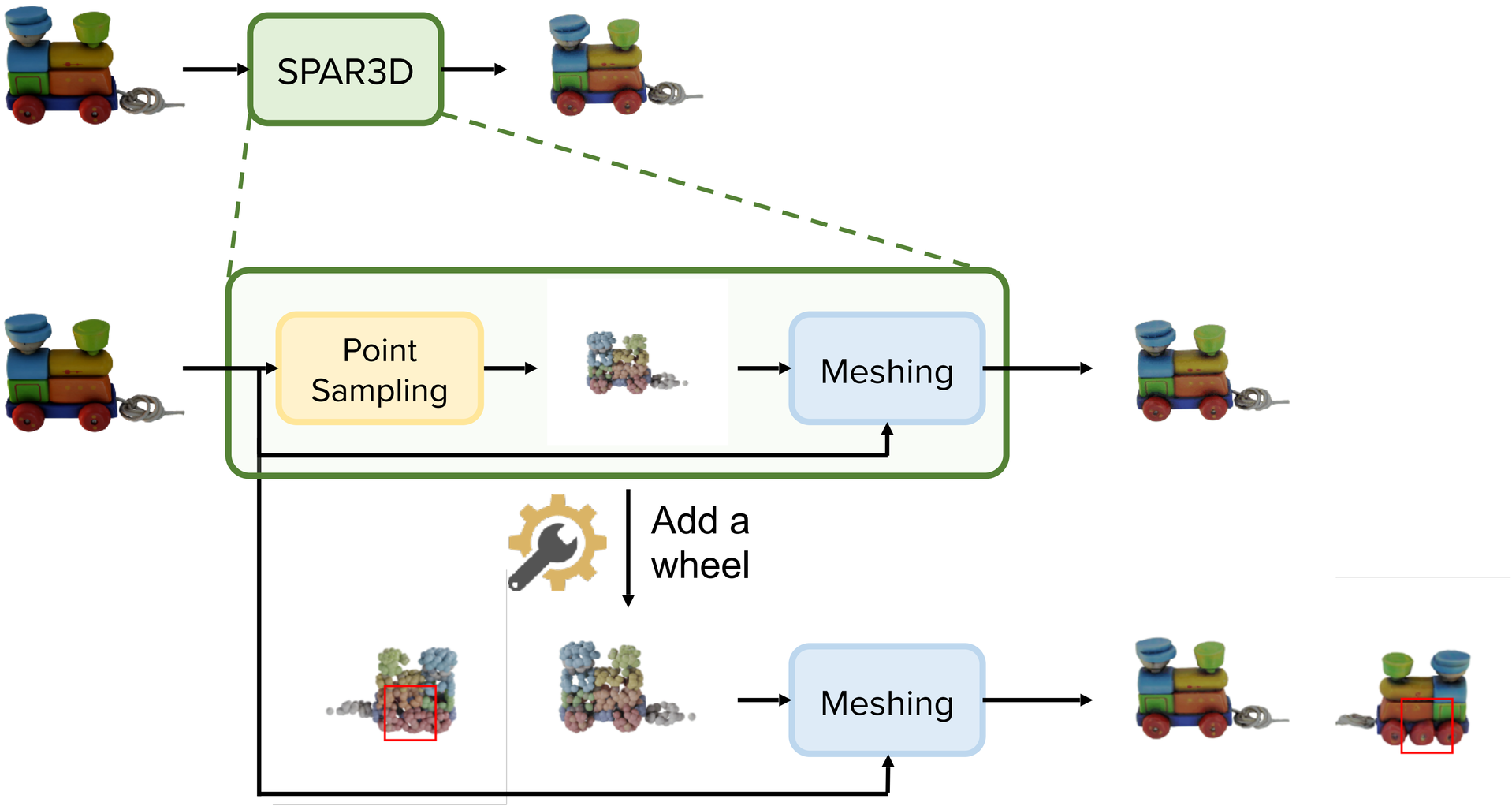


+ Fast: sparse point generation and efficient feedforward meshing

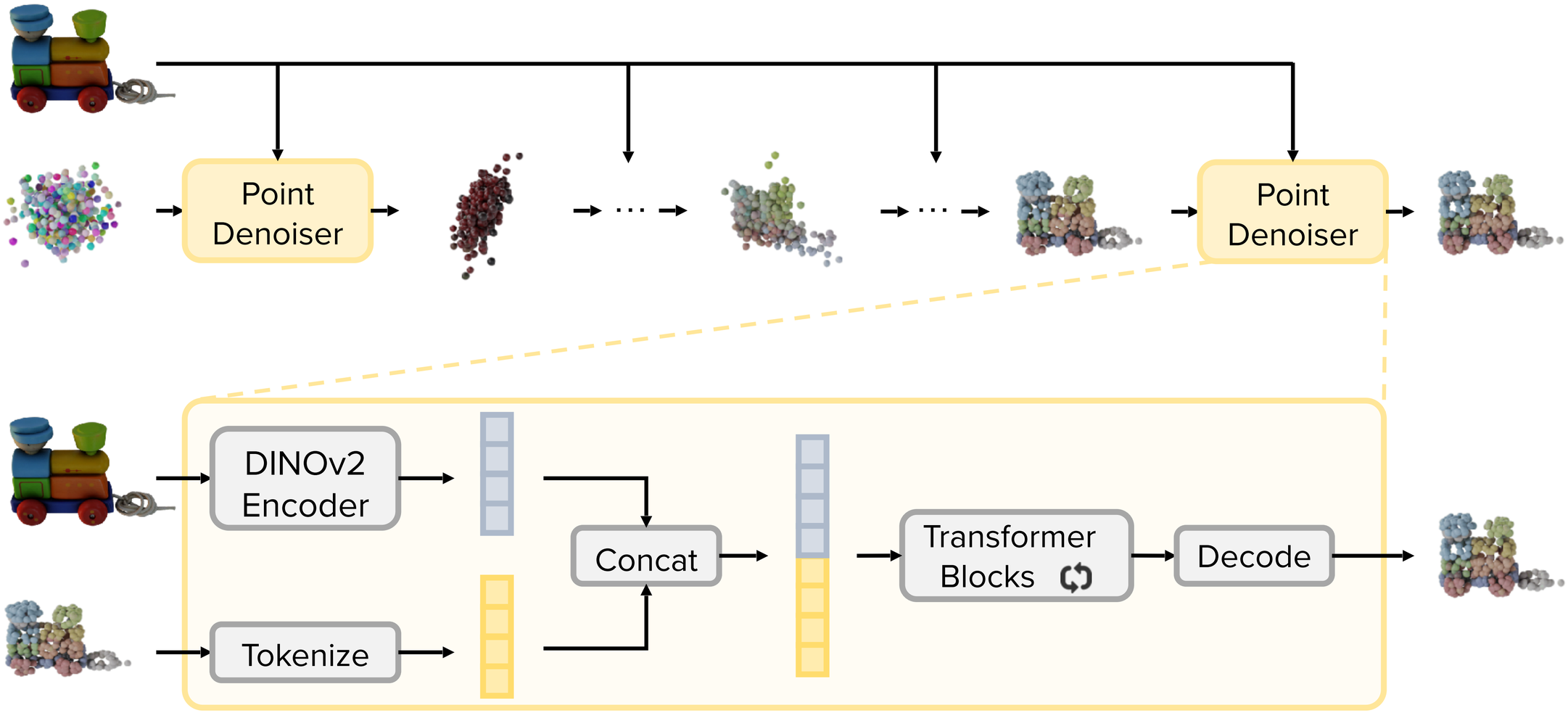
+ Better back surface: point sampling reduce back surface uncertainty

+ High output fidelity: meshing use image features to adjust visible surface

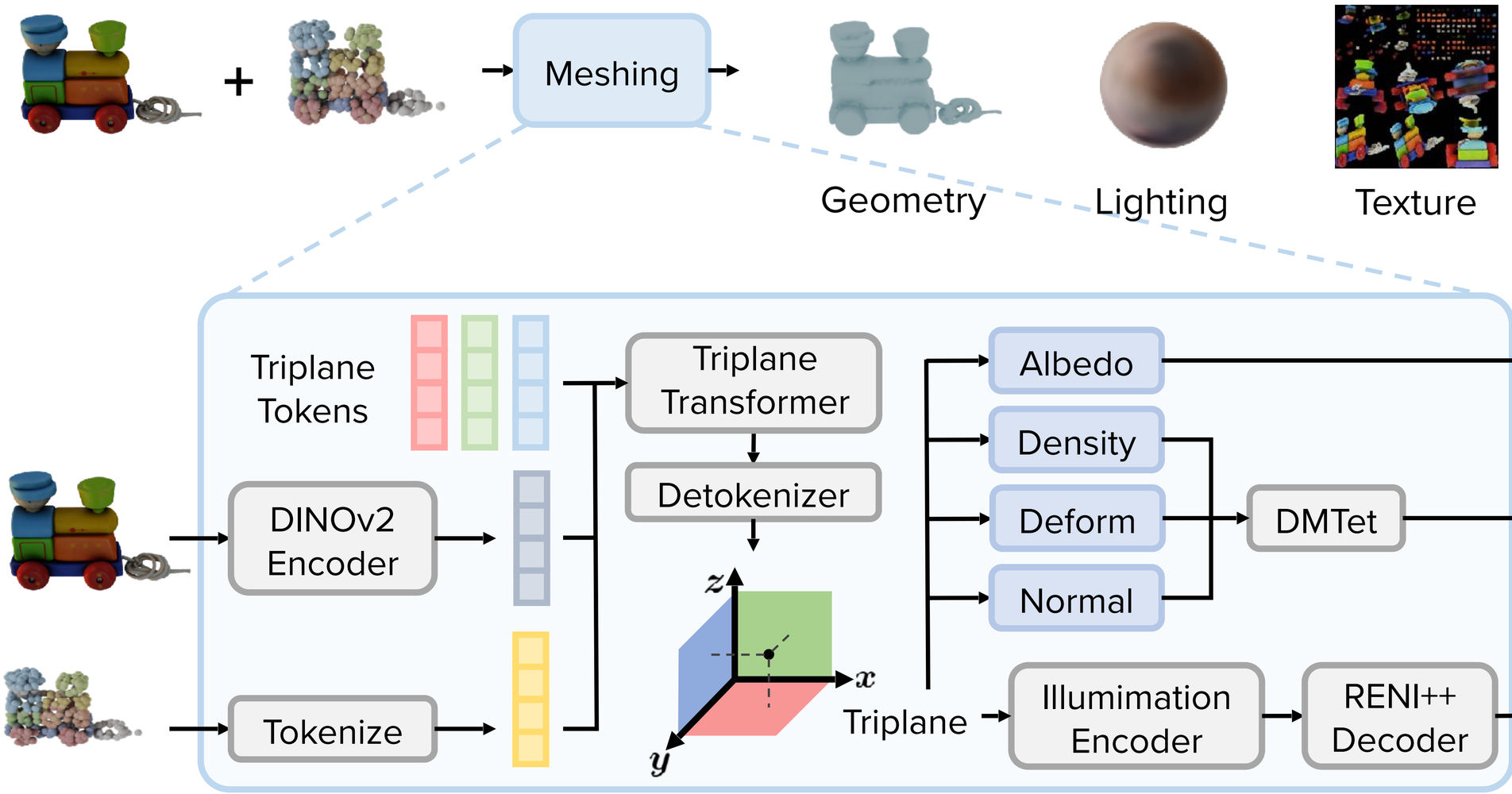
Interactive Edits with SPAR3D



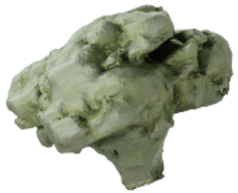
Point Sampling via Diffusion



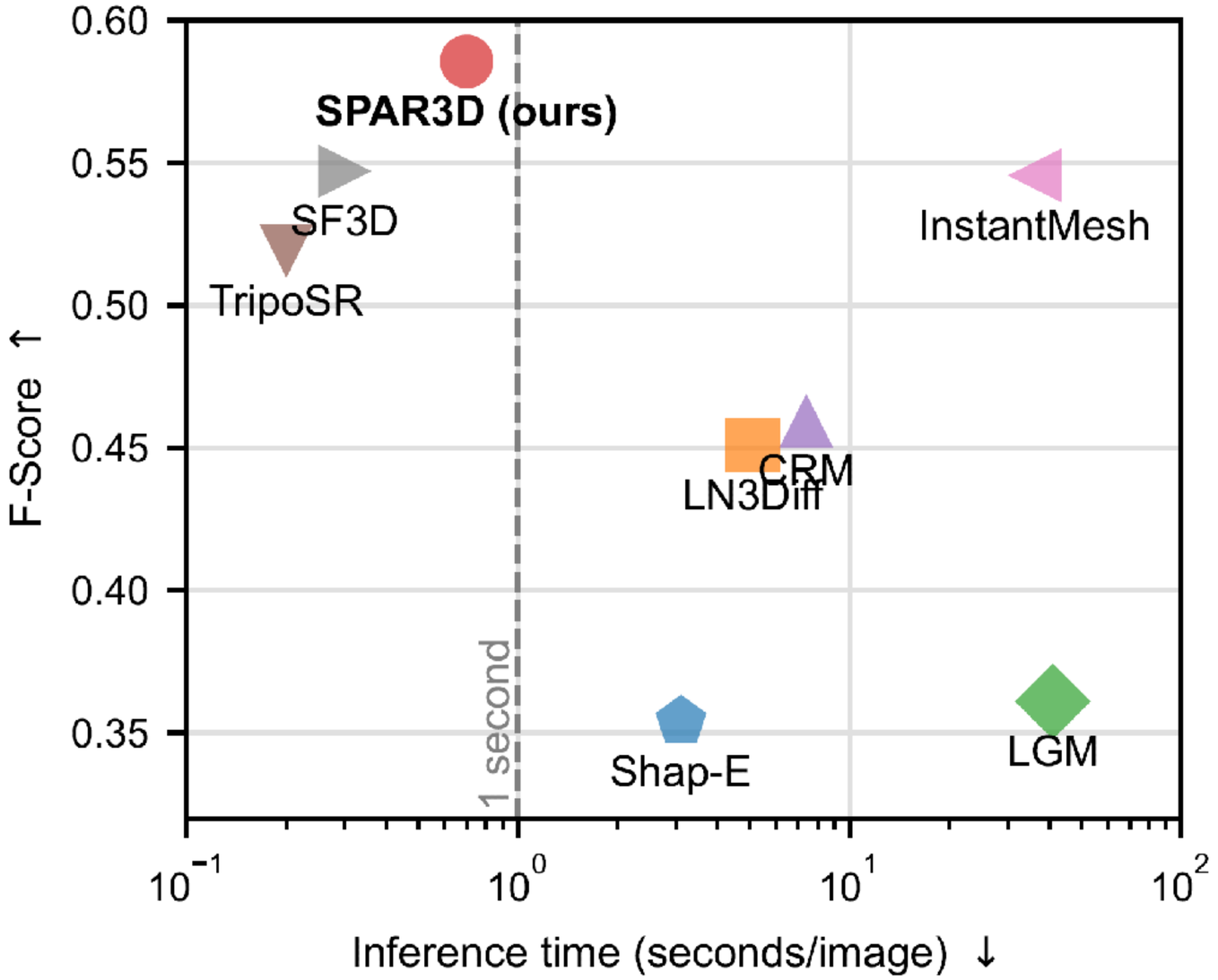
Meshing with Large Triplane Transformer



Qualitative Comparison

	CRM	TripoSR	InstantMesh	SF3D	**Ours**
					
					
					

Quantitative Comparison



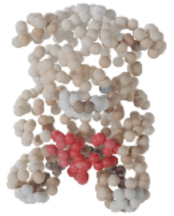
Generalization to In-the-wild Images

Editing Examples

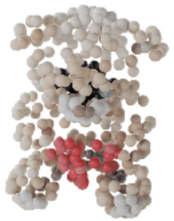


Elongate
the nose

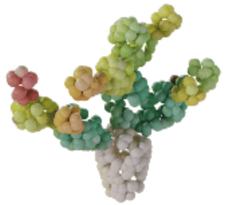


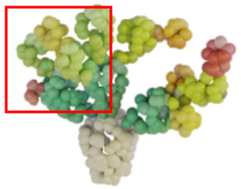
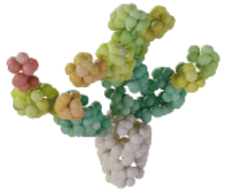

Add a hat
on the back


Editing Examples


Re-attach
floating parts



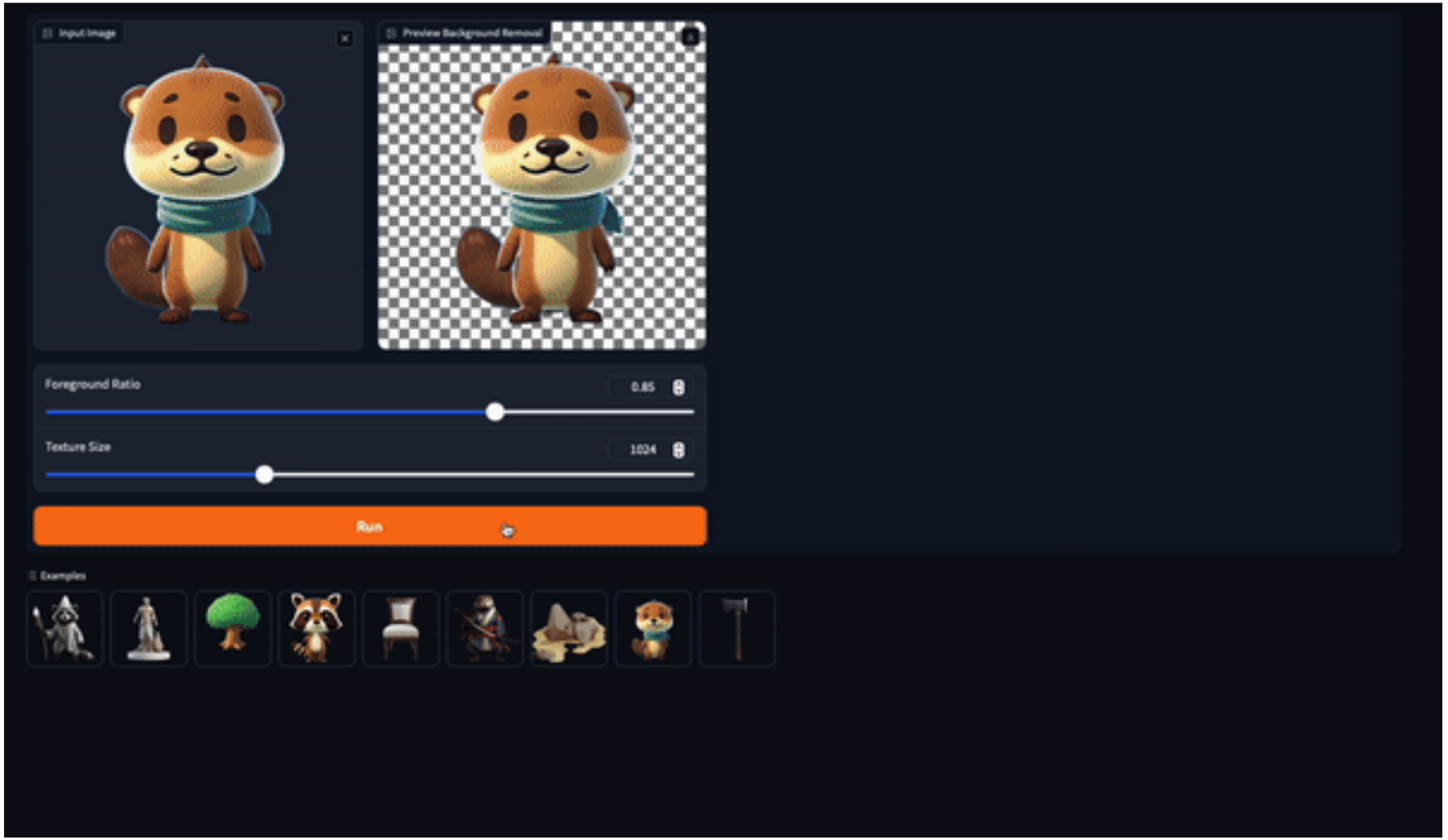

Change
minor details



Conclusion

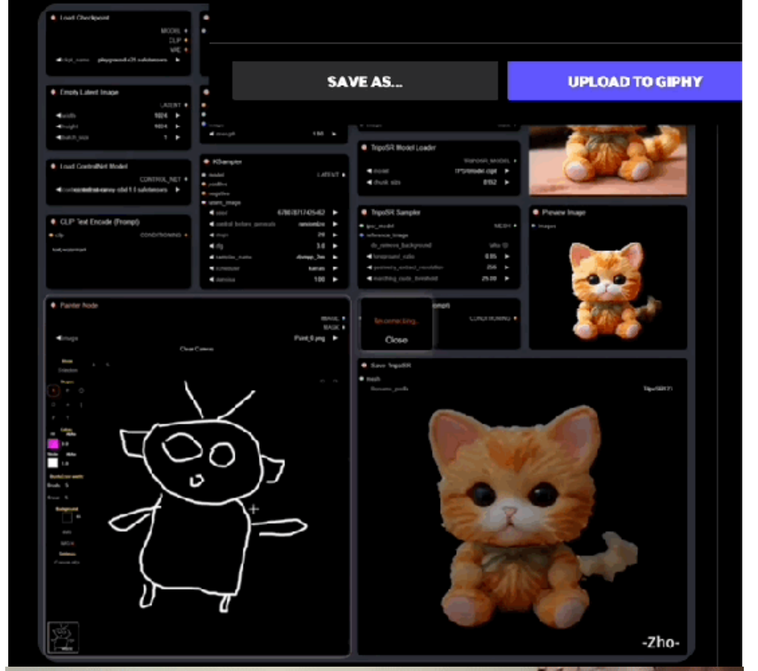
- SPAR3D -- SOTA 3D object reconstructor from single-view images:
  - Two-stage design inherits the benefit of regression and diffusion methods
  - Fast reconstruction speed under 1 second per image
  - Intermediate point clouds facilitate interactive editing

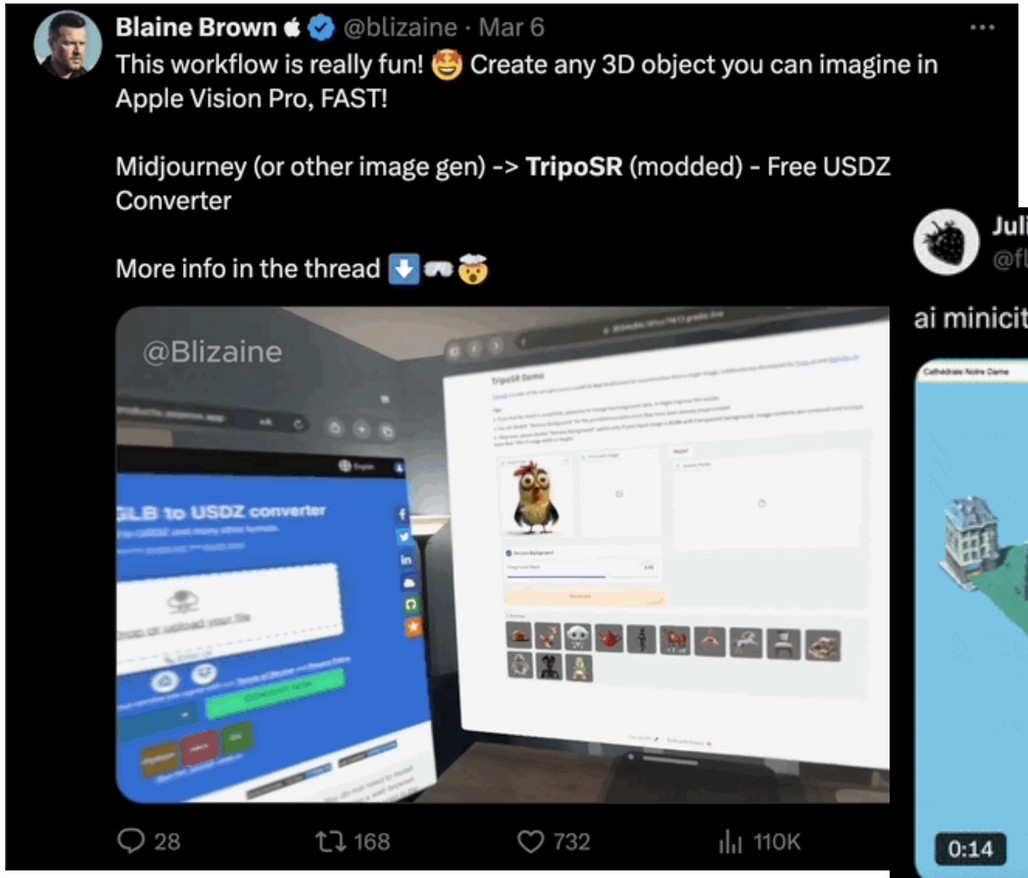
Demo - SF3D


Input Image
Preview Background Removal
Foreground Ratio
0.85
Texture Size
1024
Run
Examples

Quick adoption in the community

Several interesting use cases and workflows





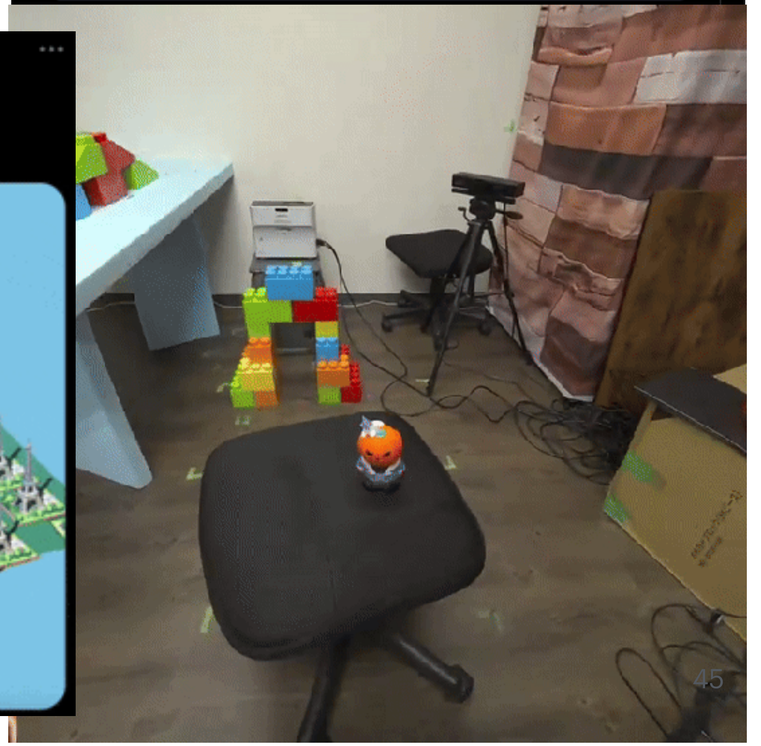


Image-based Exploration

Project Page

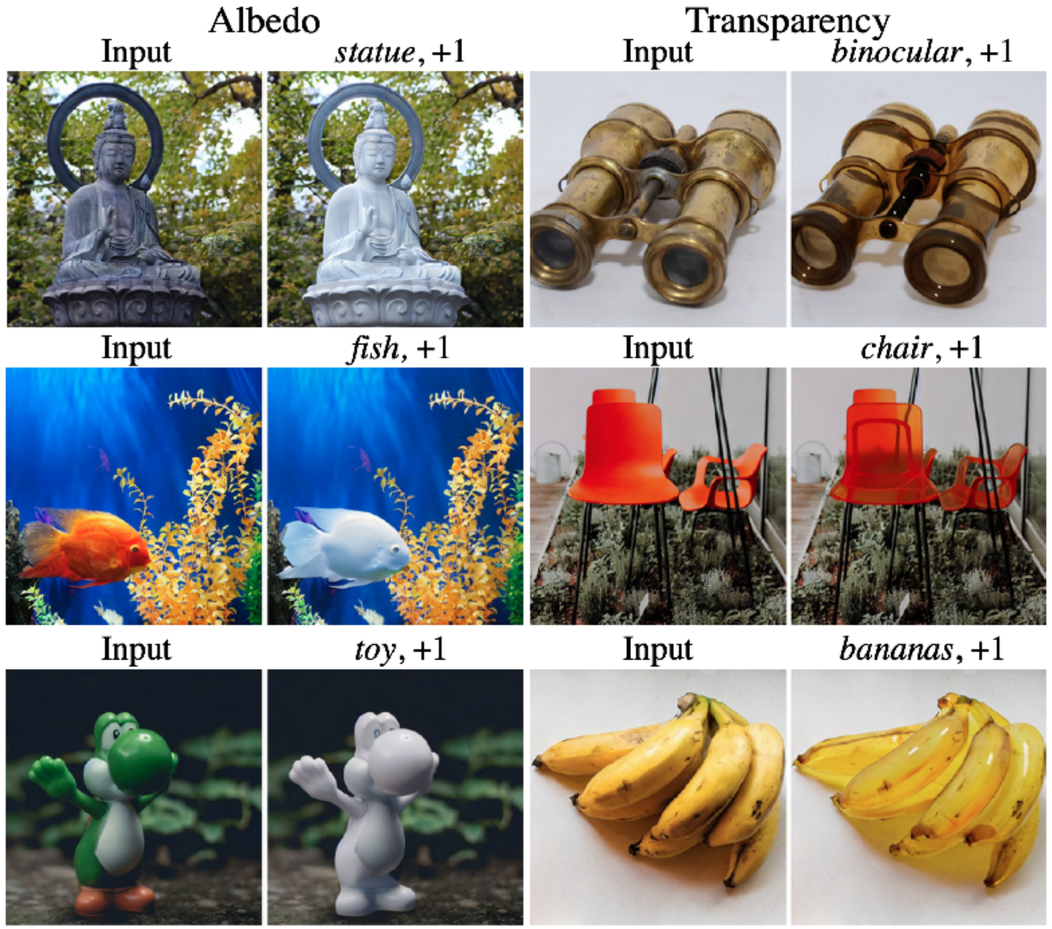
MARBLE: Material Recomposition and Blending in CLIP-Space

Ta-Ying Cheng, Prafull Sharma, Mark Boss, Varun Jampani

Saturday Morning Poster Session - Poster #230

Material Control in Images…

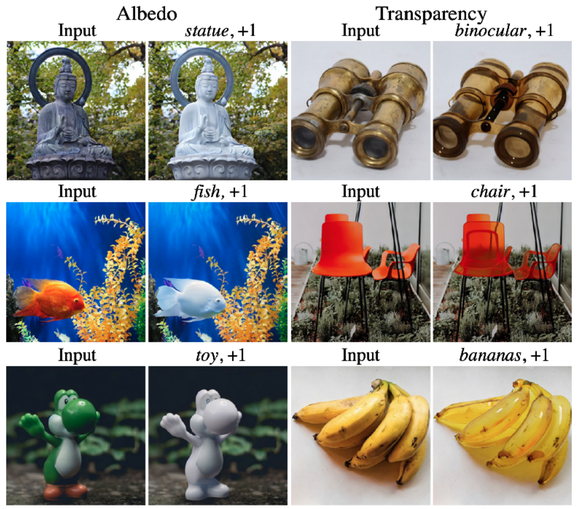
Parametric Controls



Exemplar-Based Control

Previous Works

Parametric Controls



Exemplar-Based Control

- Parametric Control
  - Finetune the entire generative model architecture
  - Overfitting on synthetic data, destroying prior knowledge of generative models
- Zero-Shot Exemplar-Based Control:
  - Only allows coarse control and does not allow parametric tuning

Can we perform a series of material controls on diffusion models by manipulating the CLIP-Space alone?

Targeted Material Block Injection

Injecting into individual blocks in Denoising UNet

Material

Context

All Blocks (ZeST)

Down, 1, 0

Down, 1, 1

Down, 2, 0

Down, 2, 1

Up, 1, 0

Up, 1, 1

Up, 1, 2

Up, 0, 0

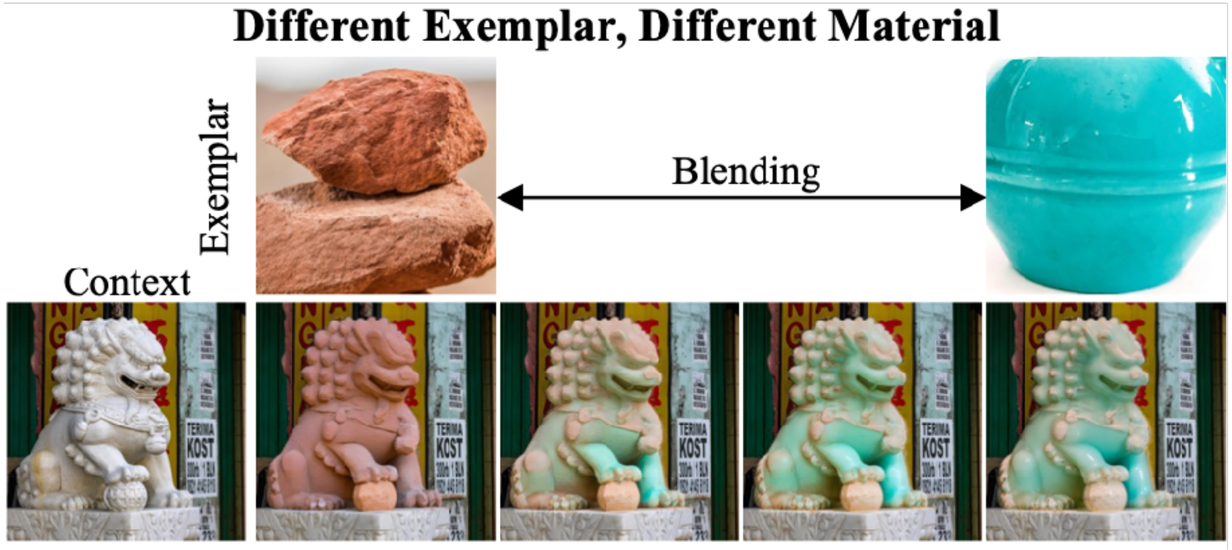
Up, 0, 1 (Material Block)

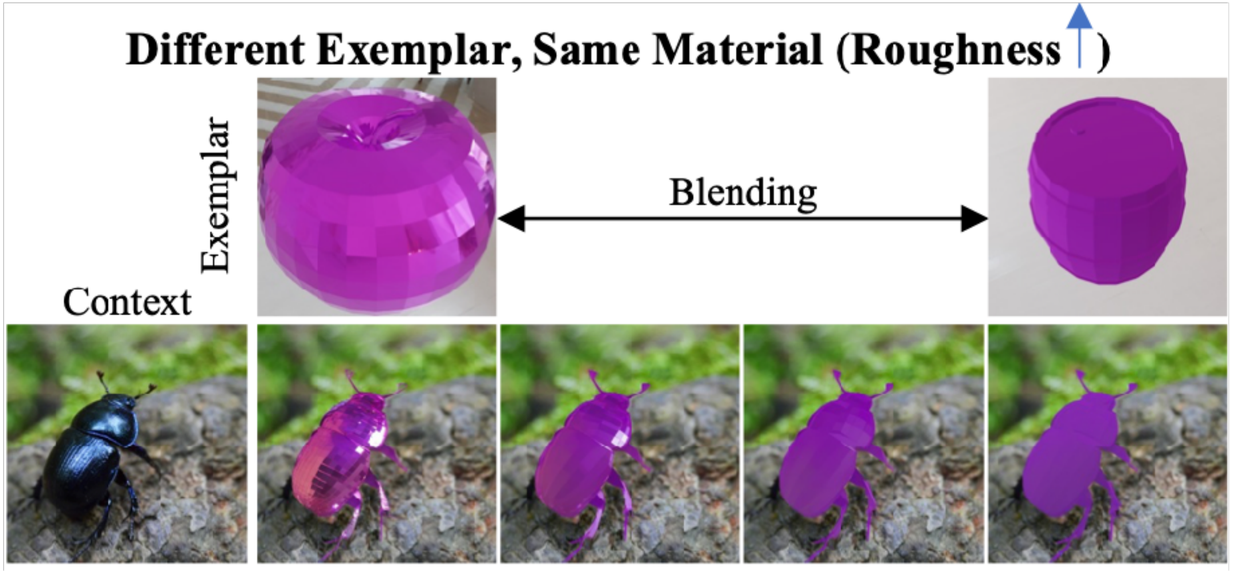
Up, 0, 2

Targeted Material Block Injection

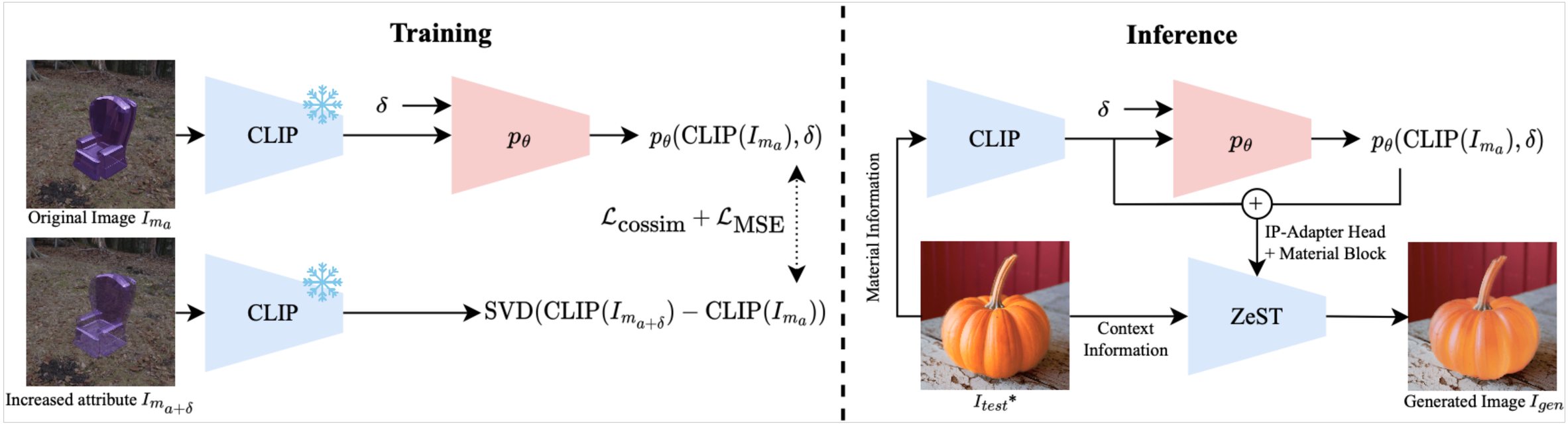
Blending Two Exemplars in CLIP Space

Material Blending Results

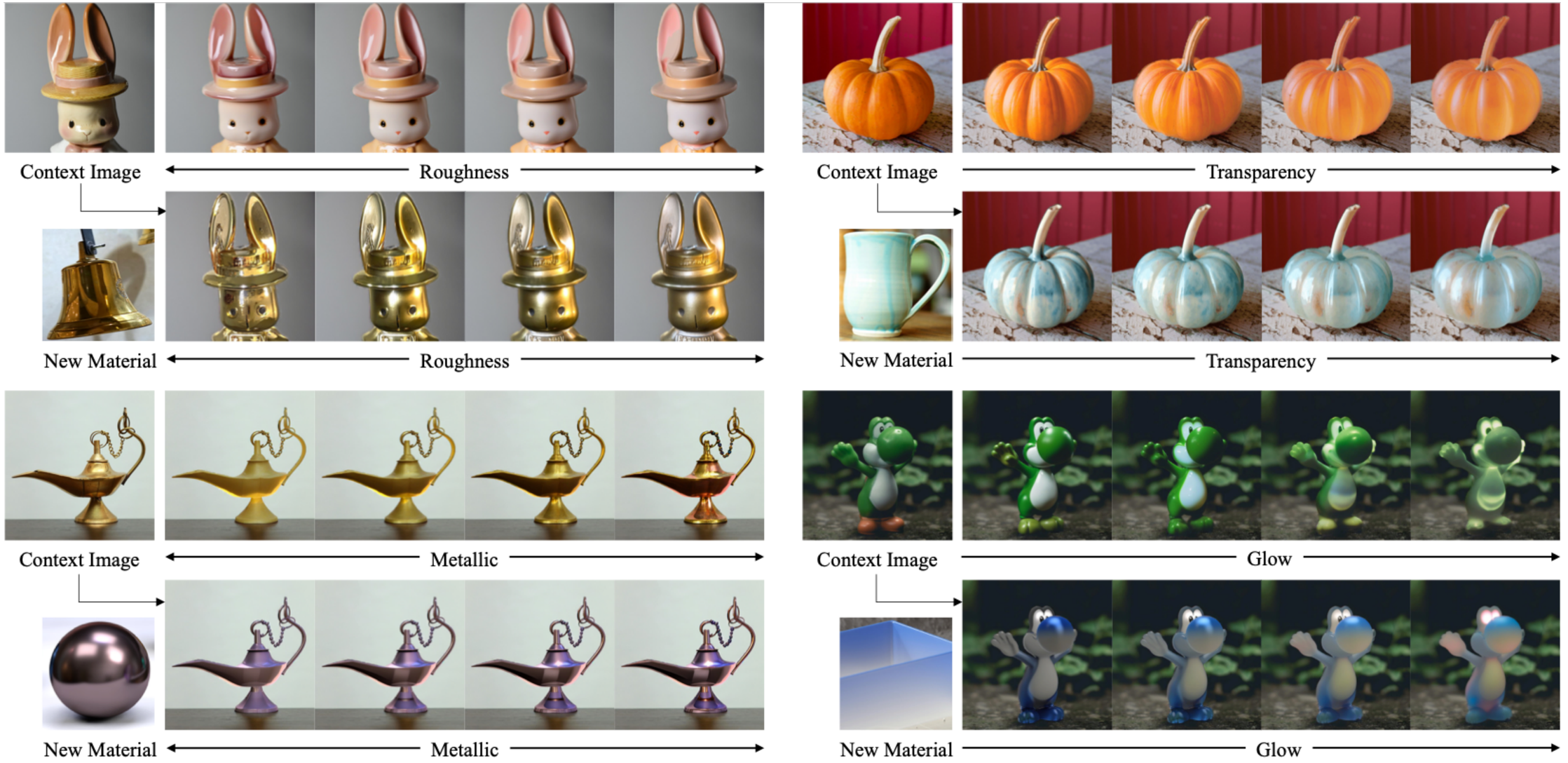




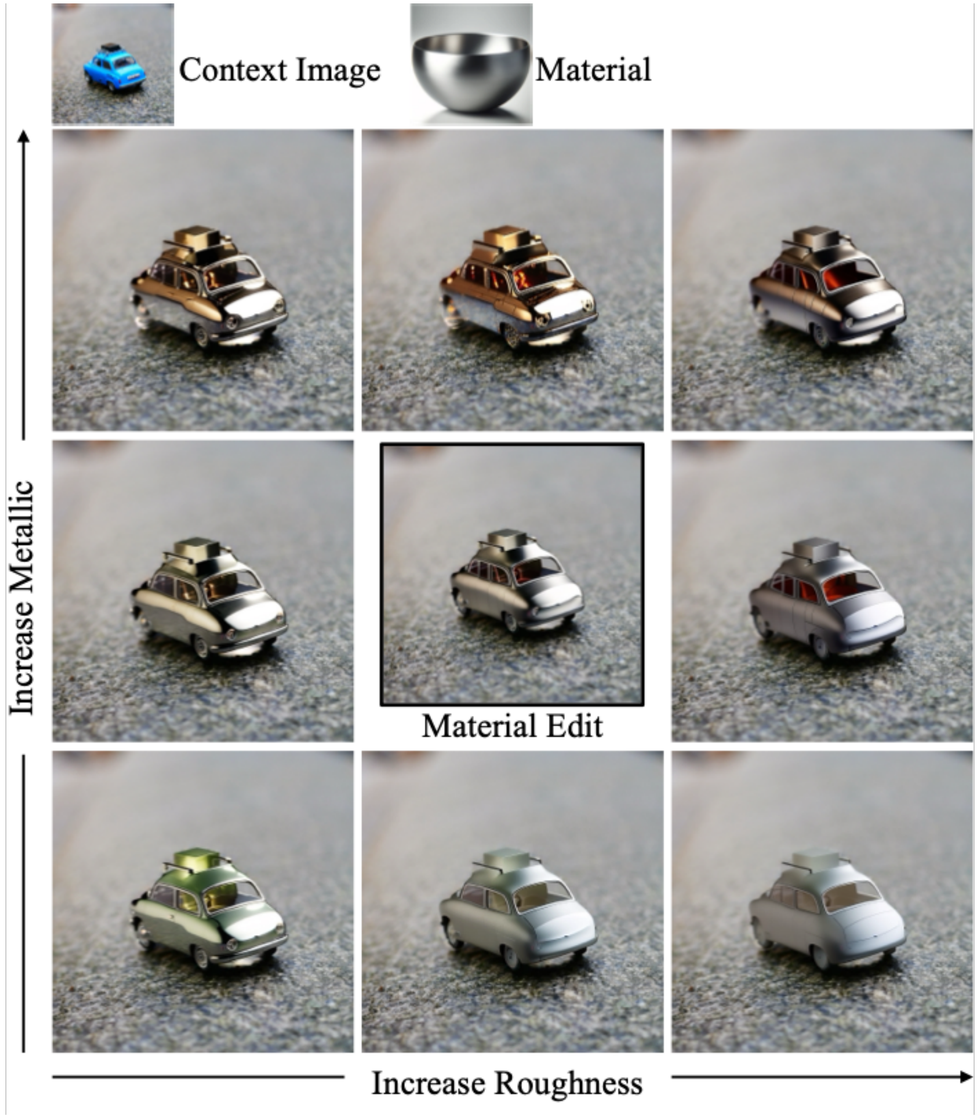
Parametric Control – Architecture



Parametric Control Results



Multiple Controls at Once



Parametric Control in Style

Increase in Transparency

'A teapot on a table, Van Gogh painting'

Increase in Roughness

'A teapot on a table, neon cyberpunk style painting'

Conclusion

- Wide range of novel editing controls
- Operates only in CLIP-space
- Robustness to various styles

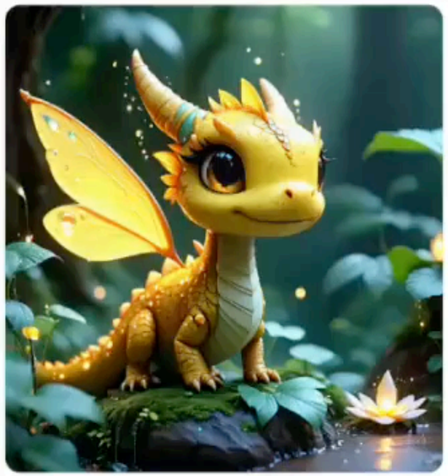
Project Page

Stable Virtual Camera: Generative View Synthesis with Diffusion Models

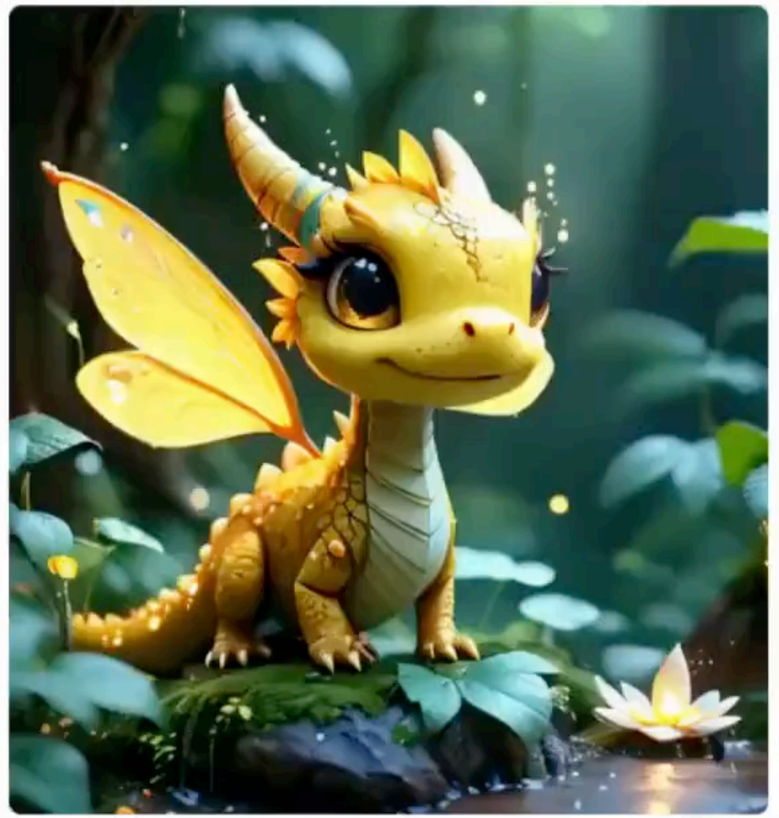
Jensen Zhou, Hang Gao, Vikram Voleti, Aaryaman Vasishta, Chun-Han Yao, Mark Boss, Philip Torr, Christian Rupprecht, Varun Jampani

NVS as a Video Generation

Input: Observed View



Output: Novel Views

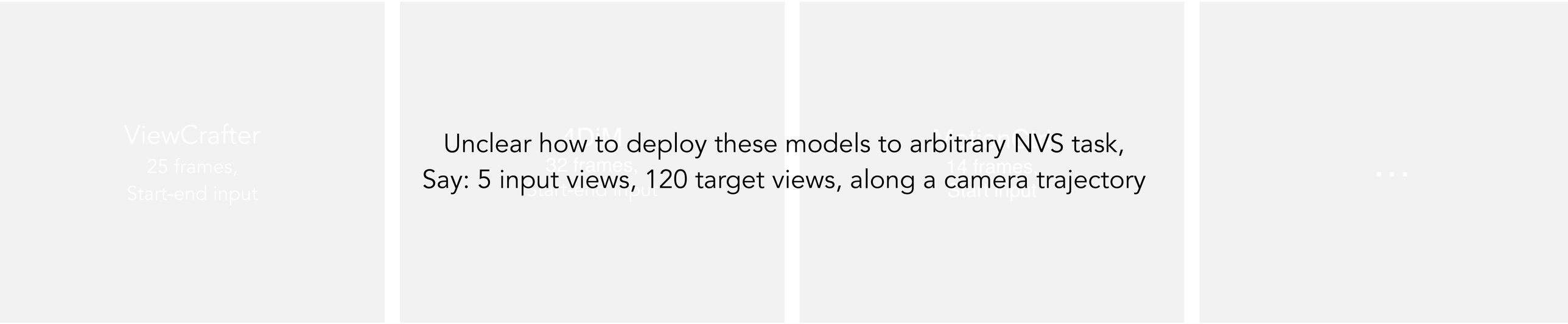


Challenges

Interpolation Smoothness

CAT3D, Gao et al., 2024 (Our reproduction)

Challenges



Unclear how to deploy these models to arbitrary NVS task,
Say: 5 input views, 120 target views, along a camera trajectory

Baked-in Task Assumptions

ViewCrafter, Yu et al., 2024; 4DiM, Watson et al., 2024; MotionCtrl, Wang et al., 2023

Goal

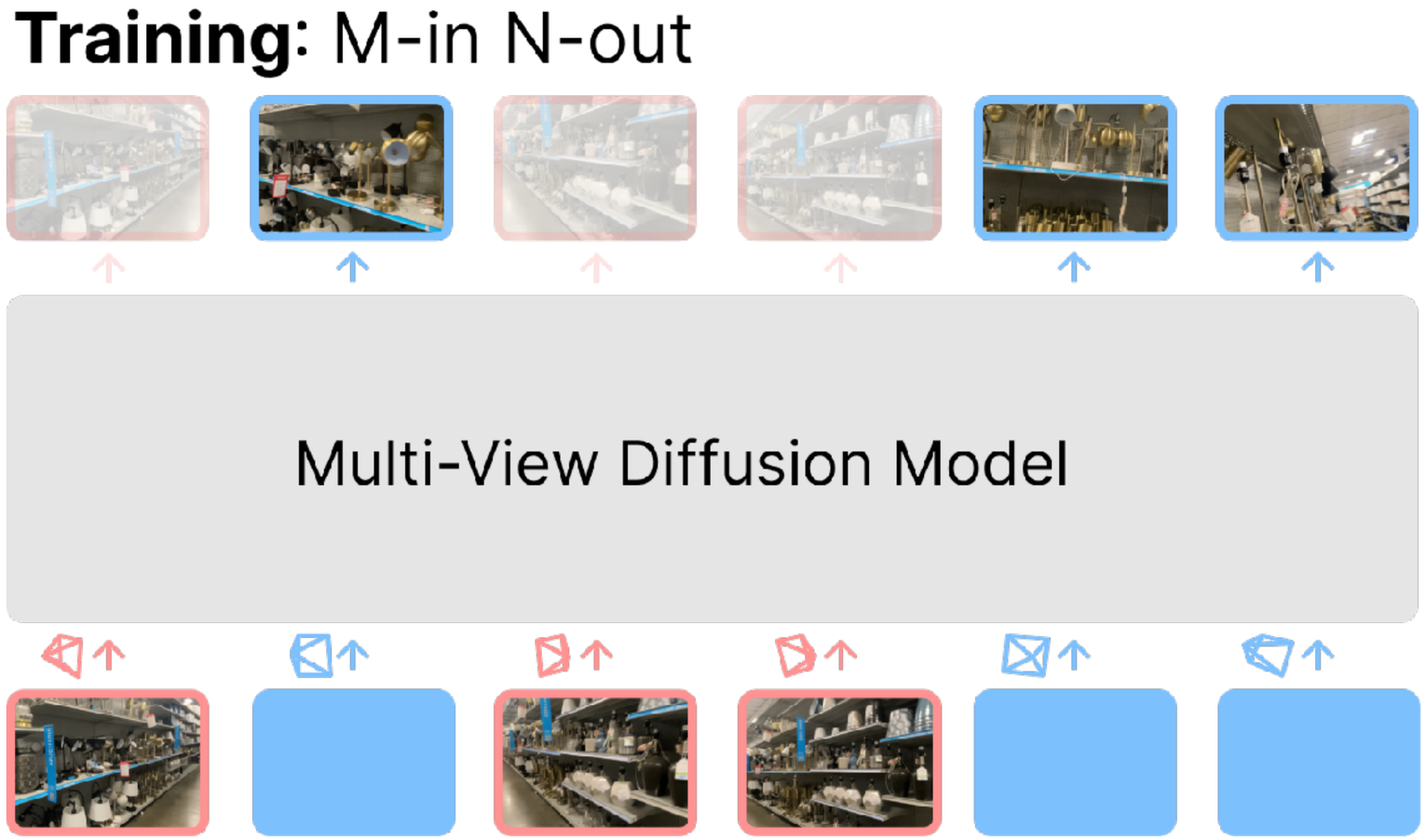
Create a single model that has

High Generation Capacity

Good Interpolation Smoothness

Versatility to any NVS task

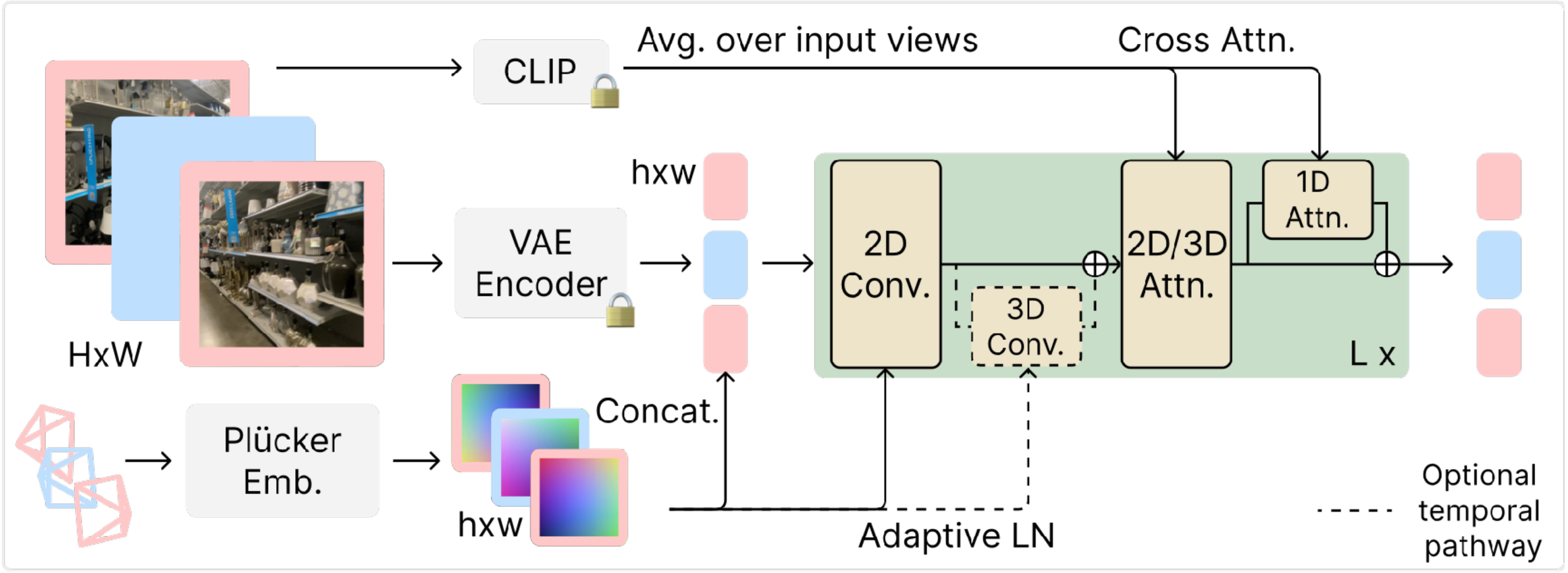
Model Architecture



Input

Target

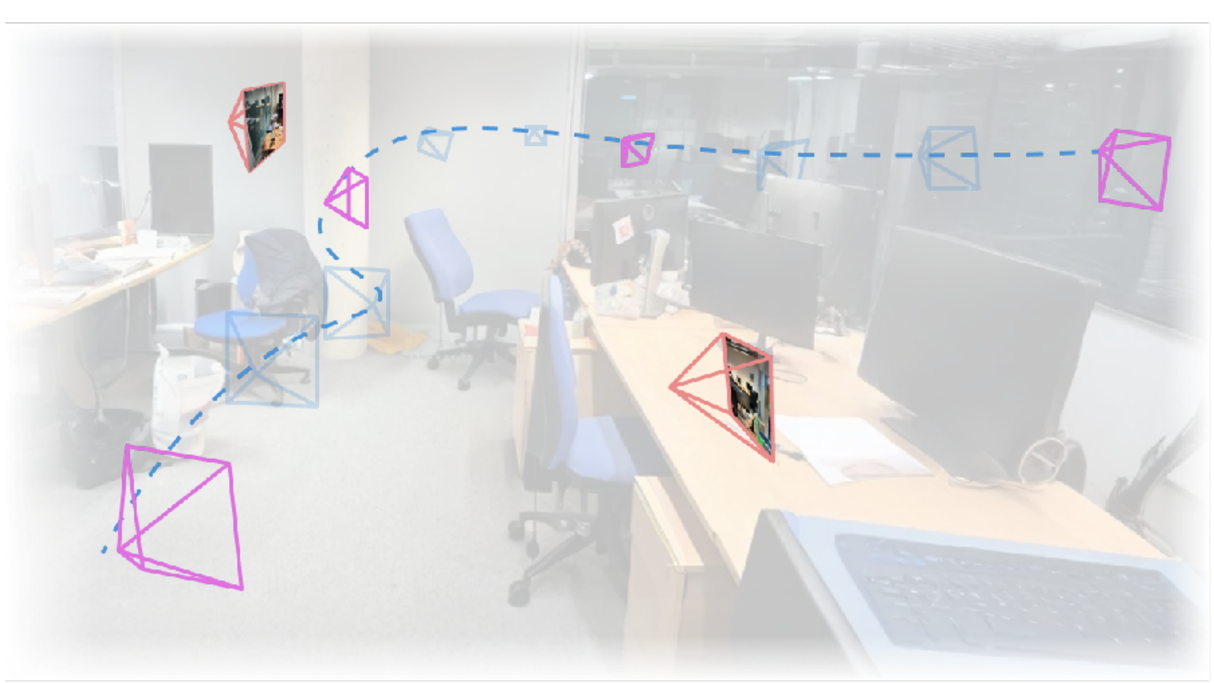
Model Architecture



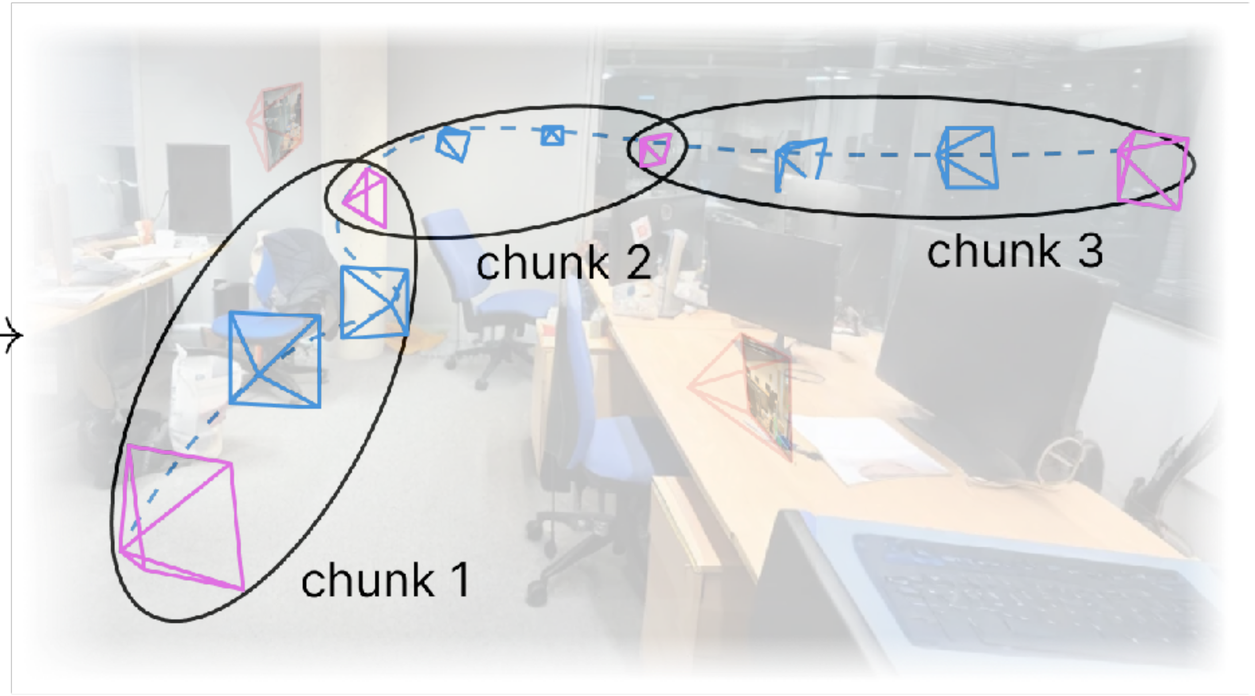
Input

Target

Procedural two-pass sampling



Anchor Generation



Target Generation

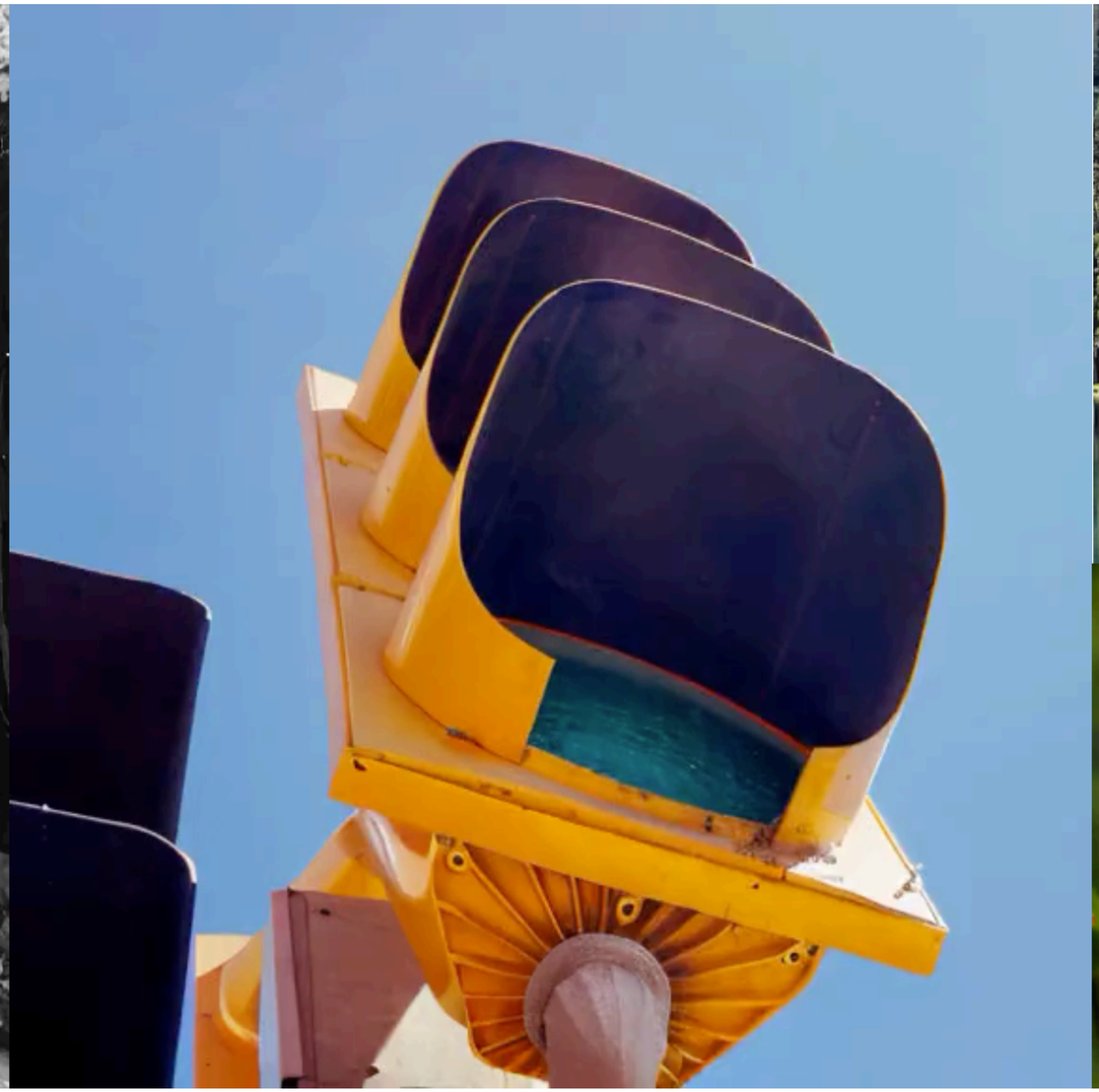
Input

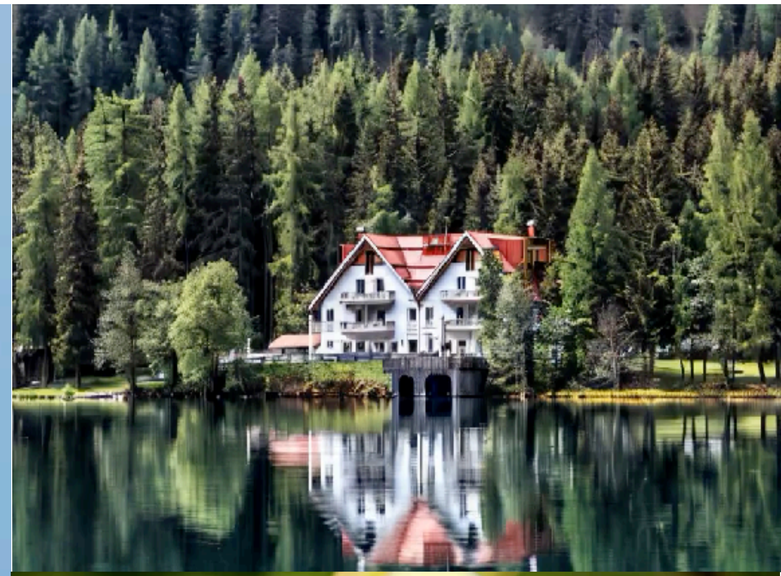
Target

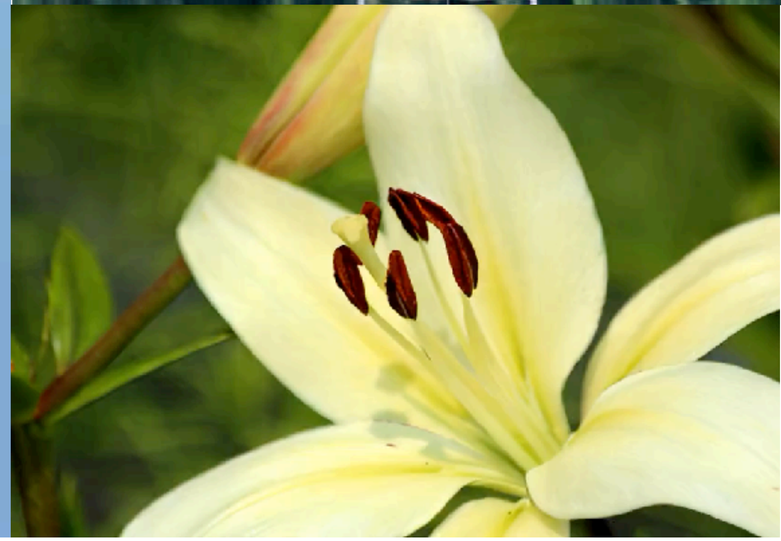
Anchor

Single Image NVS

Single Image NVS





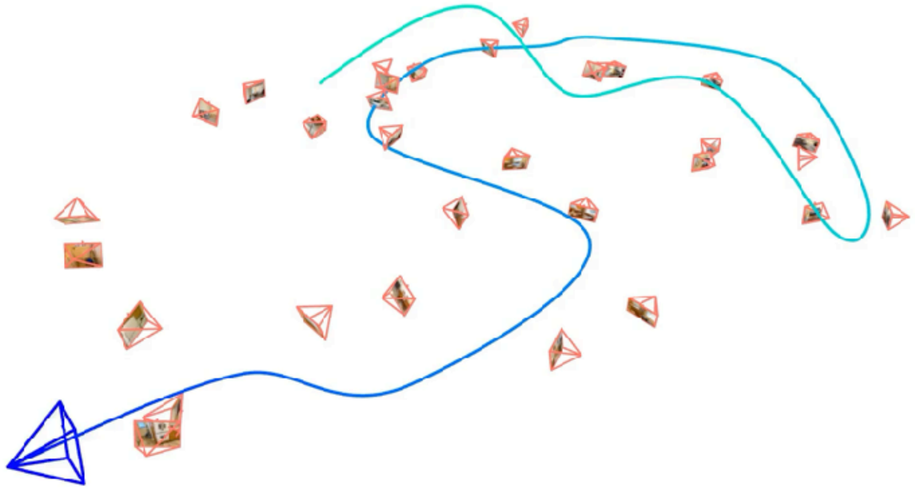


Sparse-view NVS

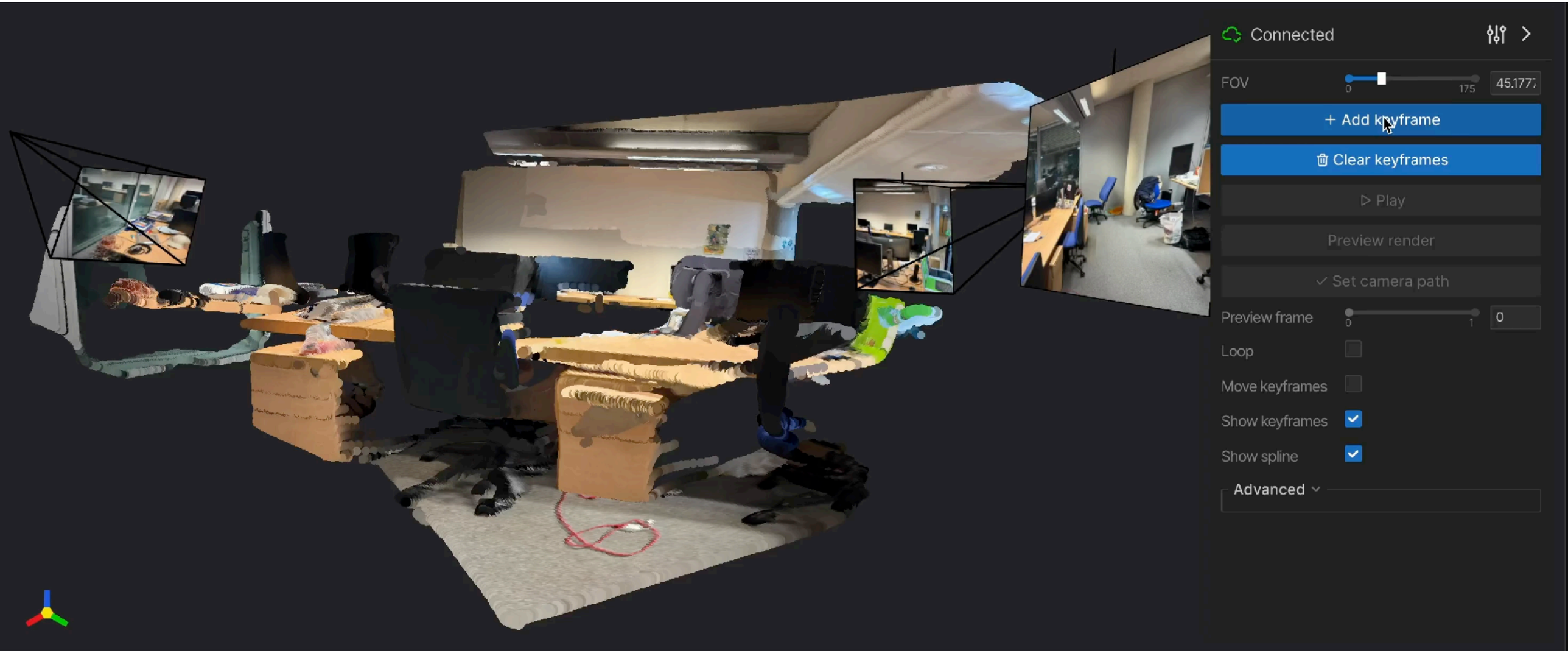
Open-world NVS



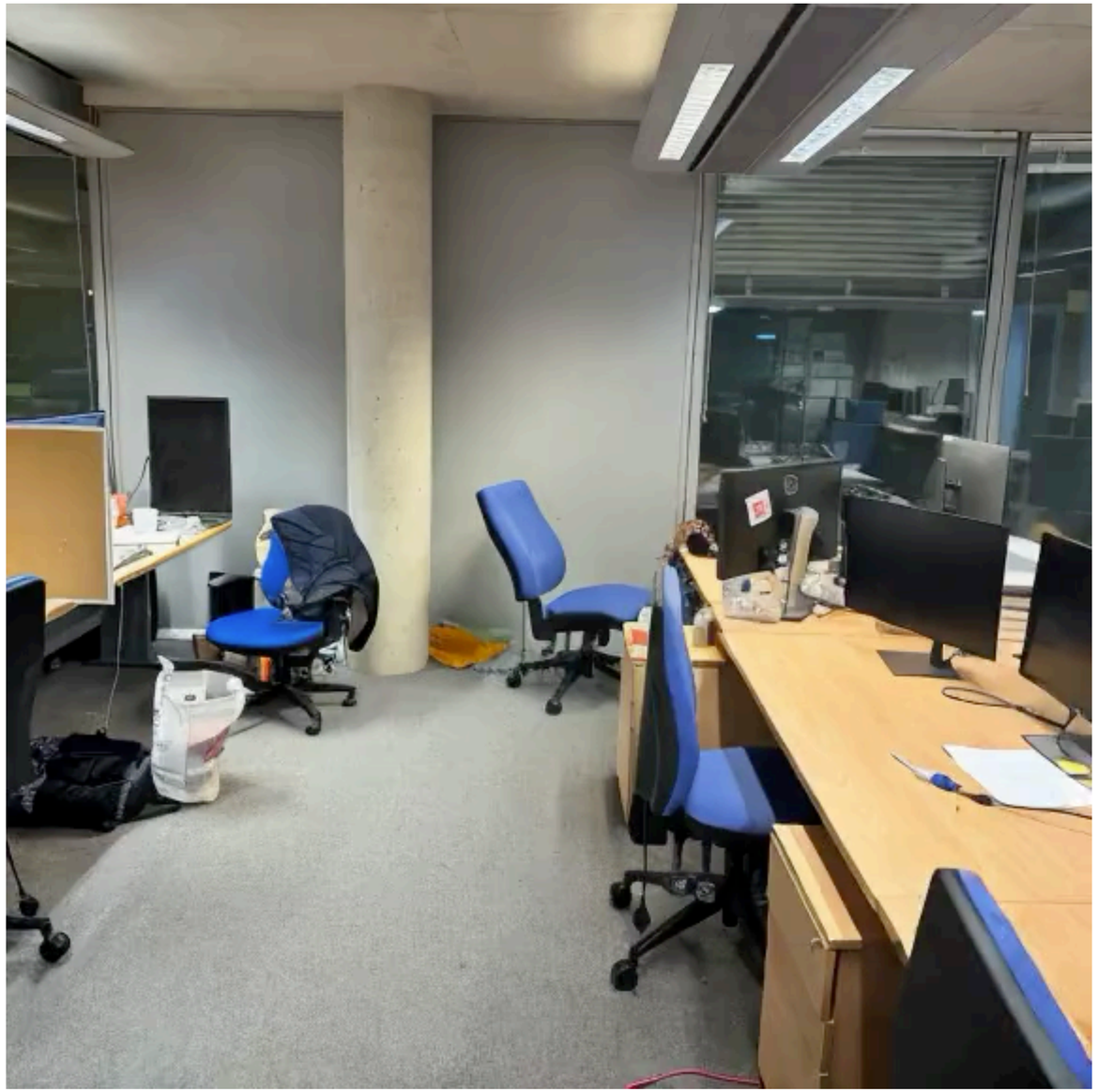
...



Demo for any camera trajectory



Demo for any camera trajectory



Comparison with prior works

ViewCrafter

CAT3D (our repro)

Ours

Limitations

Human

Animal

Dynamic texture

Conclusion

- A single versatile novel view synthesis network
  - High generation capacity with long video generation
  - Good interpolation smoothness
  - Versatility in terms of input view conditioning

Conclusion

Summary

- 3D and VFX production pipelines are long and complex
- Several subtasks are highly interesting as well
- Techniques covered:
  - Enable rapid prototyping in 3D (SF3D & SPAR3D)
  - Exploration from images (MARBLE & Stable Virtual Camera)

  Outlook

  - More control in the 3D generation and image-editing
  - Reaching post-production quality levels

Feel free to ask questions