Multi-Modal Generative AI with Foundation Models

Ziwei Liu 刘子纬

Nanyang Technological University

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE | S-LAB FOR ADVANCED INTELLIGENCE

2023

By ~~2027~~, creators won't have to be technical, just creative, thanks to automation tools.

AI-Generated Content

Movie

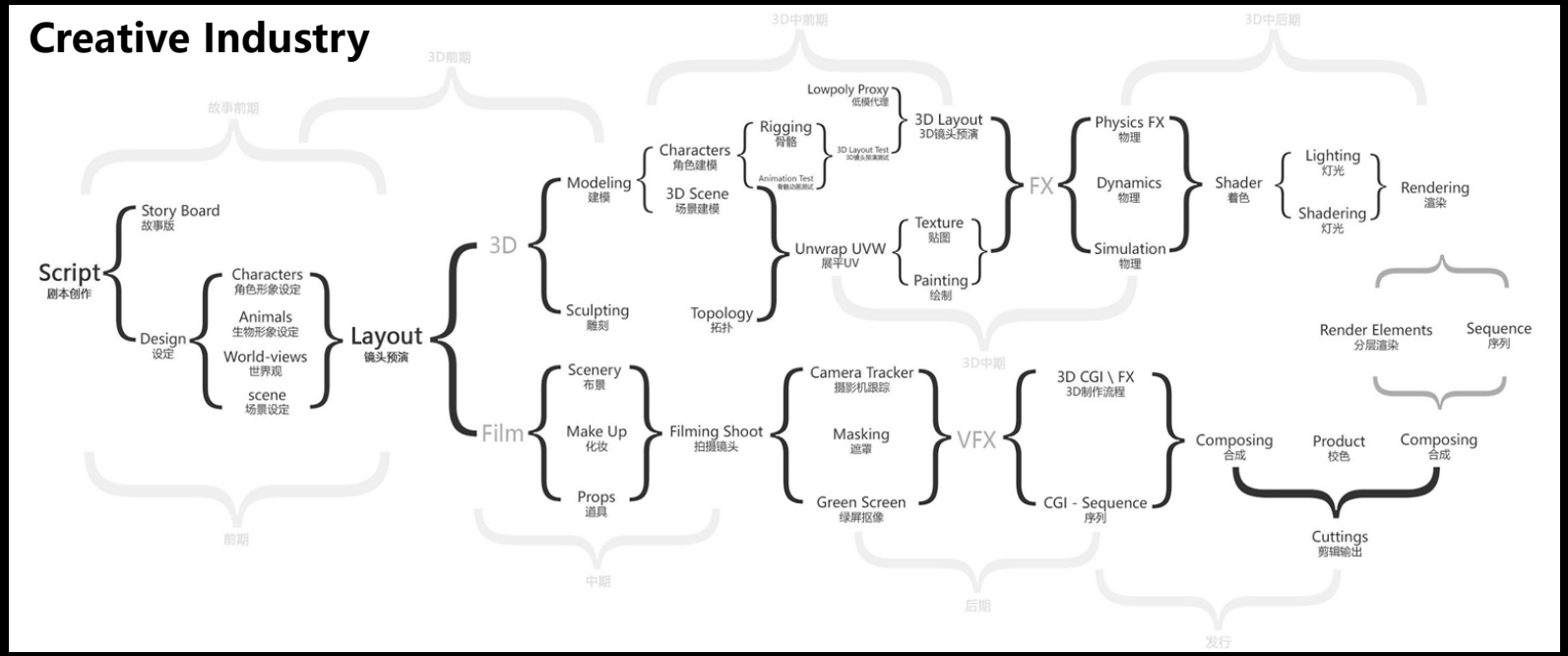
Game

Anime

VTuber

Virtual Beings

Creative Industry



Visual AIGC

Avatar

Foundation Model

Object

Scene

Visual AIGC

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE | S-LAB FOR ADVANCED INTELLIGENCE

Avatar

Foundation Model

Object

Scene

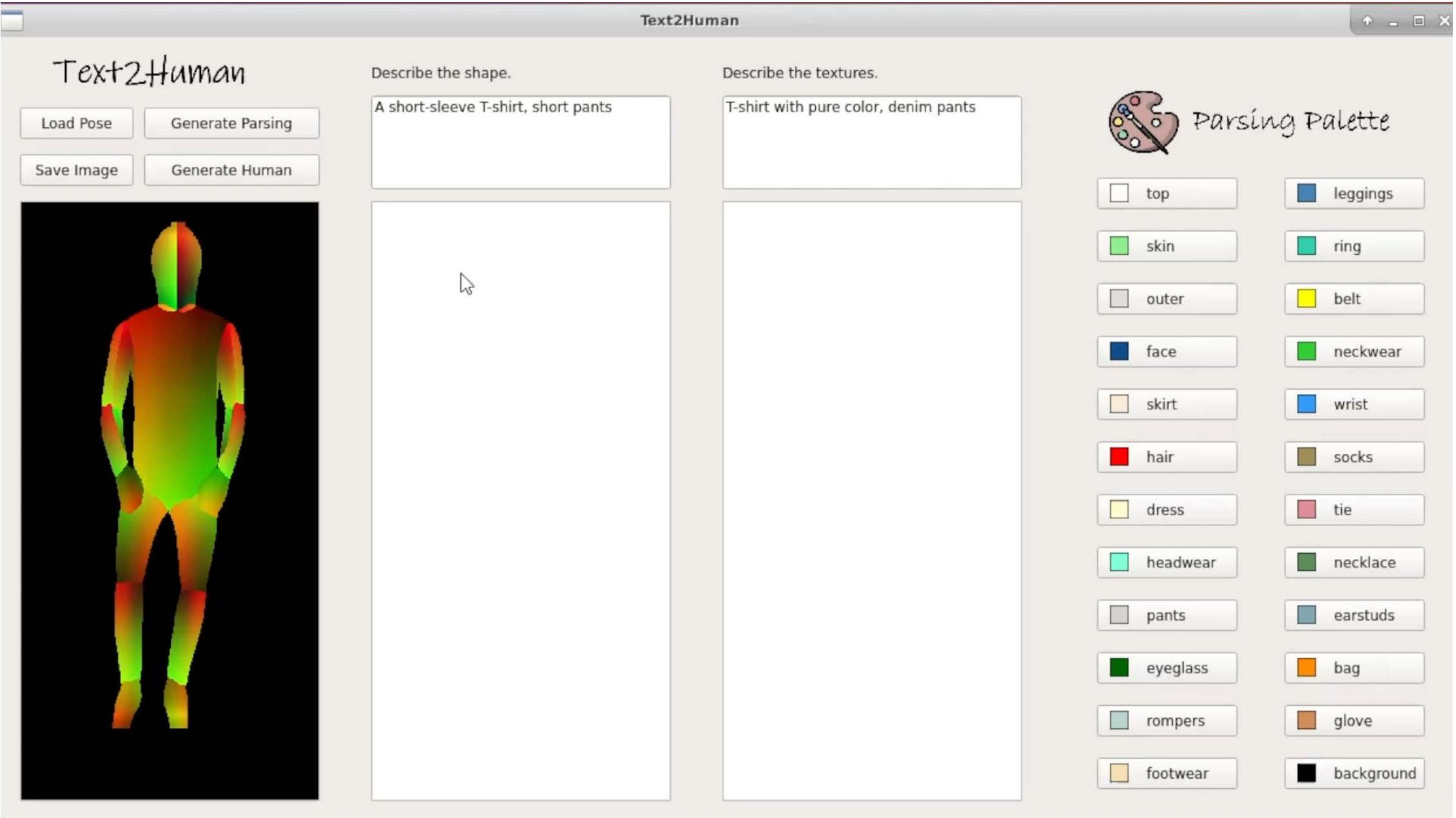
StyleGAN-Human: 2D Human Generation

HyperHuman: 2D Human Generation

A baby girl with beautiful blue eyes standing next to a brown teddy bear.

A little girl with wavy hair and smile holding a teddy bear.

Text2Human: Text-to-2D Human



Text2Performer: Text-to-2D Human Video

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE | S-LAB FOR ADVANCED INTELLIGENCE

The dress the person wears has medium sleeves and it is of short length. The texture of it is pure color.

The lady moves to the left.

She is turning right from the front to the side.

She is turning right from the side to the back.

She turns right from the back to the side.

She turns right from the side to the front.

She moves to the right.

EVA3D: 3D Human Generation

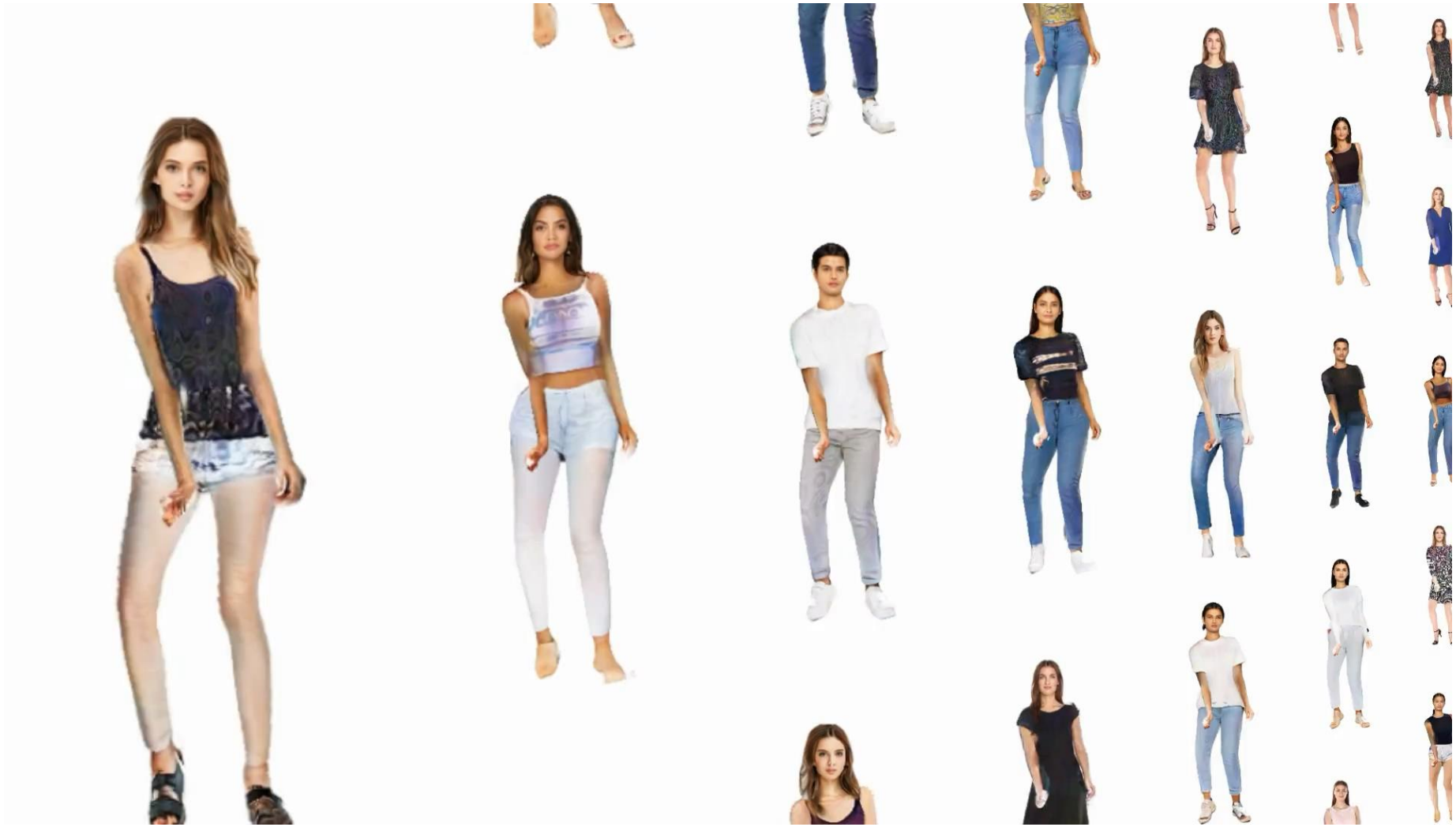
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE | S-LAB FOR ADVANCED INTELLIGENCE

- Learn 3D generation from 2D image collections

Static **?** ***Articulated***

EVA3D: 3D Human Generation

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE | S-LAB FOR ADVANCED INTELLIGENCE



AvatarCLIP: Text-to-3D Avatar

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE | S-LAB FOR ADVANCED INTELLIGENCE

I want to generate a tall and fat Iron Man that is running.

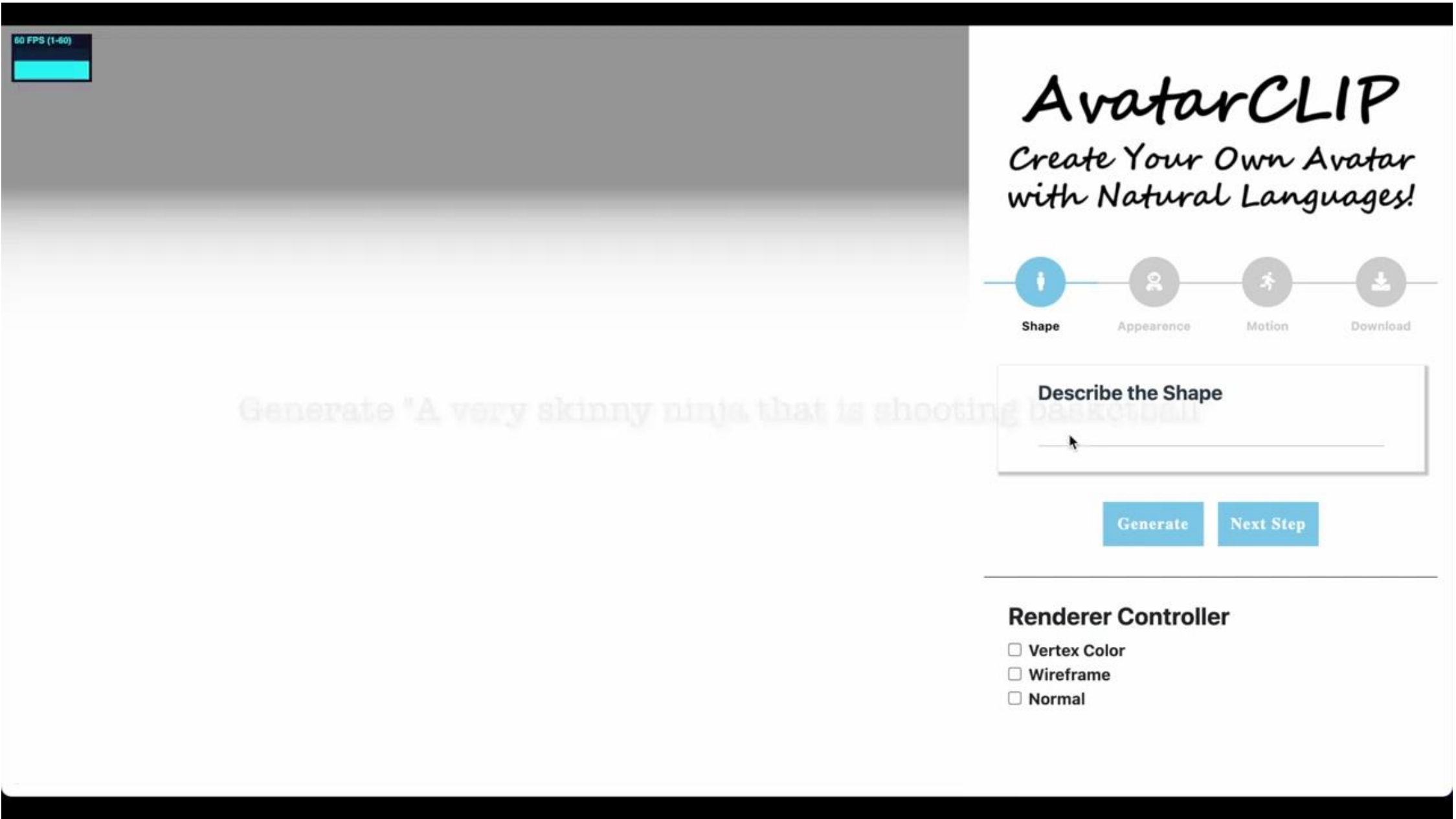
I would like to generate a skinny ninja that is raising arms.

I want to generate a tall and skinny female soldier that is arguing.

I want to generate an overweight sumo wrestler that is sitting.

AvatarCLIP: Text-to-3D Avatar

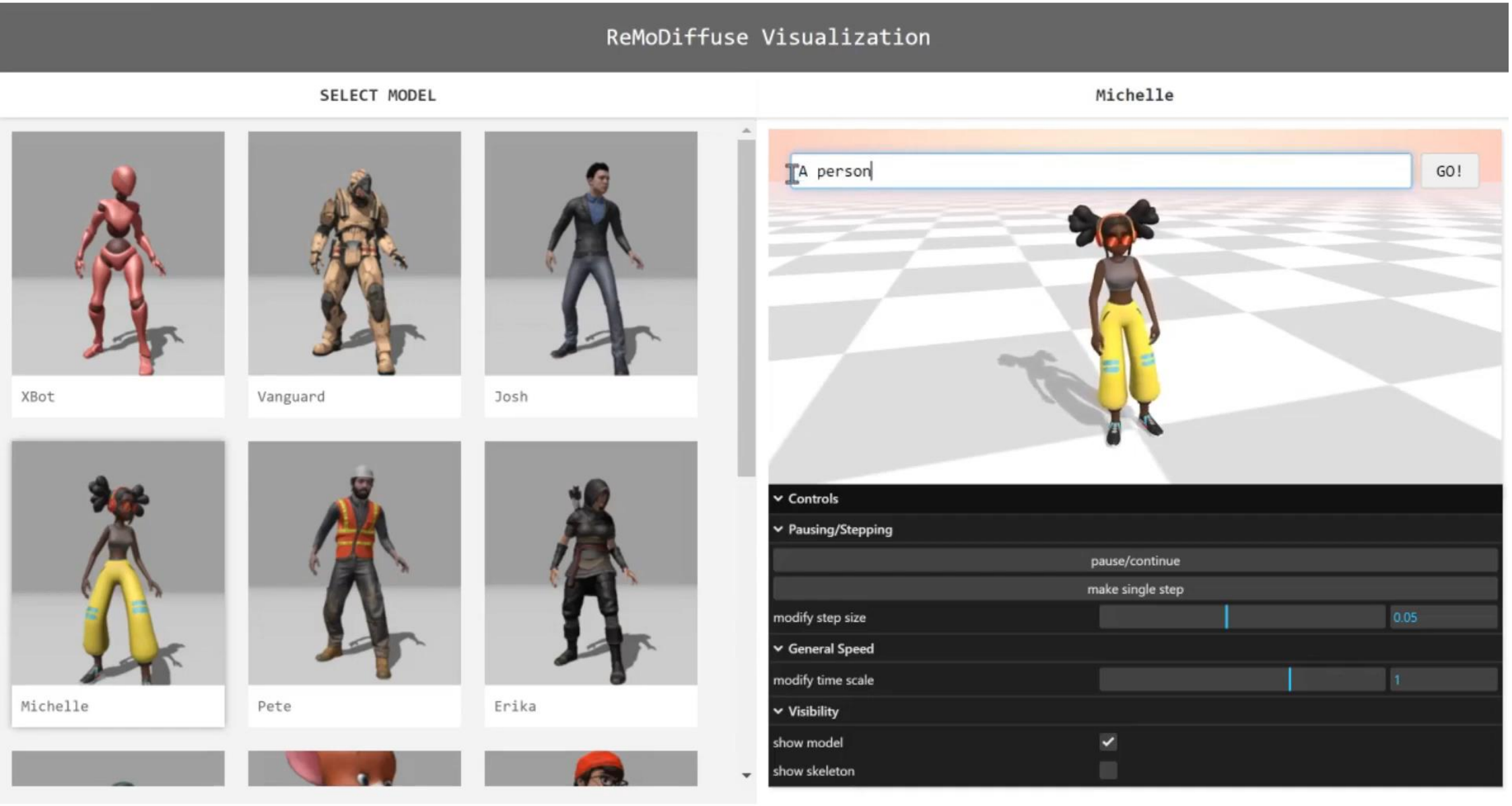
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE | S-LAB FOR ADVANCED INTELLIGENCE



MotionDiffuse: Text-to-3D Human Video

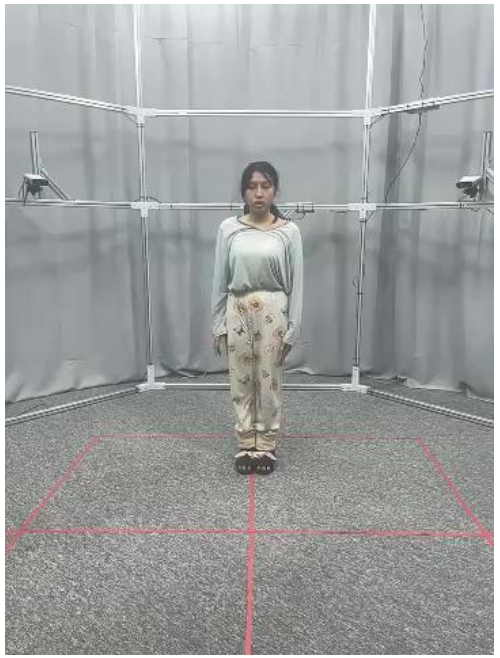
ReMoDiffuse: Text-to-3D Human Video

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE | S-LAB FOR ADVANCED INTELLIGENCE



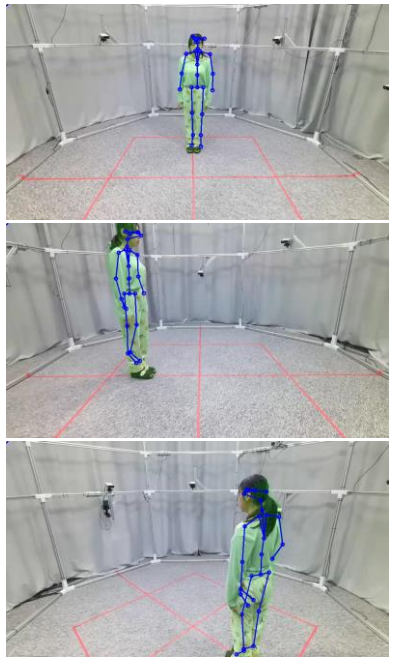
HuMMan Dataset

Artec Eva

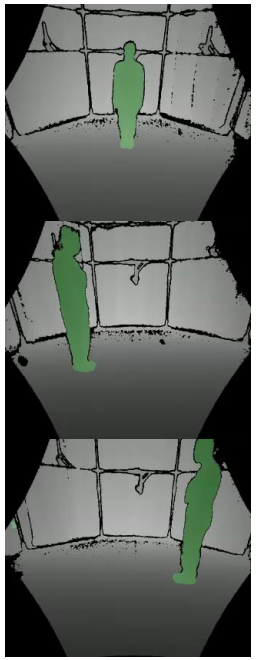


iPhone RGB

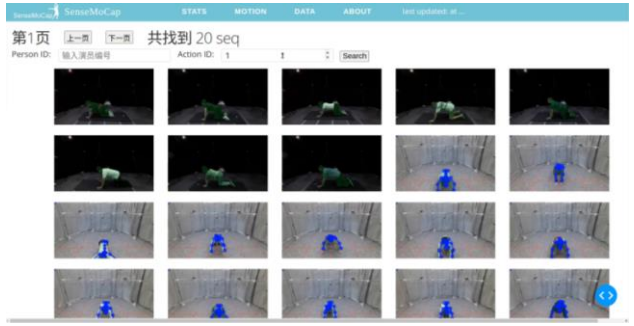
iPhone Depth



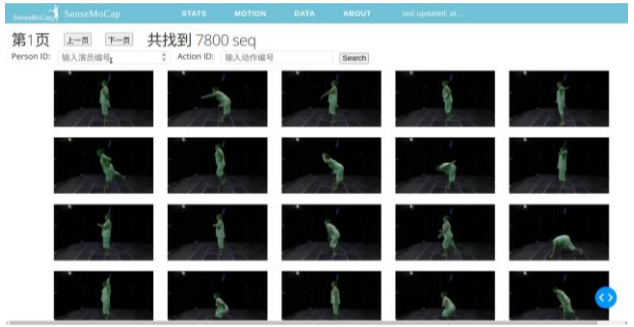
Kinect RGB



Kinect Depth



Search by Action



Search by Actor

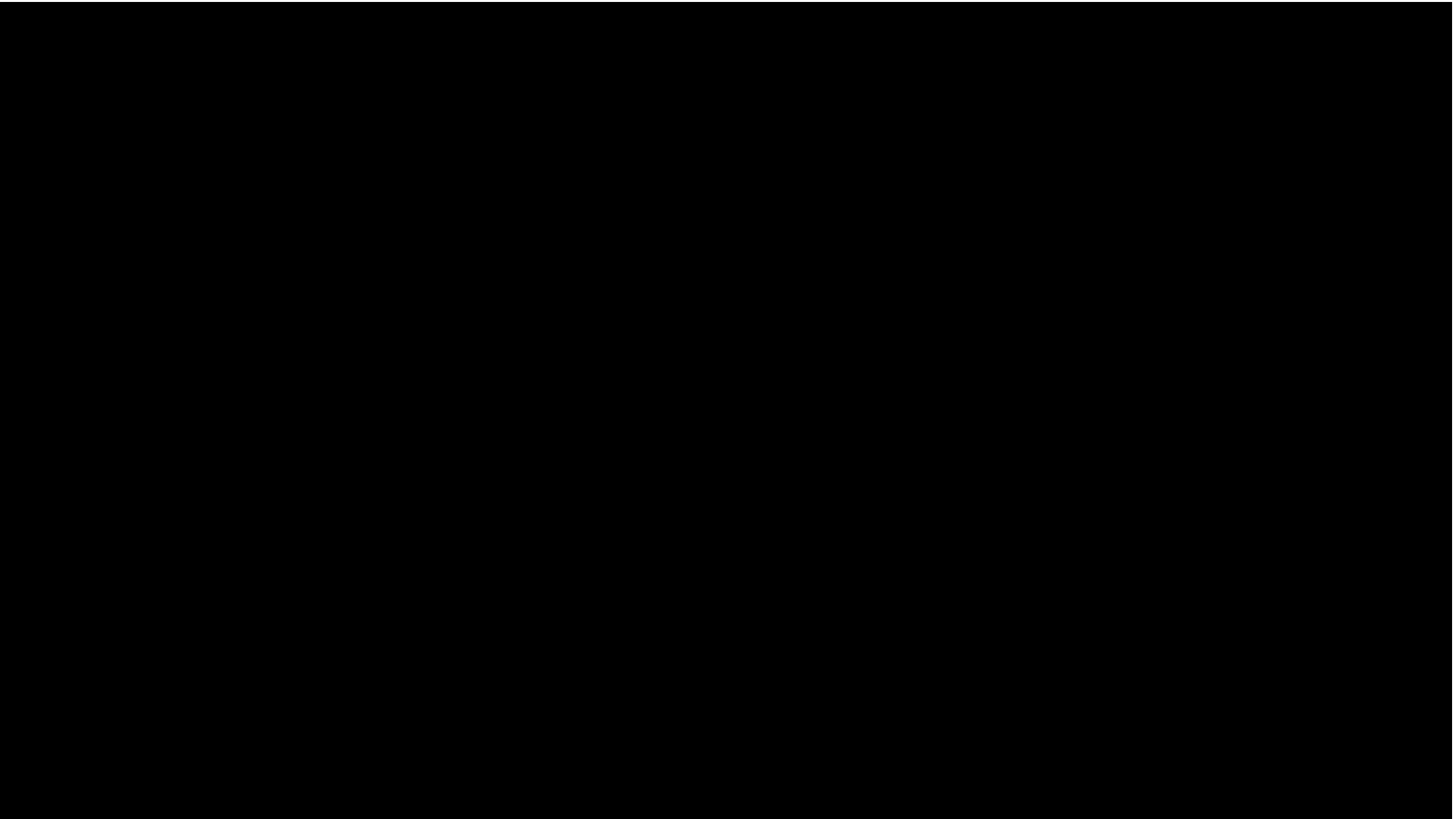
0.1mm Accuracy

11 Cameras

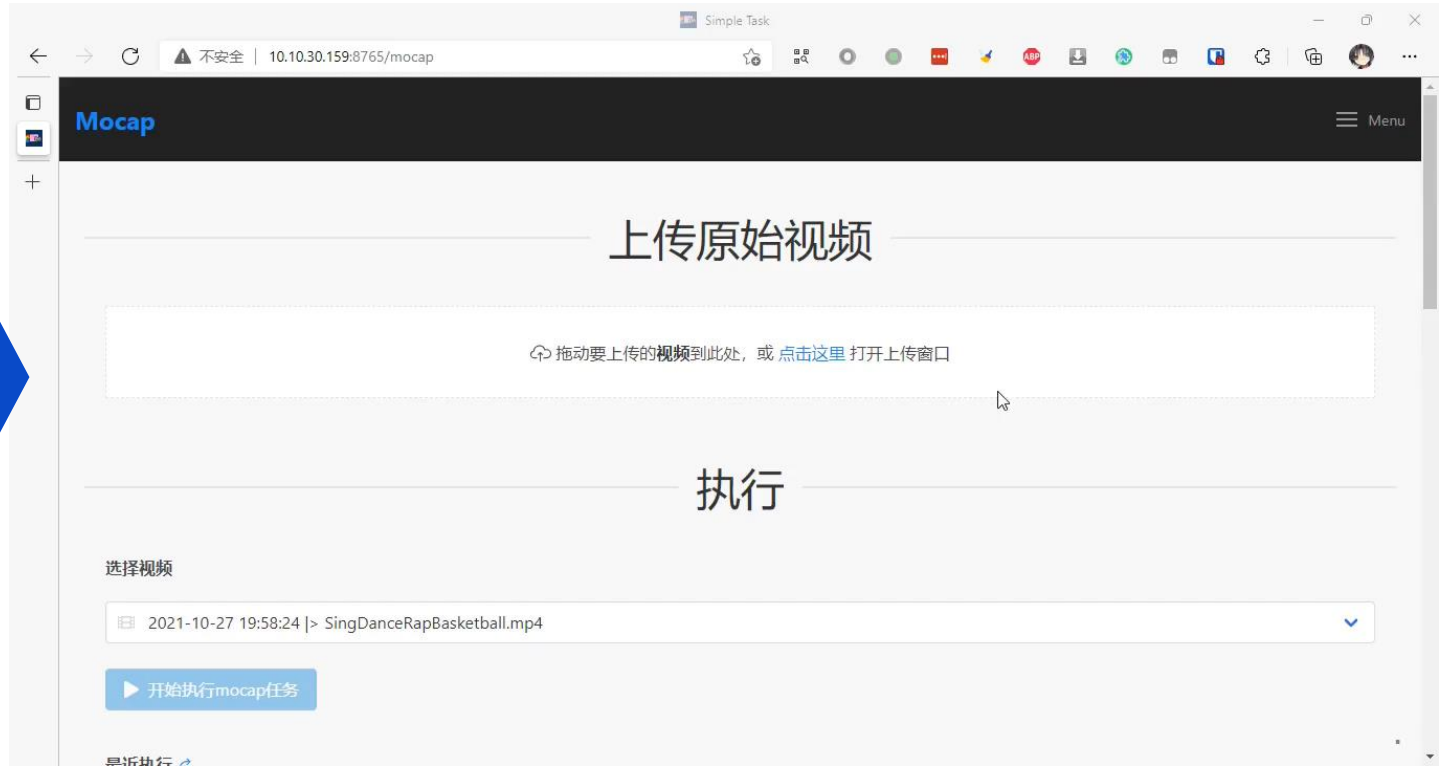
1G Data / Sec

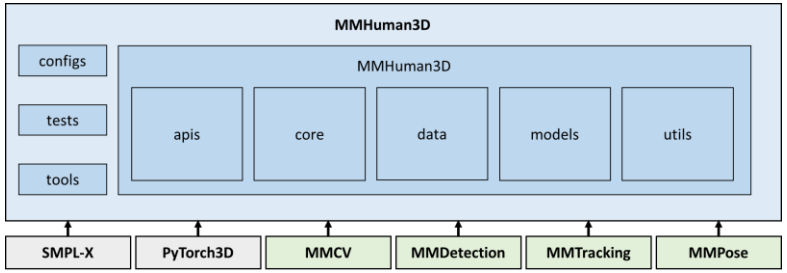
6 Actor / Day

SMPLer-X: MoCap Anybody



MMHuman3D Software





3D Animation Production:

3 days -> 30 min

Visual AIGC

Avatar

Foundation Model

Object

Scene

OmniObject3D: 3D Object Datasets

OmniObject3D is a **large-vocabulary** 3D dataset for **real-world scanned objects**.

- ✓ **6k** high-quality 3D models
- ✓ **190** categories
- ✓ **4** modalities: textured mesh, point cloud, real-captured video, synthetic multi-view images.
- ✓ Many down-stream tasks

	Dataset	Year	Real	Full 3D	Video	Num Objs	Num Cats
	ShapeNet	2015		√		51k	55
	ModelNet	2014		√		12k	40
	3D-Future	2020		√		16k	34
	ABO	2021		√		8k	63
	Toys4K	2021		√		4k	105
	CO3D	2021	√		√	19k	50
Real-world 3D scans	DTU	2014	√	√		124	NA
	GSO	2021	√	√		1k	17
	AKB-48	2022	√	√		2k	48
	Ours	2022	√	√	√	**6k**	**190**

OmniObject3D: 3D Object Datasets

Summary

It's a teacup.

Appearance

This is a relatively small teacup with a brownish-red exterior and white interior, featuring a blue line pattern at the top and a rounded white bump on the bottom, structured in an overall axisymmetric manner.

Material

Ceramic, hard, reflective, smooth surface.

Style

Simplicity.

Function

Water storage.

Summary

It's a teapot.

Appearance

This teapot is white with a gray handle positioned perpendicular to the spout, and a small round gray handle at the top of the lid; the body of the teapot is adorned with a pattern of pink lotuses, gray lotus leaves, and red buds, all structured in an asymmetric manner.

Material

Ceramic, rough surface, hard, slightly reflective.

Style

Simplicity.

Function

Tea making, water storage.

Summary

It's a glasses case.

Appearance

Overall purple, the box features a pink LinaBell on the surface wearing a dark purple flower and blue eyes, complemented by a row of purple and pink letters underneath, all structured in an axisymmetric manner.

Material

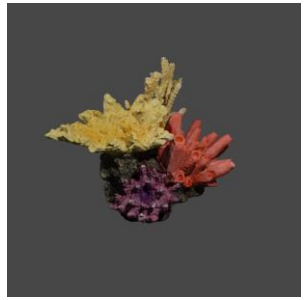
Leather, rubber, metal, smooth surface, hard, slightly reflective, metallic.

Style

Cartoon.

Function

Store glasses, decoration.



Summary

It's a coral simulation model.

Appearance

The upper part of this coral simulation model is yellow, below the yellow section, there are pink and purple corals, the purple corals have white attachments on their surfaces, several colors of corals are on a brown reef, and the entire model is asymmetrical.

Material

Plastic, rough surface, hard, slightly reflective.

Style

Reality.

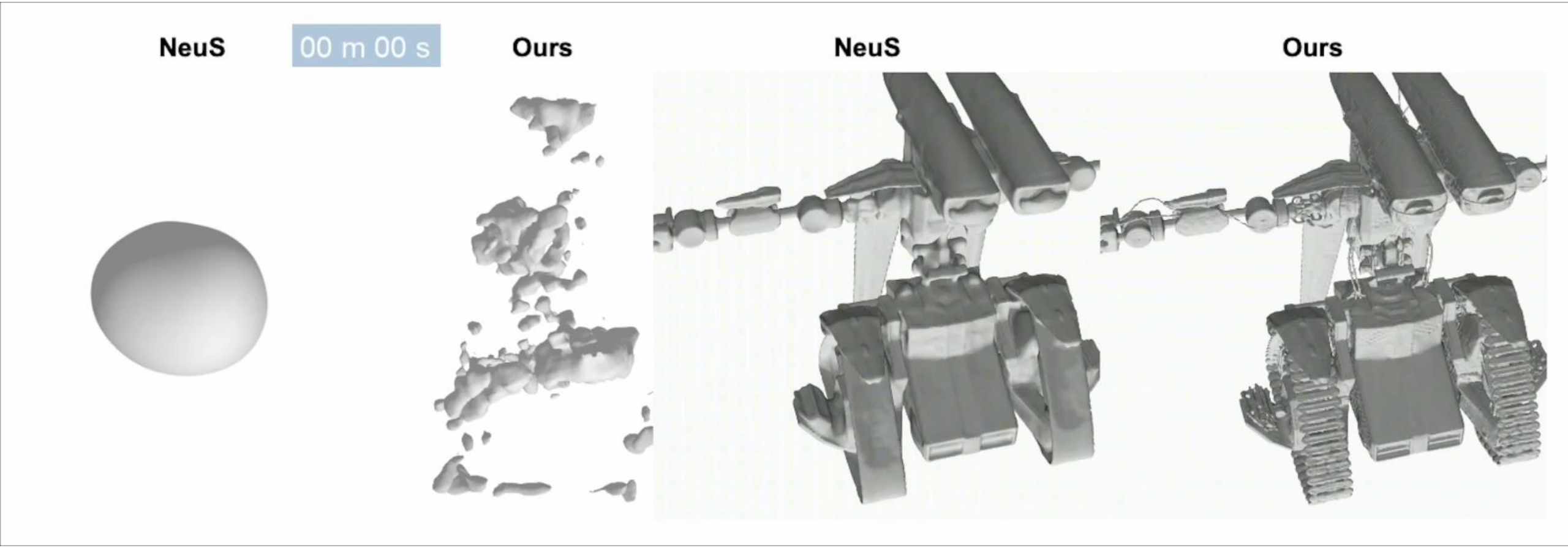
Function

Entertainment, decoration.

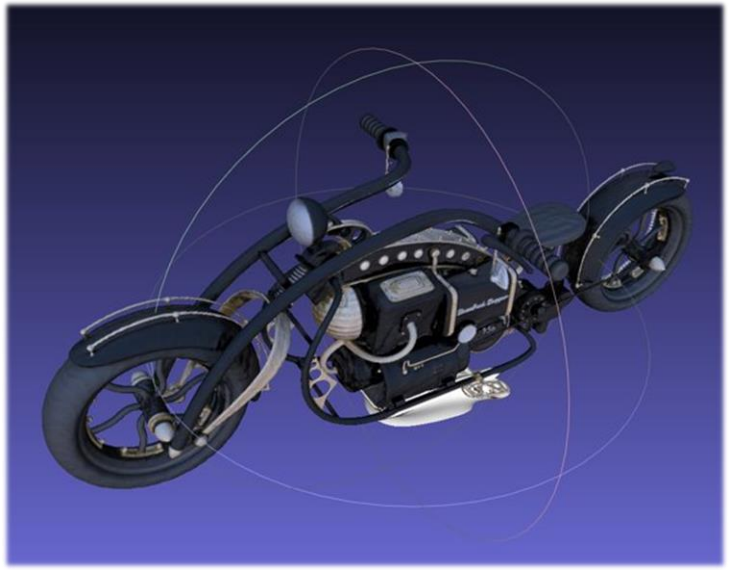
DiffTF: Large-Vocabulary 3D Diffusion Model

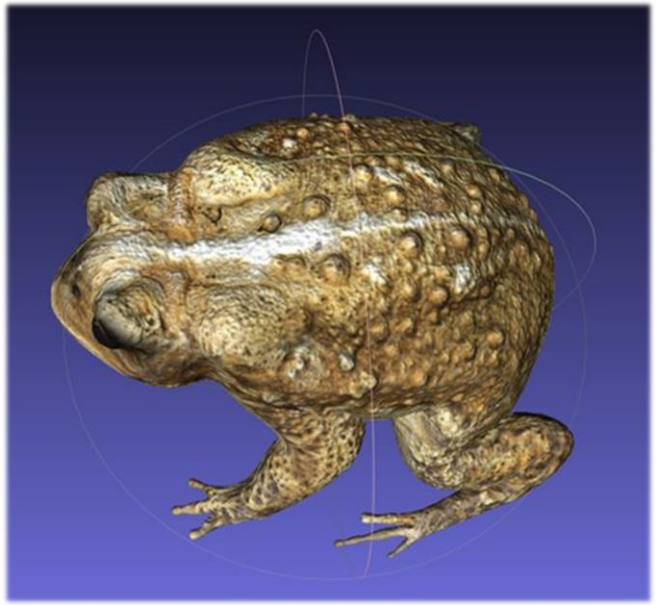
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE | S-LAB FOR ADVANCED INTELLIGENCE

Voxurf: Fast 3D Object Reconstruction



Voxurf: Fast 3D Object Reconstruction

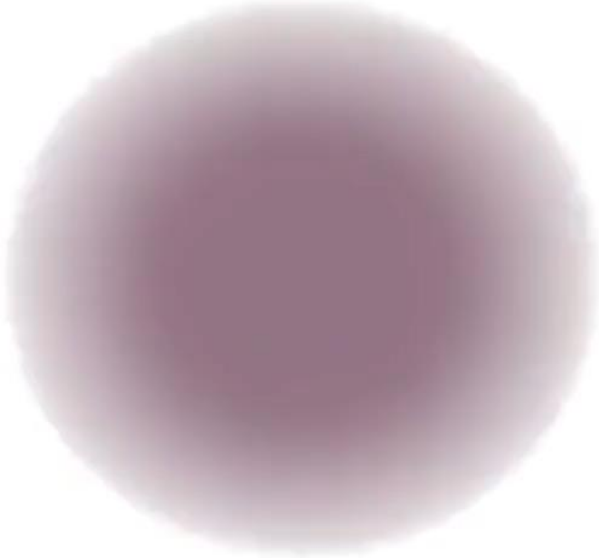





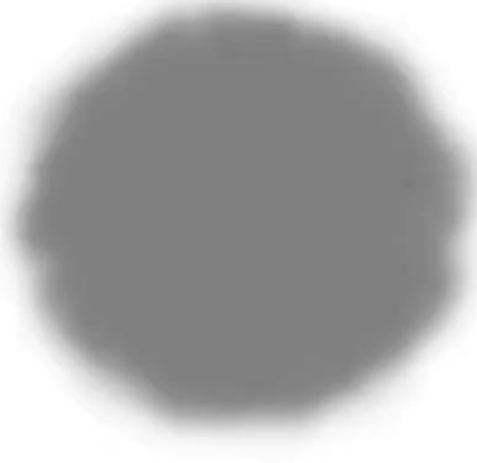

DreamGaussian: Efficient 3D Generation

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE | S-LAB FOR ADVANCED INTELLIGENCE

00:00
Minute Second



Zero-1-to-3 (NeRF)



Ours (Gaussian Splatting)

Visual AIGC

Avatar

Foundation Model

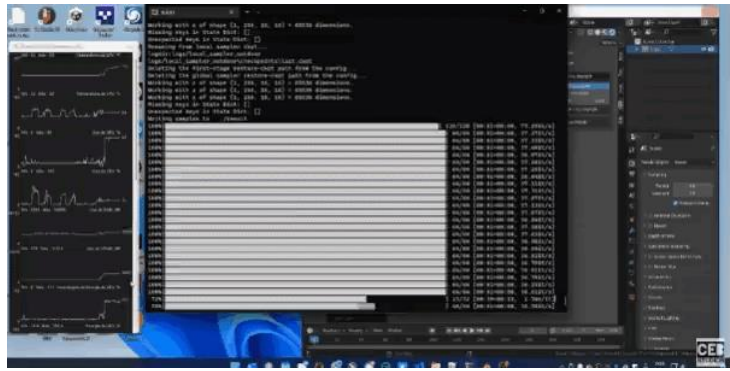
Object

Scene

Text2Light: Text-to-3D Environment

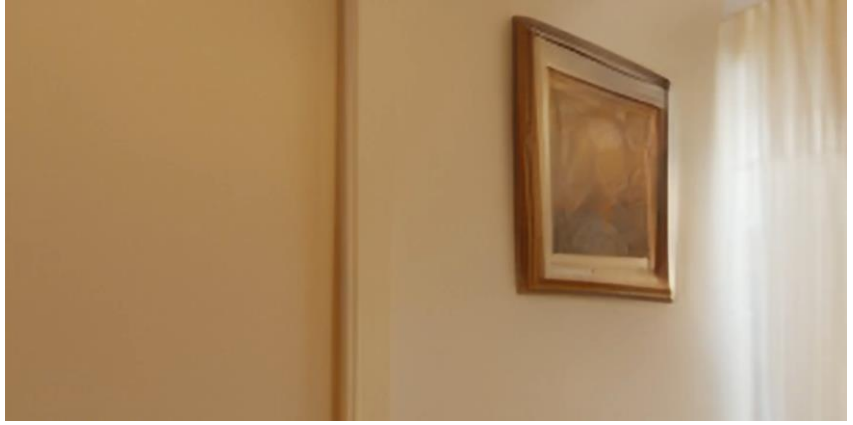
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE | S-LAB FOR ADVANCED INTELLIGENCE

"brown wooden dock on lake surrounded by green trees during daytime"

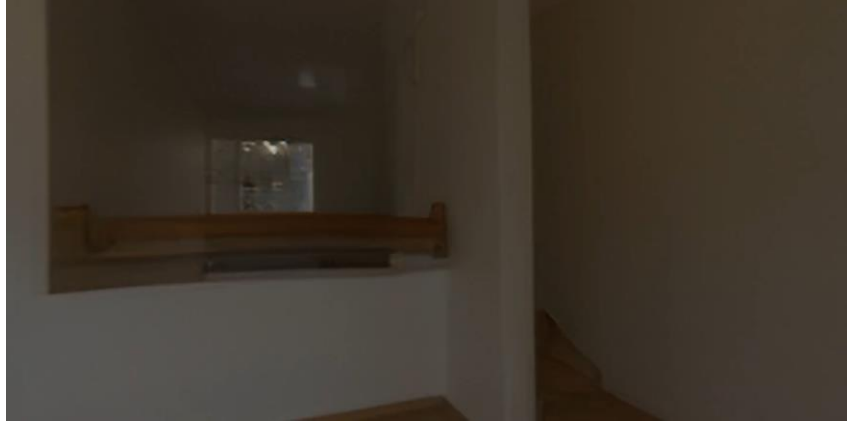


4K+ Resolution with High Dynamic Range

“white bed linen with white pillow”

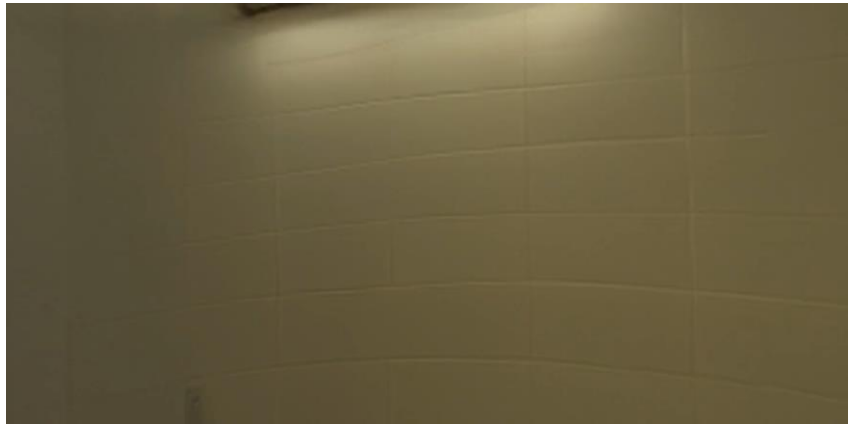


“brown wooden floor with white wall”

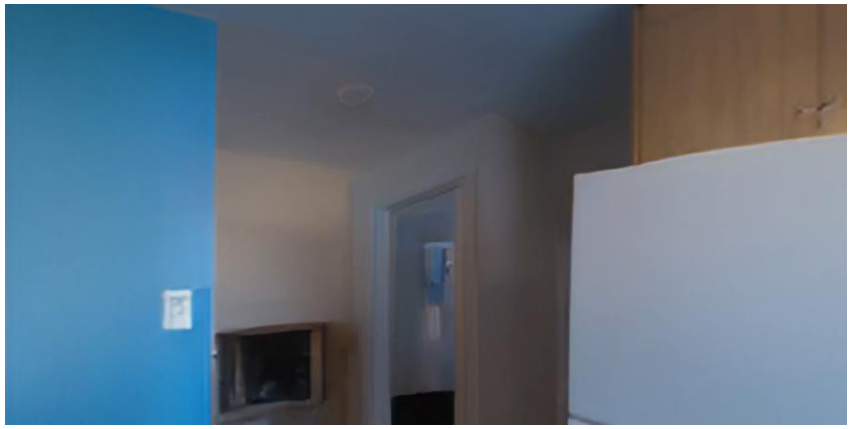


“gray concrete pathway with wall signages”

“closeup photo of concrete stair surrounded by white painted wall”



“blue and brown wooden counter”



“empty parking lot during daytime”

Suzanne Monkey: glossy Shader balls: glass, diffuse, glossy, mixture of diffuse and glossy

Text2Light: Text-to-3D Environment

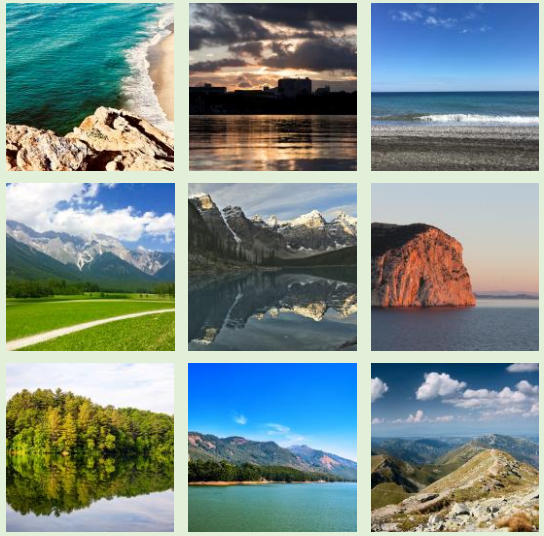
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE | S-LAB FOR ADVANCED INTELLIGENCE

Describe Your Scene

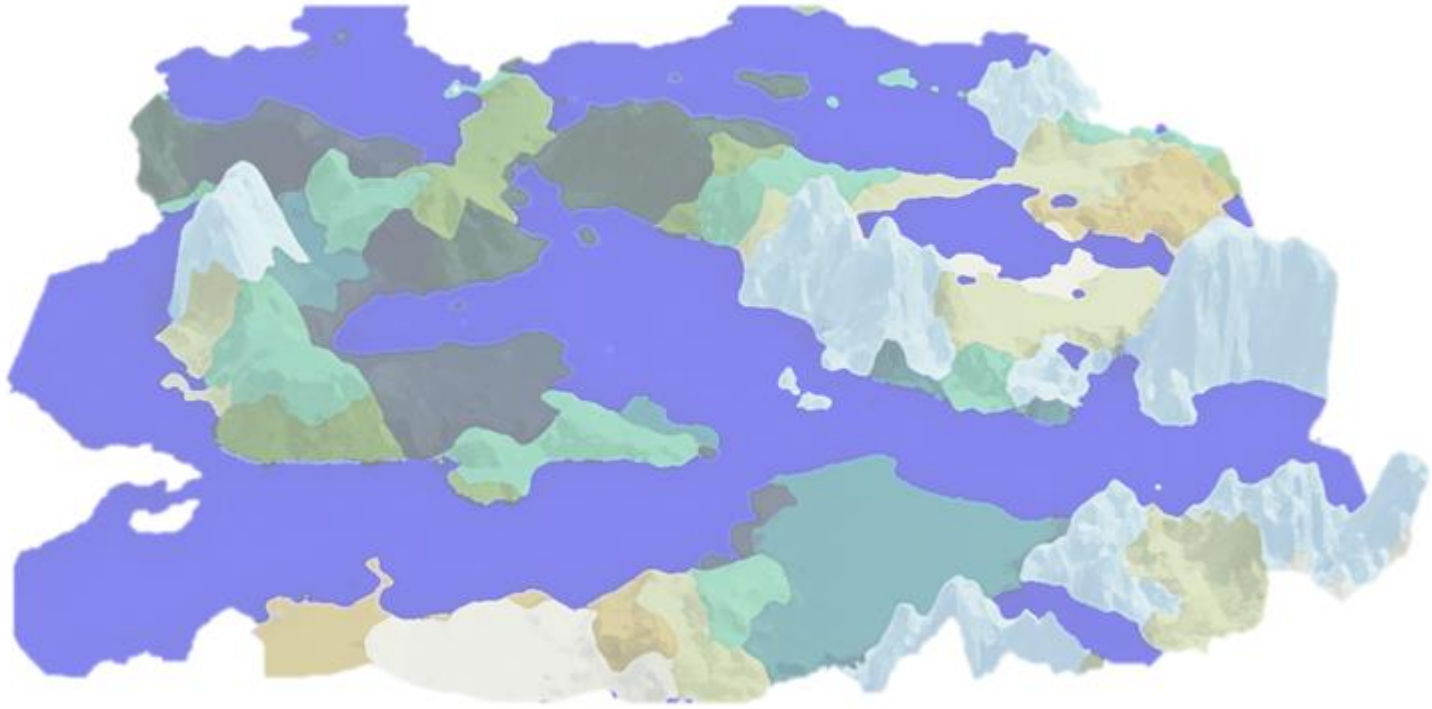
e.g. a living room

Generate Render

SceneDreamer: Unbounded 3D Scene Generation

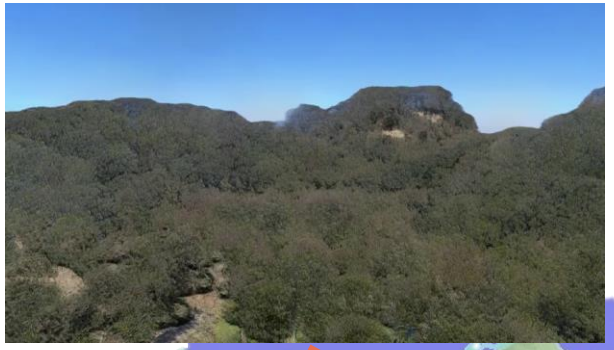


In-the-wild 2D Image Collections



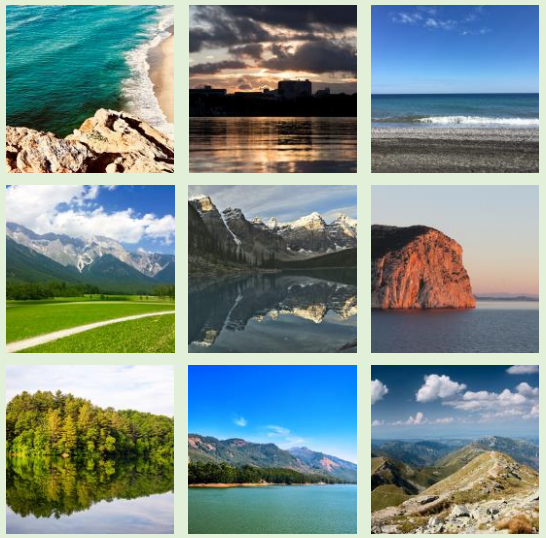
Photorealistic Unbounded 3D Scenes

SceneDreamer: Unbounded 3D Scene Generation

Multi-view consistent

Well-defined geometry



In-the-wild
Image Collections

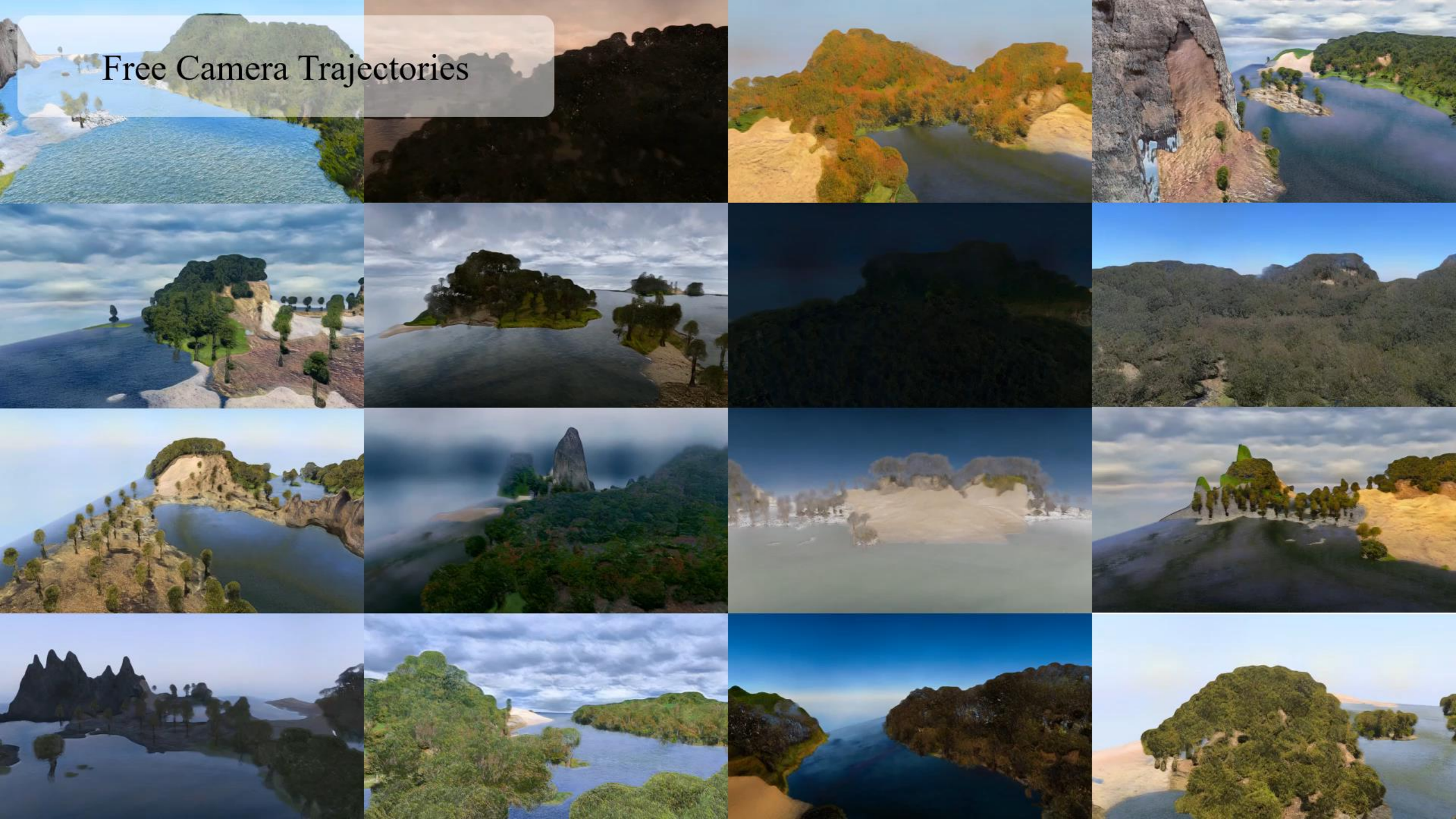
Photorealistic
Unbounded 3D Scenes

Diverse scenes and styles



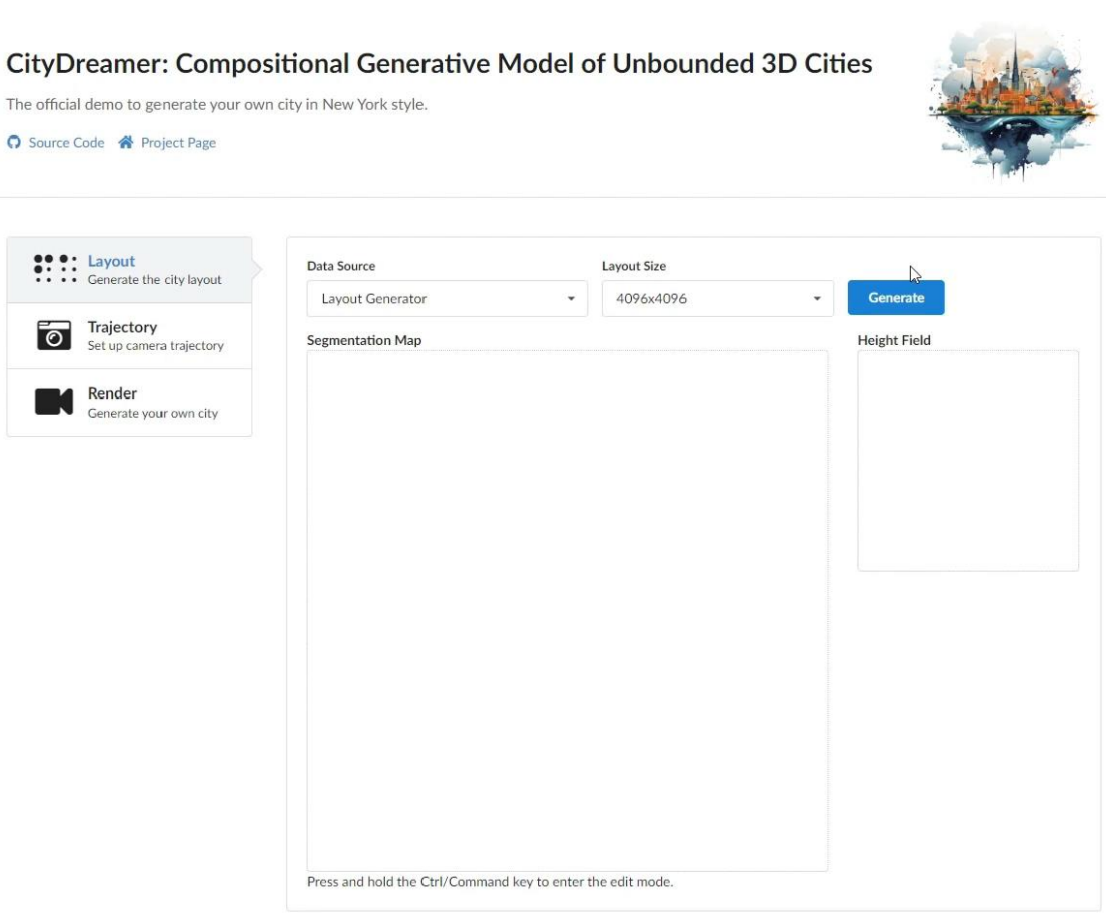
Infinite 3D World!


Generate with Different Styles


Free Camera Trajectories

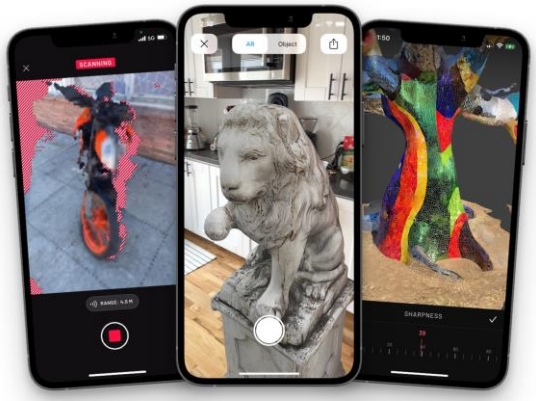
CityDreamer: Unbounded 3D City Generation

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE | S-LAB FOR ADVANCED INTELLIGENCE

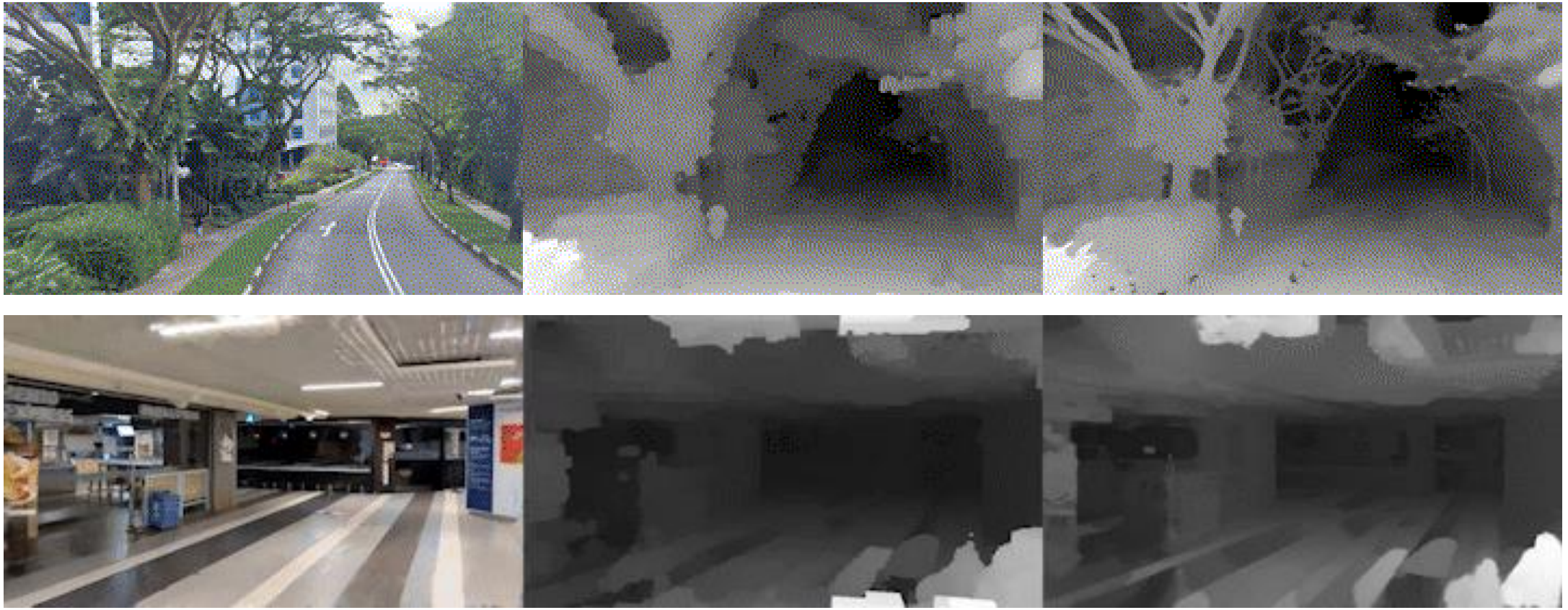


F2NeRF: Mobile 3D Scene Reconstruction

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE | S-LAB FOR ADVANCED INTELLIGENCE



F2NeRF: Mobile 3D Scene Reconstruction



Visual AIGC

Avatar

Foundation Model

Object

Scene

FreeU: Free Lunch in Diffusion U-Net

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE | S-LAB FOR ADVANCED INTELLIGENCE

FreeU:

Free Lunch in Diffusion U-Net

Chenyang Si, Ziqi Huang, Yuming Jiang, Ziwei Liu

S-Lab, Nanyang Technological University

LaVie: Text-to-Video Foundation Model

Cinematic shot of Van Gogh's selfie, Van Gogh style.

A corgi's head depicted as an explosion of a nebula.

Yoda playing guitar on the stage.

The bund Shanghai, oil painting.

The Eiffel Tower at sunset, trending on artstation, oil painting.

A fantasy landscape, trending on artstation, 4k, high resolution.

A super cool giant robot in Cyberpunk city, artstation.

A Mars rover moving on Mars.

A space shuttle launching into orbit, with flames and smoke billowing out from the engine.

Adventure of a Panda

S-LAB
FOR ADVANCED
INTELLIGENCE

SEINE: Image-to-Video Foundation Model

Image2Video

Transition

Image2Video

Transition

HyperDreamer: Image-to-3D Foundation Model

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE | S-LAB FOR ADVANCED INTELLIGENCE

Visual AIGC

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE | S-LAB FOR ADVANCED INTELLIGENCE

Avatar

Thank You!

Object

Scene