



3D by AI: Using Generative AI and NeRFs for Building Virtual Worlds [S52163]

Gavriel State, Senior Director, Simulation and AI | GTC Online - March 2023

NVIDIA Omniverse

USD Based Platform for Creating and Connecting Virtual Worlds



USD



Physics



Materials



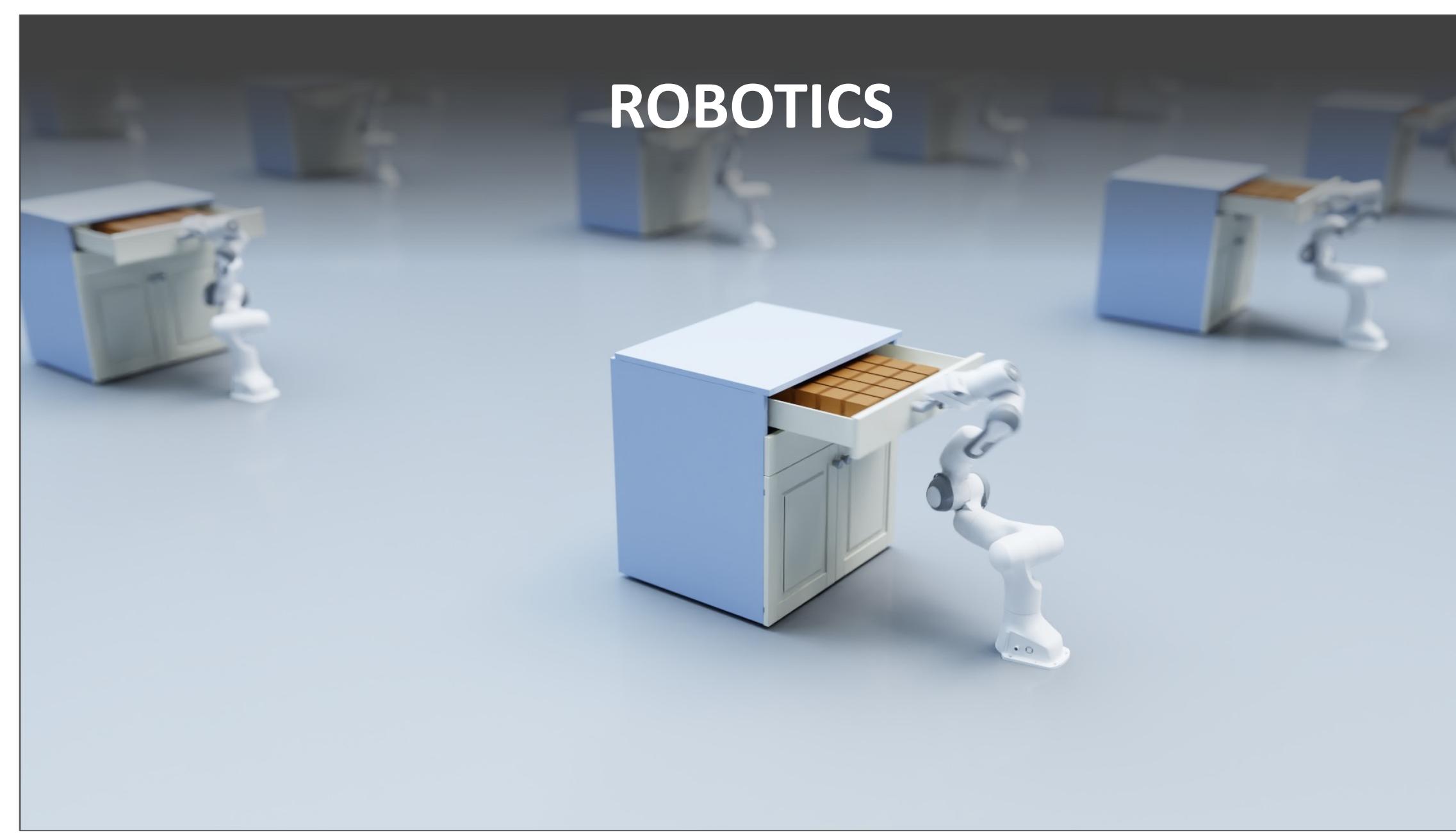
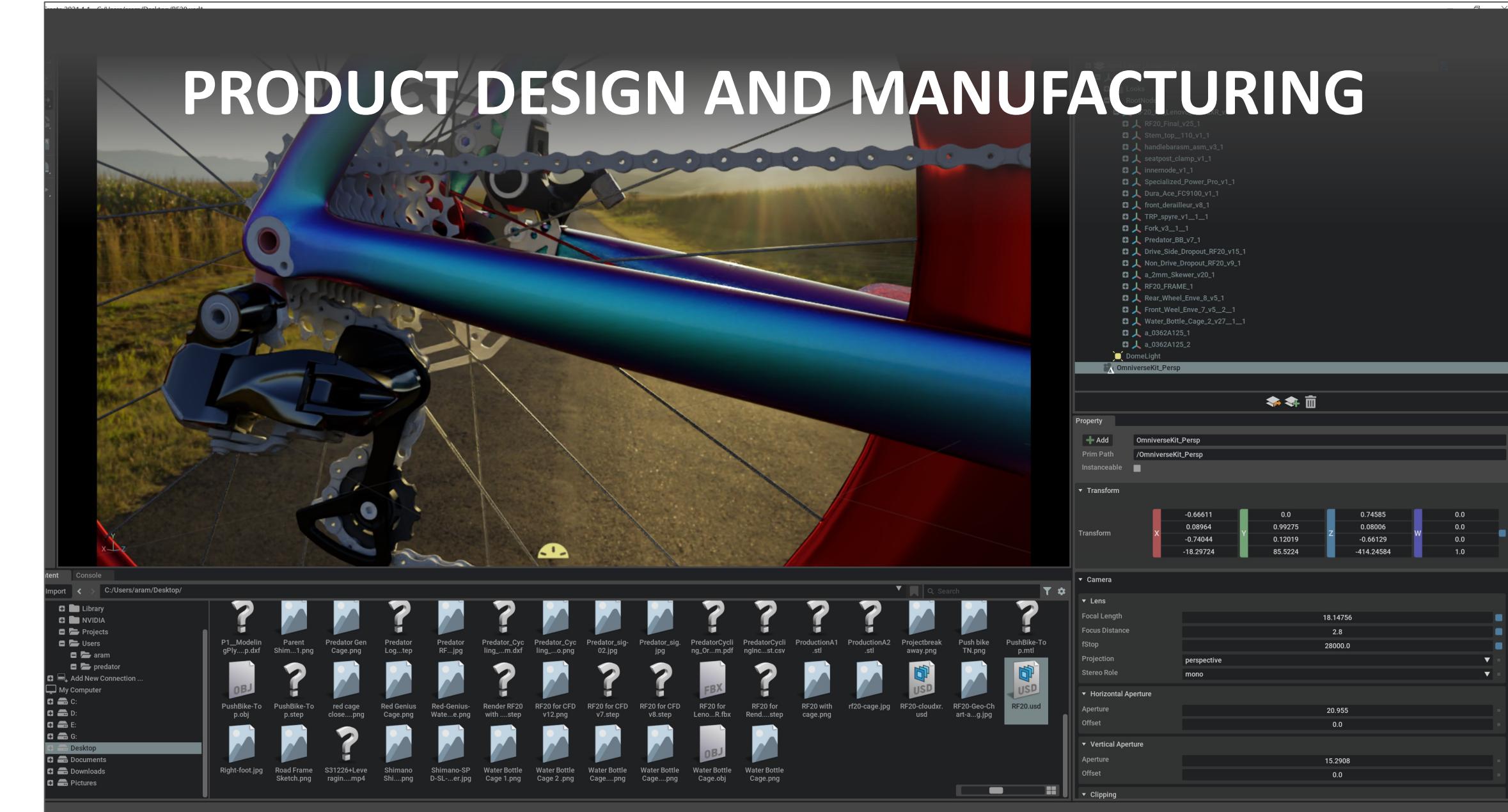
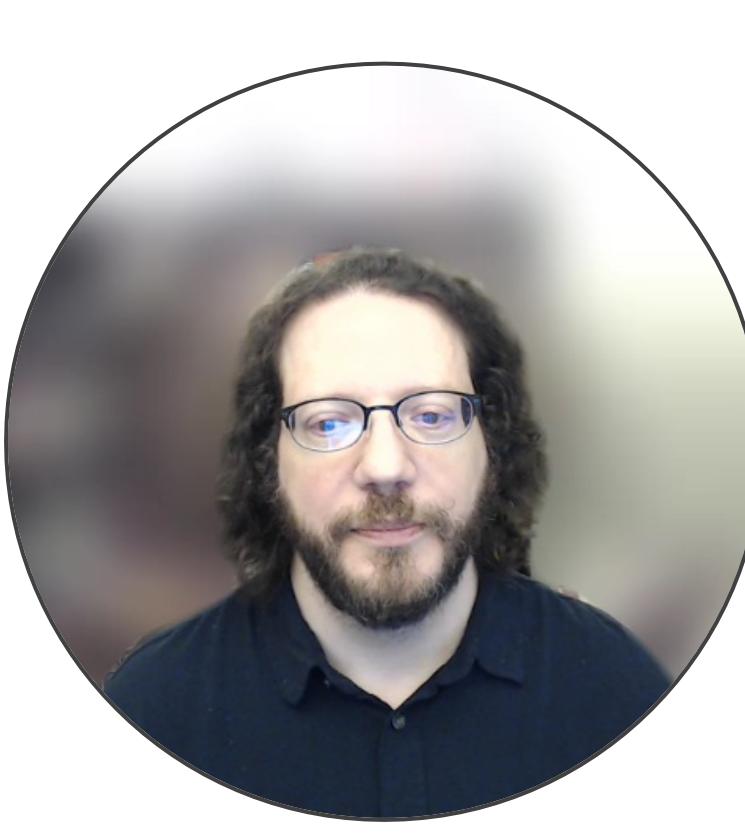
Path-Tracing



AI

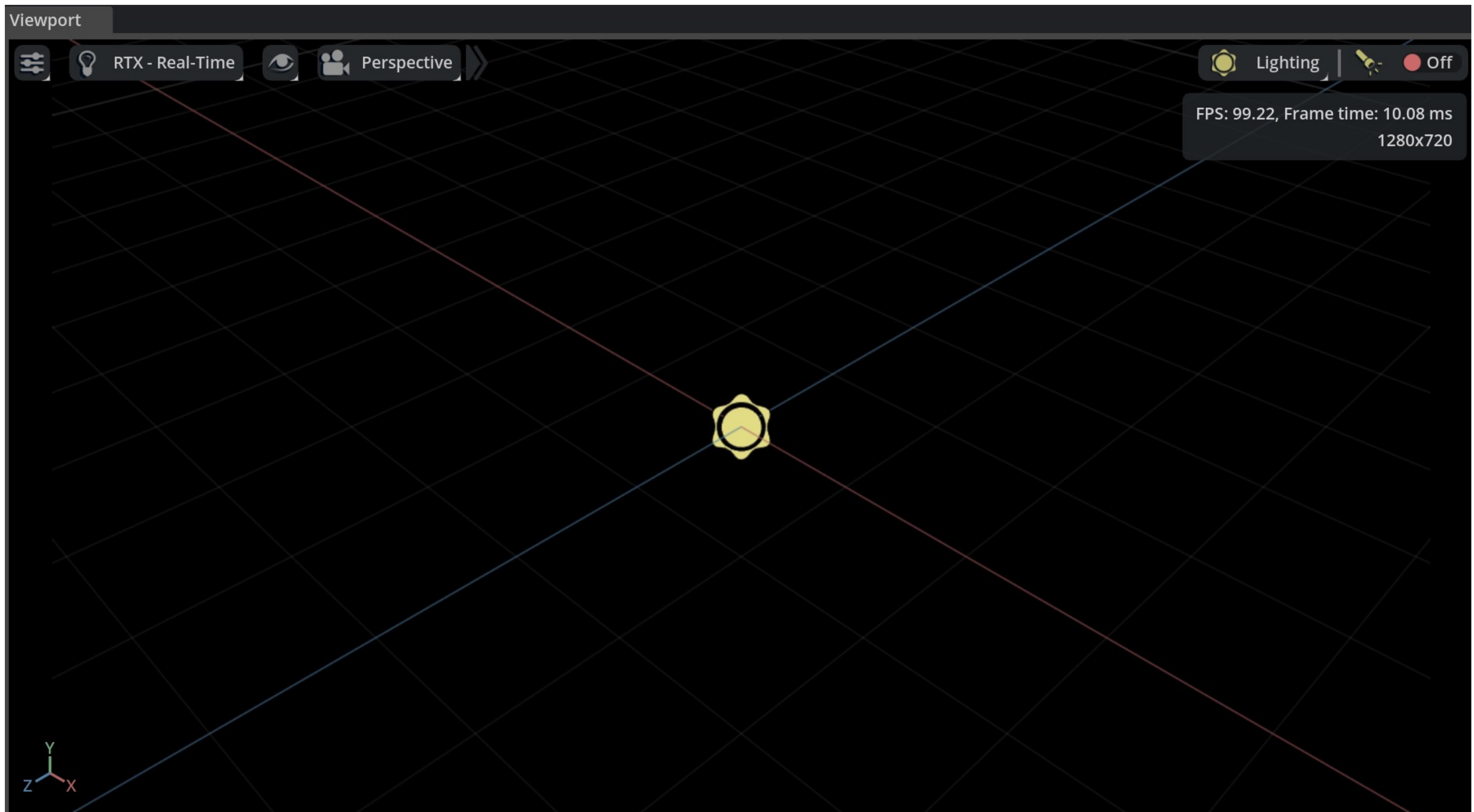


Omniverse: 3D Workflows and Tools for Every Industry



Virtual Worlds Need Content

Without it we just have a blank slate



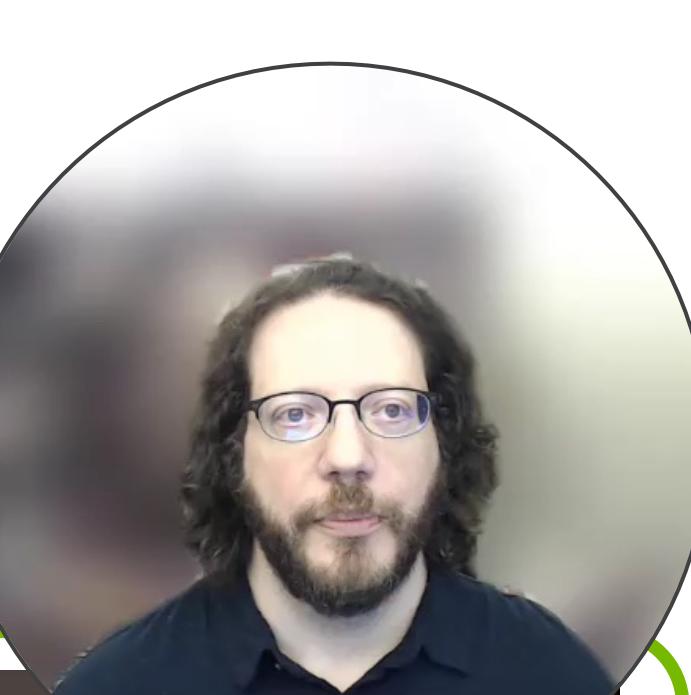


Traditional Approach: Manual Artist-Created Assets

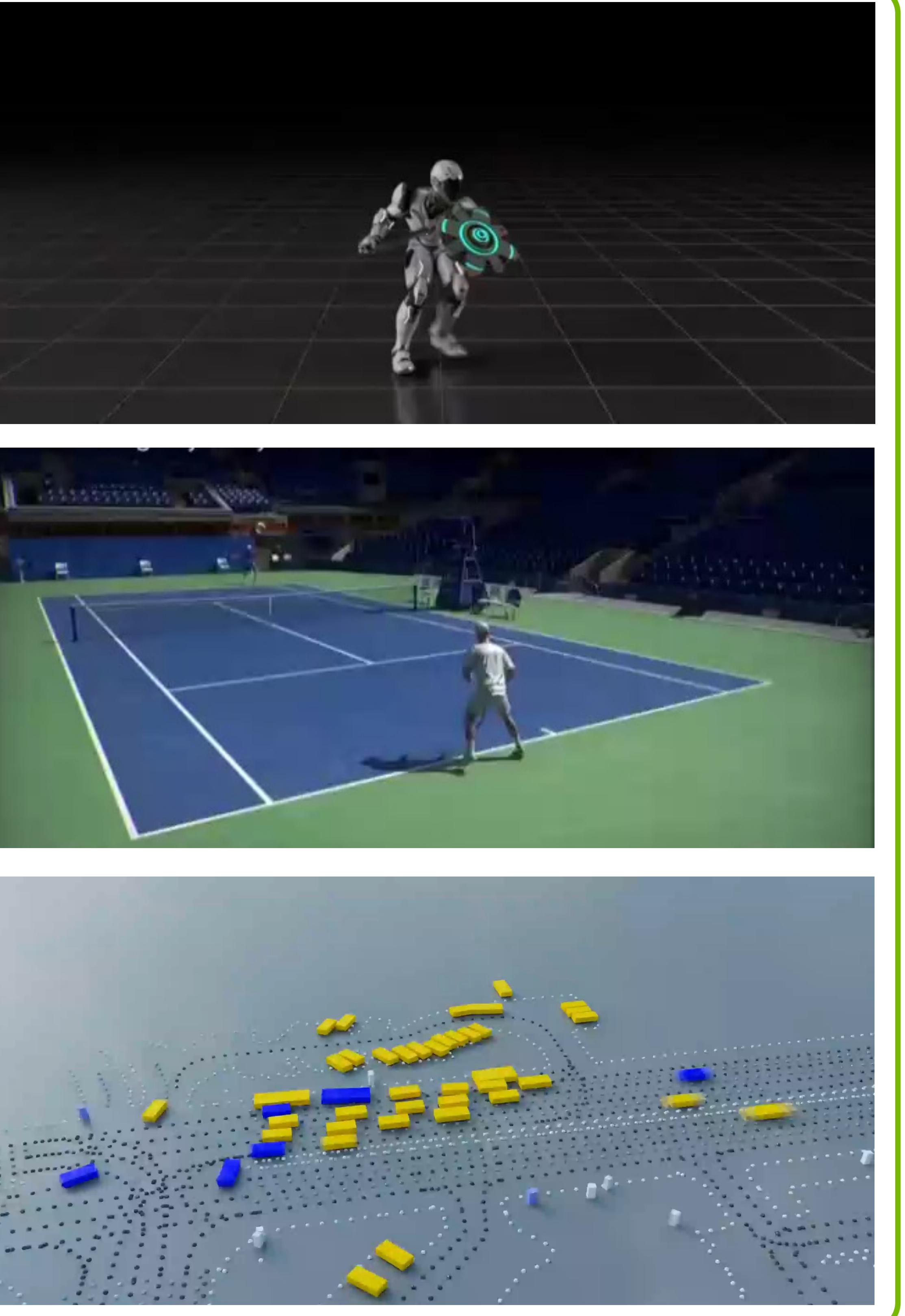
Very labour intensive – impossible to scale!



AI can Help Everywhere



Asset Creation



Behavior & Animation



World Capture + Augmentation

Neural Radiance Fields (NeRFs) introduced in 2020

NeRFs are an astonishing new method for capturing reality

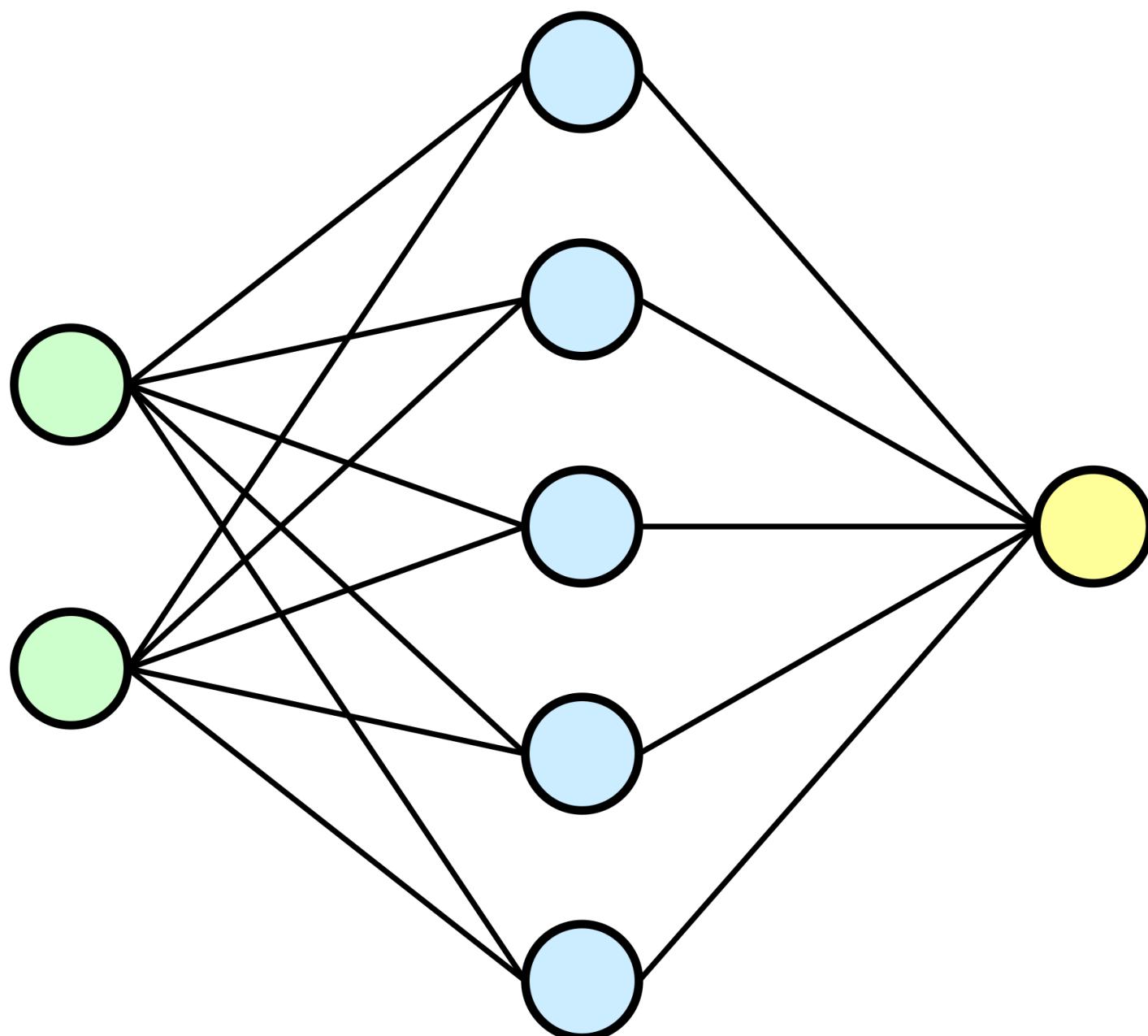


NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis

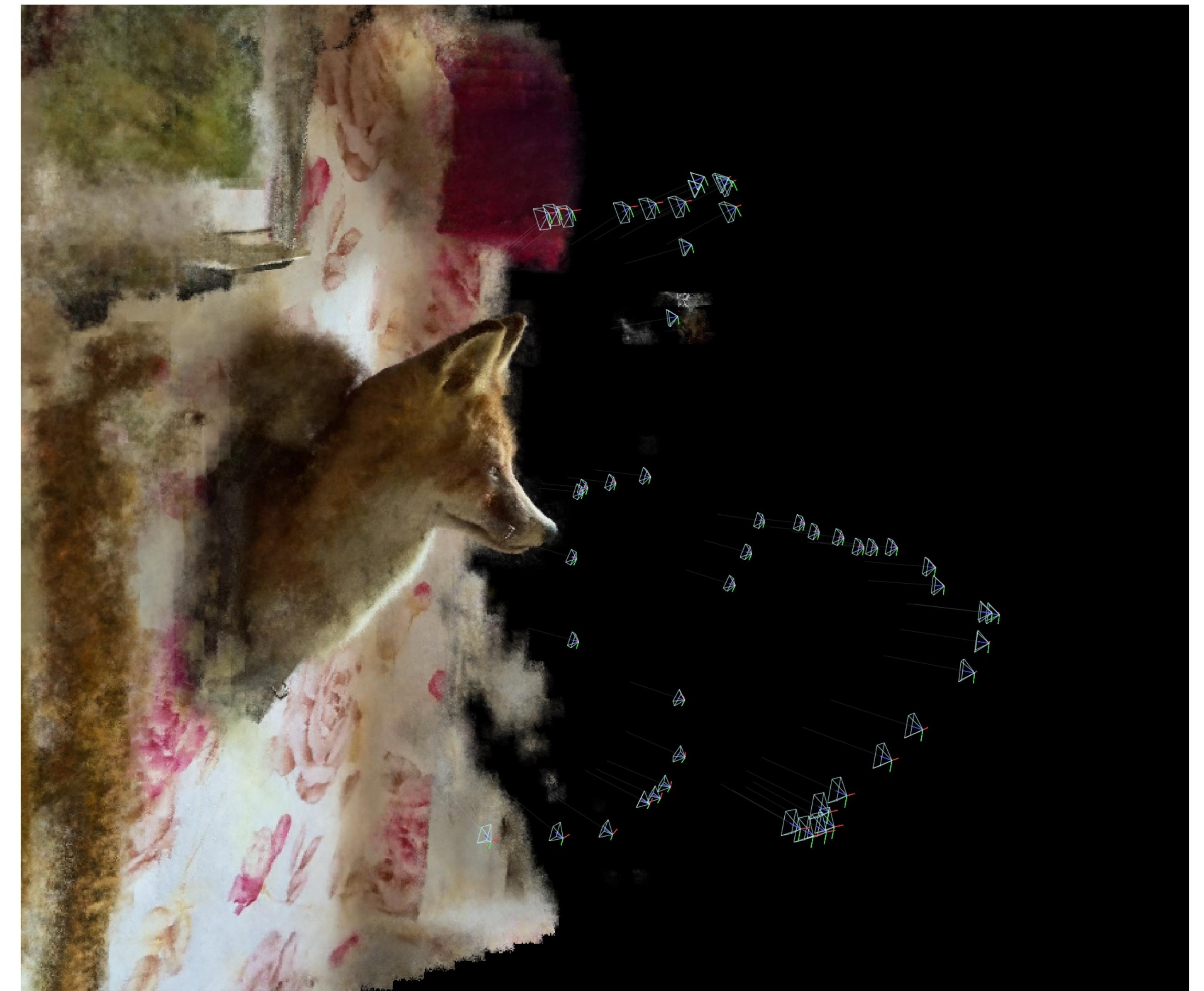
Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2020.

What is a NeRF exactly?

Light field learned by a network from 2D inputs



Note: Lighting for the scene
is baked into the NeRF
during training!

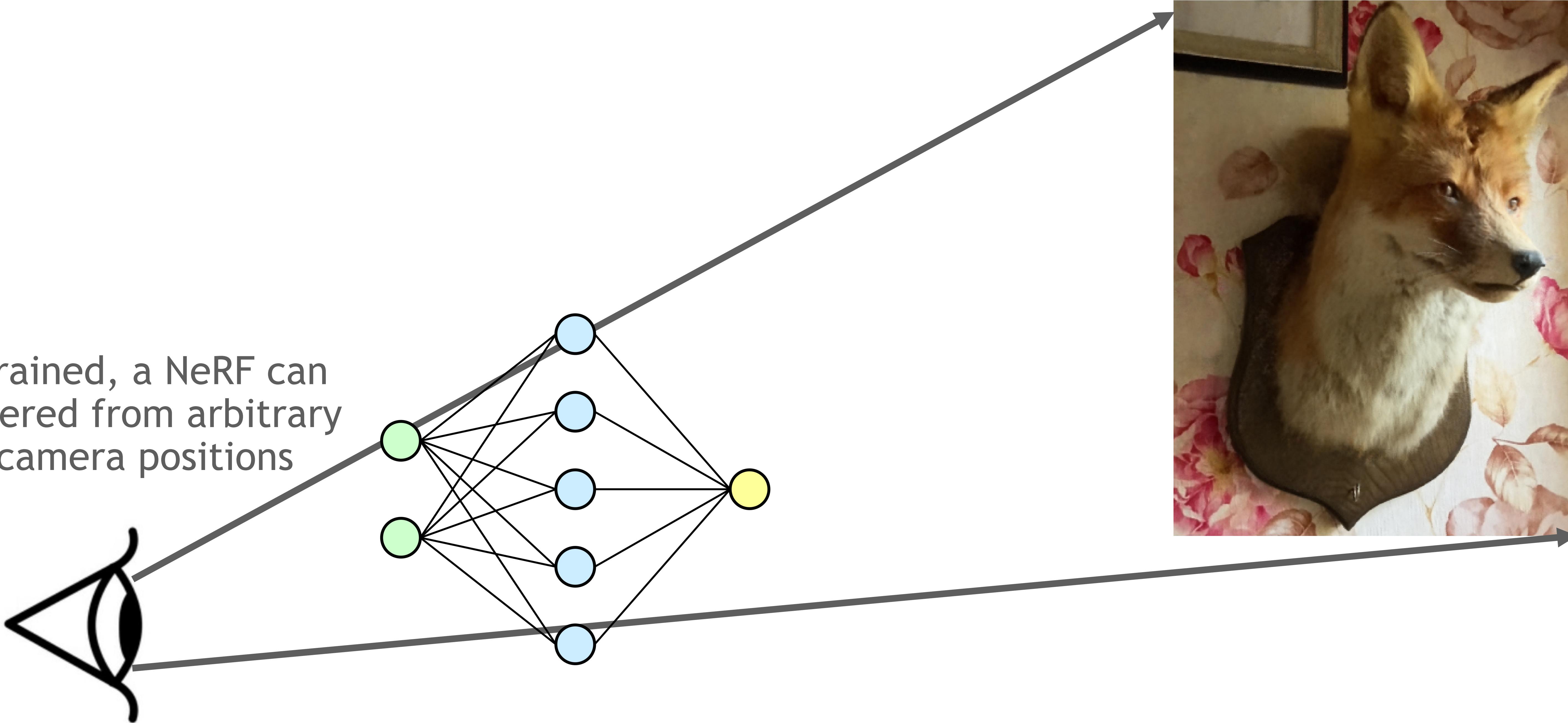


How do we render NeRFs?

Evaluate the network using a set of input rays from the camera

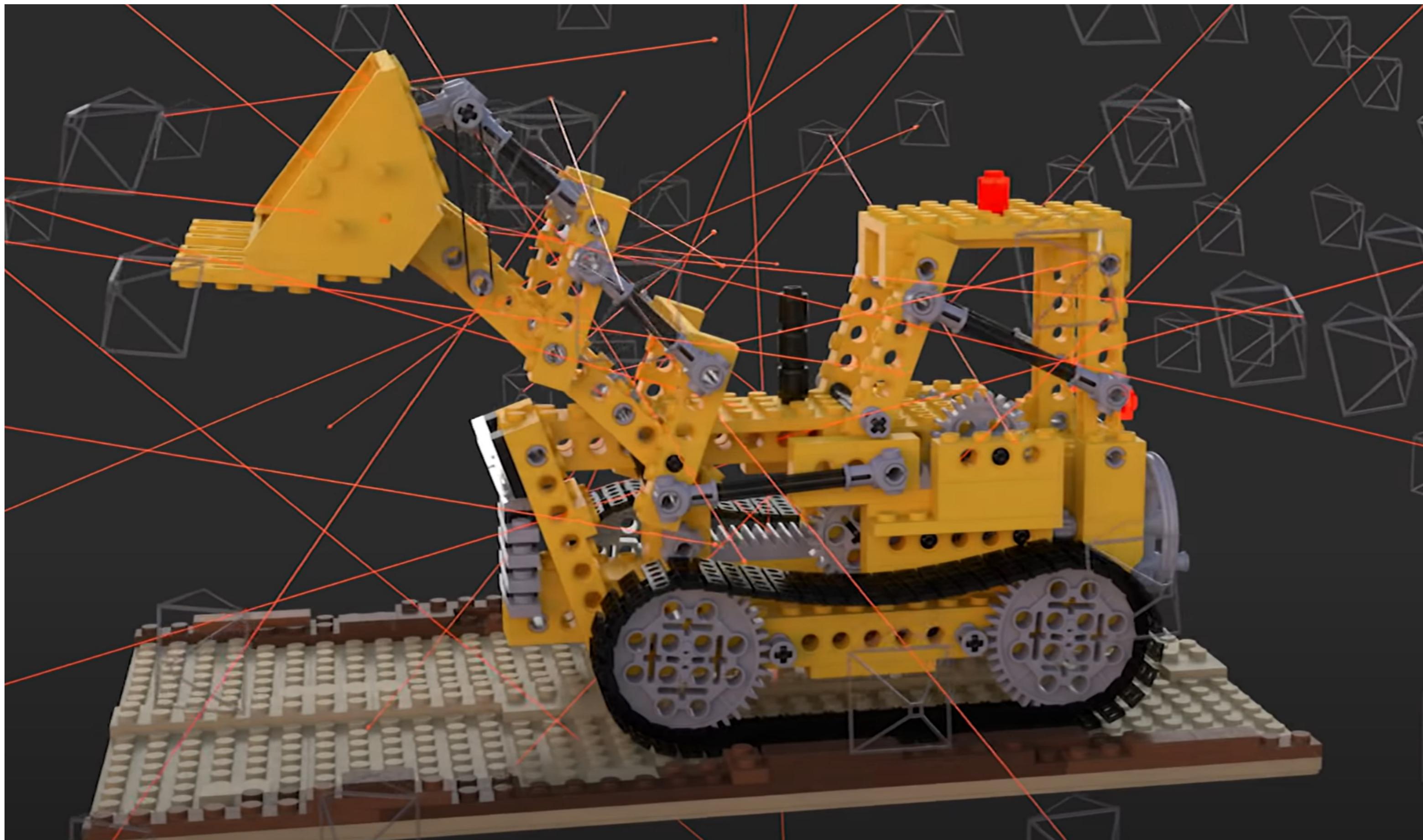
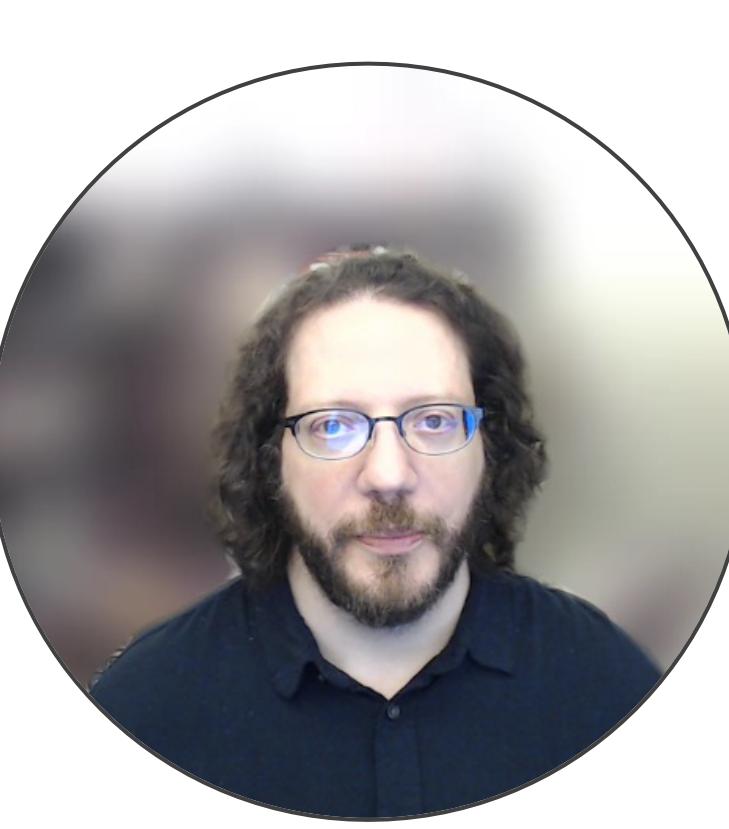


Once trained, a NeRF can
be rendered from arbitrary
new camera positions



2022: NVIDIA Instant NGP – Real Time Nerfs

TIME Magazine Named NVIDIA Instant NeRF a Best Invention of 2022



The original neural radiance fields paper from 2020 took many hours to train a model and rendered at 0.03fps

With Instant NGP, NeRFs train in seconds, and render at real time rates (30 fps or faster)

Thomas Muller, Alex Evans, Christopher Schied, and Alex Keller



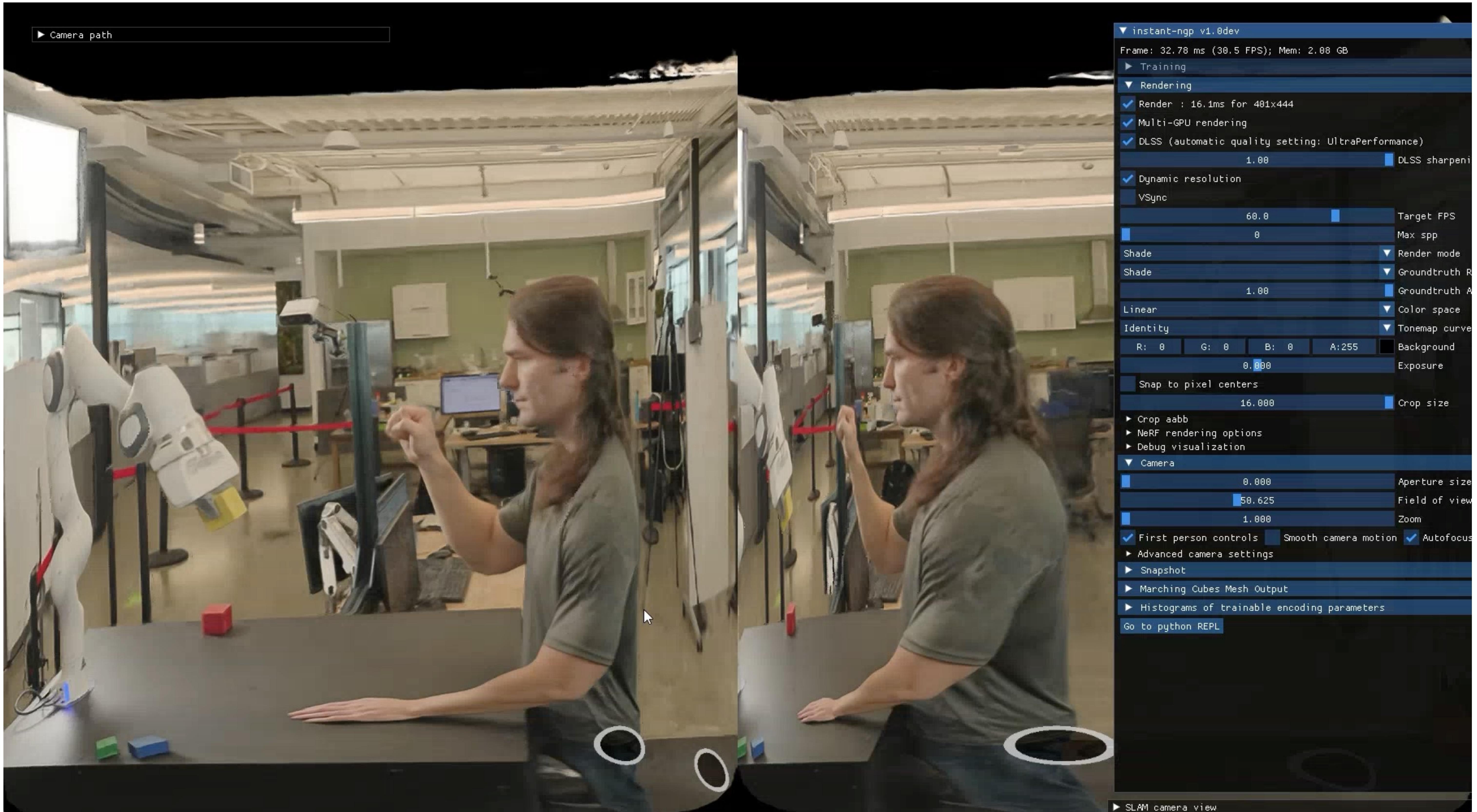
Real-time NeRF rendering



Real-time NeRF rendering

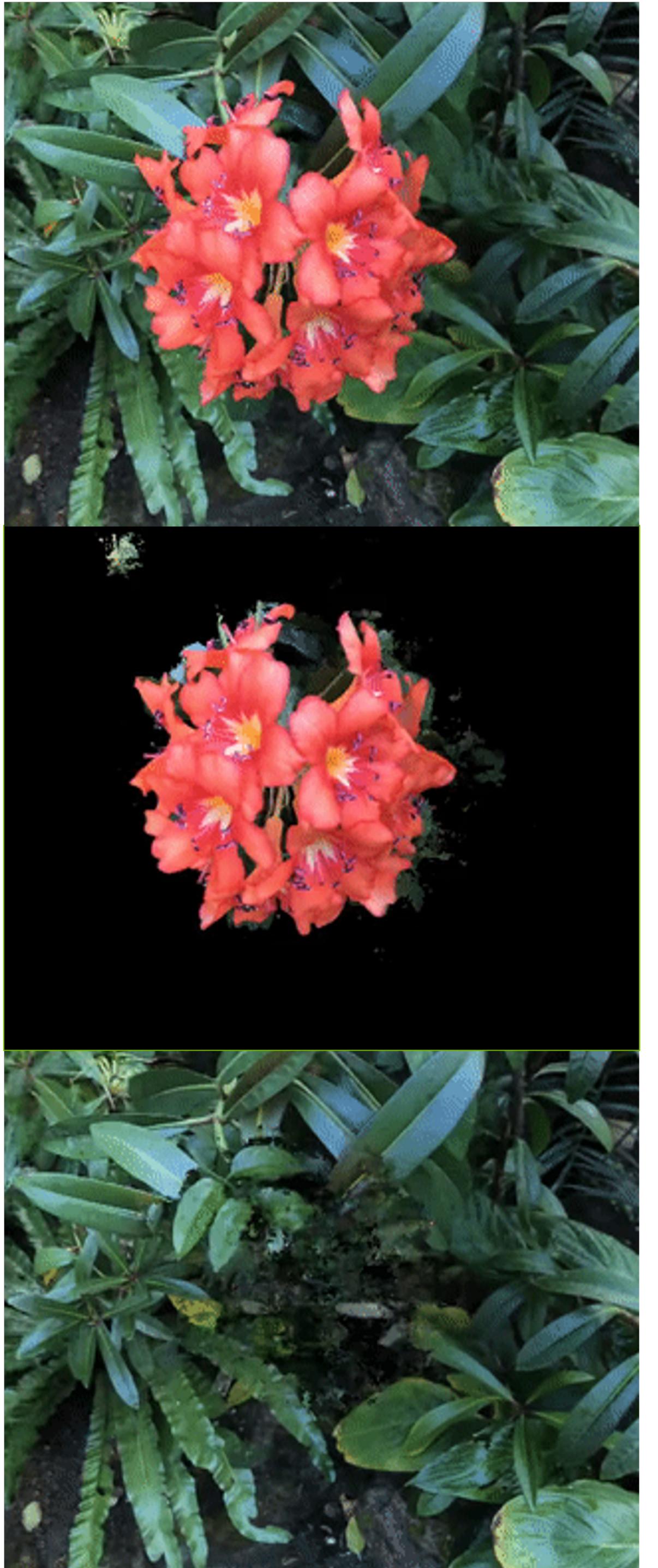
Instant NGP NeRFs in VR

With Editing!

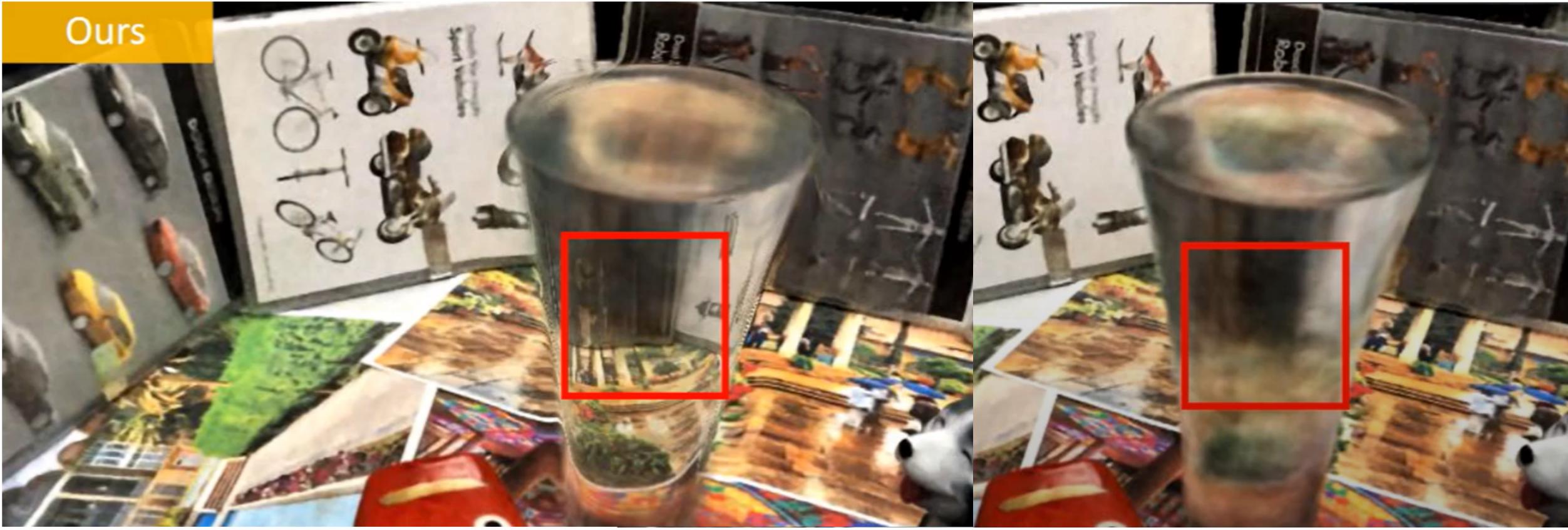


Limitations of today are tomorrow's applied research

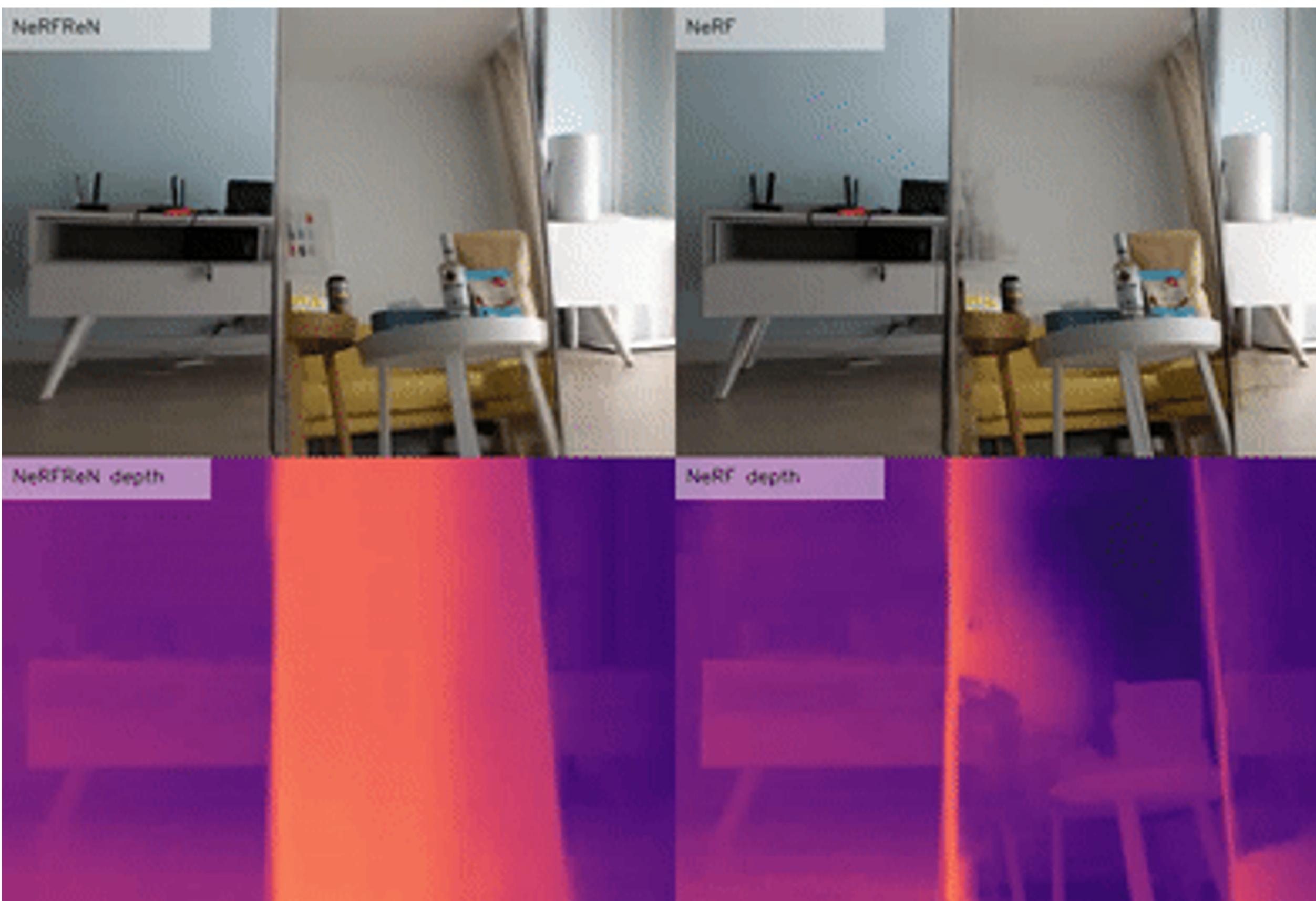
Over 140 research papers related to NeRFs in 2022 alone



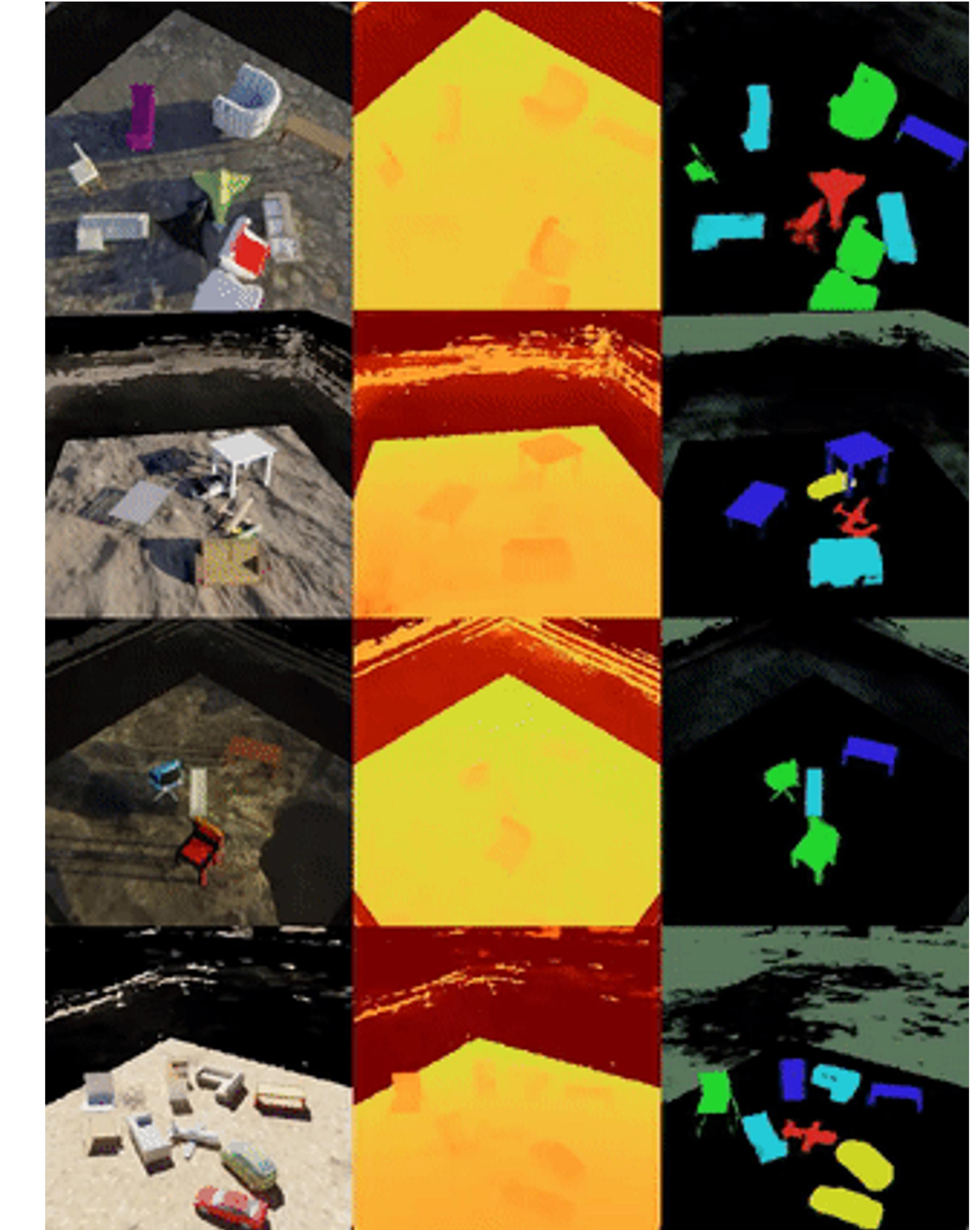
Decomposing NeRF for Editing via
Feature Field Distillation
Kobayashi et al NeurIPS 2022



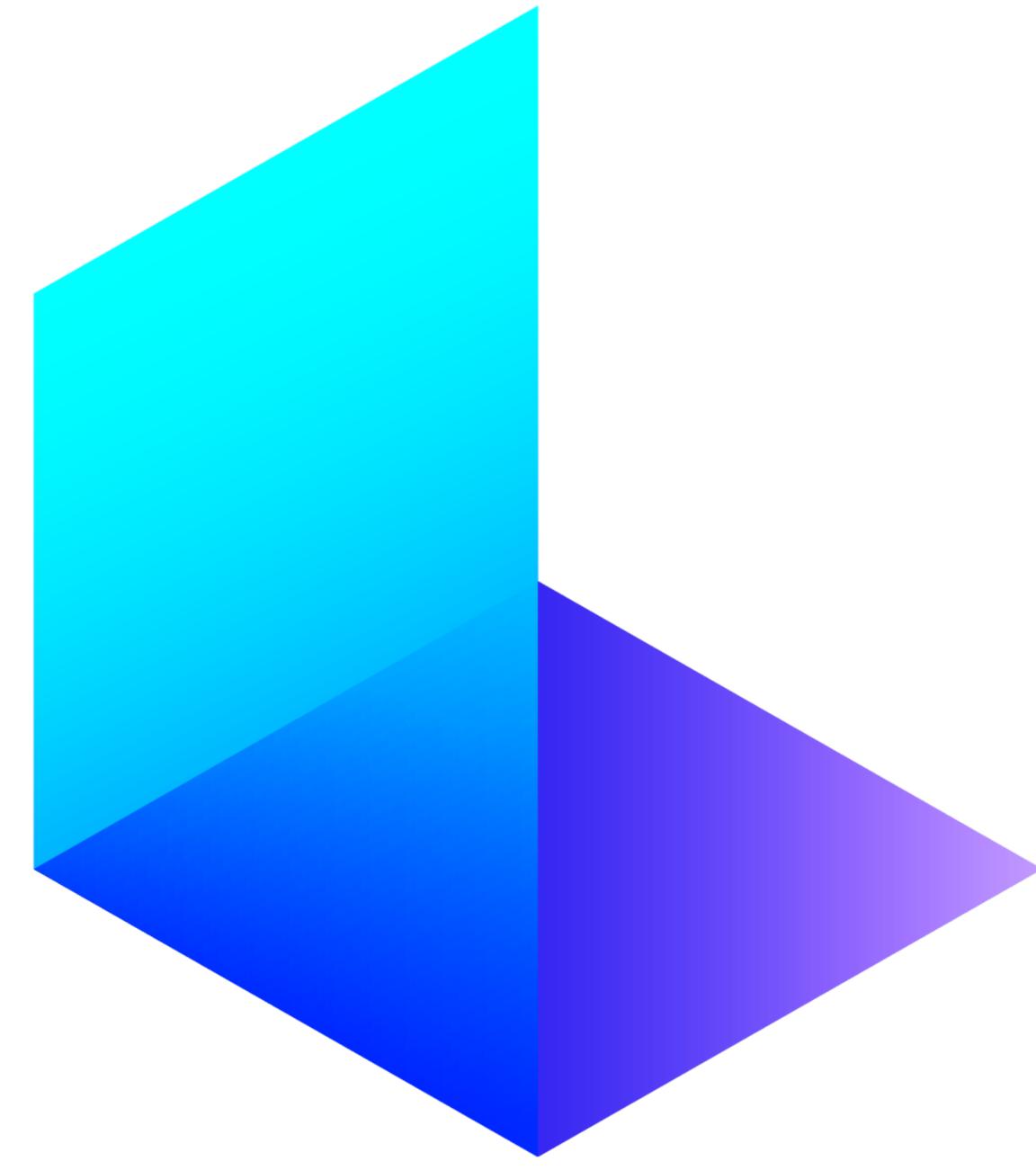
Eikonal Fields for Refractive Novel-View Synthesis
Bemana et al SIGGRAPH 2022



NeRFReN Neural Radiance Fields with Reflections
Guo et al CVPR 2022

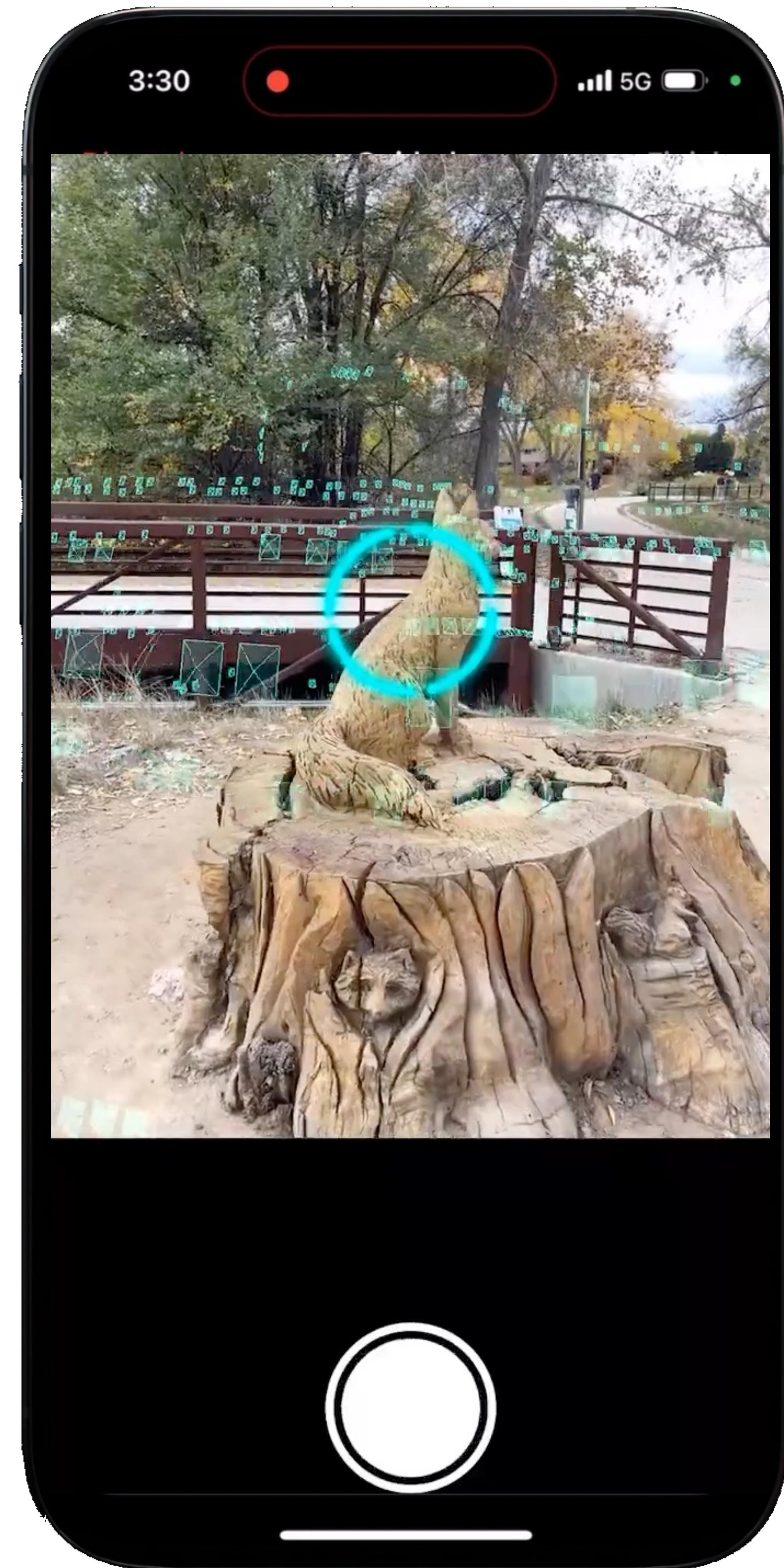


NeSF: Neural Semantic Fields for Generalizable Semantic
Segmentation of 3D Scenes
Vora, Radwan, et al TMLR 2022



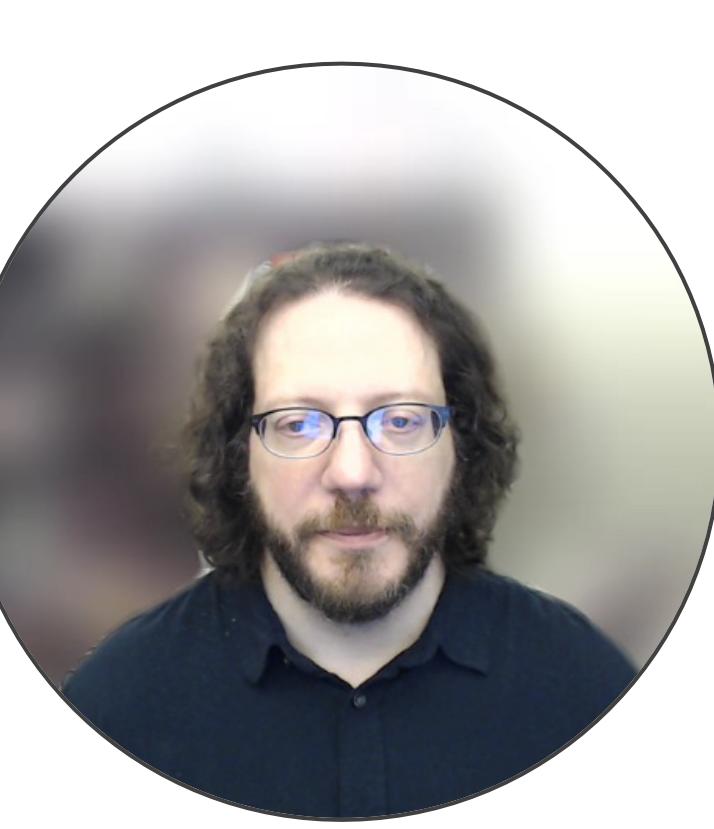
NeRF capture tools beginning to appear broadly

Luma AI iPhone Capture Tool Example



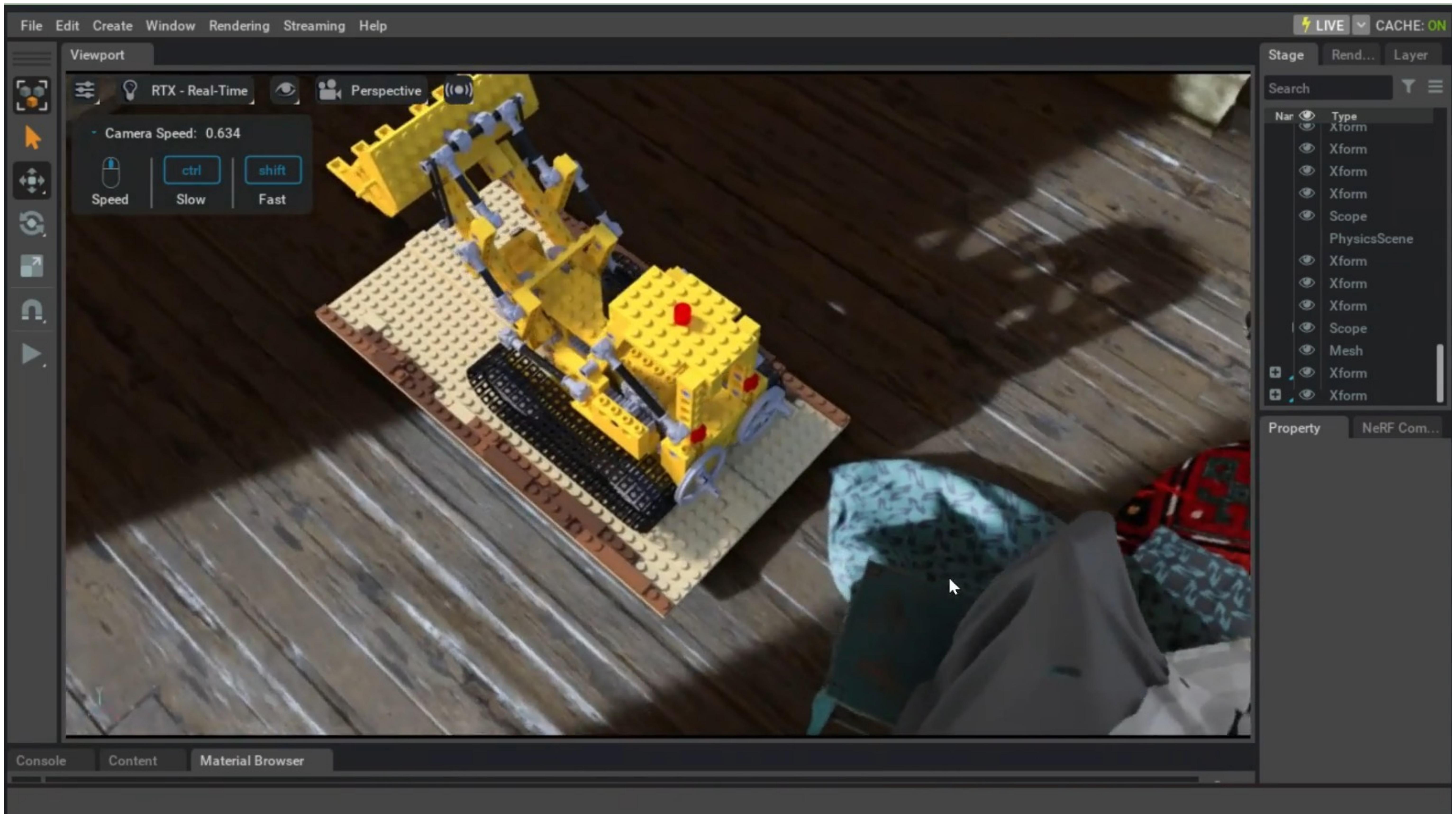
NeRF Environments + NeRF Object

Two ways to use NeRF in a virtual world



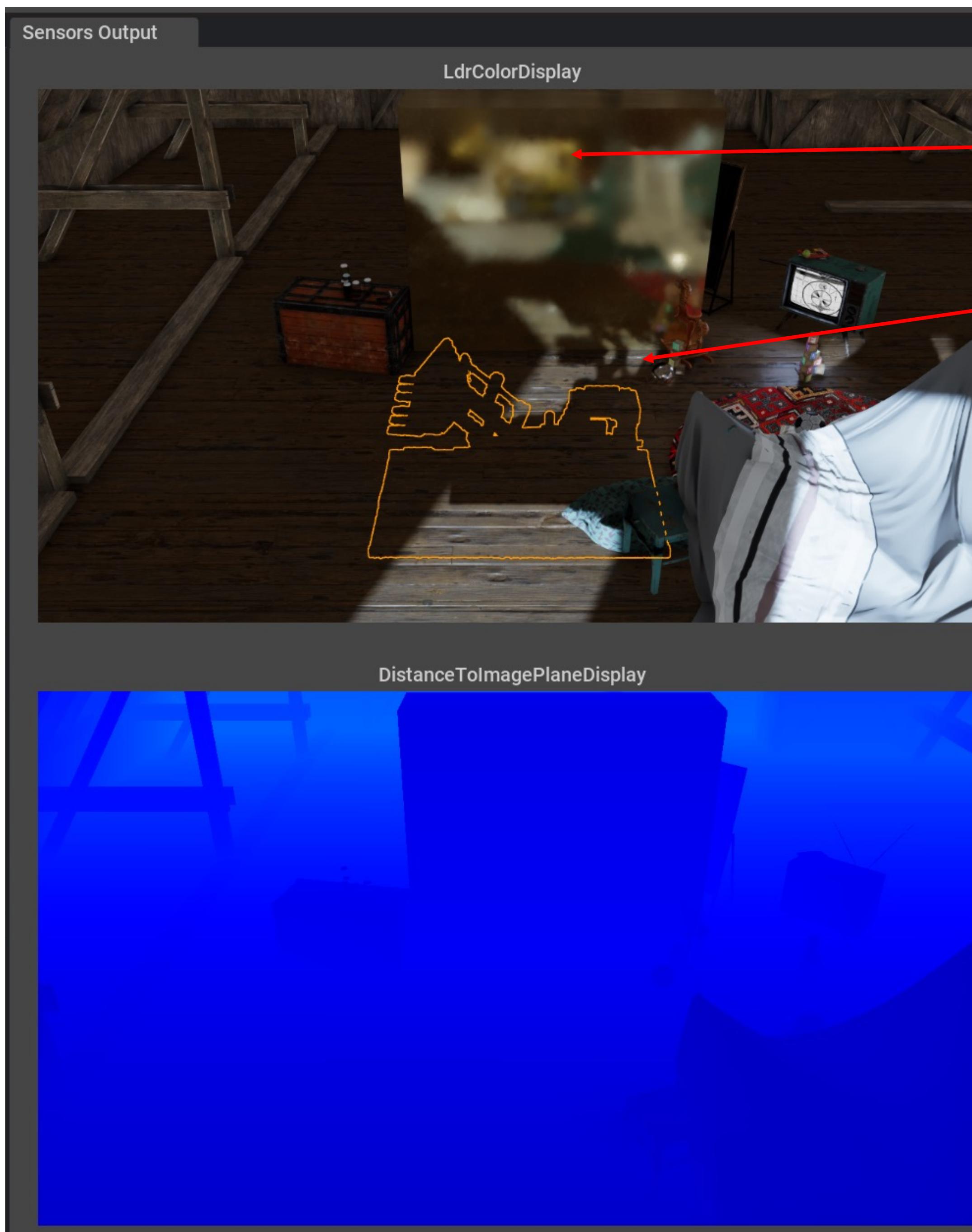
NeRF Objects In Omniverse

Development Prototype



NeRF Rendering

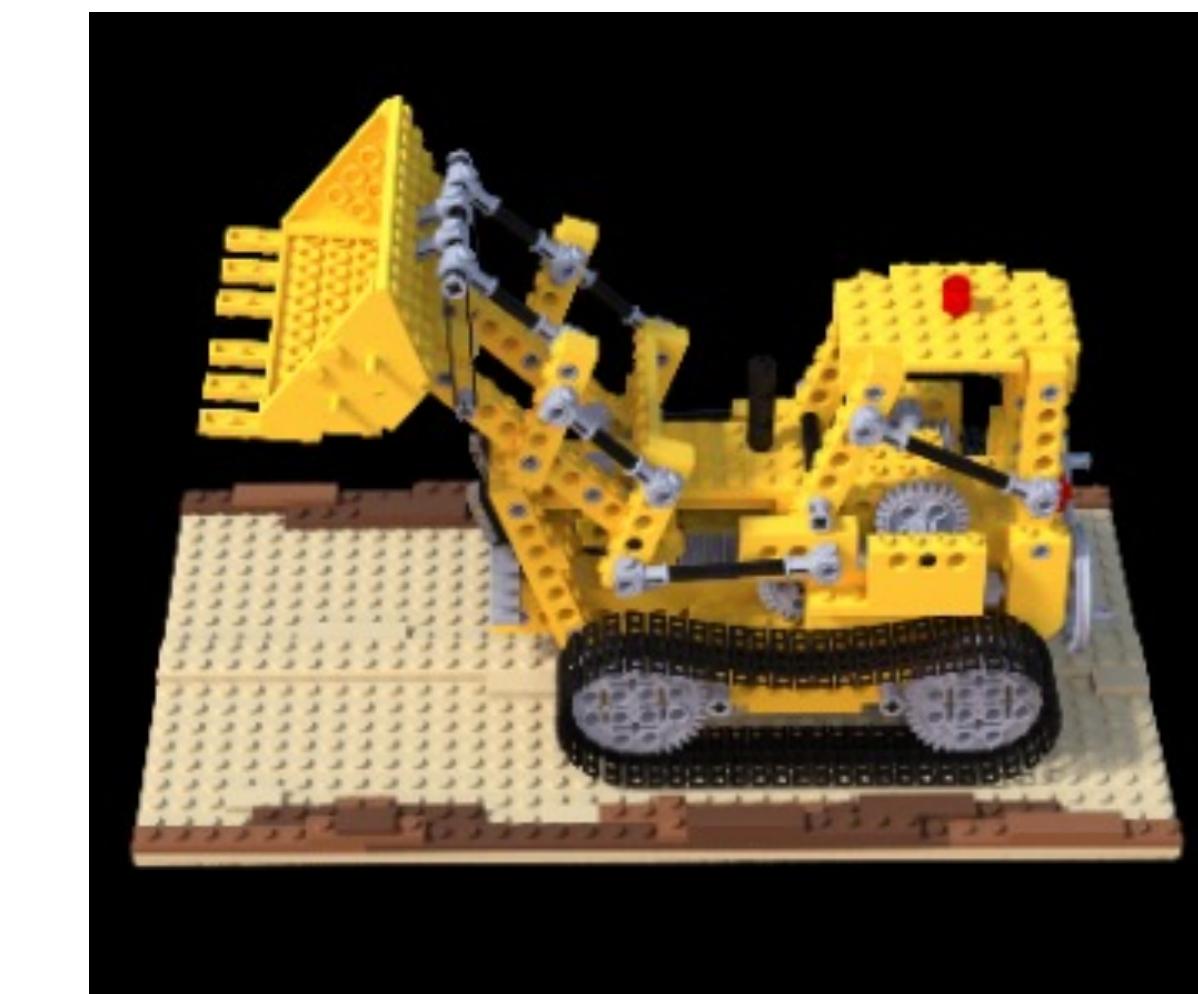
Depth Based Composition + Proxy Mesh



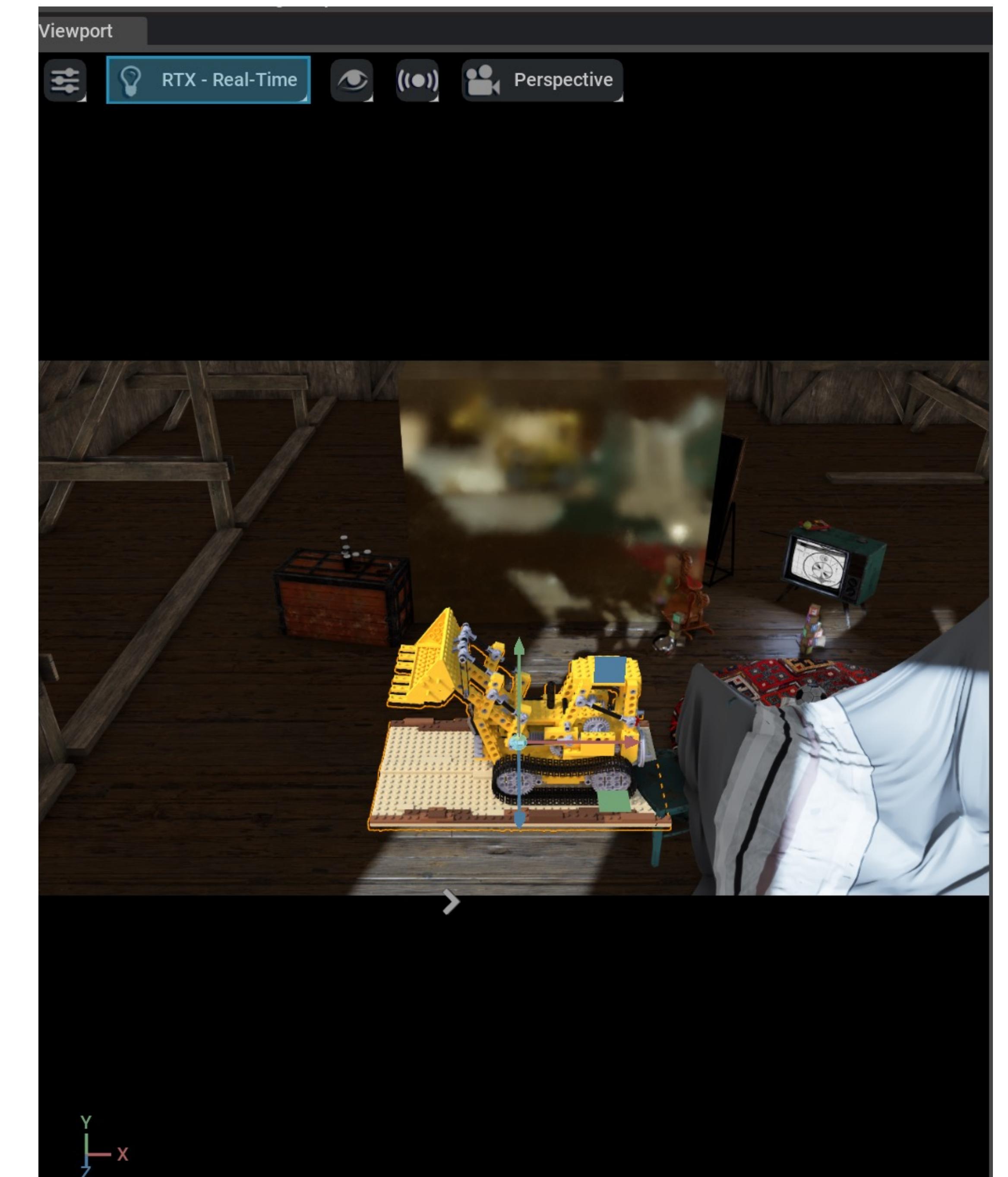
Reflection from Proxy

Shadow from Proxy

+



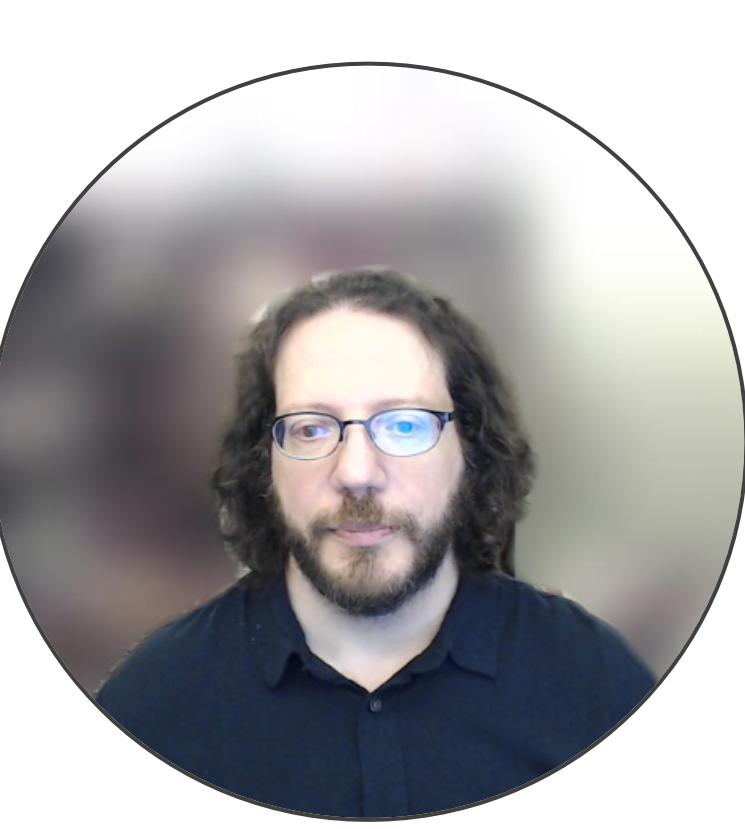
Inferenced NeRF



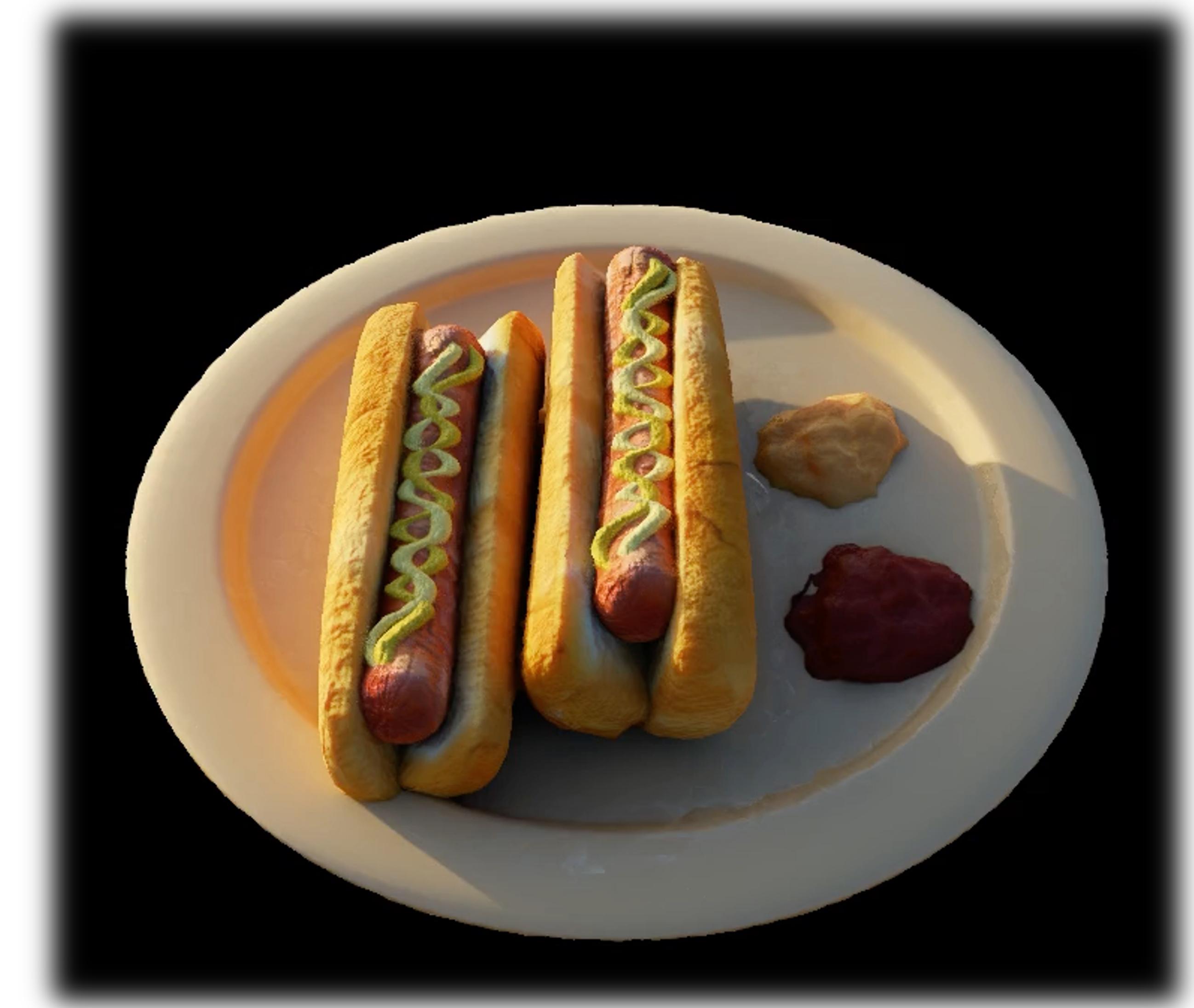
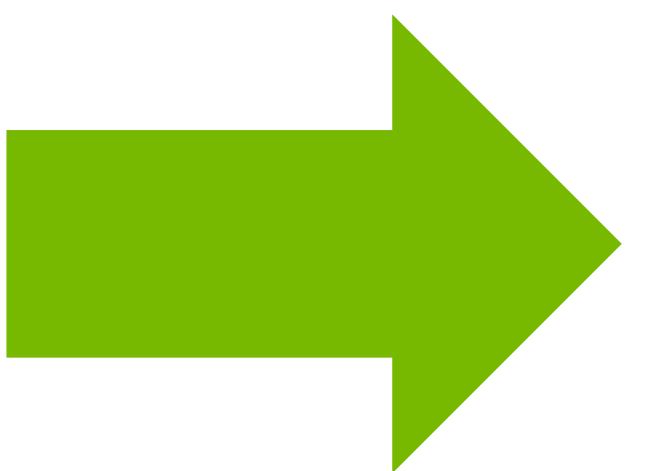
RTX Renderer RGB +
Depth Output

Composed Final
Output

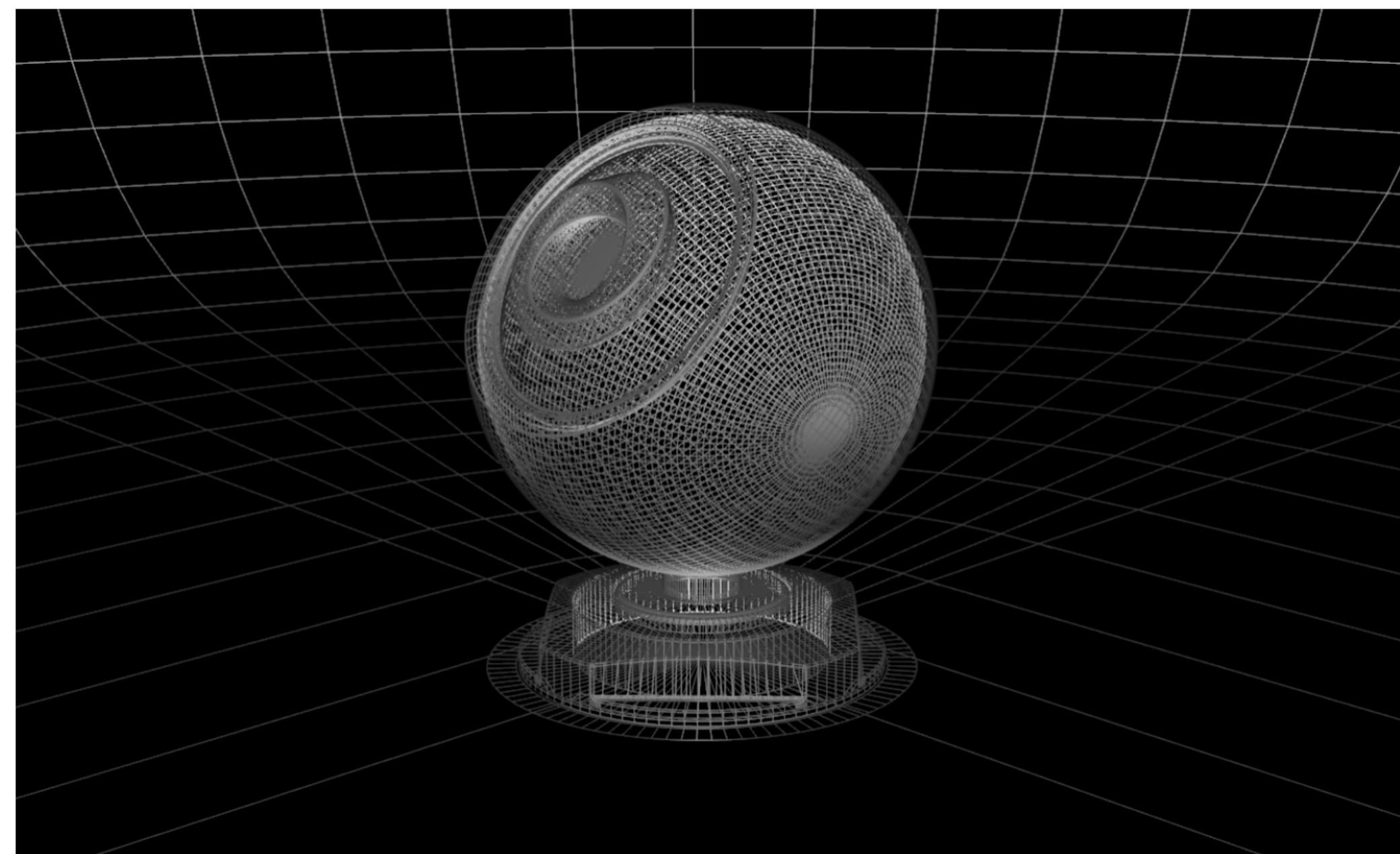
Standardization for Production



3D Capture with Traditional Asset Formats



Forward (Traditional) 3D Rendering



Mesh



Material



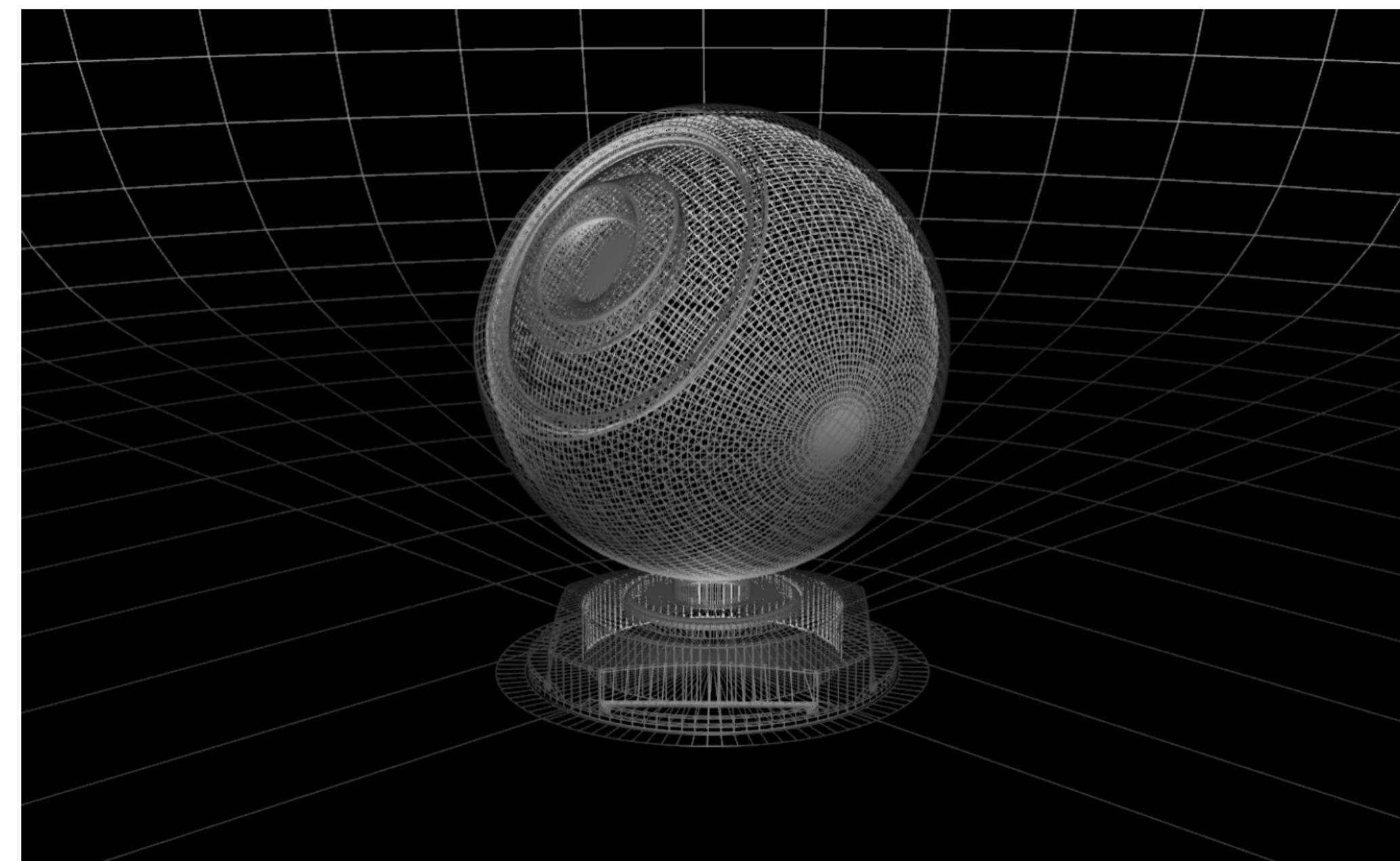
Renderer



Output



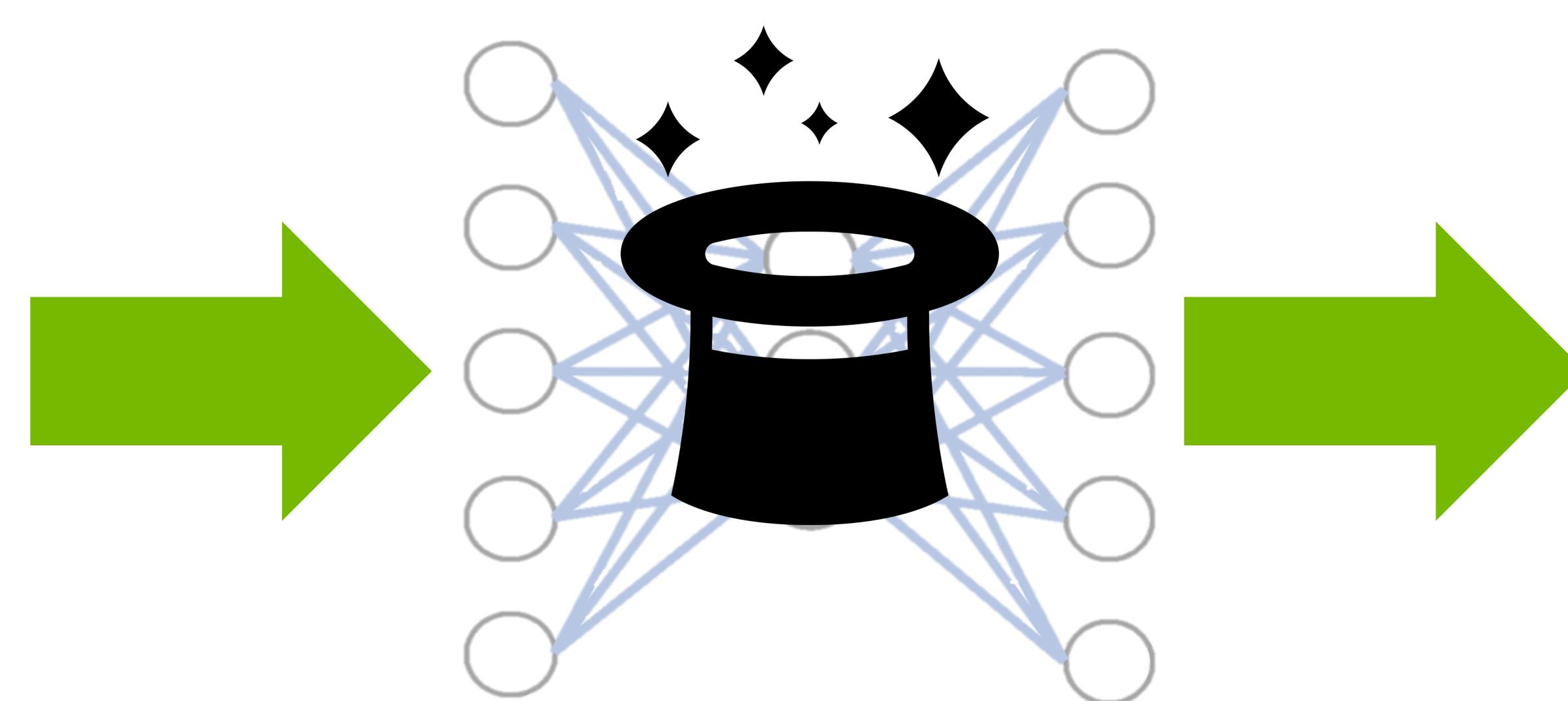
Differentiable Rendering



Mesh



Material



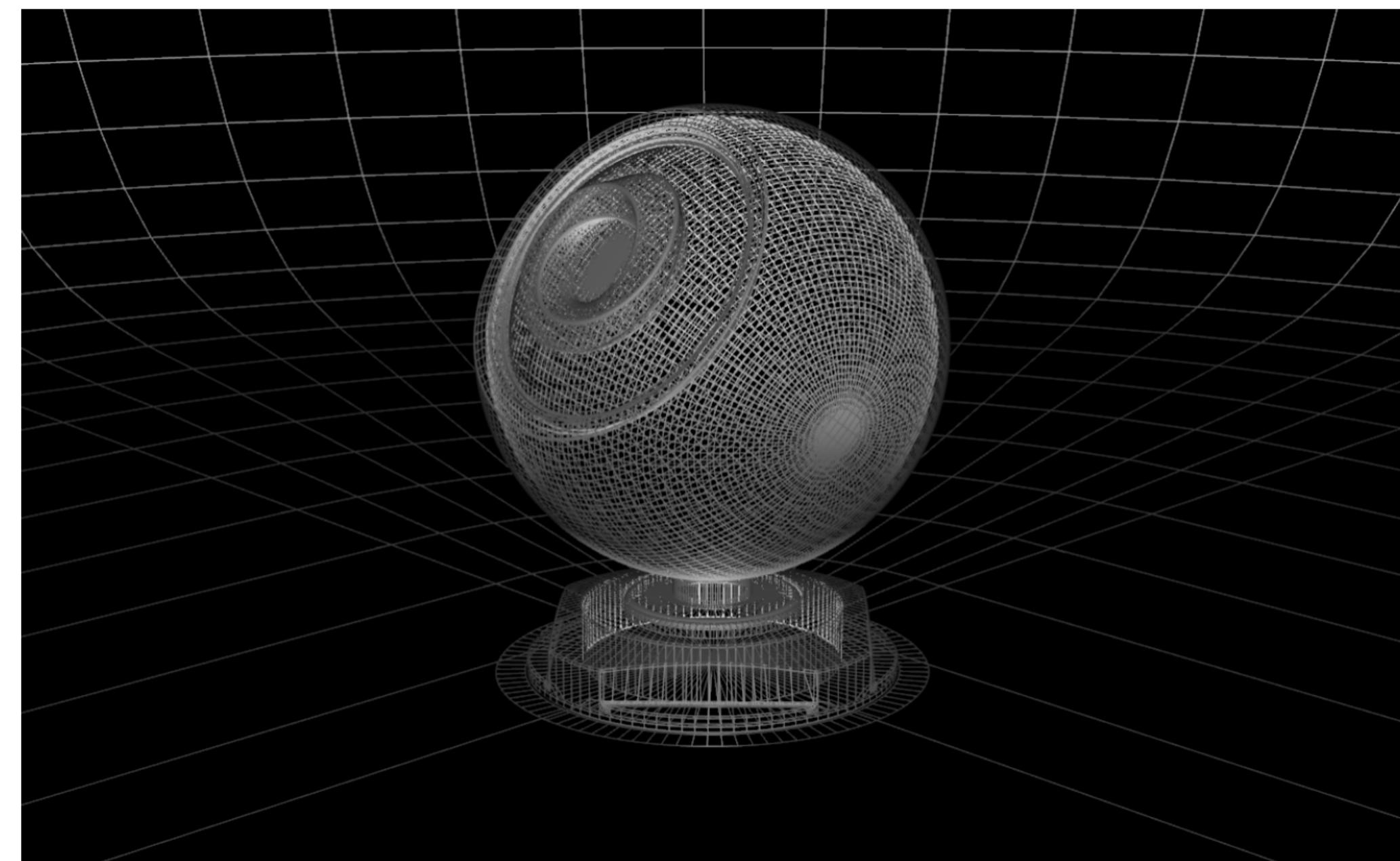
Renderer



Output

NVDiffRast: <https://nvlabs.github.io/nvdiffrast/>
DIB-R: <https://nv-tlabs.github.io/DIB-R/>

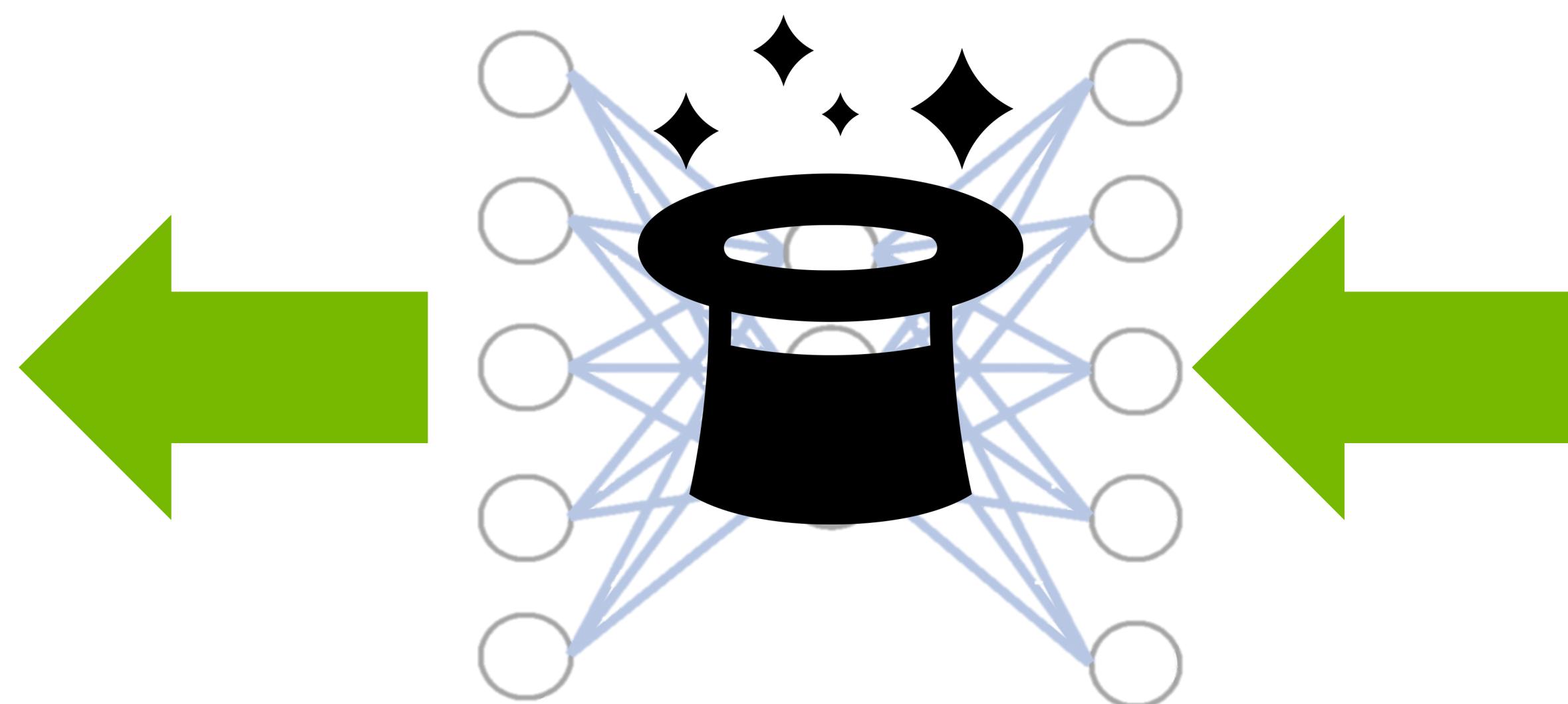
Differentiable Rendering – Inverse Graphics



Mesh



Material

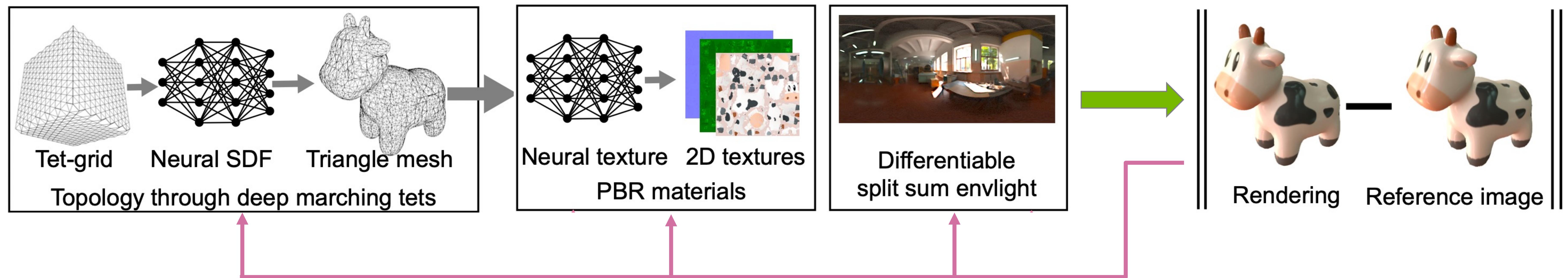


Renderer



Output

3D MoMa / nvdiffrec

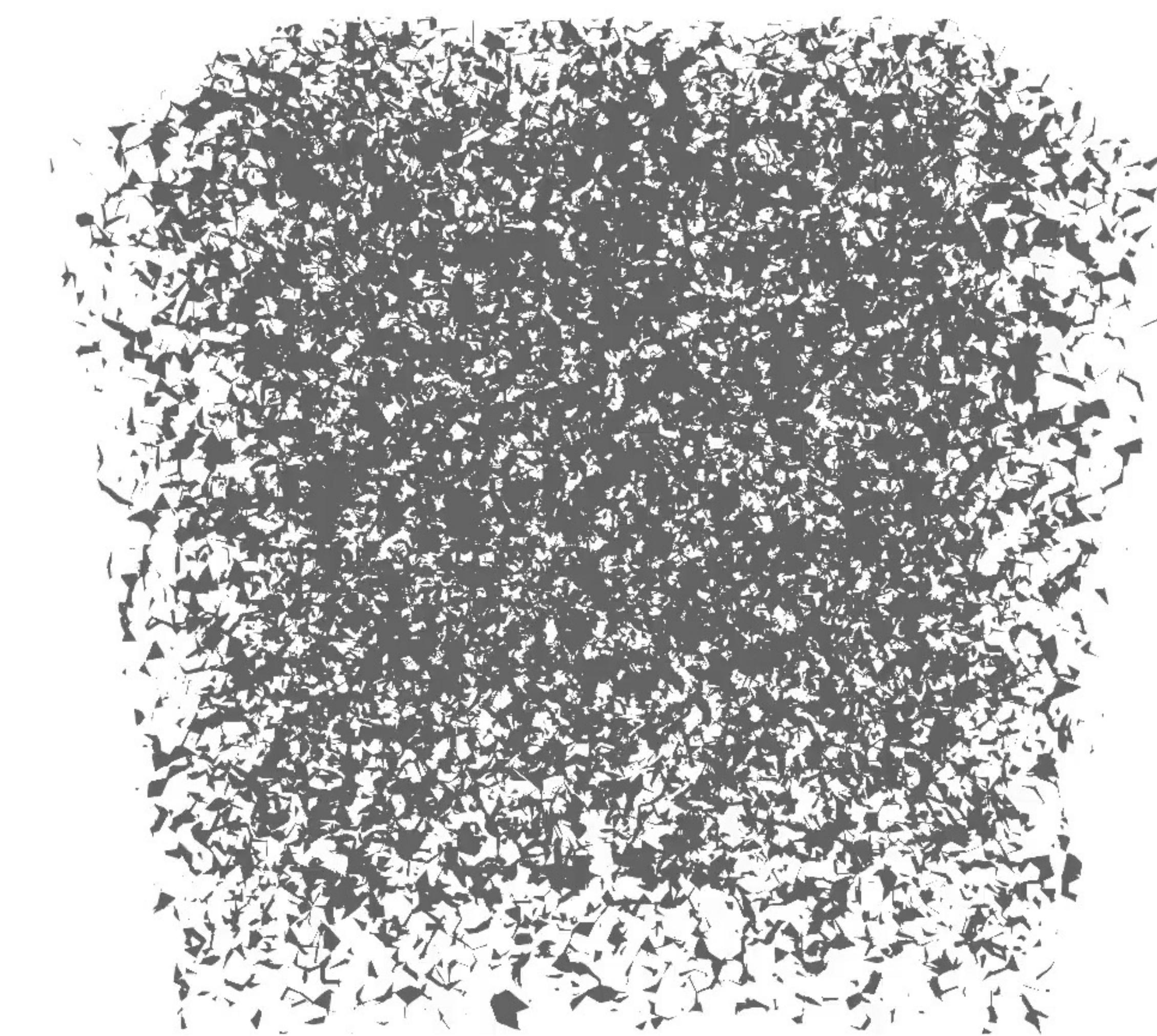


Extracting Triangular 3D Models, Materials, and Lighting From Images
Munkberg, Hasselgren, Shen, Gao, Chen, Evans, Muller, Fidler
CVPR 2022 (Oral)

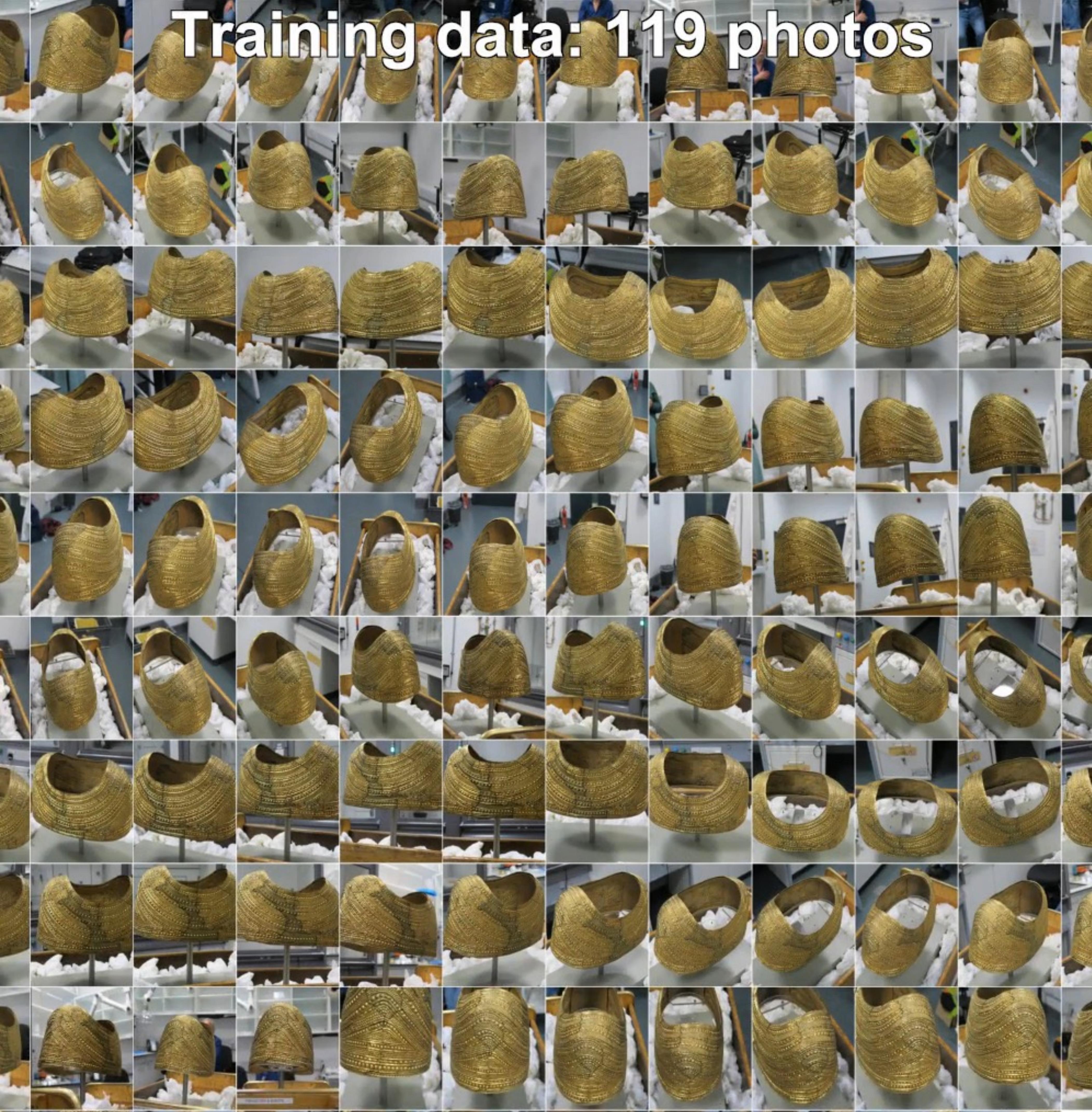


Learning Example

Iteration: 0

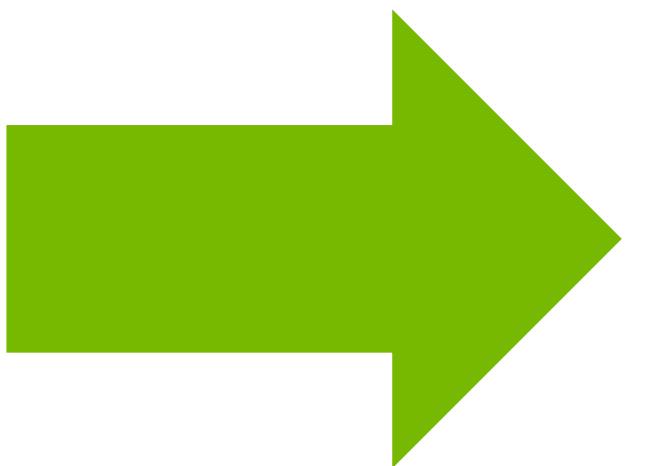
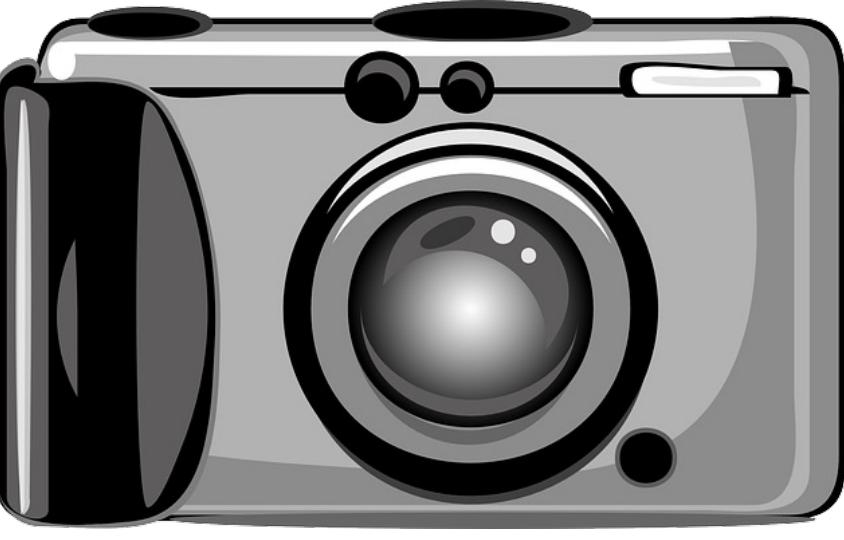
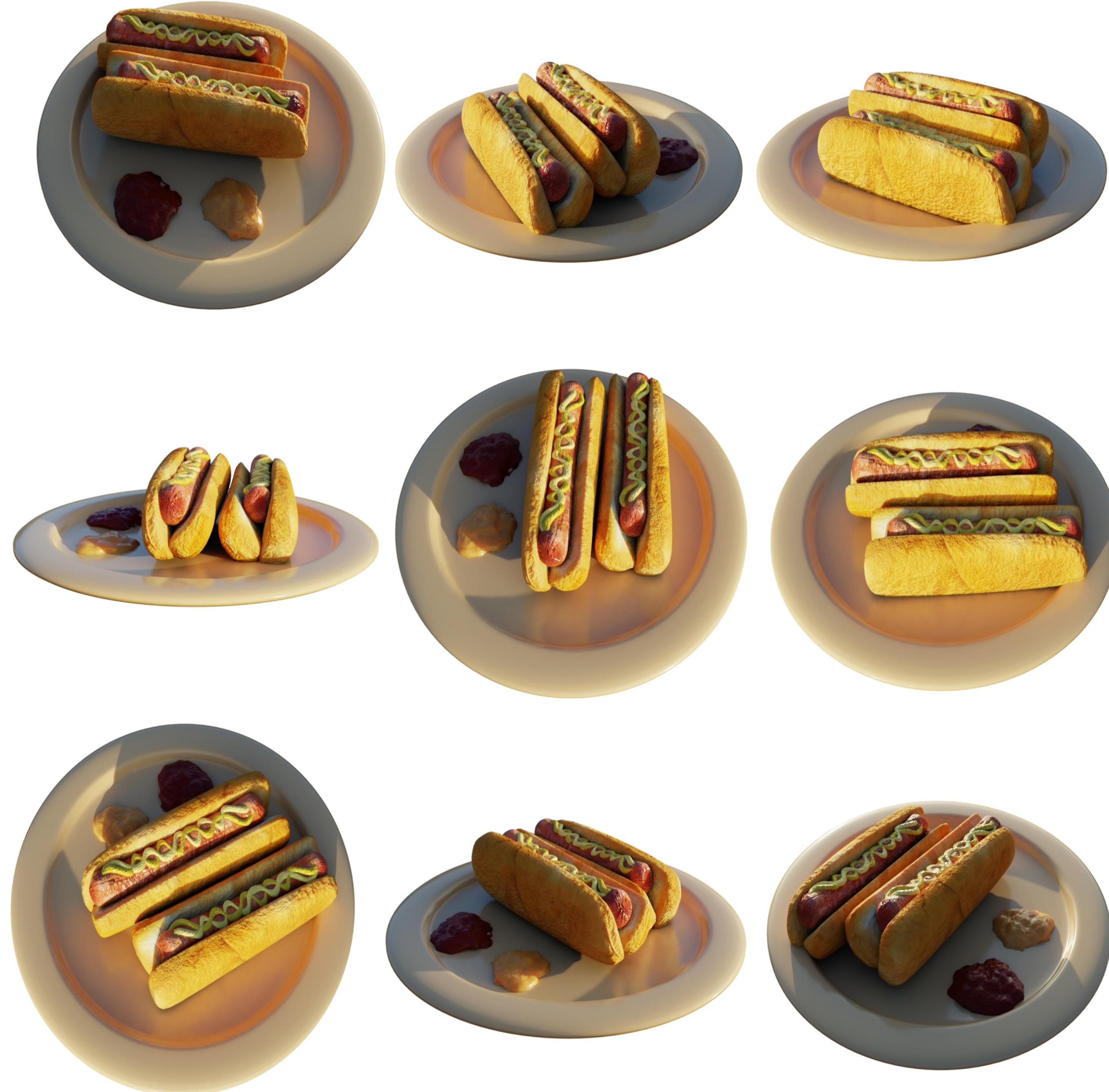


Training data: 119 photos



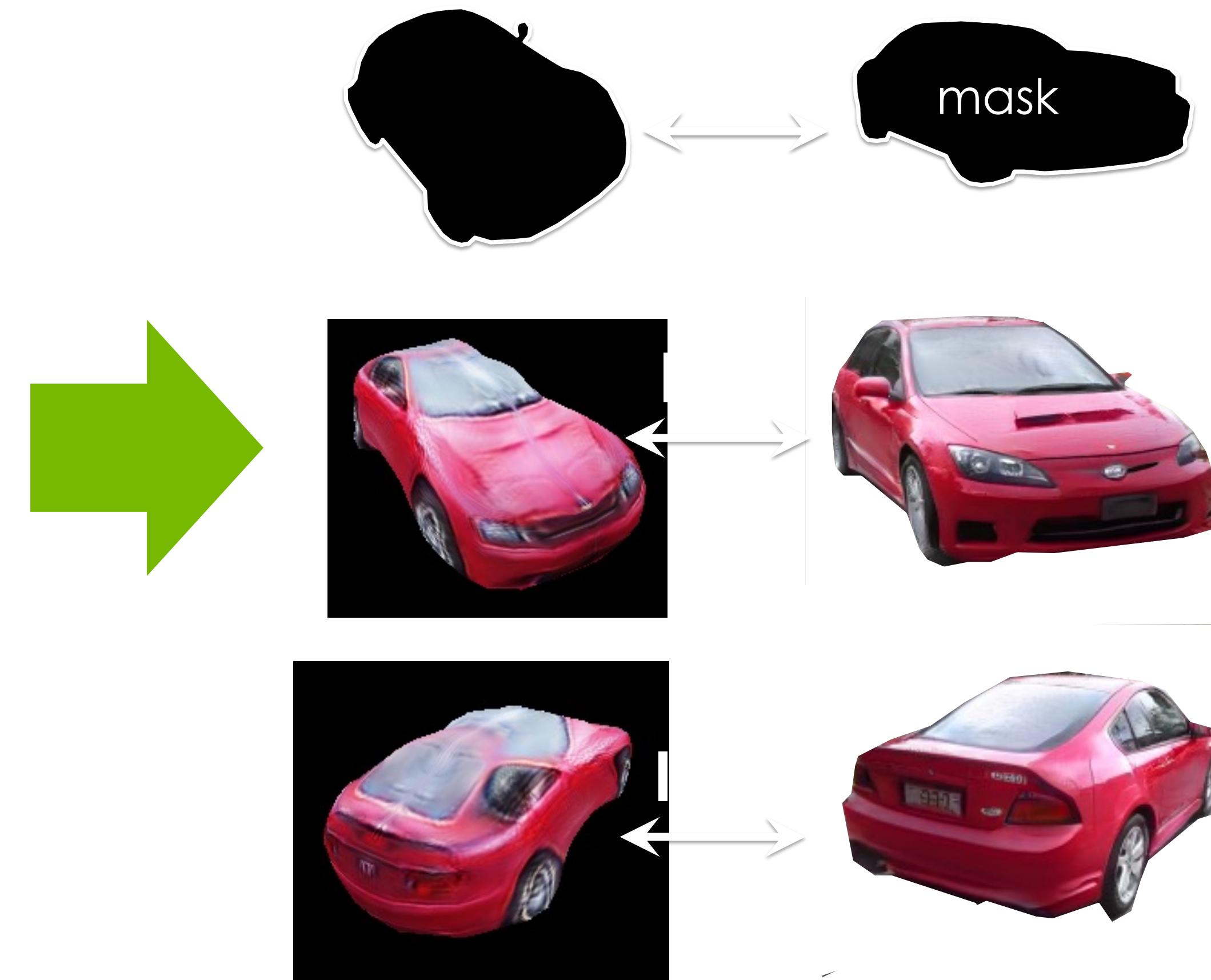
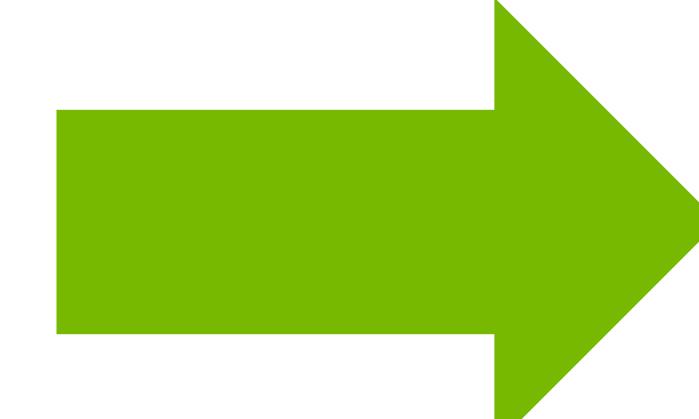
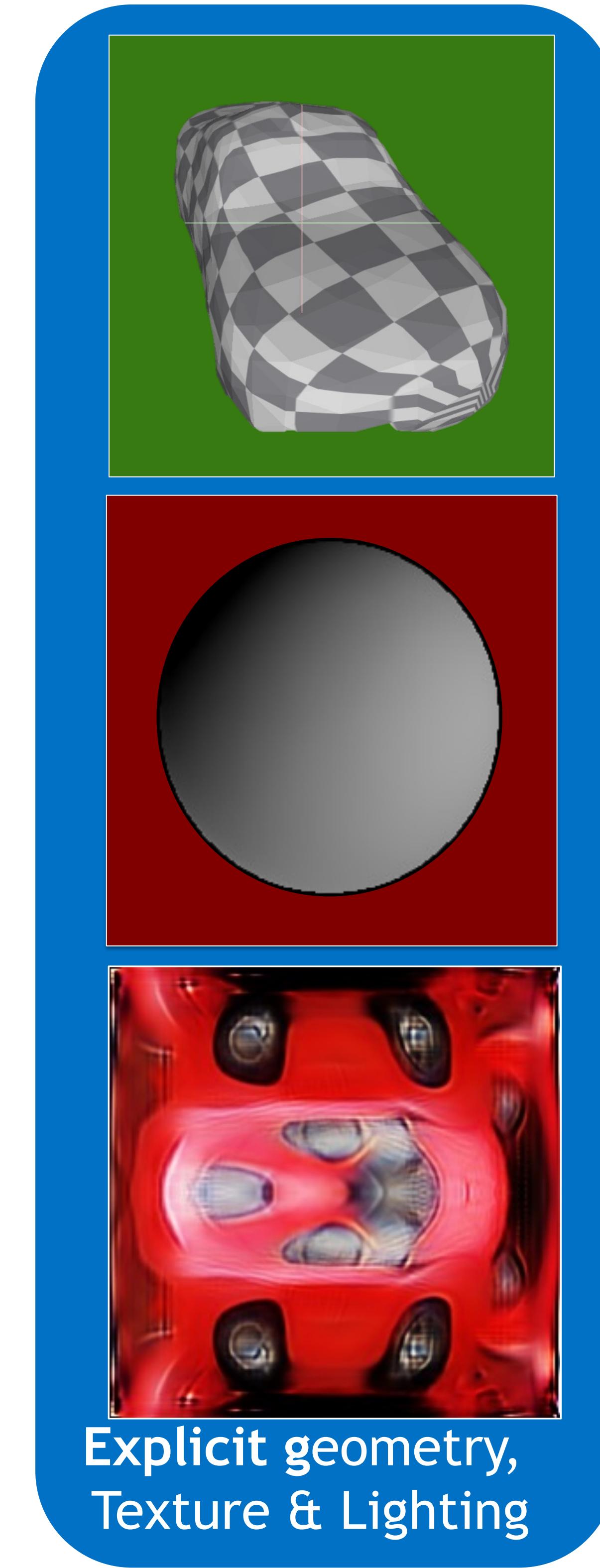
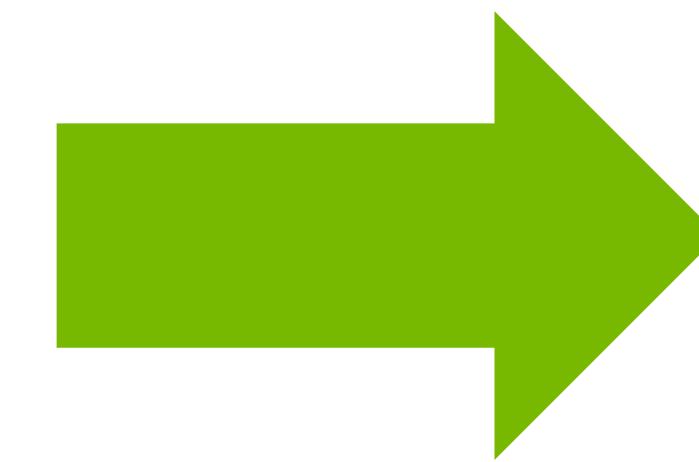
Real World Capture

3D Capture from a Single Image



GANverse 3D

Generative Adversarial Networks + Differentiable Rendering



Training data: images only!

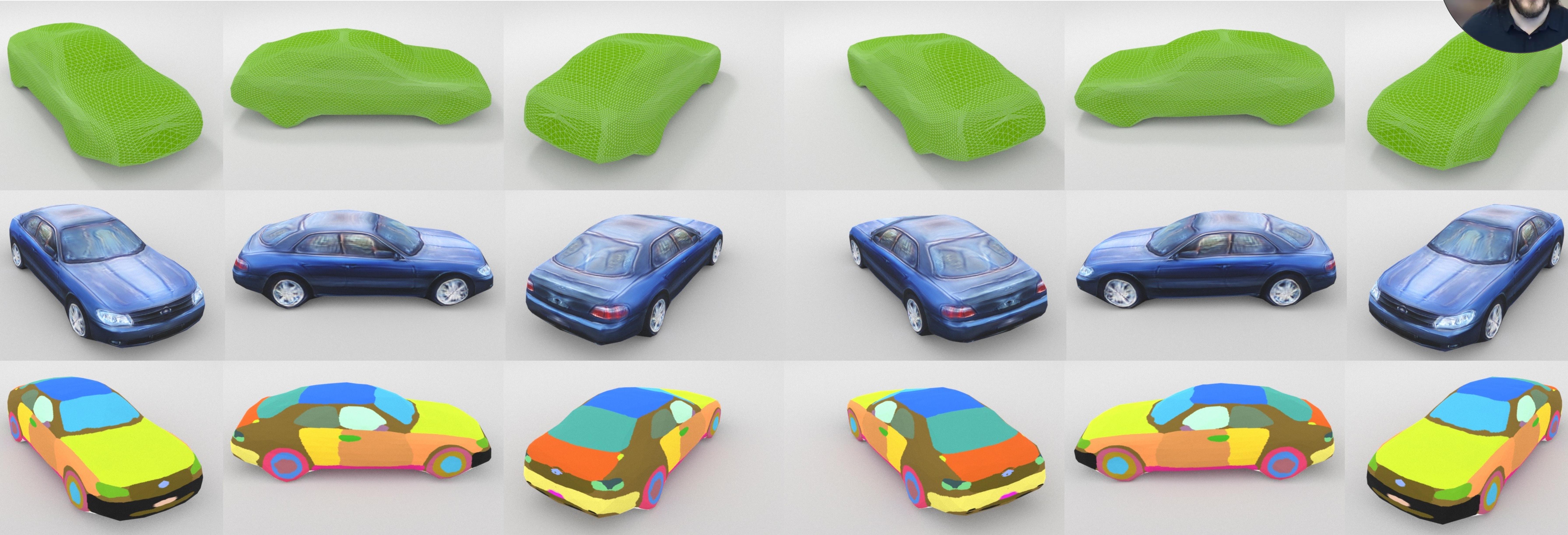
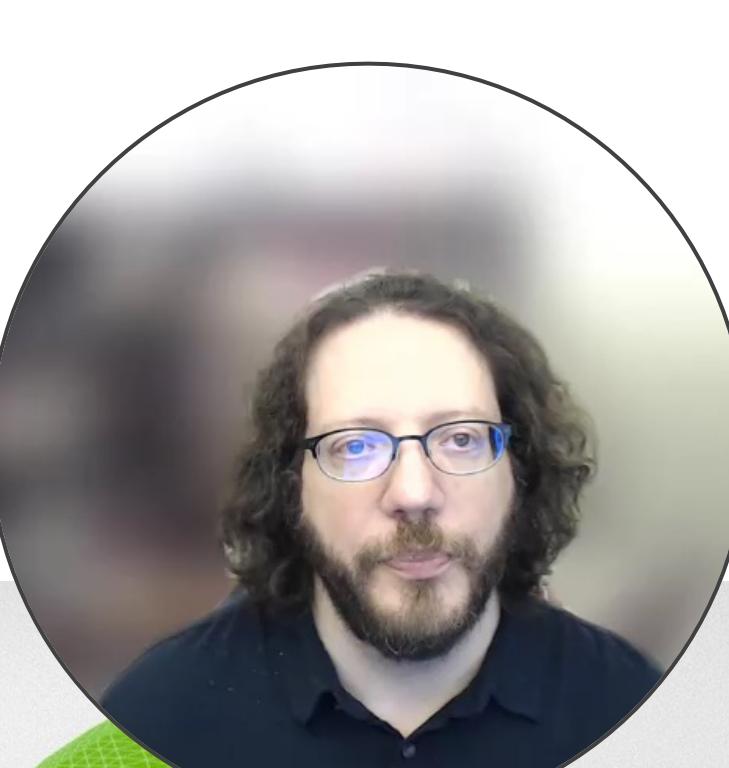
No 3D needed!



- Car: 1,422,984
- Van: 2,674,607
- Bus: 4,799,402
- Building: 1,638,425
- Bicycle: 3,071,418
- Tricycle: 111,341
- Traffic sign: 150,647
- Dog: 1,731,929
- Human: 3,276,718
- Pedestrian: 659,542
- Person: 507,506
- Skater: 2,454,728
- Skateboard: 1,073,733

The image displays four screenshots of the Flickr website interface, each showing a grid of images related to a specific search term. The top-left screenshot shows search results for 'cars', displaying various types of vehicles including sedans, SUVs, and racing cars. The top-right screenshot shows results for 'bus', featuring several double-decker buses and a yellow school bus. The bottom-left screenshot shows results for 'building', with images of architectural structures like palaces, modern skyscrapers, and rural houses. The bottom-right screenshot shows results for 'dog', with a variety of dog breeds in different settings. Each screenshot includes the Flickr logo at the top, a search bar with the query, and navigation options like 'Photos', 'People', 'Groups', and 'Advanced' filters.

GANverse3D: 3D Content Creation from Images



- 3D (mesh)
- texture
- part map

Viewport

Camera RTX Real-time

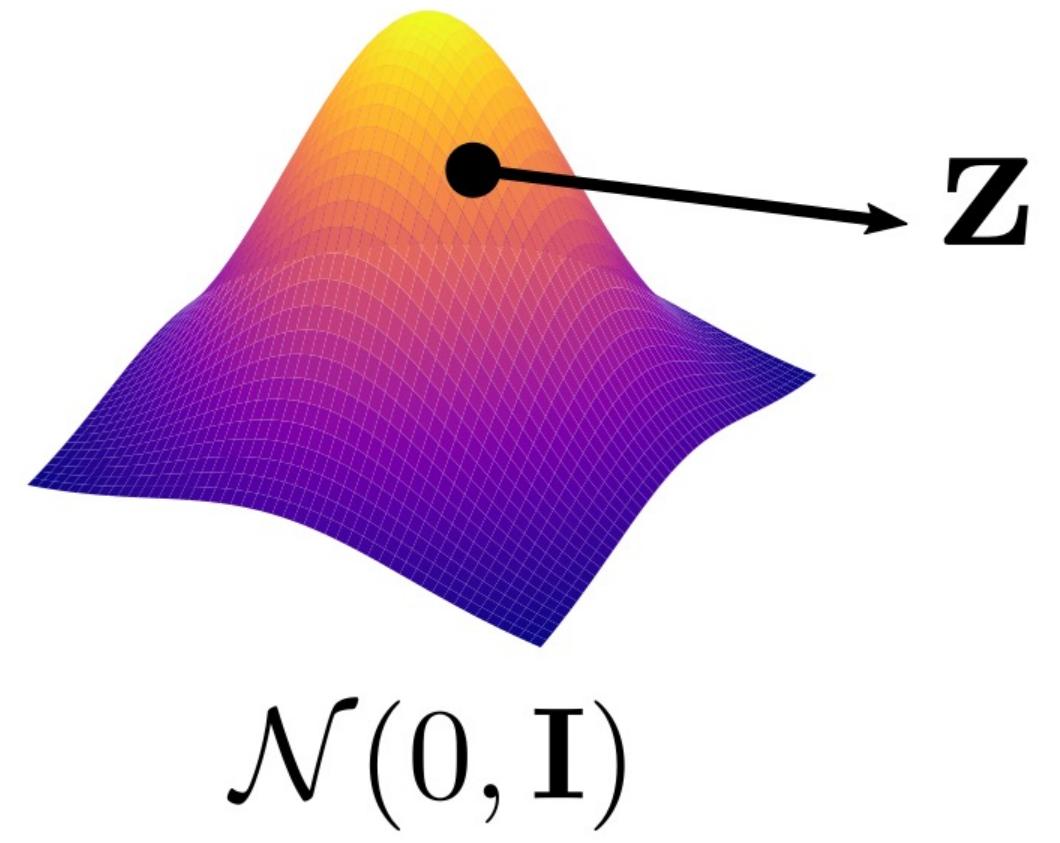
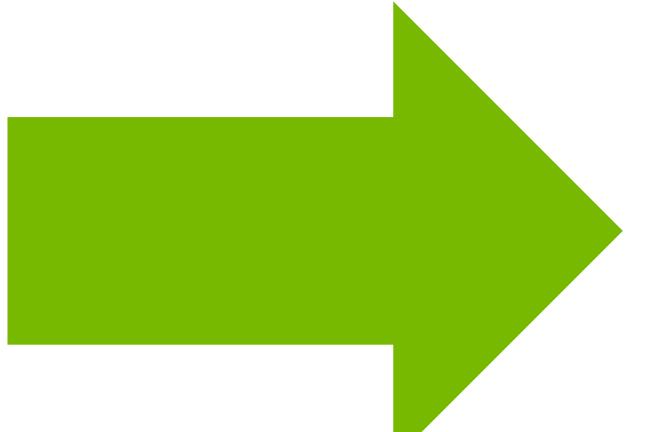
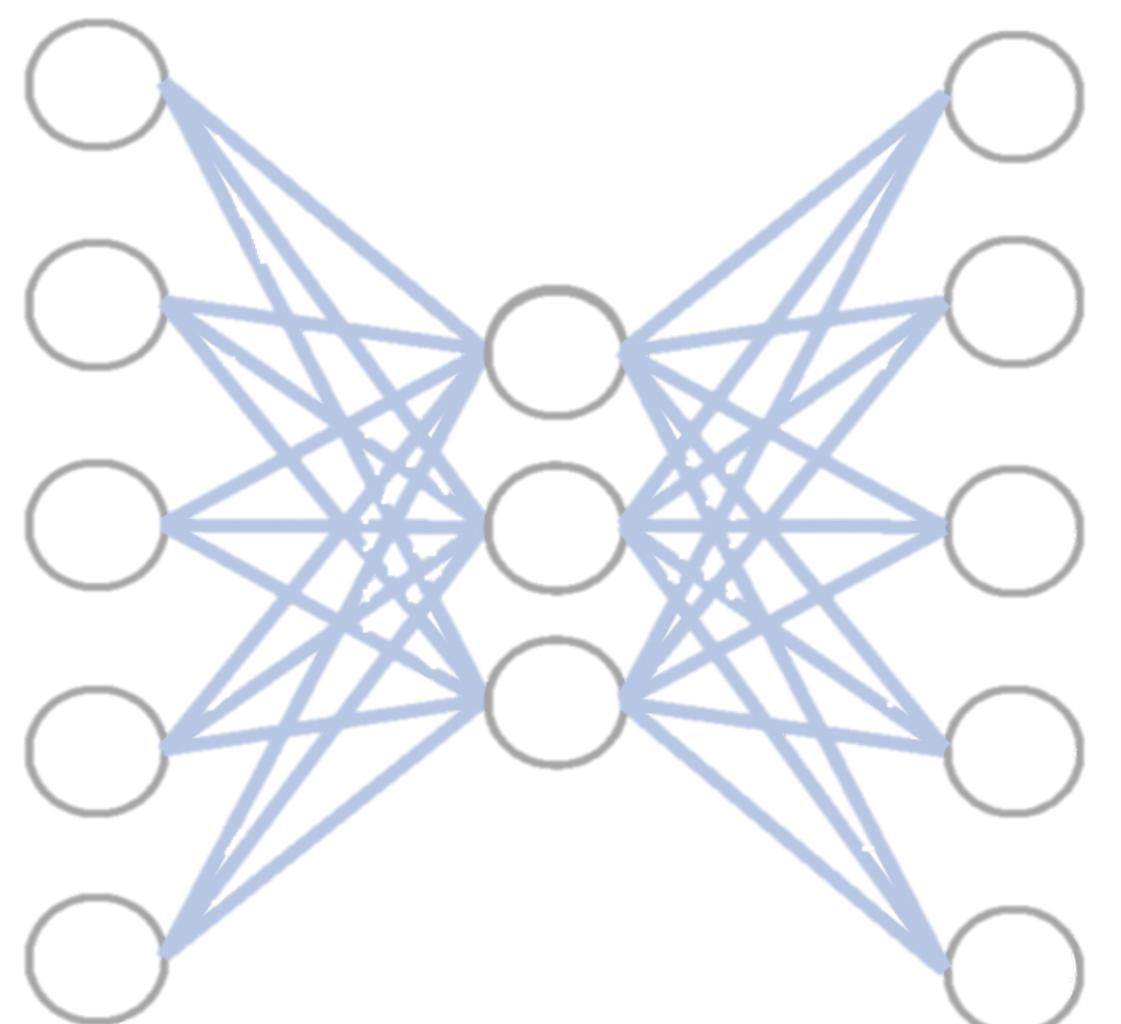
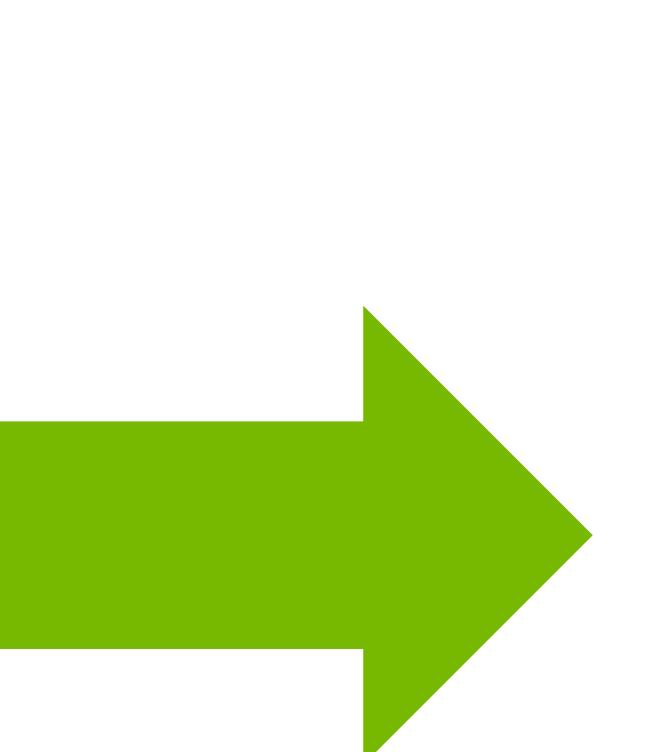
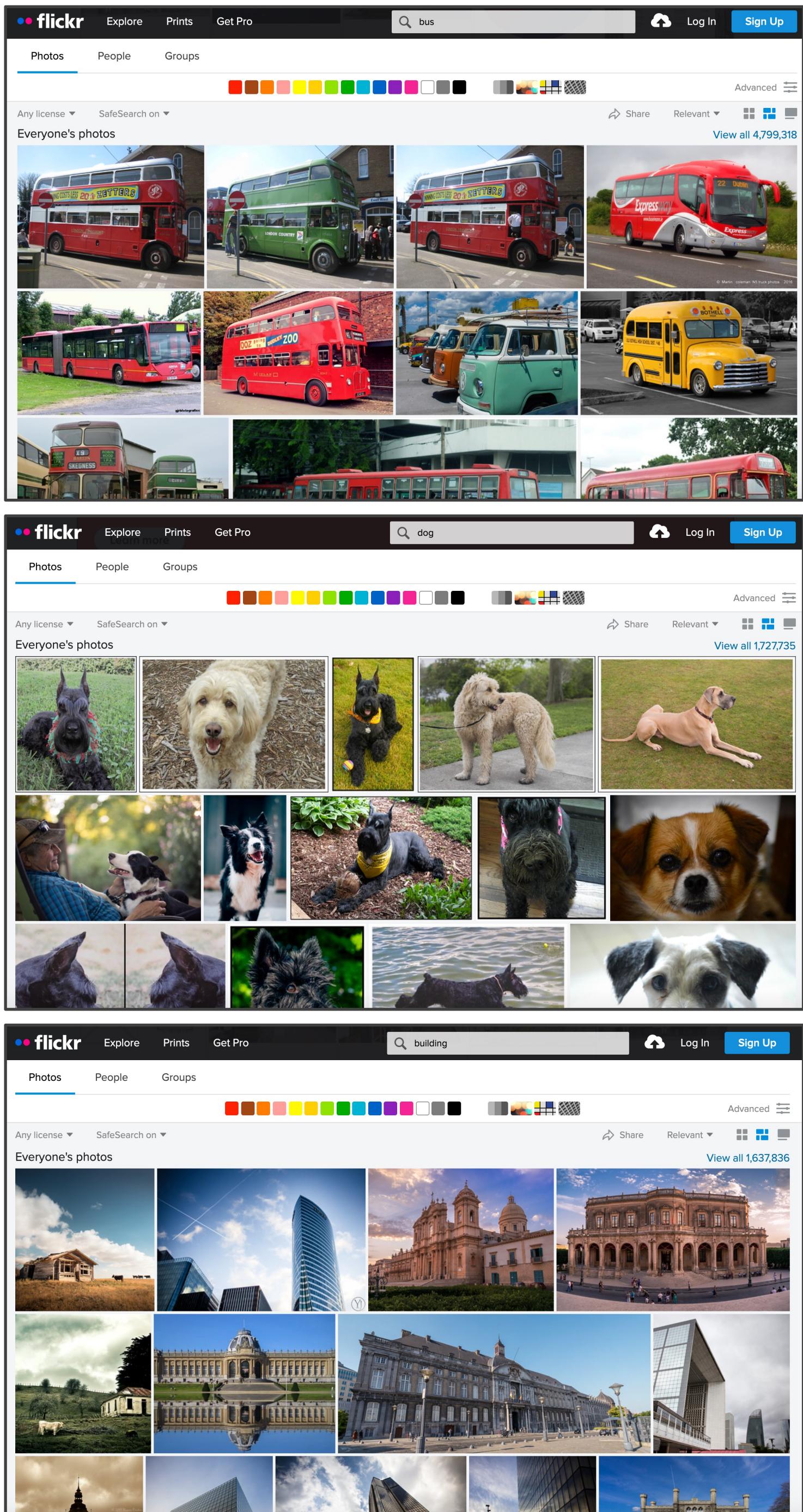
GANverse3D Omniverse Extension



Image to Mesh

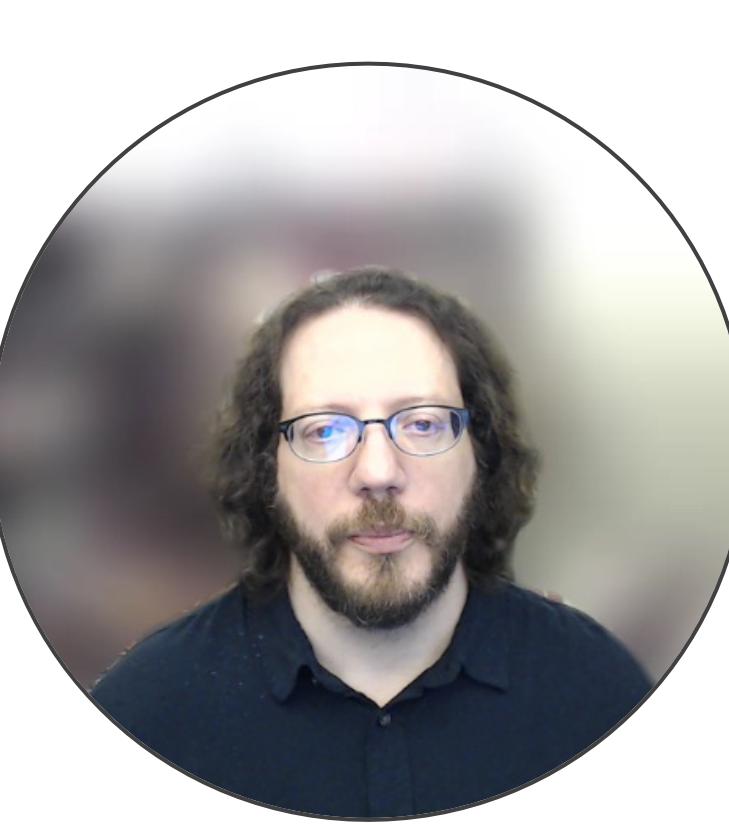


Digression: Latent Spaces / Embeddings



DeepSearch

AI Search of Nucleus Server



Omniverse Navigator

https://content.ov.nvidia.com/omni/web3/omniverse://content.ov.nvidia.com/?query=description:"car%20tire"%20ext:usd

content.ov.nvidia.com

SEARCH RESULTS (64)

Details Permissions

Date modified
Created by
Modified by

content.ov.nvidia.com

DOWNLOAD COPY URL NAVIGATE TO

Bookmarks

Omniverse

NVIDIA

NVIDIA-Staging

Library

Projects

Users

shot_06_tire_close.usd shot_06_tire_close.usd OV_Car_Config_Opene r_LookDev_03.usd wheel.usd

Michelin_Pilot_Sports_4S_RF.usd Michelin_Pilot_Sports_4S_RR.usd Michelin_Pilot_Sports_4S_LR.usd Wheel.usd

Wheel_40.usd Michelin_Pilot_Sports_4S_LF.usd wheel_1.usd shot_09a_bonnet.usd

33 NVIDIA

The screenshot displays the Omniverse Navigator interface, specifically the search results for "car tire" files in USD format. The search bar at the top shows the query: "description:'car tire' ext:usd". The results are listed in a grid format, with each item showing a thumbnail, the file name, and a small lock icon indicating it's locked or requires authentication. The interface also includes a sidebar for bookmarks and navigation, and a detailed view panel on the right showing file metadata like date modified, created by, and modified by.

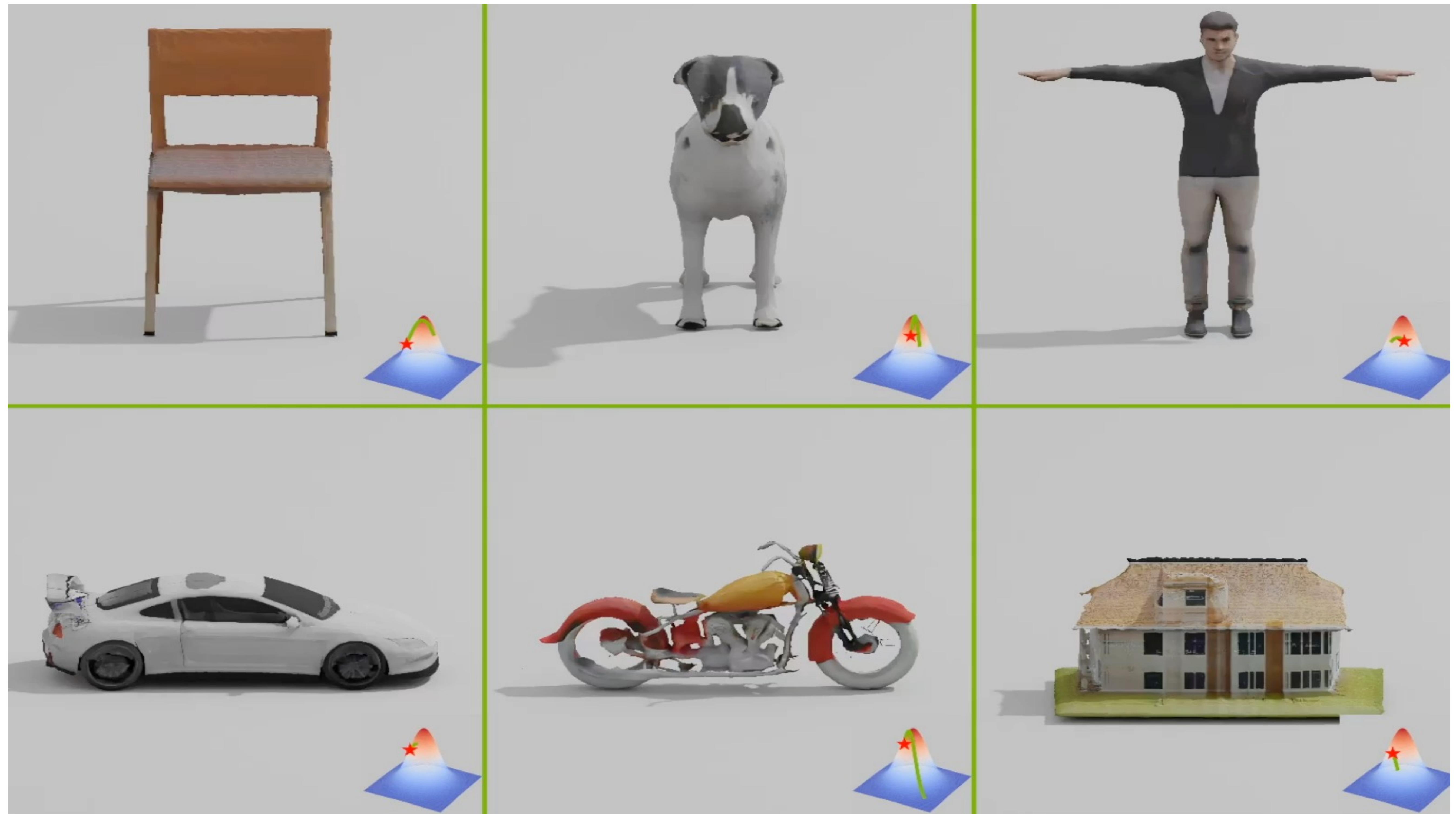
DeepSearch

Latent Space Visualization

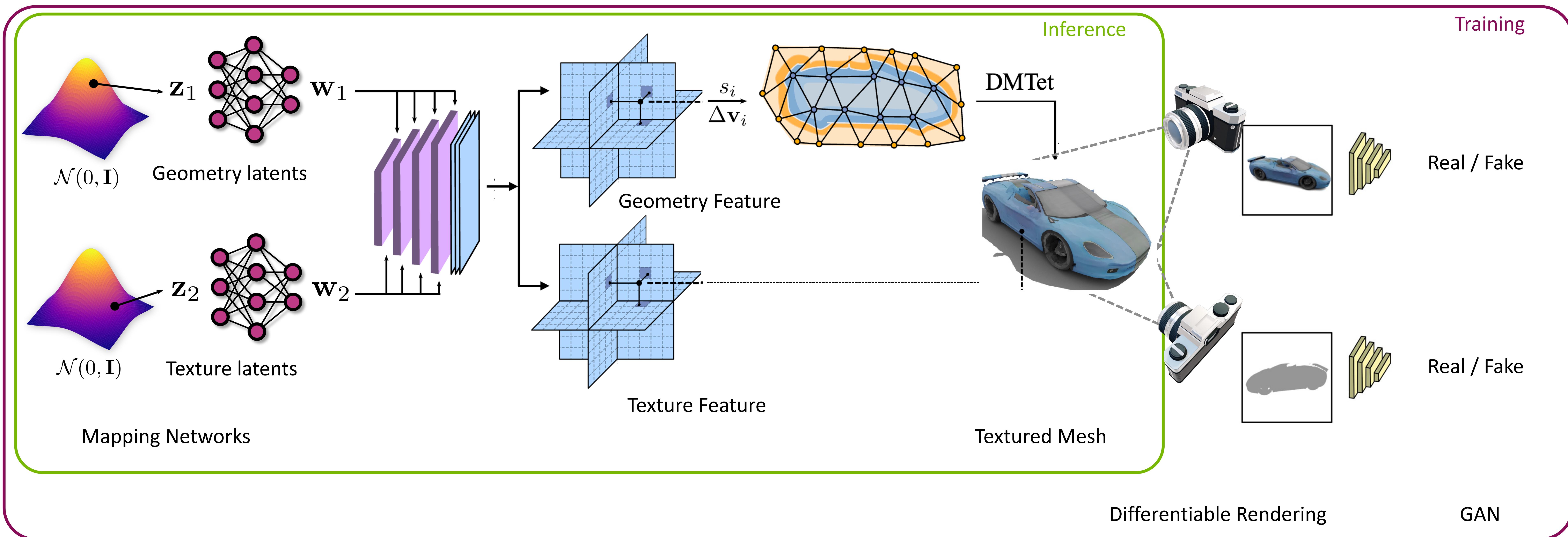


GET3D

3D Asset Creation



GET3D



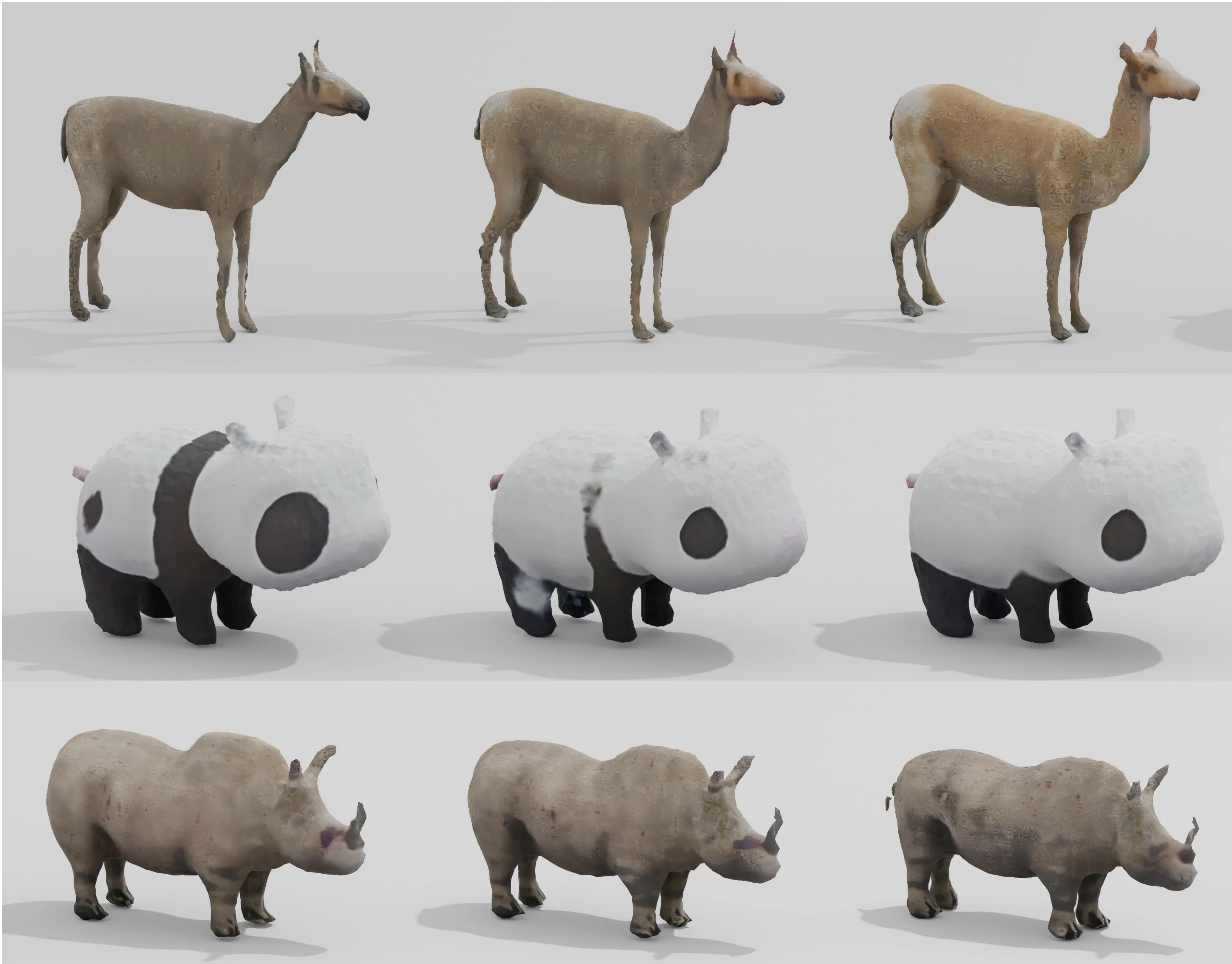
GET3D

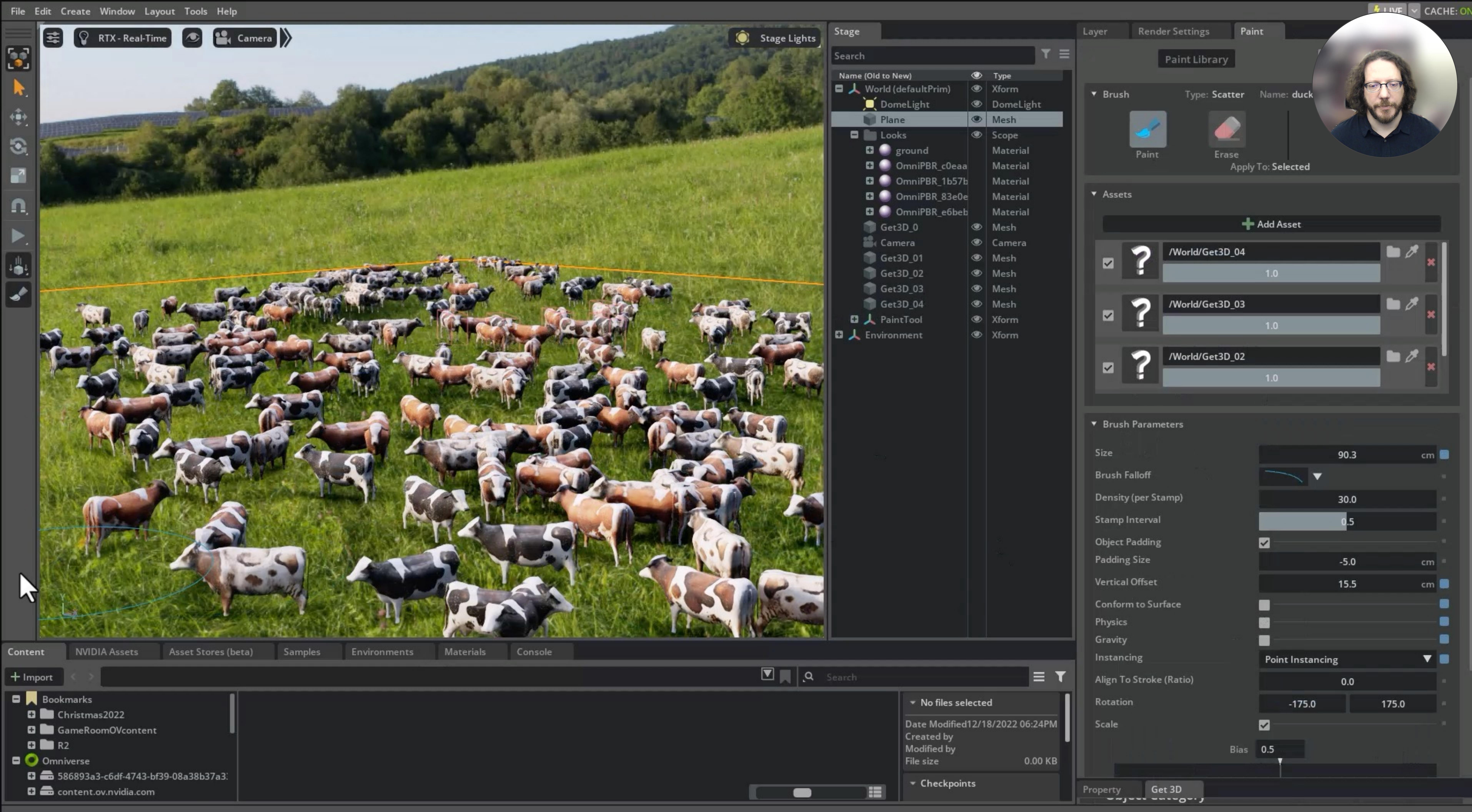
Fine grained interpolation



GET3D

Locally perturbing the latent codes: creating content variations







Back to 2D: Diffusion Models

Picasso / Edify-Image



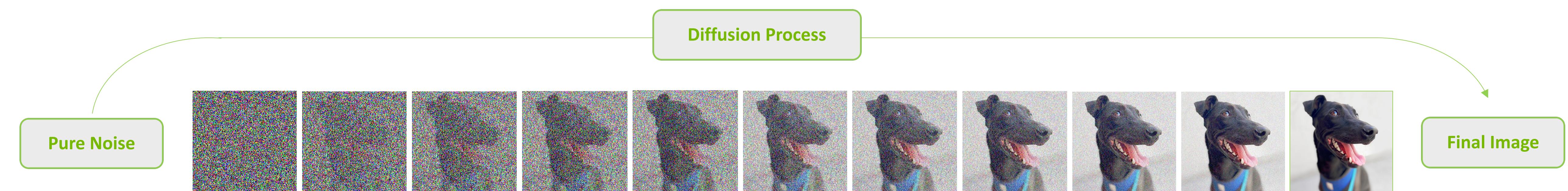
A highly detailed digital painting of a portal in a mystic forest with many beautiful trees. A person is standing in front of the portal.



A highly detailed zoomed-in digital painting of a cat dressed as a witch wearing a wizard hat in a haunted house, artstation.



An image of a beautiful landscape of an ocean. There is a huge rock in the middle of the ocean. There is a mountain in the background. Sun is setting.





Diffusion Models

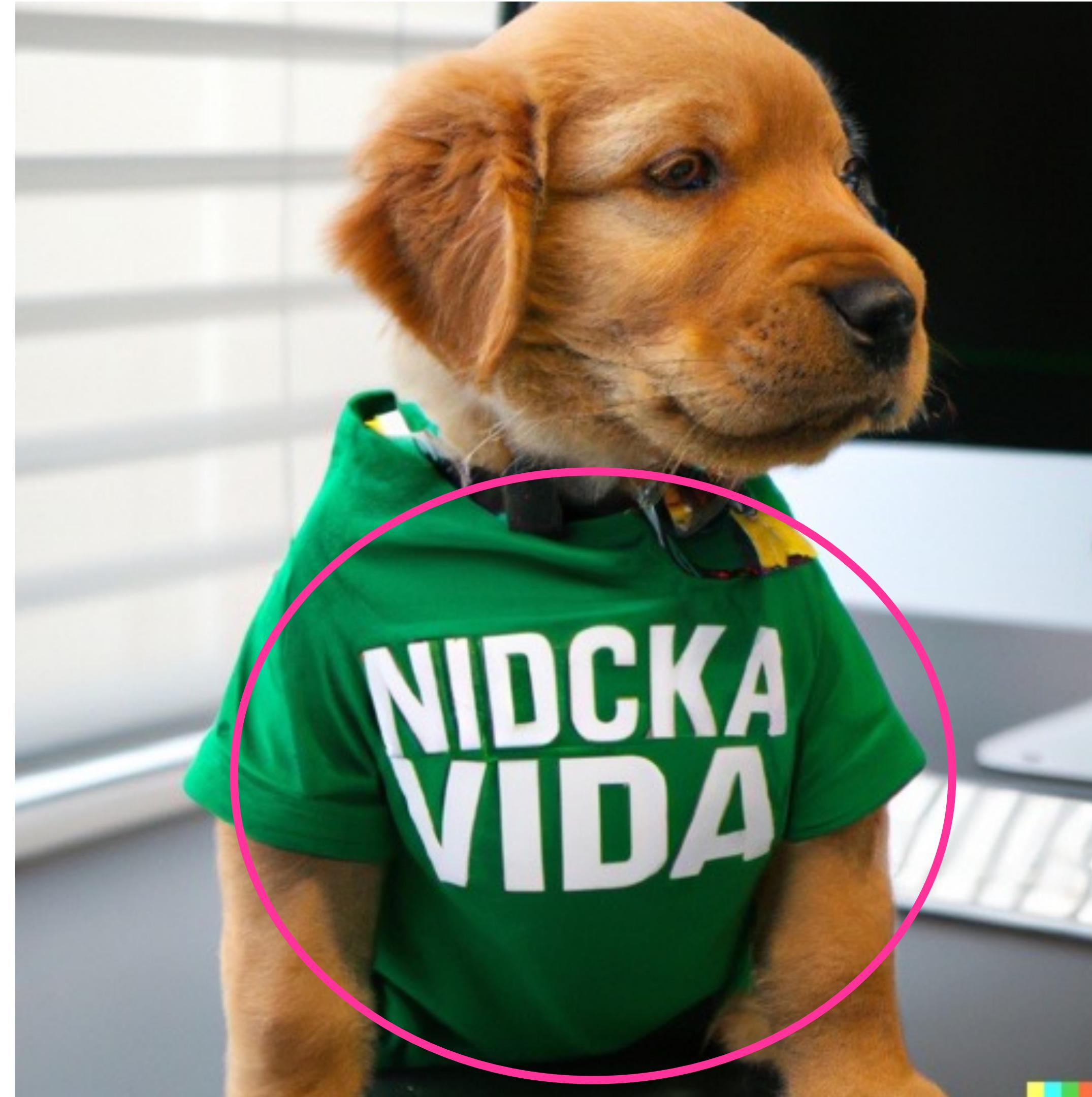
Text Generation with Edify-Image

A photo of a golden retriever puppy wearing a green shirt. The shirt has text that says “**NVIDIA rocks**”.

Background office. 4k dslr.



Stable Diffusion



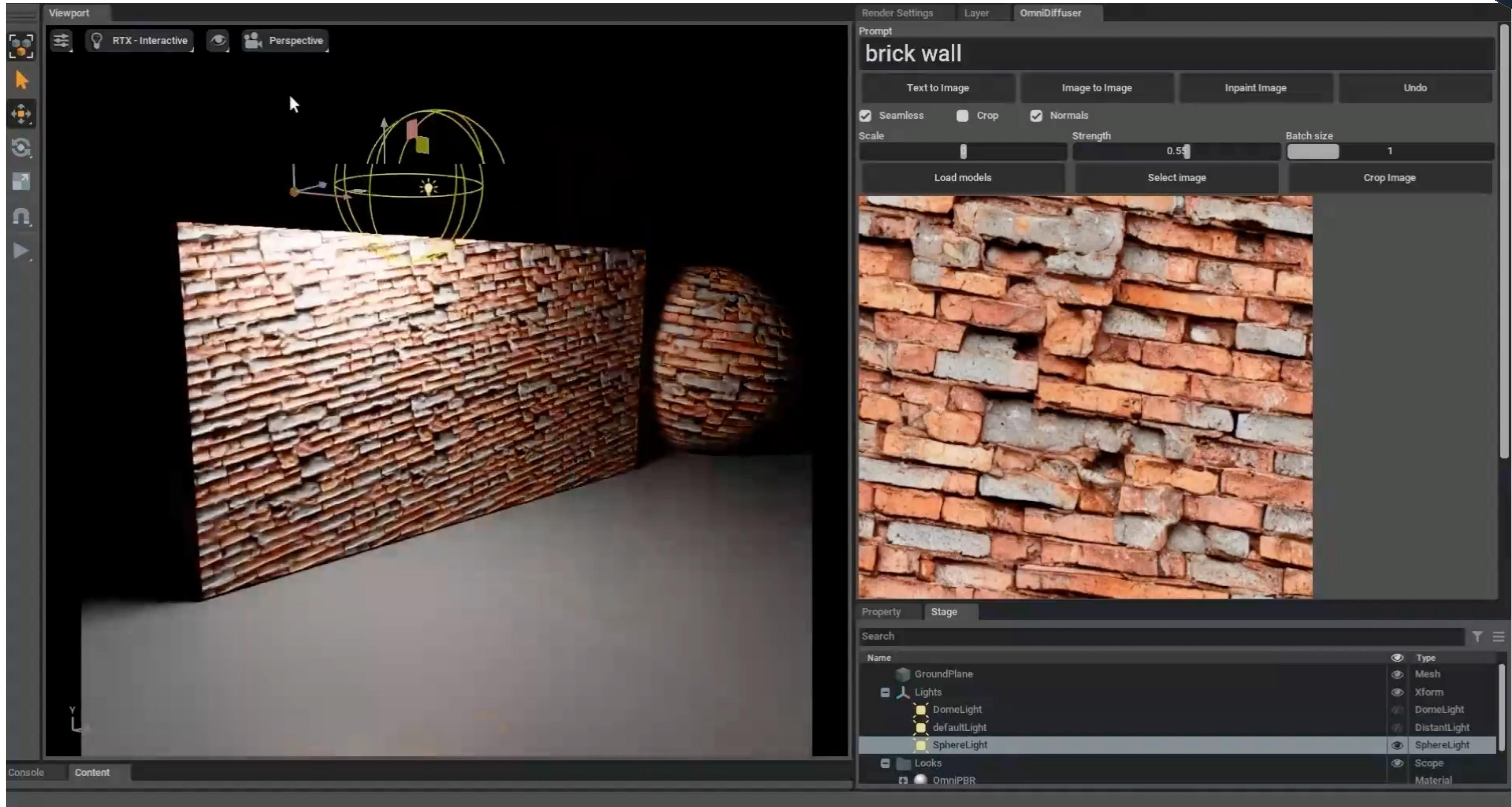
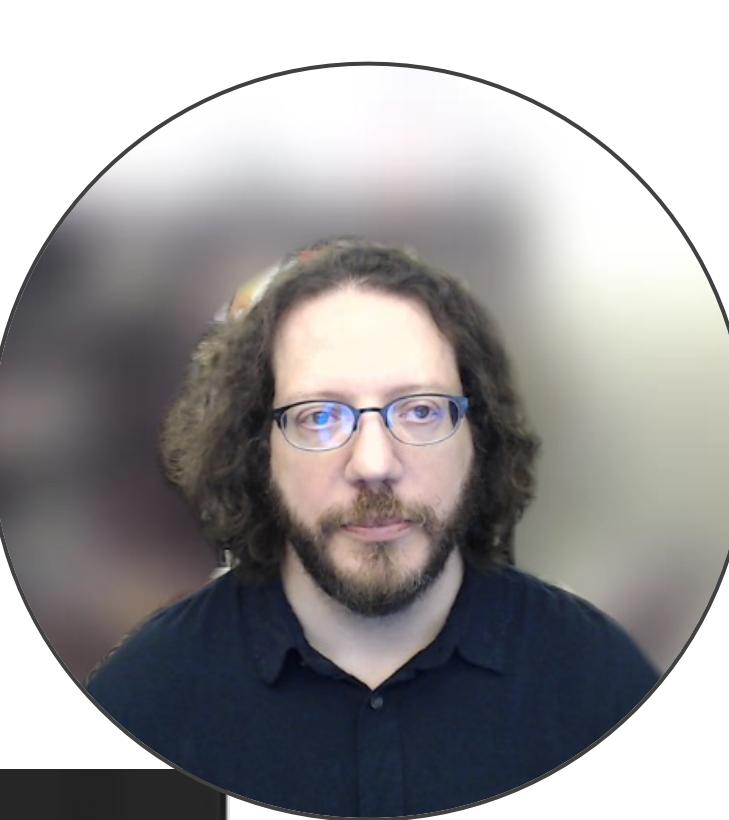
DALL·E 2



Edify-Image

AI Based Texture + Material Generation

Using 2D Generative Models for 3D Applications

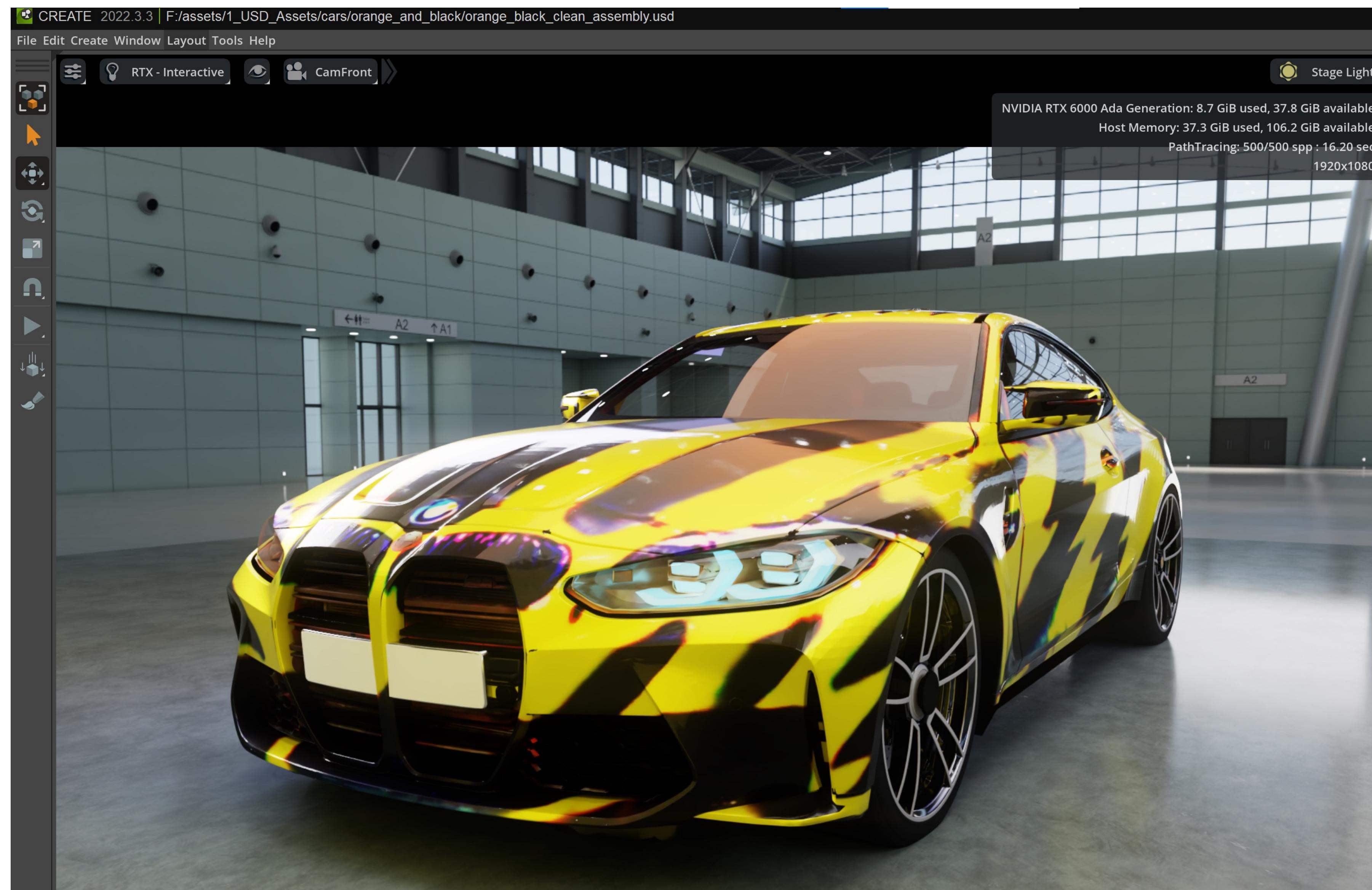


AI Based Texture + Material Generation

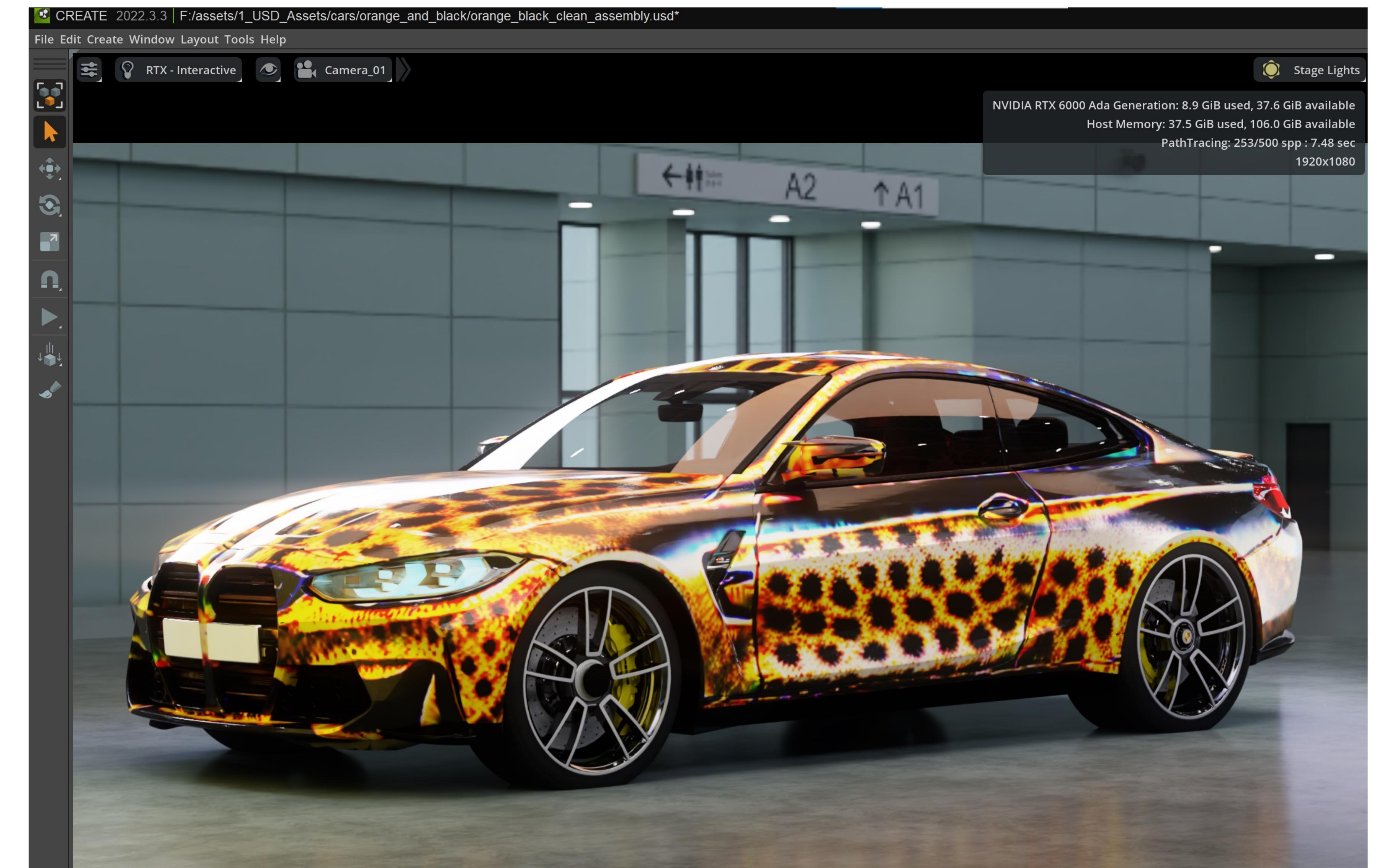
Wrapping Generated Materials around a Model



“Black and yellow wrap”



“Cheetah-inspired style”



AI Based Texture + Material Generation

Wrapping Generated Materials around a Model

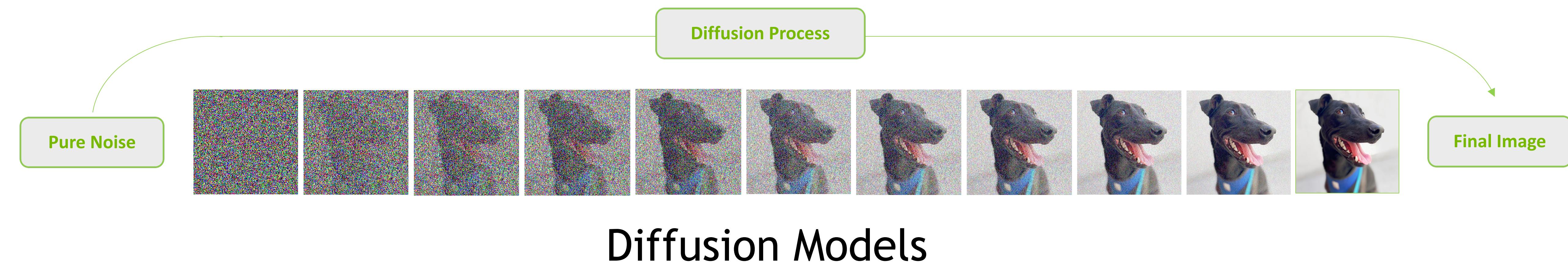




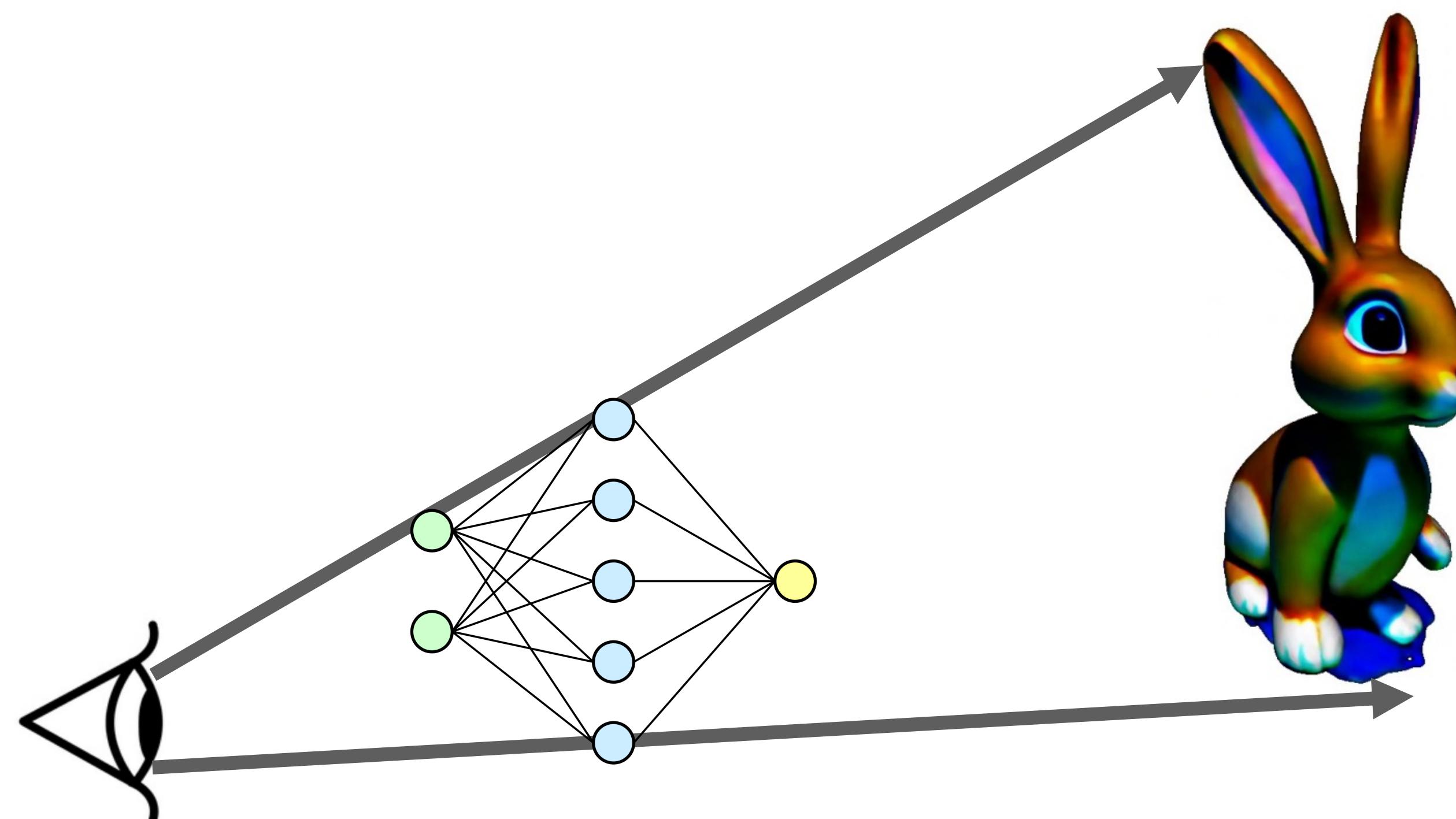
Bringing it all Together

Time for some real magic

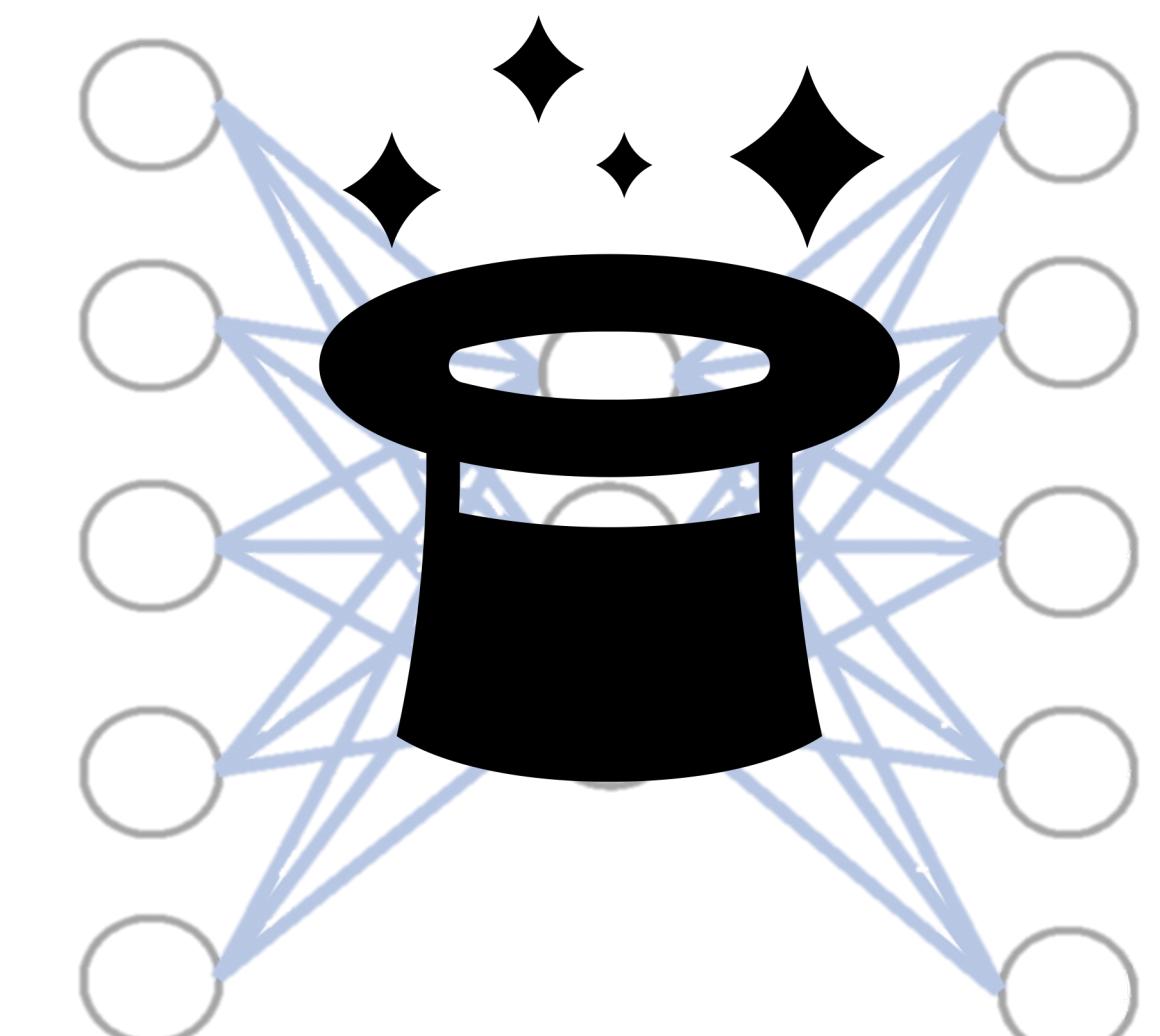
Text Input for Latent Space Embedding



Diffusion Models



Instant-NGP NeRFs + Hashgrid



Differentiable Rendering

Text to 3D Asset Generation

Magic 3D



Magic3D: High-Resolution Text-to-3D Content Creation

Chen-Hsuan Lin* Jun Gao* Luming Tang* Towaki Takikawa* Xiaohui Zeng*
Xun Huang Karsten Kreis Sanja Fidler# Ming-Yu Liu# Tsung-Yi Lin

*# : equal contributions

NVIDIA Corporation

Text to 3D Asset Generation

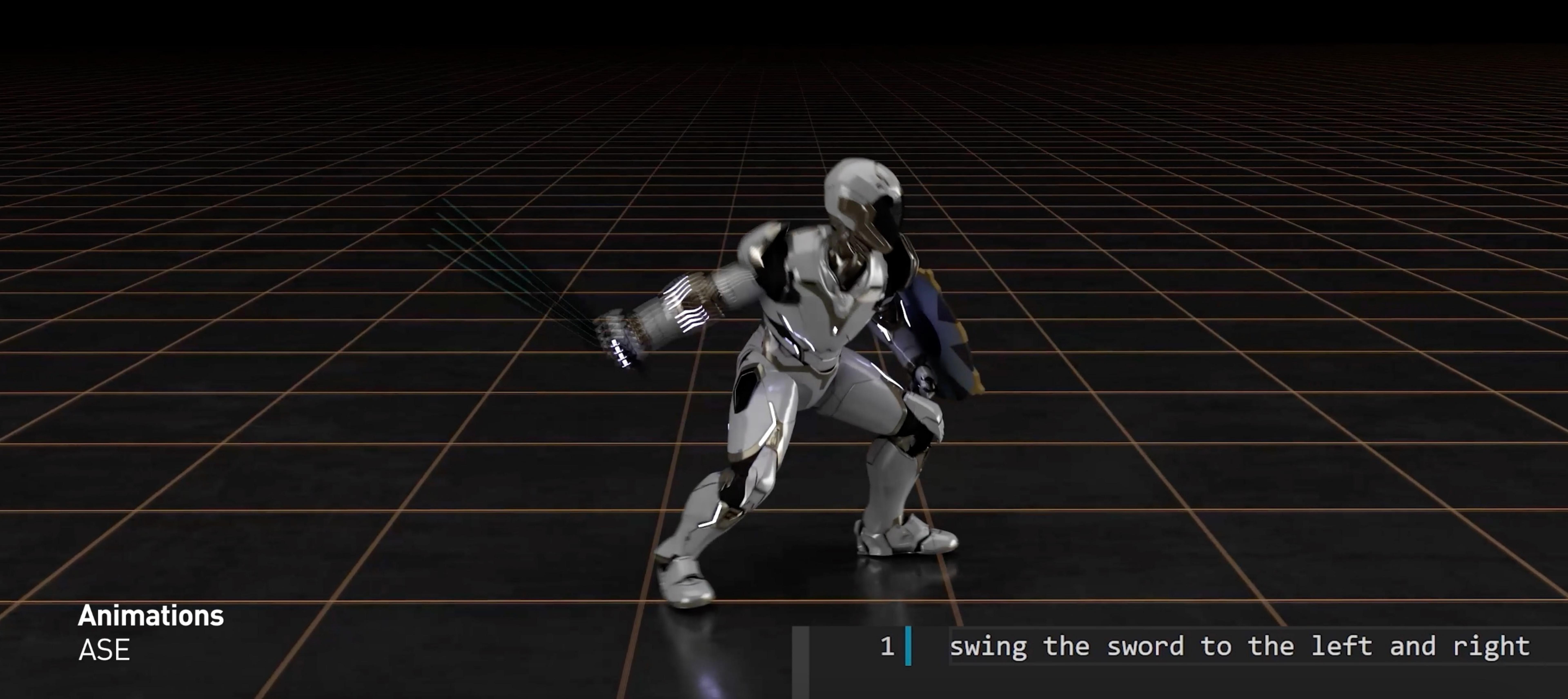
Picasso / Edify-3D





AI for Behaviour

ASE: Adversarial Skill Embedding

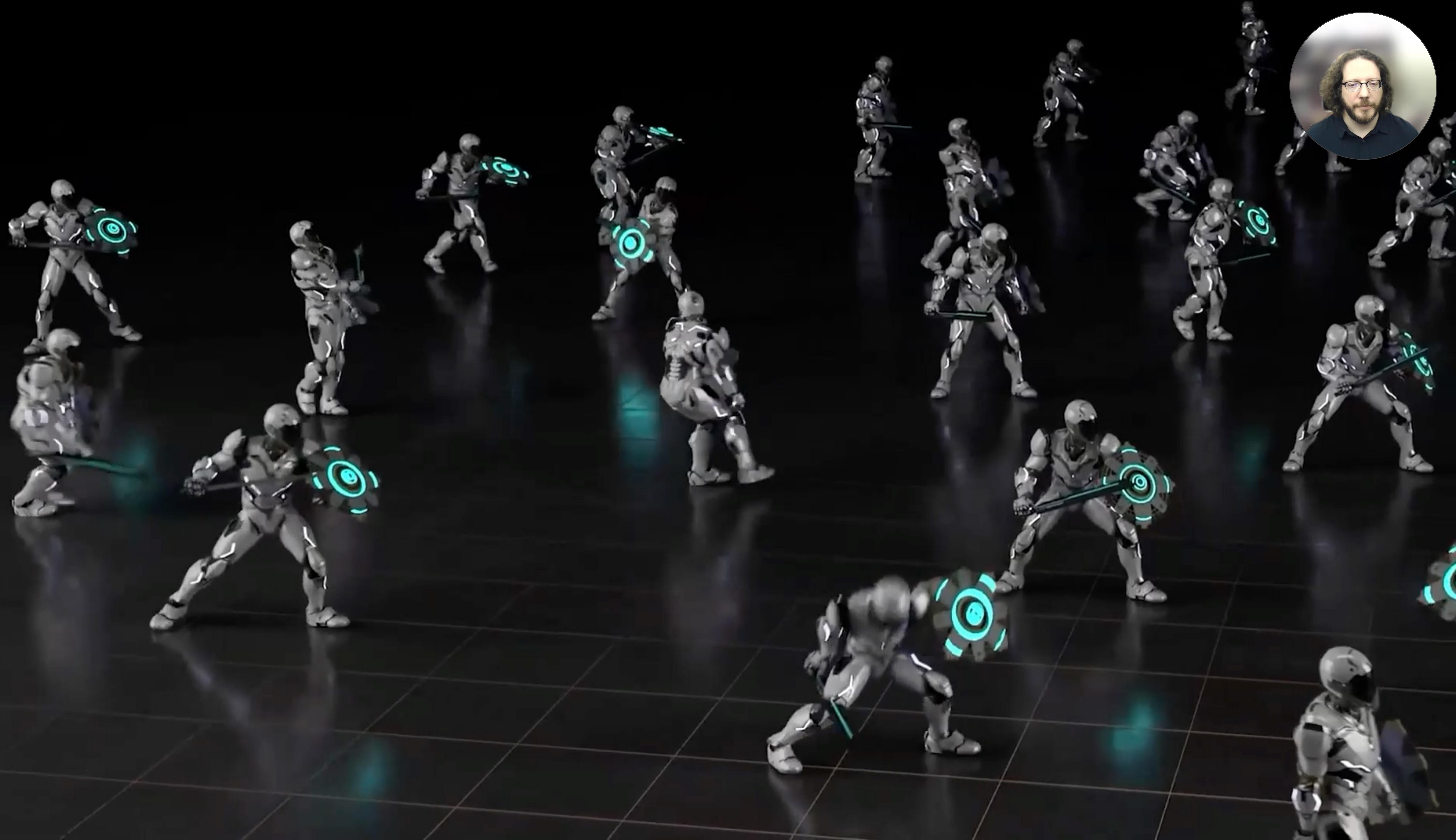


Animations

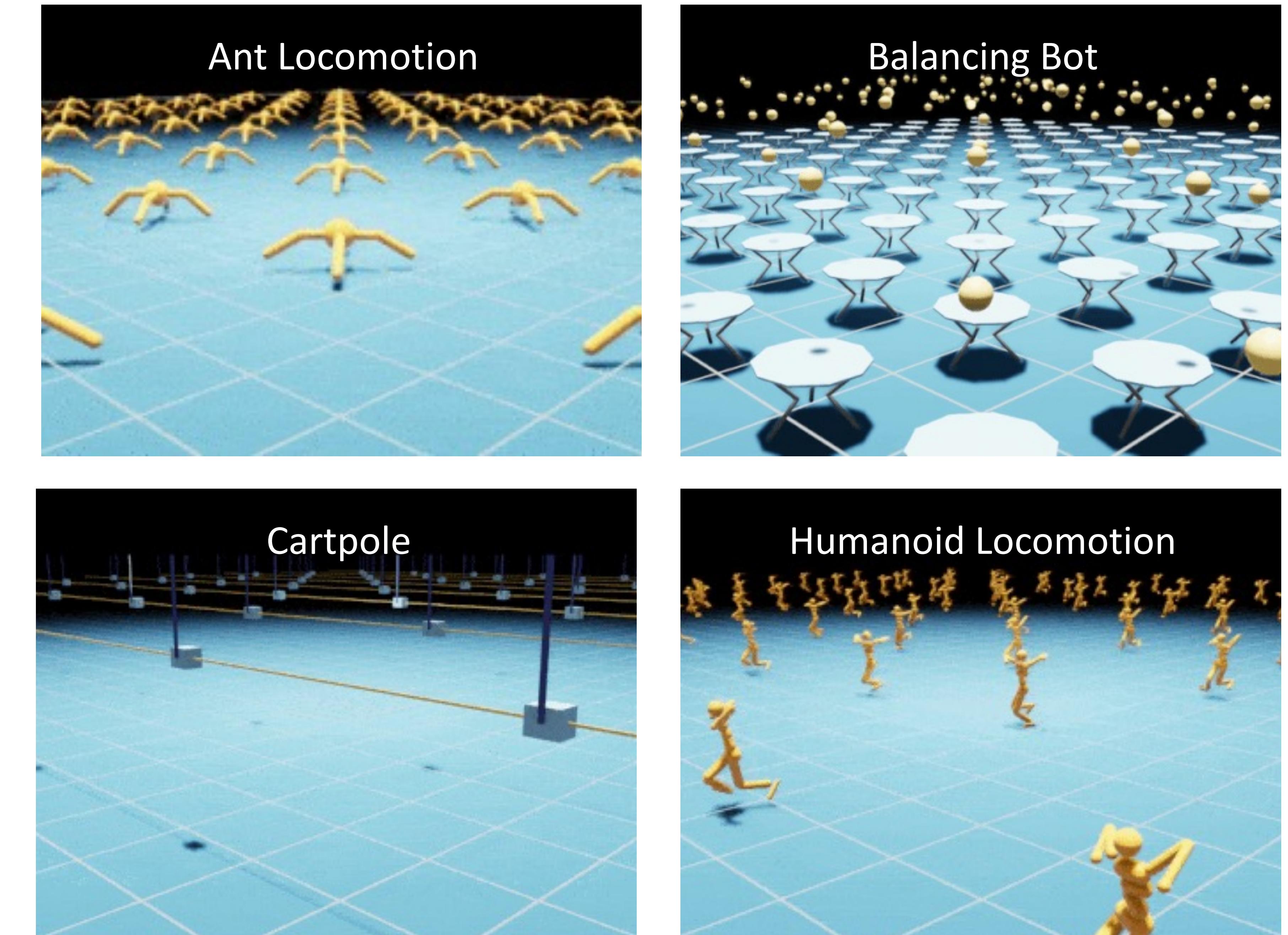
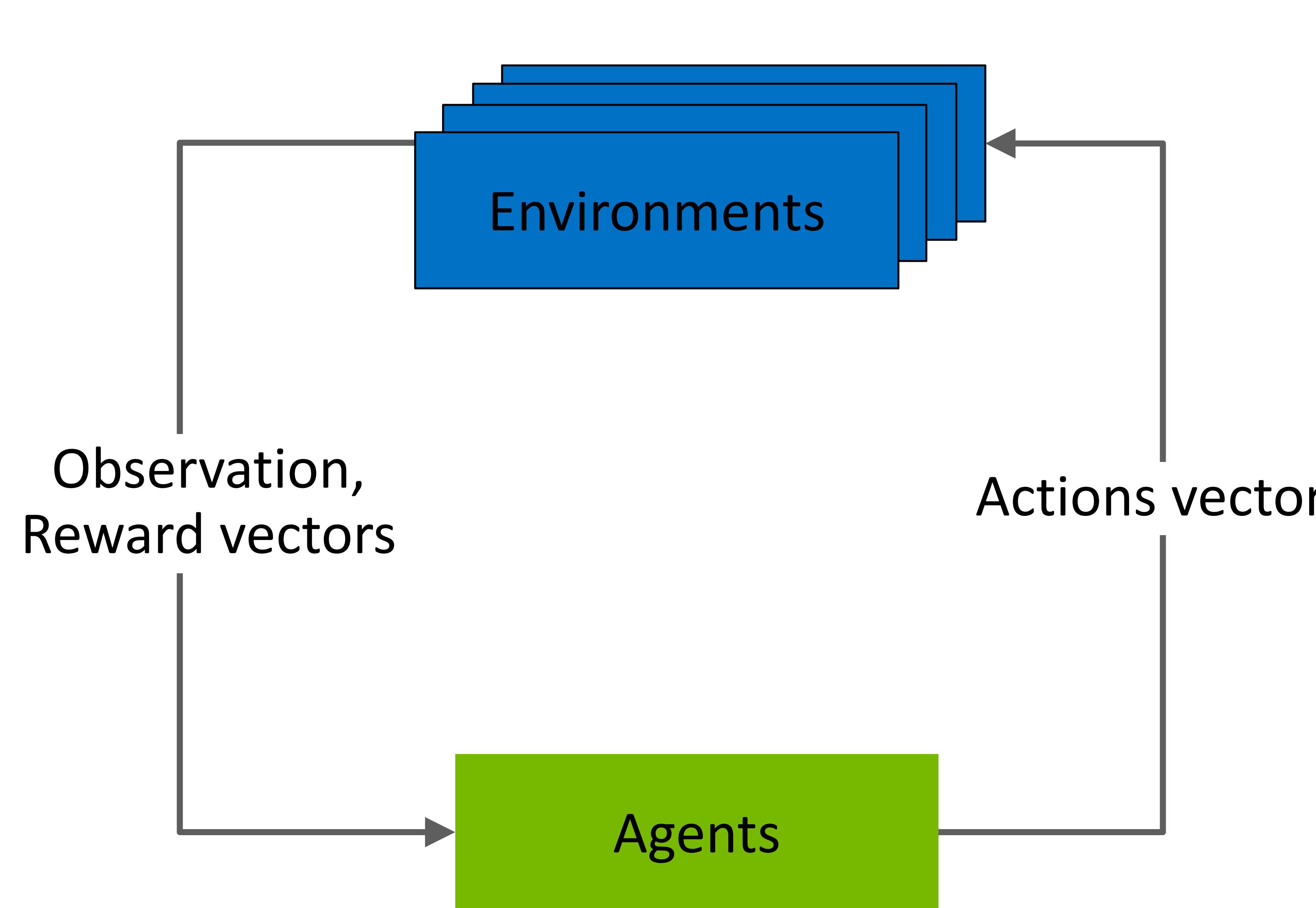
ASE

1

swing the sword to the left and right



Isaac Gym: Reinforcement Learning with GPU Parallel simulation



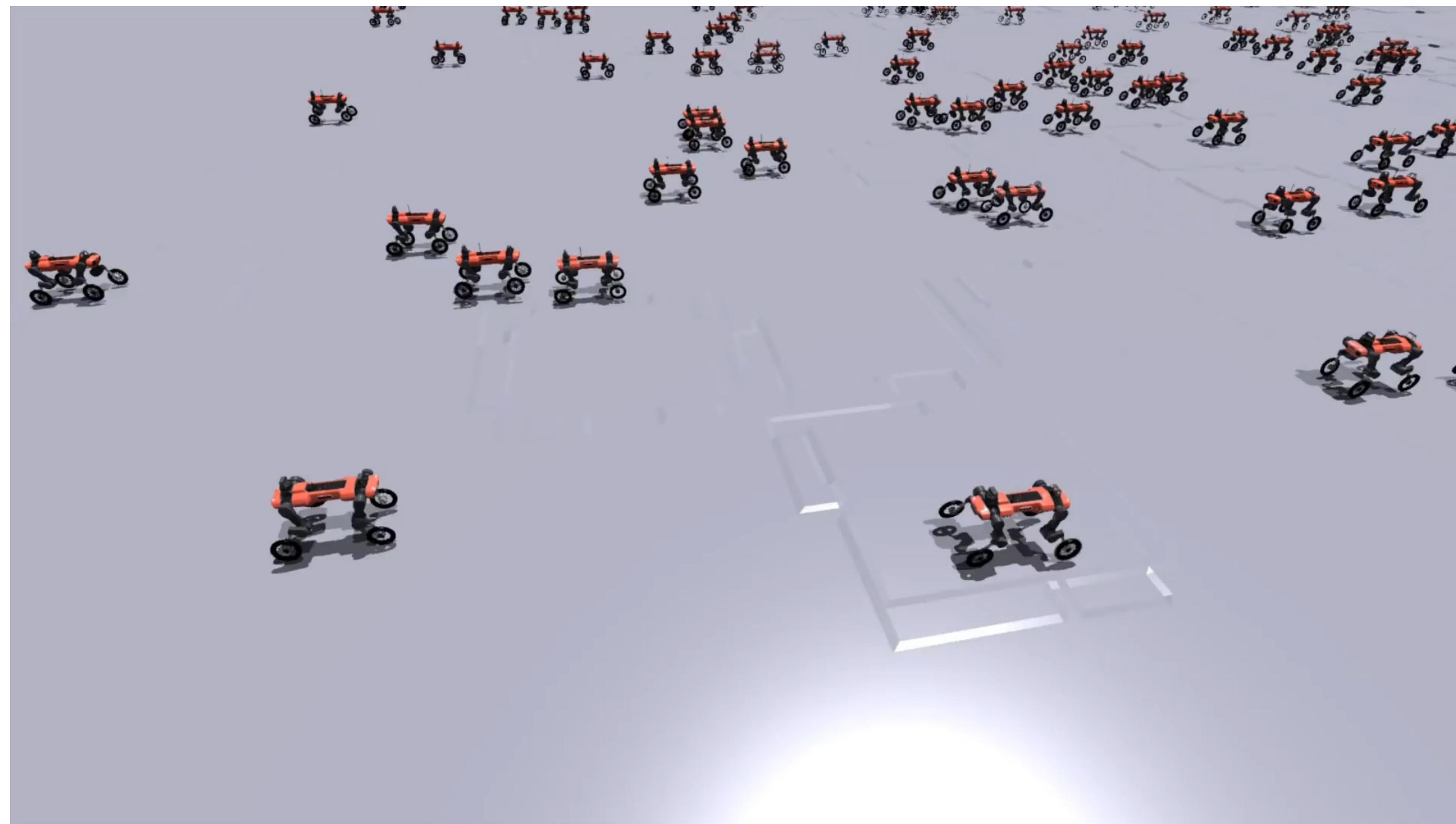
PhysX 5

Open Source USD Physics Implementation

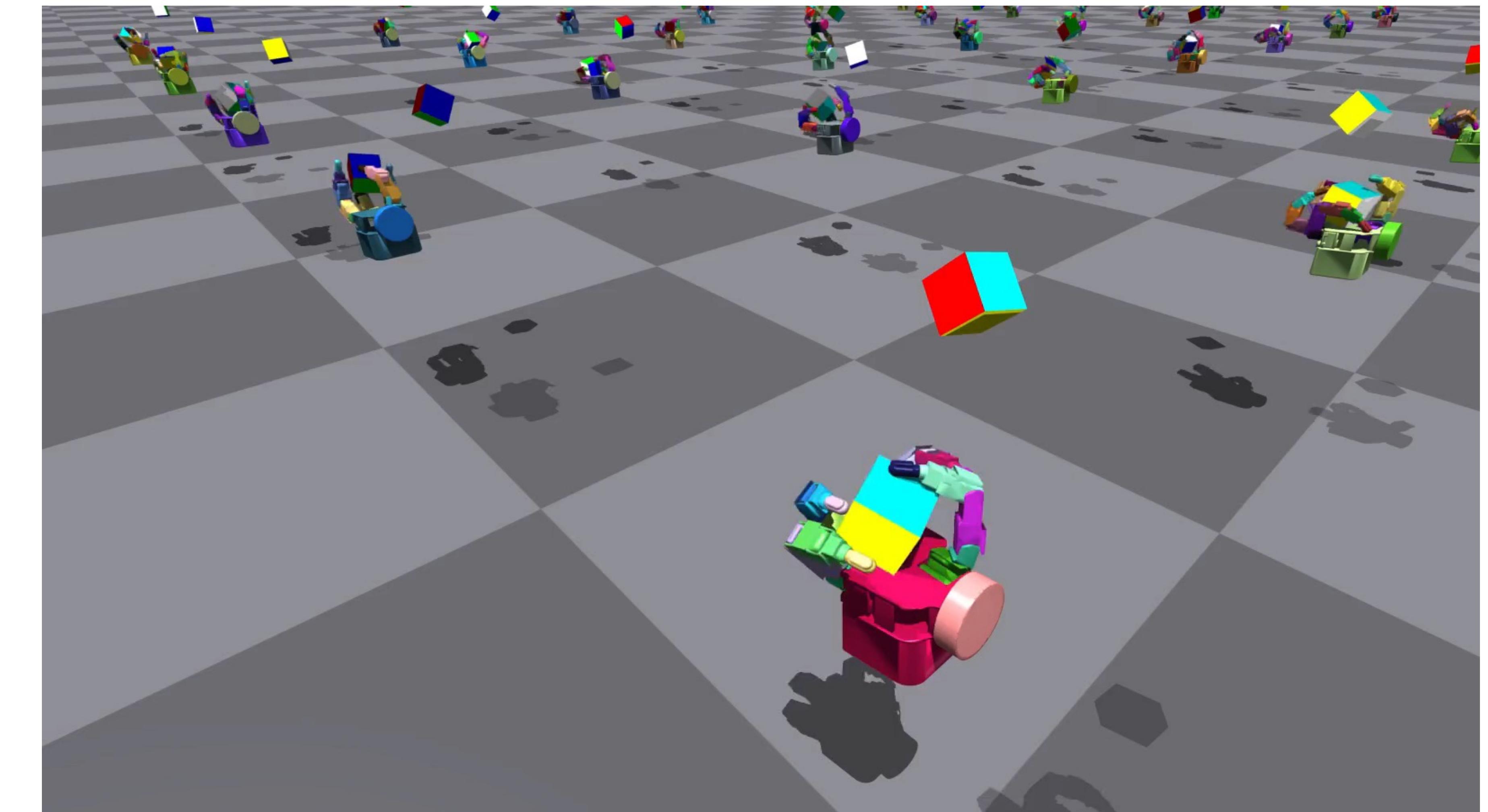


Isaac Gym

Training in Simulation



Wheeled Quadruped Locomotion



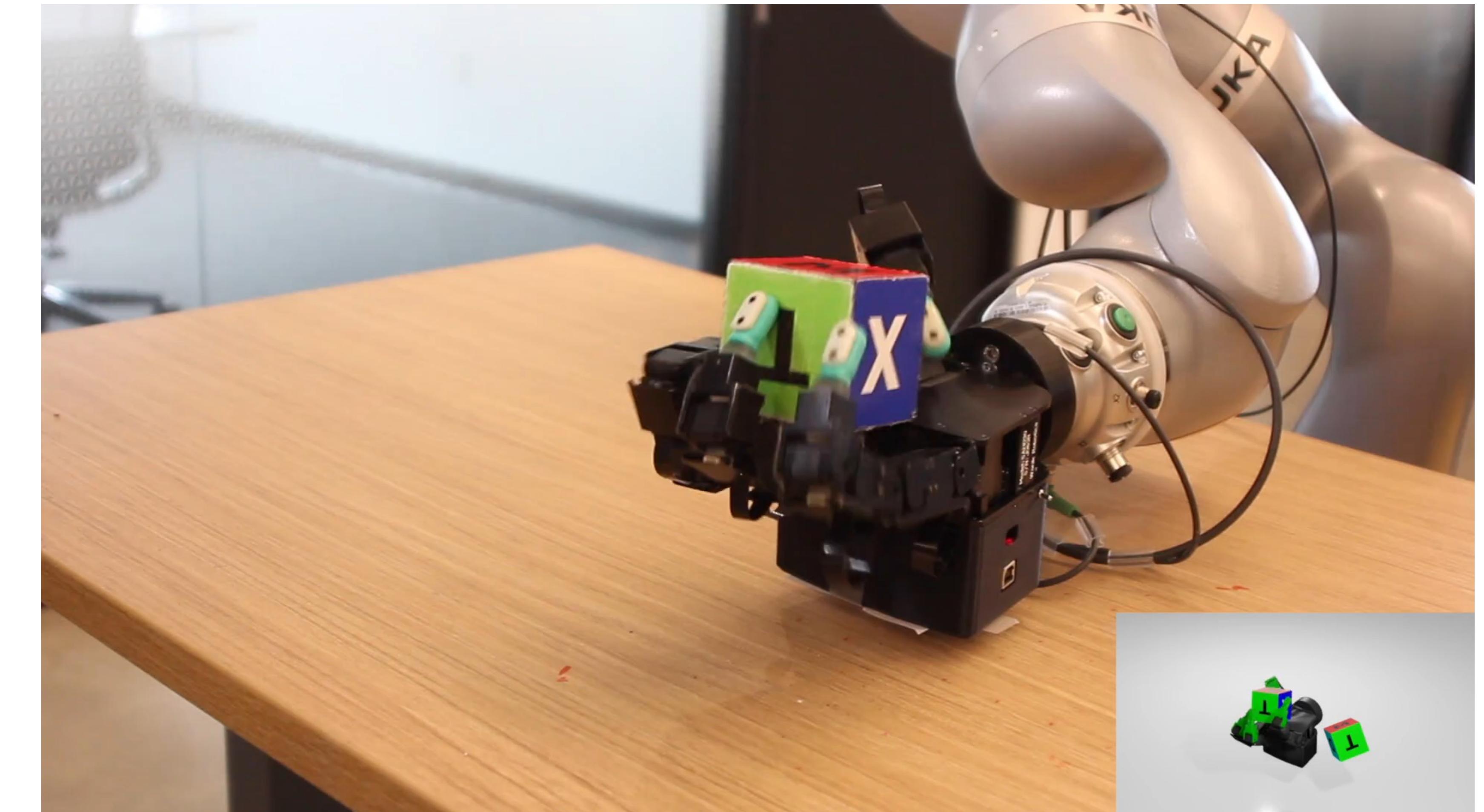
Dexterous Object Manipulation

Isaac Gym

Trained policy transfer to real robots



Wheeled Quadruped Locomotion

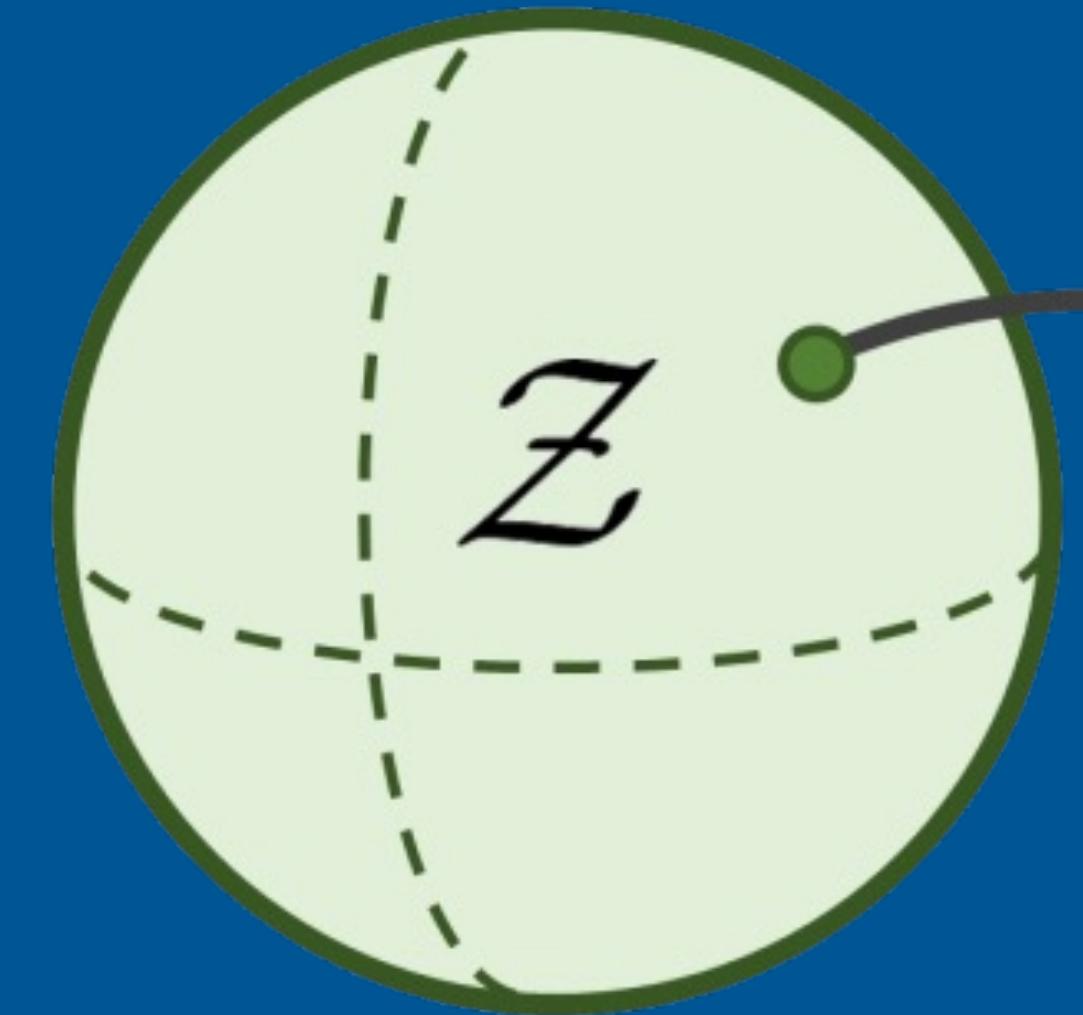


Dexterous Object Manipulation

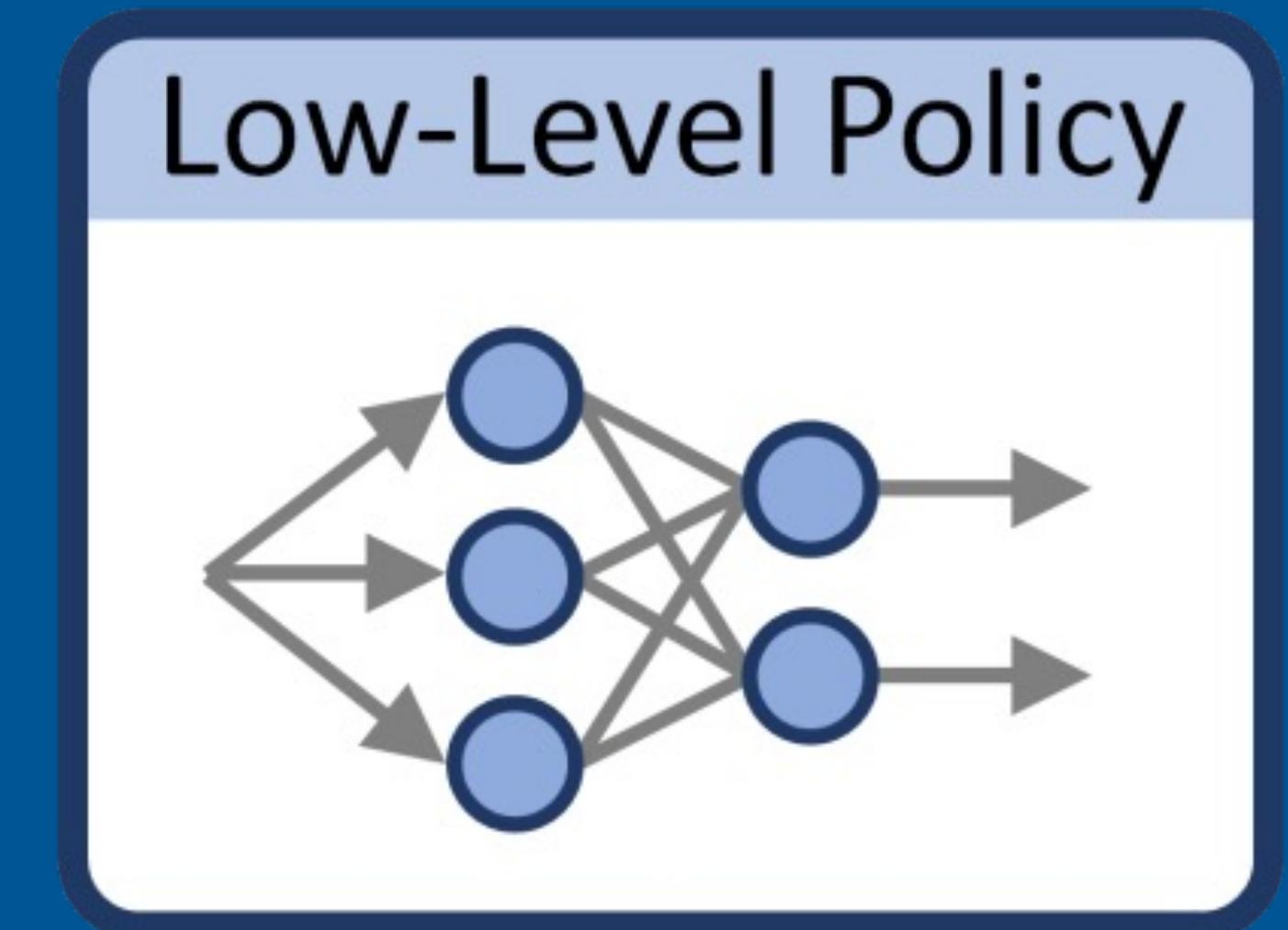
Character Animation AI: ASE



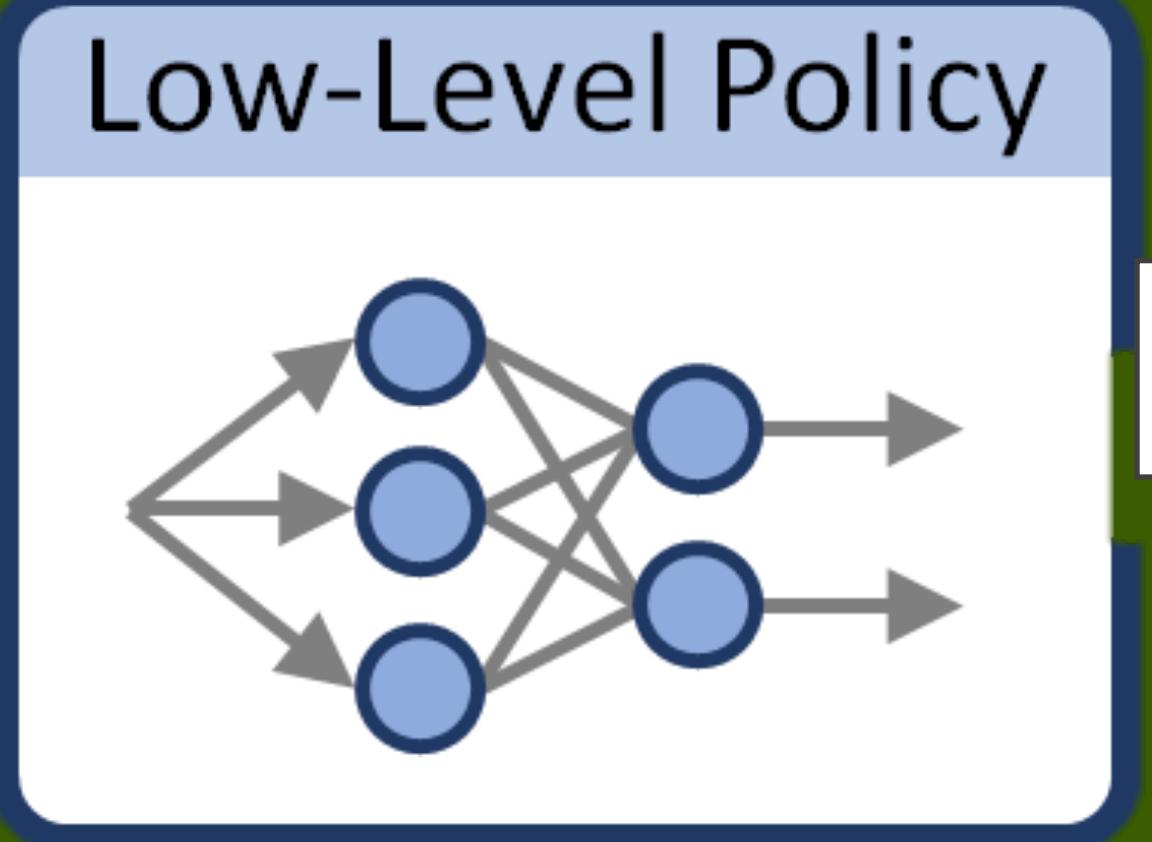
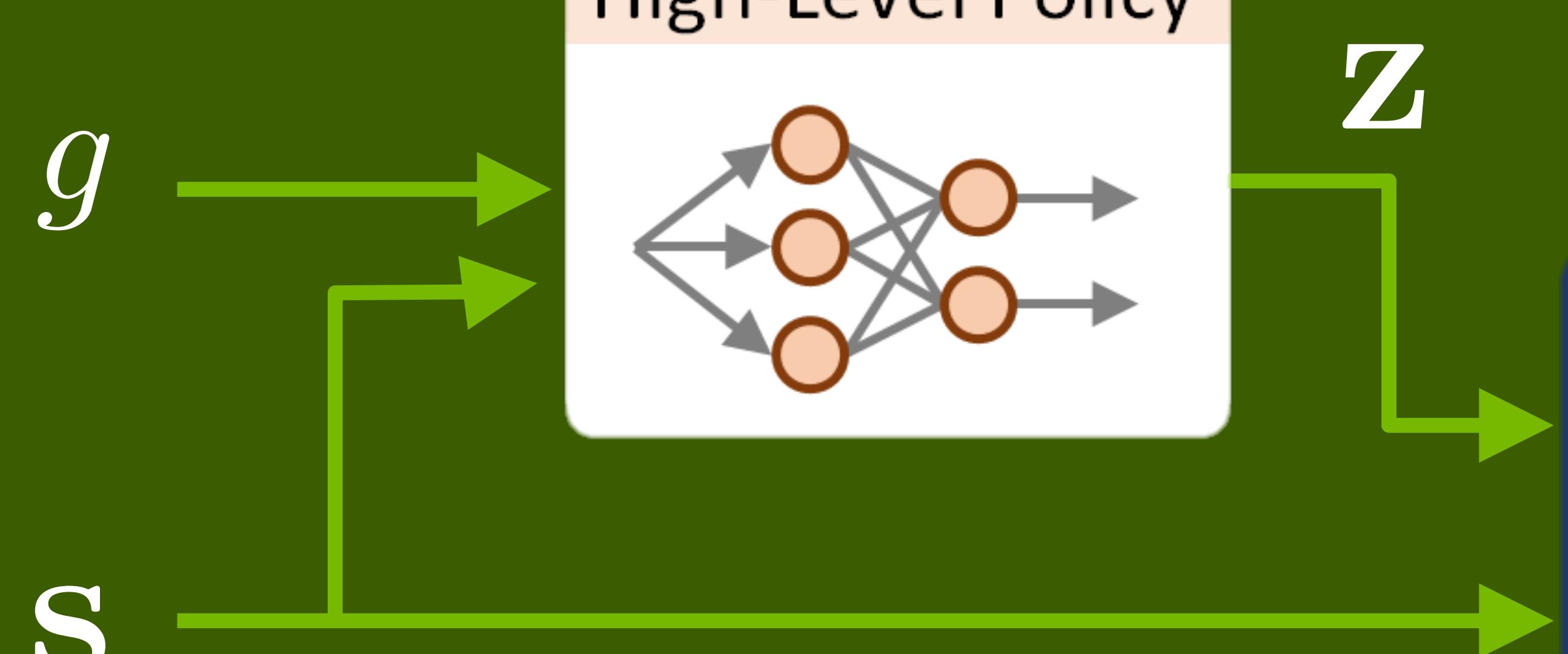
Latent space that represents reusable motion “skills”



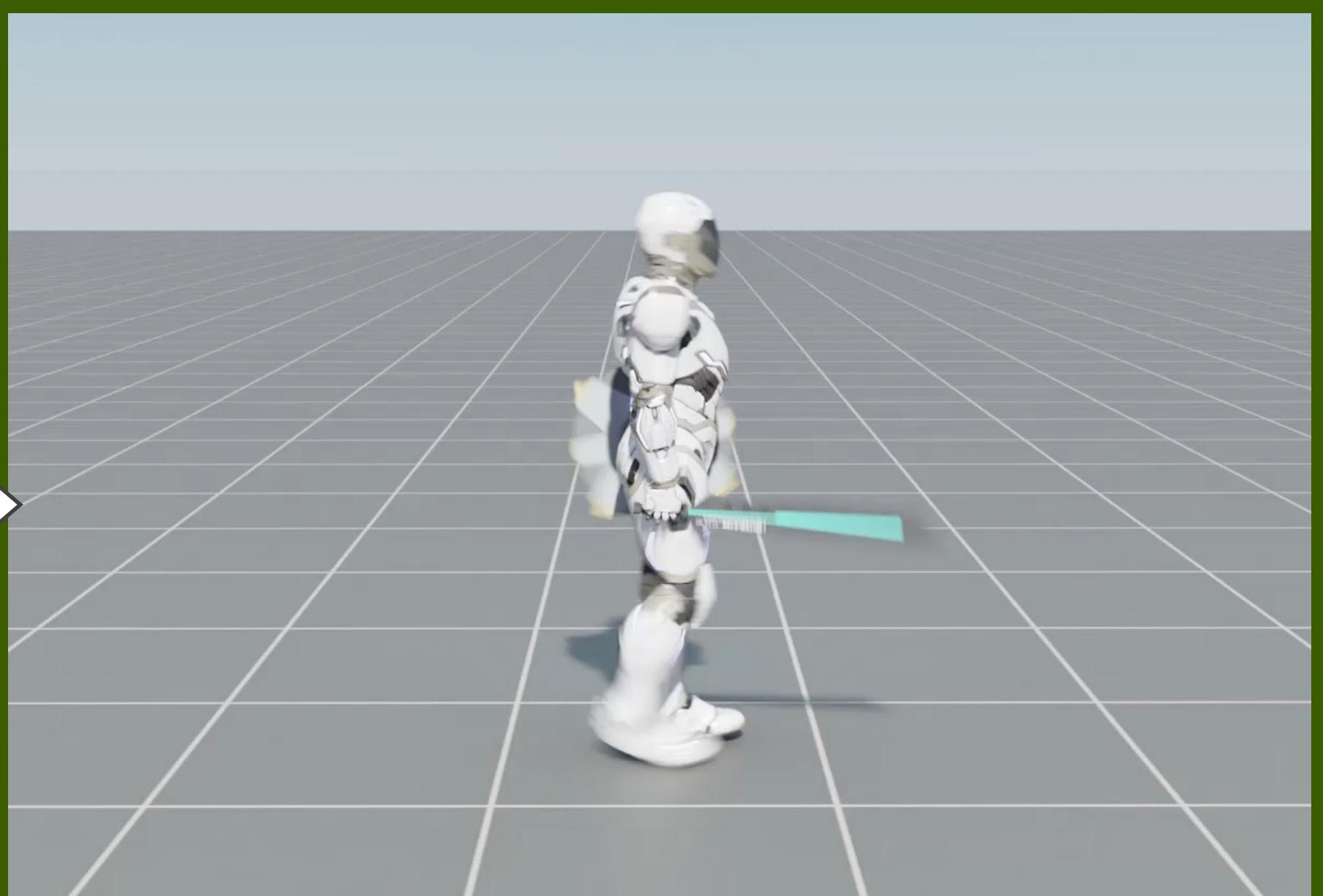
z



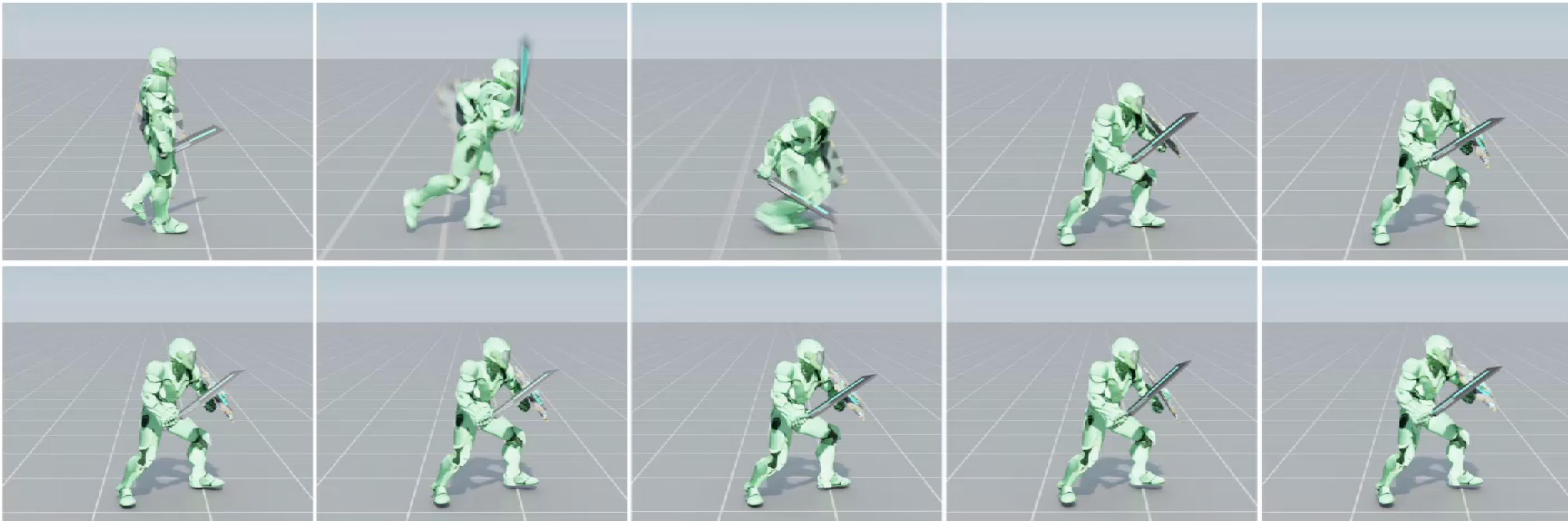
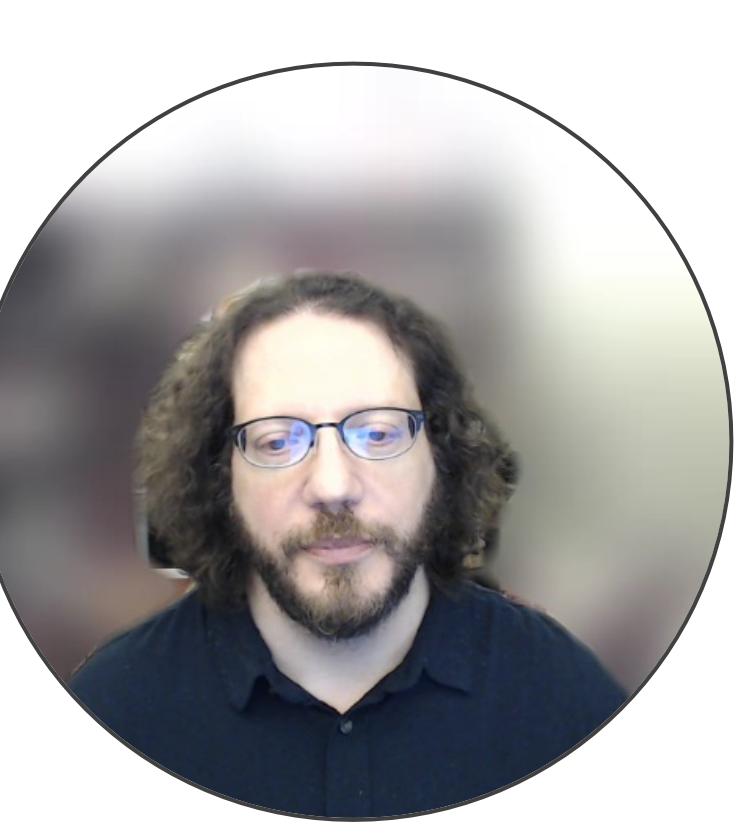
Task policy



“Strike a target”



Character Animation AI: ASE



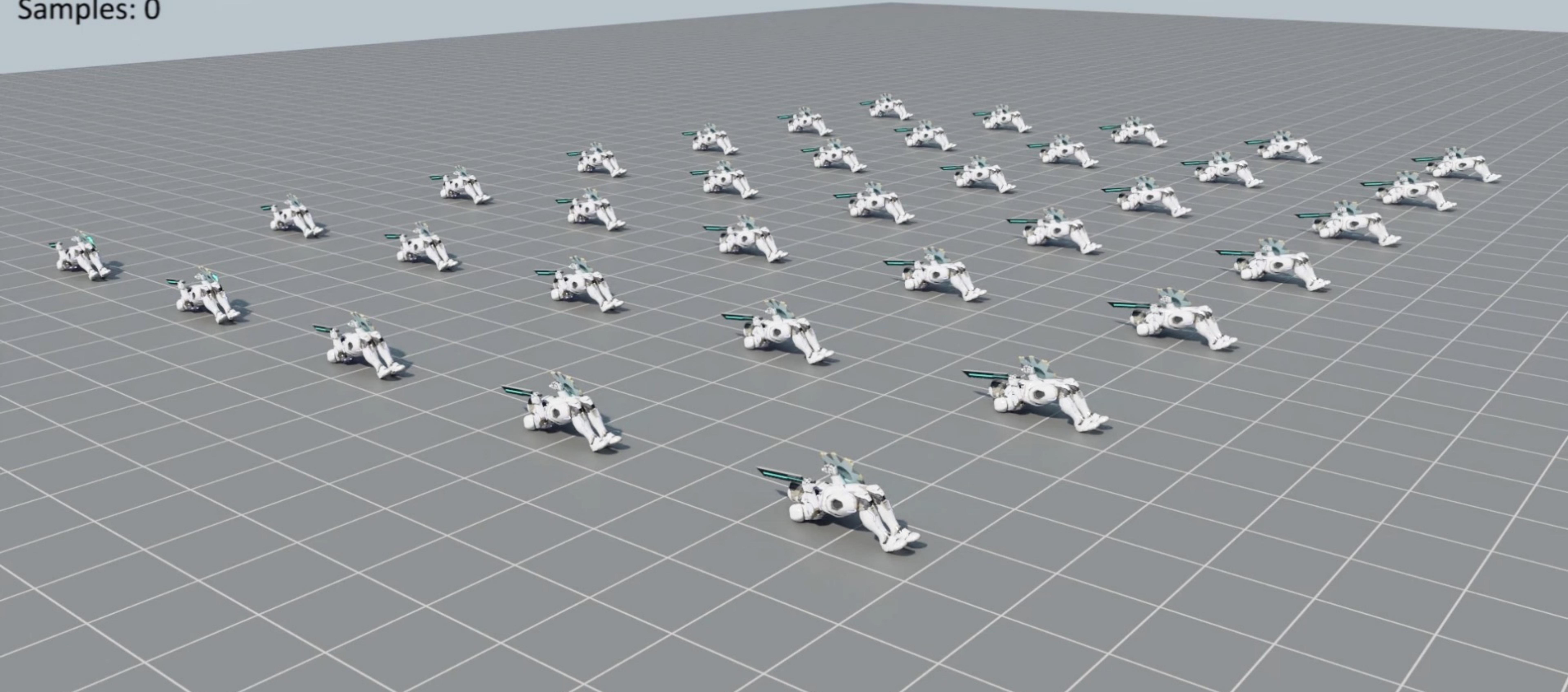


Large-Scale Training

Sim Time: 0 days

Real Time: 0 minutes

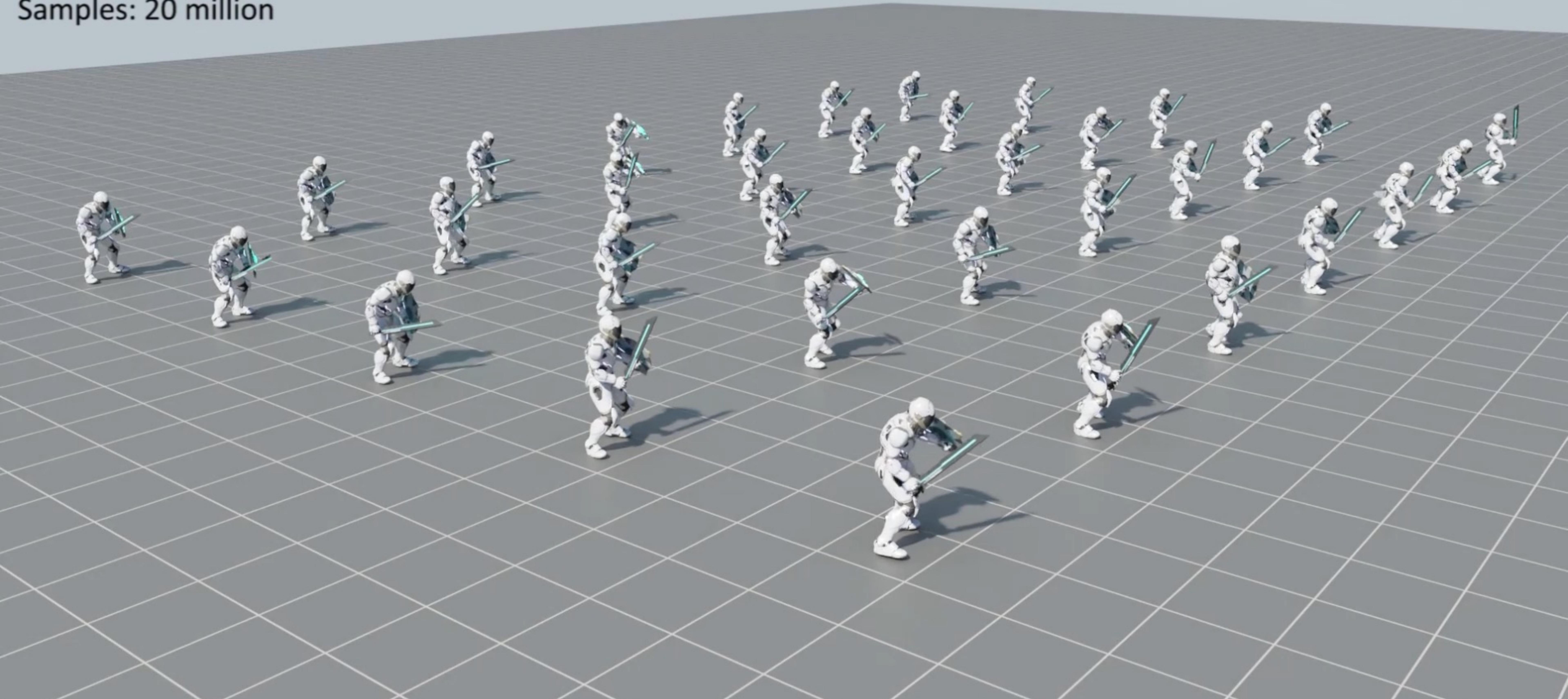
Samples: 0





Large-Scale Training

Sim Time: 1 week
Real Time: 30 minutes
Samples: 20 million





Large-Scale Training

Sim Time: 2 months

Real Time: 4 hours

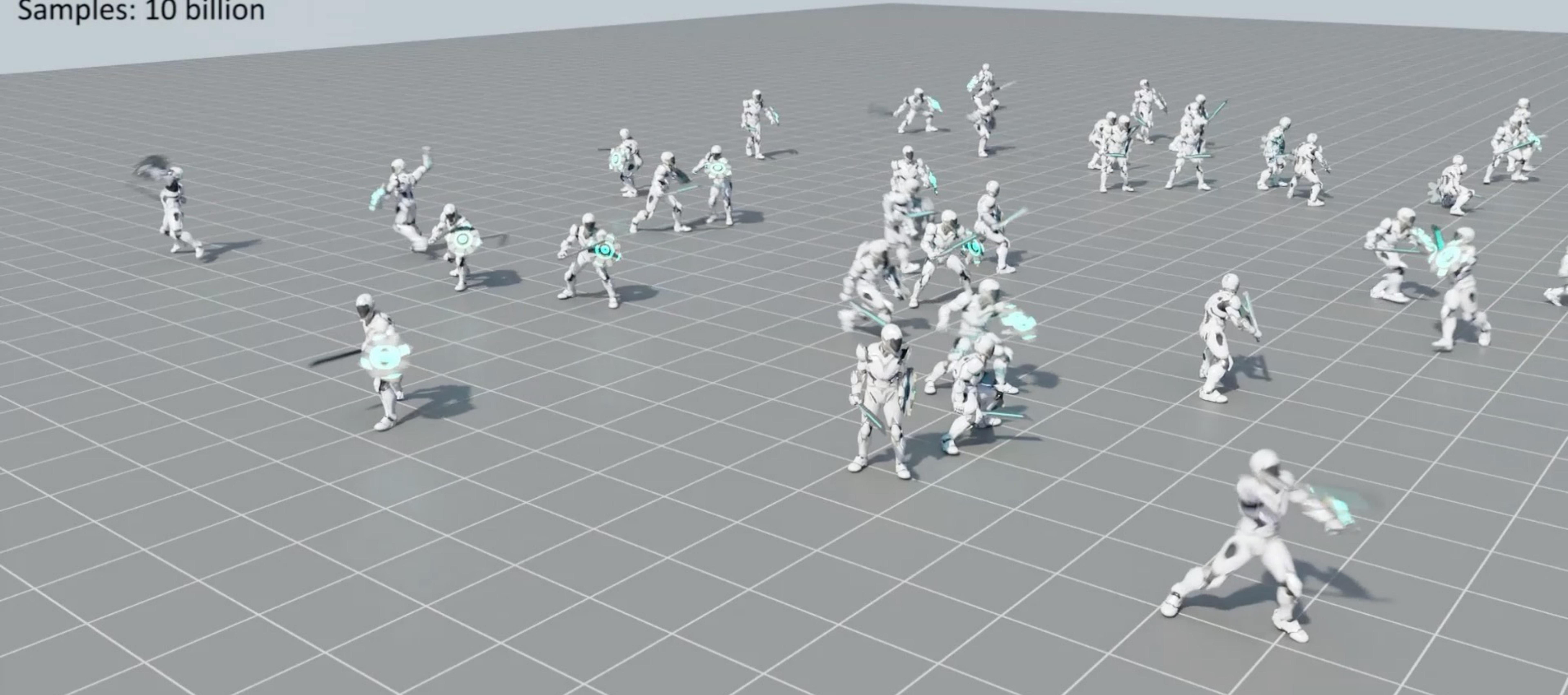
Samples: 160 million



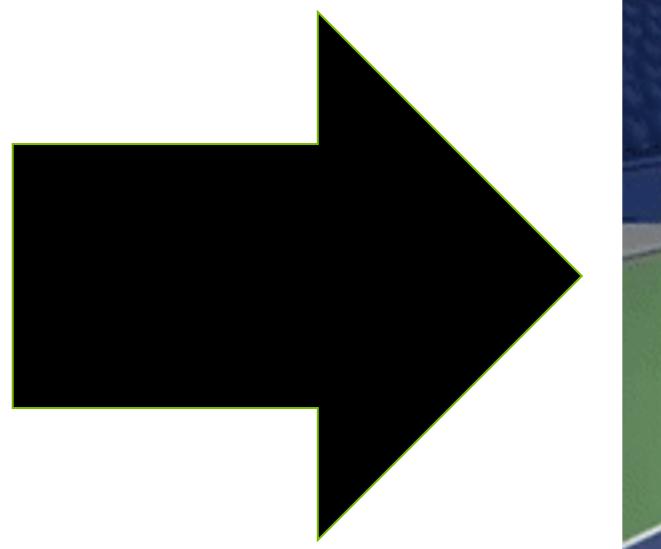
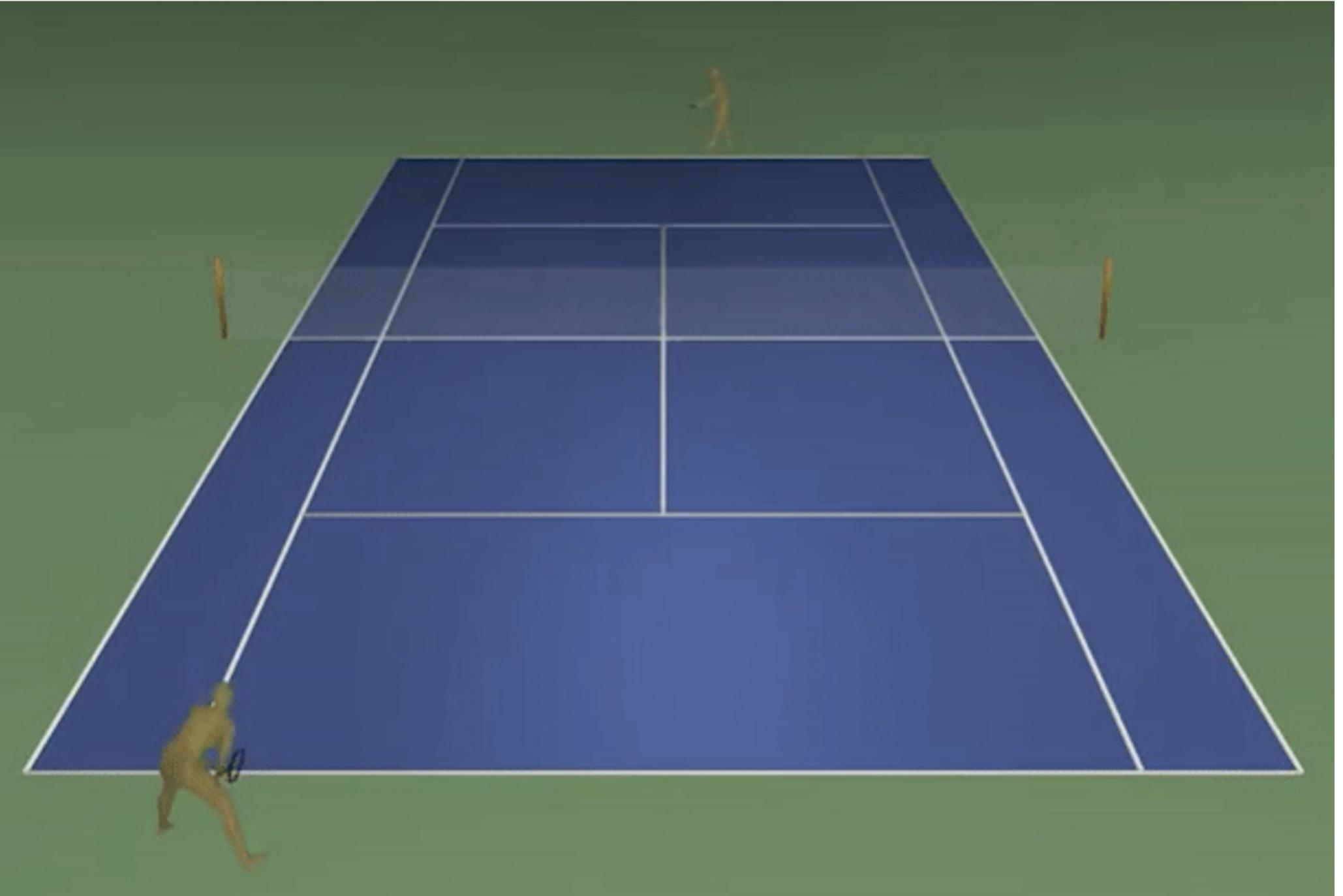
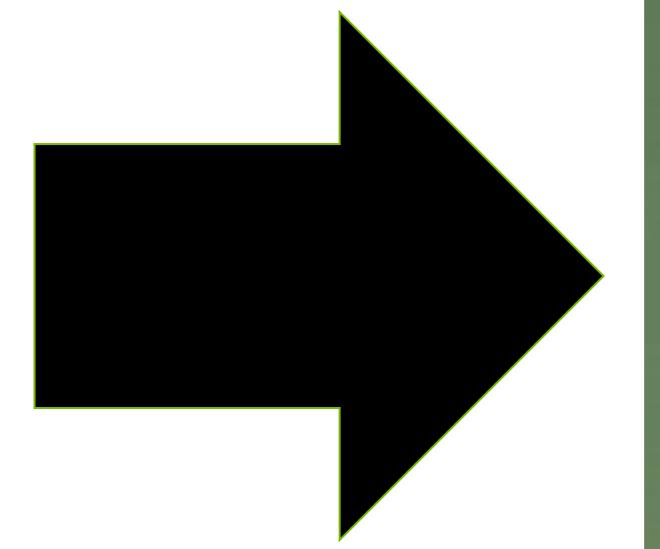


Large-Scale Training

Sim Time: 10 years
Real Time: 10 days
Samples: 10 billion



Game AI (Tennis) Learned from Broadcast videos



Video Data

Motion
Reconstruction

Physics
Simulation

Game AI (Tennis) Learned from Broadcast videos



Learning Physically Simulated Tennis Skills from Broadcast Videos



