

NeRF for Manipulation

박지호

Contents

- **What is NeRF for?**
- **Progress in NeRF Research**
 - 0. High-quality Reconstruction
 - 1. Faster Reconstruction in NeRF
 - 2. Generalization in NeRF
 - 3. Semantic Information in NeRF
- **How are these method applied to Manipulation**

NeRF



NeRF

Task: 3D Reconstruction from 2D Images

Method: Optimize 3D(MLP) with Ground Truth 2D Images

3D Representation(MLP) \rightarrow 2D Image \rightarrow Loss(rendered_img, GT)



NeRF

Donut

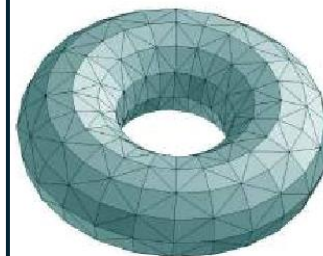


NeRF

Donut



Explicit



Mesh:

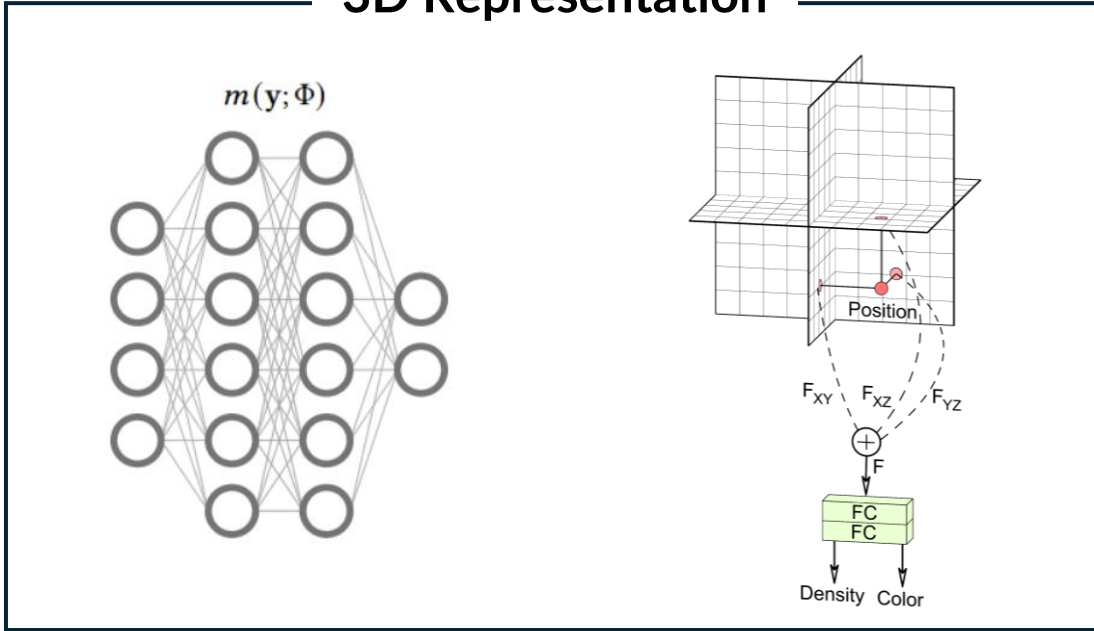
Points of triangles

Implicit

$$(R - \sqrt{x^2 + y^2})^2 + z^2 = r^2$$

NeRF

3D Representation

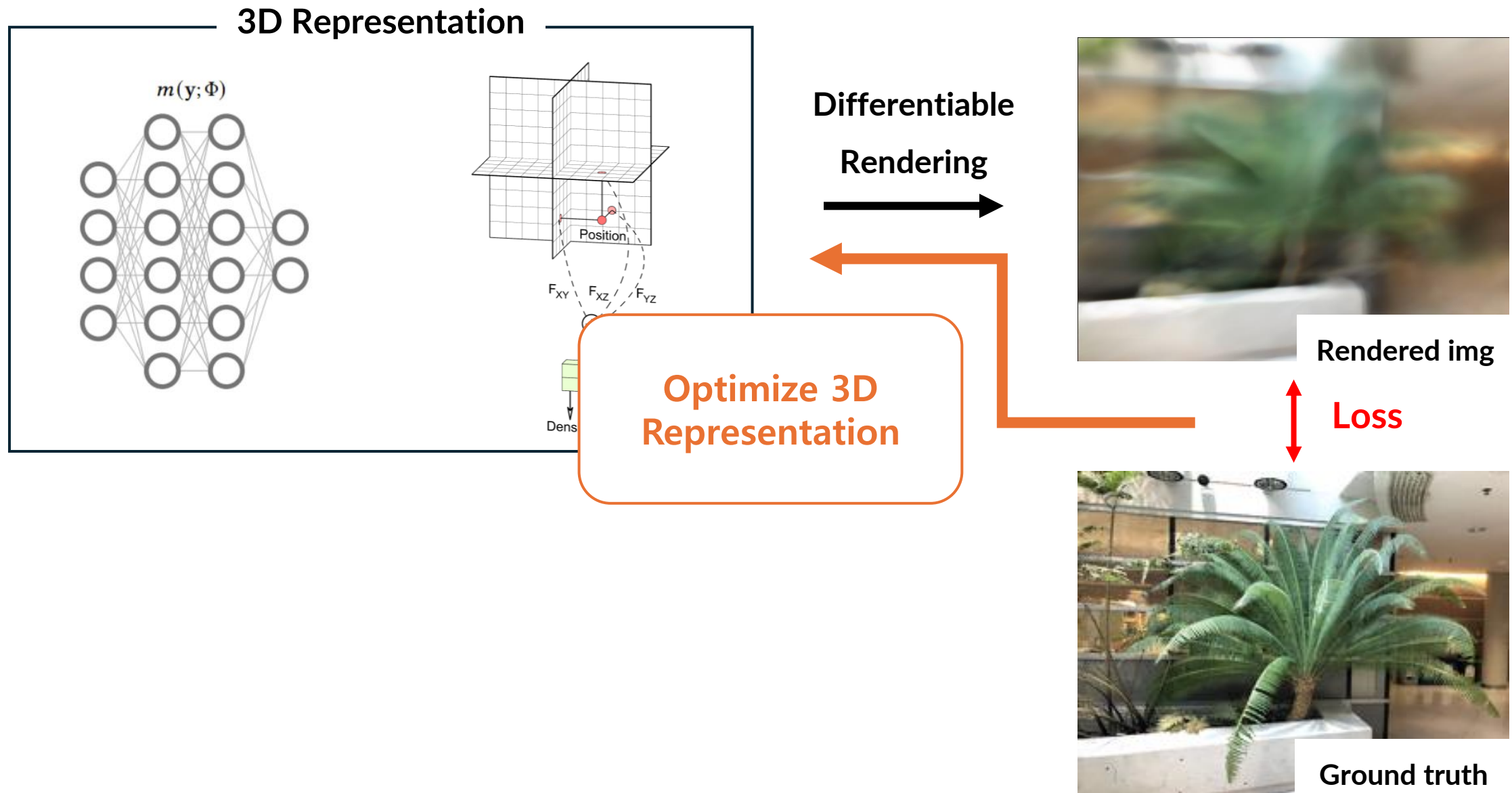


Differentiable
Rendering

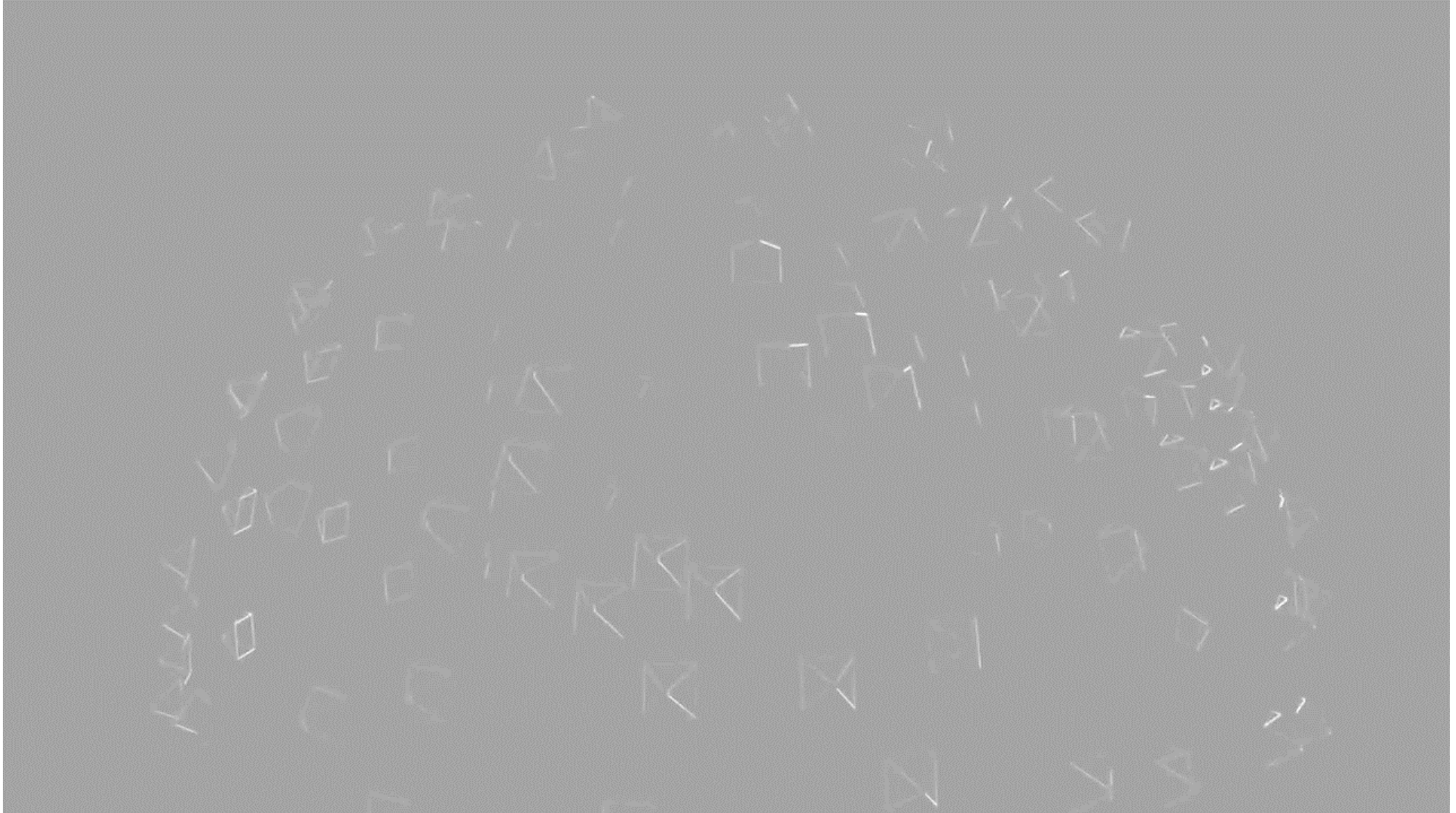


Rendered img

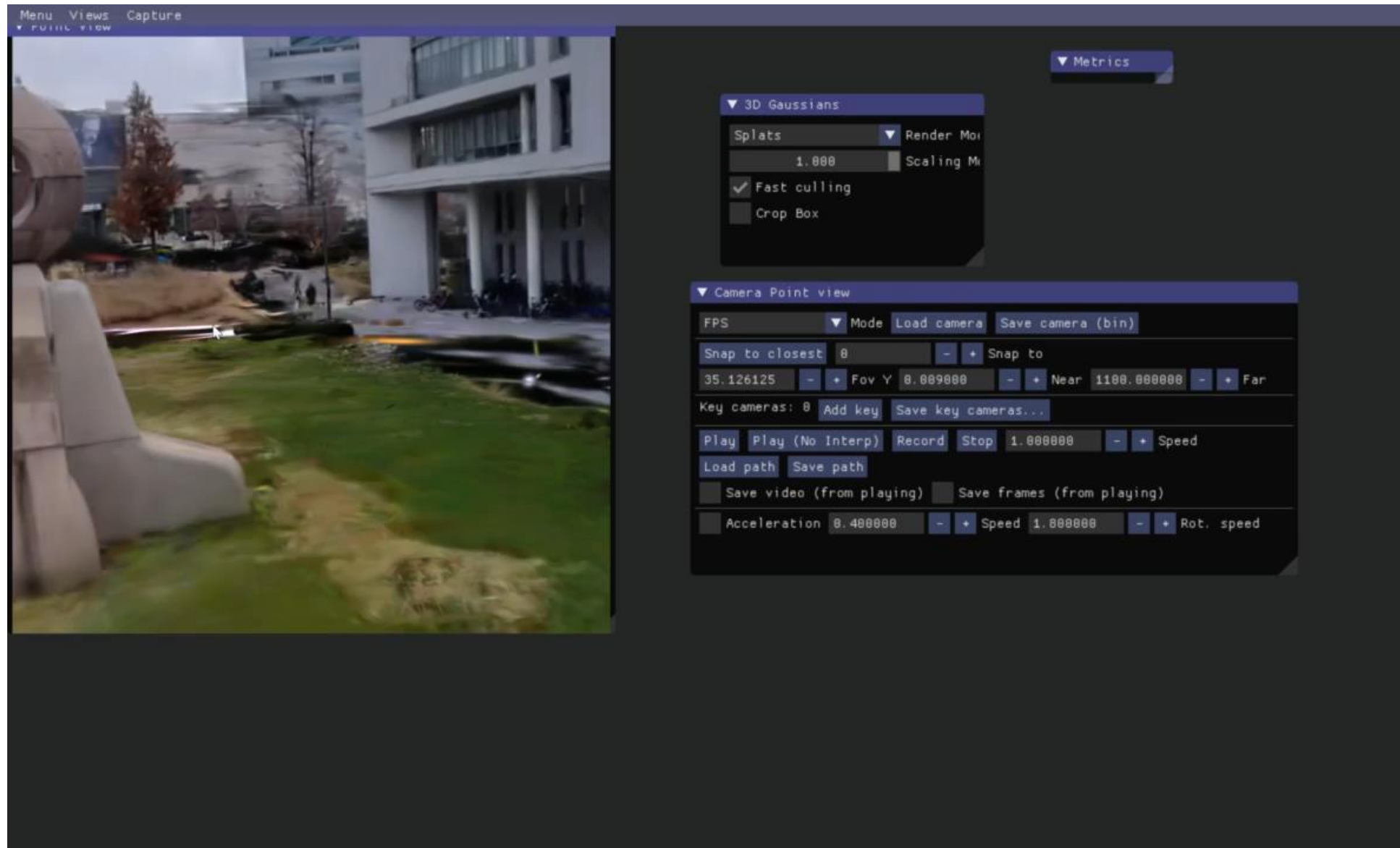
NeRF



NeRF

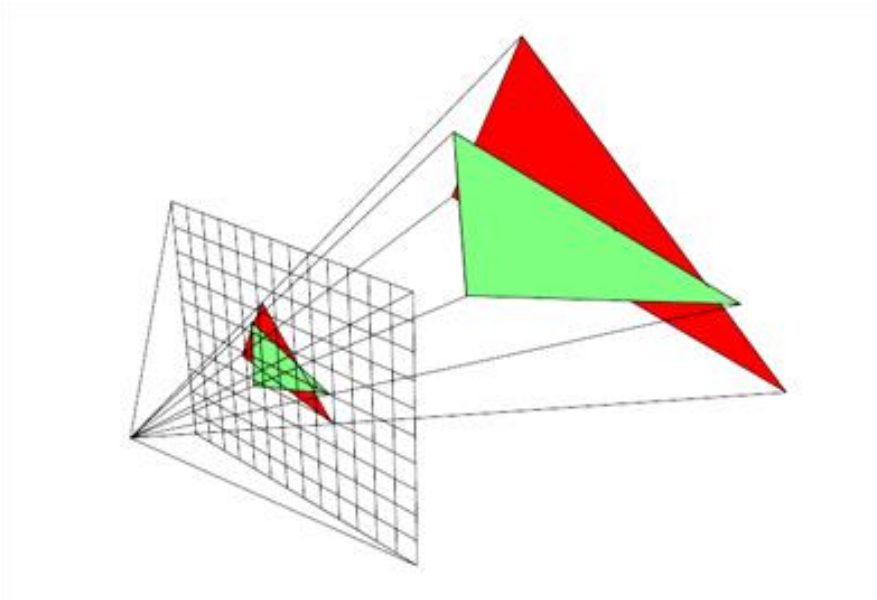


3D Gaussian Splatting



3D Gaussian Splatting

- No Neural Network, Only Point Clouds
- No Ray, Rasterization-based



Advantages

- **Fast Rendering (Real-time Rendering)**
- Fast Training
- High Quality (not best)
- Easier Code

Limitations

- Large Memory
- Splotchy Artifacts

What is this 'NeRF' for?

Obviously...

What is this 'NeRF' for?

Obviously... Entertainment! VR/AR! Metaverse!



4K4D: Real-Time 4D View Synthesis at 4K Resolution

What is this 'NeRF' for?

Obviously... Entertainment! VR/AR! Metaverse!



What is this 'NeRF' for?

Obviously... Entertainment! VR/AR! Metaverse!

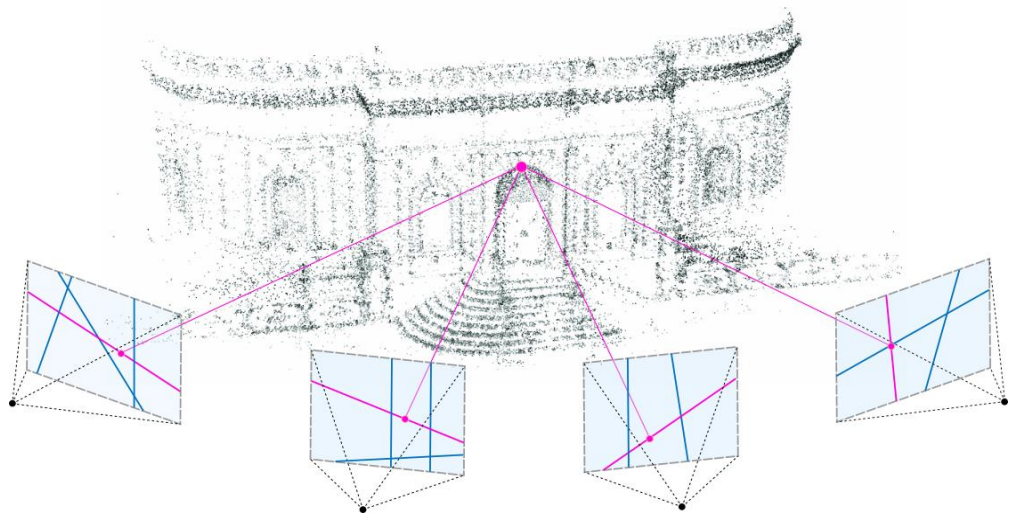


What Else...? Robotics, Simulation?

What is NeRF Doing?

3D Reconstruction = 'Understanding 3D Space/Object'

What's different from conventional 3D Reconstruction? ➔ **Photo-Realism..!**



Structure-from-Motion

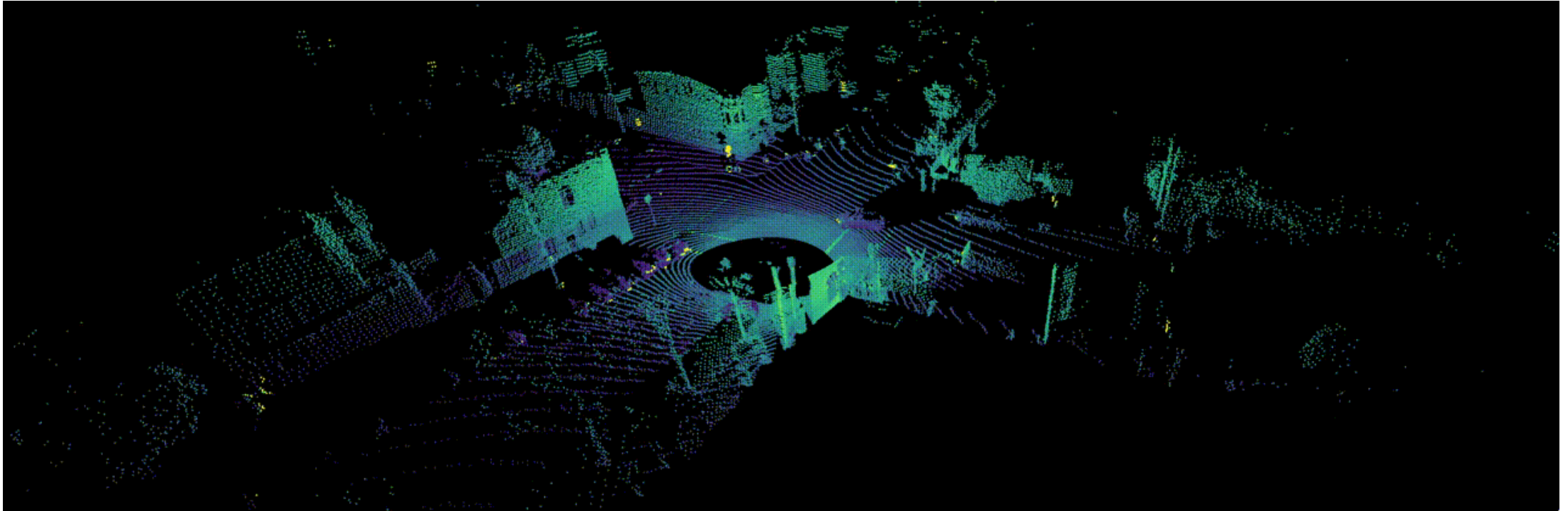
VS



NeRF

What is NeRF Doing?

Conventional 3D Recon → Focused on **Geometry**...! (Different Purpose)

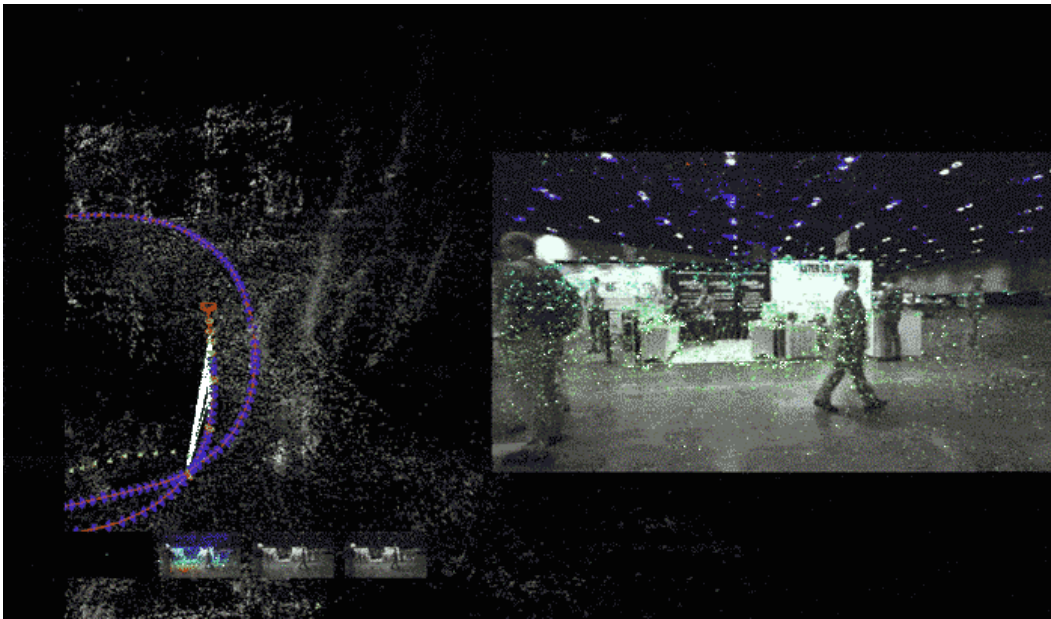


SLAM

What is NeRF Doing?

Conventional 3D Recon → Focused on **Geometry**...!

While NeRF task aims **Complete Reconstruction** (both Photo-realism & Geometry)



SLAM



NeRF Task

NeRF for Robotics

Is NeRF 'too much' for Robotics?

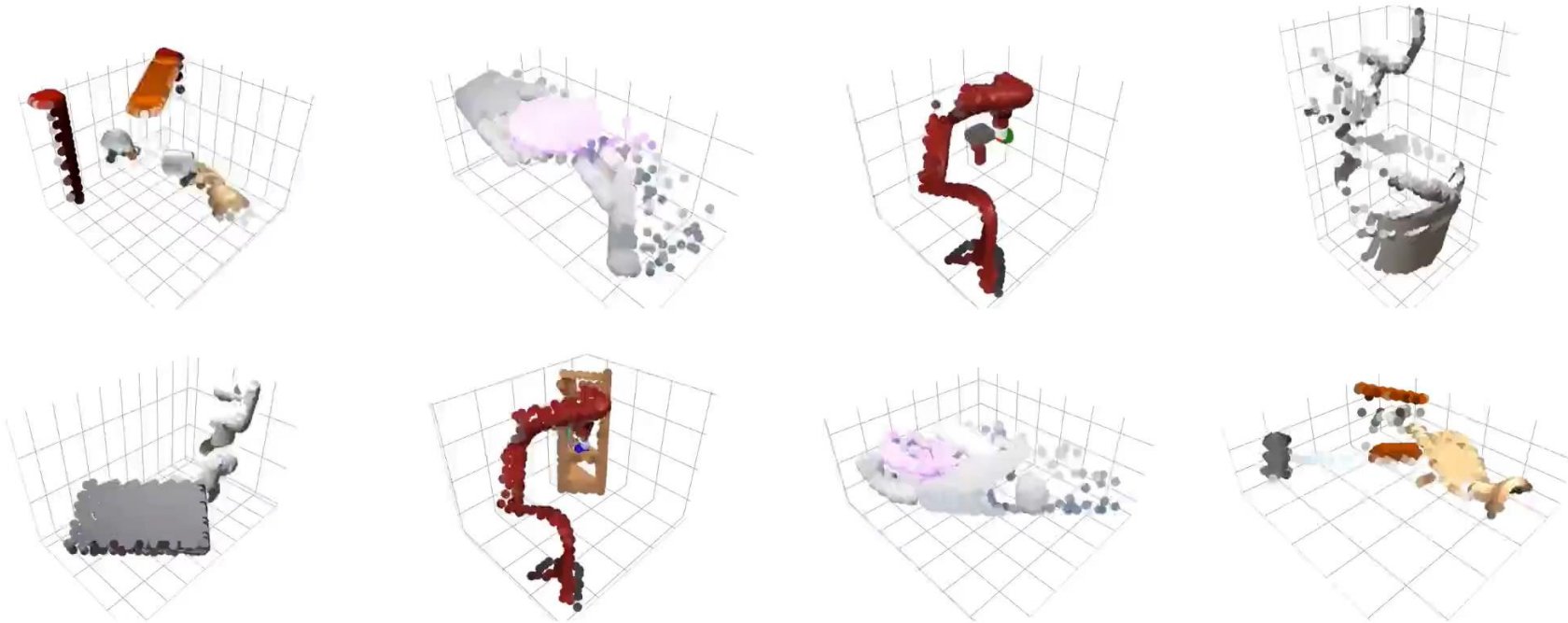
=

Full-3D is Expensive, Is it worth it?

NeRF for Robotics

3D + Manipulation

- Leveraging 3D Information
- [3D Diffusion Policy](#) (24 arxiv)
 - **Point Cloud** Conditioned Diffusion



NeRF for Robotics: Progress in NeRF Research

0. High-quality Reconstruction

- Mip-NeRF(21), Mip-Splatting(23)

1. Faster Reconstruction in NeRF

- Instant-NGP(21), GaussianSplatting(23)

2. Generalization in NeRF

- PixelNeRF(20), PixelSplat(23)

3. Semantic Information in NeRF

- LeRF(23), LangSplat(24)

NeRF for Robotics: Progress in NeRF Research

0. ~~High-quality Reconstruction~~

- Mip-NeRF(21), Mip-Splatting(23)

1. Faster Reconstruction in NeRF

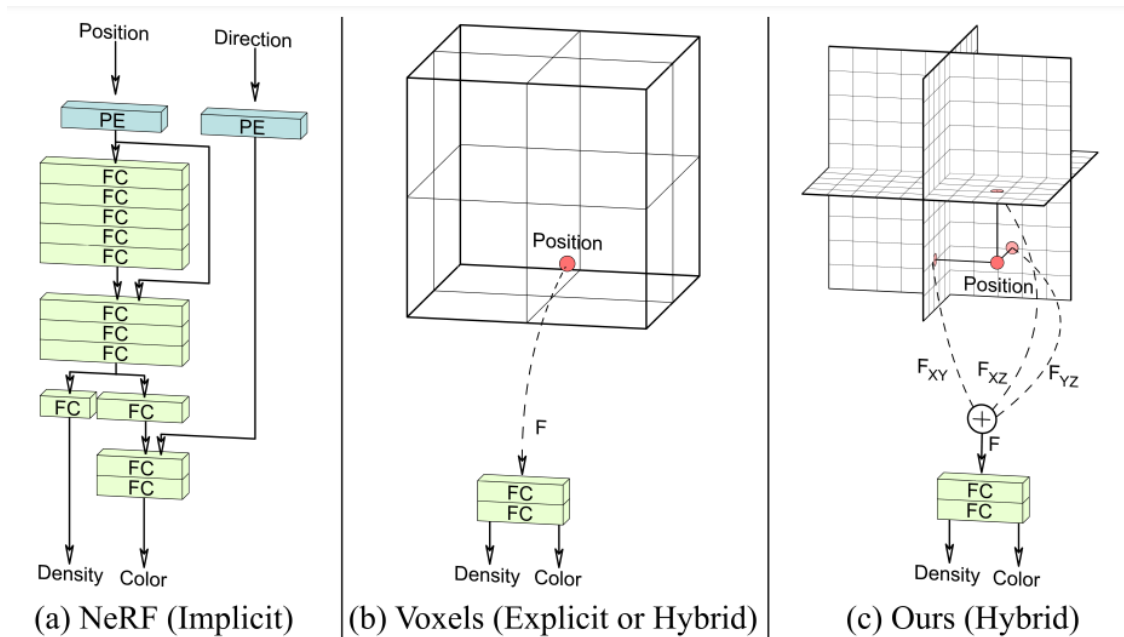
- Instant-NGP(21), GaussianSplatting(23)

2. Generalization in NeRF

- PixelNeRF(20), PixelSplat(23)

3. Semantic Information in NeRF

- LeRF(23), LangSplat(24)



Tri-plane representation from EG3D

NeRF for Robotics: Progress in NeRF Research

0. ~~High-quality Reconstruction~~

- Mip-NeRF(21), Mip-Splatting(23)

1. Faster Reconstruction in NeRF

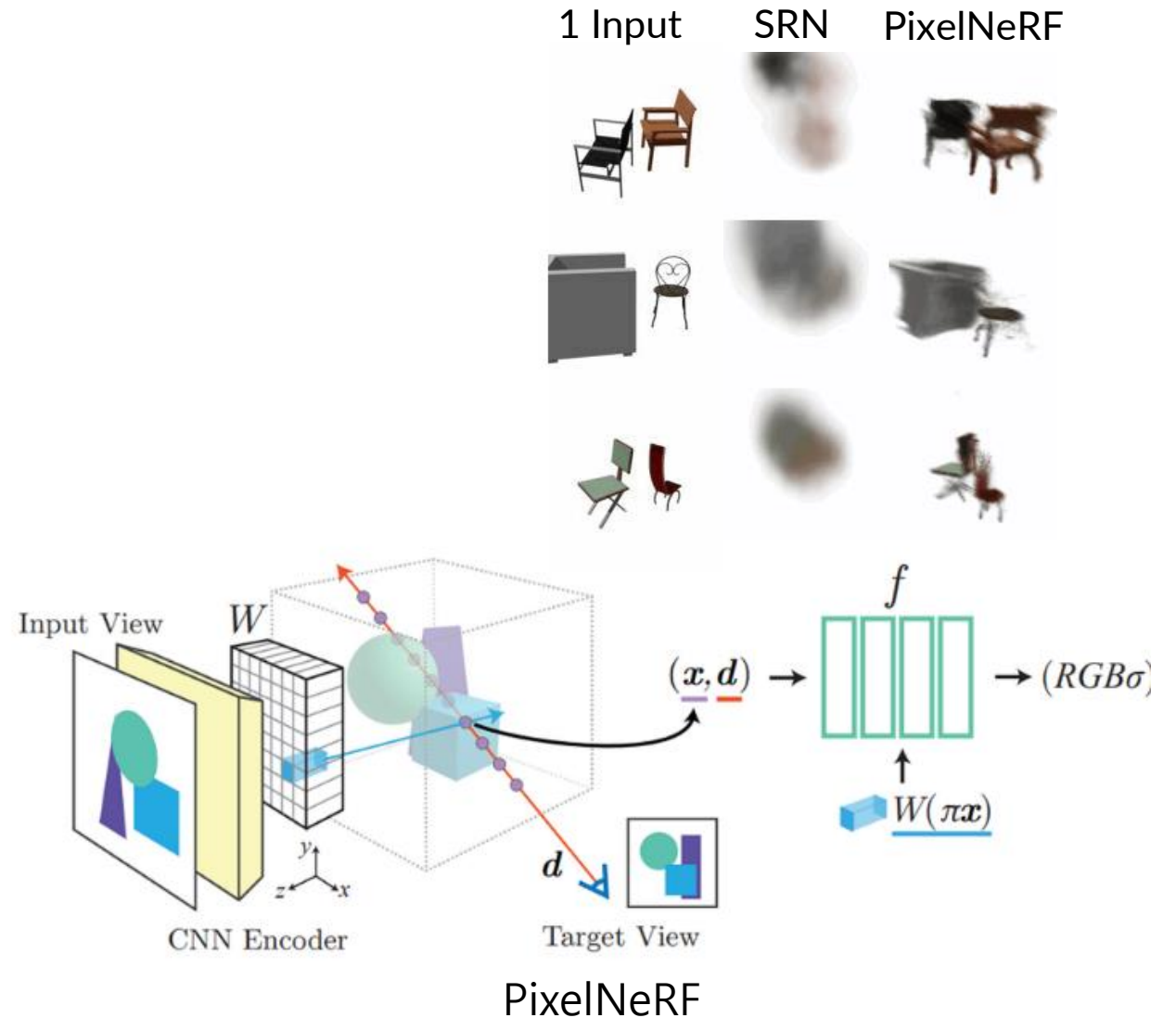
- Instant-NGP(21), GaussianSplatting(23)

2. Generalization in NeRF

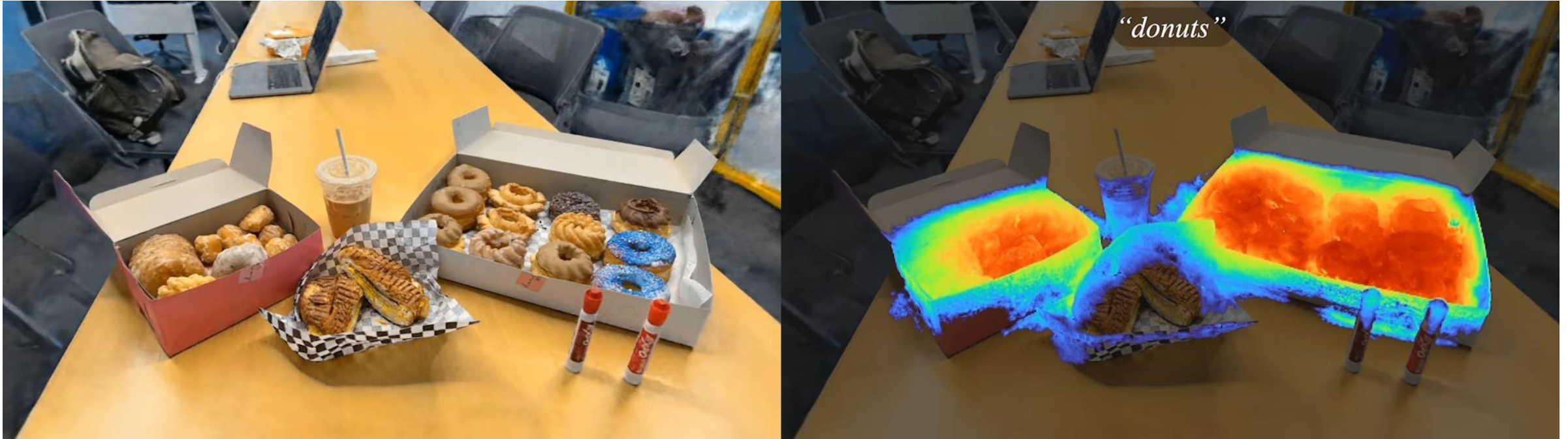
- PixelNeRF(20), PixelSplat(23)

3. Semantic Information in NeRF

- LeRF(23), LangSplat(24)



NeRF for Robotics: Progress in NeRF Research



3. Semantic Information in NeRF

- LeRF(23), LangSplat(24)

NeRF for Robotics: Progress in NeRF Research

0. High-quality Reconstruction

- Mip-NeRF(21), Mip-Splatting(23)

1. Faster Reconstruction in NeRF

- Instant-NGP(21), GaussianSplatting(23)

2. Generalization in NeRF

- PixelNeRF(20), PixelSplat(23)

3. Semantic Information in NeRF

- LeRF(23), LangSplat(24)

How can these method applied in Manipulation task?

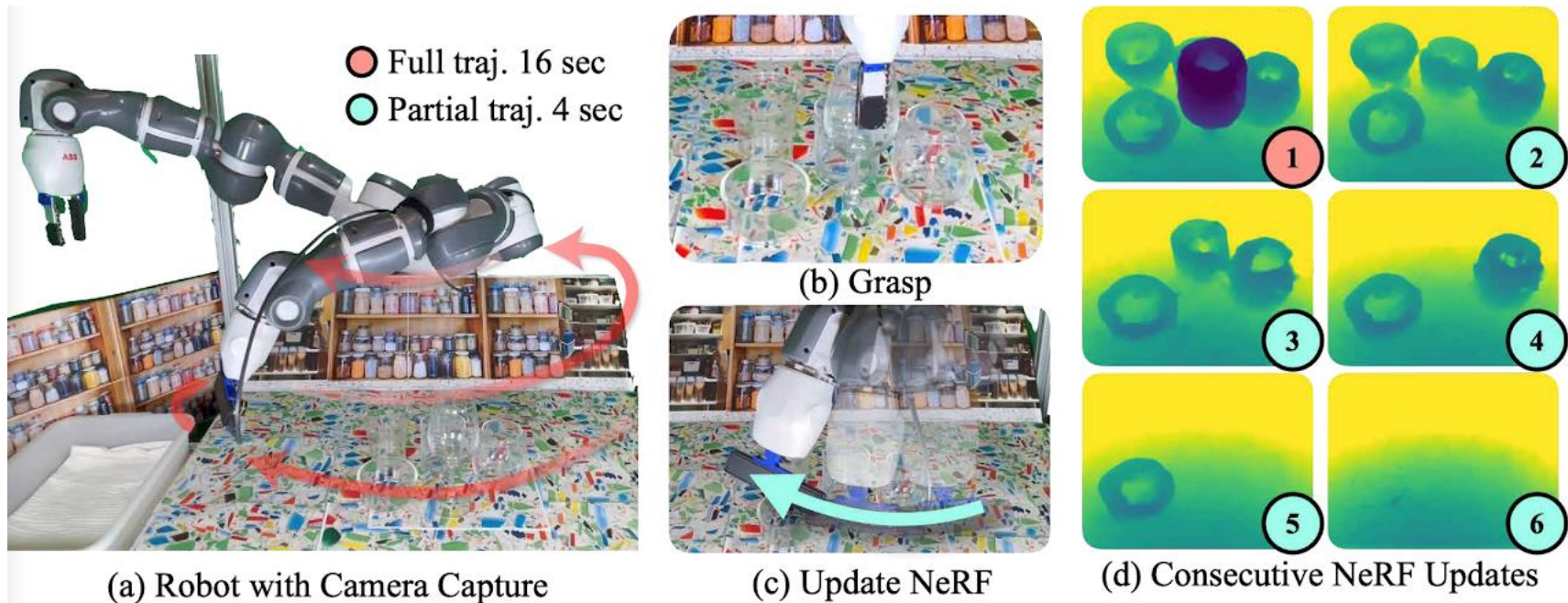
Basic NeRF Utilization:

Scan -> Recon -> Grasp

NeRF for Manipulation

1. Faster Reconstruction in NeRF

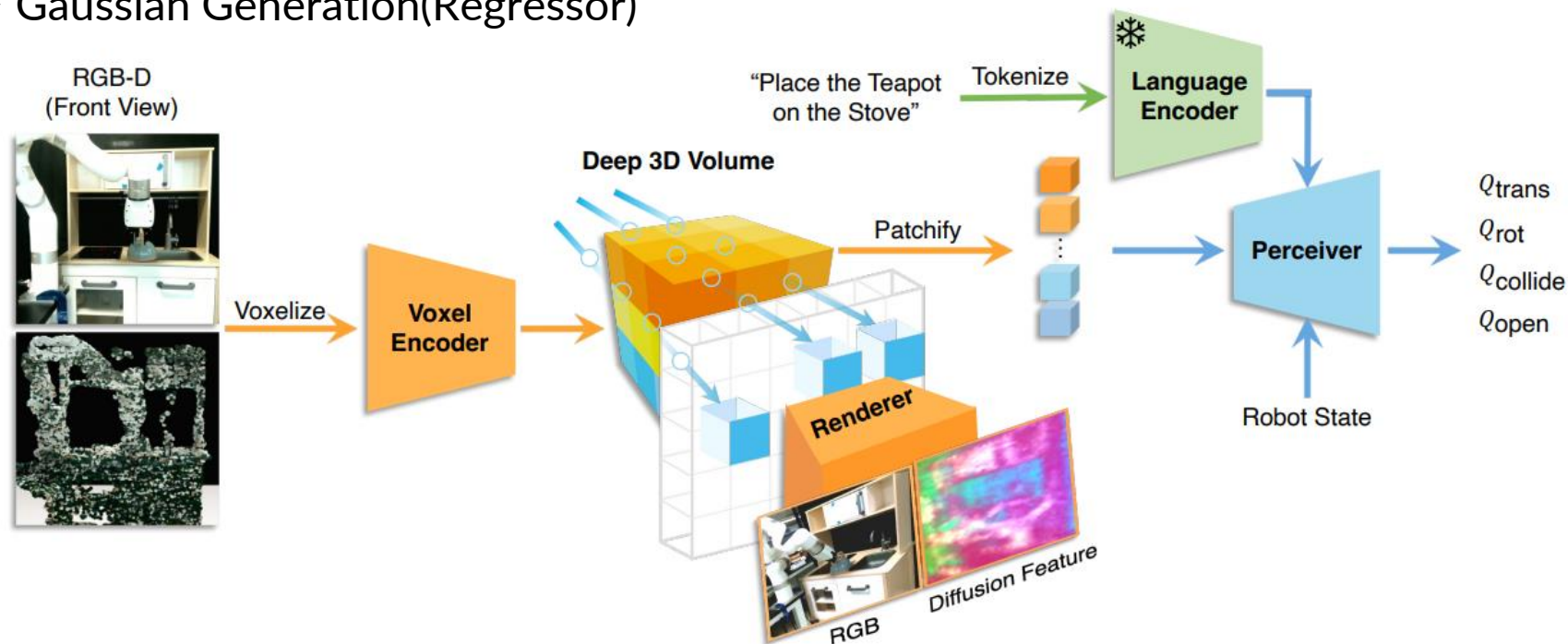
- Dex-NeRF(CoRL21)
 - Transparent Object Geometry(Depth Camera didn't work)
- Evo-NeRF(CoRL22 Oral)
 - Using Instant-NGP for faster recon



NeRF for Manipulation

2. Generalization in NeRF Reconstruction is still slow! → Generation Method

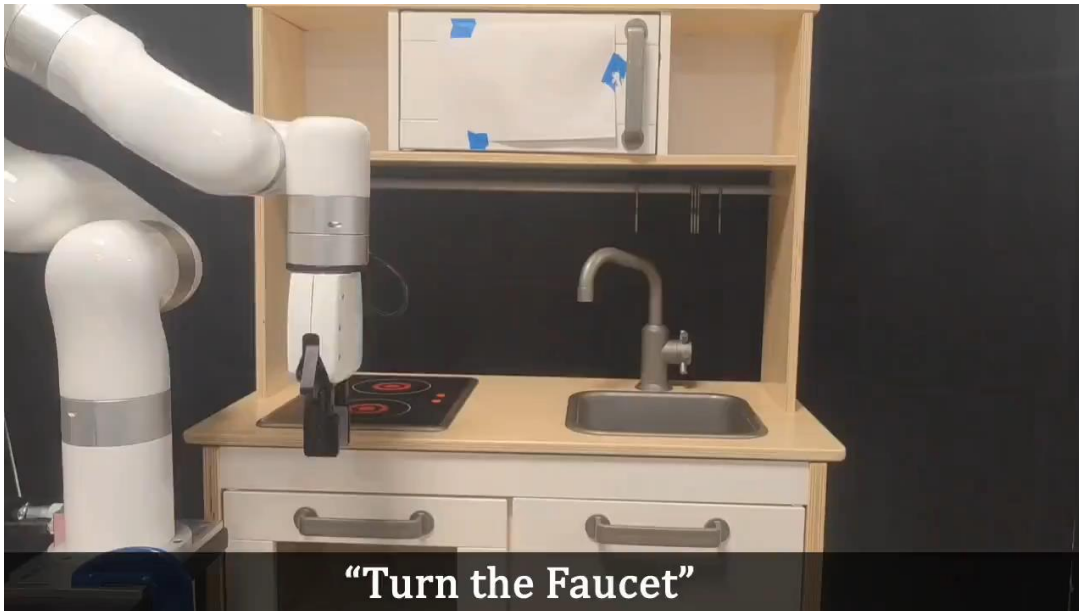
- GNFactor(CoRL 23 Oral)
 - ✓ Few-shot Img → NeRF Generation
- ManiGaussian(ECCV 24)
 - ✓ Few-shot Img → Gaussian Generation(Regressor)



NeRF for Manipulation

2. Generalization in NeRF Reconstruction is still slow! → Generation Method

- GNFactor(CoRL 23)
 - ✓ Few-shot Img → NeRF Generation



NeRF for Manipulation

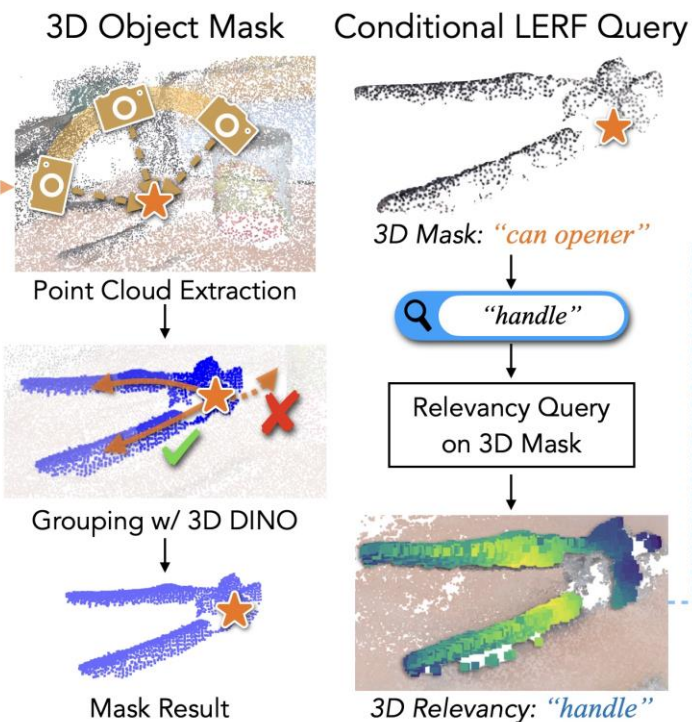
3. Semantic Information in NeRF

- LeRF-ToGo (CoRL 23 Best Paper Finalist)
- Distilled Feature Fields(CoRL 23 Best Paper)
 - ✓ Zero-shot Text2Grasp
 - ✓ Leveraging Clip Feature

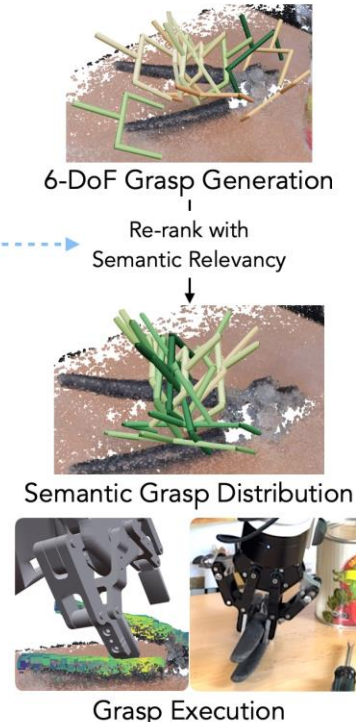
1. Scene Reconstruction



2. Object + Part Selection



3. Grasp Selection



NeRF for Manipulation

3. Semantic Information in NeRF

- LeRF-ToGo (CoRL 23 Best Paper Finalist)
- Distilled Feature Fields(CoRL 23 Best Paper)
 - ✓ Zero-shot Text2Grasp
 - ✓ Leveraging Clip Feature



NeRF for Manipulation

3. Semantic Information in NeRF

- LeRF-ToGo (CoRL 23 Best Paper Finalist)
- Distilled Feature Fields(CoRL 23 Best Paper)
 - ✓ Zero-shot Text2Grasp
 - ✓ Leveraging Clip Feature



NeRF for Manipulation

3. Semantic Information in NeRF

- LeRF-ToGo (CoRL 23 Best Paper Finalist)
- Distilled Feature Fields(CoRL 23 Best Paper)
 - ✓ Zero-shot Text2Grasp
 - ✓ Leveraging Clip Feature



Discussion

Using NeRF for Robotics?



Managing Trade-off between **Cost** and **Perception**

Real 2 Sim

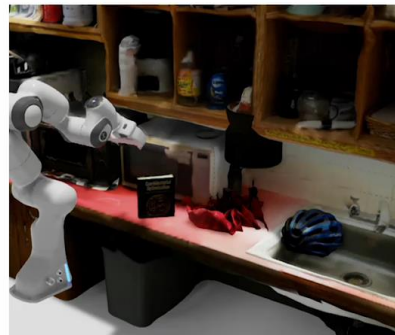
Digital Twin for Sim2Real

- Real-to-Sim-to-Real Approach for Robust Manipulation (RSS 24)
 - ✓ Hard to do RL in Real World: Real-2-Sim(Recon) \rightarrow RL \rightarrow Sim2Real

3D scene
reconstruction



RL fine-tuning
in sim



Robust policy in
the real world



Real-to-sim transfer
of policies

Thank You!