An Introduction to Implicit Neural Representations for Image-Based Modeling

Xiaowei Zhou
Zhejiang University

Overview

Image-based modeling and rendering

Traditional methods and their limitations

Implicit Neural Representations

- Neural Radiance Fields (NeRF)
- Neural SDF for surface reconstruction
- Neural dynamic scene representations

Image-based modeling and rendering

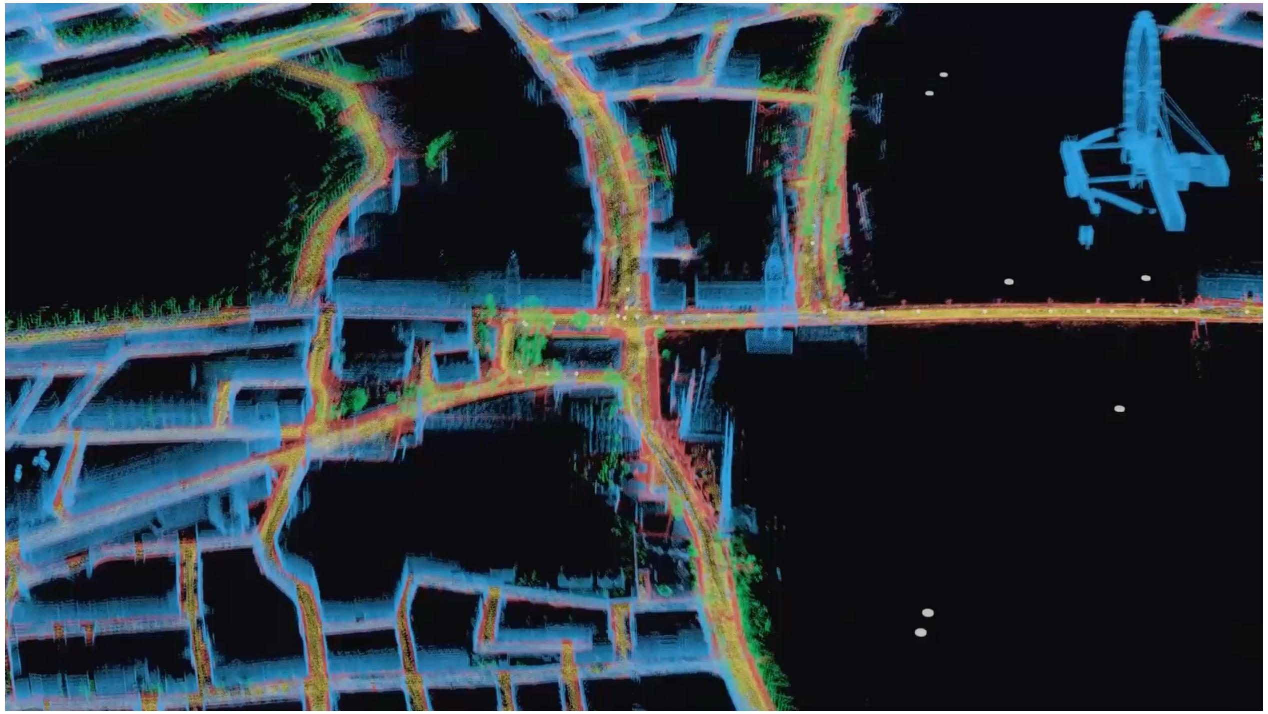
Input images 3D model Novel views

Nerf in the wild: Neural radiance fields for unconstrained photo collections. CVPR. 2021.

Applications

VR tour

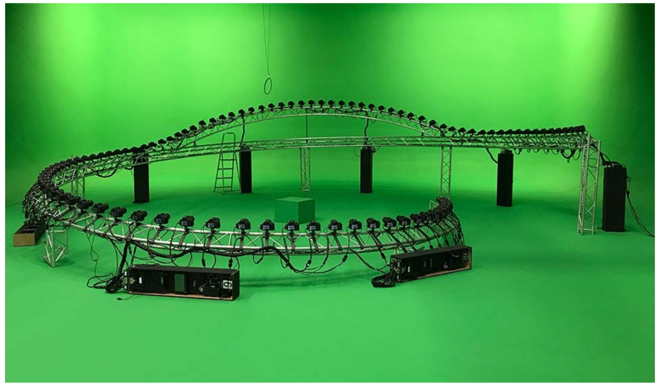
Matterport



Google Immersive View

Applications

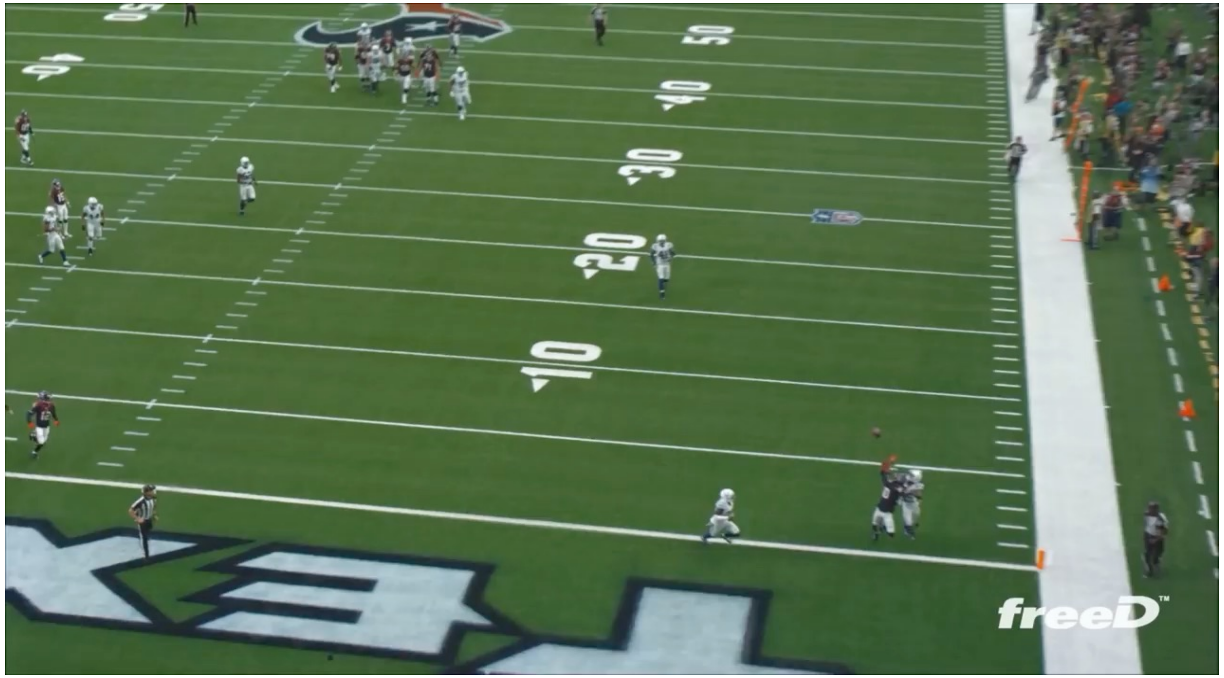
Bullet time effect



Bullet time effect in "The Matrix"

Applications

Free-viewpoint video

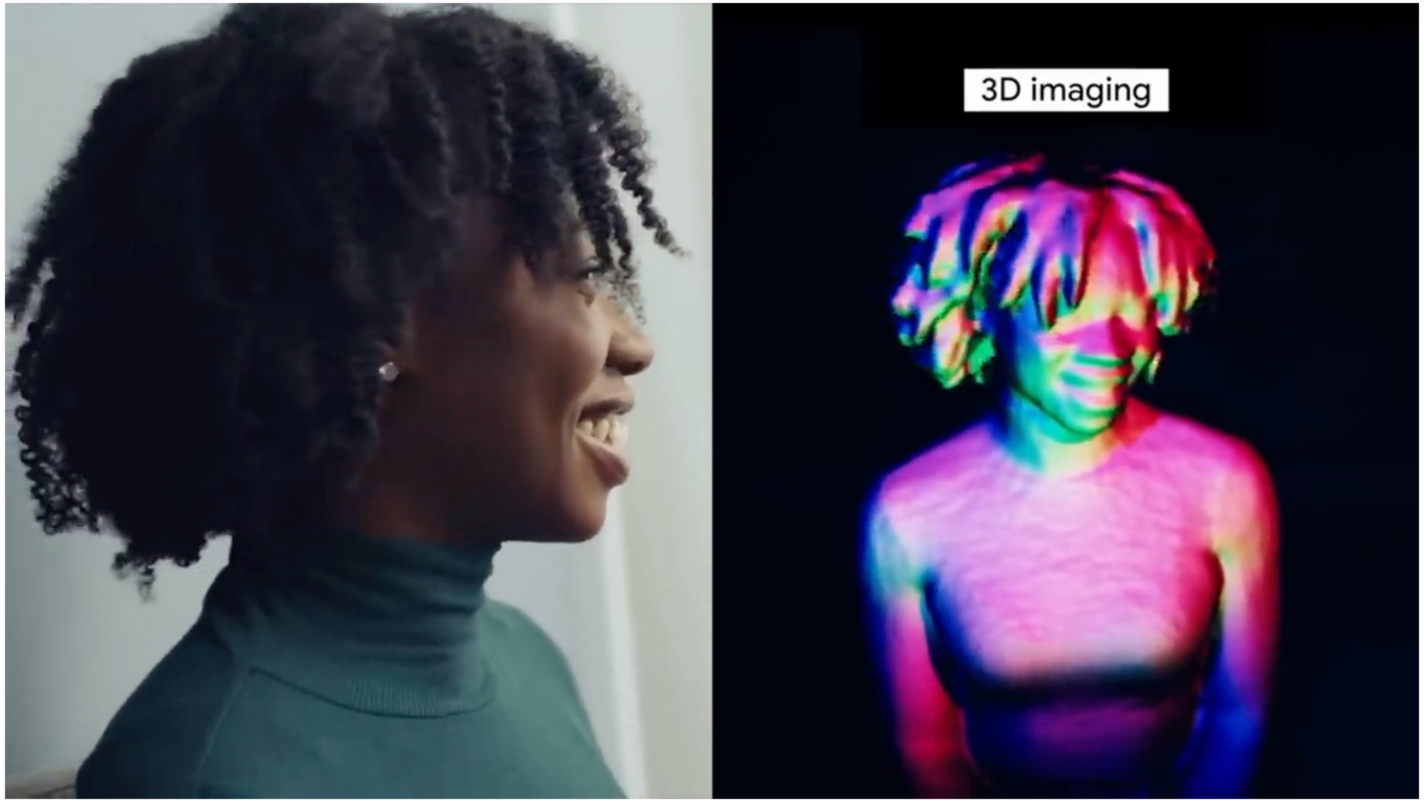


Intel TrueView

湖南卫视舞蹈风暴

Applications

Immersive telepresence



Google Project Starline

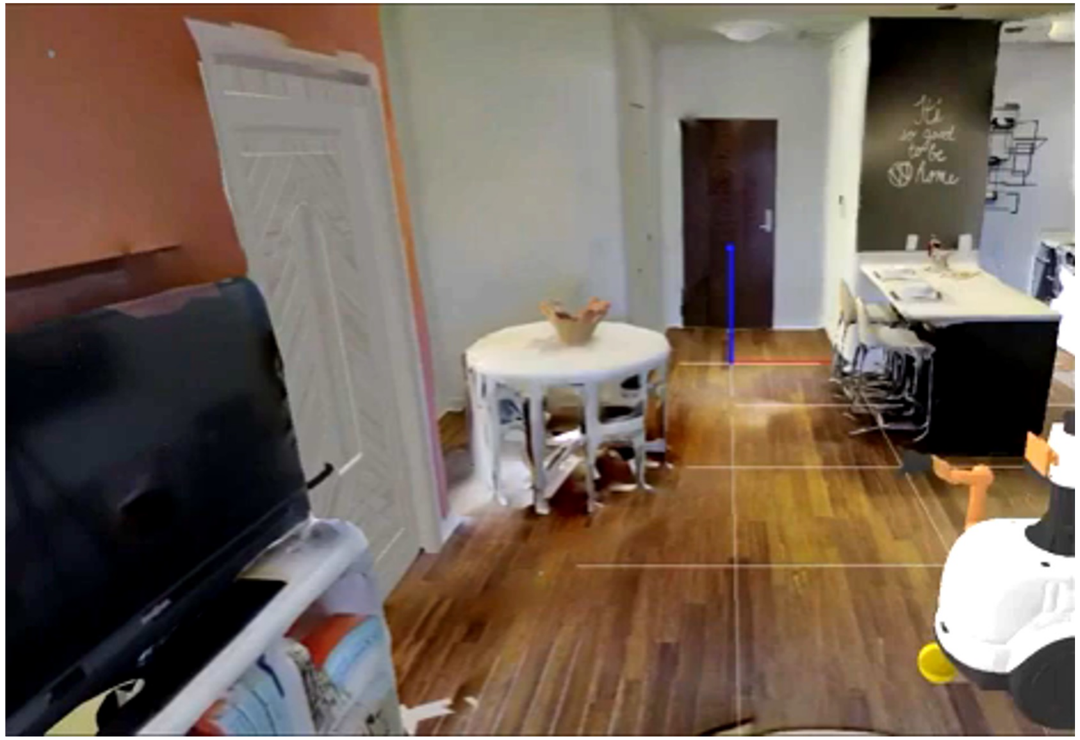
Applications

Embodied AI: training agents in simulated environments

Autonomous driving

Block NeRF

Robots



iGibson

Image-based modeling and rendering

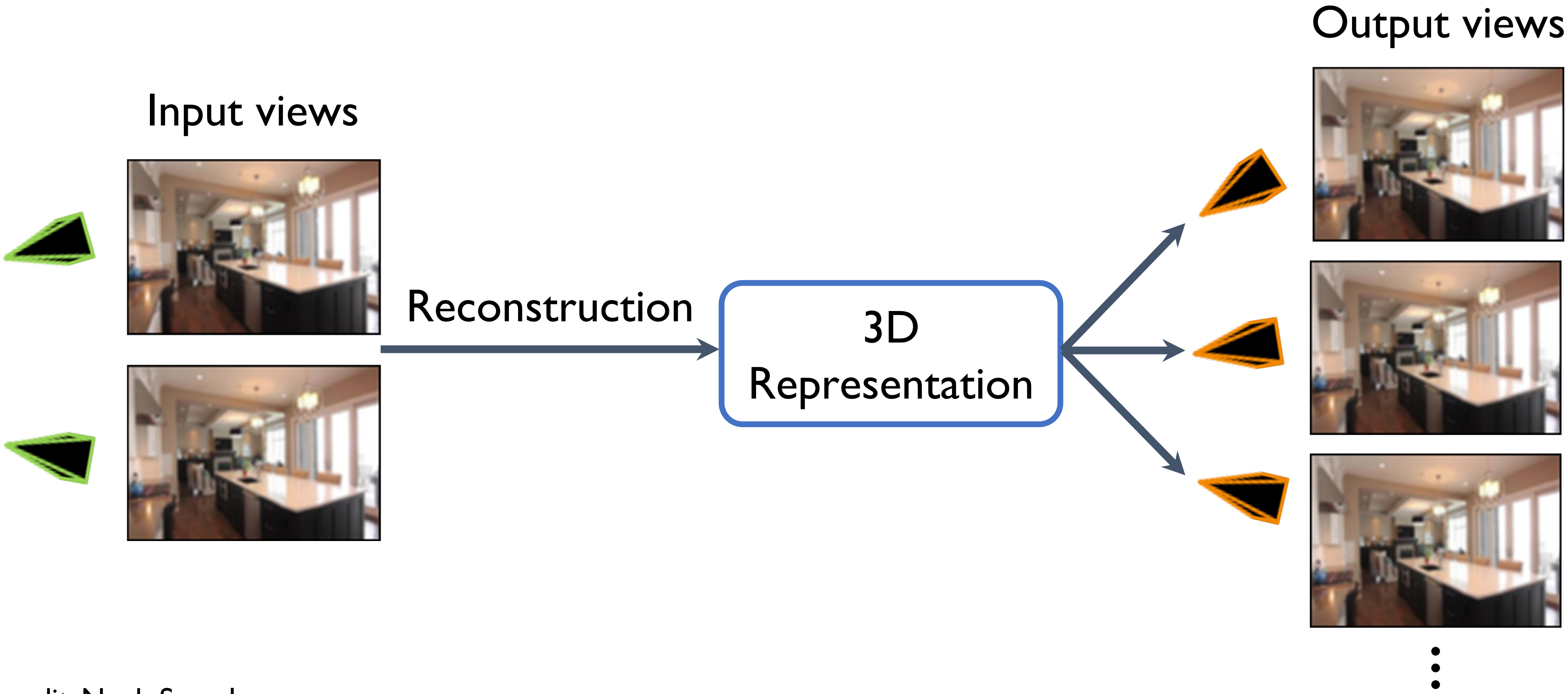
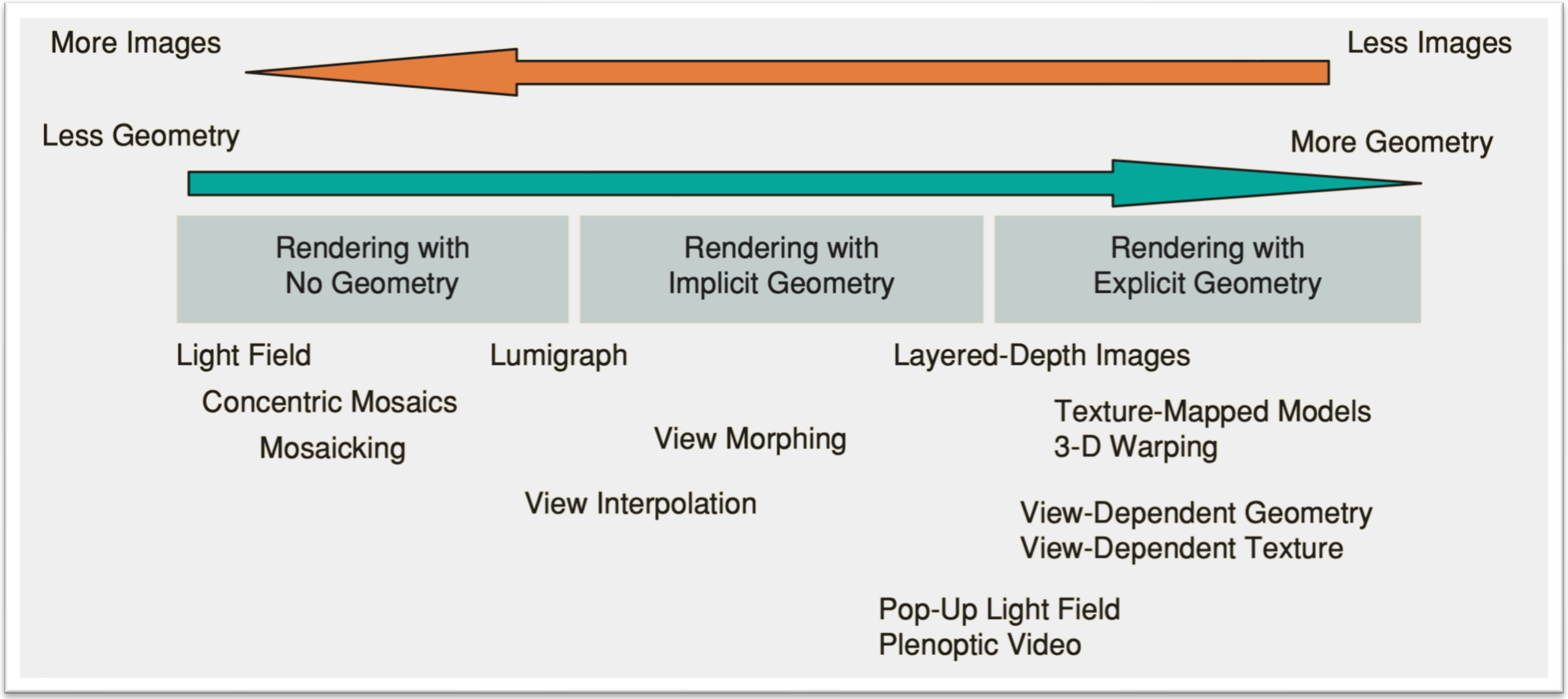


Image credit: Noah Snavely

Traditional methods



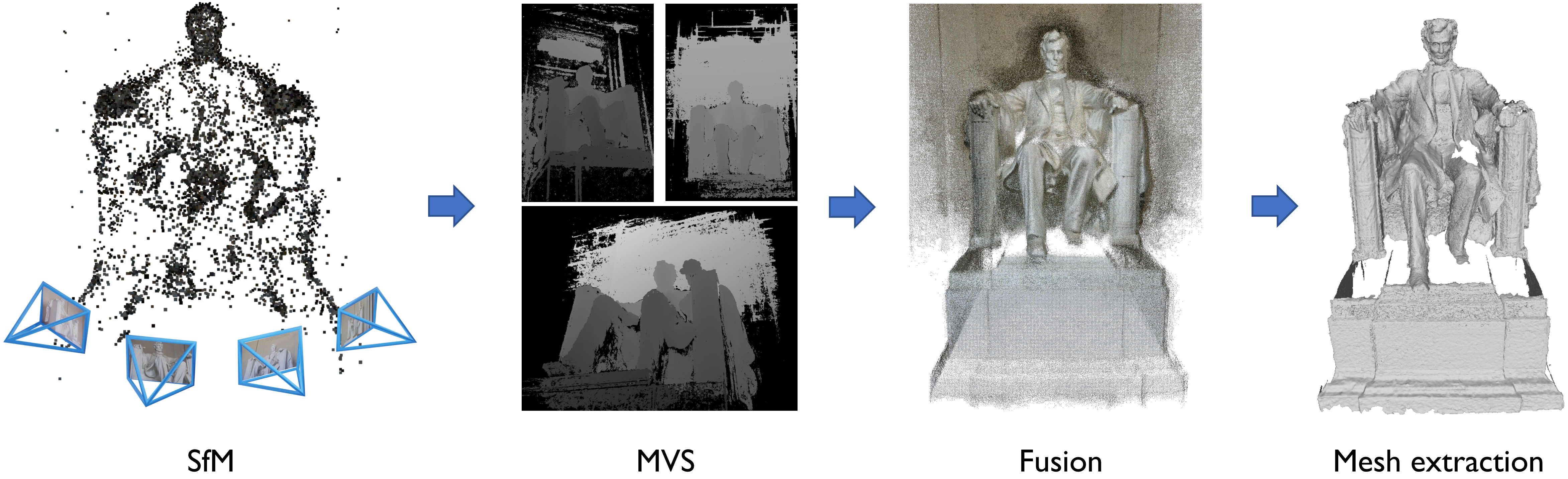
A Review of Image-based Rendering Techniques, *Visual Communications and Image Processing* 2000.

Surface-based representations

3D mesh with texture map

Surface-based representations

Mesh reconstruction pipeline



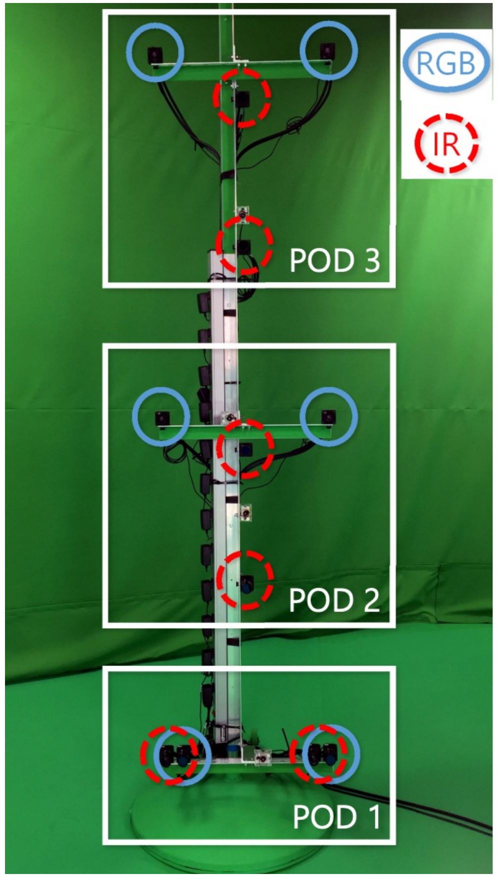
Surface-based representations

High-Quality Streamable Free-Viewpoint Video, *SIGGRAPH* 2015.

Surface-based representations

Capture system

53 RGB cameras and 53 IR cameras



High-Quality Streamable Free-Viewpoint Video, *SIGGRAPH* 2015.

Surface-based representations

Limitations:

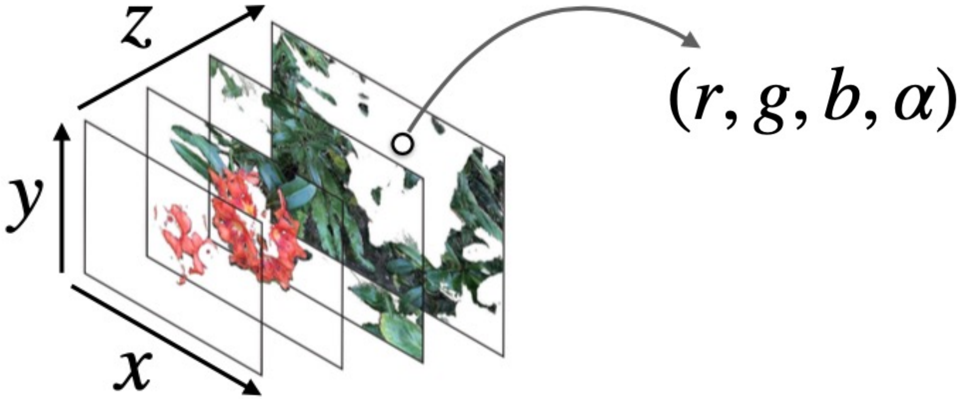
- High-quality mesh reconstruction is difficult in many cases
- Cannot represent very complex scenes

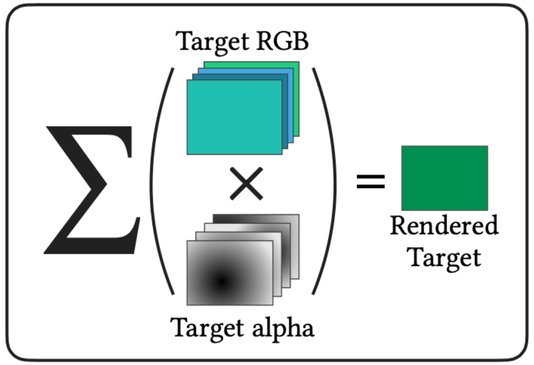
Volume-based representations

Multi-Plane Image (MPI)

A set of front-parallel planes at a fixed range of depths

Each plane encodes an RGB color image C_d and an alpha/transparency map α_d





Blend RGBA renderings together to render final output image

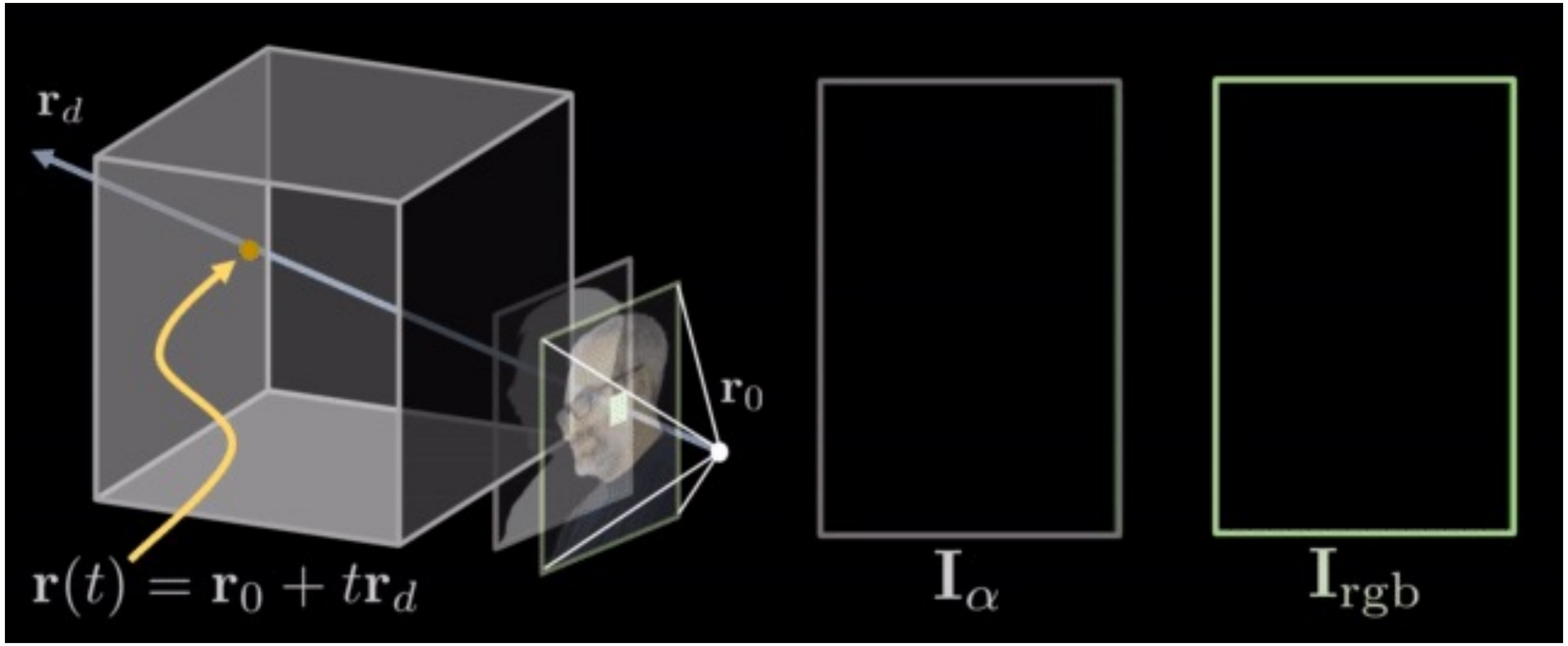
Volume-based representations

Multi-Plane Image (MPI)

DeepView: View synthesis with learned gradient descent. CVPR, 2019.

Volume-based representations

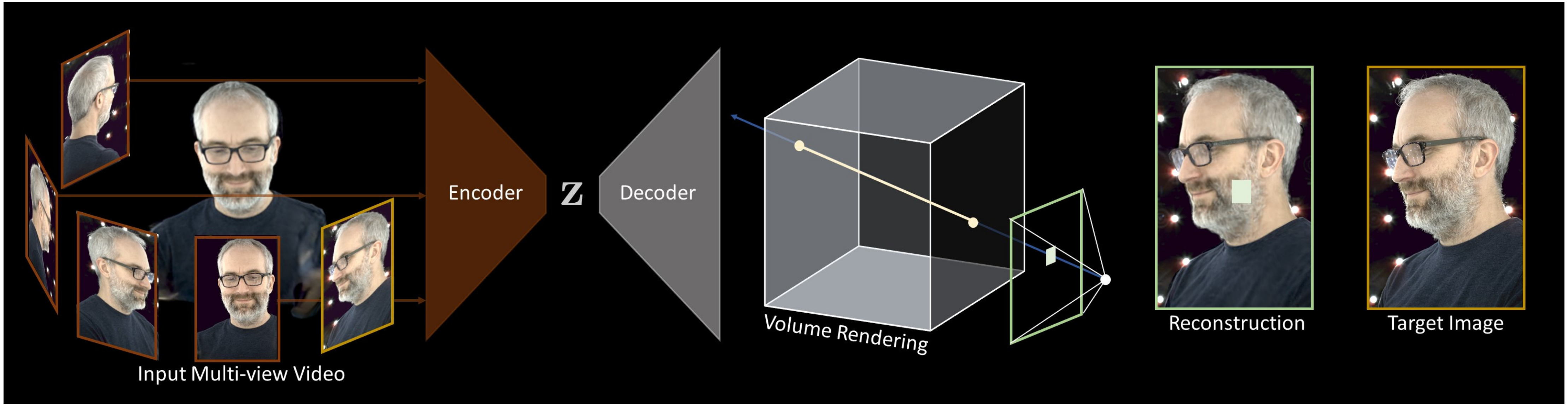
RGB-α volume



Neural Volumes: Learning Dynamic Renderable Volumes from Images, *SIGGRAPH* 2019.

Volume-based representations

Neural volumes: an encoder-decoder network that transforms input images into a 3D volume representation



Neural Volumes: Learning Dynamic Renderable Volumes from Images, *SIGGRAPH* 2019.

Volume-based representations

Neural volumes: an encoder-decoder network that transforms input images into a 3D volume representation

Neural Volumes: Learning Dynamic Renderable Volumes from Images, *SIGGRAPH* 2019.

Volume-based representations

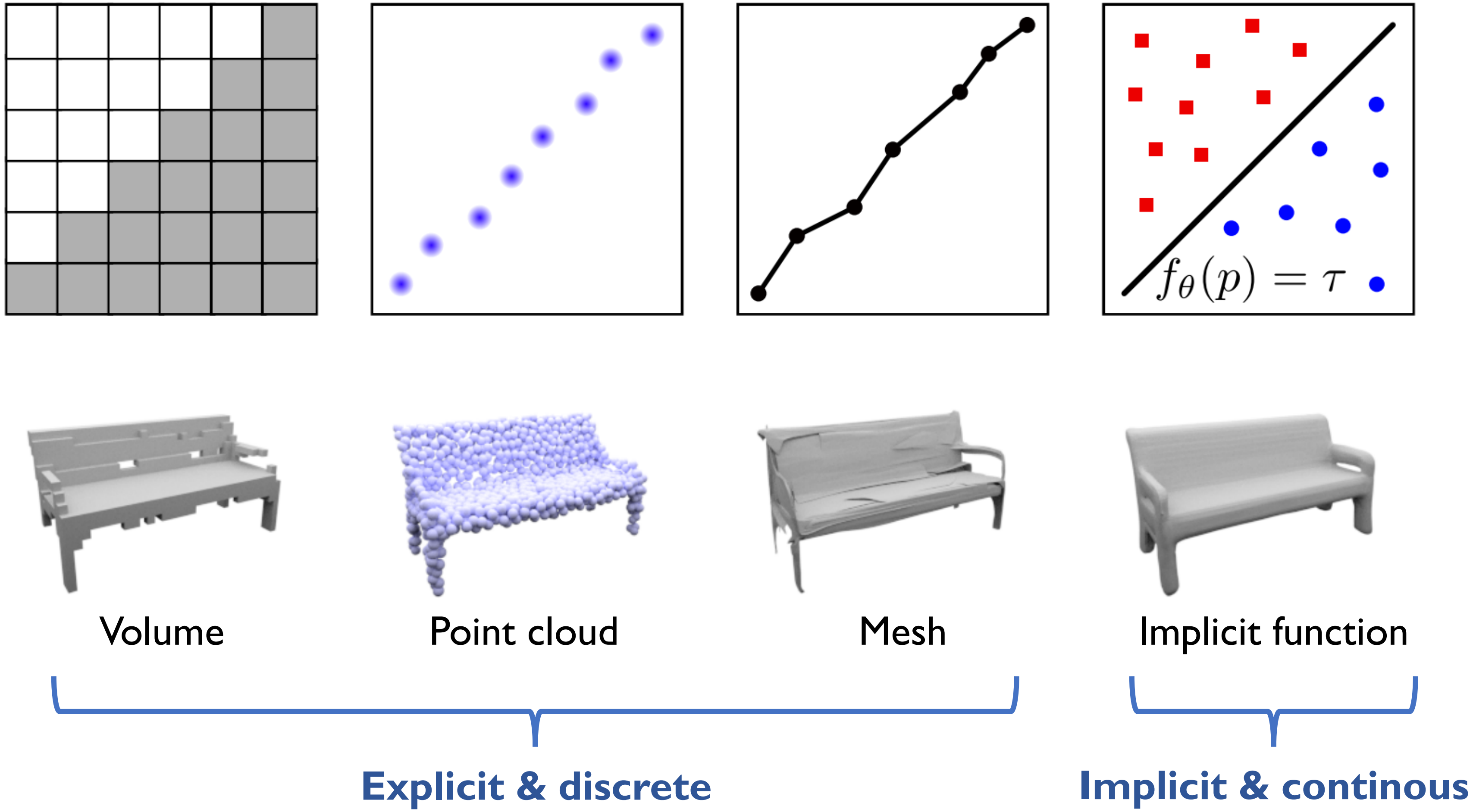
Advantages:

- Can represent very complex scenes
- Realistic reflections / specularity / transparency

Limitations:

- Discrete 3D volume requires large storage size for high-resolution rendering

Implicit Representations

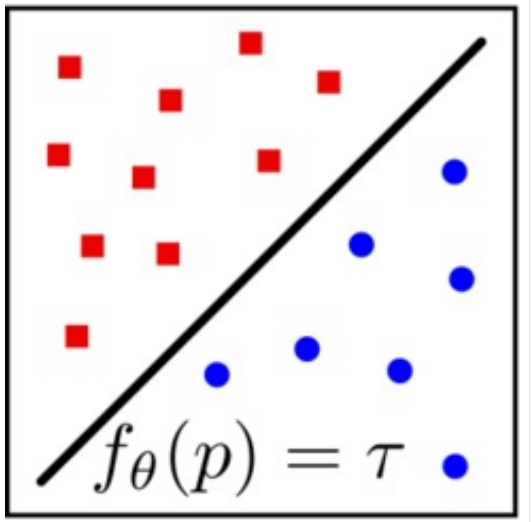


Occupancy Networks: Learning 3D Reconstruction in Function Space, *CVPR* 2019.

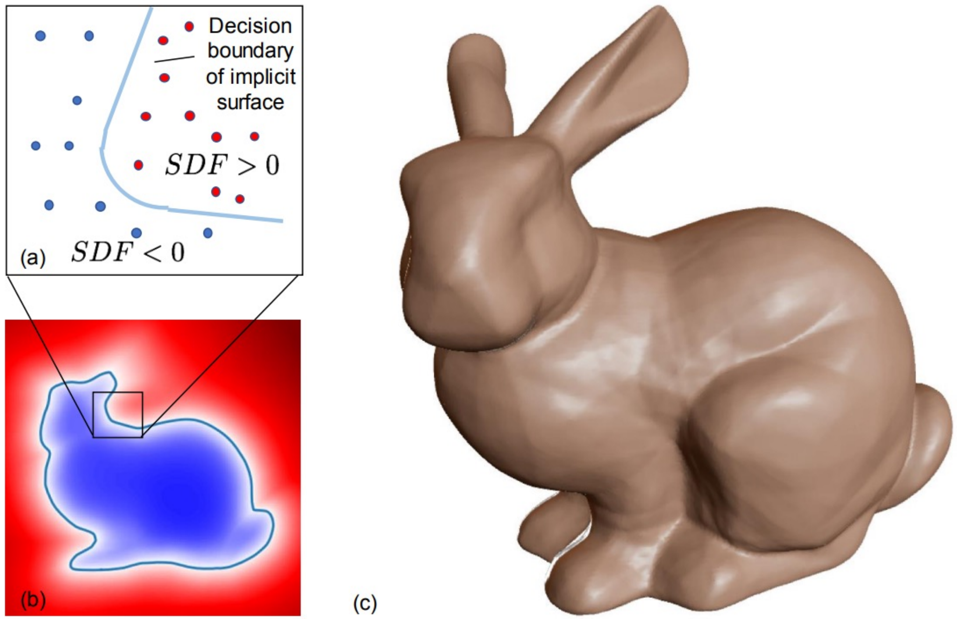
Implicit Representations

The implicit function can be:

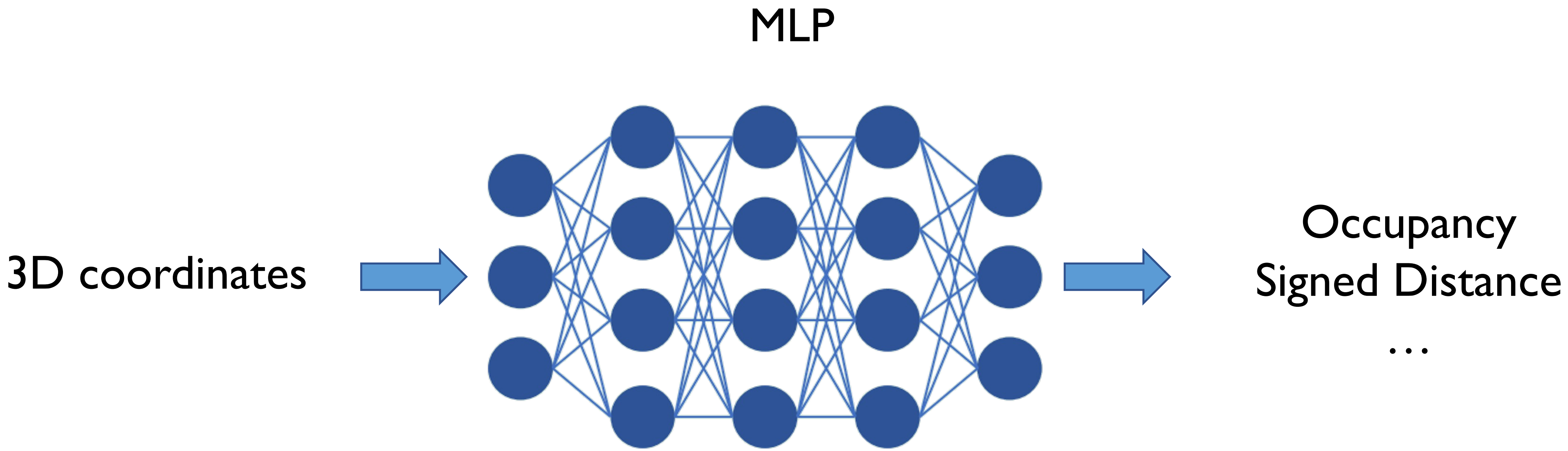
Occpupacy



Signed distance function (SDF)



Implicit Neural Representations



Occupancy Networks: Learning 3D Reconstruction in Function Space, *CVPR* 2019.

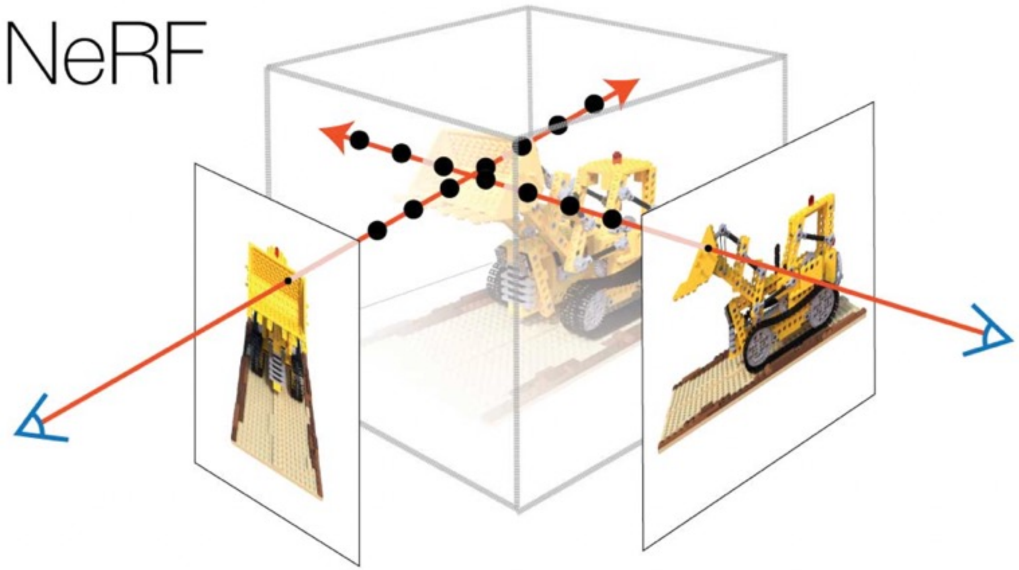
DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation, *CVPR* 2019.

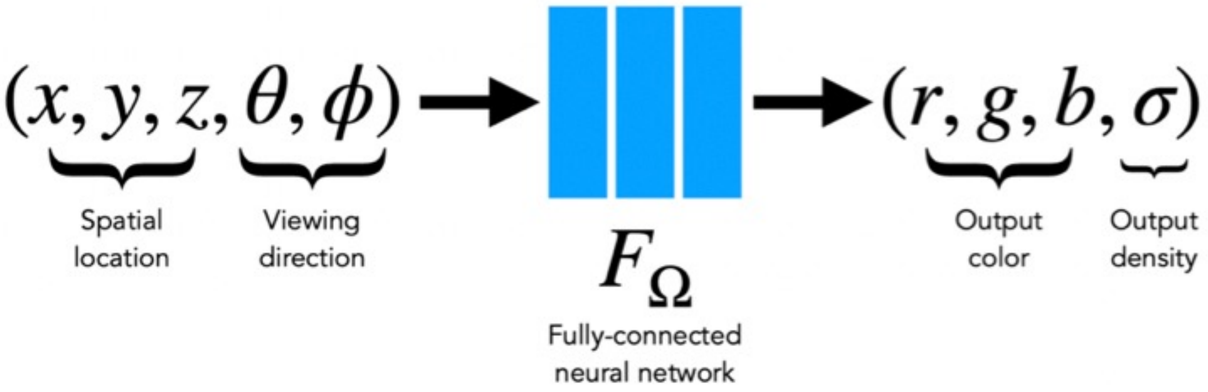
Learning implicit fields for generative shape modeling, *CVPR* 2019.

Scene Representation Networks: Continuous 3D-Structure-Aware Neural Scene Representations, NeurIPS 2019.

Neural Radiance Fields (NeRF)

Representing scenes as continuous density and color fields

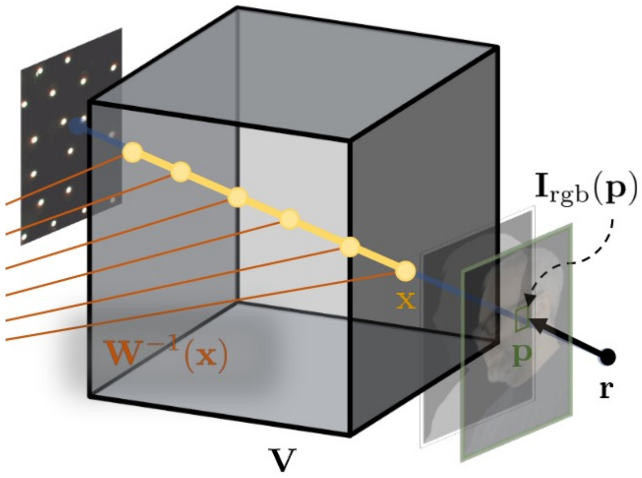




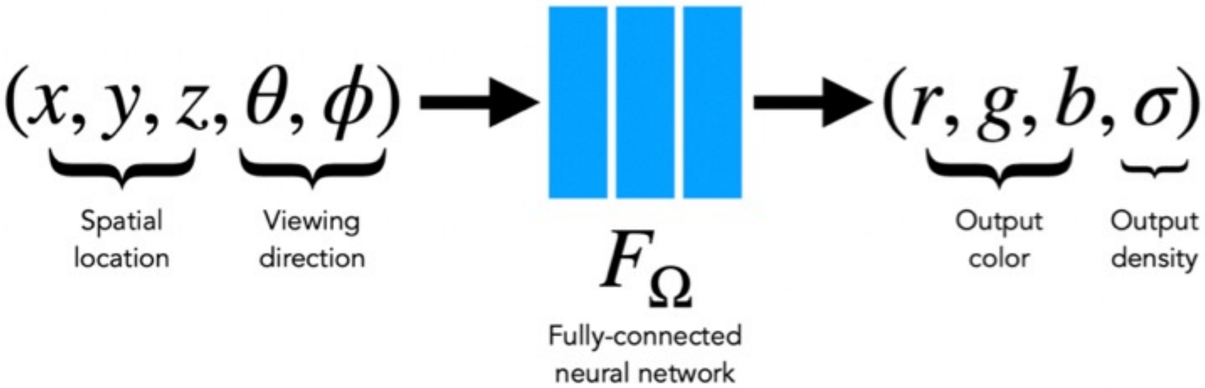
NeRF: Representing scenes as neural radiance fields for view synthesis, *ECCV* 2020.

Neural Radiance Fields (NeRF)

Representing scenes as continuous density and color fields



Discrete RGB-α volume



Continuous RGB-α field

Neural Volumes: Learning Dynamic Renderable Volumes from Images, *SIGGRAPH* 2019.

NeRF: Representing scenes as neural radiance fields for view synthesis, *ECCV* 2020.

Neural Radiance Fields (NeRF)

Volume rendering, which is differentiable

Rendering model for ray r(t) = o + td:

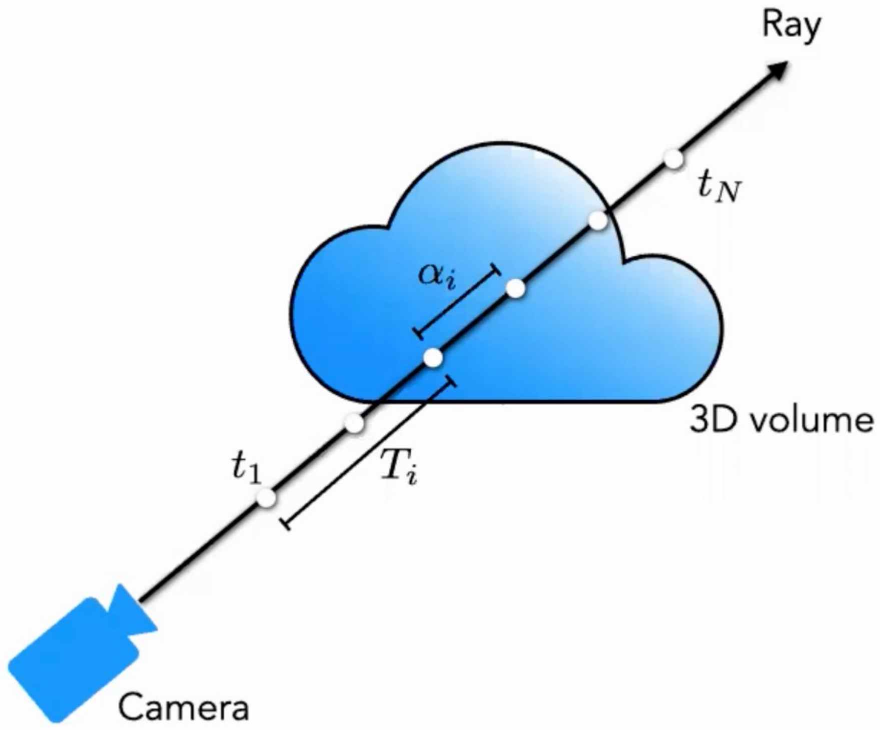
$$C \approx \sum_{i=1}^{N} T_i \alpha_i c_i$$

weights

colors

How much light is blocked earlier along ray:

$$T_i = \prod_{j=1}^{i-1} (1 - \alpha_j)$$



NeRF: Representing scenes as neural radiance fields for view synthesis, *ECCV* 2020.

Neural Radiance Fields (NeRF)

Learning NeRF from images

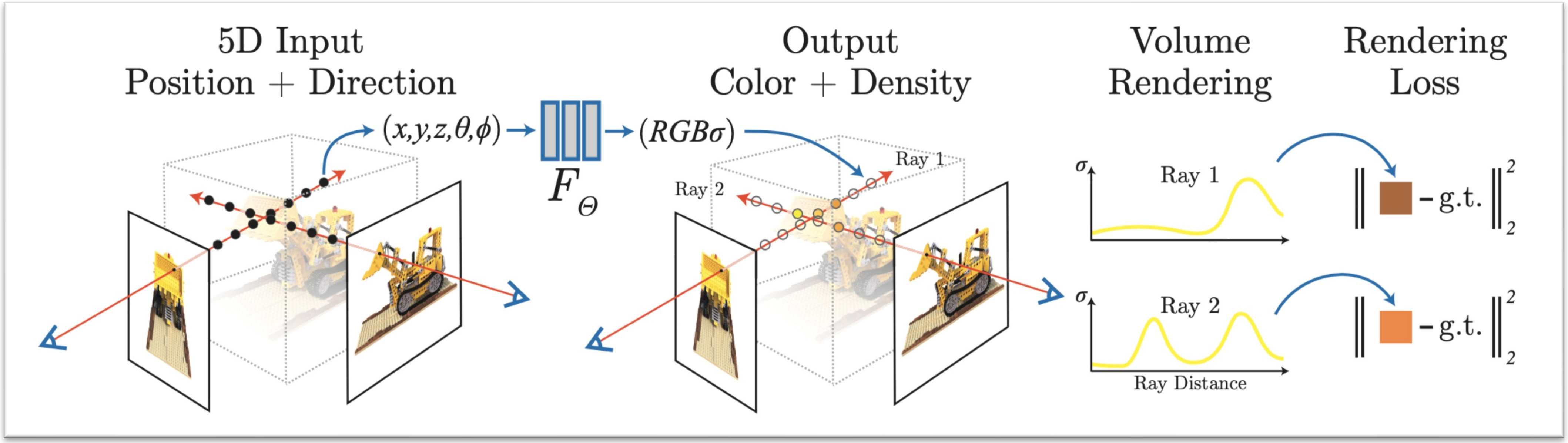
Input multi-view images

Optimizing NeRF

NeRF: Representing scenes as neural radiance fields for view synthesis, *ECCV* 2020.

Neural Radiance Fields (NeRF)

Learning NeRF from images



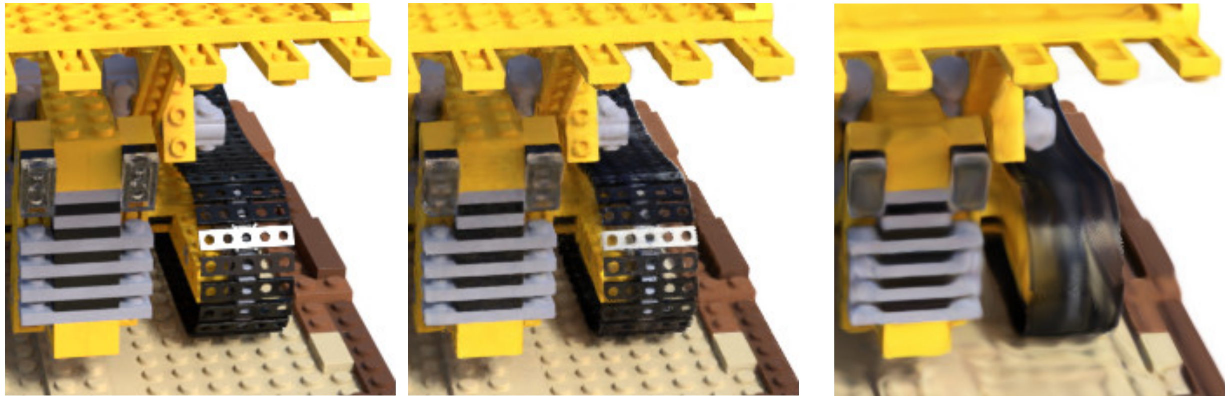
NeRF: Representing scenes as neural radiance fields for view synthesis, *ECCV* 2020.

Neural Radiance Fields (NeRF)

Positional encoding:

- Standard coordinate-based MLPs perform poorly at representing high frequency details
- Passing input coordinates through a high frequency mapping

$$\gamma(p) = \left(\sin\left(2^0 \pi p\right), \cos\left(2^0 \pi p\right), \cdots, \sin\left(2^{L-1} \pi p\right), \cos\left(2^{L-1} \pi p\right) \right)$$



Ground Truth　　Complete Model　　No Positional Encoding

NeRF: Representing scenes as neural radiance fields for view synthesis, *ECCV* 2020.

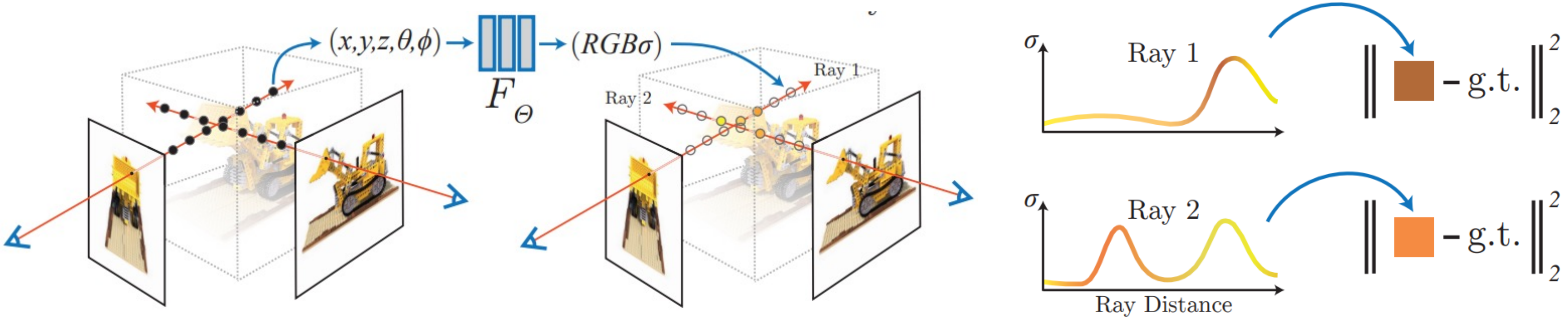
Neural Radiance Fields (NeRF)

NeRF: Representing scenes as neural radiance fields for view synthesis, *ECCV* 2020.

Neural Radiance Fields (NeRF)

Why better?

- The representation is continuous and flexible
- Optimizing rendering quality end-to-end



Neural Radiance Fields (NeRF)

Limitations:

- Computationally inefficient in terms of training and inference

  Optimizing a MLP network needs about 1 day

  Render one novel view needs 30 seconds
- Cannot model dynamic scenes
- Poor surface reconstruction quality

Neural Radiance Fields (NeRF)

Limitations

- Computationally inefficient in terms of training and inference
- Cannot model dynamic scenes
- Poor surface reconstruction quality

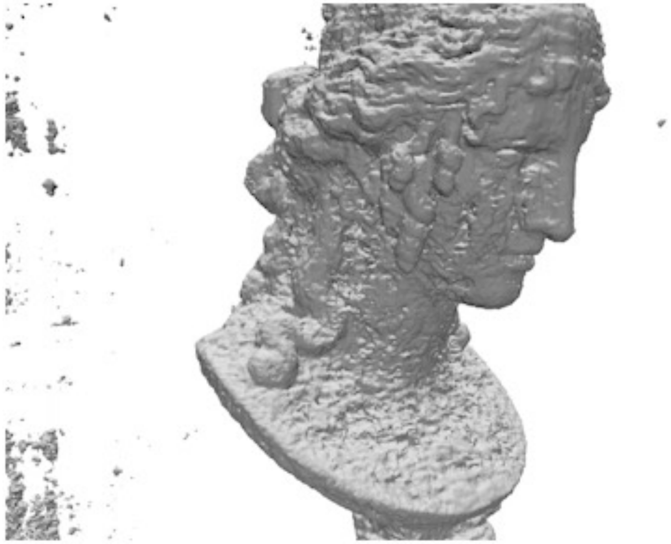
Neural Radiance Fields (NeRF)

Limitations

- Computationally inefficient in terms of training and inference
- Cannot model the motion of dynamic scenes
- Poor surface reconstruction quality

Reference image

COLMAP

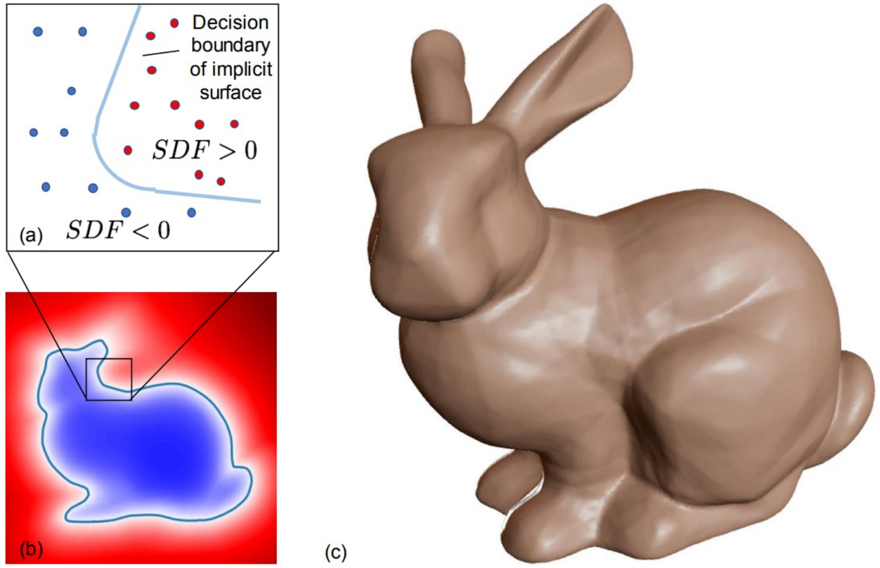


NeRF

Neural SDFs for Surface Reconstruction

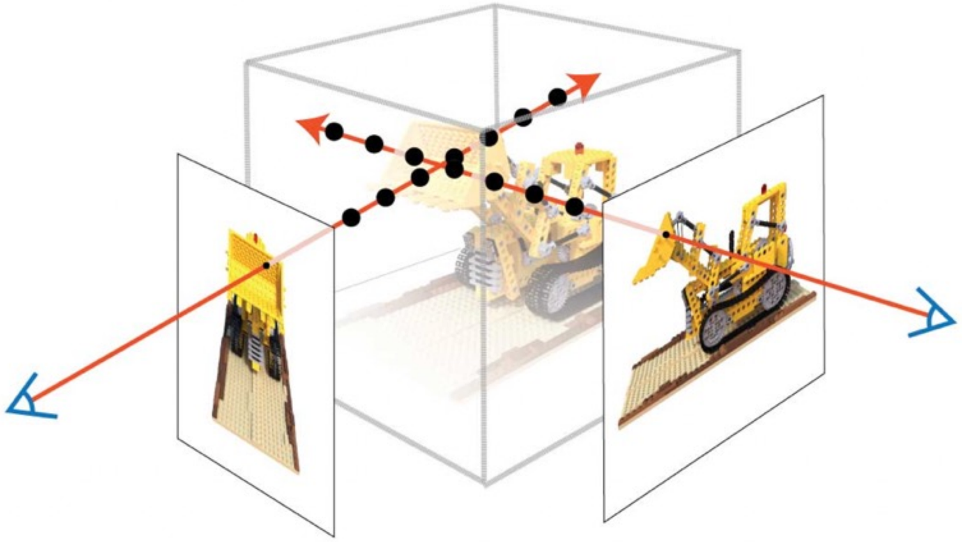
Surface reconstruction vs. Volumetric reconstruction

(x, y, z)



Signed distance

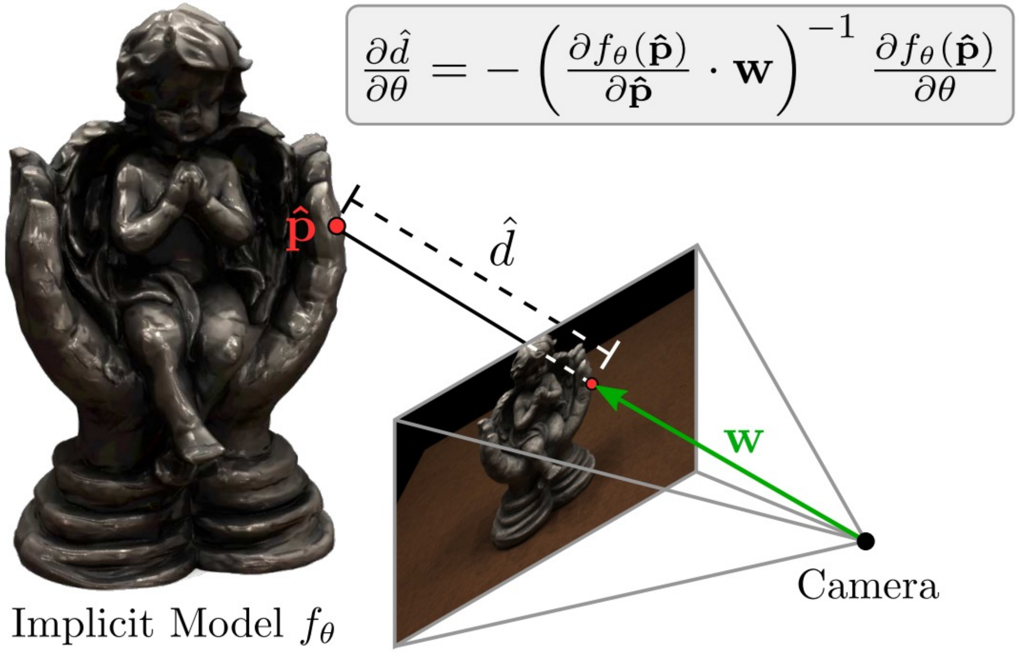
(x, y, z)



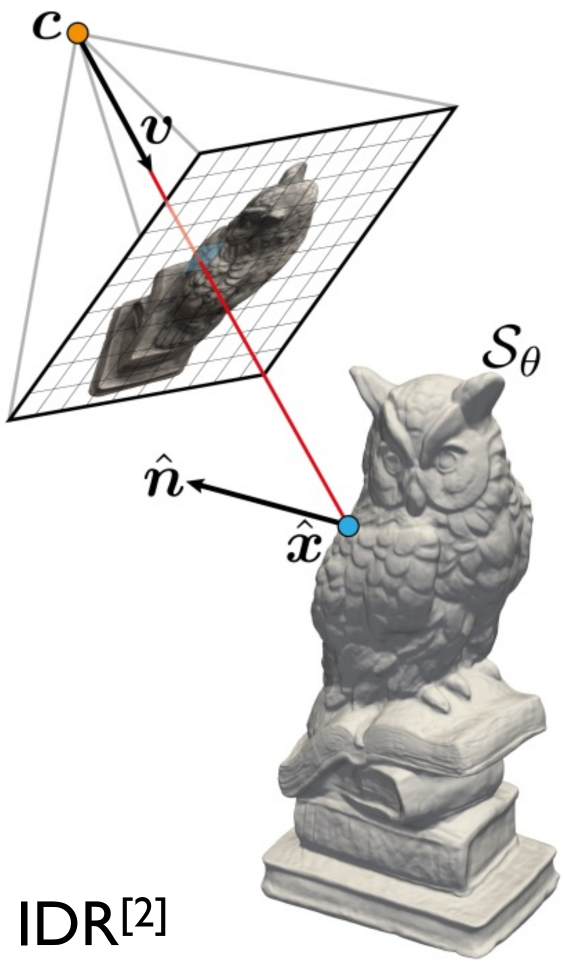
RGB + density

Neural SDFs for Surface Reconstruction

Differentiable surface rendering



DVR[1]

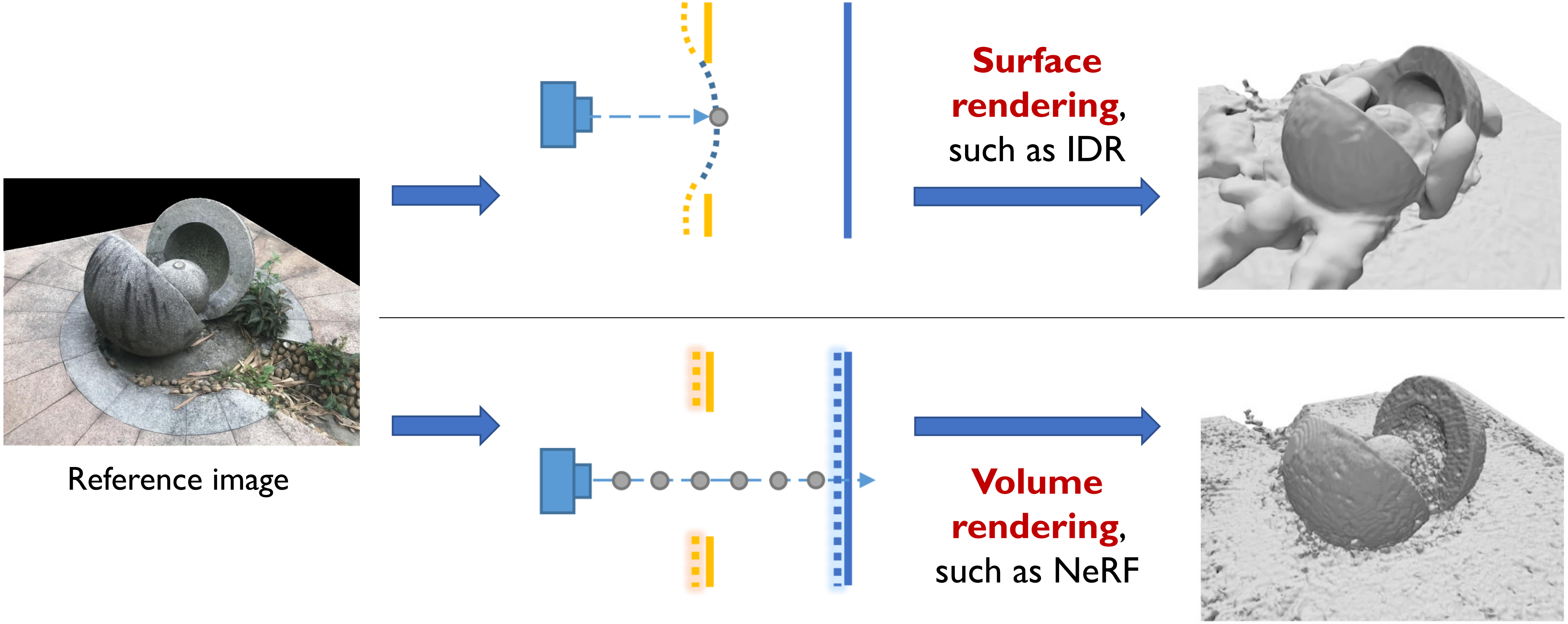


IDR[2]

[1] Differentiable Volumetric Rendering: Learning Implicit 3D Representations without 3D Supervision, *CVPR* 2020.

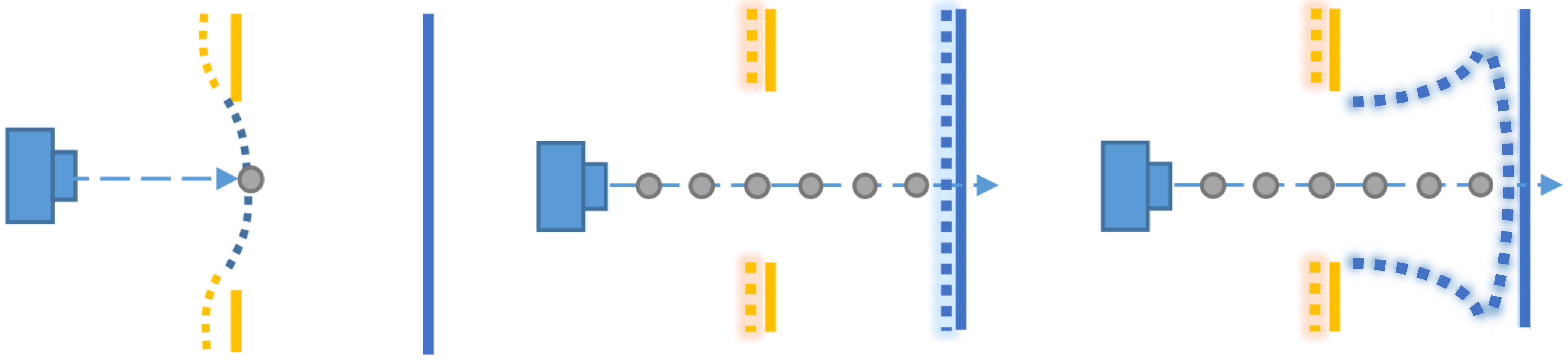
[2] Multiview Neural Surface Reconstruction by Disentangling Geometry and Appearance, *NeurIPS* 2020.

Limitation of surface rendering



NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction, *NeurIPS* 2021.

NeuS



Surface representation + Surface rendering

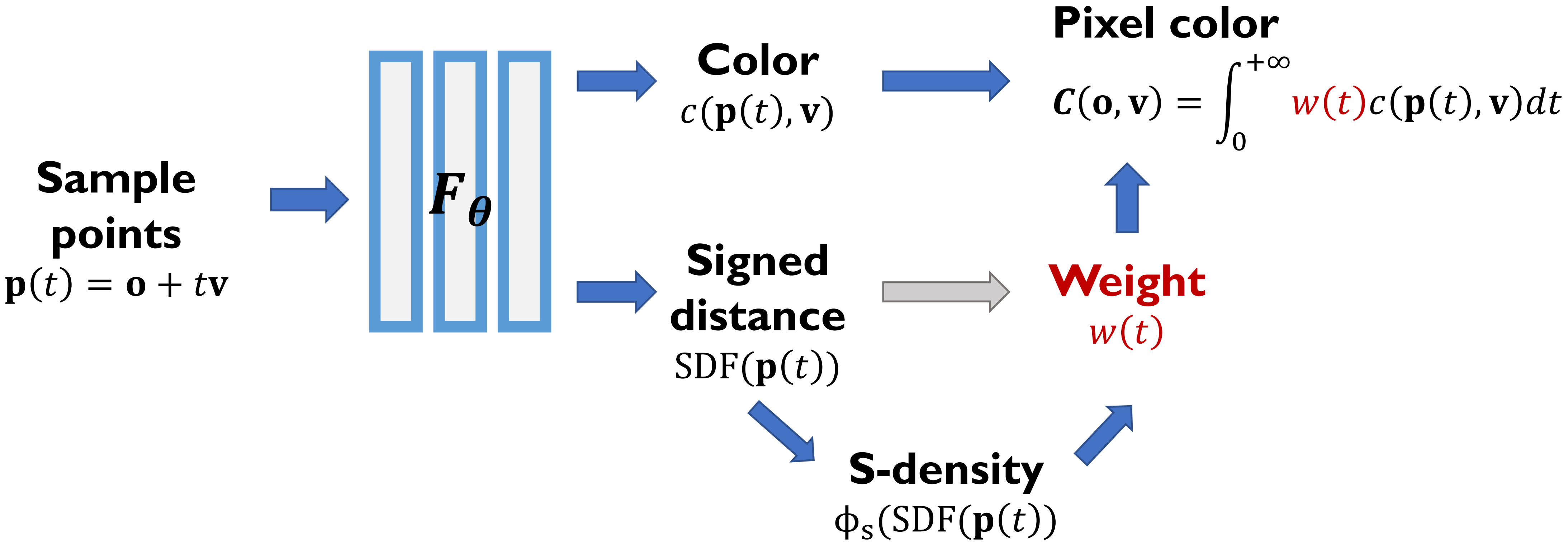
Volume representation + Volume rendering

Surface representation + Volume rendering

NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction, *NeurIPS* 2021.

NeuS

Optimizing SDF in a volumetric rendering framework



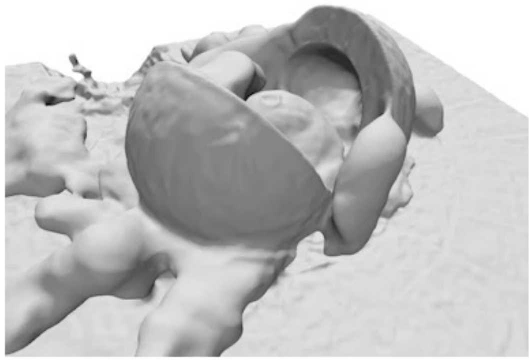
NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction, *NeurIPS* 2021.

NeuS

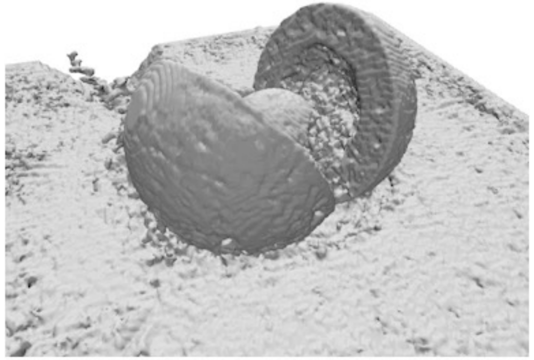
Advantages:

- Accurate 3D implicit surface reconstruction
- No need for depth or mask supervision

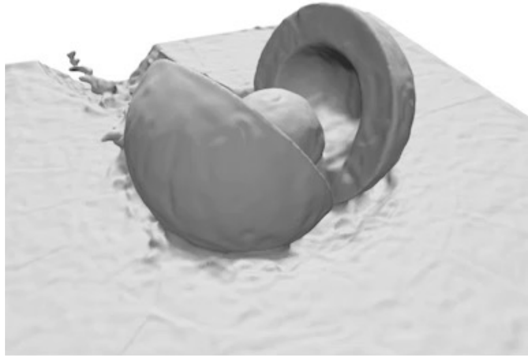
Reference image



IDR



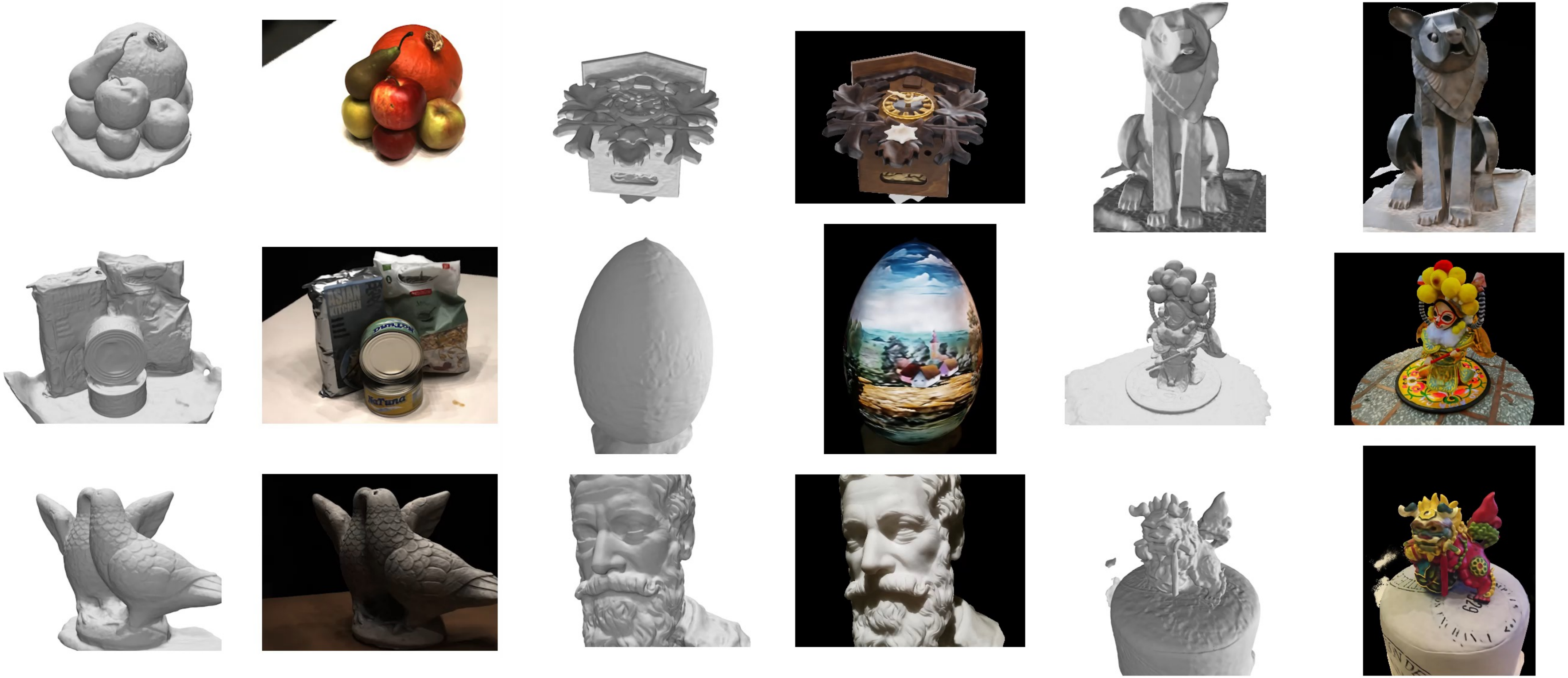
NeRF



NeuS

NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction, *NeurIPS* 2021.

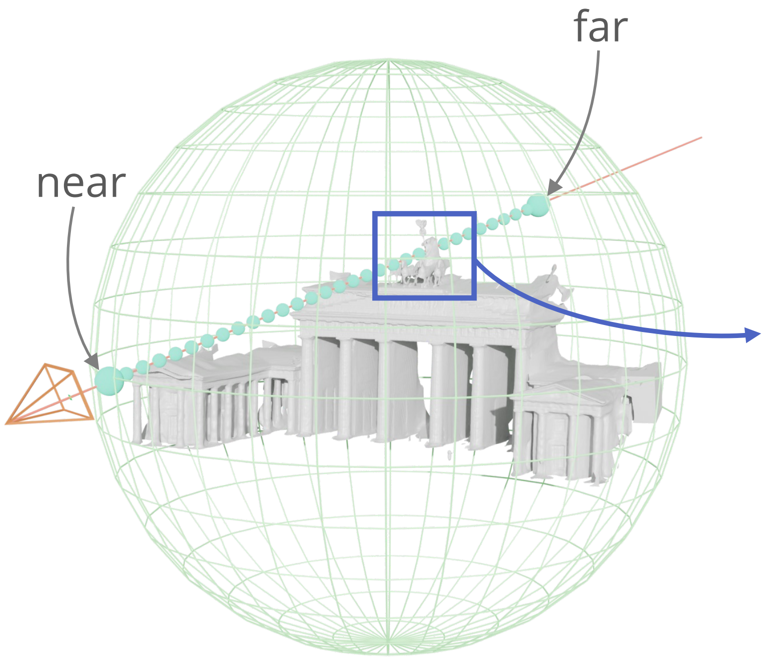
NeuS



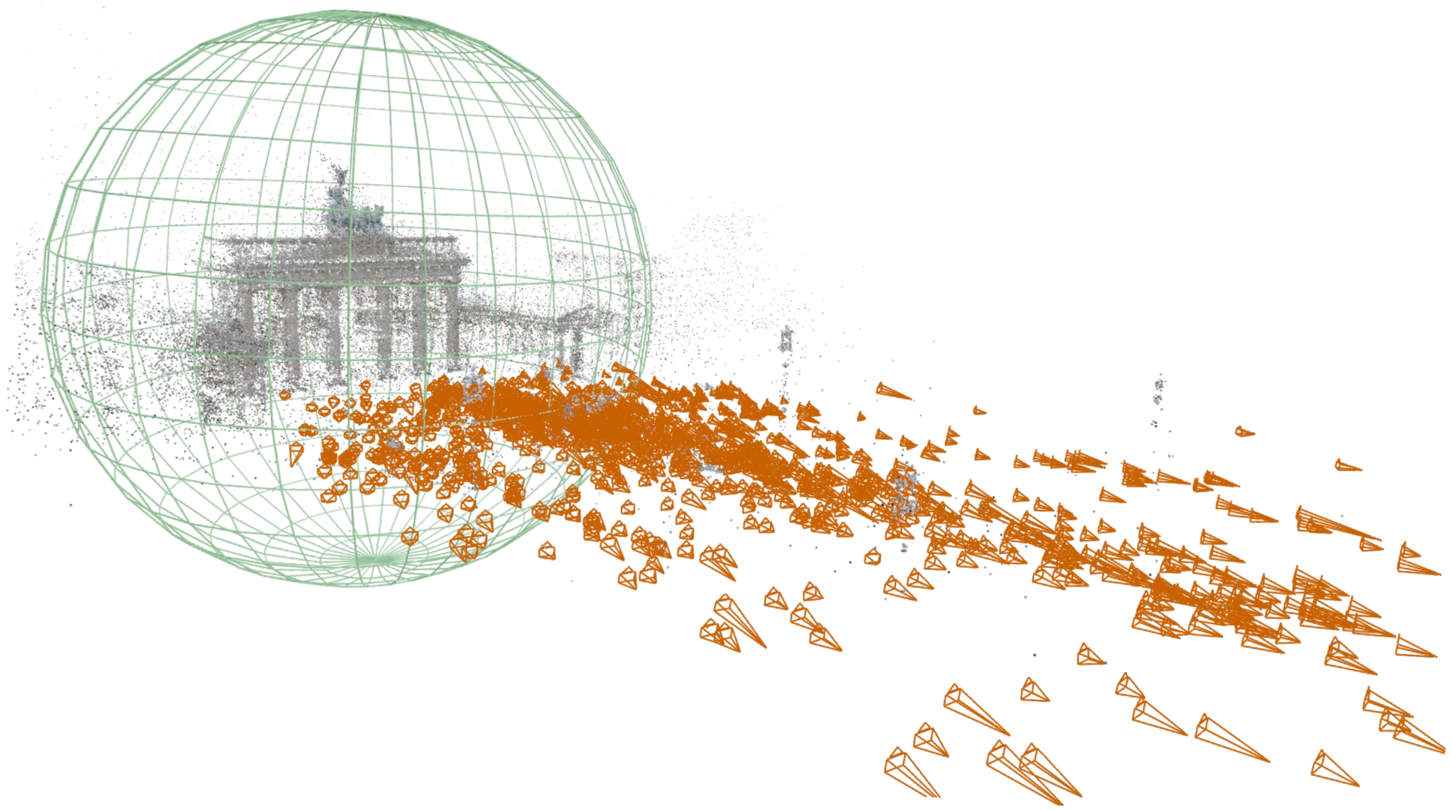
NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction, *NeurIPS* 2021.

Large-scale scene reconstruction

Challenge: large-scale scene with thousands of images

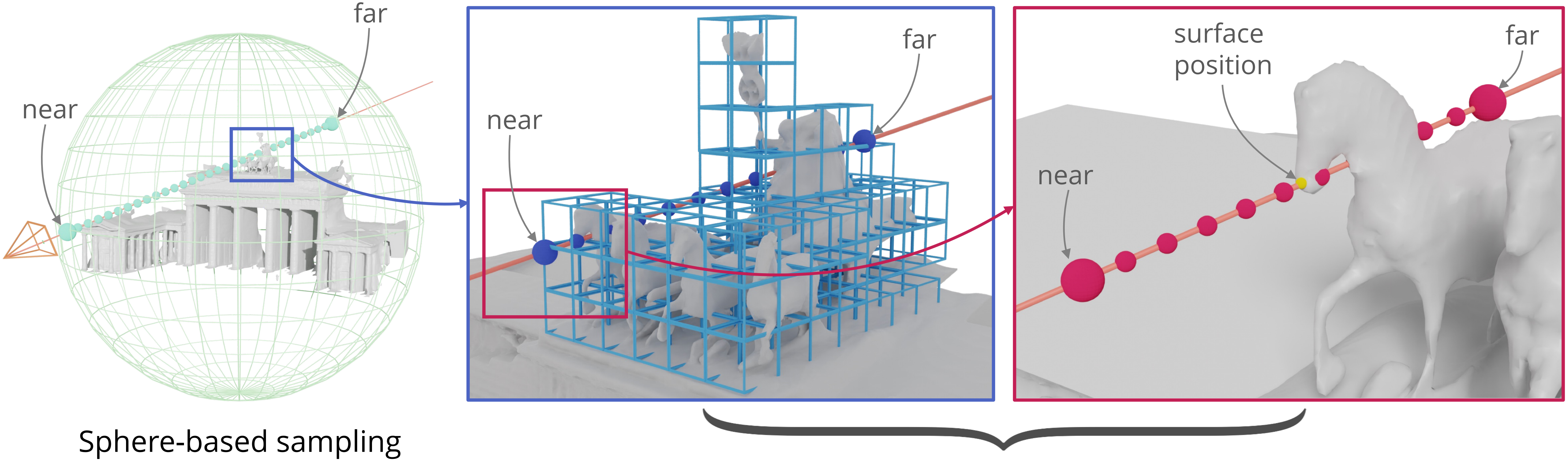


Sphere-based sampling



Large-scale scene reconstruction

Improving sampling efficiency by surface-guided sampling



Sphere-based sampling

Hybrid sampling

Neural Reconstruction in the Wild, *SIGGRAPH* 2022.

Large-scale scene reconstruction

Credits: Flickr

Neural Reconstruction in the Wild, *SIGGRAPH* 2022.

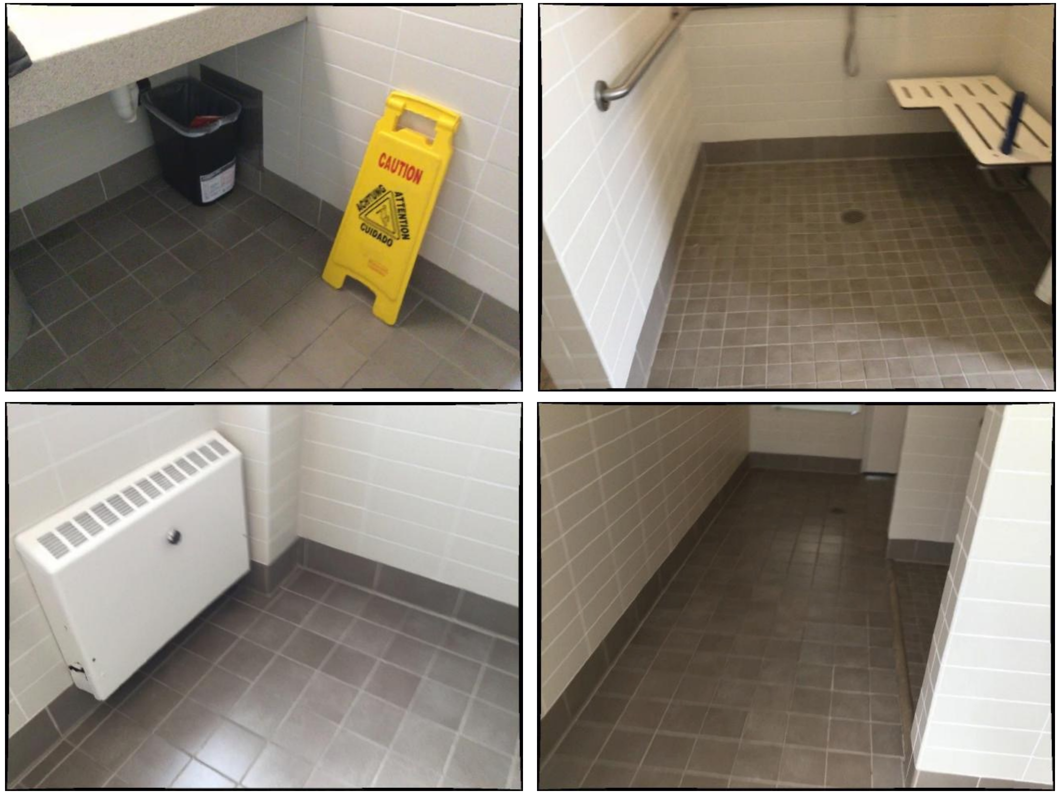
Large-scale scene reconstruction

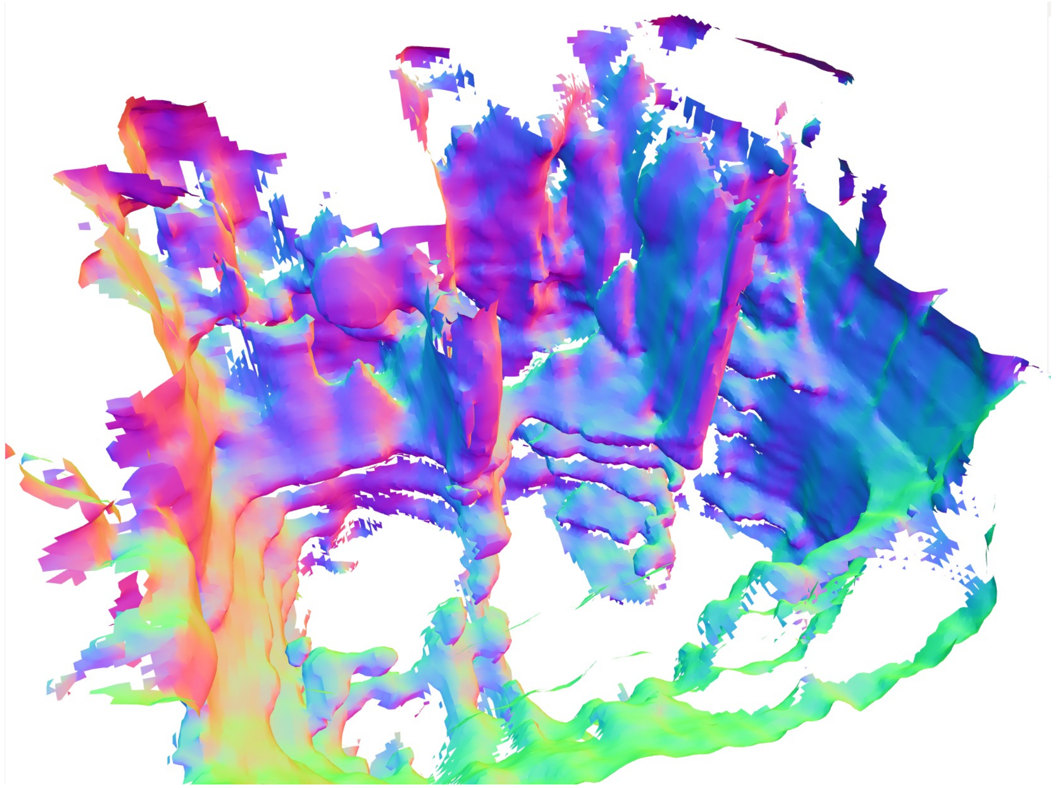
Credits: Flickr

Neural Reconstruction in the Wild, *SIGGRAPH* 2022.

Indoor scene reconstruction

Challenge: texture-low regions

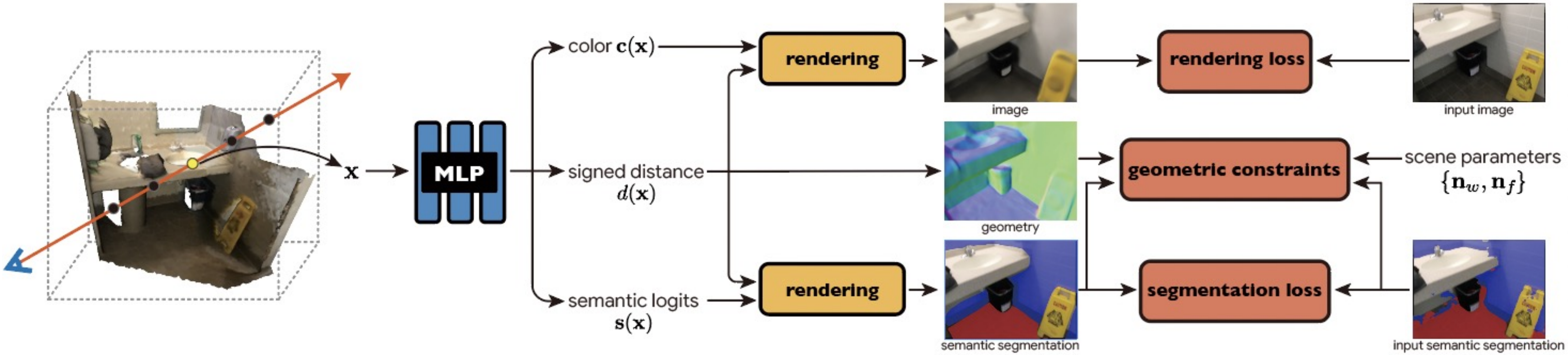




Indoor scene reconstruction

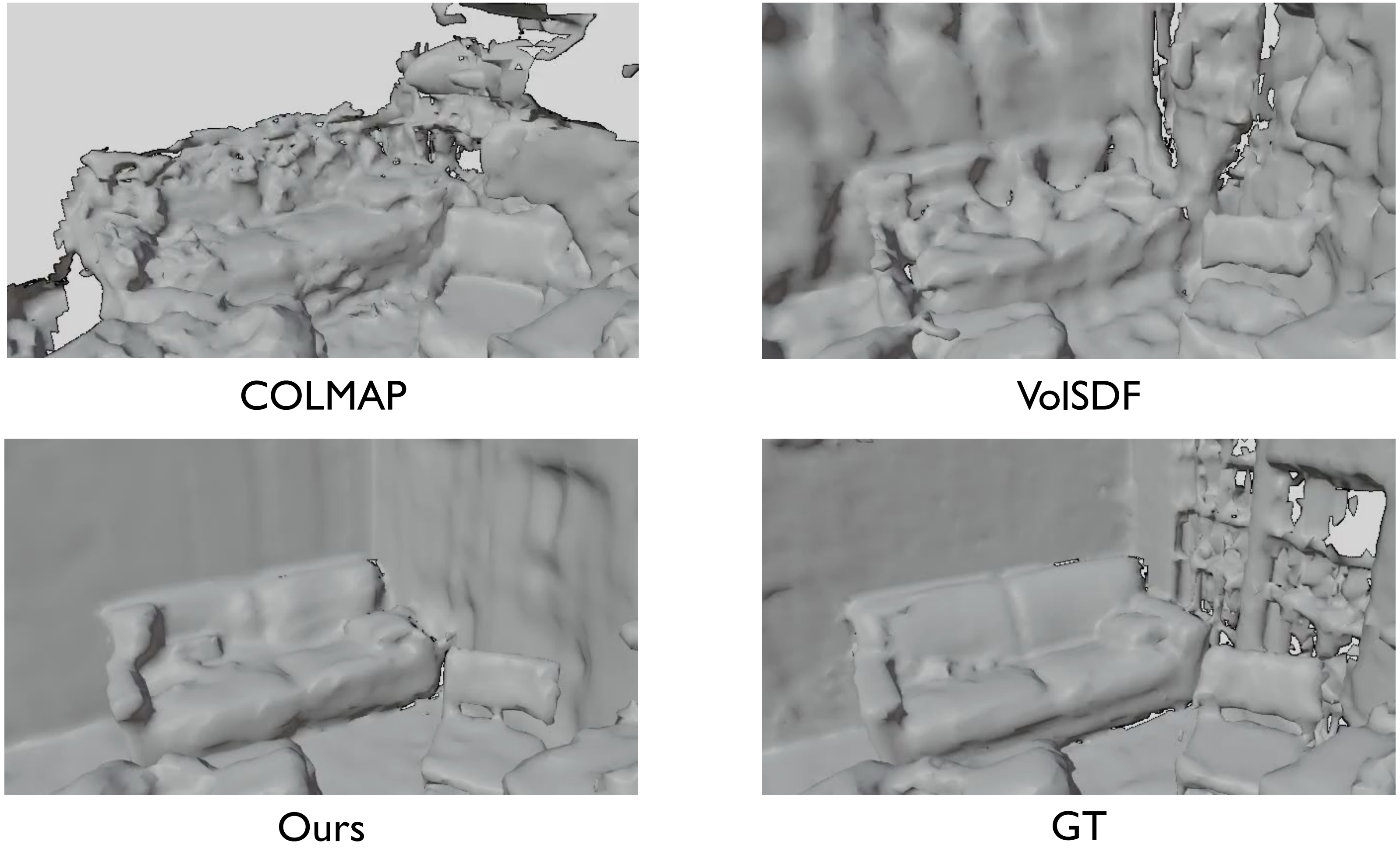
Manhattan-world assumption

Can be easily integrated when optimizing implicit neural representations



Neural 3D Scene Reconstruction with the Manhattan-world Assumption, *CVPR* 2022.

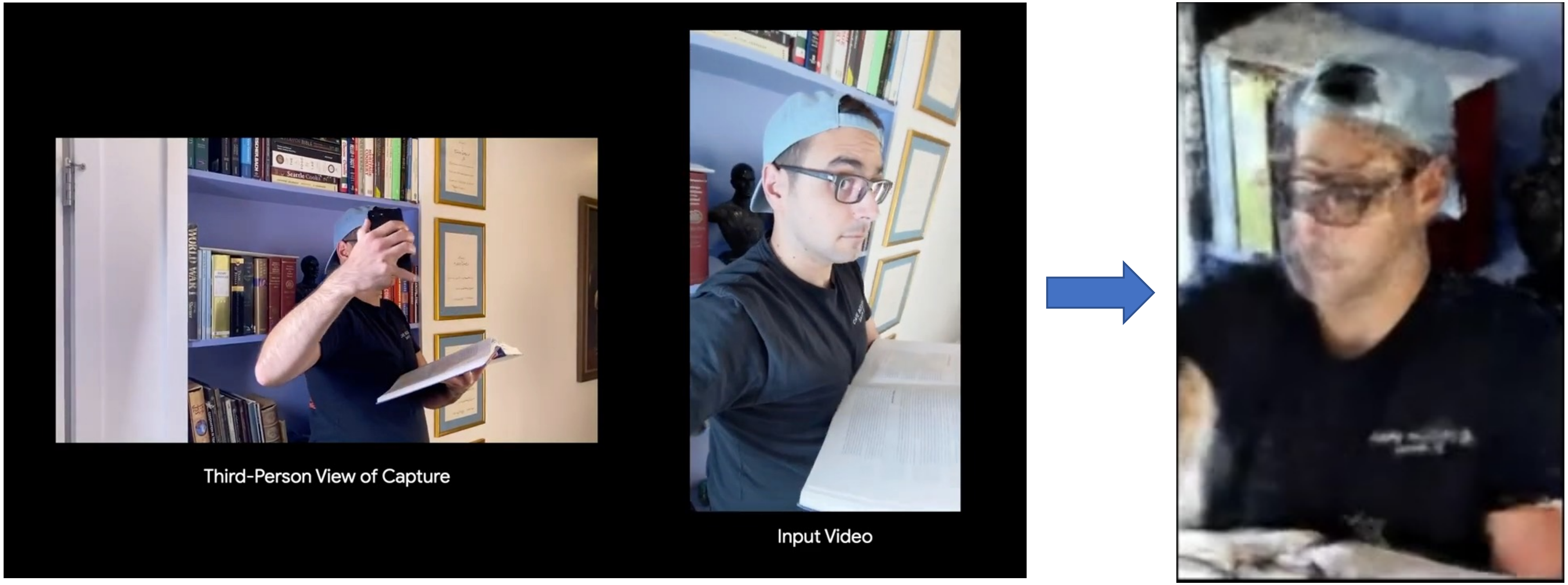
Indoor scene reconstruction



Neural 3D Scene Reconstruction with the Manhattan-world Assumption, *CVPR* 2022.
Volume Rendering of Neural Implicit Surfaces, *NeurIPS* 2021.

Neural dynamic scene representations

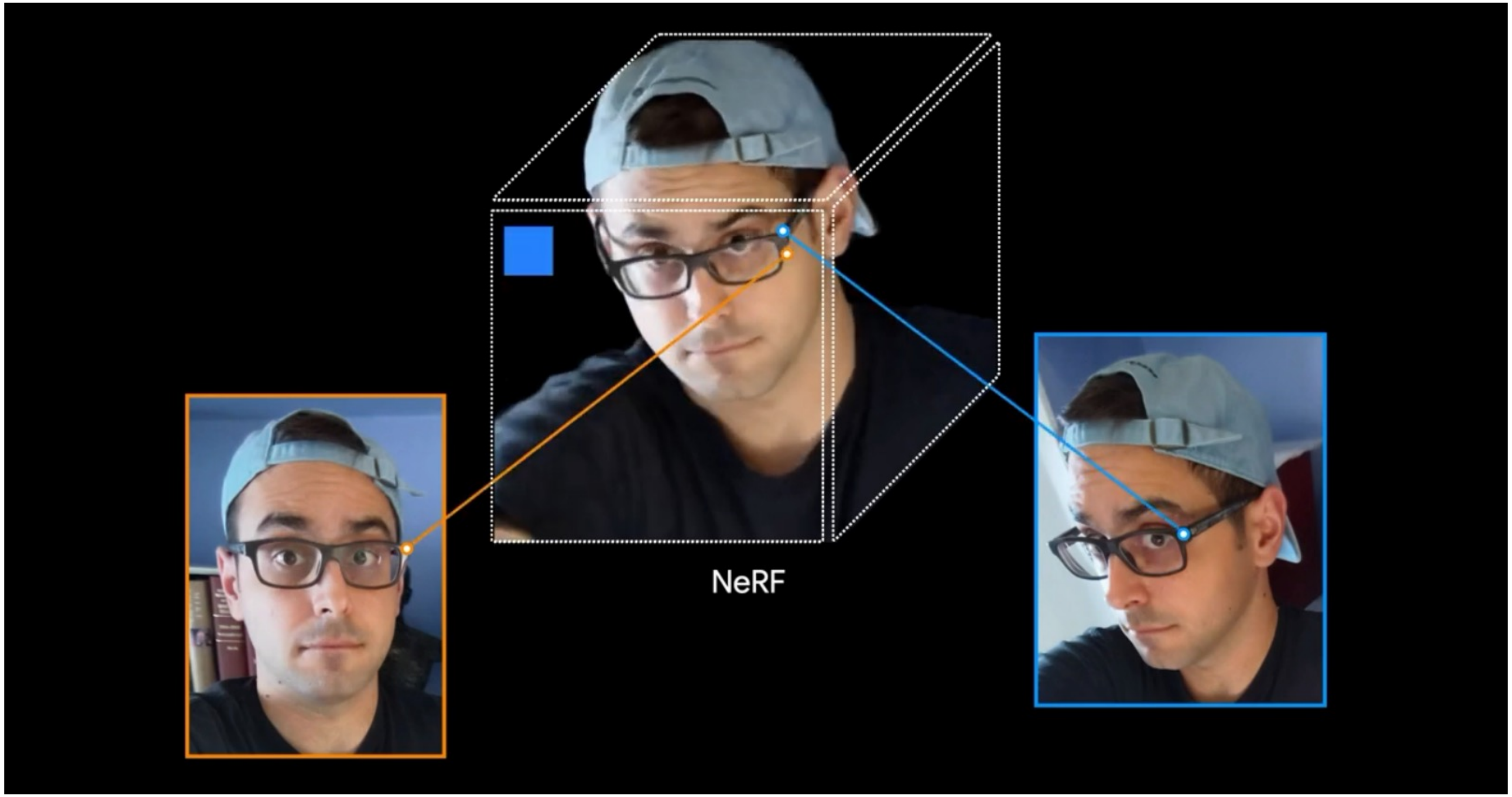
NeRF cannot model dynamic scenes



Nerfies: Deformable neural radiance fields, *ICCV* 2021.

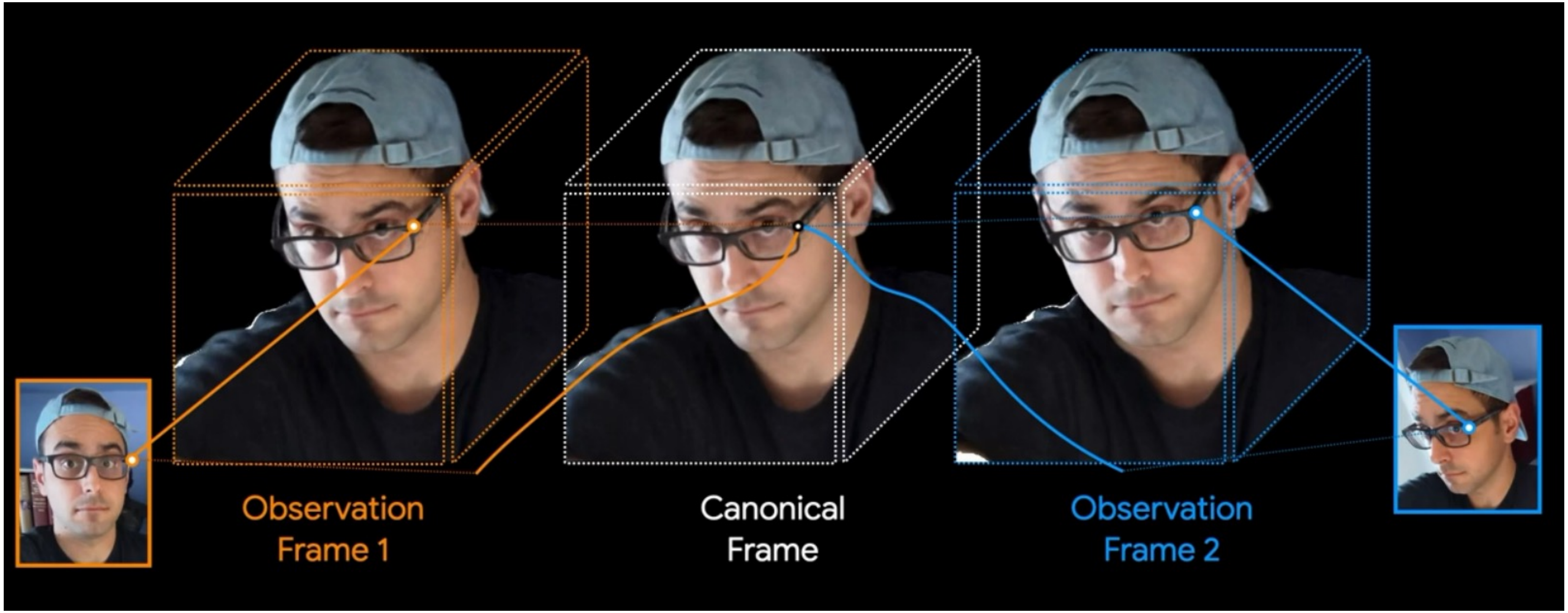
Neural dynamic scene representations

Problem: scene movements cause the rays of different frames of the same observed point do not intersect



General dynamic scenes – Deformable NeRF

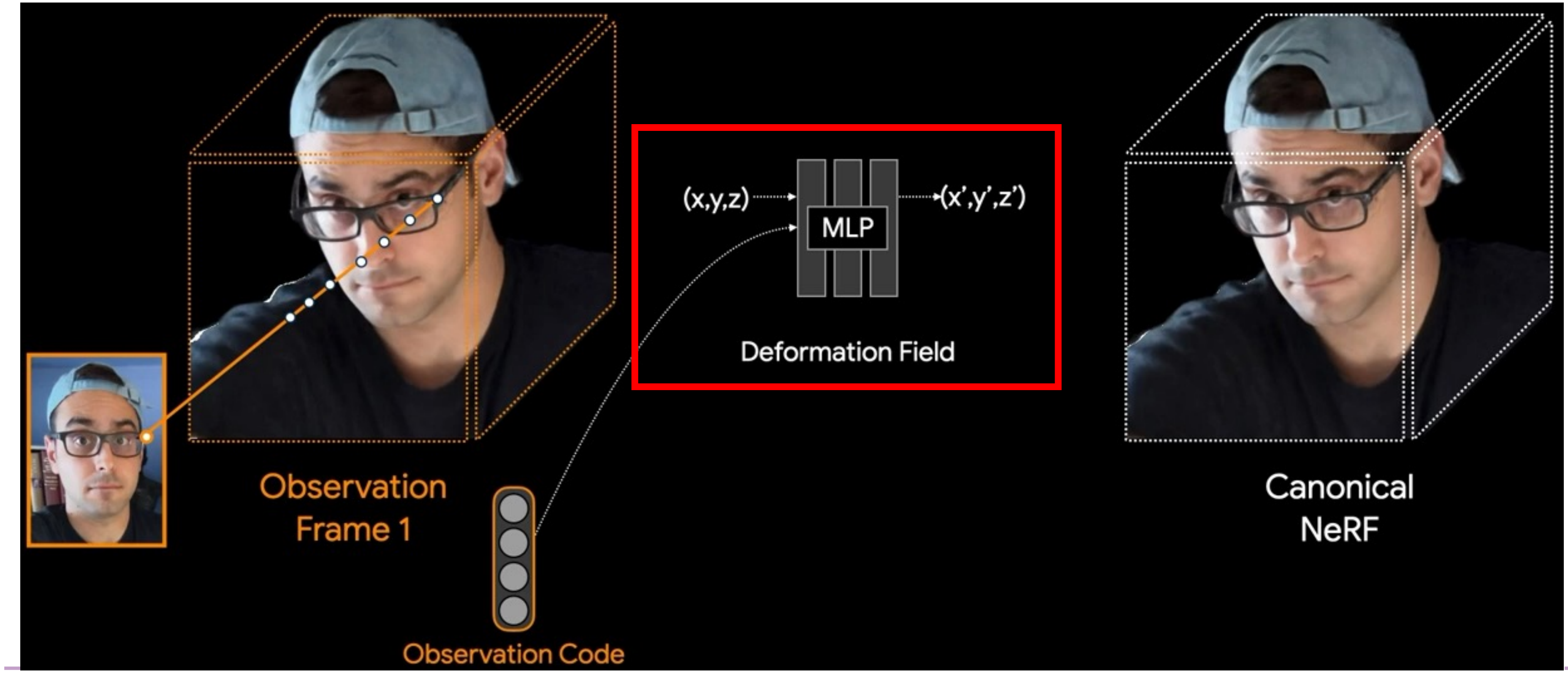
Deformable NeRF: a canonical NeRF + deformation fields



Nerfies: Deformable neural radiance fields, *ICCV* 2021.

General dynamic scenes – Deformable NeRF

The deformation field from other frames to canonical frame is learned by another MLP

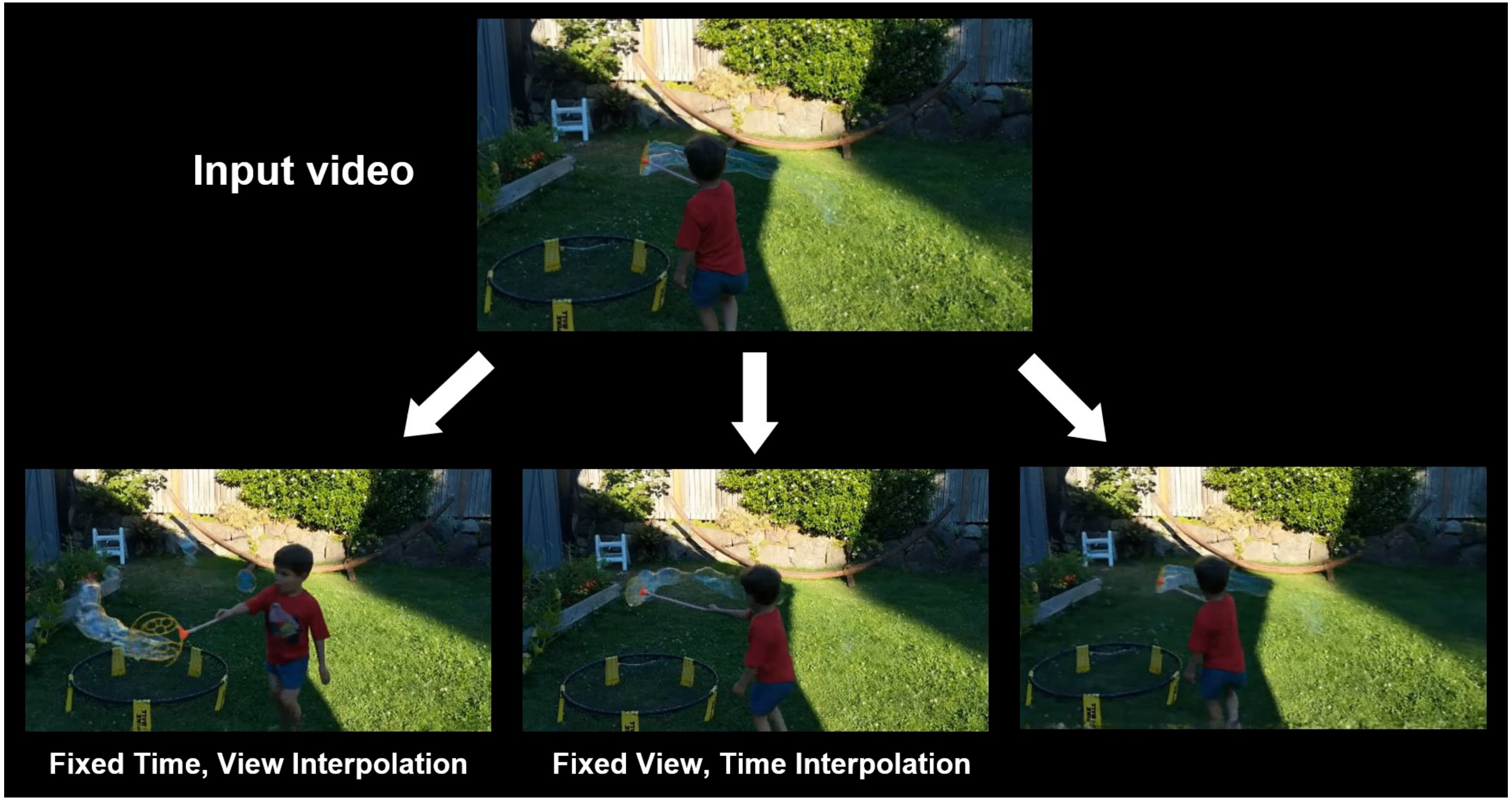


Nerfies: Deformable neural radiance fields, *ICCV* 2021.

General dynamic scenes – Deformable NeRF

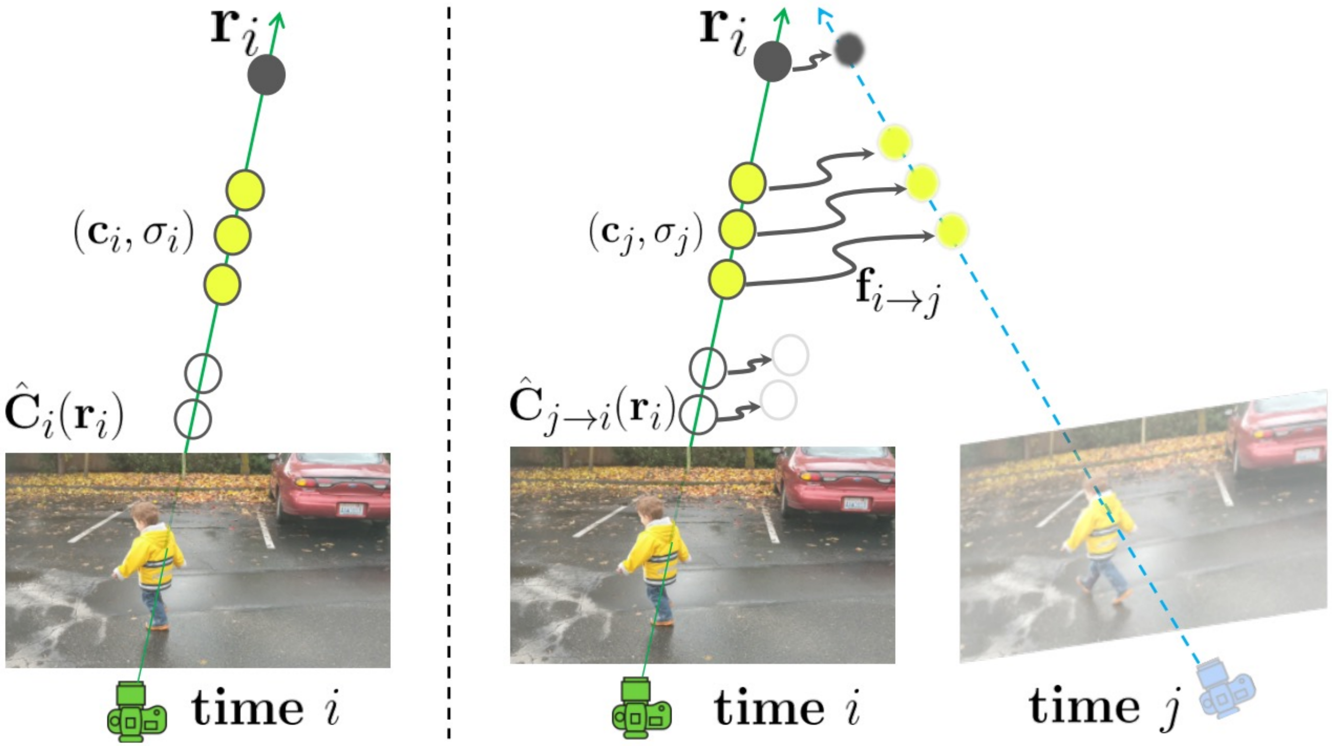
Nerfies: Deformable neural radiance fields, *ICCV* 2021.

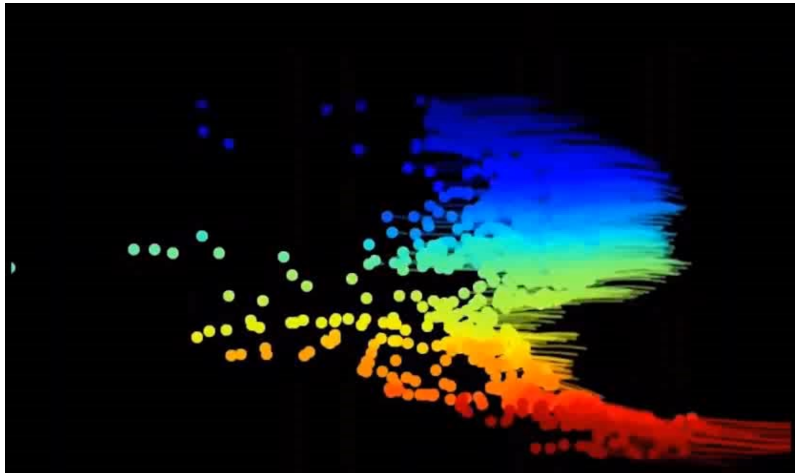
General dynamic scenes – NSFF



NSFF: Neural Scene Flow Fields for Space-Time View Synthesis of Dynamic Scenes, *CVPR* 2021

General dynamic scenes – NSFF





3D scene flow visualization

NSFF: Neural Scene Flow Fields for Space-Time View Synthesis of Dynamic Scenes, *CVPR* 2021

General dynamic scenes

Advantages:

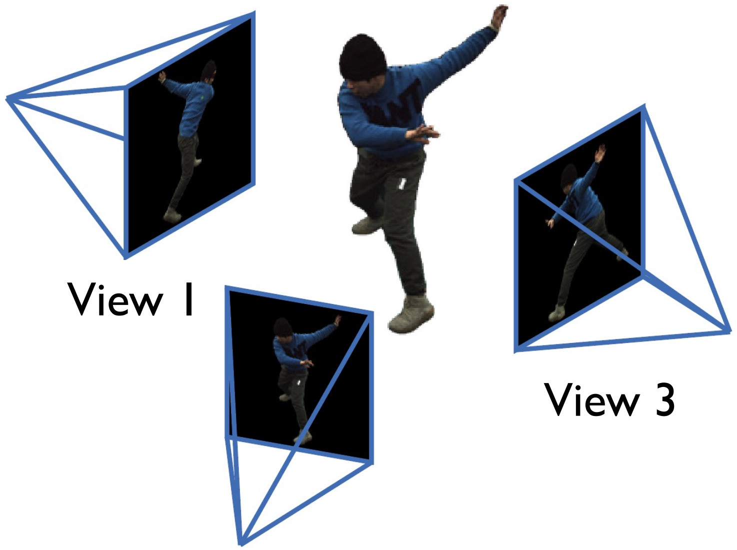
- Can model general objects and scenes, not restricted to human

Limitations:

- Need to optimize canonical NeRF and motion field simultaneously, which is prone to local optima
- It is very hard to recover large and long-range motion, e.g. fast moving human bodies

Dynamic humans – Neural body

Reconstructing dynamic human from sparse views



Neural Body: Implicit Neural Representations with Structured Latent Codes for Novel View Synthesis of Dynamic Humans, *CVPR* 2021.

Dynamic humans – Neural body

Reconstruction from sparse views is ill-posed

4 input views

NeRF reconstruction

Dynamic humans – Neural body

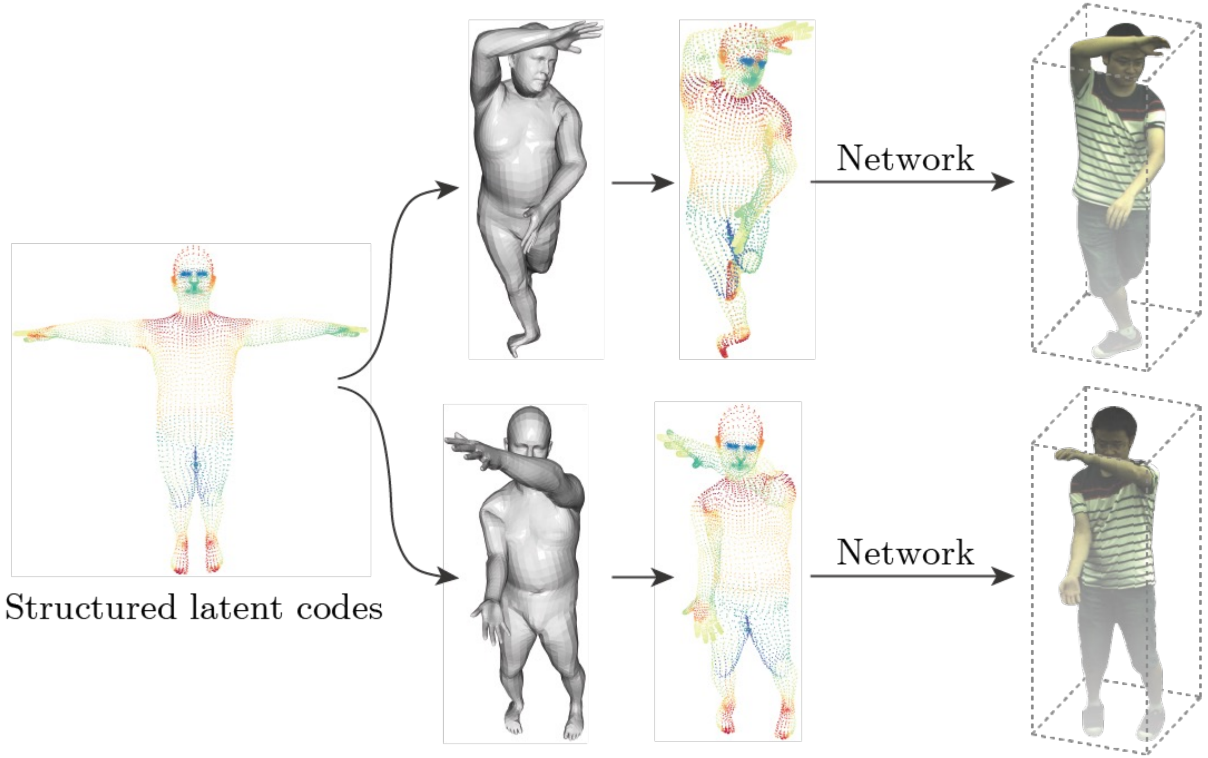
Integrating observations from multiple frames

4 input views

Our reconstruction

Dynamic humans – Neural body

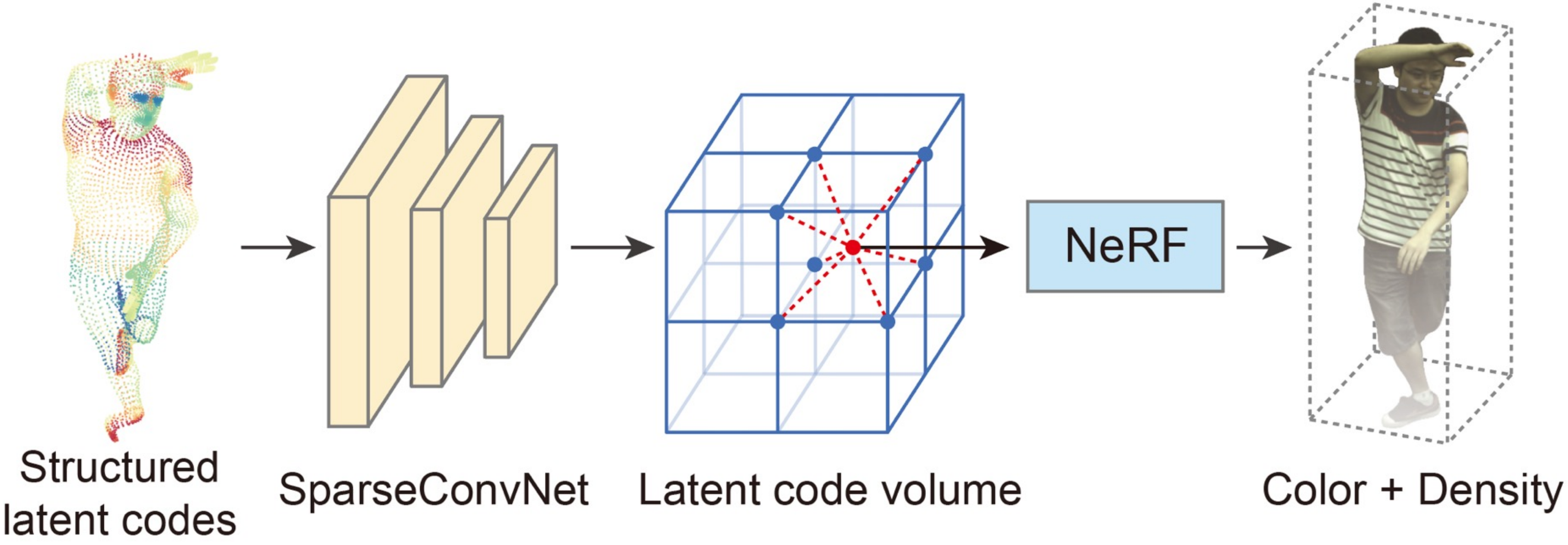
Assume NeRFs at different frames are decoded from the same set of latent codes, whose locations are pose dependent



Neural Body: Implicit Neural Representations with Structured Latent Codes for Novel View Synthesis of Dynamic Humans, *CVPR* 2021.

Dynamic humans – Neural body

The latent codes are decoded into NeRF by sparse convolution



Neural Body: Implicit Neural Representations with Structured Latent Codes for Novel View Synthesis of Dynamic Humans, *CVPR* 2021.

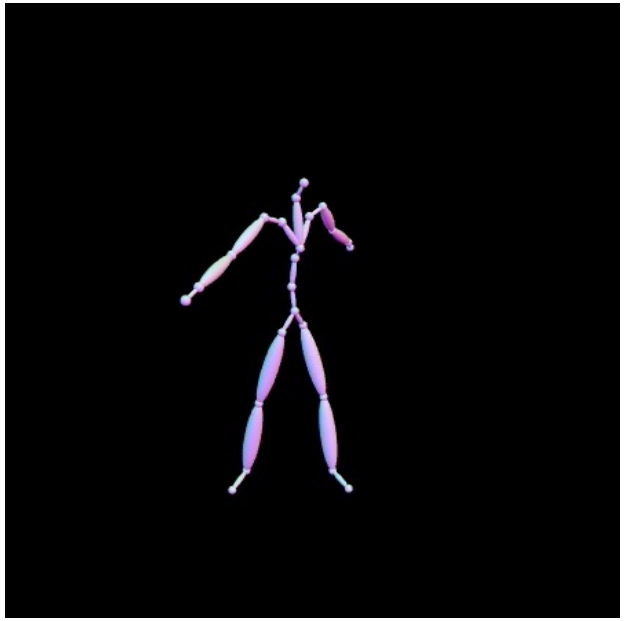
Dynamic humans – Neural body

Input 4-view video | Neural Volumes [1] | NeRF [2] | Neural Body

[1] Neural volumes: Learning dynamic renderable volumes from images, *SIGGRAPH* 2019.
[2] Nerf: Representing scenes as neural radiance fields for view synthesis, *ECCV* 2020.

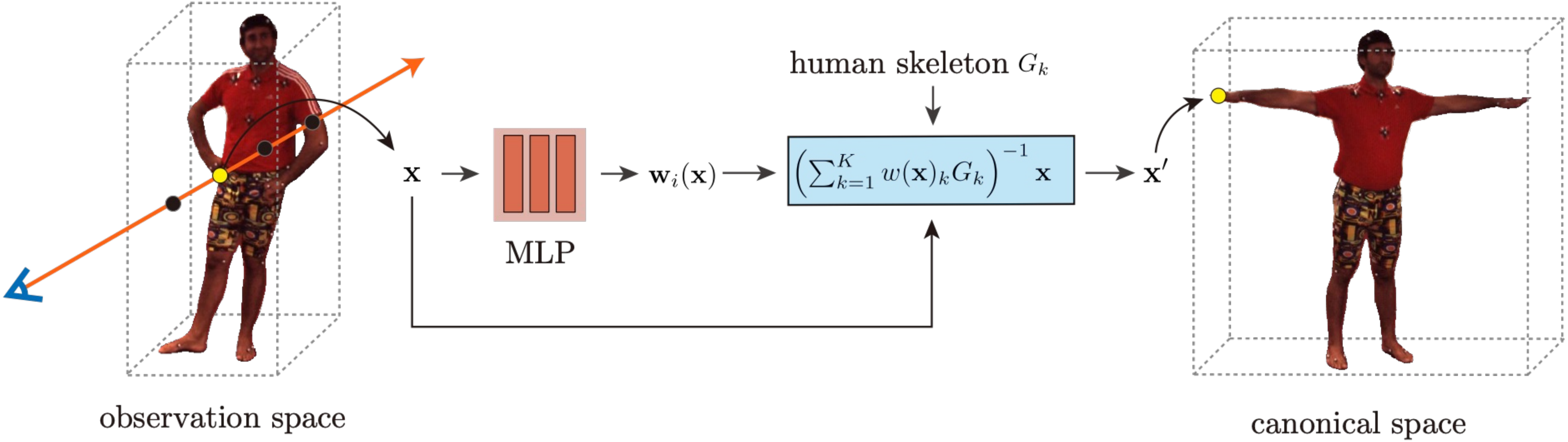
Dynamic humans – Animatable NeRF

Neural Body cannot synthesize images of **novel human poses** as the 3D convolution is **not equivariant** to pose changes



Dynamic humans – Animatable NeRF

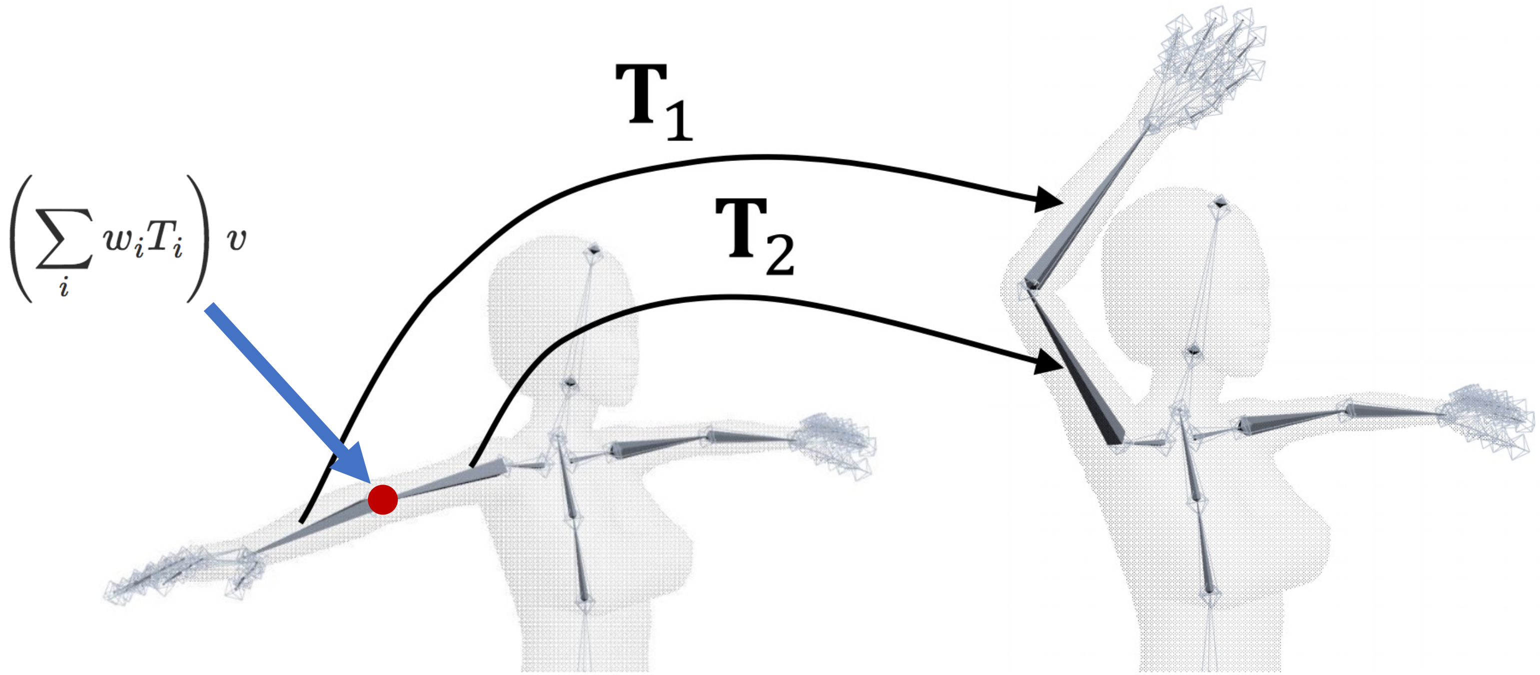
Key idea: deform NeRF with the linear blend scheme



Animatable neural radiance fields for human body modeling, *ICCV* 2021.

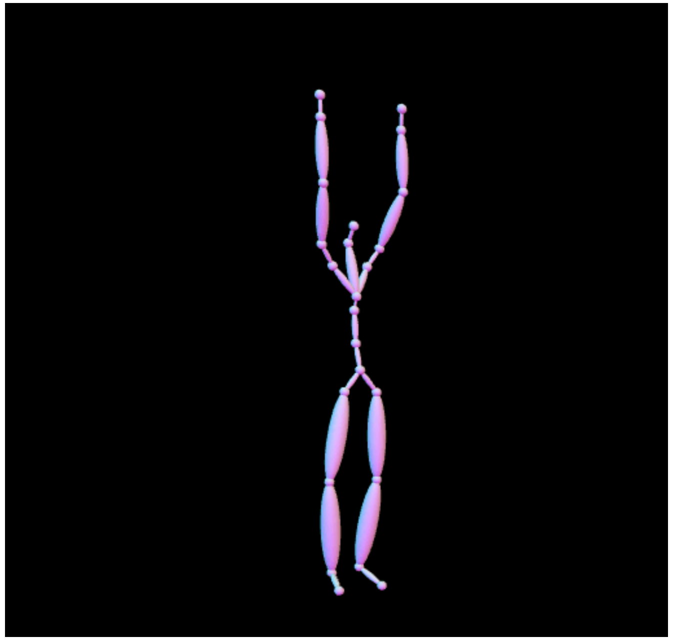
Dynamic humans – Animatable NeRF

Key idea: deform NeRF with the linear blend scheme



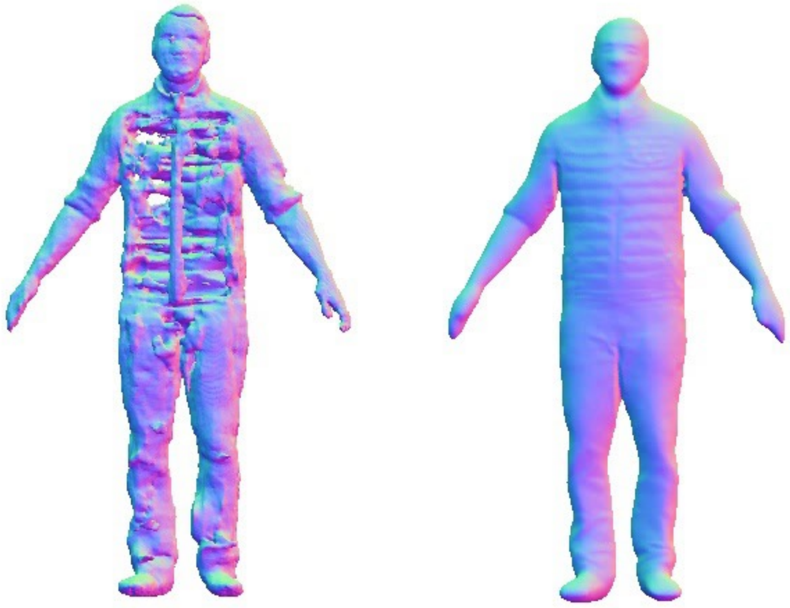
Animatable neural radiance fields for human body modeling, *ICCV* 2021.

Dynamic humans – Animatable NeRF



Dynamic humans – Animatable NeRF

Replace NeRF with Neural SDF (NeuS)



Animatable NeRF

Animatable NeuS

Animatable Implicit Neural Representations for Creating Realistic Avatars from Videos, *arXiv* 2022.

Dynamic humans – Animatable NeRF

Monocular video $\Longrightarrow$ detailed surface




Animatable Implicit Neural Representations for Creating Realistic Avatars from Videos, *arXiv* 2022.

Thanks！