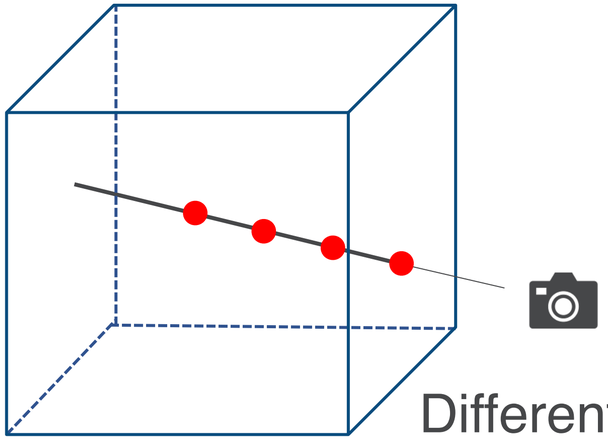
UC San Diego

Neural 3D Capturing with Better Quality, Efficiency, and Scalability

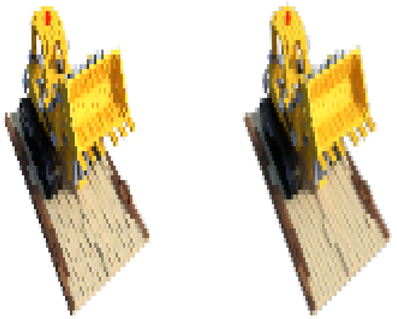
Hao Su

Neural 3D Capturing



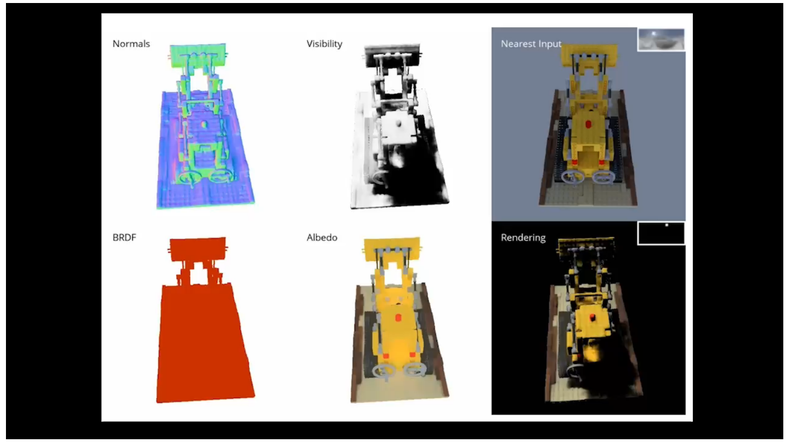
Differentiable Ray Marching

NeRF, 2020



Mip-NeRF, 2021

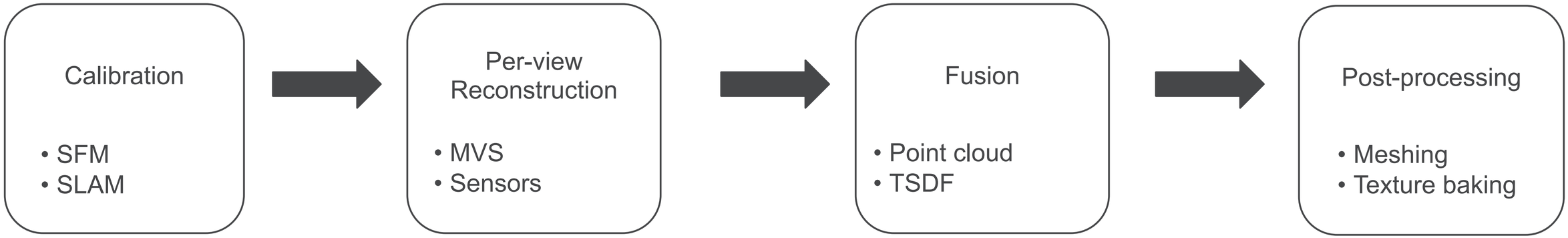
https://blogs.nvidia.com/blog/2022/03/25/instant-nerf-research-3d-ai/



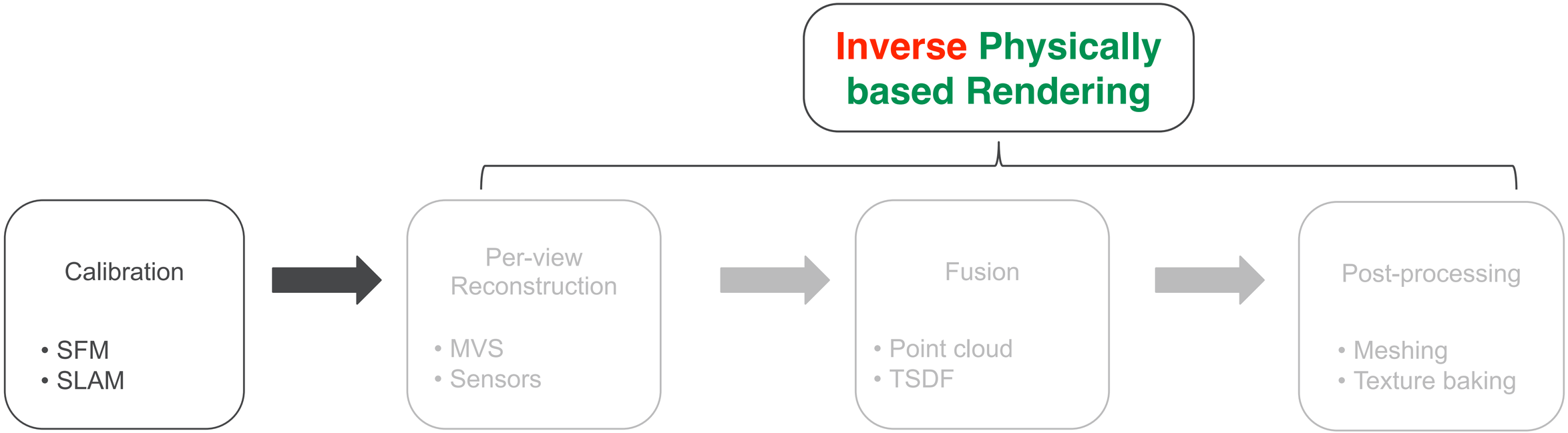
NeRFactor

Plenoxels, 2022

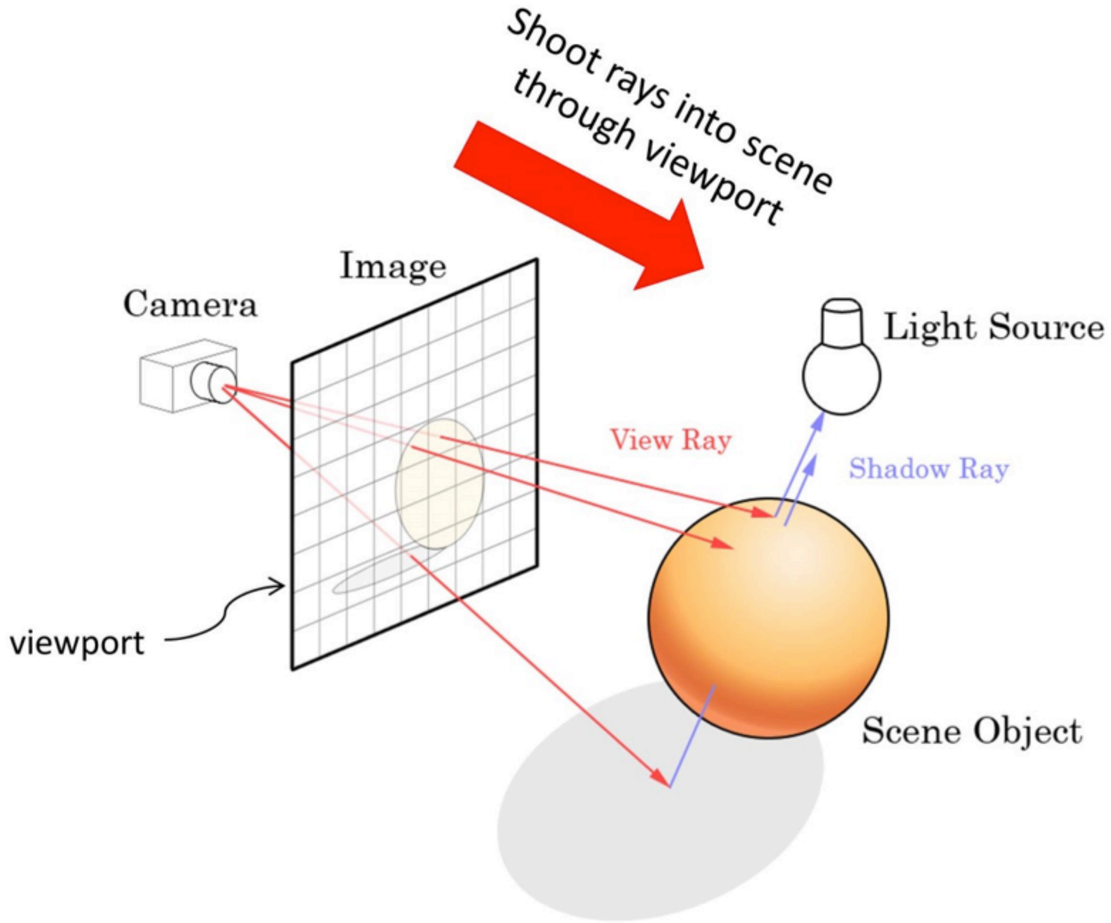
Classic 3D Capturing



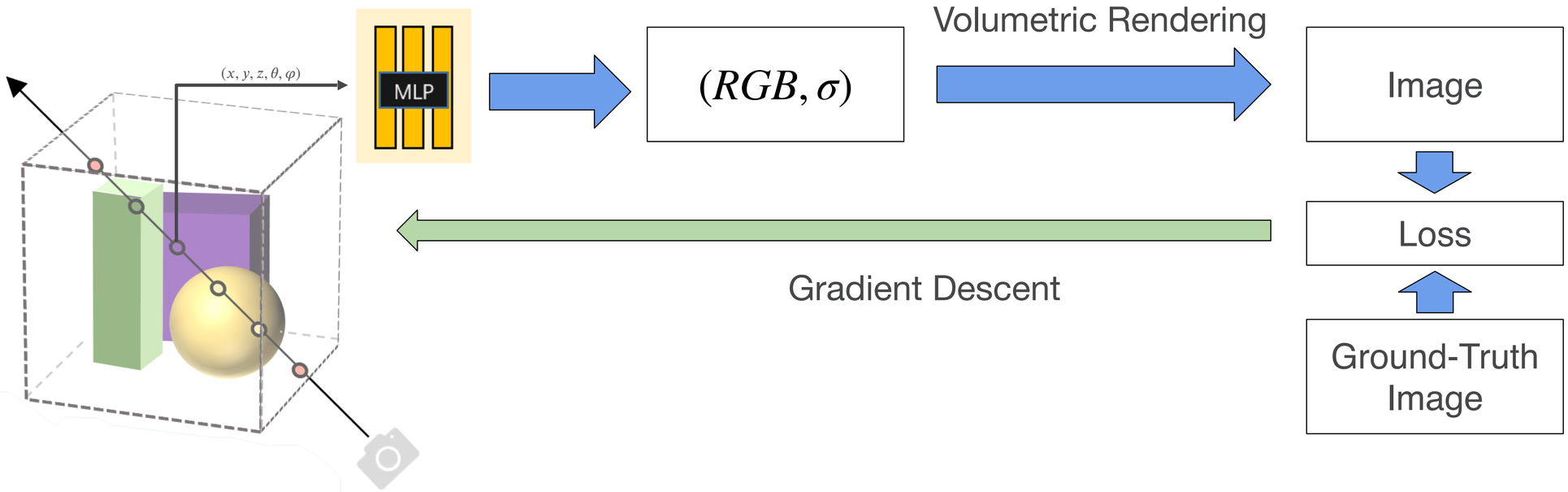
Neural 3D Capturing



Physically based Rendering

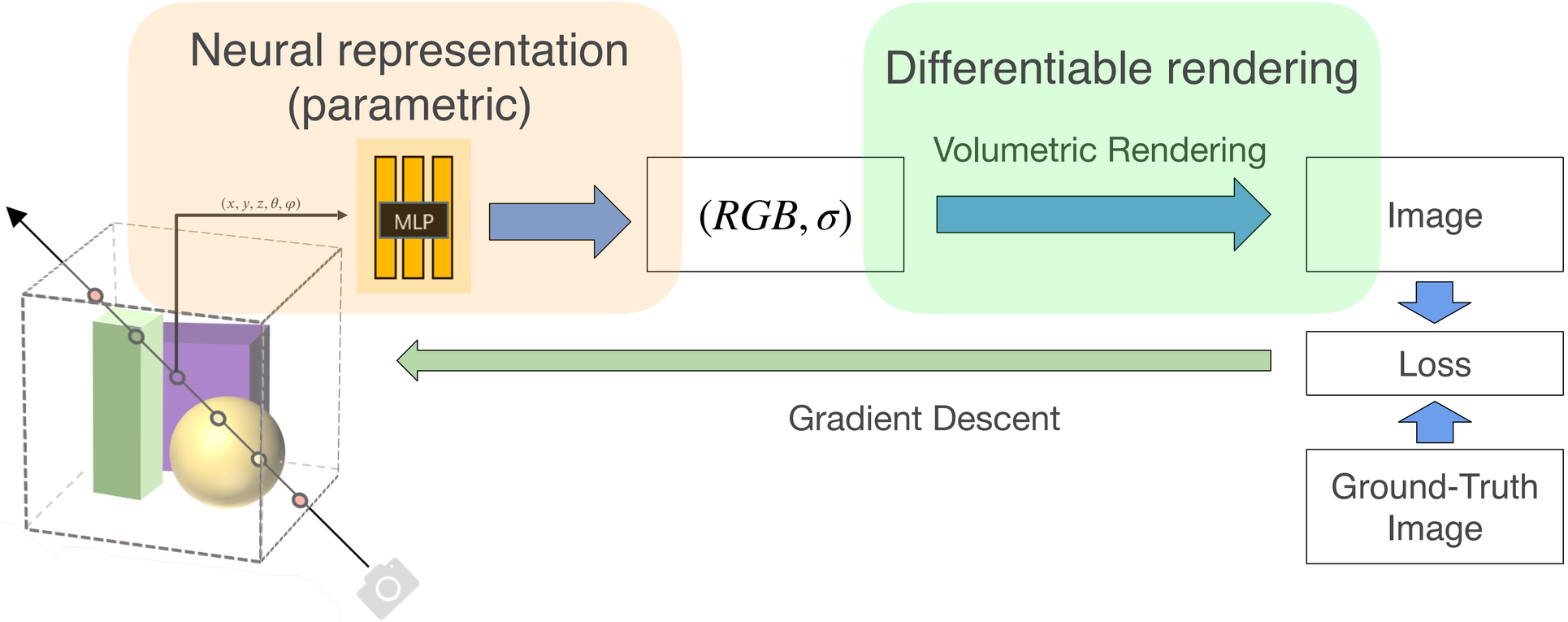


Neural 3D Capturing Example: NeRF



- Optimize on a single scene
  - store the scene in weights of the network
- Require ground-truth camera parameters

Neural 3D Capturing Example: NeRF



- Optimize on a single scene
  - store the scene in weights of the network
- Require ground-truth camera parameters

This Talk: Beyond Parametric Representation and Volumetric Rendering

- 3D capturing problem settings:

Optimization-only	v.s.	**Pretraining-based**
Offline Capturing	v.s.	**Online Capturing**
Radiance Field Estimation	v.s.	**Inverse Rendering**

Talk Outline

- 3D Capturing with Pretraining (MVSNeRF)
- Online Scalable 3D Capturing (NeRFusion)
- Tensorial Radiance Field (TensoRF)
- Tensorial Inverse Rendering (TensoVerse)

Talk Outline

- 3D Capturing with Pretraining (MVSNeRF)
- Online Scalable 3D Capturing (NeRFusion)
- Tensorial Radiance Field (TensoRF)
- Tensorial Inverse Rendering (TensoVerse)

Generalizable 3D Capturing Setup

- If a large-scale multi-view dataset can be provided for **pretraining**, can we **learn priors** to benefit **test scenes**?

Optimization-only (NeRF)		With pretraining (Our MVSNeRF)
10 hrs	→	<1s
>30 views	→	3 views

UC San Diego

MVSNeRF: Fast Generalizable Radiance Field Reconstruction from Multi-View Stereo

Anpei Chen*[1] Zexiang Xu*[2] Fuqiang Zhao[1] Xiaoshuai Zhang[3] Fanbo Xiang[3]
Jingyi Yu[1] Hao Su[3]

[1]ShanghaiTech University [2]Adobe Research [3]University of California, San Diego

CVPR2021

Scene

MVSNeRF

Goal: novel view synthesis from given a few input views without optimization

✓ A differentiable rendering procedure
✓ Appearance and geometric reasoning

- Looooong training time
- Looooong inference time
- Couldn’t generalize to a new scene

NeRF

✓ Physical meaningful volume modeling
✓ Generalize well
✓ Fast inference time

- 3D supervision
- Flying points around edges

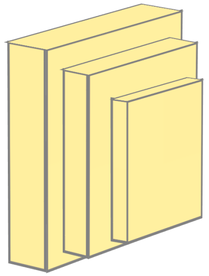
MVS

Basic idea of MVSNeRF: free viewpoint rendering without optimization via combing the advantages of nerf and mvs.

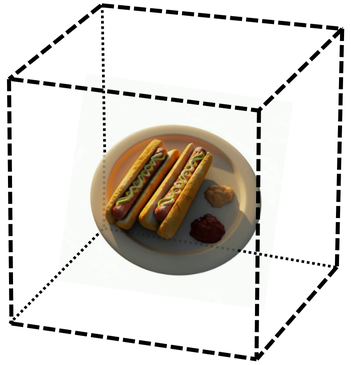
Overview

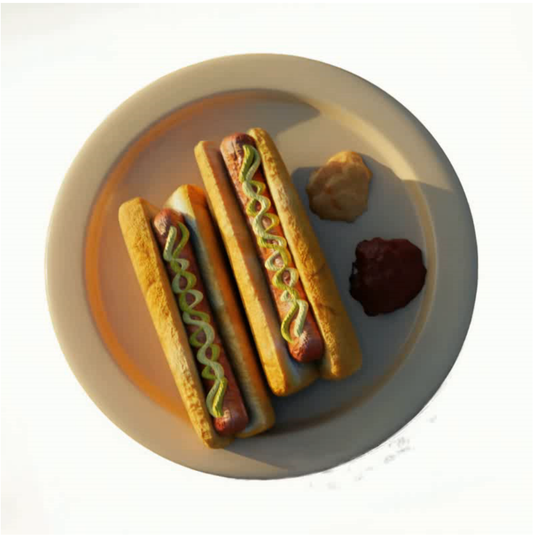
3 input images

Deep 3D Conv Network

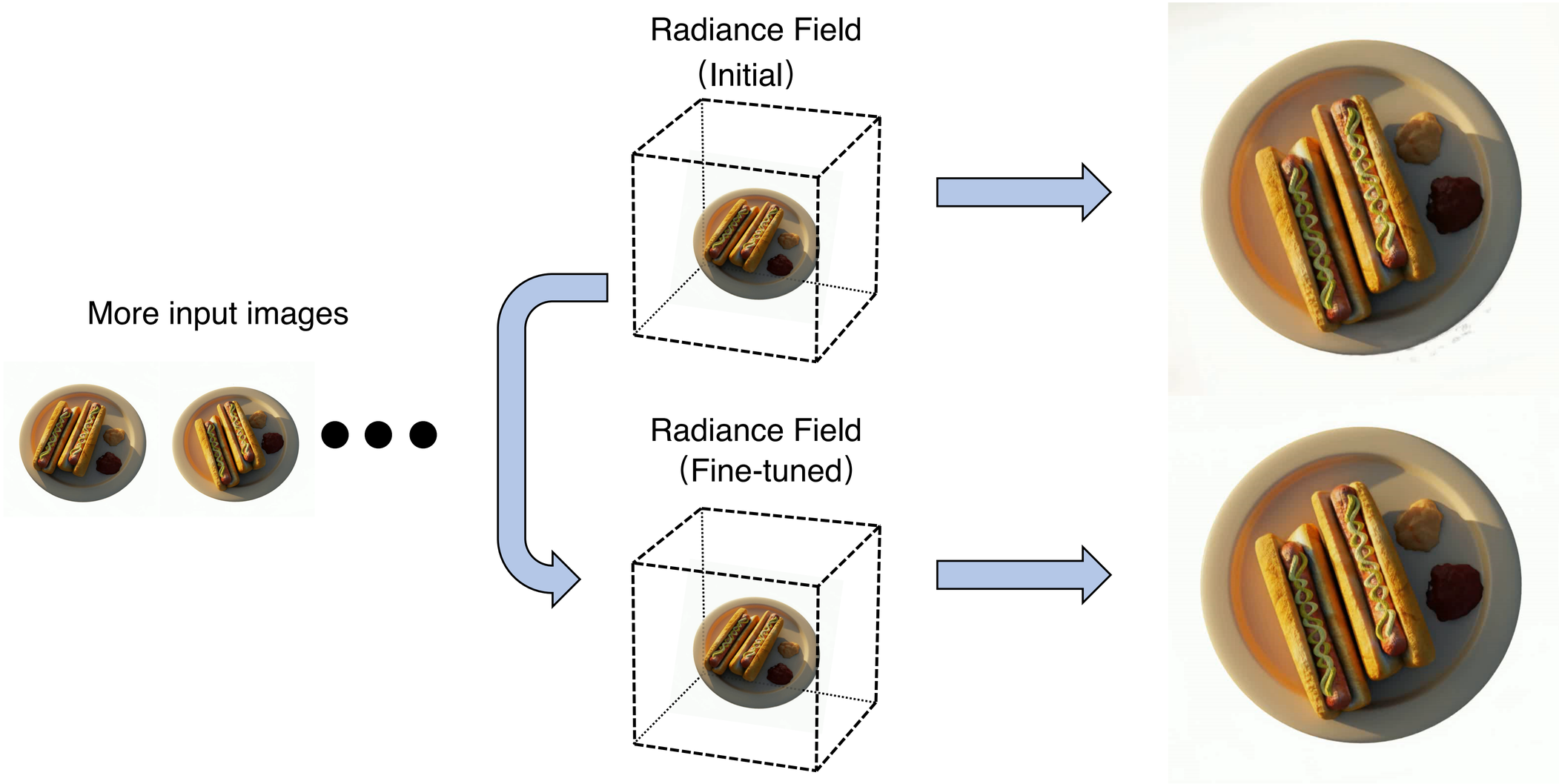


Radiance Field
(Initial)



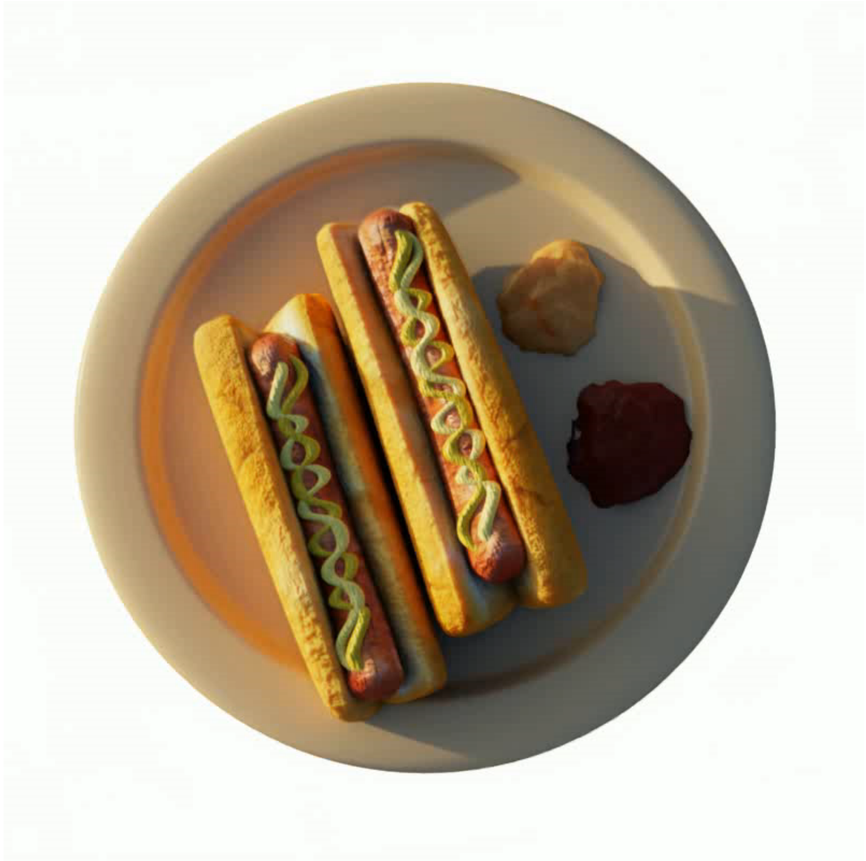


Overview


Radiance Field
(Initial)
More input images
Radiance Field
(Fine-tuned)

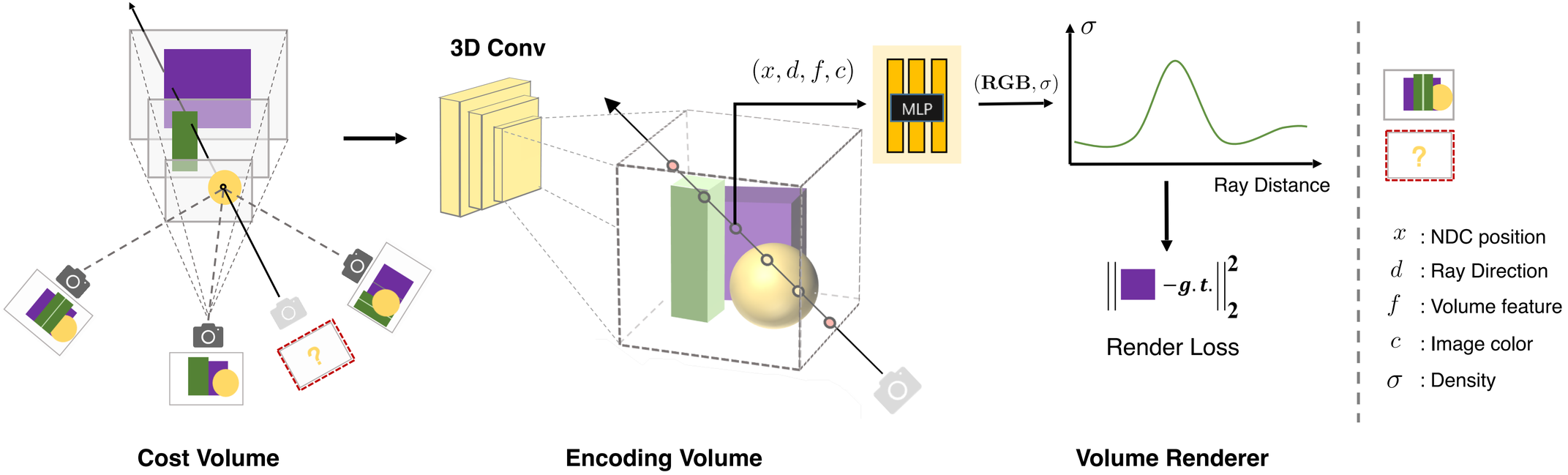
Our results w/o and w/ fine-tuning

0 min (no fine-tuning)



15 min fine-tuning

Pipeline



$$\sigma, r = \mathrm{MVSNeRF}(x, d; I_i, \Phi_i)$$

Results on FF scenes (three input images, no per-scene fine-tuning)

PixelNeRF

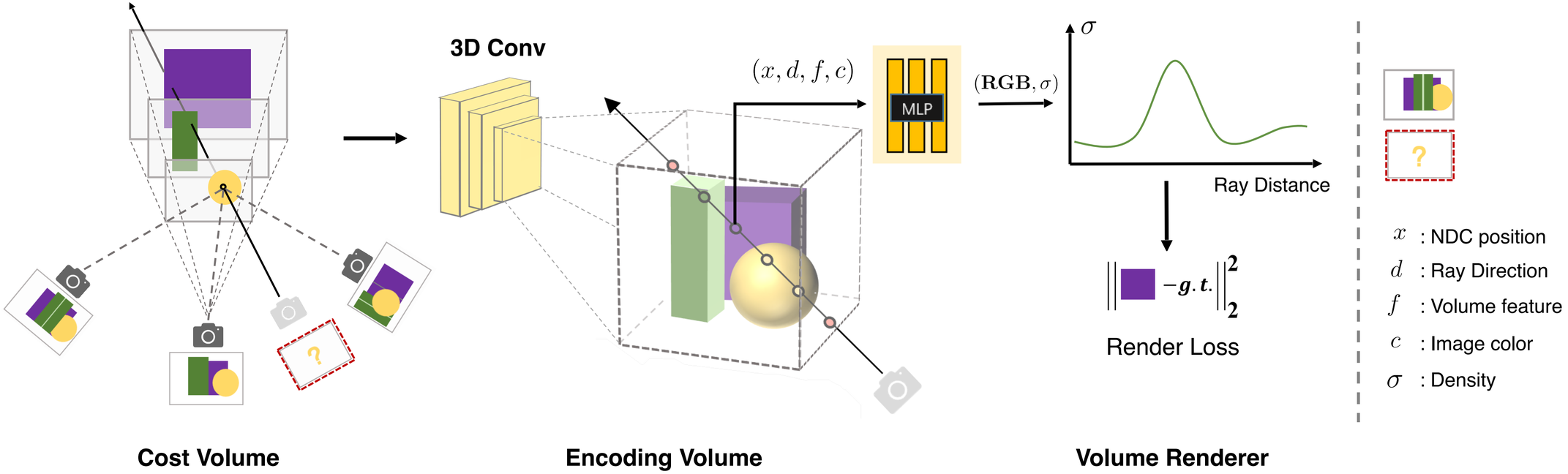
Results on FF scenes (three input images, no per-scene fine-tuning)

Source Views

IBRNet

Ours

Fine-tuning



$$\sigma, r = \mathrm{MVSNeRF}(x, d; I_i, \Phi_i)$$

Comparisons on synthetic scenes (with optimization/fine-tuning)

IBRNet
(1h optimization)

Ours
(15 min fine-tuning)

Comparisons on synthetic scenes (with optimization/fine-tuning)

NeRF
(10.2h optimization)

Ours
(15 min fine-tuning)

Talk Outline

- 3D Capturing with Pretraining (MVSNeRF)
- Online Scalable 3D Capturing (NeRFusion)
- Tensorial Radiance Field (TensoRF)
- Tensorial Inverse Rendering (TensoVerse)

Scalability Issue of NeRF/MVSNeRF

NeRF relies on a single MLP to overfit the scene

- Limited network capacity

MVSNeRF relies on a dense 3D volume to store the scene

- High GPU memory cost at training and inference

Image courtesy of *Neural Sparse Voxel Fields*

UC San Diego

NeRFusion: Fusing Radiance Fields for Large-Scale Scene Reconstruction

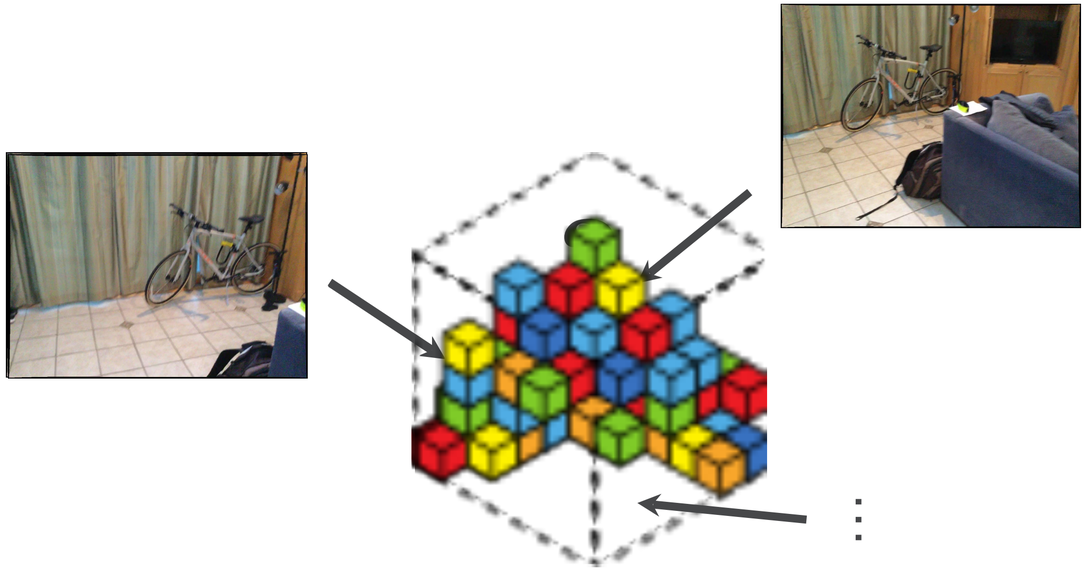
Xiaoshuai Zhang, Sai Bi, Kalyan Sunkavalli, Hao Su, Zexiang Xu

CVPR2022 (oral)

Fusing a Global Scene Representation

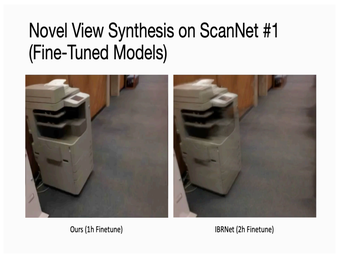
NeRFusion fuses new information from input images increasingly into a global sparse representation

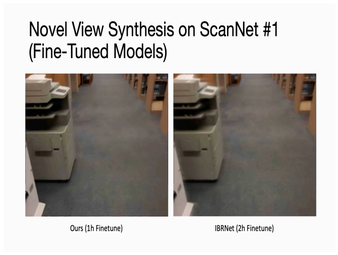
- ✓ **Generalizable:** do not need hour-long retraining for new scenes
- ✓ **Scalable:** from small object-level to large scene-level
- ✓ **Online:** new images could be added to update the model

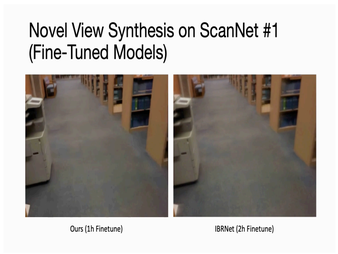


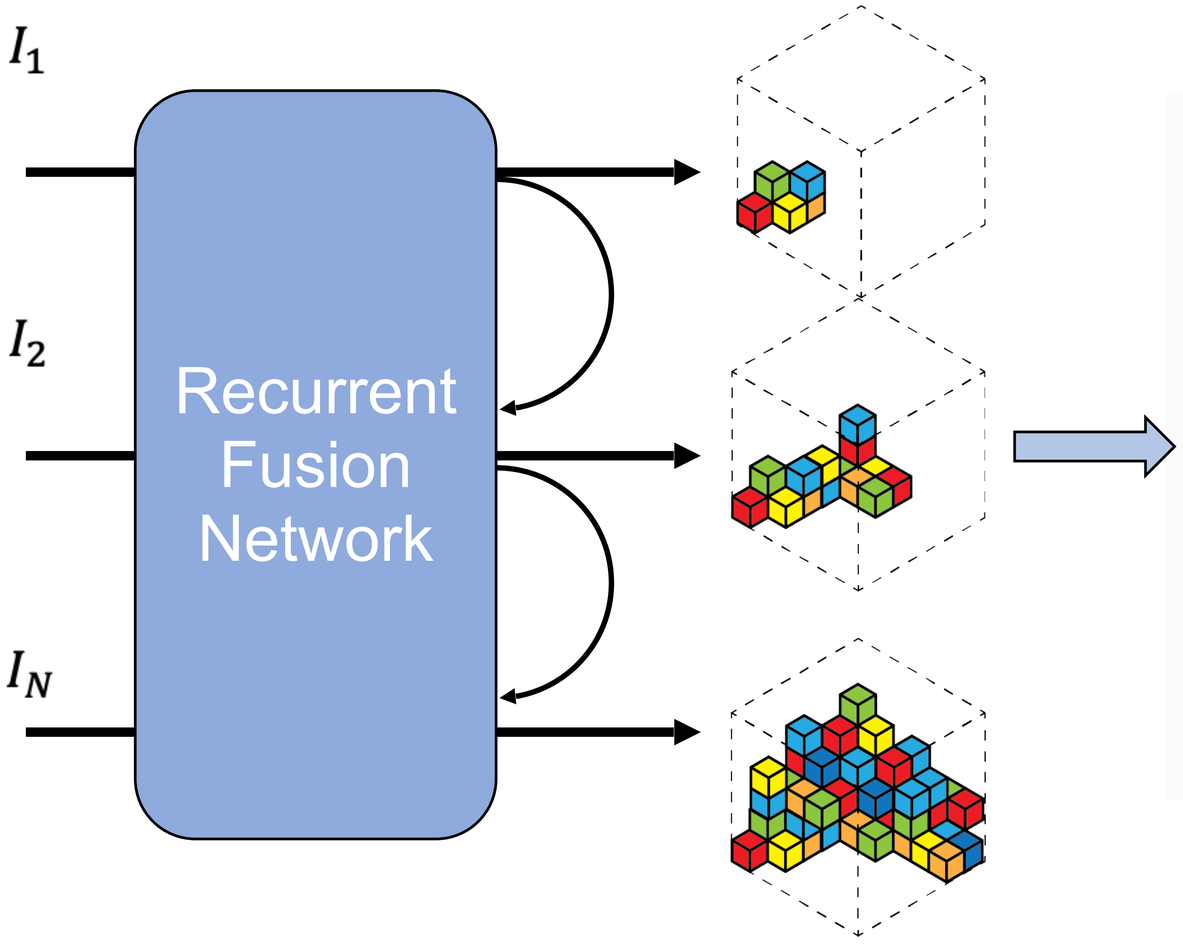
Global Scene Representation

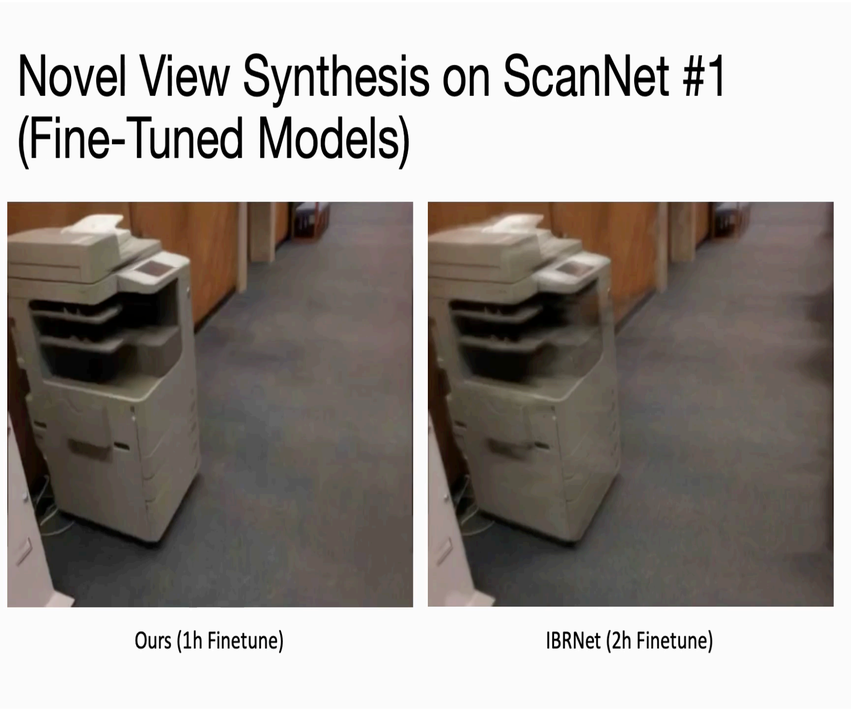
NeRFusion: Fusing Radiance Fields







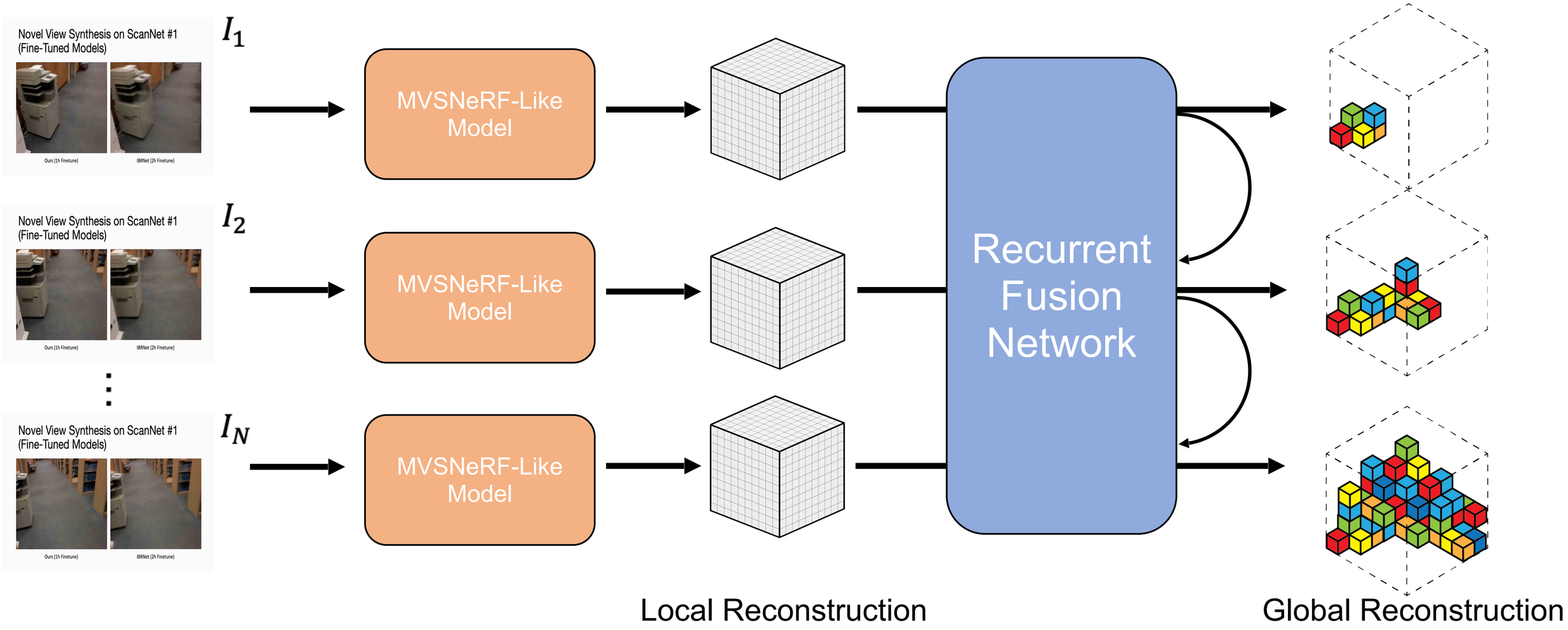




Novel view synthesis results

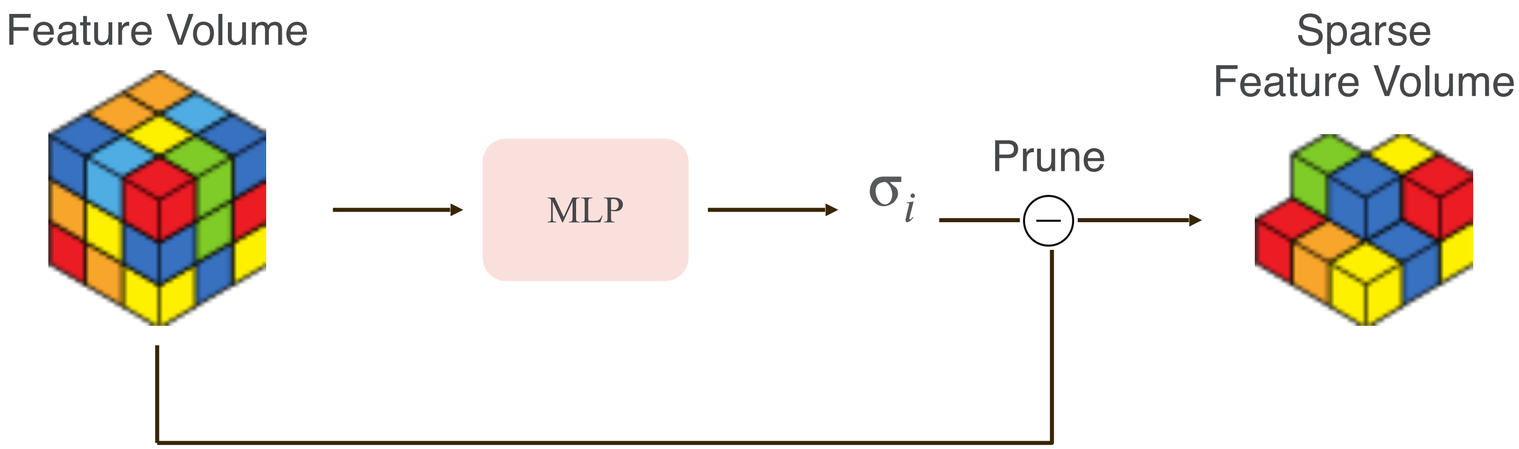
Zhang, Bi, Sunkavalli, Su, Xu; NeRFusion: Fusing Radiance Fields for Large-Scale Scene Reconstruction, *CVPR 2022*

NeRFusion: Fusing Radiance Fields



Zhang, Bi, Sunkavalli, Su, Xu; NeRFusion: Fusing Radiance Fields for Large-Scale Scene Reconstruction, *CVPR 2022*

Adaptive Sparsification using Volume Density



Online Reconstruction

Input frames

Online reconstruction at 20 fps

Zhang, Bi, Sunkavalli, Su, Xu; NeRFusion: Fusing Radiance Fields for Large-Scale Scene Reconstruction, *CVPR 2022*

ScanNet Reconstruction Through Direct-Inference

NeRF (12h Training)

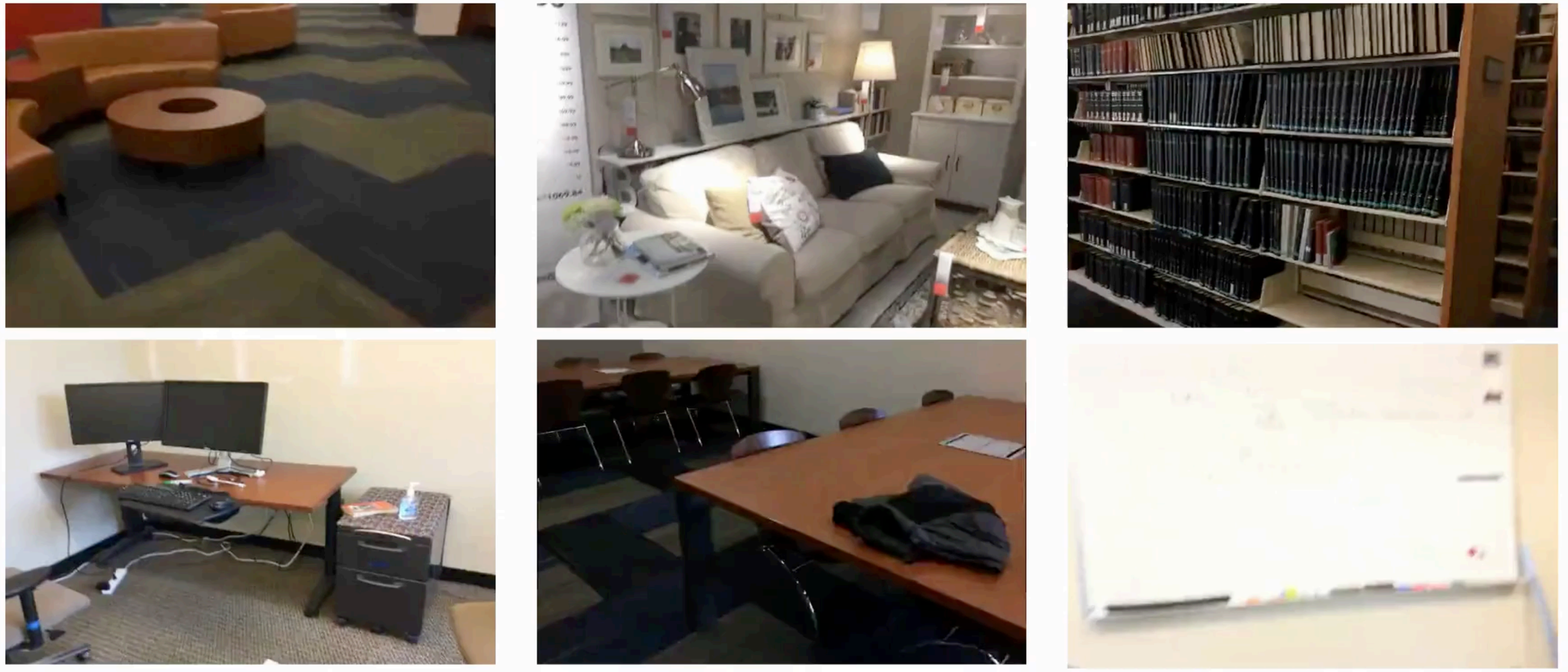
NeRFusion (Direct Inference)

Scene from ScanNet
Training On-Trajectory Frames

Zhang, Bi, Sunkavalli, Su, Xu; NeRFusion: Fusing Radiance Fields for Large-Scale Scene Reconstruction, *CVPR 2022*

ScanNet Reconstruction Through Fine-Tuning

Results: ScanNet ($\text{Ours}_{\text{ft-1h}}$)



Zhang, Bi, Sunkavalli, Su, Xu; NeRFusion: Fusing Radiance Fields for Large-Scale Scene Reconstruction, *CVPR 2022*

Summary of MVSNeRF and NeRFusion

- With per-cell features, volumetric representation provides better ability to store local information
- MVS network allows knowledge transfer from pretraining data
- MVS network allows direct inference without optimization
- We exploit the sparsity of 3D space to achieve scalable reconstruction

Talk Outline

- 3D Capturing with Pretraining (MVSNeRF)
- Online Scalable 3D Capturing (NeRFusion)
- Tensorial Radiance Field (TensoRF)
- Tensorial Inverse Rendering (TensoVerse)

Representation Design from Inductive Bias Perspective

- MLP as in NeRF is a regularizer applies globally
- Feature volume as in MVSNeRF increases flexibility but is less compact
- Sparse feature volume as in NeRFusion reduces model capacity by spatial sparsity
- Can we impose stronger regularization over feature volume?

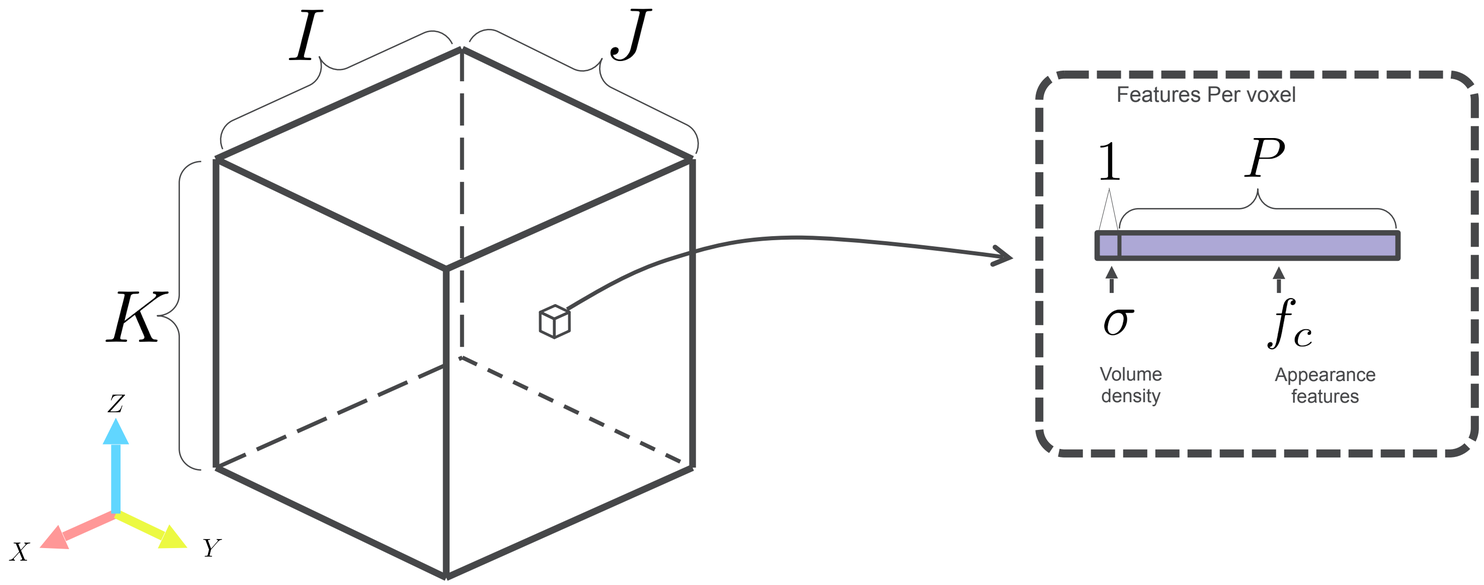
UC San Diego

TensoRF

Tensorial Radiance Fields

Anpei Chen*, Zexiang Xu*, Andreas Geiger, Hao Su

Tensorial Perspective to Radiance Fields

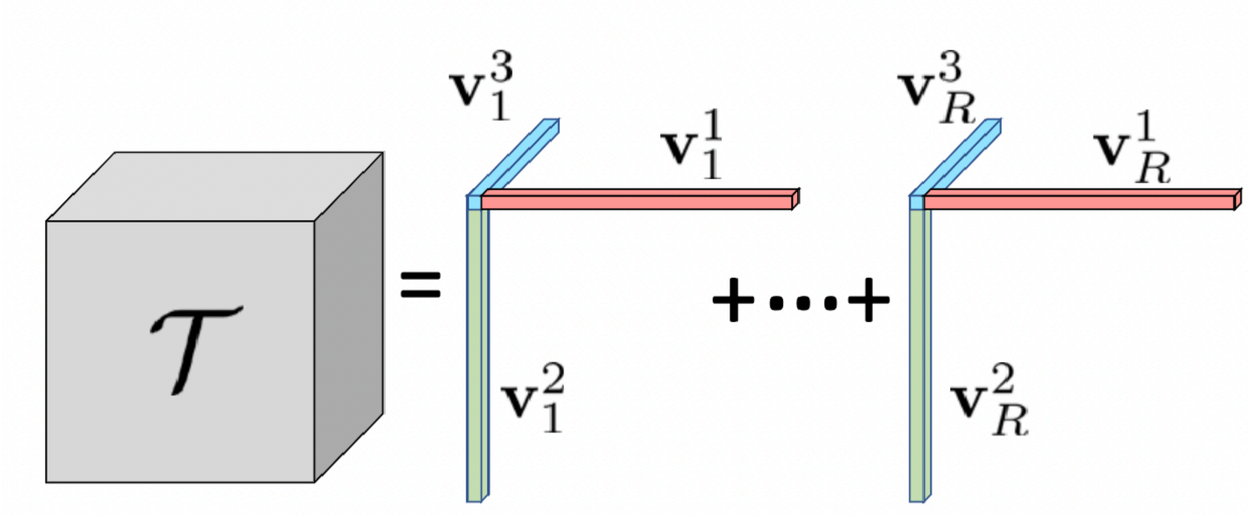


Radiance Field can be viewed as a 4D tensor

Tensorial Decomposition

CP Decomposition

- High compactness with ***low-rank*** approximation



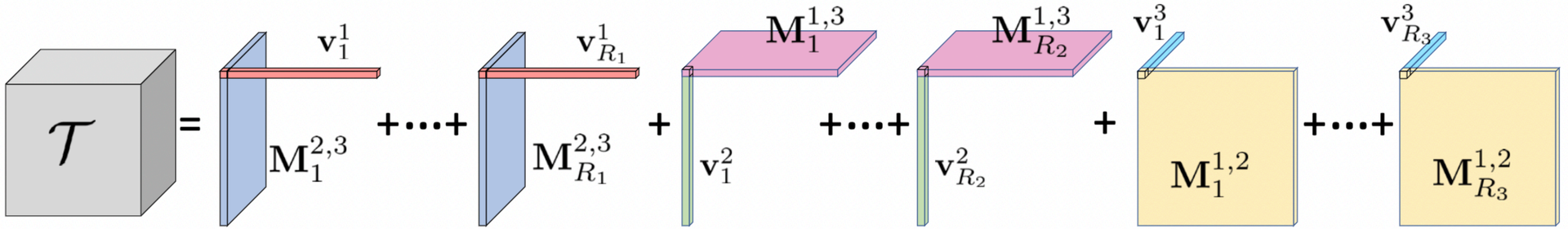
$$\mathcal{T} = \sum_{r=1}^{R} \mathbf{v}_r^1 \circ \mathbf{v}_r^2 \circ \mathbf{v}_r^3$$

$$\mathcal{T}_{ijk} = \sum_{r=1}^{R} \mathbf{v}_{r,i}^1 \mathbf{v}_{r,j}^2 \mathbf{v}_{r,k}^3$$

$$\mathcal{O}(N^3) \text{ to } \mathcal{O}(N)$$

However, CP decomposition require many components to reach high accuracy

Our Vector-Matrix Decomposition

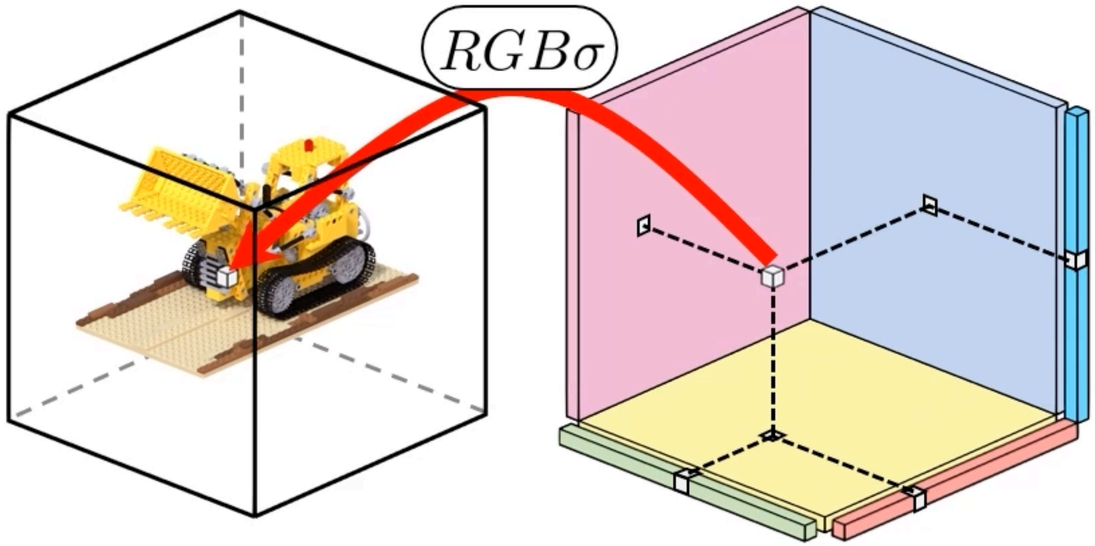


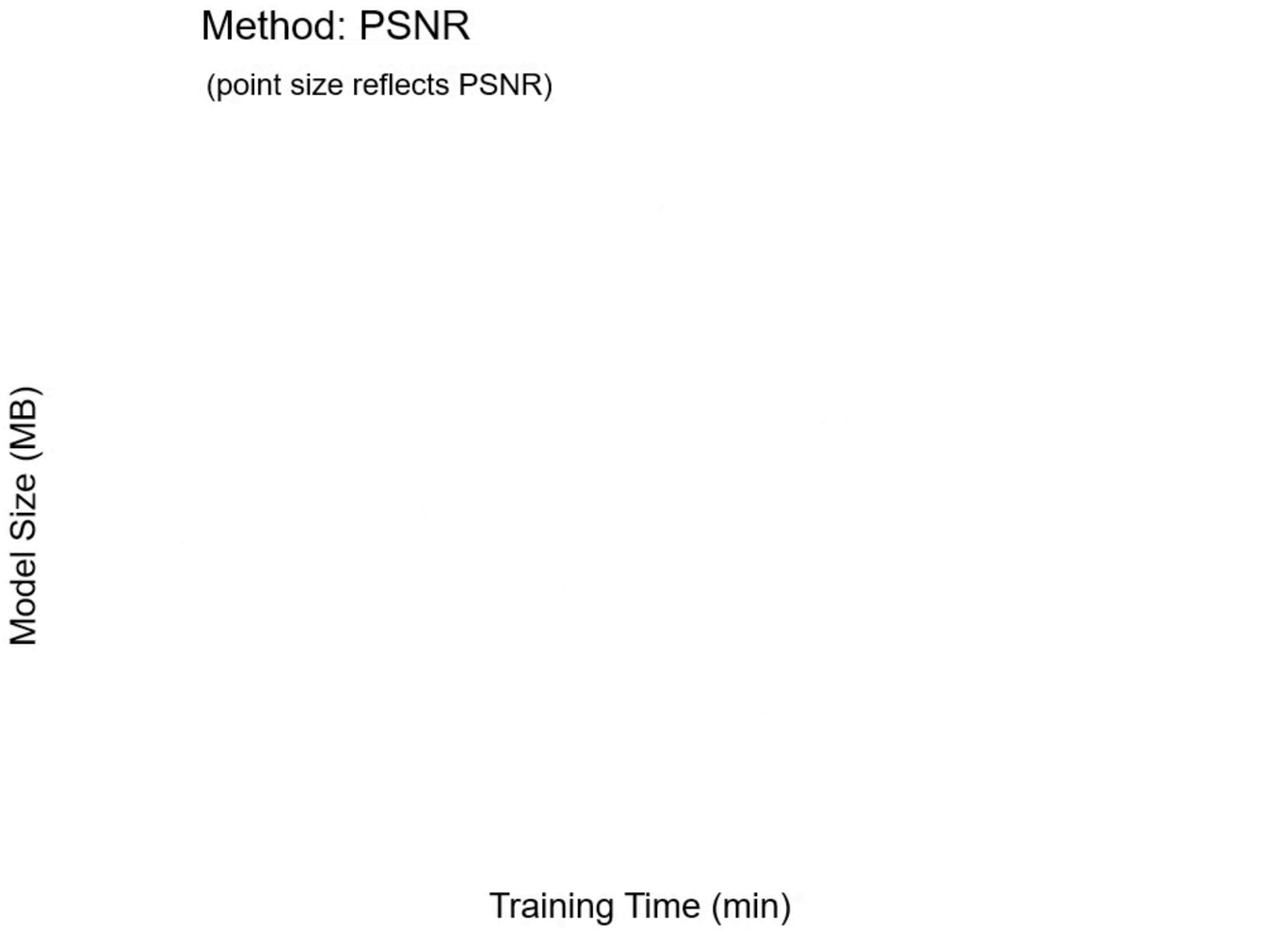
$$\mathcal{T} = \sum_{r=1}^{R_1} \mathbf{v}_r^1 \circ \mathbf{M}_r^{2,3} + \sum_{r=1}^{R_2} \mathbf{v}_r^2 \circ \mathbf{M}_r^{1,3} + \sum_{r=1}^{R_3} \mathbf{v}_r^3 \circ \mathbf{M}_r^{1,2}$$

$$\mathcal{O}(N^3) \text{ to } \mathcal{O}(N^2)$$

$$\begin{aligned}
\mathcal{T}_{ijk} &= \sum_{r=1}^{R} \mathbf{v}_{r,i}^X \mathbf{M}_{r,jk}^{YZ} + \mathbf{v}_{r,j}^Y \mathbf{M}_{r,ik}^{XZ} + \mathbf{v}_{r,k}^Z \mathbf{M}_{r,ij}^{XY} \\
&= \sum_{r=1}^{R} \mathcal{A}_{r,ijk}^X + \mathcal{A}_{r,ijk}^Y + \mathcal{A}_{r,ijk}^Z \\
&= \sum_{r=1}^{R} \sum_{m \in XYZ} \mathcal{A}_{r,ijk}^m,
\end{aligned}$$

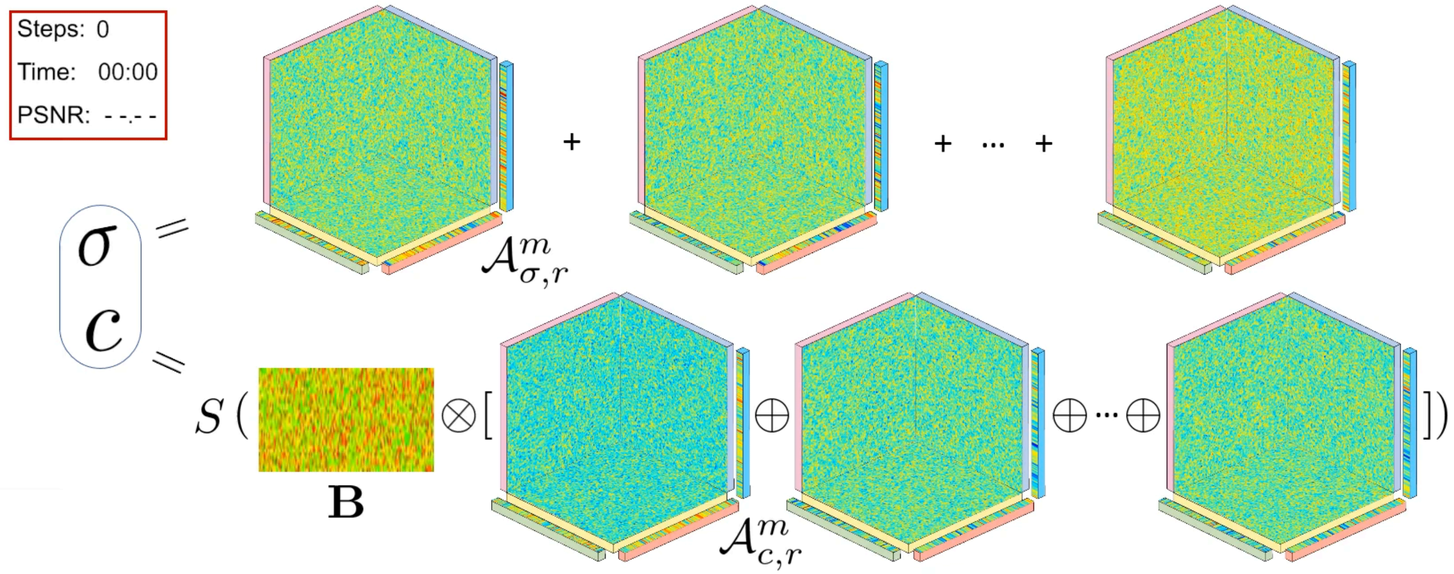
TensoRF: Tensorial Radiance Fields





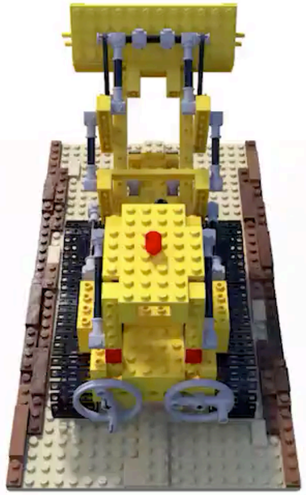
Chen, Xu, Geiger ,Yu, Su; TensoRF: Tensorial Radiance Fields, arXiv 2022

Fast Training Process



Chen, Xu, Geiger ,Yu, Su; TensoRF: Tensorial Radiance Fields, arXiv 2022

TensoRF: Tensorial Radiance Fields






Chen, Xu, Geiger ,Yu, Su; TensoRF: Tensorial Radiance Fields, arXiv 2022

Visualization of Bases