基于神经表示的三维重建与生成

Towards Real Application

崔兆鹏

浙江大学CAD&CG国家重点实验室

Outline

- Large-scale and outdoor 3D neural reconstruction
- 3D editing with neural radiance fields
- Pose estimation for 3D neural reconstruction

Large-scale and Outdoor 3D Neural Reconstruction

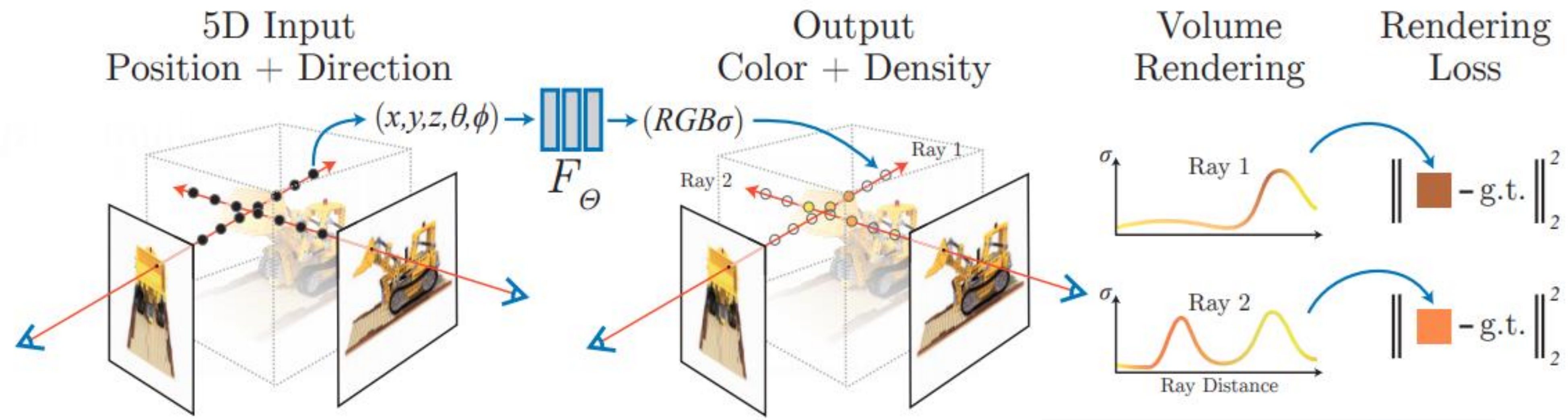
“Scalability”

Challenges for applying NeRF to large-scale scenes

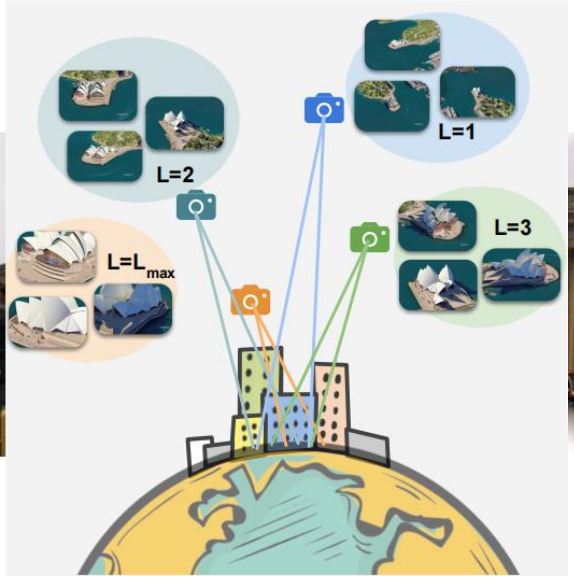
Can we directly apply NeRF to outdoor and large-scale scenes?

No!

Challenges for applying NeRF to large-scale scenes

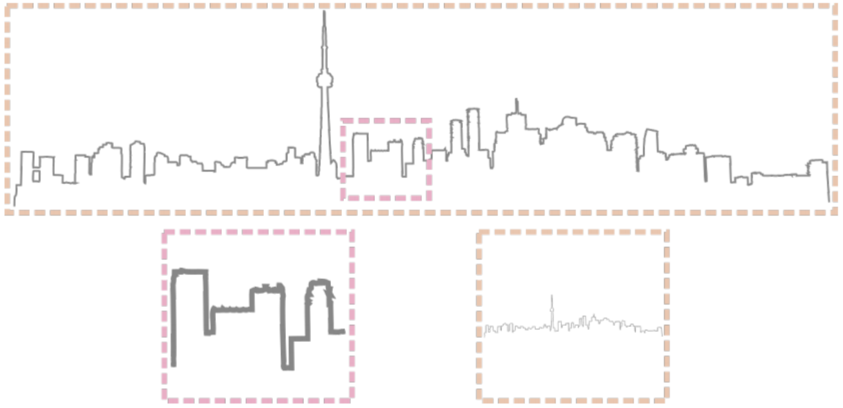


- × Unbounded scenes
- × Dynamic objects
- × Multi-scale images
- × Limited network capabilities

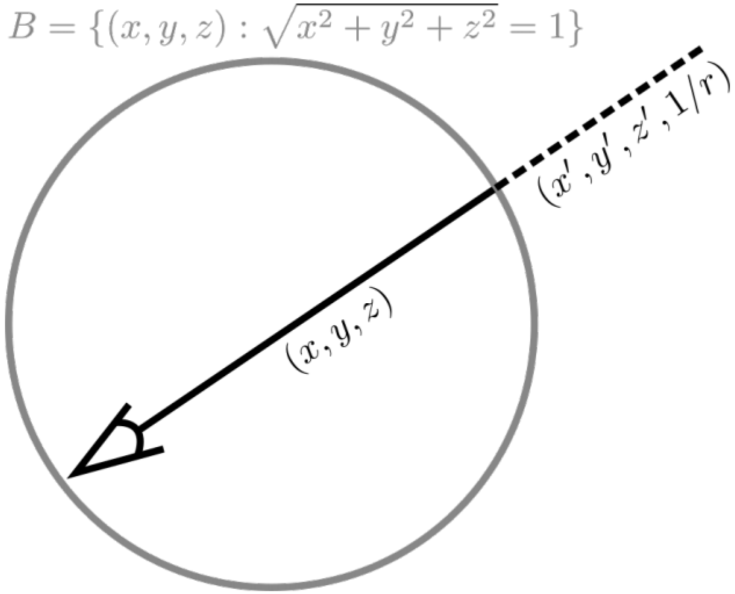


NeRF++

- **Goal:**
  - Address a parametrization issue involved in applying NeRF to 360◦ captures of objects within large-scale, unbounded 3D scenes.
- **Solution:**
  - Samples within a unit sphere enclosing all camera poses to render its foreground component and uses a different methodology for the background.



Parameterization of Unbounded Scenes

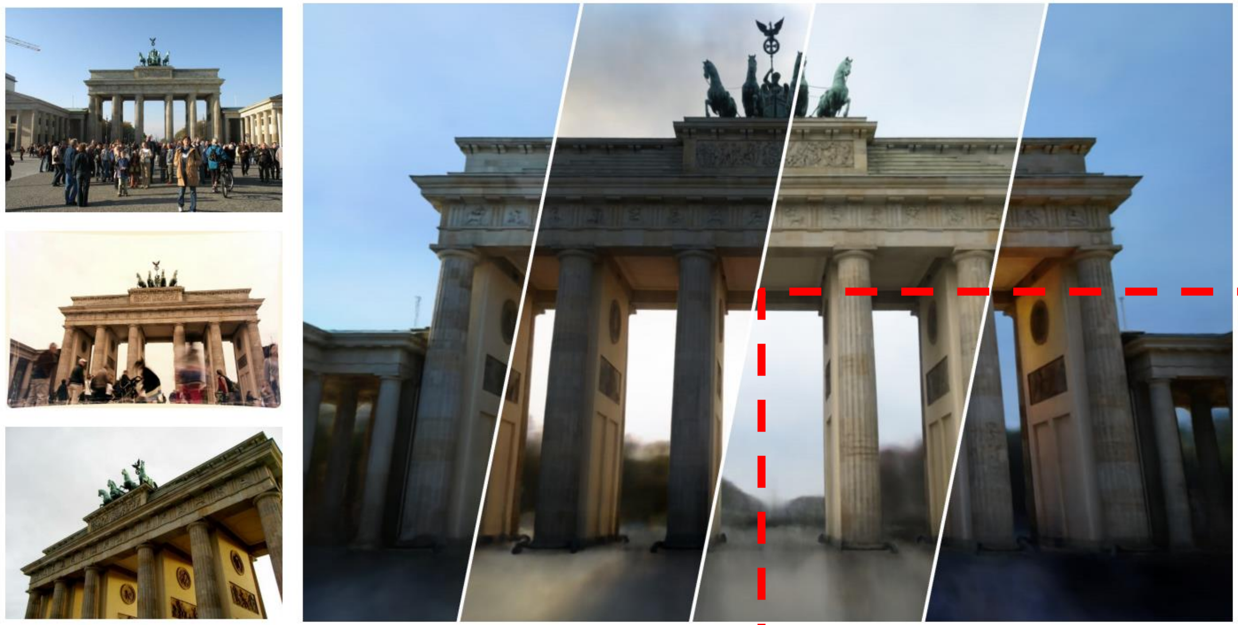


Zhang, Kai, et al. "Nerf++: Analyzing and improving neural radiance fields." arXiv preprint arXiv:2010.07492 (2020).

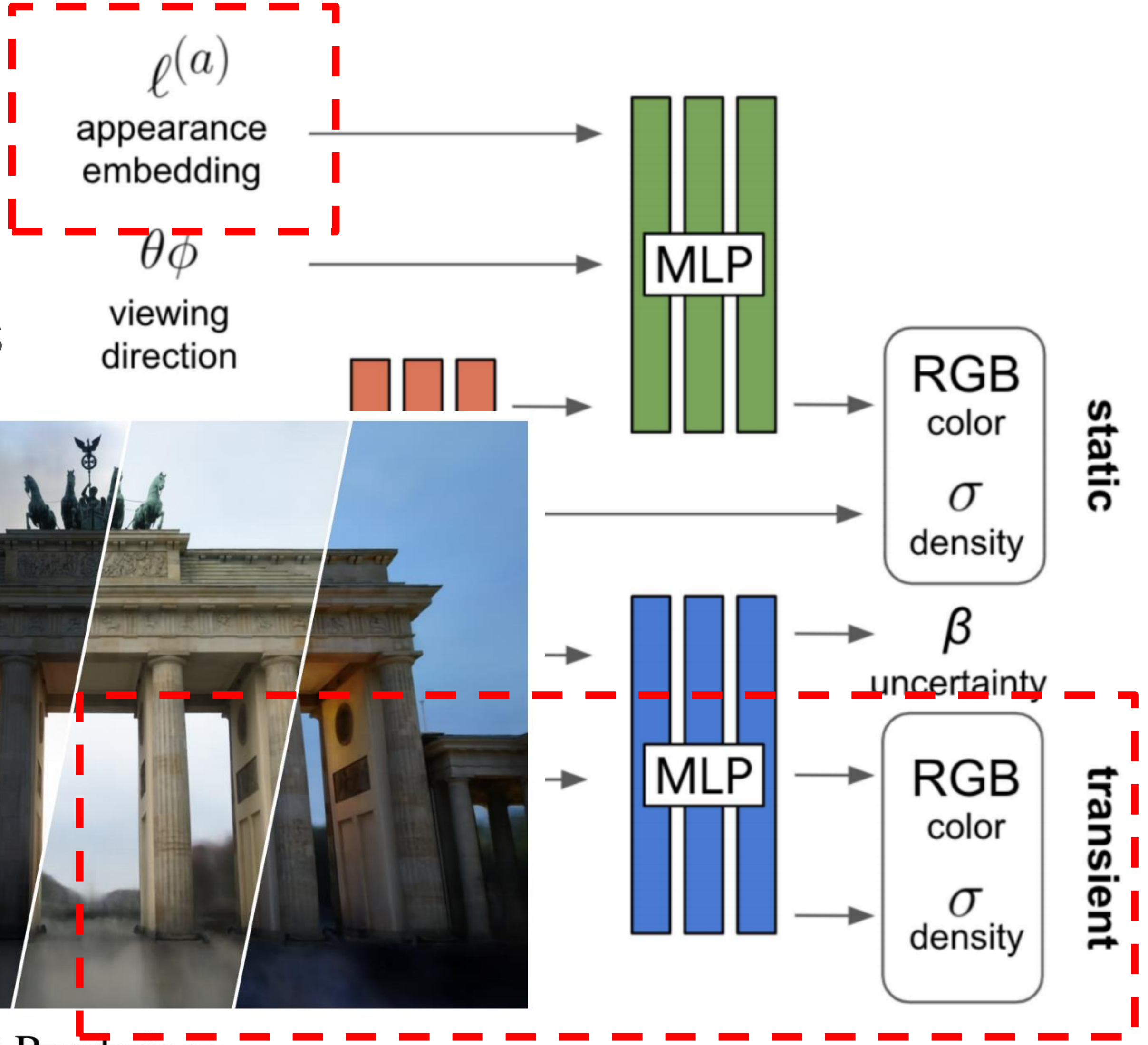
NeRF in the Wild

- **Goal:**
  - Synthesizing novel views of complex scenes using only unstructured collections of in-the-wild photographs

- **Solution:**
  - Model per-im
    in a low-dime
  - Unsupervise
    content into “
    components



(a) Photos (b) Renderings



Martin-Brualla, Ricardo, et al. “Nerf in the wild: Neural radiance fields for unconstrained photo collections.” CVPR 2021.

NeRF in the Wild

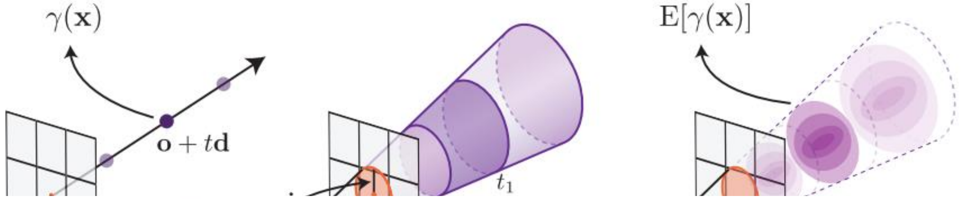


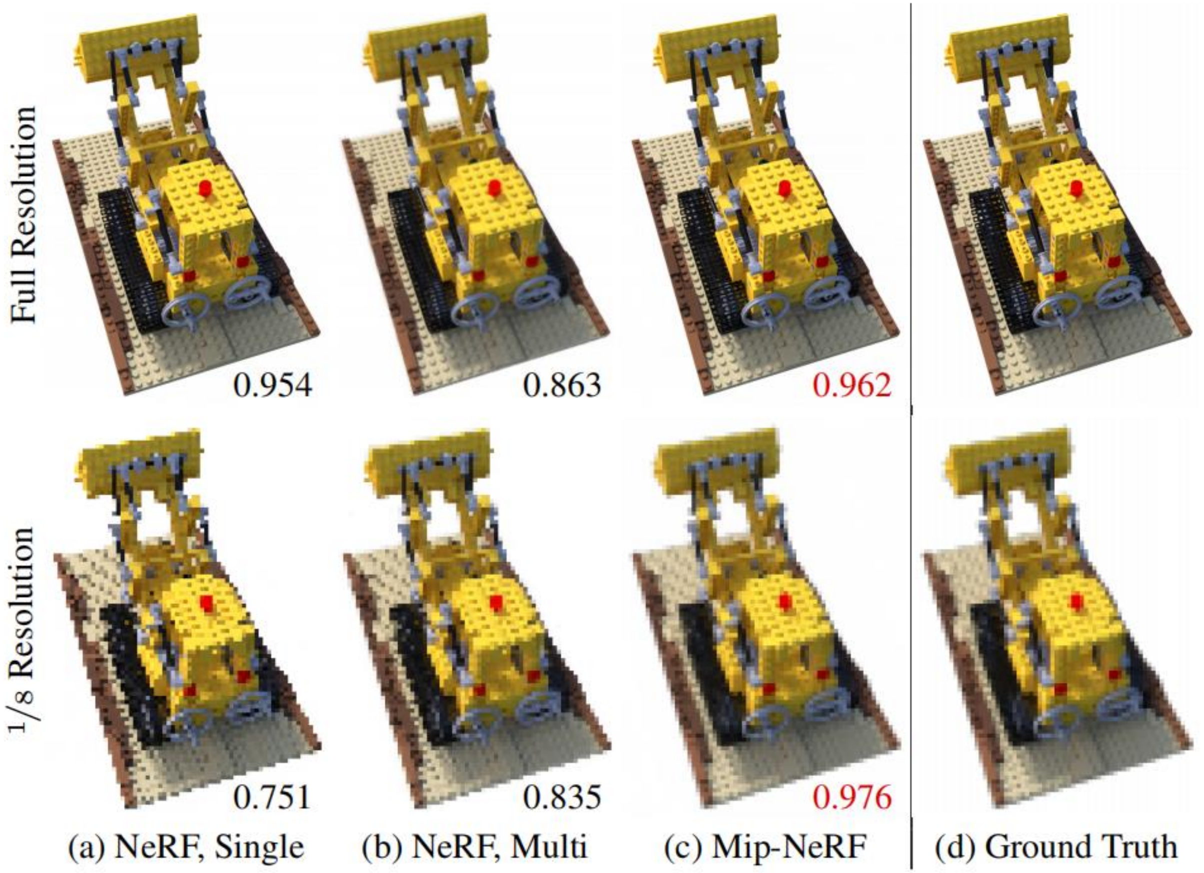
Martin-Brualla, Ricardo, et al. "Nerf in the wild: Neural radiance fields for unconstrained photo collections." CVPR 2021.

Mip-NeRF

- **Goal**:
  - Solve the blurred or aliased issue when training or testing images observe scene content at different resolutions

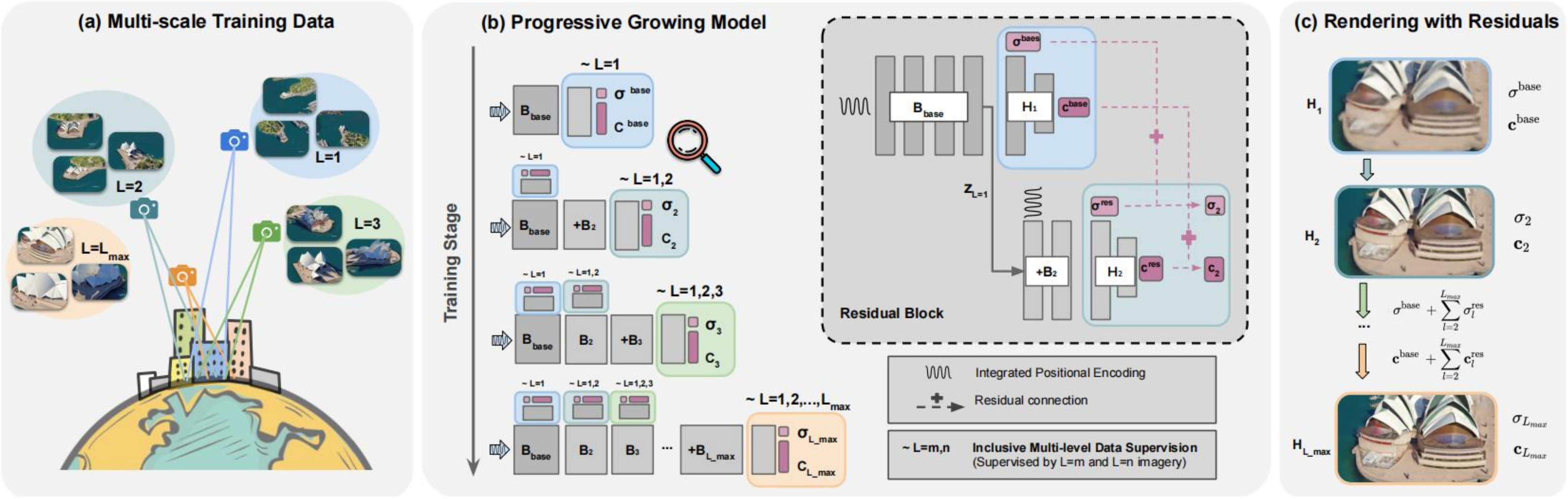
- **Solution:**
  - Replace ray tracing with cone traci
  - Replace positional encoding with integrated positional encoding (IPE





Barron, Jonathan T., et al. "Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields." . CVPR 2021..

BungeeNeRF (CityNeRF)



- **Goal:**
  - Pack extreme multi-scale city scenes into a unified model
- **Solution:**
  - Adopt a progressive neural radiance field
  - Grow model with residual block structure + Inclusive multi-level data supervision

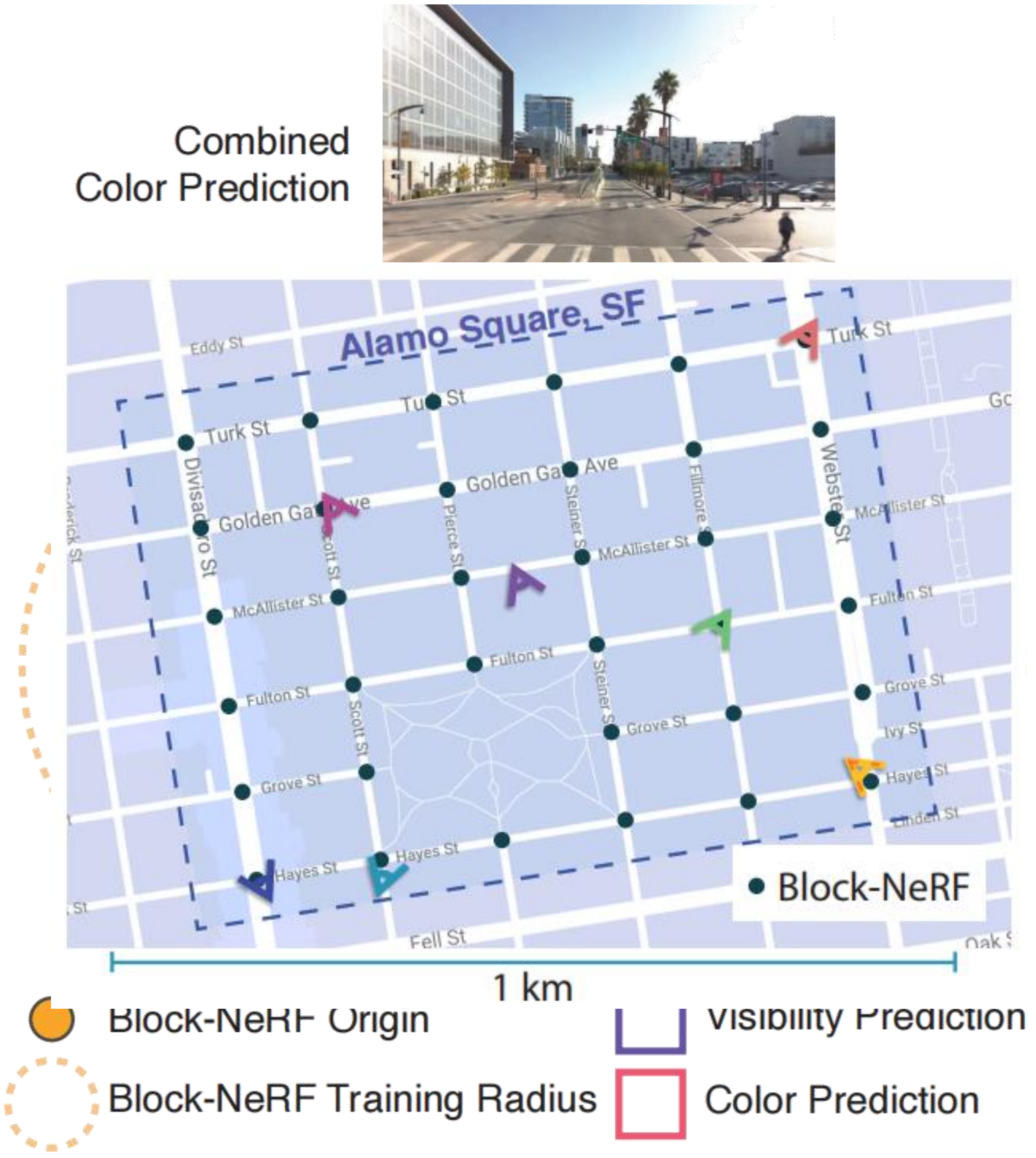
Xiangli, Yuanbo et al. "BungeeNeRF: Progressive Neural Radiance Field for Extreme Multi-scale Scene Rendering" . ECCV 2022..

BungeeNeRF (CityNeRF)

Block-NeRF

- **Goal:**
  - Enable neural radiance fields for large-scale environments

- **Solution:**
  - Dividing large environments into individually trained Block-NeRFs

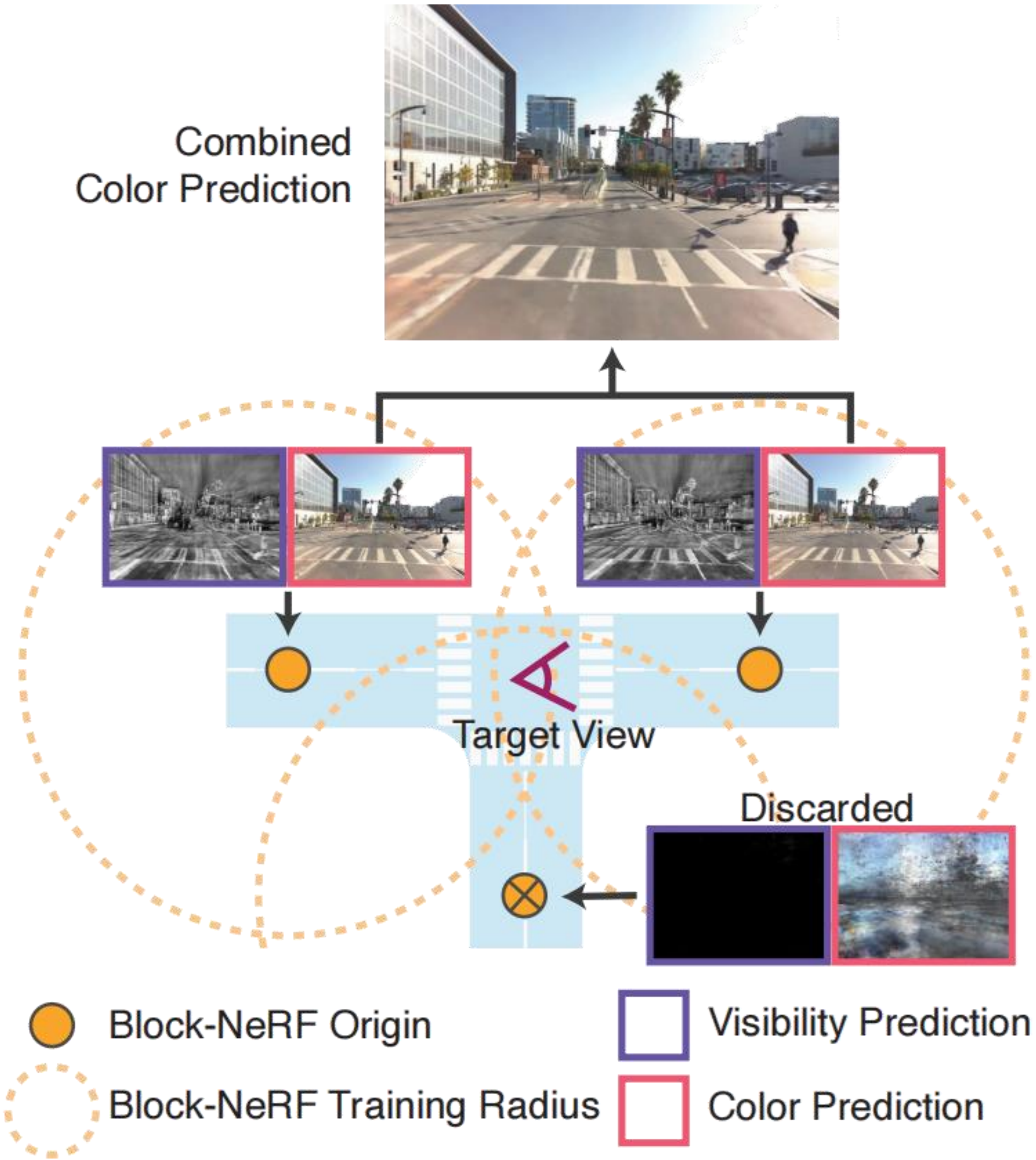


Tancik, Matthew, et al. "Block-nerf: Scalable large scene neural view synthesis." . CVPR 2022..

Block-NeRF

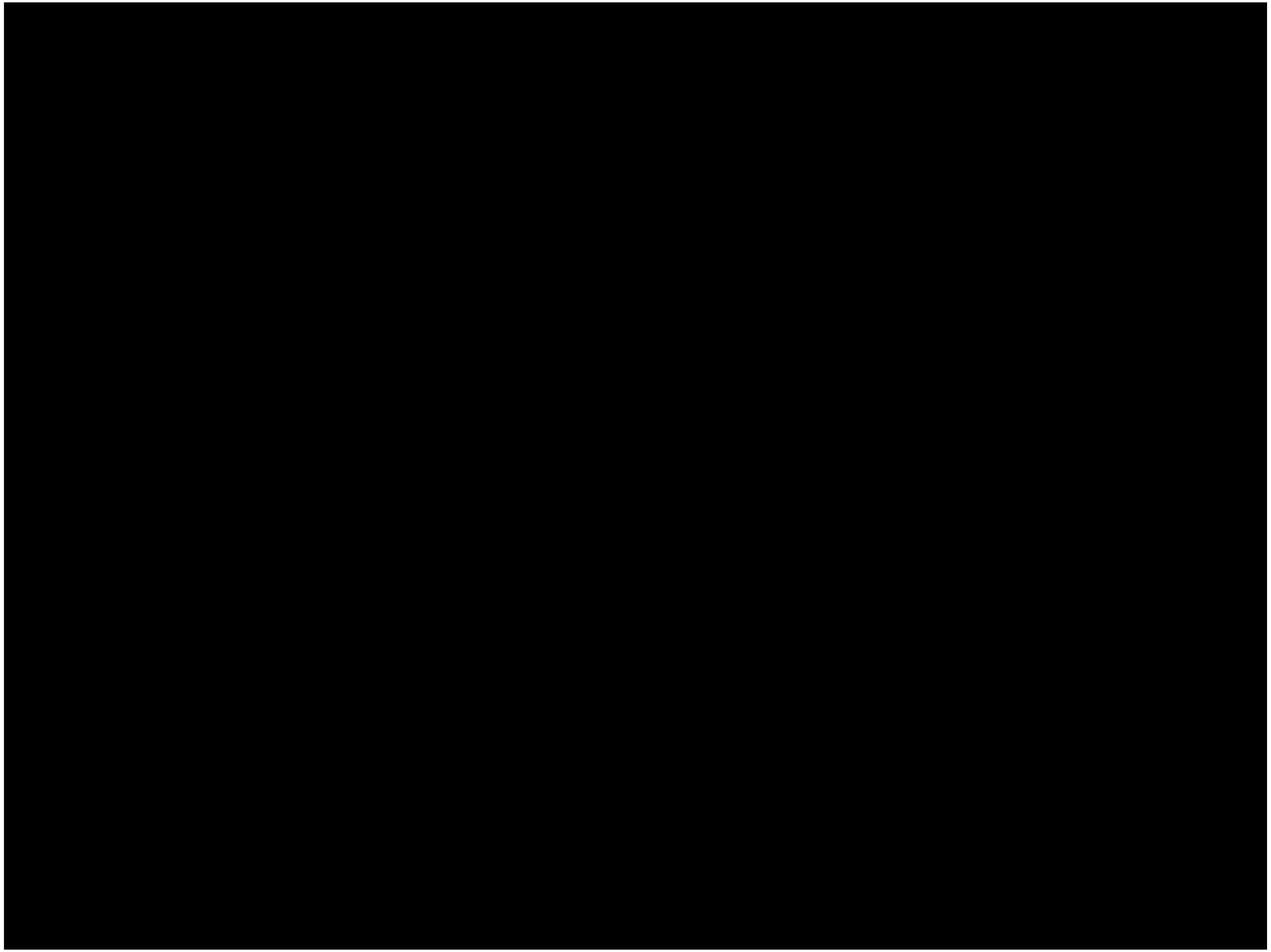
- **Goal:**
  - Enable neural radiance fields for large-scale environments

- **Solution:**
  - Dividing large environments into individually trained Block-NeRFs
  - Culling Block-NeRFs using the visibility network that predicts whether a point in space was visible in the training views



Tancik, Matthew, et al. "Block-nerf: Scalable large scene neural view synthesis." . CVPR 2022..

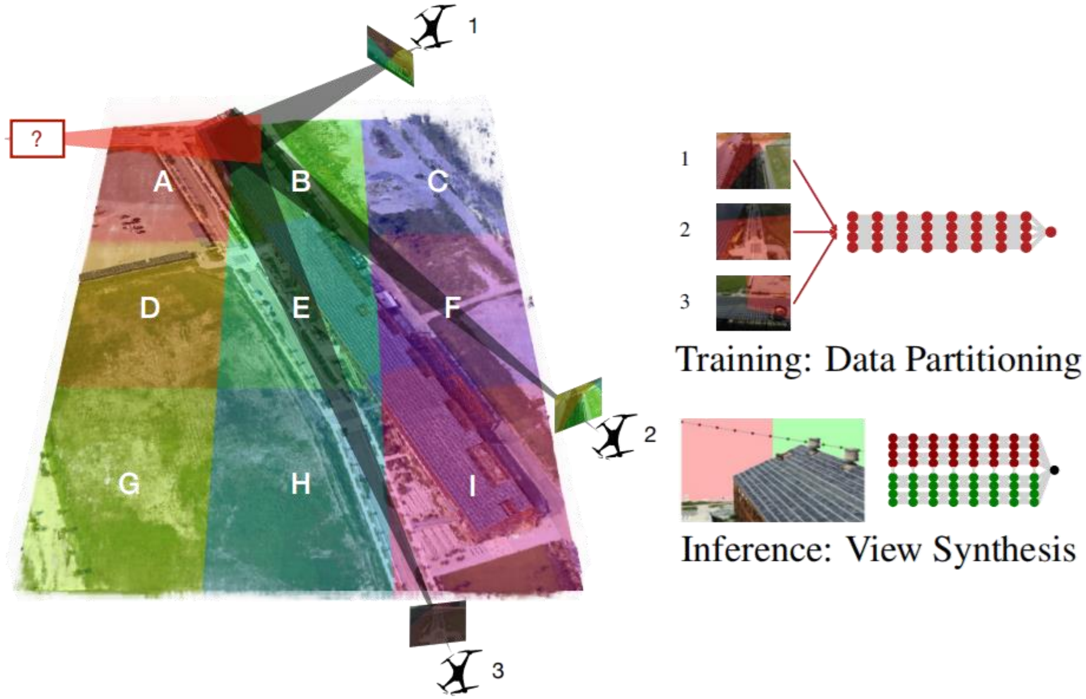
Block-NeRF



Tancik, Matthew, et al. "Block-nerf: Scalable large scene neural view synthesis." . CVPR 2022..

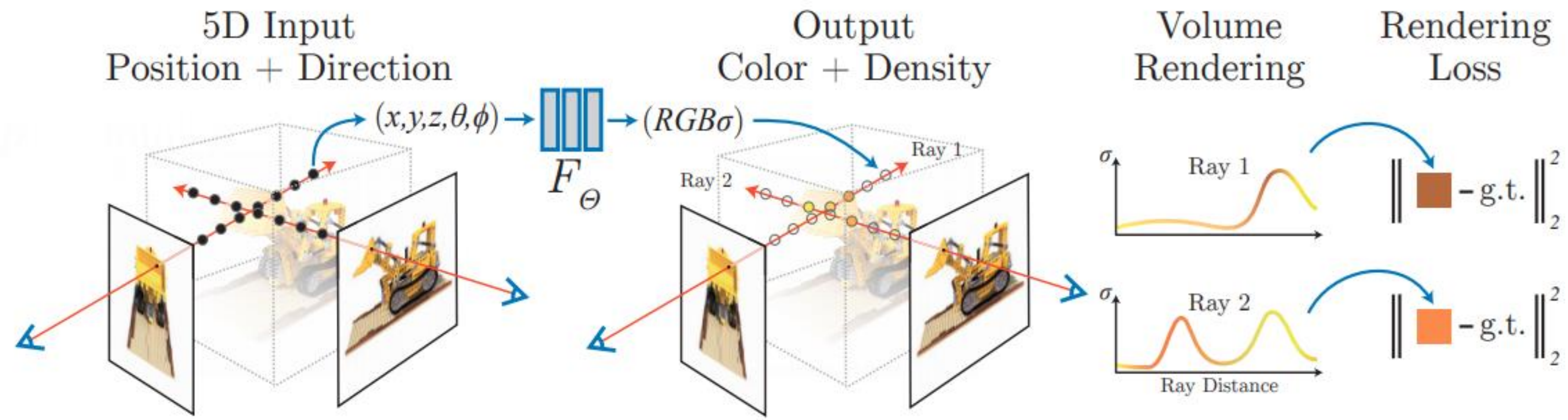
Mega-NeRF

- **Goal:**
  - Train large-scale 3D scenes efficiently
- **Solution:**
  - Exploit spatial locality and train the model subweights in a fully parallelizable manner



Turki, Haithem, et al.. "Mega-NeRF: Scalable Construction of Large-Scale NeRFs for Virtual Fly-Throughs." . CVPR 2022..

Take-home Message

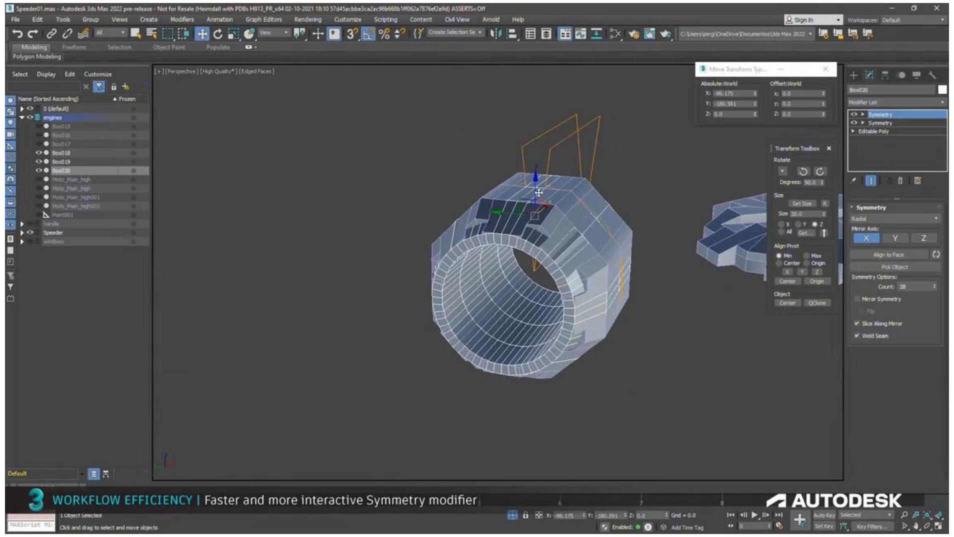


- × Unbounded scenes → Different spatial parameterizations
- × Dynamic objects → Unsupervised decomposition
- × Multi-scale images → Advance tracing and learning strategy
- × Limited network capabilities → Divide-and-conquer

3D Editing with Neural Radiance Fields

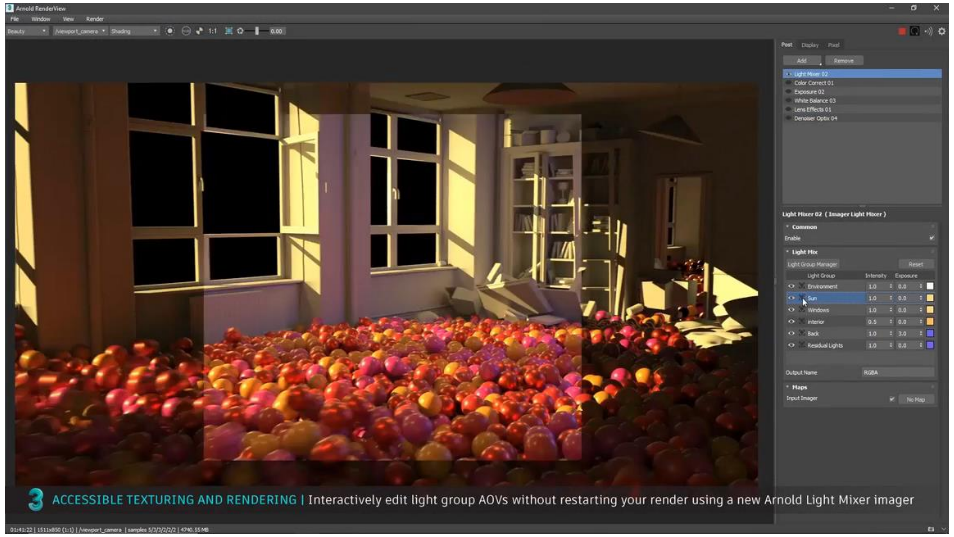
“Editing”

What does CG creation needs?

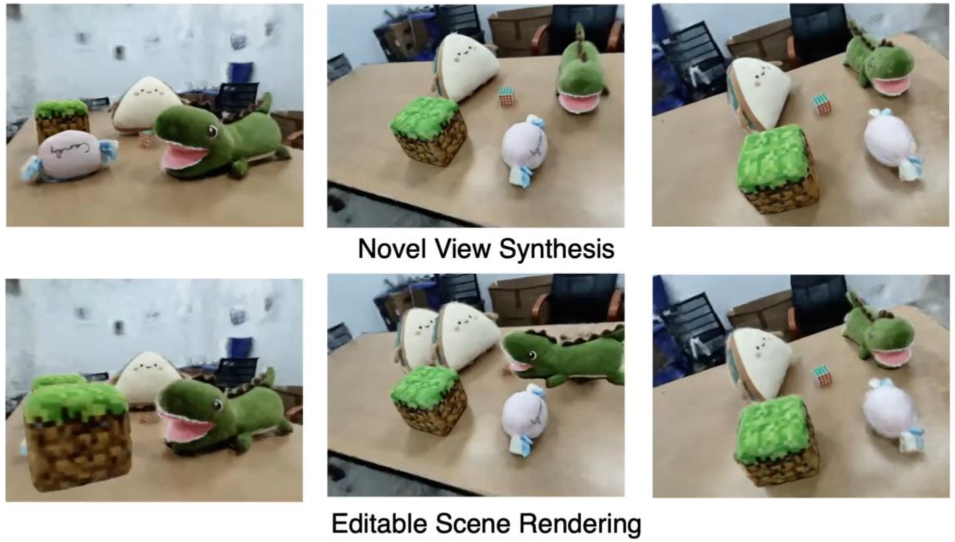


Geometry Editing

Texture Editing

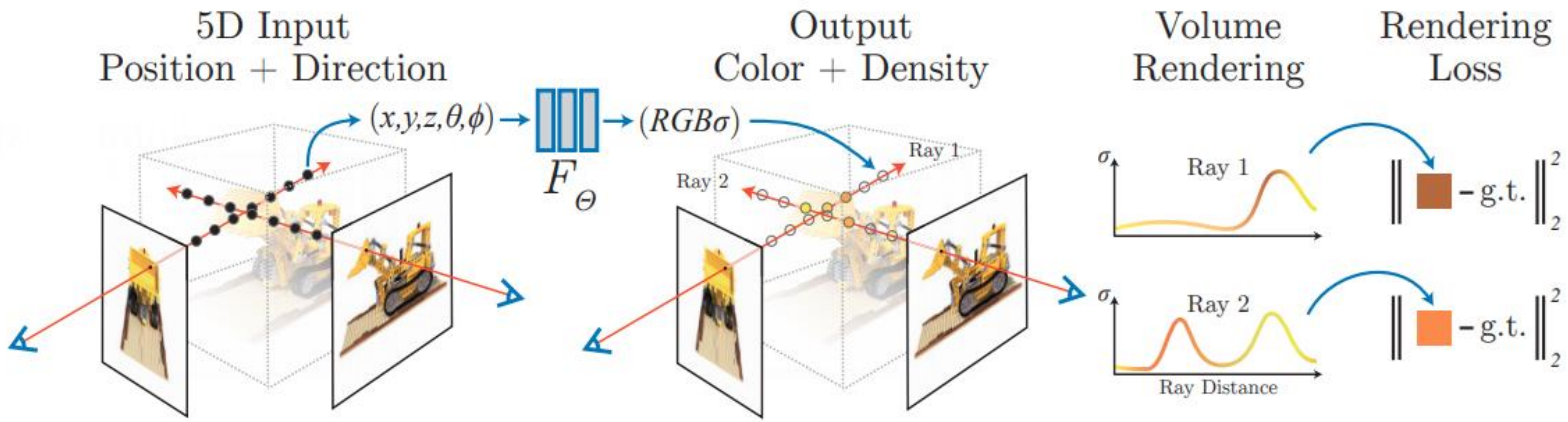


Light Editing



Scene Editing

Challenges for Editing with NeRF



× Implicit representation
× The scene is represented as a whole
× Everything is entangled within a network

Geometry Editing

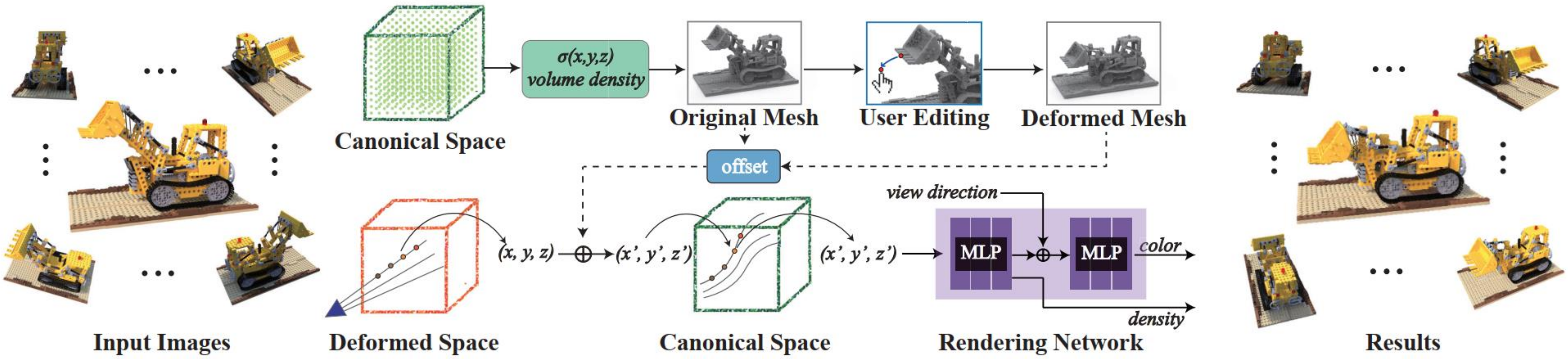
Naïve Solution: Decompose the implicit field into an explicit representation for flexible user editing

Challenge: Propagate deformation from the explicit representation back to implicit field

Key Idea: Out-of-the-box mesh-based deformation algorithm; discretize deformation into 3D space

- Nerf-Editing: use tetrahedralization method to discretize deformation
- Deforming Radiance Fields with Cages: use coarse bounding cages generation to discretize deformation

NeRF-Editing



- Explicit mesh as the editing interface
- Tetrahedralization of explicit mesh to transfer deformation to discrete volume (tetrahedron)
- Tetrahedra-based interpolation to deform radiance field

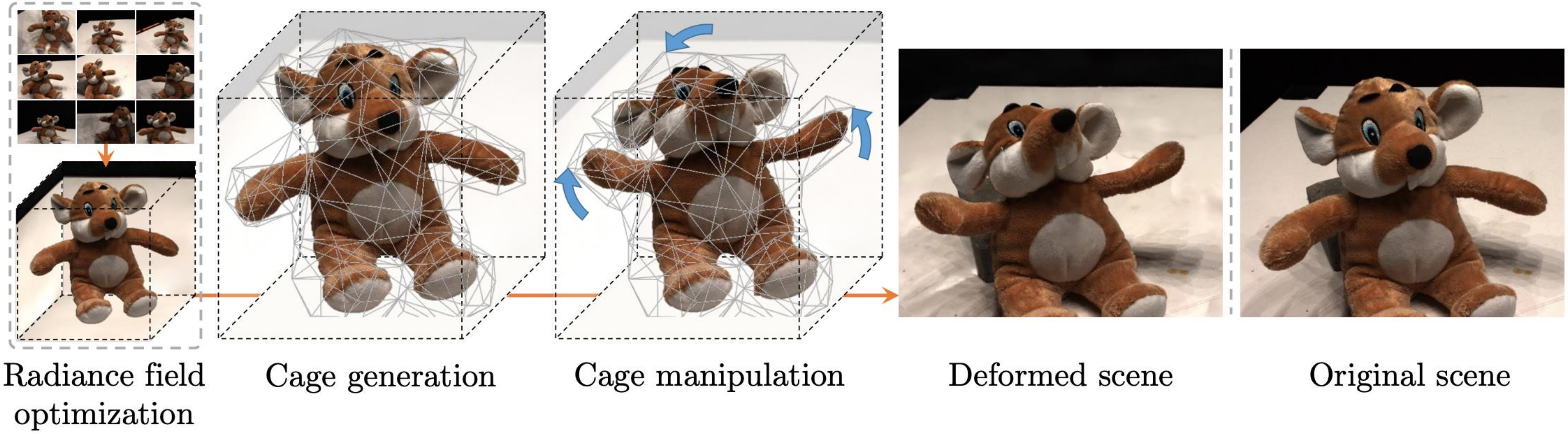
Yuan, Yu-Jie, et al. "NeRF-editing: geometry editing of neural radiance fields." CVPR. 2022.

NeRF-Editing

More results

Yuan, Yu-Jie, et al. "NeRF-editing: geometry editing of neural radiance fields." CVPR. 2022.

Deforming Radiance Fields with Cages



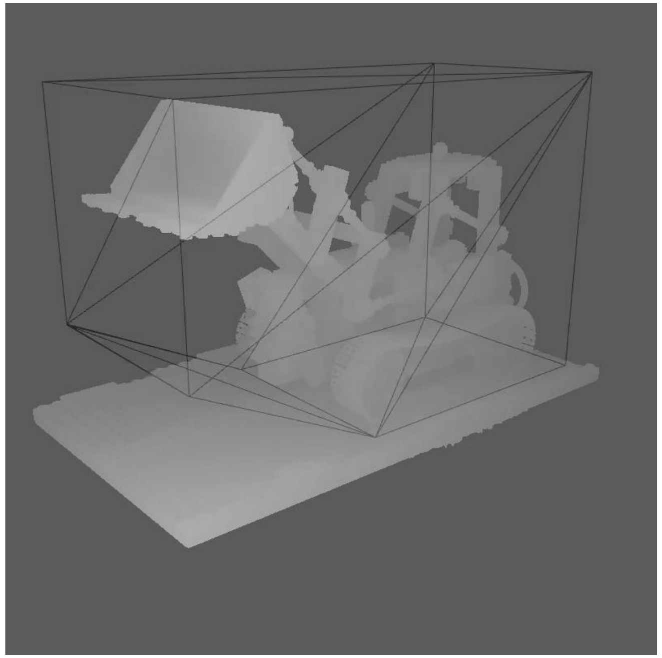
- A cage (coarse triangular mesh) as the editing interface with off-the-shelf coarse bounding cages generation
- Cage-based interpolation to deform radiance field

Xu, Tianhan, and Tatsuya Harada. "Deforming Radiance Fields with Cages." ECCV 2022

Deforming Radiance Fields with Cages

Cage

Movement



NVS

Xu, Tianhan, and Tatsuya Harada. “Deforming Radiance Fields with Cages.” ECCV 2022

Texture Editing

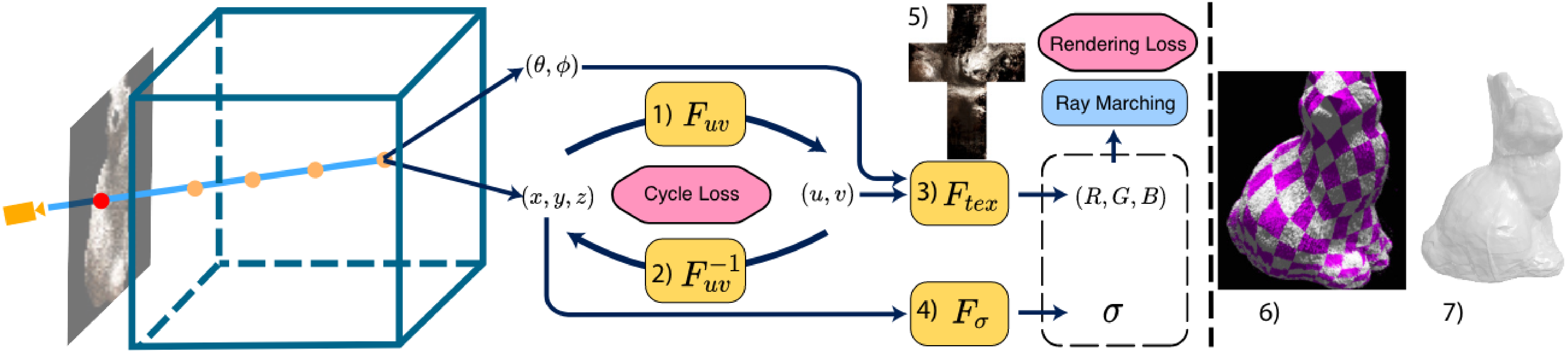
Naïve Solution: Fine-tune the color head of Nerf

Challenge: 2D texture editing lacking of view-consistency or a flexible representation for 3D texture editing

Key Idea: Disentangle geometry and texture; a flexible texture optimization strategy

- NeuTex: texture mapping network to represent texture
- Editing Conditional Radiance Fields: radiance field conditioned by shape and texture code
- NeuMesh: learn local latent features attached to raw mesh vertices

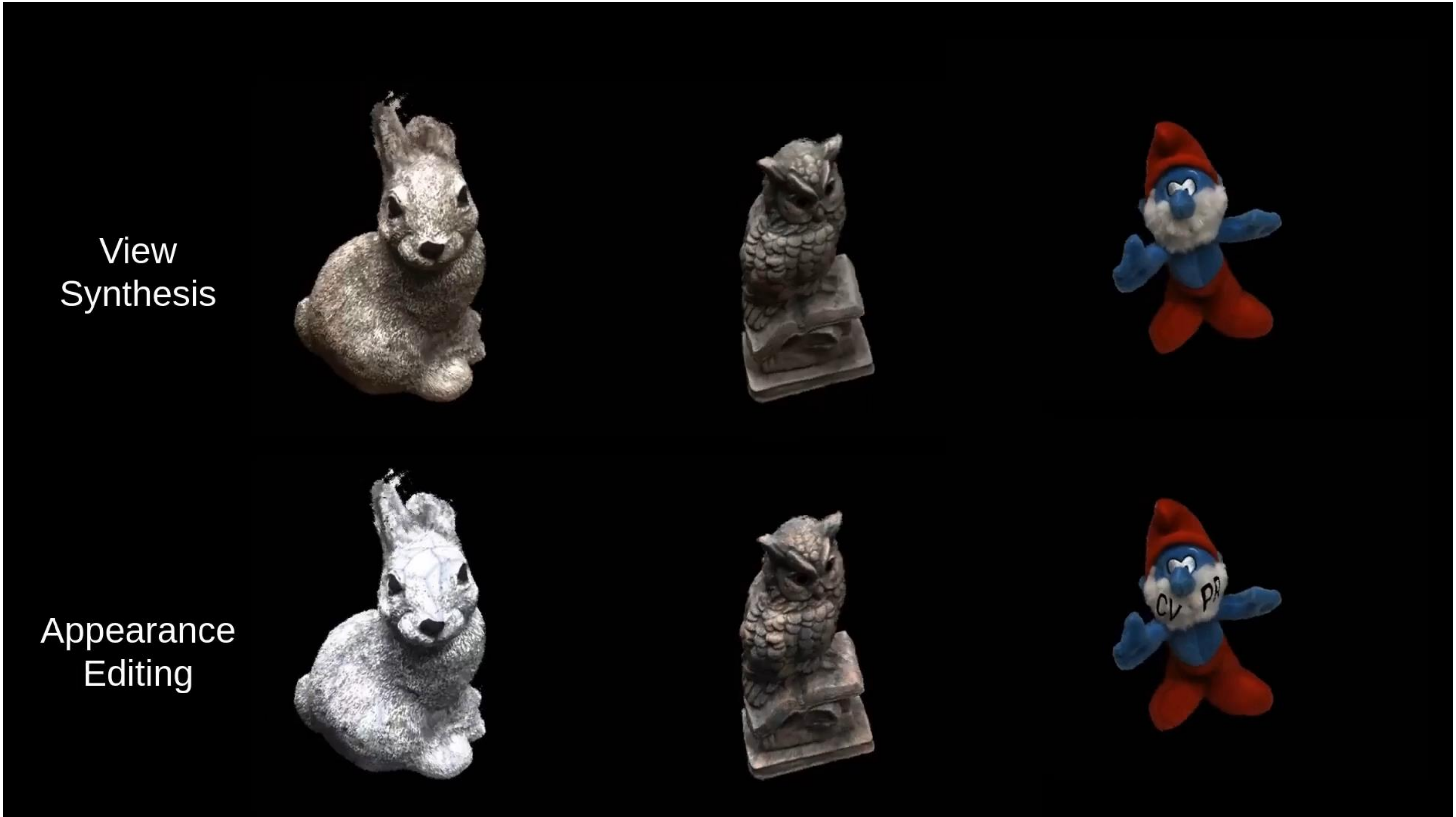
NeuTex



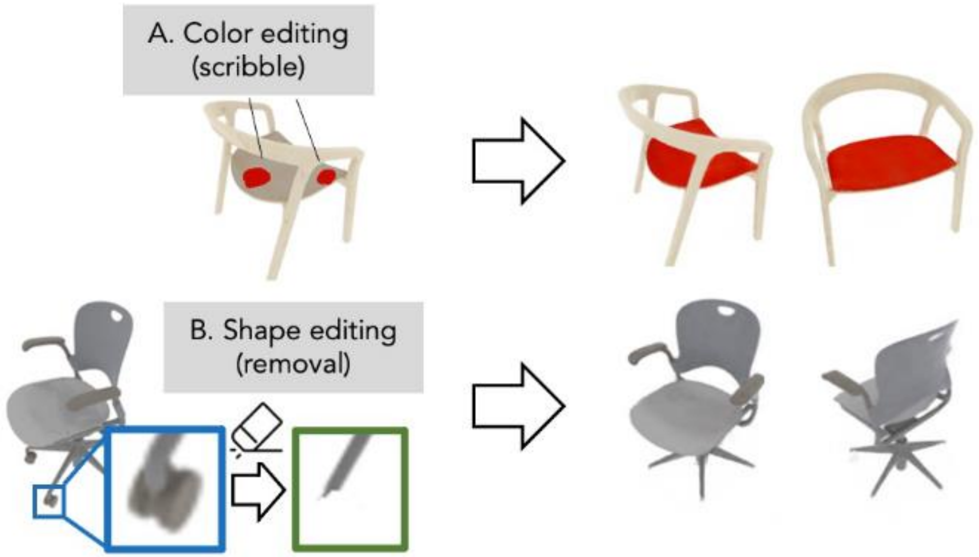
- Disentangle geometry as a continuous 3D volume F_{σ} and appearance as a continuous 2D texture map F_{tex}
- Introduce a 3D-to-2D texture mapping network F_{uv} into volumetric representations.
- Constrain this texture mapping using an additional inverse mapping network F_{uv}^{-1} and a novel cycle consistency loss

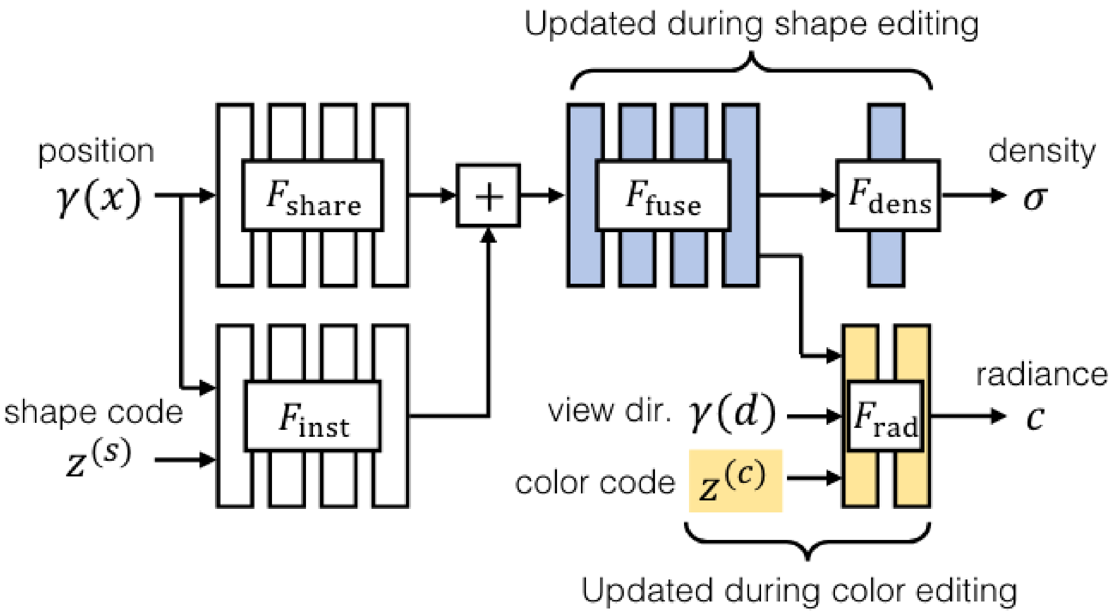
Xiang, Fanbo, et al. "Neutex: Neural texture mapping for volumetric neural rendering." CVPR. 2021.

NeuTex



Editing Conditional Radiance Fields





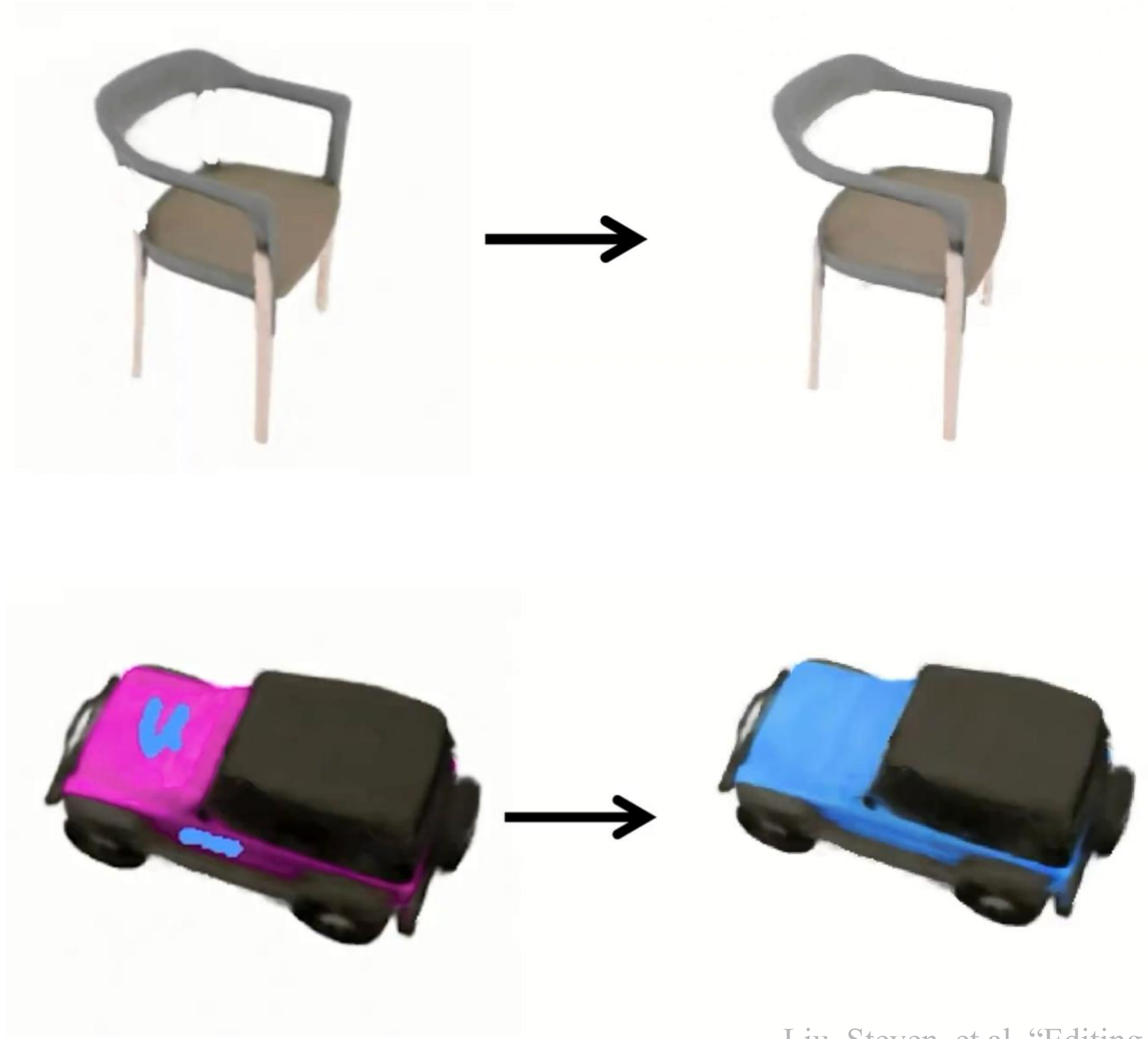
- Conditional radiance field including a shape branch that is shared across object instances
- Hybrid network update strategy balancing efficiency and accuracy: update F_{fuse} & F_{dens} during shape editing and update $Z^{(c)}$ & F_{rad} for color editing

Liu, Steven, et al. "Editing conditional radiance fields." ICCV 2021.

Editing Conditional Radiance Fields

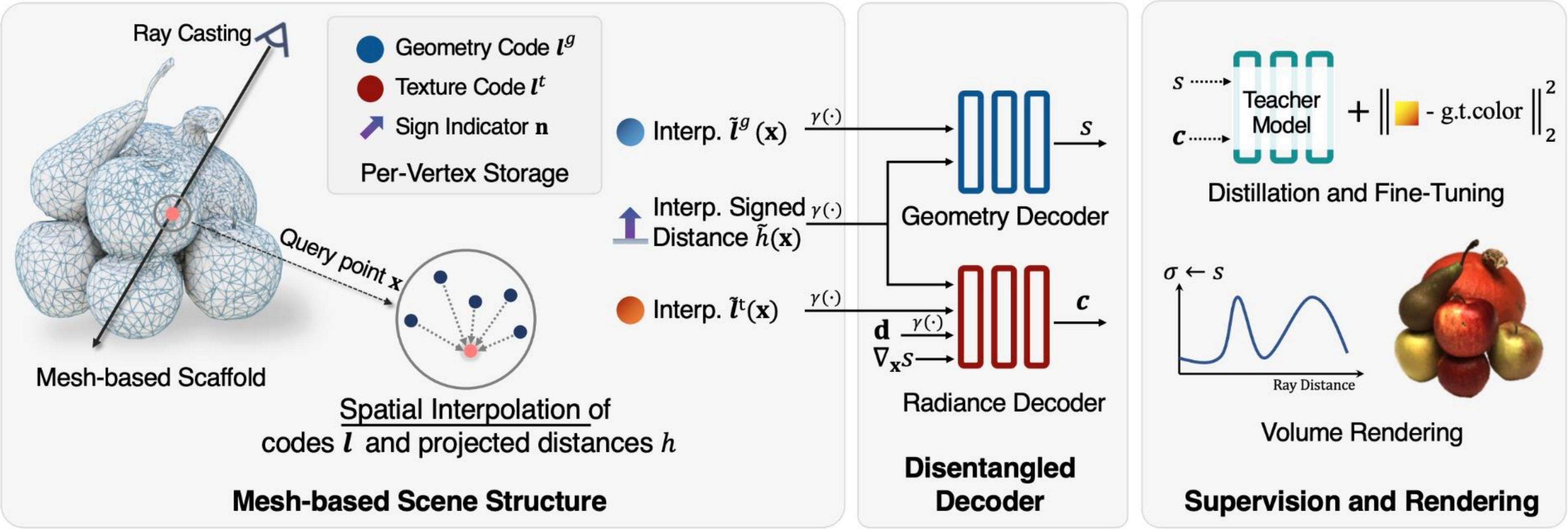
Geometry Edit

Texture Edit



Liu, Steven, et al. "Editing conditional radiance fields." ICCV 2021.

NeuMesh



- The mesh-based neural implicit field whose vertex possesses a geometry and texture code
- Support various editing functionalities: geometry deformation, texture swapping, texture filling, texture painting

Yang, Bangbang, et al. "NeuMesh: Learning Disentangled Neural Mesh-based Implicit Field for Geometry and Texture Editing." ECCV(2022).

NeuMesh: geometry deformation

- Simply deform the corresponding mesh to synchronously take effect on the implicit field, which is aligned to the mesh surface.

Yang, Bangbang, et al. "NeuMesh: Learning Disentangled Neural Mesh-based Implicit Field for Geometry and Texture Editing." ECCV(2022).

NeuMesh: texture swapping

➢ Transfer the texture from the red area to the yellow area according to user-selected vertices by swapping texture code in 3D space

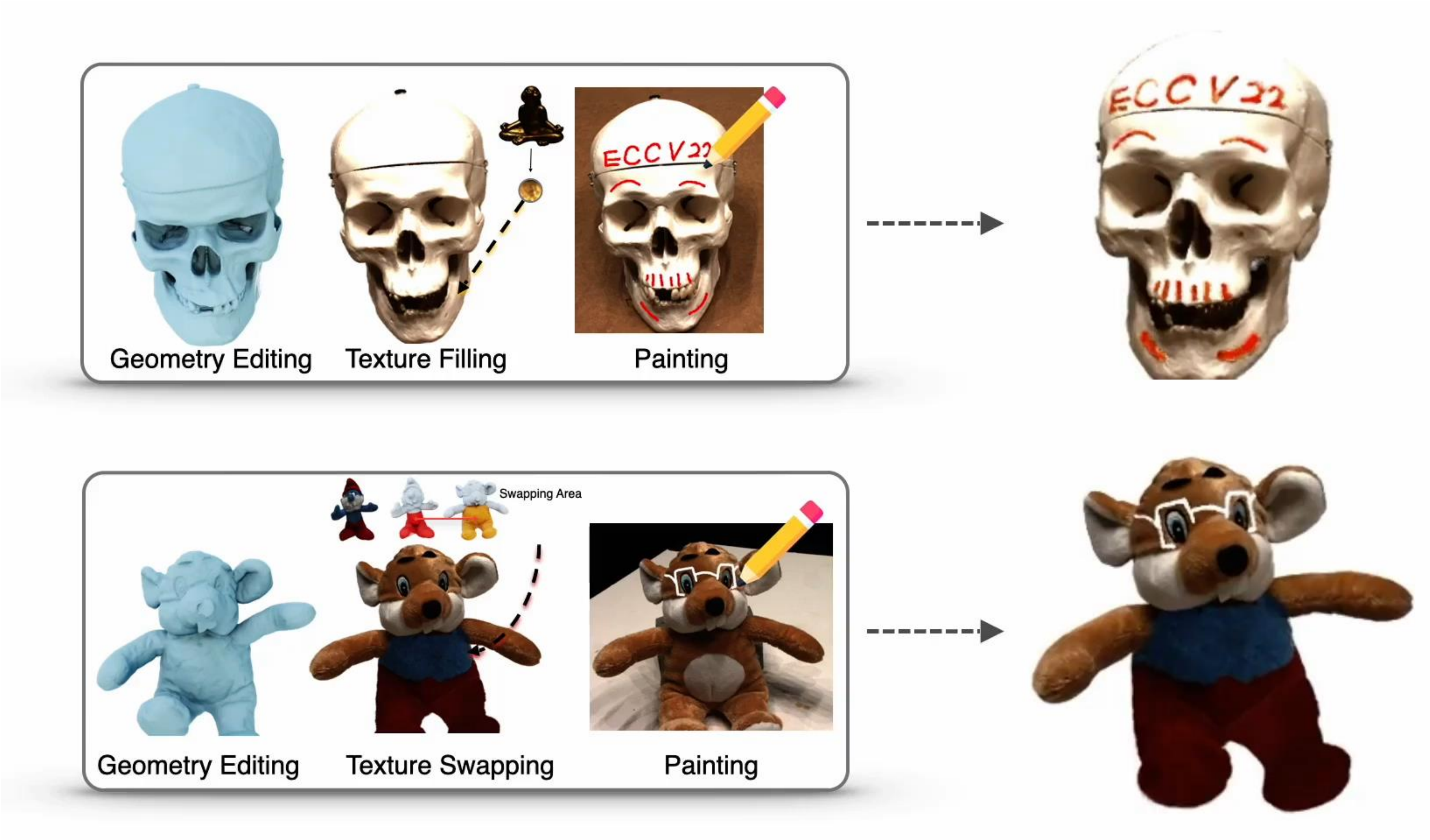
Yang, Bangbang, et al. "NeuMesh: Learning Disentangled Neural Mesh-based Implicit Field for Geometry and Texture Editing." ECCV(2022).

NeuMesh: texture filling

➤ Transfer painting from a single 2D image to the neural implicit field with proposed spatial-aware optimization

Yang, Bangbang, et al. "NeuMesh: Learning Disentangled Neural Mesh-based Implicit Field for Geometry and Texture Editing." ECCV(2022).

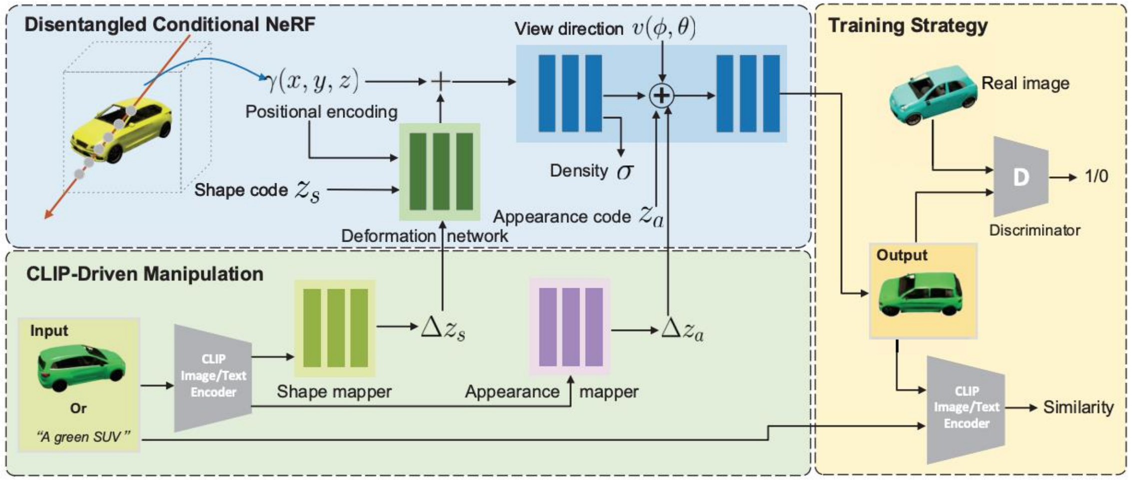
NeuMesh: Hybird Editing



Yang, Bangbang, et al. "NeuMesh: Learning Disentangled Neural Mesh-based Implicit Field for Geometry and Texture Editing." ECCV(2022).

New Editing Functionality: Text

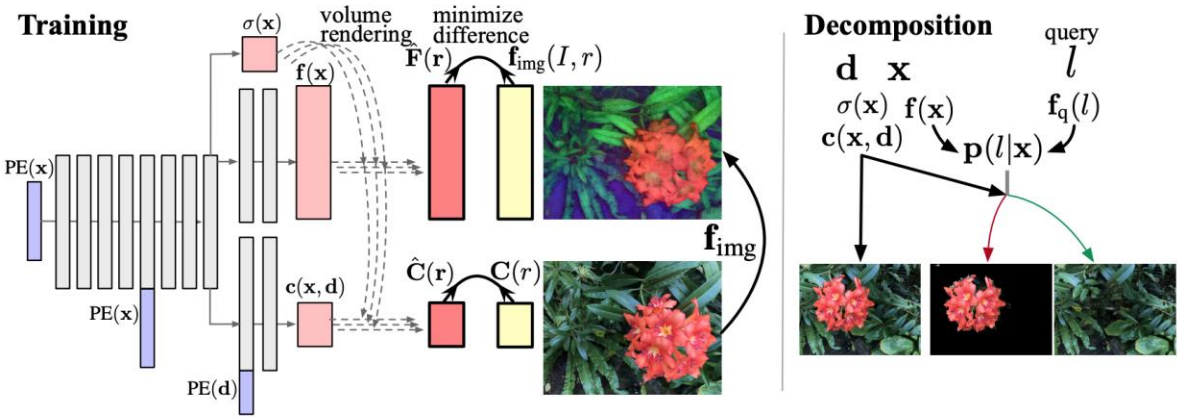
ClipNeRF



- Take a CLIP embedding as editing input with two code mappers and update the conditional Nerf to reflect the targeted editing.

Wang, Can, et al. "Clip-nerf: Text-and-image driven manipulation of neural radiance fields." CVPR2022.

Decomposing NeRF for Editing via Feature Field Distillation

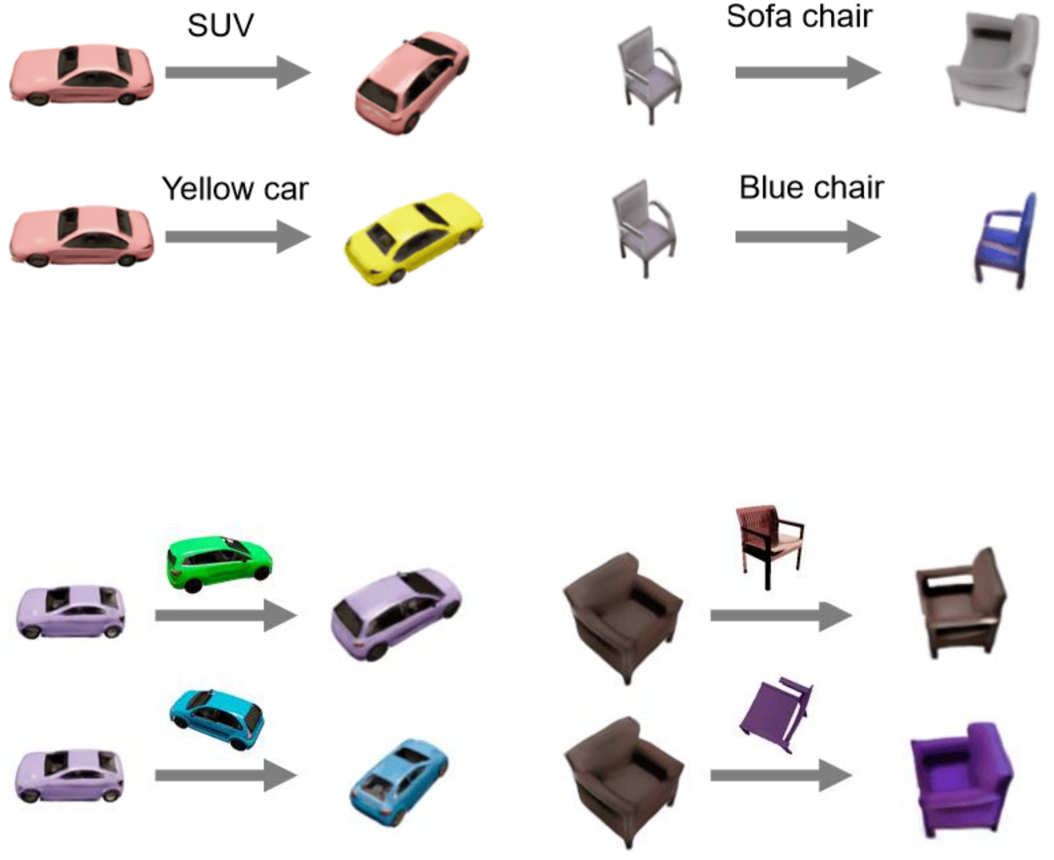


- Distill feature field from CLIP-variant model with a vanilla NeRF

Kobayashi, Sosuke, et al"Decomposing NeRF for Editing via Feature Field Distillation." arXiv 2022.

New Editing Functionality: Text

ClipNeRF



Decomposing NeRF for Editing via Feature Field Distillation

Wang, Can, et al. "Clip-nerf: Text-and-image driven manipulation of neural radiance fields." CVPR2022.

Kobayashi, Sosuke, et al"Decomposing NeRF for Editing via Feature Field Distillation." arXiv 2022.

Light Editing

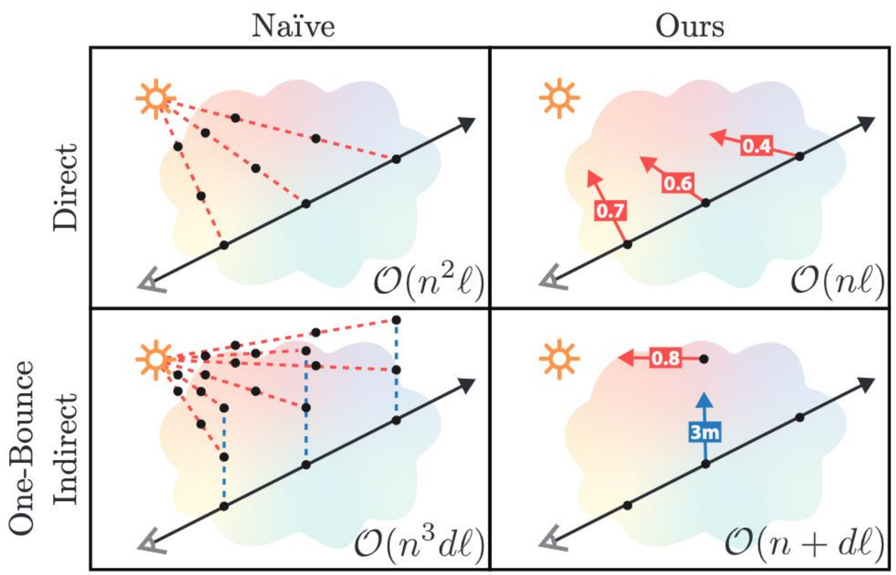
Observation: NeRF mixes the environment light effect, BRDF into color field

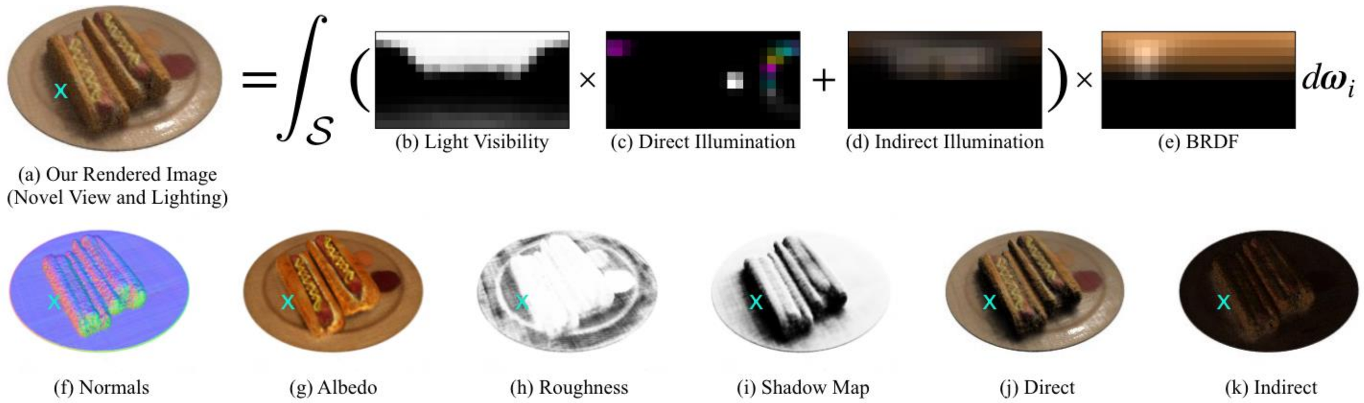
Challenge: illumination estimation, BRDF estimation, light condition of data

Key Idea: disentangle geometry, BRDF, environment light effect

- Light Estimation: spherical gaussian(NeRD), pre-baked visibility network+ an HDR light probe representation(NeRFactor)
- BRDF estimation: implicit BRDF network(NeRV) , knowledge BRDF encoder/decoder (NeRFactor, NeRD)

NeRV





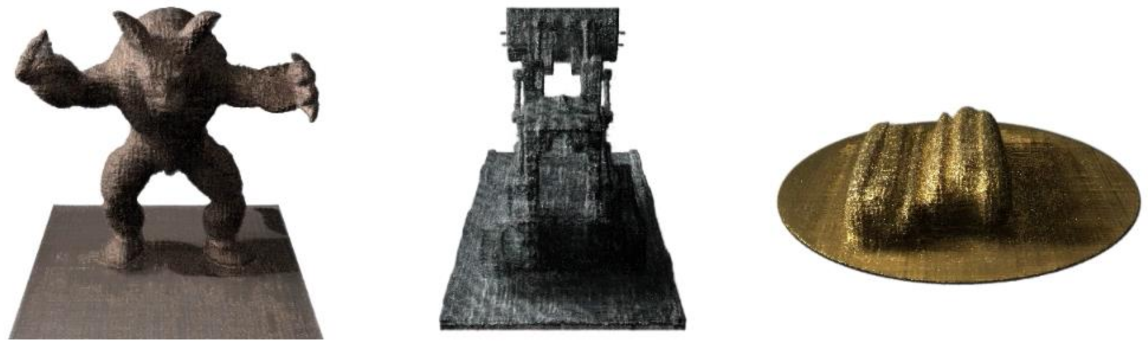
- Assume position of light source is known and use visibility network to memory the visibility of environment light
- Bounce ray once to collect the indirect light = the direct light of bounced sampled
- Volume rendering with light visibility, direct light, indirect light and BRDF (3D diffuse albedo a and 1D roughness γ from reflectance network)

Srinivasan, Pratul P., et al. "Nerv: Neural reflectance and visibility fields for relighting and view synthesis." CVPR. 2021.

NeRV

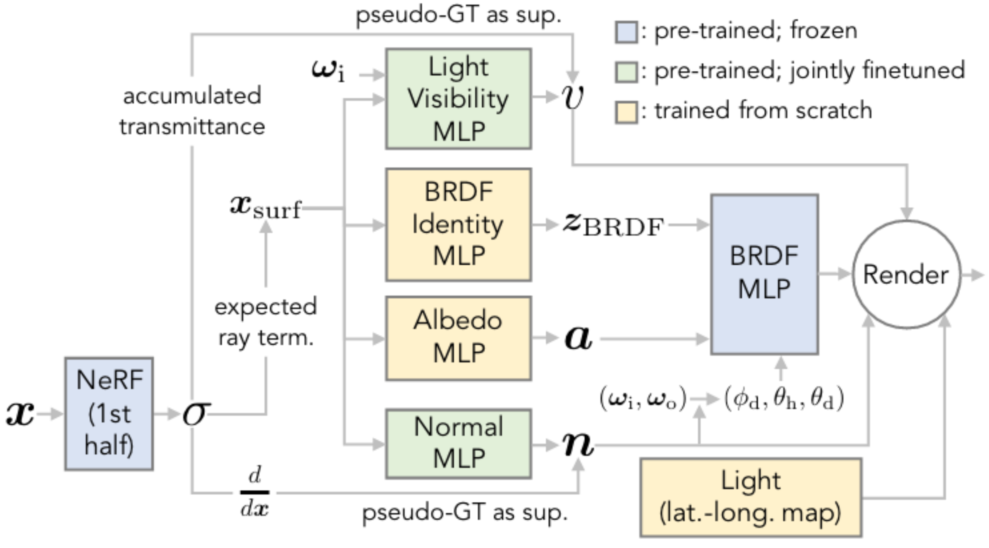
Relighting and View Synthesis

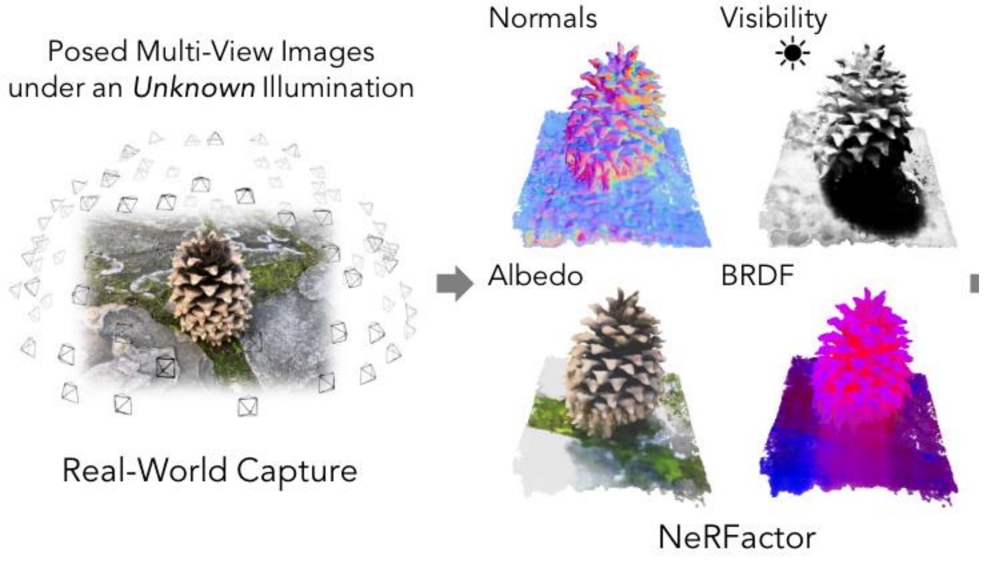
Material Editing



Srinivasan, Pratul P., et al. "Nerv: Neural reflectance and visibility fields for relighting and view synthesis." CVPR. 2021.

NeRFactor

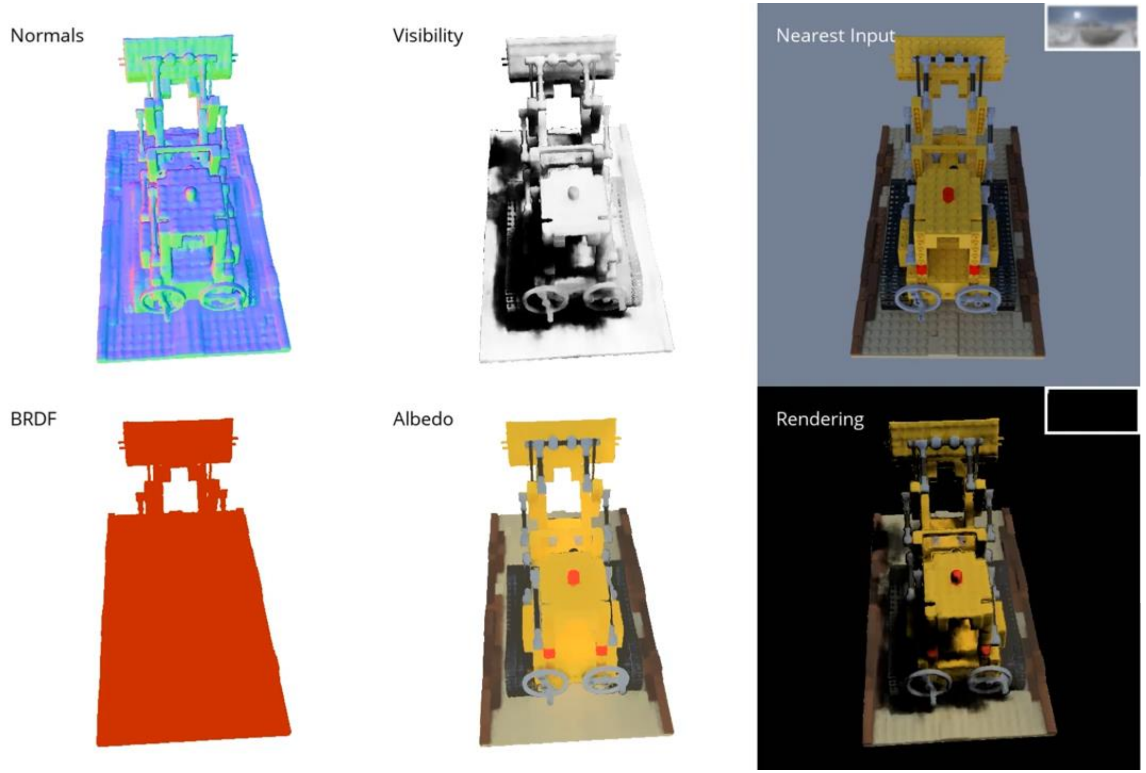




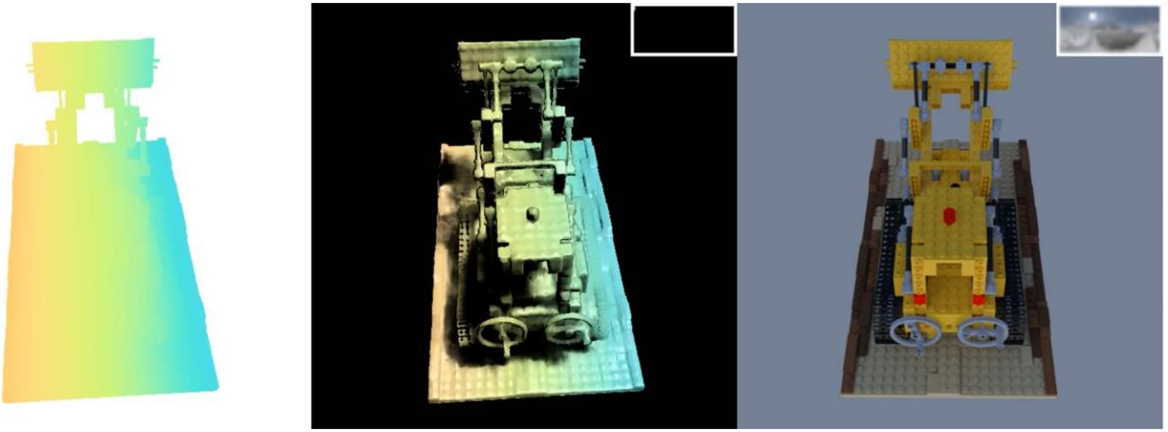
- ➢ Distill light visibility from geometry of pre-trained Nerf to model shadow
- ➢ BRDF estimation from a data-driven prior
- ➢ An HDR light probe representation to represent detailed high-frequency lighting

Zhang, Xiuming, et al. "Nerfactor: Neural factorization of shape and reflectance under an unknown illumination." TOG 40.6 (2021): 1-18.

NeRFactor



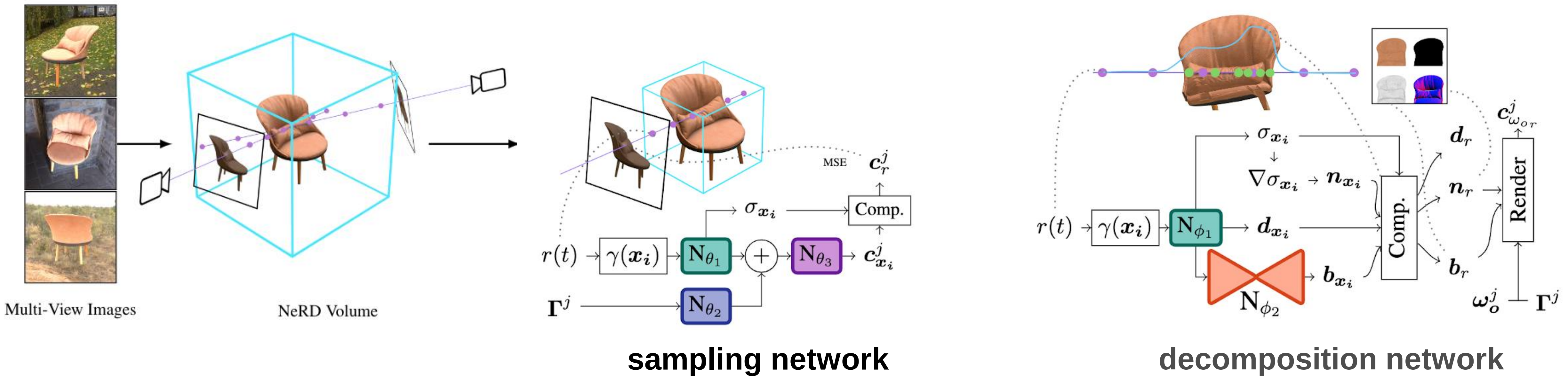
Relighting and View Synthesis



Material Editing

Zhang, Xiuming, et al. "Nerfactor: Neural factorization of shape and reflectance under an unknown illumination." TOG 40.6 (2021): 1-18.

NeRD



- ➢ Input data are collected under various light condition
- ➢ Sampling network as coarse model to learn the coarse geometry under various light condition
- ➢ Decomposition network as fine model to decompose fine geometry, direct color, BRDF of object

Boss, Mark, et al. "Nerd: Neural reflectance decomposition from image collections." ICCV. 2021.

NeRD

Boss, Mark, et al. "Nerd: Neural reflectance decomposition from image collections." ICCV. 2021.

Scene Editing

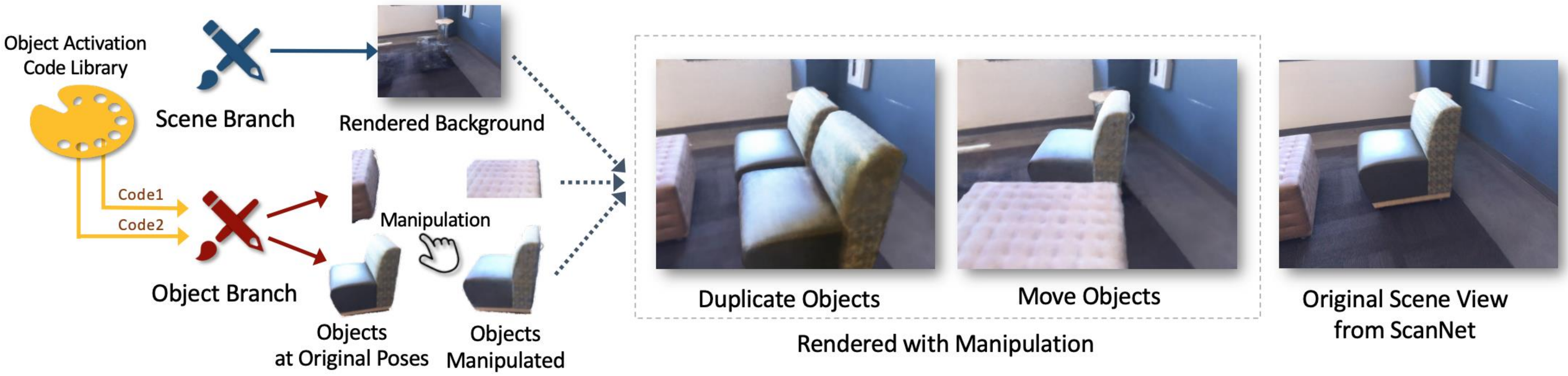
Naïve Solution: decompose scene into object level

Challenge: disentanglement of foreground and background

Key Idea: exploit explicit supervision or implicit prior to segment foreground

- Object segmentation: Object NeRF
- Implicit knowledge: UORF
- Depth: ST NeRF

Object NeRF



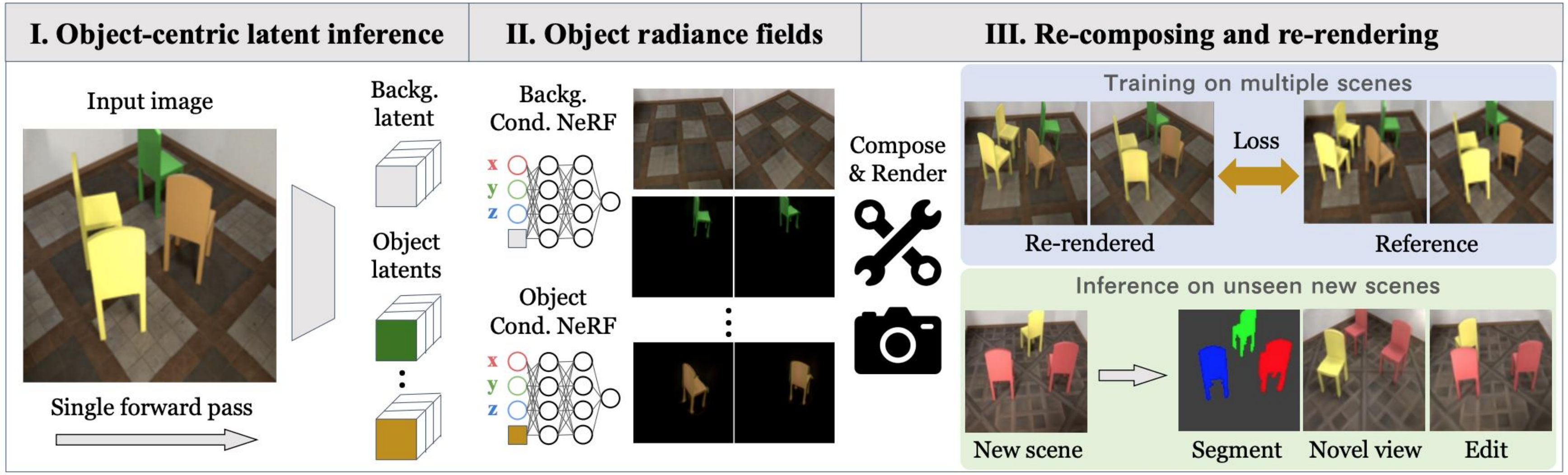
- A two-pathway architecture: scene branch (encodes the scene geometry & appearance), object branch(conditioned on learnable object activation codes)
- A scene-guided training strategy to solve the 3D space ambiguity in the occluded regions and learn sharp boundaries for each object

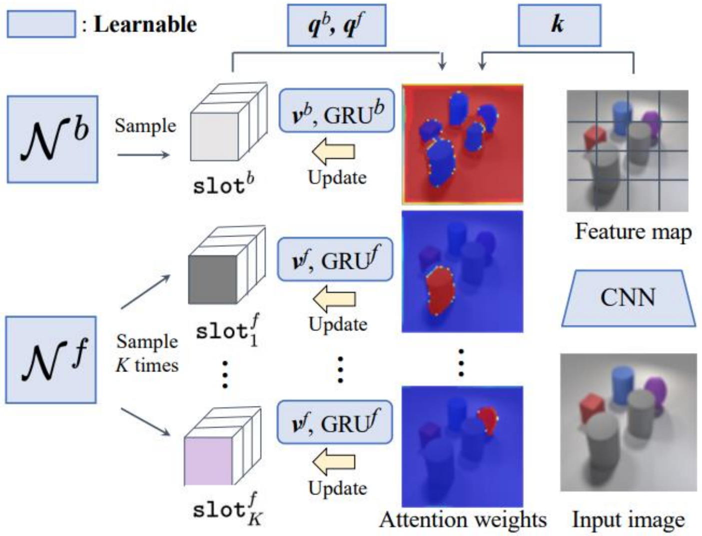
Yang, Bangbang, et al. "Learning object-compositional neural radiance field for editable scene rendering." ICCV. 2021.

Object NeRF

Yang, Bangbang, et al. "Learning object-compositional neural radiance field for editable scene rendering." ICCV. 2021.

uORF

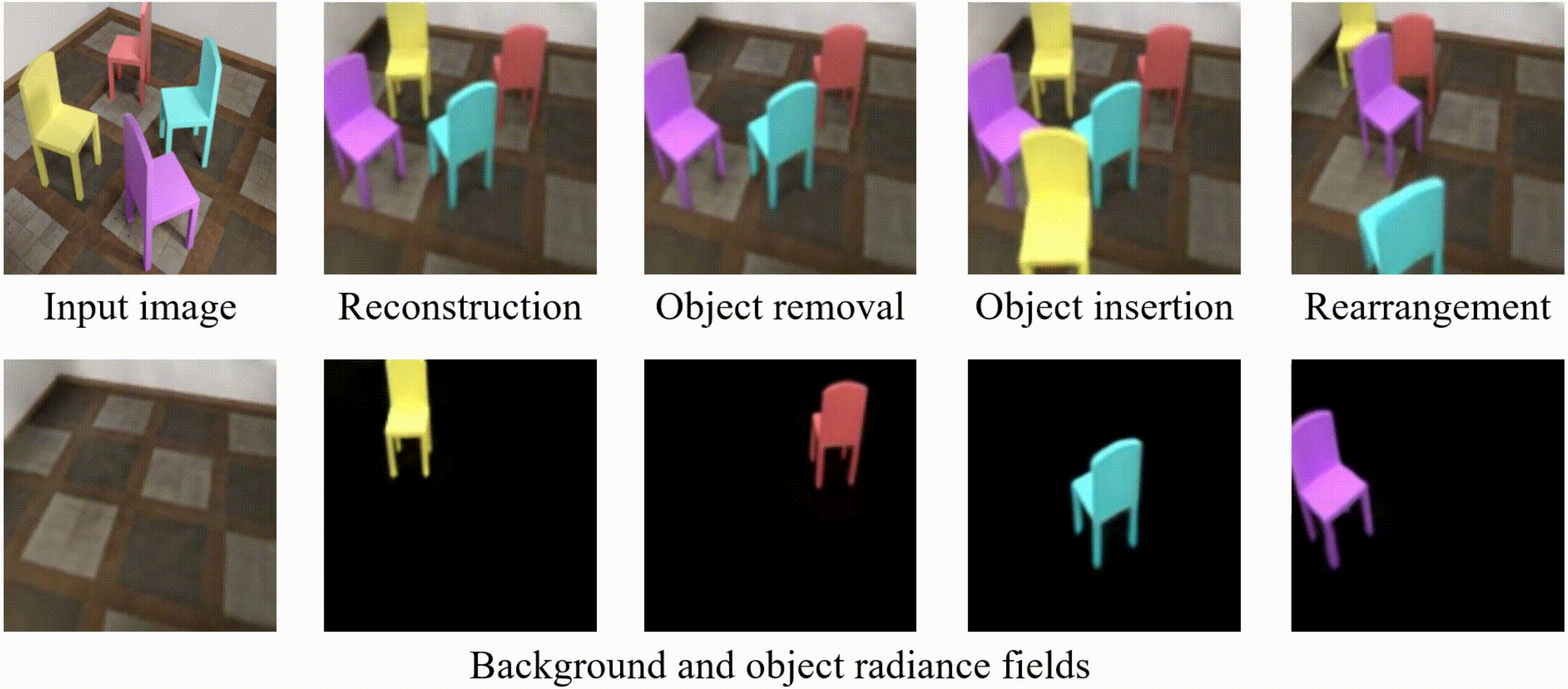




- ➢ Foreground and background decomposed design with a slot-based formulation
- ➢ Background-aware slot attention for sampling and binding to separately models objects and environment to better capture the compositional structure of 3D scenes.
- ➢ Each object slot is bound to an object region via an attention module

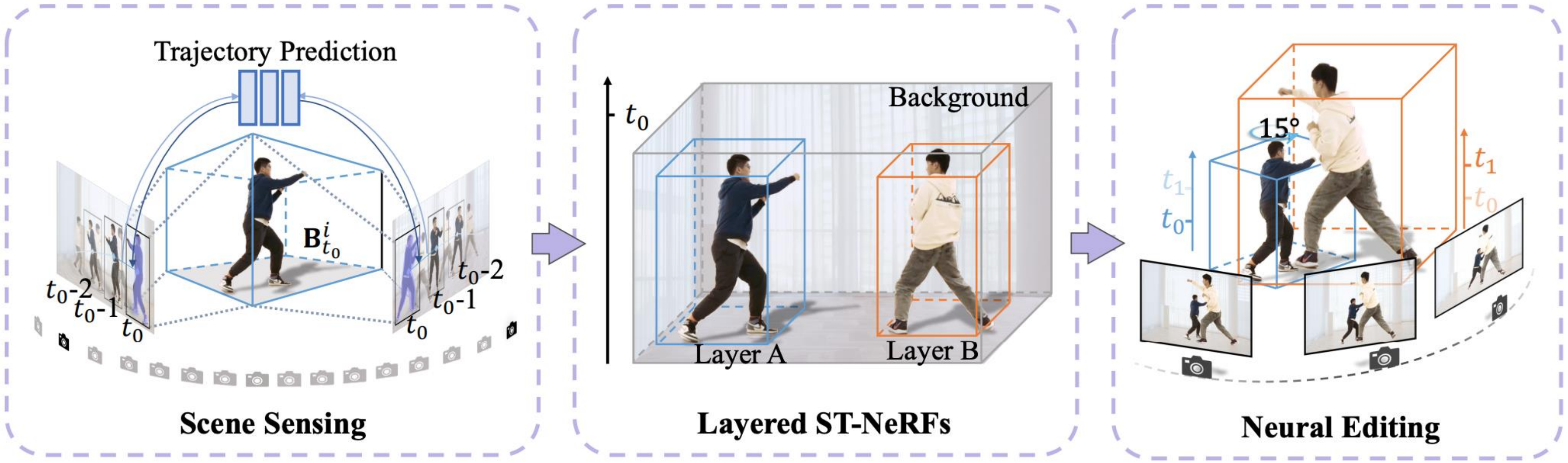
Yu, Hong-Xing, Leonidas J. Guibas, and Jiajun Wu. "Unsupervised discovery of object radiance fields."ICLR (2022).

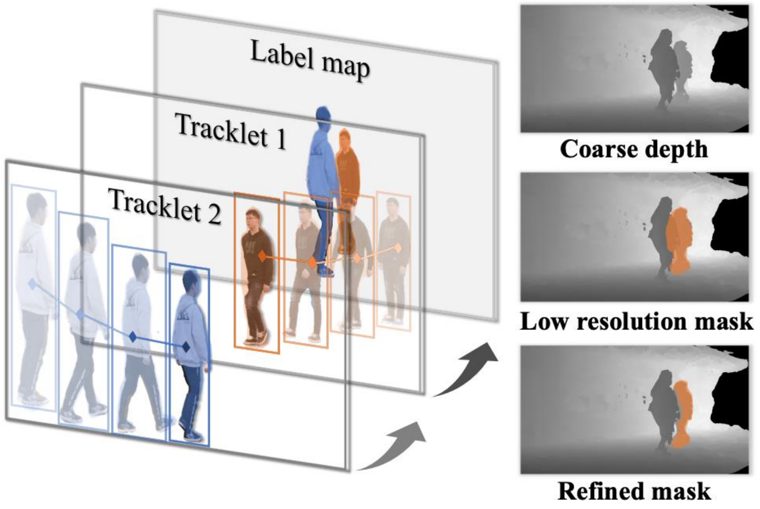
uORF



Yu, Hong-Xing, Leonidas J. Guibas, and Jiajun Wu. “Unsupervised discovery of object radiance fields.”ICLR (2022).

ST-NeRF



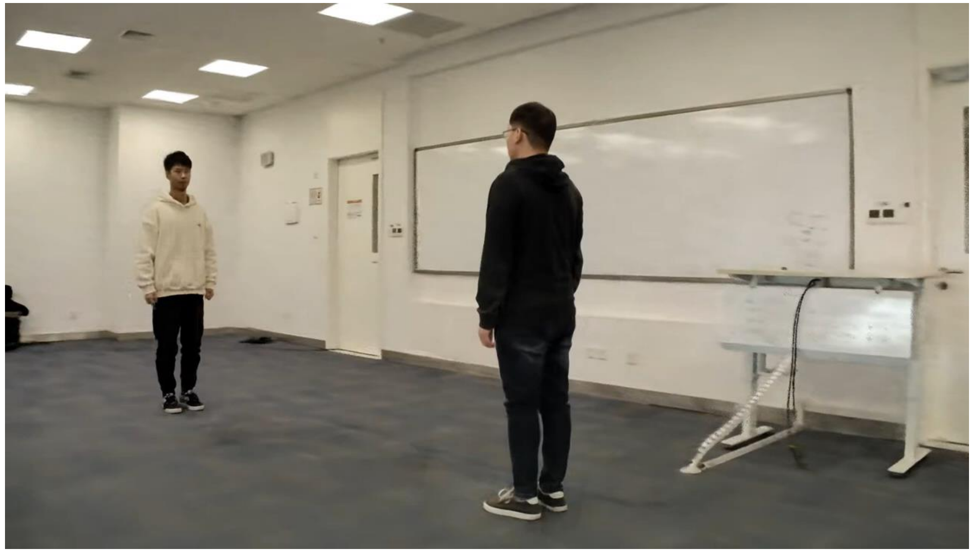


- A neural layered representation enabled by the disentanglement of location, deformation as well as the appearance of all the dynamic entities
- A layer-wise 4D label map tracking to disentangle the spatial information explicitly and a continuous deform module to disentangle the temporal motion implicitly

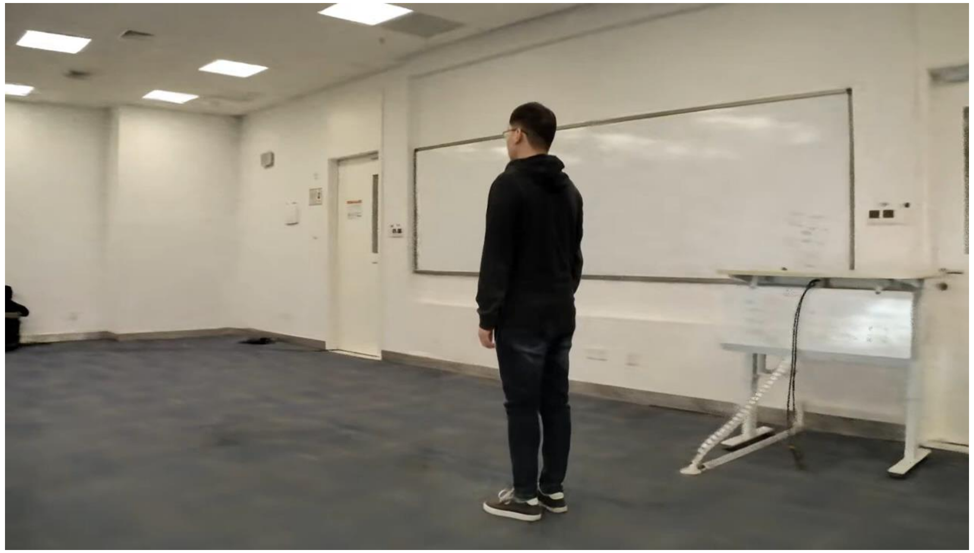
Zhang, Jiakai, et al. "Editable free-viewpoint video using a layered neural representation." TOG 40.4 (2021): 1-18.

ST-NeRF

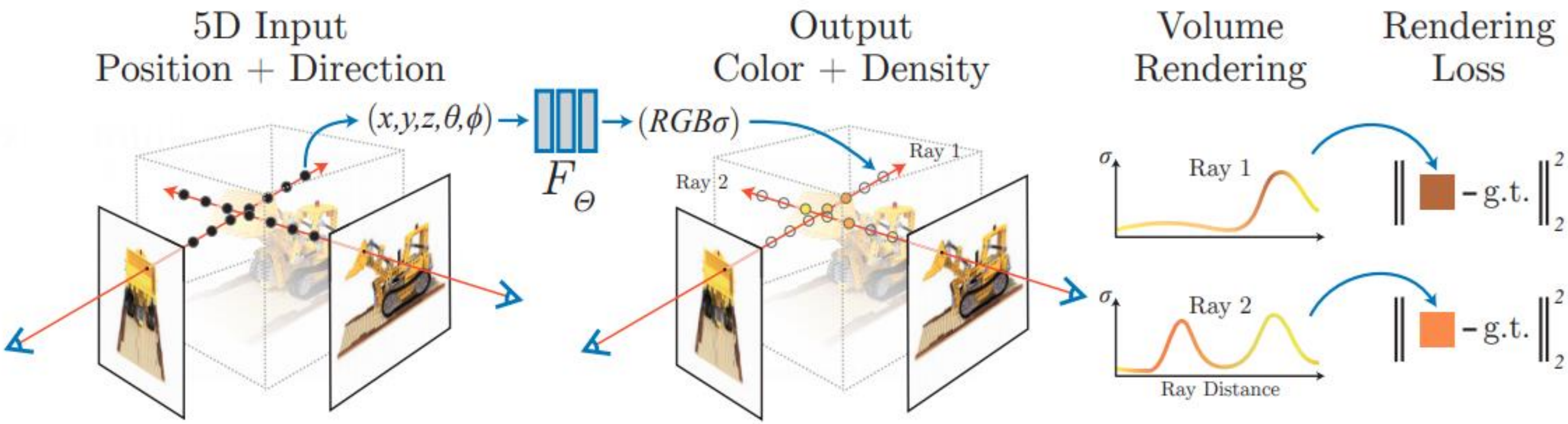
Original



Editing



Take-home Message

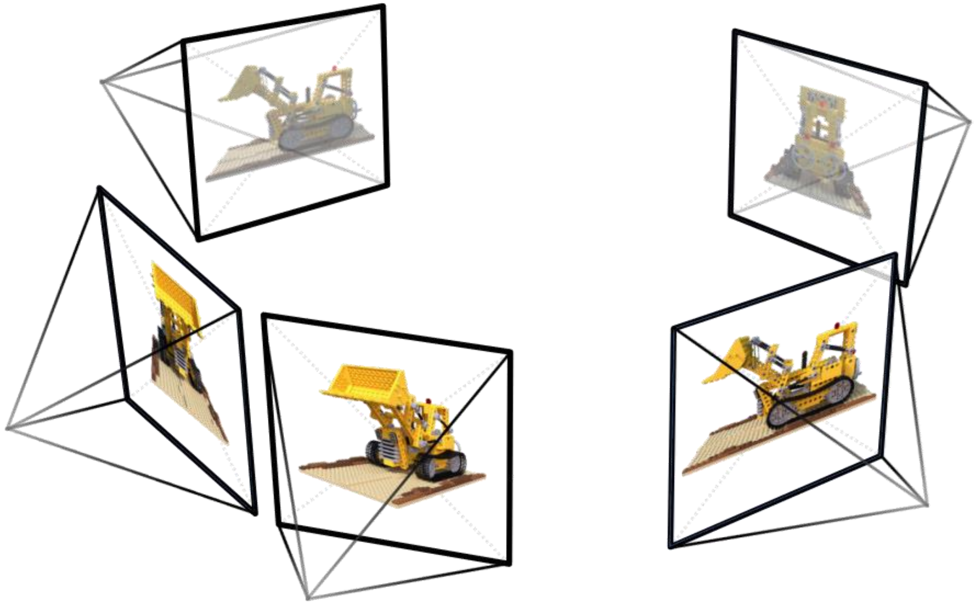


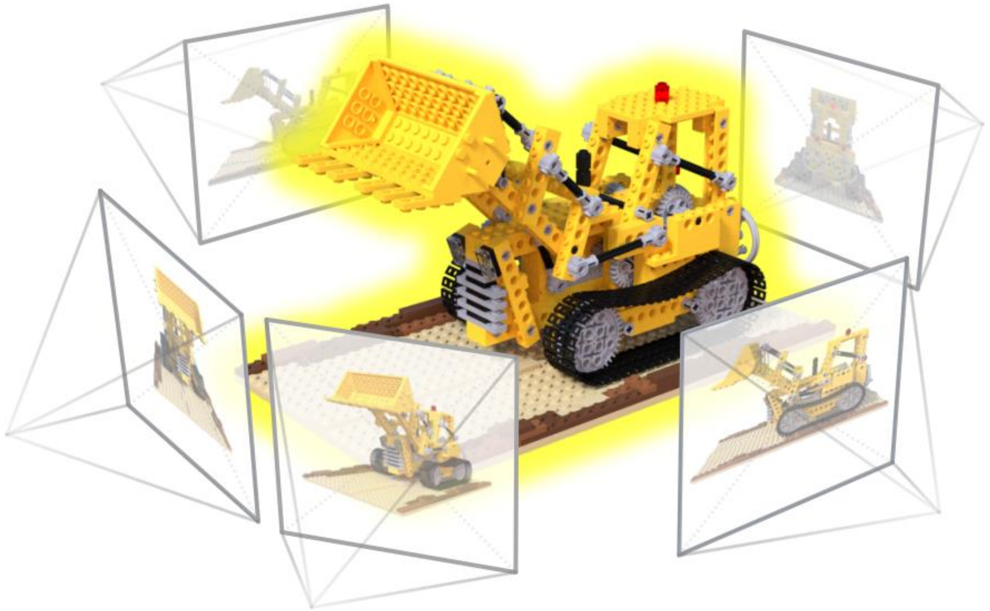
- × Implicit representation ➡ Taking explicit representation as a proxy
- × The scene is represented as a whole ➡ Learn local features
- × Everything is entangled within the network ➡ Disentangle components

Pose estimation for 3D neural reconstruction

“Pose”

Why do we need camera pose estimation?





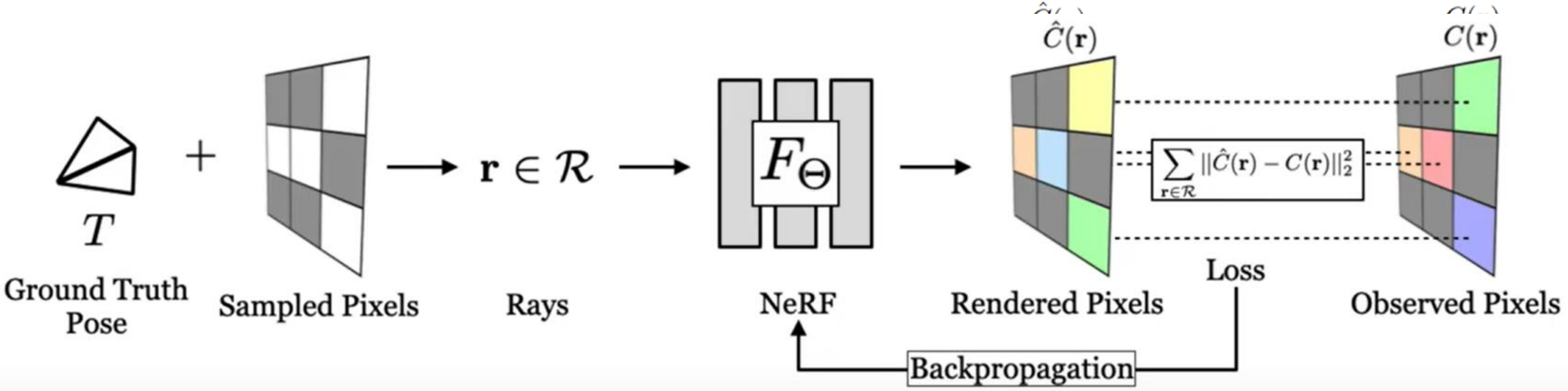
Images + accurate camera poses

3D scene representation

- × Accurate camera poses are necessary
- × Offline process

iNeRF

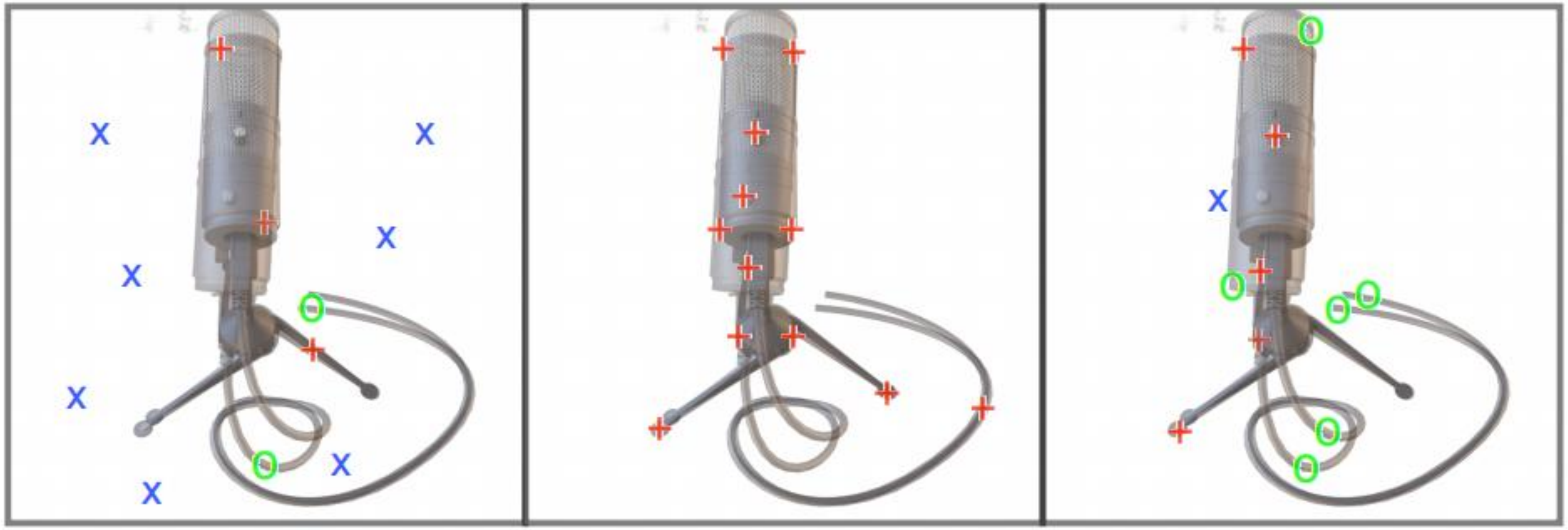
Key Idea: Inverting an optimized neural radiance field for pose estimation.



Yen-Chen, Lin, et al. "inerf: Inverting neural radiance fields for pose estimation." IROS 2021

iNeRF

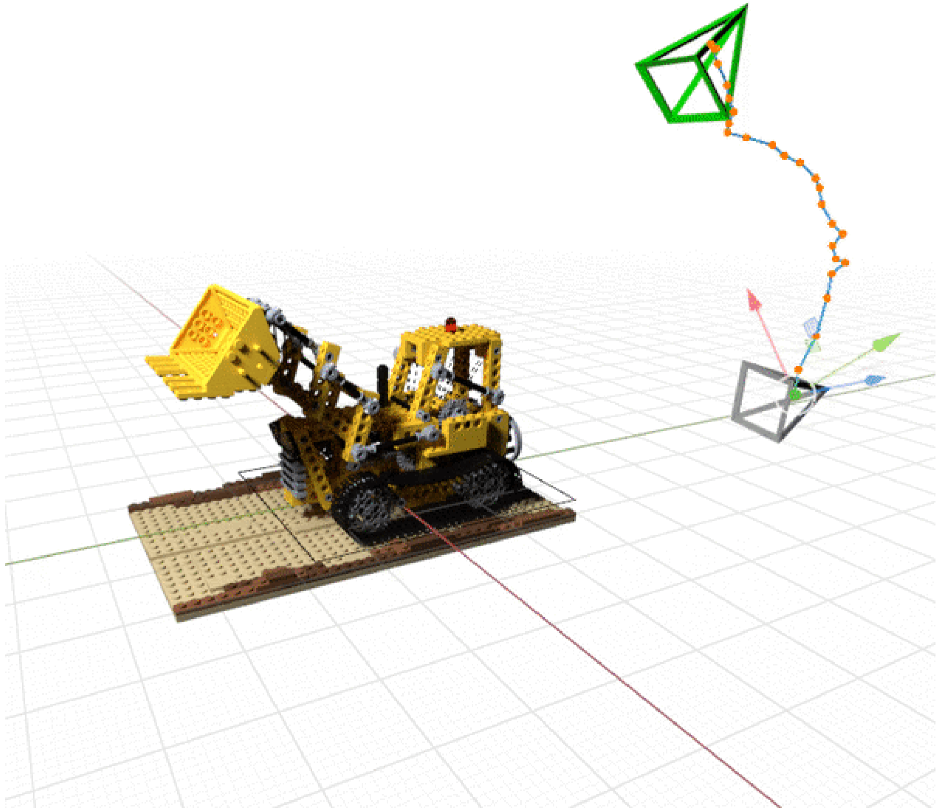
Key Observation: Sampling rays play a role in the optimization procedure.

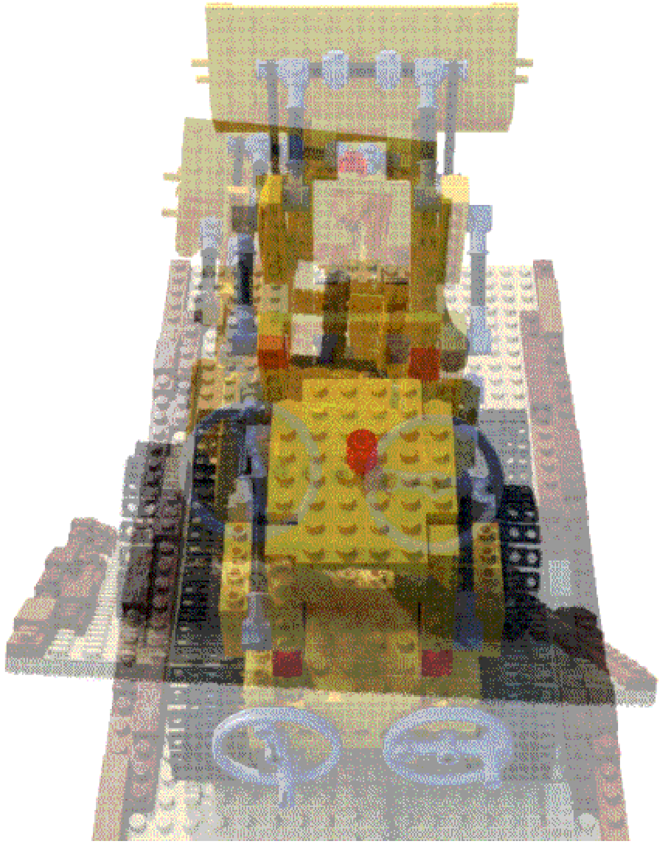


Random Interest Point Interest Region

Yen-Chen, Lin, et al. "inerf: Inverting neural radiance fields for pose estimation." IROS 2021

iNeRF



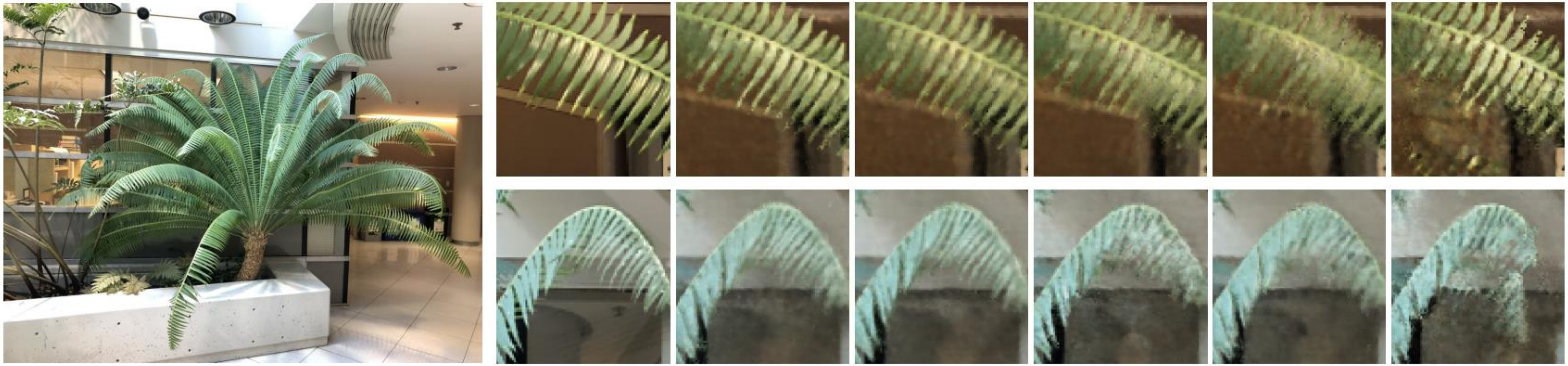


Yen-Chen, Lin, et al. "inerf: Inverting neural radiance fields for pose estimation." IROS 2021

iNeRF

Application: Self-Supervising NeRF with iNeRF

- Train a NeRF given a set of training RGB images with known camera poses;
- Use iNeRF to take in additional unknown-pose observed images and solve for estimated poses ;
- Use the self-supervised pose labels to add unknown-pose observed images into the training set.



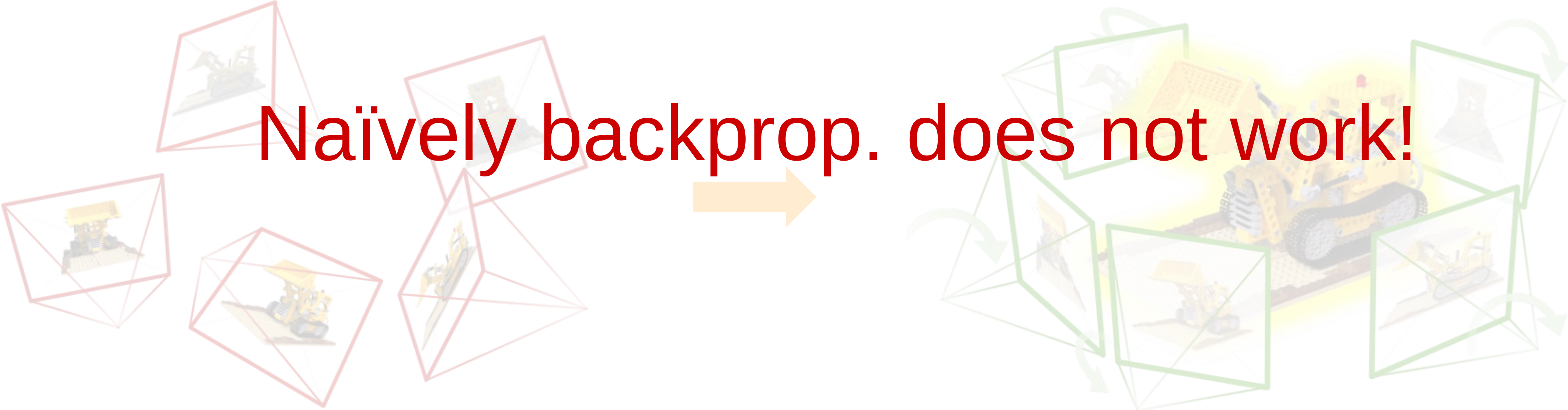
Fern   Ground Truth   100%   50%+iNeRF   50%   25%+iNeRF   25%

Yen-Chen, Lin, et al. "inerf: Inverting neural radiance fields for pose estimation." IROS 2021

BARF

Key Idea: Jointly optimizing for registration and reconstruction.

$$\min_{\mathbf{p}_1,\ldots,\mathbf{p}_M,\boldsymbol{\Theta}} \sum_{i=1}^{\overset{\text{frames}}{M}} \sum_{\mathbf{u}} \Big\| \underbrace{\hat{\mathcal{I}}}_{\substack{\text{RGB}\\\text{(rendered)}}}(\mathbf{u}; \underbrace{\mathbf{p}_i}_{\substack{\text{camera}\\\text{pose}}}, \underbrace{\boldsymbol{\Theta}}_{\substack{\text{network}\\\text{params.}}}) - \underbrace{\mathcal{I}_i}_{\text{RGB}}(\mathbf{u}) \Big\|_2^2$$



Naïvely backprop. does not work!

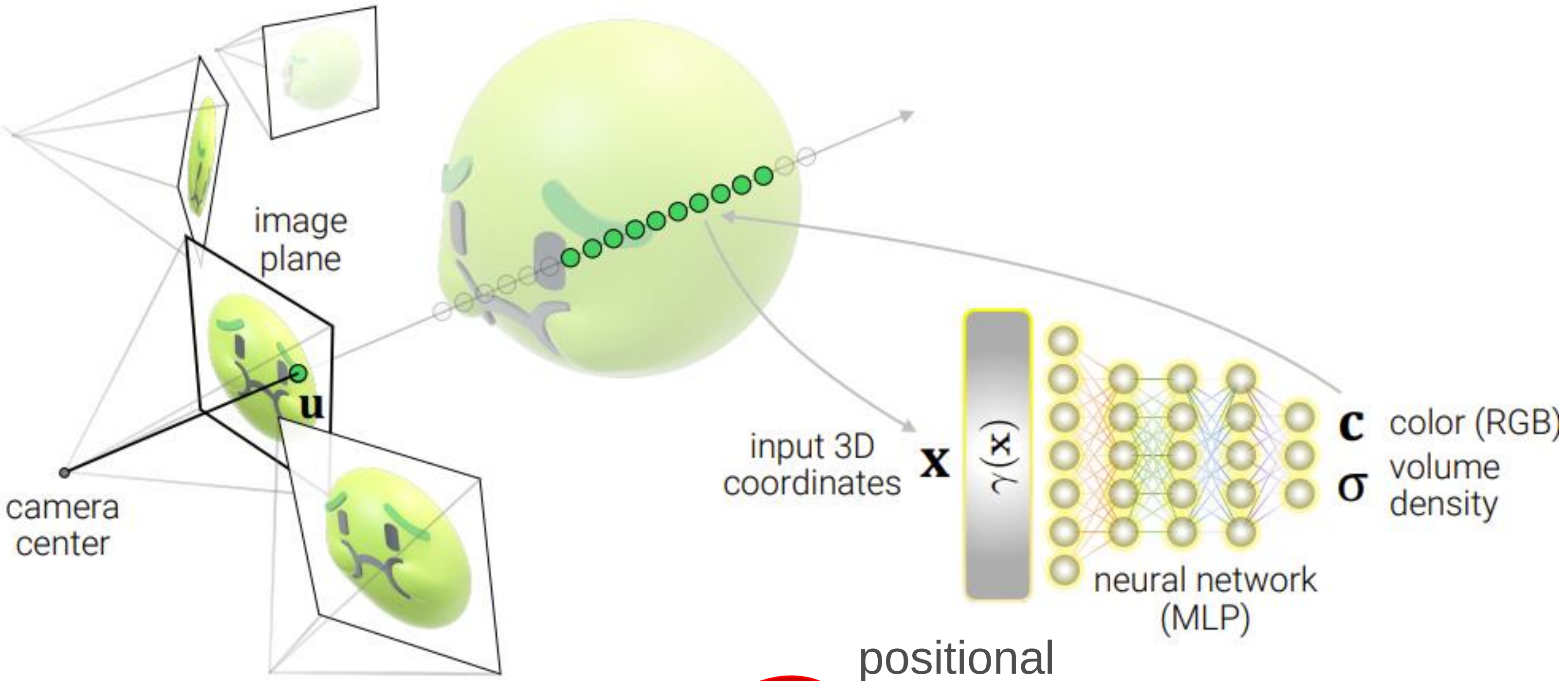
Images + **imperfect** camera poses

3D scene representation + **registered camera poses**

Lin, Chen-Hsuan, et al. "Barf: Bundle-adjusting neural radiance fields." ICCV 2021.

BARF

Key Observation:



positional
encoding

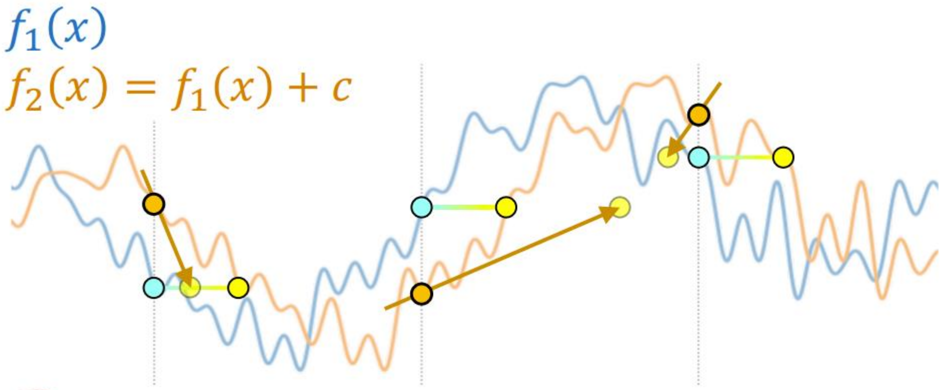
✓ encourages representation with high frequency

× detrimental to gradient-based registration!!

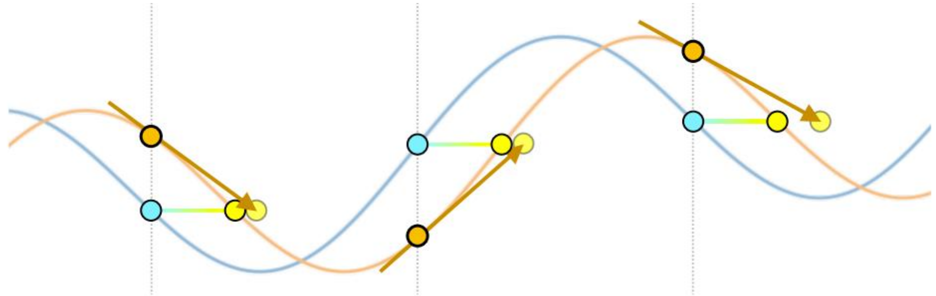
Lin, Chen-Hsuan, et al. "Barf: Bundle-adjusting neural radiance fields." ICCV 2021.

BARF

Key Solution: Making it coarse-to-fine!



× gets stuck in suboptimal solutions



✓ smooth signals -> coherent updates

Resolve large pose misalignment & coarse scene representation

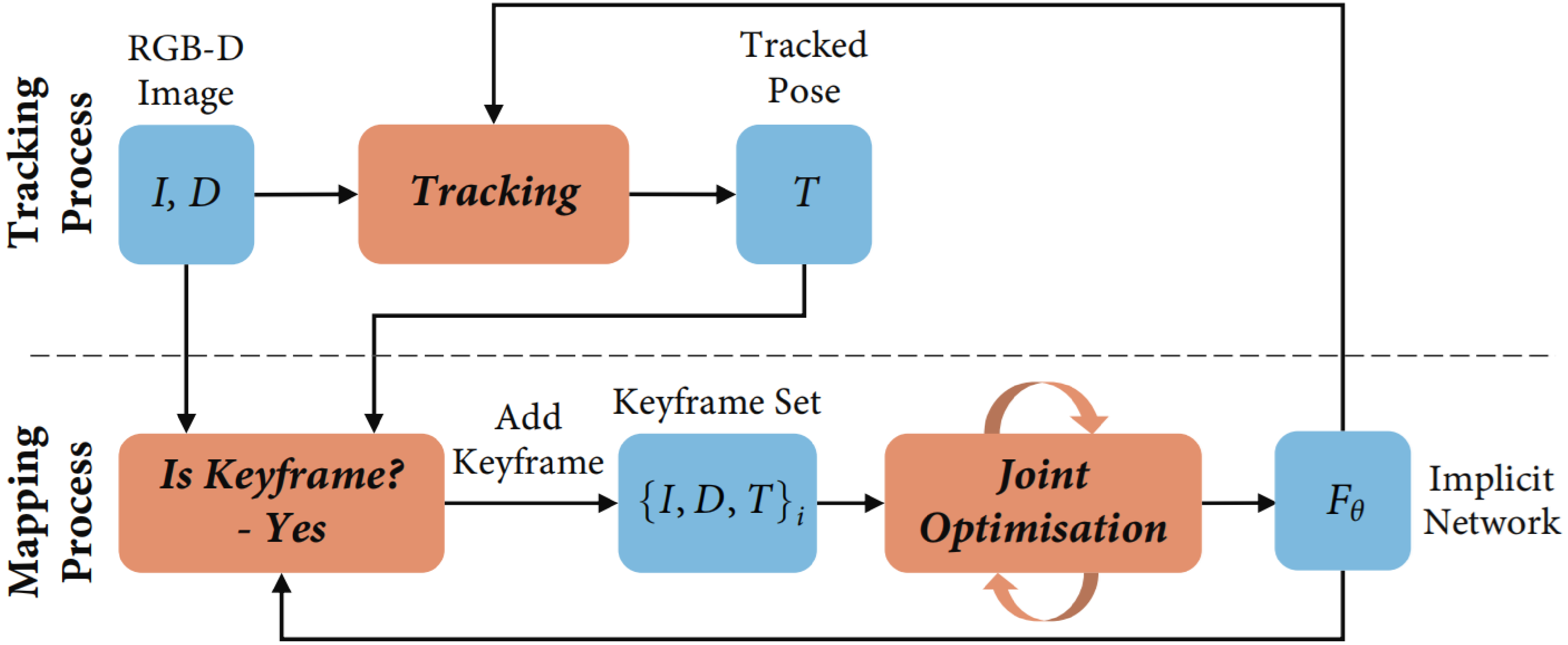
Gradually activate higher-frequency components in positional encoding

Refine granular pose misalignment & high-fidelity scene representation

Lin, Chen-Hsuan, et al. "Barf: Bundle-adjusting neural radiance fields." ICCV 2021.

iMAP

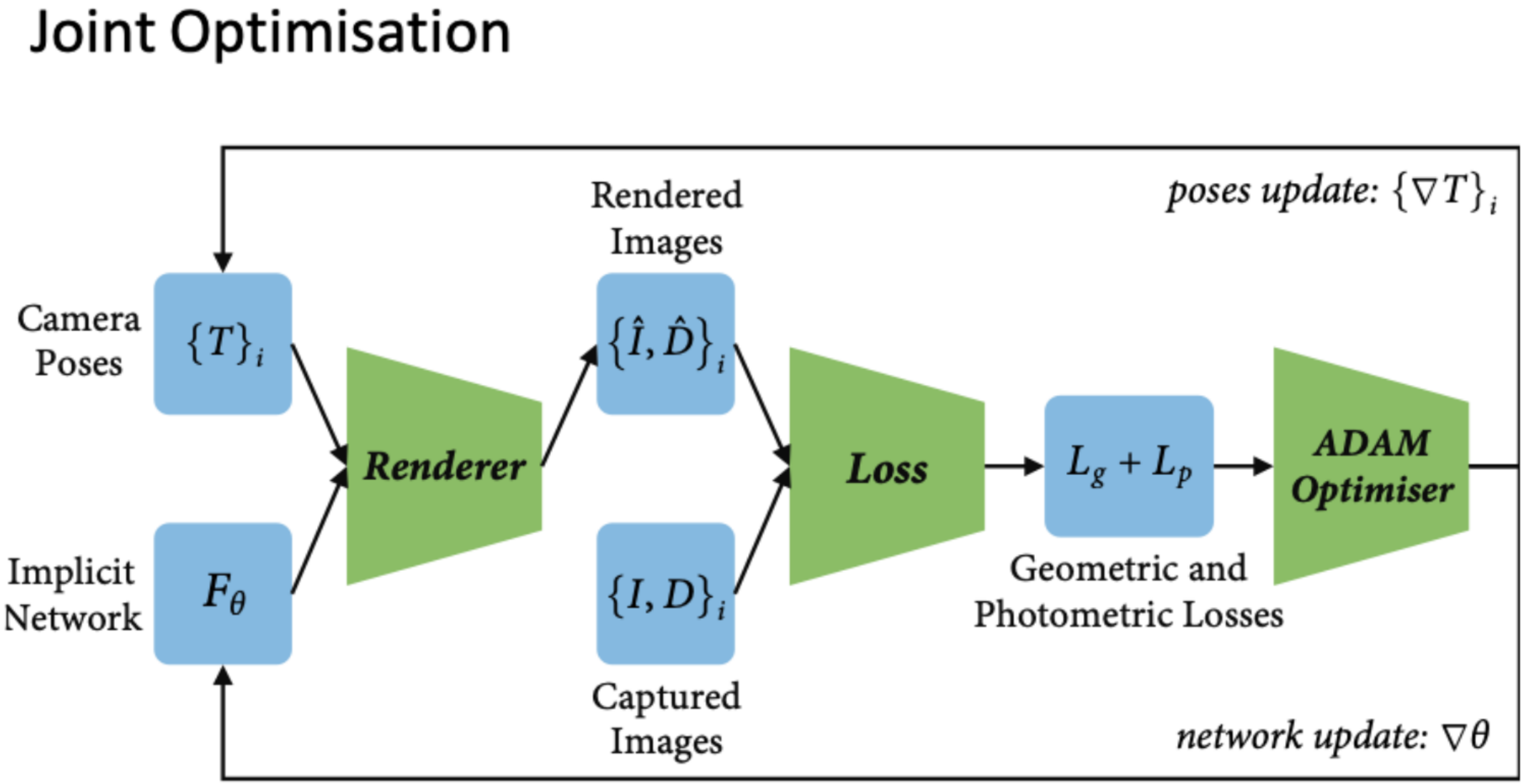
Key Idea: Use a multilayer perceptron (MLP) to serve as the only scene representation in a real-time SLAM system for a handheld RGB-D camera.



Sucar, Edgar, et al. "iMAP: Implicit mapping and positioning in real-time." ICCV 2021.

iMAP

Key Idea: Use a multilayer perceptron (MLP) to serve as the only scene representation in a real-time SLAM system for a handheld RGB-D camera.



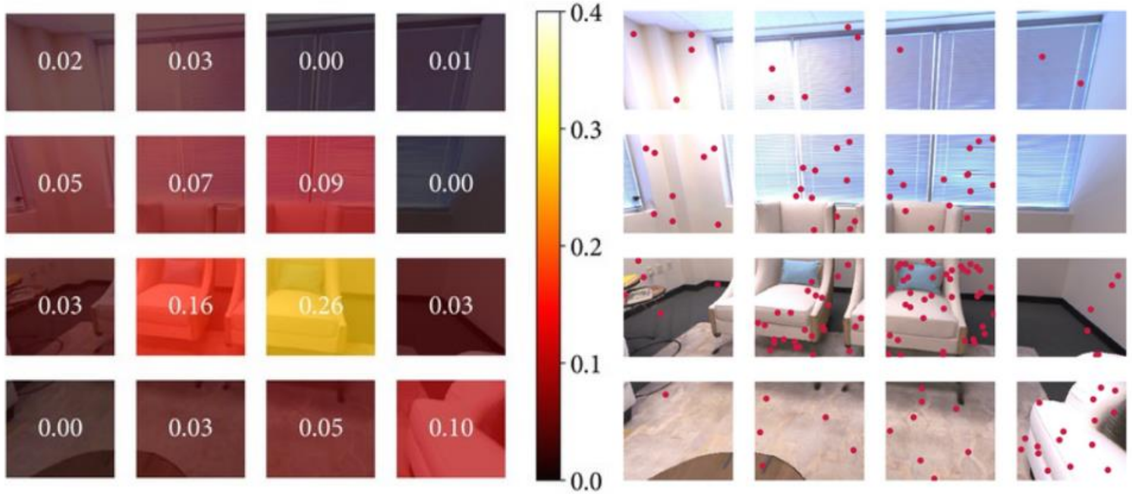
Sucar, Edgar, et al. "iMAP: Implicit mapping and positioning in real-time." ICCV 2021.

iMAP

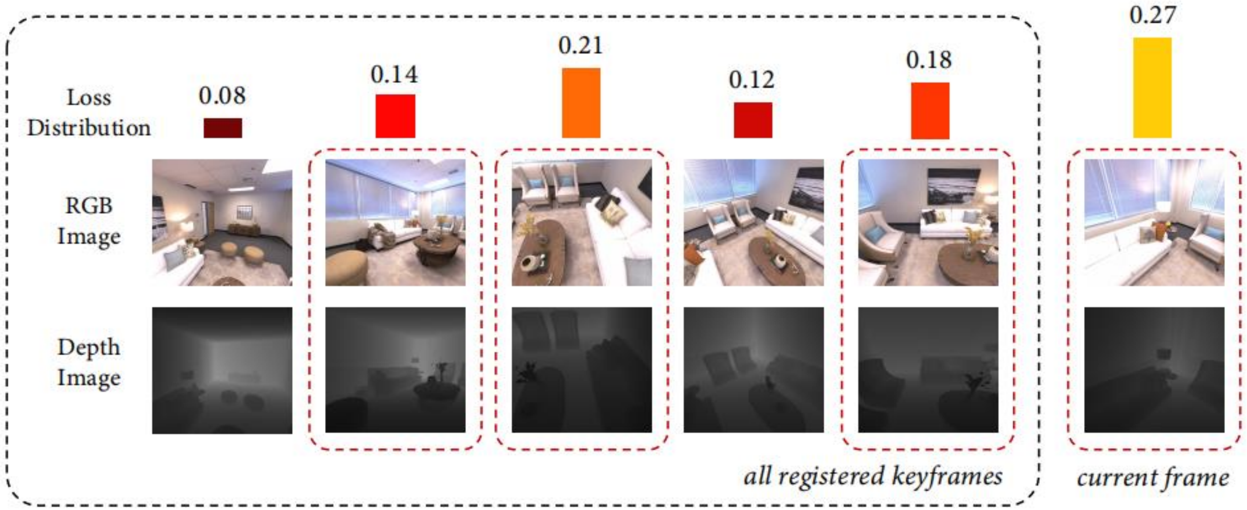
Challenges: How to make it efficient enough for the real-time application?

Solution: Active sampling

Image Active Sampling



Keyframe Active Sampling



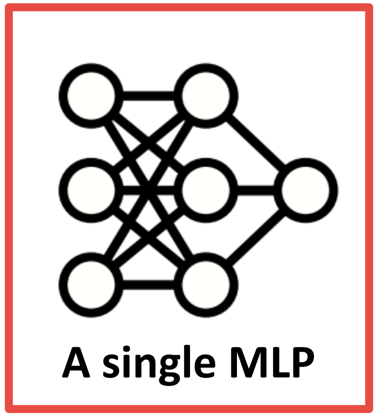
Sucar, Edgar, et al. "iMAP: Implicit mapping and positioning in real-time." ICCV 2021.

iMAP

Sucar, Edgar, et al. "iMAP: Implicit mapping and positioning in real-time." ICCV 2021.

iMAP

Problems:

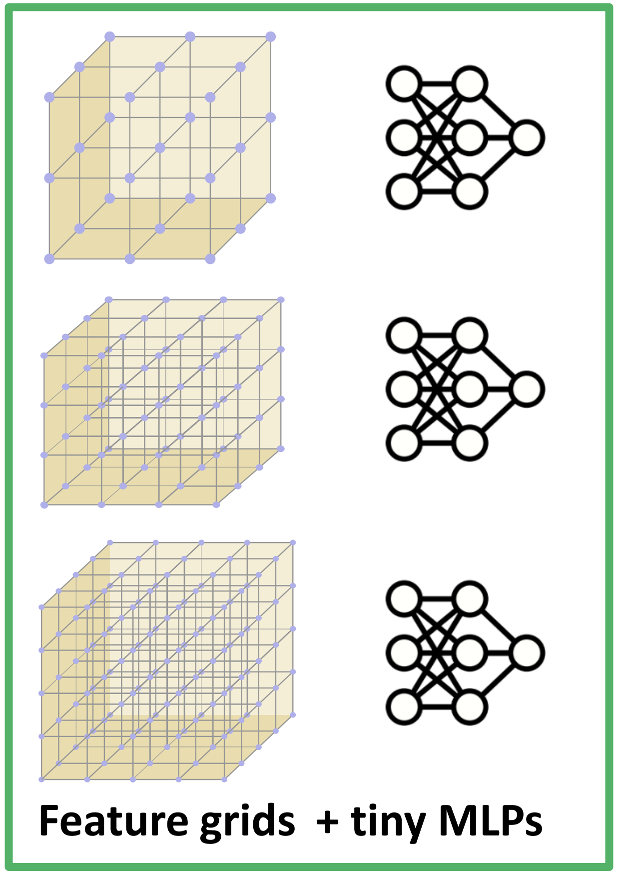


- **Fail when scaling up to larger scenes**
- Global update → **Catastrophic forgetting**
- **Slow convergence**

Predicted Poses
GT Poses

Sucar, Edgar, et al. "iMAP: Implicit mapping and positioning in real-time." ICCV 2021.

NICE-SLAM



- Applicable to **large-scale scenes**
- Local update → **No forgetting problem**
- **Fast** convergence

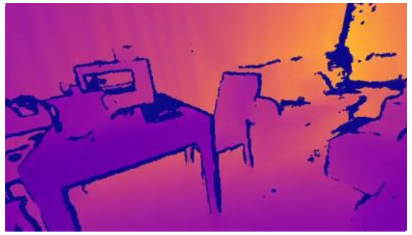
Predicted Poses
GT Poses

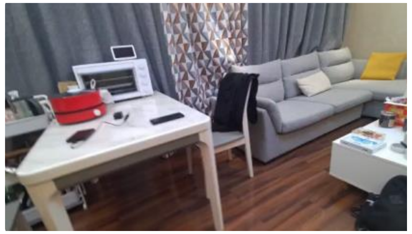
Zhu, Zihan, et al. "Nice-slam: Neural implicit scalable encoding for slam." CVPR 2022..

NICE-SLAM

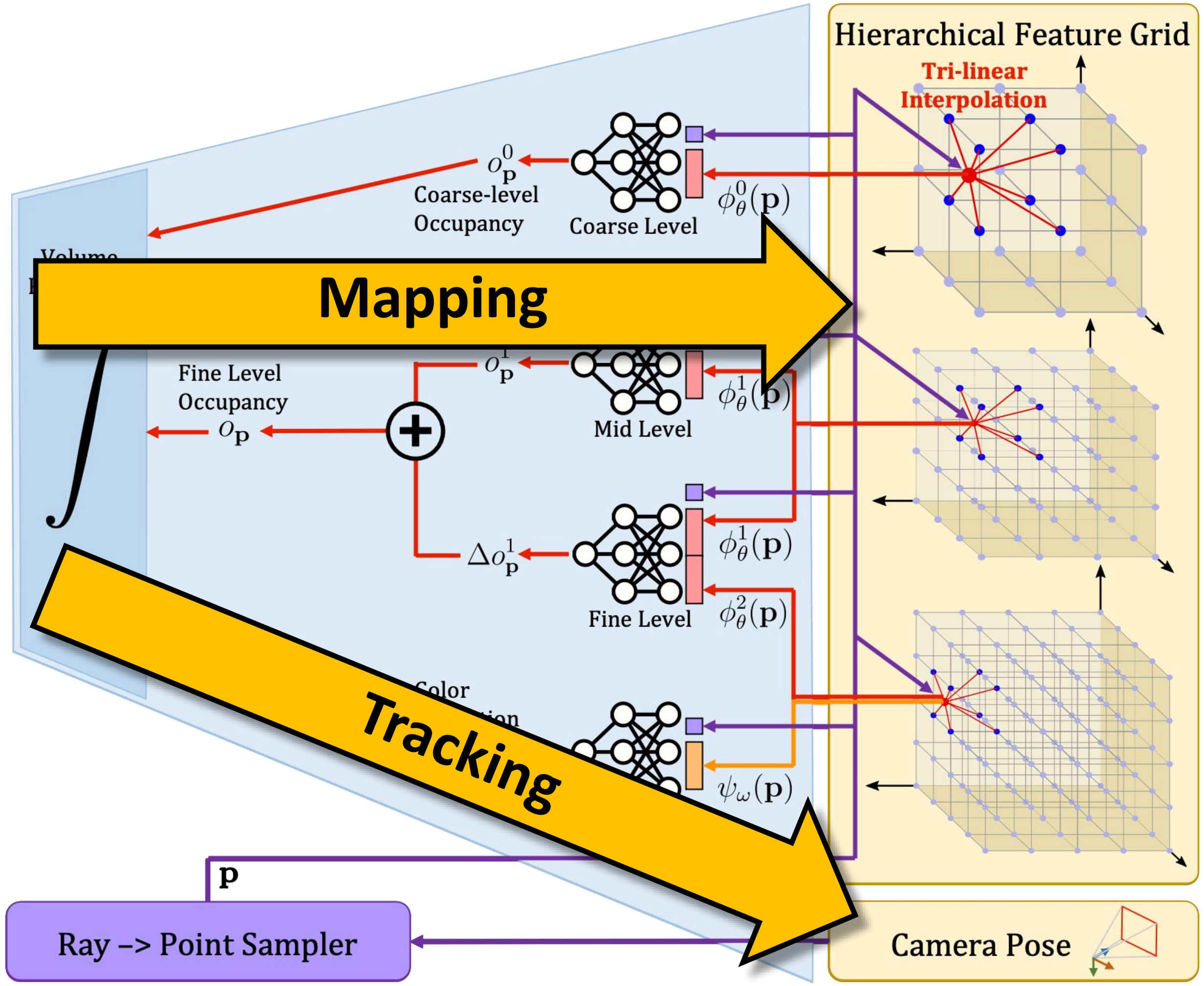
Key Idea: Hierarchical Feature Grid + Coarse-to-Fine Strategy + Shape Prior

Input Depth



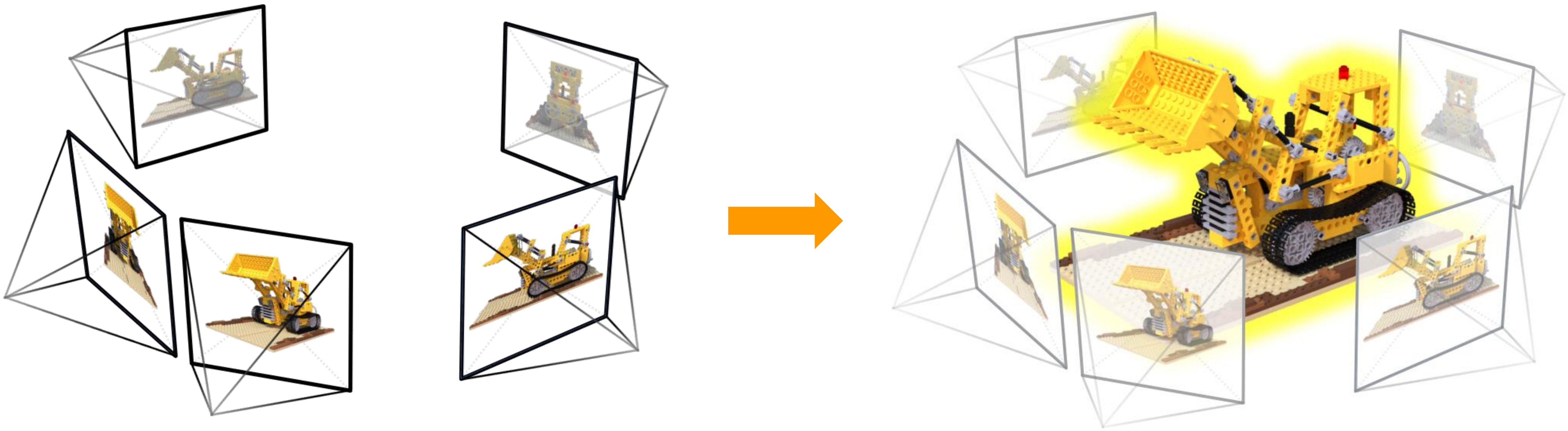


Input RGB



Zhu, Zihan, et al. "Nice-slam: Neural implicit scalable encoding for slam." CVPR 2022..

Take-home Message



Images + accurate camera poses

3D scene representation

- × Accurate camera poses are necessary ➡ Joint optimization
- × Offline process ➡ Advanced sampling + scalable representation

Thank you