Reverse Rendering for 3D Scene Reconstruction

Qiangeng (Charlie) Xu

University of Southern California

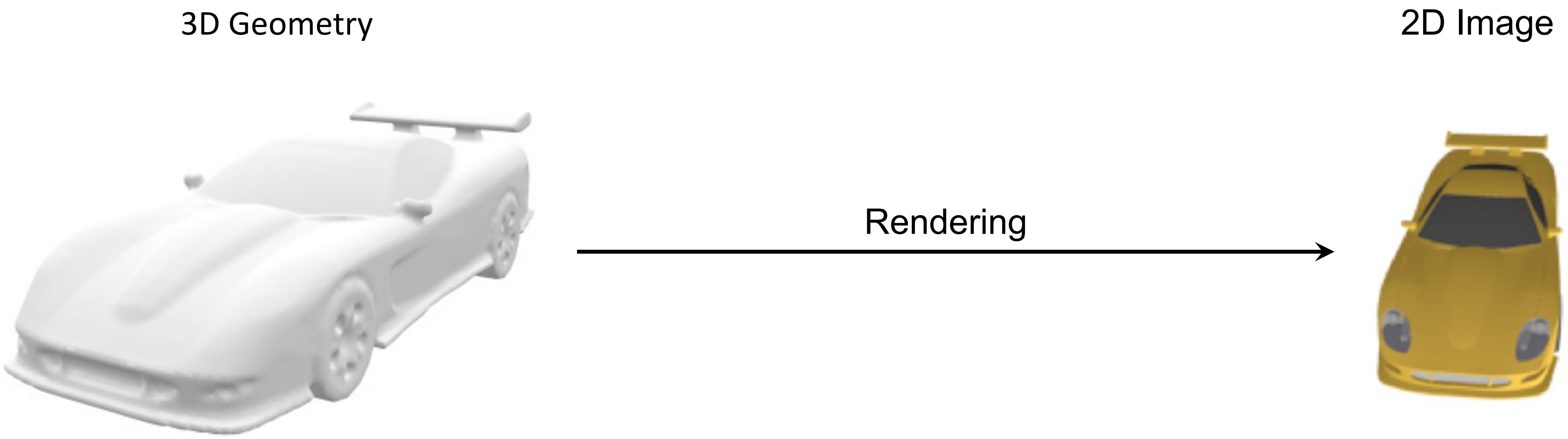
What Do We Need to Render an Image ?

3D Scene Components?

Rendering

2D Image

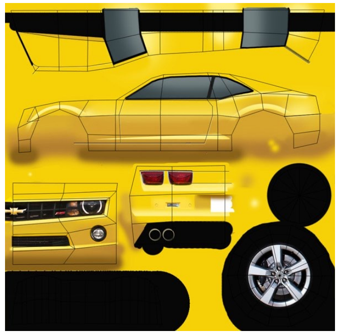
3D Scene Components to Render a RGB Image.



3D Scene Components to Render a RGB Image.

3D Geometry

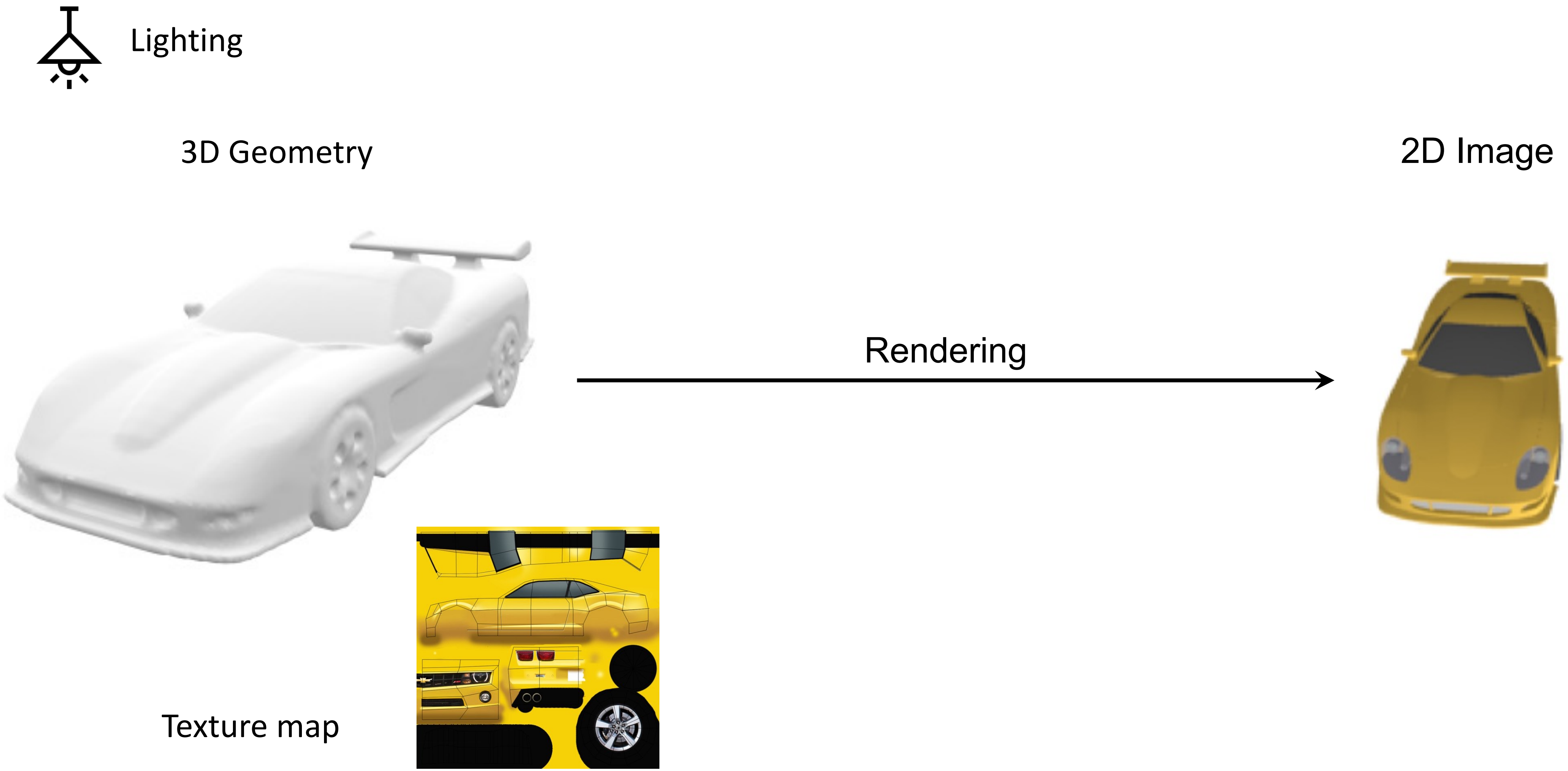
Texture map



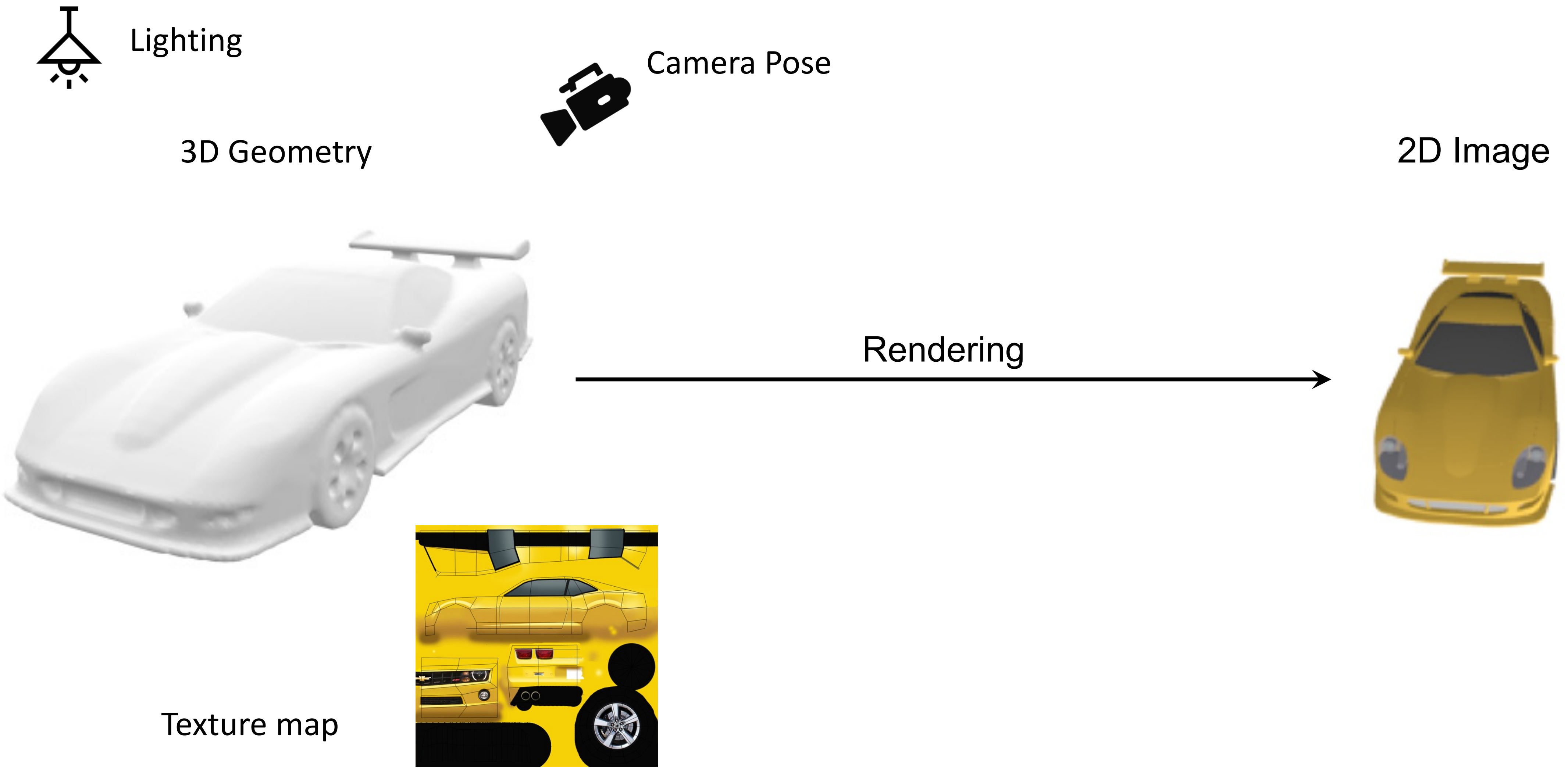
Rendering

2D Image

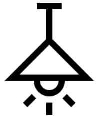
3D Scene Components to Render a RGB Image.



3D Scene Components to Render a RGB Image.



3D Scene Components to Render a RGB Image.



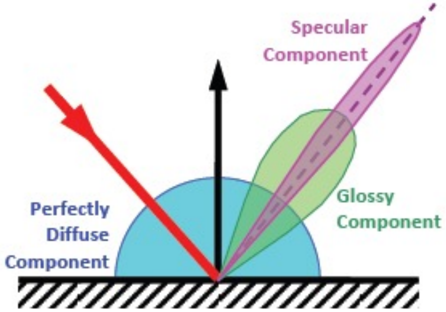
Lighting

Camera Pose

3D Geometry

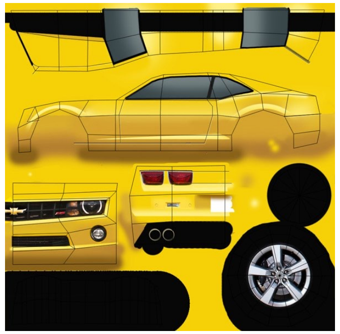
2D Image

Rendering



Surface Rendering models

Texture map



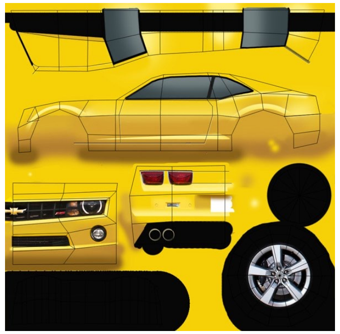
3D Scene Reconstruction from 2D Information

Entangled

Lighting

3D Geometry

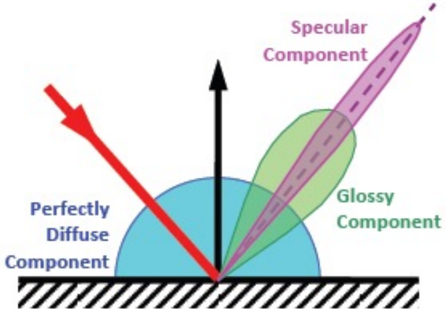
Texture map



Camera Pose

2D Image

3D Scene Reconstruction

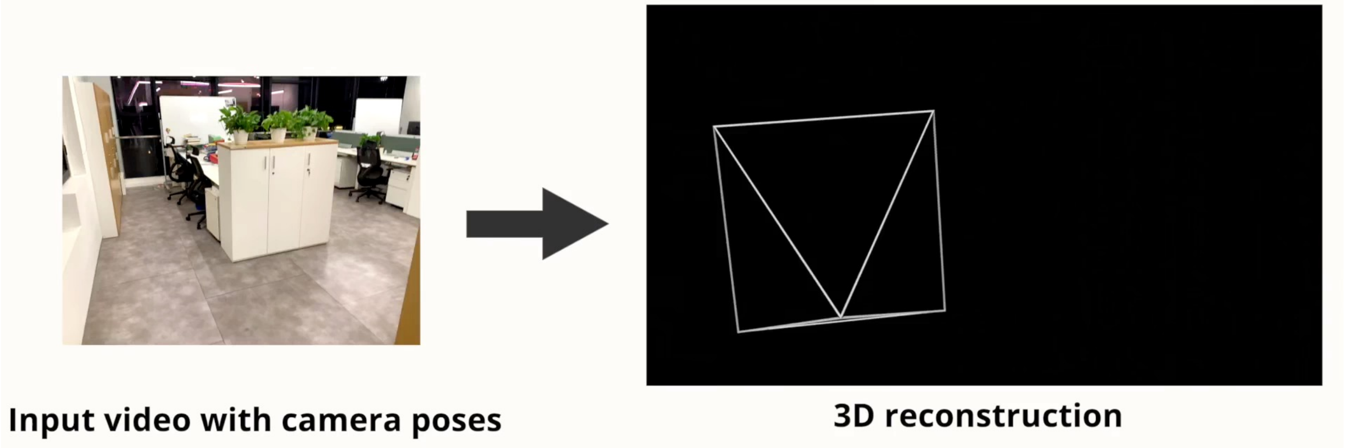


Surface Rendering models

Why Reconstruct 3D Scenes

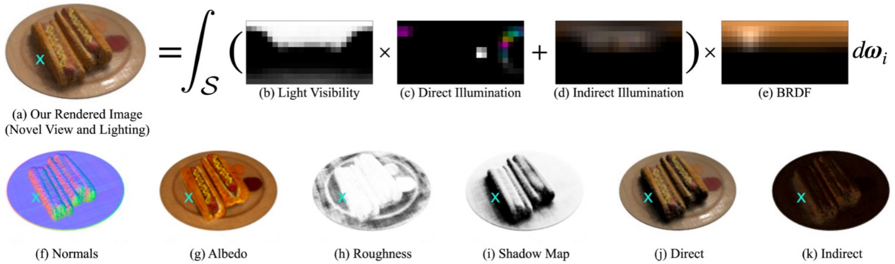
[Auto Modeling]

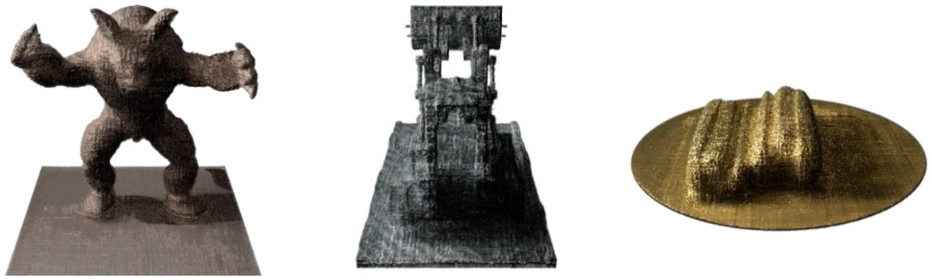
Augmented Reality



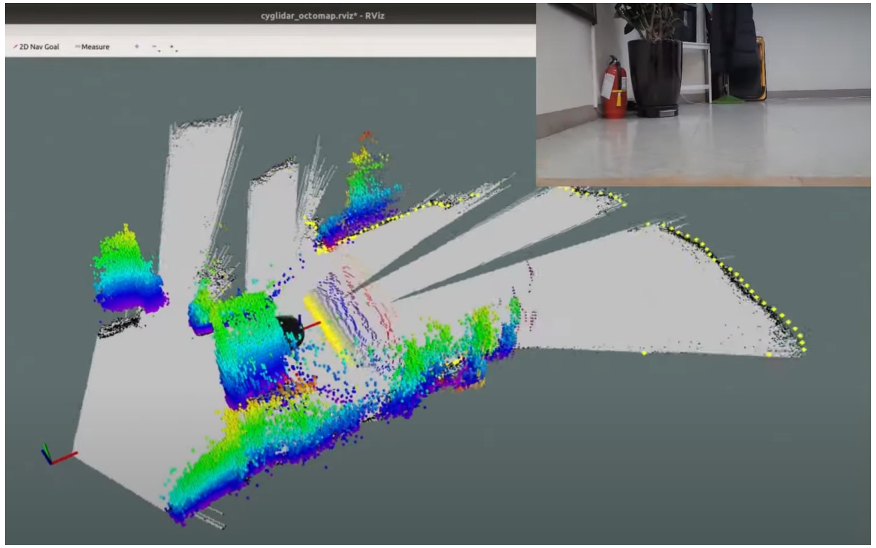
Why Reconstruct 3D Scenes

Scene Editing / Relighting





[Simultaneous Localization And Mapping (SLAM)]
Fundamental Robotic Problem



Road Map

- 3D Geometry Reconstruction from a single 2D image.
  - Implicit model
- 3D Geometry Reconstruction from Multi-view 2D images.
  - MVS method
- 3D Scene Reconstruction from Multi-view 2D images.
  - Reverse Volume Rendering

Obstacles of 3D Scene Reconstruction from a Single 2D Image.

1. Just like Recover a human body from his shadow.
Not enough information

2. Components are entangled.

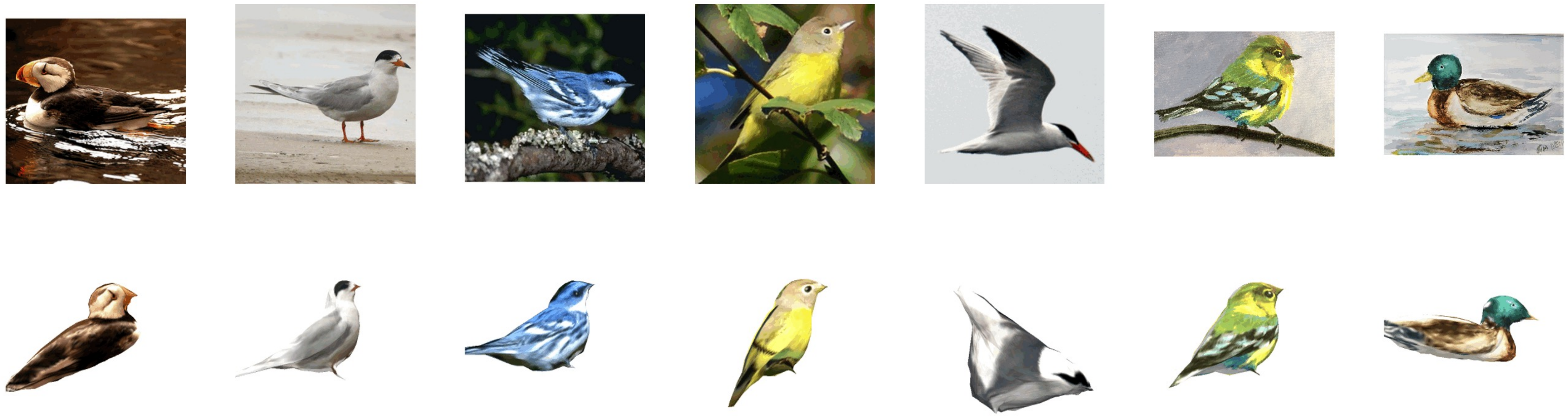
Entanglement of Geometry & Textures

3D Scene Reconstruction From a RGB Image.

Can only learn easy categories with a similar topology.

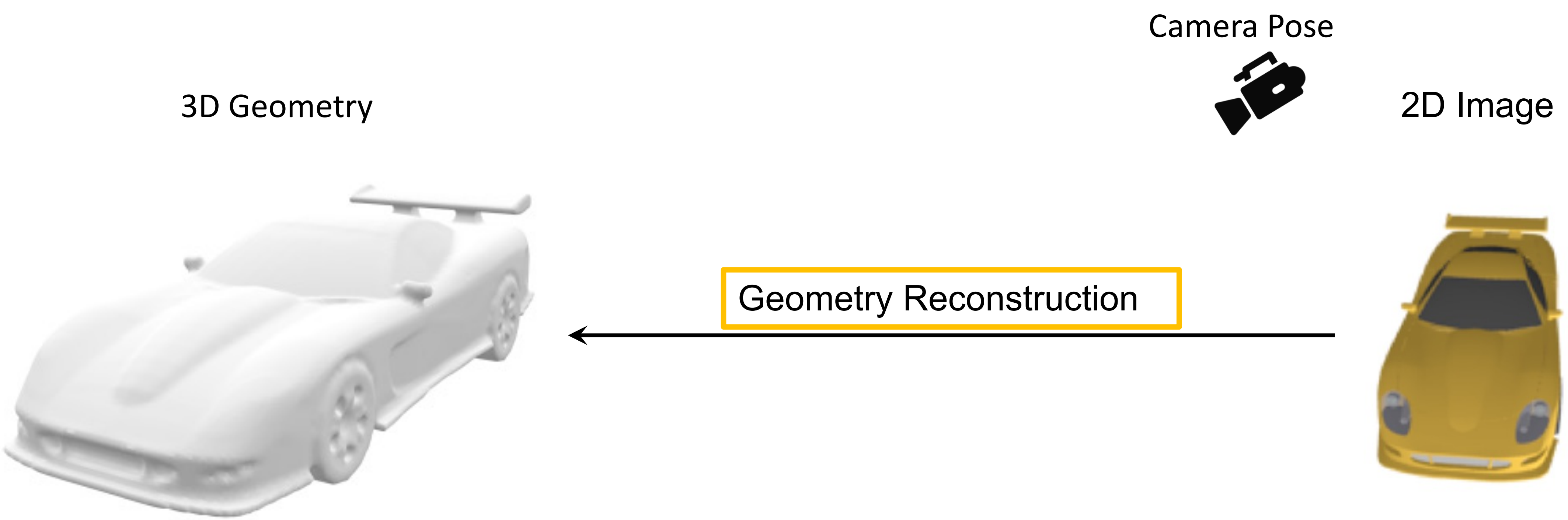
Input

Output

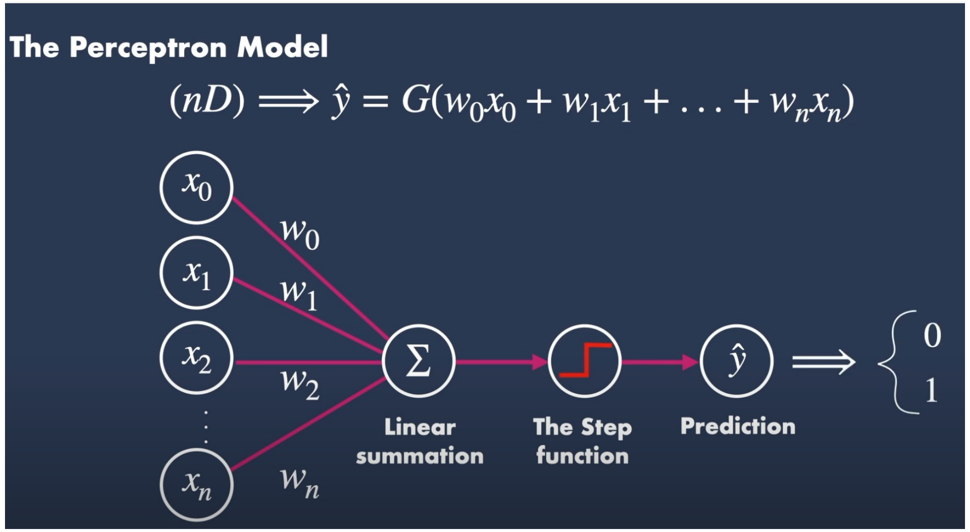


Li, X., Liu, S., Kim, K., De Mello, S., Jampani, V. and Kautz, J., Nvidia Corp, 2022. *Self-supervised single-view 3D reconstruction via semantic consistency*. U.S. Patent 11,238,650.

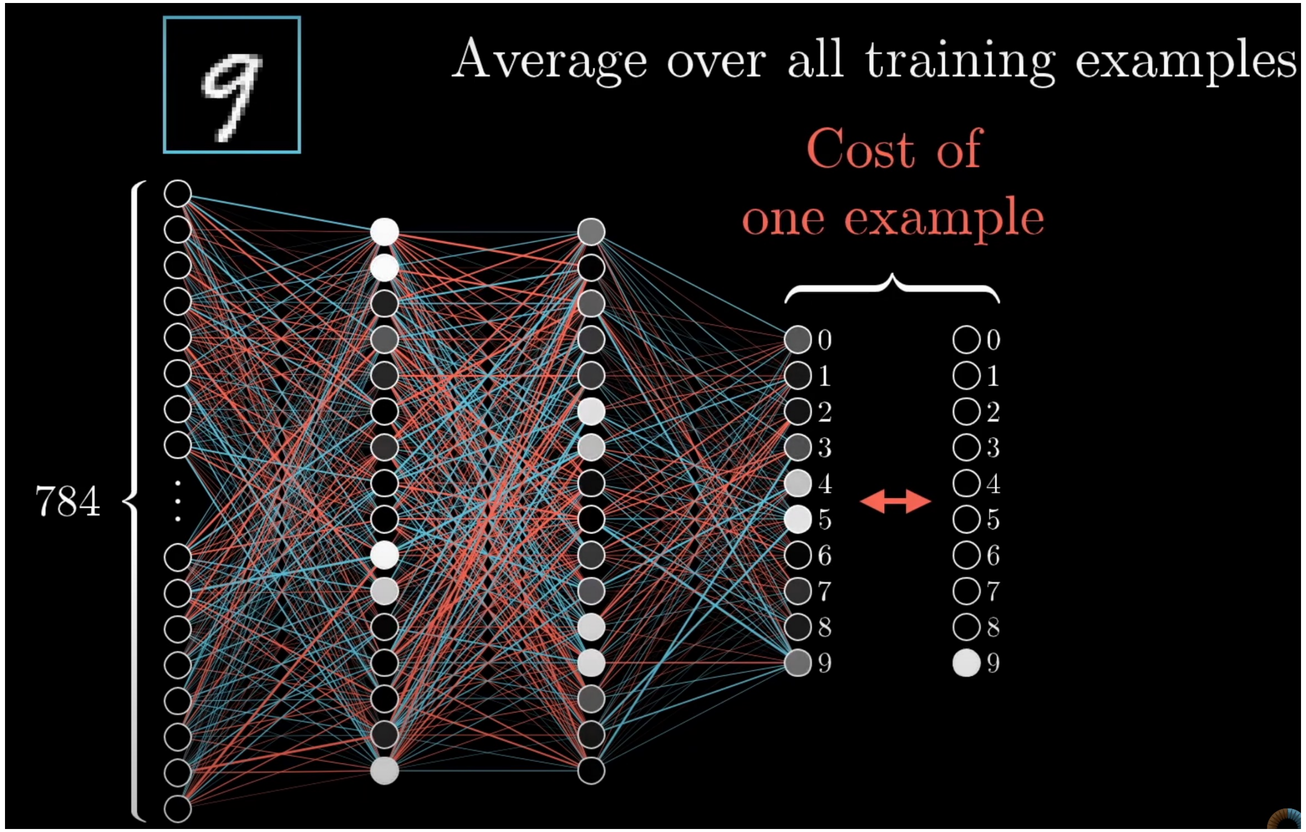
3D Geometry Reconstruction From a RGB Image.



Quick Intro to Multi-layer Perceptrons (MLPs)

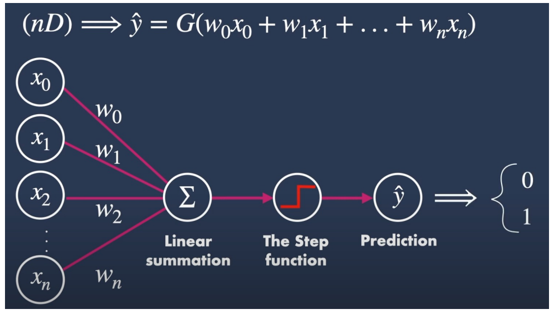


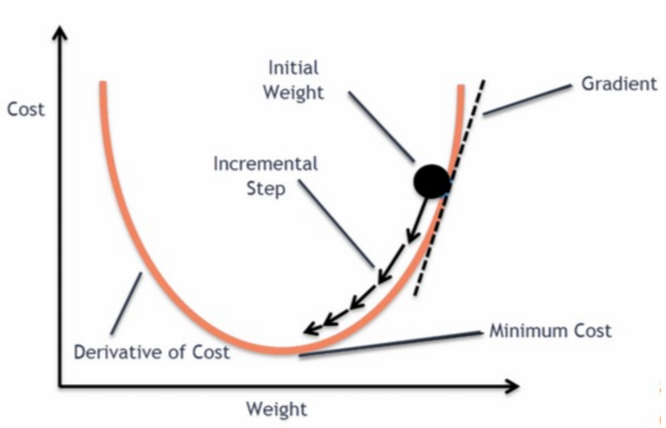
Single-layer Perceptron

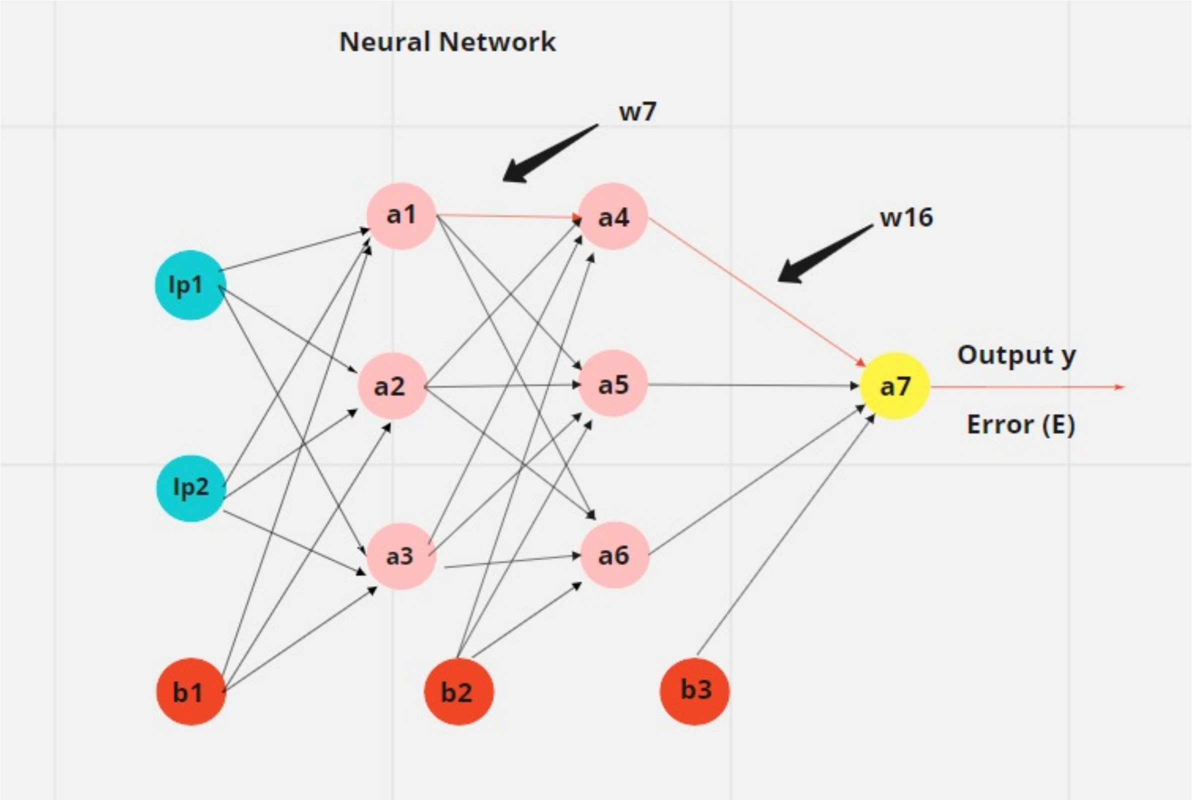


Multi-layer Perceptron

Back Propagation by Chain Rule







The derivation of output w.r.t the affine

$$\frac{\partial E}{\partial w16} = \frac{\partial E}{\partial y} * \frac{\partial y}{\partial net} * \frac{\partial net}{\partial w16}$$

The derivative of error w.r.t output

derivative of affine w.r.t the weight

The derivation of output w.r.t the affine

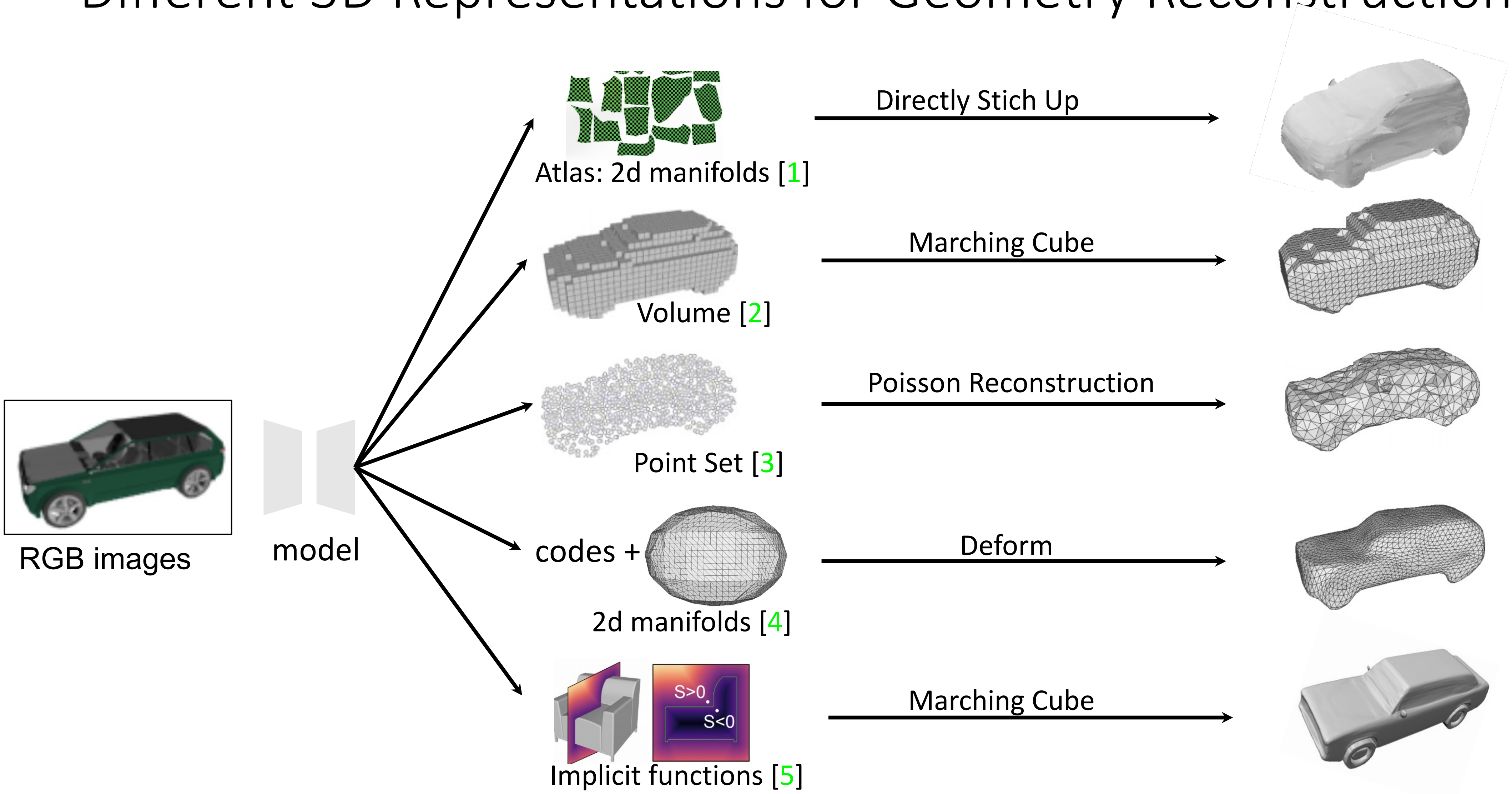
$$\frac{\partial E}{\partial a4} = \frac{\partial E}{\partial y} * \frac{\partial y}{\partial net} * \frac{\partial net}{\partial a4}$$

The derivative of error w.r.t output

derivative of affine w.r.t the activation 4

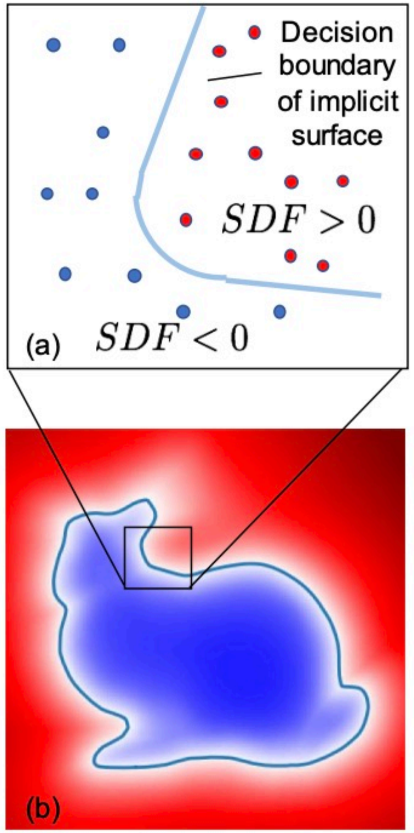
$$\frac{\partial E}{\partial w7} = \frac{\partial E}{\partial a4} * \frac{\partial a4}{\partial net7} * \frac{\partial net7}{\partial w7}$$

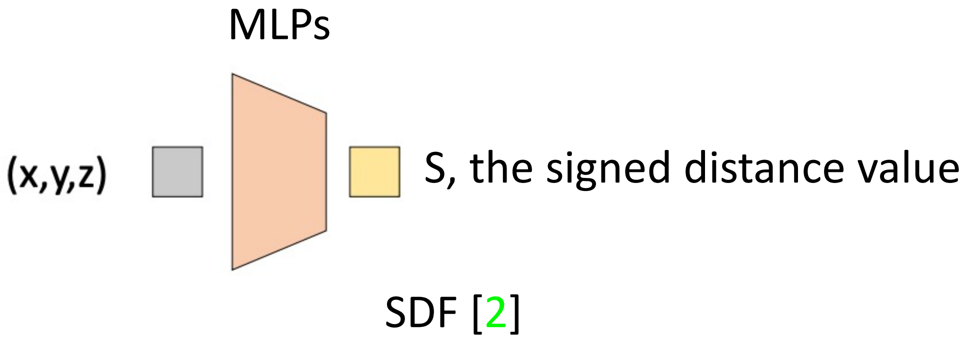
Different 3D Representations for Geometry Reconstruction



1. Thibault Groueix, Matthew Fisher, Vladimir G. Kim, Bryan Russell, and Mathieu Aubry. AtlasNet: A Papier-Mâché Approach to Learning 3D Surface Generation. In CVPR, 2018.
2. Yan, Xinchen, Jimei Yang, Ersin Yumer, Yijie Guo, and Honglak Lee. “Perspective transformer nets: Learning single-view 3d object reconstruction without 3d supervision.” NIPS, 2016.
3. Fan, Haoqiang, Hao Su, and Leonidas J. Guibas. “A point set generation network for 3d object reconstruction from a single image.” CVPR, 2017.
4. Nanyang Wang, Yinda Zhang, Zhuwen Li, Yanwei Fu, Wei Liu, and Yu-Gang Jiang. “Pixel2mesh: Generating 3d mesh models from single rgb images.” ECCV 2018
5. Xu, Q., Wang, W., Ceylan, D., Mech, R. and Neumann, U., 2019. Disn: Deep implicit surface network for high-quality single-view 3d reconstruction. In NeurIPS, 2019.

Signed Distance Functions [1]





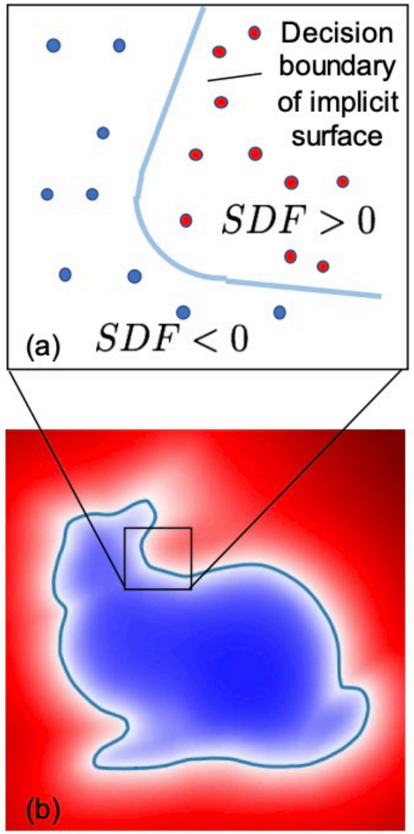
$$X := \{(\boldsymbol{x}, s) : SDF(\boldsymbol{x}) = s\}$$

The mesh surface is Implicitly represented as zero iso-surface of sdf values

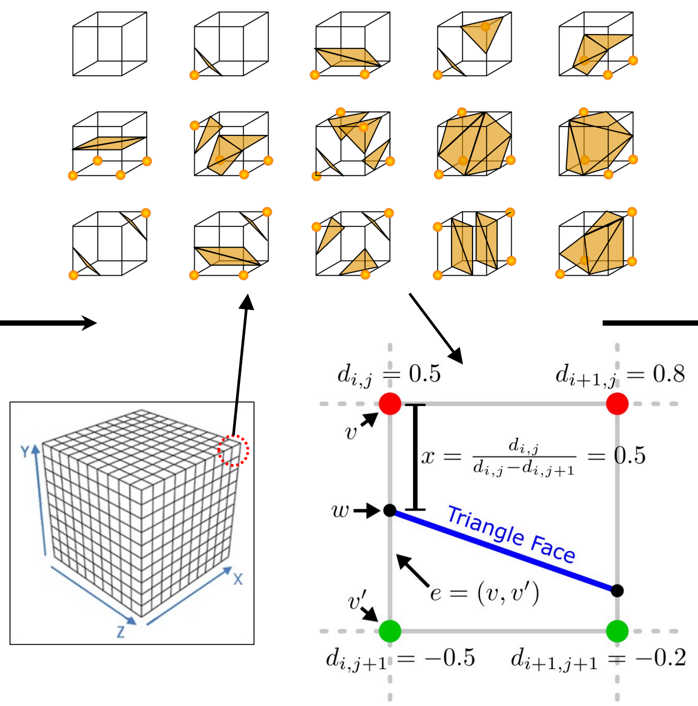
1.Malladi, R.; Sethian, J.A.; Vemuri, B.C. (1995). "Shape modeling with front propagation: a level set approach". IEEE Transactions on Pattern Analysis and Machine Intelligence. 17 (2): 158–175.
2. Park, Jeong Joon, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. “Deepsdf: Learning continuous signed distance functions for shape representation.” CVPR, 2019.

Signed Distance Functions for 3D Geometry Reconstruction

animation



Signed Distance Fields [1]



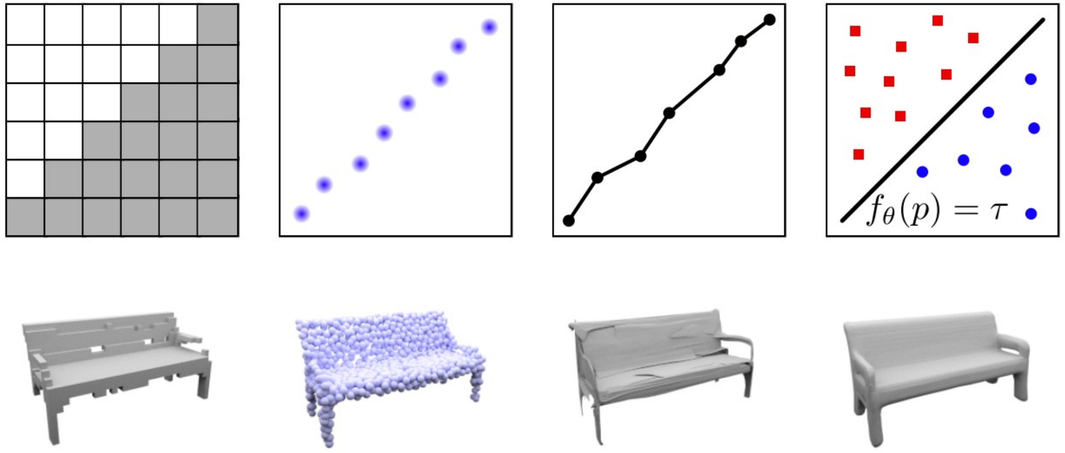
Marching Cube [2]

3D Mesh

1.Malladi, R.; Sethian, J.A.; Vemuri, B.C. (1995). "Shape modeling with front propagation: a level set approach". IEEE Transactions on Pattern Analysis and Machine Intelligence. 17 (2): 158–175.
2. Lorensen, W. E.; Cline, Harvey E. (1987). "Marching cubes: A high resolution 3d surface construction algorithm". ACM Computer Graphics. 21 (4): 163–169.

Details

Occupancy Networks: Learning 3D Reconstruction in Function Space

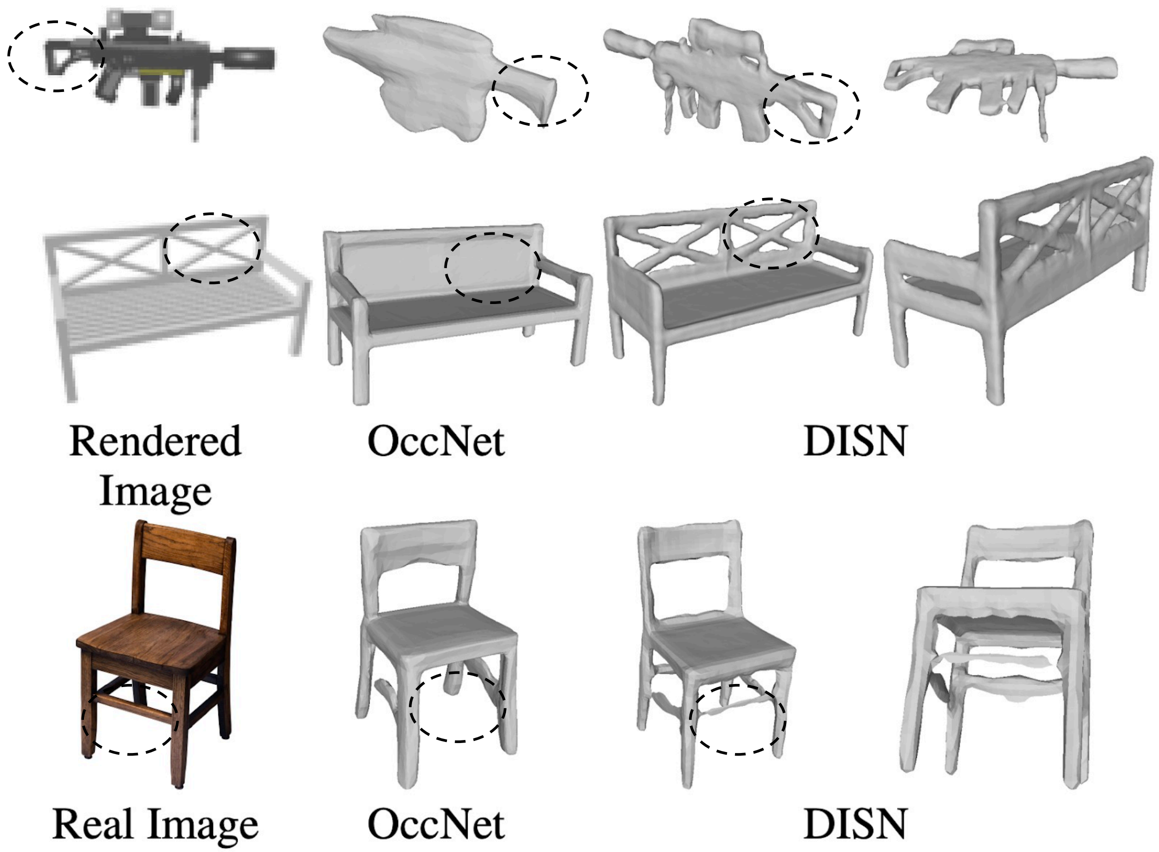


(a) Voxel (b) Point (c) Mesh (d) Ours

Input a single image

Output a 3D mesh model

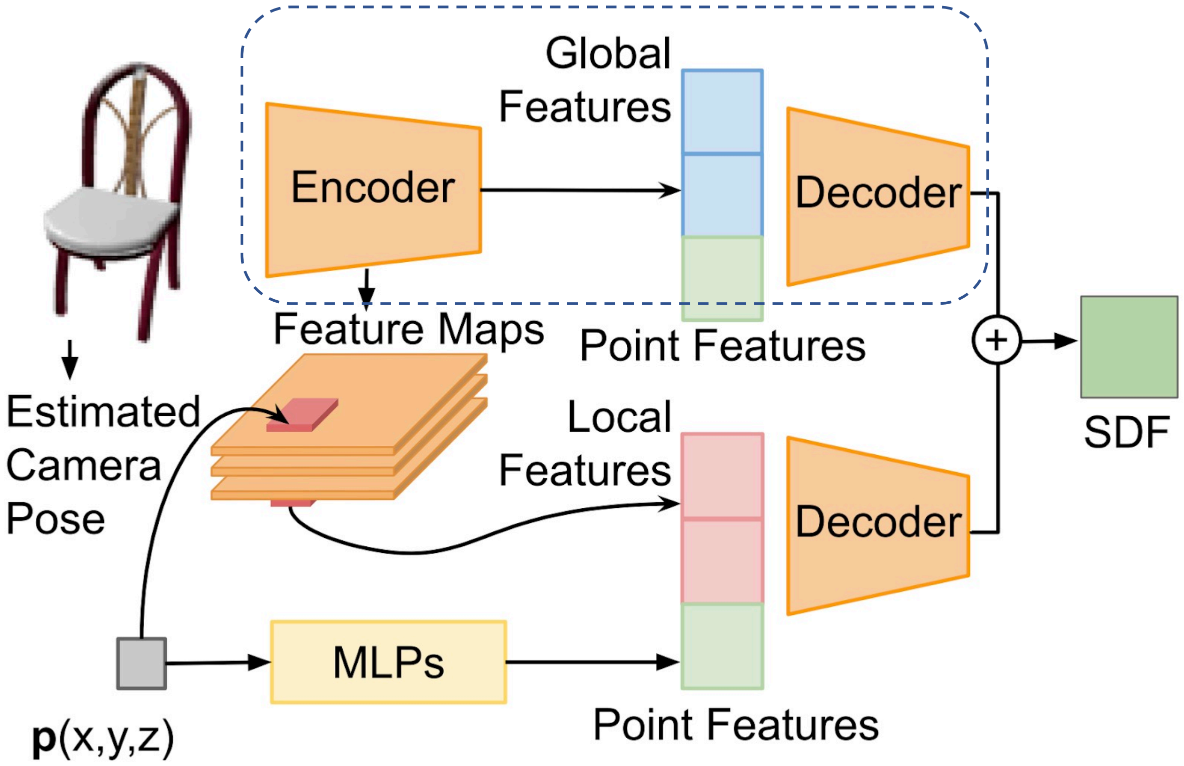
DISN: Deep Implicit Surface Network for High-quality Single-view 3D Reconstruction[1]



First method that can captures fine details & thin structures

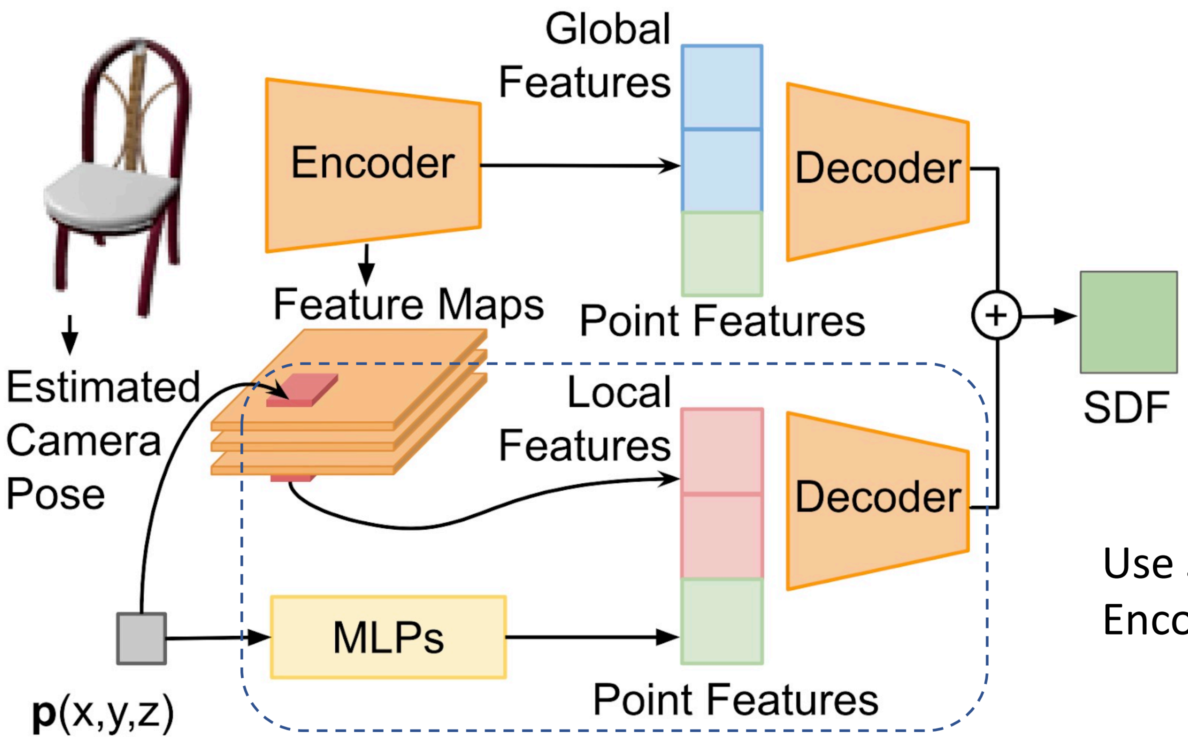
1. Xu, Q., Wang, W., Ceylan, D., Mech, R. and Neumann, U., 2019. Disn: Deep implicit surface network for high-quality single-view 3d reconstruction. In NeurIPS, 2019.

DISN Overview



DISN uses the <u>global features</u>, and the point features to predict the SDF of p.

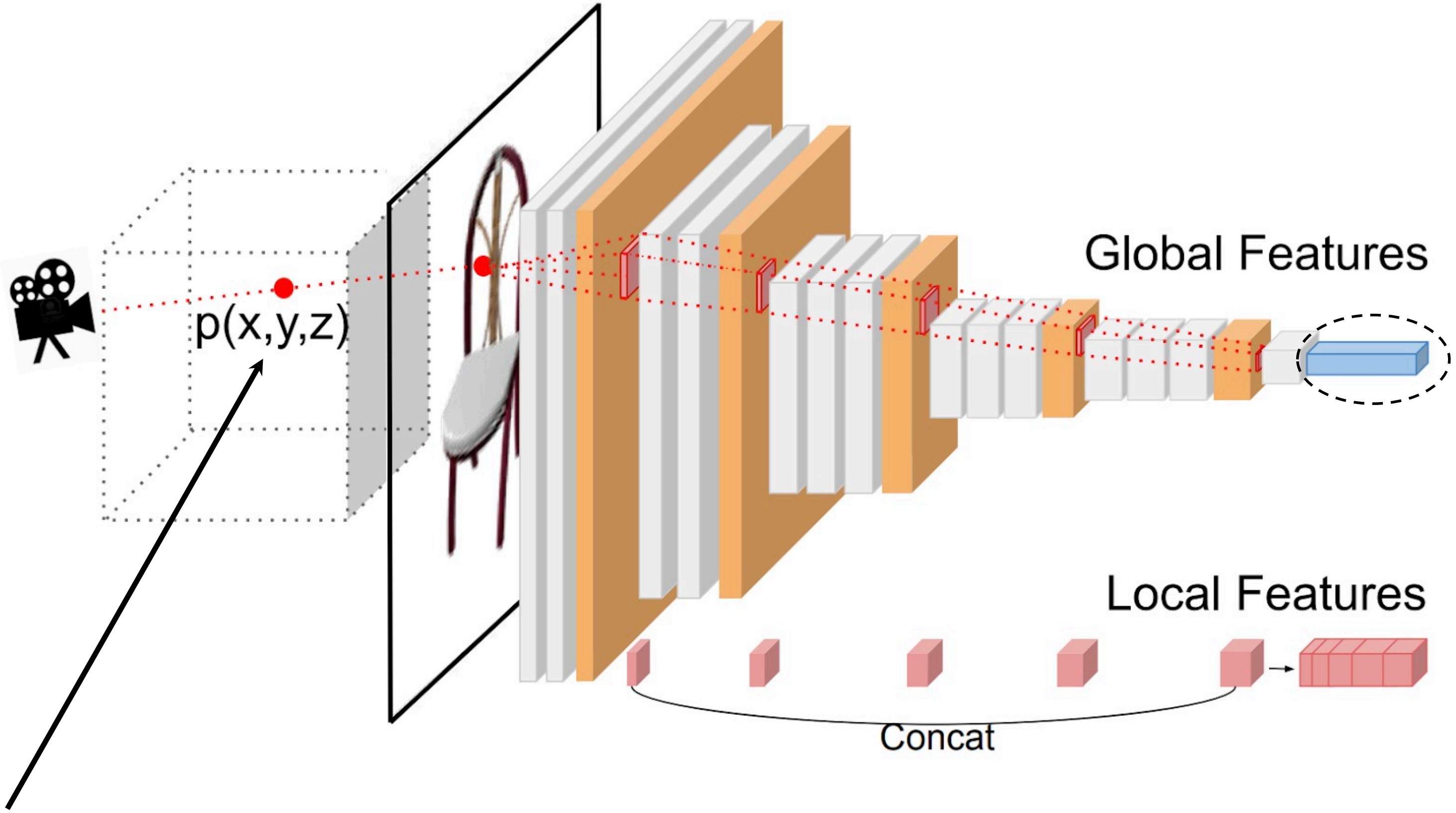
DISN Overview



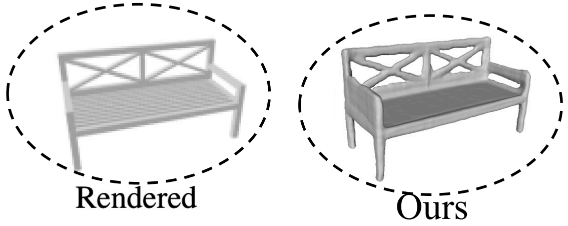
Use SDF(gt) – SDF(ours) to optimize Encoder and Decoders

DISN uses the global features, and the local features to predict the SDF of p.

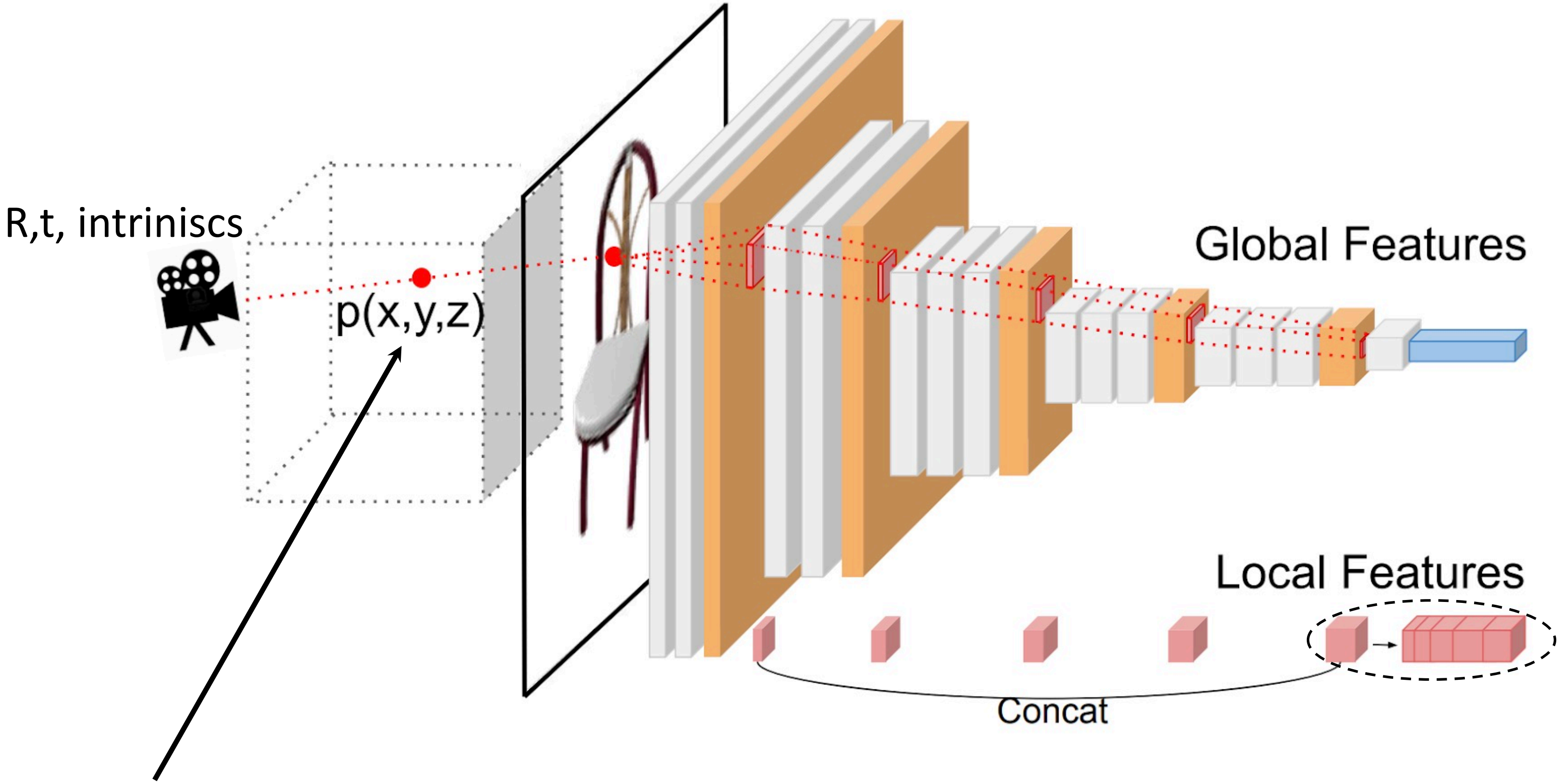
Feature Extraction



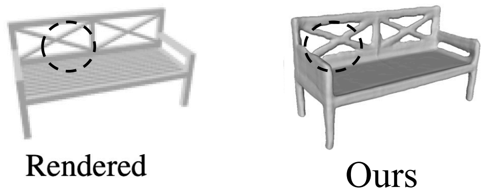
For a sampled
3D point p(x,y,z)



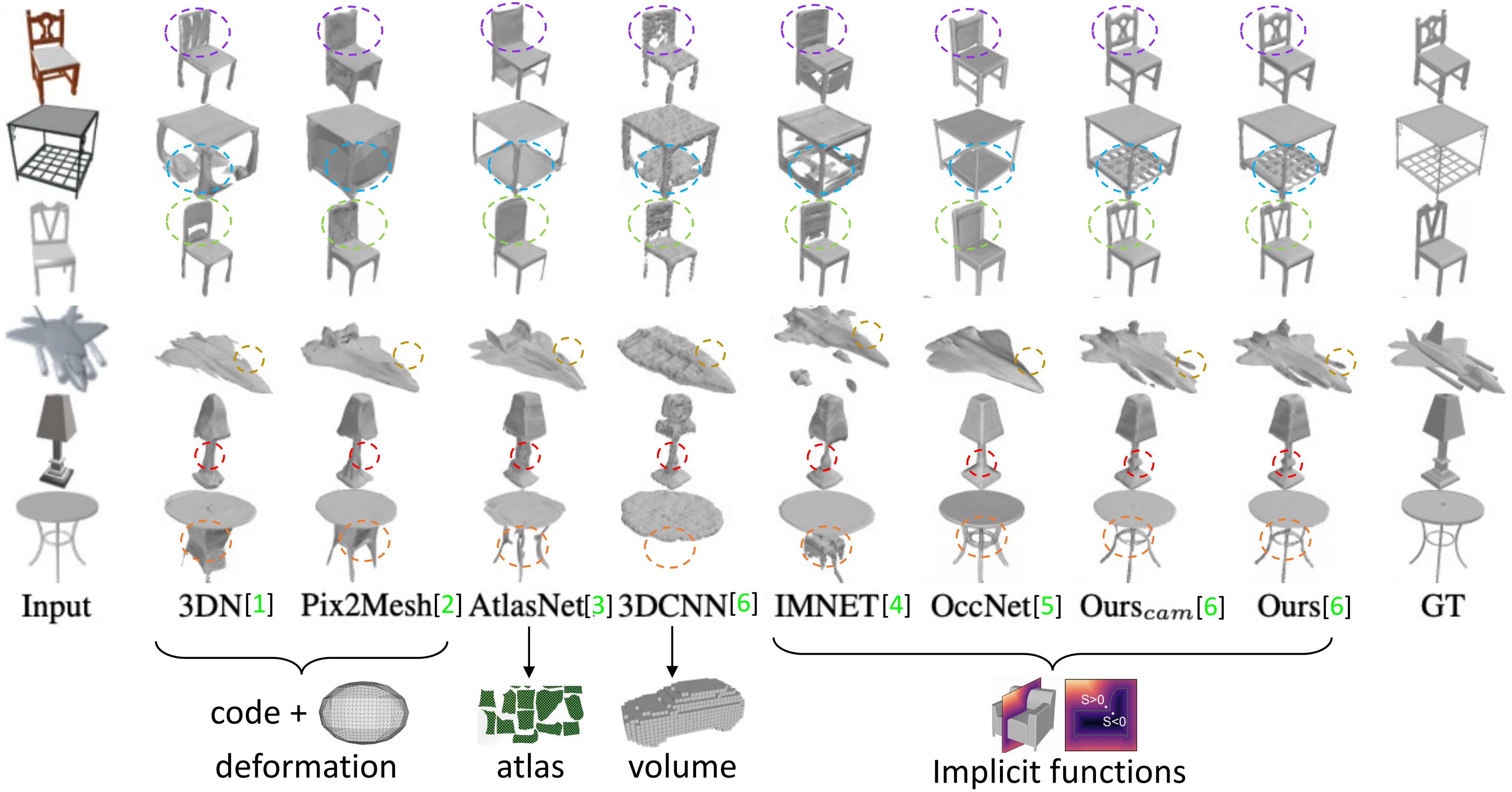
Feature Extraction



For a sampled
3D point p(x,y,z)



Results



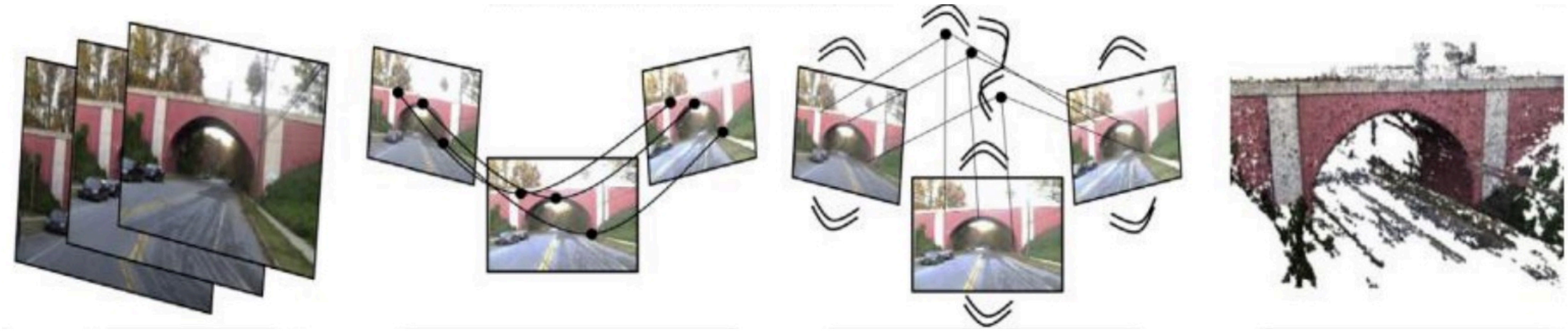
1. Weiyue Wang, Duygu Ceylan, Radomir Mech, and Ulrich Neumann. 3dn: 3d deformation network. In CVPR, 2019.
2. Nanyang Wang, Yinda Zhang, Zhuwen Li, Yanwei Fu, Wei Liu, and Yu-Gang Jiang. Pixel2mesh: Generating 3d mesh models from single rgb images.
3. Thibault Groueix, Matthew Fisher, Vladimir G. Kim, Bryan Russell, and Mathieu Aubry. AtlasNet: A Papier-Mâché Approach to Learning 3D Surface Generation. In CVPR, 2018.
4. Zhiqin Chen and Hao Zhang. Learning implicit fields for generative shape modeling. In CVPR 2019
5. Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space. In CVPR, 2019.
6. Xu, Q., Wang, W., Ceylan, D., Mech, R. and Neumann, U., 2019. Disn: Deep implicit surface network for high-quality single-view 3d reconstruction. In NeurIPS, 2019.

Section 2

3D Geometry Reconstruction from Multi-view 2D images.

- MVS method

Multi-view Reconstruction (MVS)



Multiview 2D images

Feature Matching

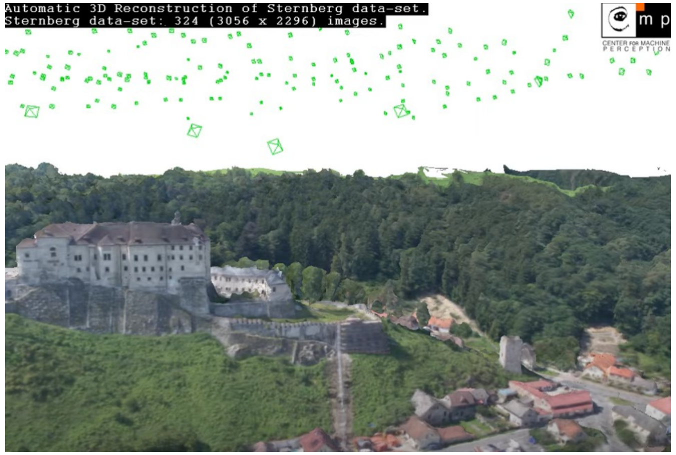
Triangulation &
Bundle Adjustment

Point Cloud
Reconstruction

(Explicit representation)

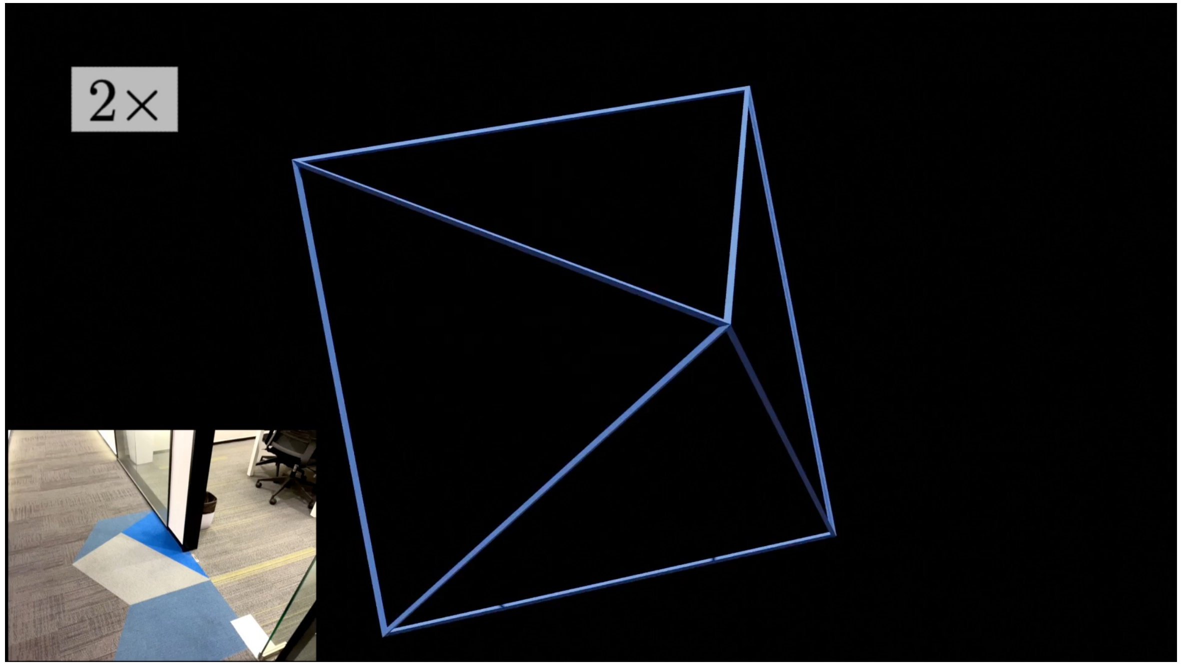
Multi-view Reconstruction (MVS)

Example 1. [Large-scale scene by drone]

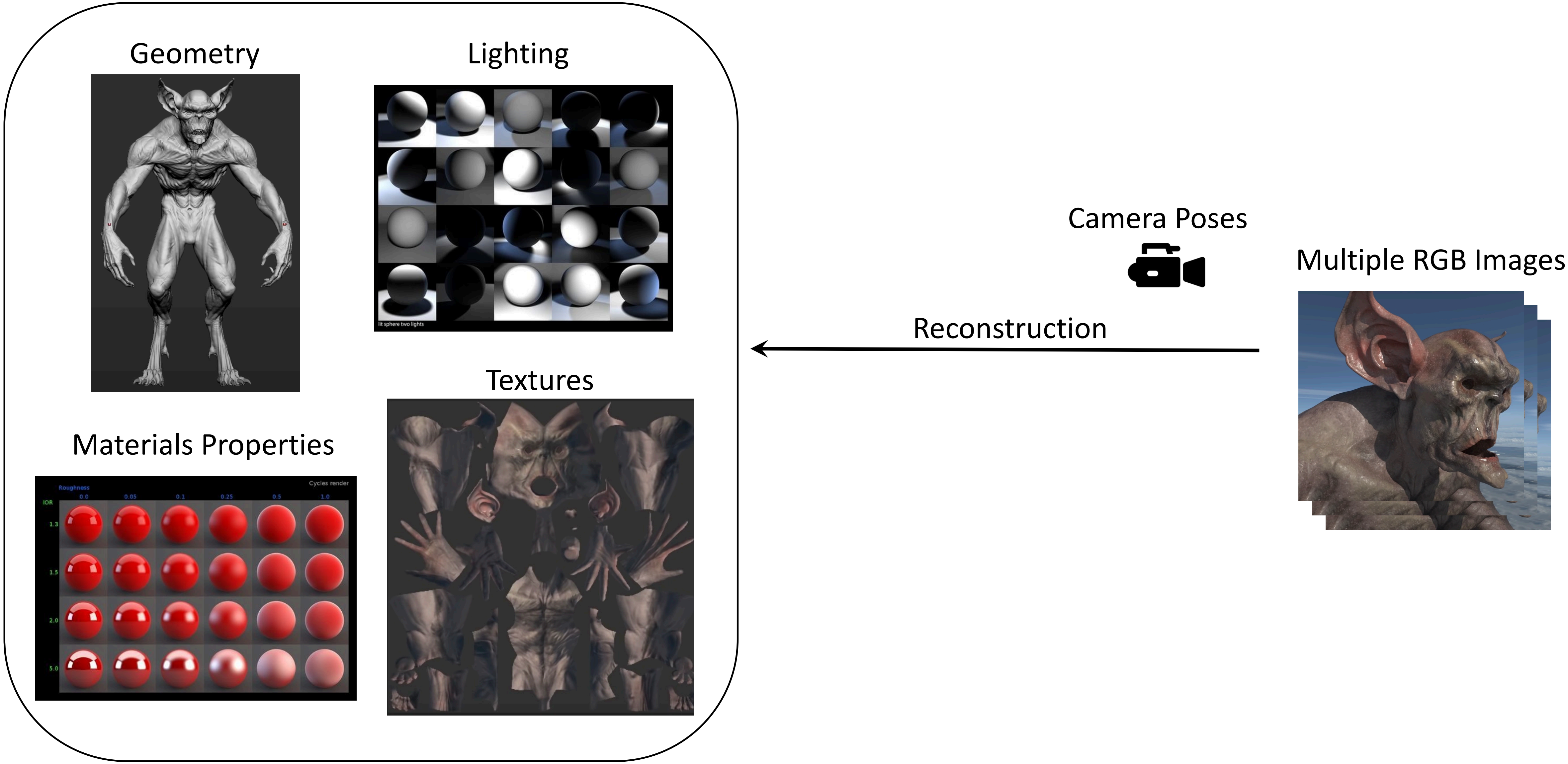


Example 2, Deep Learning Method

Sun, J., Xie, Y., Chen, L., Zhou, X. and Bao, H., 2021. NeuralRecon: Real-time coherent 3D reconstruction from monocular video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 15598-15607).



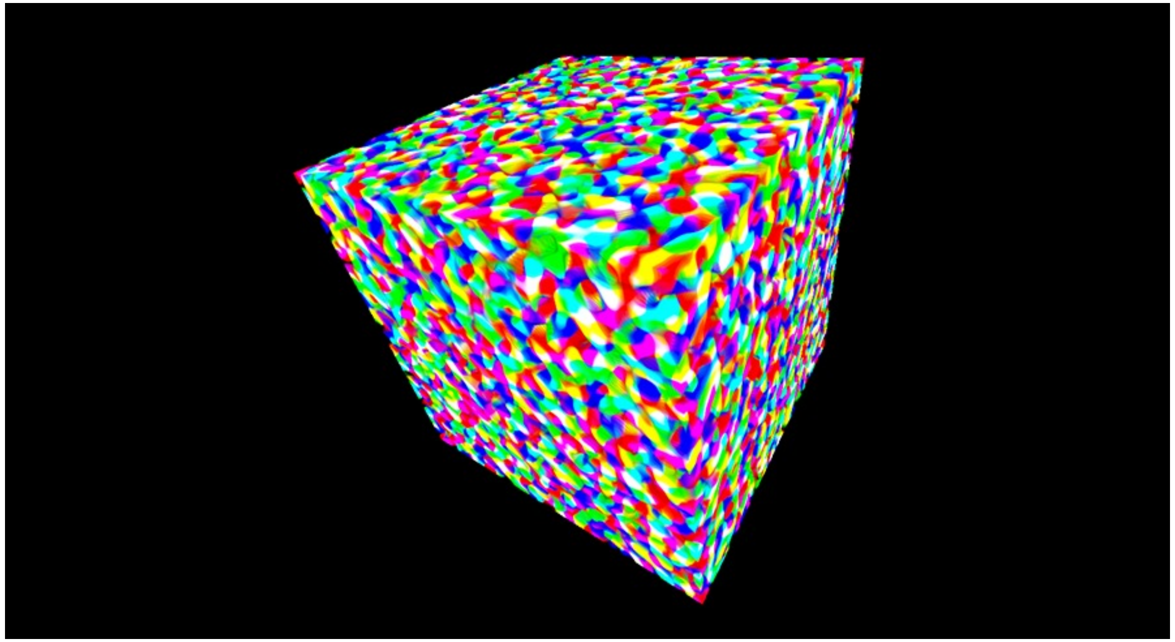
Decoupling 3D Scene Components

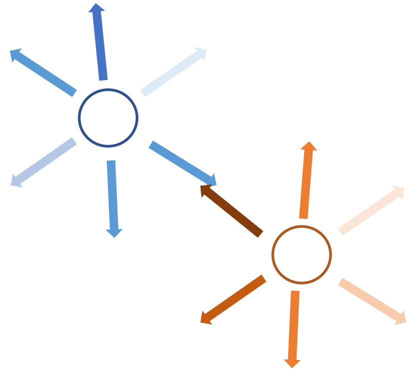


Section 3

- 3D Scene Reconstruction from Multi-view 2D images.
  - Reverse Volume Rendering

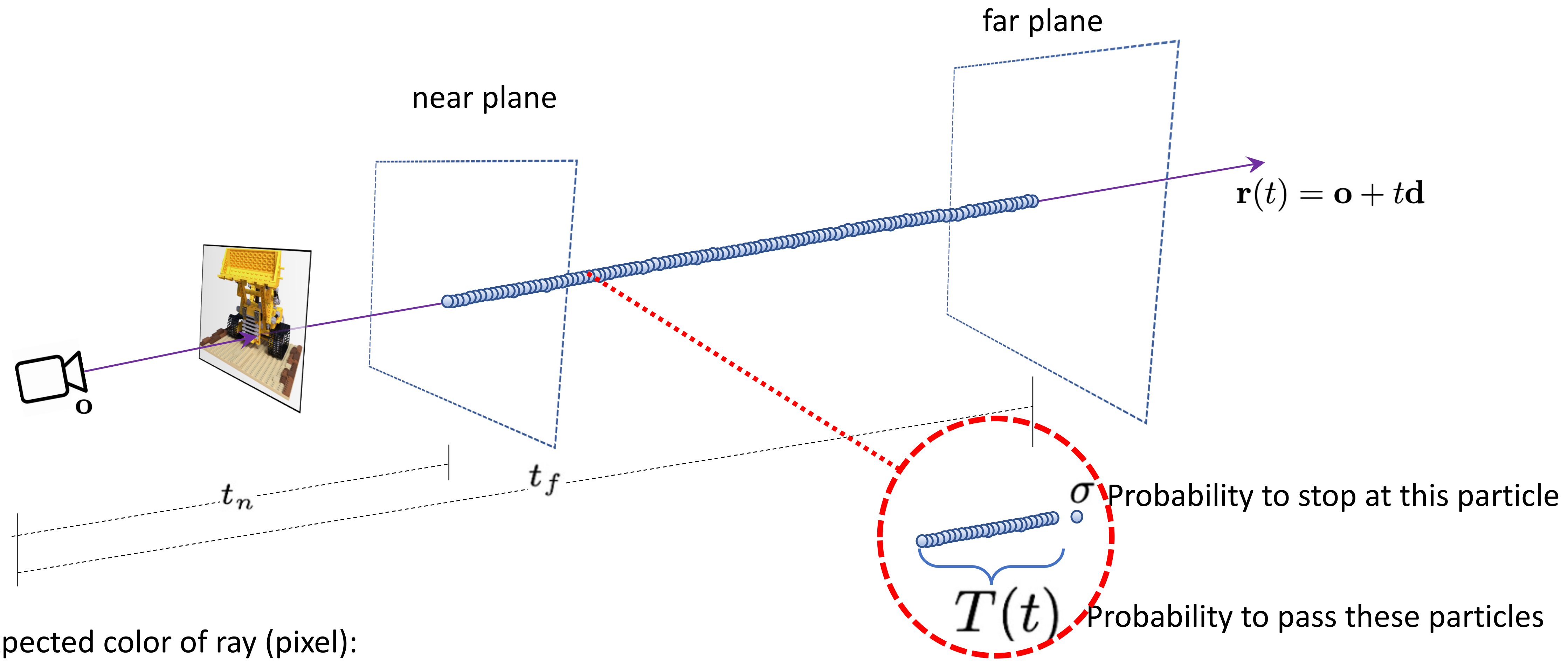
Volume Rendering (Emit Absorb Model)





1. Emit different radiance (rgb color) to all directions
2. Absorb radiance by a probability σ (density)

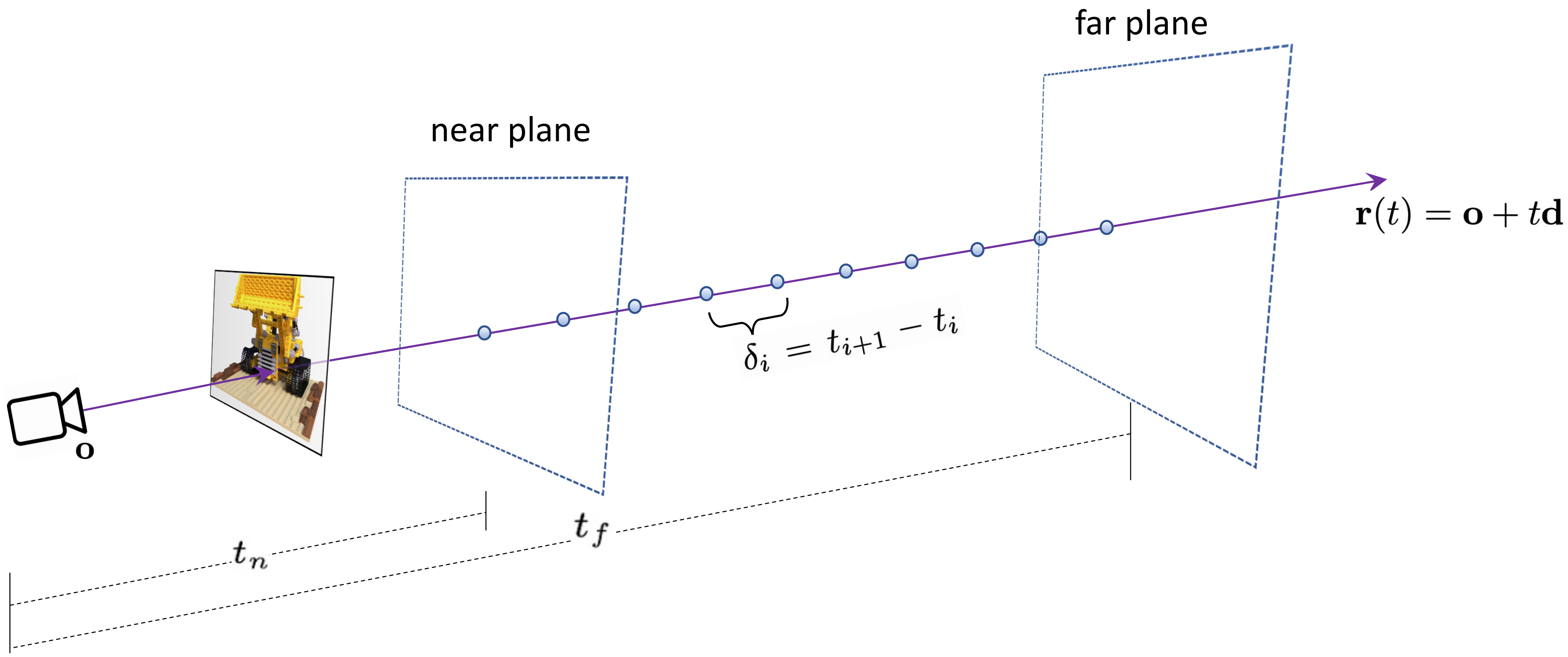
Volume Rendering (Emit Absorb Model)



Expected color of ray (pixel):

$$C(\mathbf{r}) = \int_{t_n}^{t_f} \underbrace{T(t)}_{\text{transmittance}} \overbrace{\sigma(\mathbf{r}(t))}^{\text{density}} \overbrace{\mathbf{c}(\mathbf{r}(t), \mathbf{d})}^{\text{radiance color}} dt\,, \text{ where } T(t) = \exp\left(-\int_{t_n}^{t} \sigma(\mathbf{r}(s))ds\right)$$

Volume Rendering

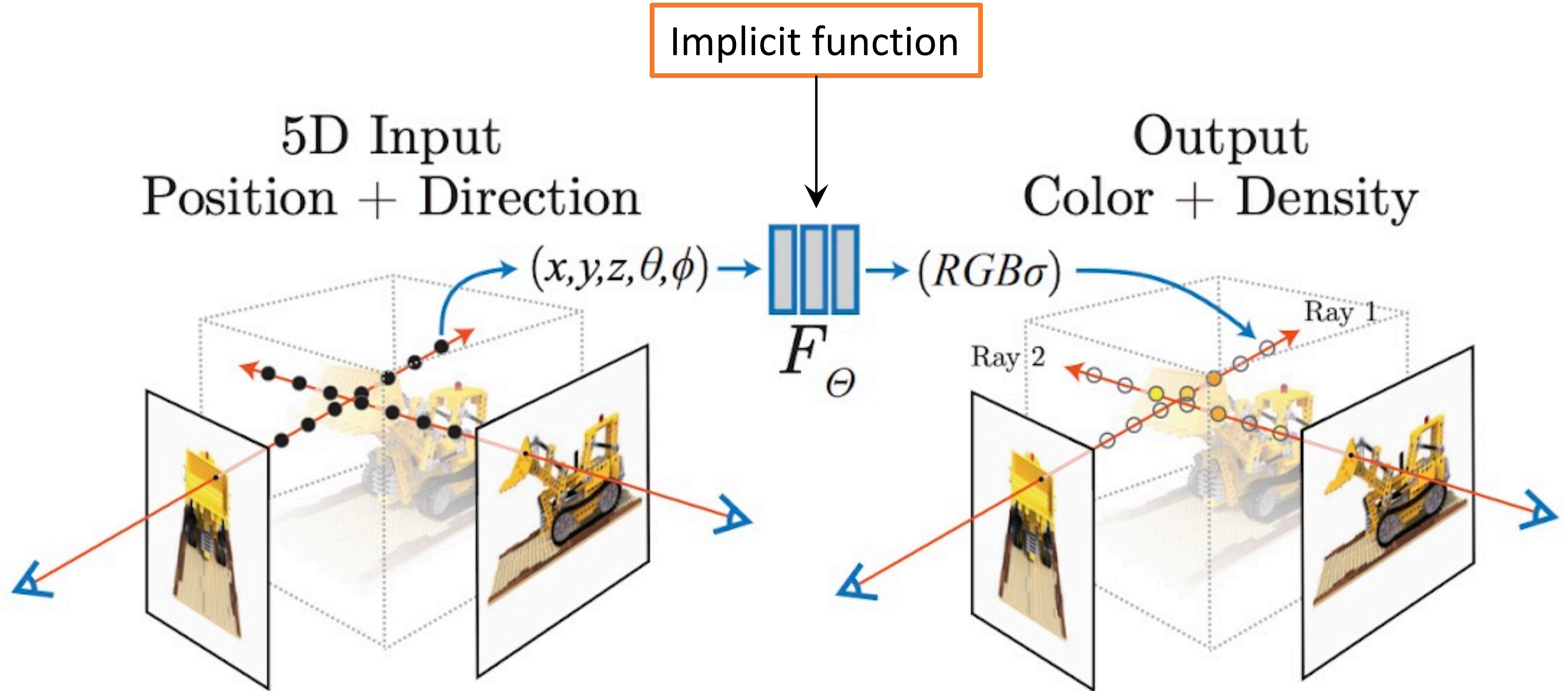


Expected color of ray [1]:

$$\hat{C}(\mathbf{r}) = \sum_{i=1}^{N} T_i(1 - \exp(-\sigma_i \delta_i))\mathbf{c}_i \,, \text{ where } T_i = \exp\left(-\sum_{j=1}^{i-1} \sigma_j \delta_j\right)$$

1. Max, N.: Optical models for direct volume rendering. IEEE Transactions on Visualization and Computer Graphics (1995)

Neural Radiance Fields [1]



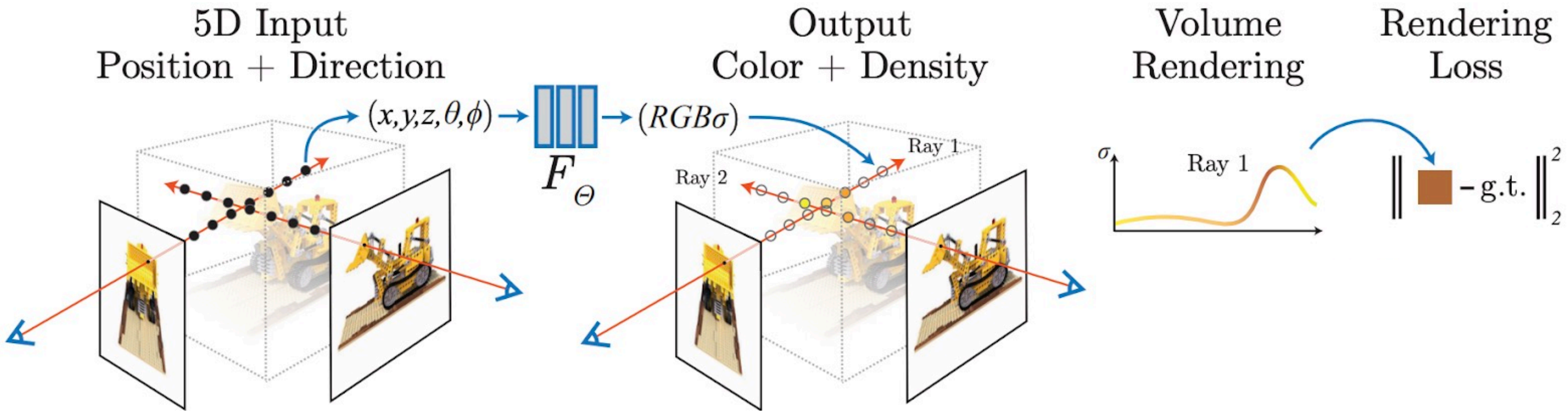
Find the ray from camera,
and sample shading points

Compute radiance at shading points
and accumulate along rays to render pixels

1. Mildenhall, Ben, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. "Nerf: Representing scenes as neural radiance fields for view synthesis." ECCV (2020)

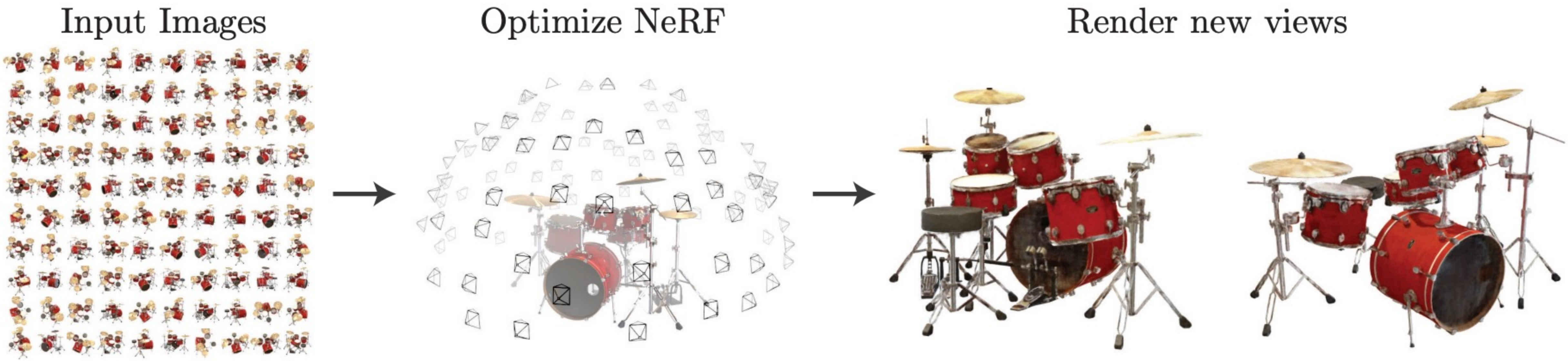
Optimizing Neural Radiance Fields

Forward — Radiance field (represented by F_Θ) $\xrightarrow{\text{Differentiable ray marching}}$ RGB Images

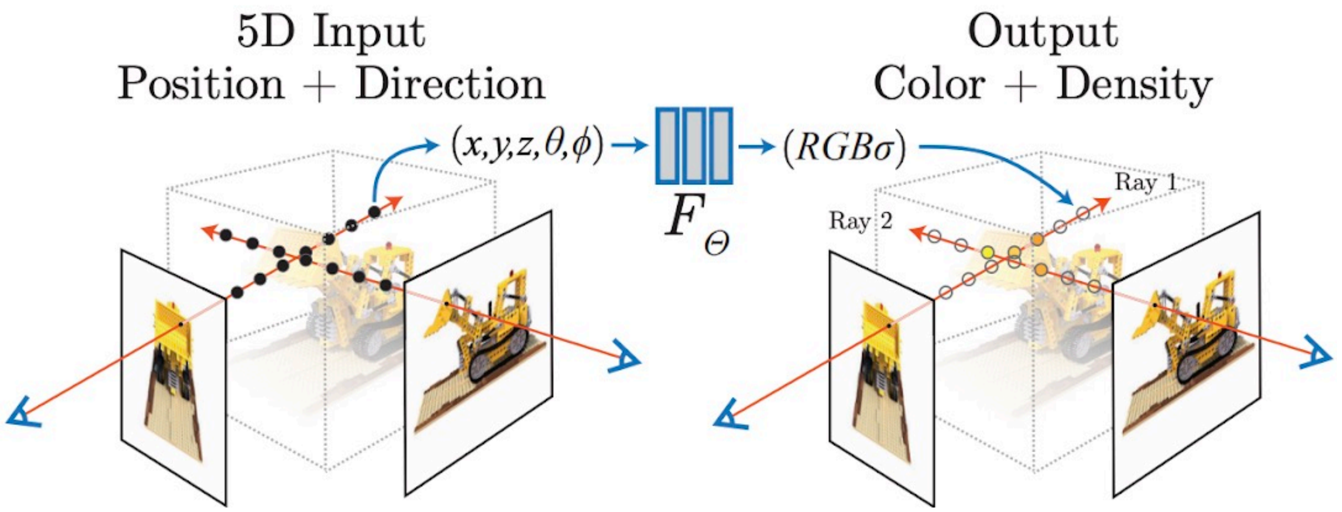


Backward — Radiance field $\xleftarrow[\text{Gradient Descent}]{\text{Back Propagation}}$ Pixel Loss on RGB Images

Per-scene Optimization



Problems with Neural Radiance Fields



A MLP encodes the entire scene

No awareness of the shape priors.

Inefficient Rendering (45s)

Slow Convergence (1-2 days).

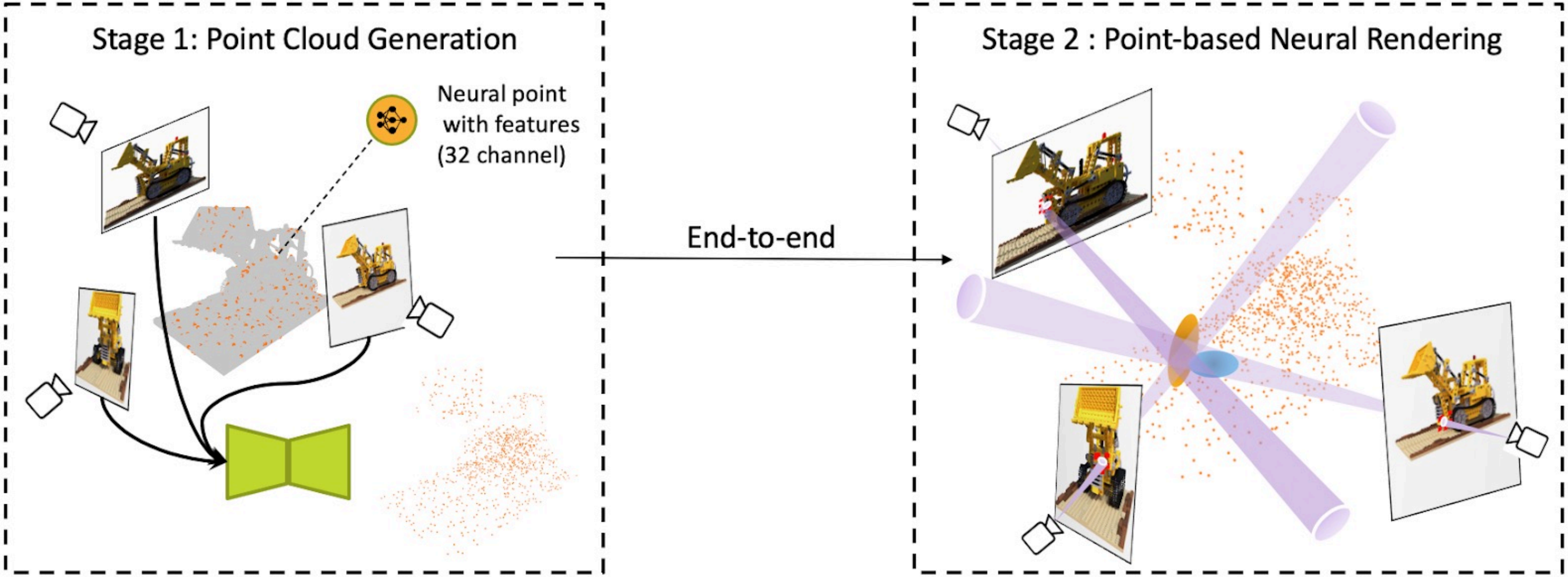
Limited Scalability

Not Generalizable

Point-based Neural Radiance Fields

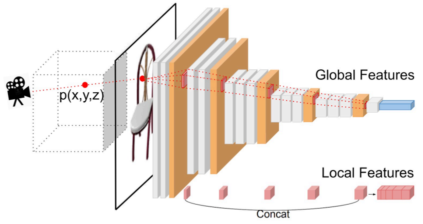
- More accurate reconstruction.
- Faster convergence.
- Able to take RGB image inputs only or both points and RGB images as input.
- Fix the holes and outliers for point inputs (enhance geometry modeling).

Point-NeRF: Point-based Neural Radiance Fields

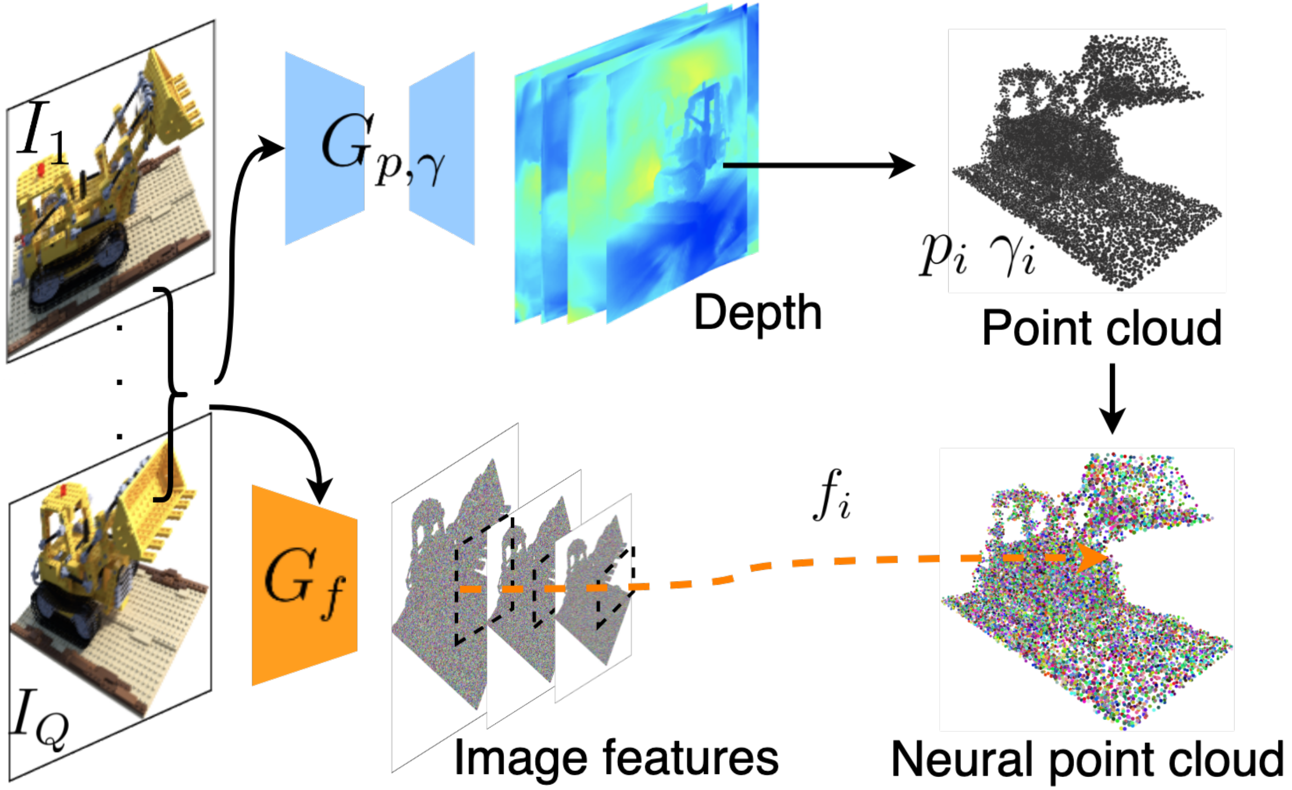


A novel local radiance representation to achieve
scalable, efficient, and generalizable neural rendering

Generating initial point-based radiance fields



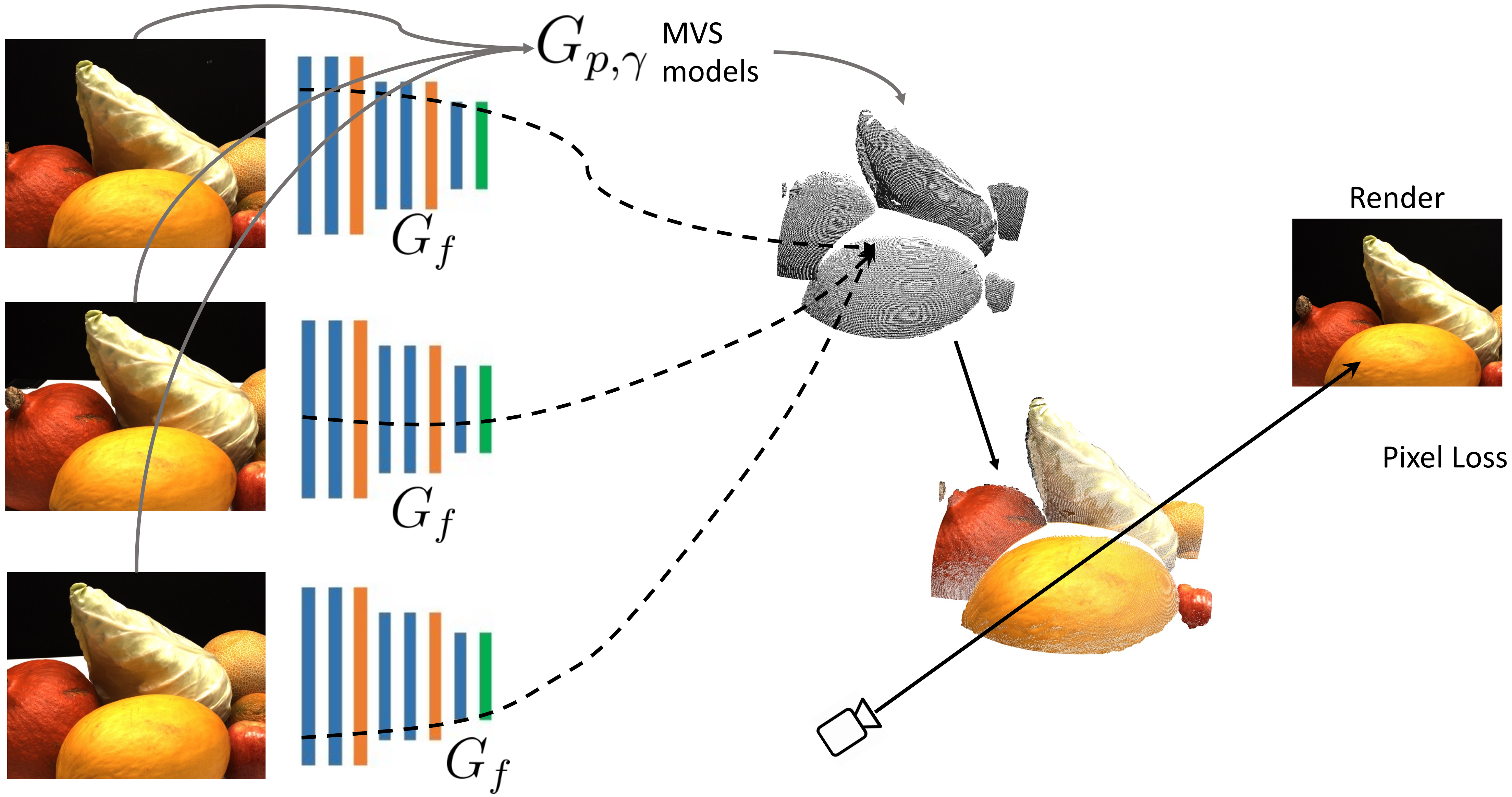
DISN



$G_{p,\gamma}$: a 3DCNN generates depth (points) p_i and confidence γ_i

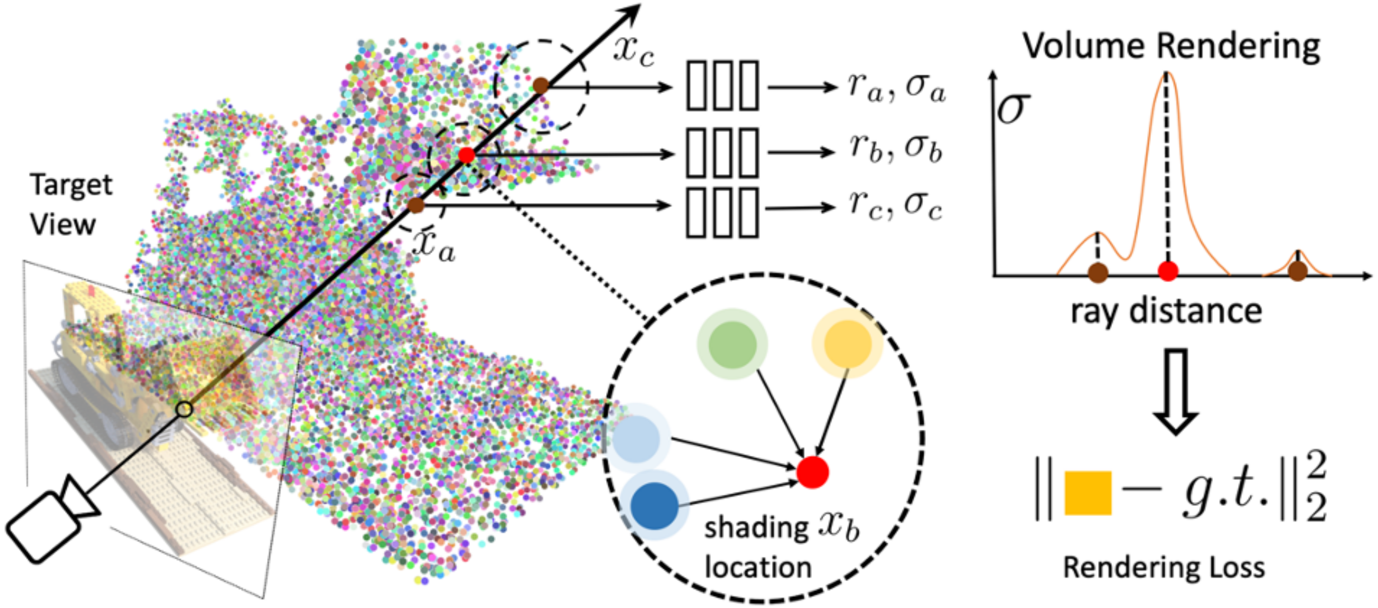
G_f : a 2D CNN extracts 2D features and project to points f_i

Next Steps: Point Initialization



Optimizing Point-based Radiance Fields

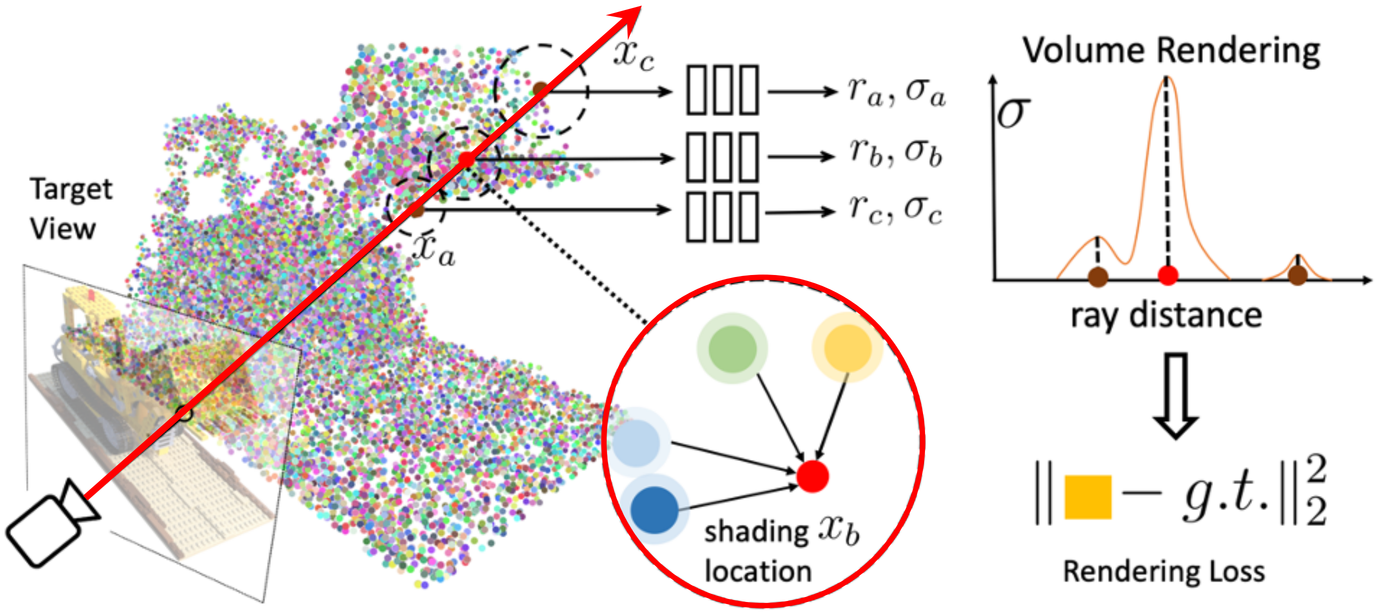
Forward — Point features / Implicit functions $\xrightarrow{\text{Shading + Differentiable Rendering}}$ RGB Images



Backward — Point features / Implicit functions $\xleftarrow{\text{Back Propagation / Gradient Descent}}$ Pixel Loss on RGB Images

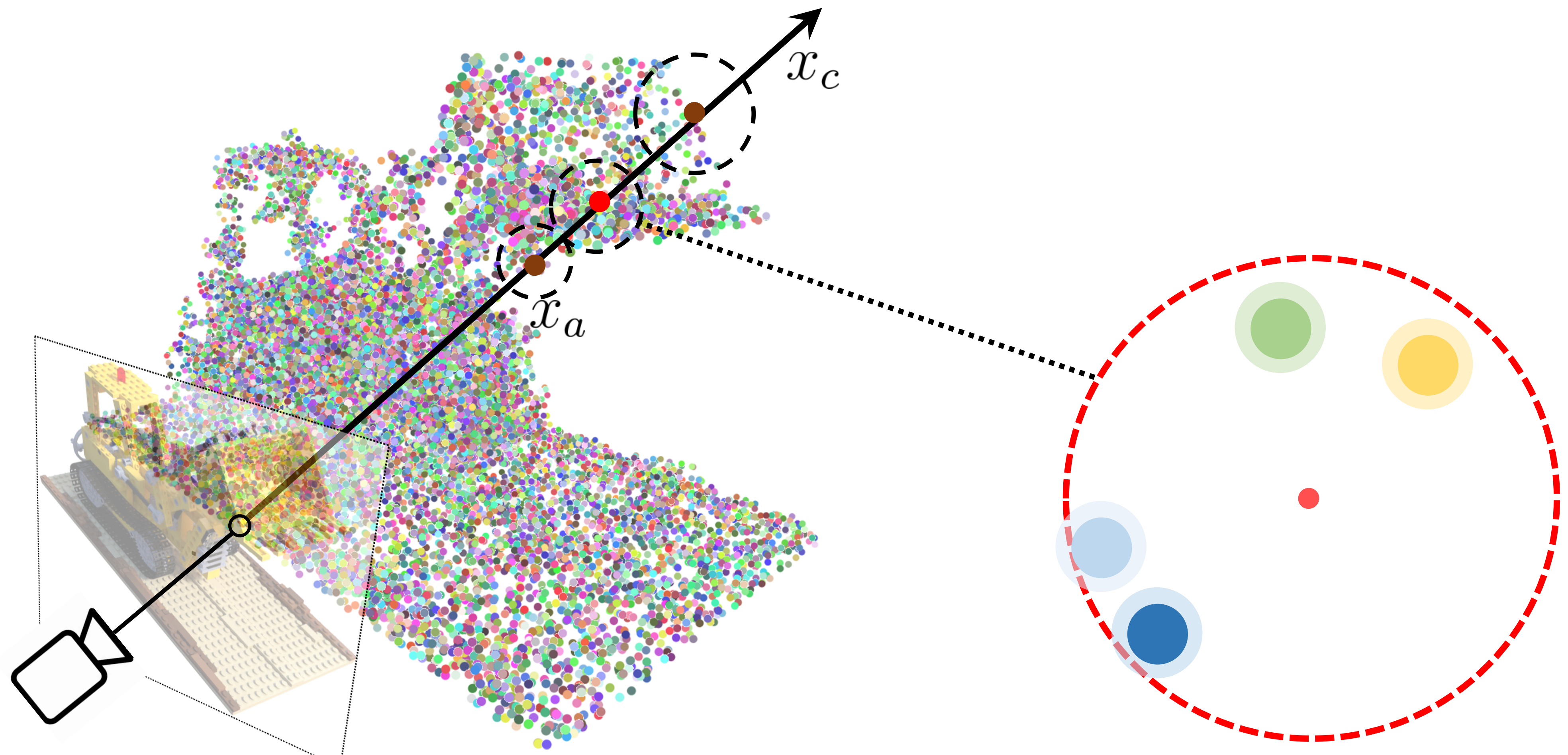
Point-based Radiance Fields

Point Neighbors Searching

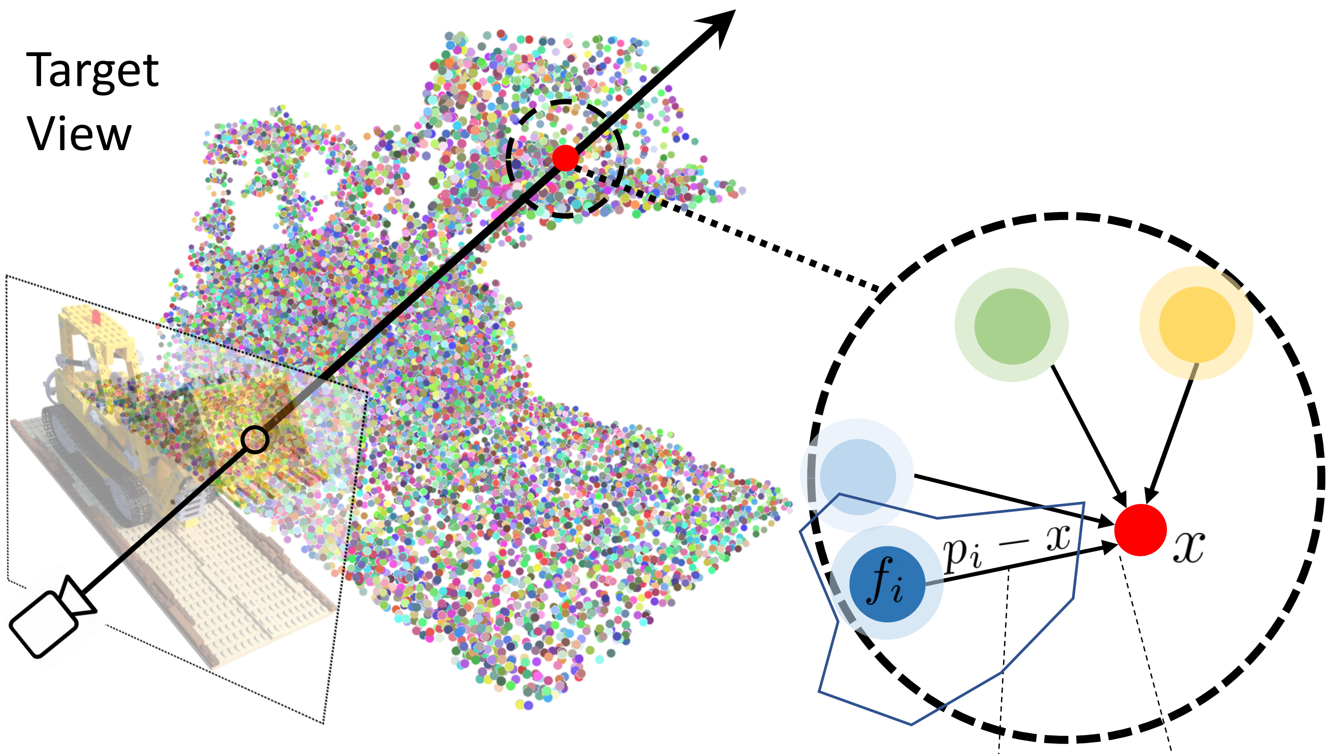


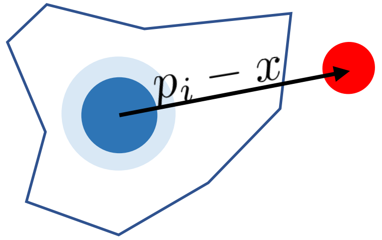
Point Neighbor Grouping

Neural Point Grouping



Compute Radiance and Density





$\mathrm{F}(f_i\,,\ p_i - x)$

Describe the local geometries and radiance colors

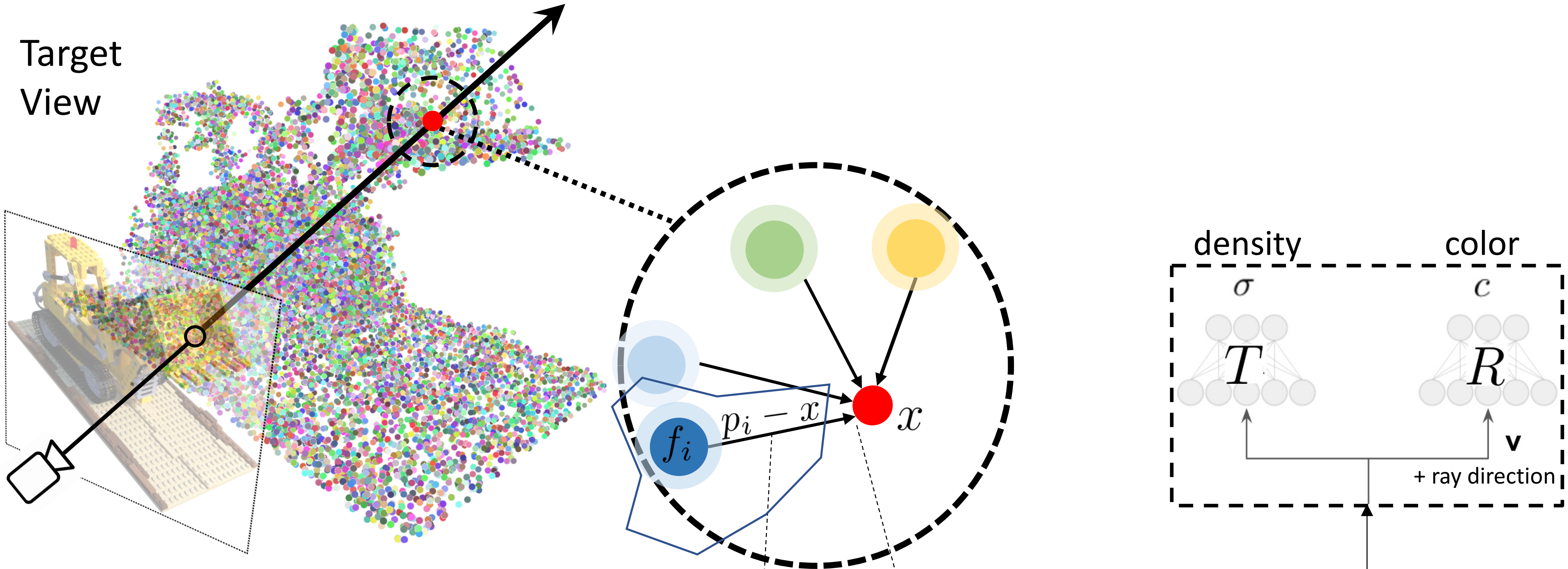
$$w_i = \frac{1}{||p_i - x||_2}$$

normalization

point features

$$\frac{1}{w_1 + \ldots + w_i}\left(w_1 \times \mathrm{F}(f_1, p_1 - x) + \ldots\ldots + w_i \times \mathrm{F}(f_i, p_i - x)\right)$$

Compute Radiance and Density



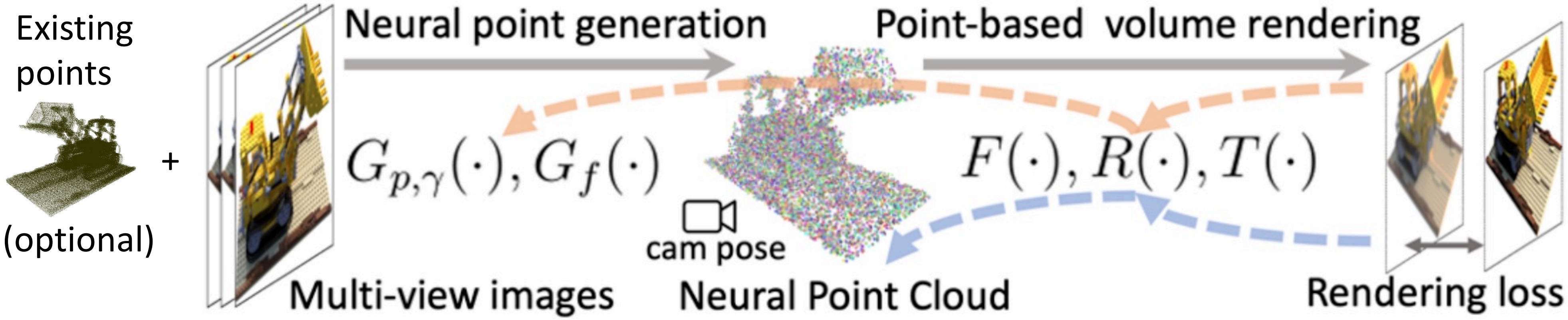
$$w_i = \frac{1}{||p_i - x||_2}$$

normalization

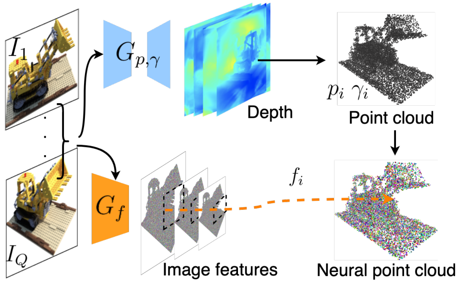
point features

$$\frac{1}{w_1 + \ldots + w_i} \left(w_1 \times \mathrm{F}(f_1, p_1 - x) + \ldots\ldots + w_i \times \mathrm{F}(f_i, p_i - x) \right)$$

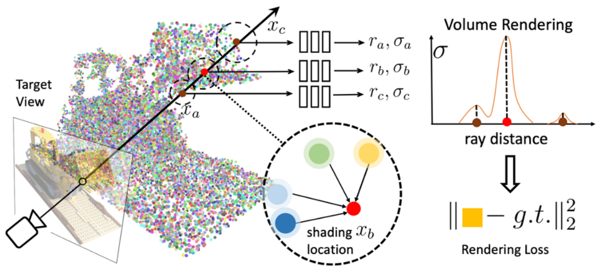
Point-NeRF Gradient Updates



Gradient updates for radiance field initialization and per-scene optimization



Stage 1. Neural Points Generation/Initialization.



Stage 2. Neural Points Per-scene optimization.

Results

PSNR 30.09 at only 20K steps

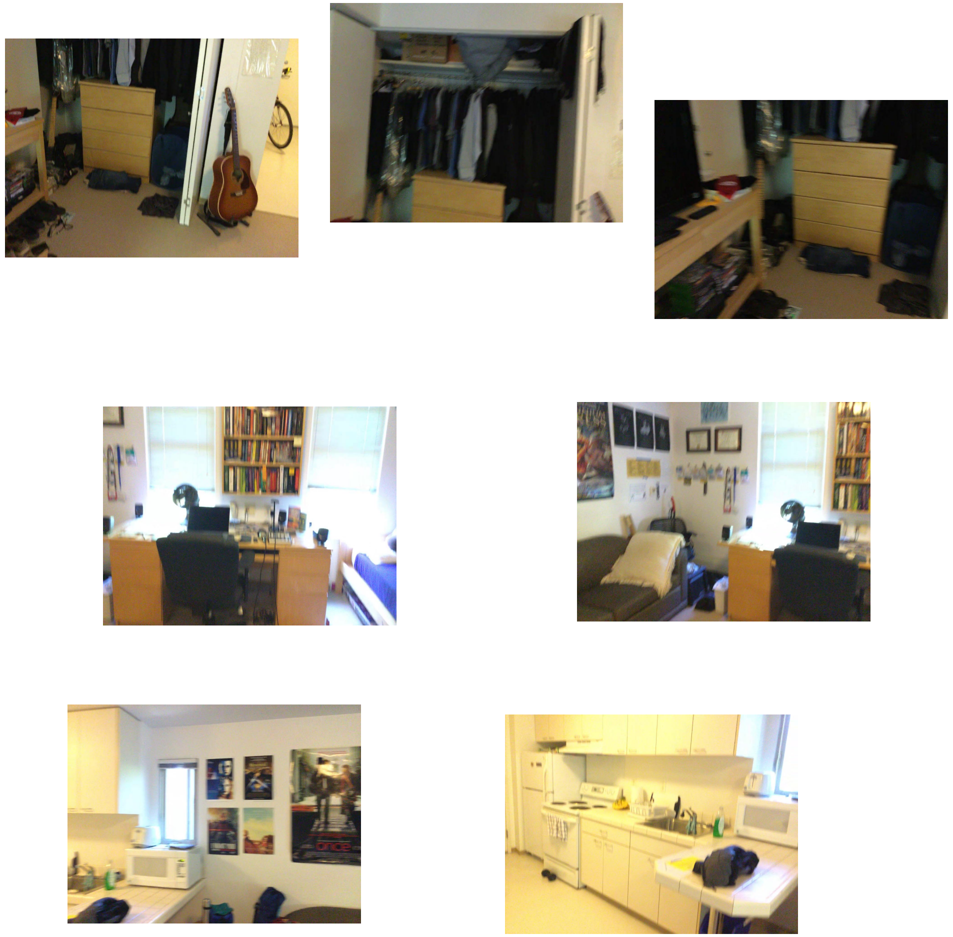
Results

Point-NeRF
PSNR: 30.82

NeRF
PSNR:28.65

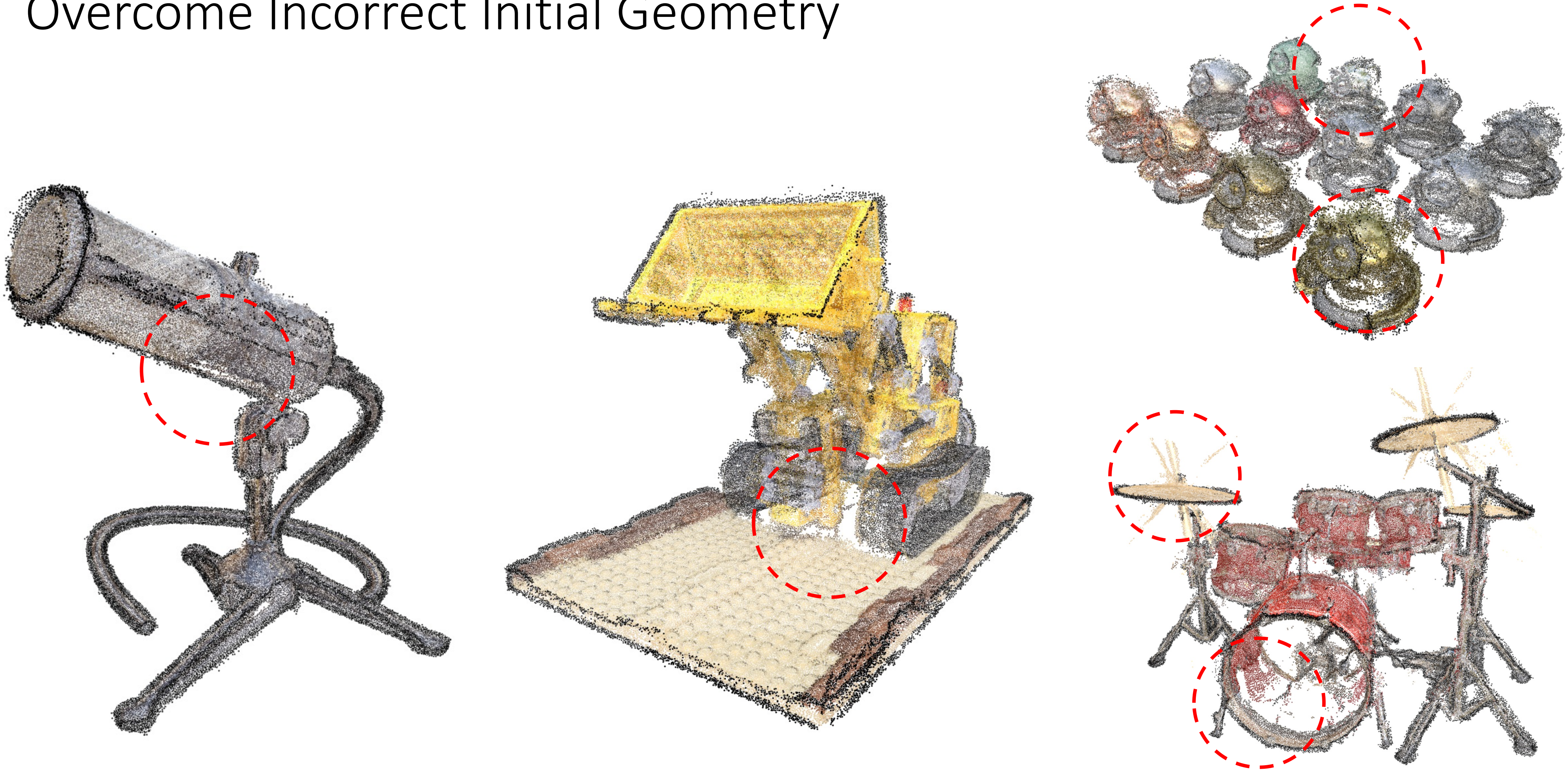
Results on ScanNet

100 static input images as input



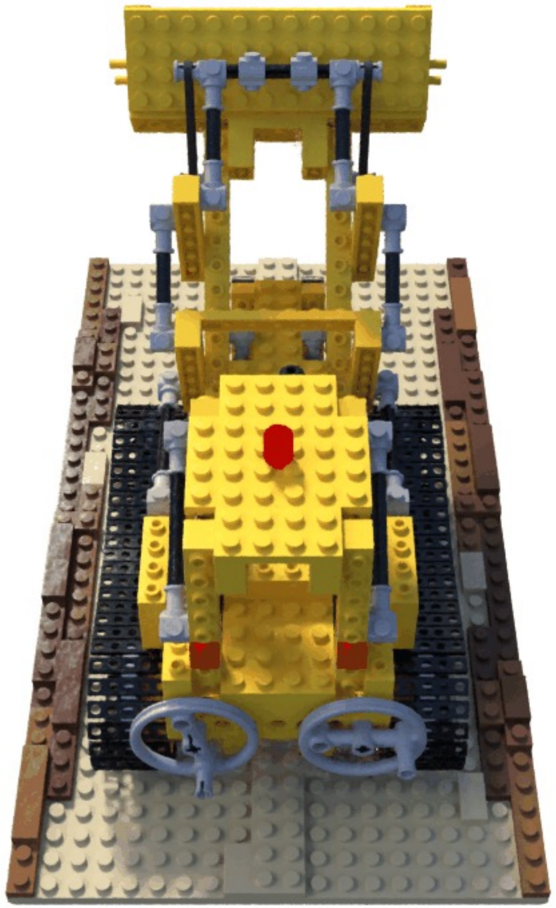
Render a video by a test camera route (1000 views)

Overcome Incorrect Initial Geometry

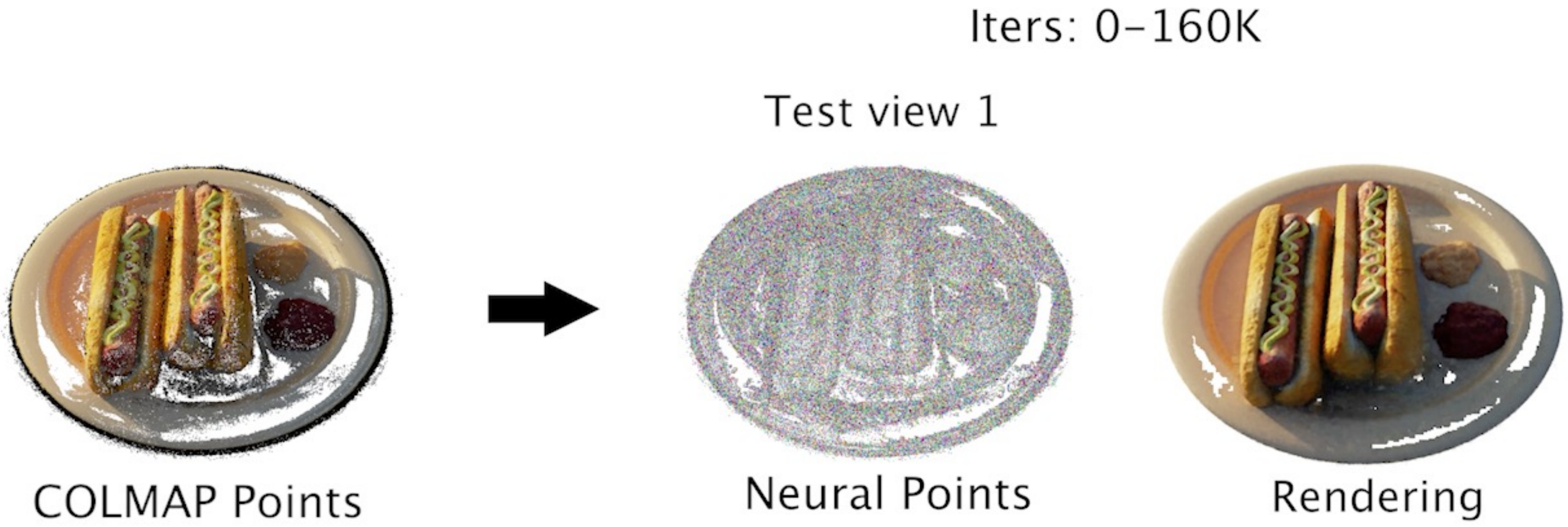


COLMAP Points (400K)

Results with Incorrect Initial Points

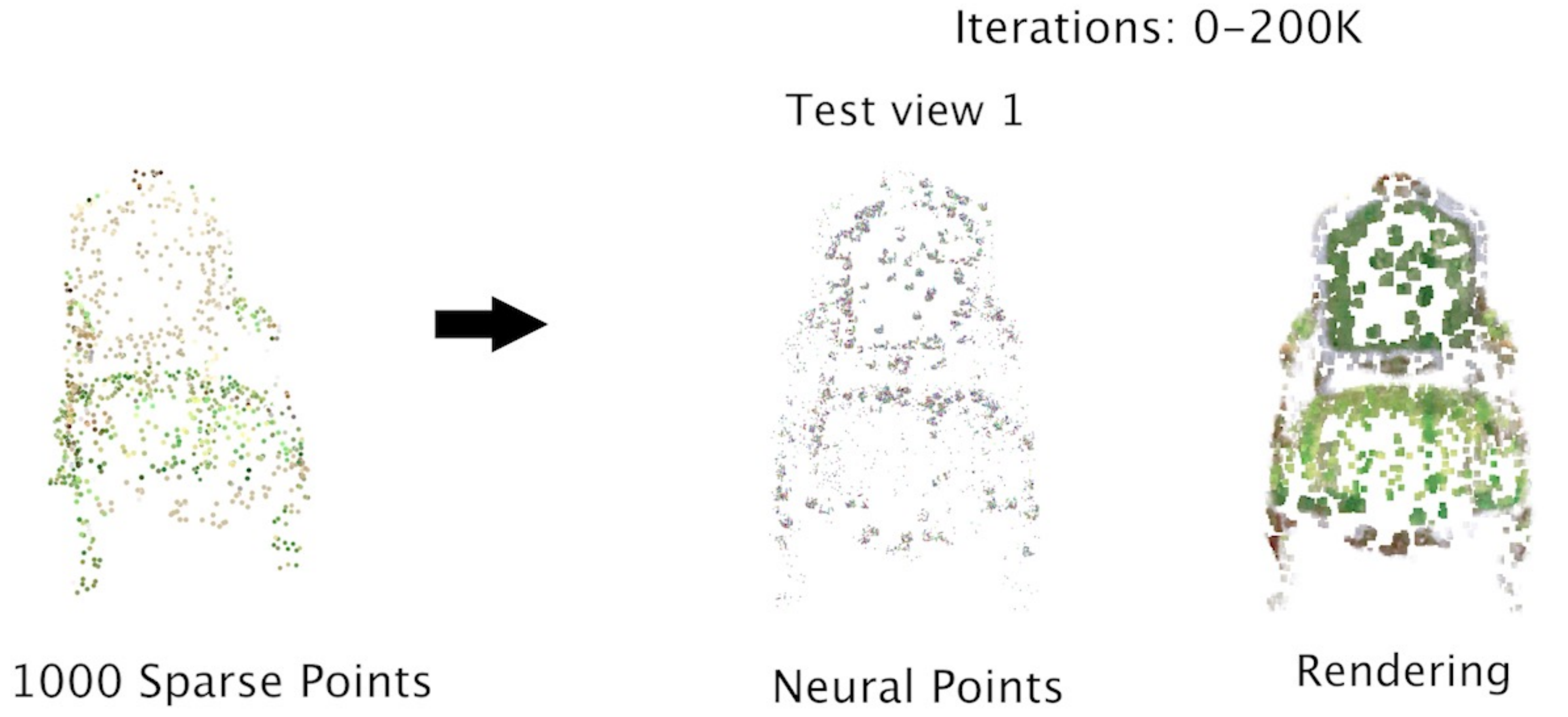


Results with Incorrect Initial Points



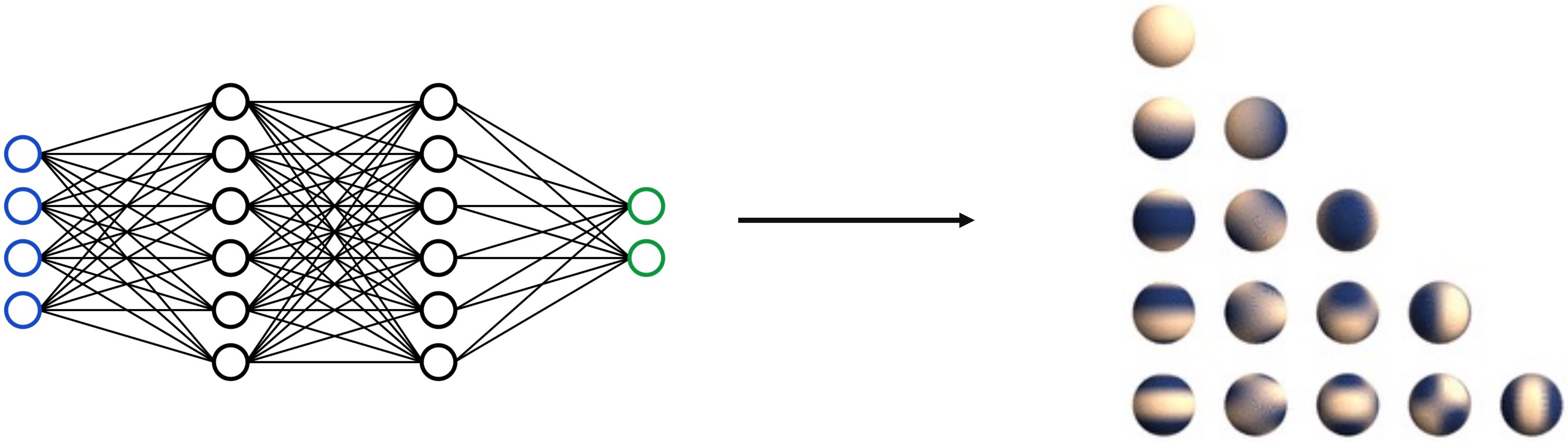
Grow and Prune Points

Grow Dense Points from 1000 Sparse Points



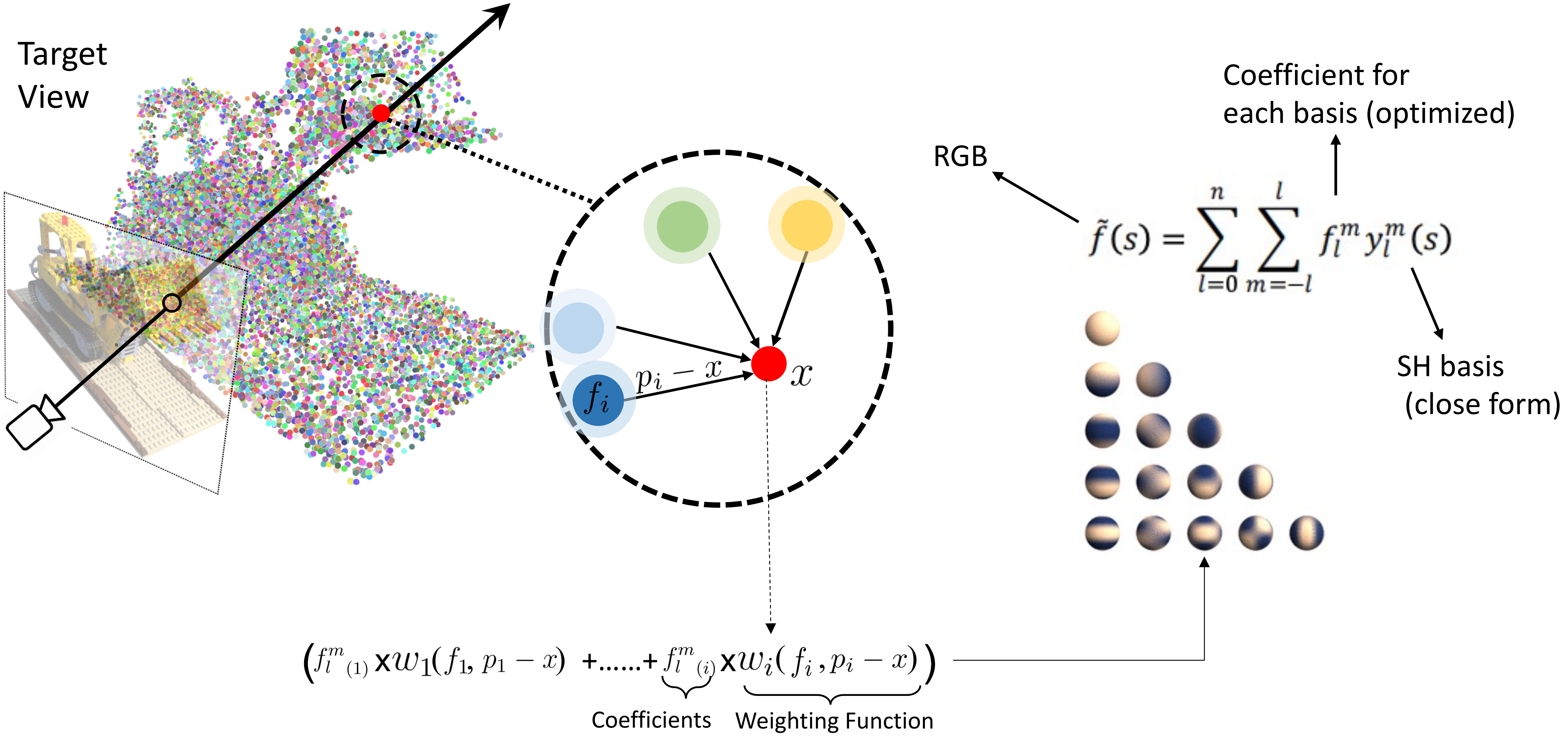
Thank you!

Future Directions: Spherical Harmonics [1] for Acceleration



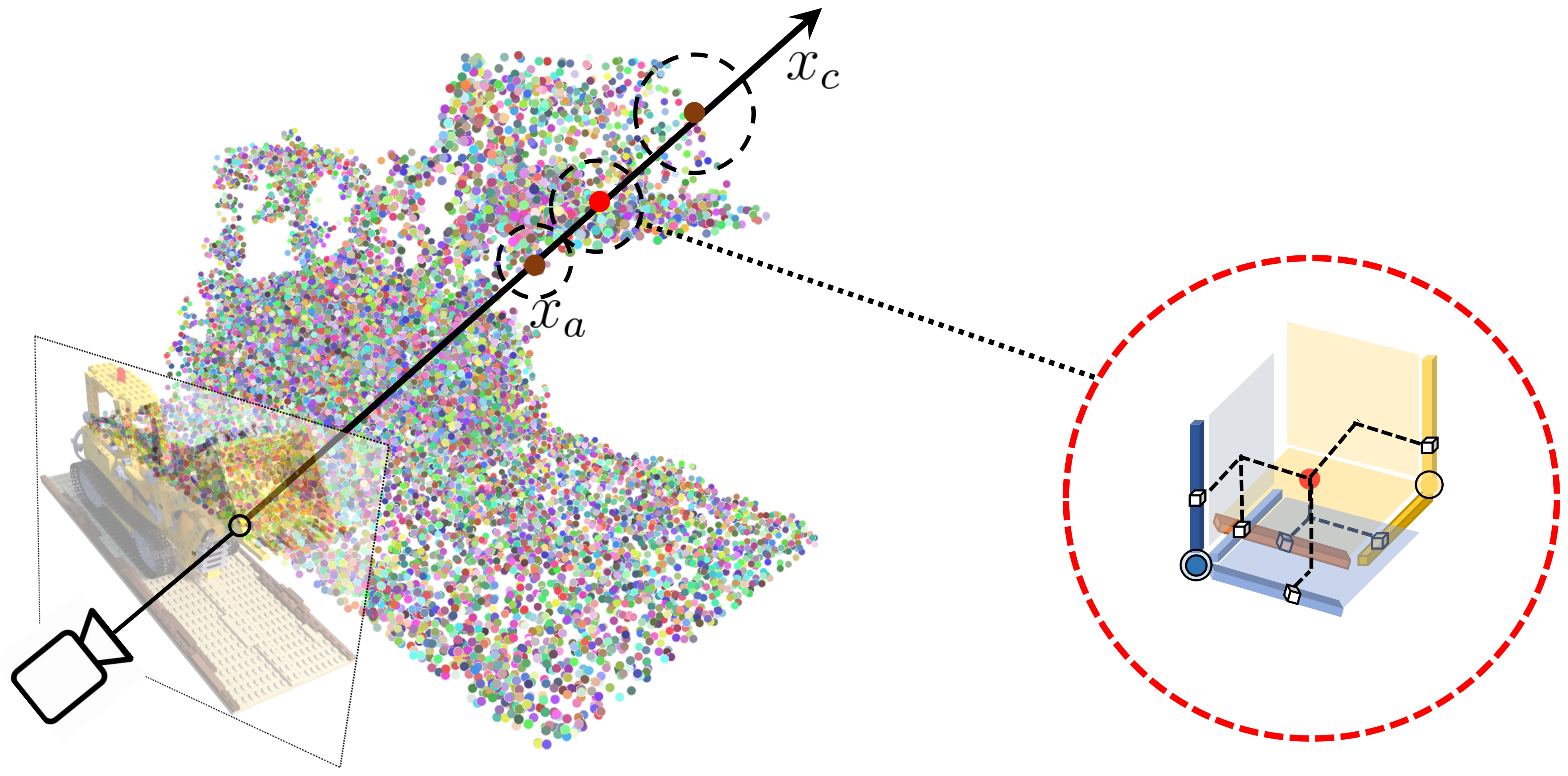
1. Ramamoorthi, Ravi, and Pat Hanrahan. "An efficient representation for irradiance environment maps." In Proceedings of the 28th annual conference on Computer graphics and interactive techniques, pp. 497-500. 2001.

Future Directions: Spherical Harmonics [1] for Acceleration

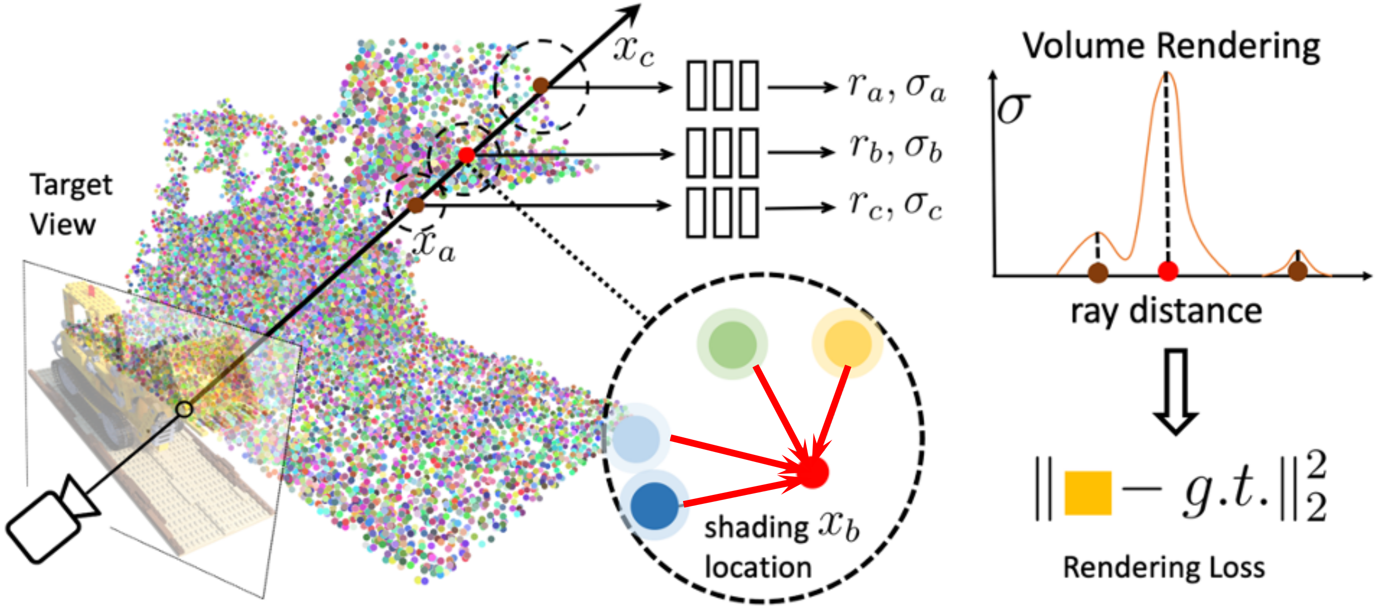


1. Ramamoorthi, Ravi, and Pat Hanrahan. "An efficient representation for irradiance environment maps." In Proceedings of the 28th annual conference on Computer graphics and interactive techniques, pp. 497-500. 2001.

Neural Point Grouping

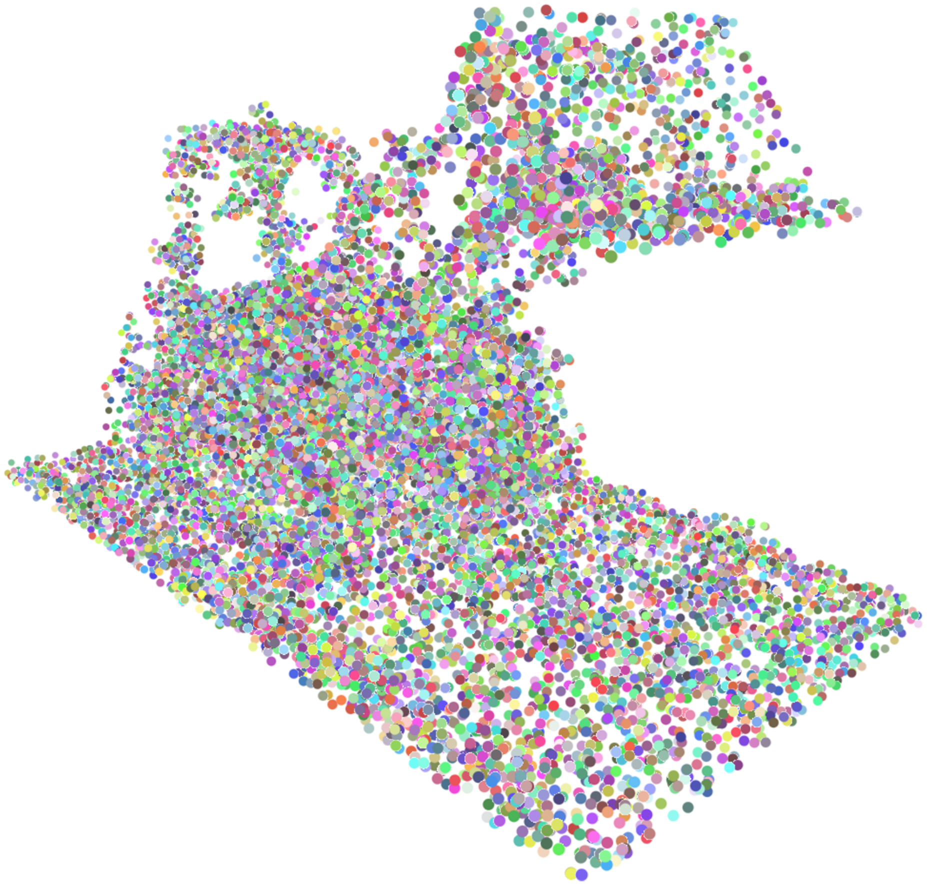


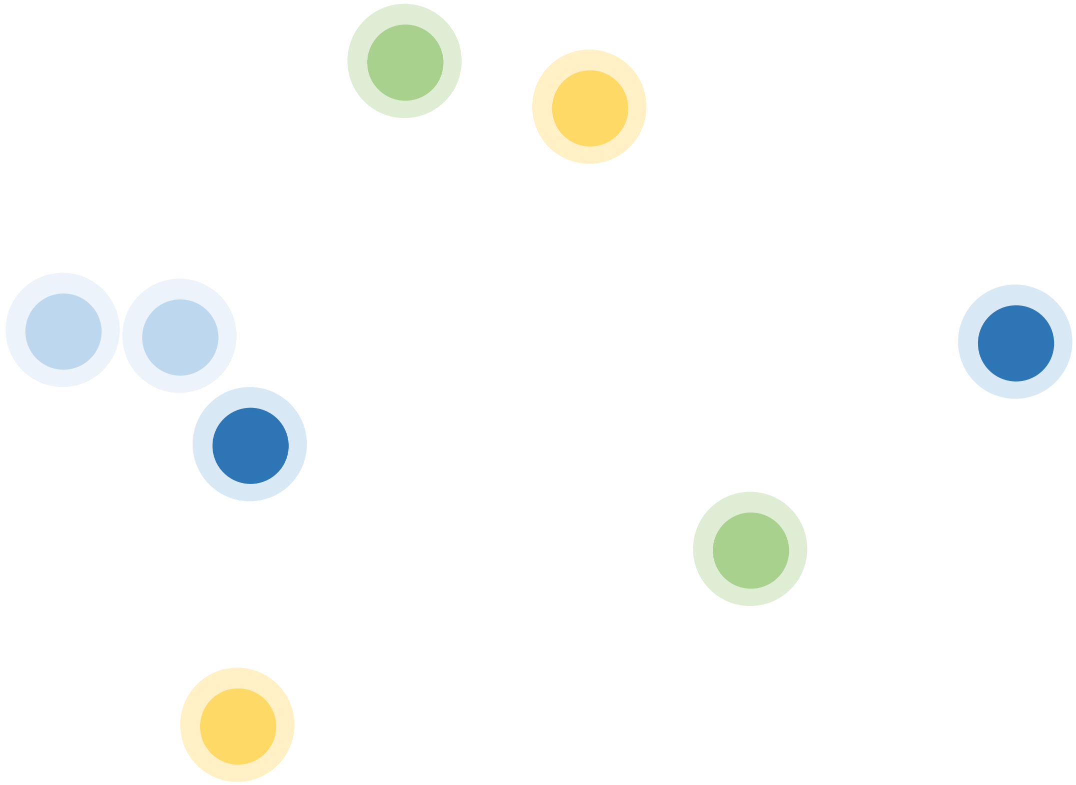
Point-based Radiance Fields



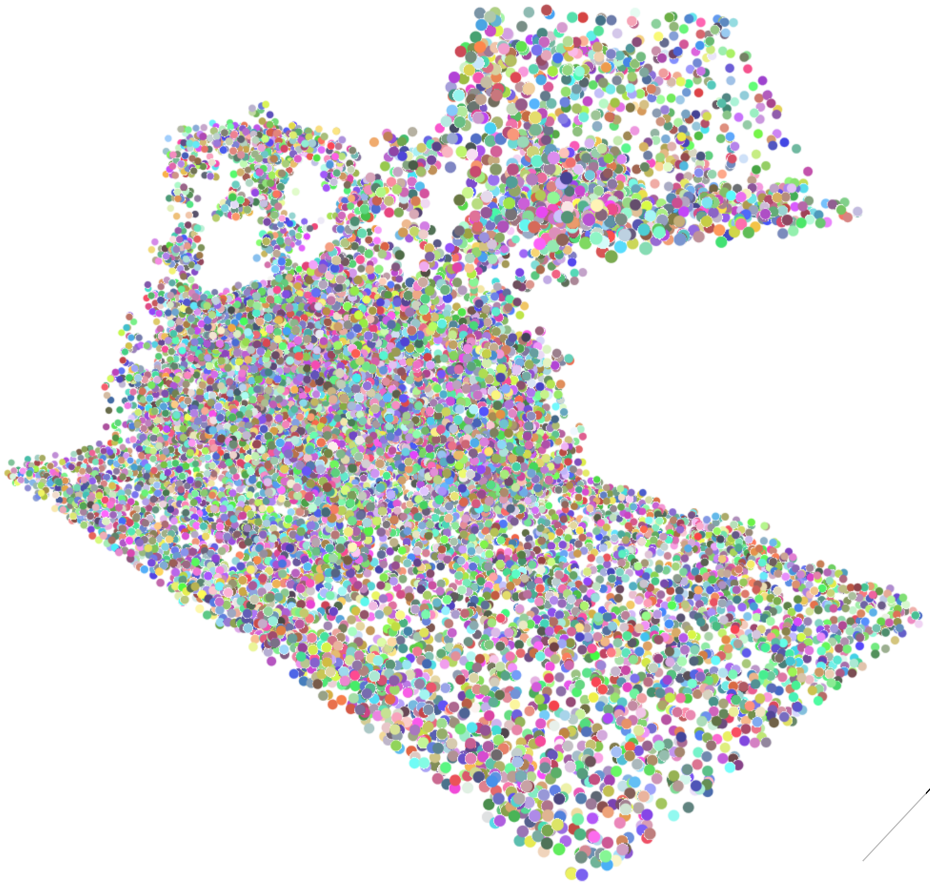
Point Feature Aggregation

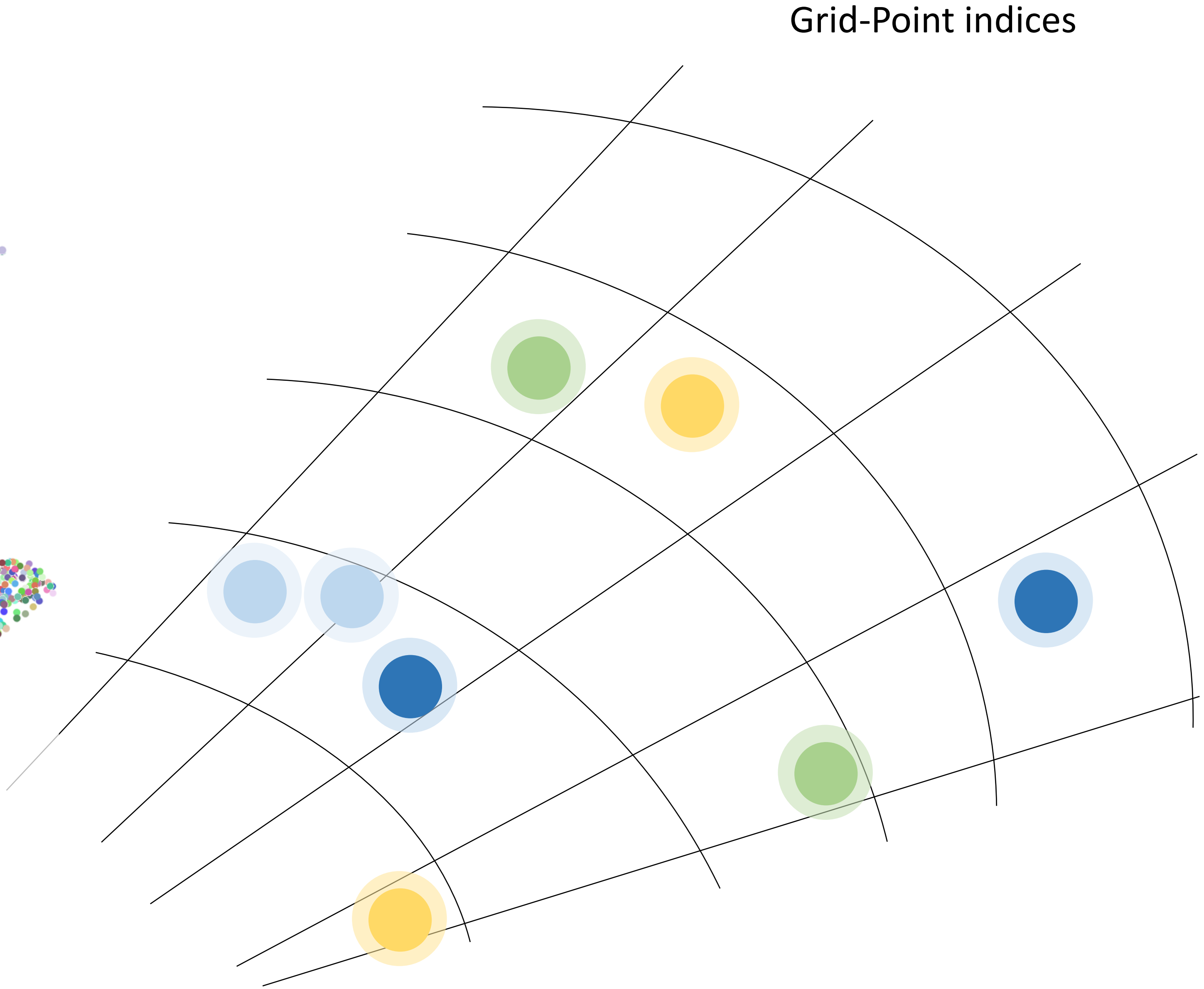
Neural Point Voxel Indices



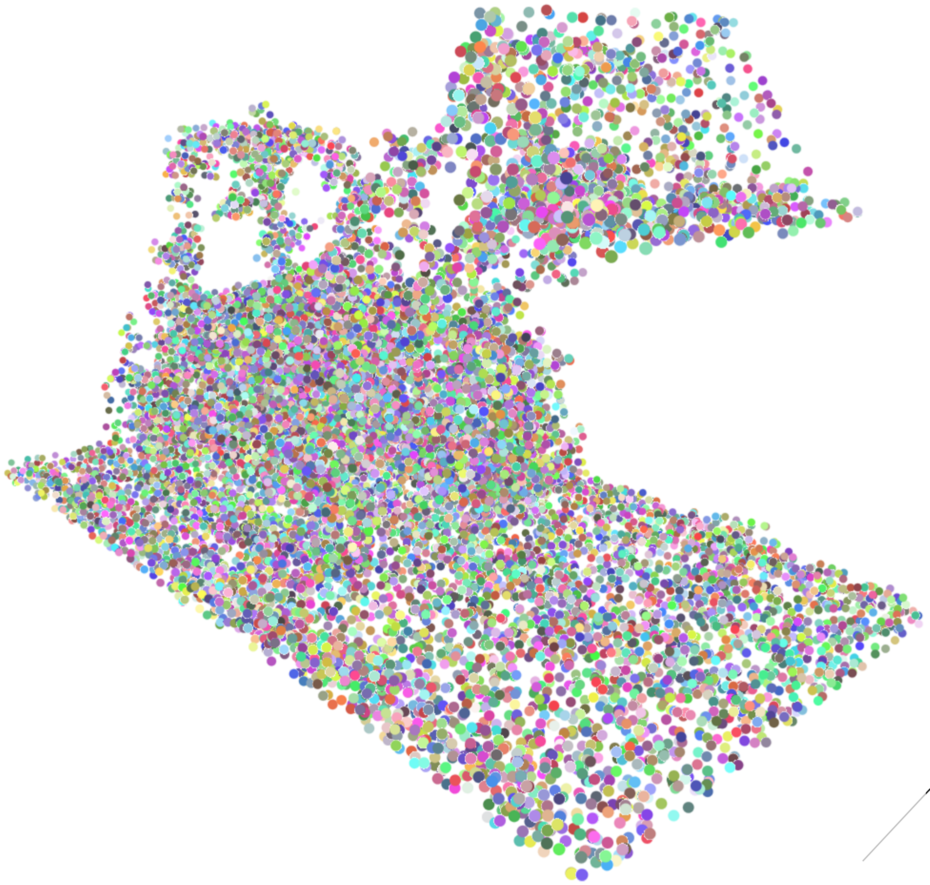


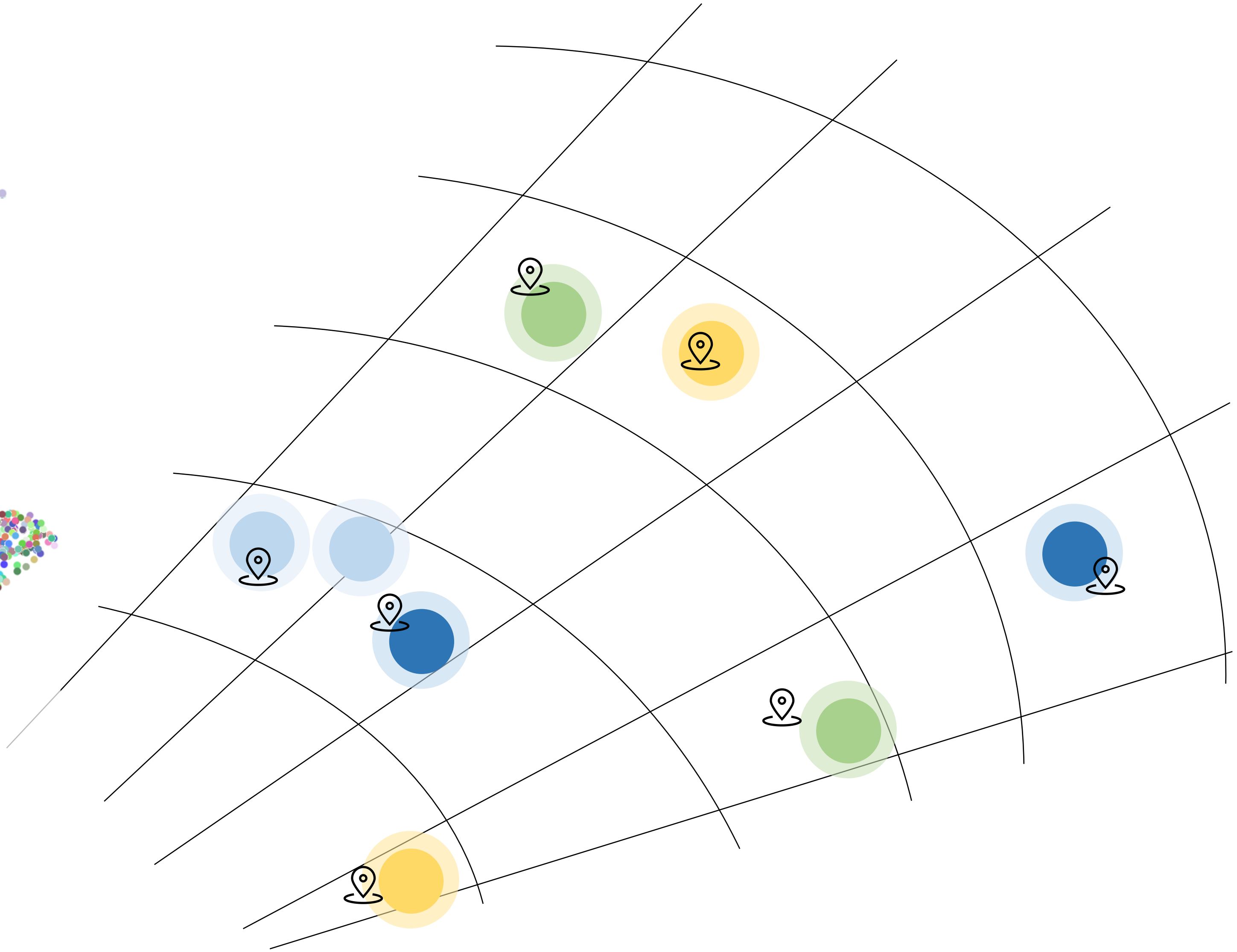
Neural Point Voxel Indices



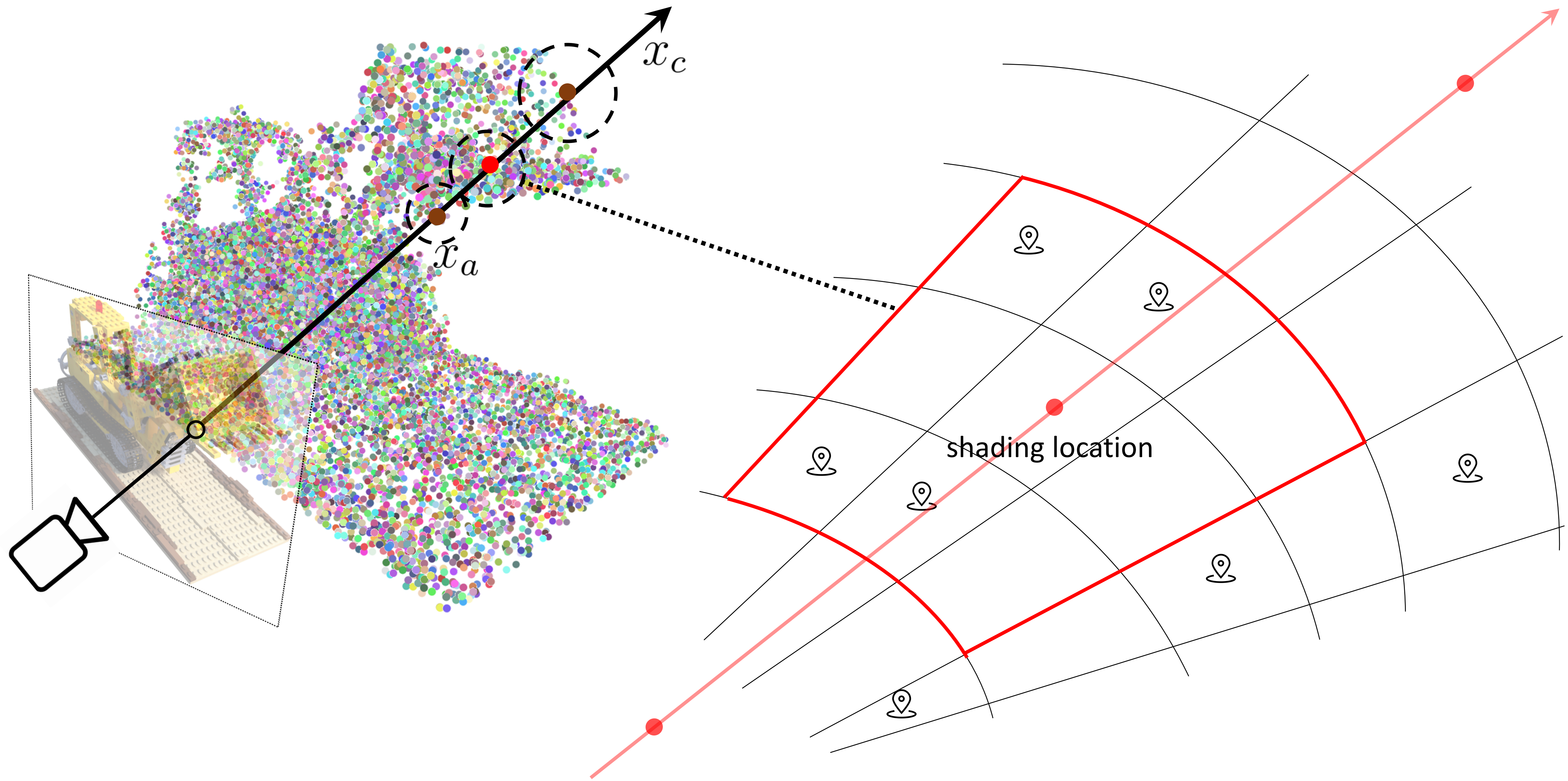


Neural Point Voxel Indices

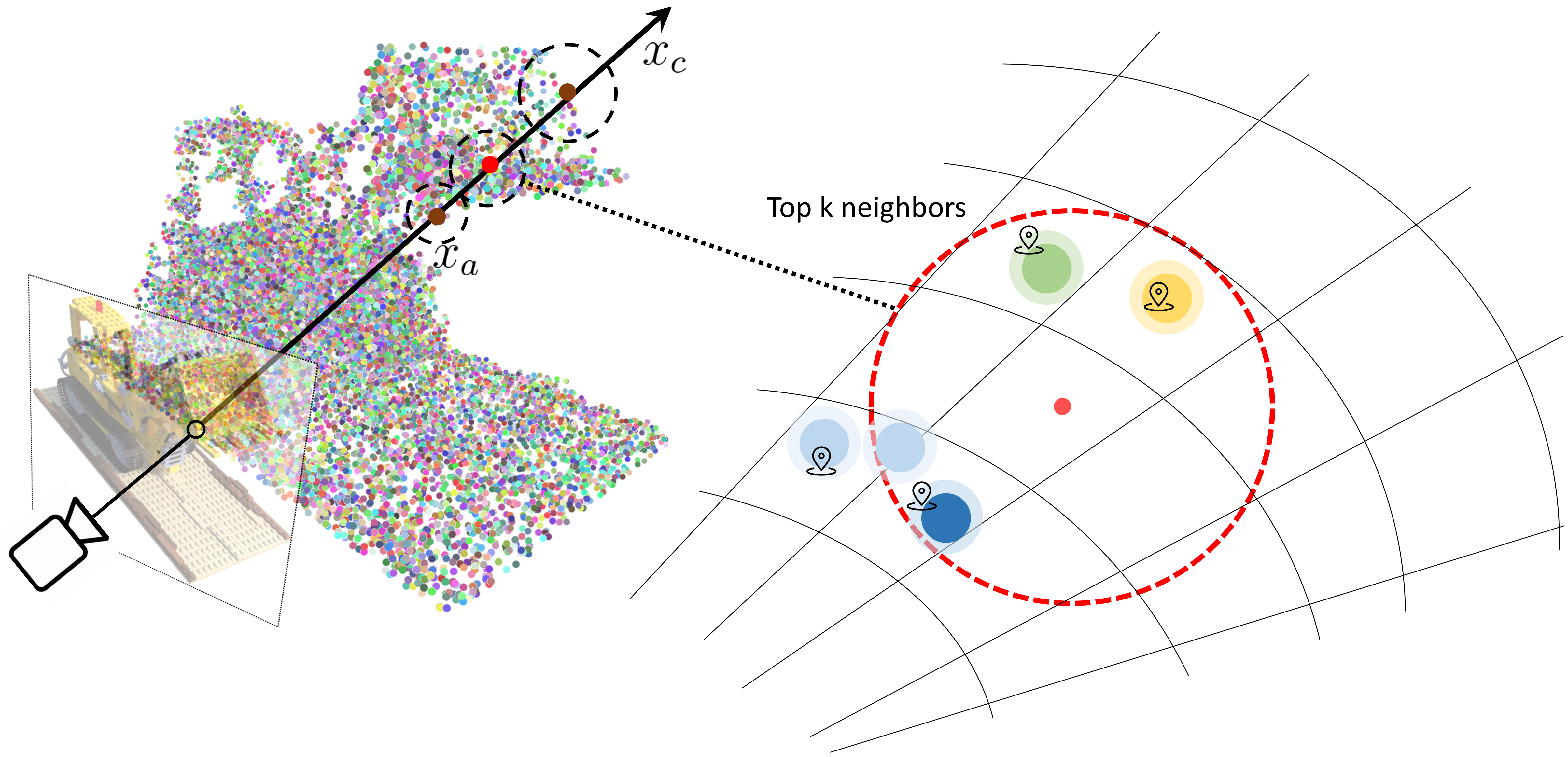




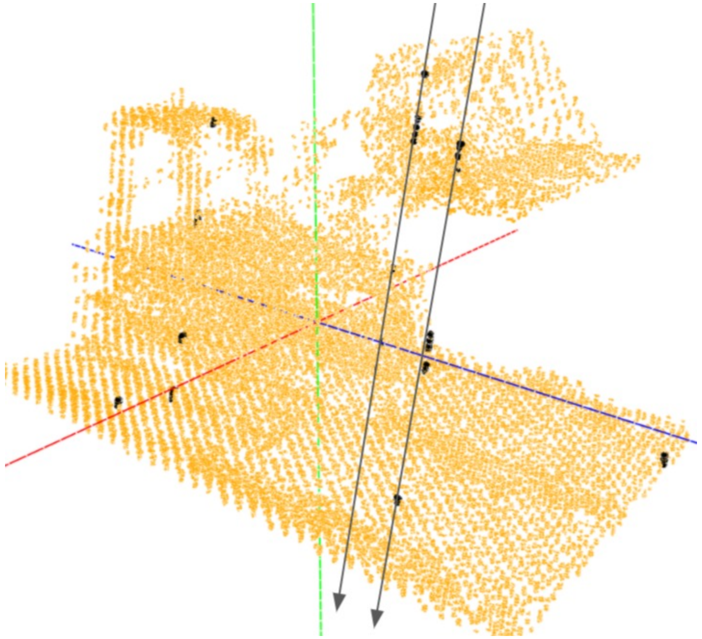
Neural Point Searching



Neural Point Searching



Query Neural Points



Visualization of valid
shading locations (black)
amount voxel centers (orange)

Lib Name	Points	Query ray	Num. shading points	K points	Elapse (ms)	Memory
kNN-CUDA	81920	32 x 32	16	128	1272.58	N/A
FRNN	5,242,875	32 x 32	10	128	44.4	8 GB
FRNN	5,242,875	800 x 800	16	128	N/A	17 GB
Open3D	5,242,875	800*800	2	128	33440	11 GB
Our CUDA	5,242,875	800*800	16	128	**8.7**	**2.1** GB