MonoSDF

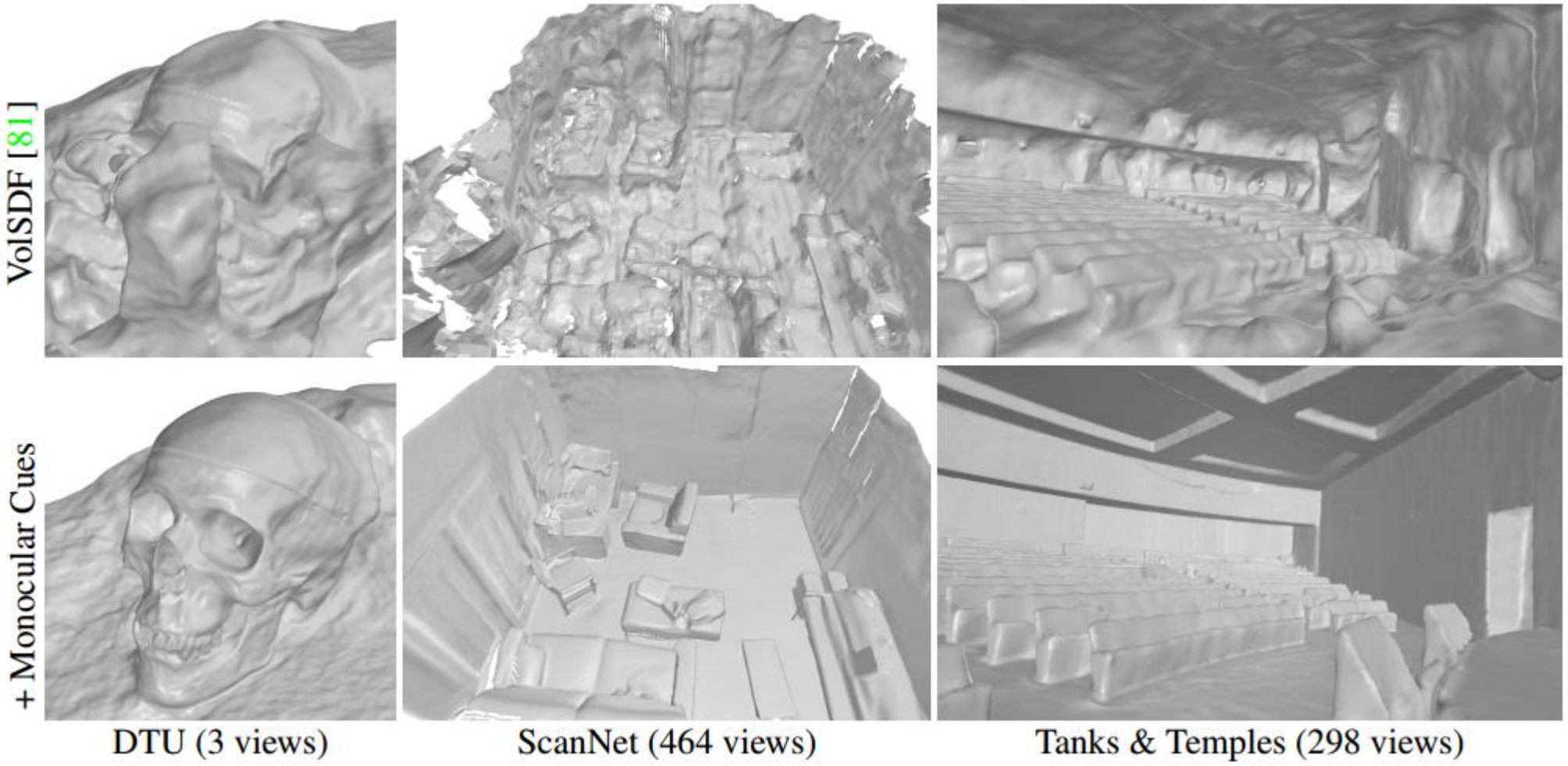
Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction

Zehao Yu, Songyou Peng, Michael Niemeyer, Torsten Sattler, Andreas Geiger

Presented by: Zeeshan Ahmad
Seminar Inverse Rendering

Date: July 4, 2023

3D Reconstruction is an ill-posed Problem



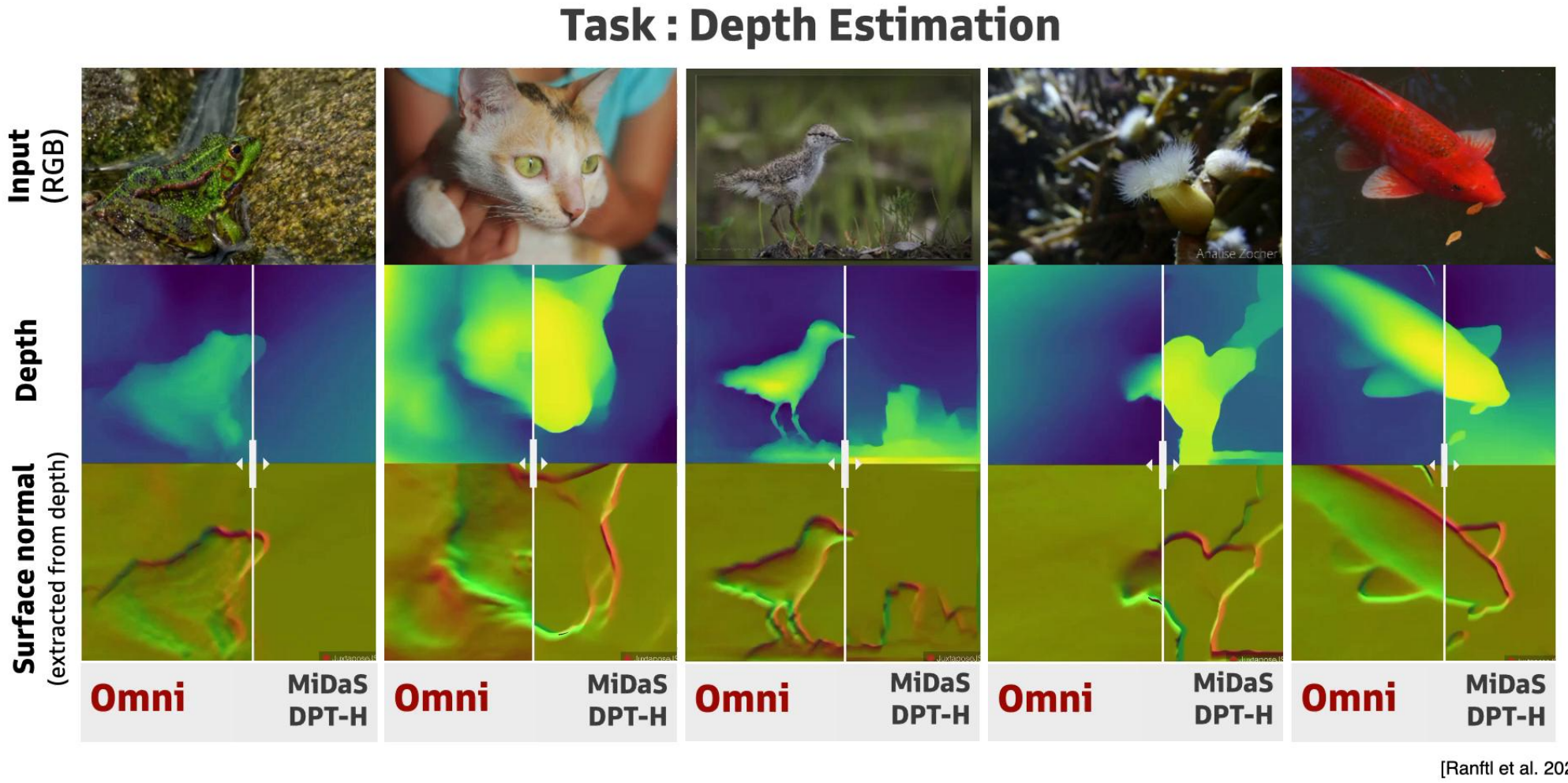
Yu, Peng, Niemeyer, Sattler, Geiger: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. NeuIPS, 2022.

Which monocular geometric cues did they use to overcome reconstruction ambiguities?

Which monocular geometric cues did they use to overcome reconstruction ambiguities?

Used Depth & Normal maps.

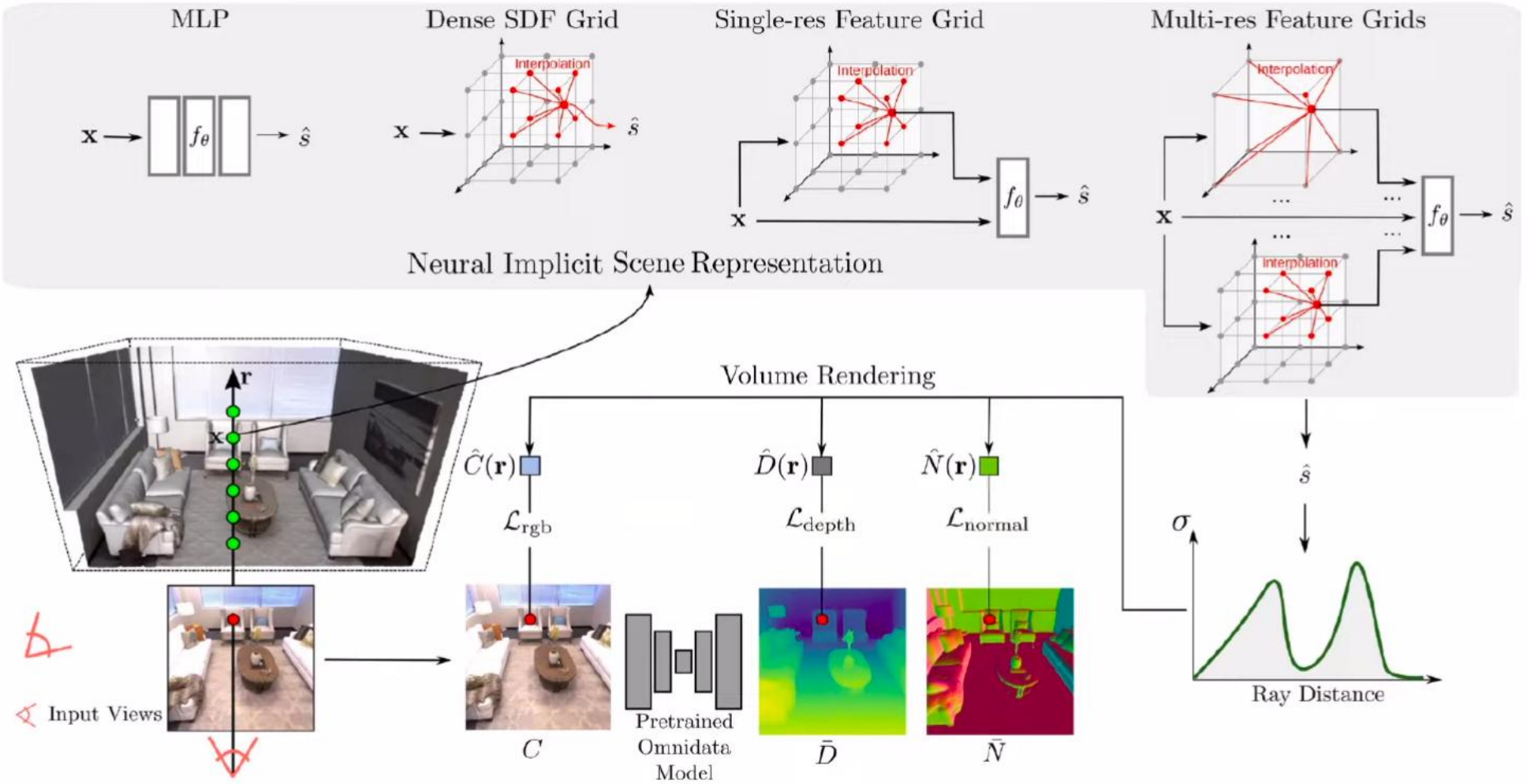
OmniData: Vision Data from 3D Scans



[Ranftl et al. 2021]

Eftekhar, Sax, Bachmann, Malik, Zamir: Omnidata: A Scalable Pipeline for Making Multi-Task Mid-Level Vision Datasets from 3D Scans. ICCV, 2021.

MonoSDF Pipeline

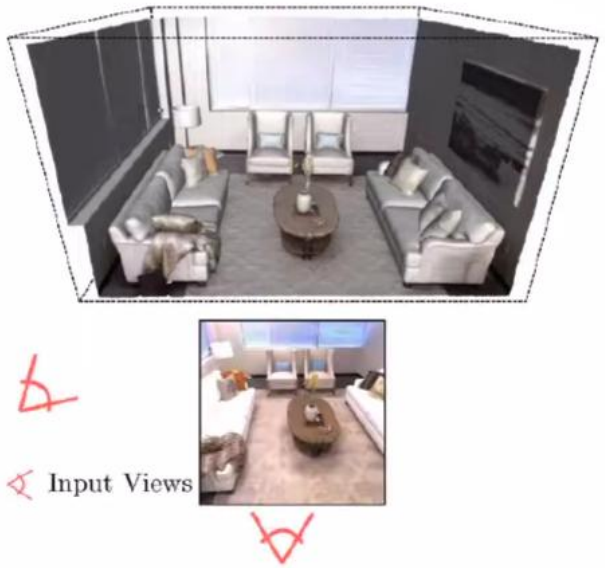


Yu, Peng, Niemeyer, Sattler, Geiger: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. NeuIPS, 2022.

MonoSDF Pipeline

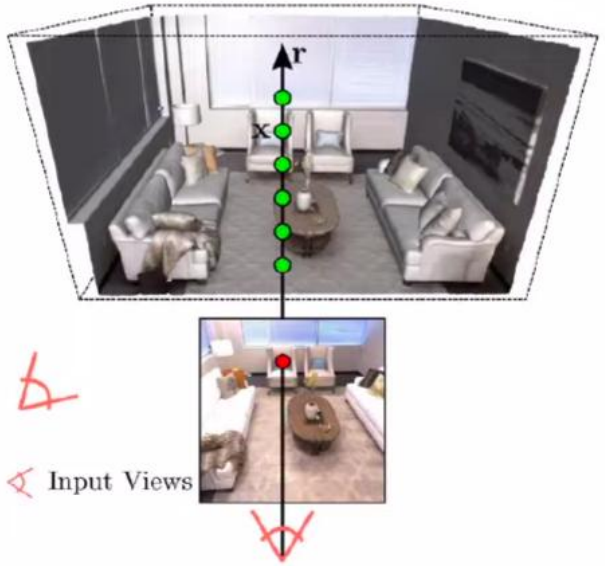
Yu, Peng, Niemeyer, Sattler, Geiger: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. NeuIPS, 2022.

MonoSDF Pipeline



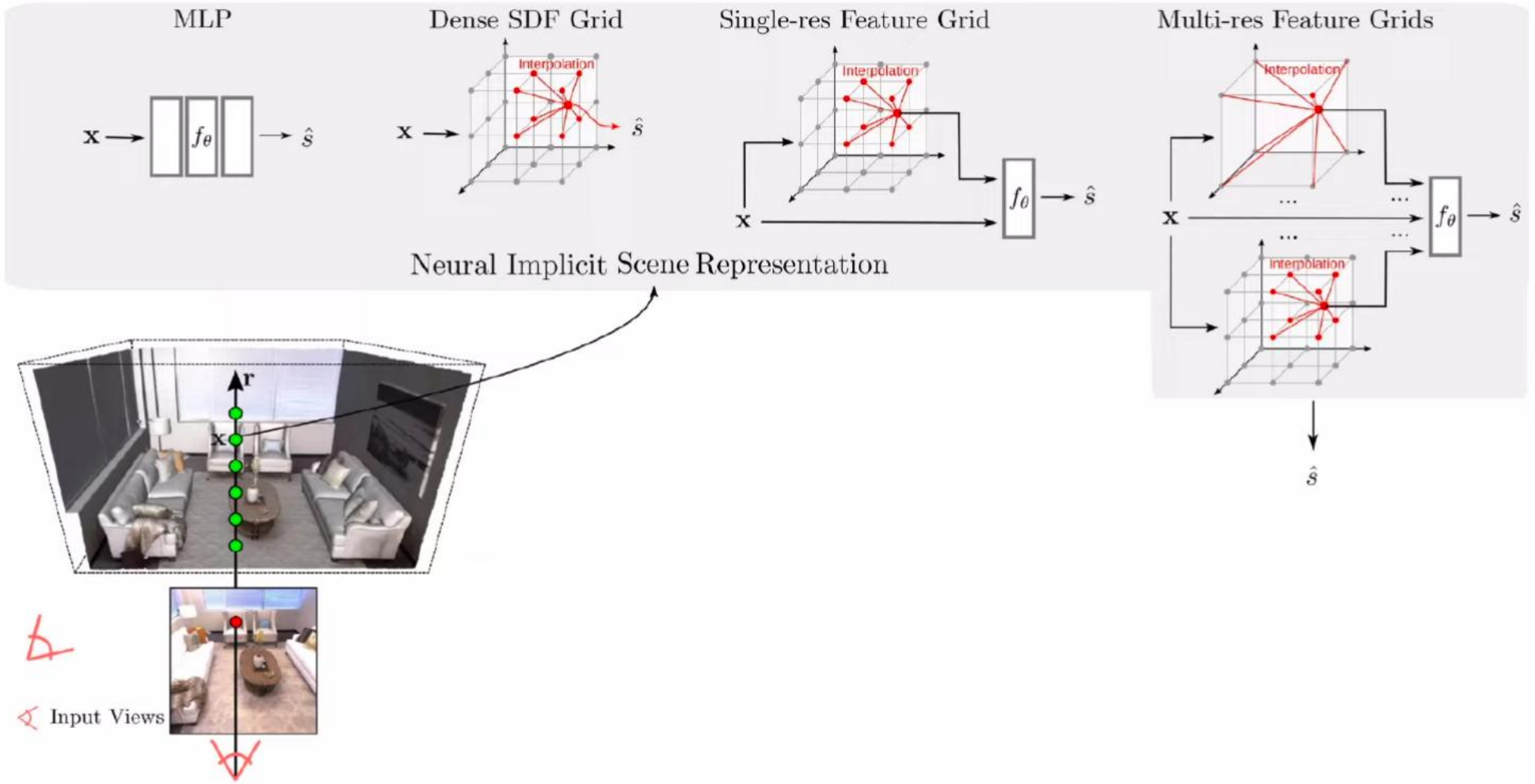
Yu, Peng, Niemeyer, Sattler, Geiger: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. NeuIPS, 2022.

MonoSDF Pipeline



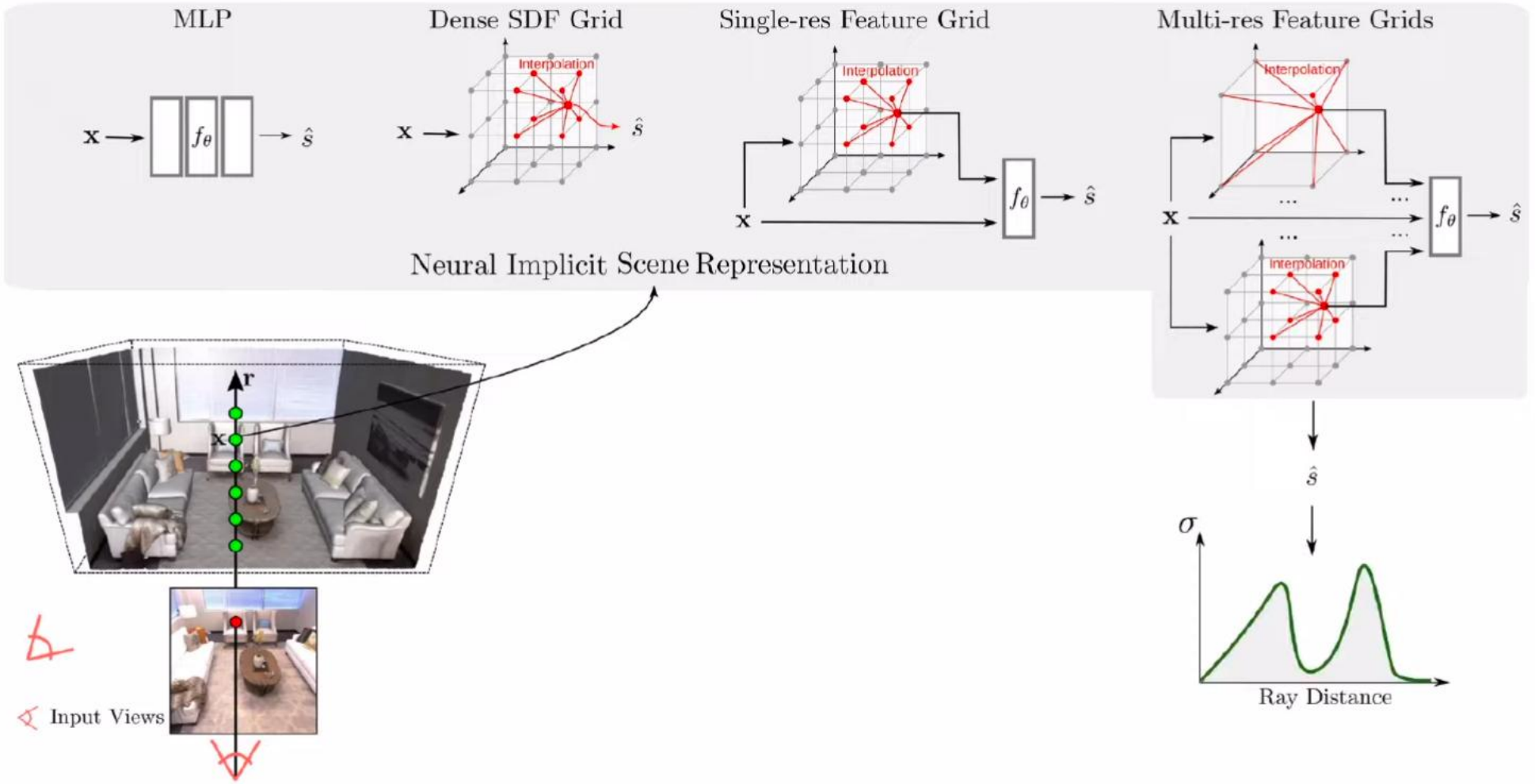
Yu, Peng, Niemeyer, Sattler, Geiger: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. NeuIPS, 2022.

MonoSDF Pipeline



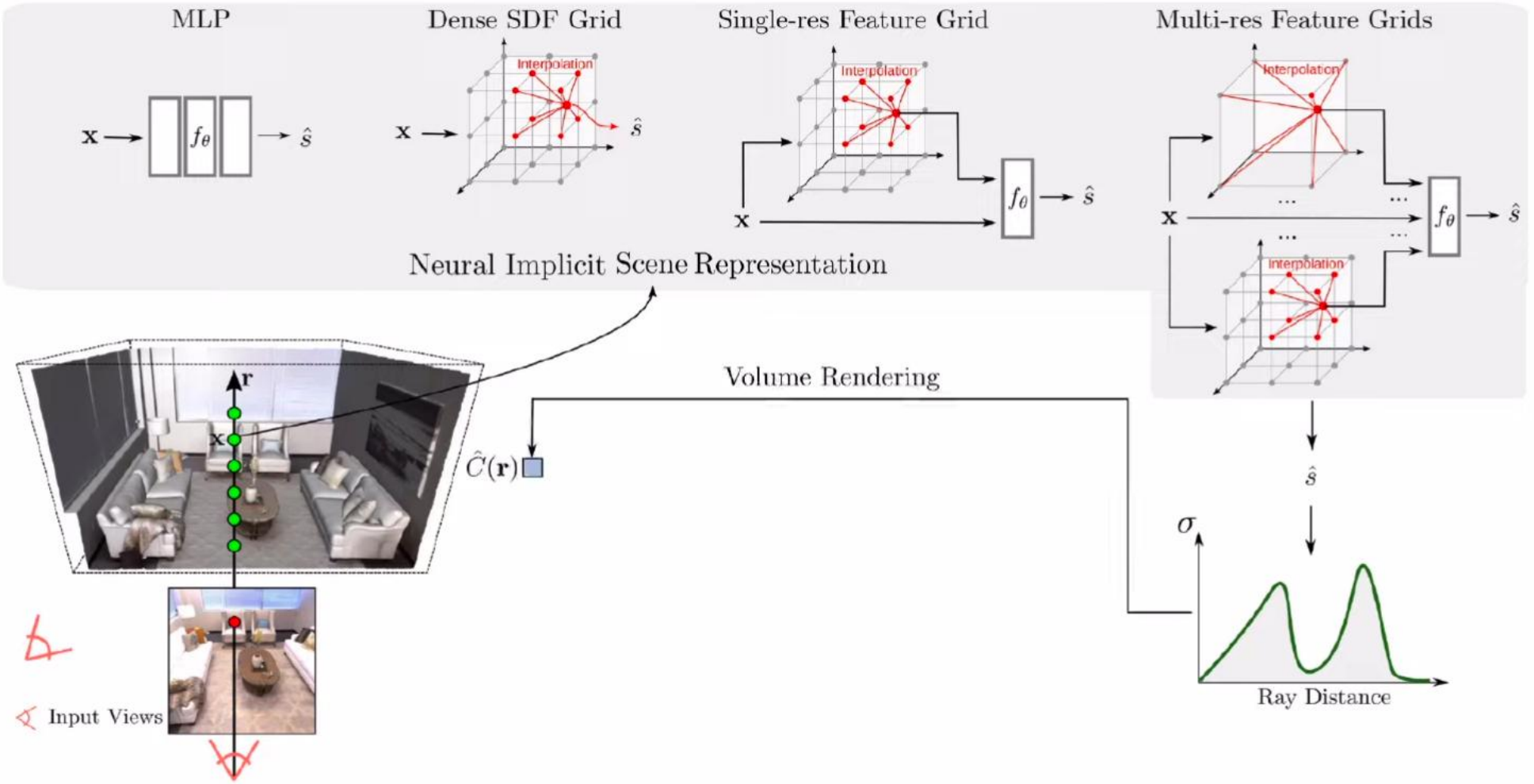
Yu, Peng, Niemeyer, Sattler, Geiger: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. NeuIPS, 2022.

MonoSDF Pipeline



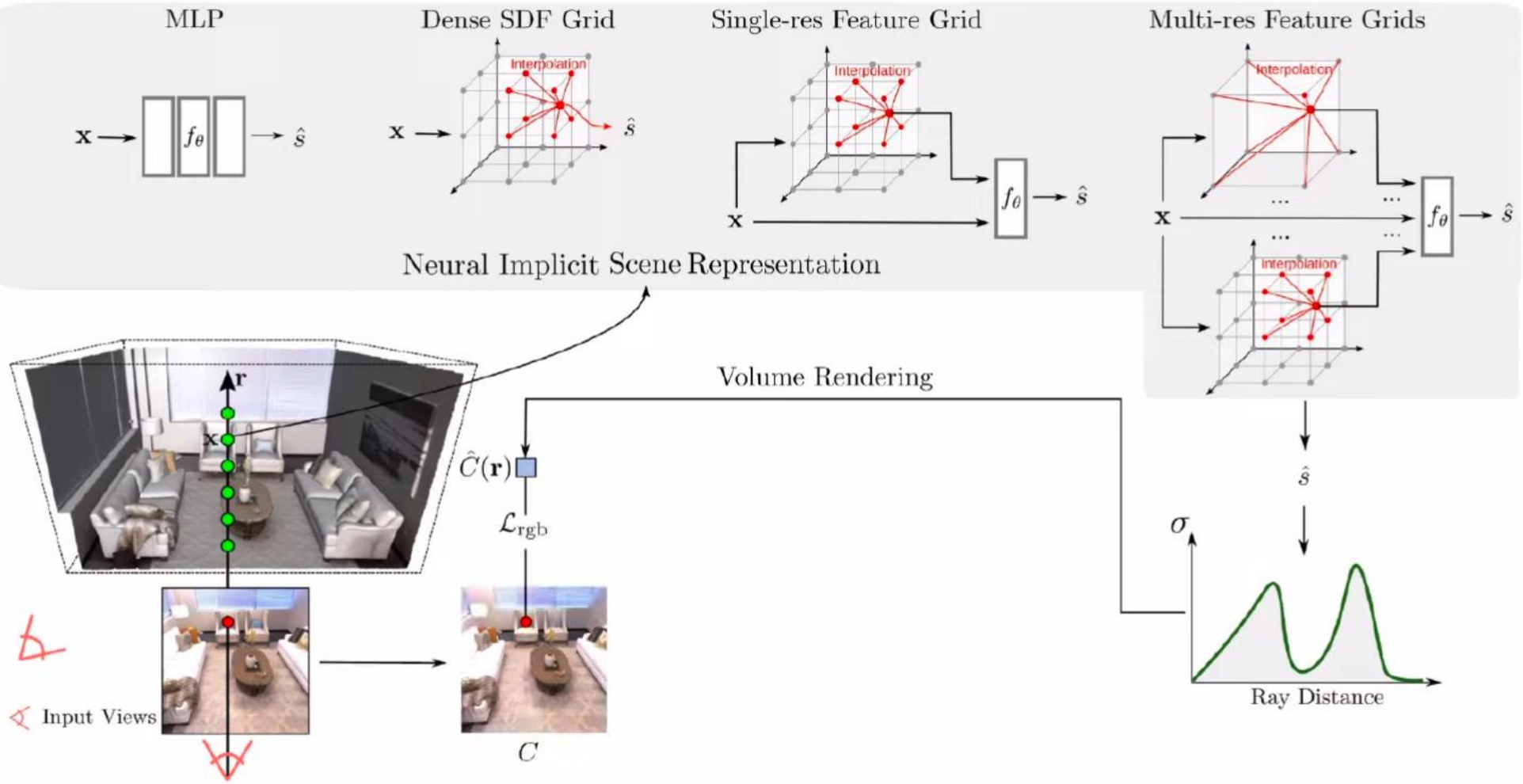
Yu, Peng, Niemeyer, Sattler, Geiger: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. NeuIPS, 2022.

MonoSDF Pipeline



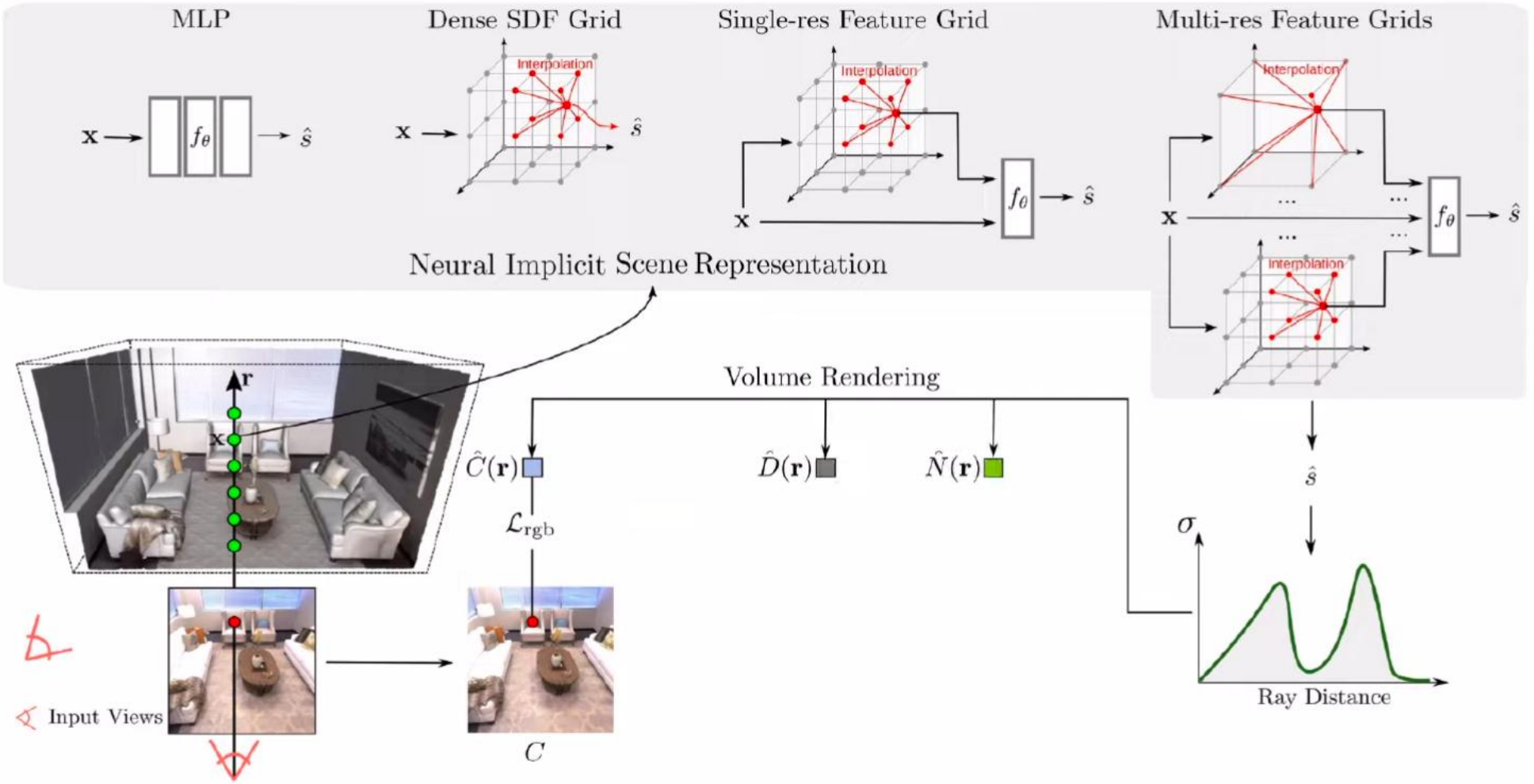
Yu, Peng, Niemeyer, Sattler, Geiger: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. NeuIPS, 2022.

MonoSDF Pipeline



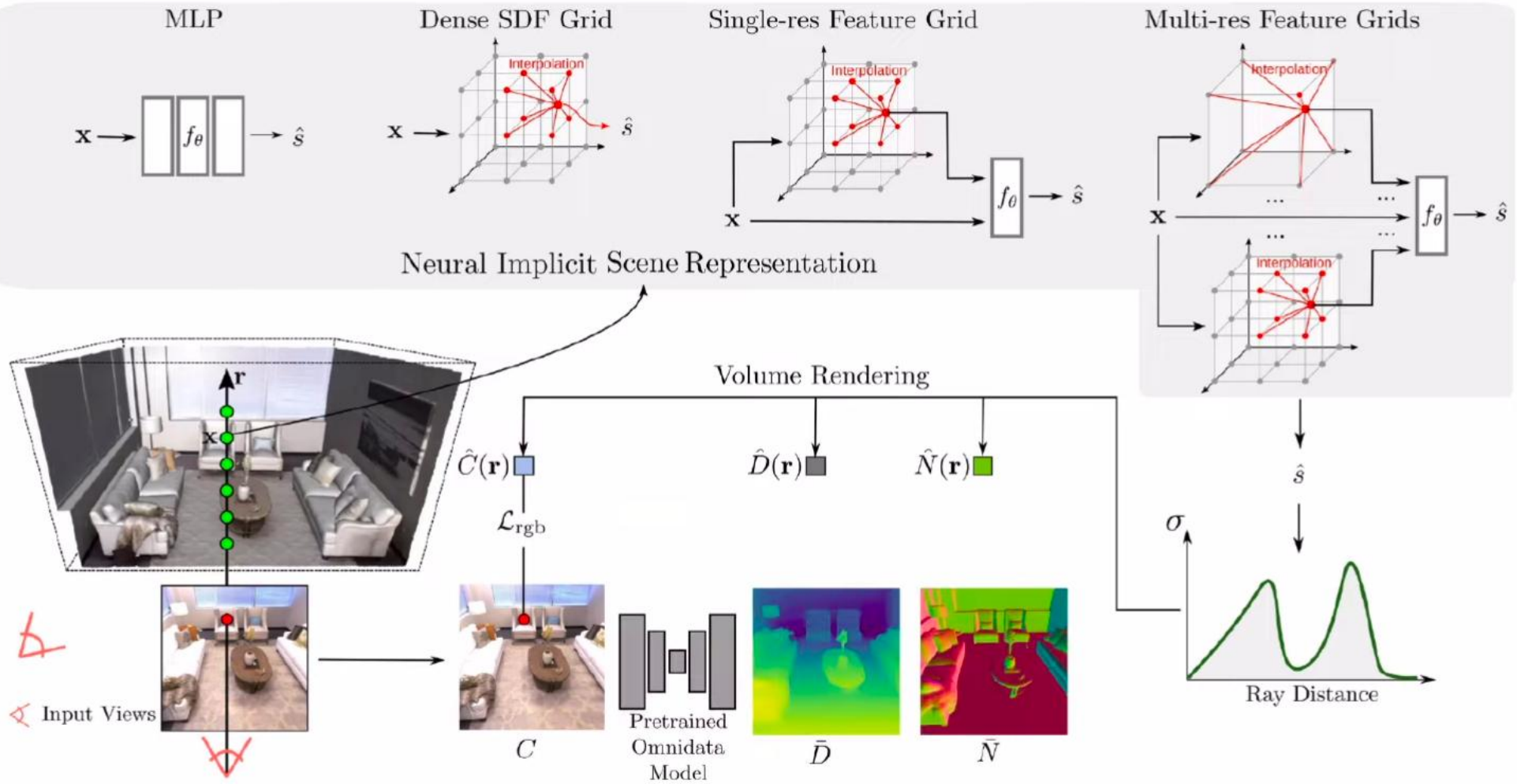
Yu, Peng, Niemeyer, Sattler, Geiger: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. NeuIPS, 2022.

MonoSDF Pipeline



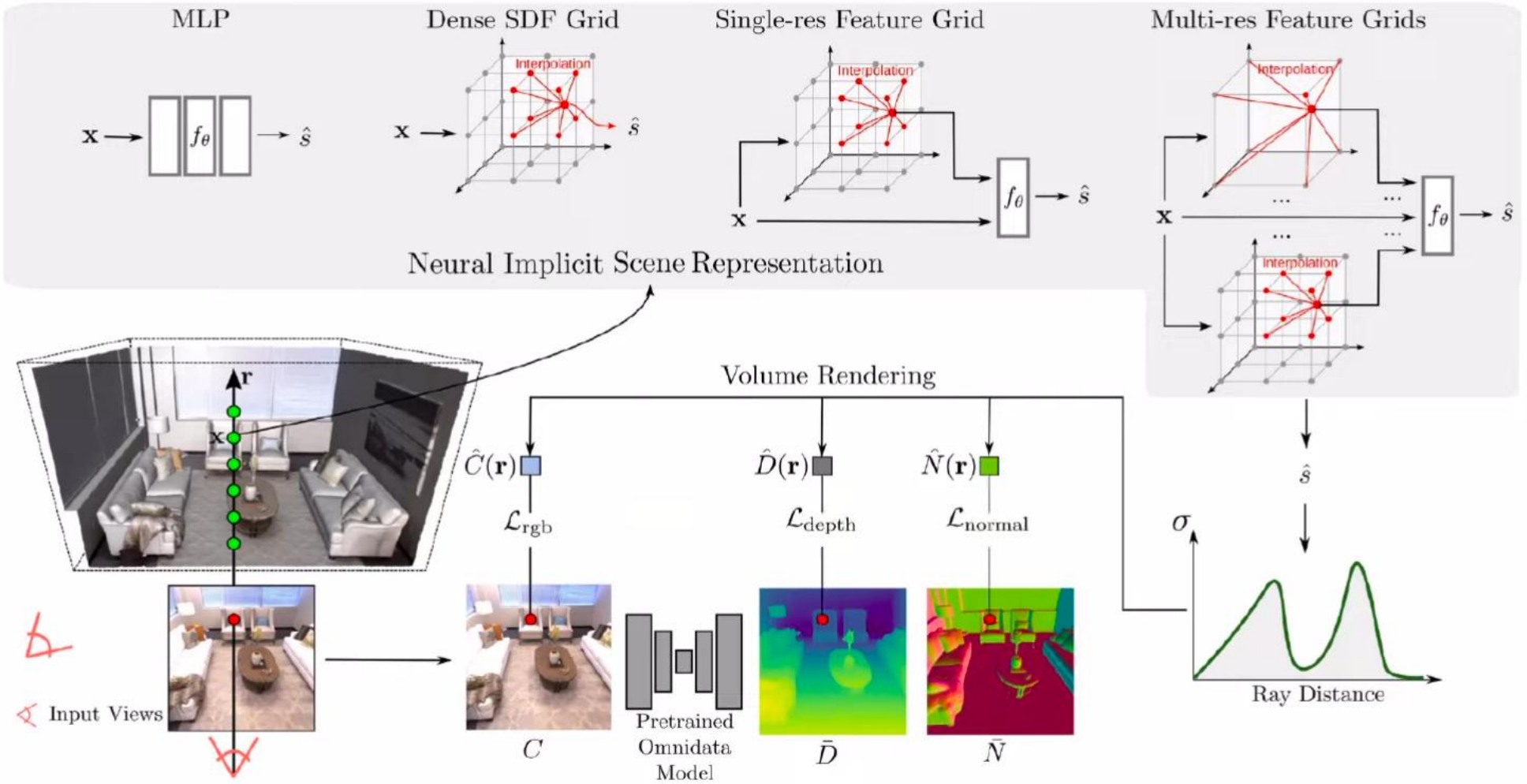
Yu, Peng, Niemeyer, Sattler, Geiger: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. NeuIPS, 2022.

MonoSDF Pipeline



Yu, Peng, Niemeyer, Sattler, Geiger: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. NeuIPS, 2022.

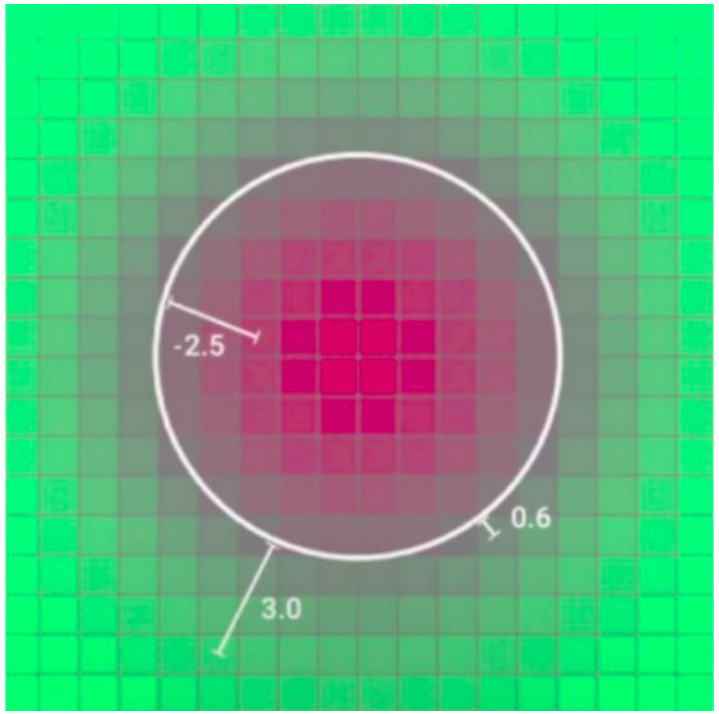
MonoSDF Pipeline



Yu, Peng, Niemeyer, Sattler, Geiger: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. NeuIPS, 2022.

Signed Distance Field

Signed distance field ‘**f**’ stores the **distance to the closest surface** at each point.

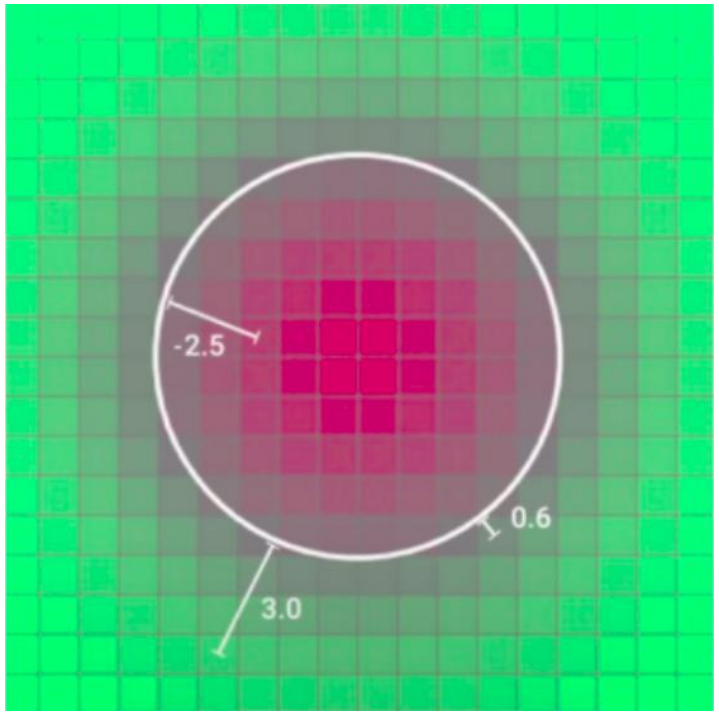


Signed Distance Field

Signed distance field ‘**f**’ stores the **distance to the closest surface** at each point.

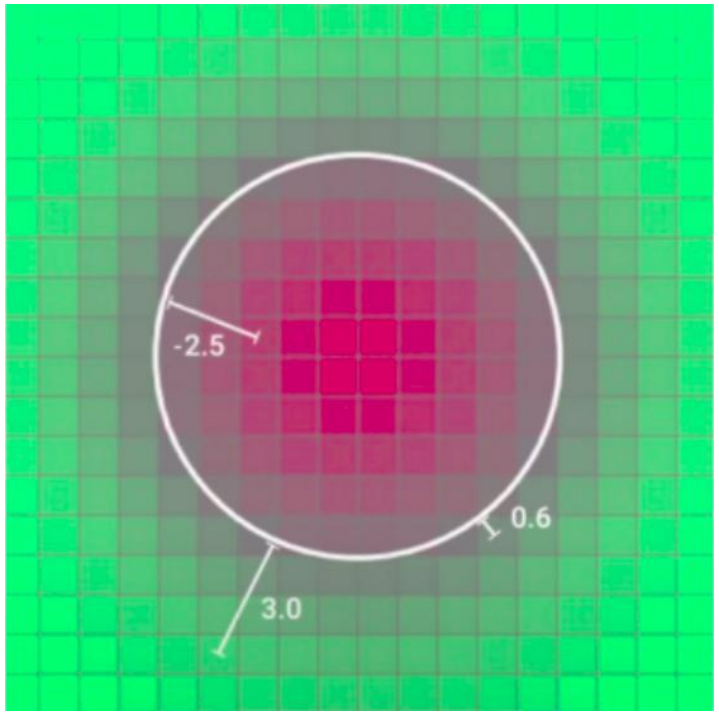
Properties:

- Surface represented by **zero-level set**.



Signed Distance Field

Signed distance field '**f**' stores the **distance to the closest surface** at each point.



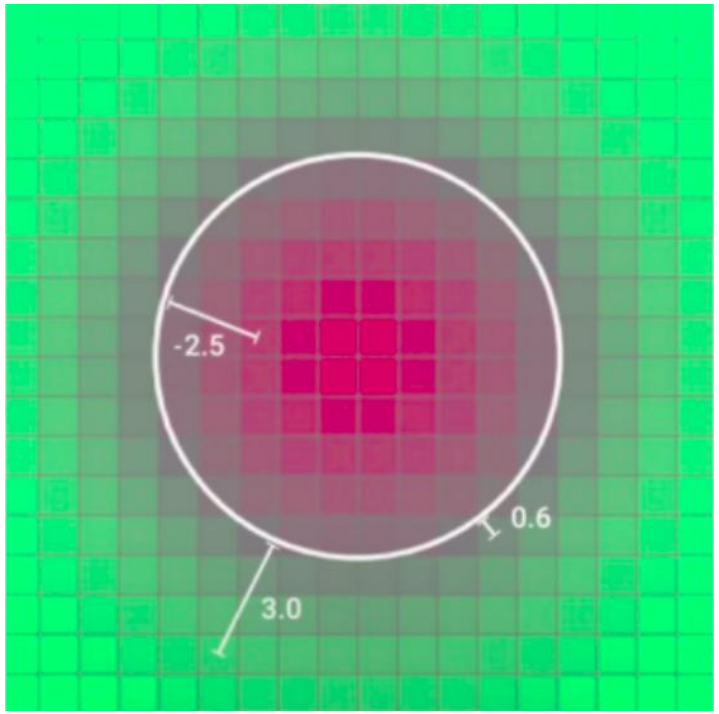
Properties:

- Surface represented by **zero-level set**.
- For any point **x** on surface, **normal** is defined as:

$$\nabla f(x) = N(x)$$

Signed Distance Field

Signed distance field '**f**' stores the **distance to the closest surface** at each point.



Properties:

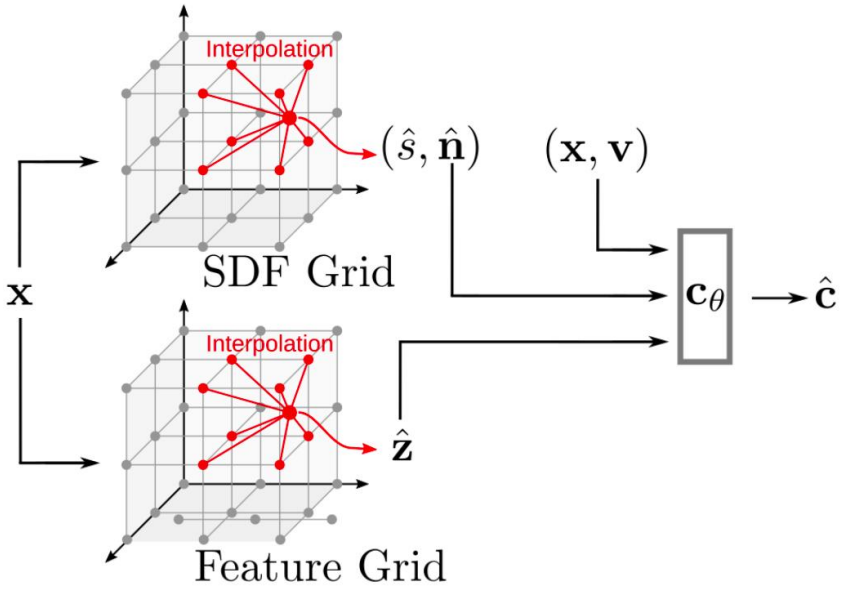
- Surface represented by **zero-level set**.
- For any point **x** on surface, **normal** is defined as:

$$\nabla f(x) = N(x)$$

- Gradient satisfies **eikonal** equation. *i.e;* $|\nabla f| = 1$.

Dense SDF Grid

– Represents geometry using **SDF volume** $\mathcal{G}_\theta$.



Yu, Peng, Niemeyer, Sattler, Geiger: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. NeuIPS, 2022.

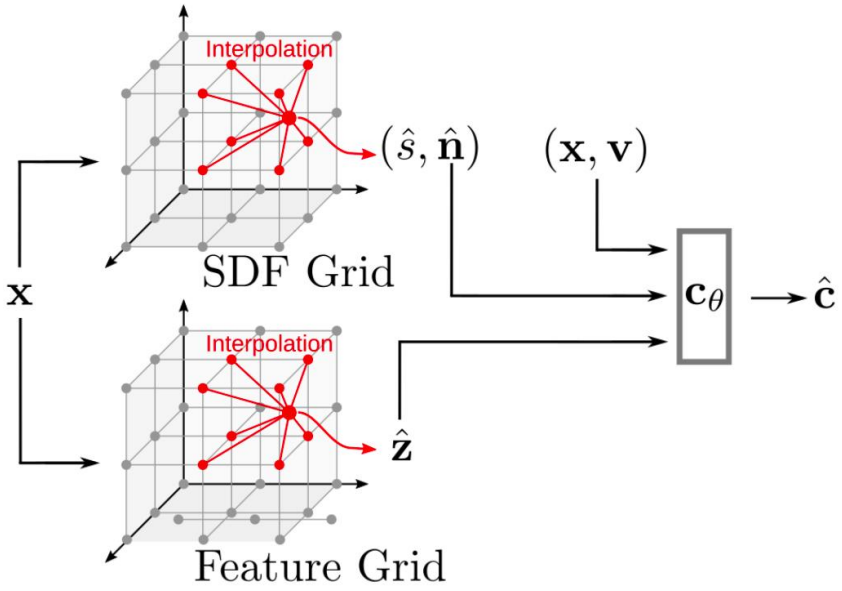
Dense SDF Grid

- Represents geometry using **SDF volume** $\mathcal{G}_\theta$.
- For a point $\mathbf{x}$ **predicted** SDF value is $\hat{s}$.

$$\hat{s} = \texttt{interp}(\mathbf{x}, \mathcal{G}_\theta)$$

where

- $\mathcal{G}_\theta$ is discretized volume storing SDF values in each cell.
- **interp** is trilinear interpolation.

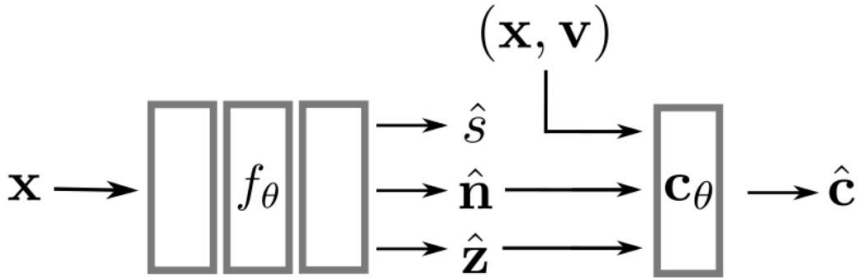


Yu, Peng, Niemeyer, Sattler, Geiger: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. NeuIPS, 2022.

Single MLP

– Parameterizes the SDF function '**f**' using MLP.

$$\hat{s} = f_\theta(\gamma(\mathbf{x}))$$



where

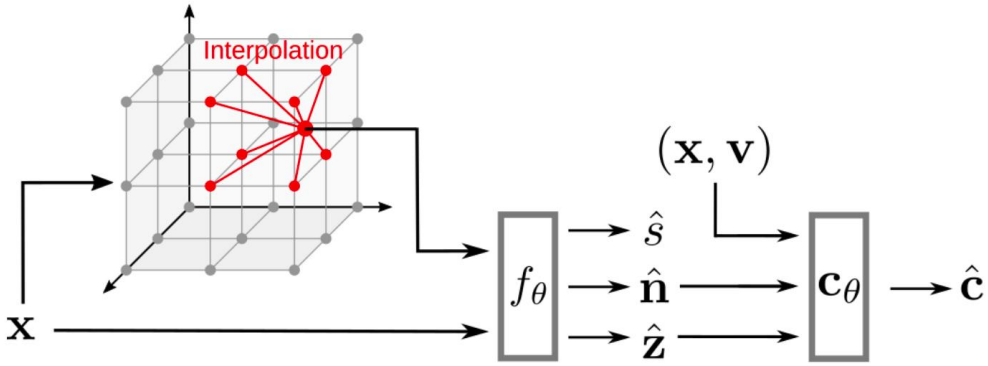
– θ is **parameter** set of function '**f**'.
– γ indicates **positional** encoding.

Yu, Peng, Niemeyer, Sattler, Geiger: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. NeuIPS, 2022.

Single Resolution Feature Grid with MLP Decoder

- Combines feature-grid and MLP.

$$\hat{s} = f_\theta(\gamma(\mathbf{x}), \texttt{interp}(\mathbf{x}, \Phi_\theta))$$



where

- f_θ is feature-conditioned MLP.
- Φ_θ is parametrized feature-grid with resolution R^3.

Yu, Peng, Niemeyer, Sattler, Geiger: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. NeuIPS, 2022.

Multi-Resolution Feature Grid with MLP Decoder

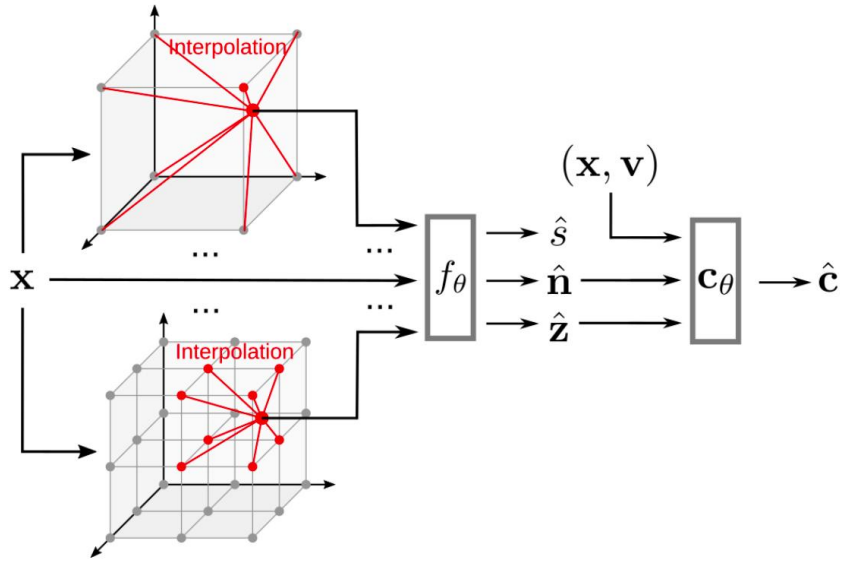
– Uses L **multi-resolution** feature-grids $\{\Phi_\theta^l\}_{l=1}^L$.

$$\hat{s} = f_\theta(\gamma(\mathbf{x}), \{\texttt{interp}(\mathbf{x}, \Phi_\theta^l)\}_l))$$

where

– $\gamma(\mathbf{x})$ is positional encoding of $\mathbf{x}$.

– Φ_θ^l is feature-grid at level l parameterized by θ.



Yu, Peng, Niemeyer, Sattler, Geiger: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. NeuIPS, 2022.

Multi-Resolution Feature Grid with MLP Decoder

- Resolution are sampled in **geometric** space.

$$R_l := \lfloor R_{\min} b^l \rfloor \qquad b := \exp\left(\frac{\ln R_{\max} - \ln R_{\min}}{L-1}\right)$$

where

- $R_{\min}, R_{\max}$ are **coarsest** and **finest** resolutions.
- b is **growth factor**.

Yu, Peng, Niemeyer, Sattler, Geiger: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. NeuIPS, 2022.

Multi-Resolution Feature Grid with MLP Decoder

- Resolution are sampled in **geometric** space.

$$R_l := \lfloor R_{\min} b^l \rfloor \qquad b := \exp\left(\frac{\ln R_{\max} - \ln R_{\min}}{L-1}\right)$$

where
  - $R_{\min}, R_{\max}$ are **coarsest** and **finest** resolutions.
  - b is **growth factor**.

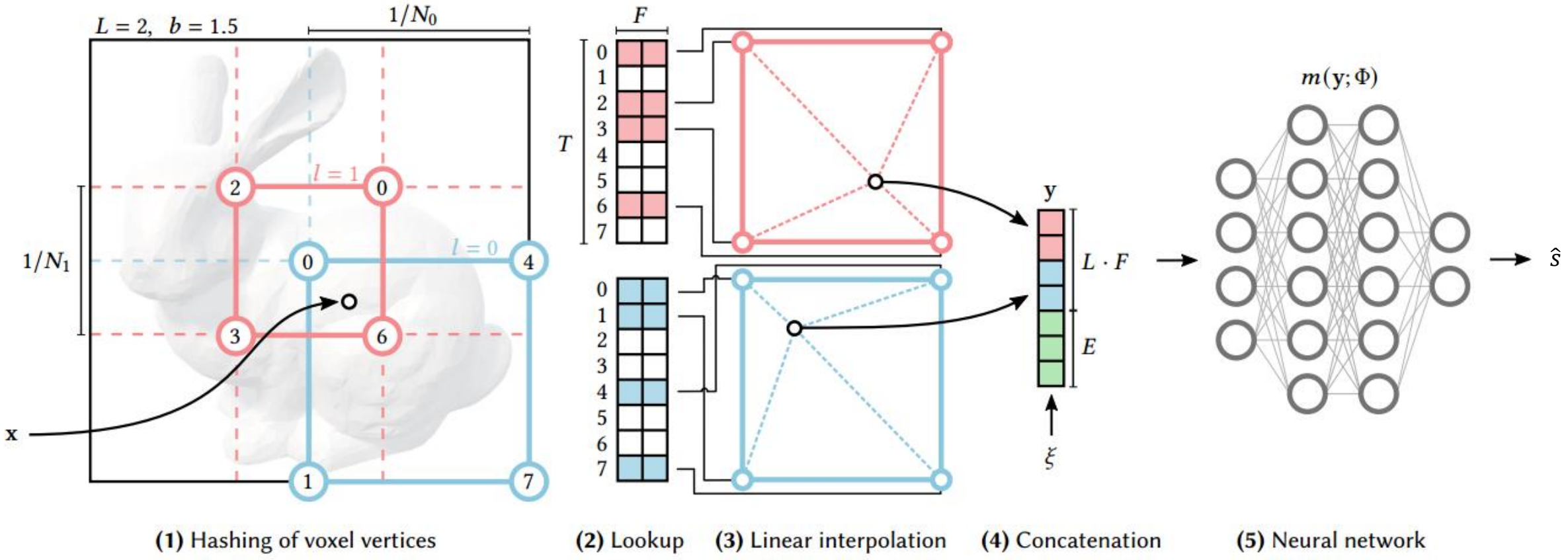
- Spatial **hash** function is used to index the feature vector.

$$h(\mathbf{x}) = \left(\bigoplus_{i=1}^{3} \mathbf{x}_i \pi_i\right) \bmod T$$

where
  - $\bigoplus$ is bit-wise **XOR**.
  - T is size of hash table, and π is large **prime number**.

Yu, Peng, Niemeyer, Sattler, Geiger: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. NeuIPS, 2022.

Multi-Resolution Feature Grid with MLP Decoder

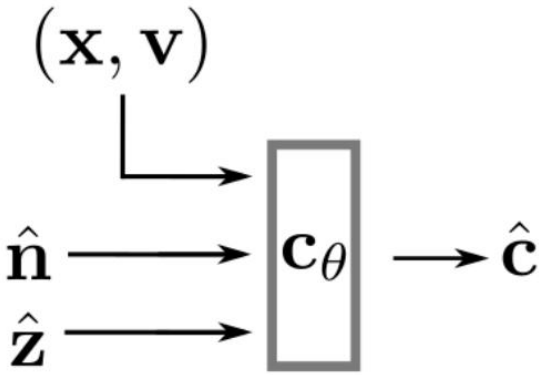


MÜLLER, EVANS, SCHIED, KELLER: Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. SIGGRAPH, 2022.

Color Prediction

- To optimize, $\mathbf{c}_\theta$ predicts the **RGB** color value $\hat{\mathbf{c}}$.

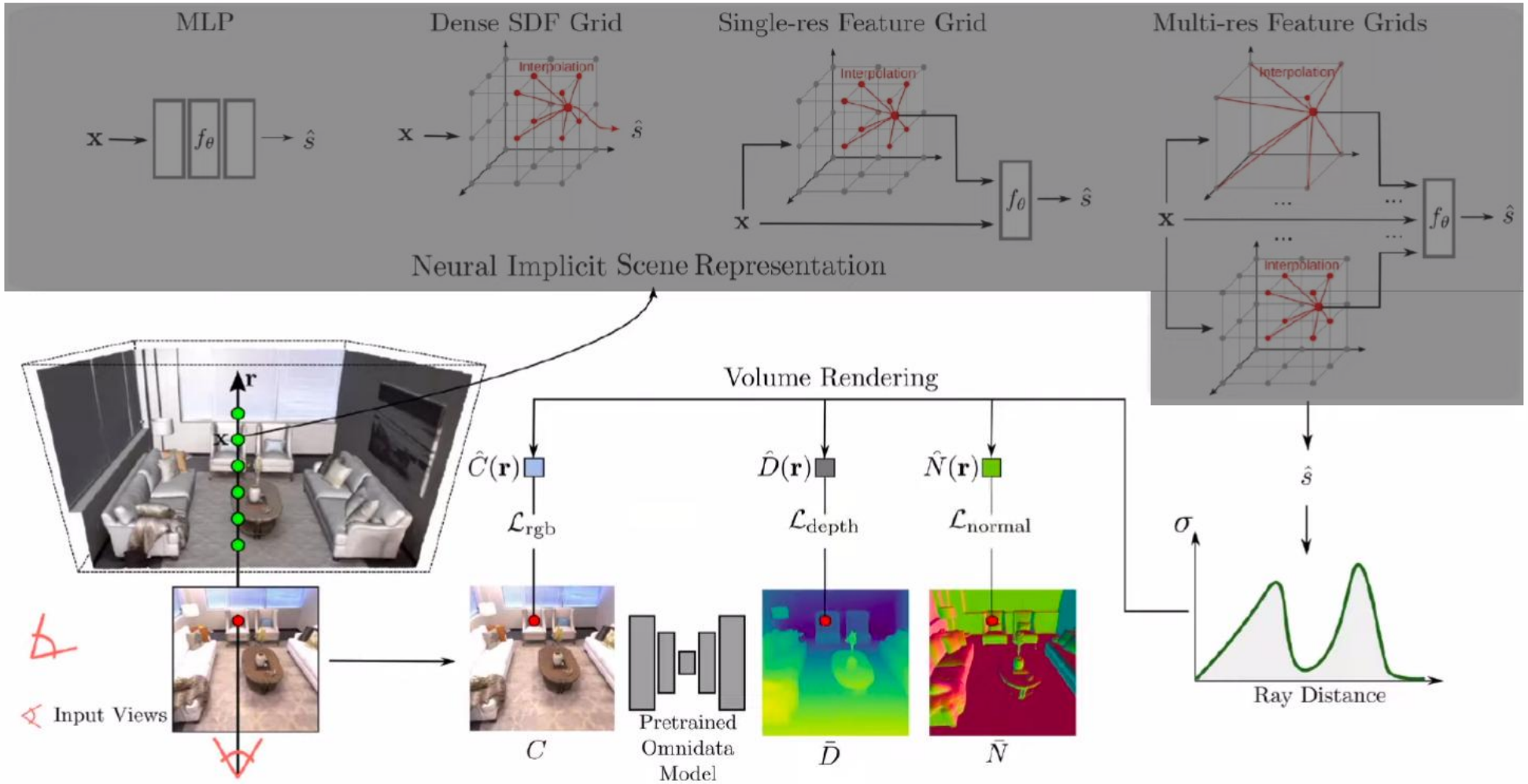
$$\hat{\mathbf{c}} = \mathbf{c}_\theta(\mathbf{x}, \mathbf{v}, \hat{\mathbf{n}}, \hat{\mathbf{z}})$$



- where
  - $\mathbf{x}$ is a 3D point, and $\mathbf{v}$ is **viewing direction.**
  - **Normal vector** $\hat{\mathbf{n}}$ is computed analytically.
  - $\hat{\mathbf{z}}$ is **feature-vector** predicted by implicit network.

Yu, Peng, Niemeyer, Sattler, Geiger: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. NeuIPS, 2022.

MonoSDF Pipeline



Yu, Peng, Niemeyer, Sattler, Geiger: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. NeuIPS, 2022.

SDF to Density Transformation

- Sample M **points** $\mathbf{x}_\mathbf{r}^i = \mathbf{o} + t_\mathbf{r}^i \mathbf{v}$ along the **ray r**.

SDF to Density Transformation

- Sample M **points** $\mathbf{x}_\mathbf{r}^i = \mathbf{o} + t_\mathbf{r}^i \mathbf{v}$ along the **ray r**.
- Predict their SDF value $\hat{s}$ and color $\hat{\mathbf{c}}$.

SDF to Density Transformation

- Sample M **points** $\mathbf{x}_\mathbf{r}^i = \mathbf{o} + t_\mathbf{r}^i \mathbf{v}$ along the **ray r**.
- Predict their SDF value $\hat{s}$ and color $\hat{\mathbf{c}}$.
- Convert SDF value to **density**:

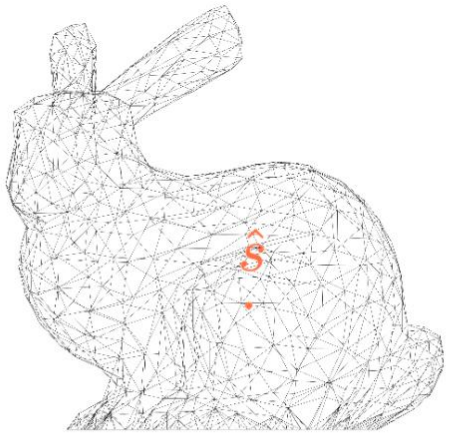
$$\sigma_\beta(s) = \begin{cases} \frac{1}{2\beta} \exp\left(\frac{s}{\beta}\right) & \text{if } s \leq 0 \\ \frac{1}{\beta}\left(1 - \frac{1}{2}\exp\left(-\frac{s}{\beta}\right)\right) & \text{if } s > 0 \end{cases}$$

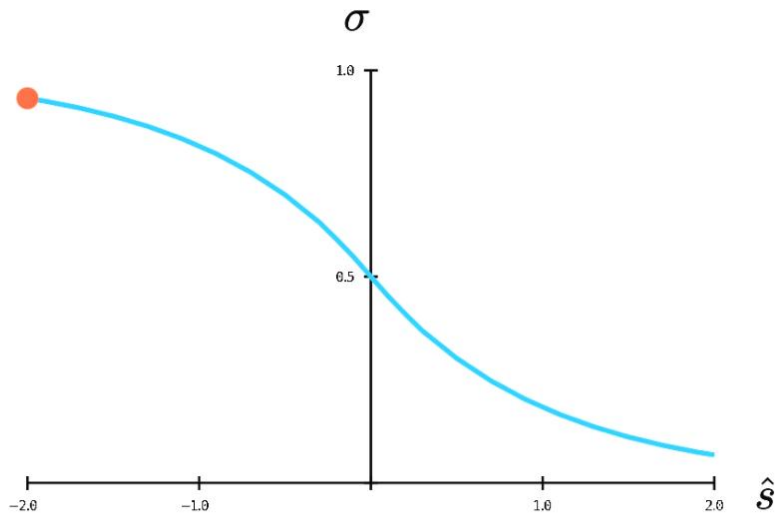
where

- β is a **learnable parameter**.
- σ indicates density.

SDF to Density Transformation

$$\sigma(s) = \begin{cases} \frac{1}{2\beta} e^{s/\beta}, & \text{if } s \leq 0 \\ \frac{1}{\beta}\left(1 - \frac{1}{2} e^{-s/\beta}\right), & \text{if } s > 0 \end{cases}$$



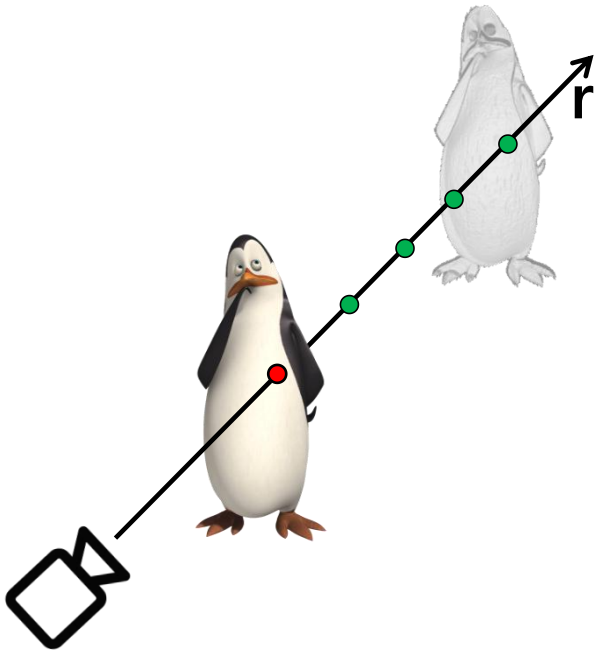


SDF to Density Transformation

Volume Rendering

– Color $\hat{C}(\mathbf{r})$ for ray **r** is rendered as below:

$$\hat{C}(\mathbf{r}) = \sum_{i=1}^{M} T_{\mathbf{r}}^{i}\, \alpha_{\mathbf{r}}^{i}\, \hat{\mathbf{c}}_{\mathbf{r}}^{i} \qquad T_{\mathbf{r}}^{i} = \prod_{j=1}^{i-1} \left(1 - \alpha_{\mathbf{r}}^{j}\right) \qquad \alpha_{\mathbf{r}}^{i} = 1 - \exp\left(-\sigma_{\mathbf{r}}^{i} \delta_{\mathbf{r}}^{i}\right)$$



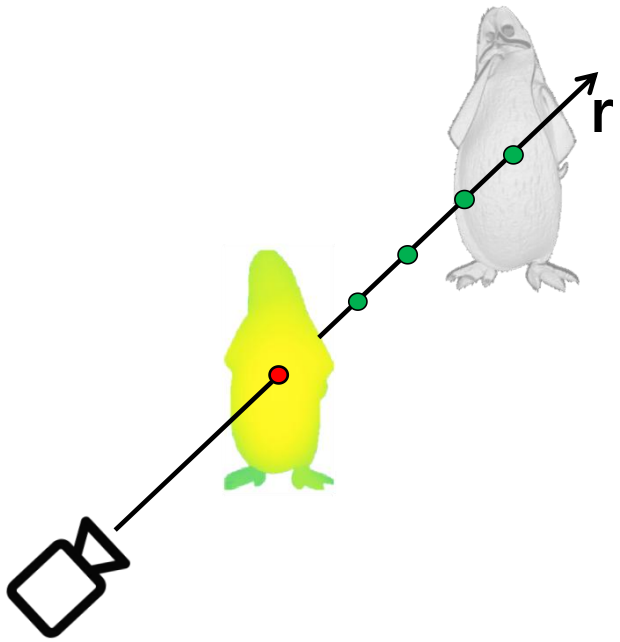
where

– $T_{\mathbf{r}}^{i}$ is transmittance, and $\alpha_{\mathbf{r}}^{i}$ is alpha value of sample i.

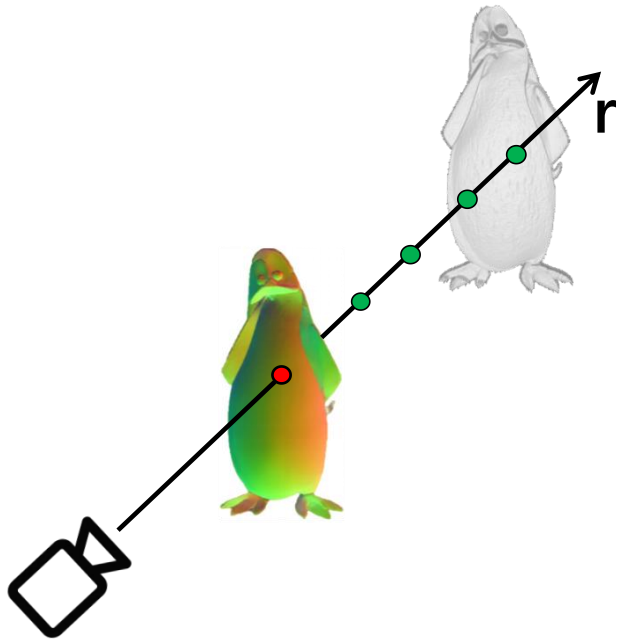
Volume Rendering

– Similarly, render depth $\hat{D}(\mathbf{r})$ and normal $\hat{N}(\mathbf{r})$:

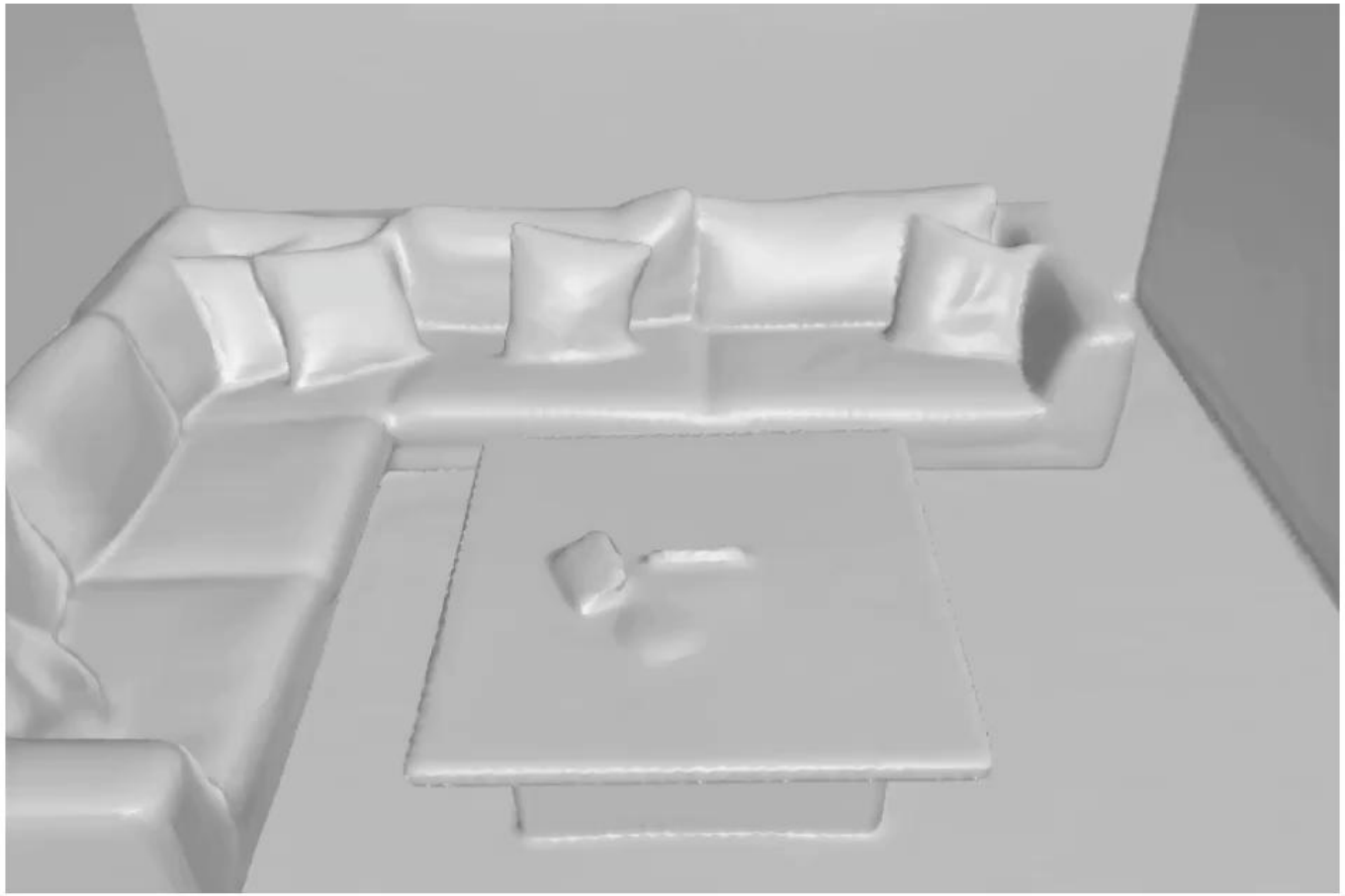
$$\hat{D}(\mathbf{r}) = \sum_{i=1}^{M} T_{\mathbf{r}}^{i}\, \alpha_{\mathbf{r}}^{i}\, t_{\mathbf{r}}^{i}$$



$$\hat{N}(\mathbf{r}) = \sum_{i=1}^{M} T_{\mathbf{r}}^{i}\, \alpha_{\mathbf{r}}^{i}\, \hat{\mathbf{n}}_{\mathbf{r}}^{i}$$



Ablation Study: Depth & Normal Cues

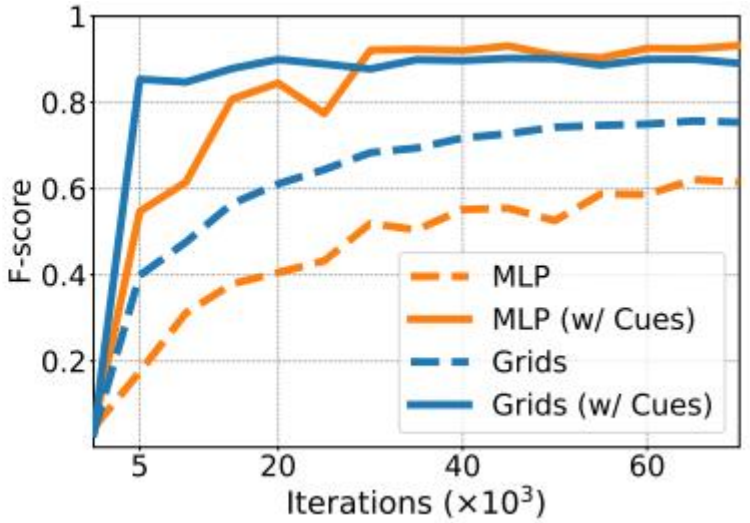


Depth & Normal Cues

Yu, Peng, Niemeyer, Sattler, Geiger: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. NeuIPS, 2022.

Ablation Study: Replica Dataset

		Normal C.↑	Chamfer-L_1 ↓	F-score ↑
MLP	No Cues	86.48	6.75	66.88
	Only Depth	90.56	4.26	76.42
	Only Normal	91.35	3.19	85.84
	Both Cues	**92.11**	**2.94**	**86.18**
Multi-Res. Grids	No Cues	87.95	5.03	78.38
	Only Depth	90.87	3.75	80.32
	Only Normal	89.90	3.61	81.28
	Both Cues	**90.93**	**3.23**	**85.91**

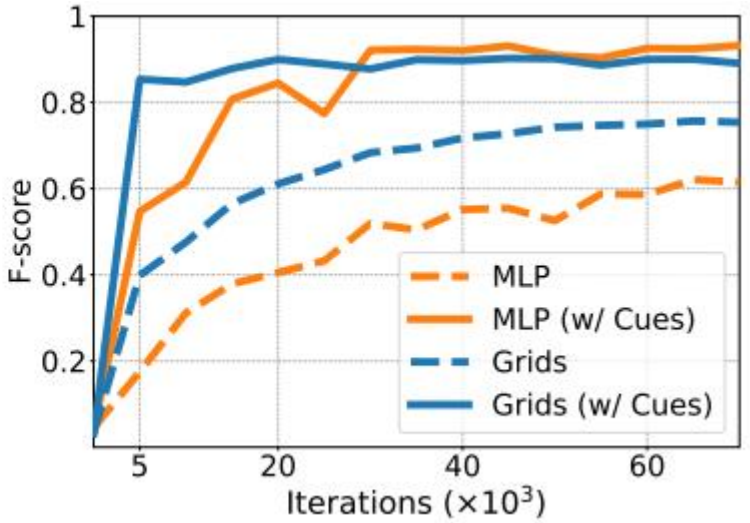


– Combining **depth** & **normal** cues lead to best result for both MLP, and multi-resolution feature grids.

Yu, Peng, Niemeyer, Sattler, Geiger: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. NeuIPS, 2022.

Ablation Study: Replica Dataset

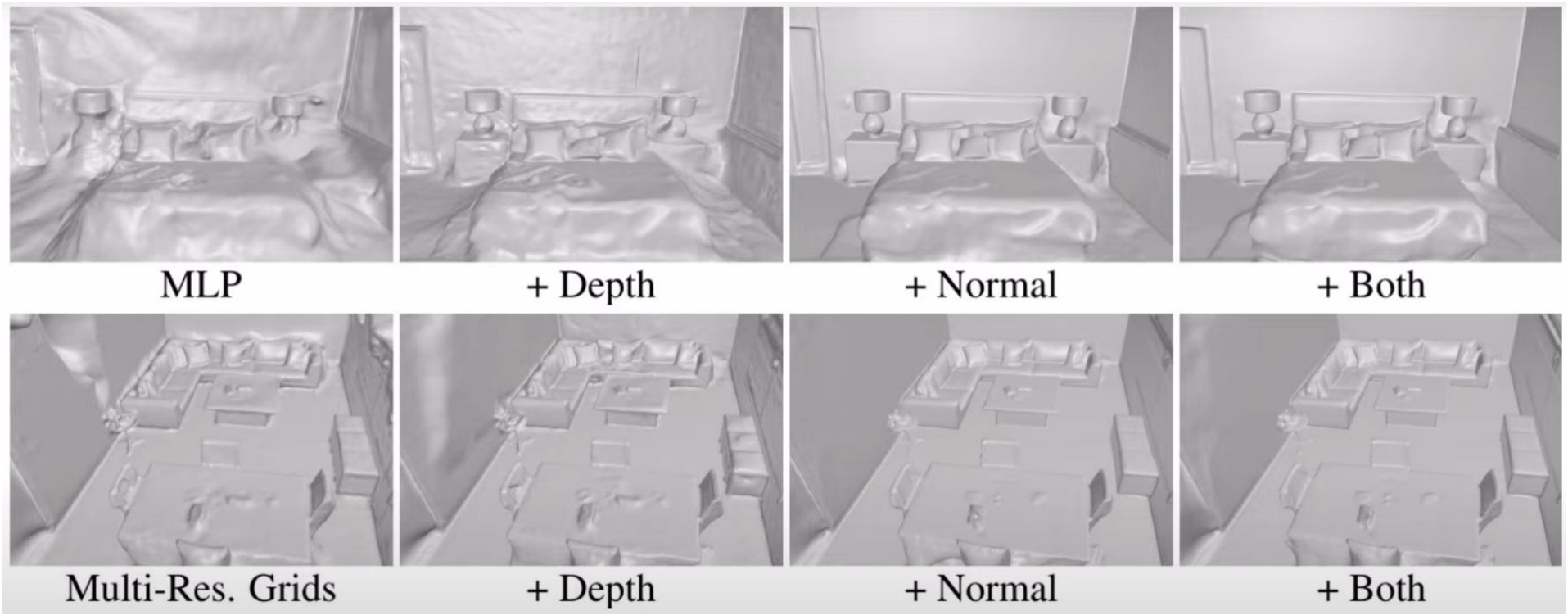
		Normal C.↑	Chamfer-L_1 ↓	F-score ↑
MLP	No Cues	86.48	6.75	66.88
	Only Depth	90.56	4.26	76.42
	Only Normal	91.35	3.19	85.84
	Both Cues	**92.11**	**2.94**	**86.18**
Multi-Res. Grids	No Cues	87.95	5.03	78.38
	Only Depth	90.87	3.75	80.32
	Only Normal	89.90	3.61	81.28
	Both Cues	**90.93**	**3.23**	**85.91**



- Combining **depth** & **normal** cues lead to best result for both MLP, and multi-resolution feature grids.
- Monocular cues improve **convergence speed**.
- Feature grids converge slightly **faster**, while MLP w/ cues yield **best accuracy**.

Yu, Peng, Niemeyer, Sattler, Geiger: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. NeuIPS, 2022.

Ablation Study: Replica Dataset



- **Depth cues** yields better **overall structure** and detailed geomtery.
- **Normal cues** help fill the missing details and remove **noise**.

Yu, Peng, Niemeyer, Sattler, Geiger: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. NeuIPS, 2022.

Ablation Study: ScanNet

	COLMAP [62]	UNISURF [49]	NeuS [76]	VolSDF [81]	M-SDF [21]	NeuRIS [75]	Ours (Grids)	Ours (MLP)
Chamfer-L_1 ↓	0.141	0.359	0.194	0.267	0.070	0.050	0.064	**0.042**
F-score↑	0.537	0.267	0.291	0.364	0.602	0.692	0.626	**0.733**

- COLMAP and **VolSDF** completely **fails** to reconstruct the geometry.
- Manhattan-SDF achieves better results, but still **less detailed** and noisy.

Yu, Peng, Niemeyer, Sattler, Geiger: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. NeuIPS, 2022.

Ablation Study: ScanNet

	COLMAP [62]	UNISURF [49]	NeuS [76]	VolSDF [81]	M-SDF [21]	NeuRIS [75]	Ours (Grids)	Ours (MLP)
Chamfer-L_1 ↓	0.141	0.359	0.194	0.267	0.070	0.050	0.064	**0.042**
F-score↑	0.537	0.267	0.291	0.364	0.602	0.692	0.626	**0.733**

- COLMAP and **VolSDF** completely **fails** to reconstruct the geometry.
- Manhattan-SDF achieves better results, but still **less detailed** and noisy.
- MLP perform better than Grids because ScanNet contains **motion-blur images** and **noisy camera poses**.

Yu, Peng, Niemeyer, Sattler, Geiger: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. NeuIPS, 2022.

Ablation Study: DTU Dense Views

Scan	24	37	40	55	63	65	69	83	97	105	106	110	114	118	122	**Mean**
COLMAP	0.81	2.05	0.73	1.22	1.79	1.58	1.02	3.05	1.40	2.05	1.00	1.32	0.49	0.78	1.17	1.36
NeRF [44]	1.90	1.60	1.85	0.58	2.28	1.27	1.47	1.67	2.05	1.07	0.88	2.53	1.06	1.15	0.96	1.49
UniSurf [49]	1.32	1.36	1.72	0.44	1.35	0.79	0.80	1.49	1.37	0.89	0.59	1.47	0.46	0.59	0.62	1.02
NeuS [77]	1.00	1.37	0.93	0.43	1.10	**0.65**	**0.57**	1.48	**1.09**	0.83	**0.52**	1.20	0.35	**0.49**	0.54	0.84
VolSDF [81]	1.14	1.26	0.81	0.49	1.25	0.70	0.72	1.29	1.18	0.70	0.66	1.08	0.42	0.61	0.55	0.86
Ours (MLP)	0.83	1.61	0.65	0.47	0.92	0.87	0.87	1.30	1.25	0.68	0.65	**0.96**	**0.41**	0.62	0.58	0.84
Ours(Grids)	**0.66**	**0.88**	**0.43**	**0.40**	**0.87**	0.78	0.81	**1.23**	1.18	**0.66**	0.66	**0.96**	**0.41**	0.57	**0.51**	**0.73**

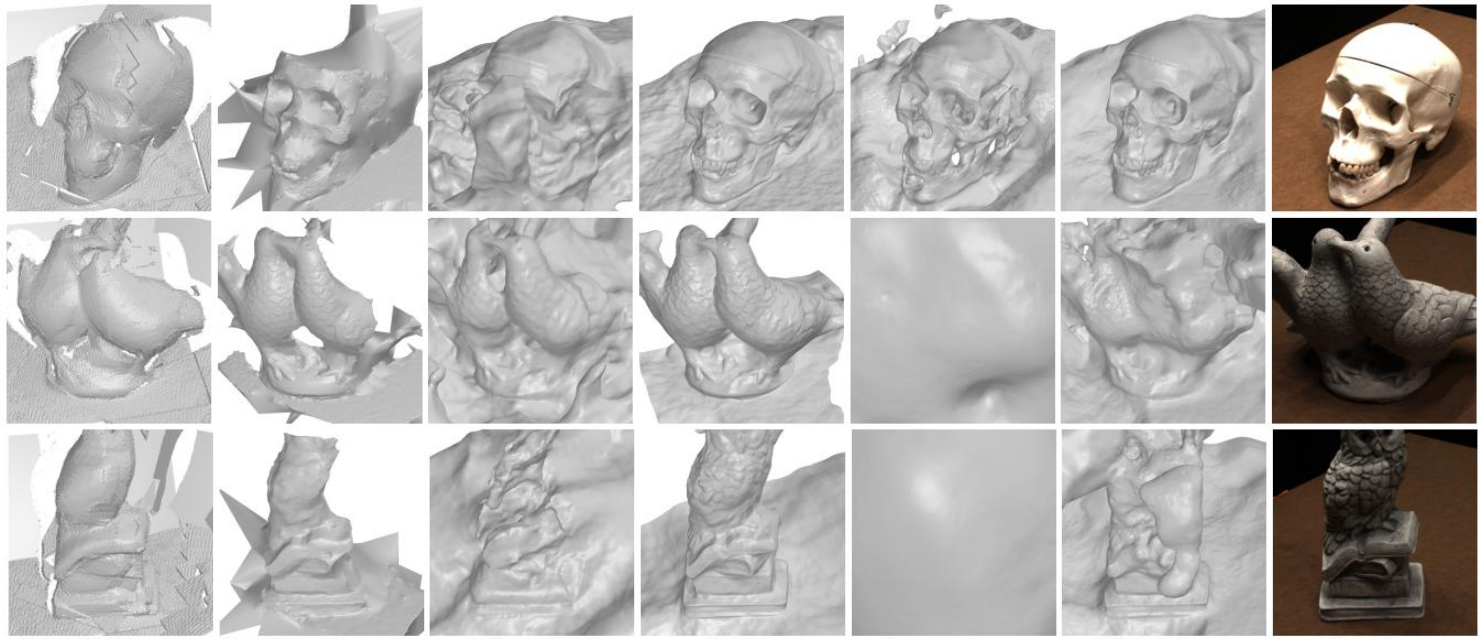
Chamfer-L_1 ↓

– For multi-resolution feature **grids** monocular cues help to suppress noise, reconstruct **smooth** surfaces.

Yu, Peng, Niemeyer, Sattler, Geiger: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. NeuIPS, 2022.

Ablation Study: DTU Sparse Views

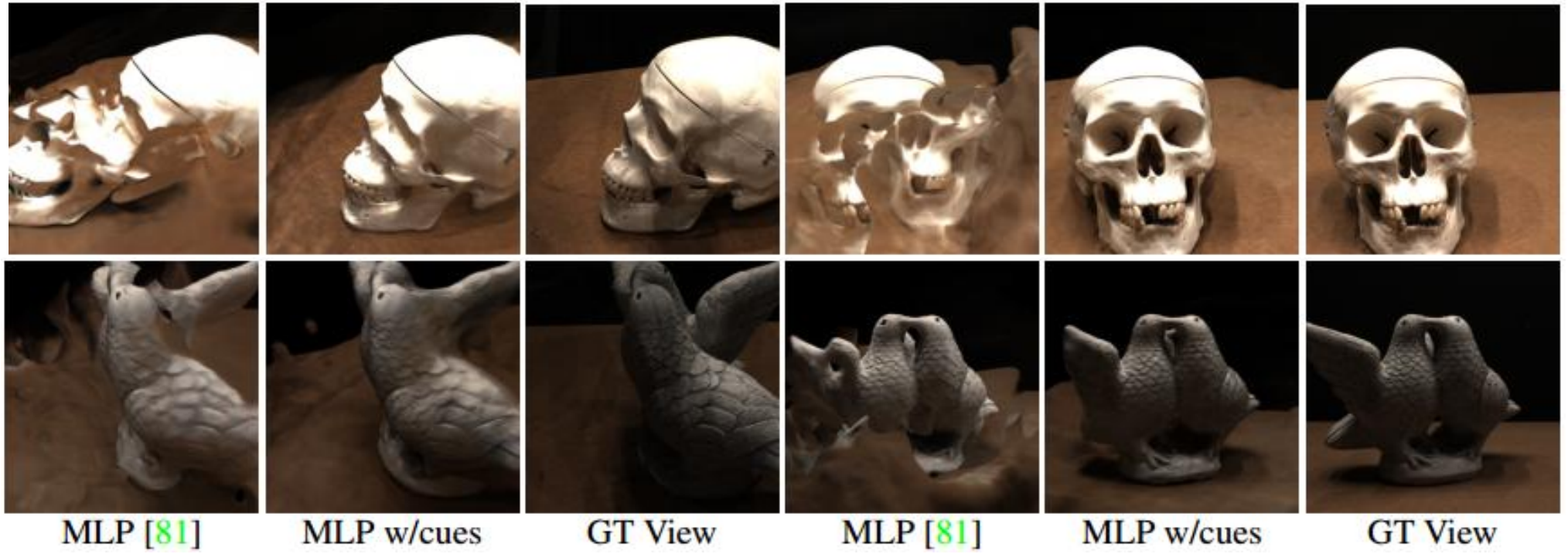
	Chamfer-L_1 ↓
TSDF-Fusion [12]	4.80
COLMAP [62]	2.56
RealityCapture	2.84
Grids	6.47
Grids w/ cues	3.68
MLP [81]	4.21
MLP w/ cues	**1.86**



- **Few-shot** reconstruction (3 input views) for MLP w/ cues perform well due to **inductive bias**.
- MLP w/ cues, more robust to **less-observed** regions.

Yu, Peng, Niemeyer, Sattler, Geiger: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. NeuIPS, 2022.

Ablation Study: DTU Novel View Systhesis



– Monocular geometric **cues** help improve **few-shot** novel view systhesis (3 input views).

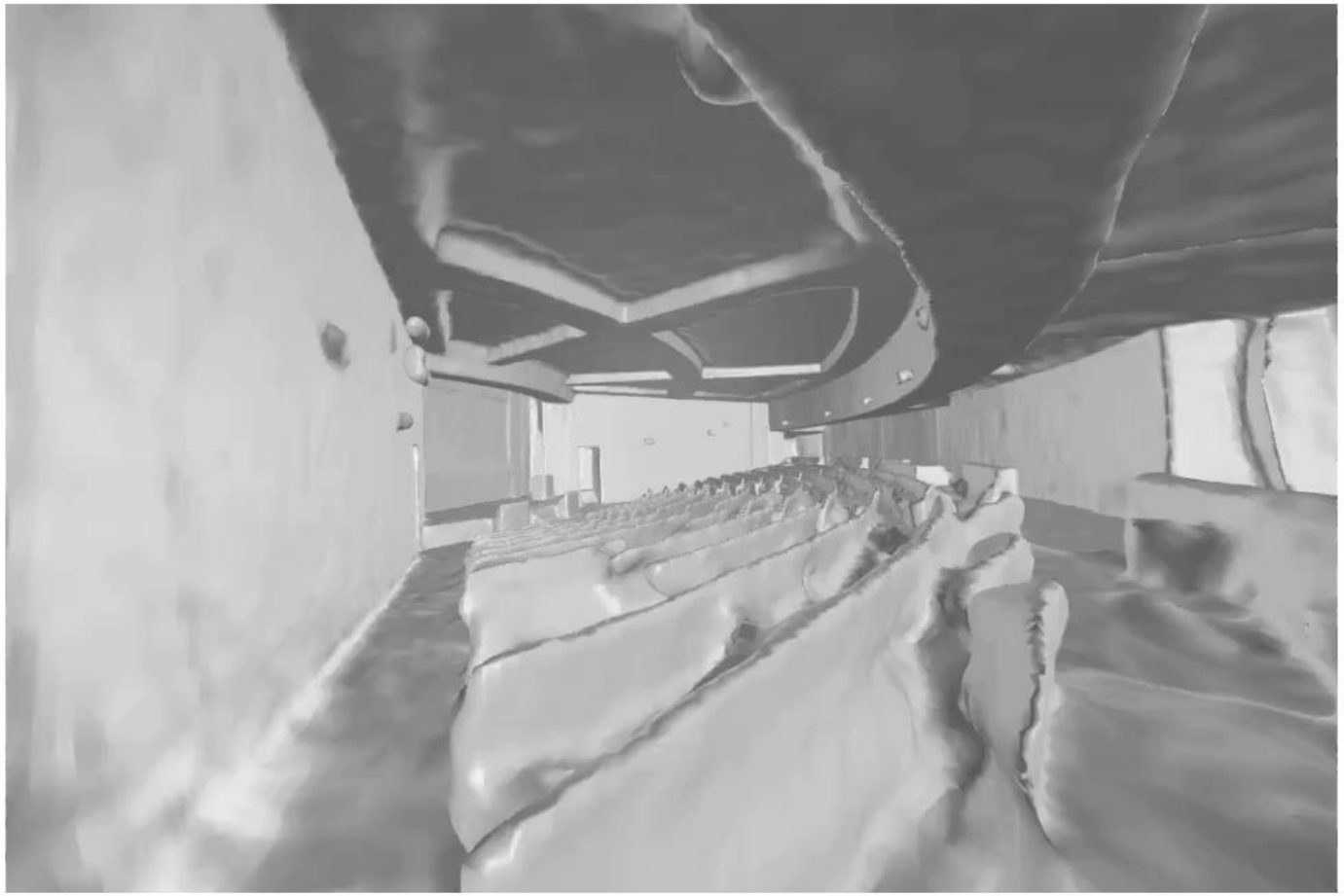
Yu, Peng, Niemeyer, Sattler, Geiger: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. NeuIPS, 2022.

Results: ScanNet

Ours

Yu, Peng, Niemeyer, Sattler, Geiger: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. NeuIPS, 2022.

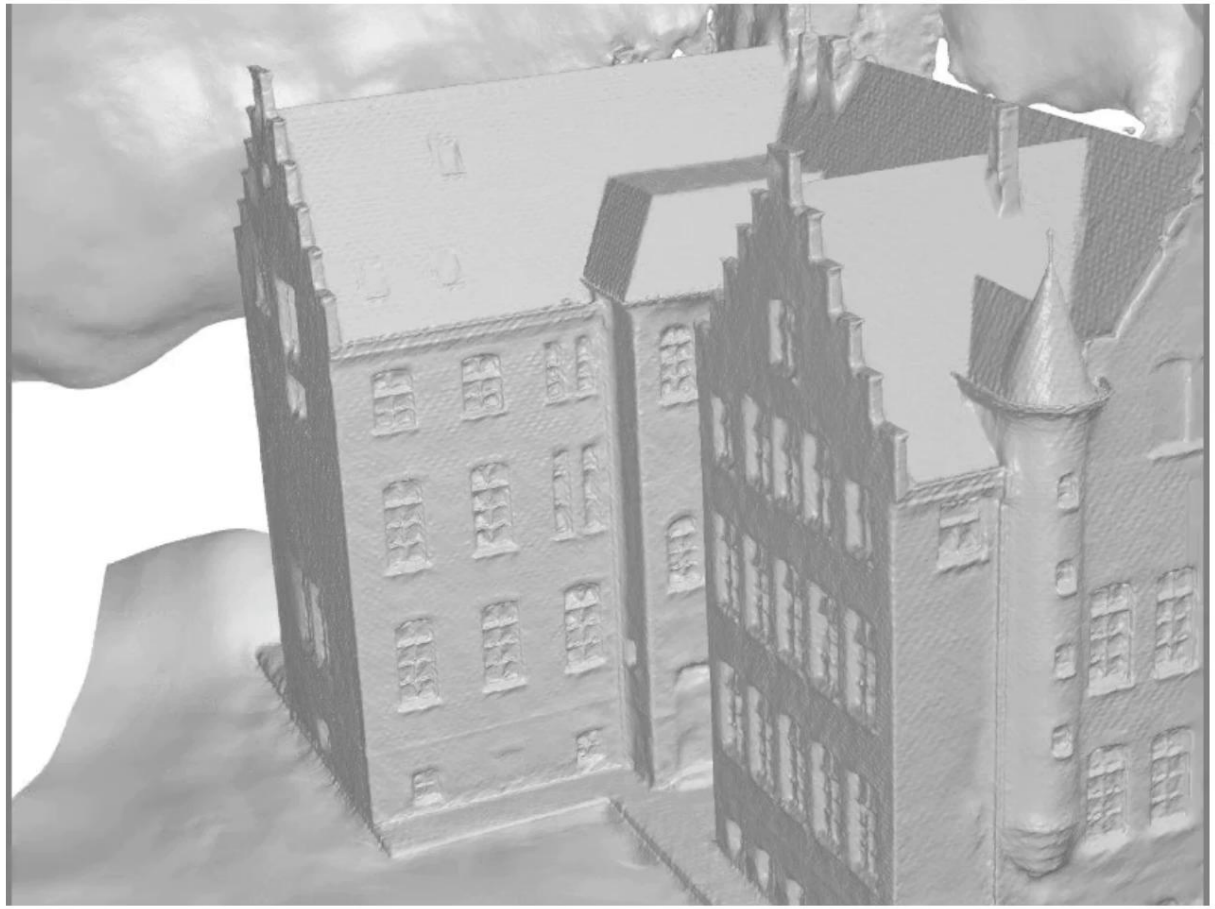
Results: Tanks & Temples



Ours

Yu, Peng, Niemeyer, Sattler, Geiger: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. NeuIPS, 2022.

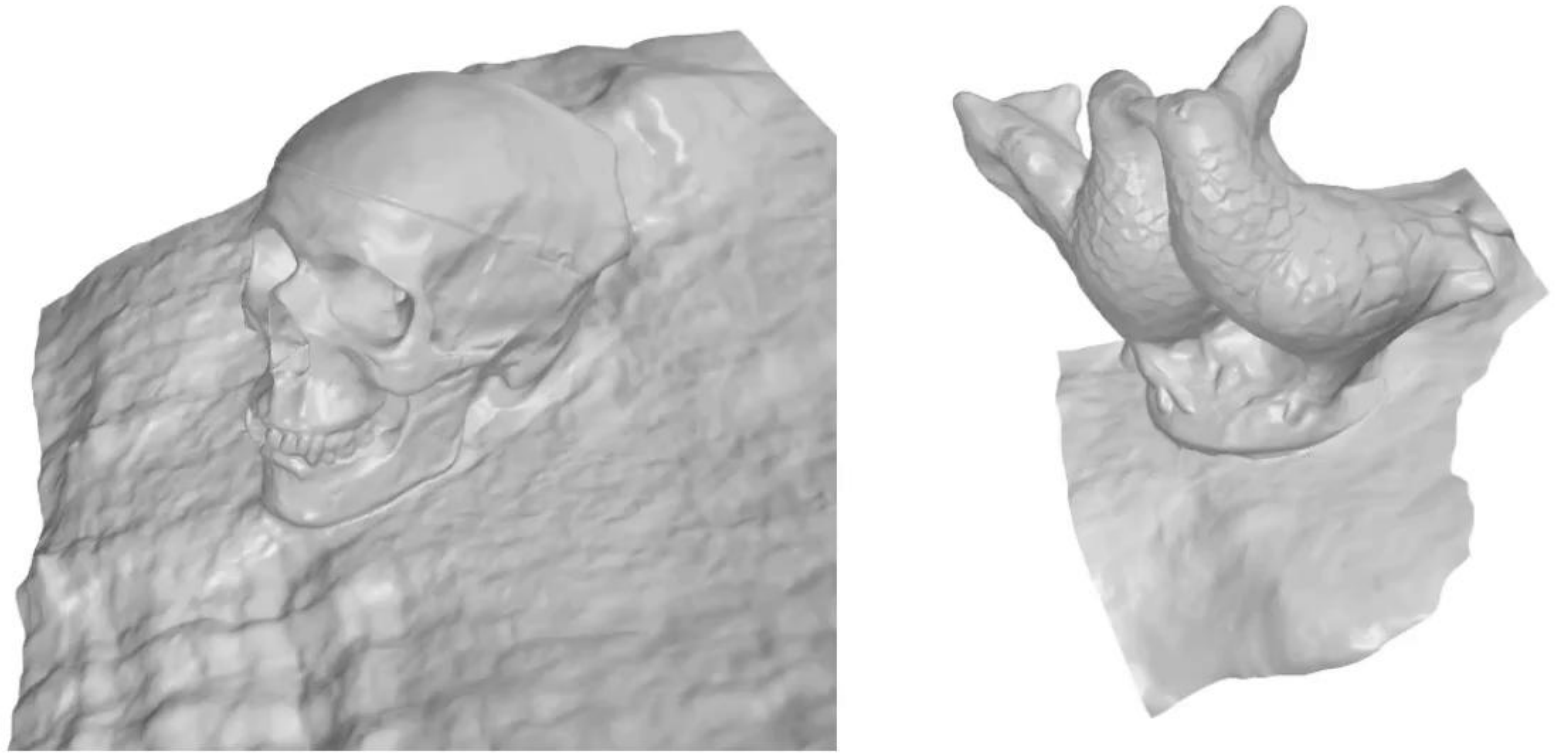
Results: DTU Dense Views



Ours (Grids)

Yu, Peng, Niemeyer, Sattler, Geiger: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. NeuIPS, 2022.

Results: DTU Sparse Views



Ours

Yu, Peng, Niemeyer, Sattler, Geiger: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. NeuIPS, 2022.

Limitations

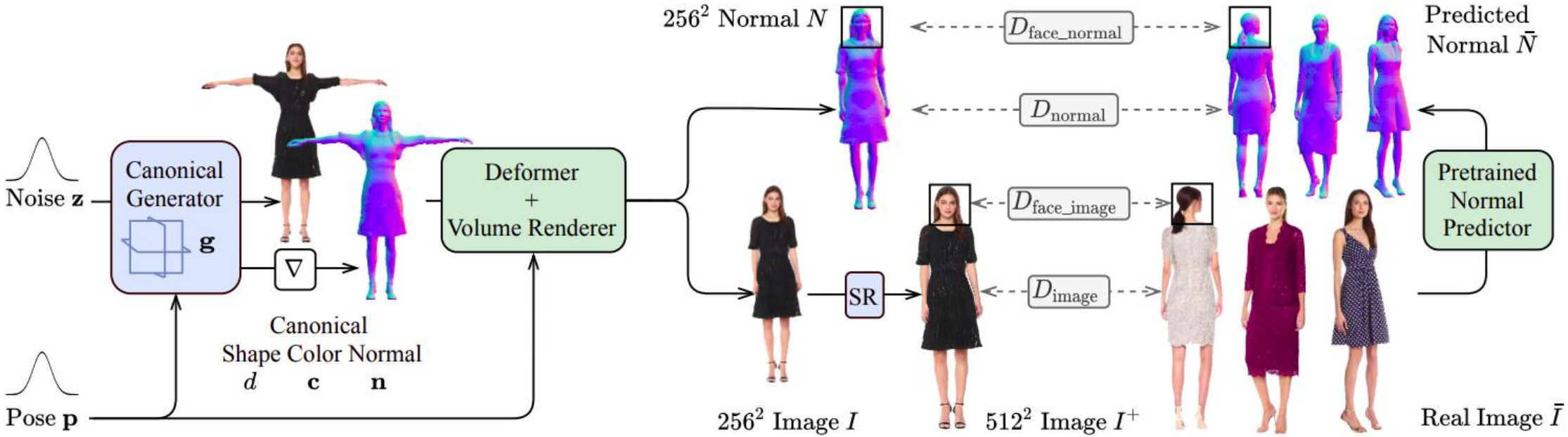
- Reconstruction **quality** depends on quality of **monocular cues**.

Limitations

- Reconstruction **quality** depends on quality of **monocular cues**.
- Low-resolution output of the **Omnidata** model.

Follow-up Work

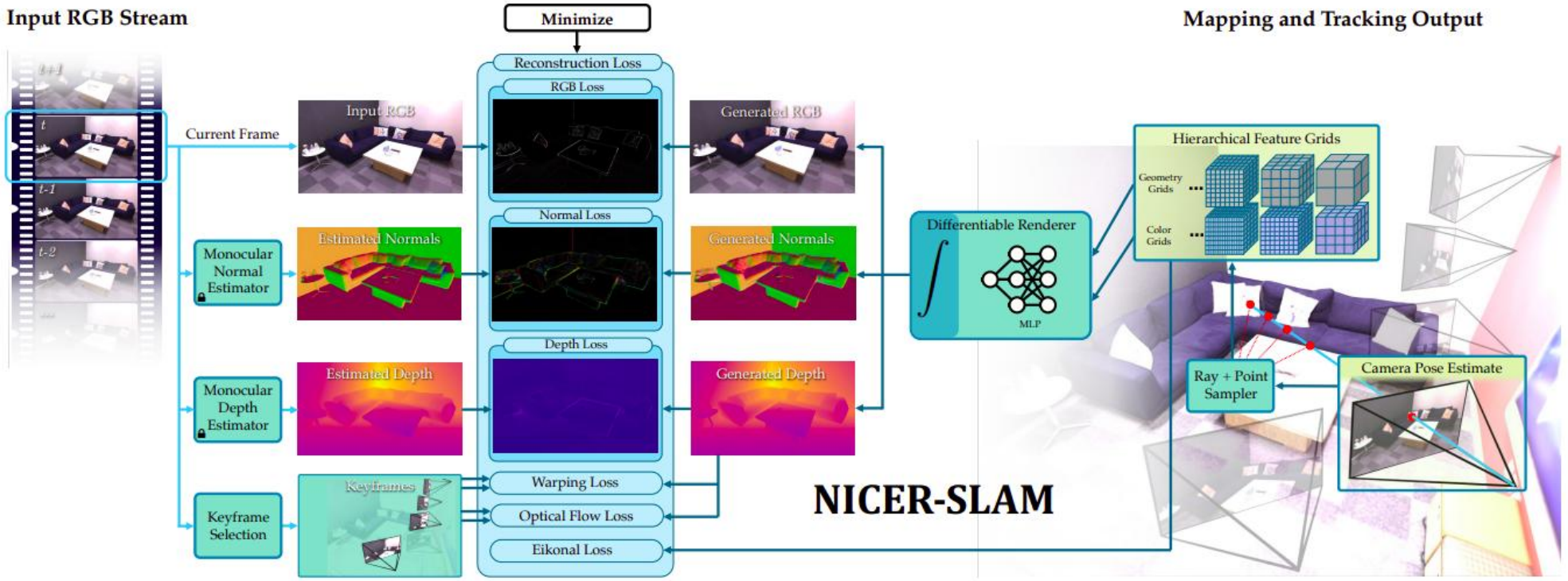
AG3D: Learning to Generate 3D Avatars from 2D Image Collections



Dong, Chen, Yang, Black, Hilliges, Geiger: AG3D: Learning to Generate 3D Avatars from 2D Image Collections. arXiv preprint, 2023.

Follow-up Work

NICER-SLAM: Neural Implicit Scene Encoding for RGB SLAM



Yu, Zhu, Peng, Larsson, Cui, Oswald, Geiger, Pollefeys: NICER-SLAM: Neural Implicit Scene Encoding for RGB SLAM. arXiv preprint, 2023.

Follow-up Work

Neuralangelo: High-Fidelity Neural Surface Reconstruction

Li, Müller, Evans, Taylor, Unberath, Liu, Lin: Neuralangelo: High-Fidelity Neural Surface Reconstruction. CVPR, 2023.

Thank you for your attention!
Questions and comments are welcome.

Thank you for your attention!
Questions and comments are welcome.