NeILF: Neural Incident Light Field for Multi-view Geometry and Material Estimation

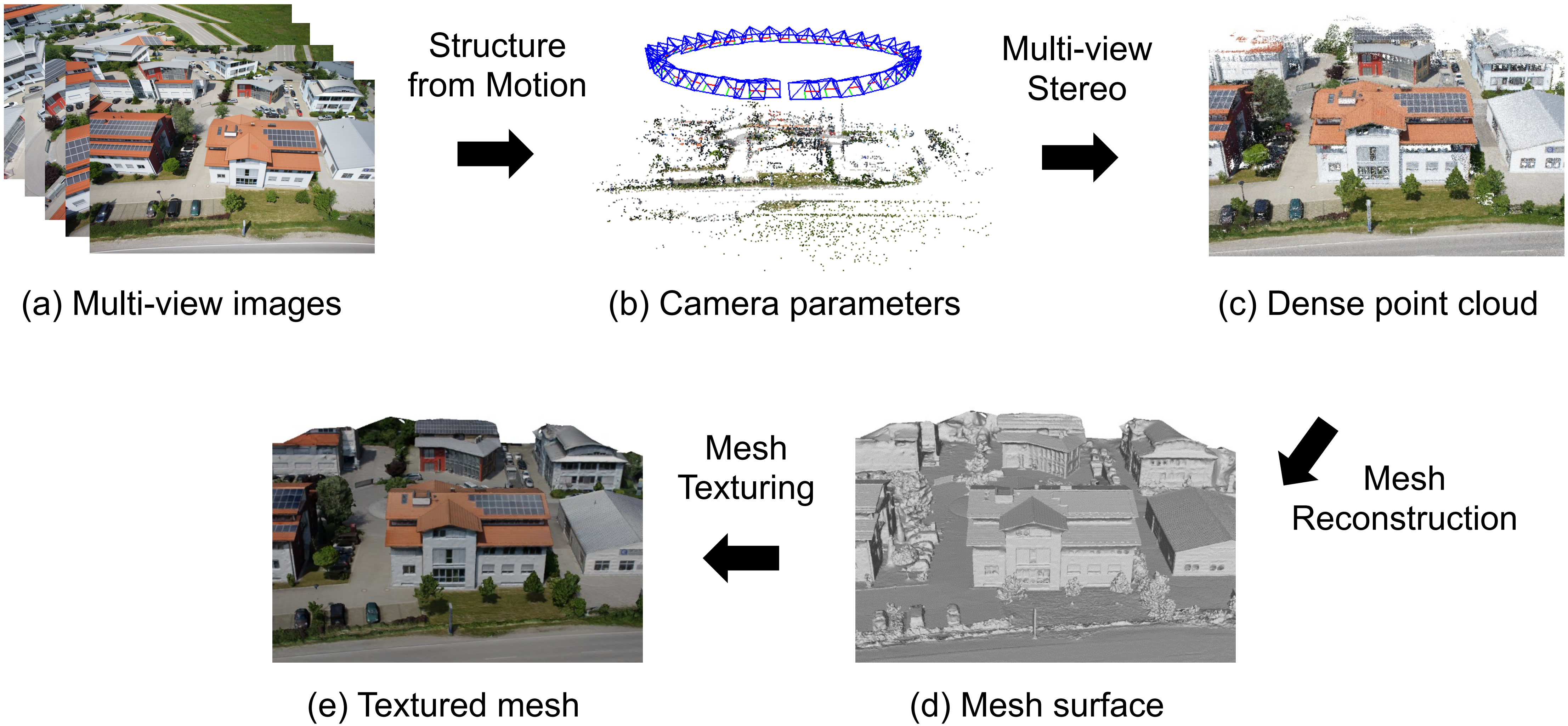
Yao Yao

3D Vision Lab @ Nanjing University

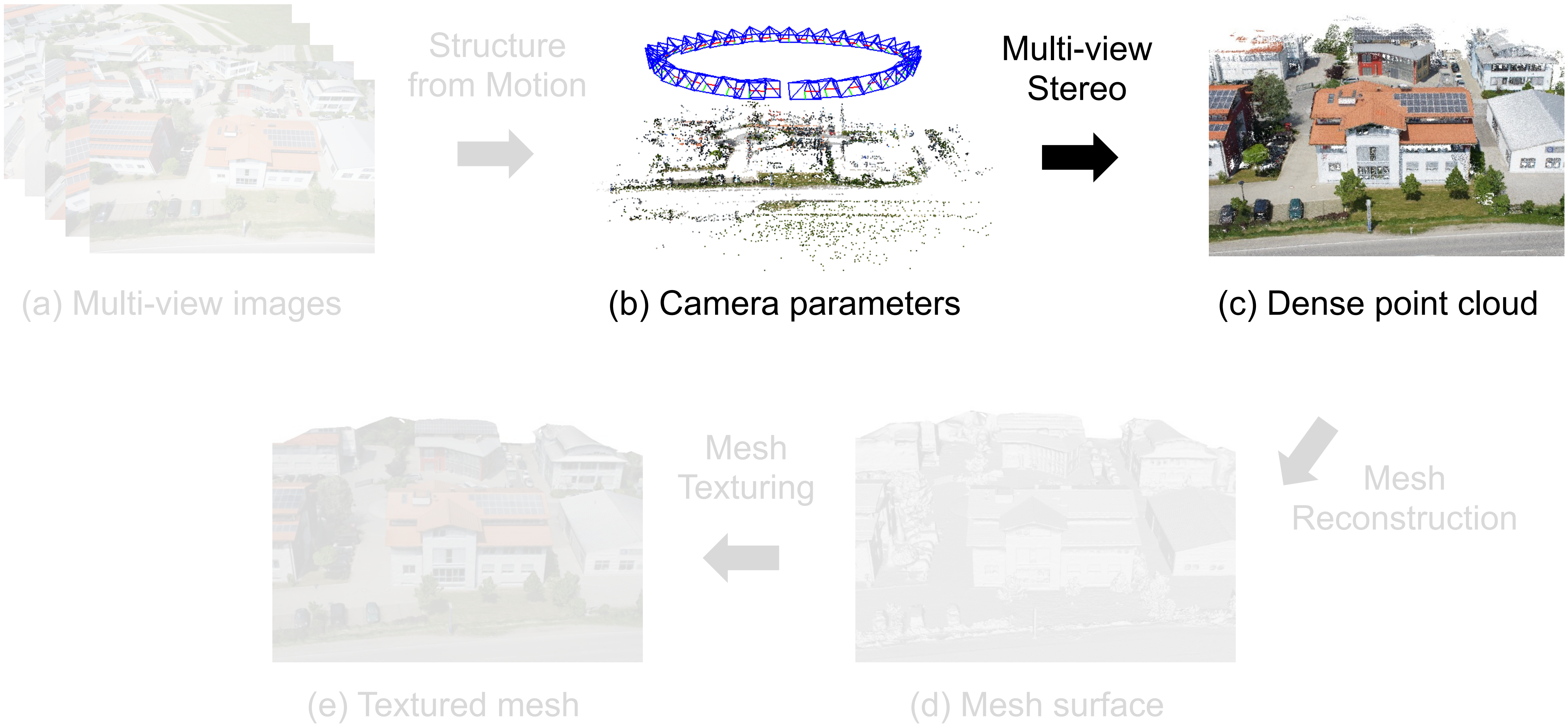
Overview

- Introduction to Multi-view 3D Reconstruction
- Introduction to Differentiable Rendering
- NeILF for Geometry and Material Estimation

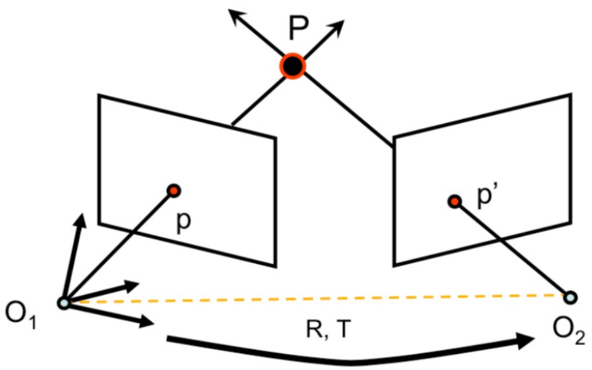
Introduction: Photogrammetry Pipeline



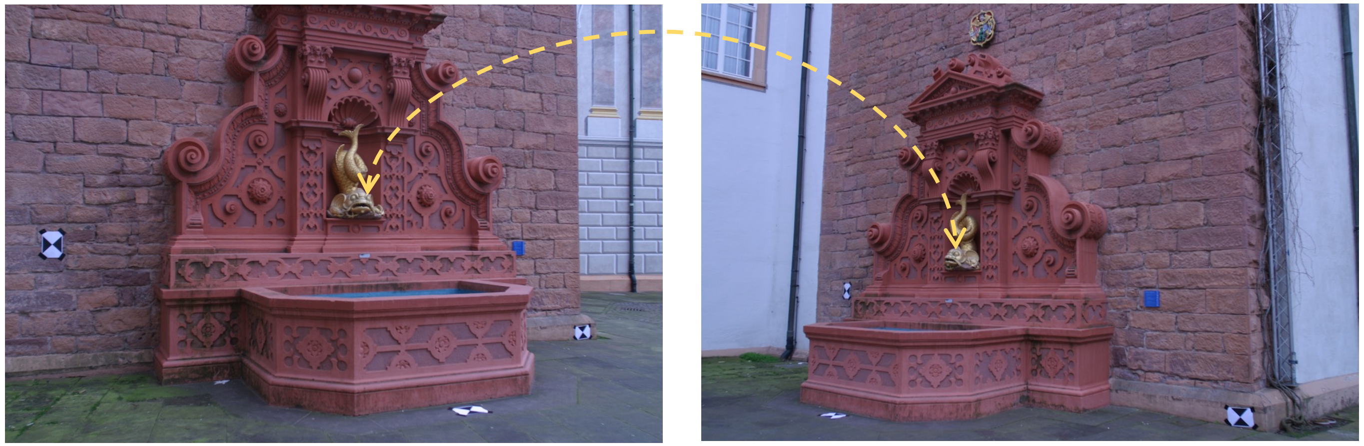
Introduction: Multi-view Stereo (MVS)



Introduction: Classical MVS



(a) 3D triangulation (from slides [1])



(b) Visual correspondences

How to find better **pixelwise** correspondences?

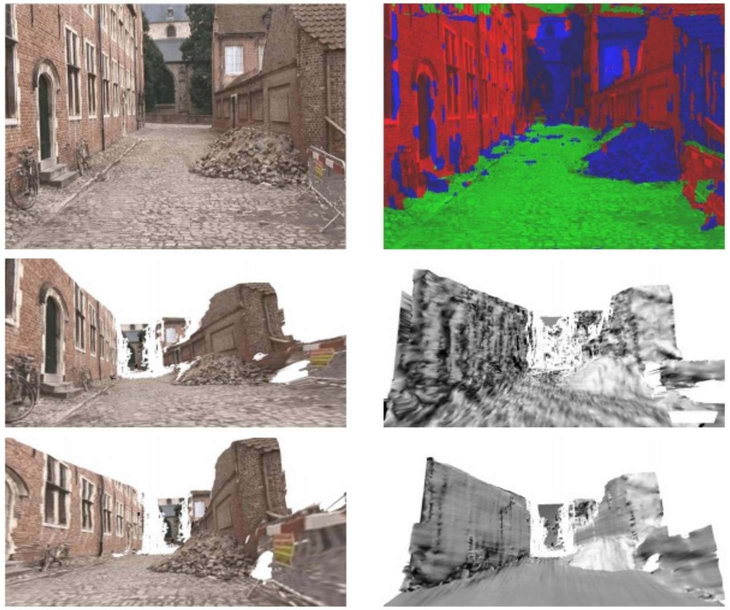
- Hand-crafted matching cost (e.g., SSD and NCC)

How to find better **dense** correspondences?

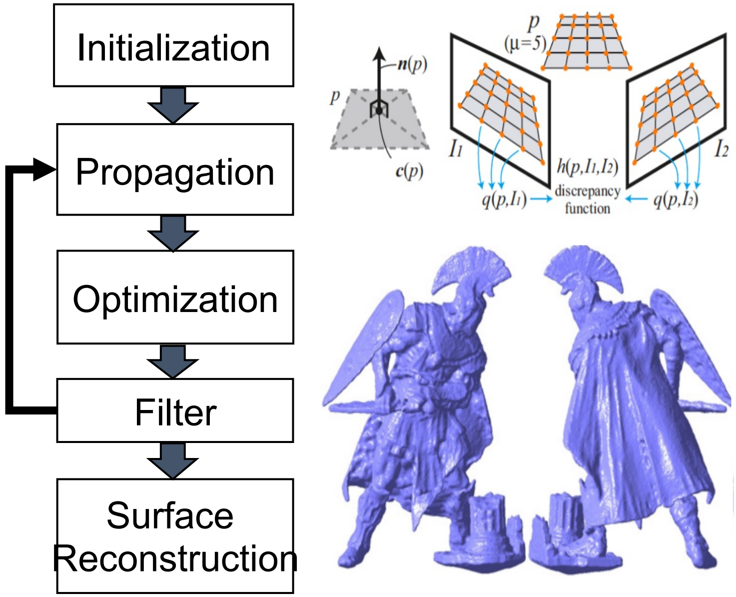
- Engineered cost regularizations (e.g., semi-global matching, propagation)

[1] https://web.stanford.edu/class/cs231a/course_notes/03-epipolar-geometry.pdf

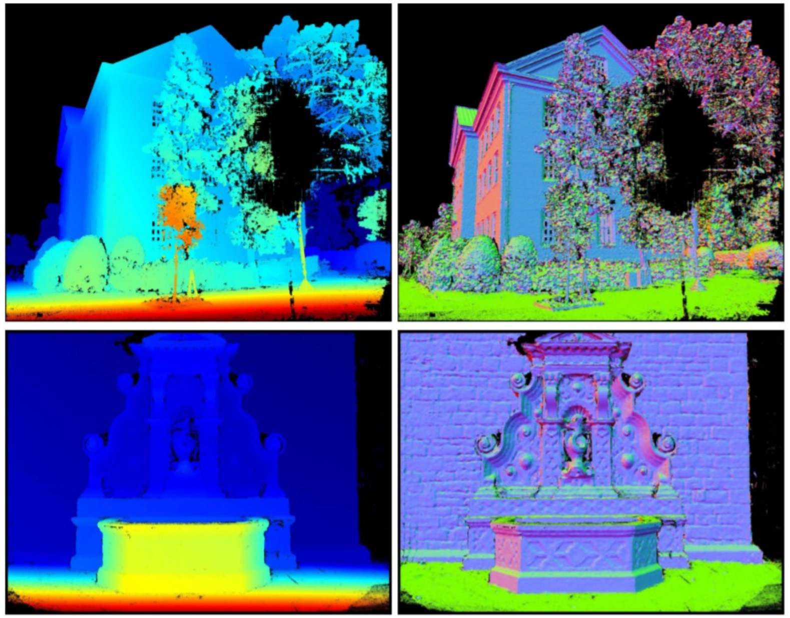
Introduction: Classical MVS



Multi-direction Planesweep
Gallup et al. CVPR2007



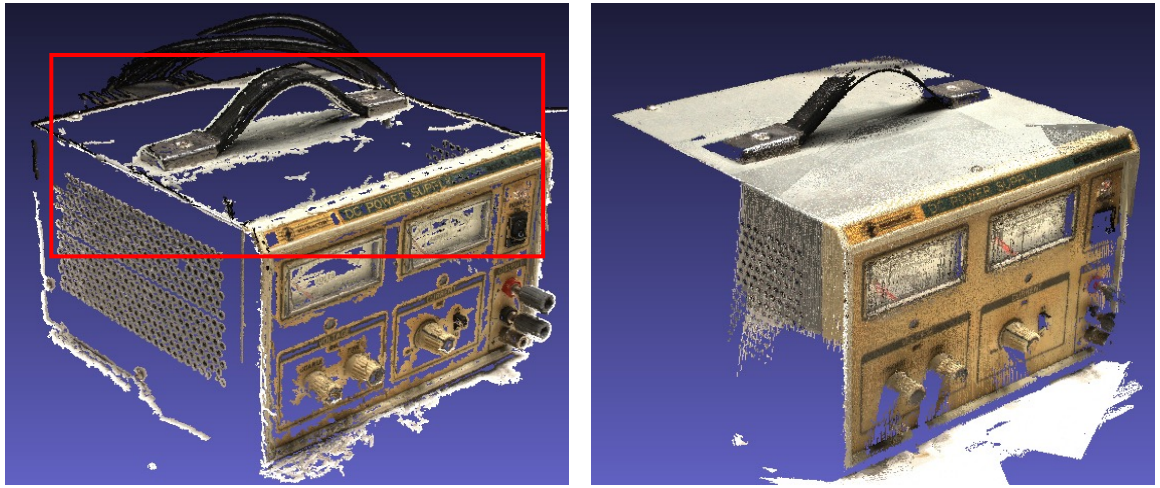
PMVS
Furukawa and Ponce. PAMI2010



Colmap
Schonberger et al. ECCV2016

Learning-based MVS: Motivation and Difficulties

✓ Well-textured areas



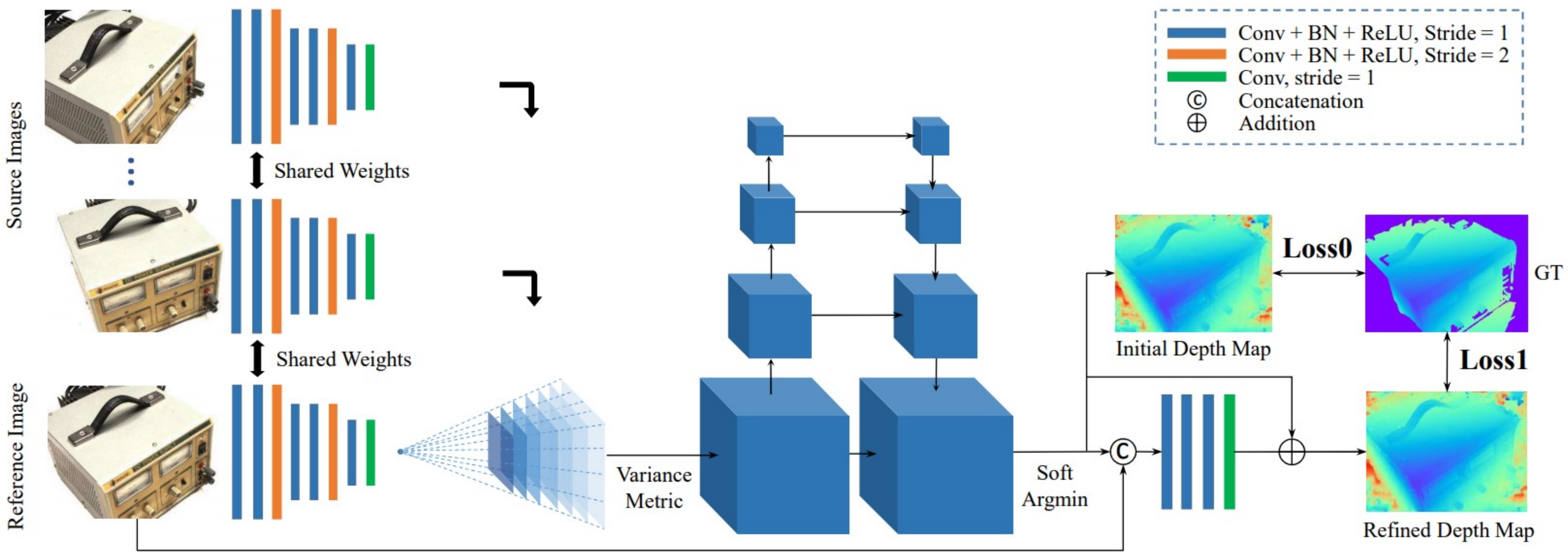
Ground Truth

× Textureless, Non-lambertian areas

MVS + Deep Learning:

- Arbitrary N-views input?
- Unstructured camera geometry?
- Generalization?

MVSNet: Network Architecture



- End-to-end MVS learning framework
- Camera geometry encoded as differentiable homography
- Variance-based Cost Metric for N-view similarity

MVSNet: Depth Inference for Unstructured Multi-view Stereo. **Yao et al**. ECCV2018 oral

MVSNet: Results on DTU Dataset

(a) MVSNet (b) Ground Truth

	Mean Distance (mm)			Percentage ($<1mm$)			Percentage ($<2mm$)		
	Acc.	Comp.	*overall*	Acc.	Comp.	*f-score*	Acc.	Comp.	*f-score*
Camp [3]	0.835	0.554	0.695	71.75	64.94	66.31	84.83	67.82	73.02
Furu [7]	0.613	0.941	0.777	69.55	61.52	63.26	78.99	67.88	70.93
Tola [35]	0.342	1.190	0.766	90.49	57.83	68.07	93.94	63.88	73.61
Gipuma [8]	**0.283**	0.873	0.578	**94.65**	59.93	70.64	**96.42**	63.81	74.16
SurfaceNet[14]	0.450	1.04	0.745	83.8	63.38	69.95	87.15	67.99	74.4
MVSNet (Ours)	0.396	**0.527**	**0.462**	86.46	**71.13**	**75.69**	91.06	**75.31**	**80.25**

- Highest **f-score**
- Significant improvement on **completeness**
- Fast running speed at ~5s / view

MVSNet: Depth Inference for Unstructured Multi-view Stereo. **Yao et al**. ECCV2018 oral

MVSNet: Results on Tanks and Temples Dataset

(a) Family

(b) Francis

(c) Train

(d) Lighthouse

- Trained on DTU **without any fine-tuning**
- Rank 3rd until April 18, 2018

MVSNet: Depth Inference for Unstructured Multi-view Stereo. **Yao et al**. ECCV2018 oral

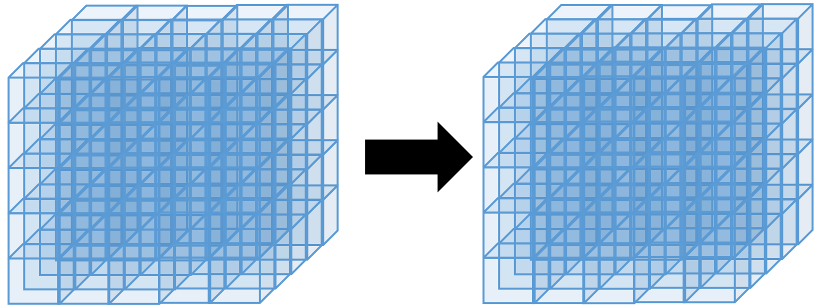
R-MVSNet: Large-scale Reconstruction

Problem of MVSNet:

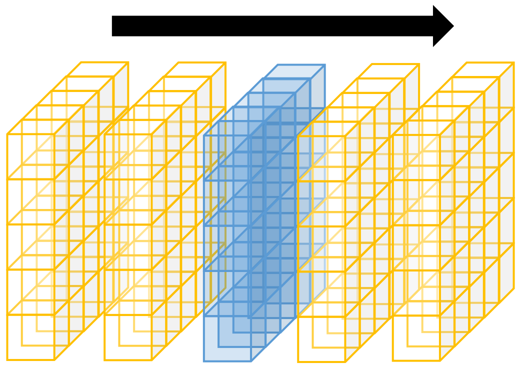
- 3D CNNs is memory consuming
- Cannot handle high-res & large-scale scenes

Solution:

- Sequential regularization



3D CNNs Regularization



Sequential Regularization

Recurrent MVSNet for High-resolution Multi-view Stereo Depth Inference. **Yao et al**. CVPR2019

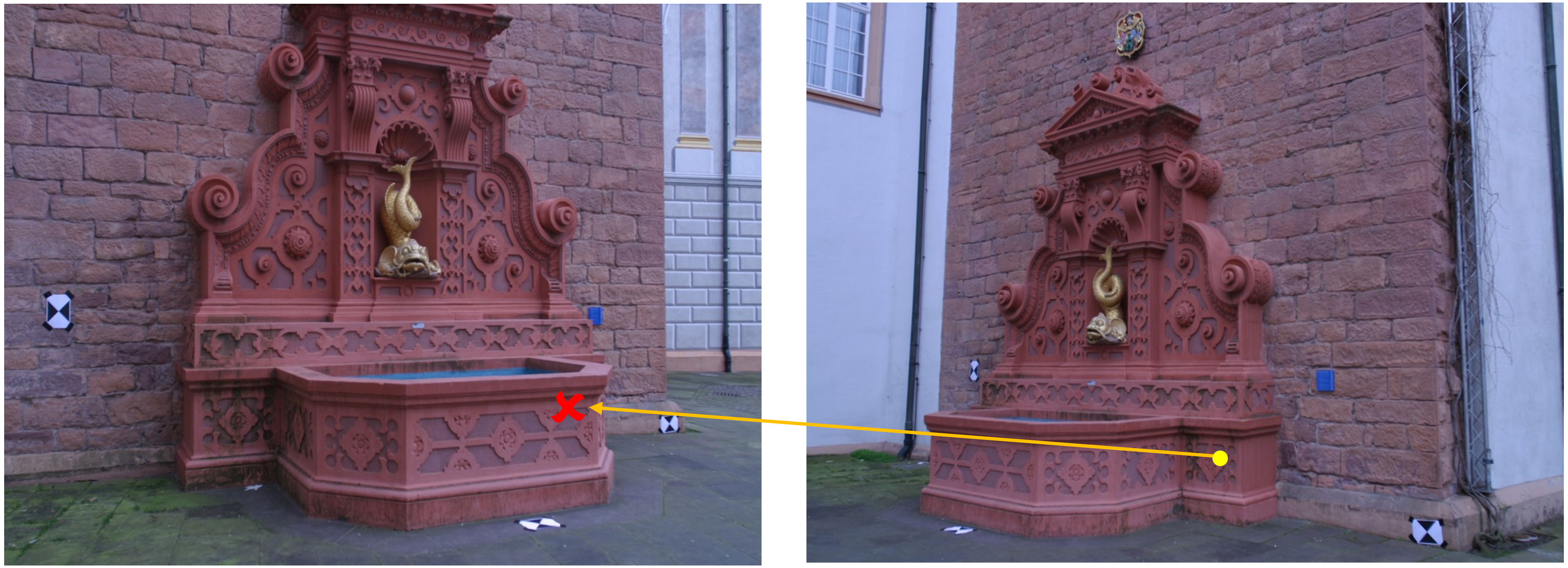
Vis-MVSNet: Visibility Handling

Matchability and Visibility:

- Is the pixel matchable and visible in all views?

Solution:

- Explicit matchability and visibility modeling



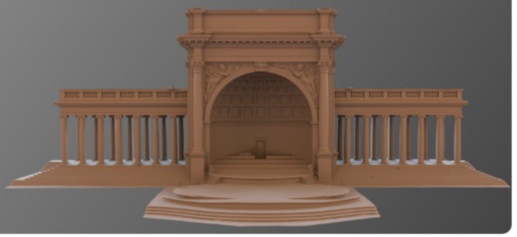
MVSNet: Other Extensions

- **Point-based method:** PointMVSNet (ICCV19), Vis-PointMVSNet (PAMI20)
- **Recurrent regularization:** R-MVSNet (CVPR19), D2HC-RMVSNet (ECCV20), AA-RMVSNet (ICCV21)
- **Coarse-to-fine**: CasMVSNet (CVPR20), UCSNet (CVPR20), CVP-MVSNet (CVPR20)
- **Visibility handling**: Vis-MVSNet (BMVC20), PVA-MVSNet (ECCV20), PVSNet (Arxiv20)
- **Un/self-supervision**: MVS^2 (3DV19), M^3VSNet (Arxiv20), JDACS (AAAI21)
- **Attention**: AA-MVSNet (CVPR20), MVS2D (CVPR22), TransMVSNet (CVPR22), MVSTER (ECCV22)
- **Satellite Image:** Sat-MVS (ICCV21)
- **Novel-view Synthesis**: MVSNeRF (CVPR21)
- **Others**: …

Multi-view 3D Reconstruction: Where are We Now?

Aug. 2023: **all top-ranking** methods are based on MVSNet

Leaderboard

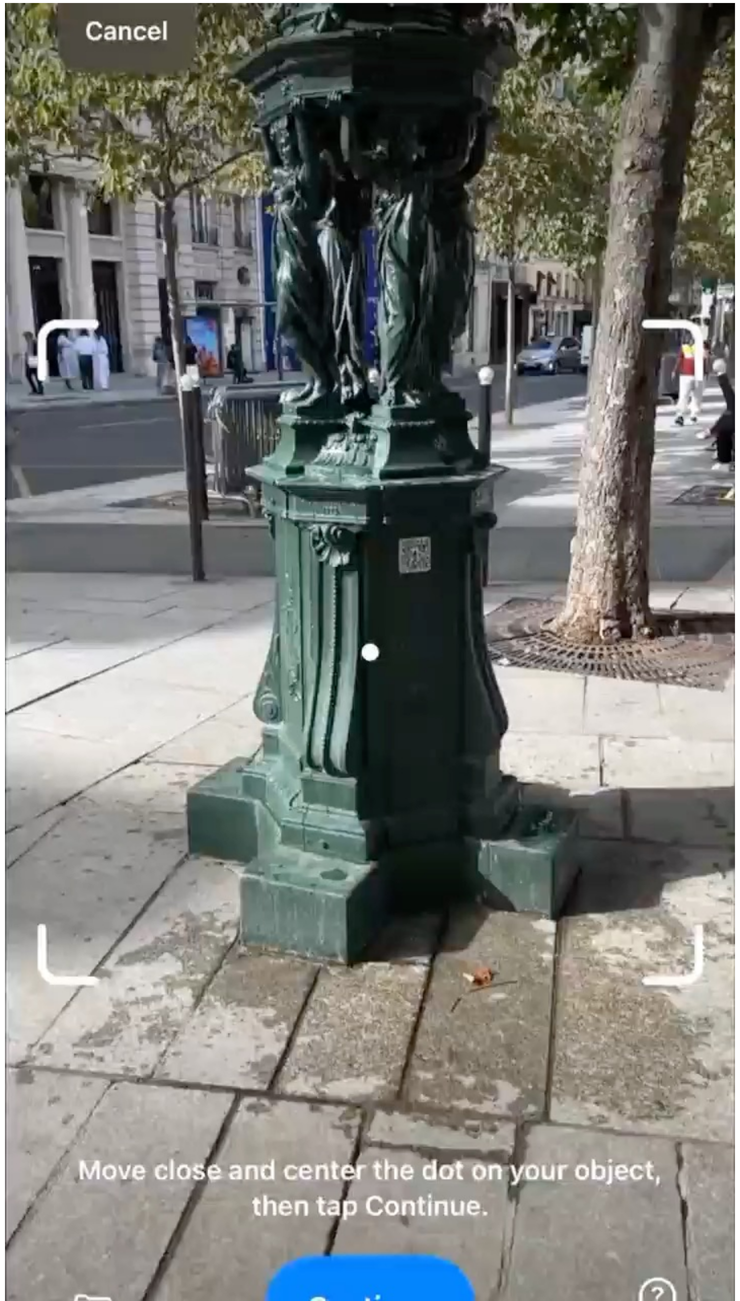


Intermediate ▾ Advanced ▾

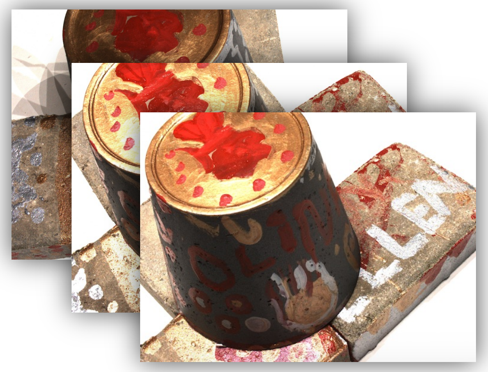
Intermediate F-score

method	rank	mean	runtime*	Family	Francis	Horse	Lighthouse	M60	Panther	Playground	Train
MVSFormer-H-Intermediate	**8.62**	**66.41**	N.A.	82.06	**69.33**	60.59	68.54	65.67	64.07	61.45	59.54
GeoMVSNet	10.50	65.89	N.A.	81.64	67.53	55.78	68.02	65.49	67.19	63.27	58.22
RA-MVSNet	11.12	65.72	N.A.	**82.44**	66.61	58.40	64.78	**67.14**	65.60	62.74	58.08
WT-MVSNet	12.25	65.34	N.A.	81.87	67.33	57.76	64.77	65.68	64.61	62.35	58.38
ET-MVSNet	13.62	65.49	N.A.	81.65	68.79	59.46	65.72	64.22	64.03	61.23	58.79
EI_MVSNet	17.00	65.52	N.A.	81.59	67.67	61.67	63.18	65.10	63.42	60.62	**60.95**
UniMVSNet_NR-MVSNet	18.62	65.17	N.A.	80.98	65.09	54.94	68.71	64.08	64.67	65.61	57.29
TF-MVSNet	20.12	64.51	N.A.	81.45	67.17	57.01	62.64	65.34	62.20	61.44	58.85
PFR-MVSNet	20.88	64.56	N.A.	81.57	65.50	54.43	63.37	66.44	65.65	62.31	57.23
UniMVSNet-intermediate	21.88	64.36	N.A.	81.20	66.43	53.11	63.46	66.09	64.84	62.23	57.53

Multi-view 3D Reconstruction: Where are We Now?



Introduction: Reconstruction and Rendering



Input Images

Reconstruction (Vision)

Geometry: Point cloud / Mesh / SDF …

Appearance: Texture map / BRDF / …

Lighting: Environment map / NeILF / …

Rendering (Graphics)

Rendered Images

Introduction: Differentiable Rendering

Bridging vision & graphics!

Input Images

Reconstruction (Vision)

Geometry: Point cloud / Mesh / SDF …

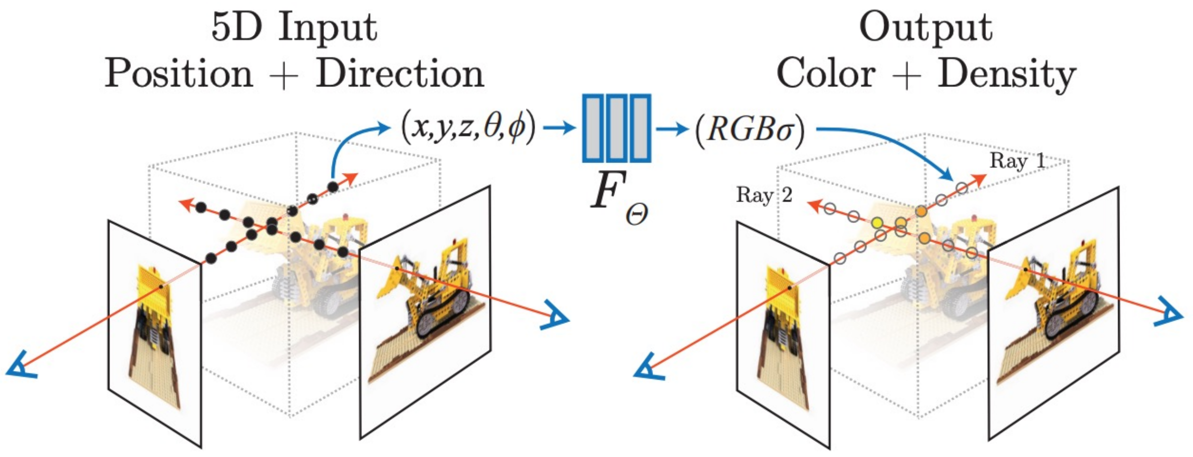
Appearance: Texture map / BRDF / …

Lighting: Environment map / NeILF / …

Rendering (Graphics)

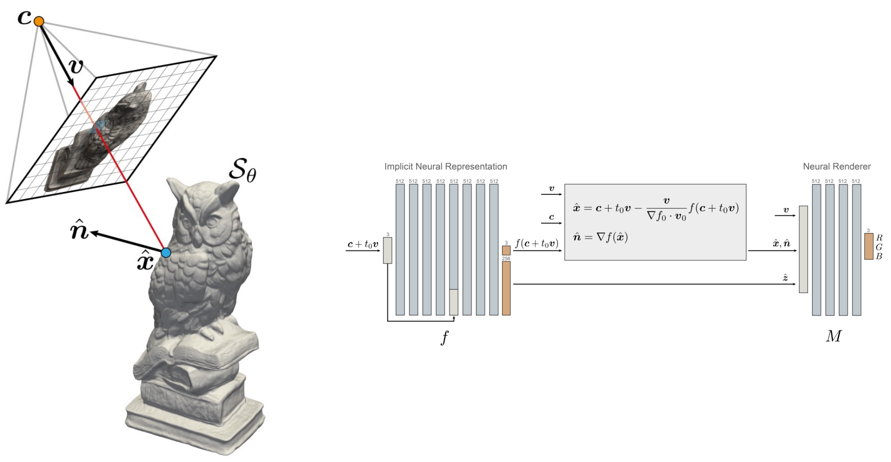
Rendered Images

Introduction: Differentiable Volume & Surface Rendering



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

Volume Rendering (e.g., NeRF)



$$L_o^P(\boldsymbol{\omega_o}, \mathbf{x}) = \int_{\Omega} f(\boldsymbol{\omega_o}, \boldsymbol{\omega_i}, \mathbf{x}) L_i(\boldsymbol{\omega_i}, \mathbf{x})(\boldsymbol{\omega}_i \cdot \mathbf{n}) d\boldsymbol{\omega}_i$$

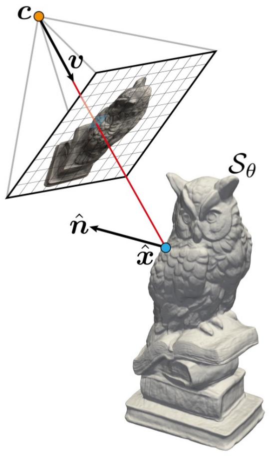
Surface Rendering (e.g., IDR)

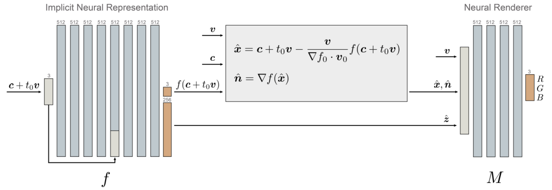
NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis
IDR: Multi-view Neural Surface Reconstruction by Disentangling Geometry and Appearance

Introduction: Differentiable Volume & Surface Rendering

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

Volume Rendering (e.g., NeRF)





$$L_o^P(\boldsymbol{\omega_o}, \mathbf{x}) = \int_\Omega f(\boldsymbol{\omega_o}, \boldsymbol{\omega_i}, \mathbf{x}) L_i(\boldsymbol{\omega_i}, \mathbf{x})(\boldsymbol{\omega}_i \cdot \mathbf{n}) d\boldsymbol{\omega}_i$$

Surface Rendering (e.g., IDR)

NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis

IDR: Multi-view Neural Surface Reconstruction by Disentangling Geometry and Appearance

MVSDF: Motivation

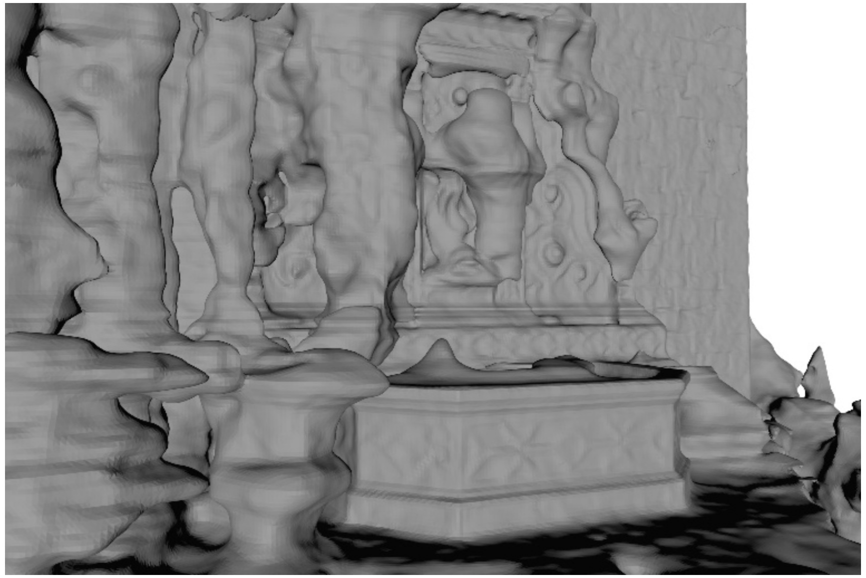
Problem of Differentiable Rendering:

- Geometry-appearance ambiguity (IDR)

Solution:

- Geometric priors
- Regularizations

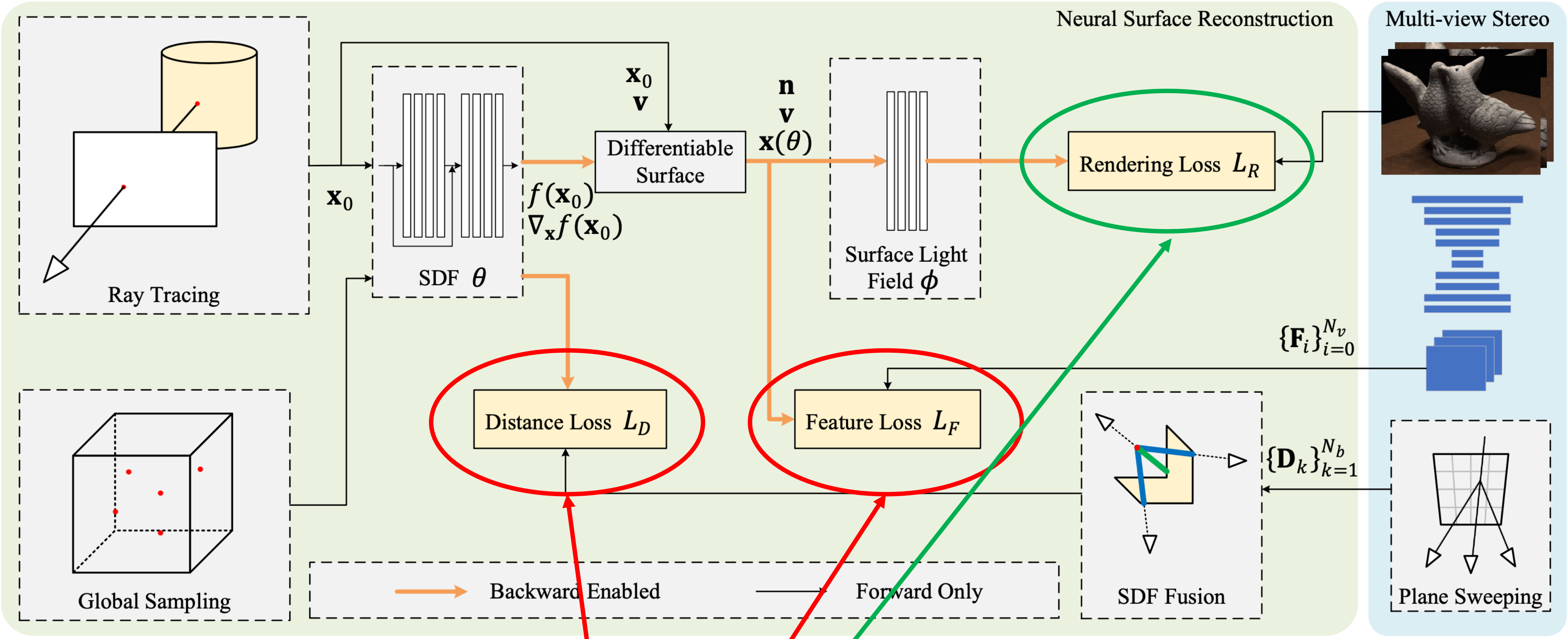
✓ Qualified Rendering



× Erroneous Geometry

IDR: Multi-view Neural Surface Reconstruction by Disentangling Geometry and Appearance

MVSDF: Geometric Priors from MVS



- Joint geometry and appearance optimization
- **Geometry initialization from Vis-MVSNet**
- Multi-view feature consistency loss

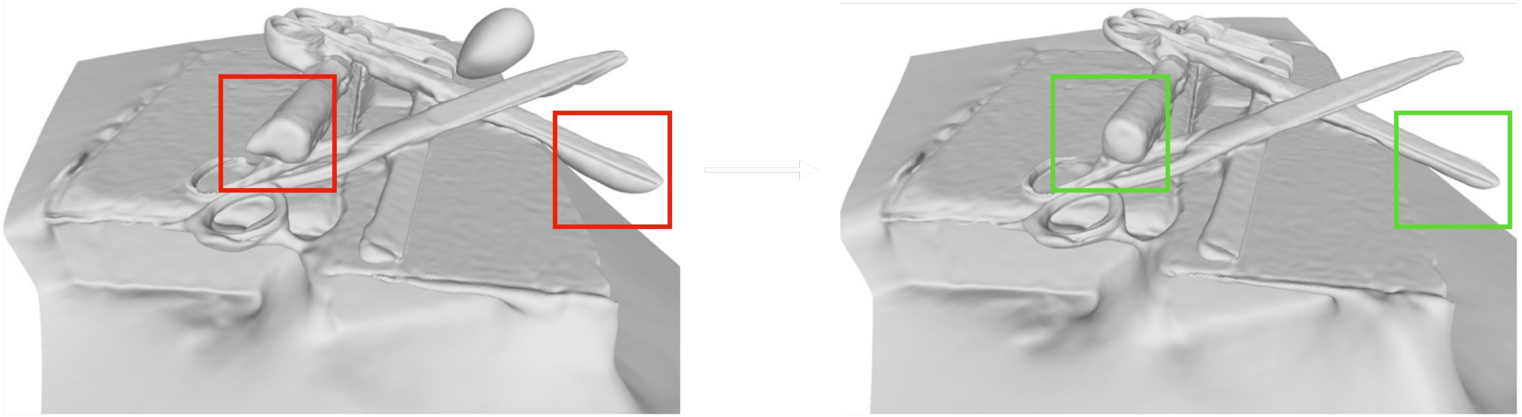
MVSDF: Learning Signed Distance Field for Multi-view Surface Reconstruction. **Zhang et al.** ICCV2021 oral

MVSDF: Geometric Priors from MVS

IDR | IDR with masks | **MVSDF (ours)**

MVSDF: Learning Signed Distance Field for Multi-view Surface Reconstruction. **Zhang et al.** ICCV2021 oral

RegSDF: Regularizations



- **Minimum Surface Constraint:**

  Minimize the surface area:

$$\begin{aligned}\text{area}(\mathcal{S}) &= \int_\Omega \|\nabla H(f(\mathbf{x};\theta))\| d\mathbf{x} \\ &= \int_\Omega \delta(f(\mathbf{x};\theta))\|\nabla f(\mathbf{x};\theta)\| d\mathbf{x},\end{aligned}$$

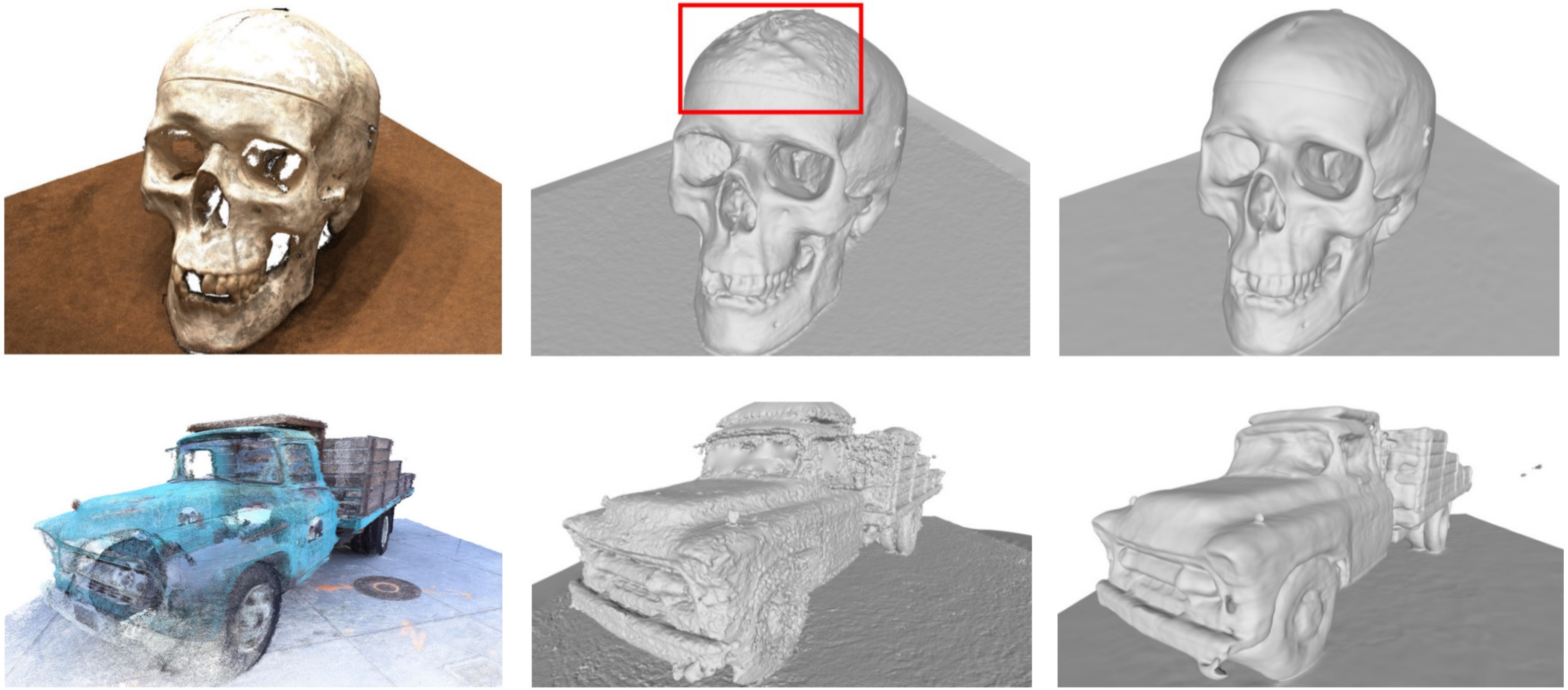
$$L_M = \frac{1}{|\mathcal{R}|}\sum_{\mathbf{x}\in\mathcal{R}} \delta_\epsilon(f(\mathbf{x})), \text{ where } \delta_\epsilon(z) = \frac{\epsilon\pi^{-1}}{\epsilon^2 + z^2}.$$

- **Hessian Smoothness:**

$$L_H = \frac{1}{|\mathcal{R}|}\sum_{\mathbf{x}\in\mathcal{R}} \|\mathbf{H}f(\mathbf{x})\|_1,$$

RegSDF: Critical Regularizations for Neural Surface Reconstruction in the Wild. **Zhang et al.** CVPR2022

RegSDF: Results



Vis-MVSNet Point Cloud

Poisson Reconstruction

RegSDF (ours)

RegSDF: Critical Regularizations for Neural Surface Reconstruction in the Wild. **Zhang et al.** CVPR2022

Introduction: Reconstruction and Rendering

Input Images

Reconstruction
(Vision)

Geometry: Point cloud / Mesh / SDF …

Appearance: Texture map / BRDF / …

Lighting: Environment map / NeILF / …

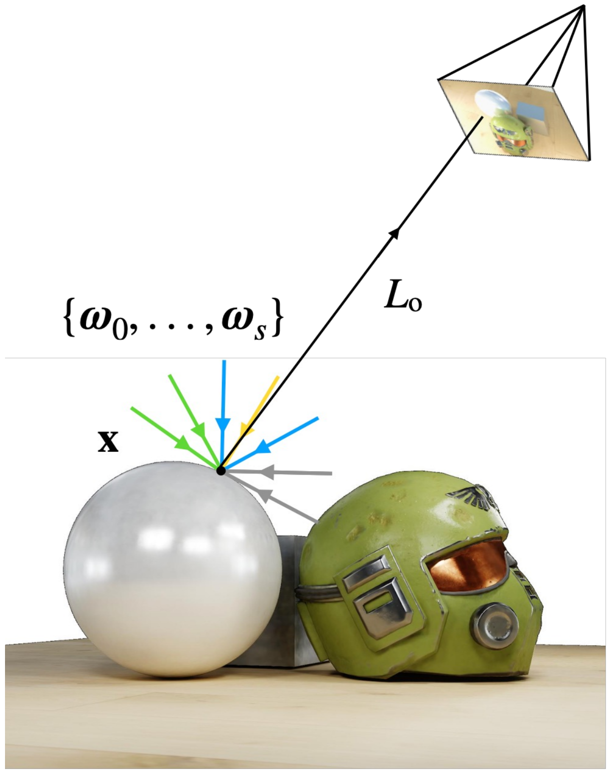
Rendering
(Graphics)

Generalized 3D Reconstruction:

- Geometry
- Surface materials
- Scene Lightings
- Other physical signals…

Rendered Images

NeILF: Neural Incident Light Field for Material Estimation



$$L(\boldsymbol{\omega_o}, \mathbf{x}) = \int_\Omega f(\boldsymbol{\omega_o}, \boldsymbol{\omega_i}, \mathbf{x}) L_i(\boldsymbol{\omega}_i, \mathbf{x})(\boldsymbol{\omega}_i \cdot \mathbf{n}) d\boldsymbol{\omega}_i$$

Observed pixel color

Surface Material:

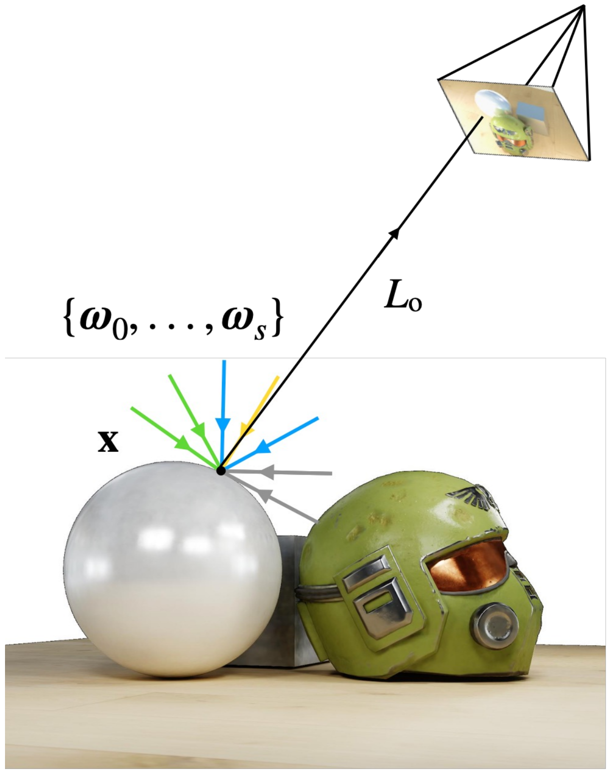
- BRDF (base color, metallic, roughness …)

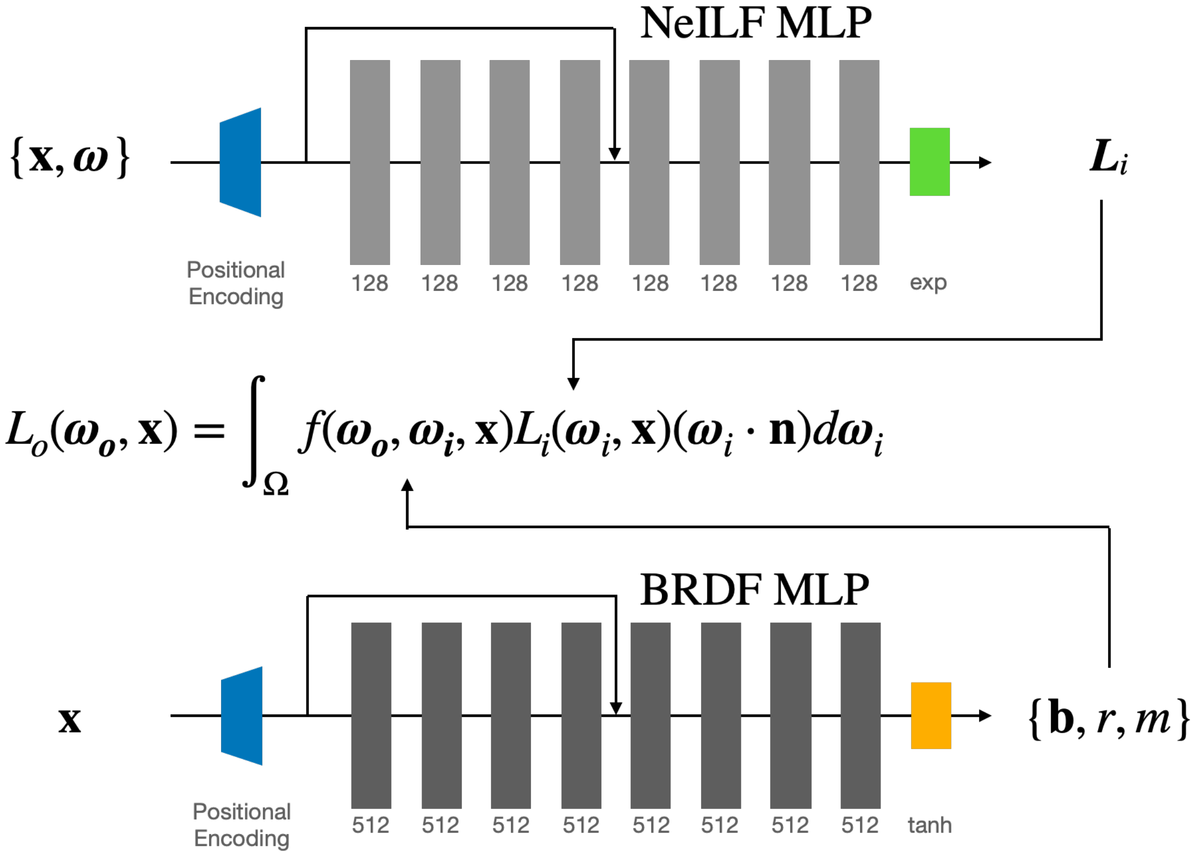
Incident Light:

- Env Map Approximation?
- Neural incident light field

NeILF: Neural Incident Light Field for Material and Lighting Estimation. **Yao et al.** ECCV2022

NeILF: Network Architecture

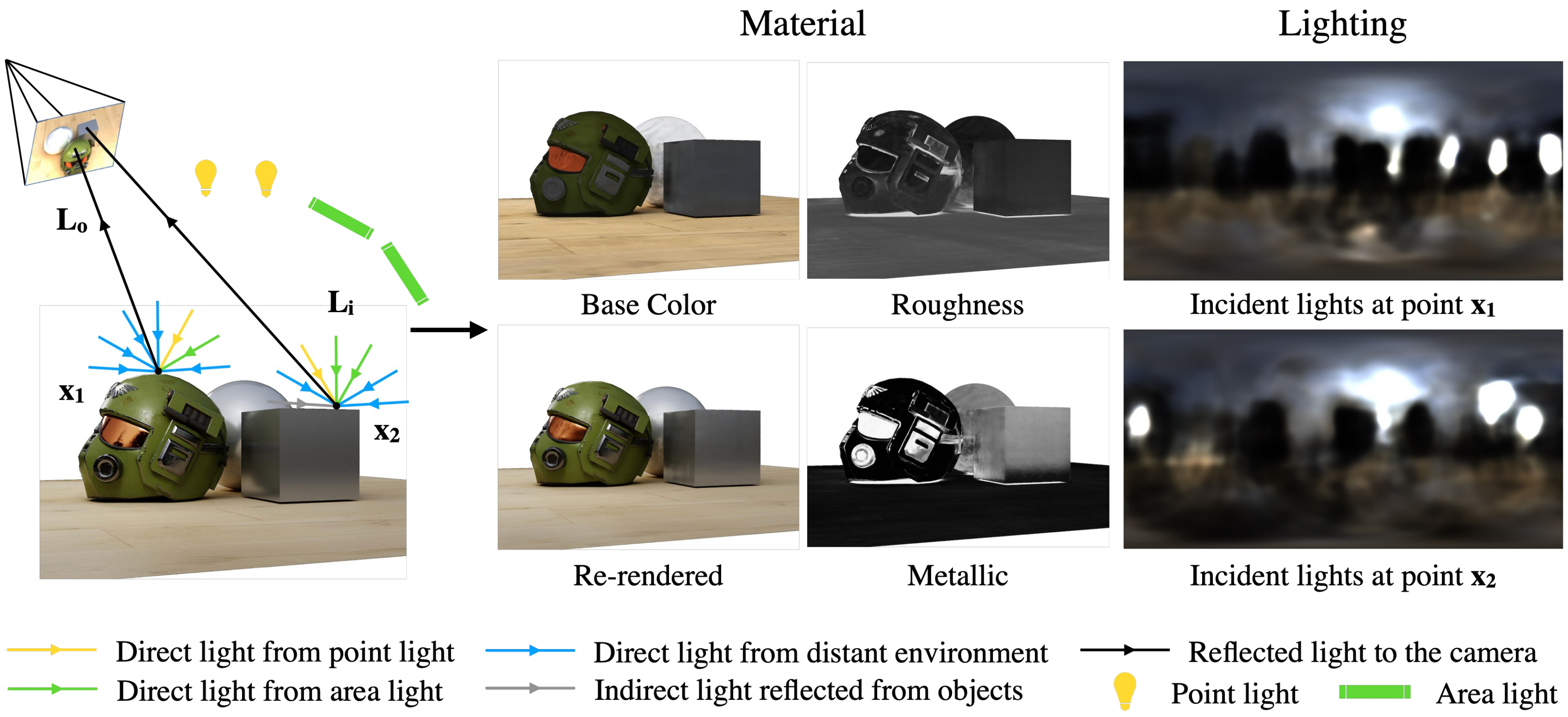




$$L_o(\boldsymbol{\omega_o}, \mathbf{x}) = \int_{\Omega} f(\boldsymbol{\omega_o}, \boldsymbol{\omega_i}, \mathbf{x}) L_i(\boldsymbol{\omega_i}, \mathbf{x})(\boldsymbol{\omega_i} \cdot \mathbf{n}) d\boldsymbol{\omega_i}$$

NeILF: Neural Incident Light Field for Material and Lighting Estimation. **Yao et al.** ECCV2022

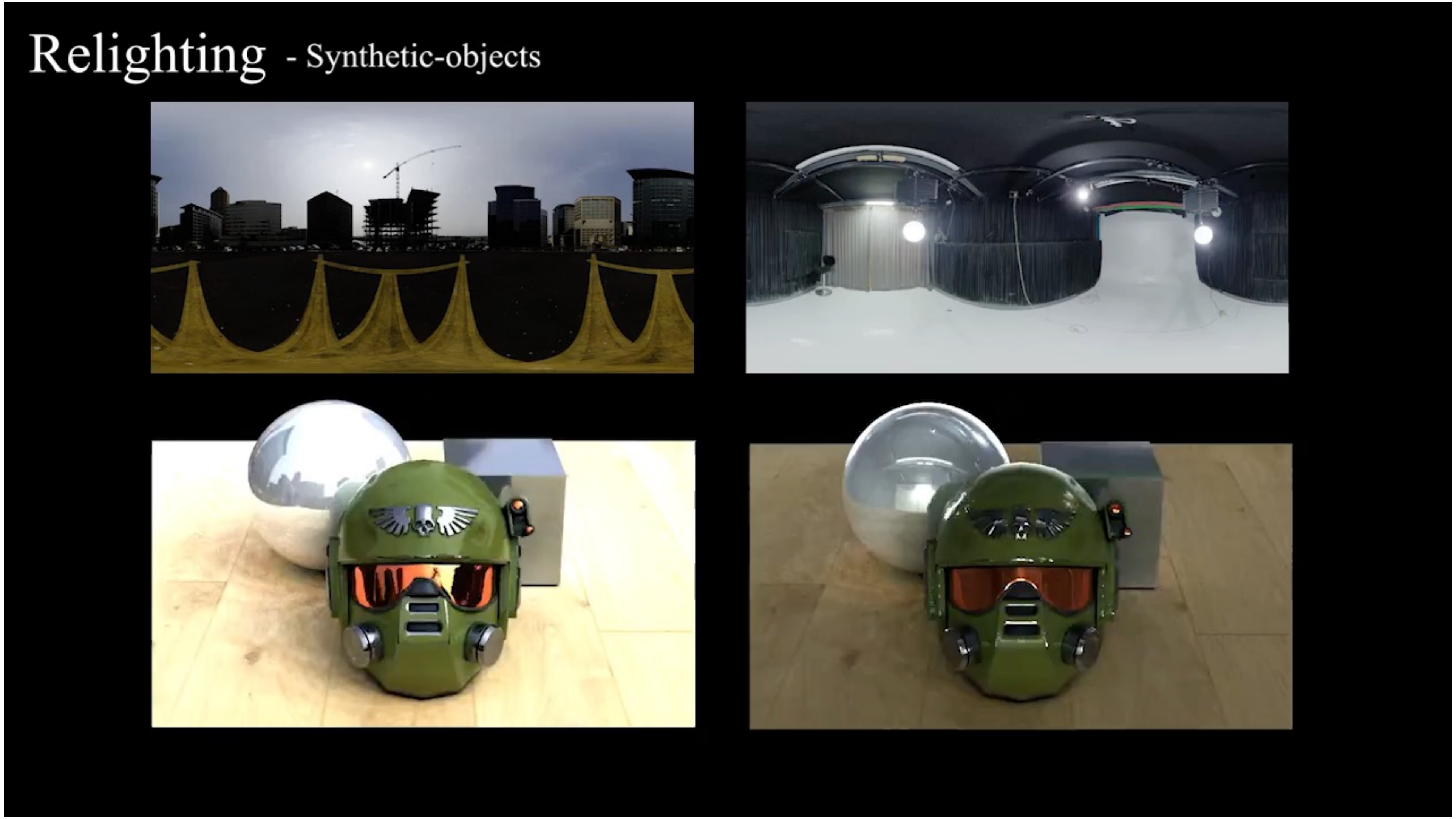
NeILF: Joint Estimation of Material and Lighting



- Fully 5D light field
- Joint illumination of direct/indirect lights of **any static scenes**
- Without the need of tracing rays of multiple bounces

NeILF: Neural Incident Light Field for Material and Lighting Estimation. **Yao et al.** ECCV2022

NeILF: Relighting Results



NeILF: Neural Incident Light Field for Material and Lighting Estimation. **Yao et al.** ECCV2022

NeILF: from a NeRF Perspective

NeRF:

- Volume Rendering
- Joint geometry and lighting estimation
- 5D out-going radiance field
- Spatially-varying geometry property (density/SDF)
- Sampling along viewing ray

NeILF:

- Surface Rendering
- Joint material and lighting estimation
- 5D incident light field
- Spatially-varying surface property (SV-BRDF)
- Sampling over surface hemi-sphere

NeILF: Neural Incident Light Field for Material and Lighting Estimation. **Yao et al.** ECCV2022

NeILF: from a NeRF Perspective

NeRF:

- Volume Rendering
- Joint geometry and lighting estimation
- 5D out-going radiance field
- Spatially-varying geometry property (density/SDF)
- Sampling along viewing ray

NeILF:

- Surface Rendering
- Joint material and lighting estimation
- 5D incident light field
- Spatially-varying surface property (SV-BRDF)
- Sampling over surface hemi-sphere

NeILF: Neural Incident Light Field for Material and Lighting Estimation. **Yao et al.** ECCV2022

NeILF: from a NeRF Perspective

NeRF:

- Volume Rendering
- Joint geometry and lighting estimation
- 5D out-going radiance field
- Spatially-varying geometry property (density/SDF)
- Sampling along viewing ray

NeILF:

- Surface Rendering
- Joint material and lighting estimation
- 5D incident light field
- Spatially-varying surface property (SV-BRDF)
- Sampling over surface hemi-sphere

NeILF: Neural Incident Light Field for Material and Lighting Estimation. **Yao et al.** ECCV2022

NeILF: from a NeRF Perspective

NeRF:

- Volume Rendering
- Joint geometry and lighting estimation
- 5D out-going radiance field
- Spatially-varying geometry property (density/SDF)
- Sampling along viewing ray

NeILF:

- Surface Rendering
- Joint material and lighting estimation
- 5D incident light field
- Spatially-varying surface property (SV-BRDF)
- Sampling over surface hemi-sphere

NeILF: Neural Incident Light Field for Material and Lighting Estimation. **Yao et al.** ECCV2022

NeILF: from a NeRF Perspective

NeRF:

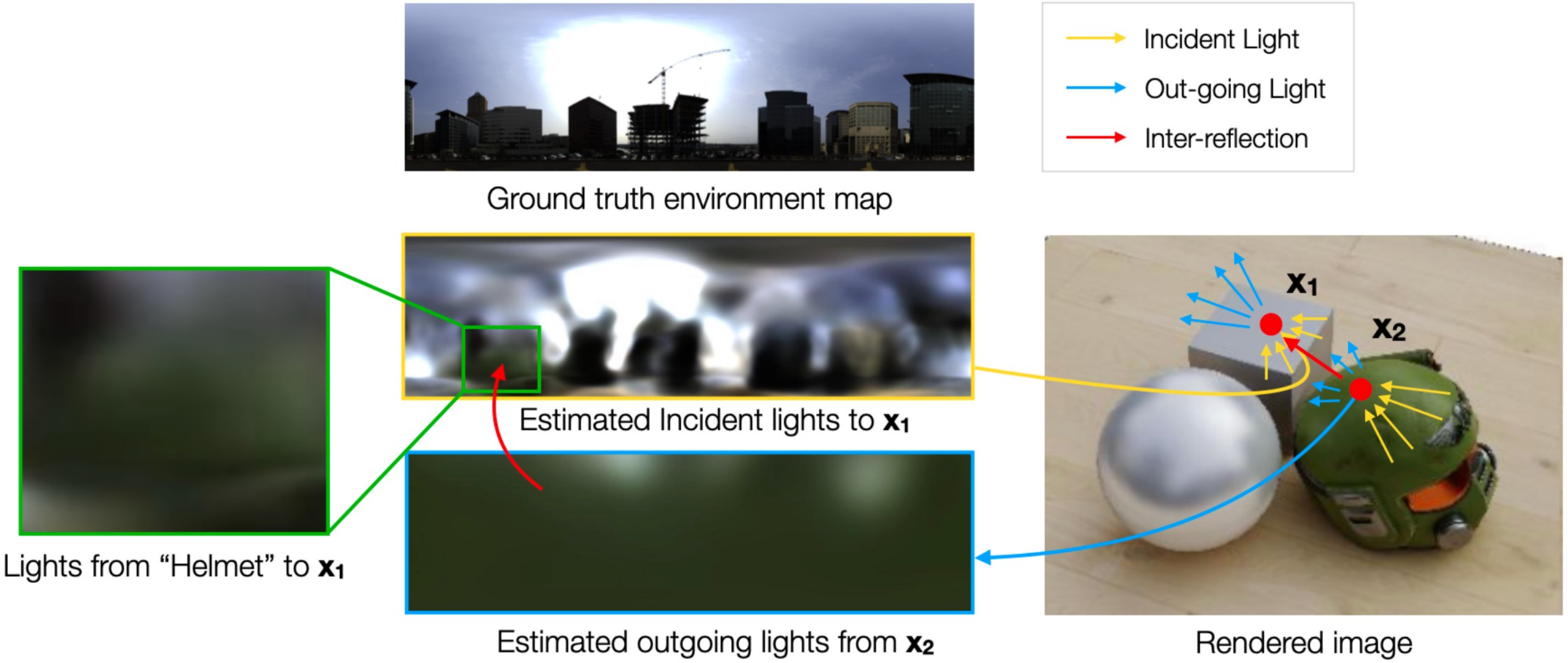
- Volume Rendering
- Joint geometry and lighting estimation
- 5D out-going radiance field
- Spatially-varying geometry property (density/SDF)
- Sampling along viewing ray

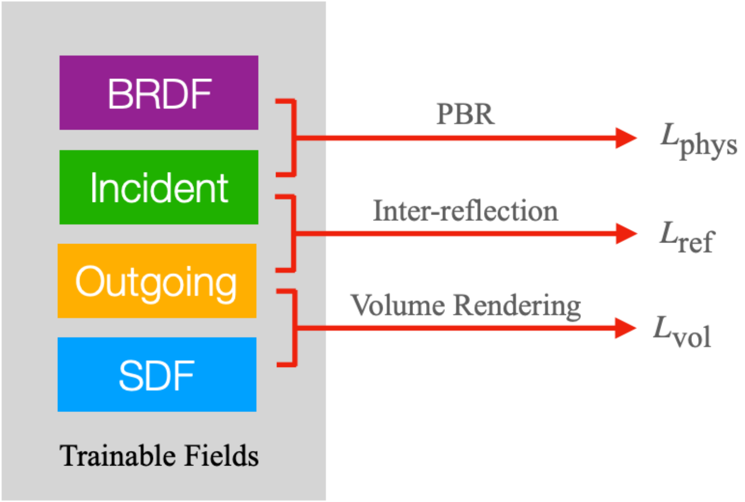
NeILF:

- Surface Rendering
- Joint material and lighting estimation
- 5D incident light field
- Spatially-varying surface property (SV-BRDF)
- Sampling over surface hemi-sphere

NeILF: Neural Incident Light Field for Material and Lighting Estimation. **Yao et al.** ECCV2022

NeILF++: Joint Geometry, Material, and Lighting Estimation

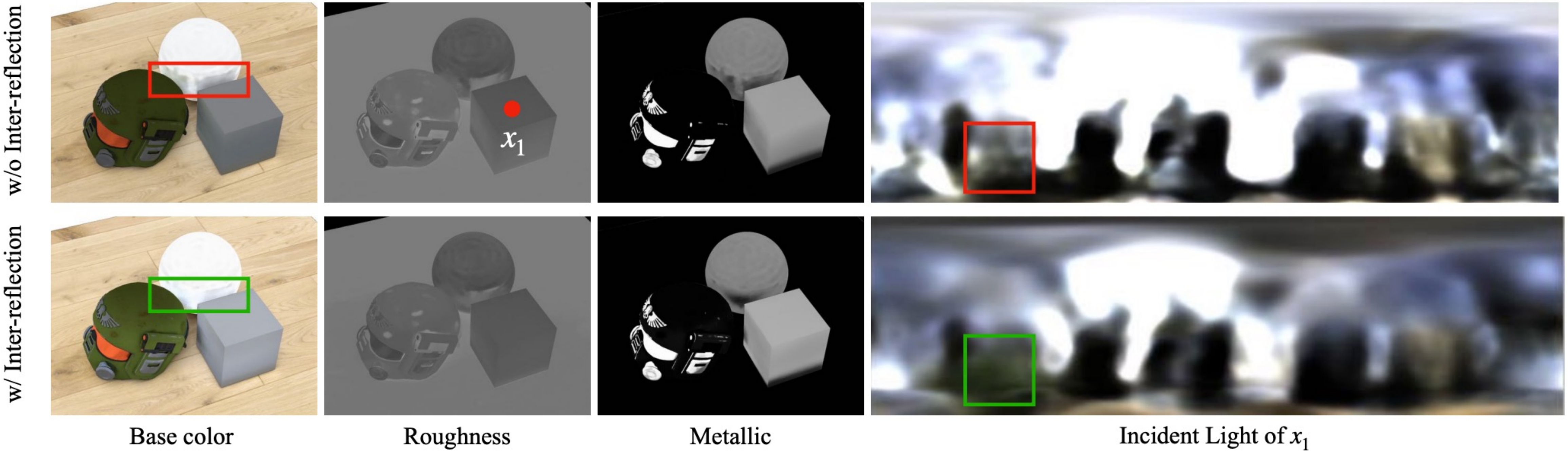




- Light field = outgoing radiance field (NeRF) + incident light field (NeILF)
- Unifying NeRF and NeILF through PBR and inter-reflection
- Joint geometry, material, and lighting estimation.

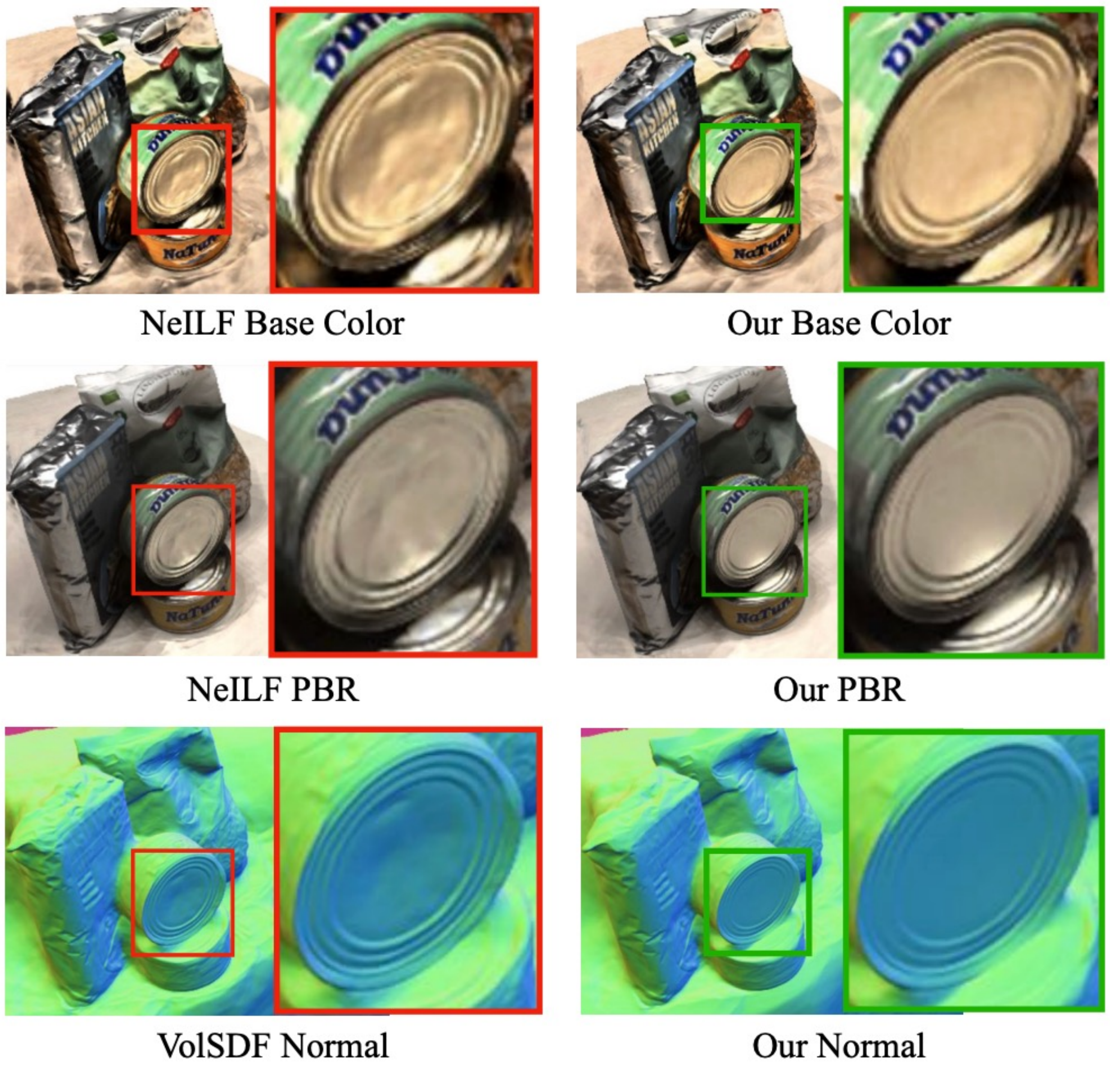
NeILF++: Inter-reflectable Light Field for Geometry and Material Estimation. **Zhang et al.** ICCV 2023.

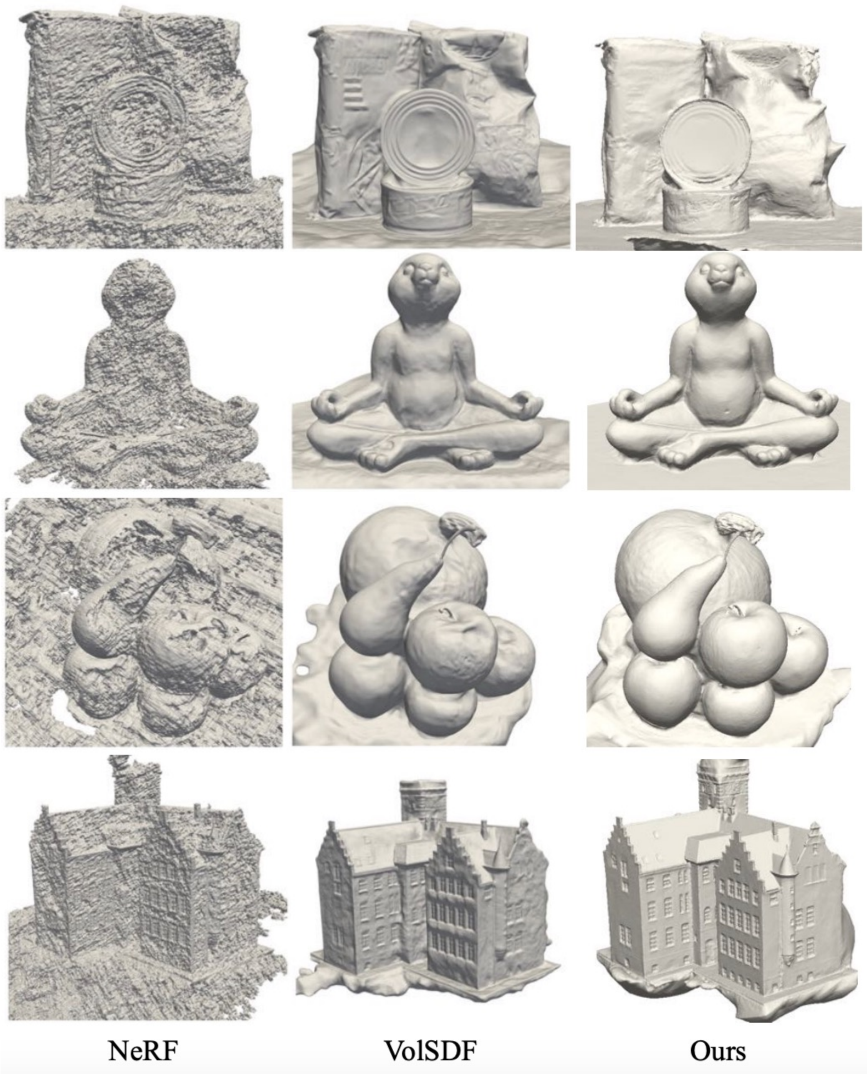
NeILF++: Material & Lighting Estimation



NeILF++: Inter-reflectable Light Field for Geometry and Material Estimation. **Zhang et al.** ICCV 2023.

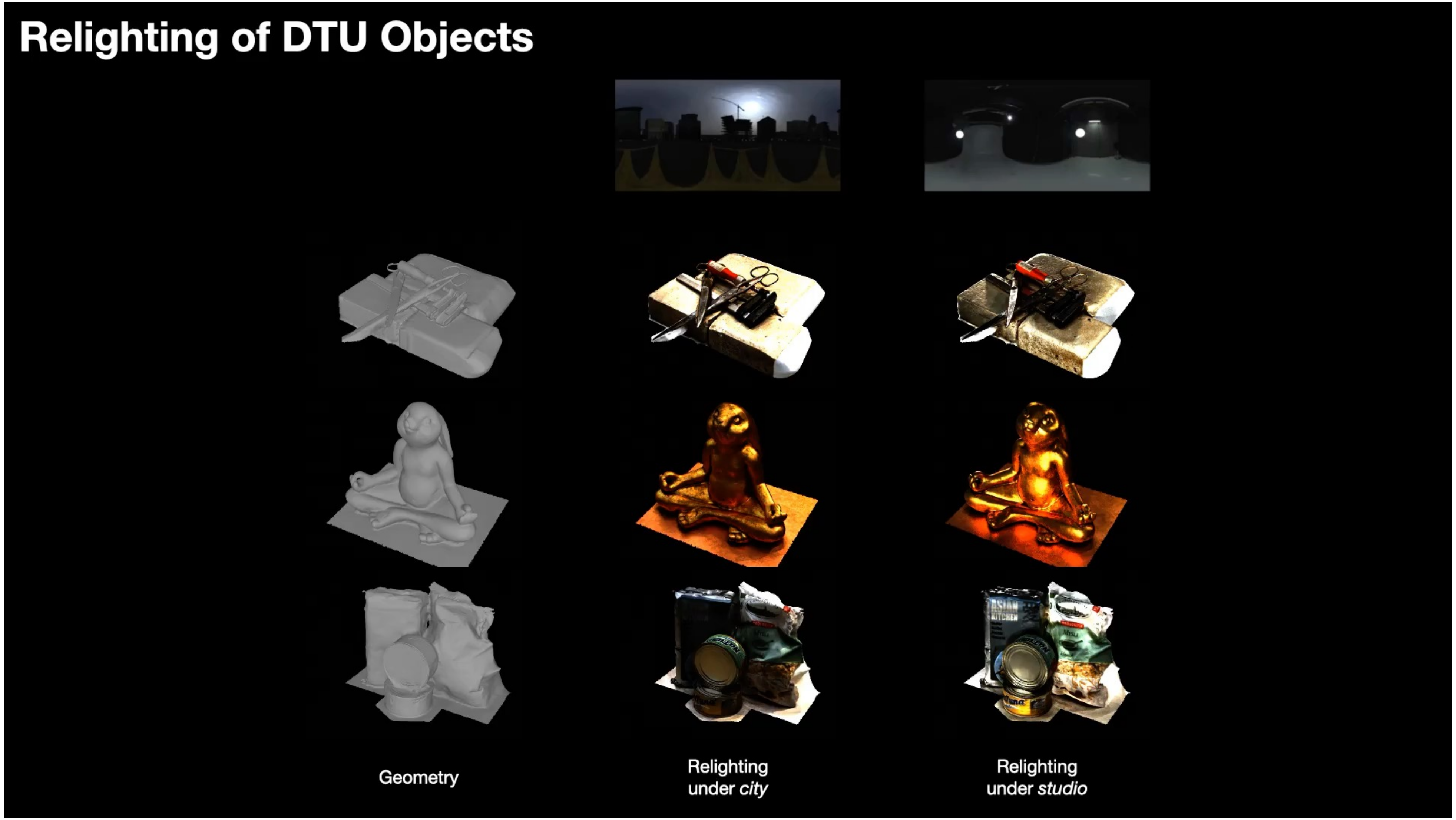
NeILF++: Geometry Refinement





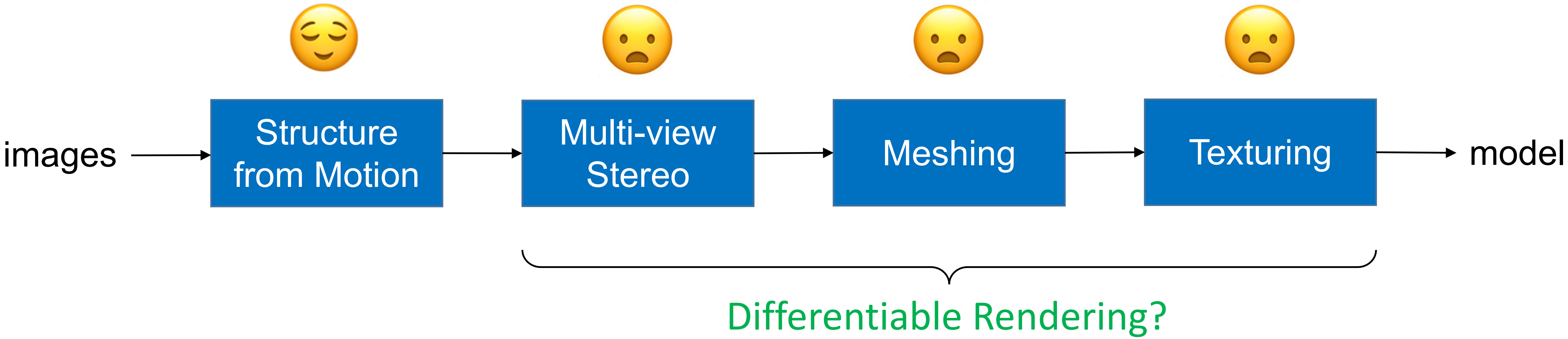
NeILF++: Inter-reflectable Light Field for Geometry and Material Estimation. **Zhang et al.** ICCV 2023.

NeILF++: Joint Geometry, Material, and Lighting Estimation

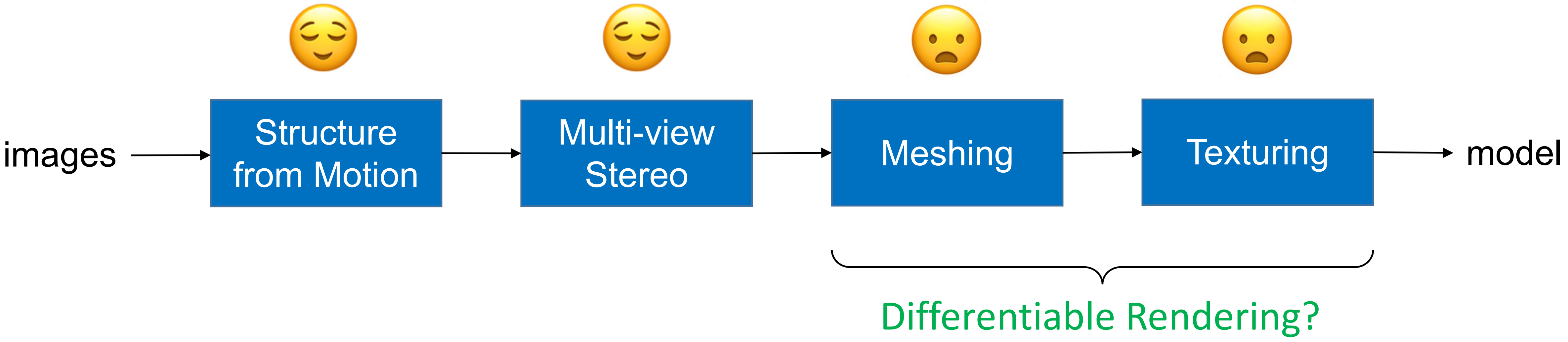


NeILF++: Inter-reflectable Light Field for Geometry and Material Estimation. **Zhang et al.** ICCV 2023.

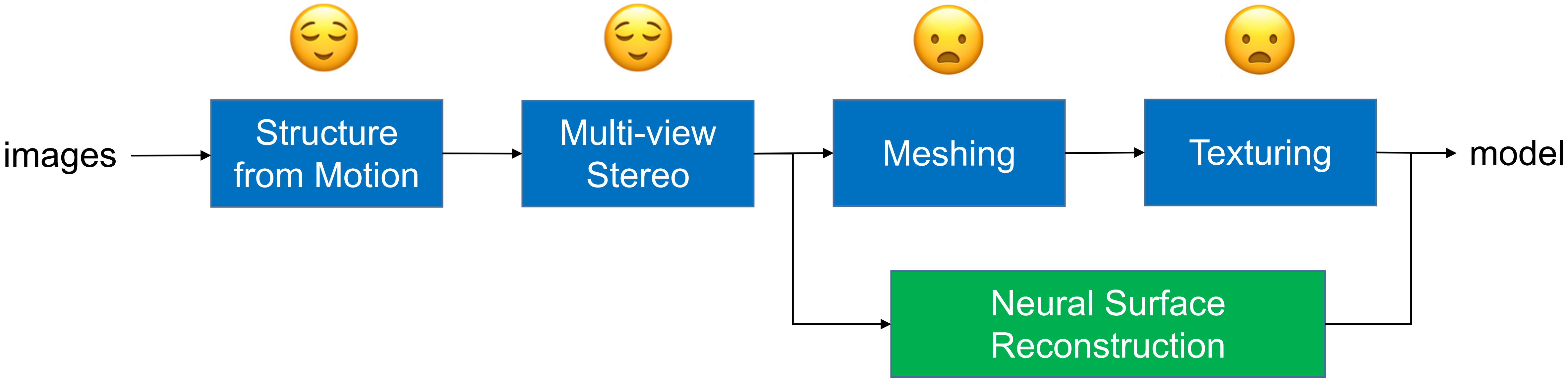
Photogrammetry with Differentiable Rendering



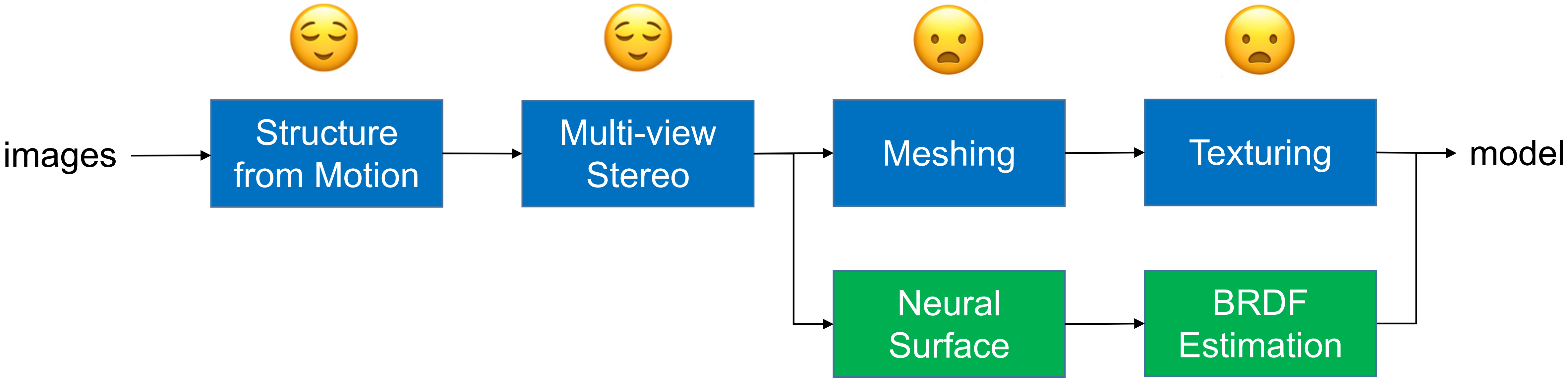
Photogrammetry with Differentiable Rendering



Photogrammetry with Differentiable Rendering

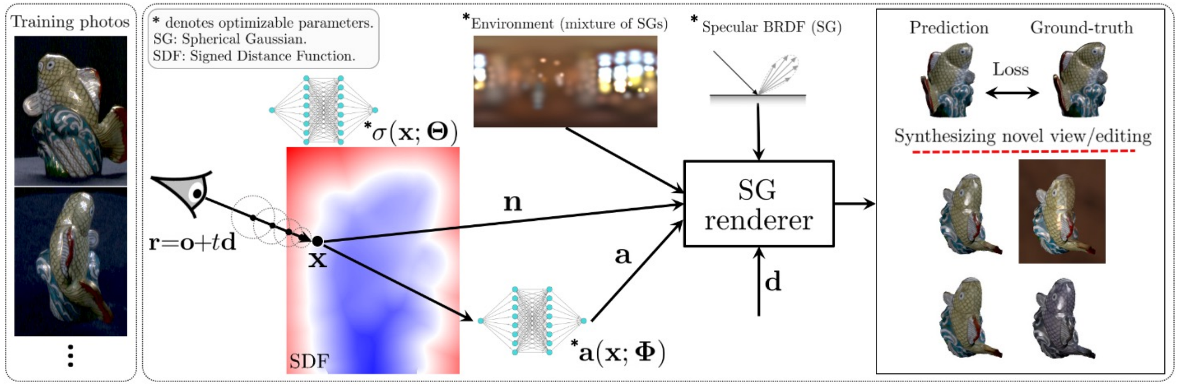


Photogrammetry with Differentiable Rendering

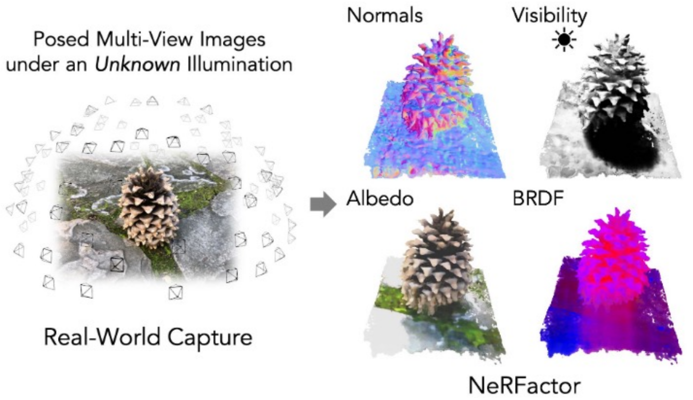


Future Works in the Era of Neural Rendering

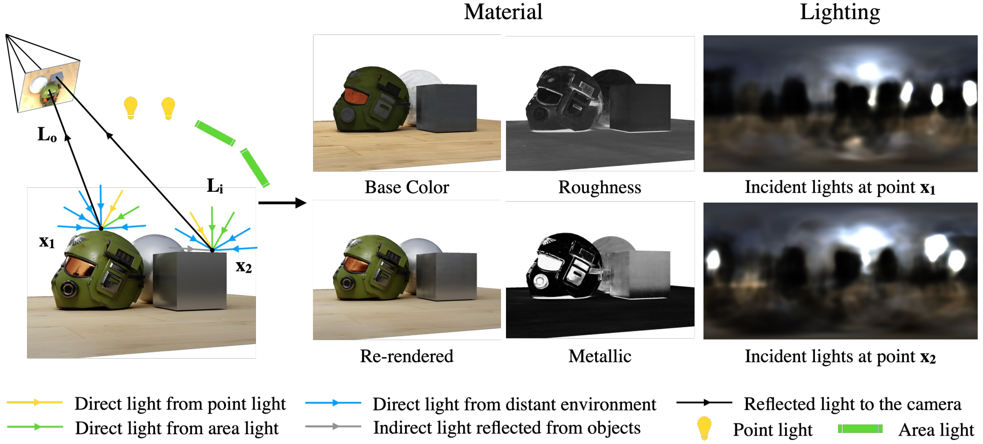
Future Works: Material & Lighting Estimation



PhySG (CVPR 21)



NeRFactor (SIGGRAPH Asia 21)



NeILF (ECCV 22)

NeRO (Siggraph 23)

Future Works: Dynamic 3D Reconstruction

- Plenoptic Function:

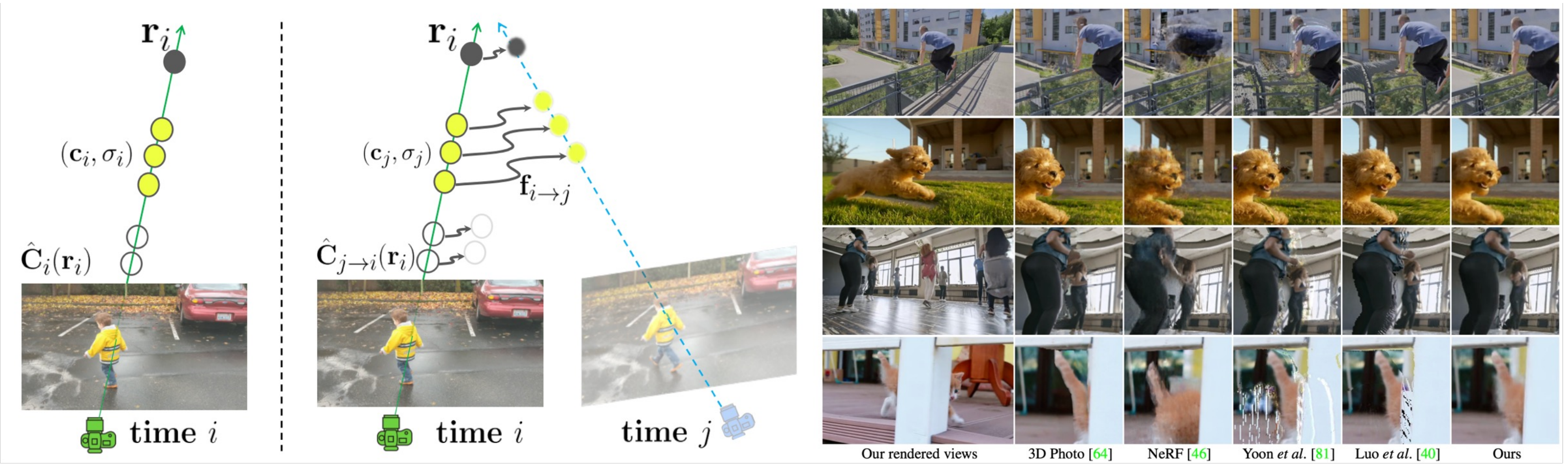
$$f(x, y, z, \theta, \varphi, \lambda, t)$$

- The Light Field:

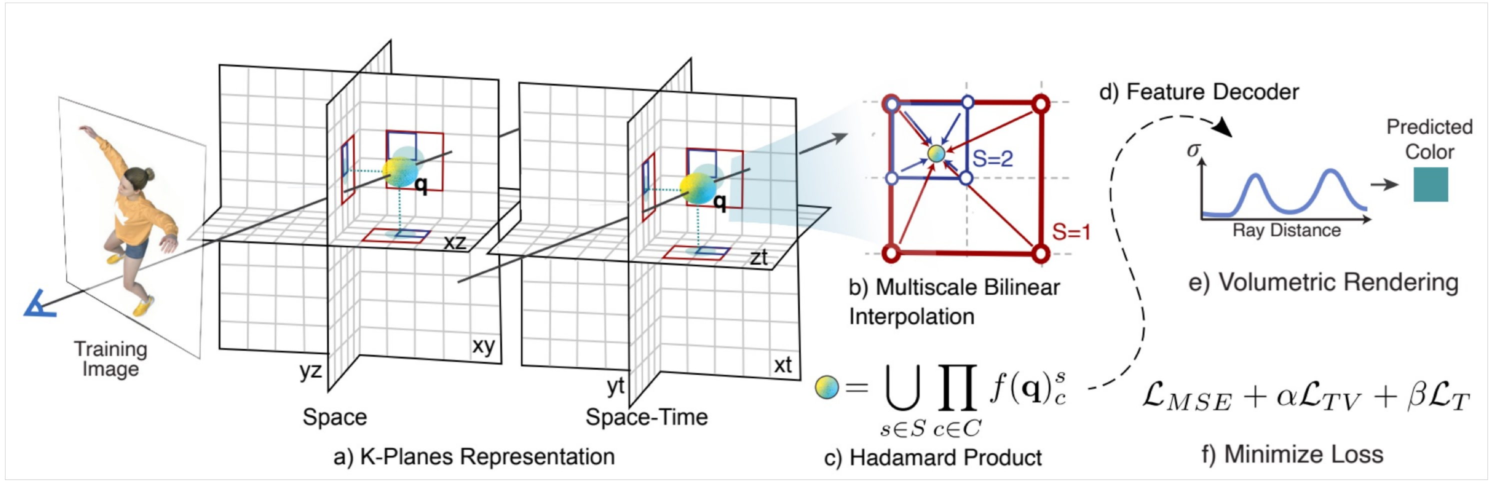
$$f(x, y, z, \theta, \varphi$$ [redacted] $$)$$

- The 4D Light Field:

$$f(x, y, z, \theta, \varphi$$ [redacted] $$t)$$

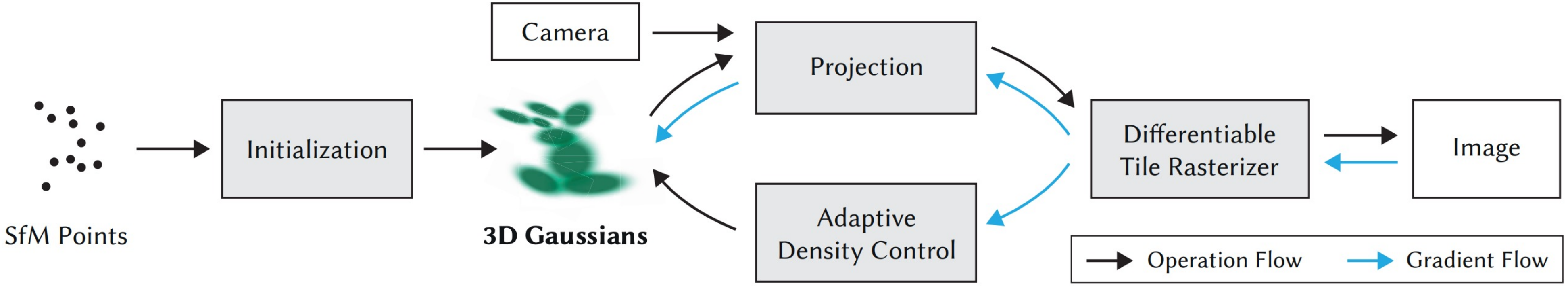


Neural Scene Flow Fields for Space-Time View Synthesis of Dynamic Scenes. CVPR 2021.



K-Planes: Explicit Radiance Fields in Space, Time, and Appearance. arXiv 2023

Future Works: Point-based Differentiable Rendering



Thanks!

Personal homepage: https://yoyo000.github.io/

Email: yaoyao@nju.edu.cn