Multi-View Photometric Stereo Revisited

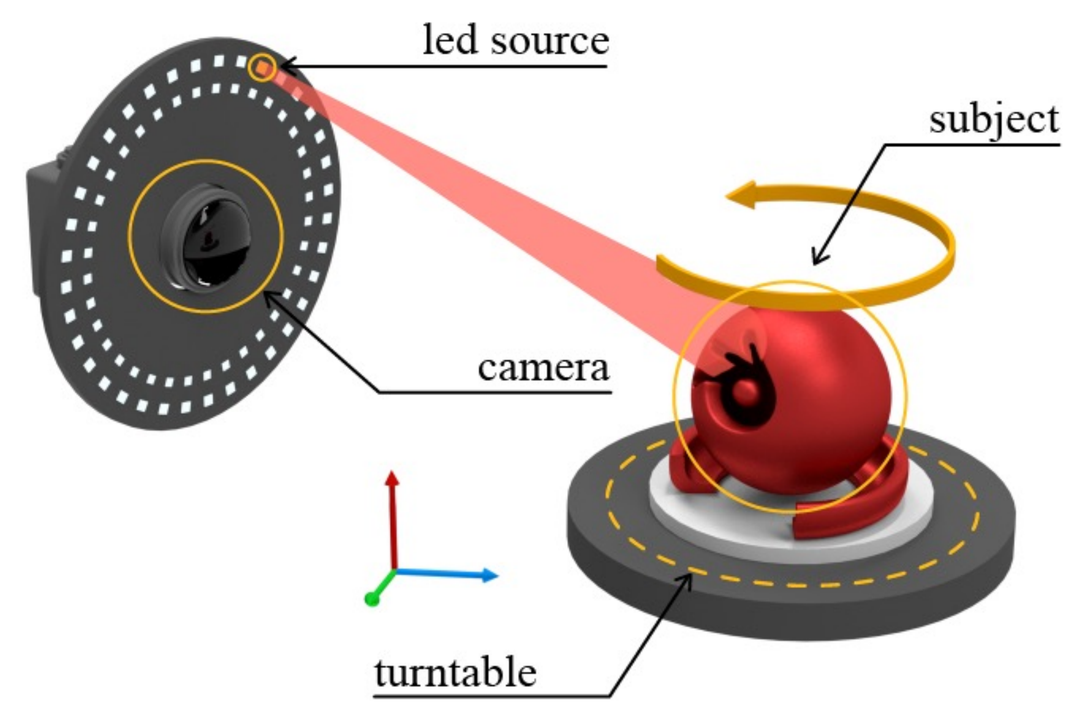
Berk Kaya[1], Suryansh Kumar[1], Carlos Oliveira[1], Vittorio Ferrari[2], Luc Van Gool[1,3]
Computer Vision Lab, ETH Zurich[1], Google Research[2], KU Leuven[3]

Overview

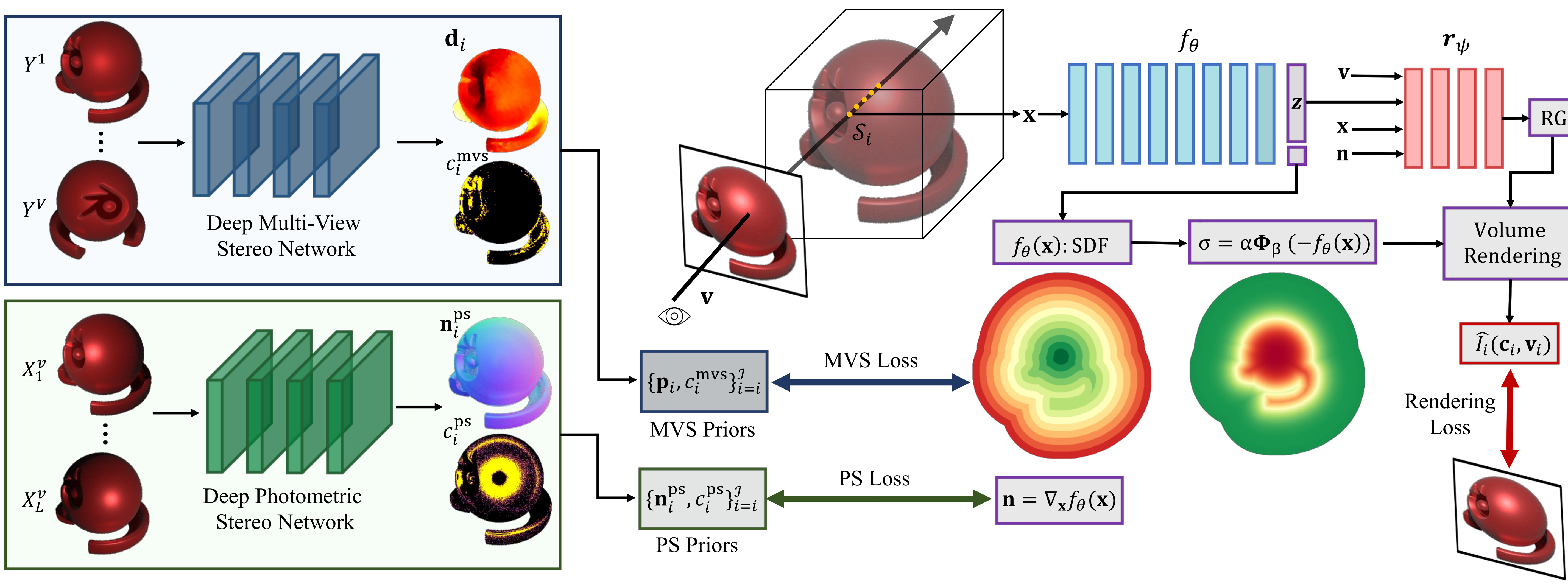
- **Multi-view photometric stereo (MVPS)** aims at recovering accurate and complete 3D reconstruction of an object using multi-view stereo (MVS) and photometric stereo (PS) images.
- Existing methods are **complex to execute** and **limited to isotropic material objects.**

Contributions:

- Simple, efficient, scalable, and effective MVPS method for the detailed and complete recovery of the object's 3D shape.
- **Uncertainty modeling** in deep-MVS and deep-PS networks → selects confident depth and suface normal predictions.
- Neural volume rendering methodology for a trustworthy fusion of MVS and PS → handles diverse object material types such as **anisotropic** or **glossy** surfaces.



Method



Deep-MVS Network

- Learning-based PatchMatch approach for coarse-to-fine depth prediction.
- It provides uncertainty measure via computed **confidence values** of depth predictions.

Deep-PS Network

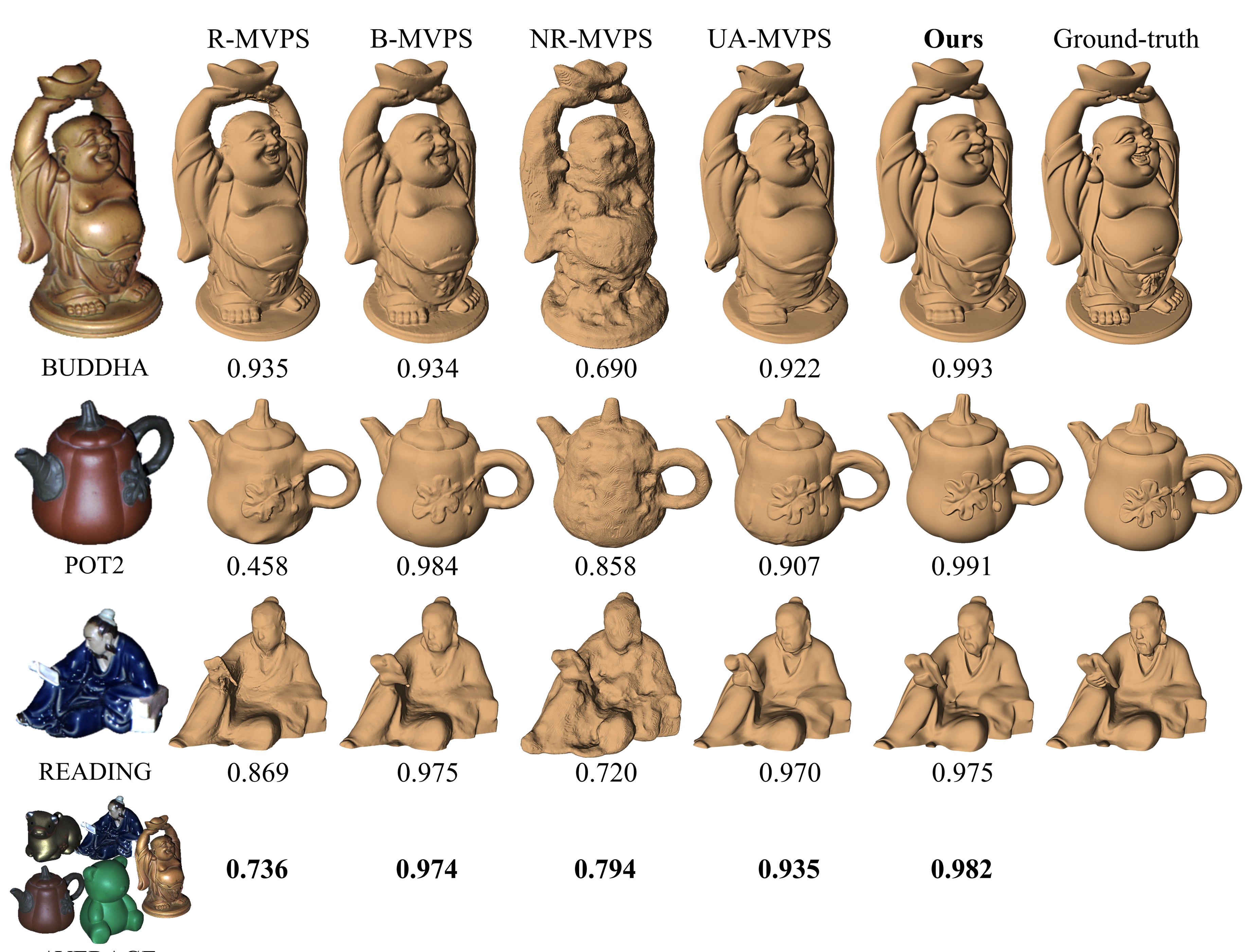
- Observation map representation for surface normal prediction.
- **Monte Carlo (MC) Dropout** technique for obtaining prediction uncertainty

Neural Volume Rendering and Optimization

- Signed distance function (SDF) of the object surface is defined by implicit function $f_\theta(\mathbf{x}): \mathbb{R}^3 \rightarrow \mathbb{R}$.
- Volume rendering recovers missing details of the surface → provides high-quality 3D reconstructions of challenging material types.

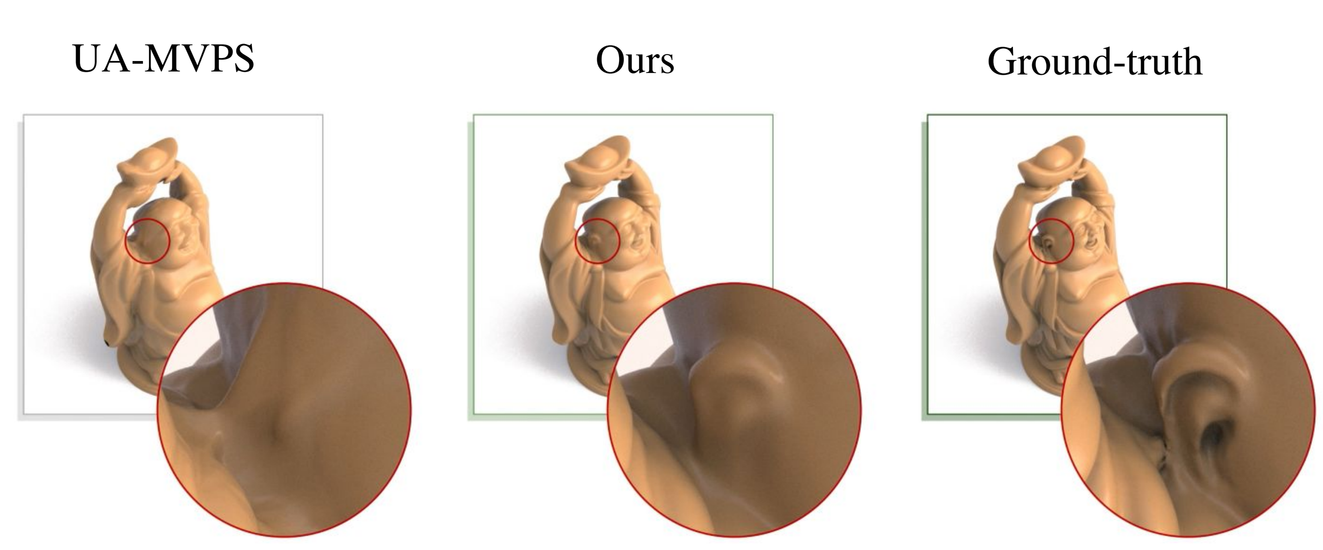
$$\mathcal{L} = \frac{1}{\mathcal{J}} \sum_{i=1}^{\mathcal{J}} \Big(\overbrace{c_i^{\text{mvs}} |f_\theta(\mathbf{p}_i)|}^{\text{MVS Loss}} + \overbrace{c_i^{\text{ps}} \|\mathbf{n}_i^r - \mathbf{n}_i^{ps}\|^2}^{\text{PS Loss}} + \overbrace{(1 - c_i^{\text{mvs}} c_i^{\text{ps}}) \|I_i - \hat{I}_i(\mathbf{c}_i, \mathbf{v}_i)\|_1}^{\text{Rendering Loss}} \Big)$$
$$+ \underbrace{\frac{\lambda_m}{|\mathcal{M}|} \sum_{i \in \mathcal{M}} CE\big(\max_{j \in \mathcal{S}_i}(\sigma_j/\alpha), 0\big)}_{\text{Mask Loss}} + \underbrace{\lambda_e \mathbb{E}_{\boldsymbol{x}}(\|\nabla_{\boldsymbol{x}} f_\theta(\boldsymbol{x})\| - 1)^2}_{\text{Eikonal Regularization}}$$

Reconstruction Results

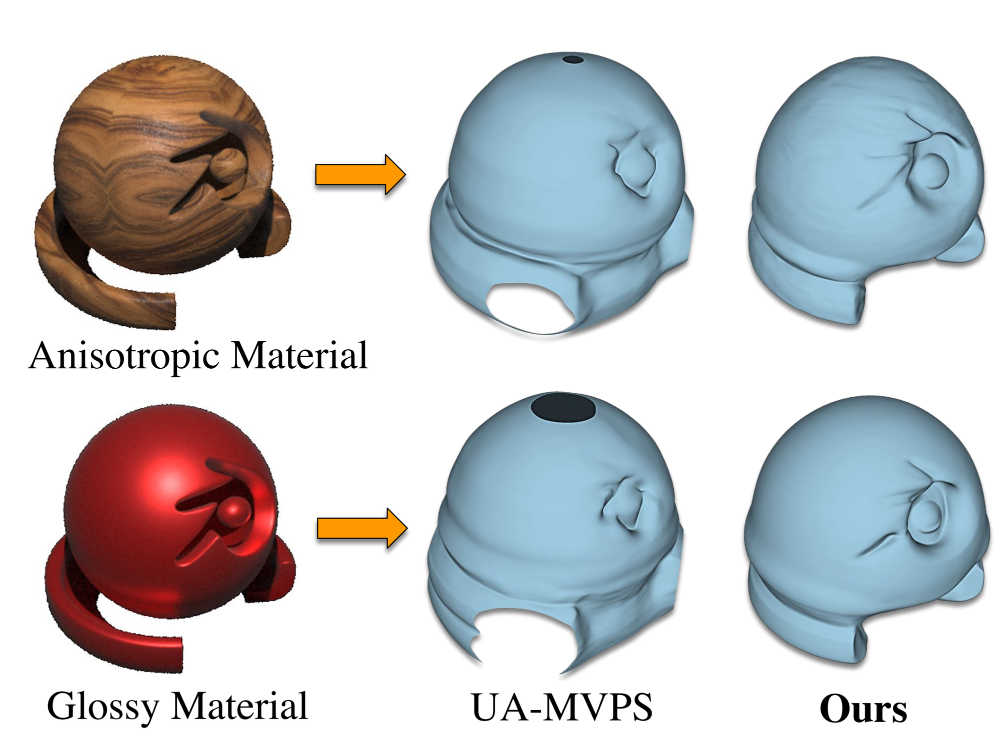


Method Category →	Deep Multi-View Stereo		Photometric Stereo			View-Synthesis		
Dataset↓ | Method →	MVSNet	PM-Net	Robust PS	SDPS-Net	CNN-PS	NeRF	VolSDF	**Ours**
BEAR	0.135	0.672	0.266	0.239	0.293	0.865	0.962	**0.965**
BUDDHA	0.147	0.799	0.367	0.298	0.363	0.713	0.786	**0.993**
COW	0.095	0.734	0.245	0.447	0.511	0.810	0.985	**0.987**
POT2	0.126	0.666	0.231	0.464	0.632	0.859	0.946	**0.991**
READING	0.115	0.834	0.242	0.188	0.508	0.673	0.683	**0.975**
AVERAGE	0.124	0.741	0.270	0.327	0.461	0.784	0.873	**0.982**

F-Score Comparison on DiLiGenT-MV



- Our method recovers details that are missing in other methods and provides outstanding 3D reconstructions.



- Our method is able to correctly recover the fine details for **anisotropic** and **glossy** material objects.

Key references:

- Ikehata, CNN-PS, ECCV 2018
- Wang et al., PM-Net, CVPR 2021
- Yariv et al., VolSDF, NeurIPS 2022
- Kaya et al., UA-MVPS, CVPR 2022
- Park et al., R-MVPS, TPAMI 2016
- Li et al., B-MVPS, TIP 2020