RNb-NeuS: Reflectance and Normal-based Multi-View 3D Reconstruction

Supplementary Material

Paper ID 12185

This supplementary material provides technicalities and detailed analysis of the experiments that were omitted in the main paper due to space limitations. We provide the reader with explicit formulations of the evaluation metrics in Section 1. We then share additional implementation details in Section 2. In Section 3, we present additional quantitative and qualitative results. In Section 4, we illustrate some limitations of our method. For a more comprehensive comparison showcasing the performance of our method, we have also included a supplementary video clip. We strongly encourage readers to watch the supplementary video for a better assessment of the proposed method's performance.

1. Evaluation

Metrics. All quantitative evaluations were carried out using Chamfer distance, F-score and mean angular error (MAE) between the reconstructed mesh $\mathcal{P}$ and the ground truth one $\mathcal{G}$. For a reconstructed point $\hat{\mathbf{x}} \in \mathcal{P}$, its distance to the ground truth is defined as follows:

$$d_{\hat{\mathbf{x}}\to\mathcal{G}} = \min_{\mathbf{x}\in\mathcal{G}} \|\hat{\mathbf{x}} - \mathbf{x}\|, \tag{1}$$

and vice versa for a ground truth point $\mathbf{x} \in \mathcal{G}$ and its distance to the reconstructed mesh.

The distance measures are accumulated over the entire meshes to define the Chamfer distance

$$CD = \frac{1}{2}\left(\frac{1}{|\mathcal{P}|}\sum_{\hat{\mathbf{x}}\in\mathcal{P}} d_{\hat{\mathbf{x}}\to\mathcal{G}} + \frac{1}{|\mathcal{G}|}\sum_{\mathbf{x}\in\mathcal{G}} d_{\mathbf{x}\to\mathcal{P}}\right) \tag{2}$$

and the F-score

$$F(\epsilon) = \frac{2P(\epsilon)R(\epsilon)}{P(\epsilon)+R(\epsilon)}, \tag{3}$$

where

$$P(\epsilon) = \frac{1}{|\mathcal{P}|}\sum_{\hat{\mathbf{x}}\in\mathcal{P}} [d_{\hat{\mathbf{x}}\to\mathcal{G}} < \epsilon] \tag{4}$$

and

$$R(\epsilon) = \frac{1}{|\mathcal{G}|}\sum_{\mathbf{x}\in\mathcal{G}} [d_{\mathbf{x}\to\mathcal{P}} < \epsilon] \tag{5}$$

are precision and recall measures, respectively, [.] is the Iverson bracket and ϵ is the distance threshold.

The mesh segmentations into low visibility and high curvature areas are performed on the ground truth meshes. Because the geometry of the reconstruction differs from that of the ground truth, the segmentation procedure yields different areas when applied to the reconstruction. For this reason, the reported results for low visibility and high curvature areas only consider the Chamfer distance term indicating the average distances between the ground truth vertices and their nearest neighbors in the reconstructed mesh.

For the MAE computation, the reconstructed and ground truth meshes are projected onto image planes and the normals are computed at each pixel. The MAE over all the pixels M is written as

$$MAE = \frac{1}{|M|}\sum_{k\in M} \cos^{-1}(\hat{\mathbf{n}}_k^\top \mathbf{n}_k). \tag{6}$$

DiLiGenT-MV dataset. All the state-of-the-art methods were evaluated from the meshes that were kindly provided by their authors. For all evaluated meshes, we eliminated all internal vertices. Then, a mesh upsampling for both estimated and ground truth meshes was then performed in order to achieve a point density of 0.1 mm. The computations of Chamfer distance and F-score were specifically conducted for distances under 5 mm in order to mitigate the impact of outliers (inspired by the DTU evaluation [5]).

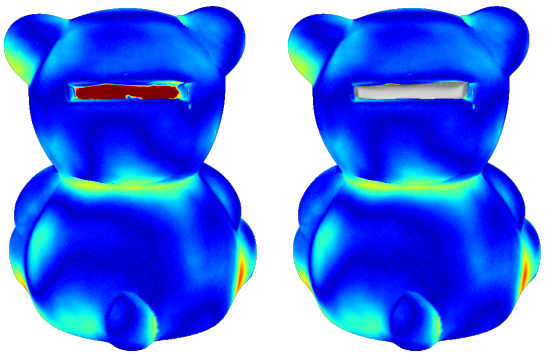
We observed a few defects in the ground truth meshes from the DiLiGenT-MV dataset in concave areas. Notably, such imperfections are well visible at the back of Bear's head (Fig. 1) and the spout's inner area of Pot2 (Fig. 2). Although these areas represent a small amount of vertices, they were discarded in all evaluations so as to avoid penalizing methods which faithfully reconstruct them.

Manual efforts in [9]. Li19 [9] is mentioned as requiring manual efforts. Indeed, the authors manually establish point correspondences in textureless areas. See [9] for details.

2. Implementation details

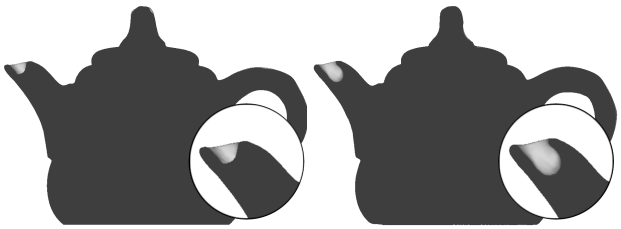
We recall that to simulate the radiance values in Step 4 described in Section 4 of the main paper, we choose as lighting triplet the one which is optimal, relatively to the normal $\mathbf{n}_k$. Following [2], this optimal triplet is equally spaced in tilt 120 degrees apart with a slant angle of 54.74 degrees. Concretely, the expression of L_k as a function of $\mathbf{n}_k$ is written:

$$\mathsf{L}_k = \mathsf{R}_k \mathsf{L}_{\text{canonic}} \tag{7}$$



(a) (b)

Figure 1. Rear view of the 3D heatmaps representing errors for the Bear dataset in terms of Chamfer distance. (a) The ground truth from DiLiGenT-MV lacks any vertices in the rectangular aperture. For that reason, any method which faithfully reconstructs this area is penalized (area shown in red). This area is thus discarded in all evaluations, providing heatmaps such as (b).



(a) (b)

Figure 2. Cross-section of Pot2's spout delivered by (a) the ground truth of the DiLiGenT-MV dataset and (b) our reconstruction method. Our method shows a deeper reconstruction of the internal wall of the spout. This area is thus discarded in all evaluations to avoid penalizing methods that faithfully reconstructs it.

where $\mathsf{R}_k = \mathsf{U}$ with $[\mathsf{U}, \Sigma, \mathsf{U}] = \mathrm{SVD}(\mathbf{n}_k \mathbf{n}_k^\top)$ and

$$\mathsf{L}_{\text{canonic}} = \begin{bmatrix} \sin(\phi) & \sin(\phi)\cos(\theta) & \sin(\phi)\cos(2\theta) \\ 0 & \sin(\phi)\sin(\theta) & \sin(\phi)\sin(2\theta) \\ \cos(\phi) & \cos(\phi) & \cos(\phi) \end{bmatrix} \quad (8)$$

with $\theta = \frac{120\pi}{180}$ and $\phi = \frac{54.74\pi}{180}$.

3. Additional Results

In this section, we extend the experiments of the main paper by providing further statistical analysis and qualitative comparisons.

Comparison with mono-illumination NeuS. We propose an additional comparison of our method against the multi-view mono-illumination 3D reconstruction method NeuS [11]. While NeuS is not directly applicable in multi-view multi-light acquisition settings in theory, it may become feasible under certain conditions. This feasibility hinges on factors such as the number, spatial distribution and types of lighting conditions, and the object material properties. One can leverage a heuristic method, initially proposed in [9] and later employed for obtaining pixel depths using MVS in [6, 8]. This heuristic involves approximating input images captured under mono-illumination for each viewpoint by taking the median of pixel intensities obtained under varying illumination. See, e.g., [9] for detailed information.

A qualitative comparison between the results of mono-illumination NeuS using this heuristic and the ones from our method is provided in Fig. 3. As can be seen, our proposed approach provides a much finer level of details. In particular, mono-illumination NeuS requires a high number of viewpoints, with a drastic decline in the reconstruction quality when using 5 viewpoints. On the contrary, our method shows stable results, only losing some fine details over concave areas. Moreover, even with all viewpoints used, mono-illumination NeuS fails in reliably reconstructing the low visibility and high curvature areas. In addition to Fig. 3 (right), this can be observed in the quantitative evaluation provided in Table 1, where mono-illumination NeuS shows a reconstruction error 62% higher than ours on low visibility areas and 46% higher than ours on high curvature areas.

Photometric stereo method. Our method can be employed with any PS method. To illustrate this flexibility, we evaluate the reconstruction accuracy on the Buddha dataset while taking as input the normal maps from CNN-PS [3], used in Kaya22-23 [6, 8], and SDPS-Net [1], used in PS-NeRF [12], in addition to the one obtained using normal maps from SDM-UniPS [4] reported in the main paper. The results are reported in Table 2. As expected, we observe that the choice of a particular PS technique influences the final outcome, yet our framework consistently improves the results in comparison with previous works, including those based on multi-objective optimizations [6, 8].

	CNN-PS		SDPS-Net		SDM-UniPS
Buddha	Kaya23	Ours	PS-NeRF	Ours	Ours
H. curv.	0.35	**0.29**	0.51	**0.31**	0.26
Low curv.	0.24	**0.22**	0.33	**0.25**	0.23
All	0.25	**0.22**	0.34	**0.25**	0.23

Table 2. Results of our method with different input normals, namely CNN-PS (used in Kaya22-23), SDPS-Net (used in PS-NeRF) and SDM-UniPS. High curvature corresponds to the results averaged over all the vertices whose absolute curvature is higher than 3.3. Our method shows to perform best irrespective of the PS method being used.

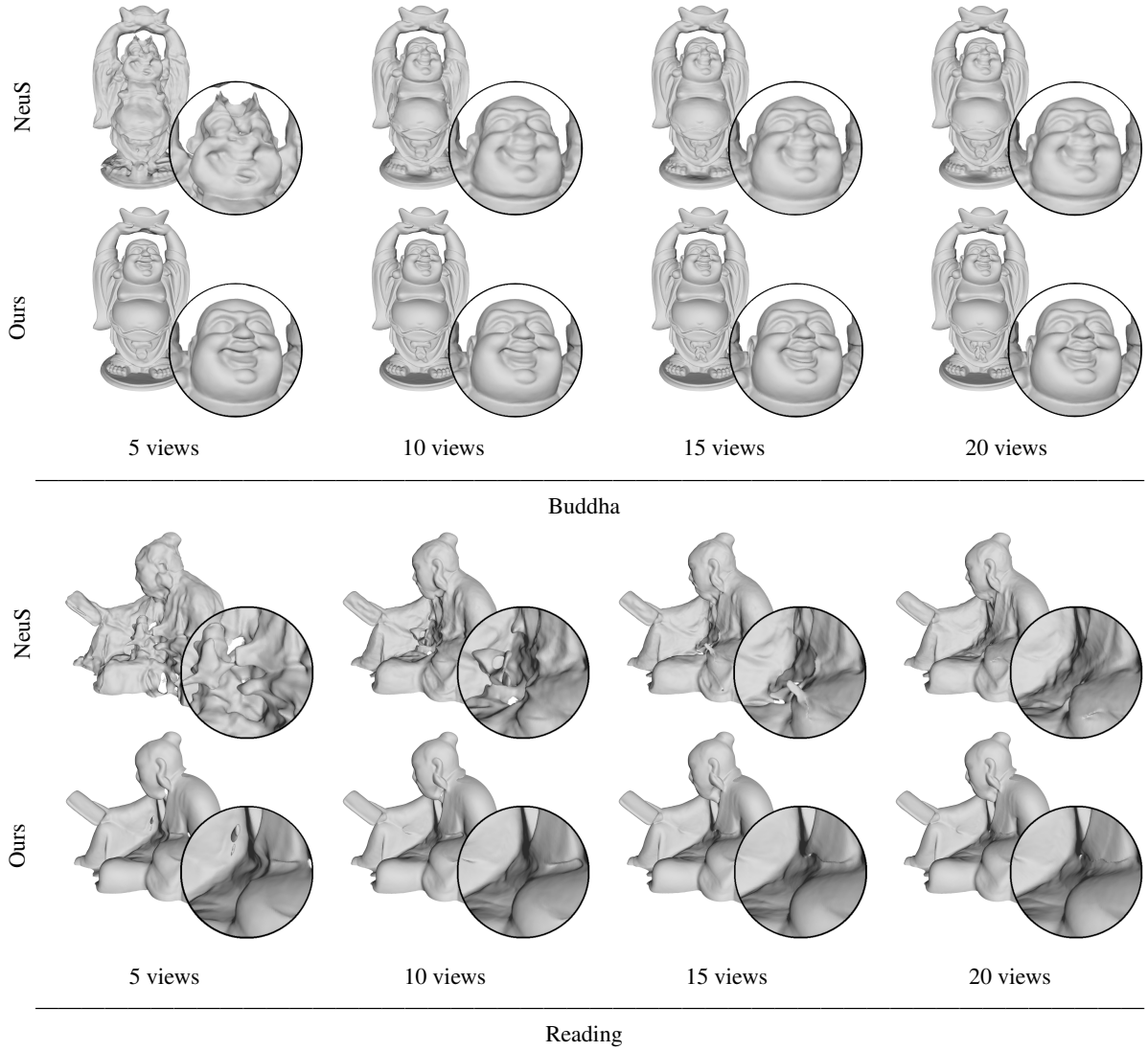


Figure 3. Qualitative comparison of Buddha and Reading between mono-illumination NeuS and our method, for different number of input viewpoints.

	Chamfer distance (visibility 1-5) ↓						Chamfer distance (high curvature) ↓					
Methods	Bear	Buddha	Cow	Pot2	Reading	Average	Bear	Buddha	Cow	Pot2	Reading	Average
Park16	1.07	0.75	0.41	0.47	0.7	0.68	1.64	0.58	0.98	0.56	0.65	0.88
Li19†	0.63	1.03	0.37	0.54	0.81	0.67	0.59	0.65	0.38	0.34	0.57	0.51
NeuS	0.58	0.52	**0.17**	0.32	0.54	0.42	0.28	0.46	**0.21**	0.39	0.38	0.35
Kaya22	0.48	0.51	0.32	0.5	0.7	0.5	0.33	0.43	0.31	0.41	0.45	0.38
PS-NeRF	0.48	0.62	0.3	0.66	0.64	0.54	0.42	0.5	0.42	0.44	0.44	0.45
Kaya23	0.46	**0.35**	0.39	0.42	**0.44**	**0.41**	0.33	**0.29**	**0.19**	**0.3**	**0.33**	**0.29**
MVPSNet	**0.43**	0.68	0.27	0.49	0.57	0.49	0.56	0.58	0.52	0.47	0.54	0.53
Ours	**0.23**	**0.27**	**0.19**	**0.19**	**0.43**	**0.26**	**0.22**	**0.23**	0.26	**0.23**	**0.25**	**0.24**

Table 1. Chamfer distance on (a) low visibility and (b) high curvature areas. **Best results**. **Second best results**.

Ablation. We complete our ablation study with qualitative results on the ear and the knot of Buddha shown in Fig. 4.

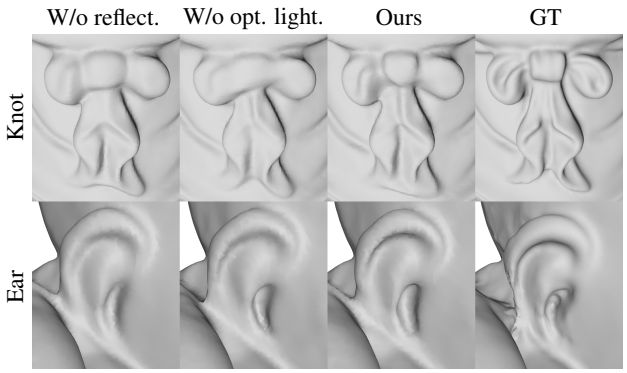


Figure 4. Qualitative comparison on the knot and the ear of Buddha between our results and those without the use of reflectance and optimal lighting, disabled individually. Our method exhibits better results in both cases.

Additional benchmarking. We provide in Fig. 7 a qualitative comparison of the angular error maps on the five objects of DiLiGenT-MV, for our method and state-of-the-art ones, namely Park16 [10], Li19 [9], Kaya22 [7], PS-NeRF [12], Kaya23 [8], MVPSNet [13] and also SDM-UniPS [4] although it does not provide a full 3D reconstruction. The recovered geometry shows to be overall more accurate with our method. Interestingly, our recovered normals overcome the PS ones, especially in concave areas where inter-reflections bias the single-viewpoint reconstruction. Lastly, we provide further quantitative comparisons, namely precision and recall in Fig. 5, and MAE on low visibility and high curvature areas in Table 3. Our proposed approach consistently yields the most accurate reconstructions.

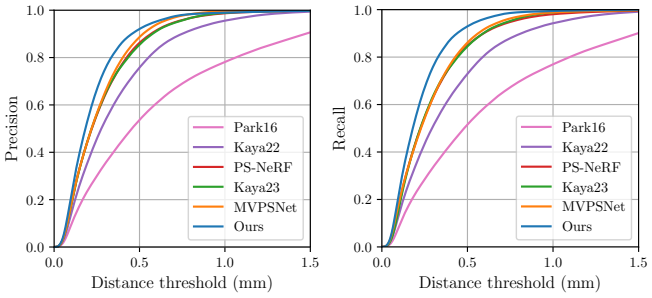


Figure 5. (a) Precision (higher is better) and (b) recall (higher is better) as functions of the distance error threshold, in comparison with other state-of-the-art methods.

4. Limitations

The reconstructions obtained through the proposed method yet exhibit a few poorly reconstructed areas, as illustrated in Fig. 6, particularly for Reading's neck and Bear's right ear. The suboptimal reconstruction of Reading's neck can be attributed, in part, to inacurracies of the normal estimates from SDM-UniPS. However, the underlying causes of these discrepancies have yet to be systematically identified.

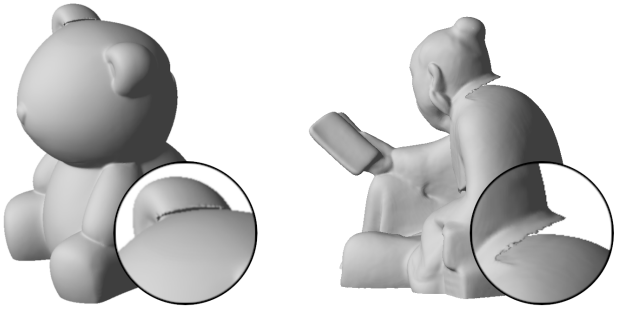


Figure 6. Regions in Bear (a) and Reading (b) where our method exhibits limitations.

References

[1] Guanying Chen, Kai Han, Boxin Shi, Yasuyuki Matsushita, and Kwan-Yee K. Wong. SDPS-Net: Self-calibrating Deep Photometric Stereo Networks. In *CVPR*, 2019. 2

[2] Ondrej Drbohlav and Mike Chantler. On optimal light configurations in photometric stereo. In *Proceedings of the 10th IEEE International Conference on Computer Vision*, pages 1707–1712, 2005. 1

[3] Satoshi Ikehata. CNN-PS: CNN-based photometric stereo for general non-convex surfaces. In *ECCV*, 2018. 2

[4] Satoshi Ikehata. Scalable, Detailed and Mask-Free Universal Photometric Stereo. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13198–13207, 2023. 2, 4

[5] Rasmus Jensen, Anders Dahl, George Vogiatzis, Engil Tola, and Henrik Aanæs. Large scale multi-view stereopsis evaluation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 406–413, 2014. 1

[6] Berk Kaya, Suryansh Kumar, Carlos Oliveira, Vittorio Ferrari, and Luc Van Gool. Uncertainty-aware deep multi-view photometric stereo. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12601–12611, 2022. 2

[7] Berk Kaya, Suryansh Kumar, Francesco Sarno, Vittorio Ferrari, and Luc Van Gool. Neural radiance fields approach to deep multi-view photometric stereo. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1965–1977, 2022. 4

[8] Berk Kaya, Suryansh Kumar, Carlos Oliveira, Vittorio Ferrari, and Luc Van Gool. Multi-View Photometric Stereo

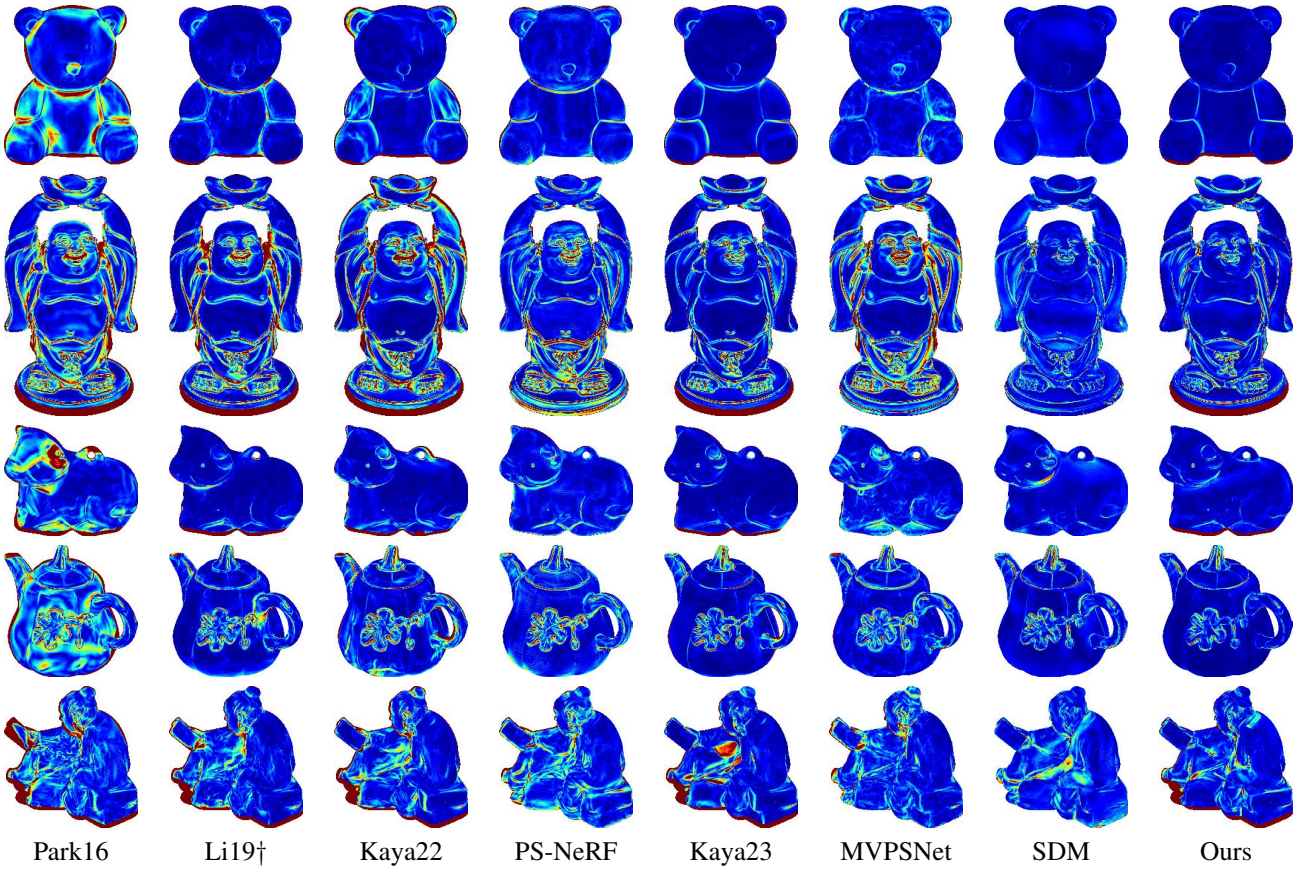


Figure 7. Normal angular error comparison over all DiLiGenT-MV dataset between state-of-the-art methods and ours.

	Normal MAE (Visibility 1-5) ↓						Normal MAE (High Curvature) ↓					
Methods	Bear	Buddha	Cow	Pot2	Reading	Average	Bear	Buddha	Cow	Pot2	Reading	Average
Park16	38.5	29.3	34.6	**25.2**	20.6	29.6	31.7	26.2	39.5	**23.3**	24.1	29.0
Li19†	41.1	33.7	29.4	39.0	23.3	33.3	26.5	26.4	30.6	23.4	24.1	26.2
Kaya22	32.0	27.6	40.5	40.0	18.4	31.7	20.2	29.1	35.9	32.8	21.8	28.0
PS-NeRF	19.4	19.6	27.4	32.2	21.1	24.0	21.2	28.0	27.9	23.9	28.3	25.8
Kaya23	19.4	17.6	**24.0**	28.1	**14.6**	20.7	24.1	24.2	**21.6**	28.5	**19.3**	**23.6**
MVPSNet	31.1	29.6	30.4	35.3	18.1	28.9	**18.7**	26.3	27.7	**23.3**	23.5	23.9
SDM	**12.9**	**14.4**	28.5	25.7	16.9	**19.7**	21.6	**21.0**	**23.4**	28.2	24.7	23.8
Ours	**13.0**	**14.1**	**26.8**	**21.5**	**13.5**	**17.8**	**18.4**	**24.2**	28.0	24.9	**19.9**	**23.1**

Table 3. Normal MAE on (a) low visibility and (b) high curvature areas. **Best results**. **Second best results**.

Revisited. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3126–3135, 2023. 2, 4

[9] Min Li, Zhenglong Zhou, Zhe Wu, Boxin Shi, Changyu Diao, and Ping Tan. Multi-view photometric stereo: A robust solution and benchmark dataset for spatially varying isotropic materials. *IEEE Transactions on Image Processing*, 29:4159–4173, 2020. 1, 2, 4

[10] Jaesik Park, Sudipta N. Sinha, Yasuyuki Matsushita, Yu-Wing Tai, and In So Kweon. Multiview photometric stereo using planar mesh parameterization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1161–1168, 2013. 4

[11] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction. In *Proceedings of the Conference on Neural Information Processing Systems*, 2021. 2

[12] Wenqi Yang, Guanying Chen, Chaofeng Chen, Zhenfang Chen, and Kwan-Yee K. Wong. PS-NeRF: Neural Inverse Rendering for Multi-view Photometric Stereo. In *Proceedings of the European Conference on Computer Vision*, pages

266–284, 2022. 2, 4

[13] Dongxu Zhao, Daniel Lichy, Pierre-Nicolas Perrin, Jan-Michael Frahm, and Soumyadip Sengupta. MVPSNet: Fast Generalizable Multi-view Photometric Stereo. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12525–12536, 2023. 4