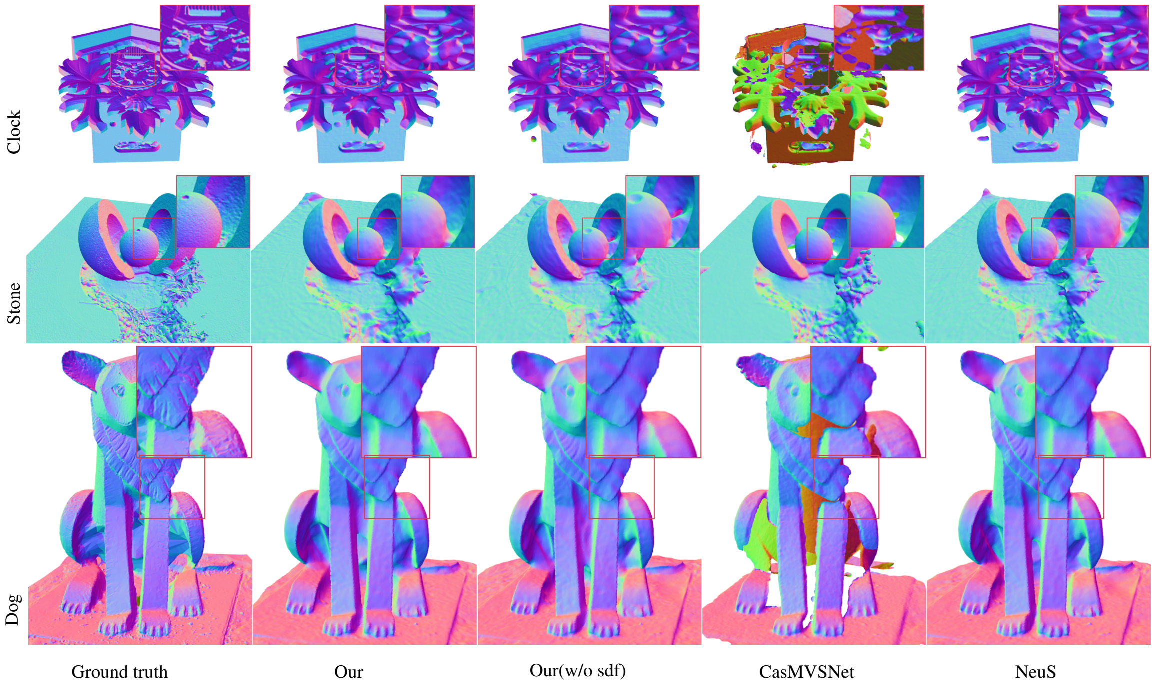

Figure 1. Qualitative comparisons of rendered normal maps on BlendedMVS.

1. Experiments on BlendedMVS

The BlendedMVS dataset [1] is another dataset used for evaluation of object-centered 3D reconstruction. There are 31-143 posed images with the resolution of 768 × 576 in each scene. Ground truth meshes and masks are supplied by the authors. We use 8 scenes selected by NeuS [5] to evaluate our method and others. We provide quantitative and qualitative comparisons with NeuS [3] and CasMVSNet [1] on the BlendedMVS [4] dataset. As for the CasMVSNet [1], we select the checkpoint pre-trained on DTU [4][1]. The fused point clouds are converted into meshes by the screened Poisson Surface Reconstruction [2] with trim parameter 8 at the resolution 512^3.

1.1. Quantitative Comparisons

Similar to the DTU evaluation pipeline, the output meshes are first cleaned with the dilated visibility masks, and then are evaluated with the chamfer distance. Since the units of different scene are unknown in this dataset, we scale the bounding box of each object to 100. As shown in Table 1, we compare our method with NeuS [3] and CasMVSNet [1]. Although CasMVSNet [1] can provide accurate point clouds, there are many holes on the recovered meshes. On the contrary, NeuS [3] can produce smooth surfaces, but its precision is limited. Compared with NeuS [3], our ablation model without SDF loss achieves better performance. Combined with the SDF loss, our full model achieves the smallest error.

ScanID	NeuS	CasMVSNet	Our(w/o sdf)	Our
bear	0.27	0.40	**0.26**	**0.26**
clock	0.93	0.53	0.33	**0.27**
dog	0.83	0.98	0.80	**0.49**
durian	0.37	0.50	0.37	**0.33**
jade	1.04	1.17	1.01	**0.82**
man	0.57	0.62	0.57	**0.55**
sculpture	0.50	0.53	0.46	**0.43**
stone	0.42	0.27	0.35	0.30
mean	0.62	0.63	0.52	**0.43**

Table 1. Quantitative evaluation of meshes on BlendedMVS.

1.2. Qualitative Comparisons

We visually compare our method with baselines in Fig. 1 and Fig. 2. Columns represent the ground truth, CasMVSNet [1], NeuS [3], our model without SDF loss and our full model, respectively. Rows represent the results of different scenes. It can be seen that there are many lost regions in the result of CasMVSNet [1]. As shown in the first row of Fig. 1 and Fig. 2, the minute hand of the clock is lost in the result of NeuS [3]. A Large area of concave error occurs in the inner surface of the scene stone. Our model without SDF loss can recover the minute hand correctly, and has smaller concave errors. More details are captured in the result of our full model.

[1] https://github.com/kwea123/CasMVSNet_pl

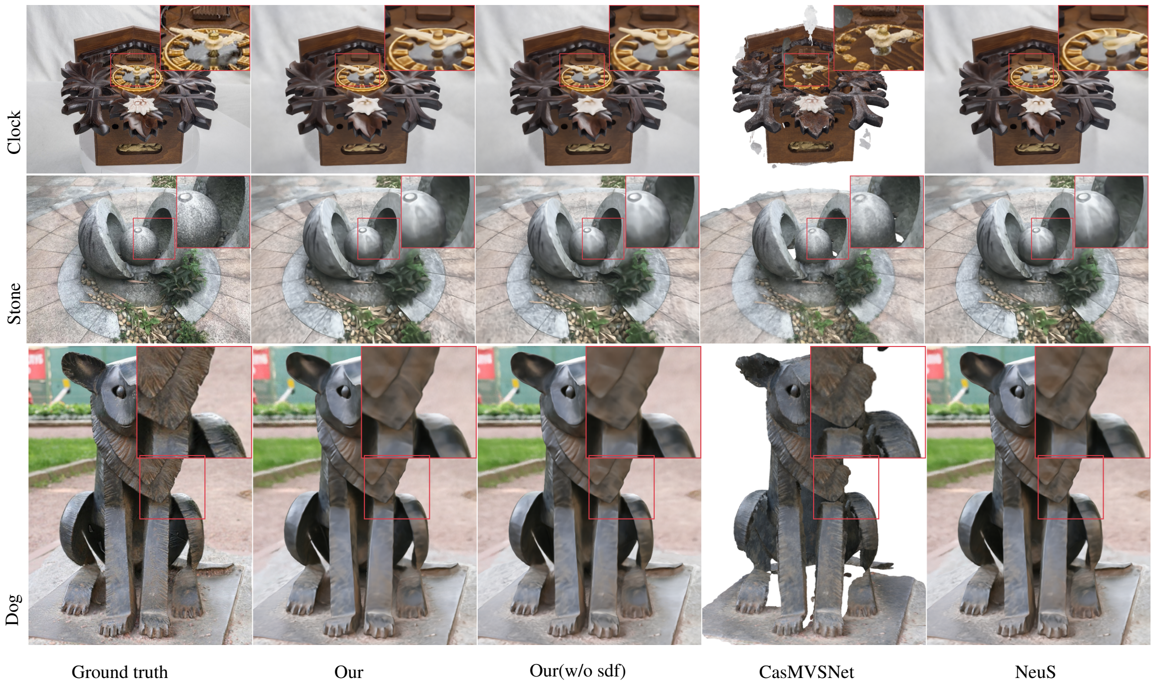

Figure 2. Qualitative comparisons of rendered images on BlendedMVS

ScanID	NeuS	NeuS-12	CasMVSNet	Our(w/o sdf)	Our
24	1.00	0.93	0.82	0.72	**0.49**
37	1.37	1.07	0.93	0.77	**0.71**
40	0.93	0.81	0.41	0.67	**0.37**
55	0.43	0.38	0.43	0.37	**0.36**
63	1.10	1.02	0.88	0.93	**0.80**
65	0.65	0.60	0.96	0.58	**0.56**
69	0.57	0.58	0.68	0.61	**0.52**
83	1.48	1.43	1.24	1.23	**1.17**
97	1.09	1.15	1.25	1.15	**0.97**
105	0.83	0.78	0.82	**0.65**	0.66
106	0.52	0.57	0.65	0.56	**0.48**
110	1.20	1.16	0.78	1.08	**0.73**
114	0.35	0.35	0.48	0.34	**0.32**
118	0.49	0.45	0.58	0.45	**0.42**
122	0.54	0.46	0.54	0.47	**0.42**
mean	0.84	0.78	0.76	0.71	**0.60**

Table 2. Additional quantitative comparisons of meshes on DTU.

2. Additional Experiments on DTU

Similar to Section 1, we compare our model with baselines on DTU [4]. Here we choose the checkpoint of CasMVSNet [1] pre-trained on BlendedMVS [4]. As shown in Table 2, our model without SDF loss outperforms NeuS [3], and our full model outperforms both of NeuS [3] and CasMVSNet [1].

3. Convergence time

The convergence steps of our model and NeuS [3] are shown in Fig. 3. It can be seen that only 20K steps are needed to achieve compared results to NeuS [3] at 300K steps. Our model can speed up the training process efficiently.

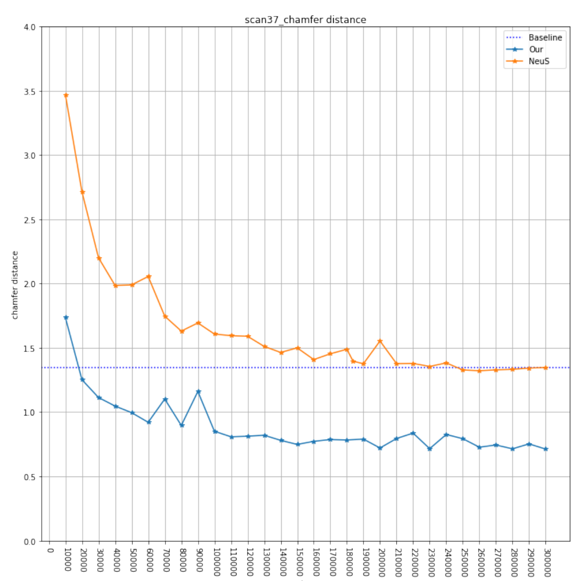

Figure 3. Convergence steps on DTU Scan37.

4. Proof of Unbiased Property of Our Model

Logistic CDF. As for the Logistic CDF $\Psi_\beta(x) = \frac{1}{1+\exp(-x/\beta)}$ and $\alpha = \frac{1}{\beta}$, the density is modeled as

$$\sigma(\mathbf{p}(t)) = \frac{1}{\beta\left(1+\exp\left(\frac{f(\mathbf{p}(t))}{\beta|f'(\mathbf{p}(t))|}\right)\right)}. \tag{1}$$

It is known that $\Psi'_\beta(0) = \frac{1}{4\beta}$, $\Psi_\beta(0) = \frac{1}{2}$. The derivative of $\frac{-f(\mathbf{p}(\bar{t}))}{|f'(\mathbf{p}(t))|}$ respect to $\bar{t}$ at any intersection point $f(\mathbf{p}(\bar{t})) = 0$ is

$$\begin{aligned} \left(\frac{-f(\mathbf{p}(\bar{t}))}{|f'(\mathbf{p}(t))|}\right)' &= \frac{-f'(\mathbf{p}(\bar{t}))|f'(\mathbf{p}(\bar{t}))|+f(\mathbf{p}(\bar{t}))|f'(\mathbf{p}(\bar{t}))|'}{|f'(\mathbf{p}(\bar{t}))|^2} \\ &= -\frac{f'(\mathbf{p}(\bar{t}))}{|f'(\mathbf{p}(\bar{t}))|}. \end{aligned} \tag{2}$$

Then it can be deduced $\sigma(\mathbf{p}(\bar{t})) = \frac{1}{2\beta}$, $\sigma'(\mathbf{p}(\bar{t})) = \frac{1}{4\beta^2}\frac{-f'(\mathbf{p}(\bar{t}))}{|f'(\mathbf{p}(\bar{t}))|}$. Finally, the derivative of $w(\bar{t})$ respect to $\bar{t}$ at any intersection point $f(\mathbf{p}(\bar{t})) = 0$ is

$$\begin{aligned} \frac{\mathrm{d}w}{\mathrm{d}t}(\bar{t}) &= T(\bar{t})\left(\sigma'(\mathbf{p}(\bar{t})) - \sigma(\mathbf{p}(\bar{t}))^2\right) \\ &= \frac{1}{4\beta^2}T(\bar{t})\left(-\frac{f'(\mathbf{p}(\bar{t}))}{|f'(\mathbf{p}(\bar{t}))|} - 1\right) \\ &= \begin{cases} 0, & if\ f'(\mathbf{p}(\bar{t})) < 0 \\ -\frac{1}{2\beta^2}T(\bar{t}), & if\ f'(\mathbf{p}(\bar{t})) > 0 \end{cases} \end{aligned} \tag{3}$$

Laplace CDF. As for the Laplace CDF $\Psi_\beta(x) = \frac{1}{2} + \frac{1}{2}\,\mathrm{sign}(x)\left(1-\exp(-|x|/\beta)\right)$ and $\alpha = \frac{2}{\beta}$, the density is modeled as

$$\sigma(\mathbf{p}(t)) = \frac{1}{\beta} + \frac{1}{\beta}\,\mathrm{sign}\left(\frac{-f(\mathbf{p}(t))}{|f'(\mathbf{p}(t))|}\right)\left(1-\exp\left(-\left|\frac{-f(\mathbf{p}(t))}{\beta|f'(\mathbf{p}(t))|}\right|\right)\right). \tag{4}$$

It is known that $\Psi'_\beta(0) = \frac{1}{2\beta}$, $\Psi_\beta(0) = \frac{1}{2}$. For any intersection point $f(\mathbf{p}(\bar{t})) = 0$, it can be deduced $\sigma(\mathbf{p}(\bar{t})) = \frac{1}{\beta}$, $\sigma'(\mathbf{p}(\bar{t})) = \frac{1}{\beta^2}\frac{-f'(\mathbf{p}(\bar{t}))}{|f'(\mathbf{p}(\bar{t}))|}$. Finally, the derivative of $w(\bar{t})$ respect to $\bar{t}$ at any intersection point $f(\mathbf{p}(\bar{t})) = 0$ is

$$\begin{aligned} \frac{\mathrm{d}w}{\mathrm{d}t}(\bar{t}) &= T(\bar{t})\left(\sigma'(\mathbf{p}(\bar{t})) - \sigma(\mathbf{p}(\bar{t}))^2\right) \\ &= \frac{1}{\beta^2}T(\bar{t})\left(-\frac{f'(\mathbf{p}(\bar{t}))}{|f'(\mathbf{p}(\bar{t}))|} - 1\right) \\ &= \begin{cases} 0, & if\ f'(\mathbf{p}(\bar{t})) < 0 \\ -\frac{2}{\beta^2}T(\bar{t}), & if\ f'(\mathbf{p}(\bar{t})) > 0 \end{cases} \end{aligned} \tag{5}$$

References

[1] Rasmus Jensen, Anders Dahl, George Vogiatzis, Engin 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, 2

[2] Michael Kazhdan and Hugues Hoppe. Screened poisson surface reconstruction. *ACM Transactions on Graphics (ToG)*, 32(3):1–13, 2013. 1

[3] 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. *Advances in Neural Information Processing Systems*, 34:27171–27183, 2021. 1, 2

[4] Yao Yao, Zixin Luo, Shiwei Li, Jingyang Zhang, Yufan Ren, Lei Zhou, Tian Fang, and Long Quan. Blendedmvs: A large-scale dataset for generalized multi-view stereo networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1790–1799, 2020. 1, 2

[5] Lior Yariv, Yoni Kasten, Dror Moran, Meirav Galun, Matan Atzmon, Basri Ronen, and Yaron Lipman. Multiview neural surface reconstruction by disentangling geometry and appearance. *Advances in Neural Information Processing Systems*, 33:2492–2502, 2020. 1