

NeUDF: Leaning Neural Unsigned Distance Fields with Volume Rendering

– Supplemental Material –

Yu-Tao Liu^{1,2} **Li Wang**^{1,2} **Jie Yang**¹ **Weikai Chen**³
Xiaoxu Meng³ **Bo Yang**³ **Lin Gao**^{1,2*}

¹Beijing Key Laboratory of Mobile Computing and Pervasive Device,
Institute of Computing Technology, Chinese Academy of Sciences

²University of Chinese Academy of Sciences

³Tencent Games Digital Content Technology Center

liuyutao17@mails.ucas.ac.cn {wangli20s, yangjie01}@ict.ac.cn chenwk891@gmail.com
{xiaoxumeng, brandonyang}@tencent.com gaolin@ict.ac.cn

1. Overview

In the main paper, we introduce a novel UDF-based volume rendering approach to achieve high-fidelity multi-view reconstruction for arbitrary shapes with both open and closed surfaces. This supplemental material consists of detailed proofs, implementation details and additional results of multi-view reconstruction. All the sections are organized as follows:

- Section 2 analyzes the inherent bias in color rendering of the naive UDF solution based on SDF renderer.
- Section 3 provides detailed proofs of the unbiased and occlusion-aware properties of our proposed NeUDF.
- Section 4 provides implementation details on network architecture (Section 4.1), training details (Section 4.2) and data preparation (Section 4.3).
- Section 5 provides additional qualitative results of multi-view reconstruction.

2. Bias in Naive UDF solution based on SDF renderer

In this section we illustrate the bias of color rendering introduced by the naive UDF solution based on SDF renderer, which directly extends the weight of NeuS to UDF. The bias causes inherent geometric error like redundant surfaces and floating noises.

To apply the naive UDF solution based on the SDF renderer of NeuS, we denote the rendered color $C(o, v)$:

$$C(o, v) = \int_0^{+\infty} w_n(t) c(p(t), v) dt, \quad (1)$$

where (o, v) are the origin and view direction of the sample ray, $c(x, v)$ the color at position x along the view direction v , and $w_n(t)$ the rendering weight of NeuS:

$$w_n(t) = \rho_s(t) e^{- \int_0^t \rho_s(u) du} \quad (2)$$

$$\rho_s(t) = \max\left\{-\frac{\partial(\Phi_s \circ \Psi \circ p)}{\partial t}(t), 0\right\} \quad (3)$$

where $\rho_s(t)$ denotes the opaque density of NeuS, $\Phi_s(d)$ the Sigmoid function, and $\Psi(x)$ the UDF value at position x . The learnable parameter s controls the distribution of the Sigmoid function, which is expected to increase to infinity during training.

Assume that the ray linearly crosses the open surface in its local neighbor, e.g., there exists an interval (t^l, t^r) , the intersection point $t^* \in (t^l, t^r)$, which satisfies:

$$\Psi \circ p(t) = |\cos \theta| \cdot |t - t^*|, \forall t \in (t^l, t^r), \quad (4)$$

where θ is the angle between the view direction and the surface normal.

In UDF, the color $C(o, v)$ rendered based on SDF renderer, Equ. 1, consists of inherent bias and inconsistency of the geometry. Denote the first intersection point t_0^* and its corresponding interval (t_0^l, t_0^r) , the bias can be formalized as below:

$$\lim_{s \rightarrow \infty} C(o, v) = 0.5c(p(t_0^*), v) + \frac{2^k - 1}{2^{k+1}} c_m + \frac{1}{2^{k+1}} c_n, \quad (5)$$

where k is the number of intersection points along the ray, c_m the undesired mixture of colors from invisible surfaces and c_n the colors from floating noise induced by the rendering bias. The parameter s decides the weight distribution of colors along the ray, and is supposed to increase towards infinity during training.

Note that the weight distribution corresponding to Equ. 5 satisfies the local maximal constraint discussed in NeuS, *i.e.* the weight attains local maxima at each intersection point (locally unbiased). But the local maximal constraint is not sufficient for an unbiased rendering for open surfaces due to the volume-surface representation discrepancy. The volume rendering relies on the volume-level color fusion for optimization, while the ground-truth color is exactly the surface color at the intersection point of the sample ray and the first intersected surface. A self-consistent rendering procedure should be able to address this volume-surface discrepancy, *i.e.* the color fusion range should be limited as close to the first intersection point as possible (globally unbiased). Otherwise the network is not able to converge to a surface representation through volume rendering. Note that the weight of NeuS is globally and locally unbiased for SDF, but not globally unbiased for UDF, and this difference comes from the difference of the value domains of SDF and UDF.

To illustrate the detailed causation of c_m and c_n , we first prove that:

$$\lim_{s \rightarrow \infty} \int_0^{t_0^l} w_n(t) dt = 0 \quad (6)$$

$$\lim_{s \rightarrow \infty} \int_{t_0^l}^{t_0^l} w_n(t) dt = 0.5, \quad (7)$$

which means that the output color consists of undesired bias whose weight sums to 0.5, and the bias cannot be corrected through training. Then we show the detailed distribution of the bias c_m and c_n for corroboration.

Proof of Equ. 6. Specifically, to prove Equ. 6, we have:

$$\begin{aligned} & \int_0^{t_0^l} w_n(t) dt \\ &= \int_0^{t_0^l} \rho_s(t) e^{- \int_0^t \rho_s(u) du} dt \\ &= \int_0^{t_0^l} -\frac{\partial}{\partial t} e^{- \int_0^t \rho_s(u) du} dt \\ &= -e^{- \int_0^{t_0^l} \rho_s(u) du} \Big|_0^{t_0^l} \\ &= -e^{- \int_0^{t_0^l} \rho_s(u) du} + 1 \\ &= -e^{- \int_0^{t_0^l} \max\{\frac{\partial(\Phi_s \circ \Psi \circ p)(u)}{\partial u}, 0\} du} + 1 \end{aligned} \quad (8)$$

It follows that:

$$\begin{aligned} & \int_0^{t_0^l} w_n(t) dt \\ &\leq -e^{- \int_0^{t_0^l} \left| \frac{\partial(\Phi_s \circ \Psi \circ p)(u)}{\Phi_s \circ \Psi \circ p(u)} \right| du} + 1 \\ &= -e^{- \int_0^{t_0^l} \left| \frac{\partial \Phi_s \circ \Psi \circ p(u) \cdot \partial \Psi \circ p(u)}{\partial \Psi \circ p(u) \cdot \Phi_s \circ \Psi \circ p(u)} \right| du} + 1 \\ &= -e^{- \int_0^{t_0^l} \left| \frac{\Phi'_s \circ \Psi \circ p(u) \cdot \partial \Psi \circ p(u)}{\Phi_s \circ \Psi \circ p(u)} \right| du} + 1 \\ &= -e^{- \int_0^{t_0^l} \frac{|\Phi'_s \circ \Psi \circ p(u)| \cdot |\partial \Psi \circ p(u)|}{|\Phi_s \circ \Psi \circ p(u)|} du} + 1 \end{aligned} \quad (9)$$

Denote that:

$$A = |\Phi'_s \circ \Psi \circ p(u)| \quad (10)$$

$$B = \left| \frac{\partial \Psi \circ p(u)}{\partial u} \right| \quad (11)$$

$$C = |\Phi_s \circ \Psi \circ p(u)| \quad (12)$$

We have:

$$\int_0^{t_0^l} w_n(t) dt = -e^{- \int_0^{t_0^l} \frac{A \cdot B}{C} du} + 1 \quad (13)$$

Because t_0^* is the first zero point of $\Psi \circ p(t)$ and $\Psi(x)$ is a continuous function, there is:

$$\exists \Psi_{min} > 0, \text{s.t., } \Psi \circ p(t) > \Psi_{min}, \forall t \in (0, t_0^l).$$

Note that $\Phi_s(x)$ is the Sigmoid function $\Phi_s(x) = (1 + e^{-s \cdot x})^{-1}$, and $\frac{\partial \Psi \circ p(u)}{\partial u}$ is the gradient of the UDF along the ray. We have:

$$C = |\Phi_s \circ \Psi \circ p(u)| \quad (14)$$

$$= (1 + e^{-s \cdot \Psi \circ p(u)})^{-1} \quad (15)$$

$$> (1 + e^{-s \cdot \Psi_{min}})^{-1} \quad (16)$$

$$> 0.5 \quad (17)$$

$$B = \left| \frac{\partial \Psi \circ p(u)}{\partial u} \right| < 1 \quad (18)$$

and $\forall \epsilon > 0, \exists S = \max\{1, \frac{-4t_0^l}{\ln(1-\epsilon) \cdot \Psi_{min}^2}\}, \text{s.t., } \forall s > S,$

there is:

$$\begin{aligned}
A &= |\Phi'_s \circ \Psi \circ p(u)| \\
&= \frac{s \cdot e^{-s \cdot \Psi \circ p(t)}}{(1 + s \cdot e^{-s \cdot \Psi \circ p(t)})^2} \\
&\leq \frac{2}{\Psi^2 \circ p(t) \cdot s} \\
&\leq \frac{2}{\Psi_{min}^2 \cdot s} \\
&\leq \frac{2}{\Psi_{min}^2 \cdot \frac{-4t_0^l}{\ln(1-\epsilon) \cdot \Psi_{min}^2}} \\
&= \frac{-0.5 \ln(1-\epsilon)}{t_0^l}
\end{aligned} \tag{19}$$

It follows that:

$$\begin{aligned}
\int_0^{t_0^l} w_n(t) dt &= -e^{-\int_0^{t_0^l} \frac{A \cdot B}{C} du} + 1 \\
&< -e^{-\int_0^{t_0^l} \frac{0.5 \ln(1-\epsilon)}{t_0^l} du} + 1 \\
&= -e^{-\int_0^{t_0^l} \frac{\ln(1-\epsilon)}{t_0^l} du} + 1 \\
&= -e^{-t_0^l \cdot \frac{\ln(1-\epsilon)}{t_0^l}} + 1 \\
&= -e^{\ln(1-\epsilon)} + 1 \\
&= -(1-\epsilon) + 1 \\
&= \epsilon
\end{aligned} \tag{20}$$

This leads to:

$$\begin{aligned}
\lim_{s \rightarrow \infty} \int_0^{t_0^l} w_n(t) dt &= \lim_{s \rightarrow \infty} (-e^{-\int_0^{t_0^l} \frac{A \cdot B}{C} du} + 1) \\
&= 0
\end{aligned} \tag{21}$$

The Equ. 21 means that the weight before the first intersection of the ray converges against zero during training, so the output color composites no color before the first intersected surface. This completes the proof of Equ. 6.

Proof of Equ. 7. Then we give the proof of Equ. 7. Same as the derivation of Equ. 8, we have:

$$\begin{aligned}
&\int_{t_0^l}^{t_0^*} w_n(t) dt \\
&= \int_{t_0^l}^{t_0^*} \rho_s(t) e^{-\int_0^t \rho_s(u) du} dt \\
&= \int_{t_0^l}^{t_0^*} -\frac{\partial}{\partial t} e^{-\int_0^t \rho_s(u) du} dt \\
&= -e^{-\int_0^{t_0^l} \rho_s(u) du} \Big|_{t_0^l}^{t_0^*} \\
&= -e^{-\int_0^{t_0^*} \rho_s(u) du} + e^{-\int_0^{t_0^l} \rho_s(u) du} \\
&= -e^{-\int_0^{t_0^l} \rho_s(u) du} - e^{-\int_{t_0^l}^{t_0^*} \rho_s(u) du} + e^{-\int_0^{t_0^l} \rho_s(u) du} \\
&= e^{-\int_0^{t_0^l} \rho_s(u) du} (-e^{-\int_{t_0^l}^{t_0^*} \rho_s(u) du} + 1)
\end{aligned} \tag{22}$$

Note that when $t \in (t_0^l, t_0^*)$, we have:

$$\frac{\partial \Psi \circ p(t)}{\partial t} = -|\cos \theta|. \tag{23}$$

It follows that:

$$\begin{aligned}
&-e^{-\int_{t_0^l}^{t_0^*} \rho_s(u) du} + 1 \\
&= -e^{-\int_{t_0^l}^{t_0^*} \max\{\frac{\partial(\Phi_s \circ \Psi \circ p)}{\partial u}(u), 0\} du} + 1 \\
&= -e^{-\int_{t_0^l}^{t_0^*} \left| \frac{\partial(\Phi_s \circ \Psi \circ p)}{\partial u}(u) \right| du} + 1 \\
&= -e^{-\int_{t_0^l}^{t_0^*} \left| \frac{\partial}{\partial u} \ln \Phi_s \circ \Psi \circ p(u) \right| du} + 1 \\
&= -e^{-\int_{t_0^l}^{t_0^*} -\frac{\partial}{\partial u} \ln \Phi_s \circ \Psi \circ p(u) du} + 1 \\
&= -e^{\ln \Phi_s \circ \Psi \circ p(t_0^*) - \ln \Phi_s \circ \Psi \circ p(t_0^l)} + 1 \\
&= -\frac{e^{\ln \Phi_s \circ \Psi \circ p(t_0^*)}}{e^{\ln \Phi_s \circ \Psi \circ p(t_0^l)}} + 1 \\
&= -\frac{\Phi_s \circ \Psi \circ p(t_0^*)}{\Phi_s \circ \Psi \circ p(t_0^l)} + 1
\end{aligned} \tag{24}$$

Since t_0^* is the intersection point, we have $\Psi \circ p(t_0^*) = 0$ and $\Phi_s \circ \Psi \circ p(t_0^*) = 0.5$. It follows that:

$$\begin{aligned}
&-e^{-\int_{t_0^l}^{t_0^*} \rho_s(u) du} + 1 \\
&= -\frac{\Phi_s \circ \Psi \circ p(t_0^*)}{\Phi_s \circ \Psi \circ p(t_0^l)} + 1 \\
&= -\frac{0.5}{\Phi_s \circ \Psi \circ p(t_0^l)} + 1 \\
&\leq -\frac{0.5}{1} + 1 = 0.5
\end{aligned} \tag{25}$$

$$\begin{aligned}
\forall \epsilon > 0, \exists S = \frac{-\ln 2\epsilon}{\Psi \circ p(t_0^l)}, \text{ s.t., } \forall s > S, \\
& -e^{-\int_{t_0^l}^{t_0^*} \rho_s(u) du} + 1 \\
& = -\frac{0.5}{\Phi_s \circ \Psi \circ p(t_0^l)} + 1 \\
& = -\frac{0.5}{(1 + e^{-s \cdot \Psi \circ p(t_0^l)})^{-1}} + 1 \\
& \geq -\frac{0.5}{(1 + e^{-\frac{-\ln 2\epsilon}{\Psi \circ p(t_0^l)} \cdot \Psi \circ p(t_0^l)})^{-1}} + 1 \quad (26) \\
& = -\frac{0.5}{(1 + e^{\ln 2\epsilon})^{-1}} + 1 \\
& = -\frac{0.5}{(1 + 2\epsilon)^{-1}} + 1 \\
& = -\epsilon + 0.5
\end{aligned}$$

The Equ. 25 and 26 derive that:

$$\lim_{s \rightarrow \infty} (-e^{-\int_{t_0^l}^{t_0^*} \rho_s(u) du} + 1) = 0.5 \quad (27)$$

It has been proved in Equ. 20 that:

$$\lim_{s \rightarrow \infty} (-e^{-\int_0^{t_0^l} \rho(u) du} + 1) = 0, \text{ i.e.,} \quad (28)$$

$$\lim_{s \rightarrow \infty} (e^{-\int_0^{t_0^l} \rho(u) du}) = 1 \quad (29)$$

The equations 22, 27 and 29 together derive that:

$$\begin{aligned}
& \lim_{s \rightarrow \infty} \int_{t_0^l}^{t_0^*} w_n(t) dt \\
& = \lim_{s \rightarrow \infty} (e^{-\int_0^{t_0^l} \rho_s(u) du} (-e^{-\int_{t_0^l}^{t_0^*} \rho_s(u) du} + 1)) \quad (30) \\
& = 0.5
\end{aligned}$$

The Equ. 30 determines that the rendered color $C(o, v)$ of NeuS in UDF cannot converge to the ground-truth color $c(p(t_0^*), v)$ as up to half of the weight is not constrained, which causes the mixed rendering color with undesired bias and inherent geometric error. This completes the proof of Equ. 7.

Distribution of Bias. Further, we illustrate the components of the bias, e.g., c_m and c_n , and show the corresponding distribution.

For $t \in (t_0^*, t_1^*)$, where t_0^* and t_1^* denotes the first and second intersection points along the ray $p(t)$. Consider that:

$$\begin{aligned}
w_n(t) & = \rho_s(t) e^{-\int_0^t \rho_s(u) du} \\
& = \rho_s(t) e^{-\int_{t_0^*}^t \rho_s(u) du} \cdot e^{-\int_0^{t_0^*} \rho_s(u) du} \quad (31)
\end{aligned}$$

As is proved, $\lim_{s \rightarrow \infty} e^{-\int_0^{t_0^*} \rho_s(u) du} = 0.5$, there is:

$$w_n(t_1^*) = 0.5 \rho_s(t) e^{-\int_{t_0^*}^{t_1^*} \rho_s(u) du} \quad (32)$$

According to the assumption that $\exists(t_1^l, t_1^r) \ni t_1^*$, the UDF value $\Psi(t)$ along the ray is linear for $t \in (t_1^l, t_1^r)$. So similarly we can prove that:

$$\begin{aligned}
& \lim_{s \rightarrow \infty} \int_0^{t_1^*} w_n(t) dt \\
& = 0.5 \lim_{s \rightarrow \infty} \int_0^{t_1^*} \rho_s(t) e^{-\int_{t_0^*}^t \rho_s(u) du} dt \\
& = 0.25
\end{aligned} \quad (33)$$

Consequently, for any given $k > 0$, we have:

$$\lim_{s \rightarrow \infty} \int_0^{t_k^*} w_n(t) dt = \frac{1}{2^{k+1}} \quad (34)$$

The colors of the k invisible surfaces are mixed to the output color $C(o, v)$, whose weight sums to $\frac{2^k - 1}{2^{k+1}}$. The mixed colors integral c_m leads to the undesired bias $\frac{2^k - 1}{2^{k+1}} c_m$, which cannot be corrected during training. The last weight $1 - 0.5 - \frac{2^k - 1}{2^{k+1}} = \frac{1}{2^{k+1}}$ comes from the disturbance besides the neighborhood of surfaces, and leads to new redundant surfaces during training. The bias c_m and c_n case inherent geometric error like redundant surfaces and floating noises in invisible space.

3. Proofs of Unbiased and Occlusion-aware properties of NeUDF

In this subsection we illustrate the capability of NeUDF for UDF learning from three aspects. First we show that different from NeuS, NeUDF avoids the c_m and c_n which cause the biased rendering color and inherent geometric error in UDF. Then we give the proofs of the unbiased and occlusion-aware properties of NeUDF respectively.

3.1. Avoidance of c_m and c_n in NeUDF.

Before providing the detailed proofs of the unbiased and occlusion-aware properties of NeUDF, we briefly show that NeUDF is free from the undesired colors c_m and c_n by introducing the new rendering weight function:

$$w_r(t) = \tau_r(t) e^{-\int_0^t \tau_r(u) du}, \quad (35)$$

$$\tau_r(t) = \left| \frac{\frac{\partial \varsigma_r \circ \Psi \circ p}{\partial t}(t)}{\varsigma_r \circ \Psi \circ p(t)} \right|, \quad (36)$$

where $\varsigma_r(d)$ satisfies that:

$$\varsigma_r(0) = 0, \lim_{d \rightarrow \infty} = 1, \quad (37)$$

$$\forall d > 0, \varsigma'_r(d) > 0, \varsigma''_r(d) < 0. \quad (38)$$

Similar to the derivation in 2, there is:

$$\lim_{r \rightarrow \infty} \int_0^{t_0^l} w_r(t) dt = 0 \quad (39)$$

and

$$\begin{aligned} & \lim_{r \rightarrow \infty} \int_{t_0^l}^{t_0^*} w_n(t) dt \\ &= \lim_{r \rightarrow \infty} e^{- \int_0^{t_0^l} \tau_r(u) du} (-e^{- \int_{t_0^l}^{t_0^*} \tau_r(u) du} + 1) \quad (40) \\ &= \lim_{r \rightarrow \infty} -e^{- \int_{t_0^l}^{t_0^*} \tau_r(u) du} + 1 \end{aligned}$$

When $t \in (t_0^l, t_0^*)$, there is:

$$\frac{\partial \Psi \circ p(t)}{\partial t} = -|\cos \theta| < 0 \quad (41)$$

We have:

$$\begin{aligned} & -e^{- \int_{t_0^l}^{t_0^*} \tau_r(u) du} + 1 \\ &= -e^{- \int_{t_0^l}^{t_0^*} \left| \frac{\partial \varsigma_r \circ \Psi \circ p(u)}{\partial u} \right| du} + 1 \\ &= -e^{- \int_{t_0^l}^{t_0^*} \left| \frac{\partial}{\partial u} \ln \varsigma_r \circ \Psi \circ p(u) \right| du} + 1 \\ &= -e^{ \int_{t_0^l}^{t_0^*} \frac{\partial}{\partial u} \ln \varsigma_r \circ \Psi \circ p(u) du} + 1 \quad (42) \\ &= -e^{\ln \varsigma_r \circ \Psi \circ p(t_0^*) - \ln \varsigma_r \circ \Psi \circ p(t_0^l)} + 1 \\ &= -\frac{\varsigma_r \circ \Psi \circ p(t_0^*)}{\varsigma_r \circ \Psi \circ p(t_0^l)} + 1 \\ &= -0 + 1 \\ &= 1 \end{aligned}$$

So we have:

$$\lim_{r \rightarrow \infty} \int_{t_0^l}^{t_0^*} w_n(t) dt = 1 \quad (43)$$

It follows that:

$$\begin{aligned} \lim_{r \rightarrow \infty} C(o, v) &= \lim_{r \rightarrow \infty} \int_{t_0^l}^{t_0^*} w_n(t) dt \cdot c(p(t_0^*), v) \\ &+ (1 - \lim_{r \rightarrow \infty} \int_{t_0^l}^{t_0^*} w_n(t) dt) \cdot c_m \quad (44) \\ &= c(p(t_0^*), v) \end{aligned}$$

It indicates that NeUDF avoids the limitation introduced by the undesired mixture c_m (and c_n). The detailed proof of unbiased property of NeUDF is provided in the next section.

3.2. Proof of Unbiased Property in NeUDF.

Intuitively, the rendering weight function should be unbiased, *i.e.*, more contribution should come from the intersection point than its neighbor. In this subsection we prove that NeUDF is unbiased:

- Given the ray $p(t)$ and the UDF $\Psi(x)$, the weight of rendering $w_r(t)$ in NeUDF attains a locally maximum value at a intersection point t^* .

Assume that the weight $w_r(t)$ is a linear function within the local neighborhood (t^l, t^r) of the zero point $t^* \in (t^l, t^r)$. We consider the intervals (t^l, t^*) and (t^*, t^r) respectively. For $t \in (t^l, t^*)$, we have:

$$\begin{aligned} w_r(t) &= \tau_r(t) e^{- \int_0^t \tau(u) du} \\ &= \tau_r(t) e^{- \int_0^{t^l} \tau(u) du} e^{- \int_{t^l}^t \tau(u) du} \\ &= \tau_r(t) e^{- \int_0^{t^l} \tau(u) du} e^{- \int_{t^l}^t \left| \frac{\partial \varsigma_r \circ \Psi \circ p(u)}{\partial u} \right| du} \\ &= \tau_r(t) e^{- \int_0^{t^l} \tau(u) du} e^{- \int_{t^l}^t \frac{\partial}{\partial u} \ln \varsigma_r \circ \Psi \circ p(u) du} \\ &= \tau_r(t) e^{- \int_0^{t^l} \tau(u) du} e^{\ln \varsigma_r \circ \Psi \circ p(t) - \ln \varsigma_r \circ \Psi \circ p(t^l)} \\ &= \tau_r(t) e^{- \int_0^{t^l} \tau(u) du} \frac{e^{\ln \varsigma_r \circ \Psi \circ p(t)}}{e^{\ln \varsigma_r \circ \Psi \circ p(t^l)}} \\ &= \tau_r(t) e^{- \int_0^{t^l} \tau(u) du} \frac{\varsigma_r \circ \Psi \circ p(t)}{\varsigma_r \circ \Psi \circ p(t^l)} \\ &= \left| \frac{\partial \varsigma_r \circ \Psi \circ p(u)}{\partial u} \right| \frac{e^{- \int_0^{t^l} \tau(u) du} \varsigma_r \circ \Psi \circ p(t)}{\varsigma_r \circ \Psi \circ p(t^l)} \\ &= \left| \frac{\partial \varsigma_r \circ \Psi \circ p(t)}{\partial \Psi \circ p(t)} \right| \cdot \left| \frac{\partial \Psi \circ p(t)}{\partial t} \right| e^{- \int_0^{t^l} \tau(u) du} \frac{\varsigma_r \circ \Psi \circ p(t)}{\varsigma_r \circ \Psi \circ p(t^l)} \\ &= \frac{|\varsigma_r' \circ \Psi \circ p(t)| \cdot |\cos \theta|}{|\varsigma_r \circ \Psi \circ p(t)|} e^{- \int_0^{t^l} \tau(u) du} \frac{\varsigma_r \circ \Psi \circ p(t)}{\varsigma_r \circ \Psi \circ p(t^l)} \\ &= \frac{\varsigma_r' \circ \Psi \circ p(t) \cdot |\cos \theta|}{\varsigma_r \circ \Psi \circ p(t)} e^{- \int_0^{t^l} \tau(u) du} \frac{\varsigma_r \circ \Psi \circ p(t)}{\varsigma_r \circ \Psi \circ p(t^l)} \\ &= \frac{\varsigma_r' \circ \Psi \circ p(t) \cdot |\cos \theta| \cdot e^{- \int_0^{t^l} \tau_r(u) du}}{\varsigma_r \circ \Psi \circ p(t^l)} \end{aligned} \quad (45)$$

For a given parameter r , $\varsigma_r \circ \Psi \circ p(t^l)$, $e^{- \int_0^{t^l} \tau_r(u) du}$ and $|\cos \theta|$ are all constant. So we have:

$$w_r(t) = A \cdot \varsigma_r' \circ \Psi \circ p(t), A = \frac{|\cos \theta| \cdot e^{- \int_0^{t^l} \tau_r(u) du}}{\varsigma_r \circ \Psi \circ p(t^l)}, \quad (46)$$

where A is a fixed positive number for any given r .

Note that $\varsigma_r'(d) > 0, \varsigma_r''(d) < 0$, it follows that:

$$w_r(t_1) > w_r(t_2), \forall t_1 > t_2, t_1, t_2 \in (t^l, t^*). \quad (47)$$

For $t \in (t^*, t^r)$, we have:

$$\tau_r(t) = \left| \frac{\frac{\partial \varsigma_r \circ \Psi \circ p}{\partial t}(t)}{\varsigma_r \circ \Psi \circ p(t)} \right| = \frac{|\varsigma'_r \circ \Psi \circ p(t)| \cdot |\cos \theta|}{\varsigma_r \circ \Psi \circ p(t)} \quad (48)$$

$\forall t_1 > t_2, t_1, t_2 \in (t^*, t^r)$, there is:

$$\tau_r(t_1) < \tau_r(t_2) \quad (49)$$

$$e^{-\int_0^{t_1} \tau_r(u) du} < e^{-\int_0^{t_2} \tau_r(u) du} \quad (50)$$

It follows that:

$$w_r(t_1) < w_r(t_2), \forall t_1 > t_2, t_1, t_2 \in (t^l, t^*). \quad (51)$$

The Equ. 47 and 51 indicates that the point closer to the zero point is with higher weight value. Note that the proof does not require a strict zero point t^* , *i.e.*, the property holds true when there is a small perturbation Δ to the zero point t^* : $\Psi \circ p(t^*) = \Delta > 0$.

Empirically, the zero point of the UDF is encoded as a small positive number, so the weight function $w_r(t)$ is continuous along the ray. Therefore we have:

$$w_r(t^*) > w_r(t), \forall t \in (t^l, t^r), t \neq t^* \quad (52)$$

This completes the proof.

3.3. Proof of Occlusion-aware Property in NeUDF.

In this subsection we prove that NeUDF is occlusion-aware. Intuitively, for two parts of the sample ray with the same UDF value, we hope that more contribution of the output colors is from the part closer to the camera. That is, the closer surfaces are more likely to have higher weight.

Specifically, given two surfaces S_1 and S_2 such that S_1 is closer to the camera, for two corresponding points $p(t_1)$ and $p(t_2)$ with the same UDF value, we have:

$$\int_{t_1}^{t_1+\delta} w_r(t) dd_1(t) > \int_{t_2}^{t_2+\delta} w_r(t) dd_2(t), \quad (53)$$

where $d_i(t)$ denotes the distance between the location $p(t)$ and the surface S_i , and δ denotes the small step length.

$$\tau_r(t) = \left| \frac{\frac{\partial \varsigma_r \circ \Psi \circ p}{\partial t}(t)}{\varsigma_r \circ \Psi \circ p(t)} \right| = \frac{|\varsigma'_r \circ \Psi \circ p(t)| \cdot |\cos \theta|}{\varsigma_r \circ \Psi \circ p(t)} \quad (54)$$

For $t_1 < t_2, \Psi(t_1) = \Psi(t_2), w_r(t_1), w_r(t_2) > 0$, we have:

$$\frac{\tau_r(t_1)}{|\cos \theta_1|} = \frac{|\varsigma'_r \circ \Psi \circ p(t_1)|}{\varsigma_r \circ \Psi \circ p(t_1)} = \frac{|\varsigma'_r \circ \Psi \circ p(t_2)|}{\varsigma_r \circ \Psi \circ p(t_2)} = \frac{\tau_r(t_2)}{|\cos \theta_2|} \quad (55)$$

$$e^{-\int_0^{t_1} \tau_r(u) du} > e^{-\int_0^{t_2} \tau_r(u) du} \quad (56)$$

There is:

$$\begin{aligned} \frac{w_r(t_1)}{|\cos \theta|} &= \frac{\tau_r(t_1) e^{-\int_0^{t_1} \tau_r(t_u) du}}{|\cos \theta|} \\ &> \frac{\tau_r(t_2) e^{-\int_0^{t_2} \tau_r(t_u) du}}{|\cos \theta|} = \frac{w_r(t_2)}{|\cos \theta|} \end{aligned} \quad (57)$$

It follows that:

$$\int_{t_1}^{t_1+\delta} w_r(t) dd_1(t) = \int_{t_1}^{t_1+\delta} \frac{w_r(t)}{|\cos \theta|} dt \quad (58)$$

$$\int_{t_2}^{t_2+\delta} w_r(t) dd_2(t) = \int_{t_2}^{t_2+\delta} \frac{w_r(t)}{|\cos \theta|} dt \quad (59)$$

$$\int_{t_1}^{t_1+\delta} w_r(t) dd_1(t) > \int_{t_2}^{t_2+\delta} w_r(t) dd_2(t), \quad (60)$$

where $d_i(t)$ denotes the distance between the location $p(t_i)$ and the surface S_i .

The Equ. 60 indicates that the cumulative weight near the the first intersected surface are higher than the second one. This means that more concentration are on the former surface. Note that no prior assumption of the existence of other intersected surfaces is required, *i.e.*, the property of occlusion-aware holds true for more than two surface intersections along the ray. This completes the proof of the occlusion-aware property.

4. Implementation Details

4.1. Network Architecture

Similar to IDR [5] and NeuS [5], we use two MLP networks to respectively encode the UDF and the color. The input of the UDF network is the spatial location $p(t)$ and the output is the corresponding UDF value along with a 256-dimensional feature vector. The UDF network $\Psi(x)$ consists of 8 hidden layers with hidden size of 256, and the activation function is chosen as the Softplus with $\beta = 100$ for all hidden layers and the output layer. A skip connection is also used to connect the input with the output of the fourth layer. The inputs of the color network are the spatial location $p(t)$, the view direction v , the gradient n of the UDF network at the spatial location $p(t)$ and the corresponding feature vector derived by the UDF network. The color network $c(x, v)$ consists of 4 hidden layers with hidden size of 256. Normal regularization is applied before the gradient n of the UDF network is used as the input of the color network. Same positional encoding and weight normalization are adopted as in Neus.

4.2. Training Details

Discretization. We adopt the α -compositing to discretize the weight function, which divides the sample ray into bins by sampling n points $p(t_i) = o + t_i | i = 1, \dots, n, t_i < t_{i+1}$ and accumulate colors within each bin according to the weight integral:

$$\begin{aligned} \alpha_i &= 1 - e^{-\int_{t_i}^{t_{i+1}} \tau_r(t) dt} \\ &= \frac{|\zeta_r \circ \Psi \circ p(t_i) - \zeta_r \circ \Psi \circ p(t_{i+1})|}{\zeta_r \circ \Psi \circ p(t_i)}. \end{aligned} \quad (61)$$

We slightly modify Equ. 61 by:

$$\alpha_i = \frac{\zeta_i^{max} - \zeta_i^{min}}{\zeta_i^{max}}, \quad (62)$$

where ζ_i^{max} and ζ_i^{min} is the maximum and minimum of the set $\{\zeta_r \circ \Psi \circ p(t_i), \zeta_r \circ \Psi \circ p(t_{i+1})\}$.

Up Sampling. We first formally sample 64 points per ray, and then hierarchically conduct importance sampling on top of the sampling weight $w_s(t)$ for another 64 points:

$$w_s(t) = \tau_s(t) e^{-\int_0^t \tau_s(u) du}, \tau_s(t) = \zeta_s \circ \Psi \circ p(t) \quad (63)$$

And $\zeta_s(\cdot)$ satisfies the rules: $\zeta_s(d) > 0$ and $\zeta'_s(d) < 0, \forall d > 0$. Intuitively, the $\tau_s(t)$ derived by the monotonically decreasing function is a view-invariant sampling density, and the density has positive correlation with the UDF value. To derive the sampling weight $w_s(t)$, the classical volume rendering scheme is applied.

The weight of the i^{th} sample point $w_s(t_i)$ is slightly modified by:

$$w'_s(t_i) = \max\{w_s(t_{i+k}), k = -1, 0, 1\} \quad (64)$$

And then the weight $w'_s(t)$ is normalized so that the integral equals to one:

$$w''_s(t) = \frac{w'_s(t)}{\sum_{i=0}^{n-1} w'_s(t_i)} \quad (65)$$

For each iteration we hierarchically conduct the importance sampling for two times, and each time 32 points are sampled. The total number of sampling points are 128. If no masks are provided, 32 points are randomly sampled in addition outside the unit sphere per ray to represent the outside scene. The outside scene is represented with NeRF++ [7], as used in NeuS [5].

Platform. The network is trained with ADAM optimizer, and the learning rate warms up to 2×10^{-4} in the first 5k iterations, and decreases to 1×10^{-5} by the end of training. For each iteration, 512 random rays are sampled from 8 input camera poses randomly selected. We train each model for 400k iterations in total for 9 hours for the setting of with mask, and 11 hours for the setting of without mask on a single Nvidia 3090 GPU.

4.3. Data Preparation

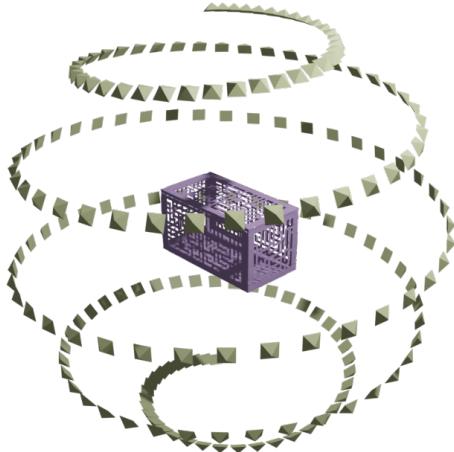


Figure 1. Poses of the camera. The camera poses are represented as the yellow pyramid, and the object to reconstruct is represented in purple.

Rendered Data. To generate the customized data, we use the pyrender package to render images from the ground-truth objects. We rendered 200 views at 800×800 pixels for each textured mesh or colored point cloud. Fig 1 visualizes the camera poses. Corresponding masks with black background are provided optionally. Only the rendered images and the masks are used as inputs of the network.

Captured Data. We additionally captured several real-world objects using the mobile phone. The captured images are extracted from the captured videos around the object. For the book object we captured 200 images at the resolution of 1920×1440 . For the fan object we captured 59 images at the resolution of 3456×4608 . For the book object we captured 200 images at the resolution of 720×1280 . All the camera poses are estimated by COLMAP [3, 4] and no masks are provided.

5. Additional Results

We visualize more reconstruction results of NeUDF on DF3D [8], MGN [1], DTU [2], BMVS [6] datasets and real-captured data. Fig. 2 shows the comparison with NeuS on the DF3D dataset without mask supervision. Fig. 3 shows the comparison with NeuS on the DF3D dataset with mask supervision. Fig. 4 shows the comparison with NeuS on the MGN dataset without mask supervision. Fig. 5 shows the comparison with NeuS on the MGN dataset with mask supervision. Fig. 6 shows the comparison with NeuS on the DTU and BMVS datasets with mask supervision. Fig. 7 shows the additional results of the real-captured scenes with open surfaces.

References

[1] Bharat Bhatnagar, Garvita Tiwari, Christian Theobalt, and Gerard Pons-Moll. Multi-garment net: Learning to dress 3d people from images. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5419–5429, 2019. 8, 11, 12

[2] Rasmus Jensen, Anders Dahl, George Vogiatzis, Engil Tola, and Henrik Aanaes. Large scale multi-view stereopsis evaluation. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 406–413. IEEE, 2014. 8, 13

[3] Johannes L. Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 4104–4113. IEEE Computer Society, 2016. 7

[4] Johannes L. Schönberger, Enliang Zheng, Jan-Michael Frahm, and Marc Pollefeys. Pixelwise view selection for unstructured multi-view stereo. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III*, volume 9907 of *Lecture Notes in Computer Science*, pages 501–518. Springer, 2016. 7

[5] 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 Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan, editors, *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pages 27171–27183, 2021. 6, 7

[6] 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. *Computer Vision and Pattern Recognition (CVPR)*, 2020. 8, 13

[7] Kai Zhang, Gernot Riegler, Noah Snavely, and Vladlen Koltun. Nerf++: Analyzing and improving neural radiance fields. *CoRR*, abs/2010.07492, 2020. 7

[8] Heming Zhu, Yu Cao, Hang Jin, Weikai Chen, Dong Du, Zhangye Wang, Shuguang Cui, and Xiaoguang Han. Deep fashion3d: A dataset and benchmark for 3d garment reconstruction from single images. In *European Conference on Computer Vision*, pages 512–530. Springer, 2020. 8, 9, 10



Figure 2. Additional results on the DF3D [8] dataset without mask supervision.



Figure 3. Additional results on the DF3D [8] dataset with mask supervision.

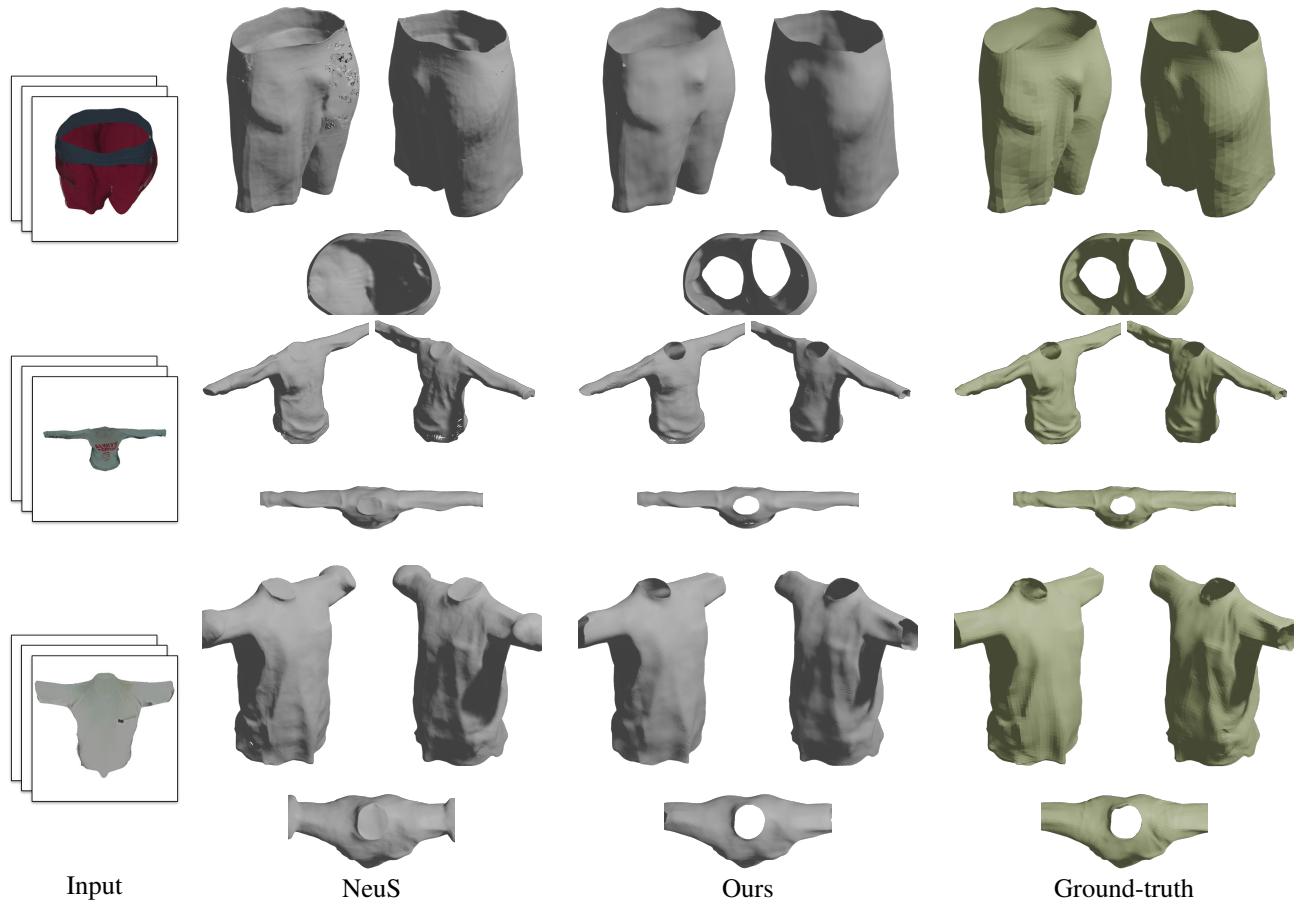


Figure 4. Additional results on the MGN [1] dataset without mask supervision.

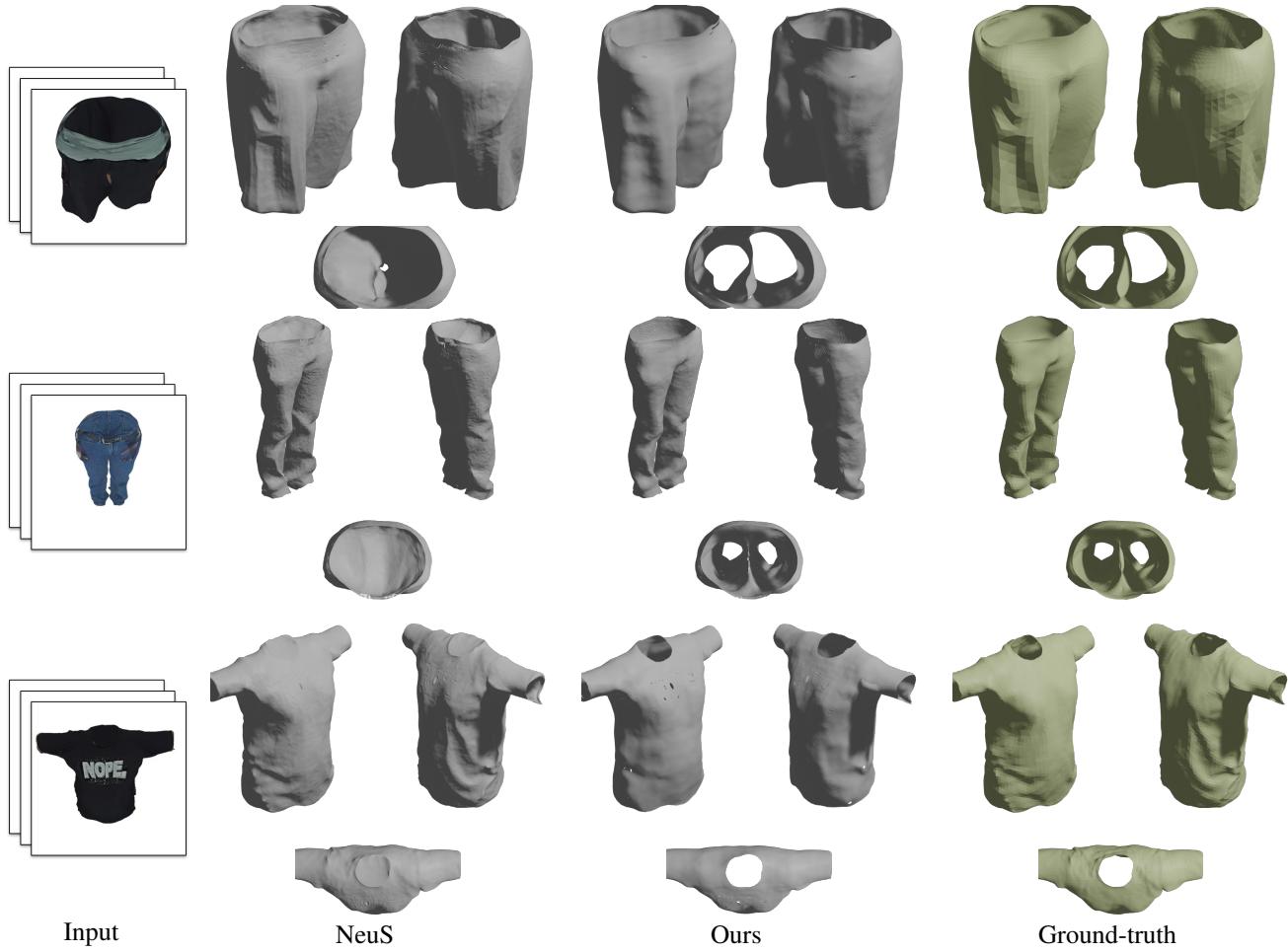


Figure 5. Additional results on the MGN [1] dataset with mask supervision.

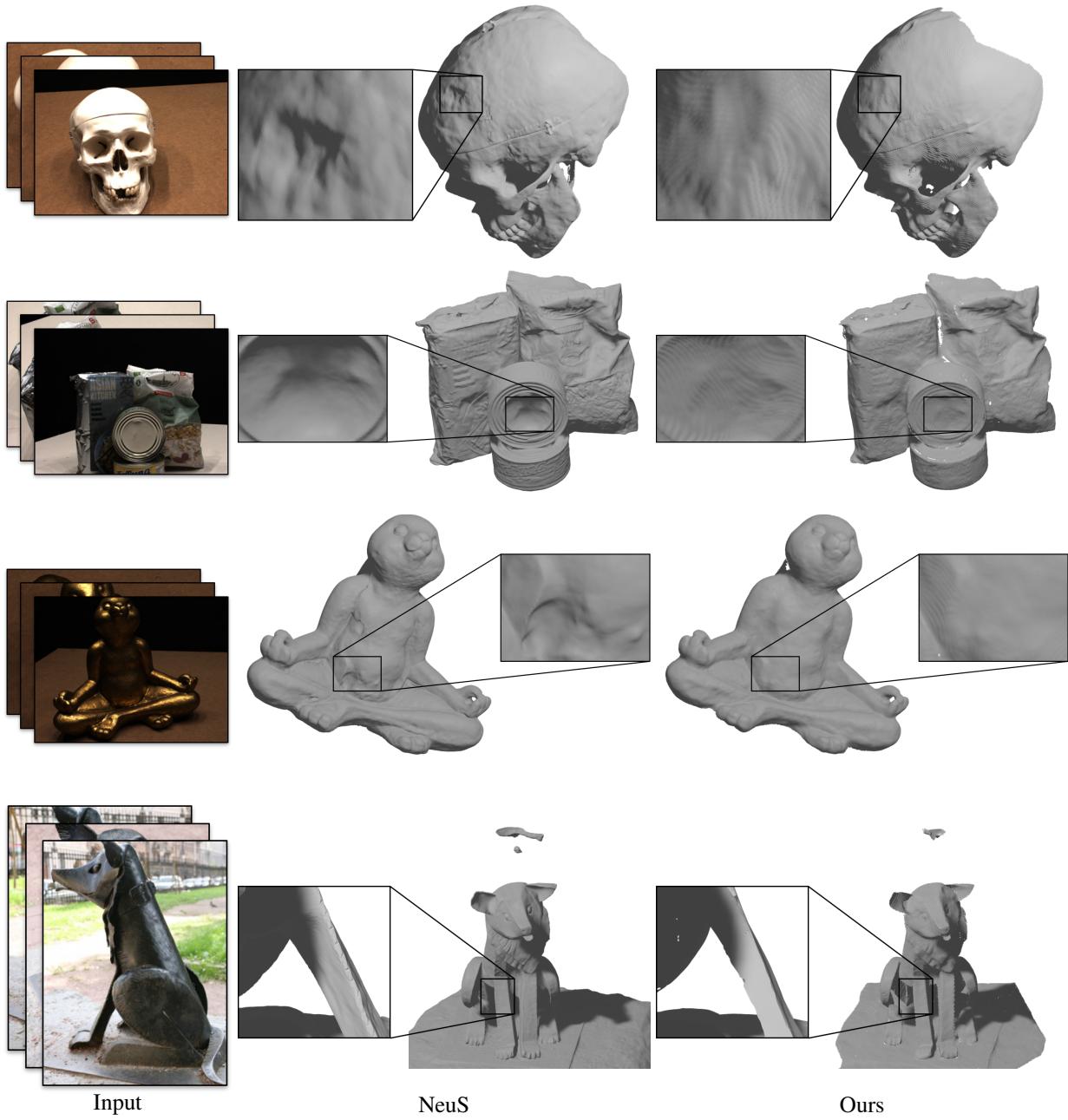


Figure 6. Additional results on the DTU [2] dataset (the first three scenes) and BMVS [6] dataset (the last one scene) with mask supervision.

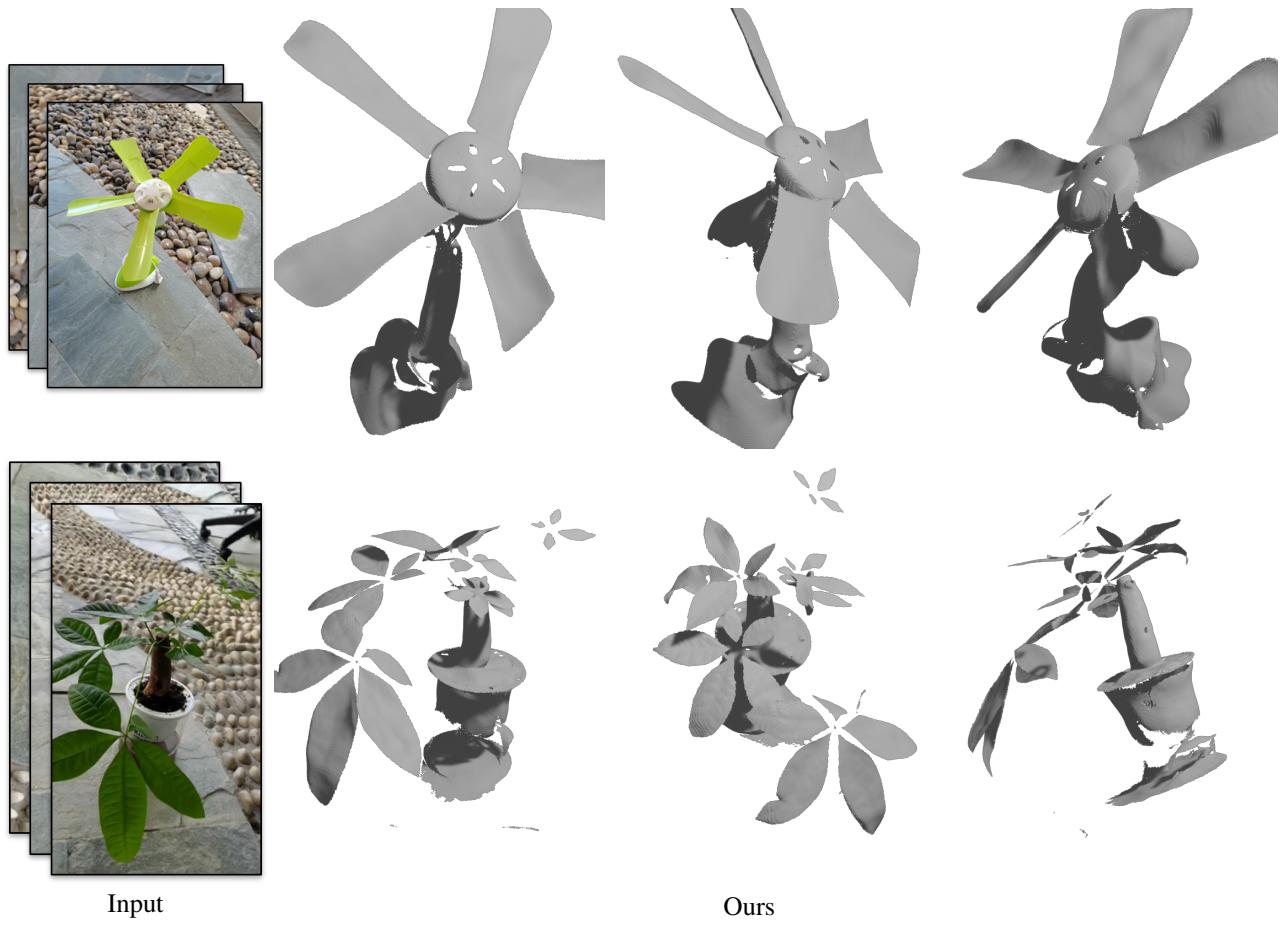


Figure 7. Additional results of the real-captured data without mask supervision.