Neural Density-Distance Fields -Appendix-

Itsuki Ueda[1][0000−0002−2681−1229], Yoshihiro Fukuhara[2][0000−0001−8892−5339],
Hirokatsu Kataoka[3][0000−0001−8844−165X], Hiroaki Aizawa[4][0000−0002−6241−3973],
Hidehiko Shishido[1][0000−0001−8575−0617], and
Itaru Kitahara[1][0000−0002−5186−789X]

[1] University of Tsukuba, 3058577, Japan
{ueda.itsuki,shishido.hidehiko,kitahara.itaru}@image.iit.tsukuba.ac.jp
[2] Waseda University, 1698050, Japan f_yoshi@ruri.waseda.jp
[3] National Institute of Advanced Industrial Science and Technology (AIST),
3058560, Japan hirokatsu.kataoka@aist.go.jp
[4] Hiroshima University, 7398511, Japan hiroaki-aizawa@hiroshima-u.ac.jp

1 Architecture details

Fig. 1 shows the details of our architecture. The network uses multilayer perceptrons (MLPs) with a width of 256, the same as in NeRF [2] and NeuS [3]. The network calculates the density using the first derivative value for the output distance D. This calculation requires careful setup of the Positional Encoding (PE) and activation functions. Using an objective function for the density field requires gradients up to the second derivative for activation functions. The architecture uses tanhExp [1] as the activation function whose second derivative is continuous. In the conventional method [2], the PE up to L dimensions utilizes values such as the following equation:

$$\gamma(\mathbf{p}) = \left[\sin(\mathbf{p}), \cos(\mathbf{p}), \cdots, \sin(2^{L-1}\mathbf{p}), \cos(2^{L-1}\mathbf{p})\right]^T . \tag{1}$$

It is the concatenation of the sin and cos of each dimension of position $\mathbf{p}$ scaled by powers of 2 from 1 to $2L-1$. This PE amplifies by the frequency in the hierarchy of first derivatives, which emphasizes the high-frequency elements in the density field. In other words, the maximum and minimum frequency components have a scale difference of 2^{L-1} in their influence on the density derivative, thus making the learning process unstable. Therefore, our architecture damps the high-frequency element so that the scale in the single-differentiation hierarchy is the same as the original PE, as shown in the following equation:

$$\gamma'(\mathbf{p}) = \left[\sin(\mathbf{p}), \cos(\mathbf{p}), \cdots, \frac{1}{2^{L-1}}\sin(2^{L-1}\mathbf{p}), \frac{1}{2^{L-1}}\cos(2^{L-1}\mathbf{p})\right]^T . \tag{2}$$

However, since the PE neglects the high-frequency component in the non-derivative hierarchy, we need to add an intermediate input of the conventional $\gamma(\mathbf{p})$ in the layers after the distance output for learning detailed color fields. Note that the performance of restoring the high-frequency component of the color field is worse than that of NeRF for the same network size.

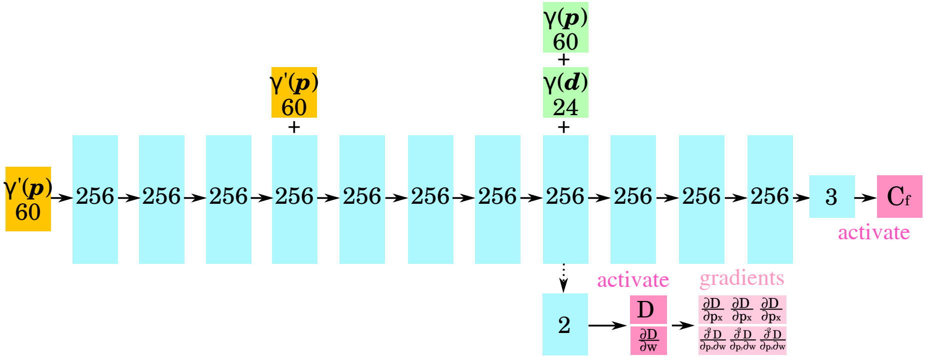

Fig. 1. Details of network architecture in NeDDF

2 Deriving the conversion from distance to density

This section describes the derivation details of Equation 10 in the main paper. For the distance field around position $\mathbf{p} \in \mathbb{R}^3$, we consider $D(\mathbf{r}(t)), \mathbf{r}(t) = \mathbf{p} + t\mathbf{v}$, which is sliced in the gradient direction $\mathbf{v}$. Calculating the derivative of the distance field in the direction of the gradient, $\frac{\partial D}{\partial t}$, we can derive an expression for σ as follows:

$$\frac{\partial D(\mathbf{r}(t))}{\partial t}|_{t=0} \tag{3}$$

$$= \lim_{\Delta t \to 0} \frac{d(\mathbf{r}(\Delta t), \mathbf{v}) - d(\mathbf{r}(0), \mathbf{v})}{\Delta t}. \tag{4}$$

The first term of Equation 4 can be deformed as follows:

$$d(\mathbf{r}(\Delta t), \mathbf{v}) \tag{5}$$

$$= \int_{t_n}^{t_f} \exp(-\int_{t_n}^{t} \sigma(\mathbf{r}(s + \Delta t)ds)\sigma(\mathbf{r}(t + \Delta t)tdt \tag{6}$$

$$= \int_{t_n}^{t_f} \exp(-\int_{t_n+\Delta t}^{t+\Delta t} \sigma(\mathbf{r}(s)ds)\sigma(\mathbf{r}(t + \Delta t)tdt. \tag{7}$$

We set $S(t_n, t) := \exp(-\int_{t_n}^{t} \sigma(\mathbf{r}(s))ds)\sigma(\mathbf{r}(t))dt$. When t_f takes a sufficiently large value until all of the light gets reflected, the Equations 8, 9, and 10 are valid.

$$T(t_f) = 0 \tag{8}$$

$$S(t_n, t_f) = 0 \tag{9}$$

$$\int_{t_n}^{t_f} S(t_n, t_f) = 1 \tag{10}$$

With Equations 8, 9, and 10, $d(\mathbf{r}(\Delta t), \mathbf{v})$ can be deformed as follows:

$$\begin{aligned}
& d(\mathbf{r}(\Delta t), \mathbf{v}) && (11)\\
&= \int_{t_n}^{t_f} S(t_n + \Delta t, t + \Delta t) t dt && (12)\\
&= \int_{t_n}^{t_f} S(t_n + \Delta t, t + \Delta t)(t + \Delta t) dt - \int_{t_n}^{t_f} S(t_n + \Delta t, t + \Delta t) \Delta t dt && (13)\\
&= \int_{t_n+\Delta t}^{t_f+\Delta t} S(t_n + \Delta t, t) t dt - \Delta t \int_{t_n}^{t_f} S(t_n + \Delta t, t + \Delta t) dt && (14)\\
&= \int_{t_n}^{t_f} S(t_n + \Delta t, t) t dt + \int_{t_f}^{t_f+\Delta t} S(t_n + \Delta t, t) t dt \\
&\qquad - \int_{t_n}^{t_n+\Delta t} S(t_n + \Delta t, t) t dt - \Delta t. && (15)
\end{aligned}$$

We calculate the first term of Equation 15 as follows:

$$\begin{aligned}
& \int_{t_n}^{t_f} S(t_n + \Delta t, t) t dt && (16)\\
&= \int_{t_n}^{t_f} \exp(- \int_{t_n+\Delta t}^{t} \sigma(\mathbf{r}(s)) ds) \sigma(\mathbf{r}(t)) t dt && (17)\\
&= \int_{t_n}^{t_f} T(t) T(t_n + \Delta t) \sigma(t_n + \mathbf{r}(t)) t dt && (18)\\
&= T(t_n + \Delta t) d(\mathbf{r}(0), \mathbf{v}). && (19)
\end{aligned}$$

Using Equation 9, the second term of Equation 15 is equal to 0. The third term of Equation 15 is calculated as follows:

$$\begin{aligned}
& \int_{t_n}^{t_n+\Delta t} S(t_n + \Delta t, t) t dt && (20)\\
&= T(t_n + \Delta t) \int_{t_n}^{t_n+\Delta t} T(t) \sigma(\mathbf{r}(t)) t dt. && (21)
\end{aligned}$$

Same as in the main paper, we assume that t_n is small enough to be valid at $T(t_n) = 1$. Equation 21 converges to $\Delta t \sigma(\mathbf{r}(t_n)) t_n$ as $\Delta t \to 0$. Therefore, the

Equation 4 can be deformed as follows:

$$
\begin{aligned}
&\frac{\partial D(\mathbf{r}(t))}{\partial t}|_{t=0} && (22)\\
&= \lim_{\Delta t \to 0} \frac{d(\mathbf{r}(\Delta t), \mathbf{v}) - d(\mathbf{r}(0), \mathbf{v})}{\Delta t} && (23)\\
&= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \left[T(t_n + \Delta t) d(\mathbf{r}(0), \mathbf{v}) - \Delta t \sigma(\mathbf{r}(t)) t_n - \Delta t - d(\mathbf{r}(0), \mathbf{v}) \right] && (24)\\
&= \lim_{\Delta t \to 0} \left[\frac{T(t_n + \Delta t) - 1}{\Delta t} d(\mathbf{r}(0), \mathbf{v}) - \sigma(\mathbf{r}(t_n)) t_n - 1 \right] && (25)\\
&= \lim_{\Delta t \to 0} \left[\frac{\exp(\Delta t \sigma(\mathbf{r}(t_n)) - \exp(0)}{\Delta t} d(\mathbf{r}(0), \mathbf{v}) - \sigma(\mathbf{r}(t_n)) t_n - 1 \right] && (26)\\
&= \sigma(\mathbf{r}(t_n)) d(\mathbf{r}(0), \mathbf{v}) - \sigma(\mathbf{r}(t_n)) t_n - 1 && (27)\\
&= -1 + \left(D(\mathbf{r}(0)) - t_n \right) \sigma(\mathbf{r}(t_n)). && (28)
\end{aligned}
$$

3 Parameter selection for the shape of the auxiliary gradient

Equation 15 in the main paper is a penalty term that constrains the shape of the auxiliary gradient by the hyperparameter α. Fig. 2 shows that the auxiliary gradient becomes active in a narrower range than the distance field when $\alpha \leq 1$, and the shape becomes more concentrated near the cusps as α is larger. This constraint leads to a unique set of auxiliary gradients.

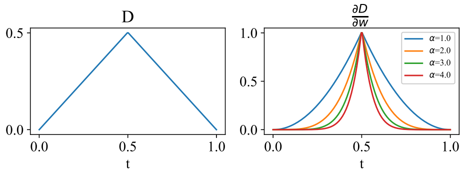

Fig. 2. Shape of auxiliary gradient for each α

Fig. 3 is a colorized visualization of the distance field, density field, and auxiliary gradient in the 2D slice. We can see that the auxiliary gradient becomes strongly activated near the cusp of the distance field, where the distances from several objects are similar. Fig. 4 shows the difference in rendering results with and without auxiliary gradients. Without auxiliary gradients, incorrect volume densities occur in the empty region.

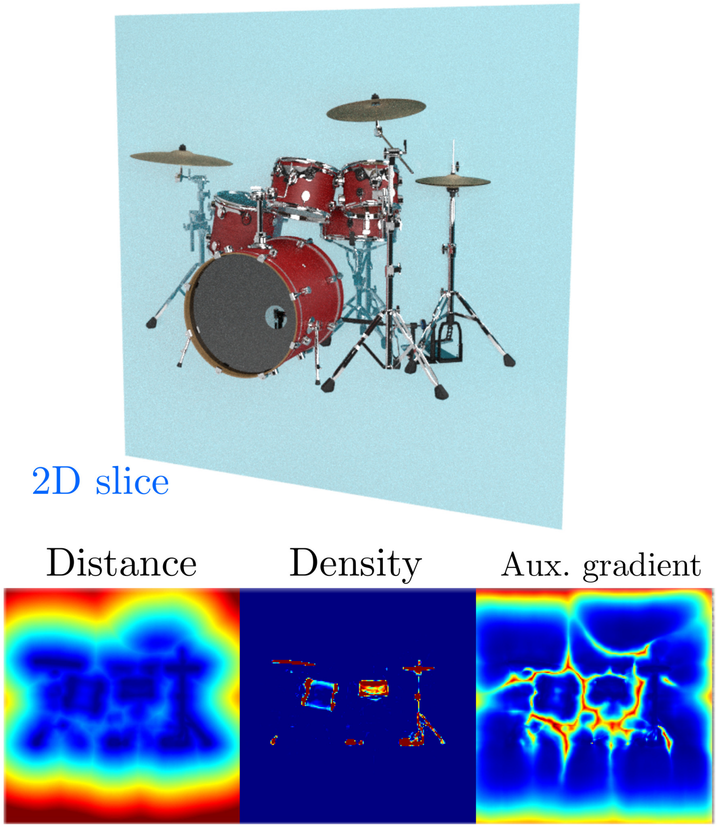

Fig. 3. Visualization of the 2D slice for distance, density, and auxiliary gradient.

4 Evaluation of reconstruction performance in smoke scenes

Most of the previous methods do not provide benchmarks for scenes with subjects containing smoke. We produce a synthetic dataset for a smoke-subject scene where the density varies over a wide area, and we qualitatively evaluate the proposed method's performance. As with the nerf synthetic dataset [2], the dataset consists of 100 viewpoints each in a hemispherical plane for train/valid data and 200 viewpoints in orbit for test data. For each shot, we record RGB and Transmittance information at a resolution of 800×800.

Fig. 5 shows the rendered image from the test viewpoint. Our method achieves high-quality Novel View Synthesis even in smoke-like scenes. Fig. 6 also shows the visualization results of the slices for the distance field, density field, and auxiliary gradient. Since the distance field shows that the distance increases again

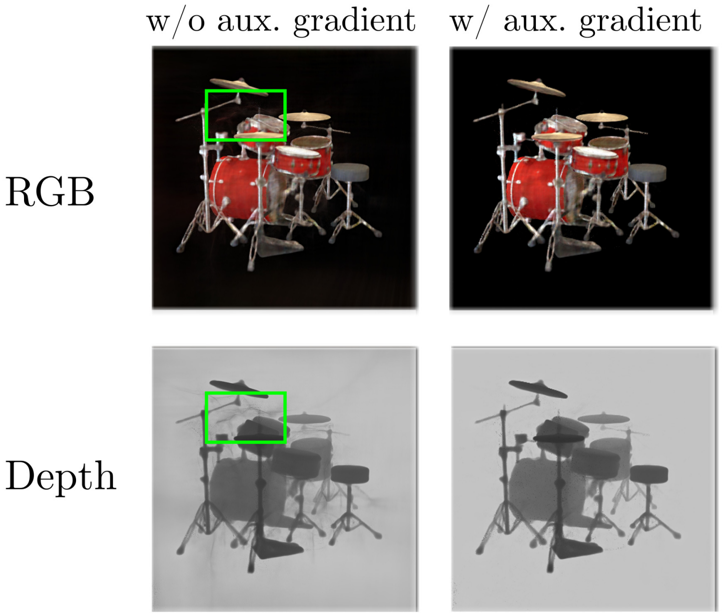

Fig. 4. The difference in rendering results with and without auxiliary gradients.

inside the object, the field's estimation also holds true inside the object. In addition, the minima of the distance field are larger than those in Fig. 3, indicating that the assumption of expressing low density with large minima of the distance field works reasonably. The density field actually expresses a translucent state rather than a bipolar one.

5 Evaluation of localization performance in other scenes

For other scenes in the NeRF synthetic dataset [2], we verify the camera localization performance in the same way as in experiment (b) of main paper. Fig. 7 plots the number of camera postures for which the position and angular errors are lower than the threshold values for each scene. In all cases, the use of the reprojection error improves performance more than the use of the photometric error alone, as in iNeRF[4]. Even in cases such as Drums and Ficus scenes, where the optimization result with only the reprojection error is worse than the initial value, we can see an improvement of performance due to the increase in the common area of the field of view. On the other hand, in scenes where the uniqueness of color information is not sufficient, such as the Materials scene, reprojection error may degrade performance due to mismatches between corresponding points. In the case where many local solutions for photometric errors exist, such as Mic and Ship scenes, the use of reprojection error does not avoid

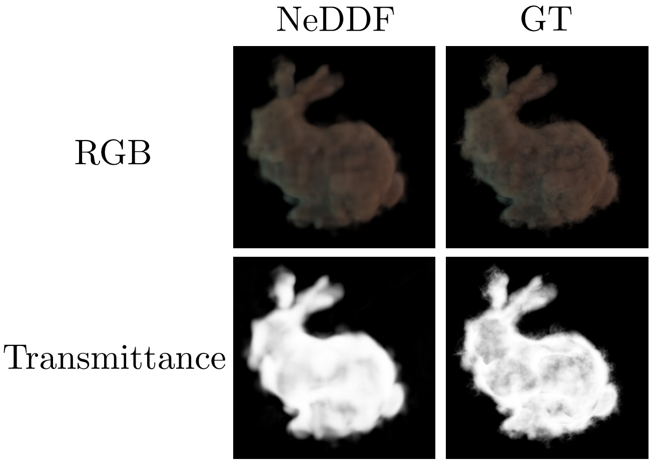

Fig. 5. Rendering results in smoke scene

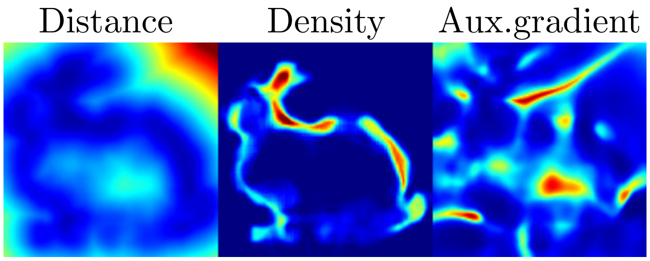

Fig. 6. Visualization of the 2D slice for distance, density, and auxiliary gradient in smoke scene.

the local solutions and does not improve the performance. We believe that we can improve such scenes by propagating unique features other than color to the empty regions in the same way as color fields.

References

1. Liu, X., Di, X.: Tanhexp: A smooth activation function with high convergence speed for lightweight neural networks. arXiv preprint arXiv:2003.09855 (2020)
2. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 405–421 (2020)
3. Wang, P., Liu, L., Liu, Y., Theobalt, C., Komura, T., Wang, W.: NeuS: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. In: Advances in Neural Information Processing Systems (NeurIPS) (2021)

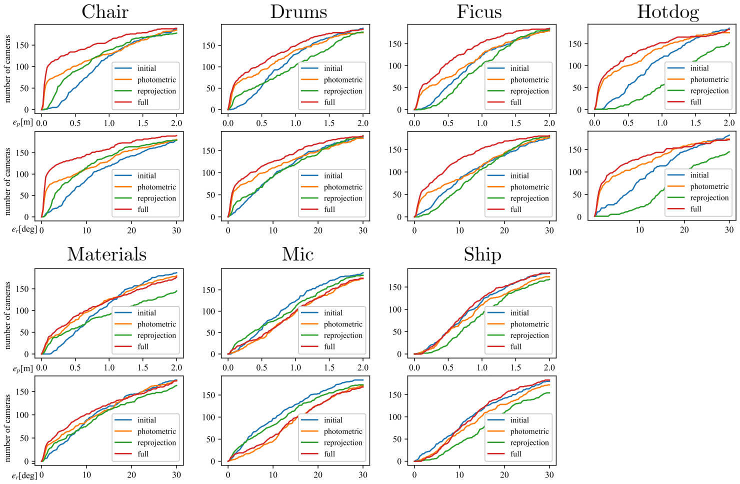

Fig. 7. Quantitative evaluation of camera poses estimation accuracy in other scenes. The horizontal axis represents the position and angle error, and the vertical axis represents the number of cameras recovered under the errors.

4. Yen-Chen, L., Florence, P., Barron, J.T., Rodriguez, A., Isola, P., Lin, T.Y.: iNeRF: Inverting neural radiance fields for pose estimation. In: Proceedings of the IEEE/RSJ Conference on Intelligent Robots and Systems (IROS). pp. 1323–1330 (2021)