

Evaluate Geometry of Radiance Field with Low-frequency Color Prior

Qihang Fang^{12*}, Yafei Song^{2*}, Keqiang Li¹², Li Shen², Huaiyu Wu¹, Gang Xiong¹ and Liefeng Bo²

¹Institute of Automation, Chinese Academy of Sciences

²XR Lab, DAMO Academy, Alibaba Group

{fangqihang2020, likeqiang2020, huaiyu.wu, gang.xiong}@ia.ac.cn

{huaihang.syf, jinyan.sl, liefeng.bo}@alibaba-inc.com

Abstract

Radiance field is an effective representation of 3D scenes, which has been widely adopted in novel-view synthesis and 3D reconstruction. It is still an open and challenging problem to evaluate the geometry, i.e., the density field, as the ground-truth is almost impossible to be obtained. One alternative indirect solution is to transform the density field into a point-cloud and compute its Chamfer Distance with the scanned ground-truth. However, many widely-used datasets have no point-cloud ground-truth since the scanning process along with the equipment is expensive and complicated. To this end, we propose a novel metric, named Inverse Mean Residual Color (IMRC), which can evaluate the geometry only with the observation images. Our key insight is that the better the geometry is, the lower-frequency the computed color field is. From this insight, given reconstructed density field and the observation images, we design a closed-form method to approximate the color field with low-frequency spherical harmonics and compute the inverse mean residual color. Then the higher the IMRC, the better the geometry. Qualitative and quantitative experimental results verify the effectiveness of our proposed IMRC metric. We also benchmark several state-of-the-art methods using IMRC to promote future related research. This project is released at <https://github.com/qihangGH/IMRC>.

1. Introduction

Radiance field has been popular for representing 3D objects or scenes since the Neural Radiance Field (NeRF) [21] demonstrated that the performance of novel-view synthesis could be dramatically improved benefiting from it. From this path-breaking work, many works aimed at improving NeRF from several aspects, such as various challenge scenarios [47, 17, 26, 2, 19], inference speed [43, 12, 28, 9, 7], training efficiency [8, 13, 42, 32, 33, 22, 5], generalization

*Equal contribution. Qihang and Keqiang were interns at Alibaba Group.

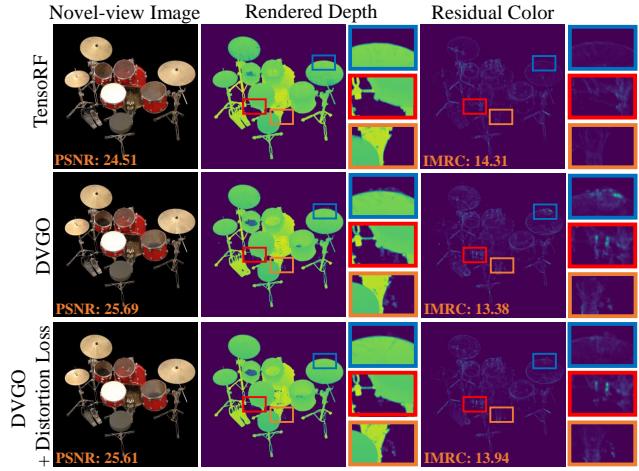


Figure 1. One example of novel-view images, rendered depth according to the reconstructed density fields, and residual color of TensoRF [5], DVGO [32], and DVGO [32] + Distortion Loss [3]. From top two rows, though DVGO [32] achieves a better PSNR↑, its geometry is qualitatively worse than TensoRF [5]. From the bottom row, distortion loss [3] could be qualitatively verified to improve the geometry. Our IMRC↑ could quantitatively evaluate these results correctly.

ability [44, 37, 6], and so on. Some other works proposed restoring the surfaces via reconstructing the latent radiance field by inverse volume rendering [36, 16, 25, 34, 45, 38].

As the ground-truth of a radiance field is hard to be obtained, we could not directly evaluate the reconstructed result. Alternatively, novel-view synthesis methods usually measure the similarity between the rendered image and its observed counterpart using image quality metrics such as PSNR. As shown in the top two rows of Fig. 1, these metrics maybe enough to evaluate the synthetic images but not appropriate to evaluate the geometry.

Besides novel-view synthesis methods, some algorithms aim at improving the reconstructed geometry, e.g., distortion loss [3], and a lot of works exploit radiance field for 3D reconstruction [36, 16, 25, 34, 45, 46, 40, 38]. These methods need a proper metric to quantitatively evaluate the

geometry result. To this end, surface reconstruction methods, *e.g.*, [36, 16], usually transform the density field into a point-cloud using the marching cubes algorithm and compute Chamfer Distance (CD) with the scanned ground-truth. However, the scanning process along with the equipment is expensive and complicated. Therefore, only a small number of datasets have this ground-truth, and many datasets, even widely-used, do not have one.

To alleviate difficulties above, our key observation is that the color of any point in an ideal radiance field tends to be low-frequency if its density is large. This phenomenon is named as low-frequency color prior. Beside this, we further design a closed-form method to compute the color field given the observation images and the density field. We name the result as computed color field to distinguish it from the reconstructed one by the radiance field reconstruction method. Via exploring the computed color field, we find that the low-frequency color prior may be invalid for the points with inaccurate density values. Therefore, the density field could be evaluated by the mean frequency of the computed color field.

However, it is difficult to directly evaluate the color frequency even for a single point, not to mention the whole field. To this end, we approximate the color with low-frequency spherical harmonics and compute the residual color. The smaller the mean residual color (MRC), the lower the color frequency, the better the geometry. Moreover, as the MRC is usually very small, it is not convenient for presentation and comparison. Therefore, we transform it into dB as PSNR does and name it as inverse mean residual color (IMRC). Then the higher the IMRC, the better the geometry. As demonstrated in Fig. 1, our IMRC metric could quantitatively evaluate the geometry correctly.

Our main contributions could be concluded as:

1. We present the low-frequency color prior via analysing the ideal radiance field. And we further find that this prior may be invalid for the computed color field if the density field is inaccurate.
2. To quantitatively evaluate the prior, we propose to approximate the color with low-frequency spherical harmonics and adopt the inverse mean residual color. Qualitative and quantitative experimental results verify its effectiveness.
3. We further benchmark several state-of-the-art radiance field reconstruction methods using inverse mean residual color to promote future related research.

2. Related Work

We firstly review two types of radiance field reconstruction methods, which aim at novel view synthesis and surface reconstruction respectively, and then discuss the geometry metrics which may inspire our work.

Novel View Synthesis. Mildenhall *et al.* [21] proposed

NeRF to synthesize novel-view images from posed images and dramatically improved the performance. They adopted a multilayer perceptron (MLP) to present the radiance field, which contains two components, *i.e.*, the density field and the color field. Each point in density field has a scalar controlling how much the color is accumulated. Each point in color field encodes the view-dependent color. The image can be rendered using the volume rendering algorithm [18]. From this path-breaking work, there have been many works aiming at improving NeRF from several aspects, such as various challenge scenarios [47, 17, 2, 26, 14, 19, 35, 39, 48, 29, 24], inference speed [43, 12, 28, 9, 7], training efficiency [8, 13, 42, 32, 33, 22, 5], generalization ability [44, 37, 6], and so on.

During this time, some works, *e.g.*, [42, 5], have found that actually the radiance field plays the key role to improve the performance other than the MLP. Since these works focused on the novel view synthesis task, to evaluate the results, they usually adopted the image quality metrics, such as PSNR, SSIM, and LPIPS. These metrics are adequate for the task. However, we still wonder how to evaluate the radiance field itself and if these metrics are still adequate. Unfortunately, as the ground-truth of the radiance field is hard to be observed, we could not directly evaluate it. Moreover, these metrics evaluate the radiance field as a whole and could not measure the quality of each component. Existing novel view synthesis works have not explored this problem.

Surface Reconstruction. Some other works proposed to restore the surfaces via reconstructing the latent radiance field by inverse volume rendering [36, 16, 25, 34, 45, 46, 40, 38]. These methods usually obtain the surfaces via transforming the latent density field into an occupancy field or signed distance field (SDF). Then the Chamfer Distance between the points on the surfaces and the ground-truth can be computed to measure the results. However, it is expensive and complicated to set up the hardware environment and scan the ground-truth point-cloud. Moreover, during the transformation process, some information would be discarded as the transformed result only contains the ISO-surface. Therefore, the CD metric may be not suitable to evaluate the density field without modification.

Geometry Metrics. Due to the unavailable density field’s ground-truth, we can not use direct evaluation metrics, such as mean squared error or mean absolute error. Therefore, only indirect metrics could be considered. In previous Structure-from-Motion (SfM) systems, *e.g.*, [11, 30, 31, 1], the mean re-projection error is well-known and widely adopted as metric and optimization objective to measure the reconstructed structures and motions. Mean re-projection error is defined as the mean distance between each observed image feature point and the re-projection point of its reconstructed 3D point. In other words, it could evaluate the reconstructed geometry without

the corresponding ground-truth. Inspired by this, we argue that the geometry of a radiance field also could be evaluated only with the observation images.

3. Low-frequency Color Prior

In this section, we first briefly present the basic formulation of radiance field, and then elaborate and analyse its low-frequency color prior.

According to NeRF [21], a radiance field \mathcal{F} is defined on a 3D space \mathbf{V} , which has two components, *i.e.*, the density field \mathcal{F}^σ and the color field \mathcal{F}^c . For a 3D point \mathbf{v} in the space \mathbf{V} , the density $\mathcal{F}_\mathbf{v}^\sigma$ is a scalar that controls how much color of this point is accumulated. And the color $\mathcal{F}_\mathbf{v}^c$ encodes all view-dependent color information. For a specific 3D direction \mathbf{d} , the color vector is denoted as $\mathcal{F}_{\mathbf{v}, \mathbf{d}}^c$. To render the color of a ray $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$ with near and far bounds t^n, t^f , we can follow the function

$$C(\mathbf{r}) = \int_{t^n}^{t^f} T(t) \mathcal{F}_{\mathbf{r}(t)}^\sigma \mathcal{F}_{\mathbf{r}(t), \mathbf{d}}^c dt, \quad (1)$$

where \mathbf{o} is the camera original point, \mathbf{d} is the direction from \mathbf{o} to the corresponding pixel center, and the transmittance

$$T(t) = \exp \left(\int_{t^n}^t -\mathcal{F}_{\mathbf{r}(s)}^\sigma ds \right). \quad (2)$$

To numerically estimate this continuous integral, NeRF [21] resorts to quadrature [18]. Then the rendering function is

$$\hat{C}(\mathbf{r}) = \sum_{i=1}^N \hat{T}(t_i) \left(1 - \exp(-\mathcal{F}_{\mathbf{r}(t_i)}^\sigma \delta_{\mathbf{r}(t_i)}) \right) \mathcal{F}_{\mathbf{r}(t_i), \mathbf{d}}^c, \quad (3)$$

where N is the number of sample points along the ray,

$$\hat{T}(t_i) = \exp \left(- \sum_{j=1}^{i-1} \mathcal{F}_{\mathbf{r}(t_j)}^\sigma \delta_{\mathbf{r}(t_j)} \right), \quad (4)$$

and $\delta_{\mathbf{r}(t_i)}$ is the distance between the point $\mathbf{r}(t_i)$ and its previous neighbour $\mathbf{r}(t_{i-1})$. Specifically, $\delta_{\mathbf{r}(t_1)} = 0$.

To illustrate the properties of a radiance field, we construct an example scene as demonstrated in Fig. 2 (a), which consists of a cube. For better understanding, we also refer to the well-known related conception, *i.e.*, light field, which describes the amount of light flowing in every direction through every point in space [23]. In our opinion, the difference between light field and the color field of a radiance field is that, if a point has a zero density, its color in color field could be arbitrary, but its color must be determined in light field. And if a point has a non-zero density, the colors in these two fields would be identical. Besides, light field has no density information. As illustrated in Fig. 2 (b), the color frequency tends to be lower and lower when the point

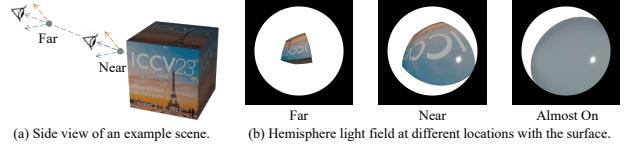


Figure 2. The color frequency tends to be lower and lower when the point approaching to a surface.

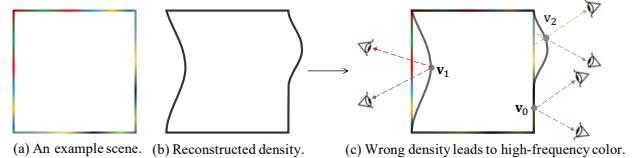


Figure 3. As the points \mathbf{v}_1 and \mathbf{v}_2 demonstrate, inaccurate density will lead to higher-frequency color whether the point inside or outside the ground-truth surface.

approaching to a surface. Moreover, the ground-truth density field should be the outer surface of the cube. Then, a straight-forward hypothesis could be concluded as that, for an ideal radiance field, the density-weighted mean of color frequency should be small. We name this as low-frequency color prior. Except for the mirrors, this prior widely exists for most types of surface whether it is rough or glossy.

Moreover, we find that wrong density will break the low-frequency color prior. To illustrate this, we present an example scene in Fig. 3 (a), its reconstructed density field in Fig. 3 (b), and combine the scene and the reconstructed density field in Fig. 3 (c). Given the observations and the reconstructed density field, the optimal color at each point could be computed. We detailedly present the computation method in Sec. 4. Intuitively, as demonstrated in Fig. 3 (c), the color of a point at the observation direction should be identical with the intersection between the observation ray and the corresponding ground-truth surface. Then, we can see that whether the point is inside or outside the ground-truth surface, as the points \mathbf{v}_1 and \mathbf{v}_2 demonstrate respectively, its color frequency tends to be higher. While the color frequency will still be low if the point is at the ground-truth surface as the point \mathbf{v}_0 demonstrates. Note that, we use a 2D graph for clarity, nevertheless, all the analyses and conclusions are applicable for 3D cases.

From these observations, if we have the color frequency field \mathcal{F}^f computed from the density field and observation images, the density-weighted mean color frequency

$$\mathbf{f}_\mathbf{V} = \frac{\int_{\mathbf{v} \in \mathbf{V}} \mathcal{F}_\mathbf{v}^\sigma \mathcal{F}_\mathbf{v}^f d\mathbf{v}}{\int_{\mathbf{v}} \mathcal{F}_\mathbf{v}^\sigma d\mathbf{v}} \quad (5)$$

should be small for a good reconstructed density field. However, it is not easy to figure out the color frequency of even a single point, not to mention the frequency field. To alleviate this difficulty, we resort to the conjugate problem. According to the frequency domain transformation

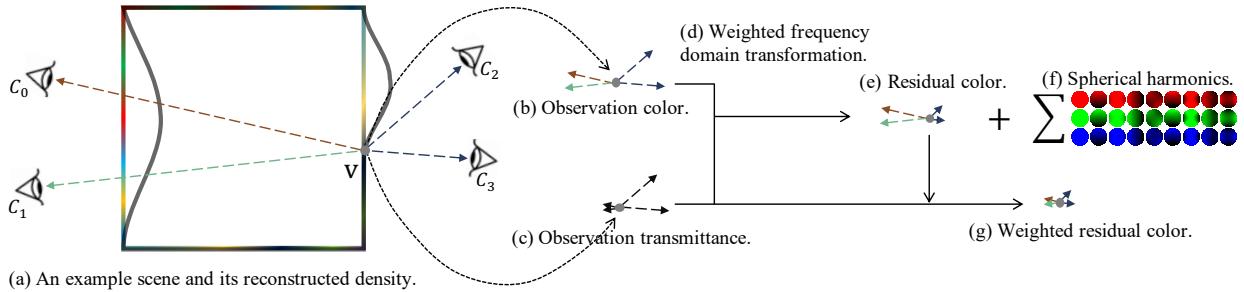


Figure 4. Illustration of the residual color computation process. (a) demonstrates an example scene (the colorful square) and its reconstructed density field (the black quadrangle). For a specific point \mathbf{v} , we can obtain its observation colors (b) according to the captured images. To tackle the occlusion, the transmittance between the point \mathbf{v} and each camera C could be taken as the confidence or weight of each observation, *i.e.*, (c). Based on observation color (b) and confidence (c), the color could be weighted-transformed (d) into frequency domain (f) with the residual color (e). Due to the confidence (c), the residual color will be large if its corresponding confidence is low. To this end, the final residual color (g) also should be weighted by observation confidence (c).

theory, if a signal is low-frequency, we could well approximate the signal with a group of low-frequency basis, and the residual would be quite small. Otherwise, if the signal is high-frequency, the residual would be large. Therefore, the residual could be adopted to replace the frequency. In other words, the density-weighted mean residual color

$$r_{\mathbf{v}} = \frac{\int_{\mathbf{v}} \mathcal{F}_{\mathbf{v}}^{\sigma} \mathcal{F}_{\mathbf{v}, \mathbf{d}}^r d\mathbf{v}}{\int_{\mathbf{v}} \mathcal{F}_{\mathbf{v}}^{\sigma} d\mathbf{v}} \quad (6)$$

should be small for a good reconstructed density field.

4. Inverse Mean Residual Color

In this section, we present the algorithm to compute the inverse mean residual color. The difficulty is that we could not use the color field reconstructed by the radiance field reconstruction method. If so, the geometry is not evaluated individually. To overcome this, we propose to utilize the observation images. For a 3D point, its projection points on all images are its observations. We can approximate its observations with a group of low-frequency basis, and the residual color could be obtained. We illustrate the whole pipeline in Fig. 4 and first introduce the observations acquisition process, then transform them into the low-frequency domain, and compute the IMRC at last.

4.1. Observations Acquisition

As illustrated in Fig. 4 (a), for any 3D point \mathbf{v} in the scene, given the reconstructed density field \mathcal{F}^{σ} and K cameras, we could obtain at most K observations corresponding to the observed images. For a specific camera, we denote its projection matrix and original point as \mathbf{P}_k , \mathbf{o}_k respectively. Then the point \mathbf{v} could be projected on the image plane following the projection equation

$$\mathbf{v} = \mathbf{P}_k \mathbf{v}, \quad (7)$$

where v is the coordinate on image. Its observed color \mathbf{c}_k as illustrated in Fig. 4 (b) could be calculated using bi-linear interpolation.

Besides the color, we should also record the observation direction since the view-dependent color is a function of direction. The observation direction could be calculated as

$$\mathbf{d}_k = \frac{\mathbf{o}_k - \mathbf{v}}{\|\mathbf{o}_k - \mathbf{v}\|}, \quad (8)$$

where $\|\cdot\|$ is the L_2 normalization operator.

Another important thing is that, due to the occlusion, observations from different cameras should not be regarded equally. As illustrated in Fig. 4 (c), it is easy to understand that, if the accumulated density between the 3D point \mathbf{v} and camera original point \mathbf{o}_k is small, the corresponding observation confidence should be high, and vice versa. Coincidentally, this measurement is actually the transmittance between the 3D point and the camera. Therefore, we directly take the transmittance to measure the confidence of each observation, which could be calculated as

$$T_k = \exp \left(\int_0^{\|\mathbf{o}_k - \mathbf{v}\|} -\mathcal{F}_{\mathbf{r}(s|\mathbf{v}, \mathbf{d}_k)}^{\sigma} ds \right), \quad (9)$$

where $\mathbf{r}(s|\mathbf{v}, \mathbf{d}_k) = \mathbf{v} + s\mathbf{d}_k$. However, this continuous integral is not easy to be computed for a computer. Therefore, we numerically estimate it following NeRF [21] as

$$\hat{T}_k = \exp \left(- \sum_{i=1}^N \mathcal{F}_{\mathbf{r}(t_i|\mathbf{v}, \mathbf{d}_k)}^{\sigma} \delta_{\mathbf{r}(t_i|\mathbf{v}, \mathbf{d}_k)} \right), \quad (10)$$

where N is the number of sample points, and $\mathbf{r}(t_i|\mathbf{v}, \mathbf{d}_k) = \mathbf{v} + t_i \mathbf{d}_k$ is the i -th point.

Overall, for any 3D point, we will have K observations where each has three cells, *i.e.*, color, direction, and confidence, which is denoted as $\langle \mathbf{c}_k, \mathbf{d}_k, \hat{T}_k \rangle$. Specially, if the point is outside the viewing cone of the k -th camera, the confidence \hat{T}_k should be set as 0.

4.2. Frequency Domain Transformation

In this subsection, we approximate the observations with a group of low-frequency basis. Spherical harmonics (SH) are natural choice since they have been widely adopted in computer graphics to represent low-frequency colors [4, 27] and also have been successfully applied on radiance field reconstruction by [43, 42]. More details about SH could be found in [10]. In the following, we briefly review the frequency coefficients estimation process following the well-known Monte Carlo method and present the difference under our situation.

From the frequency domain transformation theory, given a signal F defined on the unit sphere \mathbf{S} , we can obtain its coefficient h_l^m of the basis function Y_l^m as

$$\begin{aligned} h_l^m &= \int_{\mathbf{d} \in \mathbf{S}} \frac{F(\mathbf{d}) Y_l^m(\mathbf{d})}{p(\mathbf{d})} p(\mathbf{d}) d\mathbf{d} \\ &= \mathbb{E} \left(\frac{F(\mathbf{d}) Y_l^m(\mathbf{d})}{p(\mathbf{d})} \right) = 4\pi \mathbb{E} (F(\mathbf{d}) Y_l^m(\mathbf{d})) \end{aligned}, \quad (11)$$

where $l \in \mathbb{N}$, $m \in \mathbb{N} \cap [-l, l]$ is the degree and order of the SH basis, $p(\mathbf{d})$ is the sampling probability of direction \mathbf{d} , $\mathbb{E}()$ is the expectation. As \mathbf{d} is sampled evenly on the unit sphere \mathbf{S} , $p(\mathbf{d}) \equiv \frac{1}{4\pi}$. Based on this, given K observations which are obtained in the previous subsection, we can estimate the coefficient as

$$\hat{h}_l^m = 4\pi \frac{1}{K} \sum_{k=1}^K \mathbf{c}_k Y_l^m(\mathbf{d}_k). \quad (12)$$

However, this equation regards each observation equally, which is not consistent with the confidence of each observation. Considering the confidence, the equation could be updated as

$$\hat{h}_l^m = 4\pi \frac{1}{\sum_k T_k} \sum_{k=1}^K T_k \mathbf{c}_k Y_l^m(\mathbf{d}_k). \quad (13)$$

Unfortunately, under our situation, the estimation equation still has one problem. The method implicitly assumes that the direction \mathbf{d} distributes uniformly on the unit sphere \mathbf{S} . Therefore, the estimation process of each \hat{h}_l^m is independent. In computer graphics, this could be guaranteed via uniformly sampling the direction. However, this is not true under our settings as we cannot control the observation directions at all. In practice, the observation directions usually are not uniformly distributed.

To overcome this, we should estimate the coefficients in turn and eliminate the influence of previous estimated coefficients. Then the estimation equation could be further updated as

$$\hat{h}_l^m = 4\pi \frac{1}{\sum_k T_k} \sum_{k=1}^K T_k \cdot {}_l^m \tilde{\mathbf{c}}_k \cdot Y_l^m(\mathbf{d}_k), \quad (14)$$

where

$${}_l^m \tilde{\mathbf{c}}_k = \mathbf{c}_k - \sum_{0 \leq i \leq l}^{-l \leq j < \tilde{m}} \hat{h}_i^j \cdot Y_i^j(\mathbf{d}_k), \quad (15)$$

where $\tilde{m} = m$ if $i = l$, else $\tilde{m} = i + 1$. Actually during implementation, we increase the degree and order from low to high in turn and estimate the coefficient of corresponding SH basis. Before each estimation, we remove the color encoded by previous basis functions and only deal with the residual color ${}_l^m \tilde{\mathbf{c}}_k$.

As illustrated in Fig. 4 (e), due to the occlusion, the final residual color at a specific direction will be large if its corresponding transmittance is low. Therefore, this large residual color should be compressed. To this end, as demonstrated in Fig. 4 (g), we exploit the transmittance once again to obtain the final residual color at direction \mathbf{d}_k as

$$\tilde{\mathbf{c}}_k = T_k {}_{L+1}^{-L-1} \tilde{\mathbf{c}}_k, \quad (16)$$

where L is the maximal degree of SH. Since $\tilde{\mathbf{c}}_k$ may be negative, we take the weighted mean square of all directions

$$\tilde{c} = \frac{1}{\sum_{k=1}^K T_k} \sum_{k=1}^K T_k \left({}_{L+1}^{-L-1} \tilde{\mathbf{c}}_k \right)^2. \quad (17)$$

as the residual color of the point. And the smaller the \tilde{c} , the lower-frequency the color.

Note that, when the direction is uniformly distributed, Eq. (14) is identically equal to Eq. (12) since the basis functions are orthogonal to each other. Actually, Eq. (14) is the general equation to transform a discrete signal from original domain to another domain. If with the uniform distribution and orthogonal basis, we have $T_k = 1$, $\sum_{k=1}^K Y_i^j(\mathbf{d}_k) \cdot Y_l^m(\mathbf{d}_k) = 0$, and Eq. (14) could be transformed to Eq. (12) because

$$\begin{aligned} \hat{h}_l^m &= 4\pi \frac{1}{K} \sum_{k=1}^K \left(\mathbf{c}_k - \sum_{0 \leq i \leq l}^{-l \leq j < \tilde{m}} \hat{h}_i^j Y_i^j(\mathbf{d}_k) \right) Y_l^m(\mathbf{d}_k) = \\ &\quad \frac{4\pi}{K} \sum_{k=1}^K \mathbf{c}_k Y_l^m(\mathbf{d}_k) - \frac{4\pi}{K} \sum_{k=1}^K \sum_{0 \leq i \leq l}^{-l \leq j < \tilde{m}} \hat{h}_i^j Y_i^j(\mathbf{d}_k) Y_l^m(\mathbf{d}_k) \end{aligned}, \quad (18)$$

and

$$\begin{aligned} &\sum_{k=1}^K \sum_{0 \leq i \leq l}^{-l \leq j < \tilde{m}} \hat{h}_i^j Y_i^j(\mathbf{d}_k) Y_l^m(\mathbf{d}_k) \\ &= \sum_{0 \leq i \leq l}^{-l \leq j < \tilde{m}} \hat{h}_i^j \sum_{k=1}^K Y_i^j(\mathbf{d}_k) Y_l^m(\mathbf{d}_k) = \sum_{0 \leq i \leq l}^{-l \leq j < \tilde{m}} \hat{h}_i^j \cdot 0 = 0 \end{aligned}. \quad (19)$$

However, in our case, the directions are not uniformly distributed and weighted, therefore Eq. (14) should be adopted. We present the detail proof process and the comparison results in the Appendix.

4.3. Final Metric Computation

With the calculated residual color \tilde{c} at all locations, the MRC could be obtained following Eq. (6). As it is a continuous integral, to numerically estimate it, we resort to quadrature [18] once again. Specifically, we discretize the 3D space \mathbf{V} into a volume, and denote all voxel vertexes as \mathcal{V} . Then the mean residual color could be estimated as

$$\text{MRC} = \frac{\sum_{\mathbf{v}_i \in \mathcal{V}} (1 - \exp(-\mathcal{F}_{\mathbf{v}_i}^\sigma \cdot \delta_{\mathbf{v}_i})) \tilde{c}_{\mathbf{v}_i}}{\sum_{\mathbf{v}_i \in \mathcal{V}} (1 - \exp(-\mathcal{F}_{\mathbf{v}_i}^\sigma \cdot \delta_{\mathbf{v}_i}))}, \quad (20)$$

where $\tilde{c}_{\mathbf{v}_i}$ is the mean square residual color at vertex \mathbf{v}_i , and $\delta_{\mathbf{v}_i}$ is the voxel size.

Finally, we found that the mean residual color MRC would be very small, which makes the value not convenient to be presented and compared. To this end, we further transform it into decibel (dB) as the PSNR metric does. As the theoretically maximal residual color is 1, the final metric could be obtained as

$$\text{IMRC} = 10 \cdot \log_{10} \left(\frac{1}{\text{MRC}} \right). \quad (21)$$

As the transformation includes an inverse operation, we denote it as IMRC shorted for Inverse Mean Residual Color. The larger the IMRC, the better the geometry.

5. Experiments

In this section, we first verify the effectiveness of our IMRC metric, then explore the influence of different settings during IMRC computation. Finally, we benchmark several state-of-the-art methods on three widely used datasets to promote future related research.

5.1. Validation of IMRC

To exploit the rightness and effectiveness of the proposed IMRC metric, we conduct a series of experiments on the DTU dataset [15] as it provides the point-cloud ground-truth. Though the point-cloud is essentially different from the radiance field, it still could be taken as a reference. As previous surface reconstruction works did, *e.g.*, [41, 36], we use the selected 15 scenes from DTU. Each consists of 49 or 64 carefully calibrated images and the scanned point-cloud ground-truth. We select 5 or 6 images from the total 49 or 64 images respectively as testing ones, and take the remaining as training ones. The background of each image has been removed in advance using the given mask.

For radiance field reconstruction methods, we select 4 well-known methods, including JaxNeRF [21, 8], Plenoxels [42], DVGO [32], and TensoRF [5], which are with high impact and good performance. We adopt the code released by the authors to optimize each scene from the DTU dataset. Additionally, we add a transmittance loss to the JaxNeRF to

Table 1. The PSNR \uparrow /CD \downarrow /IMRC \uparrow /UserRank \downarrow results of 4 state-of-the-art methods on the DTU dataset. (Rank 1st, 2nd, 3rd, 4th)

Method	JaxNeRF[21, 8]	Plenoxels[42]	DVGO[32]	TensoRF[5]
Scan24	29.19 / 1.415 / 20.11 / 1	27.72 / 1.592 / 14.96 / 3	27.96 / 2.028 / 17.45 / 2	28.76 / 2.144 / 11.73 / 4
Scan37	26.77 / 1.502 / 12.99 / 1	26.22 / 1.813 / 12.99 / 3	26.47 / 1.466 / 14.05 / 2	26.00 / 1.778 / 8.86 / 4
Scan40	29.43 / 1.495 / 20.74 / 1	28.72 / 2.030 / 15.75 / 3	28.38 / 2.017 / 17.43 / 2	28.85 / 2.381 / 13.39 / 4
Scan55	31.06 / 0.637 / 19.55 / 1	30.50 / 0.892 / 15.55 / 3	31.66 / 1.247 / 19.37 / 2	29.57 / 1.765 / 9.78 / 4
Scan63	34.43 / 1.689 / 20.68 / 1	34.09 / 1.931 / 17.63 / 3	35.42 / 1.574 / 19.94 / 2	34.69 / 2.216 / 10.92 / 4
Scan65	29.90 / 1.212 / 17.59 / 1.3	31.16 / 1.548 / 14.55 / 3	30.63 / 1.482 / 17.32 / 1.7	30.63 / 1.947 / 10.78 / 4
Scan69	29.25 / 1.306 / 19.07 / 1	30.15 / 2.346 / 15.40 / 3	29.67 / 1.489 / 17.81 / 2	29.43 / 2.254 / 7.91 / 4
Scan83	35.28 / 1.478 / 17.98 / 1	37.00 / 2.245 / 16.63 / 2.7	36.14 / 1.610 / 17.95 / 2.3	35.96 / 2.712 / 7.95 / 4
Scan97	28.02 / 1.600 / 17.10 / 1	29.56 / 2.809 / 15.29 / 3	29.11 / 1.629 / 15.85 / 2	29.25 / 2.278 / 9.06 / 4
Scan105	31.86 / 1.136 / 19.99 / 1	33.36 / 1.907 / 16.67 / 3	33.07 / 1.352 / 18.85 / 2	32.92 / 2.219 / 9.19 / 4
Scan106	33.97 / 0.903 / 19.70 / 1	33.82 / 2.317 / 16.37 / 3	34.70 / 1.595 / 19.29 / 2	33.61 / 2.524 / 9.01 / 4
Scan110	32.96 / 1.842 / 16.00 / 1.7	32.49 / 3.349 / 14.86 / 3	34.00 / 1.770 / 16.13 / 1.3	33.67 / 3.396 / 7.99 / 4
Scan114	29.44 / 1.091 / 18.89 / 1	30.52 / 1.657 / 16.53 / 3	30.47 / 1.481 / 18.18 / 2	29.79 / 2.024 / 9.54 / 4
Scan118	36.68 / 0.998 / 21.77 / 1.7	36.10 / 2.719 / 17.82 / 3	37.58 / 1.297 / 22.57 / 1.3	35.76 / 2.436 / 11.34 / 4
Scan122	35.06 / 0.761 / 18.00 / 2	36.42 / 2.349 / 17.15 / 3	37.05 / 1.244 / 20.62 / 1	36.18 / 2.258 / 9.60 / 4
Average				
31.55 / 1.271 / 18.71 / 1.18				
31.86 / 2.100 / 15.88 / 2.98				
32.15 / 1.552 / 18.19 / 1.84				
31.67 / 2.289 / 9.80 / 4				

Average 31.55 / 1.271 / 18.71 / 1.18 31.86 / 2.100 / 15.88 / 2.98 32.15 / 1.552 / 18.19 / 1.84 31.67 / 2.289 / 9.80 / 4

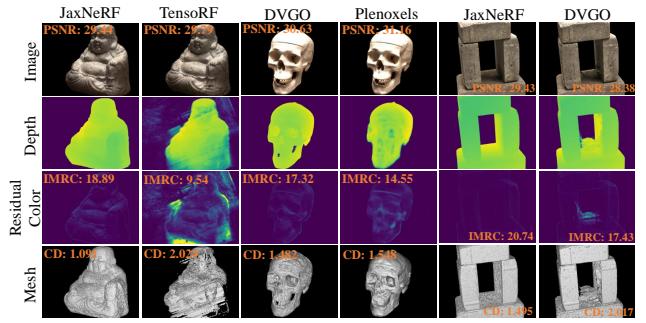


Figure 5. Examples of the CD \downarrow /IMRC \uparrow /UserRank \downarrow consistent results on DTU dataset. For each example, left is better. From left to right, Scan 114, 65, and 40.

supervise the transmittance of a ray according to the foreground mask. After the optimization process, we only export the density field obtained by each method. Notably, JaxNeRF [21, 8] models a continuous density field while other methods model discrete ones. For a fair comparison, we discrete the density field of each method via sampling a 512^3 density volume. Then the IMRC metric could be calculated via the algorithm elaborated in Sec. 4. The experimental results are presented in Tab. 1, where the PSNR on testing images and the CD calculated with the point-cloud ground-truth are also reported. Specifically, we search the optimal threshold needed by the marching cubes algorithm for each method on each scene and report the lowest CD. Because the ground-truth of the density field is not available, we perform a user study. Specifically, 3 experts evaluate and rank the scene geometry produced by the 4 methods from 1 (best) to 4 (worst) based on the depth images, residual colors, and meshes. They are unknown about the metric values and the method that produces the results. We report the mean rank of each result in Tab. 1.

In most cases, the IMRC metric, CD, and UserRank agree with each other. The average results (last row) in Tab. 1 show that they rank 4 methods consistently. We showcase three CD/IMRC/UserRank consistent cases in Fig. 5. They also indicate that PSNR cannot well evaluate a density field. Besides, there are 2 IMRC/UserRank conflict pairs and 11 CD/UserRank conflict pairs out of all $90 = 15 \cdot \binom{4}{2}$ pairs, respectively. In the left panel of Fig. 6, one

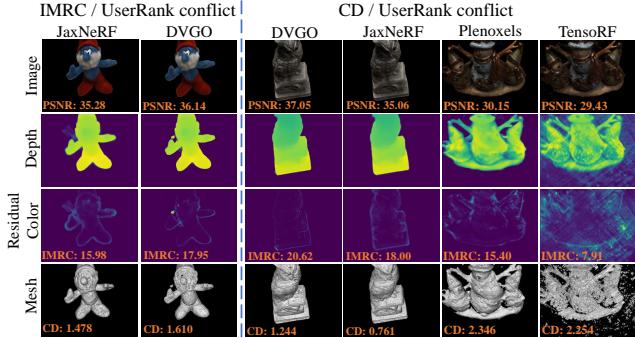


Figure 6. The IMRC \uparrow /UserRank \downarrow and CD \downarrow /UserRank \downarrow conflict results. For each example, left has better UserRank. From left to right, Scan 83, 122, and 69.

IMRC/UserRank conflict case is shown. The density field of the JaxNeRF is not sharp, and its surface is surrounded by many low density floaters. In contrast, the DVGO has a distinct floater as highlighted by the residual color. Although the IMRC of DVGO is better, such a distinct floater may lead to worse UserRank. The CD/UserRank conflicts mainly stem from two reasons. First, a NeRF model does not guarantee that its surface points all have the same density value. Therefore, by extracting one iso-surface at a typical density level with the marching cubes algorithm, some surface information and near surface floaters that have different density values are inevitably discarded. As a result, the CD for such a surface may not well reflect real geometry. As shown by the middle two columns of Fig. 6. The JaxNeRF produces many low density floaters around the surface, as highlighted by its residual color, while these floaters are missing after applying the marching cubes algorithm. On the other hand, the iso-surface of DVGO is inferior because some of its surface points that have different density values are discarded, which leads to poor CD. Another reason is that an object mask is applied in CD calculation, so points that lie outside the mask are neglected. As shown in the last two columns of Fig. 6, the TensoRF’s mesh is qualitatively worse than Plenoxels’, but lots of floaters in it are out of the mask. As a result, the CD of TensoRF is erroneously better than that of Plenoxels. The IMRC metric successfully evaluate the last two scenes and is consistent with UserRank. Due to limited space, we visualize all conflict cases in the appendix. They can be analyzed similarly.

Overall, the qualitative and quantitative results verify the effectiveness of IMRC to evaluate a density field. On one hand, PSNR could only evaluate the radiance field as a whole. On the other hand, the calculation of CD suffers from a loss of information that only an iso-surface at one specific density level is considered. The IMRC metric, in contrast, is more suitable to evaluate the whole field, rather than only a surface, as it can deal with density values at all locations. It is also more consistent with UserRank. Moreover, the CD metric needs the point-cloud ground-truth,

Table 2. IMRC results of the selected methods on the DTU dataset with different SH degrees, that indicates that higher SH degree leads to better IMRC result.

SH Degree	0	1	2	3
SH Basis #	1	4	9	16
JaxNeRF[21, 8]	15.93	17.18	18.71	19.57
Plenoxels[42]	13.81	14.78	15.88	16.52
DVGO[32]	15.66	16.80	18.19	18.97
TensoRF[5]	9.00	9.40	9.80	10.06
Average	13.60	14.54	15.64	16.28

Table 3. IMRC results of the selected methods on DTU dataset with different density volume resolutions.

Resolution	64 ³	128 ³	256 ³	512 ³	768 ³
JaxNeRF[21, 8]	17.92	18.42	18.61	18.71	18.54
Plenoxels[42]	15.96	16.04	16.04	15.88	16.00
DVGO[32]	17.05	17.72	18.14	18.19	18.27
TensoRF[5]	9.45	9.62	9.75	9.80	9.88
Average	15.09	15.45	15.64	15.64	15.67

which also limits its application as this data is complicated and expensive to be obtained and usually is not available.

5.2. Different Settings of IMRC

In this subsection, we exploit the influences of the parameters defined during IMRC computation. There are only 2 parameters that are needed to be configured manually, *i.e.*, the SH degree and the volume resolution.

The first and foremost parameter is the degree of the SH used to compute the residual color for each point. As in Tab. 2, we present the IMRC results of the selected 4 methods on the DTU dataset with different SH degrees. We can see that a higher SH degree leads to a better IMRC result. This is reasonable since the observations could be approximated better with more SH basis functions. However, we could not set a very large SH degree, because, if so, the residual color will be very small at all positions in the field, which makes the metric indistinct to measure the geometry. We also observe that the IMRC increases quickly when the SH degree is less than 2. And the increased margin becomes smaller starting from degree 3. Therefore, we set the SH degree as 2 in all other experiments. This setting is also usually used in previous works such as [42, 5]. Notably, the relative ranks of all methods remain unchanged with different SH degrees. This indicates that the IMRC is somehow stable with different SH degrees.

Another problem to be noted is that, when the number of observations are small, we should decrease the SH degree. This will happen for the methods aiming at reconstructing the radiance field from sparse-views, such as [24]. According to the frequency domain transformation theory, the SH degree should be small than the number of observations.

The second parameter is the density volume resolution. As presented in Tab. 3, with the increasing resolution, the IMRC metric is slightly better. The increase may come from the more accurate density field with higher space reso-

Table 4. The PSNR↑/IMRC↑ results on NeRF synthetic dataset. (Rank 1st, 2nd, 3rd, 4th)

Method	JaxNeRF[21, 8]	Plenoxels[42]	DVGO[32]	TensoRF[5]
Chair	30.28 / 19.21	30.66 / 17.60	33.79 / 21.59	30.46 / 19.10
Drums	24.16 / 13.41	24.26 / 12.65	25.69 / 13.38	24.51 / 14.32
Ficus	27.99 / 14.58	28.31 / 14.57	33.27 / 18.29	28.56 / 11.59
Hotdog	35.49 / 4.59	35.11 / 20.02	36.90 / 21.16	35.43 / 22.41
Lego	31.41 / 3.73	30.87 / 17.54	34.83 / 19.27	31.83 / 19.21
Materials	29.67 / 16.09	28.39 / 15.15	29.94 / 17.16	28.91 / 12.63
Mic	31.22 / 2.15	32.50 / 16.47	28.41 / 8.09	32.88 / 14.41
Ship	28.76 / 18.77	28.52 / 18.37	29.82 / 19.08	28.68 / 19.24
Average	29.87 / 11.57	29.83 / 16.55	31.58 / 17.25	30.16 / 16.61

Table 5. The PSNR↑/IMRC↑ results on LLFF dataset. (Rank 1st, 2nd, 3rd, 4th)

Method	JaxNeRF[21, 8]	Plenoxels[42]	DVGO[32]	TensoRF[5]
Fern	24.83 / 21.58	25.47 / 19.07	25.07 / 20.04	25.31 / 19.68
Flower	28.07 / 23.03	27.83 / 19.46	27.61 / 21.42	28.22 / 21.15
Fortress	31.76 / 26.97	31.09 / 22.32	30.38 / 24.82	31.14 / 24.57
Horns	28.10 / 23.58	27.60 / 19.09	27.55 / 21.20	27.64 / 21.80
Leaves	21.23 / 17.53	21.43 / 16.32	21.04 / 16.43	21.34 / 16.58
Orchids	20.27 / 18.06	20.26 / 16.03	20.38 / 16.59	20.02 / 16.37
Room	33.04 / 26.08	30.22 / 21.64	31.45 / 24.64	31.80 / 25.19
Trex	27.42 / 24.22	26.49 / 19.23	27.17 / 20.90	26.61 / 21.23
Average	26.84 / 22.63	26.30 / 19.15	26.33 / 20.76	26.51 / 20.82

Table 6. Comparison of PSNR↑/IMRC↑ results on three datasets produced by DVGO with and without distortion loss. (**Better**)

Dataset	DTU	NeRF Synthetic	LLFF
DVGO[32]	32.15 / 18.19	31.58 / 17.25	26.23 / 18.87
DVGO[32] + Distortion Loss[3]	32.20 / 18.47	31.50 / 17.94	26.33 / 20.76

lution. We also observe that the relative ranks of all methods remain unchanged at different resolutions. This indicates that the IMRC is also somehow stable with different resolutions. As the average IMRC does not change from resolution 256^3 to 512^3 , we adopt 512^3 in other experiments.

5.3. Benchmarking State-of-the-arts

Besides the DTU dataset, we also benchmark the 4 state-of-the-art methods on NeRF Synthetic [21] and LLFF [20] datasets to promote future research. For the NeRF Synthetic dataset, we use the released code and train on the black background (actually set background color as **0**) as done in the DTU dataset. For the LLFF dataset, we directly use the models released by the authors. As there is no point-cloud ground-truth on NeRF Synthetic and LLFF datasets, we could not compute the CD metric and only present the PSNR and IMRC results in Tab. 4 and Tab. 5, respectively. We also visualize some results in Fig. 7 and Fig. 8.

The IMRC metric helps quantitatively evaluate how a regularization loss improves the density field. Specifically, we train DVGO [32] with distortion loss [3] and report the results in Tab. 6. The better IMRC for all the three datasets verified that the distortion loss improves scene geometry.

6. Discussion and Conclusion

We first discuss the limitations of IMRC and then conclude this paper as following.

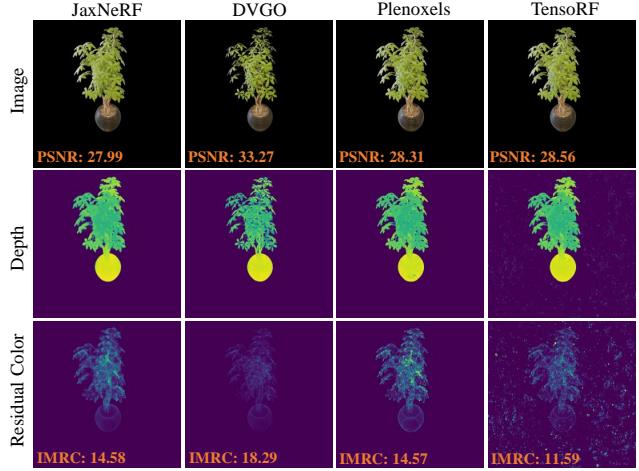


Figure 7. Example results on Ficus from NeRF synthetic dataset.

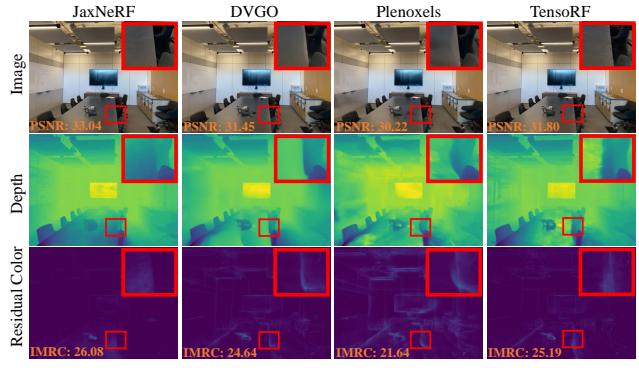


Figure 8. Example results on Room from the LLFF dataset.

Limitations. There are also some degrade situations where the proposed IMRC metric may fail. For example, if the whole density field is empty except for a few points that have been observed only by several cameras, the IMRC metric will be high. However, this density field is really bad. This type of degrade situations also exist for other geometry metrics, such as CD and re-projection error. Therefore, when using the proposed IMRC metric, it should be combined with other metrics, such as PSNR and SSIM. Only in this way, we can evaluate the results with less bias.

Conclusion. In this paper, we aim at evaluating the geometry information of a radiance field without ground-truth. This problem is important as the radiance field has been widely used not only on novel view synthesis but also 3D reconstruction tasks. For 3D reconstruction tasks, the geometry information is essential. However, due to the unavailable ground-truth and unsuitable metrics, there is no proper metric to quantitatively evaluate the geometry. To alleviate this dilemma, we propose the Inverse Mean Residual Color (IMRC) metric based on our insights on the properties of the radiance field. Qualitative and quantitative experimental results verify the effectiveness of IMRC metric. We also benchmark 4 state-of-the-art methods on 3 datasets and hope this work could promote future related research.

References

- [1] Sameer Agarwal, Noah Snavely, Ian Simon, Steven M. Seitz, and Richard Szeliski. Building Rome in a Day. In *ICCV*, pages 72–79, 2009. 2
- [2] Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *ICCV*, pages 5855–5864, 2021. 1, 2
- [3] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *CVPR*, pages 5470–5479, 2022. 1, 8
- [4] Ronen Basri and David W Jacobs. Lambertian Reflectance and Linear Subspaces. *IEEE TPAMI*, 25(2):218–233, 2003. 5
- [5] Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. TensoRF: Tensorial Radiance Fields. In *ECCV*, 2022. 1, 2, 6, 7, 8, 14
- [6] Anpei Chen, Zexiang Xu, Fuqiang Zhao, Xiaoshuai Zhang, Fanbo Xiang, Jingyi Yu, and Hao Su. MVSNeRF: Fast Generalizable Radiance Field Reconstruction From Multi-view Stereo. In *ICCV*, pages 14124–14133, 2021. 1, 2
- [7] Zhiqin Chen, Thomas Funkhouser, Peter Hedman, and Andrea Tagliasacchi. MobileNeRF: Exploiting the Polygon Rasterization Pipeline for Efficient Neural Field Rendering on Mobile Architectures. *arXiv preprint arXiv:2208.00277*, 2022. 1, 2
- [8] Boyang Deng, Jonathan T. Barron, and Pratul P. Srinivasan. JaxNeRF: an efficient JAX implementation of NeRF, 2020. 1, 2, 6, 7, 8, 14
- [9] Stephan J Garbin, Marek Kowalski, Matthew Johnson, Jamie Shotton, and Julien Valentin. FastNeRF: High-fidelity Neural Rendering at 200fps. In *ICCV*, pages 14346–14355, 2021. 1, 2
- [10] Robin Green. Spherical harmonic lighting: The gritty details. In *Archives of the game developers conference*, volume 56, page 4, 2003. 5
- [11] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004. 2
- [12] Peter Hedman, Pratul P. Srinivasan, Ben Mildenhall, Jonathan T. Barron, and Paul Debevec. Baking Neural Radiance Fields for Real-Time View Synthesis. In *ICCV*, 2021. 1, 2
- [13] Tao Hu, Shu Liu, Yilun Chen, Tiancheng Shen, and Jiaya Jia. EfficientNeRF - Efficient Neural Radiance Fields. In *CVPR*, pages 12892–12901, 2022. 1, 2
- [14] Ajay Jain, Matthew Tancik, and Pieter Abbeel. Putting Nerf on A Diet: Semantically Consistent Few-shot View Synthesis. In *ICCV*, pages 5885–5894, 2021. 2
- [15] Rasmus Jensen, Anders Dahl, George Vogiatzis, Engil Tola, and Henrik Aanæs. Large scale multi-view stereopsis evaluation. In *CVPR*, pages 406–413. IEEE, 2014. 6
- [16] Xiaoxiao Long, Cheng Lin, Peng Wang, Taku Komura, and Wenping Wang. SparseNeuS: Fast Generalizable Neural Surface Reconstruction from Sparse views. In *ECCV*, 2022. 1, 2
- [17] Ricardo Martin-Brualla, Noha Radwan, Mehdi S. M. Sajjadi, Jonathan T. Barron, Alexey Dosovitskiy, and Daniel Duckworth. NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections. In *CVPR*, 2020. 1, 2
- [18] Nelson Max. Optical Models for Direct Volume Rendering. *IEEE TVCG*, 1(2):99–108, 1995. 2, 3, 6
- [19] Ben Mildenhall, Peter Hedman, Ricardo Martin-Brualla, Pratul P Srinivasan, and Jonathan T Barron. Nerf in the dark: High dynamic range view synthesis from noisy raw images. In *CVPR*, pages 16190–16199, 2022. 1, 2
- [20] Ben Mildenhall, Pratul P. Srinivasan, Rodrigo Ortiz-Cayon, Nima Khademi Kalantari, Ravi Ramamoorthi, Ren Ng, and Abhishek Kar. Local light field fusion: Practical view synthesis with prescriptive sampling guidelines. *ACM Trans. Graph.*, 38(4), jul 2019. 8
- [21] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *ECCV*, 2020. 1, 2, 3, 4, 6, 7, 8, 14
- [22] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant Neural Graphics Primitives With A Multiresolution Hash Encoding. *arXiv preprint arXiv:2201.05989*, 2022. 1, 2
- [23] Ren Ng. Fourier slice photography. In *SIGGRAPH*, pages 735–744, 2005. 3
- [24] Michael Niemeyer, Jonathan T. Barron, Ben Mildenhall, Mehdi S. M. Sajjadi, Andreas Geiger, and Noha Radwan. RegNeRF: Regularizing Neural Radiance Fields for View Synthesis from Sparse Inputs. In *CVPR*, 2022. 2, 7
- [25] Michael Oechsle, Songyou Peng, and Andreas Geiger. UNISURF: Unifying Neural Implicit Surfaces and Radiance Fields for Multi-View Reconstruction. In *ICCV*, pages 5569–5579, 2021. 1, 2
- [26] Sida Peng, Junting Dong, Qianqian Wang, Shangzhan Zhang, Qing Shuai, Xiaowei Zhou, and Hujun Bao. Animatable Neural Radiance Fields for Modeling Dynamic Human Bodies. In *ICCV*, pages 14314–14323, 2021. 1, 2
- [27] Ravi Ramamoorthi and Pat Hanrahan. On the relationship between radiance and irradiance: Determining the illumination from images of a convex lambertian object. *JOSA A*, 18(10):2448–2459, 2001. 5
- [28] Christian Reiser, Songyou Peng, Yiyi Liao, and Andreas Geiger. KiloNeRF: Speeding up Neural Radiance Fields with Thousands of Tiny MLPs. In *ICCV*, 2021. 1, 2
- [29] Ruizhi Shao, Hongwen Zhang, He Zhang, Mingjia Chen, Yanpei Cao, Tao Yu, and Yebin Liu. DoubleField: Bridging the Neural Surface and Radiance Fields for High-fidelity Human Reconstruction and Rendering. In *CVPR*, 2022. 2
- [30] Noah Snavely, Steven M. Seitz, and Richard Szeliski. Photo tourism: Exploring photo collections in 3d. *ACM TOG*, 25(3):835–846, jul 2006. 2
- [31] Noah Snavely, Steven M. Seitz, and Richard Szeliski. Modeling the world from internet photo collections. *IJCV*, 80(2):189–210, 2008. 2
- [32] Cheng Sun, Min Sun, and Hwann-Tzong Chen. Direct Voxel Grid Optimization: Super-fast Convergence for Radiance Fields Reconstruction. In *CVPR*, 2022. 1, 2, 6, 7, 8, 14

- [33] Cheng Sun, Min Sun, and Hwann-Tzong Chen. Improved Direct Voxel Grid Optimization for Radiance Fields Reconstruction. *arXiv preprint arXiv:2206.05085*, 2022. [1](#), [2](#)
- [34] Jiaming Sun, Xi Chen, Qianqian Wang, Zhengqi Li, Hadar Averbuch-Elor, Xiaowei Zhou, and Noah Snavely. Neural 3D Reconstruction in the Wild. In *SIGGRAPH*, 2022. [1](#), [2](#)
- [35] Matthew Tancik, Vincent Casser, Xinchen Yan, Sabeek Pradhan, Ben Mildenhall, Pratul P. Srinivasan, Jonathan T. Barron, and Henrik Kretzschmar. Block-NeRF: Scalable Large Scene Neural View Synthesis. In *CVPR*, 2022. [2](#)
- [36] 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 *NeurIPS*, 2021. [1](#), [2](#), [6](#)
- [37] Qianqian Wang, Zhicheng Wang, Kyle Genova, Pratul P Srinivasan, Howard Zhou, Jonathan T Barron, Ricardo Martin-Brualla, Noah Snavely, and Thomas Funkhouser. IBRNet: Learning Multi-view Image-based Rendering. In *CVPR*, pages 4690–4699, 2021. [1](#), [2](#)
- [38] Yiming Wang, Qin Han, Marc Habermann, Kostas Daniilidis, Christian Theobalt, and Lingjie Liu. NeuS2: Fast Learning of Neural Implicit Surfaces for Multi-view Reconstruction. *arXiv*, 2022. [1](#), [2](#)
- [39] Chung-Yi Weng, Brian Curless, Pratul P. Srinivasan, Jonathan T. Barron, and Ira Kemelmacher-Shlizerman. HumanNeRF: Free-viewpoint Rendering of Moving People from Monocular Video. In *CVPR*, 2022. [2](#)
- [40] Tong Wu, Jiaqi Wang, Xingang Pan, Xudong Xu, Christian Theobalt, Ziwei Liu, and Dahu Lin. Voxurf: Voxel-based efficient and accurate neural surface reconstruction, 2022. [1](#), [2](#)
- [41] Lior Yariv, Yoni Kasten, Dror Moran, Meirav Galun, Matan Atzmon, Basri Ronen, and Yaron Lipman. Multiview neural surface reconstruction by disentangling geometry and appearance. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *NeurIPS*, volume 33, pages 2492–2502. Curran Associates, Inc., 2020. [6](#)
- [42] Alex Yu, Sara Fridovich-Keil, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. Plenoxels: Radiance Fields Without Neural Networks. In *CVPR*, 2022. [1](#), [2](#), [5](#), [6](#), [7](#), [8](#), [14](#)
- [43] Alex Yu, Ruilong Li, Matthew Tancik, Hao Li, Ren Ng, and Angjoo Kanazawa. PlenOctrees for Real-time Rendering of Neural Radiance Fields. In *ICCV*, 2021. [1](#), [2](#), [5](#)
- [44] Alex Yu, Vickie Ye, Matthew Tancik, and Angjoo Kanazawa. PixelNeRF: Neural Radiance Fields From One or Few Images. In *CVPR*, pages 4578–4587, 2021. [1](#), [2](#)
- [45] Jason Y. Zhang, Gengshan Yang, Shubham Tulsiani, and Deva Ramanan. NeRS: Neural Reflectance Surfaces for Sparse-view 3D Reconstruction in the Wild. In *NeurIPS*, 2021. [1](#), [2](#)
- [46] Kai Zhang, Fujun Luan, Zhengqi Li, and Noah Snavely. IRON: Inverse Rendering by Optimizing Neural SDFs and Materials from Photometric Images. In *CVPR*, 2022. [1](#), [2](#)
- [47] Kai Zhang, Gernot Riegler, Noah Snavely, and Vladlen Koltun. NeRF++: Analyzing and Improving Neural Radiance Fields. *arXiv preprint arXiv:2010.07492*, 2020. [1](#), [2](#)
- [48] Fuqiang Zhao, Wei Yang, Jiakai Zhang, Pei Lin, Yingliang Zhang, Jingyi Yu, and Lan Xu. HumanNeRF: Efficiently Generated Human Radiance Field from Sparse Inputs. In *CVPR*, 2022. [2](#)

Appendix

A. More Explanations of the Eq. (14) and Eq. (15) in the Main Paper

We first shed light on why estimating the coefficients in turn as done in Eq. (14) and Eq. (15) help reduce the bias caused by nonuniform sampling. Then we conduct numerical experiments to support our claim. For simplicity, the explanation is based on using Fourier series to approximate one-dimensional signals. The basic idea is the same and can be easily extended to the case that uses spherical harmonics (SH) to approximate the signals defined on a spherical surface.

Suppose that the target signal $f(x)$ defined on $[-\pi, \pi]$ is “good” enough, or satisfies the Dirichlet Conditions, thus it could be expanded into Fourier series, *i.e.*,

$$f(x) = A_0 + \sum_{n=1}^{\infty} [a_n \cos(nx) + b_n \sin(nx)]. \quad (22)$$

Given a set of samples $\{(x_t, f(x_t)) | t = 1, \dots, T\}$, we aim to estimate the Fourier coefficients up to a typical degree k_{max} . The calculation of the coefficient a_k is well-known as

$$a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(kx) dx = \frac{1}{\pi} \int_{-\pi}^{\pi} \left\{ A_0 \cos(kx) + \sum_{n=1}^{\infty} [a_n \cos(nx) \cos(kx) + b_n \sin(nx) \cos(kx)] \right\} dx. \quad (23)$$

Using the Monte Carlo method, the above integral can be approximated by T samples x_1, \dots, x_T as

$$\hat{a}_k = \frac{1}{\pi} \frac{2\pi}{T} \sum_{t=1}^T \left\{ A_0 \cos(kx_t) + \sum_{n=1}^{\infty} [a_n \cos(nx_t) \cos(kx_t) + b_n \sin(nx_t) \cos(kx_t)] \right\}. \quad (24)$$

Because of the orthogonal completeness of the trigonometric function set, if uniform sampling is performed and the number of samples is large enough, all other terms in Eq. (24) will approach 0 except for $a_k \cos(kx_t) \cos(kx_t)$. Moreover, according to the Monte Carlo method, we have

$$\int_{-\pi}^{\pi} \cos(kx) \cos(kx) dx \approx \frac{2\pi}{T} \sum_{t=1}^T \cos(kx_t) \cos(kx_t). \quad (25)$$

And then,

$$\hat{a}_k \approx \frac{1}{\pi} \frac{2\pi}{T} \sum_{t=1}^T a_k \cos(kx_t) \cos(kx_t) = a_k \frac{1}{\pi} \frac{2\pi}{T} \sum_{t=1}^T \cos(kx_t) \cos(kx_t) \approx a_k \frac{1}{\pi} \pi = a_k. \quad (26)$$

In the nonuniform sampling case, however, non-negligible estimation biases will be generated. As an example, we visualize the bias conducted by the direct current component $\frac{1}{T} \sum_{t=1}^T A_0 \cos(x_t)$ in Fig. 9. In the uniform sampling case, $\frac{1}{T} \sum_{t=1}^T A_0 \cos(x_t)$ is better approaching to 0, as the positive (red) and negative (blue) function values are well counteracted. In contrast, the nonuniform sampling case induces a larger bias as controlled by the magnitude of A_0 . We show such a bias in a numerical experiment later.

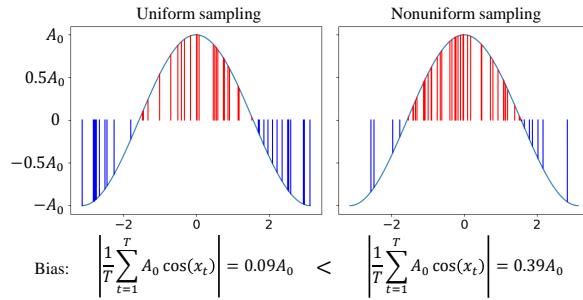


Figure 9. Nonuniform sampling generates a larger estimation bias than that of uniform sampling.

To reduce the bias caused by $\sum_{t=1}^T A_0 \cos(kx_t)$, we find it effective to minus this component in the estimation, *i.e.*,

$$\begin{aligned} a_k &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(kx) dx \\ &= \frac{1}{\pi} \int_{-\pi}^{\pi} (f(x) - A_0) \cos(kx) dx \\ &= \frac{1}{\pi} \int_{-\pi}^{\pi} \left\{ \sum_{n=1}^{\infty} [a_n \cos(nx) \cos(kx) + b_n \sin(nx) \cos(kx)] \right\} dx. \end{aligned} \quad (27)$$

This time the Monte Carlo estimator of a_k becomes

$$\hat{a}'_k = \frac{2}{T} \sum_{t=1}^T \left\{ \sum_{n=1}^{\infty} [a_n \cos(nx_t) \cos(kx_t) + b_n \sin(nx_t) \cos(kx_t)] \right\}. \quad (28)$$

Compared with \hat{a}_k , \hat{a}'_k completely eliminates the bias caused by $\sum_{t=1}^T A_0 \cos(kx_t)$. Thus, the Monte Carlo estimator

$$\frac{2}{T} \sum_{t=1}^T (f(x_t) - A_0) \cos(kx_t) \quad (29)$$

as derived from $\frac{1}{\pi} \int_{-\pi}^{\pi} (f(x) - A_0) \cos(kx) dx$ suffers less bias from nonuniform sampling than

$$\frac{2}{T} \sum_{t=1}^T f(x_t) \cos(kx_t) \quad (30)$$

as derived from $\frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(kx) dx$. Because A_0 is actually unknown, we first estimate this direct current component using the Monte Carlo method as

$$\hat{A}_0 = \frac{1}{T} \sum_{t=1}^T f(x_t). \quad (31)$$

Then, we substitute the unknown A_0 with \hat{A}_0 in Eq. (29).

Note that, the bias of other components also could be analysed and eliminated similarly. We show in the experiments later that such a treatment effectively reduces the estimation bias.

As a summary, when estimating the coefficient a_k of a frequency component, we should minus all other frequency components to reduce the estimation bias as much as possible. Notably, the coefficients for frequency components $k+1, k+2, \dots, k_{max}$ are unknown if we perform the estimation in turn from 1 to k . One feasible solution is to perform the estimation iteratively. Concretely, in the first round, we only minus the frequency components 1 to $k-1$ when estimating a_k . Then in the second round we can minus all other components. The iteration is ended until the difference of the estimated coefficients in the adjacent round becomes small enough. In the main paper, only one round of estimation is performed considering the computational overheads. **Notice that because the SH functions also has orthogonal completeness, the above strategy, *i.e.*, estimating the coefficients in turn and eliminating the influences of other frequency components, also makes a difference for estimating SH coefficients as we have done in the main paper.**

We perform numerical experiments about Fourier series to support our claim. We compare three estimators, including the original Monte Carlo estimator, Monte Carlo estimator using our bias reducing strategy, and the estimator based on the least square method. Specifically, the least square solution is

$$\begin{pmatrix} 1 & \cos(x_1) & \sin(x_1) & \cdots & \cos(k_{max}x_1) & \sin(k_{max}x_1) \\ 1 & \cos(x_2) & \sin(x_2) & \cdots & \cos(k_{max}x_2) & \sin(k_{max}x_2) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & \cos(x_T) & \sin(x_T) & \cdots & \cos(k_{max}x_T) & \sin(k_{max}x_T) \end{pmatrix}^\dagger \begin{pmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_T) \end{pmatrix}, \quad (32)$$

where \dagger denotes pseudo-inverse. We make sure that the number of samples T is larger than the number of coefficients to be estimated.

For the target signals, we randomly choose from trigonometric and polynomial functions, and determine their coefficients randomly. As a result, three random target functions are chosen as follows,

$$\begin{aligned} f_1(x) &= 2 + 0.03x^2 + 2 \sin(x) + \cos(3x), \\ f_2(x) &= 10 - 0.02x + 0.01x^2 + \cos(x) - \sin(2x), \\ f_3(x) &= 5 + 0.05x^2 - 0.001x^3 - \sin(x) + 2 \cos(2x). \end{aligned} \quad (33)$$

To perform nonuniform sampling, we sample x_t from a Normal distribution $\mathcal{N}(0, \sigma^2)$. To compare the three estimators quantitatively, we uniformly sample 1000 points from $[-2\sigma, 2\sigma]$. Then, we calculate the root mean squared error (RMSE) of this 1000 points. Because of the randomness of the estimation process, we repeat the estimation for 100,000 times and report the mean RMSE (MRMSE) as the final metric, *i.e.*,

$$\text{MRMSE} = \frac{1}{100000} \sum_{i=1}^{100000} \sqrt{\frac{1}{1000} \sum_{j=1}^{1000} (f(x_{ij}) - \hat{f}_i(x_{ij}))^2}, \quad (34)$$

where x_{ij} is the j^{th} sample in the i^{th} estimation process. $f(\cdot)$ is the ground-truth signal and $\hat{f}_i(\cdot)$ is the estimated signal in the i^{th} estimation process. The lower the MRMSE, the better.

We set $k_{max} = 3$, $\sigma = 10$, and the number of samples for estimation $T = 10, 20, \dots, 100$. The MRMSE results for $f_1(x)$, $f_2(x)$, and $f_3(x)$ are shown in Fig. 10. It is clear that with our bias reducing strategy, the estimation is consistently better than original Monte Carlo estimation. Furthermore, in the case of sparse sampling, the least square method suffers a sharp performance degradation. When only 10 samples are available, it is even worse than the original Monte Carlo estimation. When more samples are given, our method and the least square method become comparable.

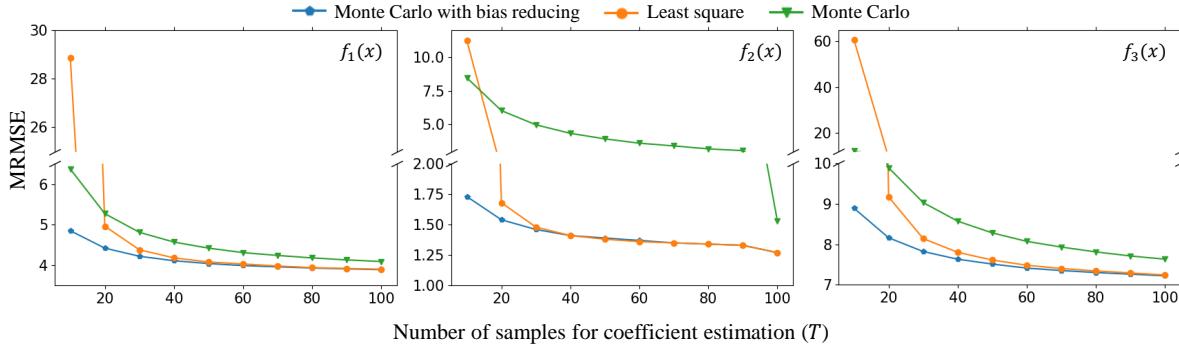


Figure 10. The $\text{MRMSE}\downarrow$ results for $f_1(x)$, $f_2(x)$, and $f_3(x)$.

In addition, we add different direct current values to $f_1(x)$ to increase its A_0 , and fix $T = 100$. The MRMSE results are listed in Tab. 7. It shows that the estimation bias of Monte Carlo estimation increases with a larger A_0 , while our method and the least square method are invariant about such additions.

Table 7. The $\text{MRMSE}\downarrow$ results when adding different direct current values to $f_1(x)$.

Addition	0	1	5	10	50	100
Monte Carlo with bias reducing	3.89	3.89	3.89	3.89	3.89	3.89
Least square	3.90	3.90	3.90	3.90	3.90	3.90
Monte Carlo	4.09	4.17	4.60	5.31	13.53	25.13

Finally, the computational cost of our method, least square method, and original Monte Carlo method for one estimation with 100 samples are 1.196×10^{-4} s, 1.052×10^{-4} s, and 0.978×10^{-4} s respectively. The experiments are performed on an Intel i7-9700 CPU. Overall, we can conclude that our method effectively reduces estimation biases with acceptable computational overheads.

B. More Experimental Details

B.1. Search for the Best Chamfer Distance

To calculate the Chamfer Distance (CD) metric on the DTU dataset, we first transform the density volume into a point-cloud using the marching cubes algorithm. The algorithm searches for iso-surfaces from the volume, where a hyper-parameter about the density level needs to be manually set. The hyper-parameter usually significantly affects the resulting CD value. Therefore, we vary this hyper-parameter and search for the best CD. Specifically, we adopt the golden section search algorithm, which effectively finds the local minimum, or probably the global minimum, of the CD in our case. The searching process is ended when the distance between the last two searched CD values is no greater than 0.001.

As in Fig. 11, we showcase the searched results of each scene reconstructed by JaxNeRF [21, 8], which are based on 512^3 density volume. The searched results first decrease and then increase. The ones of other methods, *i.e.*, DVGO [32], Plenoxels [42], and TensoRF [5], have the similar trend. It is noteworthy that such a searching process is time-consuming, which costs tens of minutes or more, whereas the calculation of our Inverse Mean Residual Color (IMRC) metric is more efficient as reported in Appendix B.2.

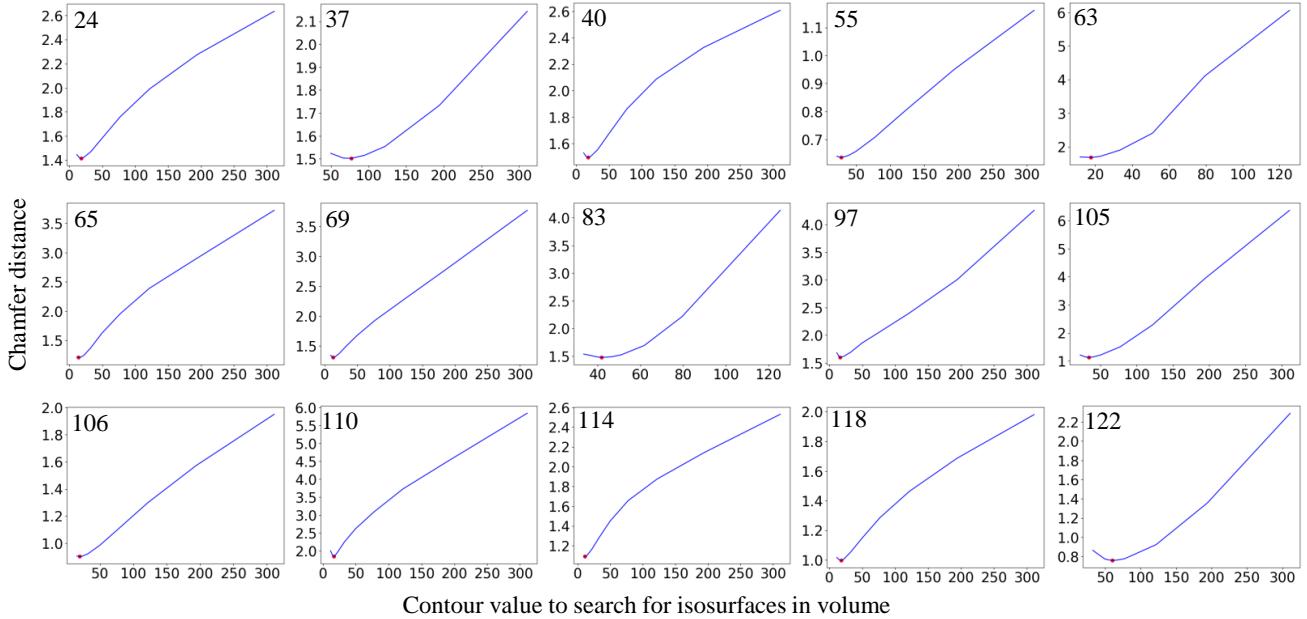


Figure 11. The searched results of the Chamfer Distance (CD) for each scene reconstructed by JaxNeRF [21, 8]. The red points indicate the best CD values. The scan ID of the scene is listed on the left-top corner of each sub-figure.

B.2. Computational Cost of the IMRC

As in Tab. 8, we report the average time cost of calculating the IMRC metric on the DTU dataset's 15 scenes with density volume resolution 512^3 and SH degree 2. All the experiments are run on a single NVIDIA A100 GPU. The average computation costs of the 4 methods are different since the sparsity of the density volume varies with different methods. More sparse the density volume is, less time the computation process costs. Compared with the CD metric, the calculation of IMRC is much more efficient as there is no searching process.

Table 8. Average time cost of calculating the IMRC on the DTU dataset.

Method	JaxNeRF [21, 8]	Plenoxels [42]	DVGO [32]	TensoRF [5]
Time cost (seconds)	17.80	16.98	10.66	16.50

C. More Experimental Results

C.1. Analysis on Reflective Objects and Typical Geometry Artifacts

C.1.1 Reflective Object

The IMRC metric is based on the low-frequency color prior. We use SH of degree 2 to approximate the view-dependent colors from different observation directions. The degree 2, rather than 0, ensures that we can deal with non-Lambertian surfaces. The higher the SH degree, the better the non-Lambertian surfaces be approximated. With degree 2, the highly reflective surface may have a high residual color. However, we are not intended to get a perfect approximation, which is not necessary. **The IMRC makes sense if it can correctly rank the geometry of the same scene produced by different methods.** We demonstrate this by a reflective object as shown in Fig. 12. The surface of the scissors is highly reflective, which leads to relatively high residual colors for a visually good geometry produced by JaxNeRF. Notice that because degree 2 is applied to all methods, it is a fair treatment that all method has high residual colors on such a surface. More importantly, if one method produces worse geometry than that of JaxNeRF, its residual color will be much higher. This makes IMRC successfully distinguish and rank scene geometry even for a reflective case.

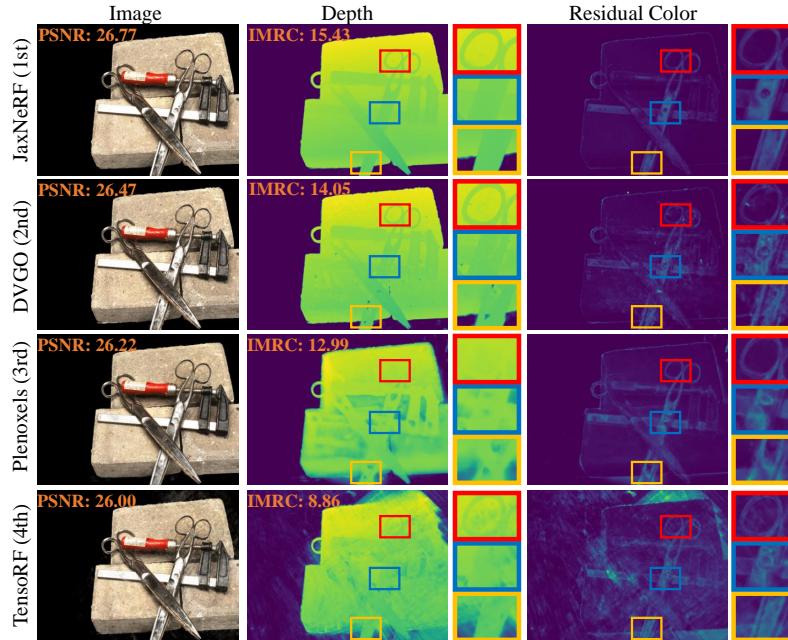


Figure 12. A reflective scene (Scan 37) in the DTU dataset. The UserRank becomes worse from top to down.

C.1.2 Thick Surface

The thick surface is a typical geometry artifact that would generate inaccurate disparity maps viewed from different locations. Because CD is only based on an iso-surface at a typical density level, it may not be aware about the thickness of a surface. In contrast, IMRC can well recognize such an artifact, because points on the thick surface always have higher residual colors than those that lie nearer to the true surface. We showcase two cases that the IMRC metric penalties thick surfaces in Fig. 13. A sharper surface is also presented for comparison.

C.1.3 Floating Surface

The floating surface is a common artifact in NeRF models. The floaters violate the low-frequency color prior, and so will have high residual colors. We showcase two cases that the IMRC metric penalties floating surfaces in Fig. 14. The better geometry without floating surfaces is also presented for comparison.

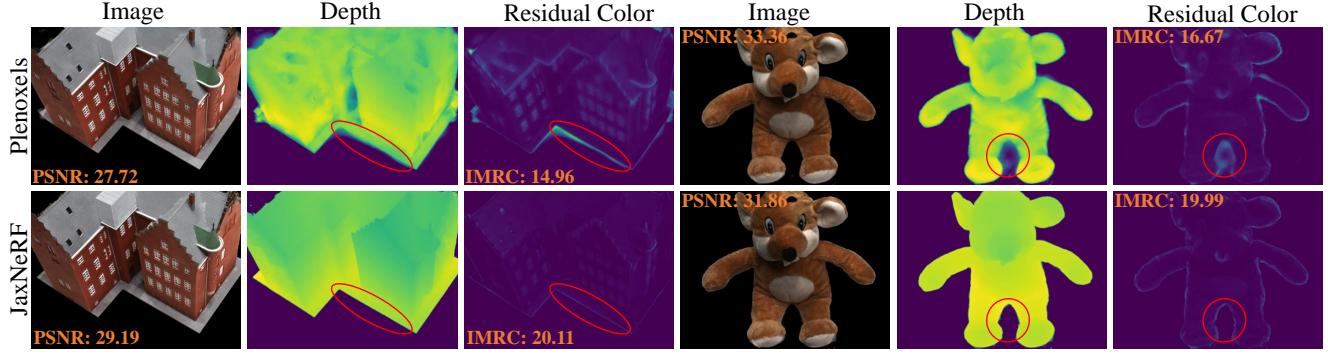


Figure 13. Two cases of thick surfaces (top row) on the DTU dataset. From left to right, Scan 24 and Scan 105. Corresponding sharp surface results (bottom row) are also presented for comparison.

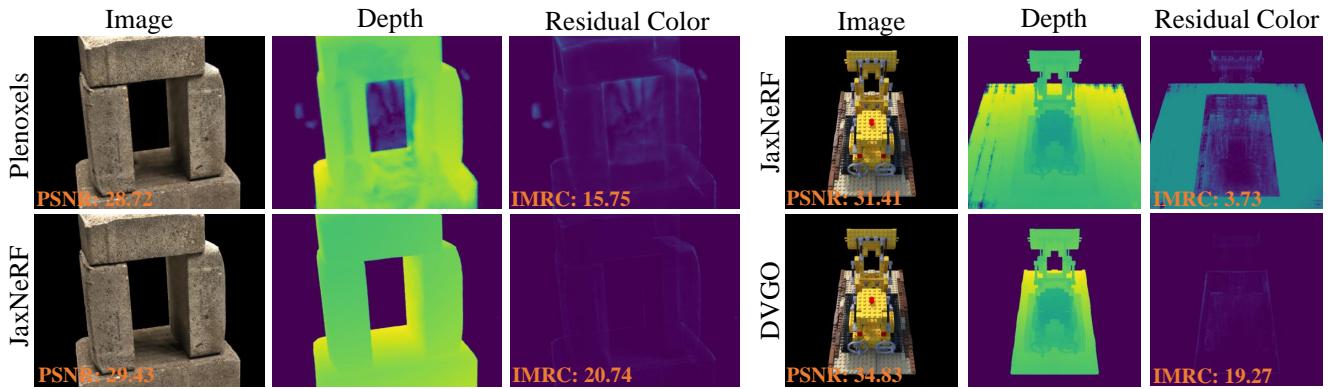


Figure 14. Two cases of floating surfaces (top row). From left to right, Scan 40 in the DTU dataset and Lego in the NeRF Synthetic dataset. Corresponding clean surface results (bottom row) are also presented for comparison.

C.2. Remaining IMRC/UserRank and CD/UserRank Conflict Results

We showcase one remaining $\text{IMRC}^\uparrow/\text{UserRank}$ conflict case in Fig. 15. On the whole, JaxNeRF produces a thicker surface than Plenoxels as its residual color is always higher surrounding the contour. The Plenoxels produces some floating surfaces and perform worse on certain local details. Overall, the density fields of both methods are not good. After applying the marching cubes algorithm, the floating surface of JaxNeRF’s density field near the right hand of the doll is missing. In contrast, some low density parts of Plenoxels’ density field are discarded, resulting in a poor mesh. Comprehensively considering the depth, residual color, and mesh, the users rank JaxNeRF as better, although the calculation results show that the Plenoxels has a lower residual color, or higher IMRC.

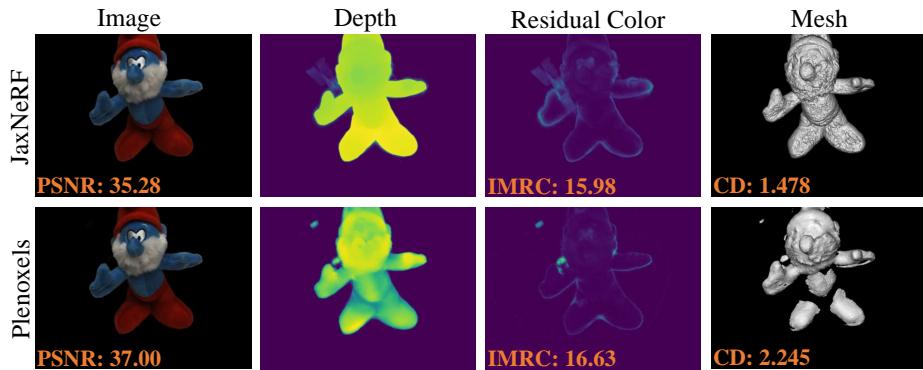


Figure 15. One remaining $\text{IMRC}^\uparrow/\text{UserRank}$ conflict case on the Scan 83 scene. The top row has a better UserRank.

We showcase 9 remaining $CD \downarrow$ /UserRank conflict cases in Fig. 16 to Fig. 18. We have analysed the two main reasons that cause conflicts in Sec. 5.1. of the main paper. On one hand, because of the marching cubes algorithm, the CD metric fails to recognize some thick surfaces, and some low density surface points are discarded. On the other hand, the object mask used in the calculation neglects some floating meshes and leads to an unfair comparison.

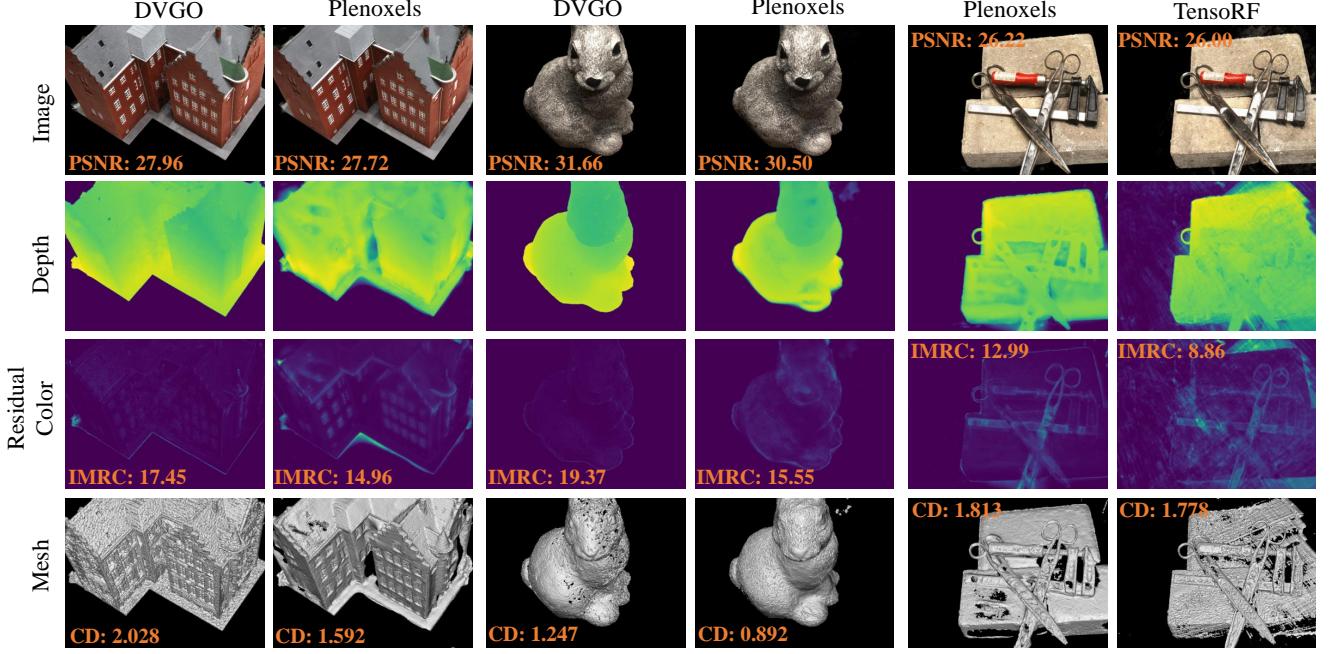


Figure 16. Three $CD \downarrow$ /UserRank conflict cases. From left to right, DTU Scan 24, 55, and 37. For each scene, left has a better UserRank.

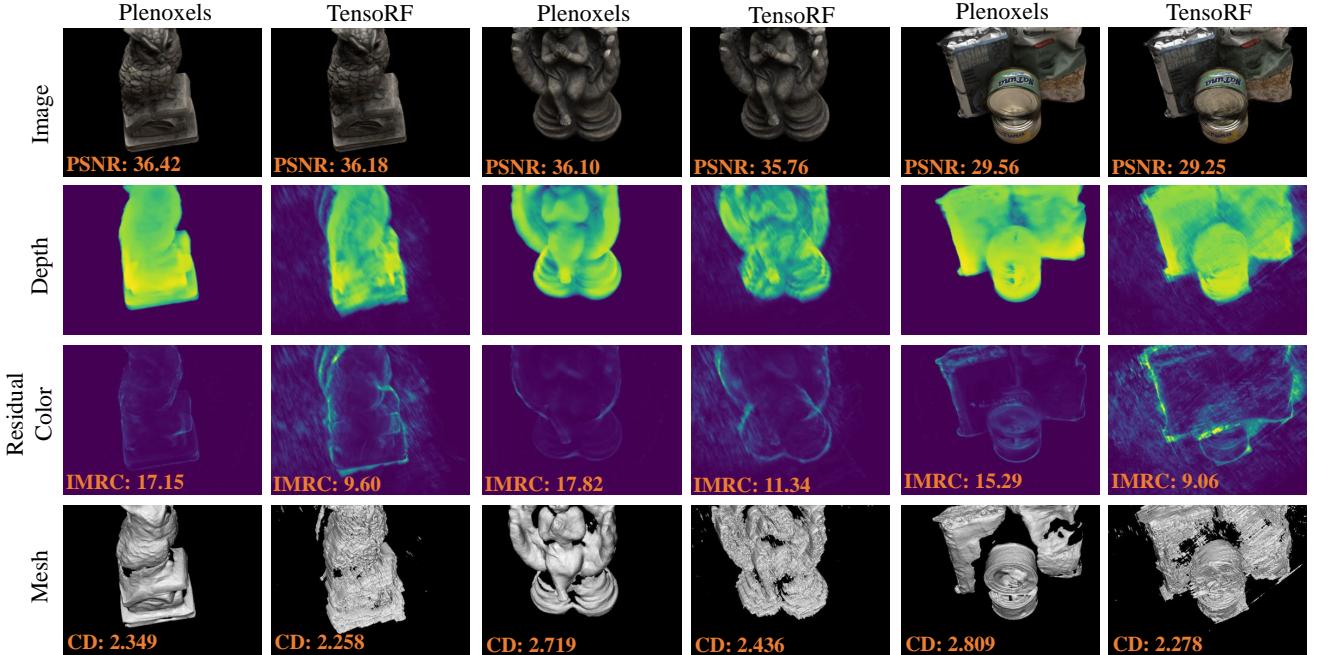


Figure 17. Three $CD \downarrow$ /UserRank conflict cases. From left to right, DTU Scan 122, 118, and 97. For each scene, left has a better UserRank.

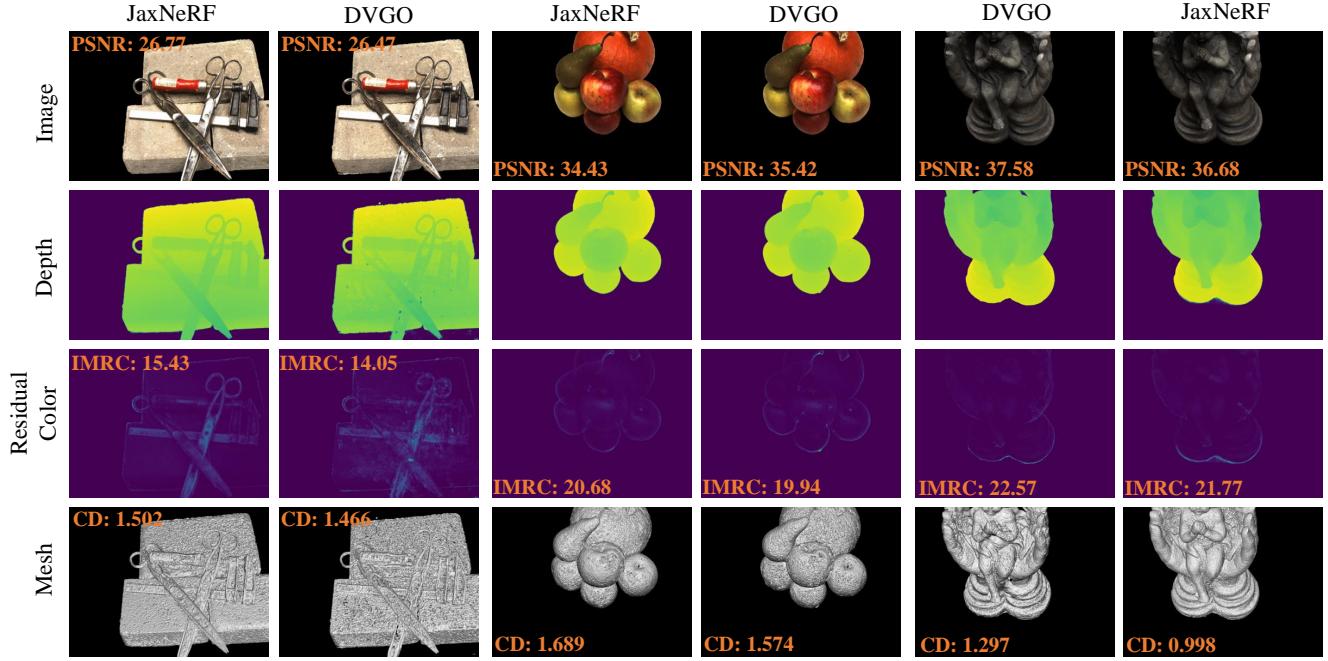


Figure 18. Three $CD \downarrow$ /UserRank conflict cases. From left to right, DTU Scan 37, 63, and 118. For each scene, left has a better UserRank.

C.3. More CD/IMRC/UserRank Consistent Results

On the DTU dataset, except for the 2 $IMRC \uparrow$ /UserRank and 11 $CD \downarrow$ /UserRank conflict cases, the remaining 77 pairs all have consistent $CD \downarrow$ / $IMRC \uparrow$ /UserRank. We illustrate more consistent cases in Fig. 19 to Fig. 23. We find that both CD and IMRC successfully reflect the quality of the density field in these cases.

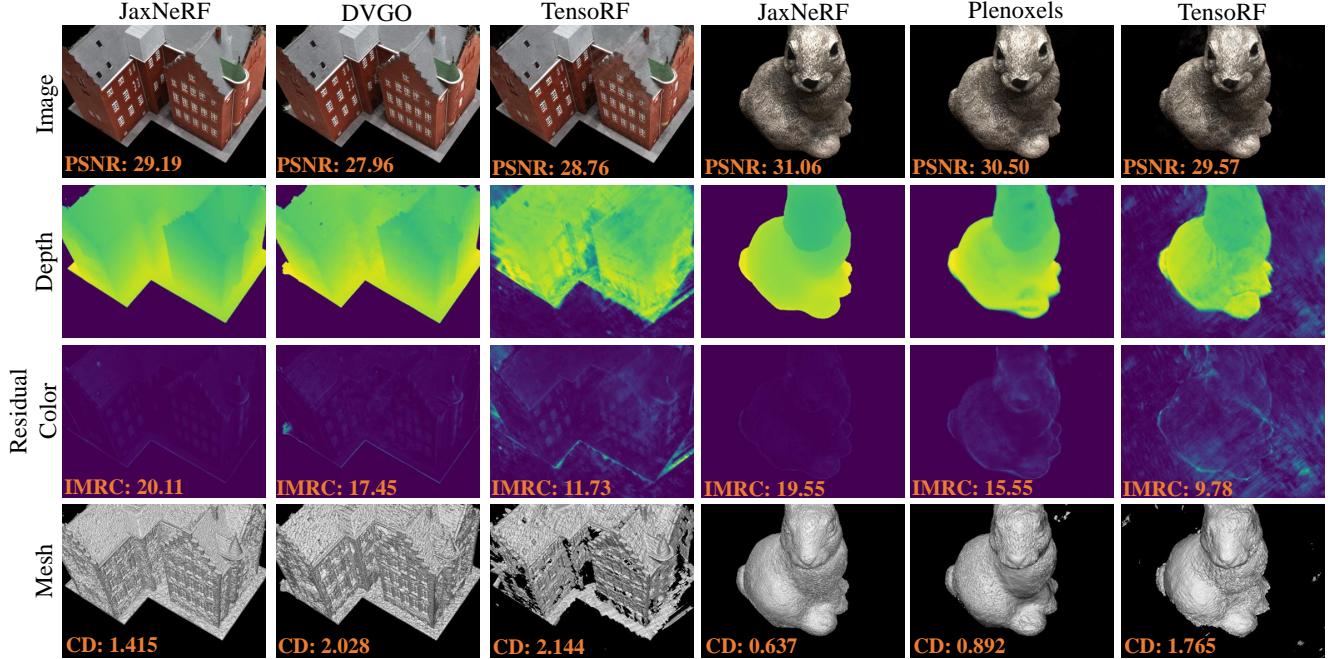


Figure 19. More $CD \downarrow$ / $IMRC \uparrow$ /UserRank consistent results on the DTU dataset. From left to right, Scan 24 and 55. For each scene, the quality of the density field decreases from left to right.

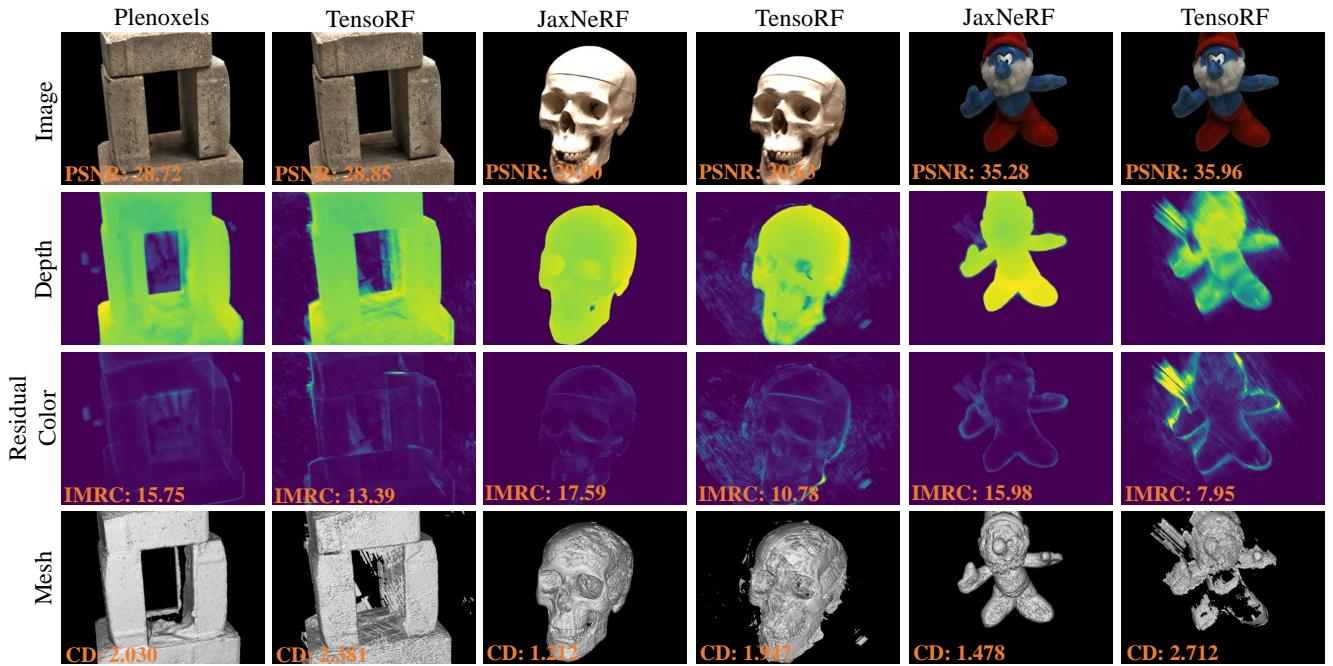


Figure 20. More CD↓/IMRC↑/UserRank consistent results on the DTU dataset. From left to right, Scan 40, 65, and 83. For each scene, the quality of the density field decreases from left to right.

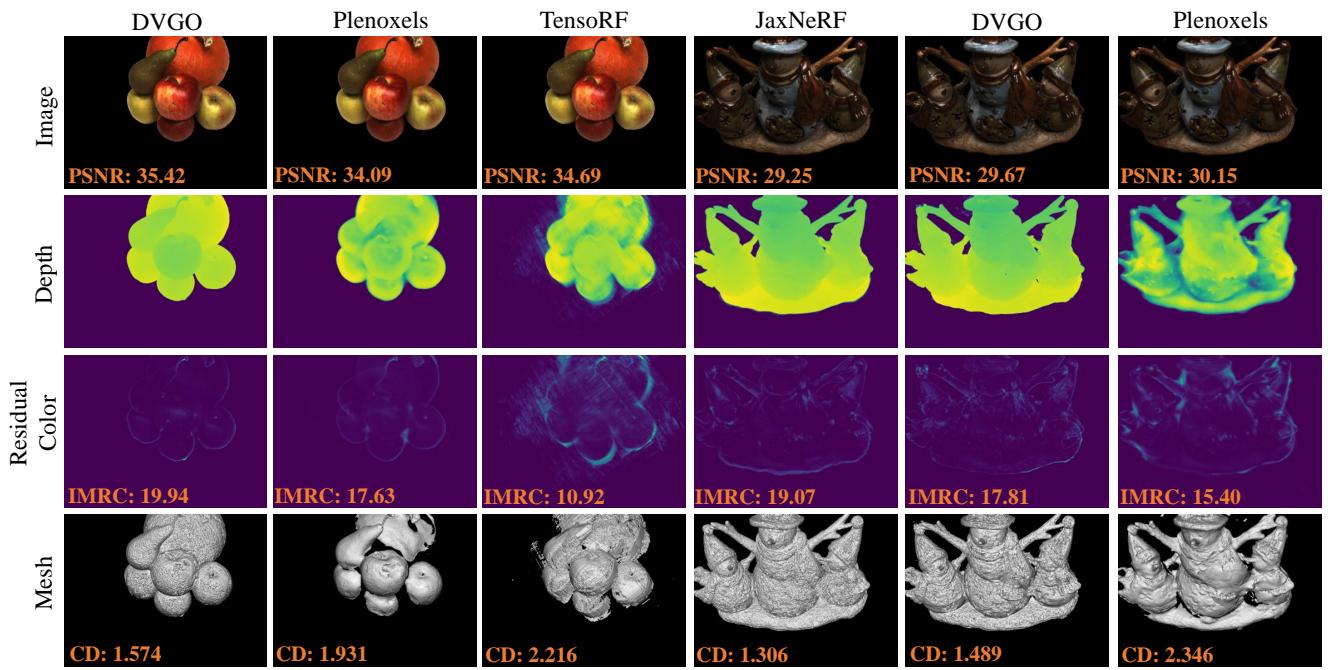


Figure 21. More CD↓/IMRC↑/UserRank consistent results on the DTU dataset. From left to right, Scan 63 and 69. For each scene, the quality of the density field decreases from left to right.

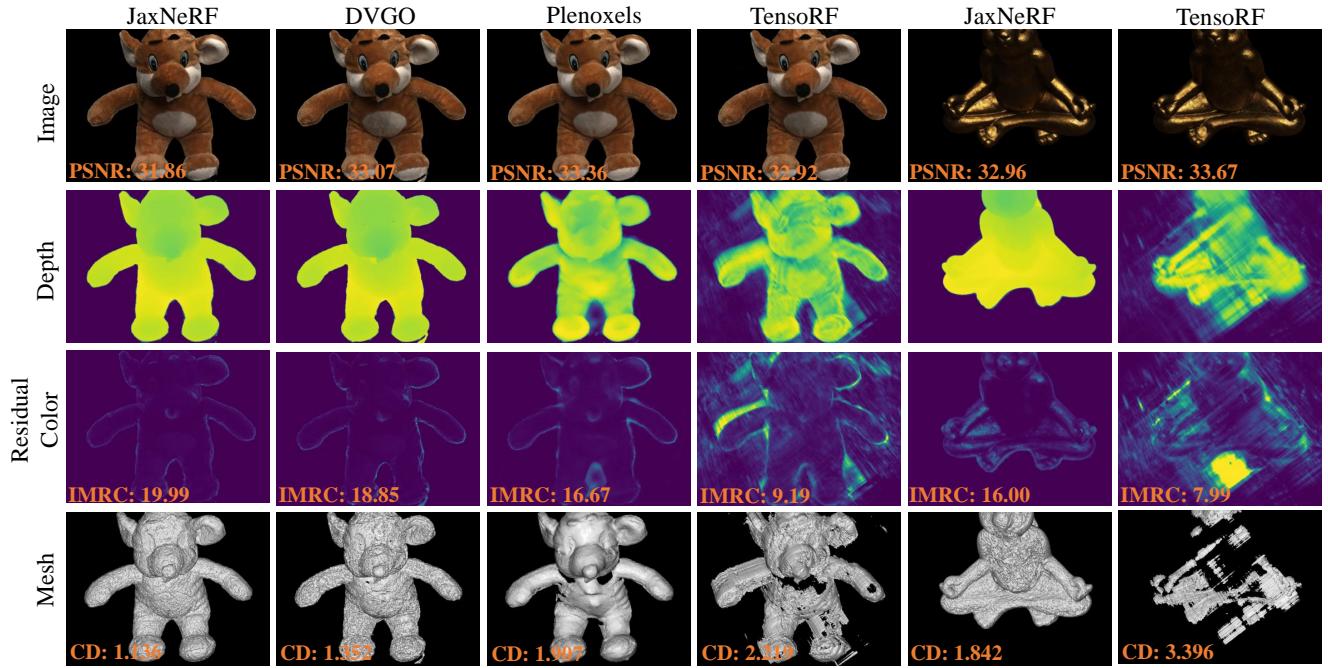


Figure 22. More CD↓/IMRC↑/UserRank consistent results on the DTU dataset. From left to right, Scan 105 and 110. For each scene, the quality of the density field decreases from left to right.

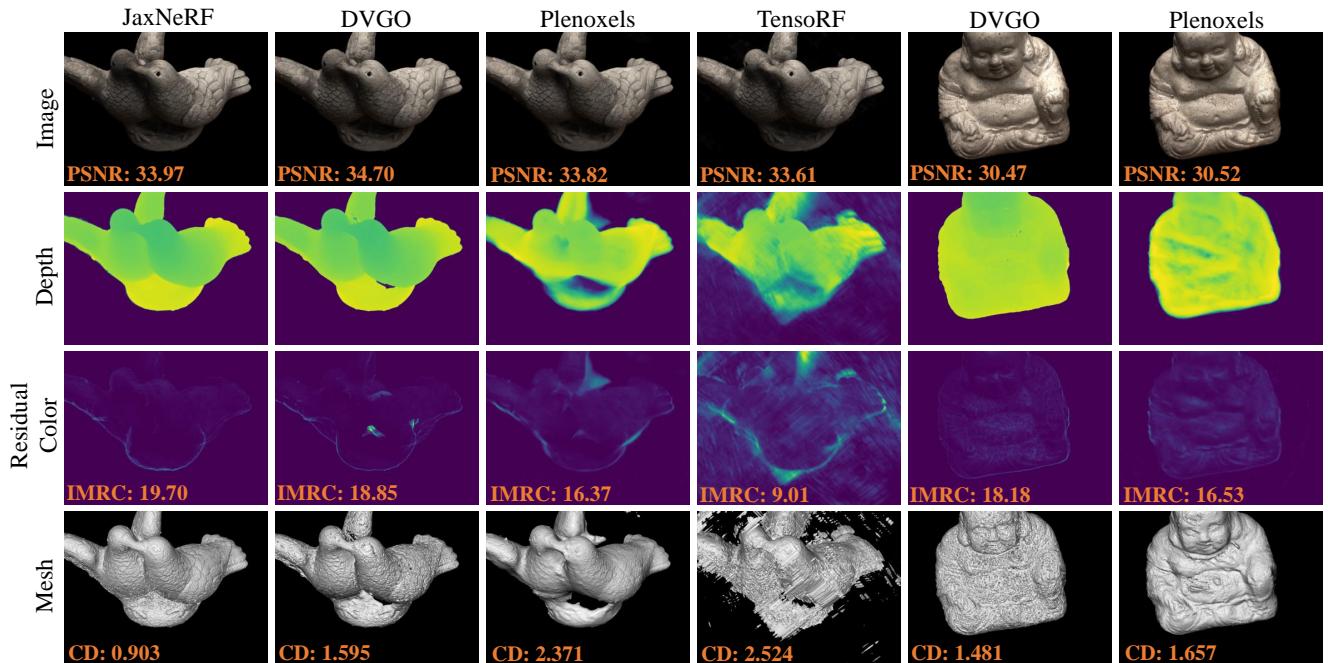


Figure 23. More CD↓/IMRC↑/UserRank consistent results on the DTU dataset. From left to right, Scan 106 and 114. For each scene, the quality of the density field decreases from left to right.