NC-SDF: Enhancing Indoor Scene Reconstruction Using Neural SDFs with View-Dependent Normal Compensation

Supplementary Material

In Sec. 1, we present the network architecture of our NC-SDF, along with details of our normal compensation model (Sec. 1.1) and hybrid geometry model (Sec. 1.2). In Sec. 2, we provide details of the training process. In Sec. 3, we define evaluation metrics used in the main paper and the supplementary material. In Sec. 4, we report additional qualitative and quantitative results. Finally, we discuss future work in Sec. 6.

1. Network architecture

Our network takes the spatial position of point $\boldsymbol{x}$, the view direction $\boldsymbol{v}$ as inputs, and outputs the signed distance s, the radiance $\boldsymbol{c}$, and the compensation rotation angles γ, β, θ. The network architecture of our NC-SDF is illustrated in Fig. 1.

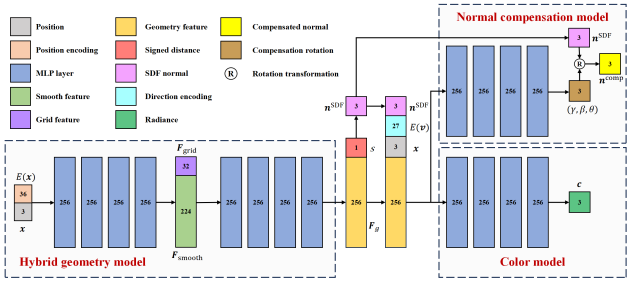

Figure 1. **Network architecture of NC-SDF.**

1.1. Details of the normal compensation model

In the normal compensation model, we utilize the ReLU function [6] as the activation function for the hidden layers and the HardTanh function [2] for the output layer. The ReLU function is defined in Eq. (1), and the HardTanh function is defined in Eq. (2). Specifically, the rotation angle can be expressed as $180° \times \text{HardTanh}(x)$. We initialize the model to produce rotations that are close to zero at the start of the training process.

$$\text{ReLU}(x) = \begin{cases} x, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases}, \tag{1}$$

$$\text{HardTanh}(x) = \begin{cases} 1, & \text{if } x > 1 \\ x, & \text{if } -1 \leq x \leq 1 \\ -1, & \text{if } x < -1 \end{cases}. \tag{2}$$

Our normal compensation model outputs the rotation angles γ, β, and θ, corresponding to the x, y, and z axes. The rotation process is described in Eq. (3) and Eq. (4).

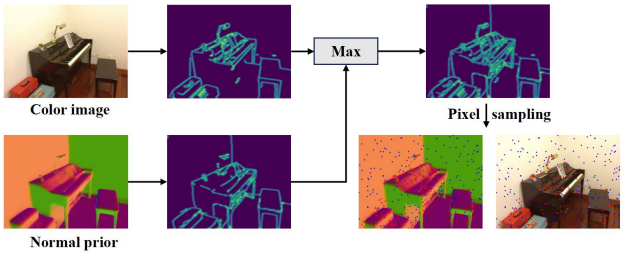

Figure 2. **The process of texture extraction.** The blue points represent the pixels from random sampling. The red points represent the pixels from informative sampling.

1.2. Details of the hybrid geometry model

Following Instant-NGP [12] and MonoSDF [19], we utilize multi-resolution voxel grids for the grid feature branch in the hybrid geometry model. Supposed grids with L layers:

$$R_l := \lfloor R_{\min} b^l \rfloor \quad b := \exp\left(\frac{\ln R_{\max} - \ln R_{\min}}{L-1}\right), \tag{5}$$

where $R_{\min}$, $R_{\max}$ represent the coarsest and finest resolutions, respectively. Each grid can accommodate up to T feature vectors with a dimensionality of F. If $R_l^3 > T$, we employ a spatial hash function [15] to index the corresponding feature vector:

$$h(\mathbf{x}) = \left(\bigoplus_{i=1}^{3} \mathbf{x}_i \pi_i\right) \bmod T, \tag{6}$$

where $\bigoplus$ denotes the bit-wise XOR operation and π_i are unique, large prime numbers. For our grid feature branch, we utilize $L = 8$, $F = 4$, $R_{\min} = 64$, $R_{\max} = 512$, $T = 256$.

2. Training details

2.1. Data preparation

For each scene in ScanNet [1] and ICL-NUIM [8], we uniformly sample one-tenth of views from the corresponding video frames. We obtain a collection of approximately 100 to 500 images per scene. The normal priors are obtained from the Omnidata model [3].

The process of informative pixel sampling is illustrated in Fig. 2. Firstly, we utilize the Canny operator to extract texture from a color image, obtaining a texture intensity map $\mathbf{I}_c$. Simultaneously, we extract a texture intensity map $\mathbf{I}_n$ from the normal prior. These two intensity maps, after

$$\boldsymbol{n}^{\text{comp}} = \mathbf{R}_{\text{ZYX}}\boldsymbol{n}^{\text{SDF}}, \tag{3}$$

where

$$\begin{aligned}\mathbf{R}_{\text{ZYX}}(\gamma,\beta,\theta) &= \mathbf{R}_{\text{Z}}(\theta)\mathbf{R}_{\text{Y}}(\beta)\mathbf{R}_{\text{X}}(\gamma) \\ &= \begin{bmatrix}\cos(\theta) & -\sin(\theta) & 0\\ \sin(\theta) & \cos(\theta) & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}\cos(\beta) & 0 & \sin(\beta)\\ 0 & 1 & 0\\ -\sin(\beta) & 0 & \cos(\beta)\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\ 0 & \cos(\gamma) & -\sin(\gamma)\\ 0 & \sin(\gamma) & \cos(\gamma)\end{bmatrix} \\ &= \begin{bmatrix}\cos(\theta)\cos(\beta) & \cos(\theta)\sin(\beta)\sin(\gamma)-\sin(\theta)\cos(\gamma) & \cos(\theta)\sin(\beta)\cos(\gamma)+\sin(\theta)sin(\gamma)\\ \sin(\theta)\cos(\beta) & \sin(\theta)\sin(\beta)\sin(\gamma)+\cos(\theta)\cos(\gamma) & \sin(\theta)sin(\beta)\cos(\gamma)-\cos(\theta)\sin(\gamma)\\ -\sin(\beta) & \cos(\beta)\sin(\gamma) & \cos(\beta)\cos(\gamma)\end{bmatrix}.\end{aligned} \tag{4}$$

normalization, are merged by selecting the maximum value for each pixel:

$$\mathbf{I}_{(i,j)} = \max(\mathbf{I}_{c,(i,j)}, \mathbf{I}_{n,(i,j)}), \tag{7}$$

where (i, j) represents the location of the pixel.

This approach is beneficial since normal priors inherently encapsulate edge information. Regions at edges in normal priors often exhibit multi-view inconsistency. By prioritizing sampling in such regions, our normal compensation model enables more consistent results through adaptive compensation for the view-dependent biases.

2.2. Training process

Our pixel sampling strategy follows a coarse-to-fine approach throughout the entire training process. As mentioned in the main paper, pixel sampling is divided into two parts: a proportion r is allocated to informative pixel sampling, while the remaining $1 - r$ is reserved for random sampling. We set an intensity threshold l_i for texture intensity maps. Specifically, we define four intensity thresholds, where $l_0 = 0$, $l_1 = 0.1$, $l_2 = 0.2$, and $l_3 = 0.3$. And we denote the current iteration count as N_{step}. In the informative sampling, we randomly select $r \times N_{\text{sample}}$ pixels from the set $\{l \mid l \geq l_i, \text{where } i = \min(\lfloor \frac{N_{\text{step}}}{10{,}000} \rfloor, 3)\}$ to create a high-information pixel set $\mathcal{P}_{\text{canny}}$. In the random sampling, we randomly sample $(1 - r) \times N_{\text{sample}}$ pixels from all the pixels, forming the set $\mathcal{P}_{\text{random}}$. Finally, the pixel set for each batch is defined as $\mathcal{P}_{\text{all}} = \mathcal{P}_{\text{canny}} \cup \mathcal{P}_{\text{random}}$. The proportion r gradually increases from 0% to 50% as the iteration count N_{step} grows. Besides, we introduce some randomness into the informative pixel sampling to ensure the sampling of pixels near the edges.

2.3. Training loss curves

In Fig. 3, we compare the training loss curves when training with and without the NC model. The results indicate that our NC model can effectively separate noise in the normal priors, leading to more accurate geometry and radiance fields. Fig. 4 in the main paper provides further confirmation.

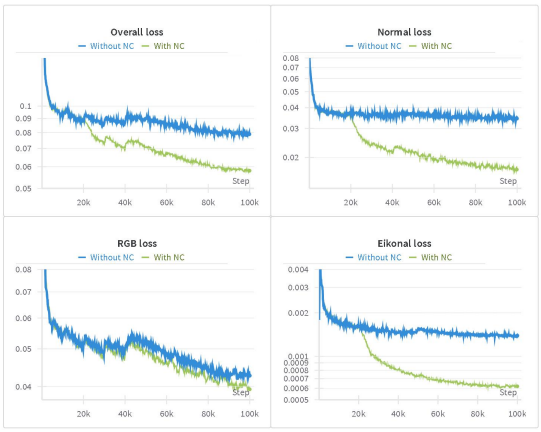

Figure 3. Training loss curves. The x-axis represents the training iteration number, the y-axis represents the **logarithmic** values of the loss. The optimization of the normal compensation model begins at the **20,000 (20k)** training iteration.

3. Evaluation metrics

3.1. Evaluation metrics used in the main paper

In Tab. 1, we provide the definitions of evaluation metrics employed in the main paper for 3D reconstruction quality.

3.2. Evaluation metrics used in the supplementary material

For a more comprehensive evaluation, we incorporate additional evaluation metrics in the supplementary material. The corresponding results are provided in Sec. 4.

Additional evaluation metrics for 3D reconstruction. We incorporate two additional commonly used evaluation metrics for 3D reconstruction in Tab. 2. The Chamfer L1 distance [13] is a measure of dissimilarity between point clouds.

Normal consistency [9] measures how well 3D reconstruction can capture higher-order information by calculating a mean absolute dot product of face-normals within the given 3D meshes.

Evaluation metrics for 2D normal and depth maps. We provide definitions of evaluation metrics for normal maps in Tab. 3 and for depth maps in Tab. 4. We employ the normal evaluation metrics, following [5]. The evaluation is based on the angular difference between predicted and ground truth (GT) normals. The two error measurements, mean error and root-mean-squared error (RMSE), quantify the disparity between predicted and GT normals, thus lower is better. We also present three pixel-accuracy metrics which represent the fraction of pixels with cosine distances to GT normals less than specified thresholds: $11.25°$, $22.5°$, and $30°$, with higher values indicating better performance. For depth evaluation, we utilize the metrics in [4]. The three error measurements, absolute relative error (Abs Rel), square relative error (Sq Rel), and RMSE, quantify the disparity between predicted and GT depths. Additionally, we provide a depth pixel-accuracy metric with a threshold $\delta = 1.25$.

Table 1. **Evaluation metrics for 3D reconstruction quality.** P and P^* are the points sampled from the predicted and GT mesh.

Metric	Definition
Acc	$\text{mean}_{\boldsymbol{p}\in P}\left(\min_{\boldsymbol{p}^*\in P^*}\|\boldsymbol{p}-\boldsymbol{p}^*\|\right)$
Comp	$\text{mean}_{\boldsymbol{p}^*\in P^*}\left(\min_{\boldsymbol{p}\in P}\|\boldsymbol{p}-\boldsymbol{p}^*\|\right)$
Prec	$\text{mean}_{\boldsymbol{p}\in P}\left(\min_{\boldsymbol{p}^*\in P^*}\|\boldsymbol{p}-\boldsymbol{p}^*\| < 0.05\right)$
Recall	$\text{mean}_{\boldsymbol{p}^*\in P^*}\left(\min_{\boldsymbol{p}\in P}\|\boldsymbol{p}-\boldsymbol{p}^*\| < 0.05\right)$
F-score	$\frac{2\times \text{Prec}\times\text{Recall}}{\text{Prec}+\text{Recall}}$

Table 2. **Additional evaluation metrics for 3D reconstruction quality.** P and P^* are the points sampled from the predicted and GT mesh. $\boldsymbol{n}$ represents unit normal vector. $\text{proj}_2(\boldsymbol{p})$ and $\text{proj}_1(\boldsymbol{p}^*)$ represent the projections of $\boldsymbol{p}$ and $\boldsymbol{p}^*$ onto the GT mesh and the predicted mesh, respectively.

Metric	Definition
Chamfer L1	$\frac{\text{Acc}+\text{Comp}}{2}$
Normal C	$\frac{1}{2}(\text{mean}_{\boldsymbol{p}\in P}(\lvert\boldsymbol{n}(\boldsymbol{p})\cdot\boldsymbol{n}(\text{proj}_2(\boldsymbol{p}))\rvert)+\text{mean}_{\boldsymbol{p}^*\in P^*}(\lvert\boldsymbol{n}(\boldsymbol{p}^*)\cdot\boldsymbol{n}(\text{proj}_1(\boldsymbol{p}^*))\rvert))$

4. Additional results

In this section, we provide additional results of comparison experiments (Sec. 4.1), ablation studies (Sec. 4.2), novel view synthesis (Sec. 4.3) and results on other dataset (Sec. 5).

Table 3. **Evaluation metrics for 2D normal maps.** n is the number of pixels with valid normal in the GT normal map. $\boldsymbol{n}$ and $\boldsymbol{n}^*$ are the predicted and GT normals.

Metric	Definition
Mean	$\frac{1}{n}\sum \cos^{-1}\left[\frac{\lvert\boldsymbol{n}\cdot\boldsymbol{n}^*\rvert}{\lvert\boldsymbol{n}\rvert\lvert\boldsymbol{n}^*\rvert}\right]$
RMSE	$\sqrt{\frac{1}{n}\sum\left(\cos^{-1}\left[\frac{\lvert\boldsymbol{n}\cdot\boldsymbol{n}^*\rvert}{\lvert\boldsymbol{n}\rvert\lvert\boldsymbol{n}^*\rvert}\right]\right)^2}$
$11.25°$	$\frac{1}{n}\#\left\{\boldsymbol{n},\boldsymbol{n}^* : \cos^{-1}\left[\frac{\lvert\boldsymbol{n}\cdot\boldsymbol{n}^*\rvert}{\lvert\boldsymbol{n}\rvert\lvert\boldsymbol{n}^*\rvert}\right] < 11.25°\right\}$
$22.5°$	$\frac{1}{n}\#\left\{\boldsymbol{n},\boldsymbol{n}^* : \cos^{-1}\left[\frac{\lvert\boldsymbol{n}\cdot\boldsymbol{n}^*\rvert}{\lvert\boldsymbol{n}\rvert\lvert\boldsymbol{n}^*\rvert}\right] < 22.5°\right\}$
$30°$	$\frac{1}{n}\#\left\{\boldsymbol{n},\boldsymbol{n}^* : \cos^{-1}\left[\frac{\lvert\boldsymbol{n}\cdot\boldsymbol{n}^*\rvert}{\lvert\boldsymbol{n}\rvert\lvert\boldsymbol{n}^*\rvert}\right] < 30°\right\}$

Table 4. **Evaluation metrics for 2D depth maps.** n is the number of pixels with valid depth in the GT depth map. d and d^* are the predicted and GT depths.

Metric	Definition
Abs Rel	$\frac{1}{n}\sum\frac{\lvert d-d^*\rvert}{d^*}$
Sq Rel	$\frac{1}{n}\sum\frac{\lvert d-d^*\rvert^2}{d^*}$
RMSE	$\sqrt{\frac{1}{n}\sum\lvert d-d^*\rvert^2}$
1.25	$\frac{1}{n}\#\left\{d,d^* : \max\left(\frac{d^*}{d},\frac{d}{d^*}\right) < 1.25\right\}$

Table 5. Quantitative comparison of reconstruction quality on ScanNet.

Method	F-score↑	Normal C↑	Chamfer L1↓
COLMAP	0.548	0.691	0.136
VolSDF	0.430	0.674	0.108
NeuS	0.320	0.639	0.175
ManhattanSDF	0.706	0.840	0.049
HelixSurf	0.756	0.866	0.039
NeuRIS	0.684	0.869	0.050
MonoSDF(MLP)	0.748	0.882	0.040
MonoSDF(Grid)	0.693	0.844	0.049
Ours	**0.781**	**0.895**	**0.036**

4.1. Comparison experiments

We provide quantitative comparison using additional evaluation metrics on ScanNet. The quantitative comparison of reconstruction quality can be found in Tab. 5, and the quantitative comparison of normal and depth maps is provided

Table 6. Ablation studies of reconstruction quality on Scannet.

Method	F-score↑	Normal C↑	Chamfer L1↓
MLP(baseline)	0.733	0.880	0.041
Grid(baseline)	0.700	0.857	0.047
Hybrid	0.745	0.880	0.040
MLP+IPS	0.742	0.882	0.040
Hybrid+IPS	0.749	0.881	0.040
Ours	**0.781**	**0.895**	**0.036**

in Tab. 7. Furthermore, we present additional visualization results. In Fig. 5, we show top views of reconstructions on ScanNet. We visualize more reconstruction details on both ScanNet (Fig. 6) and ICL-NUIM (Fig. 8). Besides, we present the corresponding rendered results of our NC-SDF on both ScanNet (Fig. 7) and ICL-NUIM (Fig. 9). Both quantitative and qualitative results demonstrate that our NC-SDF achieves the best and the most comprehensive performance.

4.2. Ablation studies

We also present the results of ablation studies using additional evaluation metrics. The ablation studies of reconstruction quality are shown in Tab. 5, and the ablation studies of normal and depth maps are shown in Tab. 7. Moreover, we visualize more reconstruction details for ablation studies in Fig. 10.

The combination of the informative pixel sampling and the hybrid geometry model yields a modest improvement in the quality of reconstructions, normal maps, and depth maps. However, the extent of this improvement is limited by multi-view inconsistent normal priors. The introduction of the normal compensation model in our NC-SDF effectively mitigates the adverse effects of inconsistent normal supervision. This leads to a substantial improvement in results, especially in terms of enhanced normal consistency and the quality of normal maps.

4.3. Novel view synthesis

We randomly select several views that are not included in the training dataset for novel view synthesis. The rendered color images and rendered normal maps are shown in Fig. 11. The combination of our informative pixel sampling and hybrid geometry model improves the quality of novel view synthesis. The introduction of our normal compensation model contributes to a more accurate geometry field and radiance field, thereby elevating the overall synthesis quality, particularly in regions with inconsistent supervision.

5. Results on other dataset

Fig. 4 shows the results on Tanks & Temples dataset [10]. Our NC-SDF outperforms baselines, providing more geometric details.

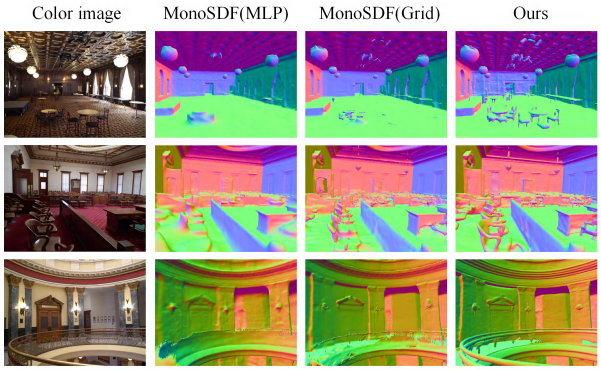

Figure 4. Reconstruction results on Tanks & Temples dataset. The ground truth meshes have not been made publicly available.

6. Future work

In this work, we assume that the poses from datasets are accurate, although, in reality, this is often not the case. Our normal compensation model is built upon the premise that color images are able to provide reliable supervision for correcting the biases in normal priors. However, inaccurate poses can lead to erroneous normal compensation, which ultimately yields suboptimal reconstructions. As part of future work, our objective is to integrate pose optimization into the framework to further enhance the reconstruction results. In addition, it is intriguing to explore the integration of other geometric priors, such as monocular depth priors and semantic priors, into neural implicit representations in an adaptive way in the future.

References

[1] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5828–5839, 2017. 1

[2] Petr Doležel and Ivan Taufer. Piecewise-linear artificial neural networks for pid controller tuning. *Acta Montanistica Slovaka*, pages 224–233, 2012. 1

[3] Ainaz Eftekhar, Alexander Sax, Jitendra Malik, and Amir Zamir. Omnidata: A scalable pipeline for making multi-task mid-level vision datasets from 3d scans. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10786–10796, 2021. 1

[4] David Eigen, Christian Puhrsch, and Rob Fergus. Depth map prediction from a single image using a multi-scale deep network. *Advances in neural information processing systems*, 27, 2014. 3

[5] David F Fouhey, Abhinav Gupta, and Martial Hebert. Data-driven 3d primitives for single image understanding. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3392–3399, 2013. 3

Table 7. Quantitative comparison of normal and depth maps on ScanNet.

Method	Normal map					Depth map			
	Mean↓	RMSE↓	11.25° ↑	22.5° ↑	30° ↑	Abs Rel↓	Sq Rel↓	RMSE↓	1.25 ↑
COLMAP [14]	42.0	56.8	0.278	0.468	0.559	0.292	0.469	0.814	0.686
VolSDF [18]	29.3	38.3	0.340	0.521	0.615	0.148	0.189	0.381	0.864
NeuS [17]	44.8	53.2	0.128	0.301	0.411	0.228	0.177	0.449	0.674
ManhattanSDF [7]	17.9	28.0	0.561	0.740	0.805	0.050	0.023	0.163	0.959
HelixSurf [11]	17.1	27.0	0.572	0.752	0.818	0.043	0.019	0.154	0.962
NeuRIS [16]	15.9	26.1	0.621	0.781	0.835	0.057	0.030	0.181	0.950
MonoSDF(MLP) [19]	14.0	23.6	0.662	0.803	0.852	0.048	0.019	0.138	0.964
MonoSDF(Grid) [19]	17.7	28.2	0.584	0.742	0.801	0.078	0.038	0.197	0.917
Ours	**13.1**	**23.0**	**0.684**	**0.820**	**0.867**	**0.037**	**0.015**	**0.136**	**0.971**

Table 8. Ablation studies of normal and depth maps on ScanNet.

Method	Normal map					Depth map			
	Mean↓	RMSE↓	11.25° ↑	22.5° ↑	30° ↑	Abs Rel↓	Sq Rel↓	RMSE↓	1.25 ↑
MLP(baseline)	14.7	25.0	0.658	0.796	0.845	0.048	0.022	0.155	0.963
Hybrid+IPS	14.6	24.8	0.655	0.799	0.848	0.047	0.021	0.153	0.963
Ours	**13.1**	**23.0**	**0.684**	**0.820**	**0.867**	**0.037**	**0.015**	**0.136**	**0.971**

[6] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 315–323. JMLR Workshop and Conference Proceedings, 2011. 1

[7] Haoyu Guo, Sida Peng, Haotong Lin, Qianqian Wang, Guofeng Zhang, Hujun Bao, and Xiaowei Zhou. Neural 3d scene reconstruction with the manhattan-world assumption. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5511–5520, 2022. 5

[8] Ankur Handa, Thomas Whelan, John McDonald, and Andrew J Davison. A benchmark for rgb-d visual odometry, 3d reconstruction and slam. In *2014 IEEE international conference on Robotics and automation (ICRA)*, pages 1524–1531. IEEE, 2014. 1

[9] Jiongchao Jin, Akshay Gadi Patil, Zhang Xiong, and Hao Zhang. Dr-kfs: A differentiable visual similarity metric for 3d shape reconstruction. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXI 16*, pages 295–311. Springer, 2020. 3

[10] Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. Tanks and temples. *ACM Transactions on Graphics (TOG)*, 36:1 – 13, 2017. 4

[11] Zhihao Liang, Zhangjin Huang, Changxing Ding, and Kui Jia. Helixsurf: A robust and efficient neural implicit surface learning of indoor scenes with iterative intertwined regularization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13165–13174, 2023. 5

[12] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (ToG)*, 41(4):1–15, 2022. 1

[13] Songyou Peng, Michael Niemeyer, Lars Mescheder, Marc Pollefeys, and Andreas Geiger. Convolutional occupancy networks. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*, pages 523–540. Springer, 2020. 2

[14] Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4104–4113, 2016. 5

[15] Matthias Teschner, Bruno Heidelberger, Matthias Müller, Danat Pomerantes, and Markus H Gross. Optimized spatial hashing for collision detection of deformable objects. In *Vmv*, pages 47–54, 2003. 1

[16] Jiepeng Wang, Peng Wang, Xiaoxiao Long, Christian Theobalt, Taku Komura, Lingjie Liu, and Wenping Wang. Neuris: Neural reconstruction of indoor scenes using normal priors. In *European Conference on Computer Vision*, pages 139–155. Springer, 2022. 5

[17] 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. *arXiv preprint arXiv:2106.10689*, 2021. 5

[18] Lior Yariv, Jiatao Gu, Yoni Kasten, and Yaron Lipman. Volume rendering of neural implicit surfaces. *Advances in Neural Information Processing Systems*, 34:4805–4815, 2021. 5

[19] Zehao Yu, Songyou Peng, Michael Niemeyer, Torsten Sattler, and Andreas Geiger. Monosdf: Exploring monocular geometric cues for neural implicit surface reconstruction. *Advances in neural information processing systems*, 35:25018–25032, 2022. 1, 5

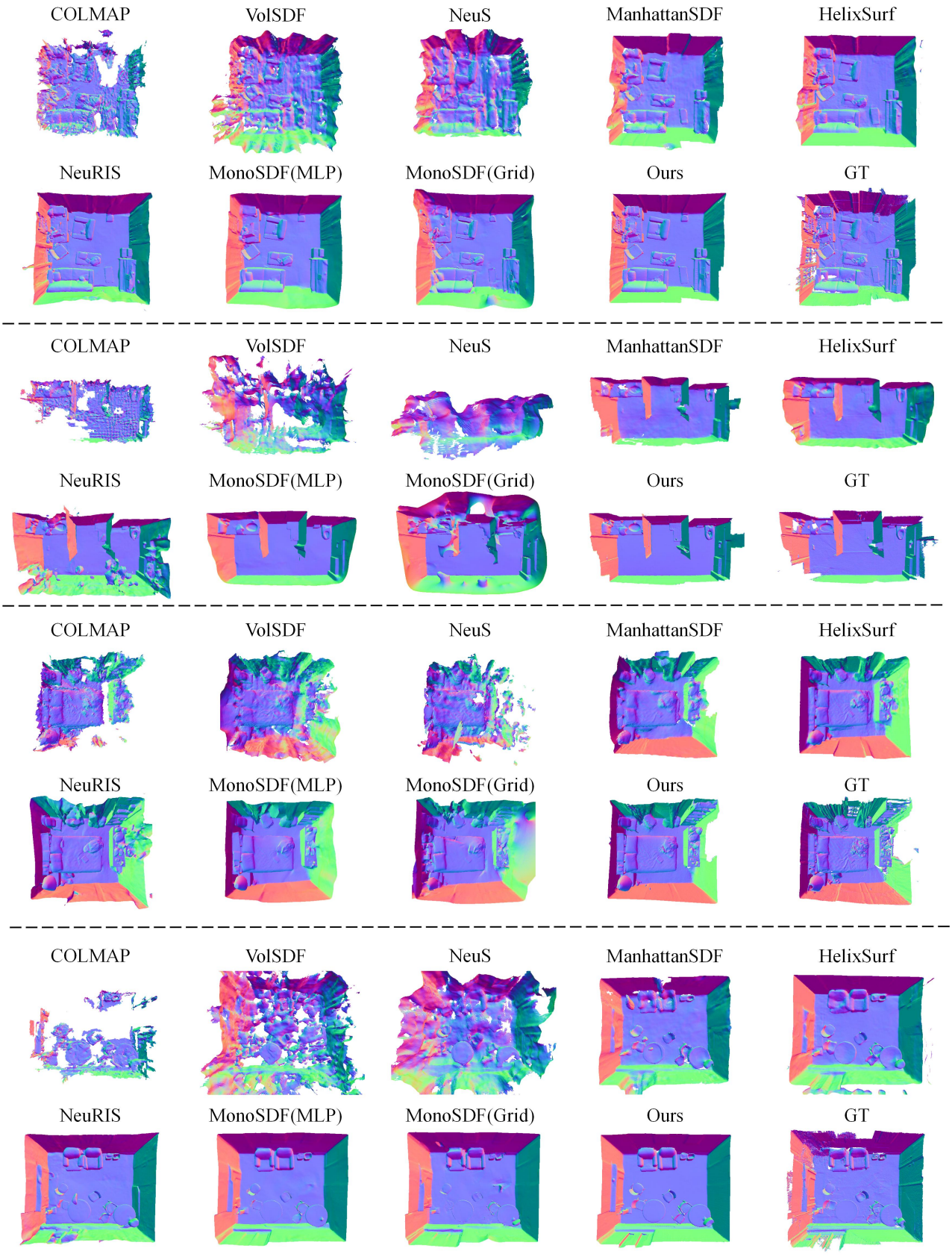

Figure 5. Top views of the reconstructions on ScanNet.

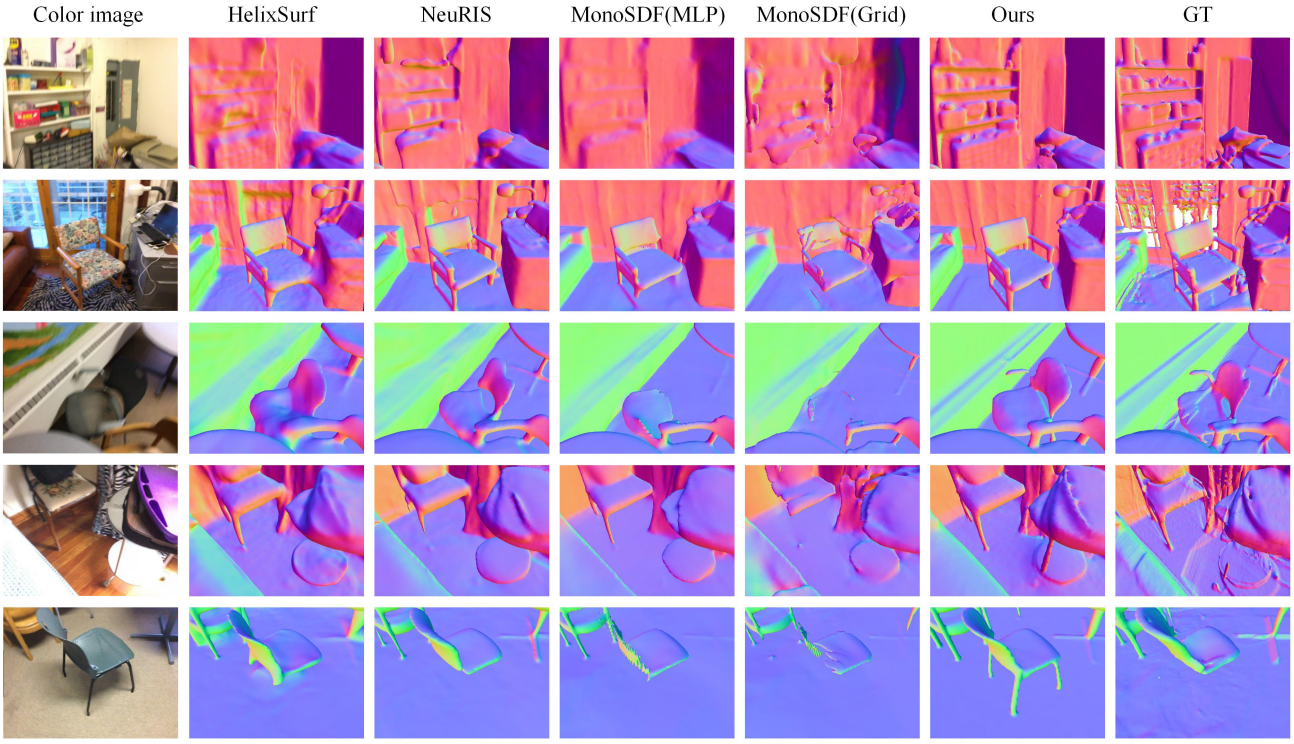

Figure 6. Comparison of reconstruction details on ScanNet.

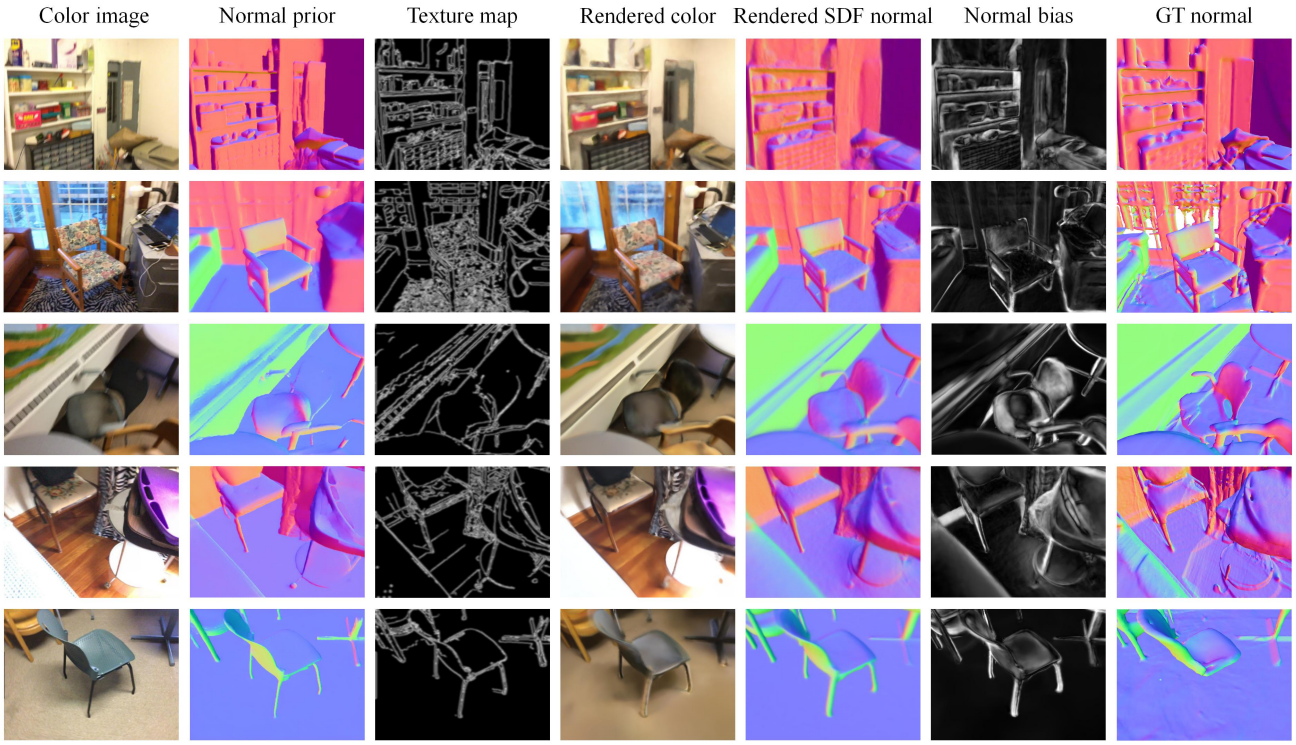

Figure 7. Rendered results of our NC-SDF on ScanNet.

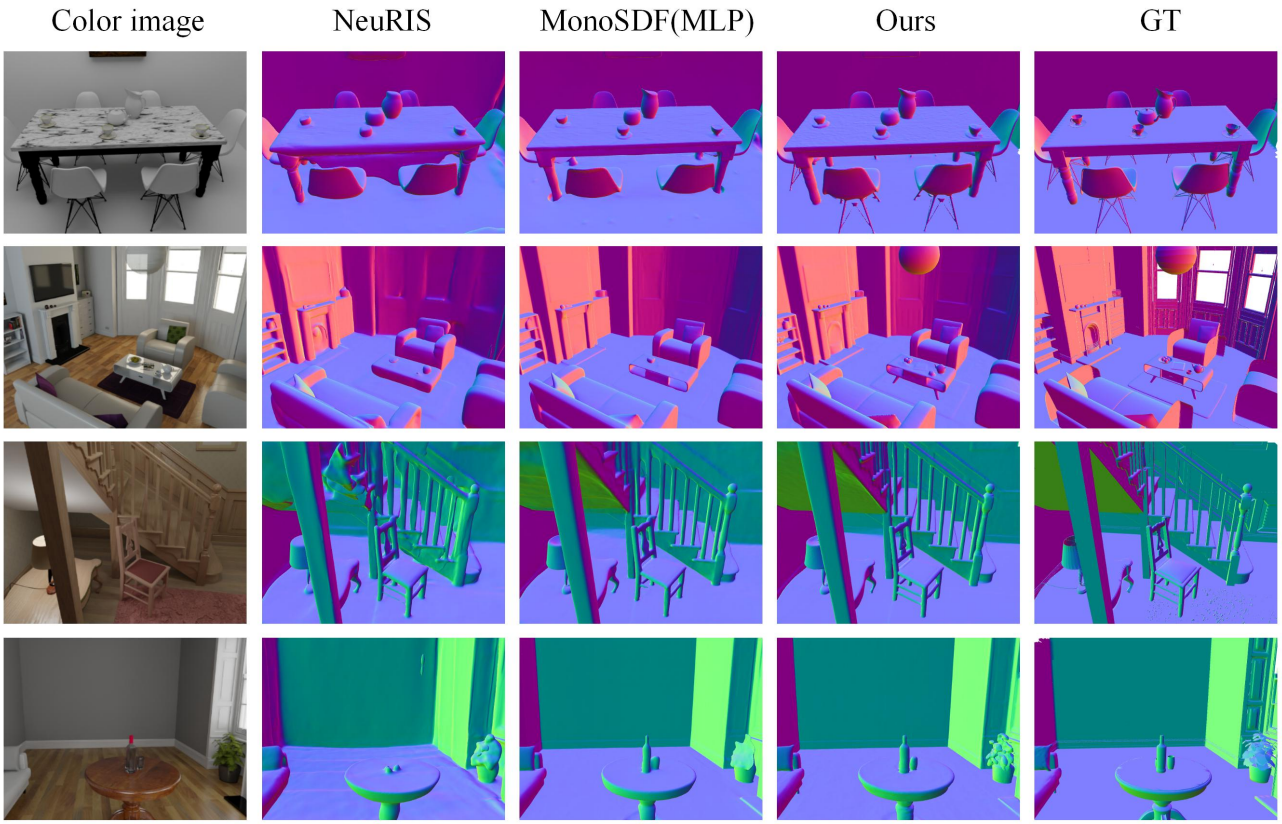

Figure 8. Comparison of reconstruction details on ICL-NUIM.

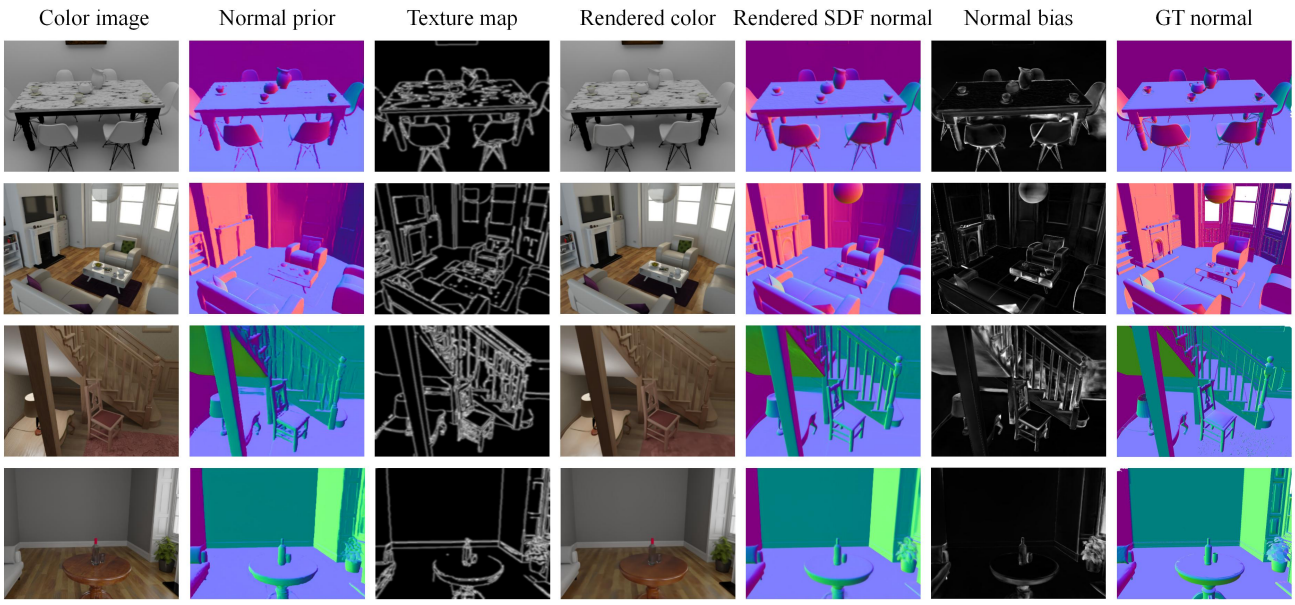

Figure 9. Rendered results of our NC-SDF on ICL-NUIM.

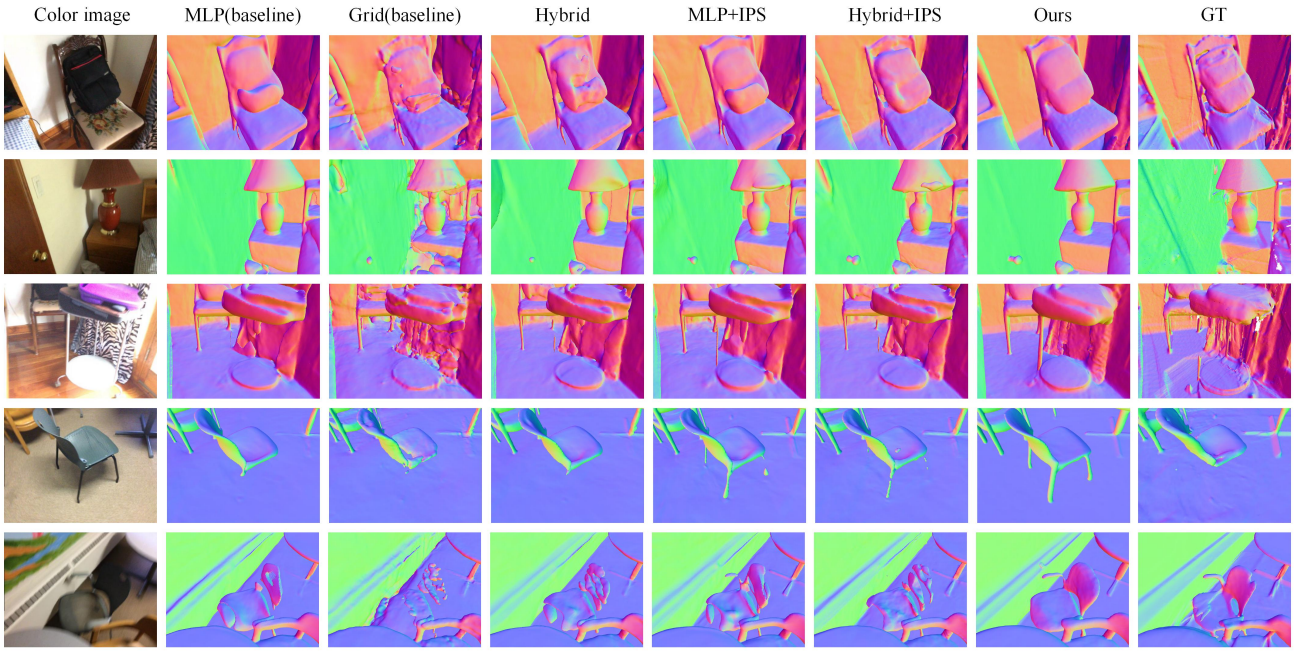

Figure 10. Visualizations for ablation studies on ScanNet.

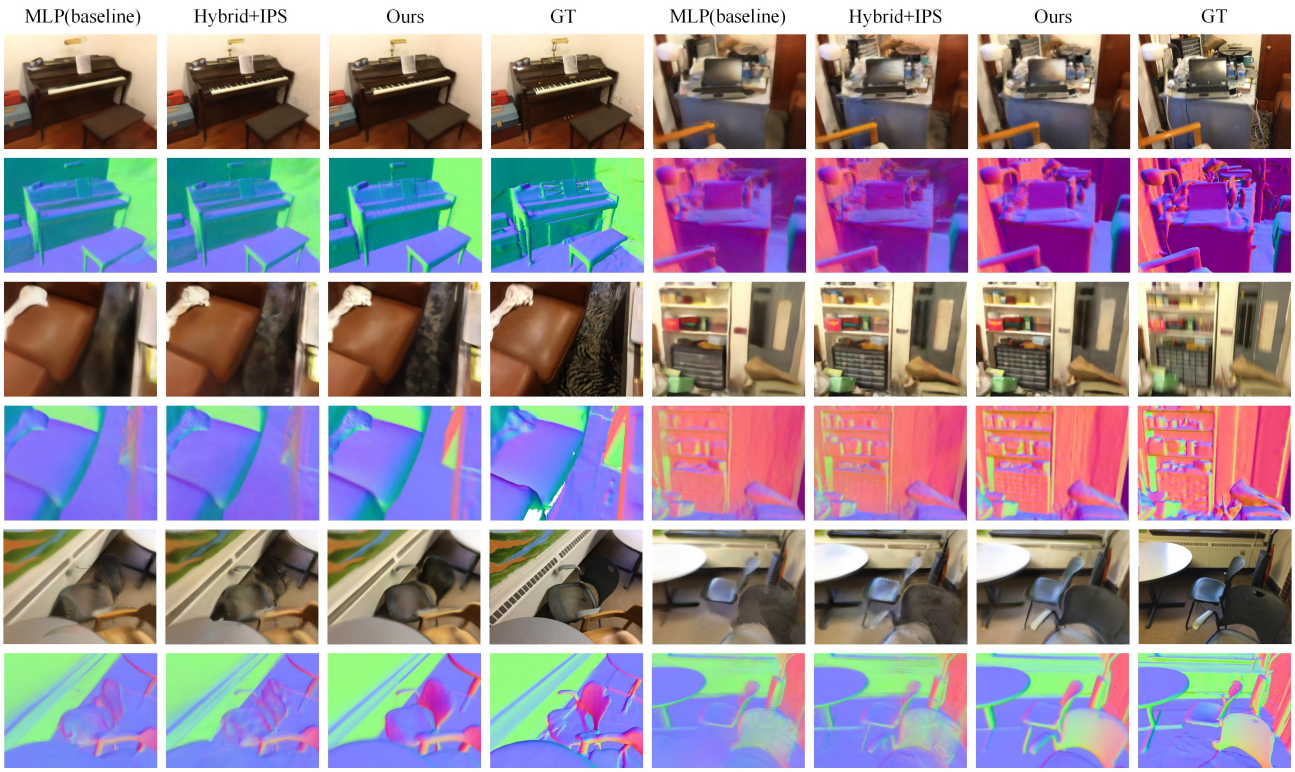

Figure 11. Comparison of novel view synthesis on ScanNet.