

Time-conditioned Illumination for Inverse Rendering of Outdoor Scenes

Supplementary Material

In this supplementary material, we provide more details and experiment results to show the effectiveness of our proposed inverse rendering framework. In the following sections, we first provide our implementation and experiment details. Afterward, we provide more visualizations for the ablation study and comparisons with the previous method.

1. Implementation Details

1.1. Network Architecture

The neural implicit surface network architecture is implemented based on Multi-resolution feature grids. The 3D position is initially mapped to a multi-resolution feature grid, utilizing tri-linear interpolation to retrieve the corresponding feature vector. This feature vector is then fed into a Multi-Layer Perceptron (MLP) to predict SDF values $s(x)$ and geometry features $z(x)$. Specifically, the network employs 16 levels for multi-resolution hash grids, with resolutions ranging from 2^5 to 2^{11} , and the dimension for the geometric feature is set to 256. A geometric initialization is applied to initialize the SDF values, setting them as a sphere. For the radiance, illumination, and material fields, MLPs are utilized to implement them. The number of layers for these networks is set to 4, with a hidden dimension of 256. Sinusoidal positional encoding is employed to encode both view direction and coordinates, using a frequency of 6. The implementation of the Bidirectional Reflectance Distribution Function (BRDF) is based on the GGX microfacet BRDF model [3].

1.2. Experiment Details

In our experiments, we utilize the PyTorch framework [1] and conduct the experiments on a single NVIDIA GeForce RTX 3090 GPU. The learning rate is set to 5×10^{-4} . The loss weights for the first stage are configured as follows: $\lambda_{\text{RGB}} = 1$, $\lambda_{\text{normal}} = 0.05$, $\lambda_{\text{eikonal}} = 0.1$, $\lambda_{\text{seg}} = 0.1$. The loss weights for the second stage are: $\lambda_{\text{render}} = 1$, $\lambda_{\text{mat}} = 0.05$, $\lambda_{\text{reg}} = 0.1$. During the training process, we normalize the RGB channels of images from a 0~255 range to a 0~1 range. For the trade-off of training speed, the sampling rate M of importance sampling for Monte Carlo ray-tracing is set to 16 during the train-

ing process, while during the inference phase, M is set to 2048 for more photorealistic rendering. The number of training iterations is set to 200,000 with a batch size of 1,600. We conduct our inverse rendering experiments on three sites from NeRF-OSR dataset [2] including: 'st', 'lk2' and 'stjacob'. The ablation studies, as demonstrated in Table 2 and Table 3 in the manuscript, are performed on the 'st' site. Besides, the relighting experiment (Table 4 in the manuscript) is conducted on the 'lk2' site using the image '20210729_203644.jpg'. Details for the corresponding sites and environment map can be accessed at <https://4dqv.mpi-inf.mpg.de/NeRF-OSR/>.

2. More Qualitative Results

Two-stage training. Fig. 1 presents the qualitative results illustrating the impact of two-stage training. Clearly, the rerendering obtained through two-stage training is more photorealistic, and the albedo exhibits more details. This enhancement is attributed to the more stable and convergent optimization process facilitated by two-stage training.

Qualitative comparison. Fig. 2 displays qualitative results showcasing the intrinsic properties compared to NeRF-OSR. In comparison with NeRF-OSR, our framework yields more reasonable results, credited to monocular geometric cues and our physically-based material and illumination representation. For additional visualizations of our inverse rendering framework, please refer to Fig. 3.

References

- [1] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019. 1
- [2] Viktor Rudnev, Mohamed Elgharib, William Smith, Lingjie Liu, Vladislav Golyanik, and Christian Theobalt. Nerf for outdoor scene relighting. In *European Conference on Computer Vision*, pages 615–631. Springer, 2022. 1
- [3] Bruce Walter, Stephen R Marschner, Hongsong Li, and Kenneth E Torrance. Microfacet models for refraction through rough surfaces. In *Proceedings of the 18th Eurographics conference on Rendering Techniques*, pages 195–206, 2007. 1



Figure 1. Impact of the training process on NeRF-OSR dataset.

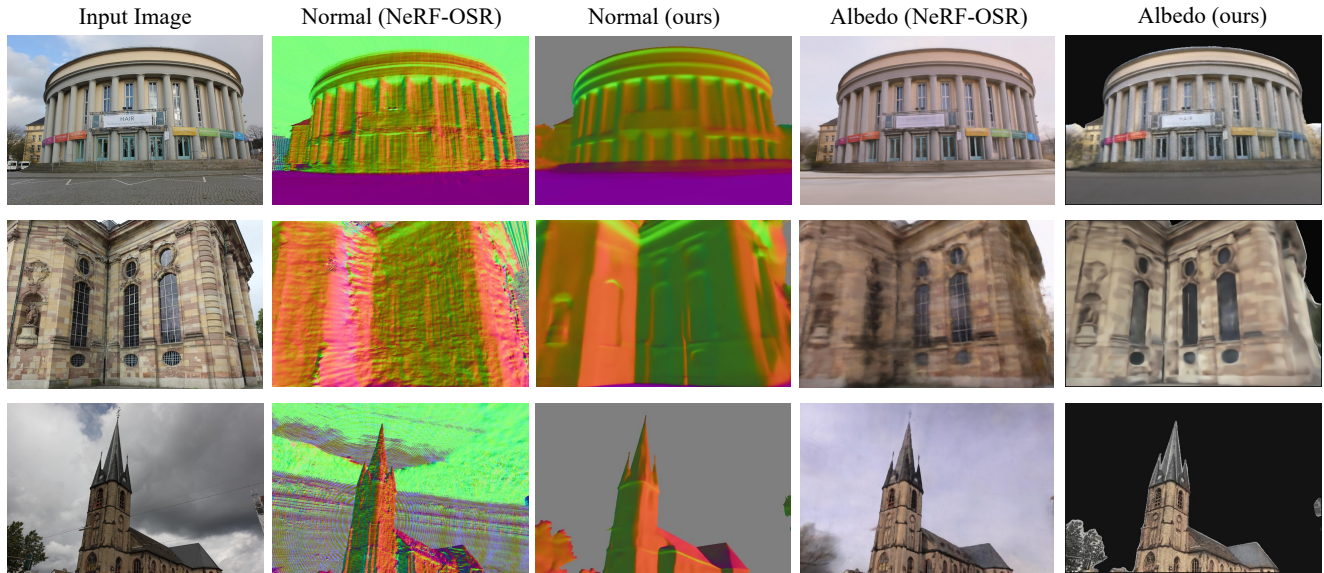


Figure 2. Qualitative comparisons with NeRF-OSR.

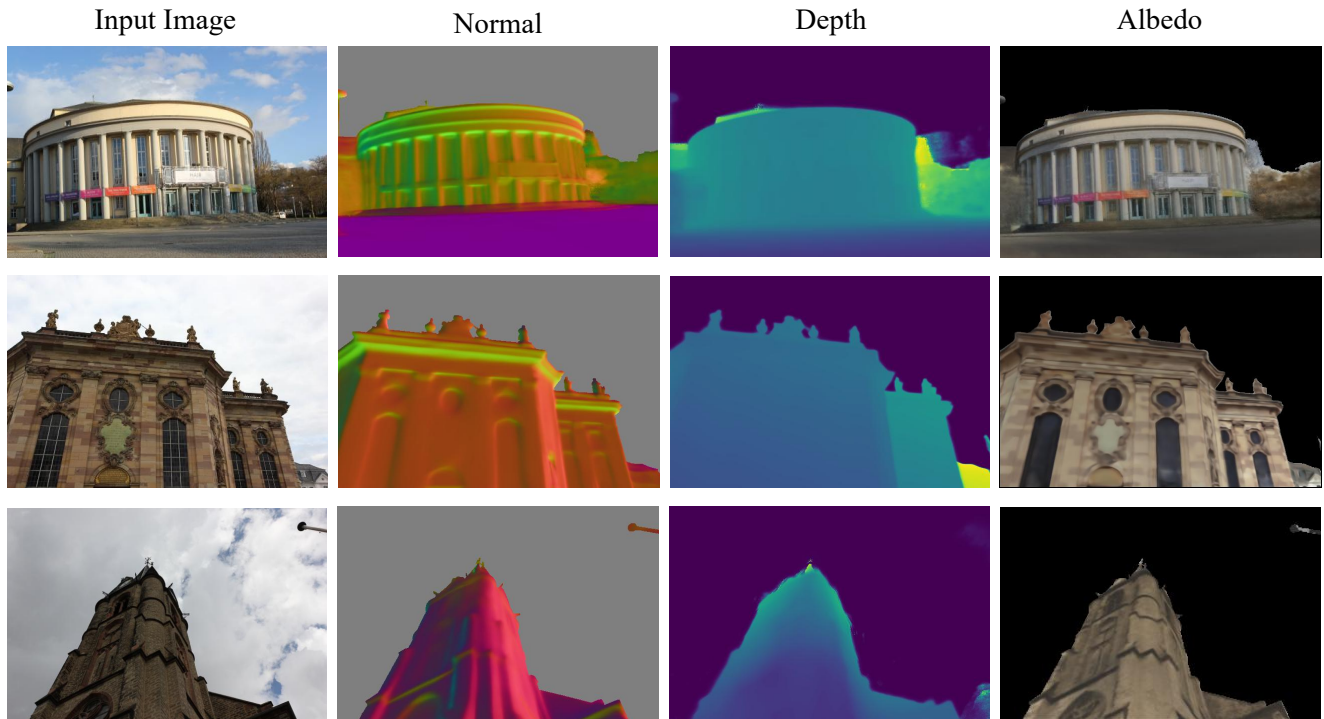


Figure 3. More qualitative results.