The Sky's the Limit: Relightable Outdoor Scenes via a Sky-pixel Constrained Illumination Prior and Outside-In Visibility
Supplementary Material

James A. D. Gardner[1], Evgenii Kashin[1], Bernhard Egger[2], and William A. P. Smith[1]

[1] Department of Computer Science, The University of York, York, YO10 5DD, UK
lncs@springer.com
http://www.springer.com/gp/computer-science/lncs
[2] Cognitive Computer Vision Lab, Friedrich-Alexander-Universität, Erlangen-Nürnberg, Erlangen, Germany
{abc,lncs}@uni-heidelberg.de

1 Introduction

Here we provide qualitative ablations of our model highlighting improvements over prior work, include further implementation details of our method and demonstrate our method on the remaining NeRF-OSR [3] scenes.

2 DDF Losses

Here we provide detailed descriptions of our DDF losses. The purpose of these losses is to encourage the geometry implied by the DDF to be consistent with the scene geometry represented by the SDF. Since the DDF is used for visibility and hence shadows which influence the appearance loss, this means that shadows can drive changes to scene geometry, i.e. the SDF, via these DDF losses.

First, the depth predicted by the DDF should match that of the scene parameterised by the SDF:

$$\mathcal{L}_{\text{ddf\_depth}} = \sum_{(\mathbf{s},\mathbf{d})\in\mathcal{B}} |d_{\text{SDF}}(\mathbf{s},\mathbf{d}) - f_{\text{DDF}}(\mathbf{s},\mathbf{d})|, \tag{1}$$

where $\mathcal{B}$ is a batch of positions ($\mathbf{s}$, with $\|\mathbf{s}\| = 1$) on the sphere and inward facing directions ($\mathbf{d}$, with $\|\mathbf{d}\| = 1$). $d_{\text{SDF}}(\mathbf{s},\mathbf{d})$ is the expected termination depth for a ray from $\mathbf{s}$ in direction $\mathbf{d}$, computed from the current SDF.

Second, travelling the distance predicted by the DDF should arrive at the SDF zero level set:

$$\mathcal{L}_{\text{ddf\_levelset}} = \sum_{(\mathbf{s},\mathbf{d})\in\mathcal{B}} f_{\text{SDF}}(\mathbf{s} + f_{\text{DDF}}(\mathbf{s},\mathbf{d})\mathbf{d})^2. \tag{2}$$

This loss penalises any non-zero SDF value at the termination point predicted by the DDF.

Third, we can impose multiview consistency on the DDF. Given an arbitrary starting point $\mathbf{s}_1$ and inward facing direction $\mathbf{d}_1$, we compute a termination point $\mathbf{x}_1 = \mathbf{s}_1 + f_{\text{DDF}}(\mathbf{s}_1, \mathbf{d}_1)\mathbf{d}_1$. Now, from an arbitrary second point $\mathbf{s}_2$, the predicted DDF depth towards $\mathbf{x}_1$ must be no greater than $\|\mathbf{x}_1 - \mathbf{s}_2\|$, since $\mathbf{x}_1$ would occlude $\mathbf{s}_2$:

$$\mathcal{L}_{\text{ddf\_multiview}} = \sum_{(\mathbf{s}_1, \mathbf{d}_1, \mathbf{s}_2) \in \mathcal{B}} \max\left(0, f_{\text{DDF}}(\mathbf{s}_2, \mathbf{d}_2) - \|\mathbf{x}_1 - \mathbf{s}_2\|\right)^2, \tag{3}$$

where $\mathbf{d}_2 = (\mathbf{x}_1 - \mathbf{s}_2)/\|\mathbf{x}_1 - \mathbf{s}_2\|$ and this time the batch comprises pairs of points and a direction.

Finally, we further take advantage of our sky segmentation maps as an additional constraint on our DDF. Rays that intersect the sky have no occlusions between the camera origin and our DDF sphere. Our DDF should therefore predict at least the distance to the camera origin for those intersecting rays:

$$\mathcal{L}_{\text{ddf\_sky}} = \sum_{\mathbf{r} \in \mathcal{R} \cap \mathcal{S}_{\text{sky}}} \max\left(0, \|\mathbf{o} - \mathbf{s}\| - f_{\text{DDF}}(\mathbf{s}, -\mathbf{r})\right) \tag{4}$$

where $\mathbf{s}$ is the point where the camera ray $\mathbf{r}$ intersects the DDF sphere and $\mathbf{o}$ the camera origin. Note that this last loss provides direct, ground truth supervision for the DDF as opposed to the previous three losses that only ensure consistency with the SDF. It plays the same role for the DDF as $\mathcal{L}_{\text{sky}}$ plays for the SDF, except it is used in an inward facing setting whereas $\mathcal{L}_{\text{sky}}$ is outward facing.

3 High Dynamic Range

As our lighting and model are both optimised in linear HDR space we implicitly reconstruct HDR sky (and scene) from the LDR input images. This enables HDR post-processing of our renderings as shown in Fig 1.

Fig. 1: HDR post-processing capabilities of our model. LDR ground truth (left), depth-of-field and HDR tonemapping (right).

4 Additional Comparisons

In Figure 2 we highlight the improvement in shadow quality our model provides over FEGR [5]. Our visibility and illumination models can represent high-quality sharp shadows whilst being trained concurrently with our geometry and albedo networks. Unlike FEGR which produces noisy artefacts and requires conversion of the SDF scene representation into a mesh for ray-tracing.

Fig. 2: FERG (L) produces noisy shadows and is trained in a cascaded manner. Ours (R) produces sharp detailed shadows and is trained end-to-end.

5 NeRF-OSR Relighting Benchmark

The relighting benchmark for NeRF-OSR [3], in which the ground truth environment map from a session is used to relight the scene and the appearance error computed from a single viewpoint per session, for pixels within a provided mask, covers sites $1-3$. To benchmark NeuSky we chose to tackle a more challenging task and instead estimate our illumination environment from a single holdout image, a test image from another viewpoint of the scene during the same capture session. From the holdout viewpoint, we hold our model static and optimise only RENI++ latent codes and scale γ for each holdout image. For this, we only optimise the appearance losses:

$$\mathcal{L}_{\text{eval\_illumination}} = \mathcal{L}_{\text{app}} + \sum_{\mathbf{r} \in \mathcal{R} \cap \mathcal{S}_{\text{sky}}} \varepsilon(\mathbf{c}_{\text{gt}}(\mathbf{r}), \mathbf{c}_{\text{sky}}(\mathbf{r}))$$

We then position the camera in the test viewpoint and evaluate within the provided mask. We decided to evaluate using this methodology for the following reasons.

1. The environment map to model alignment is unknown. SOL-NeRF [4] attempted to address this via rotations of the environment until the highest PSNR error was achieved and this was presumed to be the correct orientation.
2. The images are not HDR, we discussed with the author of NeRF-OSR and their solution is an arbitrary scaling of saturated pixels, this scaling was set to 10 for [3] and 30 in [4], to simulate HDR before fitting their illumination model.
3. Our method is more challenging as we estimate the illumination rather than being given provided with it. i.e. we simultaneously evaluate illumination estimation and relighting.

We, therefore, consider this the best tradeoff between accuracy and repeatability and recommend in the future others also use this evaluation method. We will make our fitted environment maps available for future evaluations.

Fig. 3: Renders of four other scenes in NeRF-OSR [3]. Estimated illumination of RENI++ [2], albedo and normals are shown alongside the ground truth images. Our method accurately disentangles albedo, lighting and shadows whilst producing very high-quality geometry.

Fig. 4: NeuSky lighting prediction (left), reference (right).

6 Implementation Details

During data pre-processing, we assume all cameras are looking towards the object of interest and align the average focus point of all cameras to be at the centre of our scene.

As one of the classes in our CityScapes [1] segmentation masks is ground, we have an optional ground plane alignment loss that enforces consistency between the volume rendered normal and the world-up vector for rays inside that mask. We use the normal consistency loss from MonoSDF [6]:

$$\mathcal{L}_{\text{gp}} = \sum_{\mathbf{r} \in \mathcal{R} \cap \mathcal{S}_{\text{gp}}} \|N(\mathbf{r}) - \mathbf{w}\|_1 + \left\|1 - N(\mathbf{r})^\top \mathbf{w}\right\|_1, \tag{5}$$

where $\mathcal{S}_{\text{gp}}$ is the set of ground plane pixels, $N(\mathbf{r})$ is the volume rendered normal for ray $\mathbf{r}$ and $\mathbf{w}$ is the world-up vector defined as $[0, 0, 1]$.

7 Further Results and Videos

In Figure 3 we provide more renderings of our model fit to the remaining NeRF-OSR [3] scenes, demonstrating further our model's ability to capture high-frequency geometric details, and accurately disentangle shading and albedo ambiguities. We encourage the reader to view the rendered videos on our project page demonstrating the multi-view and re-lighting capabilities of our model. In each video, we move the camera around the scene, once the camera comes to a stop we rotate the illumination environment to demonstrate the accurate shadow reproduction and relighting capabilities of our model. We also include videos of our Directional Distance Field (DDF) which we used for our sky visibility estimations and is trained concurrently with the scene representation. Each frame of this video is a single forward pass through our DDF which is able to produce the highly accurate depth maps required for accurate shadows.

References

1. Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B.: The Cityscapes Dataset for Semantic Urban Scene

Understanding. In: Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016) 5
2. Gardner, J.A.D., Egger, B., Smith, W.A.P.: Reni++ a rotation-equivariant, scale-invariant, natural illumination prior (2023) 4
3. Rudnev, V., Elgharib, M., Smith, W., Liu, L., Golyanik, V., Theobalt, C.: NeRF for Outdoor Scene Relighting. In: European Conference on Computer Vision (ECCV) (2022) 1, 3, 4, 5
4. Sun, J.M., Wu, T., Yang, Y.L., Lai, Y.K., Gao, L.: SOL-NeRF: Sunlight Modeling for Outdoor Scene Decomposition and Relighting. In: SIGGRAPH Asia 2023 Conference Papers (SA Conference Papers '23) (2023) 4
5. Wang, Z., Shen, T., Gao, J., Huang, S., Munkberg, J., Hasselgren, J., Gojcic, Z., Chen, W., Fidler, S.: Neural Fields meet Explicit Geometric Representations for Inverse Rendering of Urban Scenes. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Jun 2023) 3
6. Yu, Z., Peng, S., Niemeyer, M., Sattler, T., Geiger, A.: MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. In: Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., Oh, A. (eds.) Advances in Neural Information Processing Systems. vol. 35, pp. 25018–25032. Curran Associates, Inc. (2022), https://proceedings.neurips.cc/paper_files/paper/2022/file/9f0b1220028dfa2ee82ca0a0e0fc52d1-Paper-Conference.pdf 5