EdgeRelight360: Text-Conditioned 360-Degree HDR Image Generation for Real-Time On-Device Video Portrait Relighting (Supplementary material)

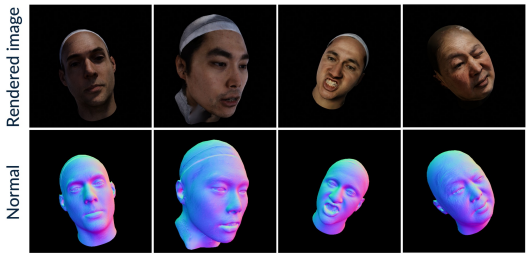

Figure 1. Examples in our synthetic normal dataset. Paired rendered images and view-dependent normal ground truths are rendered in Blender [2] given 3D head meshes and HDRI environment maps.

Geometry Net Inspired from Total Relighting [9], we adopt a similar but more lightweight UNet architecture. It contains total 13 layers, where each layer consists of a 3×3 convolution, batch normalization [8], and parametric ReLU [7]. The encoder layers contain [16, 32, 64, 64, 128, 256] filters, 256 for the bottleneck layer, and [256, 128, 64, 64, 32, 16] for the decoder layers. The encoder uses max-pooling for down-sampling, while the decoder has bilinear up-sampling layers. We use the L_1 loss between ground truth normal images in the synthetic dataset and predicted normal maps for supervision. The parameter number of Geometry Net is $2.864M$ and MACs is $7.417G$ given a 512×512 RGB input.

Synthetic Normal Dataset Inspired by [12], first, we use the 3DMD system [1] to capture head meshes of 60 subjects with 31 expressions and use the Wrap4D software [6] to automatically re-topologize them. Then, we use the ray-trace based render engine Cycles [3] in Blender [2] to render relit images and normal maps of head meshes under different environment lighting from PolyHaven [5]. To generate photorealistic renderings, we utilize tangent normals generated by the tool NormalmapGenerator [4] to add surface details to the re-topologized head meshes. In addition, different from [12] which assumes a uniform specular coefficient 0.6 during rendering, we use NormalmapGenerator [4] to synthesize reasonable specular maps, indicating that specular reflections in areas such as forehead, cheeks, and nose are stronger.

In our synthetic normal dataset, among 60 identities, 53 identities are used for training and 7 identities for testing. And for the environment lighting in Blender, we randomly select 553 HDRI environments from PolyHaven [5] for training and 71 for testing. To add more variations in the synthetic dataset, the horizontal and vertical rotation angles of the HDRI environments are randomly set within $[0°, 360°]$ and $[-60°, 60°]$. The pitch, yaw, roll rotations of the head meshes are sampled in the range of $[-15°, 15°]$, $[-45°, 45°]$, and $[-30°, 30°]$. The y-axis position of the head mesh is randomly set within $[-0.4, 0.4]$. Overall, we render 300K paired samples and examples of our synthetic normal dataset are shown in Fig. 1.

On-device Video Relighting In Fig. 2, we show screenshots of on-device video relighting of talking sequences under different generative HDRI environments and rotations. Our supplementary video results exhibits good temporal consistency and very few flickering. Note that the scaling constants s_1 and s_2 in the shading equation are empirically set to 0.29 and 0.38 for all experiments with generative HDRI environments. Additionally, in Fig. 3, we demonstrate the temporal consistency of our method and show the comparison with the web version of SwitchLight [11].

LDR text to 360-degree generation We compare additional qualitative realistic results from our LDR generative model with the LDM3D-VR approach for diverse text prompts in Fig. 4 and 5. Overall, our approach demostrates better prompt fidelity than the LDM3D-VR based on the generated images.

References

[1] https://3dmd.com/products/, 2024. [Online; accessed March-6-2024]. 1

[2] https://www.blender.org/, 2024. [Online; accessed March-6-2024]. 1

[3] https://docs.blender.org/manual/en/latest/render/cycles/index.html, 2024. [Online; accessed March-6-2024]. 1

[4] https://github.com/Theverat/NormalmapGenerator, 2024. [Online; accessed March-6-2024]. 1

[5] https://polyhaven.com/, 2024. [Online; accessed 22-February-2024]. 1

[6] https://www.russian3dscanner.com/wrap4d/, 2024. [Online; accessed March-20-2024]. 1

[7] Abien Fred Agarap. Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375*, 2018. 1

[8] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. pmlr, 2015. 1

[9] Rohit Pandey, Sergio Orts Escolano, Chloe Legendre, Christian Haene, Sofien Bouaziz, Christoph Rhemann, Paul Debevec, and Sean Fanello. Total relighting: learning to relight portraits for background replacement. *ACM Transactions on Graphics (TOG)*, 40(4):1–21, 2021. 1

[10] Gabriela Ben Melech Stan, Diana Wofk, Estelle Aflalo, Shao-Yen Tseng, Zhipeng Cai, Michael Paulitsch, and Vasudev Lal. Ldm3d-vr: Latent diffusion model for 3d vr. *arXiv preprint arXiv:2311.03226*, 2023. 5, 6

[11] SwitchLight. https://www.switchlight.beeble.ai/. [Online; accessed 22-February-2024]. 1, 4

[12] Zhibo Wang, Xin Yu, Ming Lu, Quan Wang, Chen Qian, and Feng Xu. Single image portrait relighting via explicit multiple reflectance channel modeling. *ACM Transactions on Graphics (TOG)*, 39(6):1–13, 2020. 1

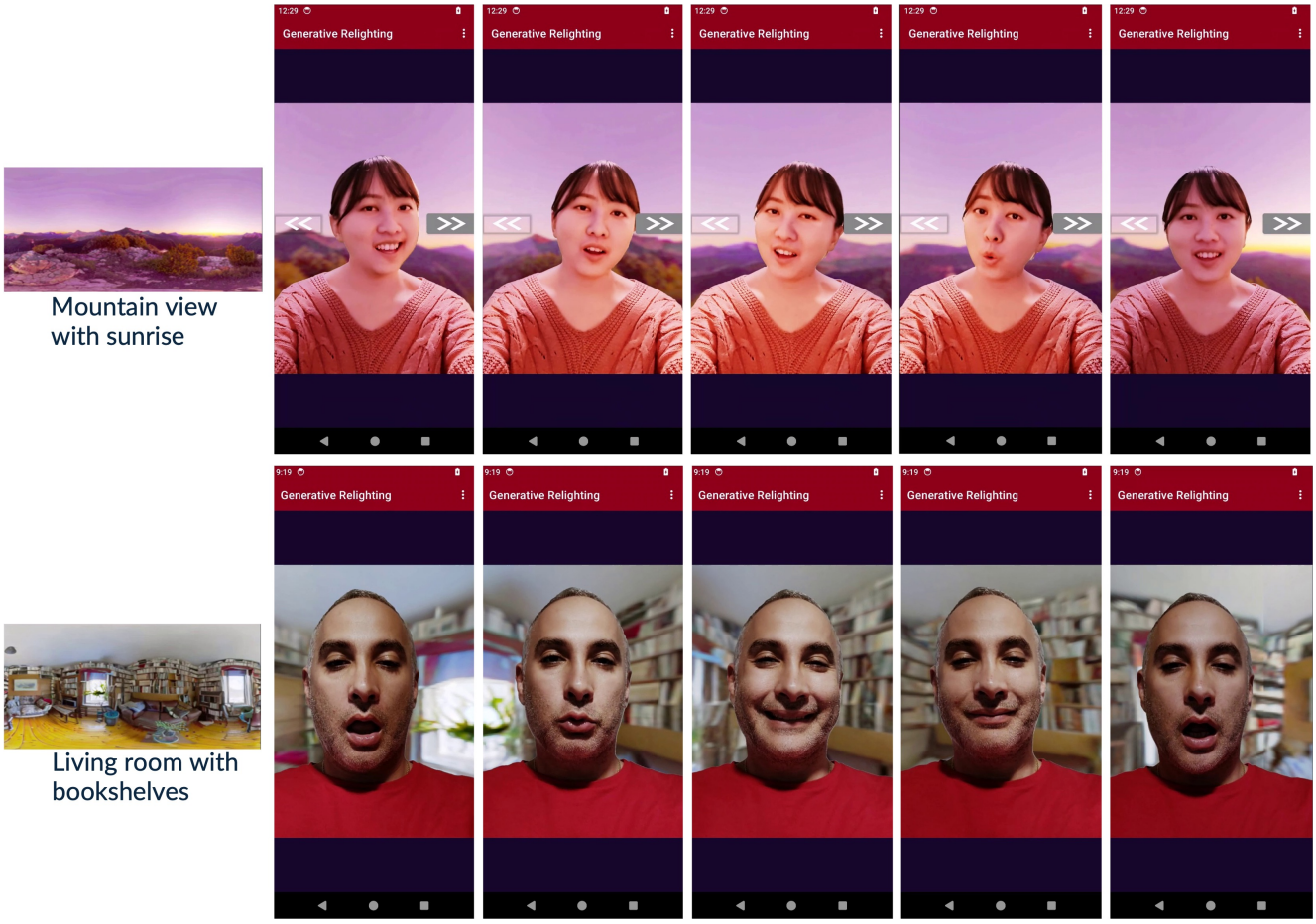

Figure 2. Our EdgeRelight360 application running on mobile device can generate temporally-consistent and realistic relit results on real talking sequences in real-time.

Figure 3. Our video relighting framework can generate more temporally-consistent relit results compared with the web version of SwitchLight[11].

Figure 4. Our text to 360-degree LDR generative model synthesize realistic images for diverse prompts compared to the LDM3D-VR [10] method.

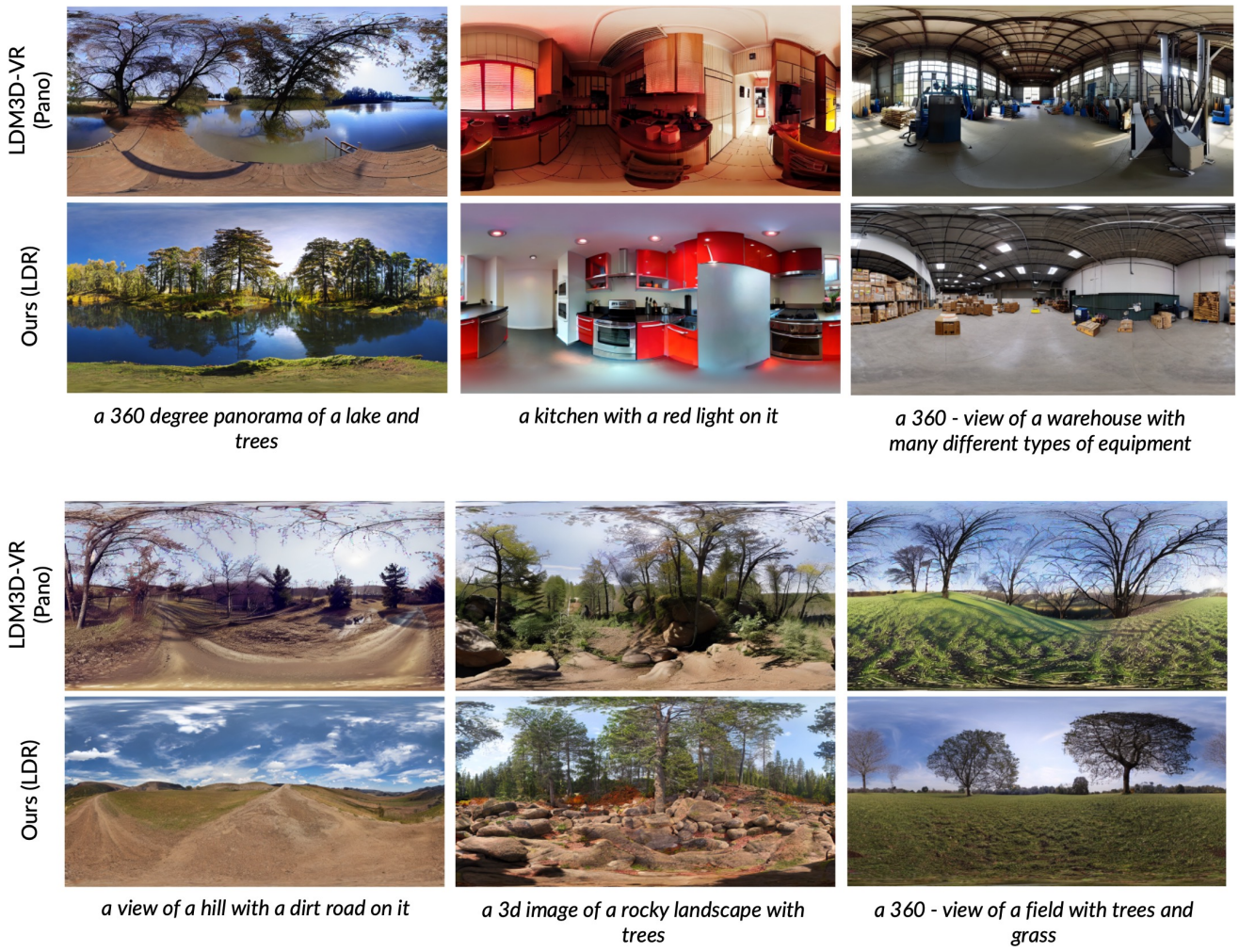

Figure 5. Ours text to 360-degree LDR generative model synthesis realistic images for diverse prompts compared to the LDM3D-VR [10] method.