

# LumiMotion: Improving Gaussian Relighting with Scene Dynamics

## Supplementary Material

### 1. Code and data repository

Code and data are included in our repository:

<https://github.com/joaxkal/LumiMotion>.

### 2. Additional videos and figures

#### 2.1. Videos

Please refer to our attached videos for more results on:

- **ENeRF dataset [3]** : real world data, moving actors with wall background. Actors cast strong shadow.
- **DNA dataset [1]**: real world multiview data, moving actors with additional items like stool, table, hair dryer.
- **our synthetic scenes.**

#### 2.2. Figures

We present additional renders for

- **DNA dataset [1]**: moving humans with various items (table, stool, hair dryer) in Fig. 3, 2, 4.
- **more comparisons with baseline methods** in Fig. 5, 6, 7.

### 3. Extended results

In Tab. 1 we present extended results, including Novel View Synthesis (NVS) and Roughness.

#### 3.1. Novel View Synthesis

We show that **the dynamic setting we use is significantly more challenging than the static setup for baselines**, as reflected in the novel view synthesis metrics. Despite this, **LumiMotion** achieves strong results for materials and re-lighting, demonstrating the effectiveness of our approach.

Please note that the high NVS scores of static baselines are also caused by **overfitting** to the training lighting conditions. When evaluated under novel illumination, their performance drops significantly, which is also consistent with our qualitative observations (Fig. 5, 6, 7) . For clarity, we report the PSNR drop in the last column of the table.

**This highlights the effectiveness of our separation strategy and the consistent behavior of our method across both training and novel lighting conditions.**

#### 3.2. Roughness

We present additional results for roughness estimation. For fair comparison, we experimented with modifying the default IRGS configuration. We found out its standard smooth constraint on roughness adversely affects material estimation - produces roughness maps that are unnaturally smooth and lack detail. See Fig. 1 for comparison. Therefore, in the table we also present results obtained by training IRGS without this loss term.

**Please note that LumiMotion consistently achieves significantly lower MSE for roughness comparing to even the closest baseline, IRGS.**

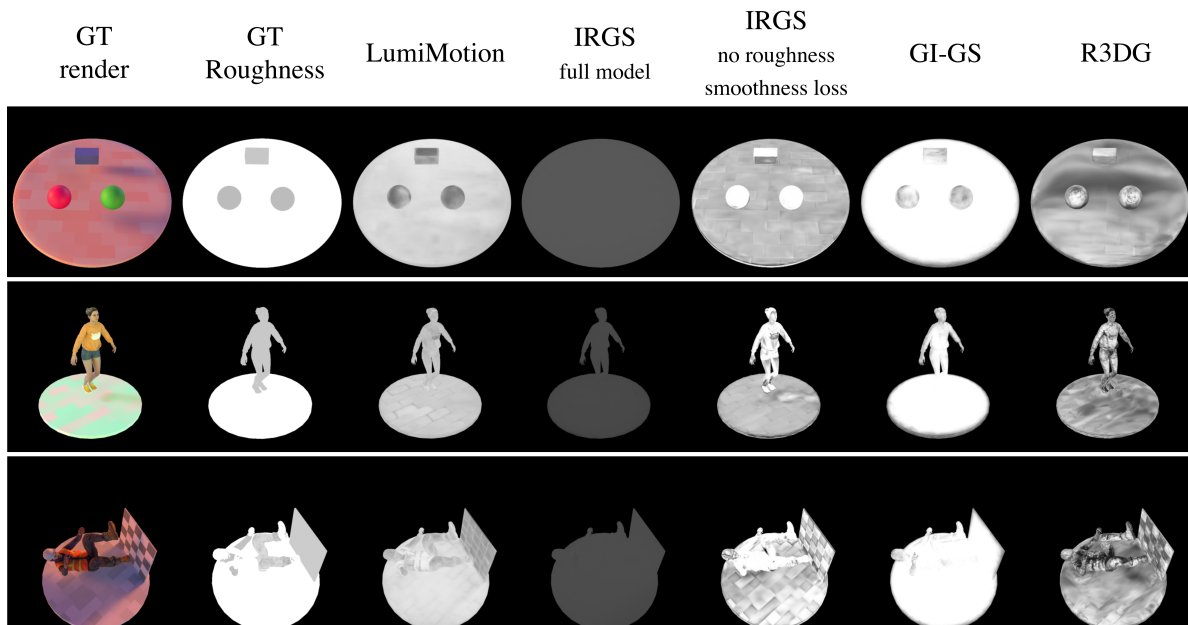


Figure 1. Roughness estimation. Baseline methods fail to recover realistic roughness maps. For IRGS, the standard loss set yields roughness values with little to no surface detail (see the ‘full model’ column). Removing smoothness constraints causes different issues, for example: in the top row, the plate is estimated as specular and the items as rough, whereas the ground truth shows the opposite. Other baselines also fail to disentangle material properties and estimate roughness reliably. In contrast, **LumiMotion** produces smooth and largely accurate roughness maps.

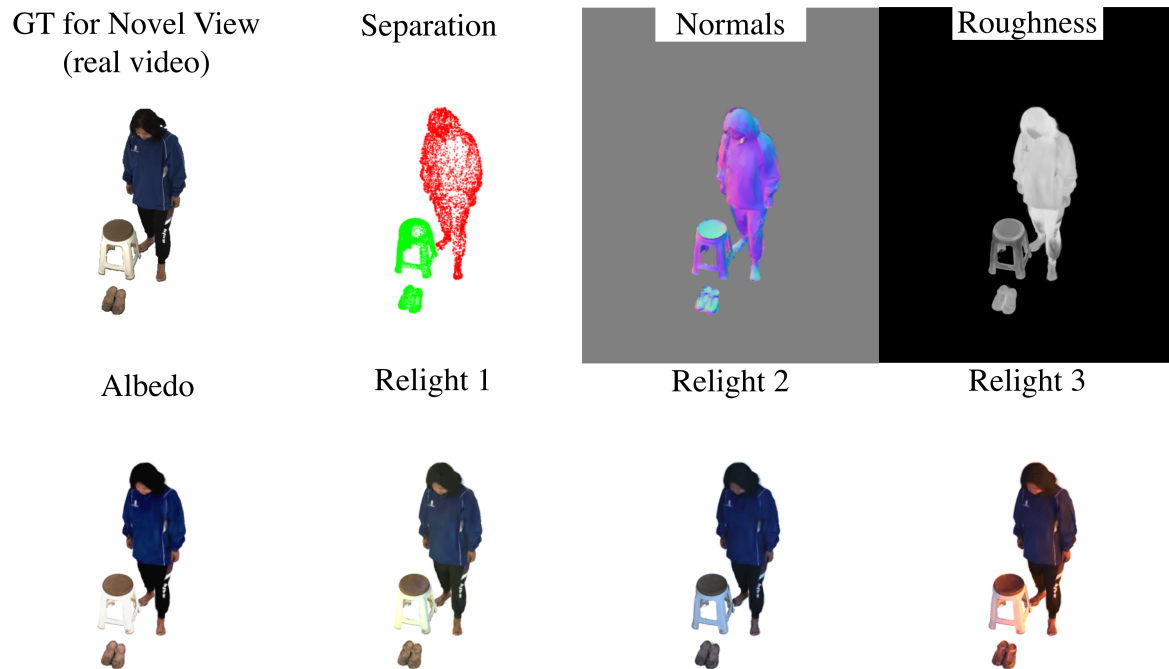


Figure 2. DNA - Actor going to sit on a stool. Please see attached video for this scene.



Table 1. **LumiMotion for novel view synthesis, material estimation, and relighting.** Static methods use only one timestep, while **LumiMotion** operates on a dynamic scene. Testing is performed on the same views and timestep as in the static setting. Results are grouped by train-test lighting conditions. For IRGS, we denote training scheme without smooth constraint on roughness by †.

**NVS:** Note that **the dynamic setting we use is significantly more challenging than static setup for baselines.** Despite this, **LumiMotion** achieves strong results for albedo and relight, demonstrating the effectiveness of our approach. Please note that the high NVS scores of static baselines are also caused by **overfitting** to the training lighting conditions. When evaluated under novel illumination, their performance drops significantly. This observation is consistent with our qualitative results. We show PSNR drop in the last column. **This highlights the effectiveness of our separation strategy and the consistent behavior of our method across both train and test lighting.**

**Material:** We achieve significantly better material estimation than the closest baseline, IRGS, regardless of its training setup. Notably, **LumiMotion** consistently produces **higher-quality albedo** and achieves at least a **2× lower roughness MSE** compared to IRGS.

Method	Novel View Synthesis			Albedo			Roughness	Relight			$\Delta$ PSNR
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	MSE $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	NVS $\rightarrow$ Relight
Dam Wall $\rightarrow$ Harbour Sunset											
R-3DGS	35.031	0.987	0.035	20.744	0.900	0.128	$0.066 \pm 0.005$	21.220	0.915	0.112	39.5%
GI-GS	26.749	0.956	0.066	20.943	0.906	0.105	$0.036 \pm 0.001$	18.431	0.868	0.139	31.1%
IR-GS	32.207	0.983	0.021	22.888	0.936	0.076	$0.136 \pm 0.040$	26.177	0.953	0.064	18.7%
IR-GS $\dagger$	32.639	0.985	0.019	23.512	0.935	0.080	$0.024 \pm 0.010$	27.156	0.954	0.067	16.8%
LumiMotion	26.948	0.952	0.025	<b>27.268</b>	<b>0.952</b>	<b>0.069</b>	<b><math>0.012 \pm 0.002</math></b>	26.037	0.928	0.060	<b>3.4%</b>
Chapel Day $\rightarrow$ Golden Bay											
R-3DGS	36.986	0.989	0.028	22.463	0.927	0.096	$0.044 \pm 0.008$	22.282	0.943	0.081	39.8%
GI-GS	29.489	0.971	0.057	24.733	0.955	0.056	$0.031 \pm 0.002$	22.673	0.880	0.125	23.1%
IR-GS	33.580	0.983	0.022	23.769	0.956	0.053	$0.128 \pm 0.045$	28.157	0.966	0.046	16.2%
IR-GS $\dagger$	34.212	0.985	0.019	24.085	0.956	0.052	$0.028 \pm 0.012$	28.702	0.968	0.046	16.1%
LumiMotion	27.636	0.952	0.022	<b>30.838</b>	<b>0.973</b>	<b>0.036</b>	<b><math>0.011 \pm 0.002</math></b>	28.563	0.939	0.041	<b>−3.3%</b>
Golden Bay $\rightarrow$ Dam Wall											
R-3DGS	36.096	0.988	0.028	19.945	0.899	0.133	$0.039 \pm 0.010$	19.563	0.918	0.118	45.8%
GI-GS	34.402	0.982	0.031	21.295	0.932	0.087	$0.031 \pm 0.003$	17.636	0.823	0.132	48.8%
IR-GS	34.404	0.980	0.026	20.910	0.937	0.082	$0.145 \pm 0.027$	25.009	0.955	0.060	27.3%
IR-GS $\dagger$	35.978	0.985	0.020	21.199	0.936	0.081	$0.021 \pm 0.008$	25.252	0.957	0.058	29.8%
LumiMotion	29.859	0.954	0.023	<b>27.929</b>	<b>0.959</b>	<b>0.058</b>	<b><math>0.010 \pm 0.002</math></b>	25.405	0.936	0.048	<b>14.9%</b>

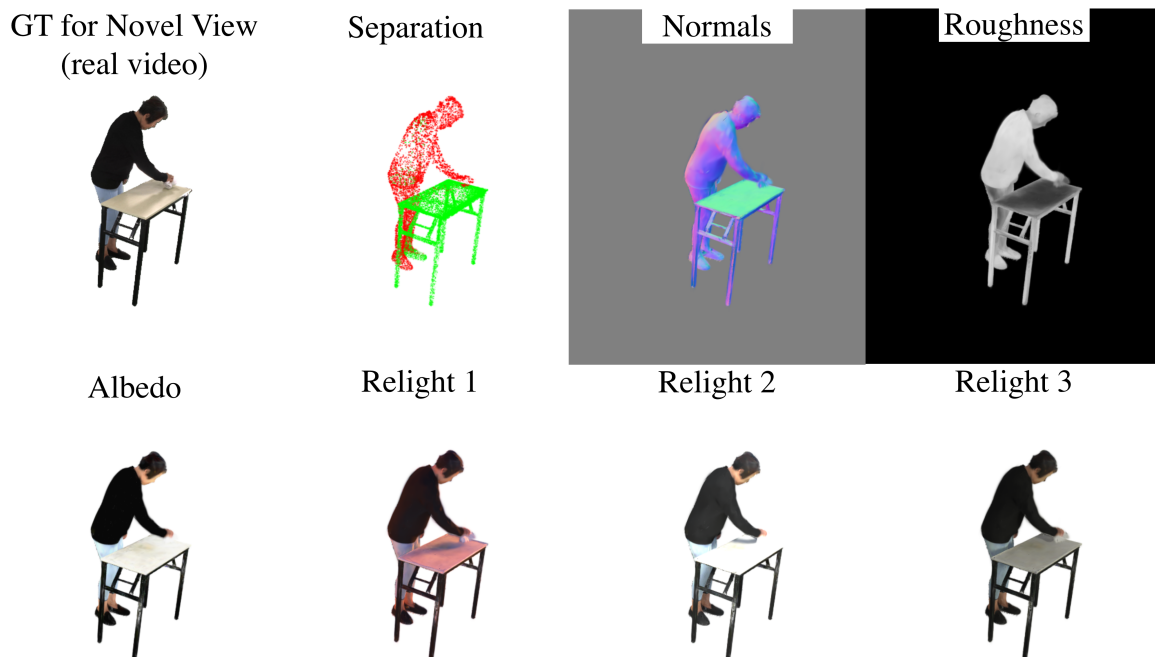


Figure 3. DNA – Actor cleaning a specular table. Although the actor’s arm casts a strong shadow on the tabletop, the Gaussians on the table are correctly classified as static. Please see attached video for this scene.

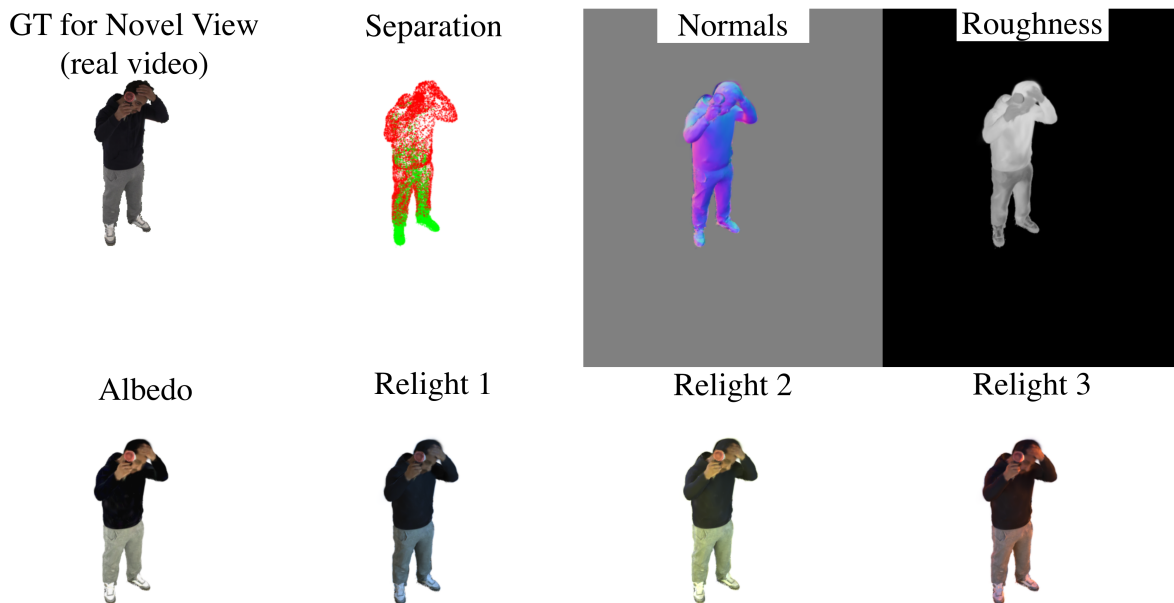


Figure 4. DNA - Actor with a hair dryer. Actor’s feet are static and only his upper body parts are moving. Please see attached video for this scene.

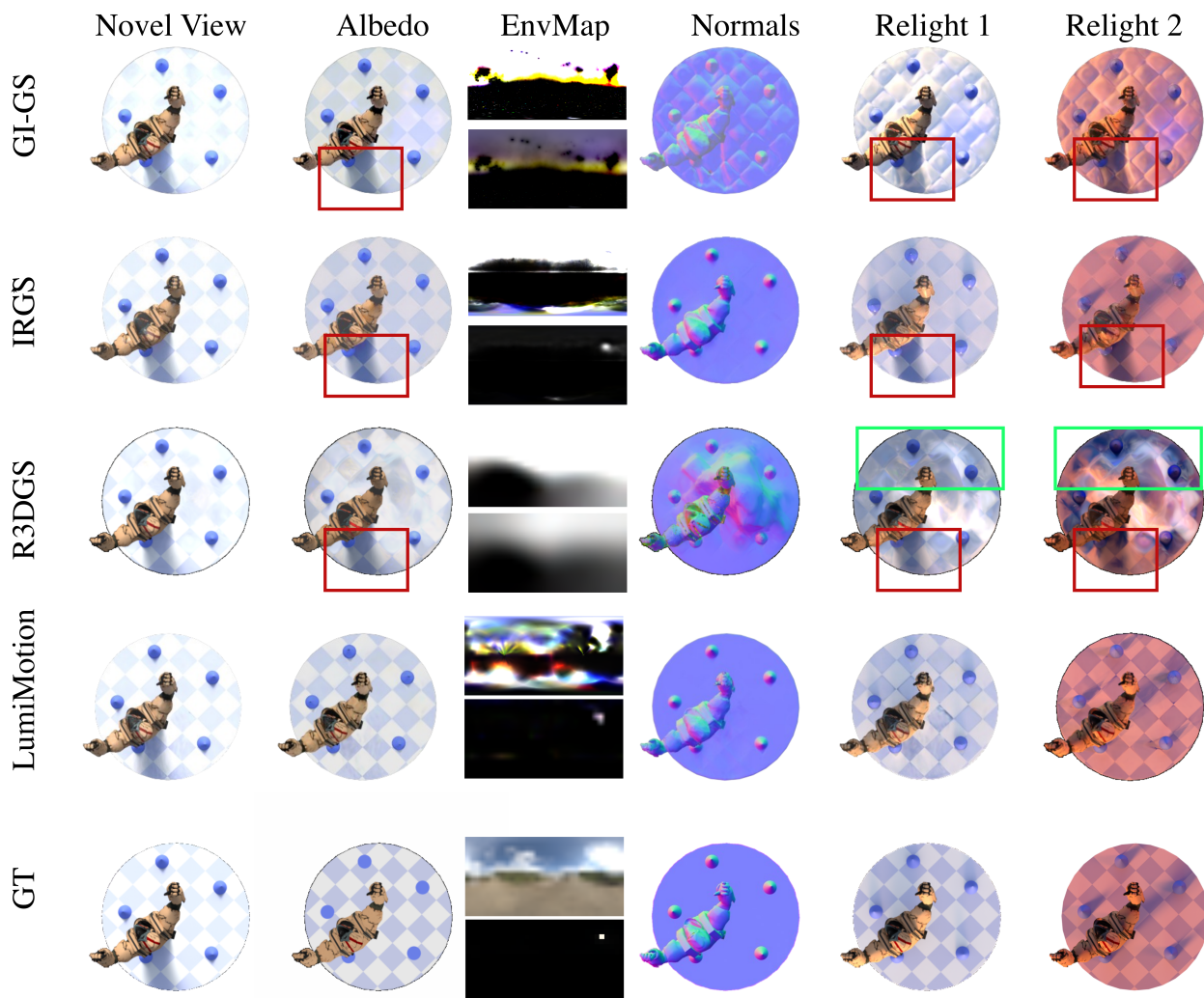


Figure 5. Scene 1. Our method produces much clearer albedo than the baselines. Although we train on dynamic scenes—more challenging than the static ones used by baselines—our results still show smooth and accurate geometry and normals. The presented environment map (top shows the map truncated to  $[0, 1]$  while bottom shows it scaled so that the maximum value is 1) captures the main light direction more accurately than environment maps estimated from static baselines. We mark in color boxes fine details like artifacts including baked-in shadows, which clearly show superiority of our method. We highlight fine details—such as artifacts and baked-in shadows—in color boxes, clearly demonstrating the superiority of our method.

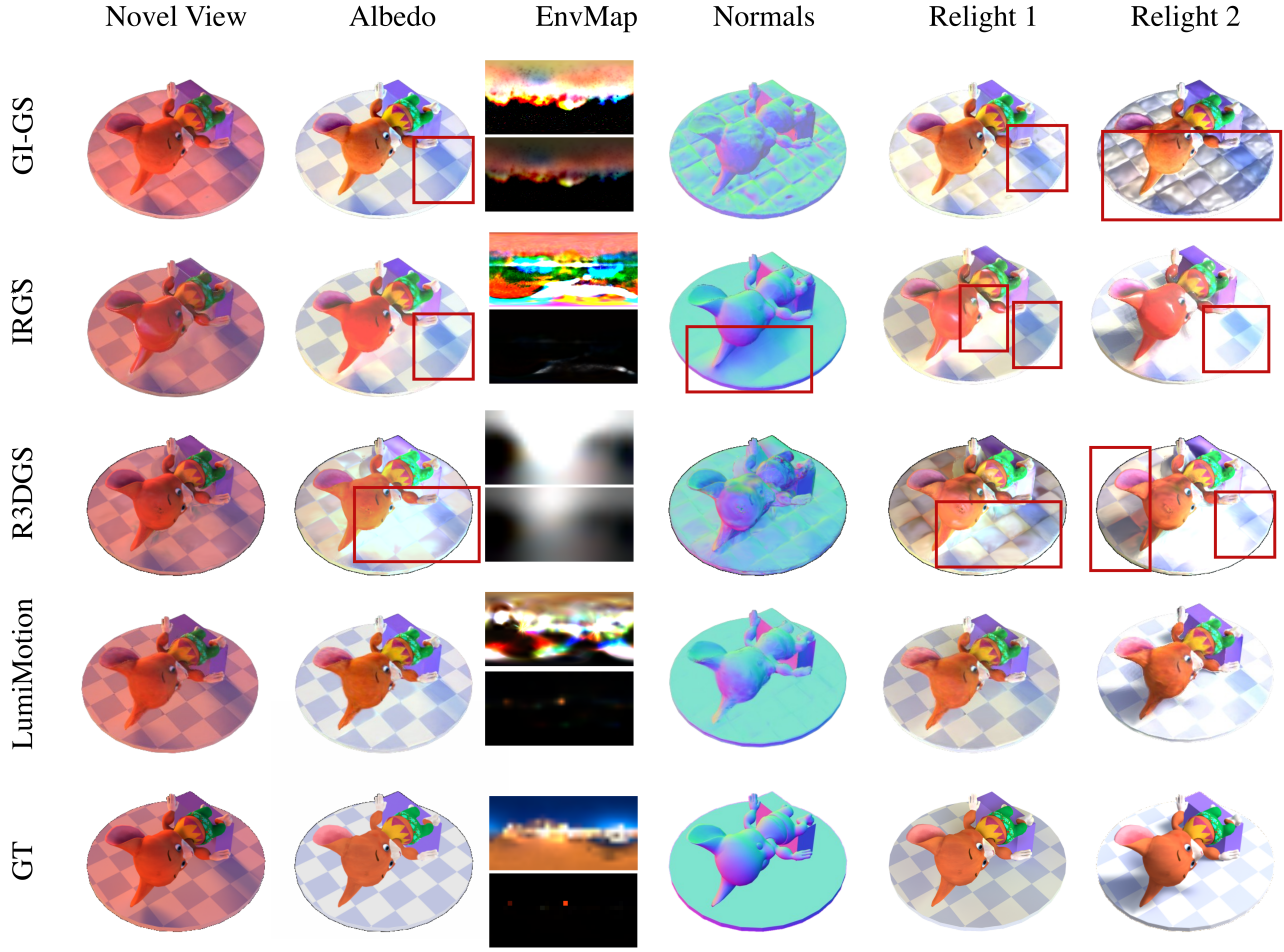


Figure 6. Scene 2. Our method produces much clearer albedo than the baselines. Although we train on dynamic scenes—more challenging than the static ones used by baselines—our results still show smooth and accurate geometry and normals. The presented environment map (top shows the map truncated to  $[0, 1]$ , while bottom shows it scaled so that the maximum value is 1) captures the main light direction more accurately than environment maps estimated from static baselines. We highlight fine details—such as artifacts and baked-in shadows—in color boxes, clearly demonstrating the superiority of our method.

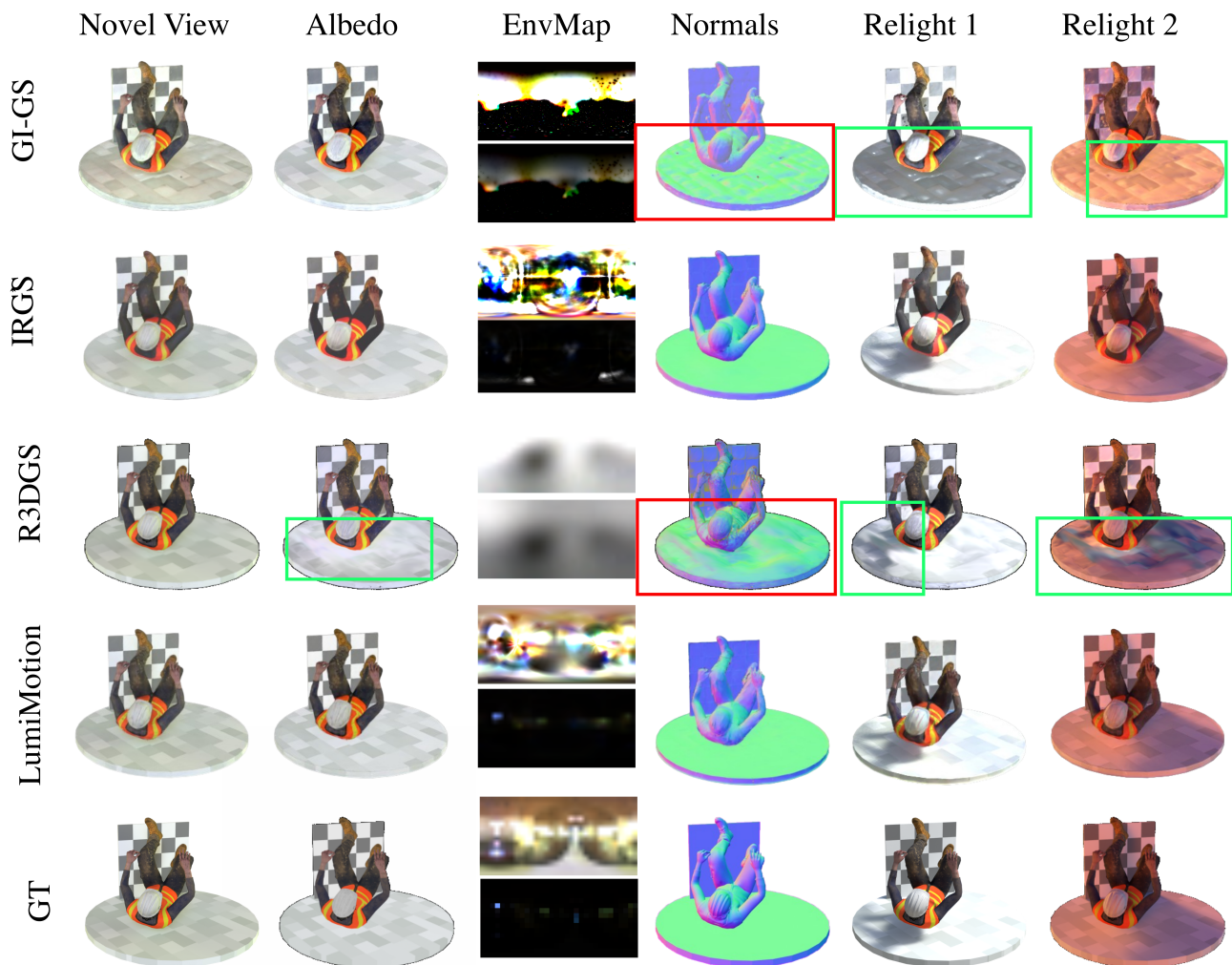


Figure 7. Scene 3. This scene was captured under mild training illumination, with no strong shadows present. As a result, the static baselines exhibit relatively well-separated albedo in the training image (no baked in shadows). We demonstrate that in this case our method achieves relighting and geometry quality competitive with the latest state-of-the-art IRGS, despite IRGS being a static baseline. The presented environment map (top shows the map truncated to  $[0, 1]$ , while bottom shows is scaled so that the maximum value is 1) captures the main light direction more accurately than environment maps estimated from static baselines.



#### 4. Separation - additional example of ablation and hyperparameter influence

In Fig. 9, we illustrate the influence of separation hyperparameters. Our separation method robustly detects moving parts of jumping actor. Depending on the scene, a delayed start or a separation value that is too low may impair the penalization of static regions. In Fig. 8 we show that without separation, strong moving shadow on the plate is modeled by moving Gaussians. Our separation strategy allows for cleaner albedo without shadow artifacts.

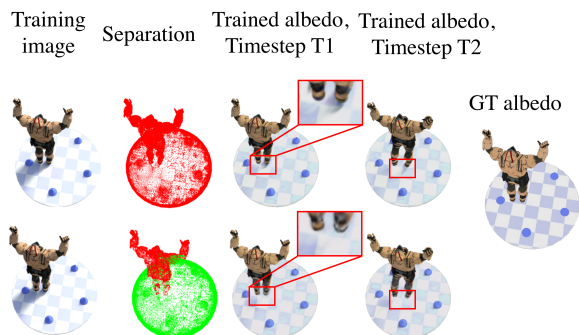


Figure 8. Separation - ablation study. **Please zoom in for details.**

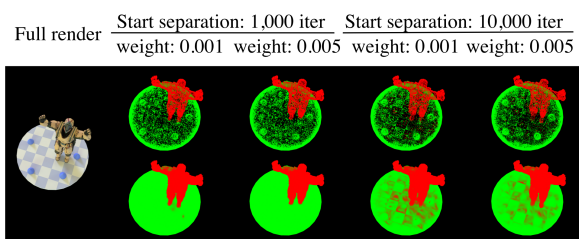


Figure 9. Separation – impact of hyperparameters. Top: Renderings using downsized Gaussians illustrate uneven Gaussian placement on the plate when the separation loss is introduced too late, in 10,000th iteration. In this case several Gaussians on the plate are incorrectly classified as dynamic and move with the shadows. Bottom: Renderings with full-sized Gaussians highlight the extent of this misclassification. **Please zoom in for details.**

## 5. Our dataset

We provide additional details about our synthetic dataset and its generation process. We build 5 synthetic datasets in Blender, using Mixamo<sup>1</sup> platform and simple Blender meshes. We prepared each scene in two versions: dynamic and static. For dynamic version, we use D-NeRF [4] like setup with different camera view for each timestep, creating a multi-view, dynamic scene suitable for evaluating relighting and novel view synthesis (see Fig. 10). For static variant, we use the same camera views and only one timestep. All scenes but ‘spheres’ contain 150 frames, for ‘spheres’ there are 100 frames.

Each scene is relit using four high-dynamic-range (HDR) environment maps from PolyHaven<sup>2</sup> selected to span diverse lighting conditions (the environment maps are rotated by us such that the dominant light source appears from various directions):

- Small Harbour Sunset
- Dam Wall
- Golden Bay
- Chapel Day

We show all environment maps in Fig. 11.

Originally the environment maps are in 4K resolution; we rescale them to  $32 \times 16$  to introduce blur and avoid extremely sharp shadows. To enable shadow analysis, each scene is composed of both dynamic and static elements

(e.g., plates and blocks), all of which cast shadows. For each dataset, we provide ground truth albedo. The amount of specular reflectance varies across scenes and objects. Example renders from our dataset are shown in Fig. 12.



Figure 10. Example visualization of the ‘jumpingjacks’ scene at time steps 6, 46, 53, and 86. Note that each time step corresponds to a different camera pose, so the renderings are shown from varying viewpoints.



Figure 11. Environment maps used to light our scenes.

<sup>1</sup><https://www.mixamo.com/>

<sup>2</sup><https://polyhaven.com/hdriis>

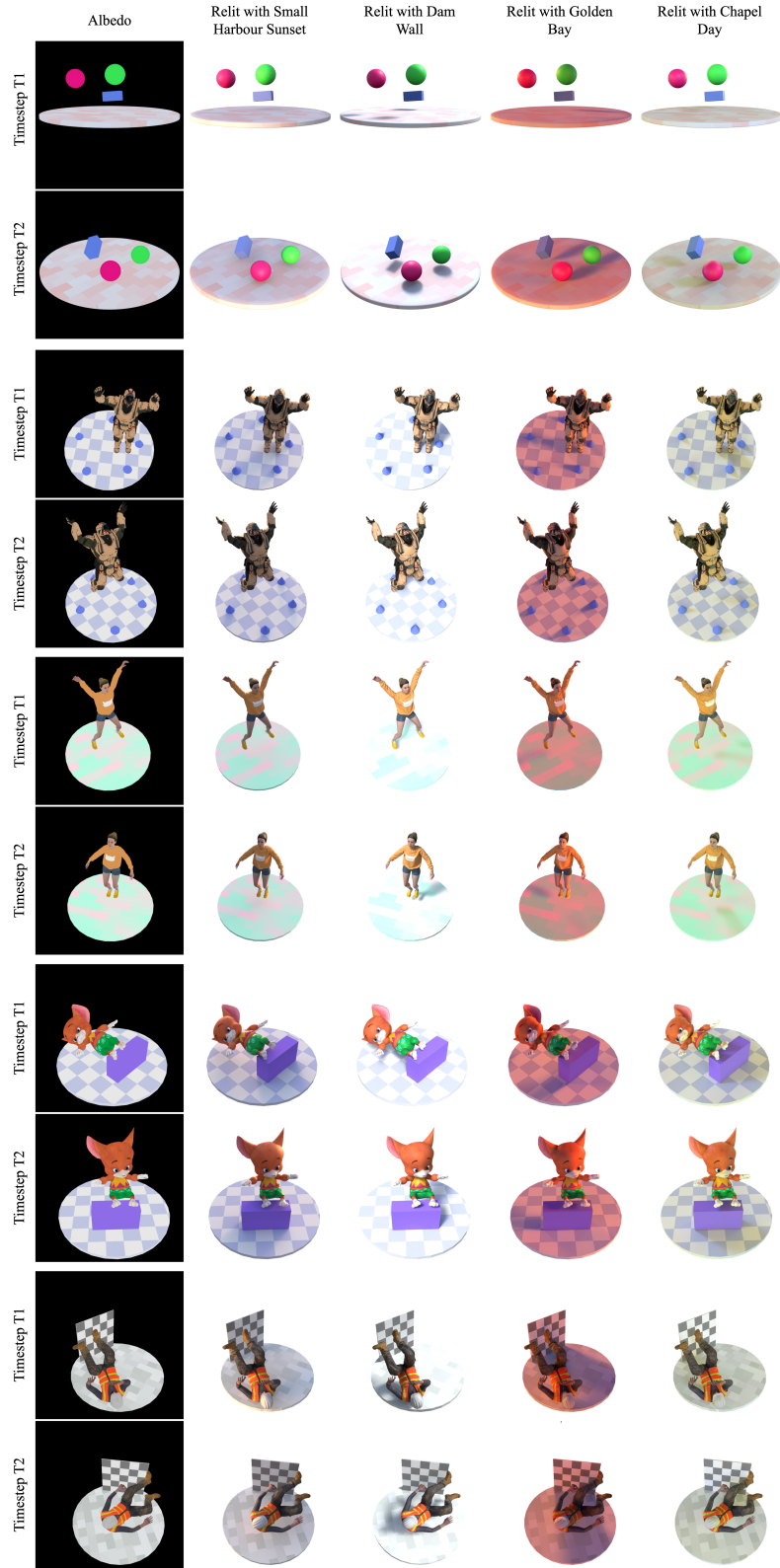


Figure 12. Example ground truth renders of our scenes.



## 6. Implementation details

We train each scene in two stages: 35,000 iterations in Stage 1 and 20,000 iterations in Stage 2. Our MLP architecture follows the design proposed in [5], consisting of an 8-layer MLP with a width of 256 units per layer. The learning rate for the MLP is set to 0.0008 and decays exponentially to 0.00008.

In Stage 1, we train using a combination of loss terms with the following weights (brackets show hyperparameter search range):

$$\lambda_n = 0.002, \quad \lambda_d = 1000, \quad \lambda_o = 0.1, \\ \lambda_P = \{0.001, 0.005\}, \quad \lambda_{\Delta c} = 0.01, \quad \lambda_{\Delta \mu} = \{0.0, 0.001\} \quad (1)$$

In Stage 2, we optimize the albedo, which is an RGB value assigned to each Gaussian, and roughness - values constant over time, and the environment map. The learning rates for the environment map, albedo and roughness are set to 0.2, 0.01, 0.005 respectively. The training environment map has a resolution of  $32 \times 16$  for synthetic data and  $128 \times 64$  for DNA scenes and ENERF data considering its very sharp shadows. We also finetune Gaussian colors from Stage 1 together with MLP head responsible for modeling  $d_{color}$ . This is important, since we use Stage 1 colors to compute indirect light for training, following [2]. We also finetune opacity to allow the model to remove some relight-related artifacts visible during training. The remaining parameters and MLP parts are frozen so the learned geometry from the Stage 1 is remained. All finetuned parameters in Stage 2 have their original lr lowered by 10 times.

During synthetic training, we sample 512 from the environment map. We randomly select  $N_r = 2^{18}$  rays per iteration, resulting in  $2^{18}/512$  pixels used to compute the  $\ell_1$  loss for synthetic data. For ENERF we use 1024 samples and  $2^{18} \cdot 16$  rays.

At inference time, we relight scenes using 1024 or 2048 sampled rays.

Please refer to our repository for the exact hyperparameter settings to reproduce our results.

## 7. Limitation - example

In Fig. 13, we illustrate the limitations of our dynamic training strategy. For more complex and detailed motions, for example near surfaces, simple separation may need to be replaced with more specialized supervision, such as optical flow.

## 8. Full affiliations

The full affiliations, abbreviated in the author section due to space constraints, are as follows: (1) Warsaw University of Technology, Poland; (2) Sano Centre for Computa-

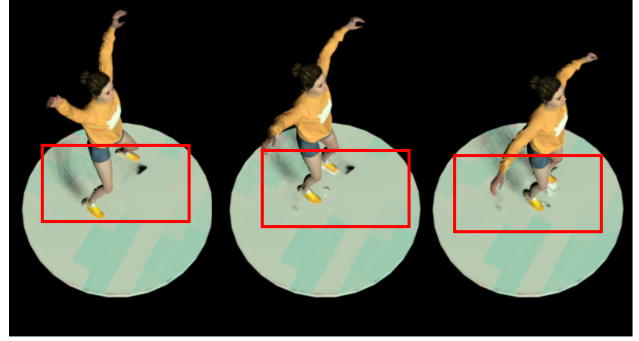


Figure 13. Limitations of simple separation strategy. Some Gaussians between the plate surface and the shoes are neither part of the static plate nor clearly part of the dynamic shoe.

tional Medicine, Kraków, Poland; (3) Institute for Biomedical Informatics, Faculty of Medicine and University Hospital Cologne, University of Cologne, Germany; (4) Faculty of Mathematics and Natural Sciences, University of Cologne, Germany; (5) Center for Molecular Medicine Cologne (CMMC), Faculty of Medicine and University Hospital Cologne, University of Cologne, Germany; (6) Jagiellonian University, Kraków, Poland; (7) IDEAS Research Institute, Warsaw, Poland; (8) Microsoft.

## References

- [1] Wei Cheng, Ruixiang Chen, Wanqi Yin, Siming Fan, Keyu Chen, Honglin He, Huiwen Luo, Zhongang Cai, Jingbo Wang, Yang Gao, Zhengming Yu, Zhengyu Lin, Daxuan Ren, Lei Yang, Ziwei Liu, Chen Change Loy, Chen Qian, Wayne Wu, Dahua Lin, Bo Dai, and Kwan-Yee Lin. Dna-rendering: A diverse neural actor repository for high-fidelity human-centric rendering. *arXiv preprint*, arXiv:2307.10173, 2023. 1
- [2] Chun Gu, Xiaofei Wei, Zixuan Zeng, Yuxuan Yao, and Li Zhang. Irgs: Inter-reflective gaussian splatting with 2d gaussian ray tracing. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 10943–10952, 2025. 11
- [3] Haotong Lin, Sida Peng, Zhen Xu, Yunzhi Yan, Qing Shuai, Hujun Bao, and Xiaowei Zhou. Efficient neural radiance fields for interactive free-viewpoint video. In *SIGGRAPH Asia Conference Proceedings*, 2022. 1
- [4] Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-nerf: Neural radiance fields for dynamic scenes. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10318–10327, 2021. 9
- [5] Ziyi Yang, Xinyu Gao, Wen Zhou, Shaohui Jiao, Yuqing Zhang, and Xiaogang Jin. Deformable 3d gaussians for high-fidelity monocular dynamic scene reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20331–20341, 2024. 11