

SAIL: Self-supervised Learning of Lighting-Invariant Representations from Real Images with Latent Diffusion

Supplementary Material

1. Latent lighting component regularization

Lighting invariance Unconditioned relighting and latent decomposition alone are not sufficient to achieve lighting invariance. The surrogate relighting task enables the model to learn changes of illumination, and the intrinsic latent decomposition separates the output into two components. However, what truly enforces lighting invariance are our latent regularizations: they ensure that one component remains identical across all lighting conditions of the same scene, while the other captures only the illumination. With these regularizations and sufficient lighting variation in the data, the model learns a lighting-invariant representation.

Analysis of Latents Distribution. To better understand the statistical behavior of the predicted lighting latent \hat{z}_i^E , we conduct an analysis under a fully supervised setup using ground-truth albedo from the MIDIntrinsics dataset [2]. We subtract the predicted albedo-like image \hat{z}^A from the encoded image z_i to isolate the illumination component \hat{z}_i^E . Fig. 7 shows the distributions of \hat{z}_i^E , \hat{z}^A , and z_i across the dataset. We observe that the illumination component \hat{z}_i^E is strongly biased toward negative values, justifying our design choice of applying a non-positivity constraint (see Eq. 9) to ensure that lighting-dependent information does not corrupt the albedo-like prediction.

2. Additional results

Figure 8 shows a qualitative comparison of predicted light-invariant images on the IIW and MAW datasets. Our method SAIL achieves more accurate color preservation, recovering the true colors of walls and objects without being biased by the ambient light color. While supervised methods [5, 7] effectively remove lighting effects, they often distort object colors and tend to erase entire regions behind transparent or reflective surfaces like mirrors and windows. In contrast, the self-supervised baseline [8] struggles to remove ambient illumination, producing albedo-like images with a strong color cast from the scene lighting. Our method overcomes these issues by producing more neutral and consistent light-invariant images, even in scenes with complex lighting and reflections.

Figure 9 illustrates albedo-like predictions across multiple lighting conditions from a scene from the MAW dataset. Even subtle changes in light intensity lead to noticeable shifts in LatentsIntrinsics [8] output, which retains color

biases from the scene illumination. IntrinsicDiffusion [5], while better at handling lighting effects, still alters the wall colors across conditions. In contrast, SAIL demonstrates strong consistency, producing stable albedo-like images that remain close to the true surface colors regardless of the lighting variation. This highlights the robustness of our approach in disentangling intrinsic appearance from external illumination.

3. Applications

Unconditioned scene relighting We show in Fig. 4 unconditioned scene relighting as an application of SAIL’s predicted albedo-like images. Our model predicts albedo-like estimates that are not affected by illumination or ambient light color, allowing for an effective separation between light-dependent and -independent properties.

We sample different relighting latents \hat{z}_i to generate multiple relit versions of the same input image. These relightings vary in ambient tone and illumination direction, producing realistic outputs under new lighting conditions. This illustrates the ability of our method to generalize relighting from a single image without any explicit supervision and conditioning.

We present in Fig. 5 unconditioned relighting results along with the corresponding decompositions predicted by SAIL. Although the light-invariant latent (Pred. Shading) does not correspond to a physically accurate shading map, it still captures lighting-related information, which allows our method to produce accurate relighting results. For comparison, we also include a pseudo-shading map (Comp. Shading) computed from the predicted albedo-like image and output relighting result, using the traditional shading equation.

Virtual scene relighting We demonstrate a downstream application of SAIL by relighting the predicted light-invariant images in Blender[3] using environment maps downloaded from PolyHaven¹. This experiment highlights the practical utility of our lighting-invariant albedo-like predictions. As shown in Fig. 3, our method produces light-invariant images that preserve object appearance and enable consistent relighting results.

Perception tasks We further demonstrate the utility of our albedo-like representation for downstream object detection in dark scenes. As shown in Fig. 6, objects are barely visible in the input and remain undetected when using prior intrinsic

¹<https://polyhaven.com/>

decomposition methods [5, 7, 8]. In contrast, our representation restores object visibility and enables YOLO [4] to detect them reliably.

Color editing In Fig. 1 and Fig. 2, we show results of color editing. We first predict the albedo-like representation of the input, then modify selected regions using a mask, and finally recompose the edited albedo-like with the original illumination using the intrinsic equation $I = A \cdot S$. This procedure yields realistic edits where the new colors remain consistent with shading and highlights. Additional examples with different colors and illumination conditions are provided in the supplementary video. We provide an accompanying video that illustrates additional qualitative results. The video shows the same scenes under different lighting conditions and with various color edits applied to the albedo-like representation. These results highlight how SAIL produces consistent decompositions across lighting changes and supports realistic edits that remain well integrated with shading and highlights.

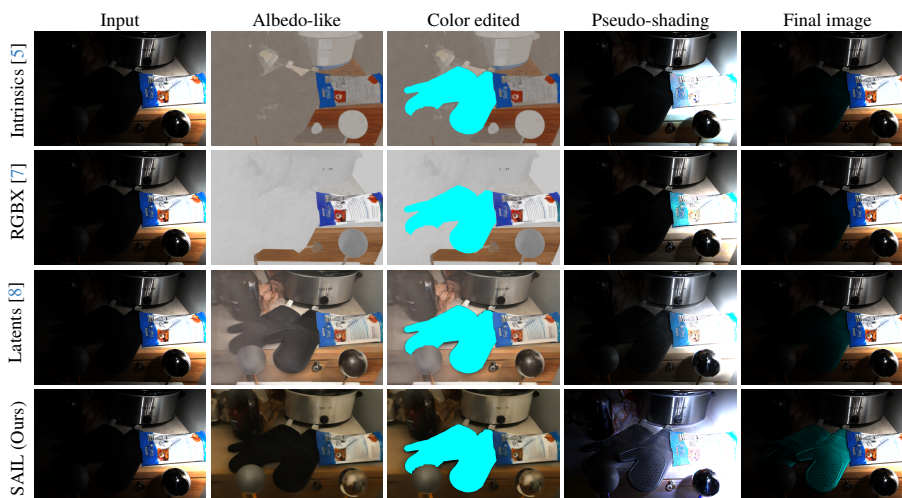


Figure 1. Our albedo-like representation enables realistic color editing in images, ensuring that the new colors remain well blended with the original lighting effects.

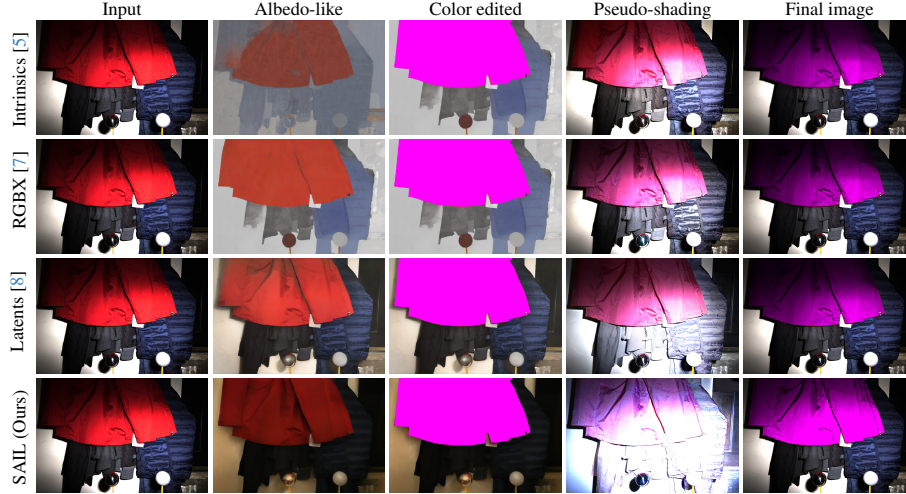


Figure 2. Our albedo-like representation enables realistic color editing in images, ensuring that the new colors remain well blended with the original lighting effects.

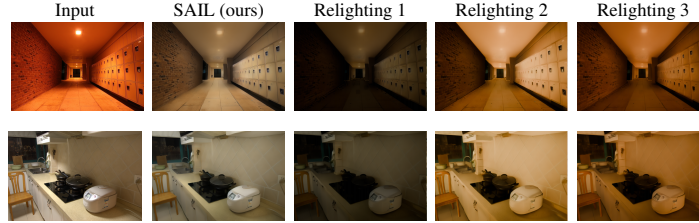


Figure 3. We show results of virtual relighting with Blender [3], where we relight the predicted albedo-like images with different environment maps.

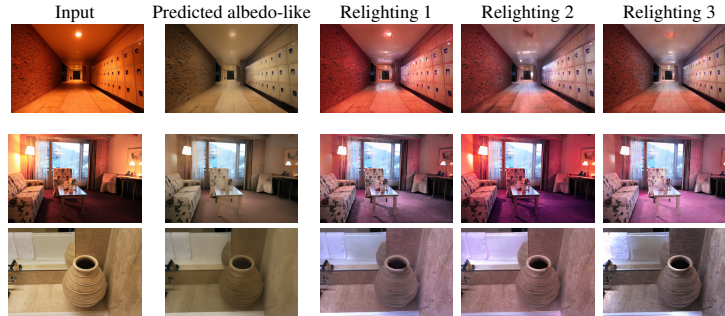


Figure 4. We show unconditioned relighting results predicted by SAIL at different inferences. Each output corresponds to a different latent relighting sample \hat{z}_i , demonstrating the ability of SAIL to generate diverse and plausible illumination conditions without any explicit supervision and conditioning.

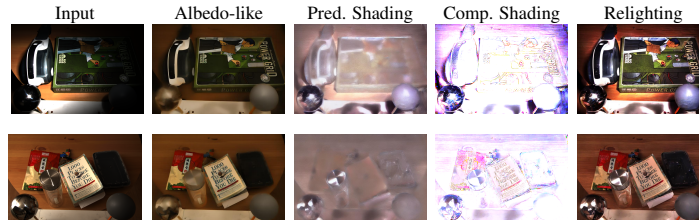


Figure 5. Unconditioned relighting results and corresponding decompositions predicted by SAIL. The light-invariant component (Pred. Shading) is not a physically accurate shading map but still captures lighting information, enabling accurate relighting. A pseudo-shading map (Comp. Shading), computed using the traditional shading equation, is shown for comparison.

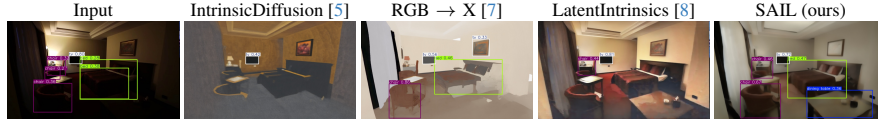


Figure 6. YOLO object detection under challenging low-light conditions. While objects are barely visible in the input, and remain undetected when using prior intrinsic decomposition methods, our albedo-like representation (SAIL) restores visibility and enables reliable detection.

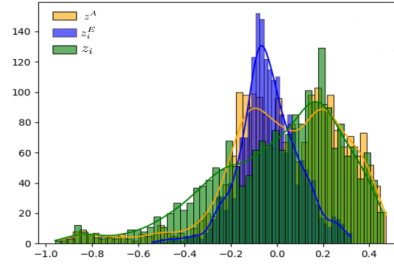


Figure 7. Analysis of of predicted latent distribution

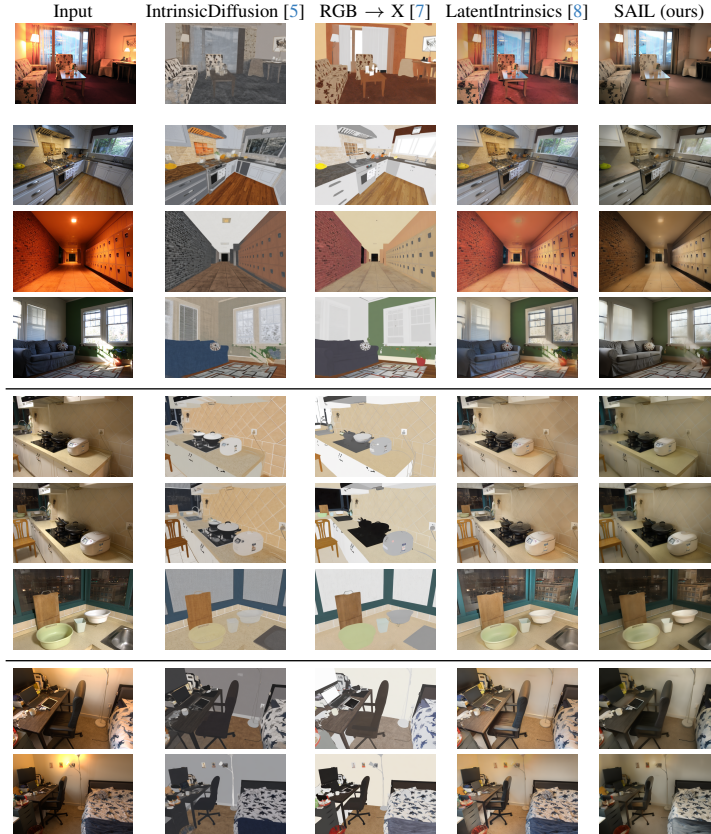


Figure 8. We qualitatively compare the predicted albedo-like images on the IIW dataset [1] and MAW dataset [6]. Our method, SAIL, produces albedo-like images that preserve the true colors of objects and walls, without being biased by ambient light color. It also more effectively removes reflections and shadows, resulting in consistent albedo-like estimates under varying lighting conditions.

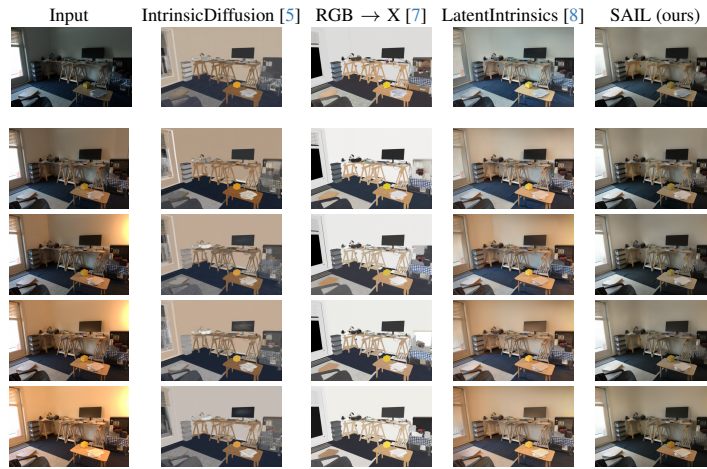


Figure 9. We qualitatively compare the predicted albedo-like images on the MAW dataset [6]. We show that SAIL predicts consistent albedo-like images from the same scene under various lighting conditions.

References

- [1] Sean Bell, Kavita Bala, and Noah Snavely. Intrinsic images in the wild. *ACM Transactions on Graphics (TOG)*, 33(4):1–12, 2014. [4](#)
- [2] Chris Careaga and Yağız Aksoy. Intrinsic image decomposition via ordinal shading. *ACM Trans. Graph.*, 2023. [1](#)
- [3] Blender Online Community. *Blender - a 3D modelling and rendering package*. Blender Foundation, 2018. [1](#), [3](#)
- [4] Glenn Jocher, Jing Qiu, and Ayush Chaurasia. Ultralytics YOLO, 2023. [2](#)
- [5] Peter Kocsis, Vincent Sitzmann, and Matthias Nießner. Intrinsic image diffusion for indoor single-view material estimation. *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*, pages 5198–5208, 2024. [1](#), [2](#), [3](#), [4](#), [5](#)
- [6] Jiaye Wu, Sanjoy Chowdhury, Hariharmano Shanmugaraja, David Jacobs, and Soumyadip Sengupta. Measured albedo in the wild: Filling the gap in intrinsics evaluation. In *2023 IEEE International Conference on Computational Photography (ICCP)*, pages 1–12, 2023. [4](#), [5](#)
- [7] Zheng Zeng, Valentin Deschaintre, Iliyan Georgiev, Yannick Hold-Geoffroy, Yiwei Hu, Fujun Luan, Ling-Qi Yan, and Miloš Hašan. $\text{Rgb} \leftrightarrow \text{x}$: Image decomposition and synthesis using material- and lighting-aware diffusion models. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers '24*, page 1–11. ACM, 2024. [1](#), [2](#), [3](#), [4](#), [5](#)
- [8] Xiao Zhang, William Gao, Seemantdar Jain, Michael Maire, David Forsyth, and Anand Bhattad. Latent intrinsics emerge from training to relight. *Advances in Neural Information Processing Systems*, 37:96775–96796, 2024. [1](#), [2](#), [3](#), [4](#), [5](#)