UniLight: A Unified Representation for Lighting

Zitian Zhang[1,2], Iliyan Georgiev[2], Michael Fischer[2], Yannick Hold-Geoffroy[2], Jean-François Lalonde[1], Valentin Deschaintre[2]

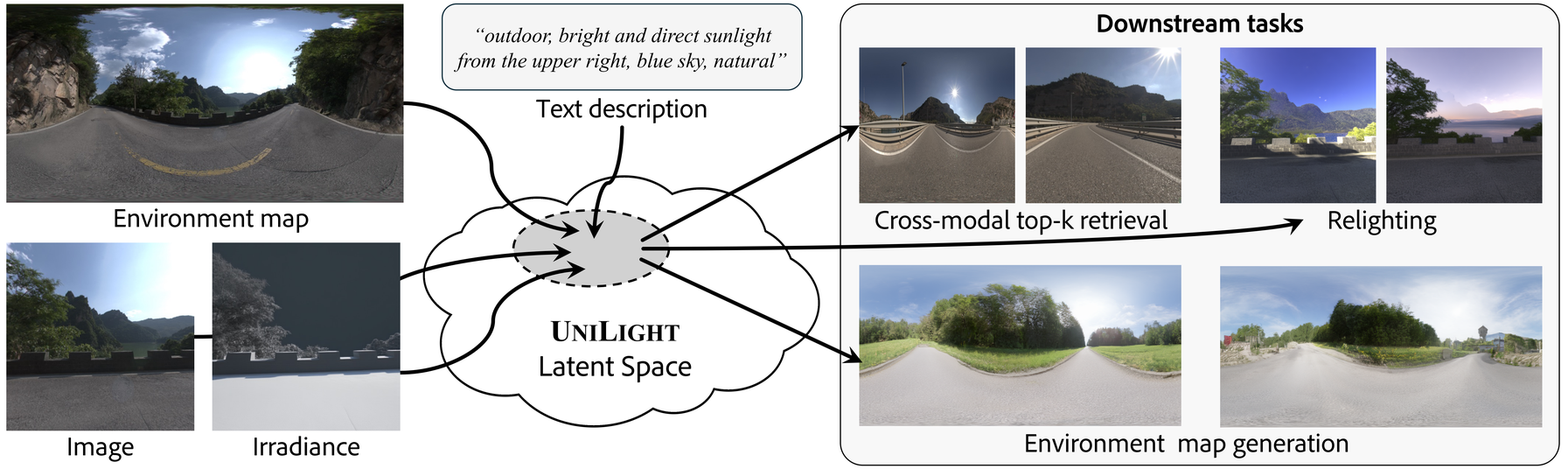
[1] Université Laval [2] Adobe

CVPR JUNE 3-7, 2026 DENVER COLORADO

TL;DR

How can we represent lighting in images?

Encode **different light modalities** into a **joint latent space**, that can be used for retrieval, relighting and env. map generation.



RETRIEVAL

Cross-modality retrieval based on the cosine similarity between latents.

RELIGHTING

Fine-tune an X→RGB model to receive original scene intrinsics (depth, albedo and normals), and UniLight latents as conditioning for relighting.

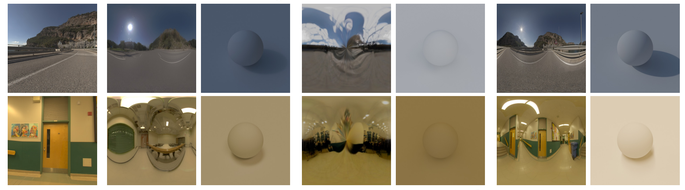
ENVIRONMENT MAP GENERATION

Fine-tune a diffusion model to receive UniLight latents as conditioning, to generate env. map from either image, irradiance, or text descriptions.

Metrics on the rendering →

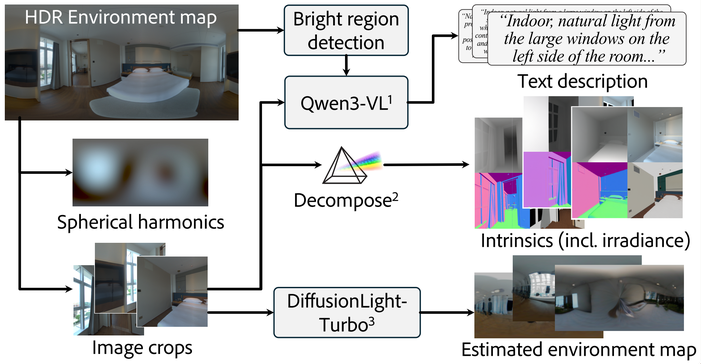
Method	PSNR↑	RMSE↓	SI-RMSE↓	SSIM↑	MAE↓	LPIPS↓
DiffusionLight-Turbo	27.77	0.157	0.062	0.902	0.148	0.088
UniLight	**28.85**	**0.133**	**0.060**	**0.915**	**0.124**	**0.079**

Visualization of the env. map and rendering ↓



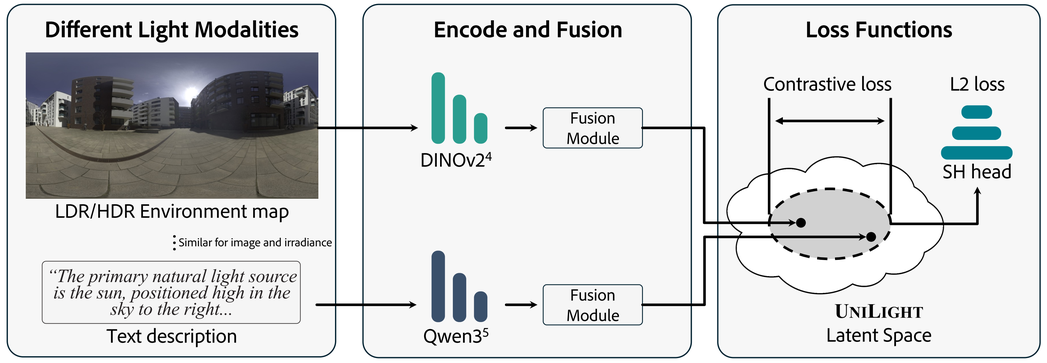
ALIGNED LIGHT MODALITIES

Starting from an HDR env. map., we extract aligned image crops, spherical harmonics, text description, intrinsics, and estimated env. map.



ARCHITECTURE

Modality-specific encoders for env. maps, images, irradiance, and text are trained contrastively to align their representations, with an auxiliary spherical-harmonics prediction task reinforcing directional understanding.



zitian.zhang.1@ulaval.ca, igeorgiev@adobe.com, mifischer@adobe.com, holdgeof@adobe.com, jflalonde@gel.ulaval.ca, deschain@adobe.com

[1] Bai et al., Qwen3-vl, arXiv 2026. [2] Dirik et al., PRISM, ICPR 2026. [3] Chinchuthakun et al., DiffusionLight-Turbo, TPAMI 2026. [4] Oquab et al., DINOv2, arXiv 2023, [5] Zhang et al., Qwen3 embedding, arXiv 2025.