Temporally Stable Generative Illumination with a One-Step Diffusion Model

Harish Anand*, Alexandr Kuznetsov*, Sungye Kim*, Wojciech Uss, Wojciech Kaliński, Rama Harihara

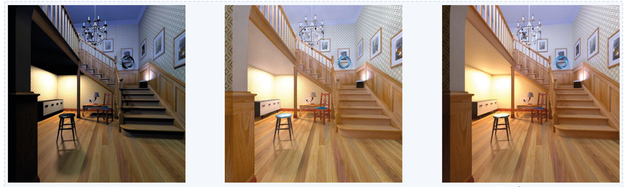
Advanced Micro Devices, Inc.

ECCV 2026 | MALMÖ | SEPT 8–12

Goal & Challenges

Generative Global Illumination

- Given sparse lighting hints grounded by materials and geometry (G-buffers), generate temporally stable global illumination without full path tracing, by leveraging strong priors present in pre-trained generative models.



Direct Lighting | Ours | Reference

Synthetic Dataset

- Render 31k frames (23k train/8k test) using the Blender Cycles renderer
- Dynamically moving camera, objects in motion and illumination from 1–30 area lights of varying sizes, intensities, and color.

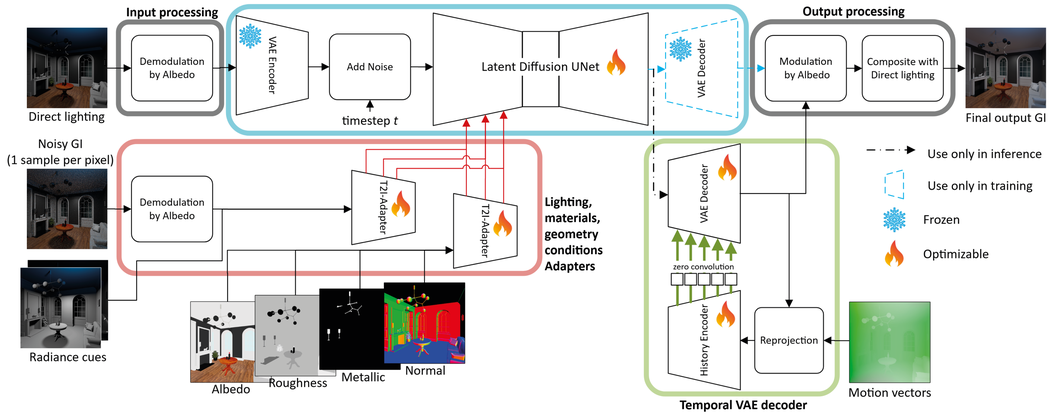
Challenges

- How to keep generation aligned to geometry and lighting?
- How to stay temporally stable without using video diffusion models?
- How to reach practical speed without iterative denoising?

Approach

Pipeline Overview

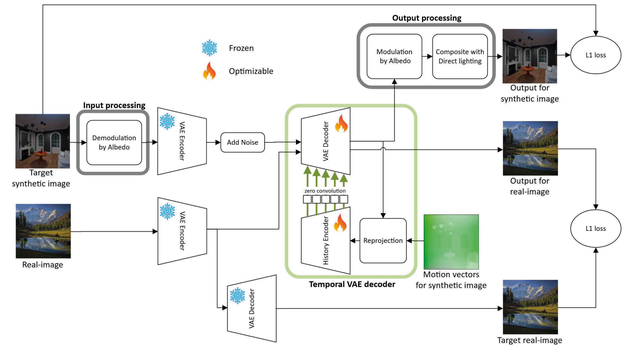
Overview of our One-Step Diffusion model-based pipeline.



- Backbone: Stable Diffusion 2.1-Turbo
- Direct lighting is demodulated by albedo, so the diffusion UNet sees illumination without material colour or textures.
- Input to the diffusion UNet is direct lighting, encoded and noised at timestep 499, giving a one-step solution.
- Two T2I-Adapters: one carries lighting cues, the other material and geometry details (G-buffers).
- Fine-tune on AMD Instinct™ MI210 GPUs for 50 epochs (batch size 32) with pixel-space optimizations.
- At inference time, we replace the SD-VAE decoder with our temporal VAE (TVAE) decoder to preserve temporal stability.

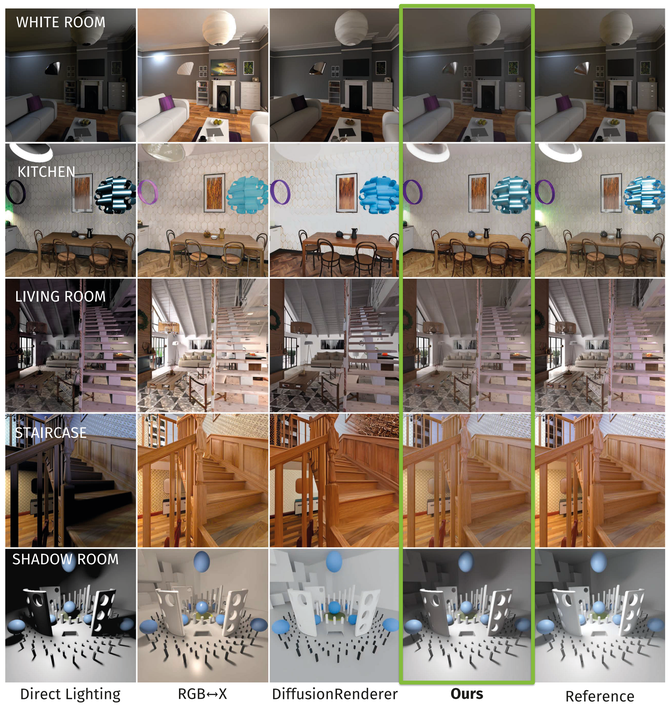
Temporal VAE Decoder

Training the Temporal VAE Decoder



- Teacher-forcing training strategy with synthetic data, and employing real image data to prevent drift
- Decodes current frame's latent with previous output, reprojected using motion vectors
- History encoder injects previous frame into decoder at multiple levels through zero convolutions
- Dataset: our synthetic dataset sequences, plus DIV2K with previous frame zeroed

Results

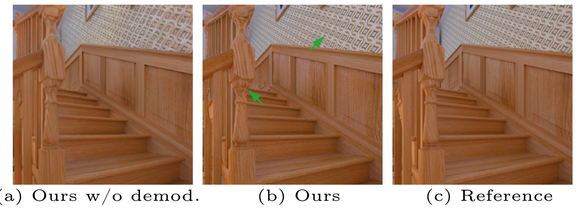


Ablation Study

- Concatenation vs T2I-Adapters

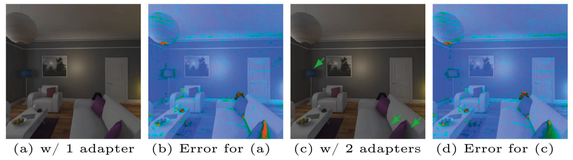
(a) Ours w/ concatenated conditions (b) Ours (w/ Adapters)

- Demodulating by albedo lets the UNet focus on illumination



(a) Ours w/o demod. (b) Ours (c) Reference

- Two adapters produce better indirect lighting



(a) w/ 1 adapter (b) Error for (a) (c) w/ 2 adapters (d) Error for (c)

- Ours-C: compact conditioning for real-world images

RGB↔X | Ours-C | Real-world data

Original images courtesy of Pexels (https://www.pexels.com)

- Runtime performance (unoptimized PyTorch model)

512×512 resolution	Ours	Ours-C	RGB↔X	DiffusionRenderer
Runtime (seconds)	0.29	0.284	3.25	52.09
GPU memory (GB)	8.68	7.97	3.76	21.81

Conclusions

Global Illumination from a One-Step Diffusion Model

Generates global illumination while preserving alignment with intrinsic scene signals by injecting lighting hints via the VAE and T2I-Adapters.

Temporal VAE decoder

Introduces a temporal VAE decoder that keeps generated sequences stable over an unlimited number of frames, and plugs into an existing diffusion pipeline at inference time.

Compact variant for real-world data

Extends the method to real-world images with compact conditioning inputs and stays robust even when the estimated intrinsics are low quality.

Scan for our research blog

* Equal contribution. Corresponding author: sungye.kim@amd.com