

Lightweight Attention-based Indirect Illumination

Wojciech Uss

Advanced Micro Devices (AMD)
Gdańsk, Poland
Wojciech.Uss@amd.com

SungYe Kim

Advanced Micro Devices (AMD)
Folsom, USA
SungYe.Kim@amd.com

Wojciech Kaliński

Advanced Micro Devices (AMD)
Warsaw, Poland
Wojciech.Kalinski@amd.com

Harish Anand

Advanced Micro Devices (AMD)
San Diego, USA
Harish.Anand@amd.com

Alexandr Kuznetsov

Advanced Micro Devices (AMD)
Seattle, USA
Alexandr.Kuznetsov@amd.com

Rama Harihara

Advanced Micro Devices (AMD)
San Jose, USA
Rama.Harihara@amd.com

Abstract

Neural global illumination methods face a fundamental tradeoff: screen-space approaches are fast but miss off-screen lighting, while scene-wide approaches require pipeline-incompatible preprocessing for input generation. We present a lightweight attention-based model that bridges this gap by augmenting screen-space G-buffers with reflective shadow map data rendered from light-source perspectives. Shifting resources from model complexity to richer inputs, our approach achieves improved indirect shadow quality and better generalization to unseen scenes.

CCS Concepts

• Computing methodologies → Rendering; Machine learning.

ACM Reference Format:

Wojciech Uss, SungYe Kim, Wojciech Kaliński, Harish Anand, Alexandr Kuznetsov, and Rama Harihara. 2026. Lightweight Attention-based Indirect Illumination. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Posters (SIGGRAPH Posters '26)*, July 19–23, 2026, Los Angeles, CA, USA. ACM, New York, NY, USA, Article 00, 3 pages. <https://doi.org/10.1145/3799825.3818739>

1 Introduction

Producing accurate global illumination (GI) in real time requires balancing fidelity against strict computational budgets. Screen-space neural GI methods [Gai et al. 2025; Nalbach et al. 2017; Xin et al. 2020] are fast but inherently lack information about off-screen light sources, which limits the quality of indirect illumination they can reconstruct. Scene-wide approaches address this limitation by gathering point-cloud data across the full scene [Xu et al. 2025] or representing triangles as a sequence of tokens [Zeng et al. 2025] to simulate light transport via attention mechanisms [Vaswani et al. 2017].

Authors' Contact Information: Wojciech Uss, Advanced Micro Devices (AMD), Gdańsk, Poland, Wojciech.Uss@amd.com; SungYe Kim, Advanced Micro Devices (AMD), Folsom, USA, SungYe.Kim@amd.com; Wojciech Kaliński, Advanced Micro Devices (AMD), Warsaw, Poland, Wojciech.Kalinski@amd.com; Harish Anand, Advanced Micro Devices (AMD), San Diego, USA, Harish.Anand@amd.com; Alexandr Kuznetsov, Advanced Micro Devices (AMD), Seattle, USA, Alexandr.Kuznetsov@amd.com; Rama Harihara, Advanced Micro Devices (AMD), San Jose, USA, Rama.Harihara@amd.com.



This work is licensed under a Creative Commons Attribution 4.0 International License. *SIGGRAPH Posters '26, Los Angeles, CA, USA*
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2548-7/26/07
<https://doi.org/10.1145/3799825.3818739>

However, they require a dedicated data-gathering process incompatible with standard rendering pipelines. LightFormer [Ren et al. 2024] strikes a middle ground, augmenting screen-space buffers with reflective shadow maps (RSM) [Dachsbacher and Stamminger 2005] rendered from light sources, but it was trained and evaluated per scene with partial overlap between training and test data, lacking generalization.

We address these limitations with a distinct lightweight attention-based approach for global illumination that improves indirect shadow quality and, unlike prior work [Ren et al. 2024], generalizes to scenes not seen during training. Our key design choice is to shift the resource budget from model complexity to richer input generation: a simpler model consuming more informative input data.

2 Approach

2.1 Data

We take the direct illumination buffer of the main camera from the rendering engine and predict the indirect illumination component only. Our approach infers indirect illumination primarily from reflected flux buffers. These buffers are generated from light-source perspectives: independent equirectangular cameras placed at the light's position and oriented along its direction, capturing G-buffers (albedo, position, normal, metallic). Together with the light source's power, these are used to compute reflected flux buffers [Dachsbacher and Stamminger 2005]. As in prior work [Ren et al. 2024], each texel of the RSM data is treated as an indirect virtual point light (iVPL)—a small light source contributing only to indirect illumination. Crucially, iVPL texels often capture geometry beyond the camera's view, enabling the model to reconstruct off-screen effects. Additionally, we use a one-bounce, one-sample-per-pixel (1spp) illumination buffer from the main camera as a guide for predicting indirect illumination. This buffer is inexpensive and very noisy, but provides a valuable hint that improves results. For high-quality ground truth (GT), frames are rendered with high sample counts (e.g., 2k–16k samples per pixel), followed by OptiX [NVIDIA 2021] denoising. We render our training dataset from 11 scenes (14,320 frames) using Blender [Community 2018], restricted to area light sources of varying size, shape, color, and intensity. Main camera buffers are at 512×512 resolution; RSM buffers are at 128×128 resolution. Data coming from various scenes is normalized with respect to the scene size. We improve robustness by augmenting with random exposure scale and scene size perturbations.

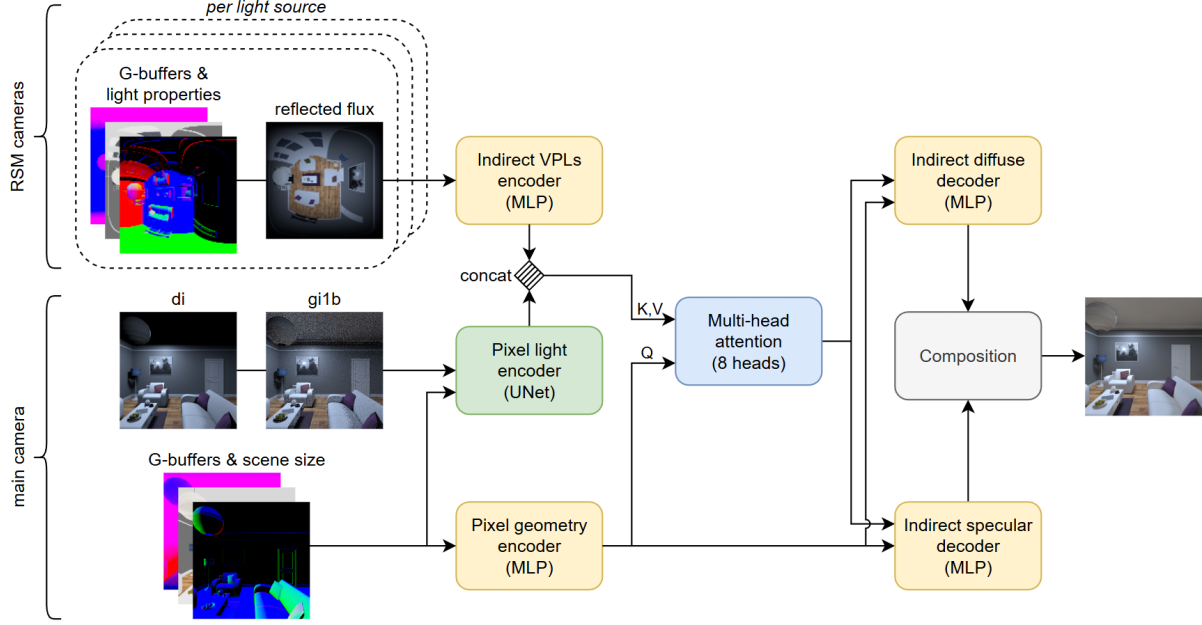


Figure 1: Overview of our approach. Input buffers rendered from the main camera perspective: direct illumination (di), one-bounce 1spp illumination ($gi1b$), G-buffers, and scene size. Input buffers captured by RSM cameras: G-buffers and light properties (reflected flux buffers). The model predicts indirect illumination components (diffuse and specular), which are composed with direct illumination to produce the final global illumination image. Connections from di and diffuse albedo to the composition module are omitted for brevity. The indirect VPLs encoder is a 5-layer MLP. The pixel geometry encoder is a 4-layer MLP. The MHA module has eight heads. The indirect diffuse and specular decoders are both 4-layer MLPs.

2.2 Model

An overview of our approach is shown in Figure 1. By exploiting the linearity of light transport, we separately predict the diffuse and specular components of indirect illumination and compose the final global illumination image using Equation (1).

$$I_{gi} = I_{di} + (\hat{I}_{diff} \cdot A_{diff} + I_{spec}), \quad (1)$$

where I_{gi} is the final global illumination image, I_{di} is the direct illumination (zero-bounce), $\hat{I}_{diff}$ is the predicted diffuse indirect illumination intrinsic (albedo-free), A_{diff} is the diffuse albedo, and I_{spec} is the predicted specular indirect illumination.

The approach consists of three stages: encoding, attention, and decoding. Three encoders generate embeddings from distinct data sources: The *iVPL encoder* is an MLP that encodes reflected flux, G-buffers (position, normal, albedo, metallic, direction to camera, distance to camera), and light properties (size, power, color, shape), similar to prior work [Ren et al. 2024]. The *pixel light encoder* is a UNet that encodes screen-space radiance buffers (direct illumination and one-bounce illumination) and G-buffers (position, normal, roughness, metallic, diffuse albedo, specular albedo). We use a UNet here, rather than a per-pixel MLP, because aggregating spatial context across larger receptive fields is essential for extracting a useful signal from the noisy one-bounce illumination buffer. The *pixel geometry encoder* is an MLP that encodes G-buffers (position, normal, roughness, metallic, diffuse albedo, specular albedo, view direction) from the main camera and scene size. The *iVPL encoder* and *pixel*

geometry encoder process data per pixel. As in prior work [Ren et al. 2024], iVPL embeddings are averaged per light and expanded to screen-space resolution, producing one buffer per light source. These are then concatenated with pixel light embeddings to form the light data embedding.

These embeddings are then processed by the core of our model, a *multi-head attention* module. It relates geometry and material properties from the main camera (pixel geometry data) with radiance information from the main camera (pixel light data) and light-source cameras (iVPL data) by taking pixel geometry embeddings as the query (Q) and the light data embedding as key (K) and value (V). Its output is a composite light embedding for each shading point. This embedding, concatenated with the pixel geometry embedding, is fed into two shading decoders: an *indirect diffuse decoder* and an *indirect specular decoder*, both MLP modules, yielding $\hat{I}_{diff}$ and I_{spec} in Equation (1), respectively.

Our model is trained with a weighted combination of loss functions: L1 in linear space (weight 0.1), and L1, MS-SSIM, Luminance, and LPIPS in tonemapped space (weights 0.2, 1.0, 0.5, and 1.0, respectively).

Compared to LightFormer [Ren et al. 2024], we shifted the resource budget from model complexity to richer input generation. The model is simpler yet consumes more informative input, yielding improved indirect shadow quality and generalization to unseen scenes.

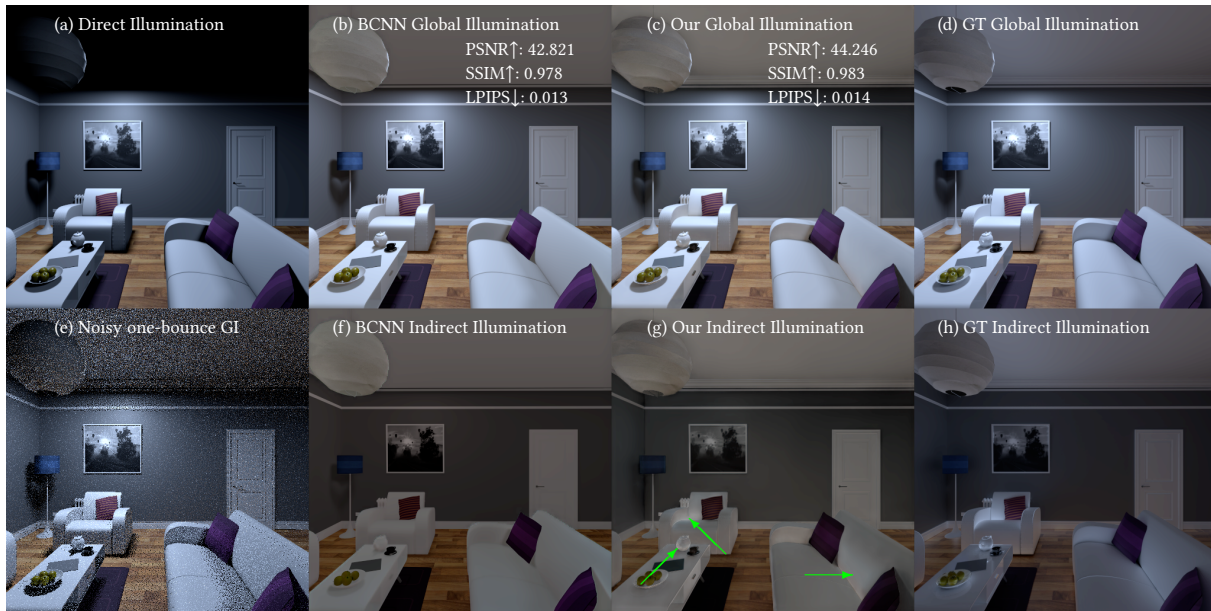


Figure 2: Result from an unseen test-only scene. We compare our final global illumination result (c) and indirect illumination result (g) to the corresponding BCNN [Xin et al. 2020] results (b) and (f), and the ground truth (GT). Arrows on our result highlight areas where our method predicts plausible indirect illumination.

3 Results

Figure 2 illustrates an inference result from THE WHITE ROOM scene [Bitterli 2016], a test-only scene entirely disjoint from our training data. Complex specular highlights on the couch, armchair, and teapot—as seen in the ground truth (GT) indirect illumination (h)—are faithfully reconstructed, even though the noisy one-bounce guide buffer provides little useful signal for these effects. For comparison, we show the result of BCNN [Xin et al. 2020] trained on the same data. Our result exhibits substantially richer indirect illumination effects. A residual color shift in the output remains an open issue. We observe empirically that adding more scenes to the training set consistently reduces this bias on held-out test scenes, suggesting that limited scene diversity is a contributing factor. This also suggests the current architecture is not yet capacity-constrained, leaving room to scale the dataset further before revisiting model size.

Our model has 2.19 million trainable parameters and requires 432 GFLOPs. Without inference optimizations like layer fusion and low-precision quantization in our PyTorch implementation, the model runs in 45.56 ms at 512×512 resolution on an AMD Instinct™ MI250 GPU.

4 Conclusions & Future Work

We demonstrated that shifting the resource budget from model complexity to richer input enables a lightweight attention-based architecture to generalize to unseen scenes while reconstructing complex indirect illumination effects. Future work will aim to address the residual color shift by extending the training dataset and explore reducing the dependency on the one-bounce guide buffer through curriculum learning and a dropout strategy.

References

- Benedikt Bitterli. 2016. Rendering resources. <https://benedikt-bitterli.me/resources/>.
- Blender Online Community. 2018. Blender - a 3D modelling and rendering package. <http://www.blender.org>.
- Carsten Dachsbacher and Marc Stamminger. 2005. Reflective shadow maps. In *Proceedings of the 2005 Symposium on Interactive 3D Graphics and Games* (Washington, District of Columbia) (*I3D '05*). Association for Computing Machinery, New York, NY, USA, 203–231. doi:10.1145/1053427.1053460
- Meng Gai, Guoping Wang, and Sheng Li. 2025. Geometry-Aware Global Feature Aggregation for Real-Time Indirect Illumination. doi:10.48550/arXiv.2508.08826
- O. Nalbach, E. Arabadzhiyska, D. Mehta, H.-P. Seidel, and T. Ritschel. 2017. Deep Shading: Convolutional Neural Networks for Screen Space Shading. *Computer Graphics Forum* 36, 4 (2017), 65–78. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.13225 doi:10.1111/cgf.13225
- Nvidia. 2021. OptiX AI-Accelerated Denoiser. <https://developer.nvidia.com/optixdenoiser>.
- Haocheng Ren, Yuchi Huo, Yifan Peng, Hongtao Sheng, Weidong Xue, Hongxiang Huang, Jingzhen Lan, Rui Wang, and Huijun Bao. 2024. LightFormer: Light-Oriented Global Neural Rendering in Dynamic Scene. *ACM Trans. Graph.* 43, 4, Article 75 (July 2024), 14 pages. doi:10.1145/3658229
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems*, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), Vol. 30. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf
- Hanggao Xin, Shaokun Zheng, Kun Xu, and Ling-Qi Yan. 2020. Lightweight Bilateral Convolutional Neural Networks for Interactive Single-Bounce Diffuse Indirect Illumination. *IEEE transactions on visualization and computer graphics* PP (09 2020). doi:10.1109/TVCG.2020.3023129
- Bing Xu, Mukund Varma T, Cheng Wang, Tzuma Li, Lifan Wu, Bartłomiej Wronski, Ravi Ramamoorthi, and Marco Salvi. 2025. A Generalizable Light Transport 3D Embedding for Global Illumination. arXiv:2510.18189 [cs.GR] <https://arxiv.org/abs/2510.18189>
- Chong Zeng, Yue Dong, Pieter Peers, Hongzhi Wu, and Xin Tong. 2025. RenderFormer: Transformer-based Neural Rendering of Triangle Meshes with Global Illumination. In *ACM SIGGRAPH 2025 Conference Papers*. Association for Computing Machinery, New York, NY, USA, 11 pages.