IRISformer: Dense Vision Transformers for Single-Image Inverse Rendering in Indoor Scenes
Rui Zhu1 · Zhengqin Li1 · Janarbek Matai2 · Fatih Porikli2 · Mannmohan Chandraker1 · 1UC San Diego · 2Qualcomm AI Research

Overview
Datasets
• Training: our in-house developed OpenRooms dataset [2] for large-scale photorealistic renderings of indoor scenes, with ground truth full 3D geometry, material, lighting and semantics. OpenRooms is used to train the entire pipeline from scratch, with full supervision on all tasks.
• Evaluation:
  • Estimation:
    • albedo estimation after finetuning: IW dataset.
    • geometry estimation (depth, normals) after finetuning: NYUV2 dataset.
  • object insertion/material editing: natural images dataset from Garon et al. [3].

Method

Evaluation: synthetic images

Evaluation: object insertion/material editing

Jointly evaluate geometry and lighting via rendering virtual objects into the scene.

Sample 1: better highlights by ours, on the center object.
Sample 2&3: more globally uniform illumination across multiple locations, in both the lighting intensities and directions.

We also carry out an A/B study using the insertion results-L1 and material editing-

Evaluation: real-world images

Table 5. Mean squared error for the illumination estimates 

Table 6. Errors on BRDF, geometry and lighting with a base of $10^{-4}$ on OpenRooms [27]. Lower is better. For lighting estimation, $R$ is the lighting reconstruction error and $L$ is the combined lighting loss for which LightNet is trained.

We also carry out an A/B study using the insertion results-L1 and material editing-

Evaluation: object insertion/material editing

Jointly evaluate geometry and lighting via rendering virtual objects into the scene.

Sample 1: better highlights by ours, on the center object.
Sample 2&3: more globally uniform illumination across multiple locations, in both the lighting intensities and directions.

We also carry out an A/B study using the insertion results-L1 and material editing-

Table 5. Mean squared error for the illumination estimates 

Table 6. Errors on BRDF, geometry and lighting with a base of $10^{-4}$ on OpenRooms [27]. Lower is better. For lighting estimation, $R$ is the lighting reconstruction error and $L$ is the combined lighting loss for which LightNet is trained.

We also carry out an A/B study using the insertion results-L1 and material editing-

Evaluation: object insertion/material editing

Jointly evaluate geometry and lighting via rendering virtual objects into the scene.

Sample 1: better highlights by ours, on the center object.
Sample 2&3: more globally uniform illumination across multiple locations, in both the lighting intensities and directions.

We also carry out an A/B study using the insertion results-L1 and material editing-

Table 5. Mean squared error for the illumination estimates 

Table 6. Errors on BRDF, geometry and lighting with a base of $10^{-4}$ on OpenRooms [27]. Lower is better. For lighting estimation, $R$ is the lighting reconstruction error and $L$ is the combined lighting loss for which LightNet is trained.

We also carry out an A/B study using the insertion results-L1 and material editing-

Table 5. Mean squared error for the illumination estimates 

Table 6. Errors on BRDF, geometry and lighting with a base of $10^{-4}$ on OpenRooms [27]. Lower is better. For lighting estimation, $R$ is the lighting reconstruction error and $L$ is the combined lighting loss for which LightNet is trained.

We also carry out an A/B study using the insertion results-L1 and material editing-

Table 5. Mean squared error for the illumination estimates 

Table 6. Errors on BRDF, geometry and lighting with a base of $10^{-4}$ on OpenRooms [27]. Lower is better. For lighting estimation, $R$ is the lighting reconstruction error and $L$ is the combined lighting loss for which LightNet is trained.

We also carry out an A/B study using the insertion results-L1 and material editing-

Table 5. Mean squared error for the illumination estimates 

Table 6. Errors on BRDF, geometry and lighting with a base of $10^{-4}$ on OpenRooms [27]. Lower is better. For lighting estimation, $R$ is the lighting reconstruction error and $L$ is the combined lighting loss for which LightNet is trained.

We also carry out an A/B study using the insertion results-L1 and material editing-

Table 5. Mean squared error for the illumination estimates 

Table 6. Errors on BRDF, geometry and lighting with a base of $10^{-4}$ on OpenRooms [27]. Lower is better. For lighting estimation, $R$ is the lighting reconstruction error and $L$ is the combined lighting loss for which LightNet is trained.

We also carry out an A/B study using the insertion results-L1 and material editing-

Table 5. Mean squared error for the illumination estimates 

Table 6. Errors on BRDF, geometry and lighting with a base of $10^{-4}$ on OpenRooms [27]. Lower is better. For lighting estimation, $R$ is the lighting reconstruction error and $L$ is the combined lighting loss for which LightNet is trained.

We also carry out an A/B study using the insertion results-L1 and material editing-

Table 5. Mean squared error for the illumination estimates 

Table 6. Errors on BRDF, geometry and lighting with a base of $10^{-4}$ on OpenRooms [27]. Lower is better. For lighting estimation, $R$ is the lighting reconstruction error and $L$ is the combined lighting loss for which LightNet is trained.

We also carry out an A/B study using the insertion results-L1 and material editing-

Table 5. Mean squared error for the illumination estimates 

Table 6. Errors on BRDF, geometry and lighting with a base of $10^{-4}$ on OpenRooms [27]. Lower is better. For lighting estimation, $R$ is the lighting reconstruction error and $L$ is the combined lighting loss for which LightNet is trained.

We also carry out an A/B study using the insertion results-L1 and material editing-

Table 5. Mean squared error for the illumination estimates 

Table 6. Errors on BRDF, geometry and lighting with a base of $10^{-4}$ on OpenRooms [27]. Lower is better. For lighting estimation, $R$ is the lighting reconstruction error and $L$ is the combined lighting loss for which LightNet is trained.

We also carry out an A/B study using the insertion results-L1 and material editing-

Table 5. Mean squared error for the illumination estimates 

Table 6. Errors on BRDF, geometry and lighting with a base of $10^{-4}$ on OpenRooms [27]. Lower is better. For lighting estimation, $R$ is the lighting reconstruction error and $L$ is the combined lighting loss for which LightNet is trained.

We also carry out an A/B study using the insertion results-L1 and material editing-

Table 5. Mean squared error for the illumination estimates 

Table 6. Errors on BRDF, geometry and lighting with a base of $10^{-4}$ on OpenRooms [27]. Lower is better. For lighting estimation, $R$ is the lighting reconstruction error and $L$ is the combined lighting loss for which LightNet is trained.

We also carry out an A/B study using the insertion results-L1 and material editing-

Table 5. Mean squared error for the illumination estimates 

Table 6. Errors on BRDF, geometry and lighting with a base of $10^{-4}$ on OpenRooms [27]. Lower is better. For lighting estimation, $R$ is the lighting reconstruction error and $L$ is the combined lighting loss for which LightNet is trained.

We also carry out an A/B study using the insertion results-L1 and material editing-

Table 5. Mean squared error for the illumination estimates 

Table 6. Errors on BRDF, geometry and lighting with a base of $10^{-4}$ on OpenRooms [27]. Lower is better. For lighting estimation, $R$ is the lighting reconstruction error and $L$ is the combined lighting loss for which LightNet is trained.