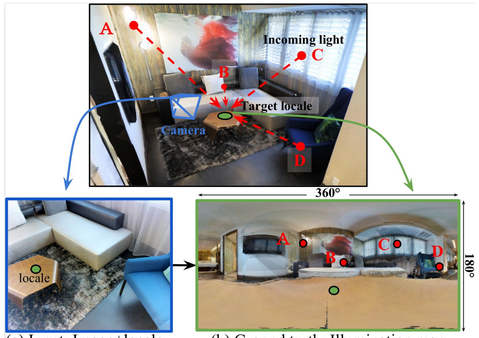
Neural Illumination: Lighting Prediction for Indoor Environments

Princeton University

Shuran Song and Thomas Funkhouser

Google Research

Goal



(a) Input: Image+locale (b) Ground truth: Illumination map

Goal: Given a single LDR image and a selected 2D pixel, we want to infer a panoramic HDR illumination map representing the light arriving from all directions at the target locale.

The illumination map: is encoded as a spherical image parameterized horizontally by φ (0-360) and vertically by θ (0-180), where each pixel stores the RGB intensity of light arriving at the "locale" from the direction (φ, θ).

Challenges

The task requires a **comprehensive** understanding of the scene, which includes:

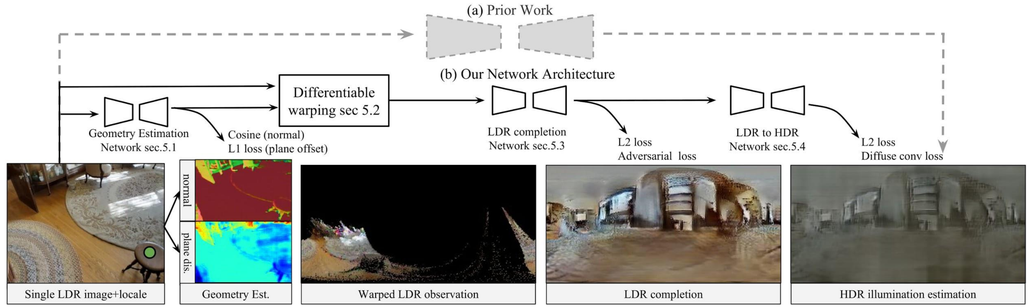
- The 3D location of that selected pixel,
- The 3D scene geometry to reason about occlusions.
- the distribution of unobserved light sources.
- The missing high dynamic range information.

References

[1] A. Chang, A. Dai, T. Funkhouser, M. Halber, M. Nießner, M. Savva, S. Song, A. Zeng, and Y. Zhang. Matterport3d:Learning from rgb-d data in indoor environments. 3DV, 2017.

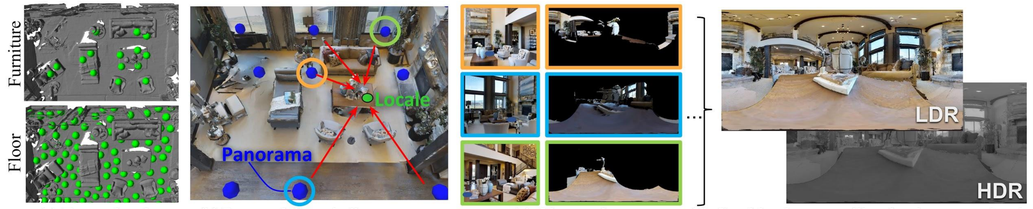
[2] M.-A. Gardner, K. Sunkavalli, E. Yumer, X. Shen, E. Gam-baretto, C. Gagné, and J.-F. Lalonde. Learning to predict indoor illumination from a single image. ACM Transactions on Graphics (TOG).

Method



Neural Illumination. Instead of using a single black-box network [2], we decompose the network into three differentiable sub-modules. Through this decomposition each sub-module is able to focus on a relatively easier task and can be trained with direct supervision. The entire network is differentiable and is trained with supervision end-to-end as well as for each intermediate sub-module

Training Data Generation



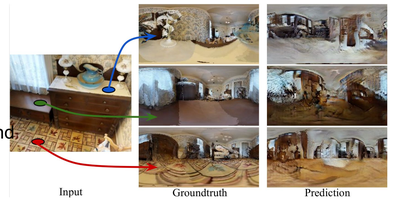
(a) Target locales (b) Resample"nearby" panoramas (c) Observations (d) Combined pano as illumination groundtruth

From these 3D reconstructions in Matterport3D dataset[1], we sample target locales (a) on supporting surfaces. For each locale, we use HDR and 3D geometric information from nearby camera to generate ground truth illumination maps.

Comparison of warping methods.

We use the scene geometry to generate geometrically accurate ground truth illumination, which accounts for complex 3D structures and more accurate than spherical warping.

Spatially varying illumination.



By using the geometry, we can generate ground truth illumination for any target locale. As a result, our model is also able to infer spatially varying illumination.

Results

Qualitative Results (Row 1) input image and selected locale. (Row 2,3) the warped observation using ground truth an predicted geometry. (Row 4,5) the completed LDR. (Row 6-10) show the final HDR illumination.

Evaluation

Method	$\ell 2(\log)$	$\ell 2$	diffuse
Gardner *et al.* [2]	0.375	0.977	1.706
Im2Im network	0.229	0.369	0.927
Nearest Neighbour	0.296	0.647	1.679
Ours	**0.202**	**0.280**	**0.772**

Comparon to alternative approach

loss	$\ell 2(\log)$	$\ell 2$	diffuse
l2	**0.116**	0.235	0.691
l2+gan	0.224	0.275	0.713
l2+gan+df	0.131	**0.212**	**0.619**

Effects of different losses.

Application: object re-lighting