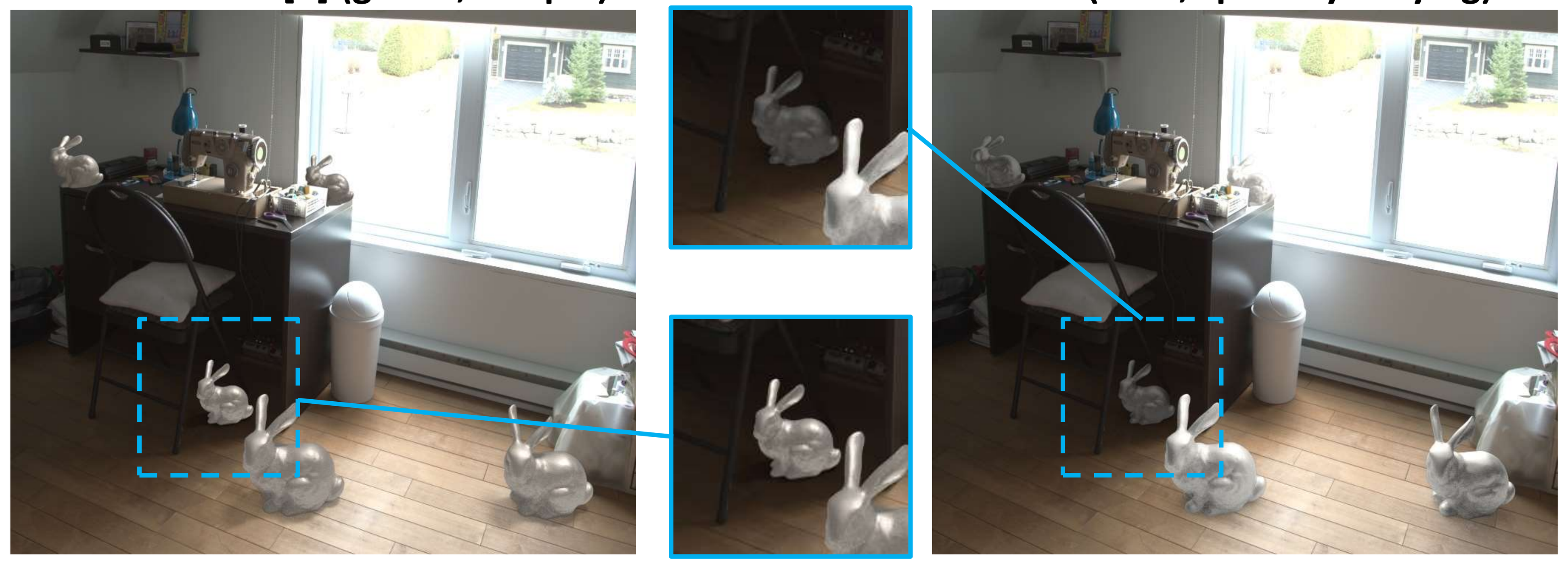
Fast Spatially-Varying Indoor Lighting Estimation

Mathieu Garon[1], Kalyan Sunkavalli[2], Sunil Hadap[2], Nathan Carr[2], Jean-François Lalonde[1]
[1]Université Laval, [2]Adobe Research
https://lvsn.github.io/fastindoorlight/

LONG BEACH CALIFORNIA
June 16-20, 2019

Context

Estimating the illumination conditions of a scene is a challenging problem. Indoor lighting is *spatially-varying*: light sources are in close proximity, and scene geometry affects local light conditions.

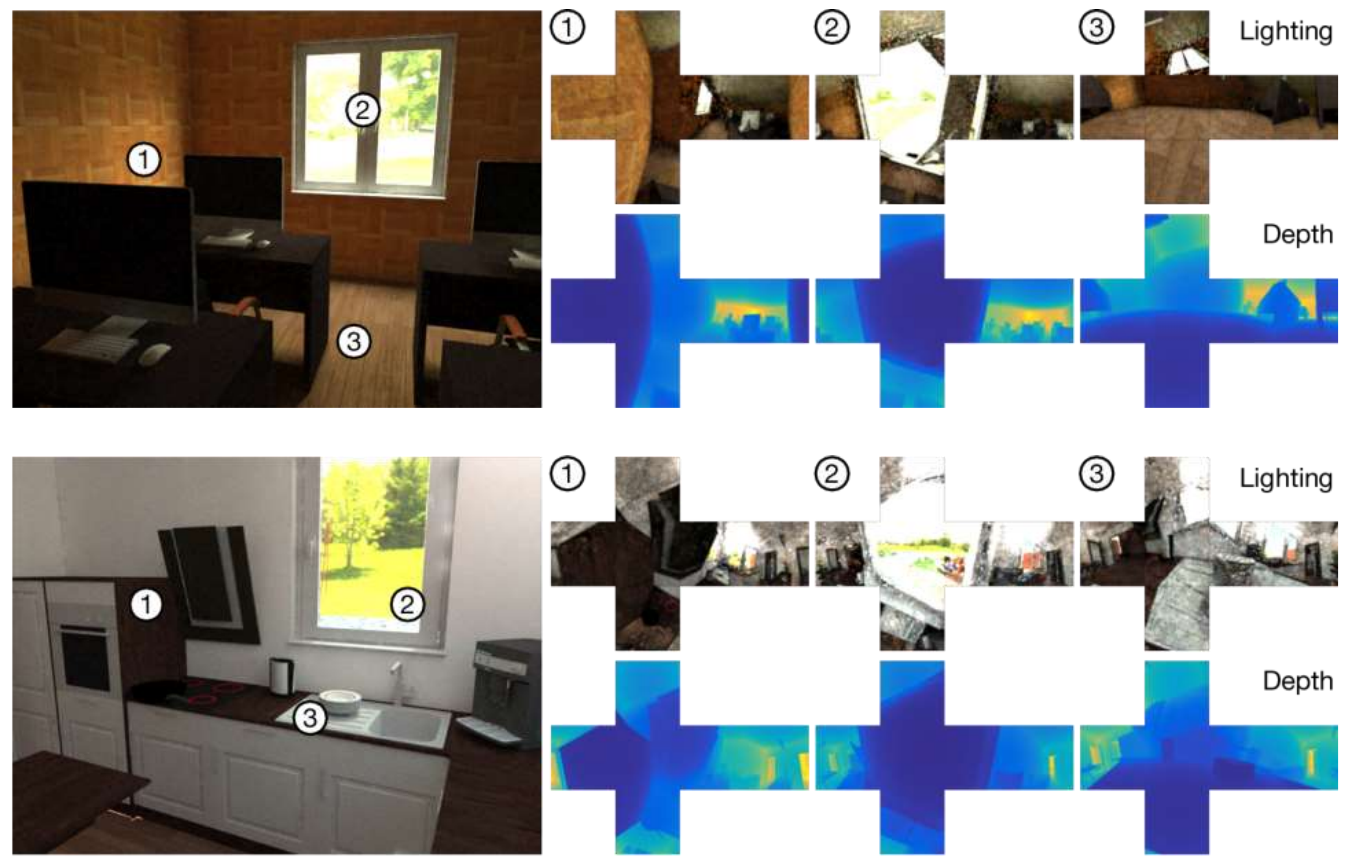


Typical limitations	Our approach
Need depth or full scene scan	Single RGB image
Estimate a single global light condition	Spatially-varying
Not real-time	Real-time: 20ms per image

Dataset

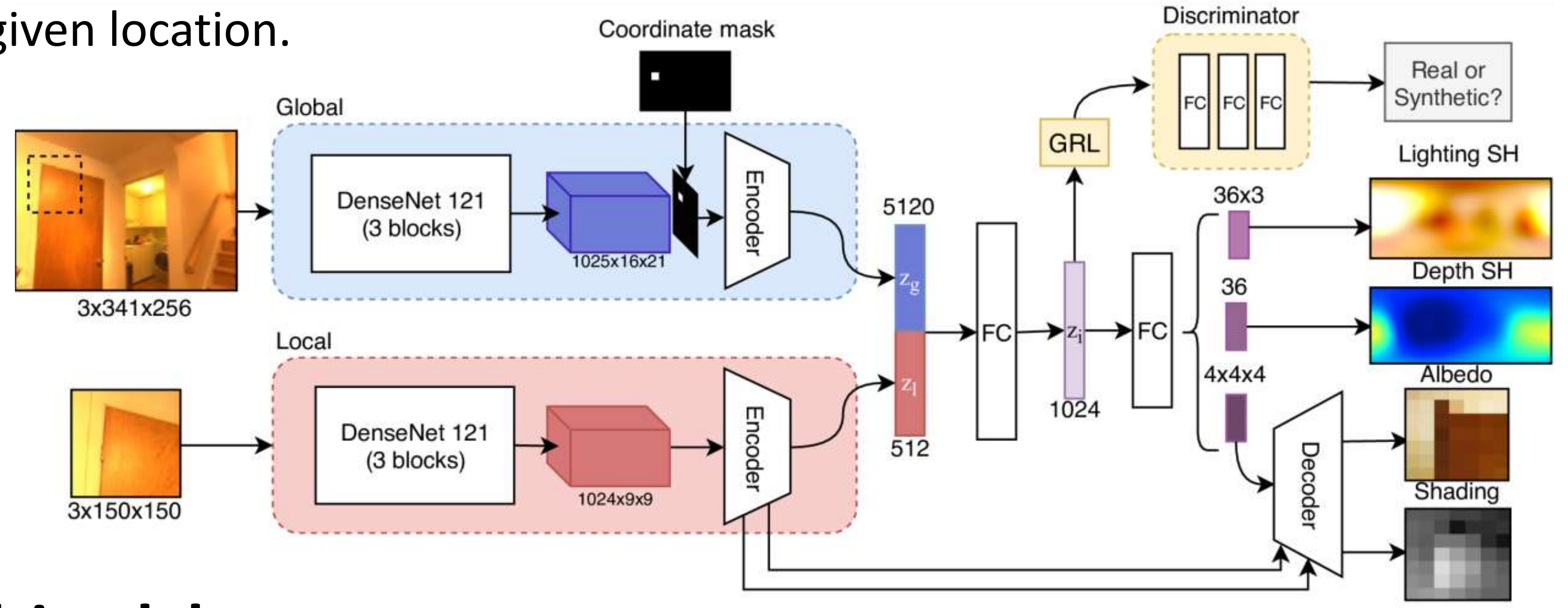
Physically-based renders from SUNCG [4].

- 26 800 viewpoints;
- Real outdoor HDR panoramas [1];
- Indoor light with various intensities;
- 4 cubemaps with light and depth information per scenes. (Only 3 shown here)

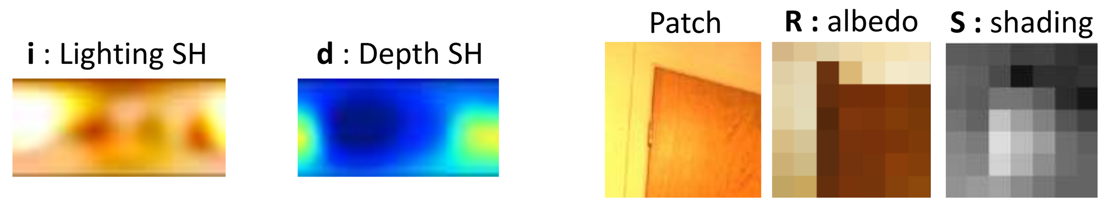


Approach

A neural network takes as input an RGB image with a pixel coordinate and outputs 5th-order Spherical Harmonics coefficients in RGB approximating the local lighting at the given location.



Multitask loss

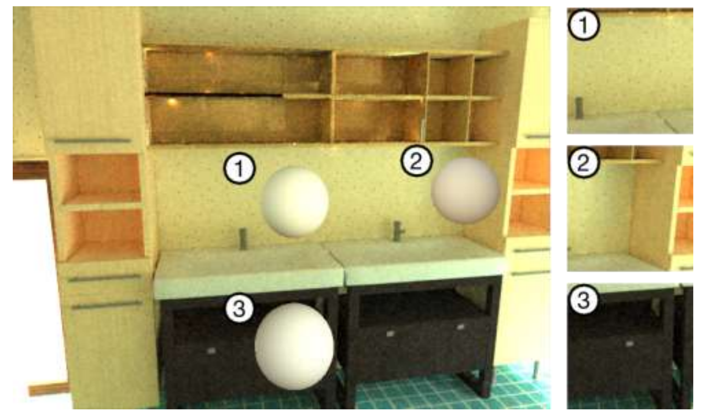


$$\mathcal{L}_{\text{i-sh}} = \frac{1}{36 \times 3} \sum_{c=1}^{3} \sum_{l=0}^{4} \sum_{m=-l}^{l} (i_{l,c}^{m*} - i_{l,c}^{m})^2$$

$$\mathcal{L}_{\text{d-sh}} = \frac{1}{36} \sum_{l=0}^{4} \sum_{m=-l}^{l} (d_l^{m*} - d_l^{m})^2$$

$$\mathcal{L}_{\text{rs-recons}} = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{P}_i^* - (\mathbf{R}_i + \mathbf{S}_i))^2$$

$$\mathcal{L}_{\text{rs-mse}} = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{R}_i^* - \mathbf{R}_i)^2 + (\mathbf{S}_i^* - \mathbf{S}_i)^2$$



Our network does not strictly rely on average patch brightness to estimate ambient lighting.

Ablation study

SH Degree	Global (w/o mask)	Global (w mask)	Local	Local + Global (w mask)
0	0.698	0.563	0.553	**0.520**
1	0.451	0.384	0.412	**0.379**
2–5	0.182	**0.158**	0.165	0.159

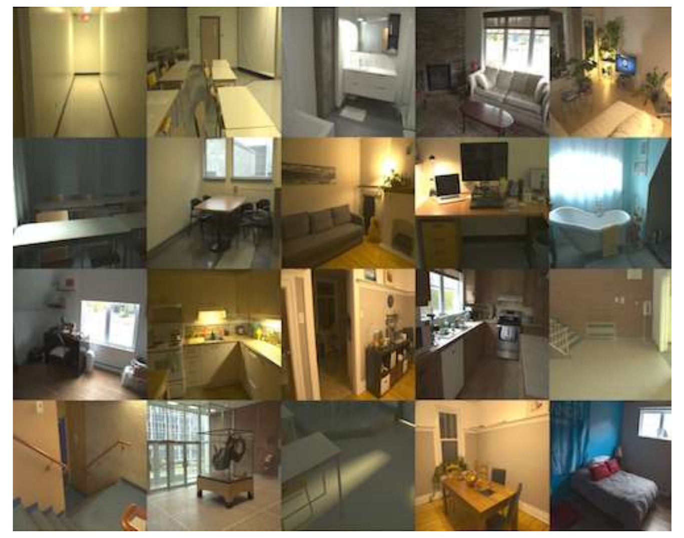
SH Degree	$\mathcal{L}_{\text{i-sh}}$	$+\mathcal{L}_{\text{d-sh}}$	$+\mathcal{L}_{\text{rs-mse}}$ $+\mathcal{L}_{\text{rs-recons}}$	All
0	0.520	0.511	0.472	**0.449**
1	0.379	0.341	0.372	**0.336**
2–5	0.159	0.149	0.166	**0.146**
Degree 1 angle	0.604	0.582	0.641	**0.541**

Qualitative comparison

Qualitative comparison to Barron and Malik [3] on the NYU-v2 dataset. Their method requires depth and takes 1.5h per image.

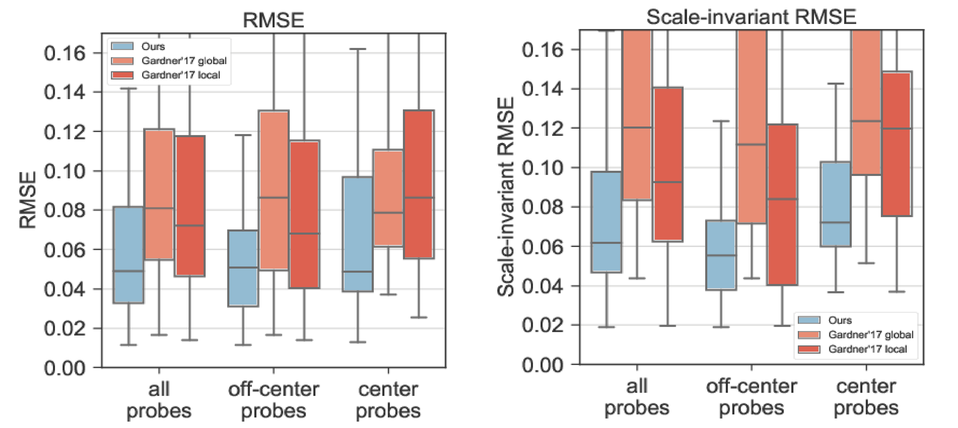
References:
[1] Y. Hold-Geoffroy, A. Athawale, and J.-F. Lalonde. Deep sky modeling for single image outdoor lighting estimation. CVPR, 2019
[2] M.-A. Gardner, K. Sunkavalli, E. Yumer, X. Shen, E. Gambaretto, C. Gagné, and J.-F. Lalonde. Learning to predict indoor illumination from a single image. SIGGRAPH Asia 2017
[3] J. T. Barron and J. Malik. Shape, illumination, and reflectance from shading. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015
[4] S. Song, F. Yu, A. Zeng, A. X. Chang, M. Savva, and T. Funkhouser. Semantic scene completion from a single depth image. CVPR, 2017

Dataset and evaluation



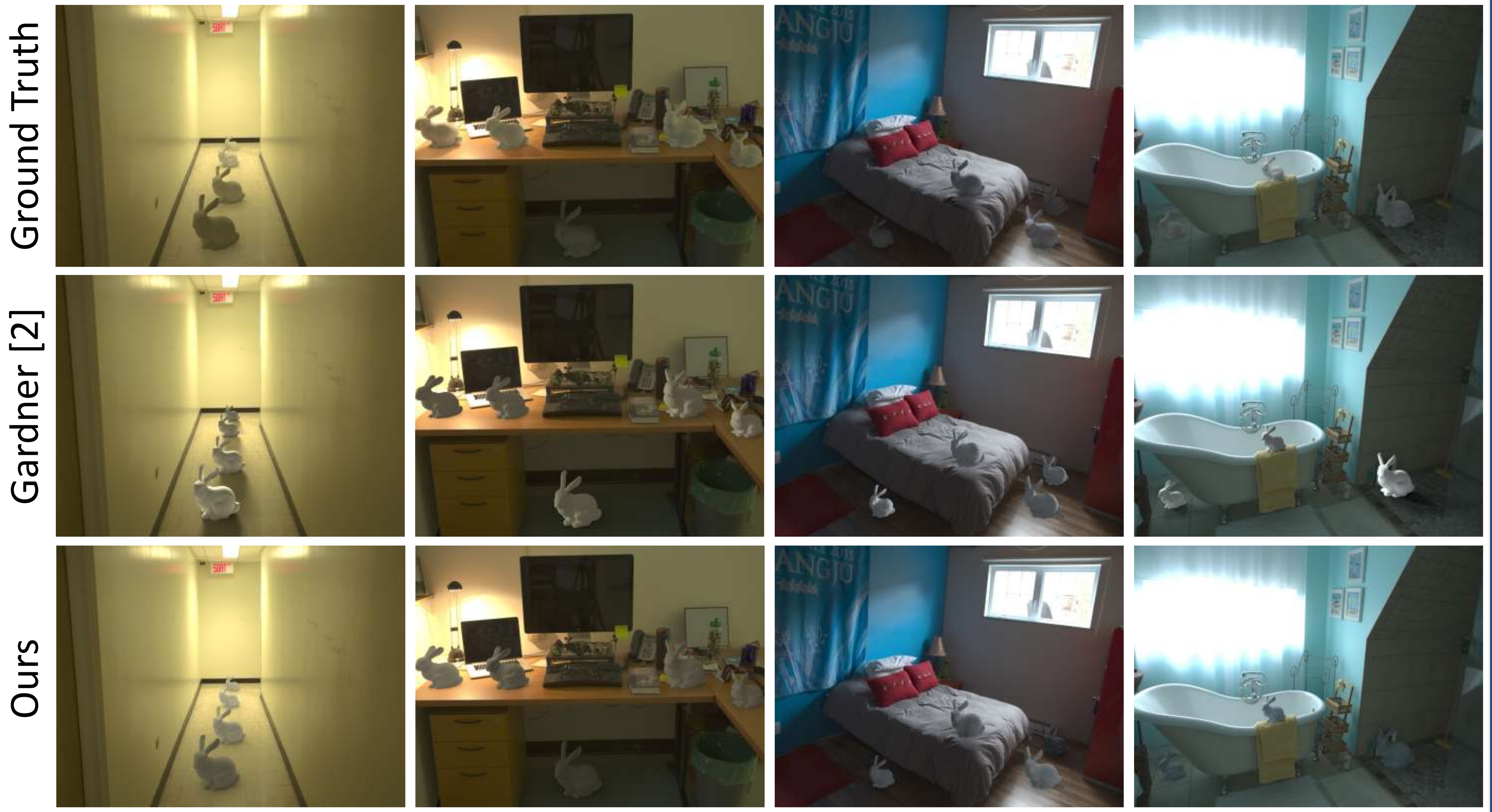
79 HDR Probes in 20 HDR scenes
Download at **indoorsv.hdrdb.com**

Render Comparison



User study

	All	Center	Off-center
Gardner et al. global	31.0%	**39.8%**	27.1%
Gardner et al. local	28.0%	25.2%	29.5%
Ours	**35.8%**	38.3%	**34.5%**



Augmented Reality

Advantages over popular AR frameworks:
- Lighting *direction* is estimated;
- Light source can be hidden;
- Adapts to local scene geometry such as occlusion and light source position.