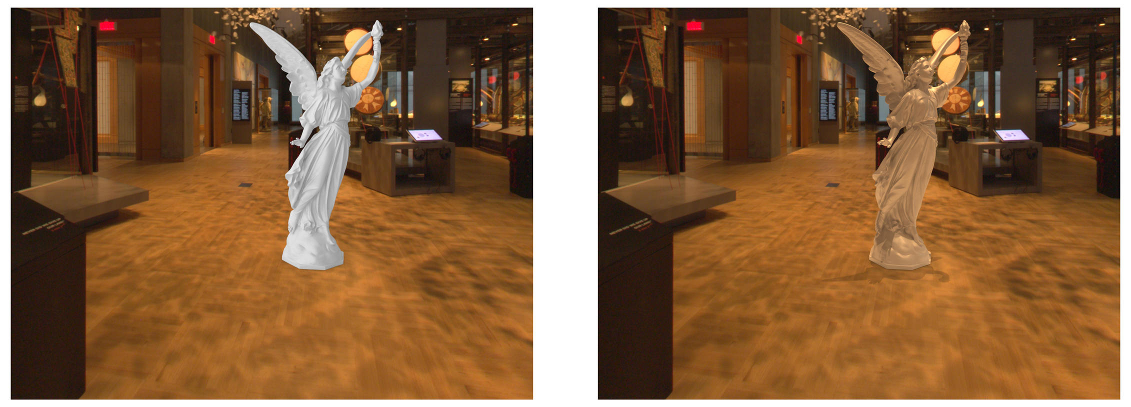
Deep Parametric Indoor Lighting Estimation

M-A. Gardner*, Y. Hold-Geoffroy†, K. Sunkavalli†, C. Gagné*, J.-F. Lalonde*
marc-andre.gardner.1@ulaval.ca, {christian.gagne, jean-francois.lalonde}@gel.ulaval.ca
*Université Laval, †Adobe Research

ICCV 2019 Seoul, Korea

Motivation

- Lighting is crucial for realistic rendering



Incorrect lighting | Correct lighting

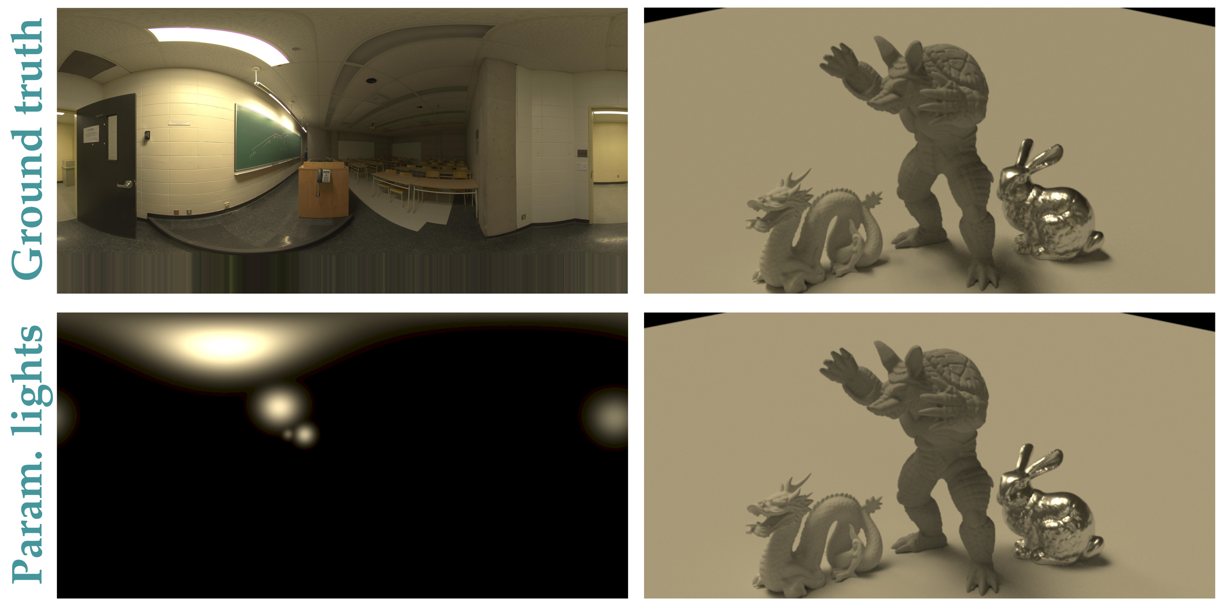
- We aim at recovering all **light sources** (in red) from a **single, limited FoV** image (in purple):

Spherical Gaussians

- We model each light as a **spherical gaussian**
- Each light source $\mathcal{S}_i$ is parametrized according to its position $\mathbf{l}_i$, size $\mathbf{s}_i$, and intensity/color $\mathbf{c}_i$:

$$f(\mathcal{P}, \mathbf{u}) = \sum_{i=1}^{N} \mathbf{c}_i \exp \frac{\mathbf{l}_i \cdot \mathbf{u} - 1}{\frac{1}{4\pi}\mathbf{s}_i} . \qquad (1)$$

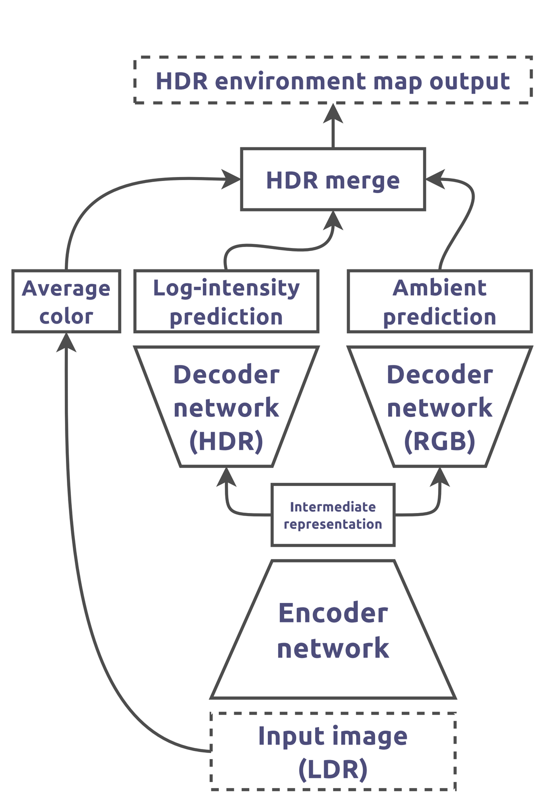
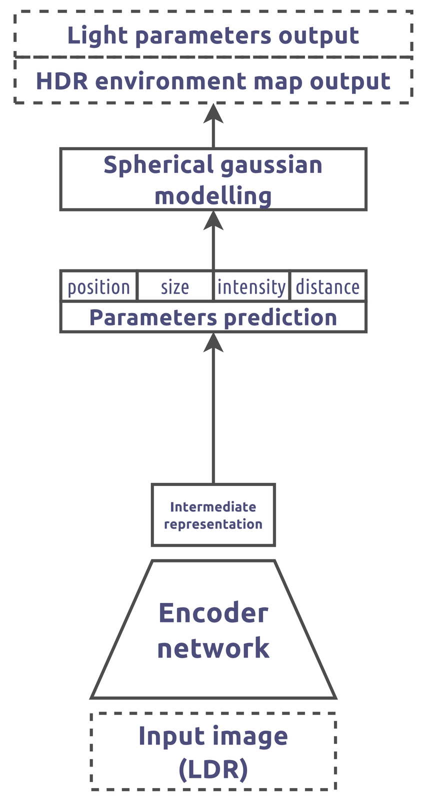
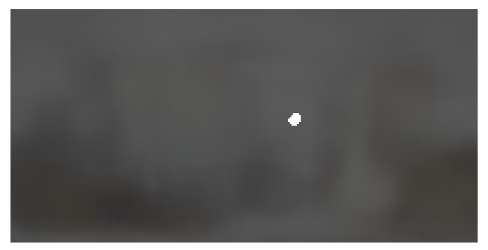
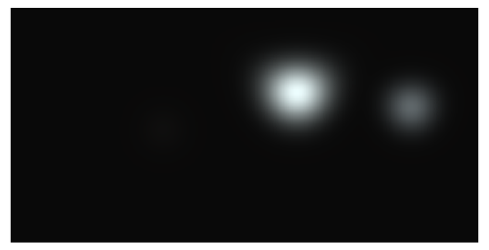
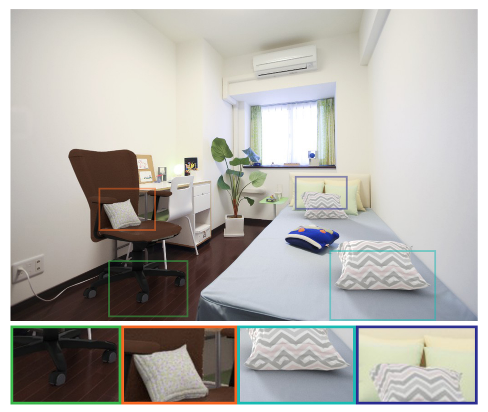
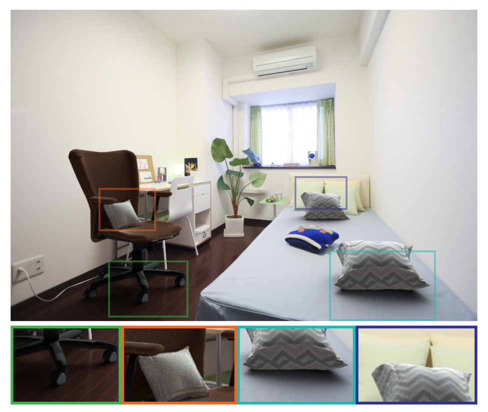
- Close to the actual panorama $\mathcal{P}$, energy-wise



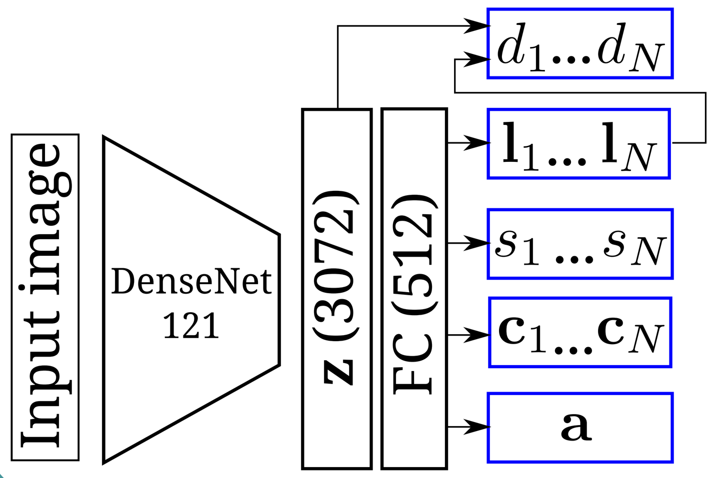
Environment map | Render

Why a parametric output?

- Previous work obtained good results by directly predicting a 360° panorama
- Why choose a parametric representation?

	Non-parametric [1]	Parametric (ours)
Learning and inference process		
Output	 32 768 pixels	 27 light parameters
Light locality	 Single, **global** lighting estimation	 **Local** lighting estimation
Complexity	30.2M network params 100 000 panoramas needed at training time 127 ms/image Non-editable	7.9M network params 2000 panoramas needed at training time 51 ms/image Intuitive to edit

Network architecture and training



Training step 1

- Parametric ground truth and predictions are projected onto panoramas
- L2 loss is used to compare prediction and ground truth

Training step 2

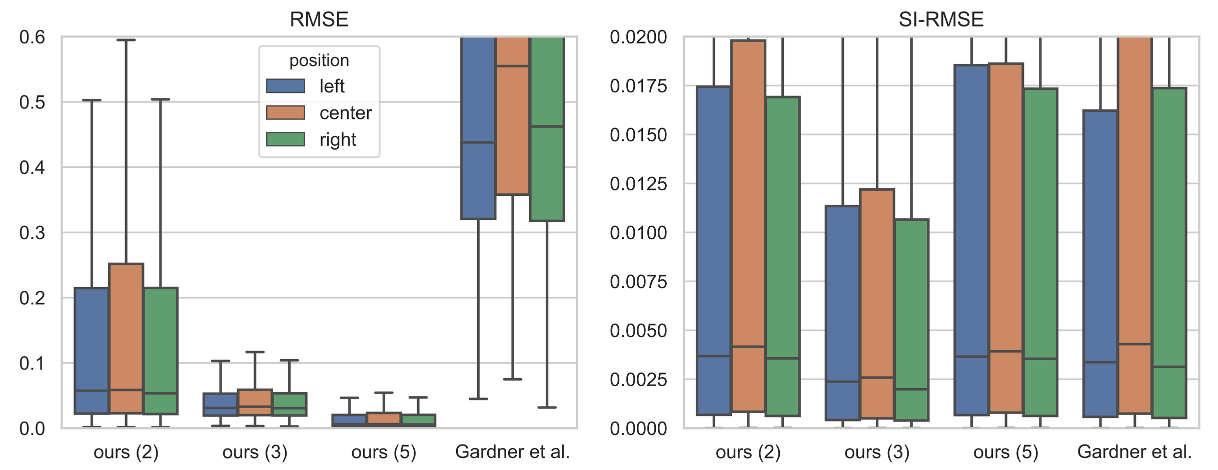
- Direct loss on the parameters
- Predicted positions $\mathbf{l}$ are used to *assign* predictions to the closest light
- Depth head estimation is trained

Results on artistic images (inset: original stock photo)

Real-world comparison

- We compare the appearance of a real and a virtual bunny in the same scene
- Can you spot which is which?

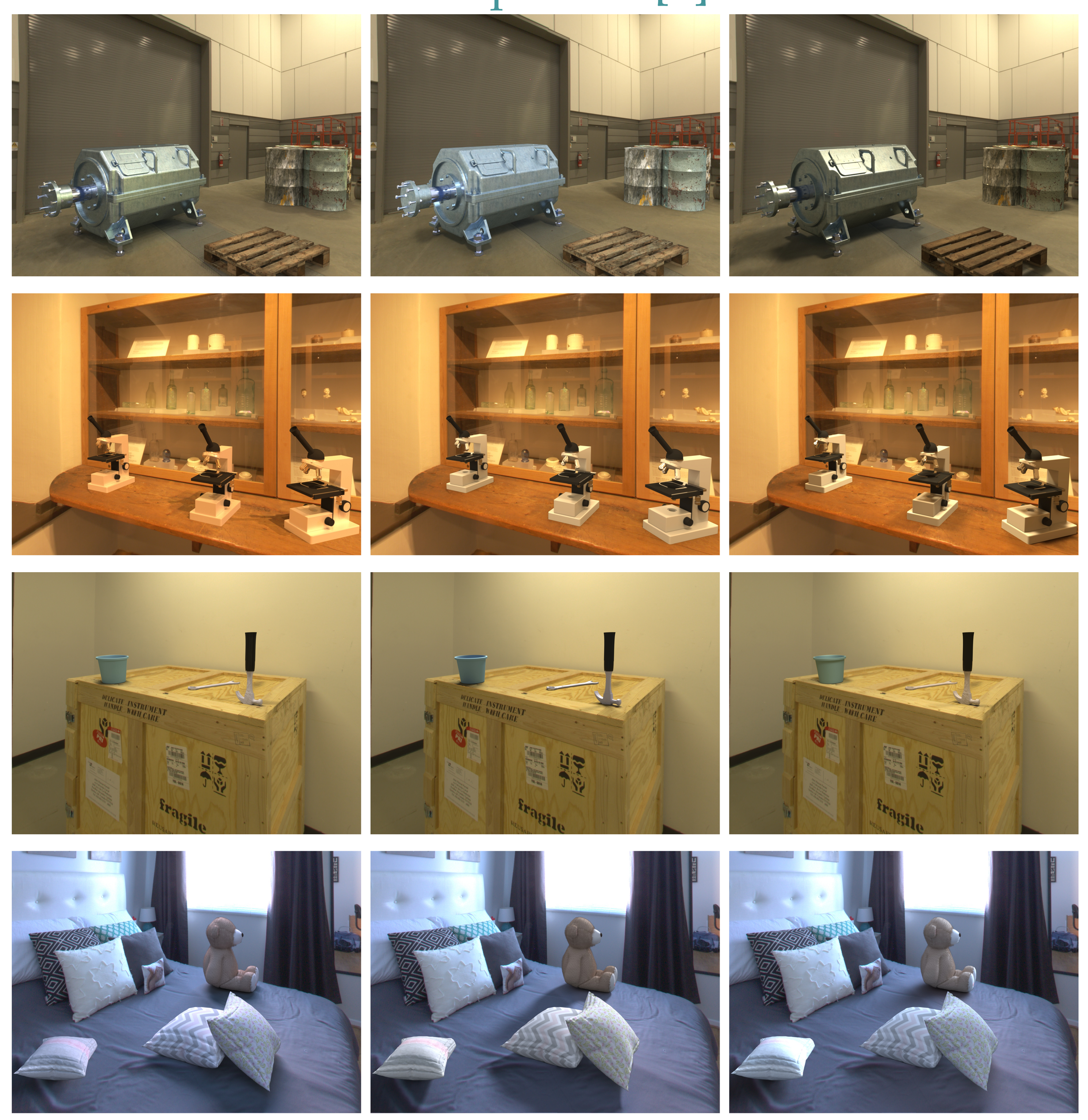
- Comparison with ground truth lighting and non-parametric approach



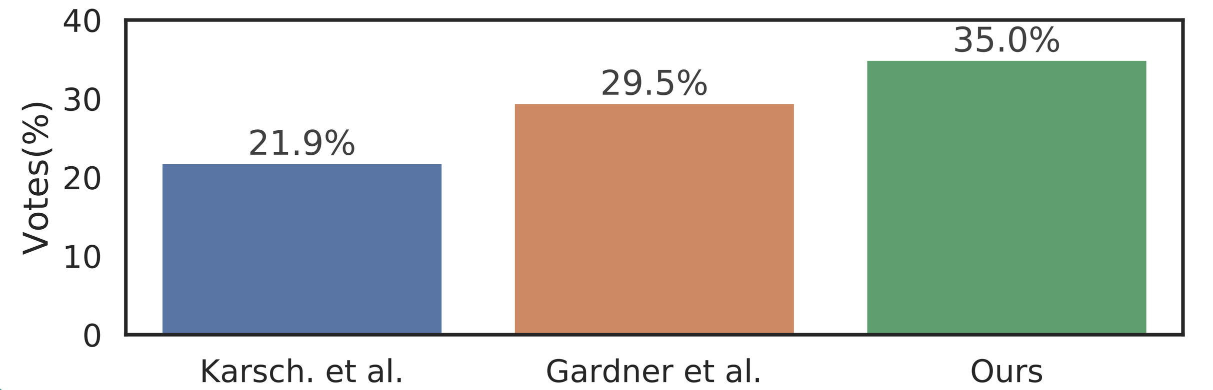
User study

- We compare to other methods with a user study
- 19 scenes and 49 participants
- **Task:** select the most *realistic* render

Ground truth | Non-param. [1] | Ours



- Percentage of time each method was preferred to the ground truth (perfect confusion=50%)



[1] Gardner, M.-A. *et al.*, Learning to Predict Indoor Illumination from a Single Image, ACM ToG, 2017.

This work was supported by FRQNT and NSERC grants. We thank Adobe for funding and nVidia for the donation of a Titan V. Project page: http://lvsn.github.io/deepparametric