Microsoft Research

YONSEI UNIVERSITY

Unified Depth Prediction and Intrinsic Image Decomposition from a Single Image via Joint Convolutional Neural Fields

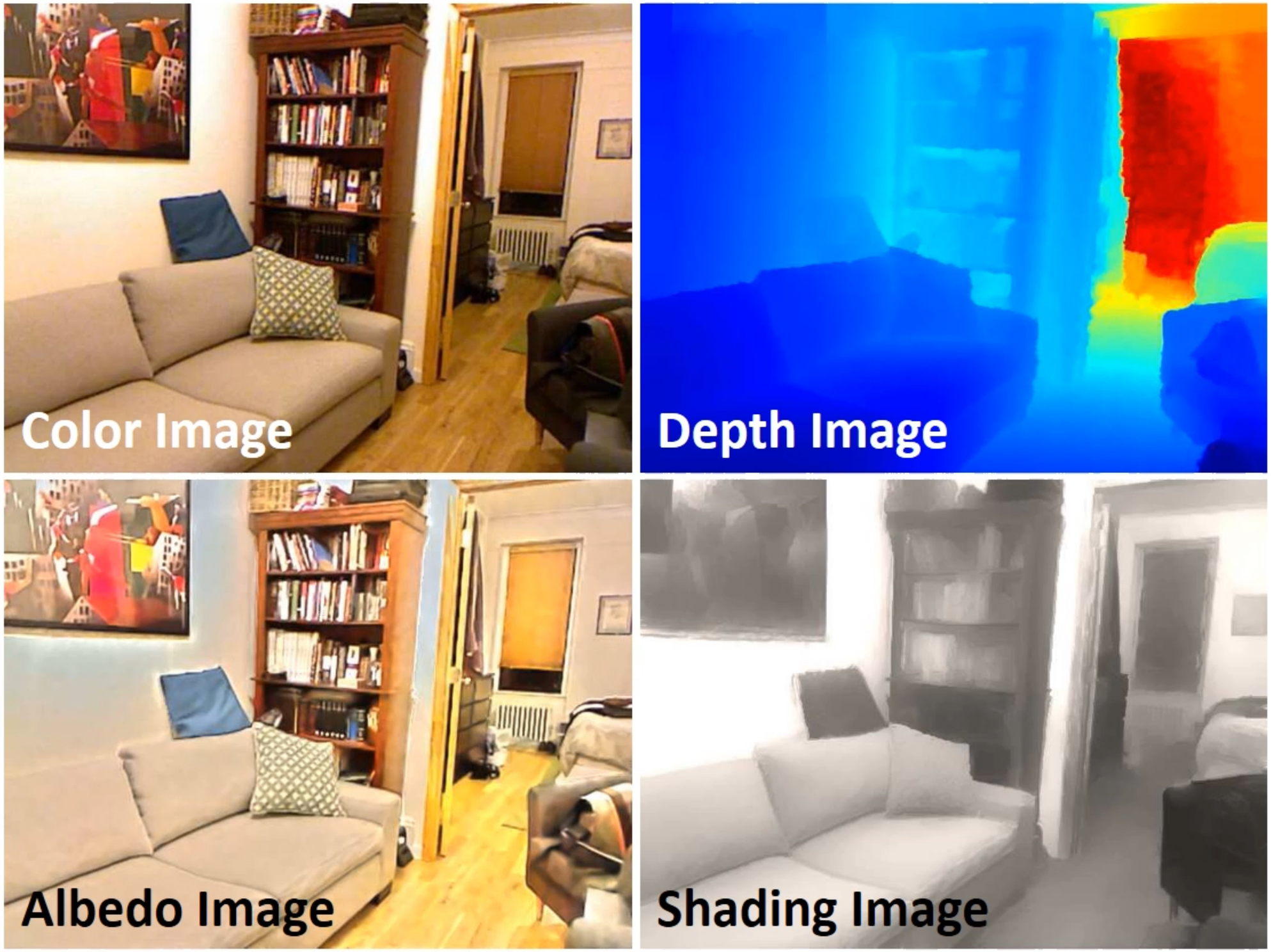
Seungryong Kim, Kihong Park, Stephen Lin,
and Kwanghoon Sohn

Department of Electrical and Electronic Engineering
Yonsei University

Oct. 13, 2016

Motivation

- **Deep Structure Understanding from Single Image**



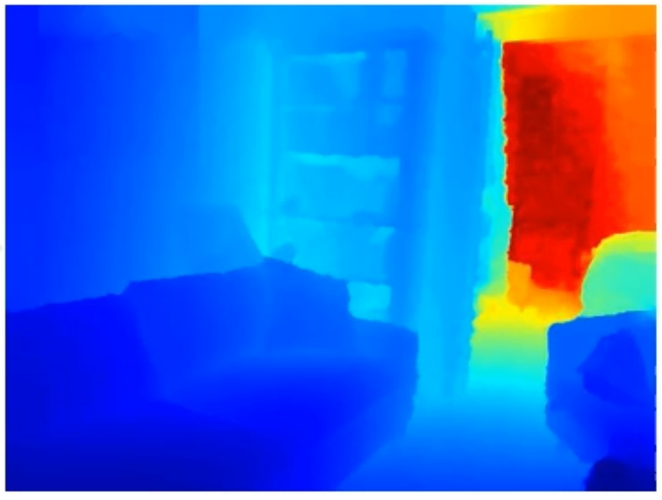
Challenges

- **Depth Prediction Problem**

RGB

Why Difficult?

- ✓ No explicit depth cue
- ✓ Depth ambiguity
- ✓ No appearance-depth mapping assumption



Depth

- **Intrinsic Image Decomposition Problem**

RGB

Why Difficult?

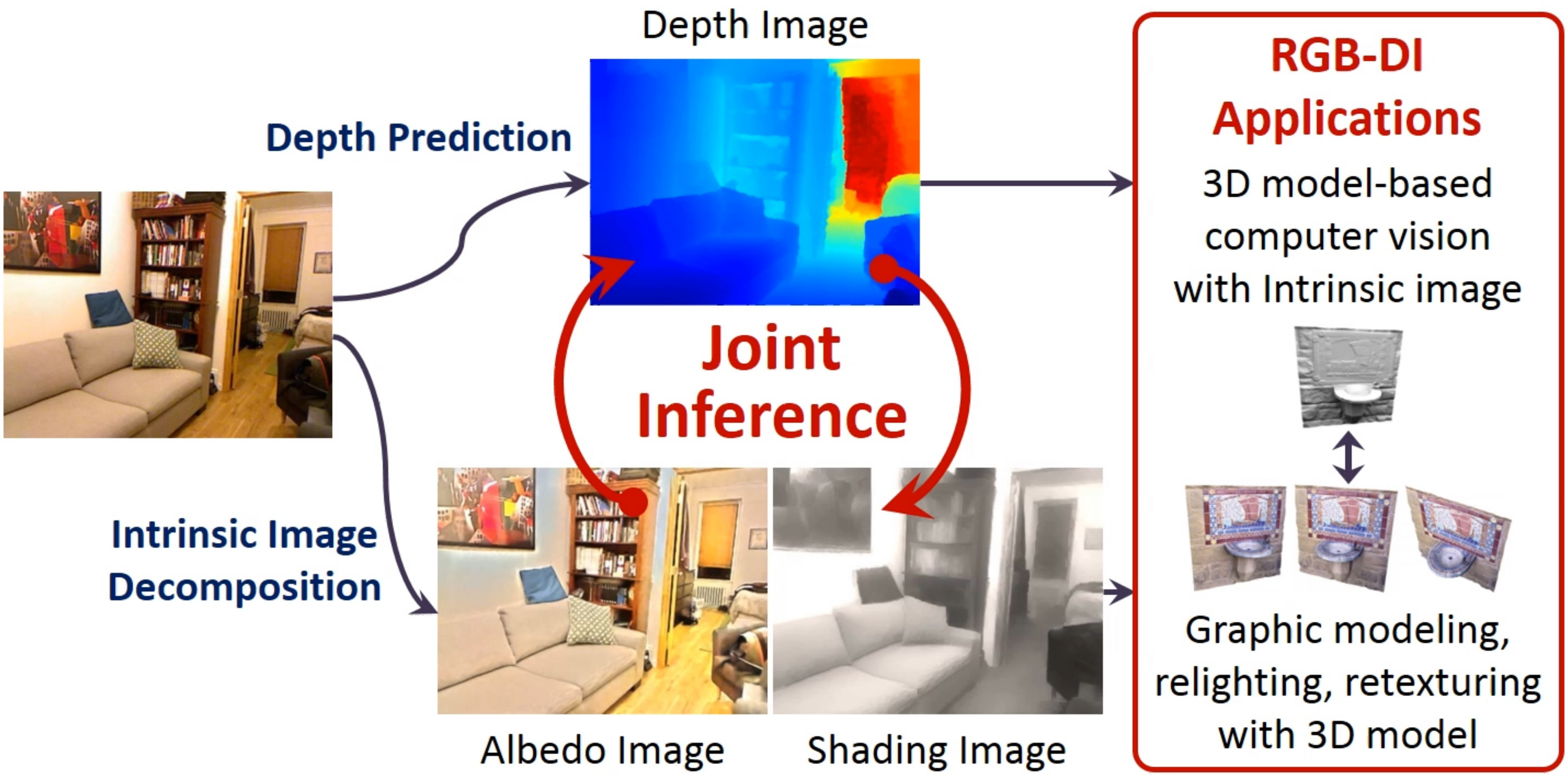
- ✓ No prior
- ✓ Retinex model
- ✓ Occlusion
- ✓ Shadows

Albedo

Shading

Our Solution

- **Joint Depth and Intrinsic Image Inference**



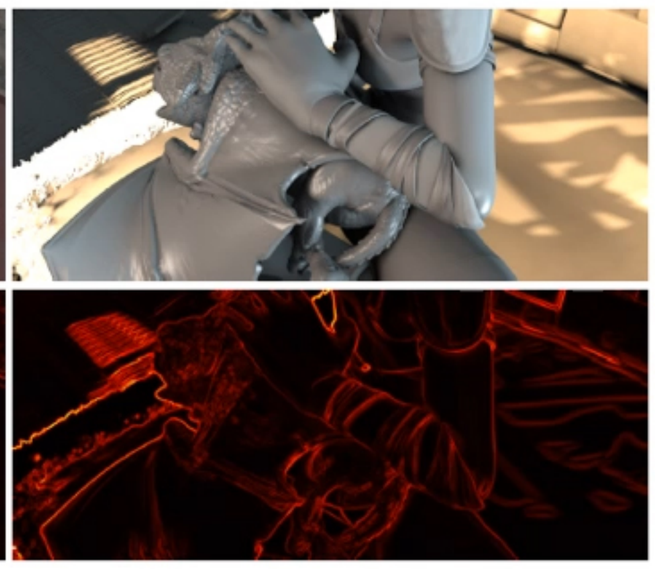
Formulation

- **Joint Convolutional Neural Field (JCNF) Model**
  - **Key-Insights**: Compared to quantities in the value domain, correlations are stronger among gradient domain

Color image
Its gradients

Depth
Its gradients

Albedo
Its gradients



Shading
Its gradients

Depth and Intrinsic Image Inference

- **Joint Convolutional Neural Field (JCNF) Model**
  - **Joint Conditional Random Field (CRF) Energy Function**

$$\mathbf{E}(D, A, S|I) = \mathbf{E}_u(D|I) + \mathbf{E}_u(A, S|I) + \lambda_D \mathbf{E}_s(D|I, A, S) + \lambda_A \mathbf{E}_s(A|I, D, S) + \lambda_S \mathbf{E}_s(S|I, D, A),$$

  - **Unary Potentials**

$$\mathbf{E}_u(D|I) = \sum_p \left(D_p - \boxed{\mathcal{F}(I_{\mathcal{P}}; \mathbf{w}_{\mathcal{F}}^D)}\right)^2 \rightarrow \text{Depth Value}$$

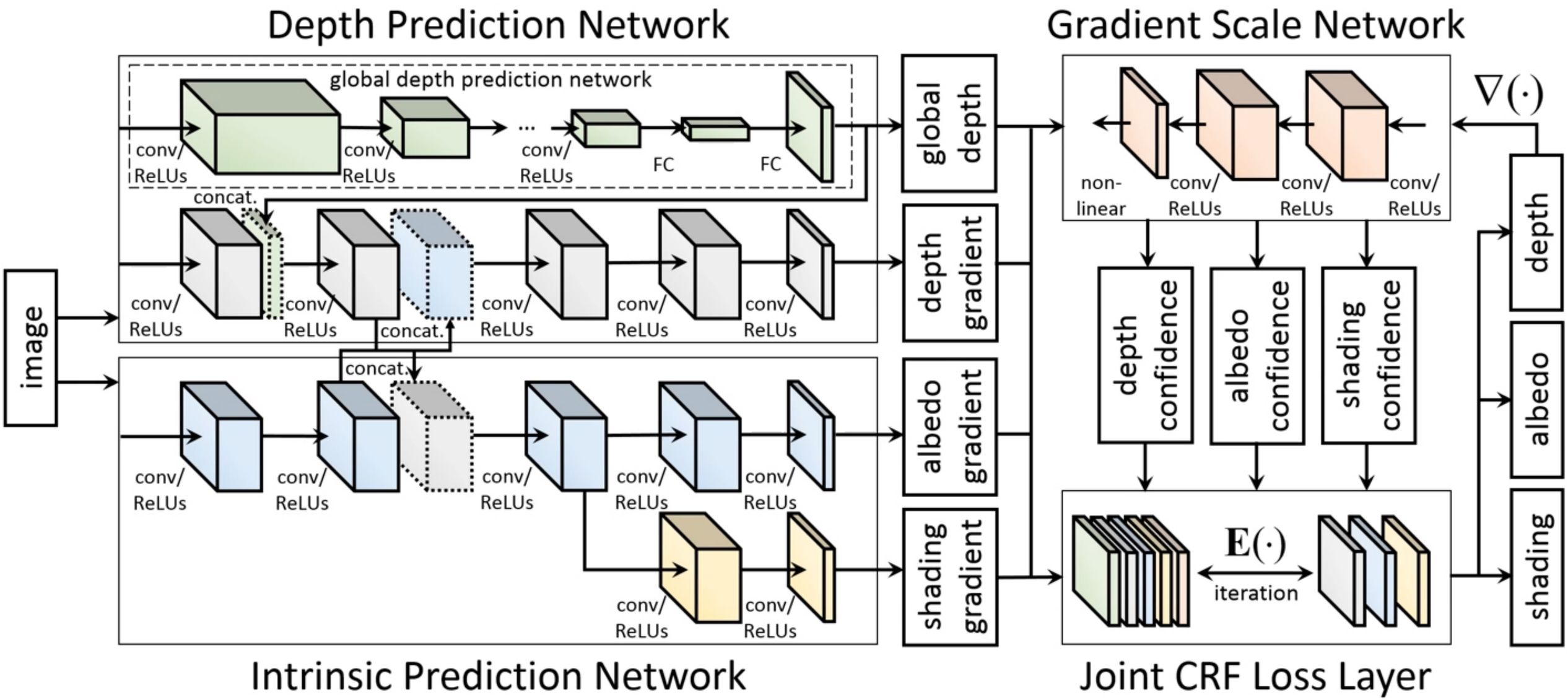
$$\mathbf{E}_u(A, S|I) = \sum_p \left(L_p(I_p - A_p - S_p)\right)^2$$

  - **Pairwise Potentials**

$$\mathbf{E}_s(D|I, A, S) = \sum_p \|\nabla D_p - \underbrace{\boxed{\mathcal{G}(\nabla I_{\mathcal{P}}, \nabla A_{\mathcal{P}}, \nabla S_{\mathcal{P}}; \mathbf{w}_{\mathcal{G}}^{\nabla D})}}_{\text{Confidence of Gradient}} \circ \underbrace{\boxed{\mathcal{F}(I_{\mathcal{P}}; \mathbf{w}_{\mathcal{F}}^{\nabla D})}}_{\text{Depth Gradient}}\|^2$$

Depth and Intrinsic Image Inference

- **Joint Convolutional Neural Field (JCNF) Model**
  - **Network Architecture**



Depth and Intrinsic Image Inference

- **Joint Convolutional Neural Field (JCNF) Model**
  - **Training Procedure**
    - Global Depth Prediction

$$\mathcal{L}(\mathbf{w}_{\mathcal{F}}^{D}) = \sum\nolimits_{\{i,p\}} \left(D_p^i - \mathcal{F}(I_{\mathcal{P}}^i; \mathbf{w}_{\mathcal{F}}^{D})\right)^2$$

    - Gradient Prediction + Gradient Scale Prediction

$$\mathcal{L}(\mathbf{w}_{\mathcal{G}}^{\nabla D}, \mathbf{w}_{\mathcal{F}}^{\nabla D}) = \sum\nolimits_{\{i,p\}} \|\nabla D_p^i - \mathcal{G}(\nabla I_{\mathcal{P}}^i, \nabla A_{\mathcal{P}}^i, \nabla S_{\mathcal{P}}^i; \mathbf{w}_{\mathcal{G}}^{\nabla D}) \circ \mathcal{F}(I_{\mathcal{P}}^i; \mathbf{w}_{\mathcal{F}}^{\nabla D})\|^2$$

Iterative Training

  - **Testing Procedure**
    - Iterative Joint Prediction

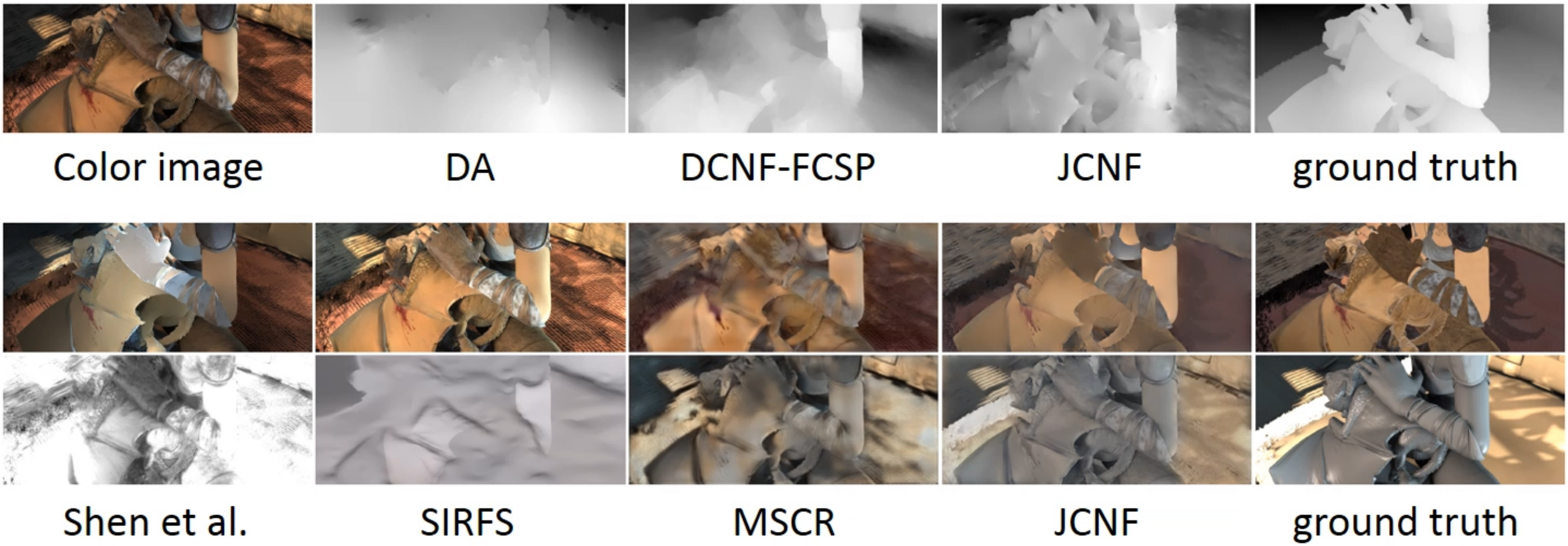
$$\mathbf{E}(D|I) = \sum\nolimits_p \left(D_p - D_p^*\right)^2 + \lambda_D \sum\nolimits_p \|\nabla D_p - C(\nabla D_p^*) \circ \nabla D_p^*\|^2$$

$$\mathbf{E}(A, S|I) = \sum\nolimits_p \left(L_p(I_p - A_p - S_p)\right)^2$$
$$+ \sum\nolimits_p \lambda_A \|\nabla A_p - C(\nabla A_p^*) \circ \nabla A_p^*\|^2 + \lambda_S \|\nabla S_p - C(\nabla S_p^*) \circ \nabla S_p^*\|^2$$

Joint Inference

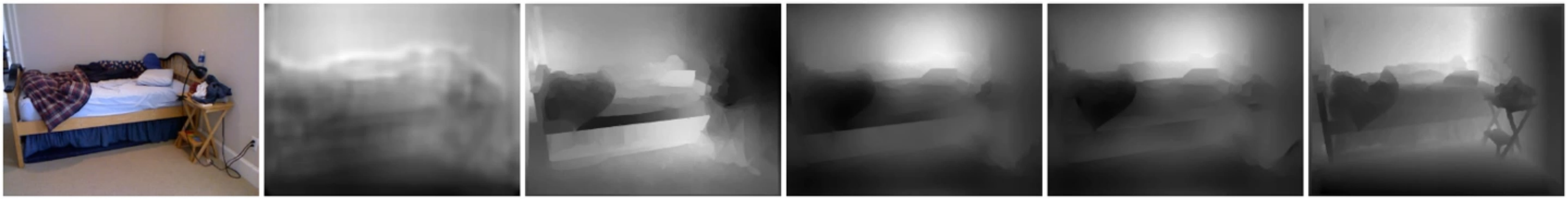
Experimental Results

- **MPI SINTEL Benchmark**

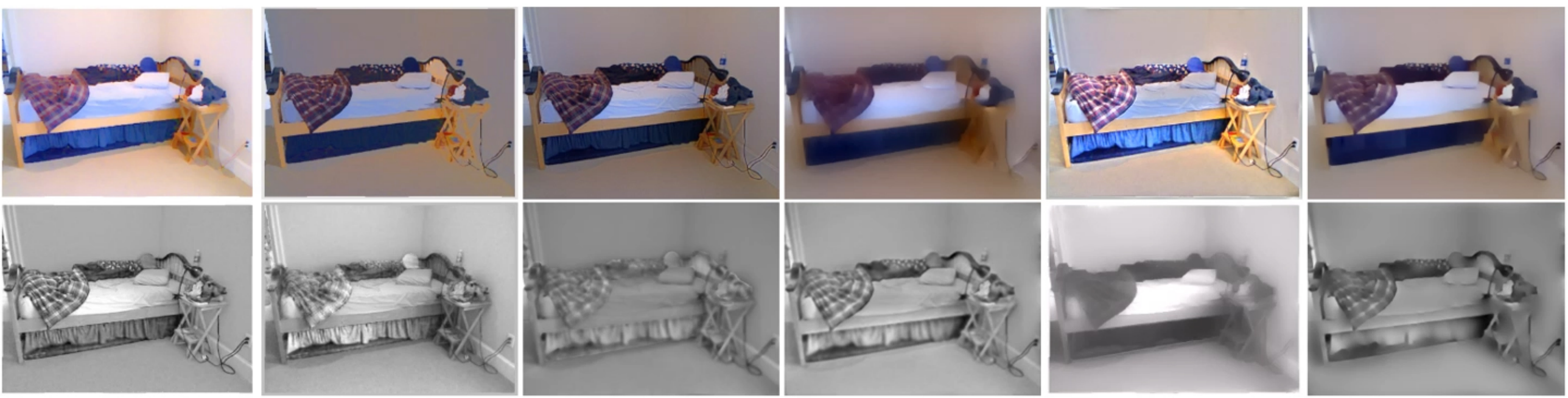


Experimental Results

- **NYU v2 RGB-D Benchmark**



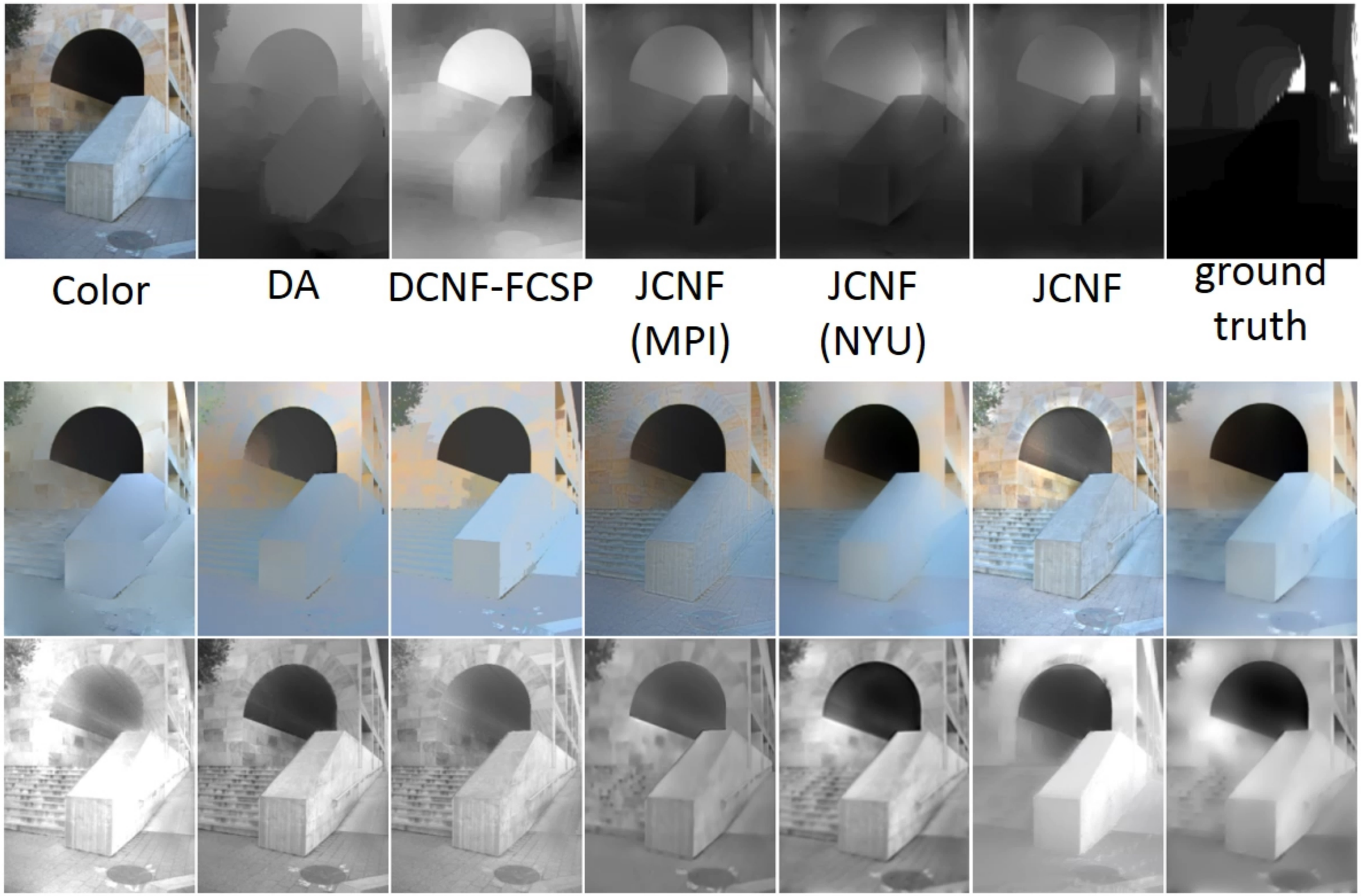
Color image | MS-CNNs | DCNF-FCSP | JCNF(MPI) | JCNF | ground truth



Li et al. | IIW | Jeon et al. | JCNF with Jeon et al. | Chen et al. | JCNF with Chen et al.

Experimental Results

- **Make3D RGB-D Benchmark**



Thank you !

More Details in

My Poster Booth (S-3C-07)

Seungryong Kim, Ph.D. Candidate
Digital Image Media Lab.
Yonsei University, Seoul, Korea
Tel: +82-2-2123-2879
E-mail: srkim89@yonsei.ac.kr
Homepage: http://diml.yonsei.ac.kr/~srkim/