NIID-Net: Adapting Surface Normal Knowledge for Intrinsic Image Decomposition in Indoor Scenes – Supplementary Material

Jundan Luo, Zhaoyang Huang, Yijin Li, Xiaowei Zhou, Guofeng Zhang, and Hujun Bao
State Key Lab of CAD&CG, Zhejiang University
ZJU-SenseTime Joint Lab of 3D Vision

Figure 1: An edited image sequence. The original image sequence is picked from the BIGTIME [3] dataset. The images are captured under different lighting conditions. The inserted virtual posters are rendered with estimated colorful shading.

In this supplementary document, we first explain the energy function for estimating colorful shading and present an application of image sequence editing. Then we present hyperparameters used in each loss term. Finally, we present more visual comparisons of intrinsic image decomposition on the IIW/SAW test sets.

1 Estimating Colorful Shading

Our proposed NIID-Net predicts a colorful reflectance image $\widehat{\mathbf{R}}$ and a gray-scale shading intensity image $\widehat{S}$ from an input image $\mathbf{I}$. The color of lighting is required by realistic image editing. Therefore, we propose to recover the global color of shading and refine network predictions by minimizing:

$$\begin{cases} \mathscr{E} = \sum_i w_{e1} \left\| \widetilde{\mathbb{R}}_i - \widehat{\mathbb{R}}_i \right\|_2 + w_{e2} \left\| \widetilde{R}_i - \widehat{R}_i \right\|_2 \\ \qquad + w_{e3} \left\| \nabla \widetilde{S}_i - \nabla \widehat{S}_i \right\|_2 + w_{e4} \left\| \mathbf{I}_i - \widetilde{\mathbf{R}}_i \times \widetilde{\mathbf{S}}_i \right\|_2, \\ \widetilde{\mathbf{R}}_i = \widetilde{\mathbb{R}}_i \times \widetilde{R}_i, \\ \widetilde{\mathbf{S}}_i = \widetilde{\alpha} \times \widetilde{\mathbf{c}} \times \widetilde{S}_i, \end{cases} \tag{1}$$

where $\widehat{}$ indicates that the images are predicted by the NIID-Net, and $\widetilde{}$ indicates the variables to be optimized. The refectance image is furtherly decomposed into the chromaticity $\widehat{\mathbb{R}}$ and the intensity $\widehat{R}$ by the following Equation:

$$\begin{cases} \widehat{R}_i = (\widehat{R}_i^{red} + \widehat{R}_i^{green} + \widehat{R}_i^{blue})/3, \\ \widehat{\mathbb{R}}_i = \widehat{\mathbf{R}}_i / \widehat{R}_i. \end{cases} \tag{2}$$

These two components of reflectance are optimized separately. We optimize shading intensity in the gradient domain to avoid extreme shading variations. $\widetilde{\alpha}$ is a scalar. $\widetilde{\mathbf{c}}$ is the global chromaticity of colorful shading. $w_{e1} = 1.0$, $w_{e2} = 0.5$, $w_{e3} = 5.0$, and $w_{e4} = 9.0$ are the weights of the terms. An optimized result is shown in Fig. 2.

A demonstration of photorealistic editing of an image sequence is presented in Fig. 1. This application can be used in augmented reality for advertising and scene refurnishing.

Note that the optimization (*i.e.*, Equation 1) is only employed for image sequence editing (Fig. 1 and the supplementary video). The shading and reflectance images presented in comparisons (including Sect. 3) are prediction results from the NIID-Net.

2 Training Hyperparameters

The detailed hyperparameters in loss functions are shown in Table 1.

Table 1: Hyperparameters in loss functions.

Loss term	Hyperparameters
$\mathscr{L}_{shading}$	$w_{s1} = 1.0, w_{s2} = 9.0, w_{s3} = 3.0$
$\mathscr{L}_{smooth\_A}$	$w_{a1} = 0.1, \delta_N = 0.15$
$\mathscr{L}_{reflect}$	$w_{r1} = 1.0, w_{r2} = 9.0, w_{r3} = 9.0$
$\mathscr{L}_{consist\_R}$	$\delta_R = 0.6$

3 Comparisons

We present more visual comparisons between our NIID-Net and previous methods [1, 2] on the IIW/SAW test sets in Fig. 3, Fig. 4 and Fig. 5. The odd rows present predicted reflectance images, and the even rows are predicted shading images, except that the first column shows input images. We show results of Bi *et al.* [1], Li and Snavely [2] (trained on CGI+IIW+SAW), and our NIID-Net in each row in order. Our shading images have the most realistic visual effects.

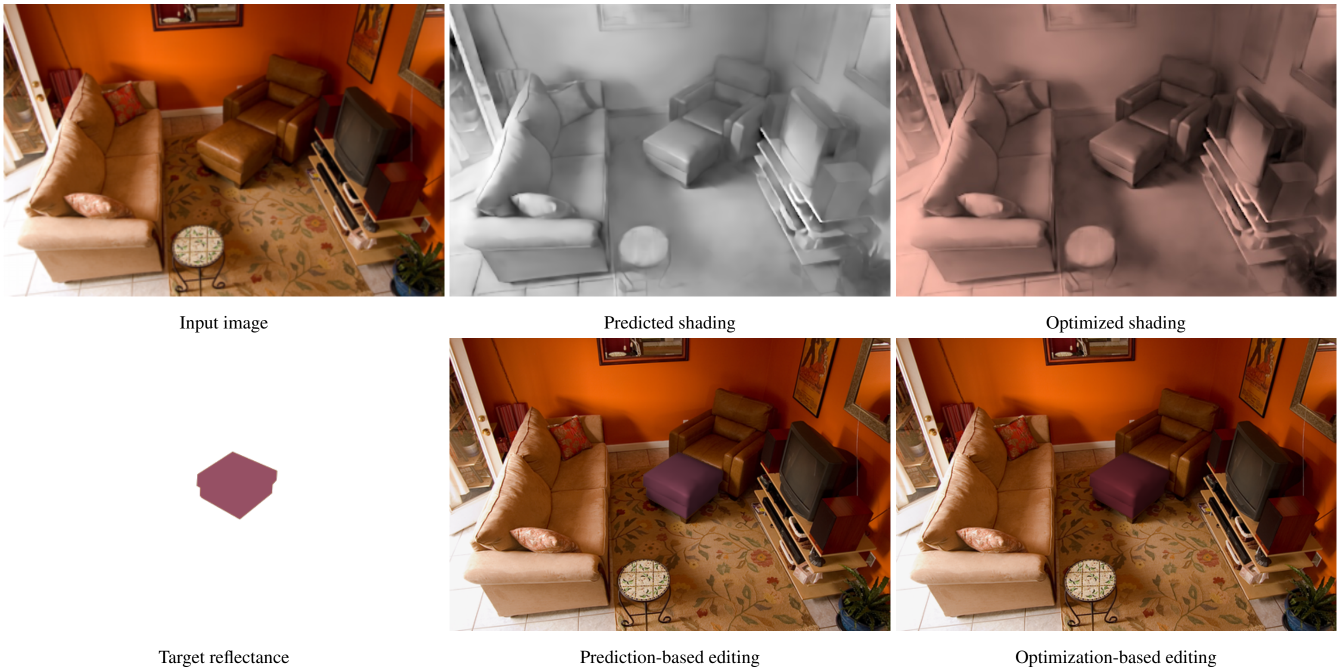

Figure 2: Image editing results based on predicted and optimized shading. We estimate the color of shading for realistic image editing. The edited image based on optimized shading is more realistic than that based on the purely predicted shading.

References

[1] S. Bi, X. Han, and Y. Yu. An L_1 image transform for edge-preserving smoothing and scene-level intrinsic decomposition. *ACM Transactions on Graphics*, 34(4):78, 2015.

[2] Z. Li and N. Snavely. CGIntrinsics: Better intrinsic image decomposition through physically-based rendering. In *Proceedings of the European Conference on Computer Vision*, pp. 381–399, 2018.

[3] Z. Li and N. Snavely. Learning intrinsic image decomposition from watching the world. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9039–9048, 2018.

Figure 3: Visual comparisons on the IIW/SAW test sets.

Input image | Bi *et al.* [1] | Li and Snavely [2] | Ours

Figure 4: Visual comparisons on the IIW/SAW test sets.

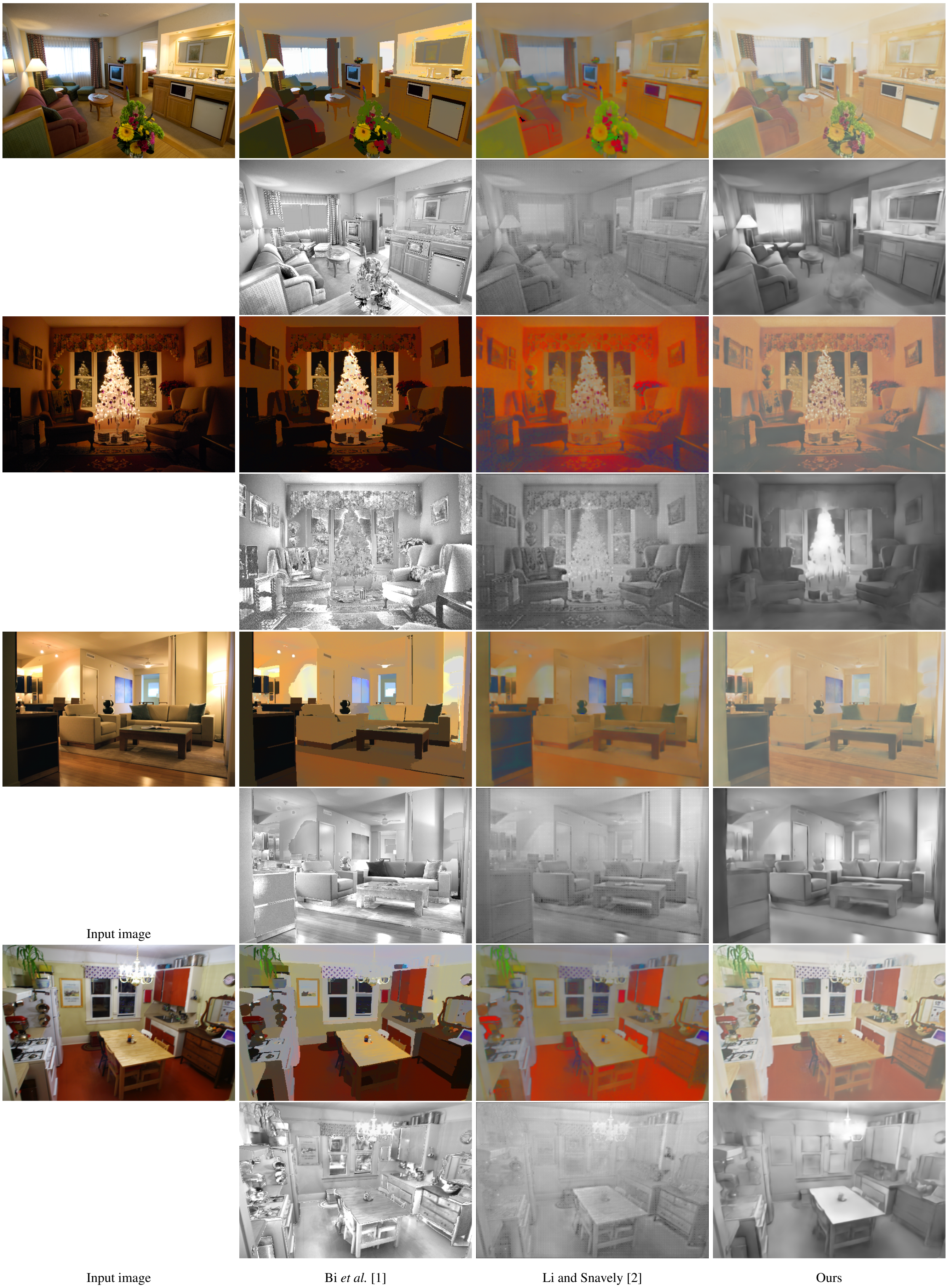

Figure 5: Visual comparisons on the IIW/SAW test sets.