

Supplementary Material: CGIntrinsics: Better Intrinsic Image Decomposition through Physically-Based Rendering

Zhengqi Li and Noah Snavely

Department of Computer Science & Cornell Tech, Cornell University

In this supplementary material, we first provide additional details about the ordinal loss we use when training with our CGINTRINSICS dataset (Equation 2 in the main paper), as well as detailed hyperparameter settings used in training. Next, we show additional visual comparisons between the images in our CGINTRINSICS dataset and the original SUNCG/PBRS dataset [1]. Finally, we provide additional qualitative prediction results and comparisons to Bell *et al.* [2] and Zhou *et al.* [3] on the IIW and SAW test sets.

1 Additional details for training losses

1.1 Ordinal term for CGINTRINSICS

Recall \mathcal{L}_{CGI} (Equation 2 in the main paper), the loss defined for our CGINTRINSICS training data:

$$\mathcal{L}_{CGI} = \mathcal{L}_{sup} + \lambda_{ord}\mathcal{L}_{ord} + \lambda_{rec}\mathcal{L}_{reconstruct} \quad (1)$$

We now provide the full formula for \mathcal{L}_{ord} , the ordinal loss term. In particular, for a given CGINTRINSICS training image and predicted reflectance R , we accumulate losses for each pair of pixels (i, j) generated from a set of pixels \mathcal{P} , where one pixel is sampled at random from oversegmented regions in that image:

$$\mathcal{L}_{ord}(R) = \sum_{\substack{(i,j) \in \mathcal{P} \times \mathcal{P} \\ i \neq j}} f_{i,j}(R), \quad (2)$$

where

$$f_{i,j}(R) = \begin{cases} (\log R_i - \log R_j)^2, & -\tau_1 < P_{i,j}^* < \tau_1 \\ (\max(0, \tau_2 - \log R_i + \log R_j))^2, & P_{i,j}^* > \tau_2 \\ (\max(0, \tau_2 - \log R_j + \log R_i))^2, & P_{i,j}^* < -\tau_2 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where R^* is the rendered ground truth reflectance, $P_{i,j}^* = \log R_i^* - \log R_j^*$, $\tau_1 = \log(1.05)$ and $\tau_2 = \log(1.5)$. The intuition is that we categorize pairs of ground truth reflectances at pixels (i, j) as having an “equal,” “greater than,” or “less than” relationship, and then add a penalty if the predicted reflectances at those pixels do not satisfy the same relationship.

1.2 Additional hyperparameter settings

In all experiments described in the main paper, following [2] and [3], we predict grayscale reflectance and shading images from our networks. When we train our networks on CGI dataset, we experiment using CGI data (correspond to Ours[†] described in main paper) or real IIW data (correspond to Ours described in main paper) as validation set. Additionally, we set $\lambda_{IIW} = \lambda_{SAW} = 2$, $\lambda_{ord} = \lambda_{rs} = \lambda_{S/NS} = 1$, $\lambda_{rec} = 2$ and $\lambda_{ss} = 4$. The number of image scales $L = 4$. The margin in Equation 7 in the main paper $m = 0.425$. For simplicity, the covariance matrix Σ defined in $\mathcal{L}_{rsmooth}$ (Equation 10 in the main paper) is a diagonal matrix, defined as:

$$\Sigma = \begin{bmatrix} \sigma_p^2 & & & & & \\ & \sigma_p^2 & & & & \\ & & \sigma_I^2 & & & \\ & & & \sigma_c^2 & & \\ & & & & \sigma_c^2 & \\ & & & & & \sigma_c^2 \end{bmatrix}$$

where $\sigma_p = 0.1$, $\sigma_I = 0.12$, $\sigma_c = 0.03$.

2 Visual comparisons between CGI and SUNCG/PBRs renderings

In Figure 1 we provide additional visual comparisons between rendered images from our CGINTRINSICS dataset and the SUNCG/PBRs dataset, illustrating the greater signal-to-noise ratio and realism in our renderings.

3 Additional qualitative results on IIW/SAW

In this section, we provide a large number of additional qualitative comparisons of intrinsic image decompositions from the IIW/SAW test sets. We include decompositions predicted from our network trained solely on the CGINTRINSICS dataset, as well as decompositions from our network trained on CGINTRINSICS plus training data from IIW and SAW. We compare our results to those of two state-of-the-art algorithms, Bell *et al.* [2] and Zhou *et al.* [3]. Predictions are shown in Figures 2-13. Note that the images in Figure 12 and 13 are from the NYUv2 [4] dataset, and are also included in the SAW dataset.

We observe that our decompositions are consistently better than those of the two state-of-the-art algorithms [2, 3] across a large number of photos of a variety of indoor environments. Note in particular how our method is better able to avoid attributing reflectance texture to the shading image. These results suggests the strong generalization abilities of our models and the surprising effectiveness of our proposed synthetic CGINTRINSICS dataset. However, our predictions still make mistakes such as residual textures in shading channels and residual shadows in reflectance channels.



Fig. 1. Additional visual comparisons between SUNCG/PBRs renderings and our CGI renderings. Odd columns: images from SUNCG/PBRs. Even columns: images from our CGI dataset. The images in our dataset have higher SNR and are generally more realistic.

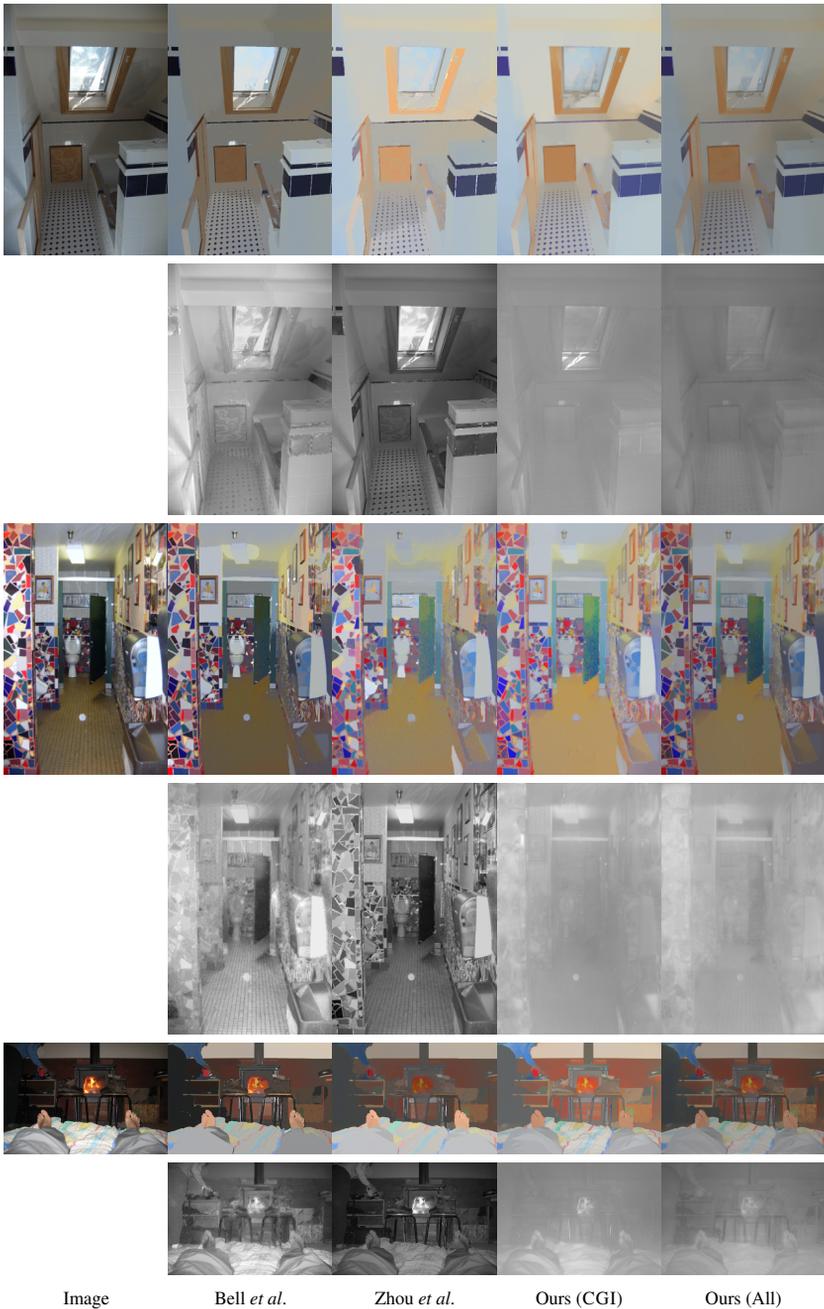


Fig. 2. Additional qualitative comparisons on the IIW/SAW test sets. Odd rows: predicted reflectance images. Even rows: predicted shading images. Columns from left to right: input image, results of Bell *et al.* [2], results of Zhou *et al.* [3], results of our CGI-trained network, results of our network trained on CGI+IIW+SAW.

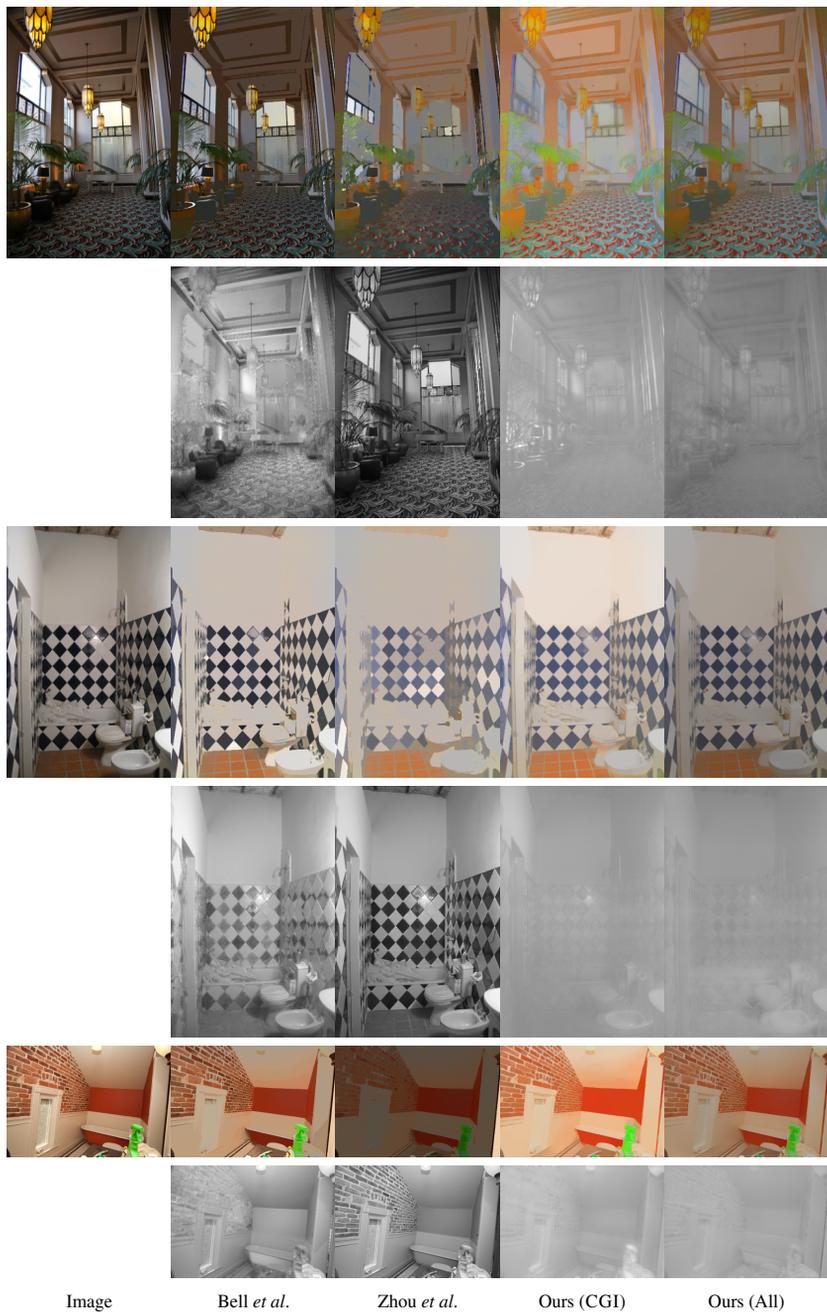


Fig. 3. Additional qualitative comparisons on the IIW/SAW test sets. Odd rows: predicted reflectance images. Even rows: predicted shading images. Columns from left to right: input image, results of Bell *et al.* [2], results of Zhou *et al.* [3], results of our CGI-trained network, results of our network trained on CGI+IIW+SAW.

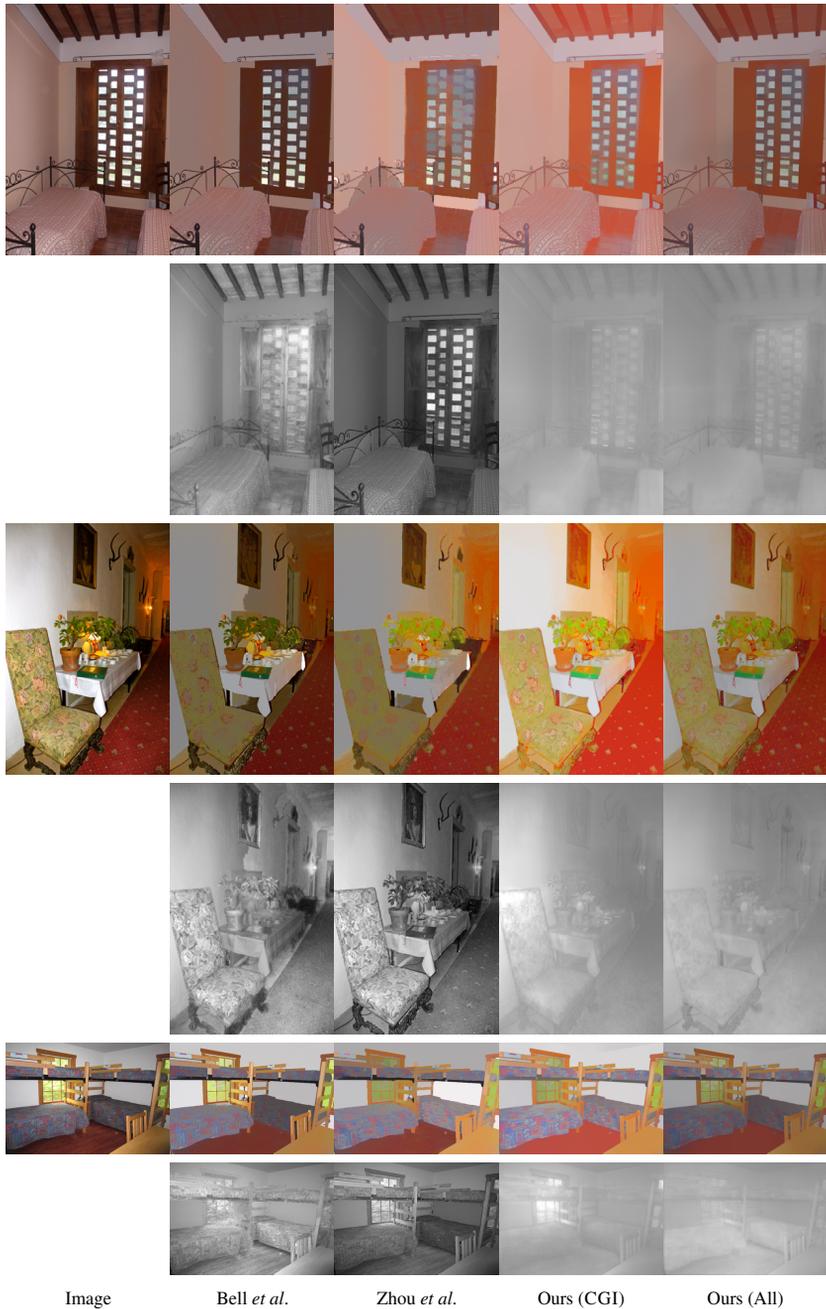


Fig. 4. Additional qualitative comparisons on the IIW/SAW test sets. Odd rows: predicted reflectance images. Even rows: predicted shading images. Columns from left to right: input image, results of Bell *et al.* [2], results of Zhou *et al.* [3], results of our CGI-trained network, results of our network trained on CGI+IIW+SAW.

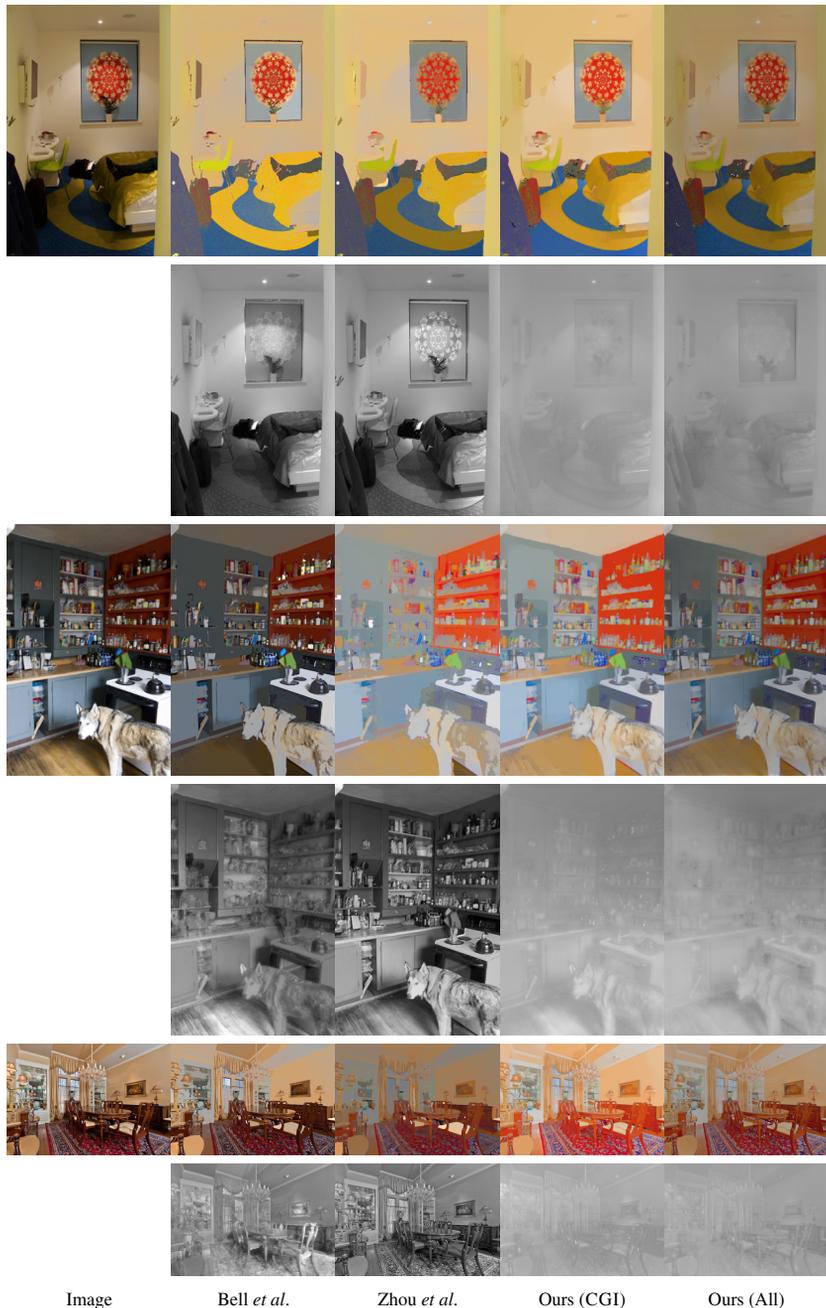


Fig. 5. Additional qualitative comparisons on the IIW/SAW test sets. Odd rows: predicted reflectance images. Even rows: predicted shading images. Columns from left to right: input image, results of Bell *et al.* [2], results of Zhou *et al.* [3], results of our CGI-trained network, results of our network trained on CGI+IIW+SAW.

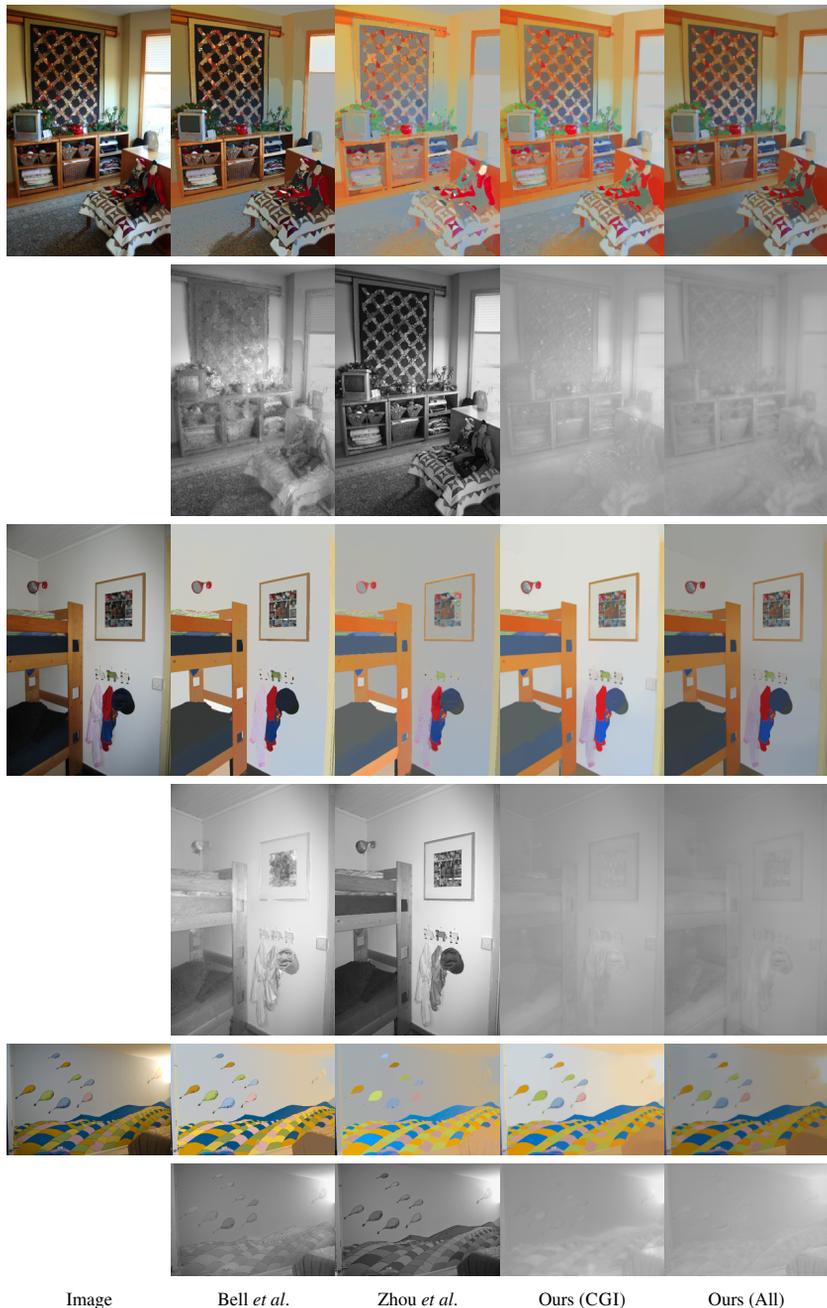


Fig. 6. Additional qualitative comparisons on the IIW/SAW test sets. Odd rows: predicted reflectance images. Even rows: predicted shading images. Columns from left to right: input image, results of Bell *et al.* [2], results of Zhou *et al.* [3], results of our CGI-trained network, results of our network trained on CGI+IIW+SAW.

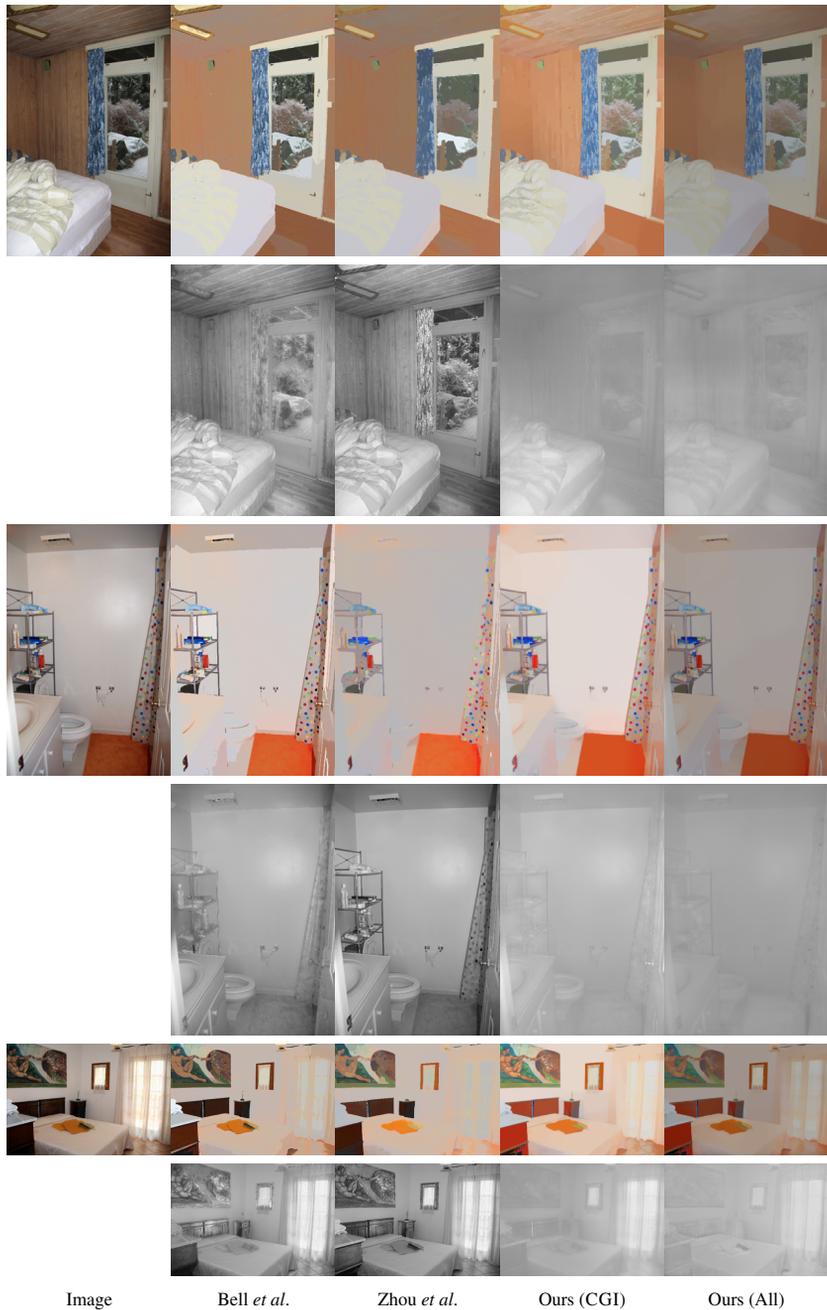


Fig. 7. Additional qualitative comparisons on the IIW/SAW test sets. Odd rows: predicted reflectance images. Even rows: predicted shading images. Columns from left to right: input image, results of Bell *et al.* [2], results of Zhou *et al.* [3], results of our CGI-trained network, results of our network trained on CGI+IIW+SAW.

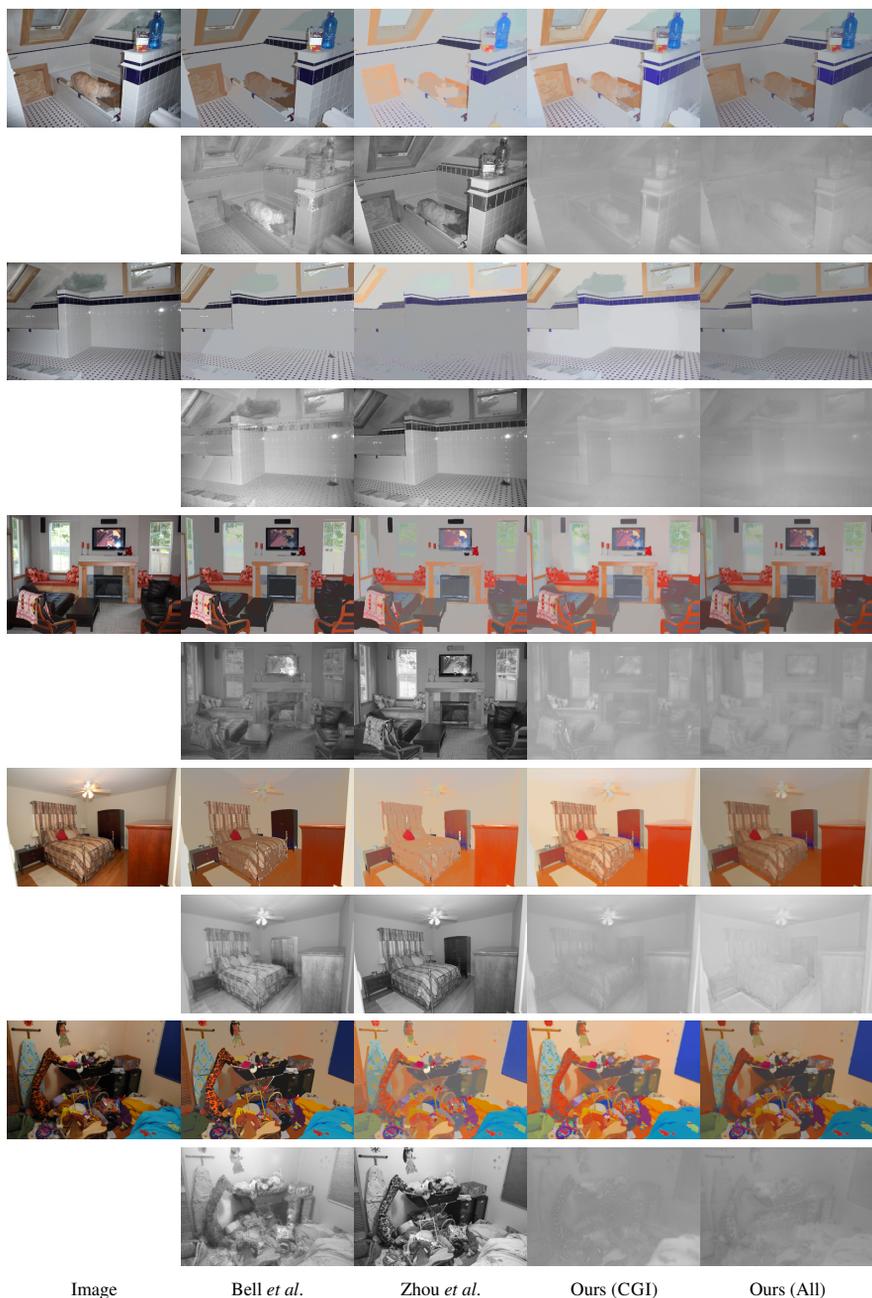


Fig. 8. Additional qualitative comparisons on the IIW/SAW test sets. Odd rows: predicted reflectance images. Even rows: predicted shading images. Columns from left to right: input image, results of Bell *et al.* [2], results of Zhou *et al.* [3], results of our CGI-trained network, results of our network trained on CGI+IIW+SAW.

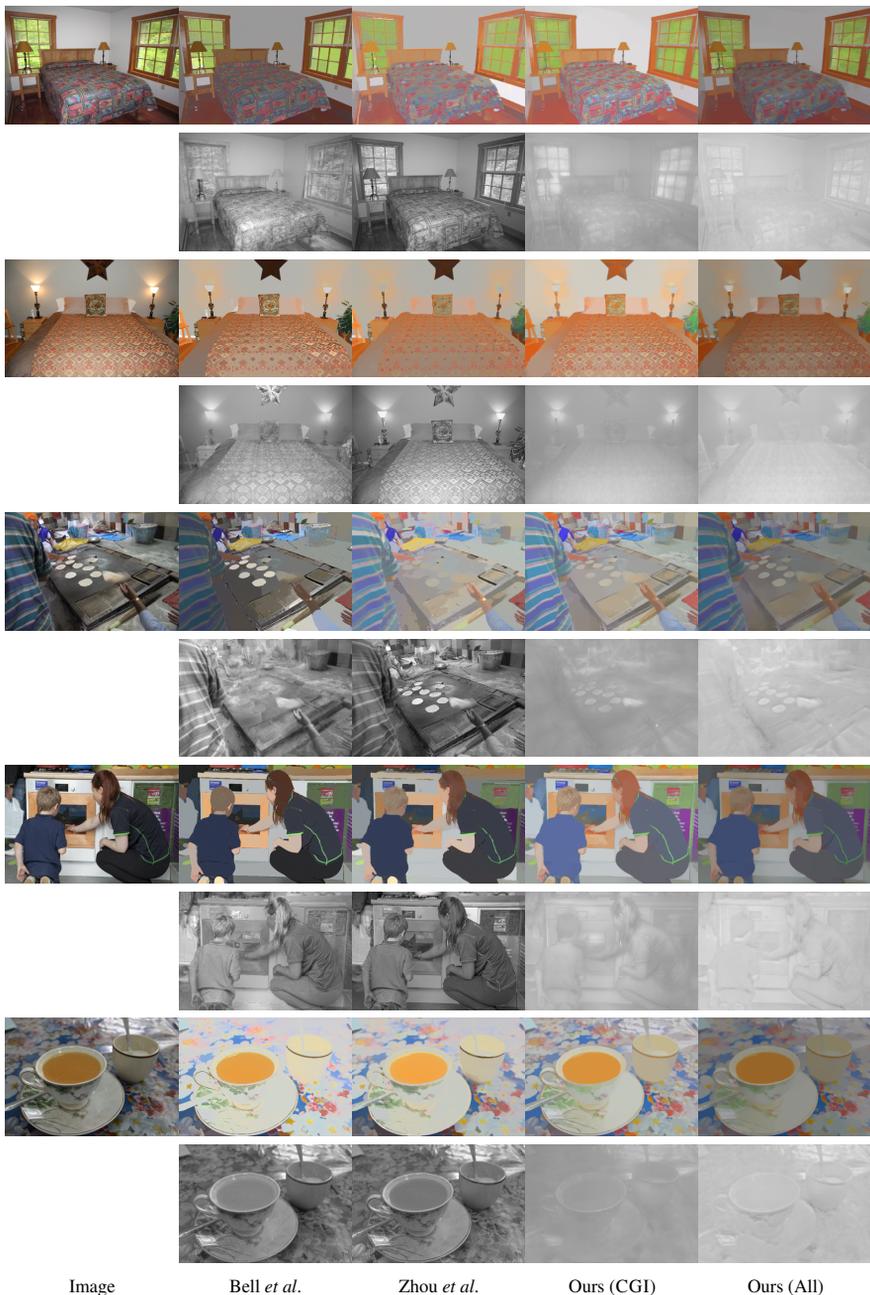


Fig. 9. Additional qualitative comparisons on the IIW/SAW test sets. Odd rows: predicted reflectance images. Even rows: predicted shading images. Columns from left to right: input image, results of Bell *et al.* [2], results of Zhou *et al.* [3], results of our CGI-trained network, results of our network trained on CGI+IIW+SAW.

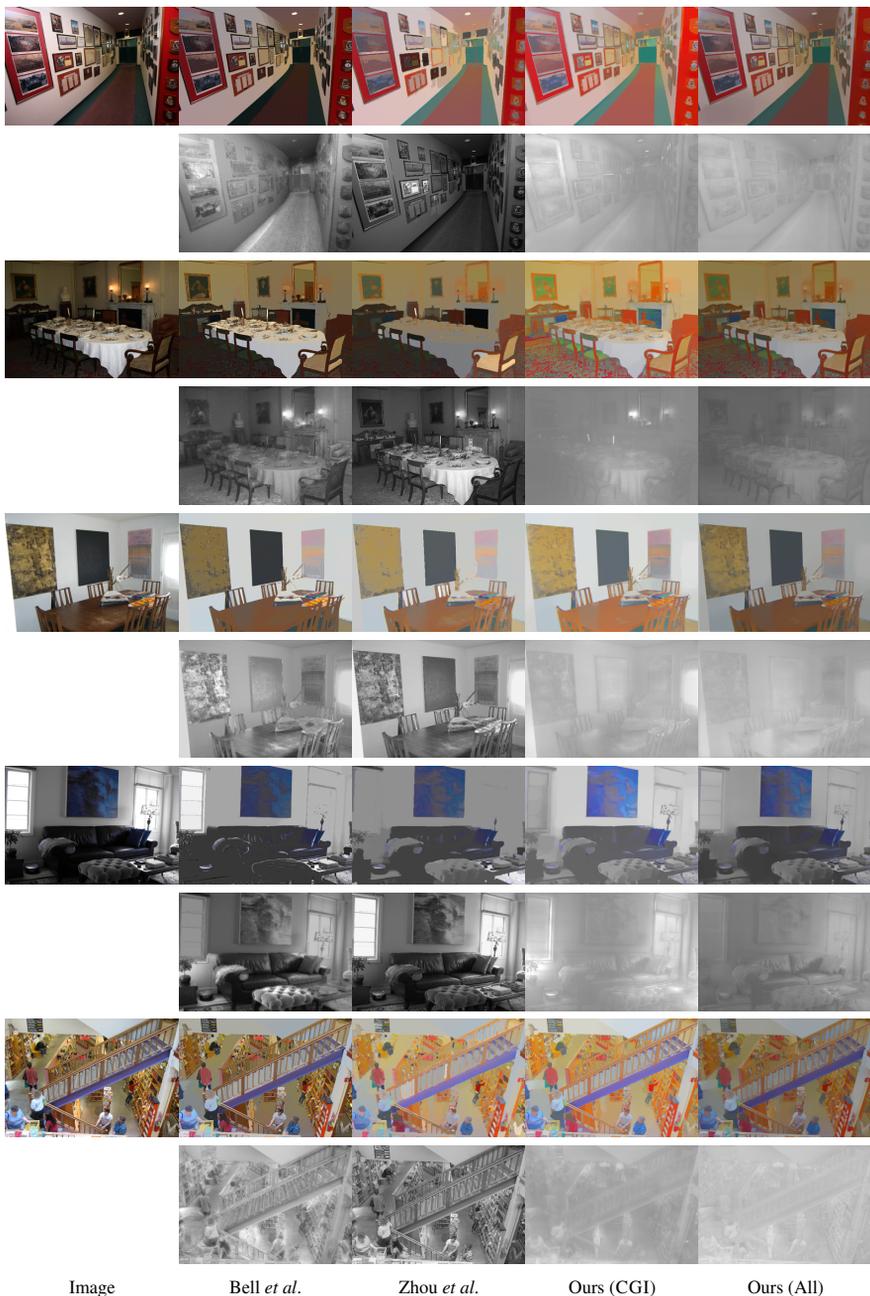


Fig. 10. Additional qualitative comparisons on the IIW/SAW test sets. Odd rows: predicted reflectance images. Even rows: predicted shading images. Columns from left to right: input image, results of Bell *et al.* [2], results of Zhou *et al.* [3], results of our CGI-trained network, results of our network trained on CGI+IIW+SAW.



Fig. 11. Additional qualitative comparisons on the IIW/SAW test sets. Odd rows: predicted reflectance images. Even rows: predicted shading images. Columns from left to right: input image, results of Bell *et al.* [2], results of Zhou *et al.* [3], results of our CGI-trained network, results of our network trained on CGI+IIW+SAW.

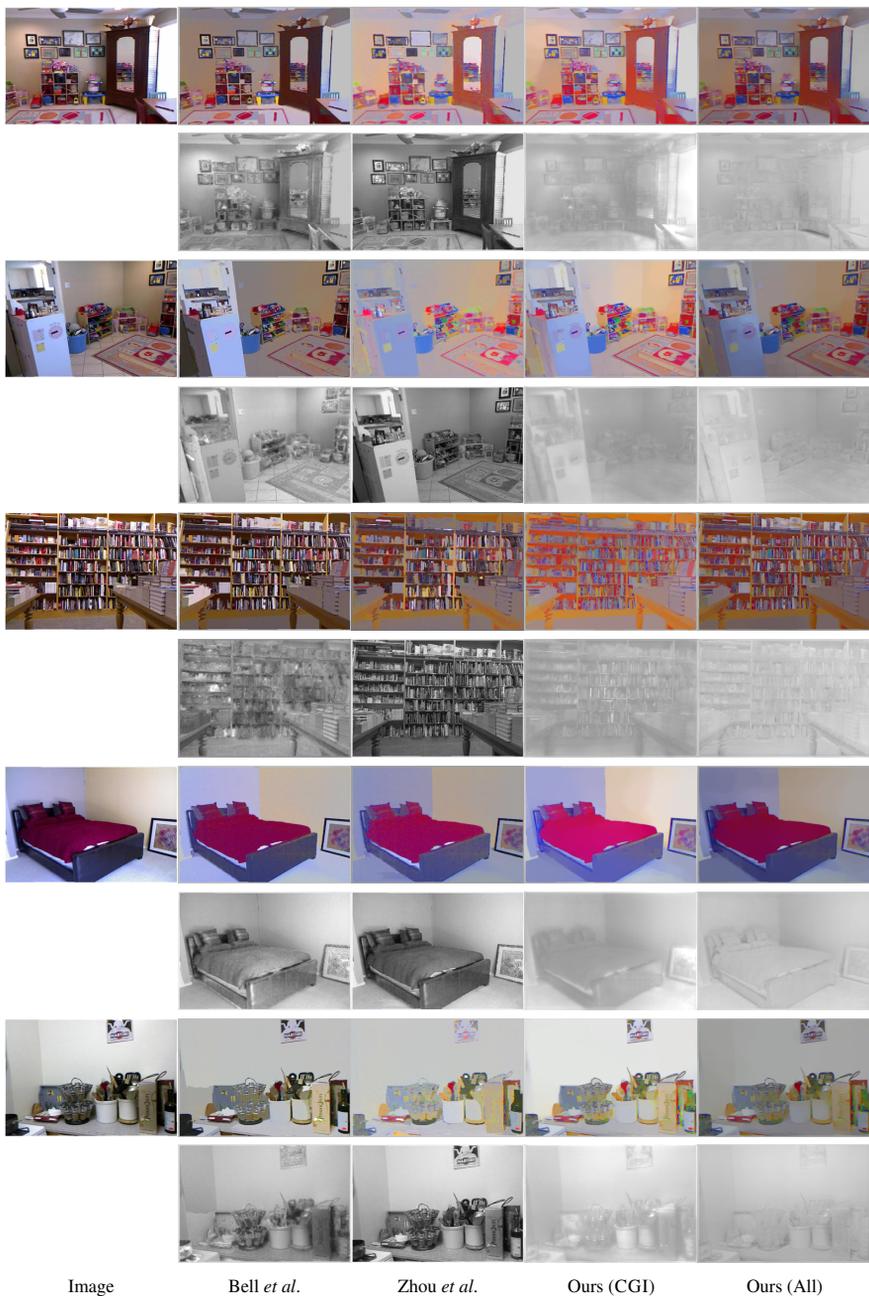


Fig. 12. Additional qualitative comparisons on the NYUv2 dataset. Note that the images in NYUv2 are also included in the SAW testset. Odd rows: predicted reflectance images. Even rows: predicted shading images. Columns from left to right: input image, results of Bell *et al.* [2], results of Zhou *et al.* [3], results of our CGI-trained network, results of our network trained on CGI+IIW+SAW.



Fig. 13. Additional qualitative comparisons on the NYUv2 dataset. Note that the images in NYUv2 are also included in the SAW testset. Odd rows: predicted reflectance images. Even rows: predicted shading images. Columns from left to right: input image, results of Bell *et al.* [2], results of Zhou *et al.* [3], results of our CGI-trained network, results of our network trained on CGI+IIW+SAW.

References

1. Zhang, Y., Song, S., Yumer, E., Savva, M., Lee, J.Y., Jin, H., Funkhouser, T.: Physically-based rendering for indoor scene understanding using convolutional neural networks. In: Proc. Computer Vision and Pattern Recognition (CVPR). (2017) 5057–5065
2. Bell, S., Bala, K., Snavely, N.: Intrinsic images in the wild. *ACM Trans. Graphics* **33**(4) (2014) 159
3. Zhou, T., Krahenbuhl, P., Efros, A.A.: Learning data-driven reflectance priors for intrinsic image decomposition. In: Proc. Int. Conf. on Computer Vision (ICCV). (2015) 3469–3477
4. Silberman, N., Hoiem, D., Kohli, P., Fergus, R.: Indoor segmentation and support inference from rgb-d images. In: Proc. European Conf. on Computer Vision (ECCV). (2012)