Supplementary Material: Learning Intrinsic Image Decomposition from Watching the World

Zhengqi Li Noah Snavely

Department of Computer Science & Cornell Tech, Cornell University

In this supplementary material, we first provide details for the hyperparameter settings we use during training (Section 1) and a detailed derivation of our proposed efficient All-Pairs Weighted Least Squares (APWLS) computation (Section 2). Then, in Section 3, we provide additional details for our modified Shading Annotations in the Wild (SAW) evaluation metrics and include a full precision-recall (PR) curve for all the methods we evaluated (see Section 6.2 of the main paper). Finally, we provide additional qualitative prediction results on the IIW and SAW test sets from our BIGTIME-trained CNN (Section 4).

1. Hyperparameters Setting

For all experiments, we set our hyperparameters as follows. For the overall energy function defined in Equation 3 in the main paper, we set $w_1 = 1, w_2 = 6$ and $w_3 = 2$. For Equation 8 describing the affinity between pixels, we define a covariance matrix Σ between reflectance feature vectors $\mathbf{f}_p$ and $\mathbf{f}_q$ as follows: we set Σ to be a diagonal matrix for simplicity, and define $\Sigma = \text{diag}(0.1^2, 0.1^2, 0.1^2, 0.025^2, 0.025^2)$. Lastly, for Equations 11 and 12 relating to shading smoothness, we set $\lambda^{\text{med}} = 20$ and $\overline{\lambda^{\text{med}}} = 4$.

2. All-Pairs Weighted Least Squares (APWLS)

In this section, we provide a detailed derivation of our proposed All-Pairs Weighted Least Squares computation (APWLS), as described in Section 5.5 of the main paper. Suppose that we have a image sequence with m images, and each image has n pixels. Now suppose each image I^i is associated with two matrices P^i and Q^i and two predictions X^i and Y^i. We then can write APWLS as

$$\text{APWLS} = \sum_{i=1}^{m}\sum_{j=1}^{m} ||P^i \otimes Q^j \otimes (X^i - Y^j)||_F^2 \tag{1}$$

$$= \sum_{p=1}^{n}\sum_{i=1}^{m}\sum_{j=1}^{m} \left(P_p^i Q_p^j (X_p^i - Y_p^j)\right)^2 \tag{2}$$

$$= \sum_{p=1}^{n}\left(\sum_{i=1}^{m}(P_p^i)^2\left(\sum_{j=1}^{m}(Q_p^j)^2(X_p^i - Y_p^j)^2\right)\right) \tag{3}$$

$$= \sum_{p=1}^{n}\left(\sum_{i=1}^{m}(P_p^i)^2\left(\left(\sum_{j=1}^{m}(Q_p^j)^2\right)(X_p^i)^2 + \sum_{j=1}^{m}(Q_p^j)^2(Y_p^j)^2 - 2X_p^i\sum_{j=1}^{m}(Q_p^j)^2 Y_p^j\right)\right) \tag{4}$$

$$= \sum_{p=1}^{n}\left(\left(\sum_{j=1}^{m}(Q_p^j)^2\right)\left(\sum_{i=1}^{m}(X_p^i)^2(P_p^i)^2\right) + \sum_{i=1}^{m}(P_p^i)^2\left(\sum_{j=1}^{m}(Q_p^j)^2(Y_p^j)^2\right) - 2\left(\sum_{i=1}^{m}(P_p^i)^2 X_p^i\right)\left(\sum_{j=1}^{m}(Q_p^j)^2(Y_p^j)\right)\right) \tag{5}$$

$$= \mathbf{1}^\top(\Sigma_{Q^2} \otimes \Sigma_{P^2X^2} + \Sigma_{P^2} \otimes \Sigma_{Q^2Y^2} - 2\Sigma_{P^2X} \otimes \Sigma_{Q^2Y})\mathbf{1} \tag{6}$$

where $\Sigma_{Q^2} = \sum_{i=1}^{m} Q^i \otimes Q^i$; $\Sigma_{P^2} = P^i \otimes P^i$; $\Sigma_{P^2X^2} = \sum_{i=1}^{m} P^i \otimes P^i \otimes X^i \otimes X^i$; $\Sigma_{Q^2Y^2} = Q^i \otimes Q^i \otimes Y^i \otimes Y^i$; $\Sigma_{P^2X} = P^i \otimes P^i \otimes X^i$; $\Sigma_{Q^2Y} = Q^i \otimes Q^i \otimes Y^i$.

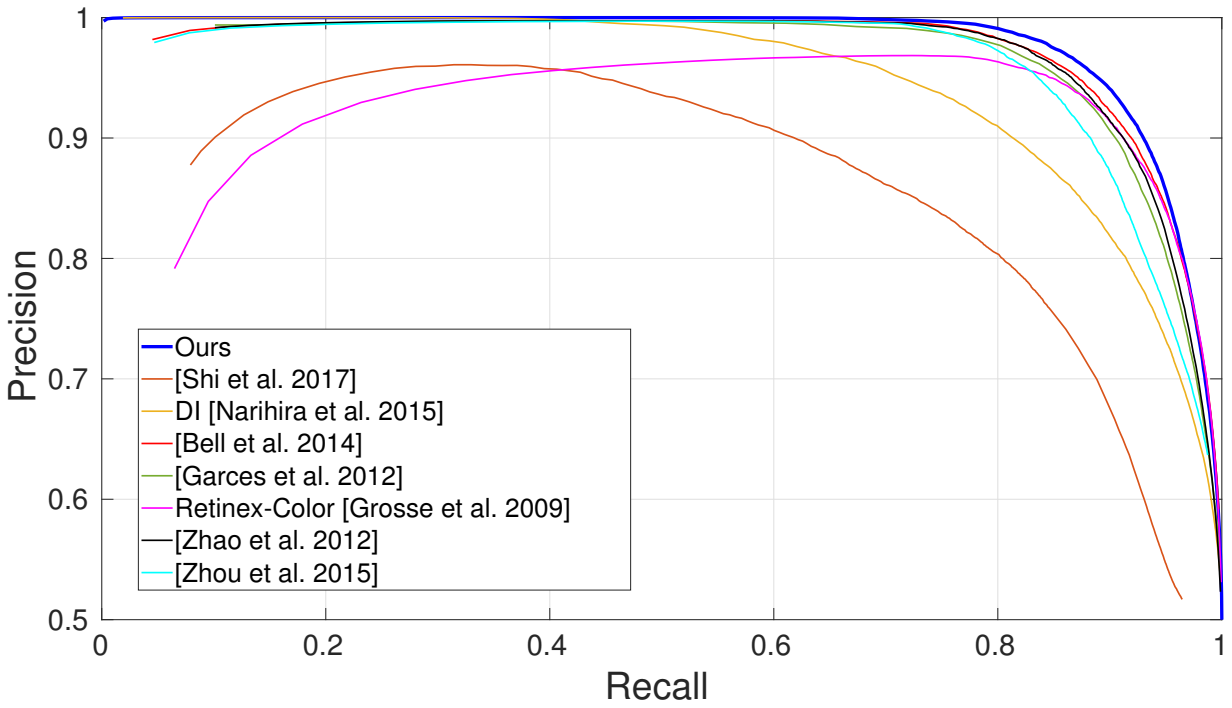


Figure 1: **Precision-recall curves for shading predictions on the SAW test set.** Please see the main paper for a description of each method.

3. Additional details for SAW evaluation metrics

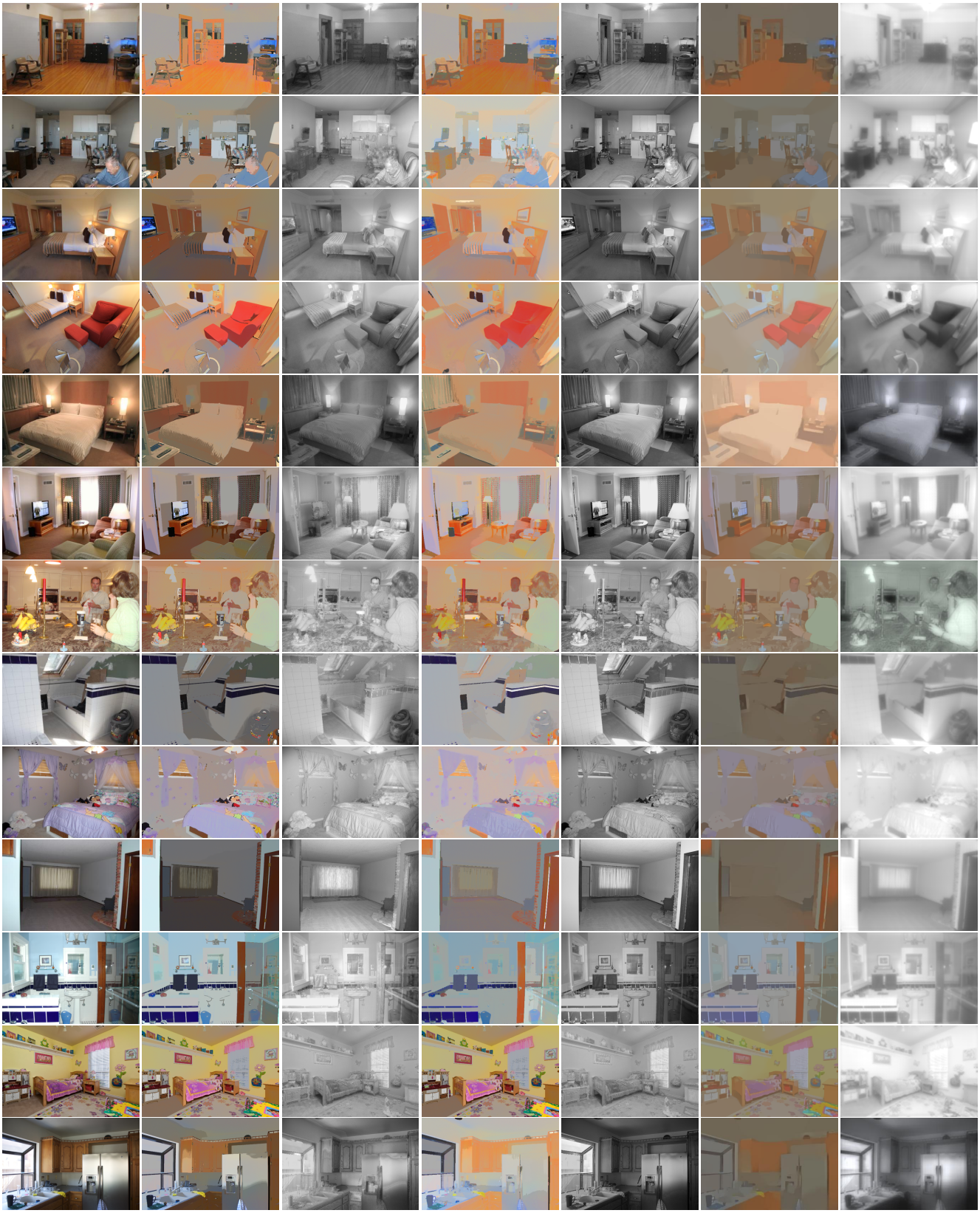
In this section, we reiterate the two improvements we made to the metric used to evaluate results on SAW annotations (described in Section 6.2 of the main paper) and provide more detailed explanations.

First, the original SAW error metric, as described by Kovacs *et al*. [2], is based on classifying a pixel p as having smooth/nonsmooth shading based on the gradient magnitude of the predicted shading image, $||\nabla S||_2$, normalized to the range $[0, 1]$. Instead, we measure the gradient magnitude in the *log* domain. We do this because of the scale ambiguity inherent to shading and reflectance, and because it is possible to have very bright values in the shading channel (e.g., due to strong sunlight), and in such cases if we normalize shading to $[0, 1]$ then most of the resulting values will be close to 0. In contrast, computing the gradient magnitude of log shading $||\nabla \log S||_2$ achieves scale invariance, resulting in fairer comparisons for all methods. As in [2], we sweep a threshold τ to create a precision-recall (PR) curve that captures how well each method captures smooth and non-smooth shading. PR curves for all methods described in the main paper are shown in Figure 1.

Second, Kovacs *et al*. [2] apply a 10×10 maximum filter to the shading gradient magnitude image before computing PR curves, because many shadow boundary annotations are not precisely localized. However, this maximum filter can result in degraded performance for smooth shading regions. Consider adding 1% salt-and-pepper noise to the shading estimate. Applying a maximum filter to this noisy gradient magnitude image would make it seem as if there are large changes everywhere. Moreover, we found several annotated smooth regions are close to the boundaries of shading changes caused by depth/normal discontinuities, and if we apply a maximum filter, we might integrate incorrect shading information out of annotated regions into our evaluation. Instead, we create two maps, the original $||\nabla \log S||_2$, and the 10×10 maximum filtered to $||\nabla \log S||_2$, which we denote $||\nabla \log S||_2^{\max}$. We use $||\nabla \log S||_2$ to classify smooth shading annotations and $||\nabla \log S||_2^{\max}$ to classify non-smooth annotations.

4. Qualitative Results

In this section, we provide additional qualitative results on IIW/SAW test set and compare our network predictions with two state-of-art intrinsic image decomposition algorithms, Bell *et al*. [1] and Zhou *et al*. [3]. These qualitative comparisons are shown in Figure 2.



(a) Image (b) Bell *et al*. (R) (c) Bell *et al*. (S) (d) Zhou *et al*. (R) (e) Zhou *et al*. (S) (f) Ours (R) (g) Ours (S)

Figure 2: **Additional qualitative comparisons for intrinsic image decomposition on the IIW and SAW test sets.** We compare our network predictions with two state-of-art intrinsic image decomposition algorithms (Bell *et al*. [1] and Zhou *et al*. [3]).

References

[1] S. Bell, K. Bala, and N. Snavely. Intrinsic images in the wild. *ACM Trans. Graphics*, 33(4):159, 2014.

[2] B. Kovacs, S. Bell, N. Snavely, and K. Bala. Shading annotations in the wild. *Proc. Computer Vision and Pattern Recognition (CVPR)*, 2017.

[3] T. Zhou, P. Krahenbuhl, and A. A. Efros. Learning data-driven reflectance priors for intrinsic image decomposition. In *Proc. Int. Conf. on Computer Vision (ICCV)*, pages 3469–3477, 2015.