-Supplementary Material- RAWtoBit: A Fully End-to-end Camera ISP Network

Wooseok Jeong and Seung-Won Jung*

Department of Electrical Engineering, Korea University, Seoul, Korea
{561wesd, swjung83}@korea.ac.kr

1 Additional Experimental Results

In addition to the LiteISPNet[4], we also trained SID[1] and InvISP[3] for the cascaded structure and tested under the same dataset settings. As can be seen on Fig. S-1, RBN outperformed the cascaded structure with these two models by a large margin. These ISP-Net models resulted in 32.68 dB (SID) and 33.16 dB (InvISP), respectively, without compression; however, our RBN achieved above 33.5 dB even with compression.

Regarding the cascaded structure, combining lossy compression of RAW image and ISP-Net is also a valid approach. For this experiments, pretrained compression teacher network for RBN and LiteISPNet are cascaded. The result is shown in Fig. S-1. While this structure is valid and necessary for completeness, its RD performance is limited since lossy compression of information-rich RAW data is critical to the ISP-Net.

2 Computational Cost

We report the number of parameters for the experimented models in Table S-1. While the proposed method has more parameters than the unified structure or cascaded structure, two teacher networks are only used to guide the training and detached during evaluation. Hence, once the training is done, RBN+KD has less parameters than the cascaded structure and is quite comparable to the unified structure.

Unified	Cascade	Proposed		
14.1M	23.0M	ISP Teacher	Comp Teacher	RBN
		8.7M	13.9M	16.4M

Table S-1. Number of parameters for the experimented models.

* Corresponding author.

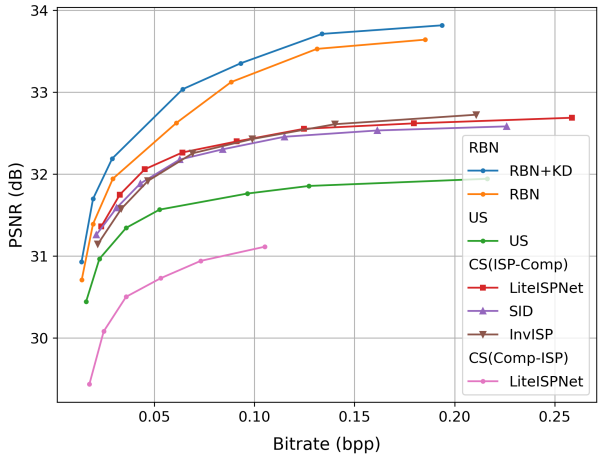

Fig. S-1. RD performance comparison with additional experiments.

3 Visual Comparisons

We provide additional visual comparisons and generalization results in this section. Figs. S-2 and S-3 show the error maps between the ground truth and reconstructed sRGB images obtained by different models. Note that RBNs (w/ and w/o KD) resulted in fewer errors compared to the cascaded structure (CS) and unified structure (US), even at the expense of fewer bitrates. It is also evident from Fig. S-4 that RBN with KD can render colors more accurately compared to the other models. The generalization test results are shown in Figs. S-5 and S-6, where the test RAW images are from the Nikon D7000 and Canon 5D subsets in the RAISE dataset [2], respectively. The results on Nikon D7000 reveal that our model can produce visually pleasant sRGB images, whereas the results on Canon 5D are unsatisfactory, especially in global contrast and color rendering. In other words, a trained model can be reused for the sensors from the same manufacturer since they share similar characteristics such as linearization and color correction parameters; however, sensor-specific network training is somewhat inevitable for the sensors from different manufacturers.

References

1. Chen, C., Chen, Q., Xu, J., Koltun, V.: Learning to see in the dark. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3291–3300 (2018)
2. Dang-Nguyen, D.T., Pasquini, C., Conotter, V., Boato, G.: Raise: A raw images dataset for digital image forensics. In: Proceedings of the ACM Multimedia Systems Conference. pp. 219–224 (2015)

3. Xing, Y., Qian, Z., Chen, Q.: Invertible image signal processing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6287–6296 (2021)
4. Zhang, Z., Wang, H., Liu, M., Wang, R., Zhang, J., Zuo, W.: Learning RAW-to-sRGB mappings with inaccurately aligned supervision. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4348–4358 (2021)

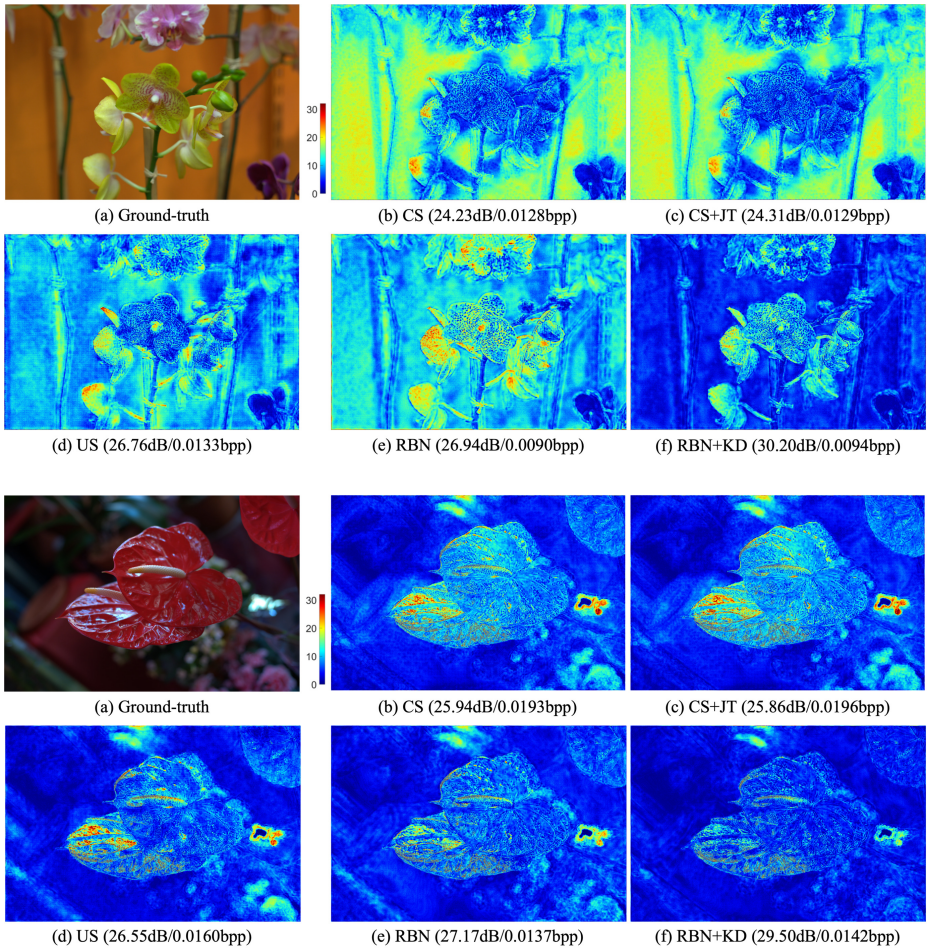

Fig. S-2. Visual comparisons: (a) Ground-truth sRGB image and the error maps between the ground-truth and reconstructed sRGB images for (b) cascaded structure, (c) cascaded structure with joint fine-tuning, (d) unified structure, (e) RBN, and (f) RBN+KD.

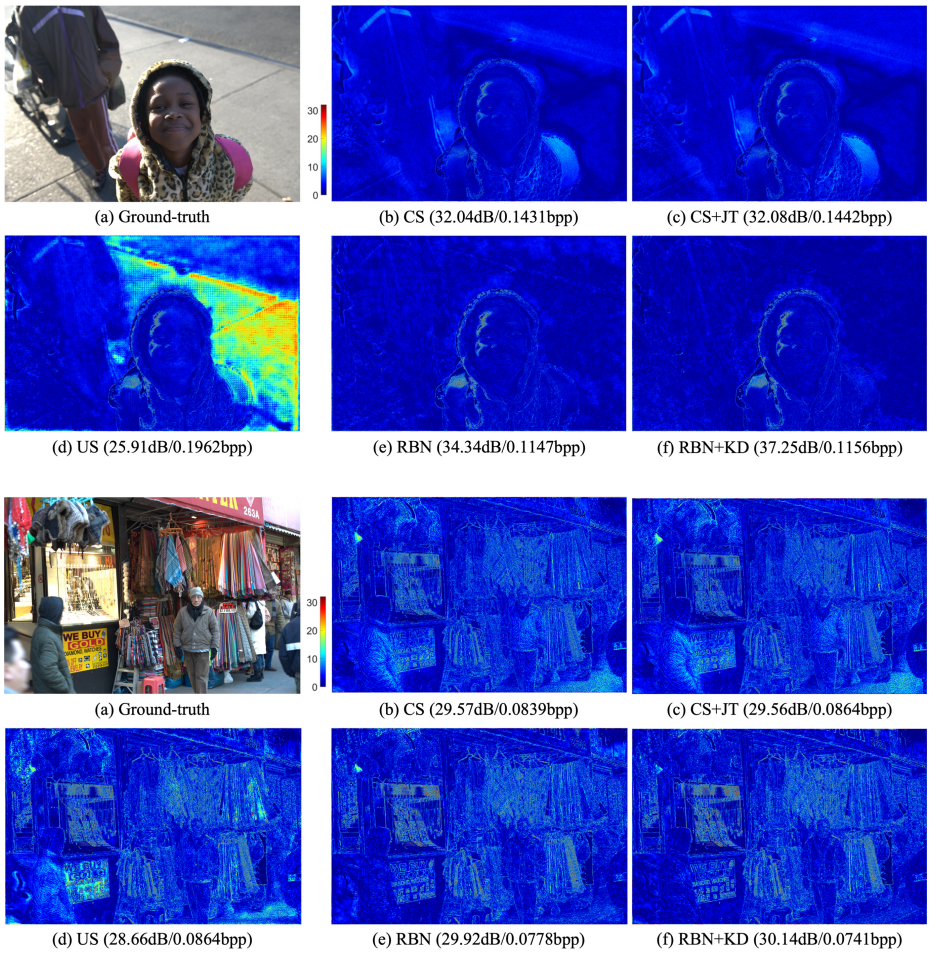

Fig. S-3. Visual comparisons: (a) Ground-truth sRGB image and the error maps between the ground-truth and reconstructed sRGB images for (b) cascaded structure, (c) cascaded structure with joint fine-tuning, (d) unified structure, (e) RBN, and (f) RBN+KD.

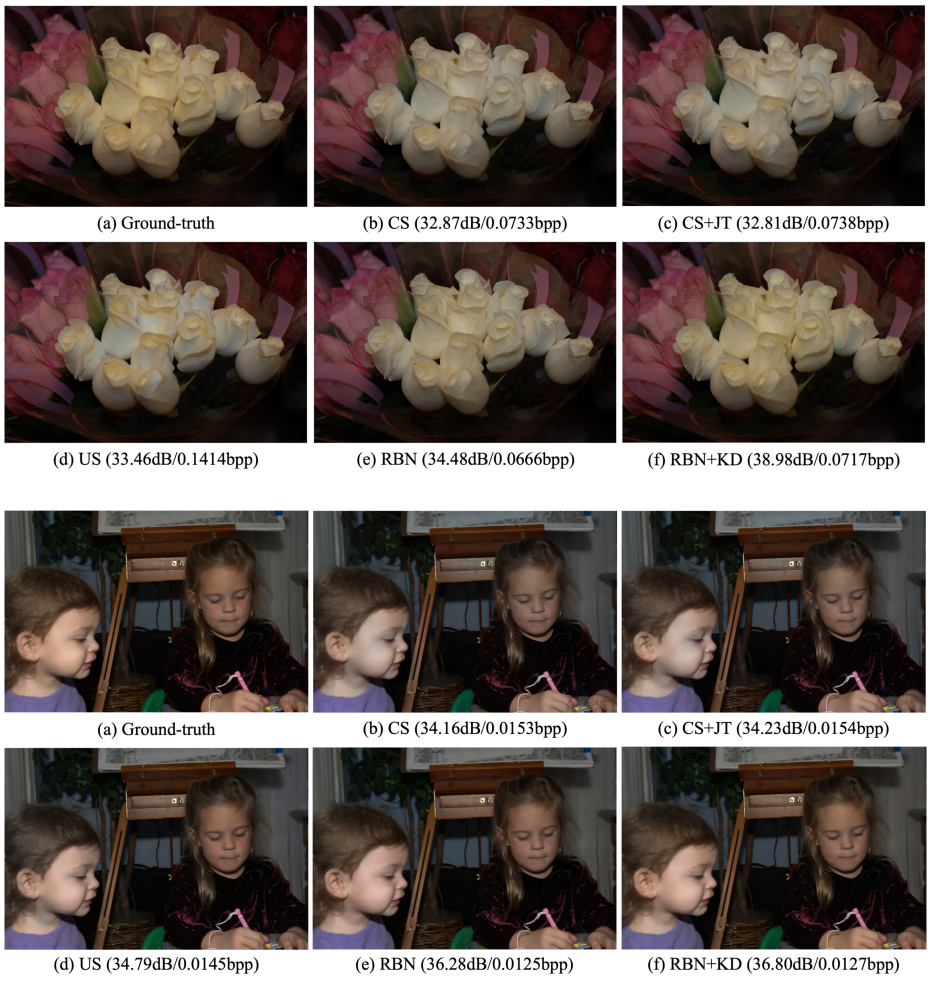

Fig. S-4. Visual comparisons: (a) Ground-truth sRGB image and reconstructed sRGB images obtained by (b) cascaded structure, (c) cascaded structure with joint fine-tuning, (d) unified structure, (e) RBN, and (f) RBN+KD.

Fig. S-5. Samples of the sRGB images rendered from RAW captured with Nikon D7000. The RBN+KD model used for testing was trained with RAW from Nikon D700. Bilinear demosaicing is applied to the RAW for visualization.

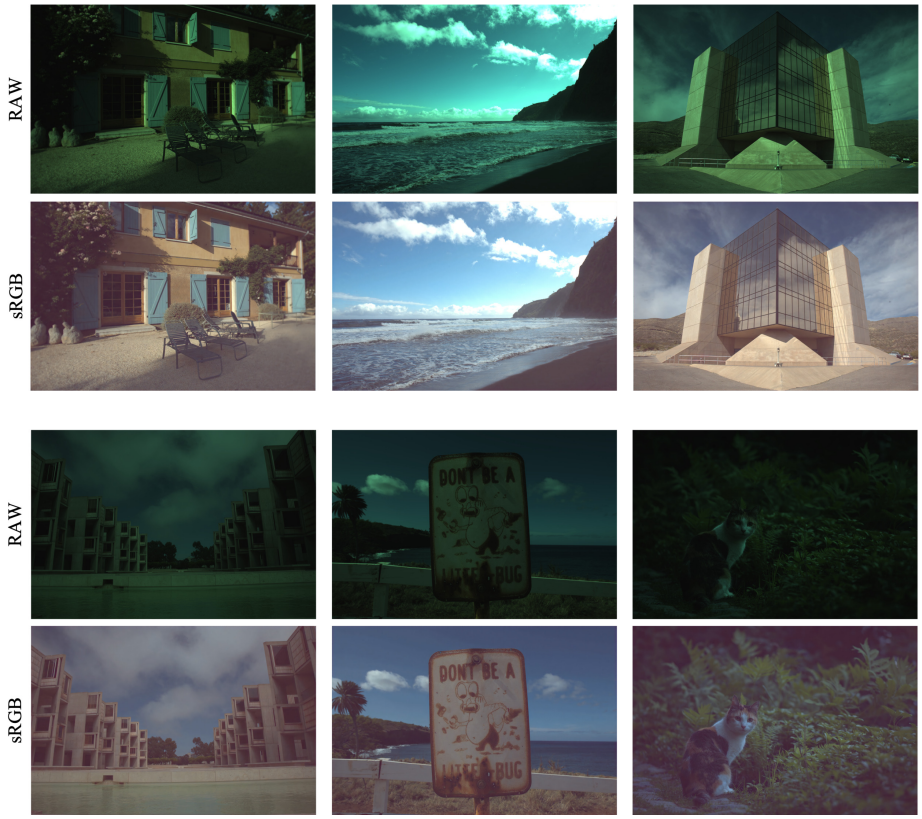

Fig. S-6. Samples of the sRGB images rendered from RAW captured with Canon 5D. The RBN+KD model used for testing was trained with RAW from Nikon D700. Bilinear demosaicing is applied to the RAW for visualization.