Supplementary Material
Learning sRGB-to-Raw-RGB De-rendering with Content-Aware Metadata

Seonghyeon Nam[1*] Abhijith Punnappurath[2] Marcus A. Brubaker[1,2] Michael S. Brown[2]
[1]York University [2]Samsung AI Center – Toronto
{shnnam,mab}@eecs.yorku.ca, {abhijith.p,michael.b1}@samsung.com

This supplementary material provides additional results and details that could not be included in the main paper due to space constraints.

S1. Results on camera ISP images

In the main paper, we had evaluated our raw reconstruction accuracy using the NUS [2] dataset with the sRGB images rendered using a software ISP emulator [3]. The NUS dataset also contains the sRGB-JPEG images rendered by each individual camera's hardware ISP. We used these sRGB images instead of the software ISP emulator [3], and the results are presented in Table S1. Our method generalizes well to different ISPs, and outperforms competitors. We do note that compared to Table 1 of our main paper, there is a drop in performance for all methods due to the more complex ISPs.

S2. Other sampling rates

We report PSNR (dB) for our method (with fine-tuning) at different sampling rates in Table S2. The results for $k = 1.5\%$ are reproduced from Table 1 of the main paper. There is a significant improvement from 0% i.e., no metadata, to 0.4%. Performance improves with higher k values but at the expense of larger metadata size.

S3. Comparison with SLIC superpixel

We performed an experiment where we replaced our learned superpixel with SLIC [1] for sampling, and trained our reconstruction network under the same settings. On the Samsung camera, we obtained PSNR/SSIM values of 45.94 / 0.9958 as against 49.57 / 0.9975 produced by our method, demonstrating the superiority of an end-to-end learnable superpixel sampler.

S4. Additional experiments

In Fig. S1, we compare the error maps of our outputs before and after fine-tuning. Fig. S2, Fig. S3, and Fig. S4 show additional qualitative results on three cameras. For visibility, we omit output raw-RGB images and ground truth. Fig. S5 to Fig. S7 show visualizations of learned superpixels and sampling masks.

Method	Fine-tuning	Samsung NX2000		Olympus E-PL6		Sony SLT-A57	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
RIR [4]	N/A	37.62	0.9696	42.19	0.9865	45.22	0.9916
SAM [5]	N/A	38.80	0.9725	43.15	0.9881	46.02	0.9921
Ours	No	40.80	0.9812	46.89	0.9938	48.51	0.9947
Ours	Yes	**41.59**	**0.9818**	**47.76**	**0.9944**	**49.58**	**0.9954**

Table S1. Quantitative evaluation on raw reconstruction using the camera ISP sRGB images.

[*]Work done while an intern at the Samsung AI Center – Toronto.

Percentage Samples k	0%	0.4%	1.5%	6.25%
Samsung NX2000	38.86	48.56	49.57	50.31
Olympus E-PL6	42.30	50.62	51.54	52.20
Sony SLT-A57	44.79	51.09	53.11	53.44

Table S2. An ablation on the sampling rate k.

References

[1] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE TPAMI*, 34(11):2274–2282, 2012. 1

[2] Dongliang Cheng, Dilip K Prasad, and Michael S Brown. Illuminant estimation for color constancy: Why spatial-domain methods work and the role of the color distribution. *JOSA A*, 31(5):1049–1058, 2014. 1

[3] Hakki Can Karaimer and Michael S Brown. A software platform for manipulating the camera imaging pipeline. In *ECCV*, 2016. 1

[4] Rang M. H. Nguyen and Michael S. Brown. RAW image reconstruction using a self-contained sRGB-JPEG image with only 64 KB overhead. In *CVPR*, 2016. 1, 3, 4, 5

[5] Abhijith Punnappurath and Michael S. Brown. Spatially aware metadata for raw reconstruction. In *WACV*, 2021. 1, 3, 4, 5

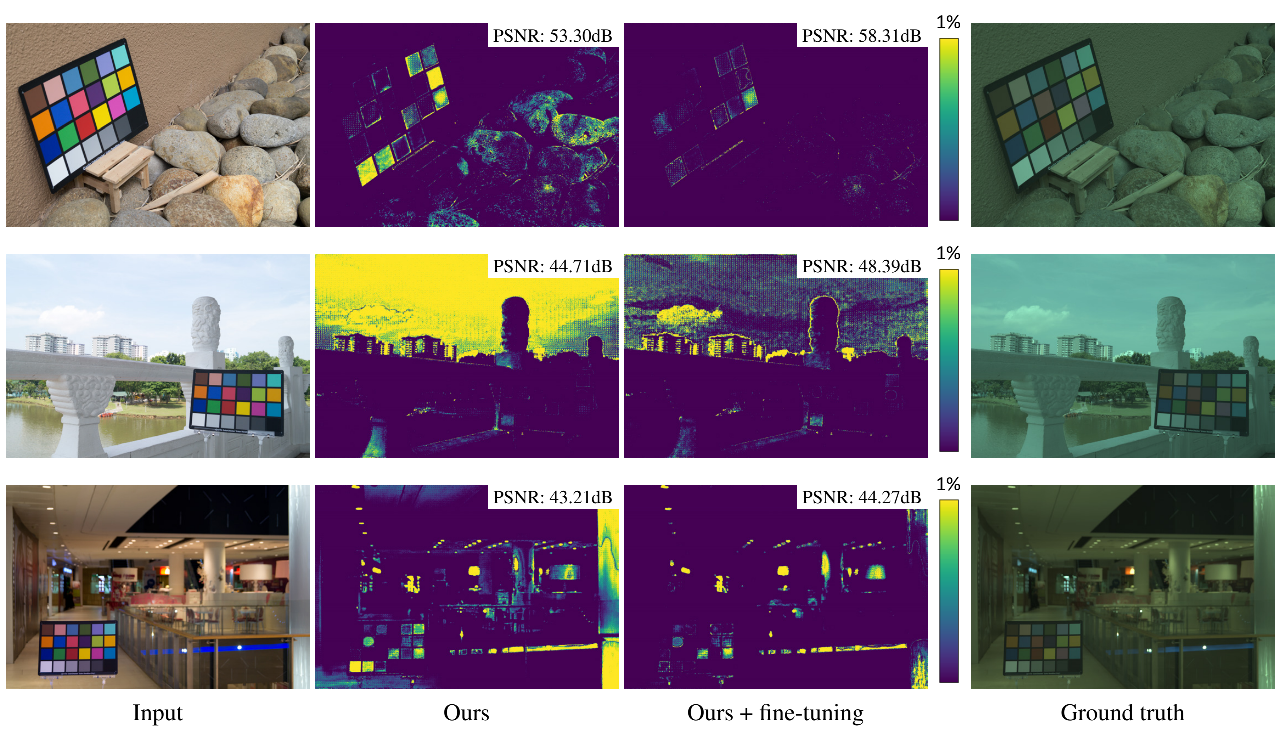

Figure S1. Qualitative evaluation of our online fine-tuning.

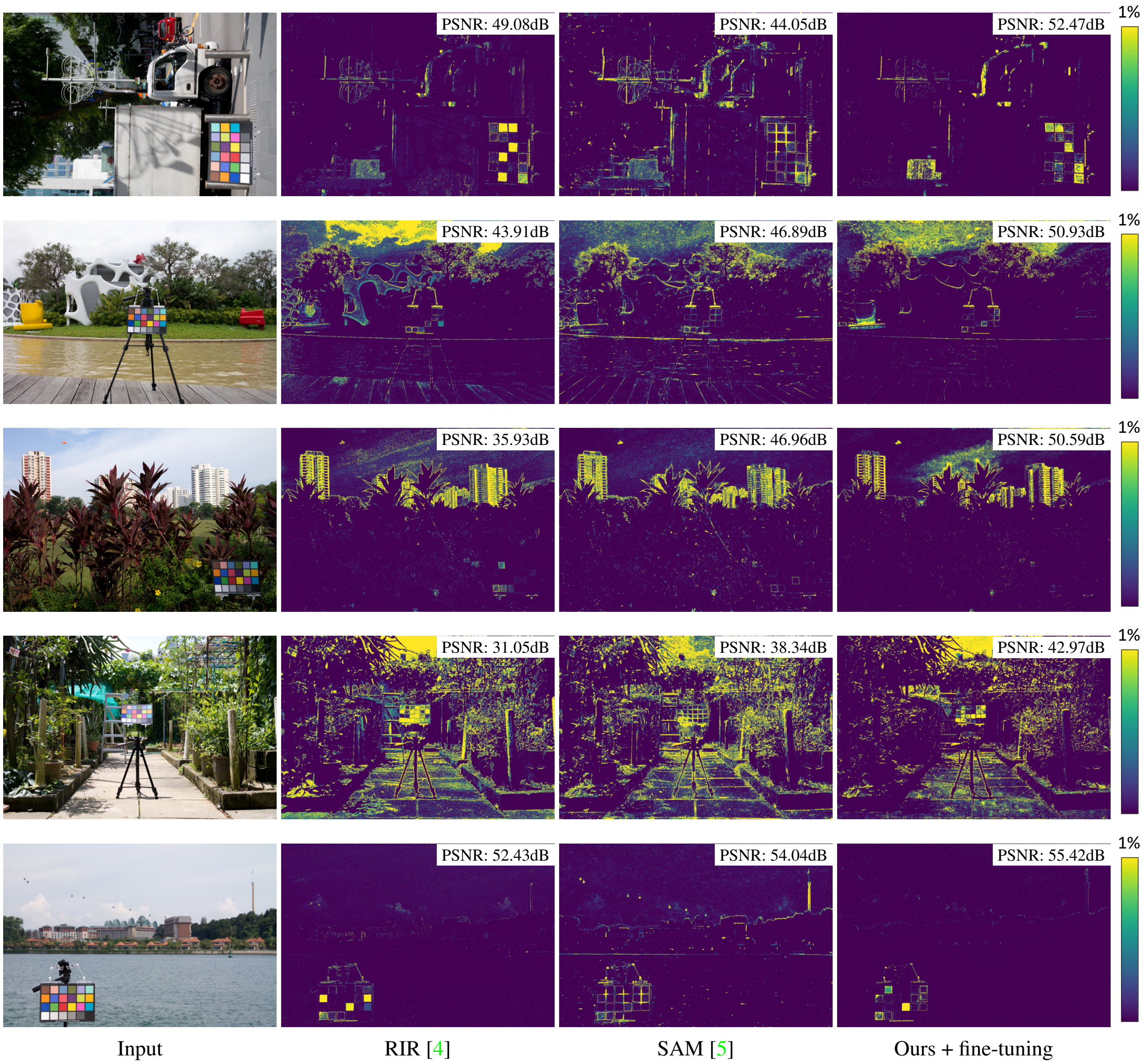

Figure S2. Qualitative comparison on Samsung NX2000.

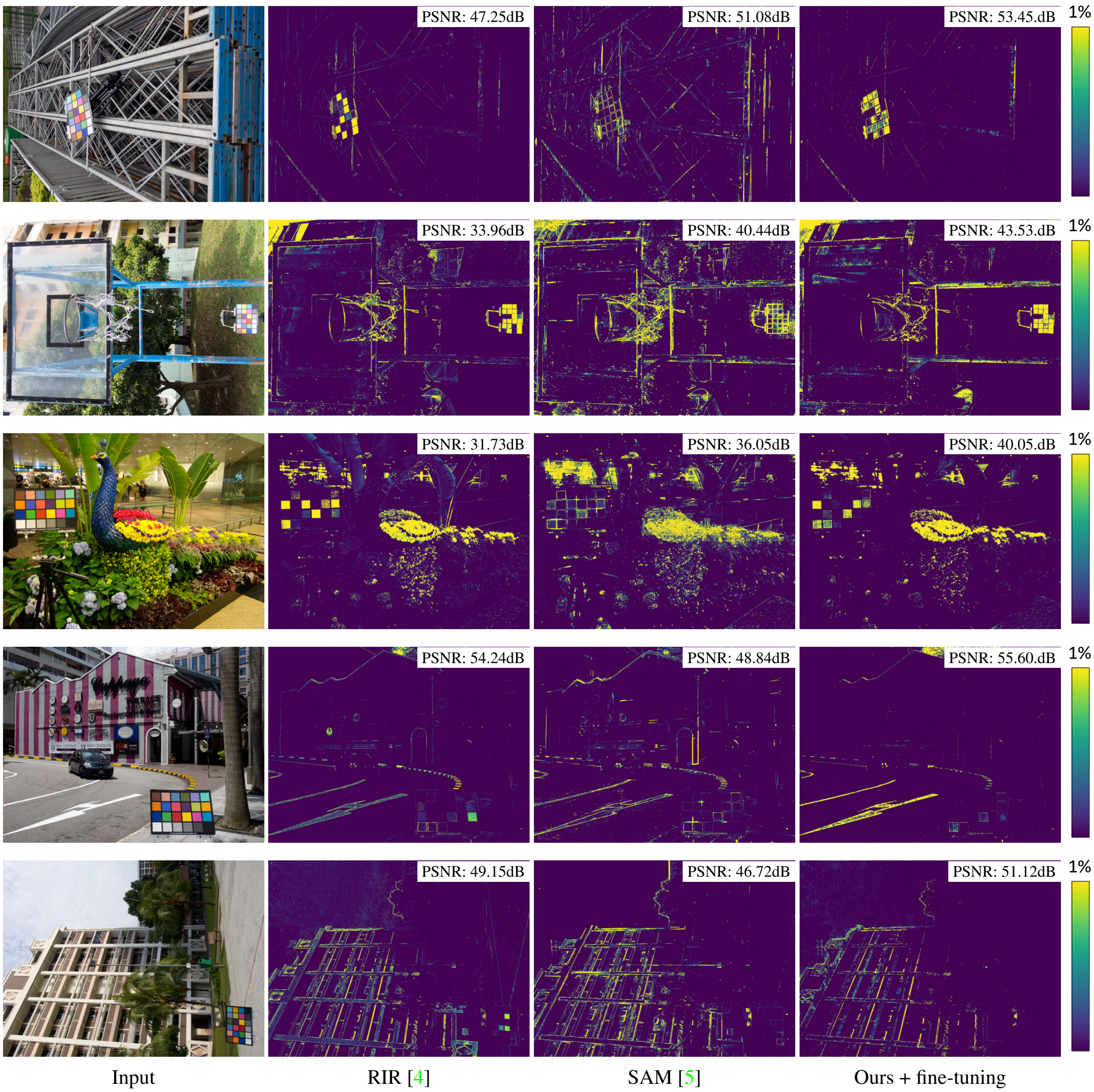

Figure S3. Qualitative comparison on Olympus E-PL6.

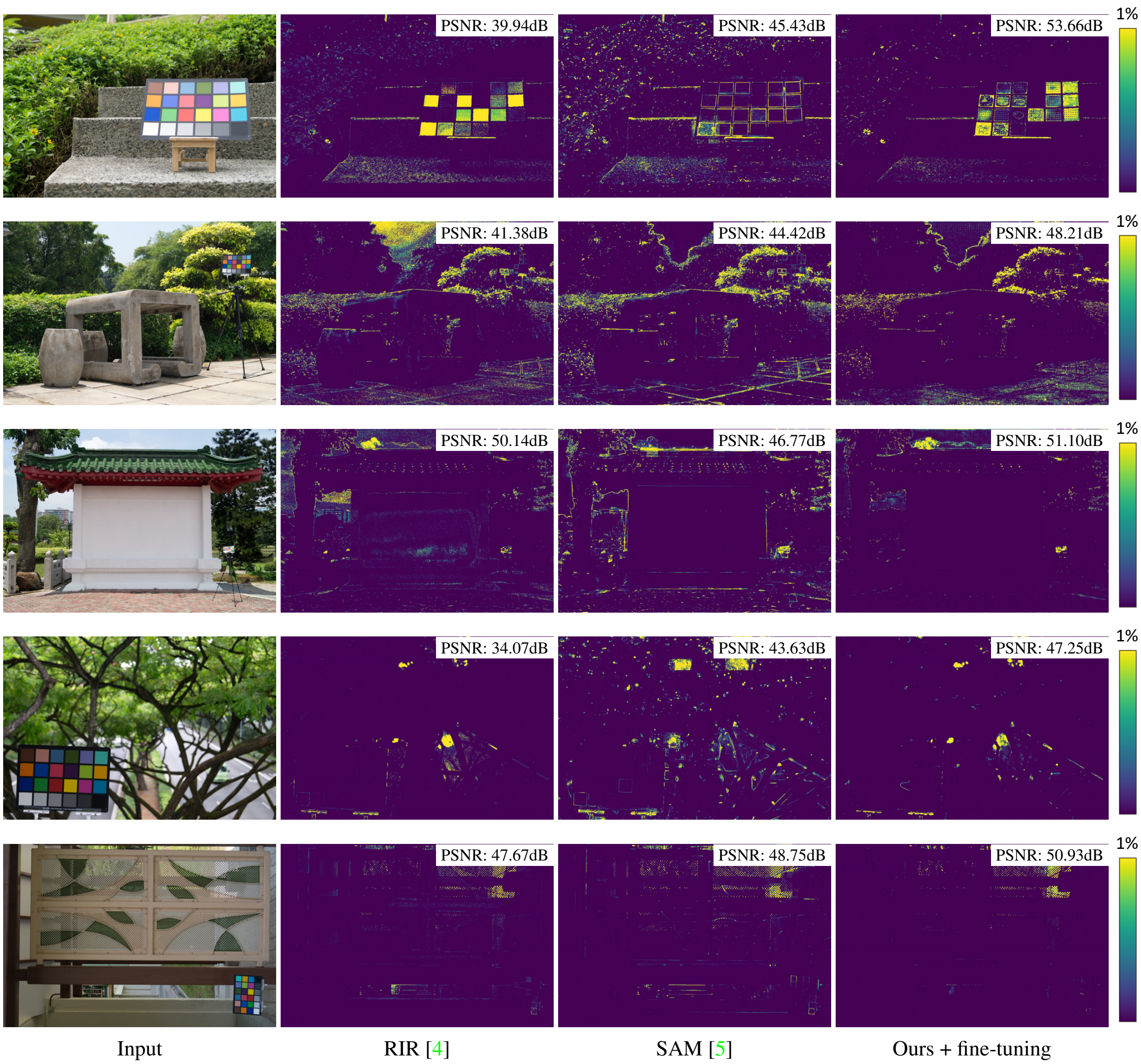

Figure S4. Qualitative comparison on Sony SLT-A57.

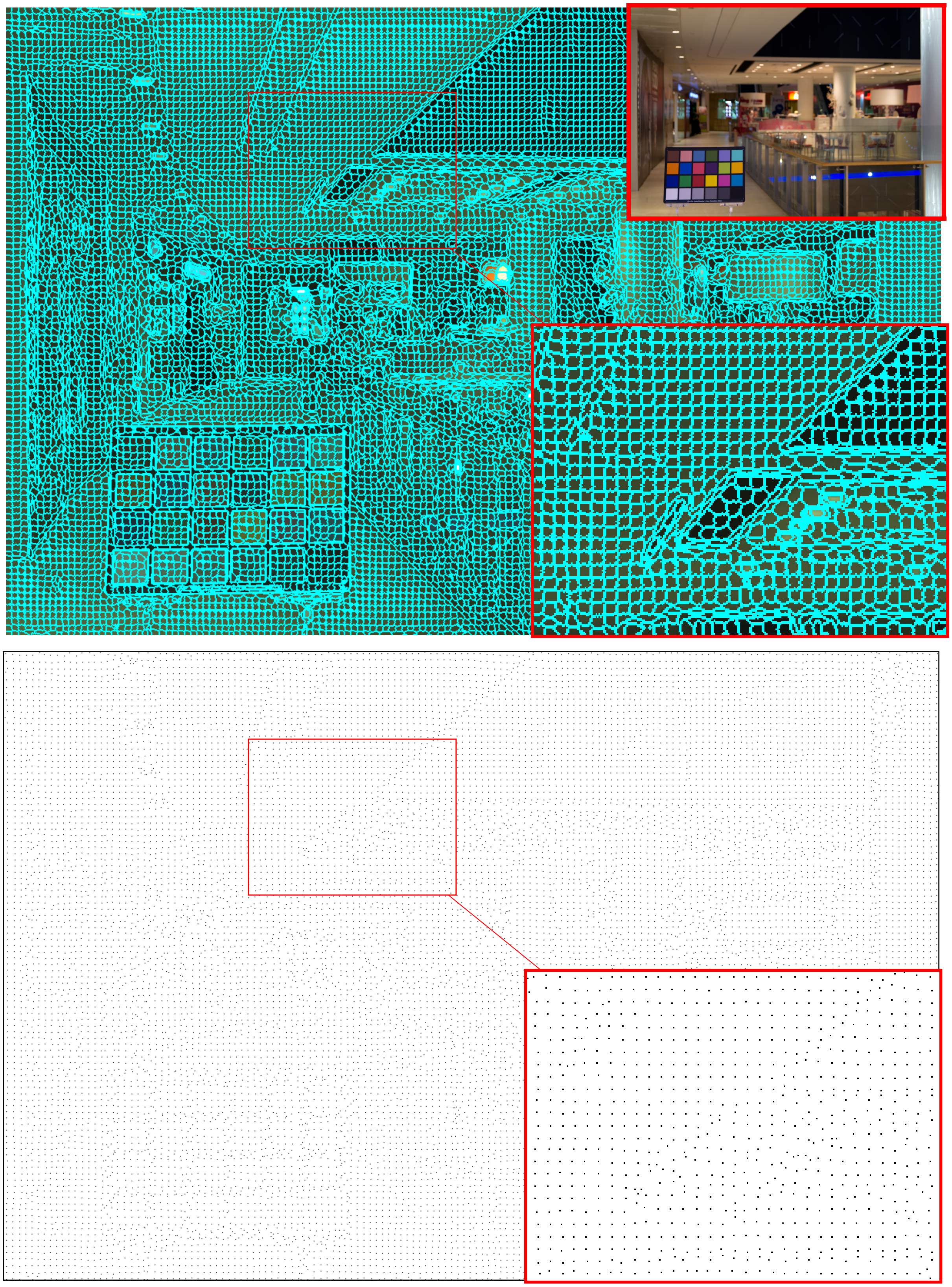

Figure S5. Visualization of learned superpixels and sampling mask.

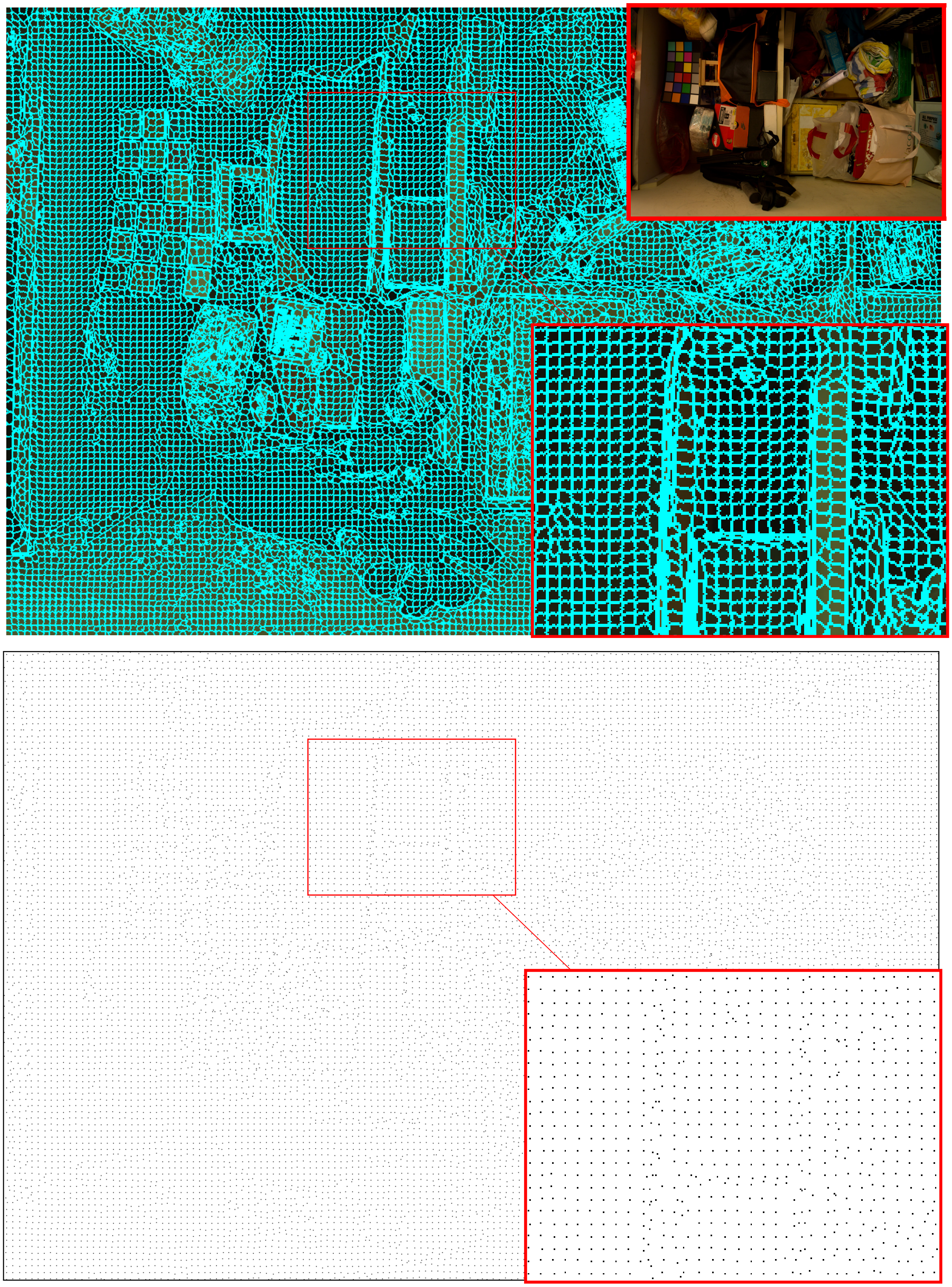

Figure S6. Visualization of learned superpixels and sampling mask.

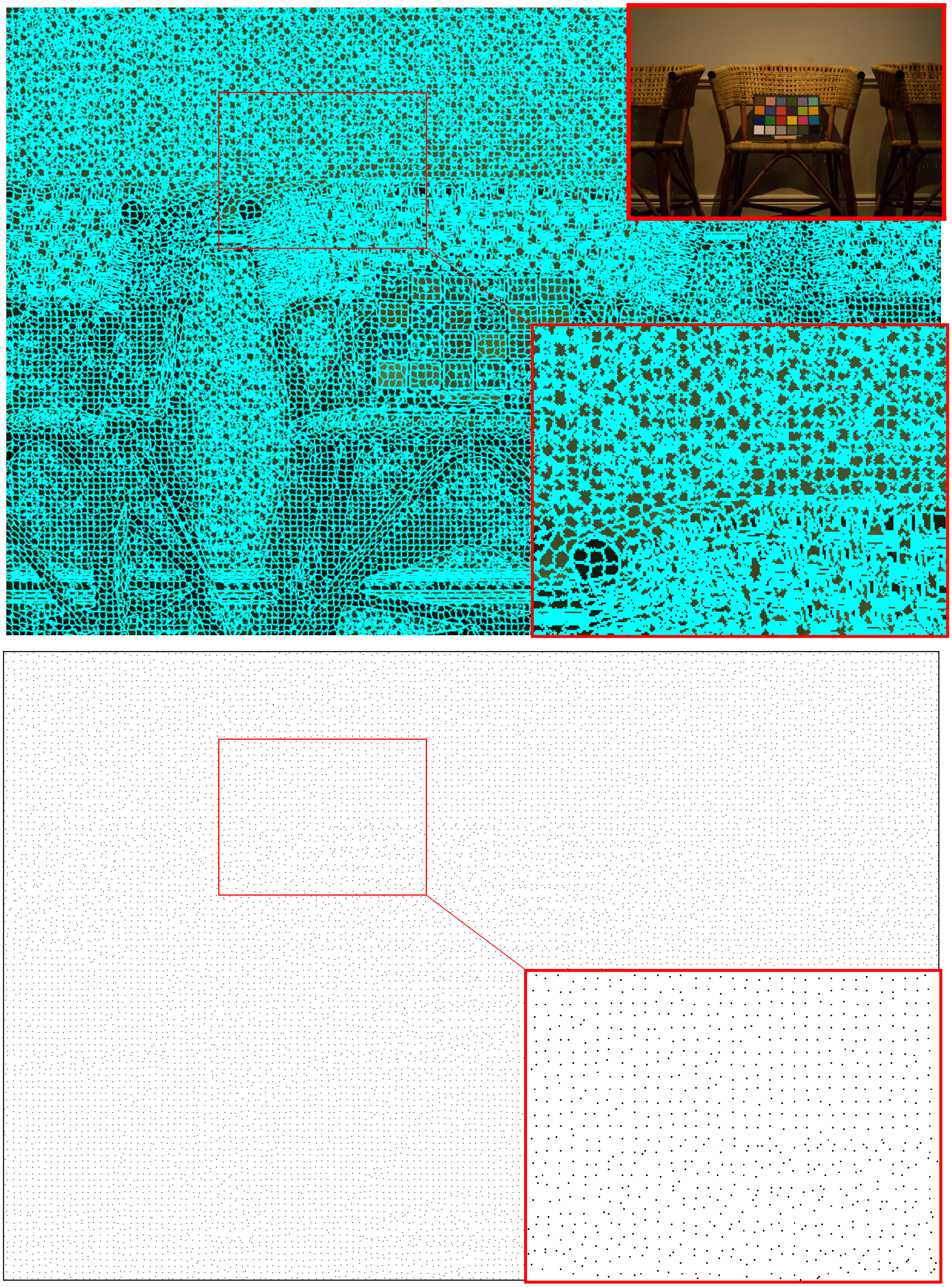

Figure S7. Visualization of learned superpixels and sampling mask.