SUPPLEMENTARY MATERIAL
Day-to-Night Image Synthesis for Training Nighttime Neural ISPs

Abhijith Punnappurath[1] Abdullah Abuolaim[2*] Abdelrahman Abdelhamed[1]
Alex Levinshtein[1] Michael S. Brown[1]
[1]Samsung AI Center – Toronto [2]York University
{abhijith.p,a.abdelhamed,alex.lev,michael.b1}@samsung.com, abuolaim@eecs.yorku.ca

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

S1. Illumination sampling

In our proposed day-to-night image synthesis framework, one of the main steps is to relight the image with night illuminants as explained in Section 3 of our main paper. To this end, we first construct a night time illuminant dictionary $\mathcal{L}$ by imaging gray cards under different nighttime illuminations. The blue markers in the plot of Fig. S1 show the $[\frac{r}{g}, \frac{b}{g}]$ chromaticity values of these ground truth illuminations recorded by the gray card. To apply the relighting step of our pipeline, we first fit a 2D multivariate Gaussian distribution of joint chromaticity values around our database of night illuminations $\mathcal{L}$, and then sample from this distribution, as described in Equations (1) and (2) of our main paper. The red markers in the plot show a few such randomly drawn samples. We use these samples to locally relight the scene and generate a synthetic nighttime image.

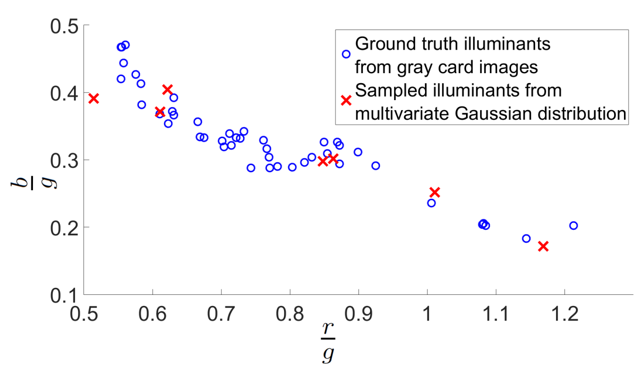

Figure S1. To relight the scene using our proposed day-to-night image synthesis framework, we draw random samples (red markers) from a 2D multivariate Gaussian distribution of joint $[\frac{r}{g}, \frac{b}{g}]$ chromaticity values fit on a database of night illuminations (blue markers) measured using a gray card.

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

S2. Comparison to burst denoising

For each scene in our nighttime dataset, we have a burst of 10 frames each at ISOs 1600 and 3200. Our experiments in the main paper were performed using only the first image in the high ISO bursts. Here, we perform a comparison to a burst denoising pipeline using all 10 frames. Since our images are already aligned, we directly average the Bayer frames. Then, we render the averaged Bayer image through the software ISP in [1]. Note that this is an idealized burst denoising pipeline that benefits from perfect alignment, which is not available in practice. Quantitative results are presented in Table S1. For ease of comparison, the results of our method are reproduced from the main paper. It can be observed that we outperform the burst denoising pipeline by a sound margin. Qualitative comparisons are provided in Fig. S2.

Table S1. Comparison to burst denoising.

Method	ISO 1600		ISO 3200	
	PSNR	SSIM	PSNR	SSIM
Burst denoising	34.91	0.8021	31.39	0.6526
Ours	37.41	0.9368	35.70	0.9142

S3. Additional qualitative results

We provide additional qualitative results of our neural ISP tasks without and with noise in Fig. S3 and Fig. S4. These figures extend Fig. 5 and Fig. 6 of our main paper, respectively.

S4. Ablation on the loss function

A comparison between L2 and L1 loss (proposed) is presented in Table S2 for our day-to-night model. PSNR

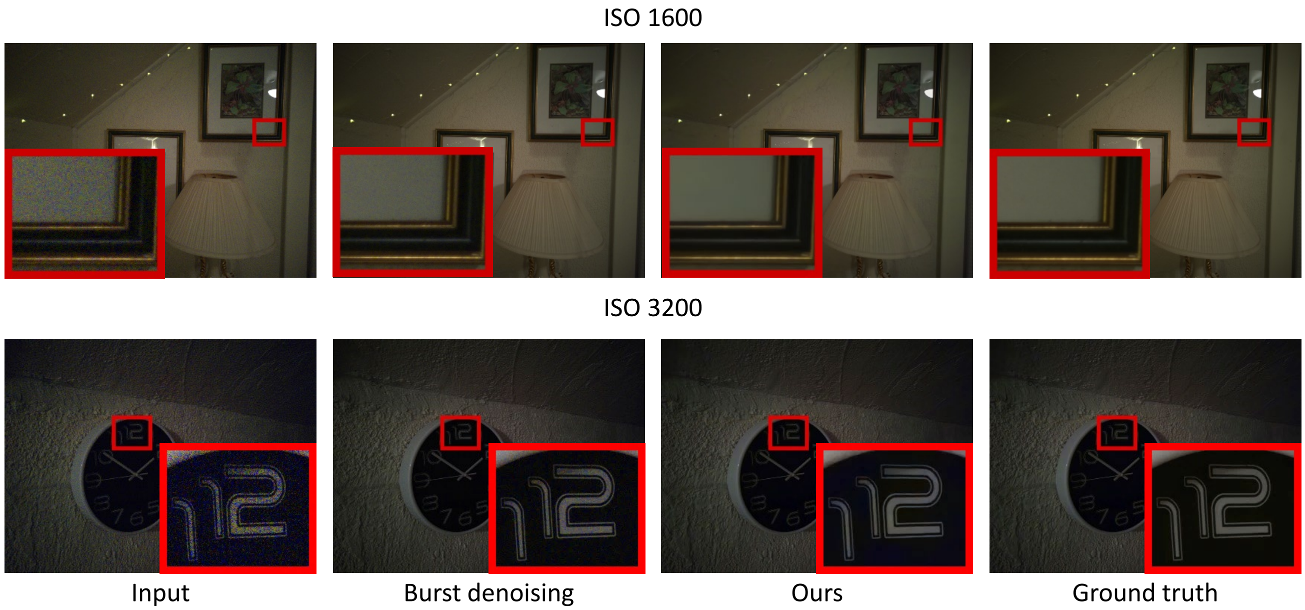

Figure S2. Comparison to burst denoising.

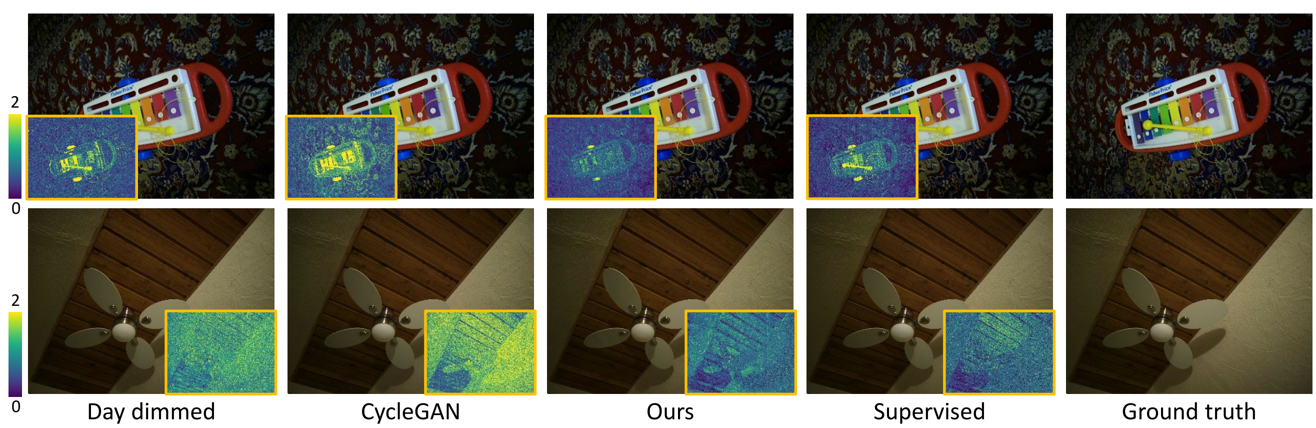

Figure S3. Additional qualitative results for our neural ISP task assuming noise-free inputs. Inset shows ΔE [3] error map.

(dB) and SSIM values are reported. L1 loss results have been reproduced from Table 1 (No noise) and Table 2 (ISO 1600/3200) of our main paper. As seen from the results, the proposed L1 loss is generally more accurate than L2.

Table S2. An ablation on the loss function.

Loss	No noise	ISO 1600	ISO 3200
L2	45.15 / 0.9887	37.10 / 0.9348	35.41 / **0.9157**
L1	**45.28 / 0.9893**	**37.41 / 0.9368**	**35.70** / 0.9142

S5. Comparison with EnlightenGAN

In the main paper, we had compared our method to CycleGAN [4]. In Table S3, we also report the results of EnlightenGAN [2], another popular image-to-image translation technique. We used the same training setup as used for CycleGAN in the main paper. PSNR (dB) and SSIM values are reported. Our results have been reproduced from Table 1 (No noise) and Table 2 (ISO 1600/3200) of our main paper. It can be observed that we outperform EnlightenGAN by a sound margin.

Table S3. Comparison with EnlightenGAN [2].

Method	No noise	ISO 1600	ISO 3200
EnlightenGAN	38.95 / 0.9652	35.24 / 0.9203	32.63 / 0.8879
Ours	**45.28 / 0.9893**	**37.41 / 0.9368**	**35.70 / 0.9142**

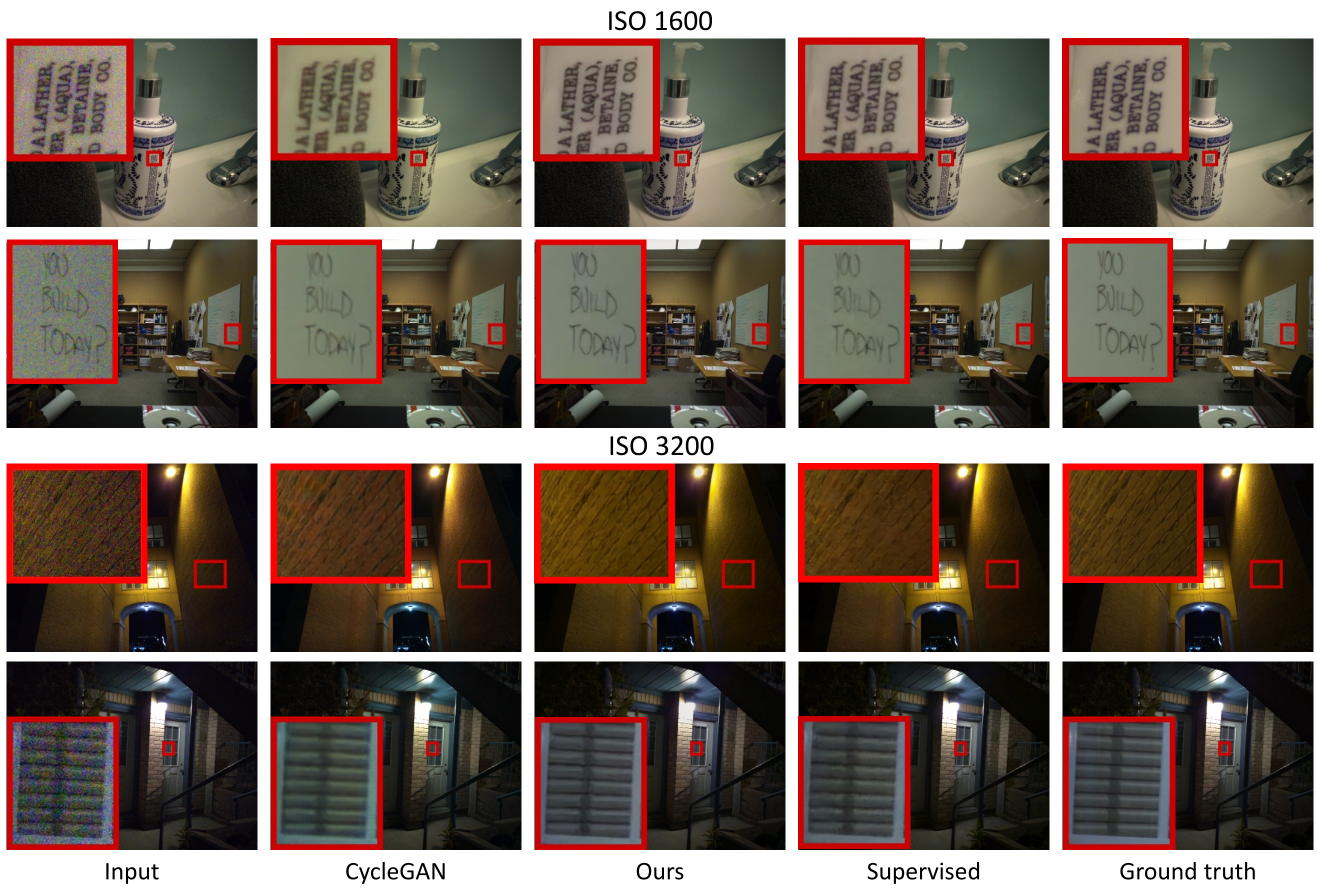

Figure S4. Additional qualitative results for our neural ISP task with real noisy inputs. Inset shows zoomed-in regions.

References

[1] Abdelrahman Abdelhamed, Stephen Lin, and Michael S. Brown. A high-quality denoising dataset for smartphone cameras. In *CVPR*, 2018. 1

[2] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. EnlightenGAN: Deep light enhancement without paired supervision. *TIP*, 30:2340–2349, 2021. 2

[3] Gaurav Sharma and Raja Bala. *Digital Color Imaging Handbook*. CRC Press, 2nd edition, 2013. 2

[4] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*, 2017. 2