Supplementary Material
Edit-aware RAW Reconstruction

Abhijith Punnappurath Luxi Zhao* Ke Zhao*
Hue Nguyen Radek Grzeszczuk Michael S. Brown
AI Center–Toronto, Samsung Electronics
{abhijith.p, lucy.zhao, k.zhao, h.nguyen1, radek.g, michael.b1}@samsung.com

This supplementary material provides additional experiments and analyses that complement the main paper. In Section S1, we show more qualitative examples of adding our edit-aware loss to different RAW reconstruction frameworks and provide implementation details for the UNet-based model. Section S2 explores target-edit-aware fine-tuning using the CAM [7] method. In Section S3, we evaluate our approach on more diverse post-processing edits, such as dehazing and local tone mapping. Section S4 shows results on the NUS [3] dataset. Finally, Section S5 presents further ablation studies. Collectively, these results offer deeper insights into the effectiveness and versatility of our edit-aware loss.

S1. Additional results

Figs. S1 and S2 provide additional results of our edit-aware loss integrated into the RAW recovery methods in CAM [7] and RAW Diffusion [9]. These figures extend Figs. 3 and 4 of our main paper. Across these examples, our method consistently reduces color and tone discrepancies, producing reconstructions that better reflect the applied edits.

Fig. S3 shows similar experiments on the metadata-assisted UNet [10] model. Compared to the baseline UNet, which has banding artifacts and color shifts, incorporating our edit-aware loss improves both tonal consistency and color fidelity in the rendered sRGB outputs. The UNet was trained using the Adam optimizer [5] with a learning rate of 0.001, batch size of 32, and patch size of 304. Training ran for 50 epochs. The fine-tuning experiments in Section 4.1 of the main paper were conducted at a reduced learning rate of 0.0001. Following the original UNet design, the first convolutional block used 32 filters.

S2. Target-edit-aware fine-tuning on CAM [7]

In the main paper, we had demonstrated our method's target-edit-aware fine-tuning capabilities using the metadata-assisted UNet model. Here, we extend this experiment to the CAM [7] framework. At capture time, CAM employs a learned sampler that generates a binary sampling mask $\mathbf{m}$ to select a set of RAW pixels. The mask $\mathbf{m}$ and the RAW samples $\mathbf{x_s}$ are saved as metadata, where the sample map $\mathbf{x_s}$ is computed by simply multiplying $\mathbf{m}$ with the RAW image $\mathbf{x}$ as $\mathbf{x_s} = \mathbf{m} \odot \mathbf{x}$, where $\odot$ denotes element-wise multiplication [7]. During inference, the reconstruction network receives the sRGB image $\mathbf{y}$, the mask $\mathbf{m}$, and the RAW sample map $\mathbf{x_s}$ as input producing a reconstructed RAW image $\hat{\mathbf{x}} = f_\theta(\mathbf{y}, \mathbf{m}, \mathbf{x_s})$. In the original CAM implementation, the fine-tuning loss is computed in RAW space over the sampled locations as:

$$\mathcal{L}_{\text{RAW}-\text{FT}} = \|\mathbf{m} \odot (\mathbf{x} - \hat{\mathbf{x}})\|_2^2. \tag{S1}$$

We augment this with our edit-aware loss applied to the same sampled RAW values as:

$$\mathcal{L}_{\text{sRGB}-\text{FT}} = \|\mathbf{m} \odot (g_\phi(\mathbf{x}) - g_\phi(\hat{\mathbf{x}}))\|_2^2, \tag{S2}$$

where g_ϕ, as described in the main paper, is our differentiable ISP with tunable parameters ϕ. Thus, the total fine-tuning objective becomes:

$$\mathcal{L}_{\text{total}-\text{FT}} = \mathcal{L}_{\text{RAW}-\text{FT}} + \lambda \mathcal{L}_{\text{sRGB}-\text{FT}}. \tag{S3}$$

Table S1 reports results for the same three scenarios described in the main paper: (i) baseline CAM fine-tuned on the test image, (ii) our edit-aware model fine-tuned on the test image with randomly sampled ϕ, and (iii) our model fine-tuned with ϕ fixed to the target edit. Even standard fine-tuning (scenarios i–ii) provides improvements over the baseline, while fixing ϕ to the target edit further enhances sRGB reconstruction fidelity. Qualitative examples in Fig. S4 show that target-edit-aware fine-tuning produces outputs that are more consistent with the intended photofinishing adjustments.

S3. Other edits

In the main paper, we had showed that training with our lightweight differentiable ISP yields strong generalization

*Equal contribution.

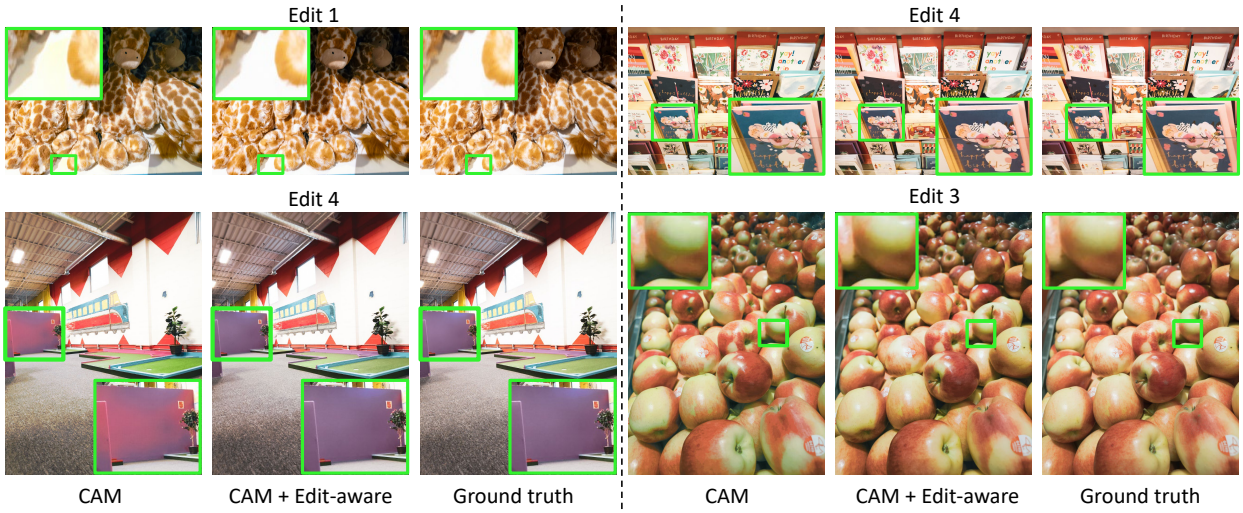

Figure S1. Additional qualitative results of adding our edit-aware loss to the RAW reconstruction method in CAM [7]. This figure complements Fig. 3 of the main paper. See Table 1 of the main paper for details of the edits. The baseline method exhibits noticeable color shifts, whereas our approach produces smoother tones and more accurate colors that better match the ground truth.

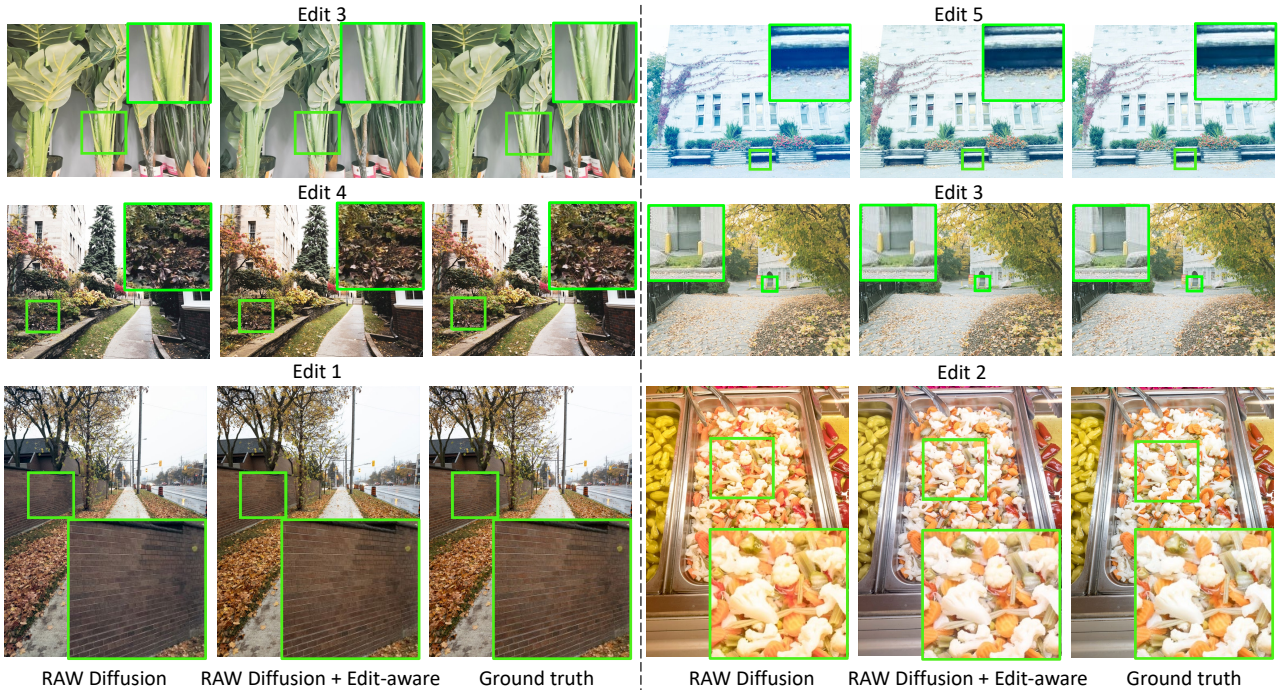

Figure S2. Additional qualitative results of adding our edit-aware loss to the RAW recovery method in RAW Diffusion [9]. These examples expand upon Fig. 4 of our main paper. See Table 1 of the main paper for details of the edits. The baseline model often shows inconsistent or shifted colors, whereas our method yields more faithful color reproduction and balanced tones that better align with the ground truth.

to more sophisticated software ISPs, such as Adobe Photoshop. Although the tone and color operations implemented in Photoshop are substantially more complex than those represented in our pipeline, the model trained with our edit-aware loss produces consistent results under these transformations. In Table 1 of the main paper, the selected edits were aligned with the same broad class of operations modeled by our loss pipeline. To further examine generalization beyond these settings, we evaluate our method on two additional edits—dehazing and local tone mapping—that are not modeled in our loss formulation.

Dehazing is applied using Photoshop with the dehazing

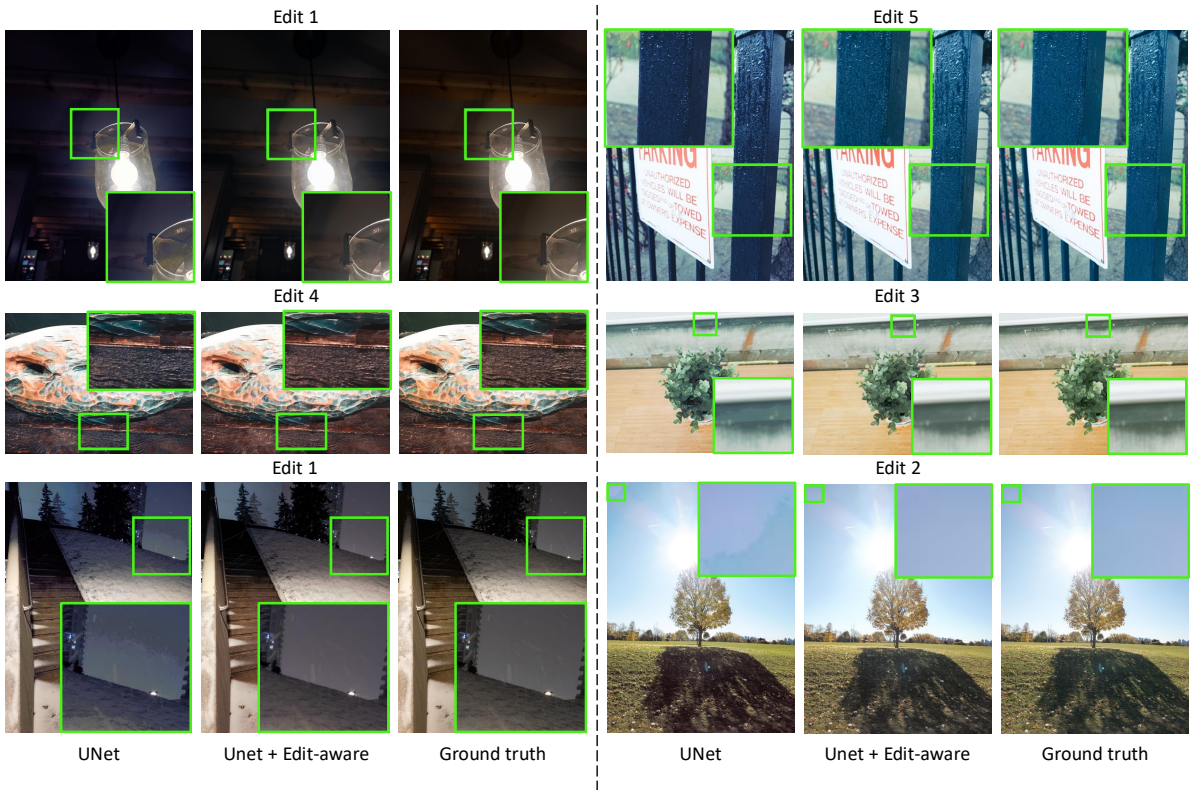

Figure S3. Results of adding our edit-aware loss to a UNet-based [10] RAW reconstruction method. See Table 1 of the main paper for details of the edits. Compared to the baseline, which exhibits banding artifacts and shifted colors due to collapsed tones, our method produces smooth tonal transitions without quantization artifacts and more accurate, faithful colors in the rendered sRGB outputs.

Method	EV +2			CCT 3000K			EV +2 & CCT 3000K		
	PSNR	SSIM	ΔE	PSNR	SSIM	ΔE	PSNR	SSIM	ΔE
CAM: FT image	28.71	0.8750	4.39	29.11	0.8774	4.99	28.59	0.8725	4.47
CAM + Edit-aware: FT image	30.57	0.8883	3.73	30.40	0.8885	4.34	30.30	0.8858	3.83
CAM + Edit-aware: FT image & FT edit	**30.95**	**0.8892**	**3.49**	**30.67**	**0.8896**	**4.18**	**30.73**	**0.8870**	**3.61**

Table S1. Quantitative results of fine-tuning (FT) the metadata-assisted CAM [7] RAW reconstruction model. Results are averaged over the 400 test images from the RAW smartphone dataset of [2]. Metrics include PSNR (dB), SSIM, and ΔE for the edited sRGB renderings. EV denotes exposure value and CCT represents correlated color temperature. The **best** and second-best results are highlighted.

Method	Dehazing			Local tone mapping		
	PSNR	SSIM	ΔE	PSNR	SSIM	ΔE
CAM	26.08	0.8258	8.28	25.12	0.8514	8.31
CAM + Edit-aware	**27.71**	**0.8565**	**6.22**	**26.58**	**0.8692**	**6.15**
RAWDiff	23.49	0.8246	10.09	23.18	0.8604	12.14
RAWDiff + Edit-aware	**24.59**	**0.8445**	**8.90**	**24.04**	**0.8638**	**10.60**
UNet	27.10	0.8376	**6.62**	25.76	0.8531	6.93
UNet + Edit-aware	**27.12**	**0.8547**	6.73	**26.29**	**0.8675**	**6.25**

Table S2. Quantitative evaluation of the proposed edit-aware loss integrated into different RAW reconstruction frameworks. Results are reported for CAM [7], RAW Diffusion [9], and a UNet–based [10] model, evaluated on 400 test images from the RAW smartphone dataset of [2]. Metrics include PSNR (dB), SSIM, and ΔE for the sRGB renderings. The **best** results are highlighted.

strength parameter in Adobe Camera RAW set to a value of 75. In the case of local tone mapping, we observed that Photoshop's output has relatively mild local tonal adjustments. To create a more challenging evaluation, we applied the contrast limited adaptive histogram equalization (CLAHE) algorithm [8], which is a classical local tone mapping technique, on the output of Photoshop. In particular, we first rendered the RAW images to 16-bit uncompressed sRGB images using Photoshop under the Adobe Camera RAW default settings, and then applied the CLAHE local tone mapping algorithm on the results.

Quantitative results for both edits are reported in Table S2. Across all reconstruction frameworks, incorporating the edit-aware loss leads to improved performance, indicat-

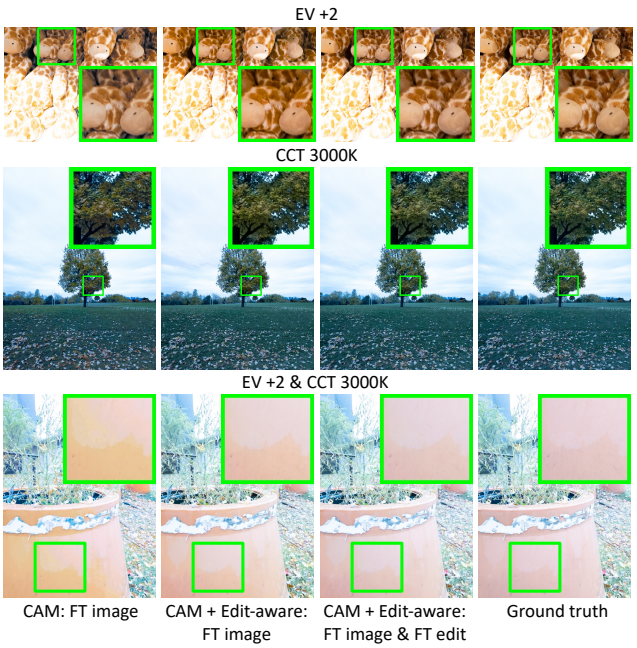

Figure S4. Examples of target-edit-aware fine-tuning applied to CAM [7]. Standard CAM fine-tuning refines the reconstruction using only RAW-domain supervision, while our edit-aware objective allows the model to incorporate information about post-processing adjustments. When the fine-tuning edit is matched to the target adjustment, the sRGB rendering of the reconstructed RAW better aligns with the specified edit.

Model	sRGB PSNR
Fixed pipeline	24.20
Large sampling	24.64
Ours	**25.15**

Table S3. Ablation study on sampling configurations. Results are reported using the UNet reconstruction model on a subset of 50 challenging images from the test split of the RAW smartphone dataset of [2]. We report sRGB PSNR (dB) for Edit 5 (see Table 1 of the main paper for edit details).

ing that the proposed training strategy enhances robustness even for operations not explicitly modeled during training. Representative qualitative examples for dehazing and local tone mapping are shown in Figs. S5 and S6, respectively. The baseline reconstructions suffer from color distortions and inconsistent tones, while our method delivers outputs that align closely with the applied edits.

S4. Results on NUS [3] dataset

We evaluated our method on three cameras from the NUS dataset [3] following the CAM [7] training/testing protocol. Results for two edits are shown (Edit A: 1.5 EV, Cool, 3500K; Edit B: 2.5 EV, Blossom, As-shot WB). Identical loss hyperparameters as in Section 4 of the main paper and a fixed $\lambda = 2$ (Eqn. 11) were used for all cameras. As shown in Table S4, our edit-aware method exhibits robust generalization across diverse cameras and edits.

Figure S5. Dehazing examples with our edit-aware loss integrated into different RAW reconstruction frameworks. These results show that our approach generalizes effectively to a dehazing edit that was never included or approximated during training.

S5. Additional ablations

Table 4 of the main paper had examined the contributions of individual loss components and the effect of parameter sampling. In those experiments, using all loss terms but without sampling yielded weaker performance than our full configuration that includes randomized sampling of all edit parameters. It is equally important, however, that the sampled parameter ranges remain consistent with edits that users typically apply. For instance, our chosen configuration draws exposure adjustments from a normal distribution centered at zero and samples illuminants near the image's ASN estimate, reflecting the fact that common edits tend to be moderate and anchored around default settings.

To further probe this sensitivity, we evaluate a setting in which the sampling distributions are broadened. Specifically, the exposure value is sampled from a uniform distribution $\varepsilon \sim \mathcal{U}(-3, 3)$; illuminant samples are drawn anywhere within the convex hull of the illuminant dictionary

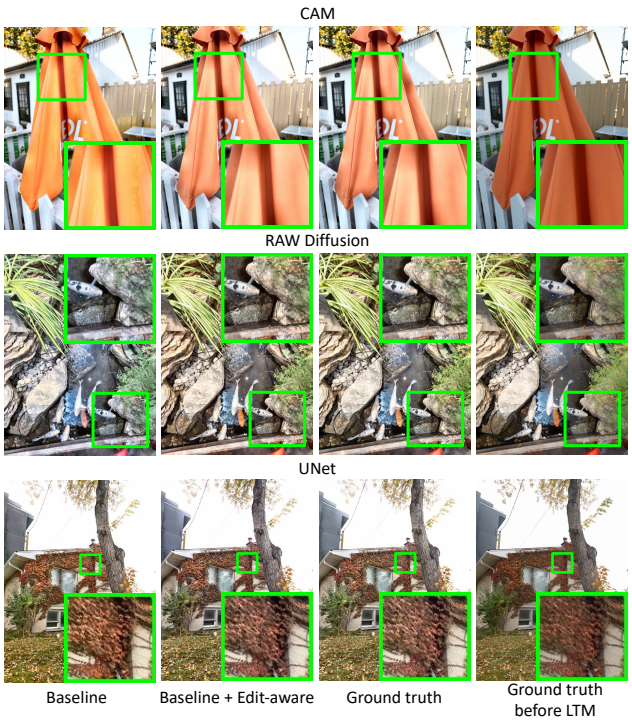

Figure S6. Local tone mapping results with our edit-aware loss integrated into different RAW reconstruction frameworks. Despite not being explicitly modeled during training, spatially varying tone adjustments are handled effectively by our approach.

rather than being restricted to the neighborhood of the ASN (l_1 norm $\leq$ 0.05 was used as the threshold for all experiments); and tone mapping uses polynomials with higher degree $d = 7$. These changes induce substantially stronger edits, and the resulting performance degradation is reflected in the second row of Table S3. For ease of comparison, the fixed-pipeline baseline (sampling disabled) and our configuration are reproduced from Table 4 of the main paper.

Table S5 shows an ablation on λ on the challenging 50-image set in Section 4.2 of the main paper. It can be observed that performance is not highly sensitive to λ.

Method	RAW	Edit A			Edit B		
	PSNR	PSNR	SSIM	ΔE	PSNR	SSIM	ΔE
Samsung	**46.18**	32.86	0.9633	3.95	32.80	0.9708	3.05
Samsung + EA	43.29	**33.73**	**0.9738**	**2.75**	**33.11**	**0.9769**	**2.53**
Olympus	**50.01**	33.61	0.9533	4.13	33.42	0.9556	3.24
Olympus + EA	48.76	**35.88**	**0.9699**	**3.37**	**34.66**	**0.9656**	**3.19**
Sony	**51.02**	34.10	0.9573	4.05	**33.26**	0.9582	3.94
Sony + EA	49.28	**34.55**	**0.9657**	**3.12**	**33.26**	**0.9626**	**2.99**

Table S4. Results of CAM [7] on three cameras (Samsung NX2000, Olympus E-PL6, Sony SLT-A57) from the NUS dataset [3]. Metrics include PSNR (dB), SSIM, and ΔE for the edited sRGB renderings. The **best** results are highlighted. Our edit-aware method (EA) generalizes across cameras and edits.

Method	CAM [7]					RAWDiff [9]				
λ value	w/o EA	1	2	4	8	w/o EA	1	2	4	8
sRGB PSNR	22.4	23.7	24.6	23.9	23.8	21.7	22.9	23.2	23.5	23.0

Table S5. Ablation study on λ. Results are reported on a subset of 50 challenging images from the test split of the RAW smartphone dataset of [2]. We report sRGB PSNR (dB) for Edit 5 (see Table 1 of the main paper for edit details).

Model	Edit 5		
	PSNR	SSIM	ΔE
Cyclic loss	20.54	0.7141	13.13
Our edit-aware loss	**27.13**	**0.8010**	**5.82**

Table S6. Ablation study on using a cyclic consistency loss versus our edit-aware loss. Results are reported using the CIE XYZ Net [1] model on a subset of 50 challenging images from the test split of the RAW smartphone dataset of [2]. We report metrics for Edit 5 (see Table 1 of the main paper for edit details).

As mentioned in the main paper, certain blind methods, such as [1, 4, 6, 11, 12], incorporate cyclic consistency constraints to encourage the reconstructed RAW to render back to the original sRGB input. This behaves similarly to using a fixed pipeline during training (row one of Table S3), which we observed to be less effective when evaluating under diverse edits. To further scrutinize our edit-aware loss, we conduct an experiment using a forward–inverse network that incorporates a cyclic consistency objective. For this study, we adopt the CIE XYZ Net [1] architecture. CIE XYZ Net consists of two sub-networks, a reverse-rendering module and a forward-rendering module, both jointly trained, where the forward-rendering sub-network is optimized using a supervised per-pixel sRGB fidelity loss to reproduce the input image. We first train a baseline model following this original setup. We then compare it against a variant in which we take only the reverse-rendering network and optimize it using the original per-pixel reconstruction loss and our edit-aware loss. As shown in Table S6, our edit-aware variant achieves substantially stronger performance under edits, highlighting that cyclic consistency alone does not encourage edit-compatible RAW recovery.

References

[1] Mahmoud Afifi, Abdelrahman Abdelhamed, Abdullah Abuolaim, Abhijith Punnappurath, and Michael S Brown. CIE XYZ Net: Unprocessing images for low-level computer vision tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9):4688–4700, 2021. 5

[2] Mahmoud Afifi, Luxi Zhao, Abhijith Punnappurath, Mohamed A Abdelsalam, Ran Zhang, and Michael S Brown. Time-aware auto white balance in mobile photography. In *ICCV*, 2025. 3, 4, 5

[3] 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, 4, 5

[4] Woohyeok Kim, Geonu Kim, Junyong Lee, Seungyong Lee, Seung-Hwan Baek, and Sunghyun Cho. ParamISP: Learned forward and inverse ISPs using camera parameters. In *CVPR*, 2024. 5

[5] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *ICLR*, 2014. 1

[6] Seonghyeon Nam and Seon Joo Kim. Modelling the scene dependent imaging in cameras with a deep neural network. In *ICCV*, 2017. 5

[7] Seonghyeon Nam, Abhijith Punnappurath, Marcus A Brubaker, and Michael S Brown. Learning sRGB-to-raw-RGB de-rendering with content-aware metadata. In *CVPR*, 2022. 1, 2, 3, 4, 5

[8] Stephen M Pizer, E Philip Amburn, John D Austin, Robert Cromartie, Ari Geselowitz, Trey Greer, Bart ter Haar Romeny, John B Zimmerman, and Karel Zuiderveld. Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing*, 39(3):355–368, 1987. 3

[9] Christoph Reinders, Radu Berdan, Beril Besbinar, Junji Otsuka, and Daisuke Iso. RAW-diffusion: RGB-guided diffusion models for high-fidelity RAW image generation. In *WACV*, 2025. 1, 2, 3, 5

[10] O. Ronneberger, P.Fischer, and T. Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015. 1, 3

[11] Yazhou Xing, Zian Qian, and Qifeng Chen. Invertible image signal processing. In *CVPR*, 2021. 5

[12] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. CycleISP: Real image restoration via improved data synthesis. In *CVPR*, 2020. 5