FastGaMer: Efficient GainMap Learning for Practical Inverse Tone Mapping ——Supplementary Material——

Yuanshen Guan[1*,†] Ruikang Xu[2*] Chang Chen[2*] Yinuo Liao[1]
Dehua Song[2] Fenglong Song[2] Zhiwei Xiong[1,‡]
[1]University of Science and Technology of China [2]Huawei Noah's Ark Lab
{guanys, yinuoliao}@mail.ustc.edu.cn, zwxiong@ustc.edu.cn
{xuruikang1, chenchang25, dehua.song, songfenglong}@huawei.com

The supplementary material is organized as follows:

- **Section 1** presents a mathematical proof showing why a single-channel gain map cannot extend color gamut.
- **Section 2** provides additional quantitative comparisons on real-world benchmark [6].
- **Section 3** presents visualizations of the intermediate results of FastGaMer.
- **Section 4** presents additional qualitative comparisons on the proposed dataset.
- **Section 5** details our dataset construction and visualization examples of both the synthetic and real-capture datasets.
- **Section 6** evaluates the robustness of FastGaMer to diverse tone mapping styles.
- **Section 7** provides additional ablation studies on module order and loss weight sensitivity.

1. Mathematical Analysis of Single-Channel Gain Maps

Proposition: Single-Channel GM Cannot Change Chromaticity. A single-channel gain map, applied as a per-pixel scalar multiplication on RGB, preserves chromaticity in the CIE Yxy space and therefore cannot expand the RGB color gamut of an image.

Setup. Let

$$S_{\mathrm{BT.709}}(a,b) = (R(a,b),\, G(a,b),\, B(a,b))^{\top}$$

be the linear BT.709 RGB value at pixel (a,b). Define the BT.2020 RGB value as

$$\mathbf{c}_{2020}(a,b) = M_{709}^{2020}\, S_{\mathrm{BT.709}}(a,b),$$

where $M_{709}^{2020} \in \mathbb{R}^{3\times 3}$ is the BT.709–BT.2020 transform.

The corresponding CIE XYZ coordinate is

$$\mathbf{v}(a,b) = \begin{pmatrix} v_X(a,b) \\ v_Y(a,b) \\ v_Z(a,b) \end{pmatrix} = M_{2020}^{XYZ}\, \mathbf{c}_{2020}(a,b),$$

with v_X, v_Y, and v_Z denoting the scalar components of $\mathbf{v}$.

Applying a single-channel gain map. Let $g(a,b)$ be the scalar gain. Ignoring offsets, the HDR RGB becomes

$$\mathbf{c}'_{2020}(a,b) = (1+g(a,b))\, \mathbf{c}_{2020}(a,b).$$

By linearity of the BT.2020–XYZ transform,

$$\mathbf{v}'(a,b) = M_{2020}^{XYZ}\, \mathbf{c}'_{2020}(a,b) = (1+g(a,b))\, \mathbf{v}(a,b).$$

*Equal contribution.
†Work done during an internship at Huawei Noah's Ark Lab.
‡Corresponding author.

Let $\alpha(a,b) = 1 + g(a,b) > 0$. Then the XYZ components satisfy

$$v'_X = \alpha v_X, \qquad v'_Y = \alpha v_Y, \qquad v'_Z = \alpha v_Z.$$

Chromaticity invariance in Yxy space. Chromaticity is defined as

$$x = \frac{v_X}{v_X + v_Y + v_Z}, \qquad y = \frac{v_Y}{v_X + v_Y + v_Z}.$$

After applying the gain:

$$x' = \frac{v'_X}{v'_X + v'_Y + v'_Z} = \frac{\alpha v_X}{\alpha(v_X + v_Y + v_Z)} = x,$$

and similarly

$$y' = y.$$

Thus every pixel satisfies

$$(x'(a,b),\, y'(a,b)) = (x(a,b),\, y(a,b)).$$

Consequence for color gamut. Let

$$\mathcal{C}_{\text{SDR}} = \{(x(a,b), y(a,b))\}, \qquad \mathcal{C}_{\text{HDR}} = \{(x'(a,b), y'(a,b))\}.$$

Since $(x', y') = (x, y)$ for all pixels,

$$\mathcal{C}_{\text{HDR}} = \mathcal{C}_{\text{SDR}}.$$

Hence,

1. All chromaticity points of the HDR image still lie within the original SDR chromaticity set.
2. On the BT.2020 chromaticity diagram, this set remains confined to the triangle spanned by the transformed BT.709 primaries.
3. A single-channel gain map only scales pixels along the luminance axis and cannot move any pixel outside the original primary triangle; therefore, it cannot achieve any genuine gamut expansion (e.g., from Rec. 709 to saturated BT.2020 primaries).

Conclusion. A single-channel gain map preserves chromaticities and therefore cannot expand the image's color gamut; only a multi-channel (per-channel) gain map can change chromaticity and achieve HDR/WCG expansion.

2. Additional Benchmark Comparisons

For a more comprehensive comparison, we additionally evaluate our method on the public single-channel GM dataset [6]. Following the protocol of [6], we retrain all baselines and our model on the real-world data and report performance on its corresponding test set. To adapt our three-channel output to the single-channel GM setting, we average the predicted channels to obtain a single-channel gain map. As shown in Tab. 1, our method clearly outperforms ITM-LUT [3] and other LUT-based baselines. Compared with heavy network-based methods such as GMNet [6] and FMNet [8], our model remains much more lightweight in both parameters and FLOPs, so a moderate performance gap to the strongest network-based baselines is expected and reflects the efficiency–quality trade-off.

3. FastGaMer Pipeline Visualization

We provide additional visualizations of intermediate outputs from different stages of FastGaMer to illustrate its working process in Fig. 1. The generated bilateral grid first performs fast local adaptation on the input SDR image, while remaining in the linear SDR domain. The learnable LUT then translates the adapted SDR to an initial GM. Next, the neural modulator corrects the global scale and shift for each channel to ensure accurate estimation. Finally, the normalized GM is combined with the estimated absolute dynamic range to produce the final GM. Since none of these operators involve high-resolution network processing, the runtime and computational overhead remain minimal.

Method	Params (M)	FLOPs (G)	Linear Domain			PQ Domain			HDR Domain	
			PSNR↑	SSIM↑	SRSIM↑	PSNR↑	SSIM↑	SRSIM↑	ΔE_{ITP}↓	HDR-VDP3↑
HDRUNet [1]	1.65	2946	33.16	0.9911	0.9866	37.39	0.9969	0.9984	5.732	9.8455
KUNet [7]	1.14	2666	33.32	0.9907	0.9880	33.78	0.9690	0.9979	9.062	9.6893
FMNet [8]	1.30	2935	32.62	0.9921	0.9828	39.76	0.9988	0.9986	4.551	9.8110
GMNet [6]	1.92	3155	33.94	0.9928	0.9896	40.21	0.9993	0.9989	4.034	9.8757
LUTwithGrid [4]	0.46	0.25	27.63	0.9342	0.9688	31.01	0.9843	0.9726	12.43	9.2448
SVDLUT [5]	0.27	0.02	28.27	0.9347	0.9732	30.98	0.9839	0.9731	11.26	9.3021
ITM-LUT [3]	0.60	41.9	31.66	0.9842	0.9871	37.90	0.9958	0.9985	5.421	9.6831
Ours	0.64	0.48	32.25	0.9917	0.9876	38.93	0.9974	0.9982	4.913	9.5736

Table 1. Quantitative comparisons on the real-world dataset from [6].

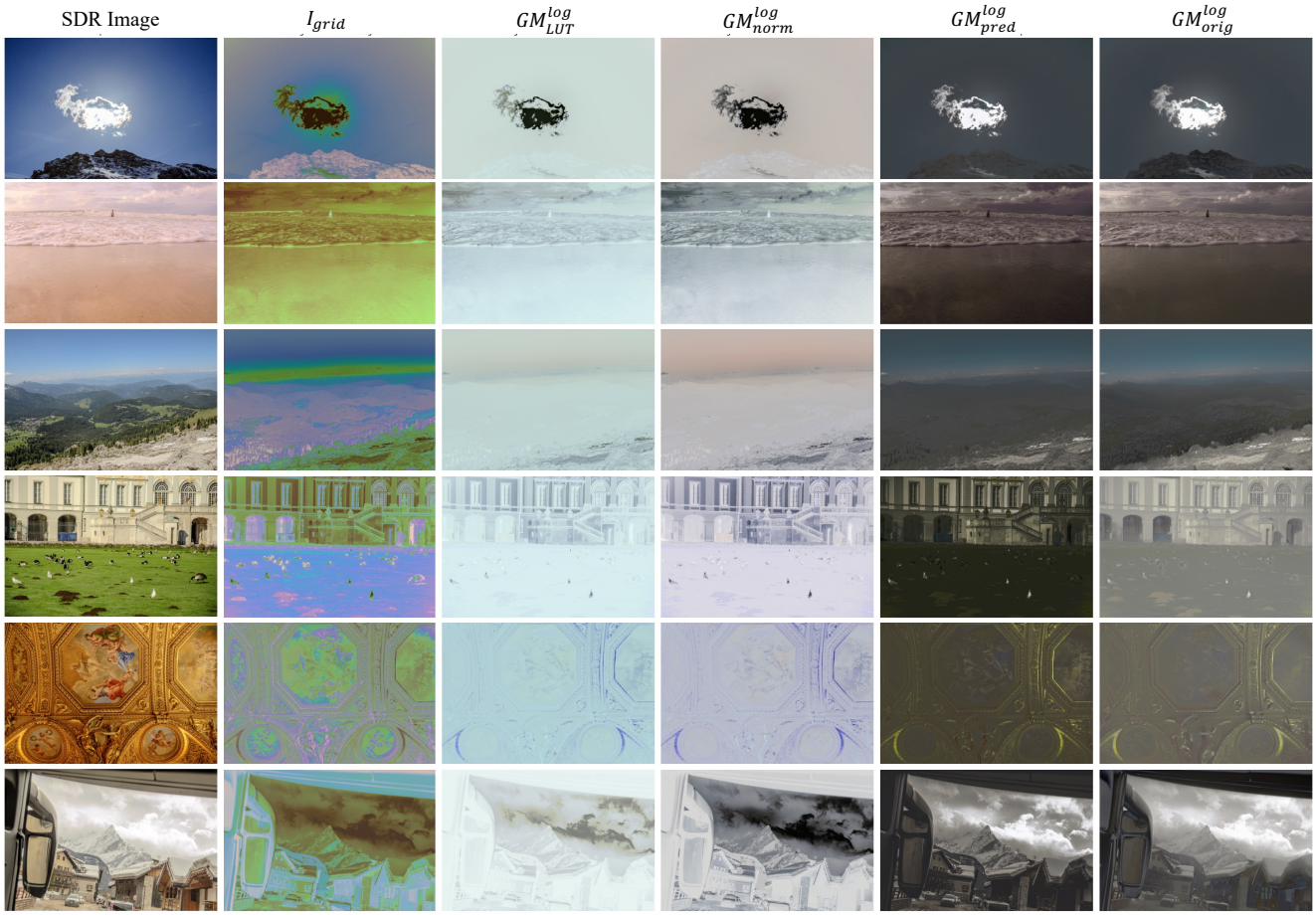

Figure 1. Intermediate results of FastGaMer. We illustrate the intermediate stages of the FastGaMer process, including bilateral grid processing, LUT translation, and neural modulation. For clear visualization, we also present the final predicted log-domain GM alongside the corresponding ground truth GM.

4. Additional Visual Comparison

We further provide more visual comparisons on both the synthetic dataset and the real-capture test set in Fig. 2 and 3, respectively. It is clear that FastGaMer is able to produce more accurate illuminance recovery and gamut expansion under

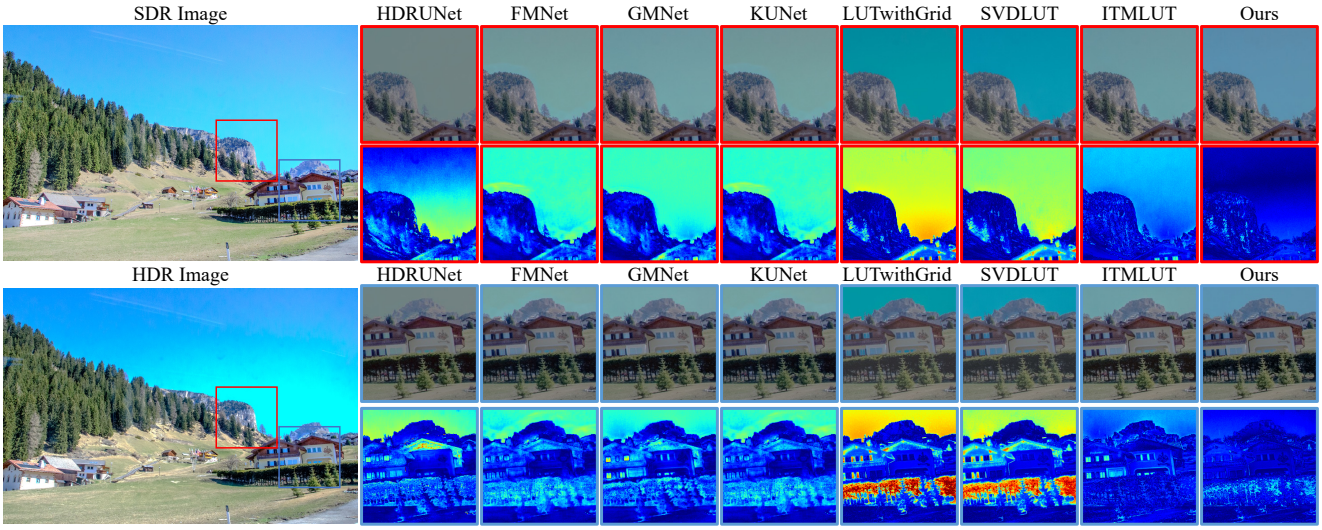

Figure 2. Visual comparison on the synthetic test set. Each column shows the predicted HDR result and its corresponding Y-channel error map (bottom) computed on PQ-encoded HDR. Our method produces more accurate and perceptually faithful results than existing methods.

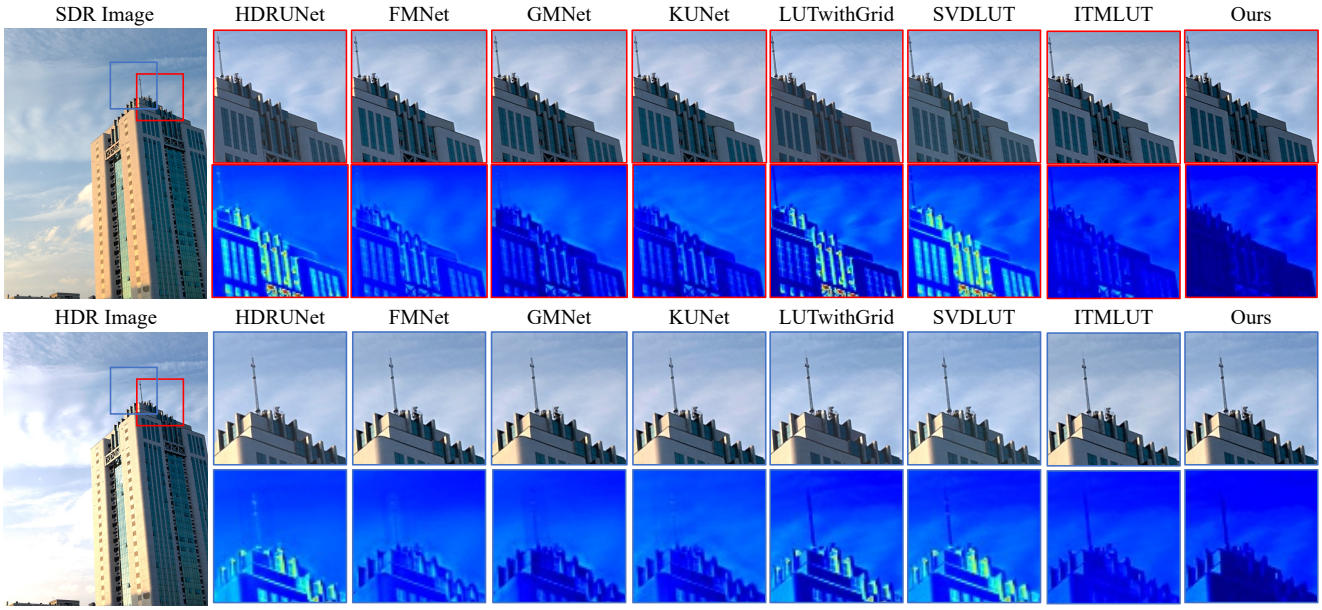

Figure 3. Visual Comparison on Real Capture Test Set. Each column shows the predicted HDR result and its Y-channel error map (bottom) computed on PQ-encoded HDR.

diverse conditions for practical usage.

5. Dataset Specification and Visualization

We visualize representative samples from the constructed synthetic dataset and the real-capture mobile dataset in Fig. 4 and 5, respectively. The synthetic dataset is derived from the RAISE dataset [2], which contains 8,156 high-resolution, uncompressed RAW images collected over three years by four photographers using three different cameras across more than 80 European locations. All images are guaranteed to be camera-native and unaltered. We process the RAW data in Adobe Photoshop to generate SDR–GM pairs by applying the "HDR" function with "Auto" settings and limiting the "HDR" parameter to three, thereby producing locally tone-mapped SDR base images. The SDR–GM pairs are saved in JPEG format

Figure 4. Visualization of the synthetic dataset. The synthetic dataset contains 8,000+ 4K SDR–GM pairs covering landscapes, nature, people, objects, and assorted indoor/outdoor scenes, delivering high-quality training data for practical ITM methods.

with the P3 color gamut.

6. Robustness to Diverse Tone Mapping

While our method and baselines are primarily designed for known tone mapping, FastGaMer achieves state-of-the-art performance on the real-capture test set (Tab. 2 in the main paper). To further evaluate generalization, we test on Adobe5K images retouched by five different experts, which undergo complex mappings and exhibit distinct styles. As shown in Fig. 6, our method effectively extends dynamic range without introducing artifacts, even in challenging cases such as recovering tiny eye highlights despite thumbnail aliasing during operator generation.

7. Additional Ablation Studies

7.1. Impact of Module Order

We conduct an ablation study to evaluate the impact of module ordering within FastGaMer. As shown in Tab. 2, the local-then-global order (Grid → LUT → Neural) achieves optimal results, confirming the design choice adopted in our pipeline. Intermediate feature maps visualizing module interactions are provided in Fig. 1 (Section 3).

Table 2. Impact of module order.

Order	**PQ-PSNR** ↑	ΔE_{ITP} ↓	**HDR-VDP-3** ↑
LUT → Grid → Neural **(Global → Local)**	29.77	25.18	8.804
Grid → Neural → LUT	29.84	24.98	8.831
Neural → Grid → LUT	29.63	25.16	8.798
Grid → LUT → Neural (Ours)	**30.02**	**24.61**	**8.859**

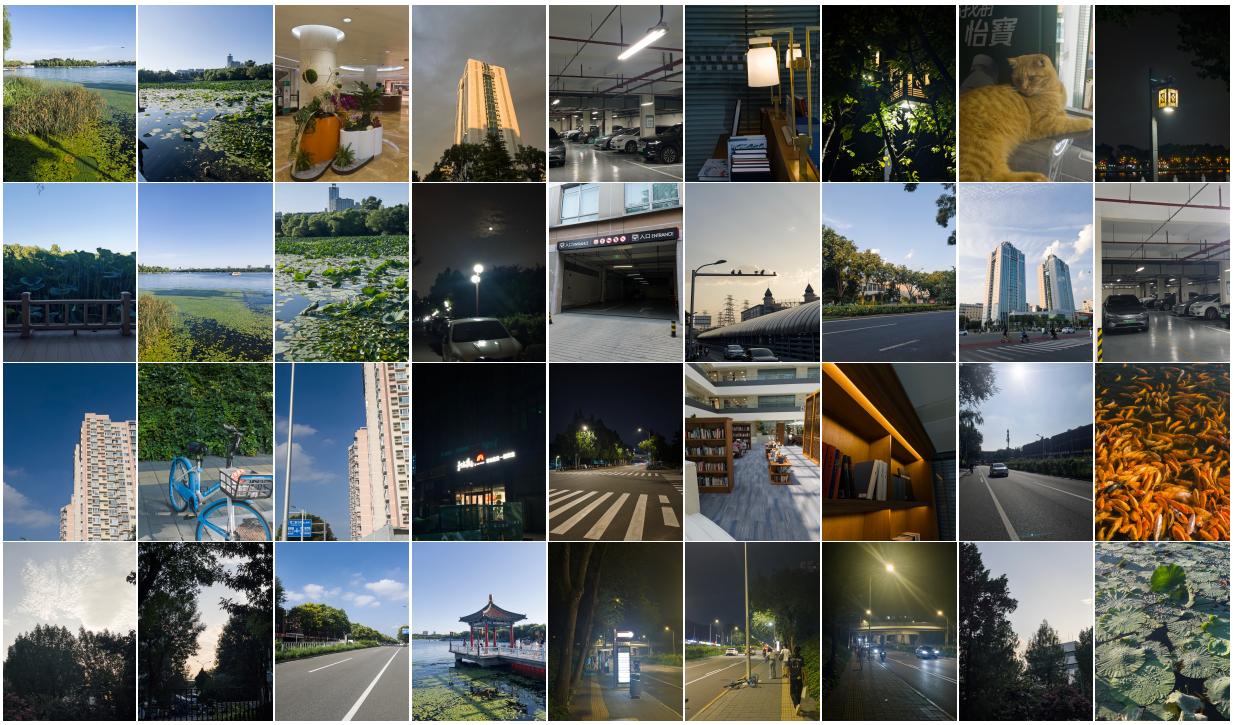

Figure 5. Visualization of the real-capture mobile dataset by iPhone 12 Pro Max with Adobe indigo software. The dataset spans diverse conditions, including nighttime and daytime, indoor and outdoor scenes, as well as static and dynamic content.

7.2. Loss Weight Sensitivity Analysis

We conduct a sensitivity analysis of the loss weights in Eq. (5) of the main paper, including the scale loss weight (λ_1) and LUT regularization weight (λ_2). As shown in Tab. 3, our default setting (λ_1=3.0, λ_2=0.1) yields the best balance between recovery accuracy and perceptual quality.

Table 3. Sensitivity analysis of loss weights.

λ_1 **(Scale)**	λ_2 **(Reg)**	**PQ-PSNR**↑	ΔE_{ITP}↓	**HDR-VDP-3**↑
0.1	0.1	29.35	25.37	8.353
1.0	0.1	29.76	24.83	8.827
5.0	0.1	29.98	24.73	8.854
3.0	0.0	29.96	24.88	8.845
3.0	1.0	29.74	25.03	8.832
3.0 (Default)	**0.1 (Default)**	**30.02**	**24.61**	**8.859**

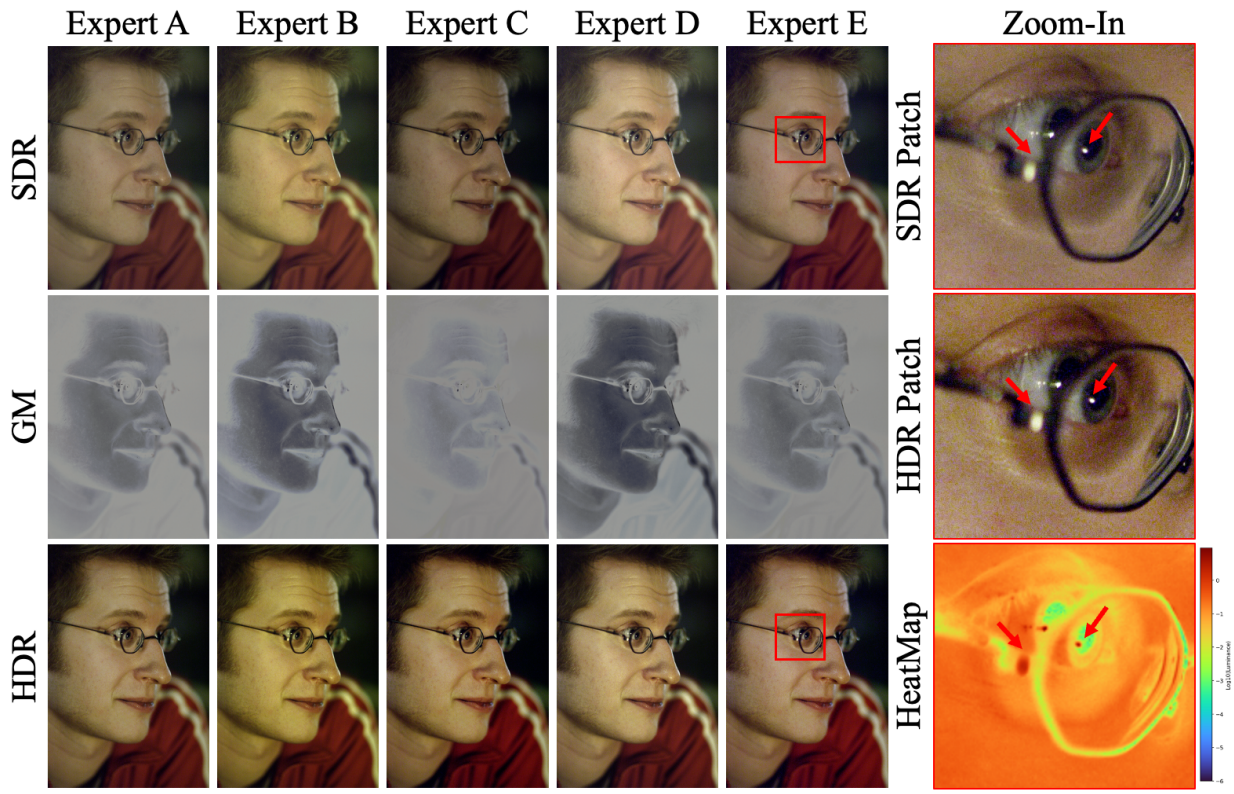

Figure 6. Visual results on the Adobe5K dataset. Our method generalizes well to diverse tone-edited SDR images, effectively extending dynamic range without artifacts.

References

[1] Xiangyu Chen, Yihao Liu, Zhengwen Zhang, Yu Qiao, and Chao Dong. Hdrunet: Single image hdr reconstruction with denoising and dequantization. In *CVPRW*, 2021. 3

[2] Duc-Tien Dang-Nguyen, Cecilia Pasquini, Valentina Conotter, and Giulia Boato. Raise: A raw images dataset for digital image forensics. In *ACM MMSys*, 2015. 4

[3] Cheng Guo, Leidong Fan, Qian Zhang, Hanyuan Liu, Kanglin Liu, and Xiuhua Jiang. Redistributing the precision and content in 3d-lut-based inverse tone-mapping for hdr/wcg display. In *ACM SIGGRAPH CVMP*, 2023. 2, 3

[4] Wontae Kim and Nam Ik Cho. Image-adaptive 3d lookup tables for real-time image enhancement with bilateral grids. In *ECCV*, 2024. 3

[5] Wontae Kim, Keuntek Lee, and Nam Ik Cho. Lightweight and fast real-time image enhancement via decomposition of the spatial-aware lookup tables. In *ICCV*, 2025. 3

[6] Yinuo Liao, Yuanshen Guan, Ruikang Xu, Jiacheng Li, Shida Sun, and Zhiwei Xiong. Learning gain map for inverse tone mapping. In *ICLR*, 2025. 1, 2, 3

[7] Hu Wang, Mao Ye, Xiatian Zhu, Shuai Li, Ce Zhu, and Xue Li. Kunet: Imaging knowledge-inspired single hdr image reconstruction. In *IJCAI*, 2022. 3

[8] Gang Xu, Qibin Hou, Le Zhang, and Ming-Ming Cheng. Fmnet: Frequency-aware modulation network for sdr-to-hdr translation. In *ACM MM*, 2022. 2, 3