Gain-MLP: Improving HDR Gain Map Encoding via a Lightweight MLP

Supplementary Material

This supplementary material provides (1) ablation experiments to characterize the performance of our proposed method and related works in terms of reconstruction quality; (2) additional figures demonstrating the different tone mapping methods employed in the dataset; (3) an expanded results table demonstrating method performance on individual tone mapping strategies.

1. Ablations

In addition to network size, which is examined in the main text, MLP performance is dependent on optimization time. In Fig. 1, we test the MLP methods addressed in the main paper on the manually tone mapped images of Cyriac et al. [2] at different optimization times. In these experiments, network size is matched to the default settings of the embedded ITM baseline, MLP-ITM, since its performance sharply decreases outside of this setting. The results show that this network is roughly as computationally efficient as the proposed Gamma-MLP. When compared to Direct-MLP [13], Gain-MLP has higher performance at low optimization times but Direct-MLP sees consistent improvements with increased optimization time.

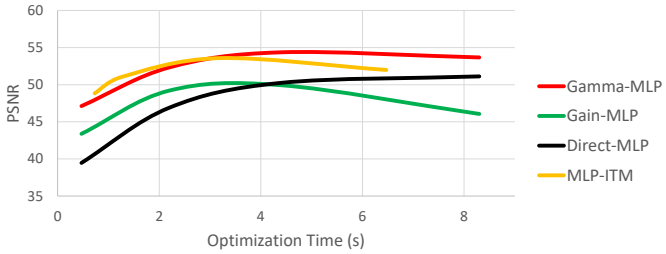

Figure 1. MLP performance as a function of optimization time on the manually tone-mapped data from the Cyriac et al. dataset [2]. All networks were set to match the metadata size of MLP-ITM in its default setting [14]. Gain-MLP has higher performance than Direct-MLP at low optimization times, but Direct-MLP improves across tested optimization times. Similarly, MLP-ITM has best performance at low optimization times but Gamma-MLP has an extended lattitude for improvement.

Next, we compare Gain-MLP performance with and without the SDR base image as input. As mentioned in the main paper, applying MLPs directly for compression [4, 5, 16] can still be computationally expensive to optimize, as this approach relies only on positional coordinates (x, y) to predict image values (r, g, b). Our application has a significant advantage, given our access to the SDR image to guide the reconstruction of the gain map. Table 1 shows how Gamma-MLP's performance reduces significantly if the input is only positional coordinates. This experiment also shows the value of the positional embedding (x,y) in addition to the SDR input. While a global (r,g,b) MLP can achieve relatively close reconstructions, local tone mapping operations can be accounted for with additional (x,y) input.

Model Inputs	PSNR ↑	SSIM ↑	HDR-VDP3 ↑
(x,y)	36.6	0.969	7.95
(r,g,b)	48.4	0.996	9.53
(x,y,r,g,b)	51.4	0.998	9.74

Table 1. Ablation of MLP (2-layer, 16 nodes) performance when predicting gamma map with just position information (x,y) or SDR pixel values (r,g,b) compared to having access to both for prediction (x,y,r,g,b). We report results on the manually tone-mapped dataset. The experiment demonstrates the benefit of the SDR guide image and the positional embedding in our application. While the image can be closely reconstructed with just a global transform of an (r,g,b) MLP, this experiment also demonstrates that the local operations described with positional embeddings (x,y) can have a large impact on reconstruction fidelity

Finally, in Table 2 the Gamma-MLP results are compared against the deep learning based inverse tone mapping and image translation methods tested by Chen et al. [1] on their 117-image UHD test set. We show that when trained on a particular tone mapper, these methods are outperformed in the reproduction of that function by standard Gain map encoding using HEIC, so they are not a viable replacement for gain map encoding.

2. Experiments

In the experimental section of the main paper, the dataset of Cyriac et al. [2] is employed to evaluate the performance of the proposed method (Gamma-MLP) against the existing framework. We augment this dataset for our quantitative comparisons by processing the HDR image through additional automatic tone mapping methods to complement the manual tone mapping results. In Fig. 3, the performance of the different tone mapping methods is evaluated qualitatively. The gain maps demonstrate that these represent significantly different reconstruction tasks, explaining the variation in their respective results. In addition, it is demonstrated that the automatically tone-mapped images represent challenging cases as they are more likely to saturate detail and amplify noise.

Since the bar for reconstruction quality in encoding ap-

Method	PSNR ↑	SSIM ↑
ResNet [8]	37.3	0.972
Pix2pix [9]	25.8	0.878
CycleGAN [18]	21.3	0.850
HDRNet [6]	35.7	0.966
CSRNet [7]	35.0	0.963
Ada-3DLUT [17]	36.2	0.966
Deep SR-ITM [10]	37.1	0.969
JSI-GAN [11]	37.0	0.969
AGCM+LE [1]	37.6	0.973
AGCM [1]	36.9	0.966
Gain-HEIC	39.2	0.972
MLP-ITM [14]	**41.6**	**0.988**
Gain-MLP	41.2	0.986
Gamma-MLP	41.5	**0.988**

Table 2. Using the 117 UHD image test dataset of Chen et al. [1], where HDR images are automatically converted to SDR using Youtube's HDR10 pipeline, the proposed MLP (Gain-MLP) is compared to deep learning methods for HDR generation (optimized on the associated training set). This experiment confirms that predictive methods do not meet the quality standards of the encoding application, even when trained on the same tone mapping function as the test set.

plications is very high it can be difficult to qualitatively differentiate competing results. In Fig. 2 we supplement main body Fig. 6 with ΔE_{00} error maps, which clip at a value of 10. These maps show several cases where the proposed method preserves image details where others fail (e.g., the sky of column 1, the legs of column 2, the head and sky of column 3 and the hands of the last column).

While the results in the main paper were averaged over all tone mapping methods from the dataset of Cyriac et al. and Chen et al. [1], Table 3 splits the results between these subsets. It can be observed that in the conventional approach, reconstruction quality varies significantly between tone mapping methods for the Cyriac dataset, but the proposed MLP is relatively consistent between these subsets. Comparing results from different HDR sources (between Cyriac et al. and Chen et al. datasets) the conventional approach maintains reconstruction quality but its size increases, while the proposed MLP has a lower quality reconstruction if the network size is maintained.

Finally, representative examples of the artifacts that can occur with the proposed MLP are demonstrated in Fig. 4. When large areas of detail are lost in the SDR image during the tone mapping process, more information is required to reconstruct the area, pushing the representation limits of the proposed and conventional methods.

References

[1] Xiangyu Chen, Zhengwen Zhang, Jimmy S Ren, Lynhoo Tian, Yu Qiao, and Chao Dong. A new journey from SDRTV to HDRTV. In *CVPR*, 2021. 1, 2, 5

[2] Praveen Cyriac, Trevor Canham, David Kane, and Marcelo Bertalmío. Vision models fine-tuned by cinema professionals for high dynamic range imaging in movies. *Multimedia Tools and Applications*, 80:2537–2563, 2021. 1, 5

[3] Frédéric Drago, Karol Myszkowski, Thomas Annen, and Norishige Chiba. Adaptive logarithmic mapping for displaying high contrast scenes. In *Computer Graphics Forum*, pages 419–426, 2003. 4, 5

[4] Emilien Dupont, Adam Golinski, Milad Alizadeh, Yee Whye Teh, and Arnaud Doucet. COIN: COmpression with Implicit Neural representations. In *ICLR*, 2021. 1

[5] Emilien Dupont, Hrushikesh Loya, Milad Alizadeh, Adam Golinski, Yee Whye Teh, and Arnaud Doucet. Coin++: Neural compression across modalities. *TMLR*, 2022. 1

[6] Michaël Gharbi, Jiawen Chen, Jonathan T Barron, Samuel W Hasinoff, and Frédo Durand. Deep bilateral learning for real-time image enhancement. *ACM TOG*, 36(4):1–12, 2017. 2

[7] Jingwen He, Yihao Liu, Yu Qiao, and Chao Dong. Conditional sequential modulation for efficient global image retouching. In *ECCV*, 2020. 2

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *ECCV*, 2016. 2

[9] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *CVPR*, 2017. 2

[10] Soo Ye Kim, Jihyong Oh, and Munchurl Kim. Deep SR-ITM: Joint learning of super-resolution and inverse tone-mapping for 4k UHD HDR applications. In *CVPR*, 2019. 2

[11] Soo Ye Kim, Jihyong Oh, and Munchurl Kim. JSI-GAN: GAN-based joint super-resolution and inverse tone-mapping with pixel-wise task-specific filters for uhd hdr video. In *AAAI*, 2020. 2

[12] Gregory Ward Larson, Holly Rushmeier, and Christine Piatko. A visibility matching tone reproduction operator for high dynamic range scenes. *IEEE TVCG*, 3(4):291–306, 1997. 4, 5

[13] Hoang M Le, Brian Price, Scott Cohen, and Michael S Brown. GamutMLP: A lightweight mlp for color loss recovery. In *CVPR*, 2023. 1

[14] Panjun Liu, Jiacheng Li, Lizhi Wang, Zheng-Jun Zha, and Zhiwei Xiong. Mlp embedded inverse tone mapping. In *MM*, 2024. 1, 2

[15] E. Reinhard and K. Devlin. Dynamic range reduction inspired by photoreceptor physiology. *IEEE TVCG*, 11(1):13–24, 2005. 4, 5

[16] Yannick Strümpler, Janis Postels, Ren Yang, Luc Van Gool, and Federico Tombari. Implicit neural representations for image compression. In *ECCV*, 2022. 1

[17] Hui Zeng, Jianrui Cai, Lida Li, Zisheng Cao, and Lei Zhang. Learning image-adaptive 3d lookup tables for high perfor-

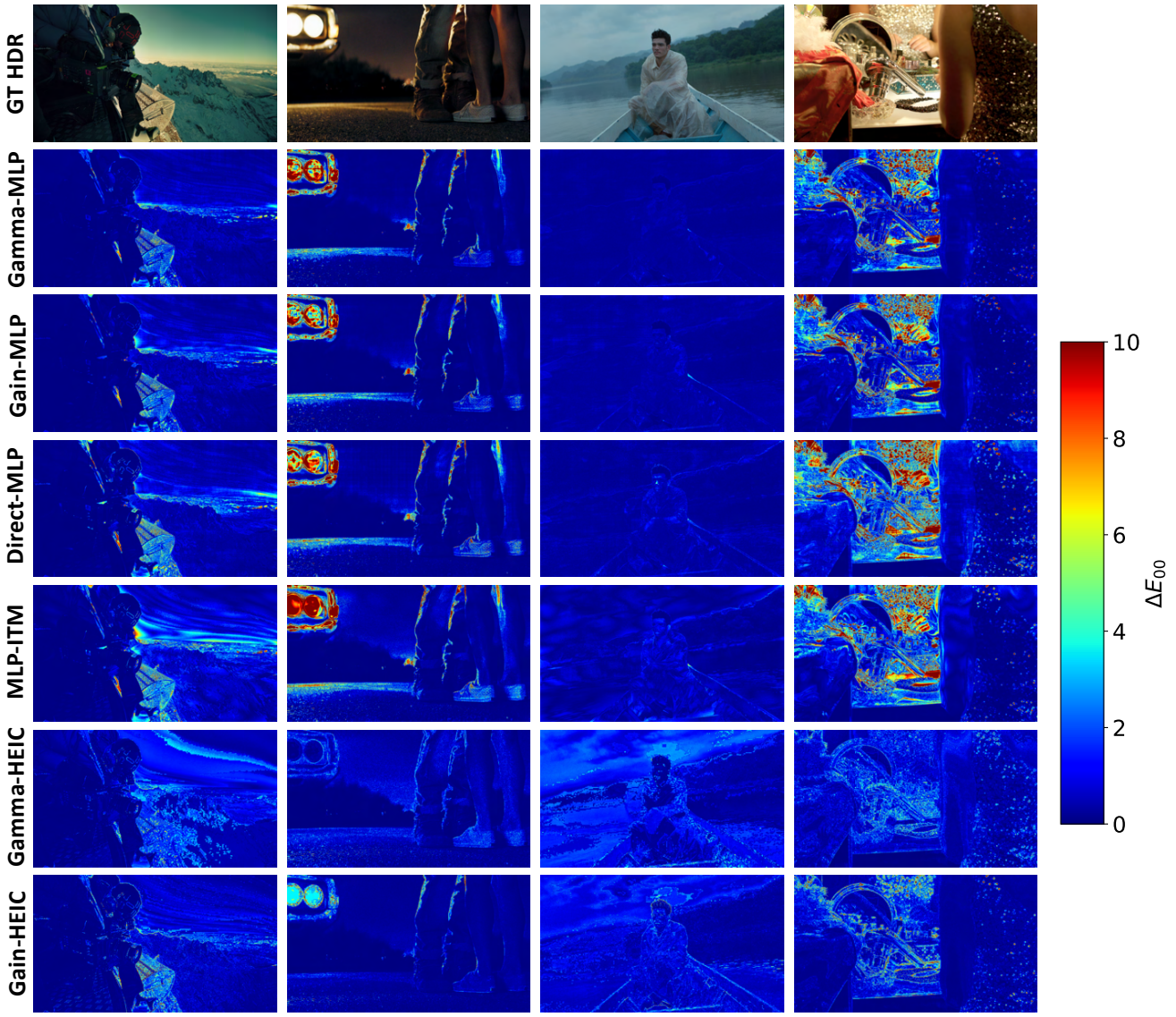

Figure 2. Qualitative comparison between the best performing methods of the quantitative comparison on images of the dataset of Cyriac et al. Errors are visualized as ΔE_{00} heatmaps, clipped at a value of 10. Since the present application is encoding, much of the error is at or below the threshold of visibility for all methods (1 ΔE_{00}). However, the tested methods differ in their accuracy in representing various problematic areas (e.g., the sky in column 1, the headlights and legs in column 2, the head in column 3 and the hands in the last column).

mance photo enhancement in real-time. *IEEE TPAMI*, 44 (4):2058–2073, 2020. 2

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

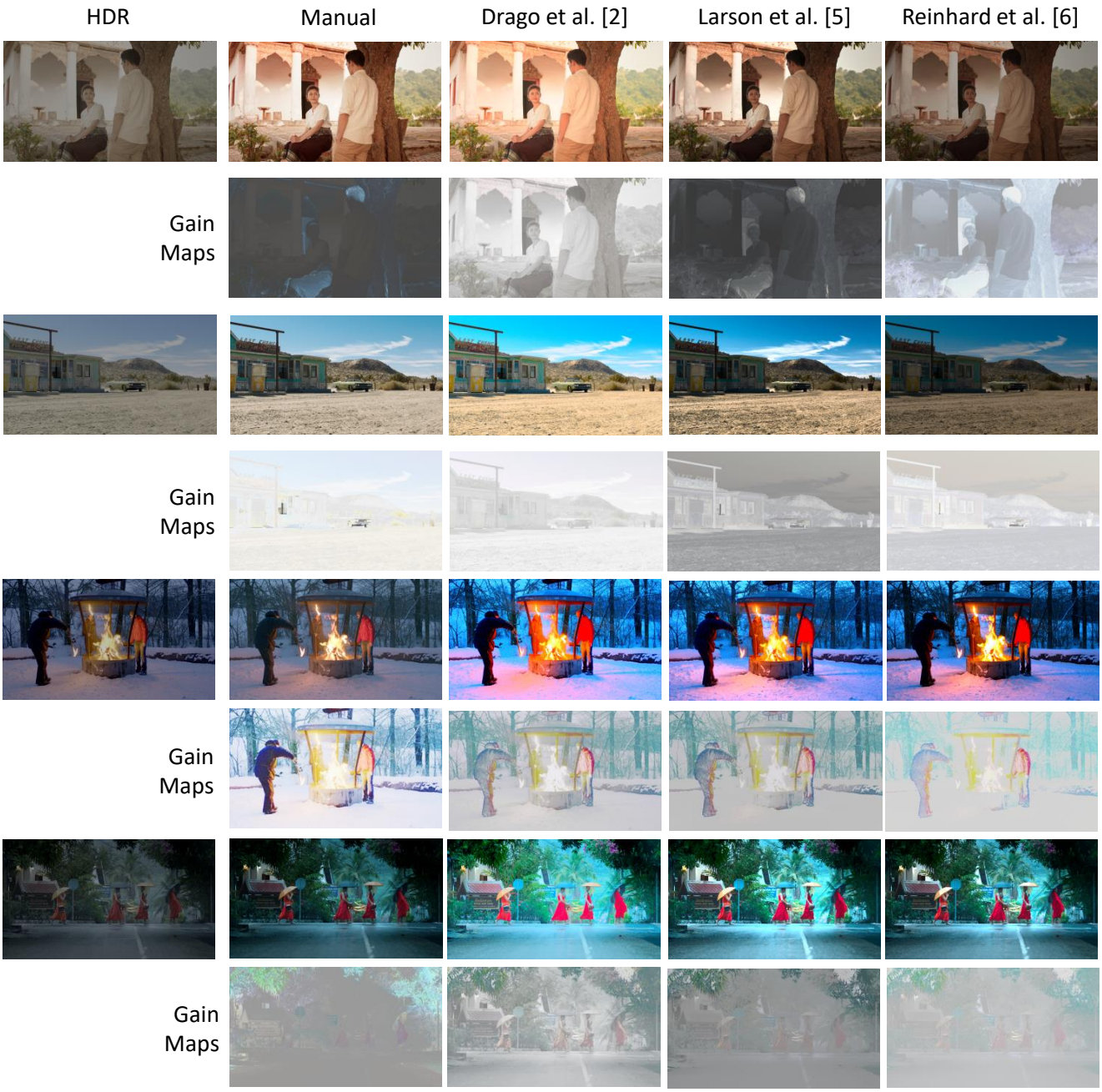

Figure 3. The tone mapping methods tested in the experimental section [3, 12, 15] result in qualitatively different SDR renditions of the HDR source. The methods' gain maps are visualized to illustrate how the HDR reconstruction task varies for different tone mapping conditions. In some instances, the results are brighter or darker, sharper, or more saturated. In general, image detail is preserved better in manual tone mapping, while automatic tone mapping tends to over-saturate dark images and amplify noise (e.g., the last two rows.)

		PSNR ↑	ΔE_{00} ↓	SSIM ↑	ΔE_{ITP} ↓	HDR-VDP3 ↑	Size (KB) ↓
Cyriac et al. [2] Manual TM	Gain-JPEG	38.1	**1.81**	**0.9721**	**8.86**	8.13	15
	Gain-HEIC	39.0	1.63	0.9768	7.93	8.42	14
	Gain-JPEG-XL	**37.2**	1.44	0.9435	7.71	**7.94**	10
	Gamma-JPEG	44.3	0.89	0.9874	5.17	9.26	14
	Gamma-HEIC	45.2	0.80	0.9891	4.68	9.37	12
	Gamma-JPEG-XL	43.1	0.76	0.9553	4.78	8.96	**9**
	MLP-iTM	51.1	**0.53**	0.9974	**2.32**	**9.83**	**34**
	Direct-MLP	47.6	0.92	0.9918	3.61	9.50	10
	Gain-MLP	48.8	0.99	0.9945	3.59	9.15	10
	Gamma-MLP	**51.4**	**0.53**	**0.9978**	**2.32**	9.74	10
Cyriac et al. [2] Reinhard et al. [15]	Gain-JPEG	39.0	**2.08**	0.9691	**9.08**	8.01	15
	Gain-HEIC	39.9	1.92	0.9731	8.22	8.24	13
	Gain-JPEG-XL	**38.2**	1.55	**0.9482**	7.23	**7.99**	**9**
	Gamma-JPEG	42.5	1.23	0.9814	6.28	8.85	15
	Gamma-HEIC	43.3	1.13	0.9837	5.72	8.98	13
	Gamma-JPEG-XL	41.4	0.97	0.9550	5.33	8.66	11
	MLP-iTM	48.8	0.86	0.9908	4.09	9.22	**34**
	Direct-MLP	47.7	0.91	0.9883	4.48	9.18	10
	Gain-MLP	49.8	0.79	**0.9933**	3.47	9.25	10
	Gamma-MLP	**50.6**	**0.69**	**0.9933**	**3.34**	**9.29**	10
Cyriac et al. [2] Larson et al. [12]	Gain-JPEG	36.0	**2.43**	0.9510	**11.23**	**7.33**	16
	Gain-HEIC	36.8	2.29	0.9565	10.33	7.60	14
	Gain-JPEG-XL	**35.2**	1.87	**0.9305**	9.31	7.34	10
	Gamma-JPEG	41.0	1.28	0.9753	6.75	8.56	16
	Gamma-HEIC	41.6	1.20	0.9776	6.26	8.71	15
	Gamma-JPEG-XL	39.9	1.01	0.9491	5.78	8.39	**9**
	MLP-iTM	48.3	**0.70**	**0.9948**	**3.47**	**9.39**	**34**
	Direct-MLP	47.6	0.76	0.9912	3.77	9.32	10
	Gain-MLP	48.0	0.75	0.9955	3.40	9.18	10
	Gamma-MLP	**49.6**	0.73	0.9930	3.84	9.14	10
Cyriac et al. [2] Drago et al. [3]	Gain-JPEG	41.2	**1.89**	0.9731	**8.24**	8.54	12
	Gain-HEIC	42.3	1.70	0.9769	7.34	8.69	10
	Gain-JPEG-XL	**40.6**	1.39	**0.9509**	6.47	**8.38**	**7**
	Gamma-JPEG	42.6	1.33	0.9780	6.66	8.80	13
	Gamma-HEIC	43.3	1.24	0.9803	6.15	8.91	11
	Gamma-JPEG-XL	41.7	1.06	0.9513	5.67	8.63	9
	MLP-iTM	46.5	1.05	0.9881	5.48	9.01	**34**
	Direct-MLP	47.6	0.86	0.9861	4.75	9.06	10
	Gain-MLP	**49.6**	0.74	0.9920	3.99	9.11	10
	Gamma-MLP	**49.6**	**0.73**	**0.9930**	**3.84**	**9.14**	10
Chen et al. [1] Youtube TM	Gain-JPEG	37.0	**2.58**	**0.9743**	10.73	7.57	37
	Gain-HEIC	38.0	2.35	0.9780	9.75	7.76	41
	Gain-JPEG-XL	37.1	2.40	0.9757	10.32	**7.45**	21
	Gamma-JPEG	36.9	2.10	0.9747	**10.74**	7.66	39
	Gamma-HEIC	37.7	2.03	0.9774	10.05	7.77	**40**
	Gamma-JPEG-XL	**36.8**	1.95	0.9729	10.47	7.15	22
	MLP-iTM	**41.6**	**1.22**	**0.9879**	**6.04**	8.25	34
	Direct-MLP	41.1	1.35	0.9861	6.67	8.22	10
	Gain-MLP	41.2	1.56	0.9861	6.88	8.19	10
	Gamma-MLP	41.5	1.25	0.9876	6.19	**8.26**	**10**

Table 3. Quantitative comparison between traditional encoding techniques and the proposed MLP broken up for tone mapping (TM) variations of Cyriac et al. [2] (Manual, Drago et al. [3], Larson et al. [12], Reinhard et al. [15]), and Chen et al. [1] (Youtube) datasets.

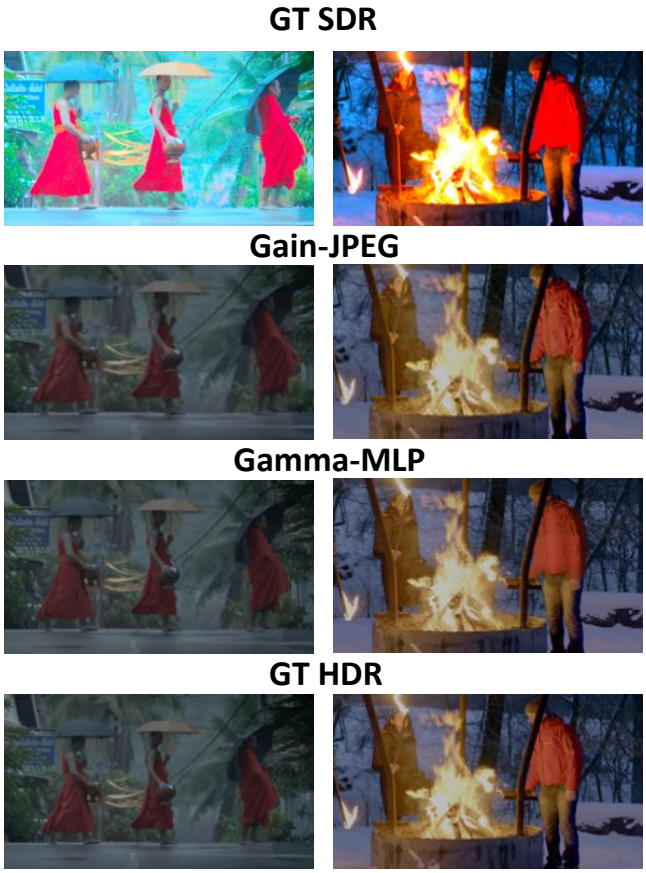

Figure 4. The proposed method (Gamma-MLP) produces artifacts when the ground truth SDR image has lost significant detail due to over-saturation in the tone mapping stage. As a result, details are lost (e.g., the monk's robes in the top row and the red jacket in the bottom row). The conventional framework results also suffer in these scenarios (Gain-JPEG included for comparison).