HDR Image Generation via Gain Map Decomposed Diffusion
——Supplementary Material——

The supplementary material is organized as follows:

- Section 1 presents additional fundamental details about the gain map (GM) and analyzes its resolution scalability under our framework.
- Section 2 details the unsupervised triplet dataset construction, including implementation details and more visual samples in the constructed dataset.
- Section 3 describes the adapted diffusion, presenting detailed configuration and ablation studies examining denoising step impacts on performance.
- Section 4 provides extended text-to-HDR image generation results and more visual comparisons on the HDRTV1K dataset [3] for the SDR-to-HDRTV up-conversion task.

1. GM Investigation

We introduce the background of GM in Sec. 1.1. Further, we investigate the resolution scalability of GM in our framework with a quantitative experiment in Sec. 1.2

1.1. GM Background

GainMap solves the problem of inconsistent HDR image rendering across devices by combining an SDR base image with a compact "Gain Map" that defines per-pixel brightness adjustments [1, 2, 4, 6]. This map, derived from the ratio between SDR and HDR version, allows dynamic adaptation at display time: devices with limited HDR capabilities blend the SDR base image and GM proportionally, while full-HDR screens apply the full GM for maximum detail. Metadata (Q_{max}) guides this scaling, ensuring consistent appearance regardless of hardware. GM preserves creative intent by letting authors explicitly define HDR-SDR transitions, supports localized adjustments (e.g., brightening only specific regions), and maintains compatibility with legacy software. Its efficient compression (via downsampling) and GPU-friendly design enable practical deployment. The ongoing ISO 21496-1 standardization highlights its viability as a unified solution for HDR content distribution. Our framework generates high-quality HDR images of this pioneering, back-compatible HDR format.

1.2. Resolution Scalability Investigation

As discussed in Sec.1.1, GM storage at reduced resolution [1, 2] maintains effectiveness in dynamic range expansion. To quantitatively verify this characteristic, we conduct ablation studies on HDRTV1K test set [3] using progressive downsampling factors (2×, 4×, 8×, 16×). More specifically, we downsample the SDR image with varying factors and generate the corresponding GM through our adapted diffusion model. The generated downsampled GM is applied to the original SDR image to obtain the predicted HDR image. As evidenced by Tab. 1, while showing moderate performance degradation in both HDR/WCG and SDR metrics, the predicted HDR images maintain satisfactory HDR/WCG characteristics and perceptual quality across all resolutions. This resolution scalability property enables efficient HDR generation by reducing computational overhead in GM processing without compromising perceptual quality.

2. Unsupervised Triple Dataset Construction

This section elaborates on the implementation details of unsupervised data construction, including VAE LoRA configuration details (Sec. 2.1) and additional visual samples from our constructed triple dataset (Sec. 2.2).

Table 1. Ablation Study on Resolution Scalability of GM on HDRTV1K [3] Test Set.

	HDR/WCG Metrics				SDR Metrics	
Factors	FHLP↑	EHL↑	FWGP↑	EWG↑	BRISQUE↓	NIQE↓
×2	0.4248	0.0809	0.0419	0.0312	45.8215	5.3809
×4	0.4302	0.0781	0.0390	0.0291	46.0363	5.4304
×8	0.4368	0.0761	0.0365	0.0280	46.6642	5.5597
×16	0.4263	0.0758	0.0421	0.0292	46.9757	5.5978

2.1. VAE Configuration

The unsupervised data construction involves employing a parameter-efficient LoRA fine-tuning strategy [5] for the pretrained image VAE. The adaptation mechanism systematically applies to all convolutional and linear layers to ensure the capability to conduct SDR-GM conversion. We configure the rank dimension and the scaling factor as 64, maintaining a unity ratio to modulate update magnitudes. Bias parameters remain frozen to preserve base model integrity while minimizing trainable parameters. This design maintains the representational capacity of the base model. We also update the parameters of the final convolutional layer of the VAE decoder to ensure effective output space feature learning.

2.2. Dataset Thumbnail

We further provide more representative samples in our training dataset in Fig. 2. The constructed triple dataset enables the framework to effectively learn the GM distribution and achieve decomposed HDR image generation.

3. Decomposed HDR Image Generation

This section first describes the implementation details of our adapted diffusion configuration and then analyzes through ablation studies how denoising steps affect the HDR generation quality.

3.1. Diffusion Configuration

The proposed diffusion pipeline leverages an original Stable Diffusion UNet with an adapted Stable Diffusion UNet, featuring an expanded 8-channel input convolution and a hierarchical block structure characterized by progressively increasing channel dimensions [320, 640, 1280, 1280]. In our experiments, we employ a DDIM noise scheduling [7] with 50 inference steps, utilizing group normalization and a SiLU activation function. The architecture integrates cross-attention mechanisms with a 768-dimensional embedding space, supporting precise latent representation learning. Implemented with half-precision computational strategy, the pipeline demonstrates enhanced computational efficiency while maintaining high-fidelity image transformation capabilities.

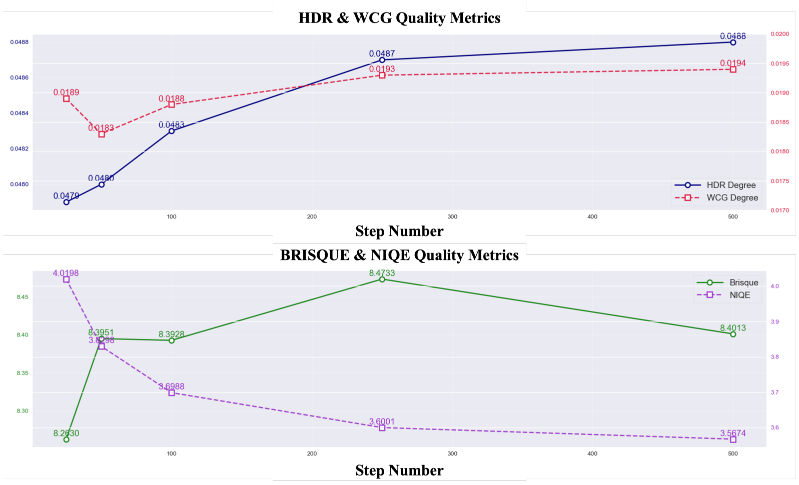


Figure 1. Ablation Study on Denoising Step Numbers.

3.2. Analysis of Denoising Steps

To investigate the impact of denoising steps in the noise scheduler, we designed a systematic experimental study. Specifically, we generated five HDR image groups using 500 diverse prompts, systematically varying the denoising steps from 25 to 500. As depicted in Fig. 1, our analysis reveals a consistent and progressive enhancement in both HDR/WCG fidelity and subjective visual quality as the number of denoising steps increases.

4. More Visual Results

We showcase representative visual outcomes for text-to-HDR image generation in Fig. 3 and SDR-to-HDR image up-conversion in Fig. 4 and 5. These results demonstrate that our framework achieves high-quality HDR image generation while also exhibiting effective dynamic range expansion capabilities in real-world applications.

References

[1] Adobe. Gain map specification. `https://helpx.adobe.com/camera-raw/using/gain-map.html`, 2024.

[2] Apple. Explore hdr rendering with edr. `https://developer.apple.com/videos/play/wwdc2021/10161`, 2021.

[3] Xiangyu Chen, Zhengwen Zhang, Jimmy S Ren, Lynhoo Tian, Yu Qiao, and Chao Dong. A new journey from sdrtv to hdrtv. In *ICCV*, 2021.

[4] Google. Ultra-hdr image format. `https://developer.android.com/media/platform/hdr-image-format`, 2024.

[5] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, et al. Lora: Low-rank adaptation of large language models. In *ICLR*, 2022.

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

[7] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *ICLR*, 2021.

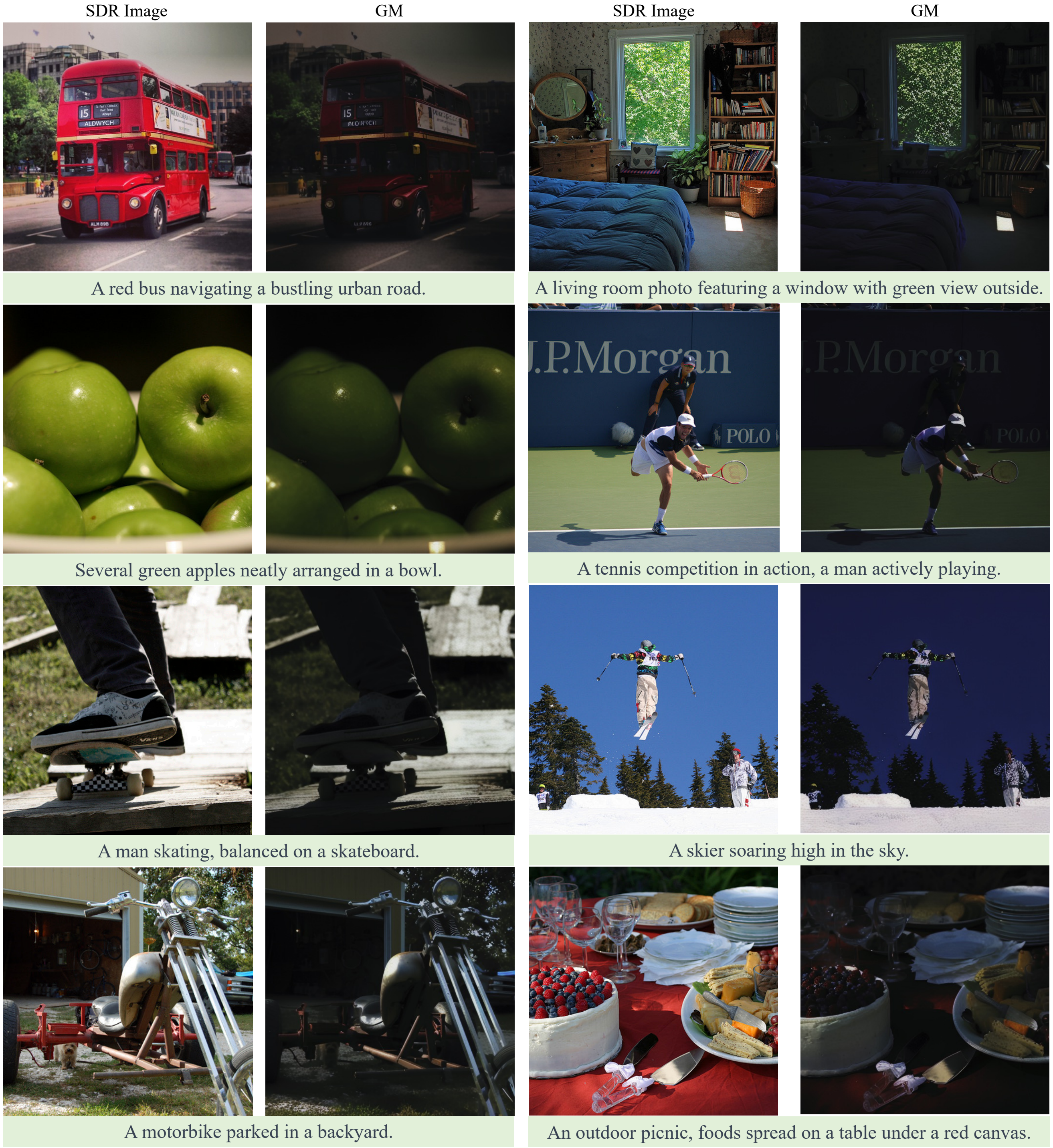
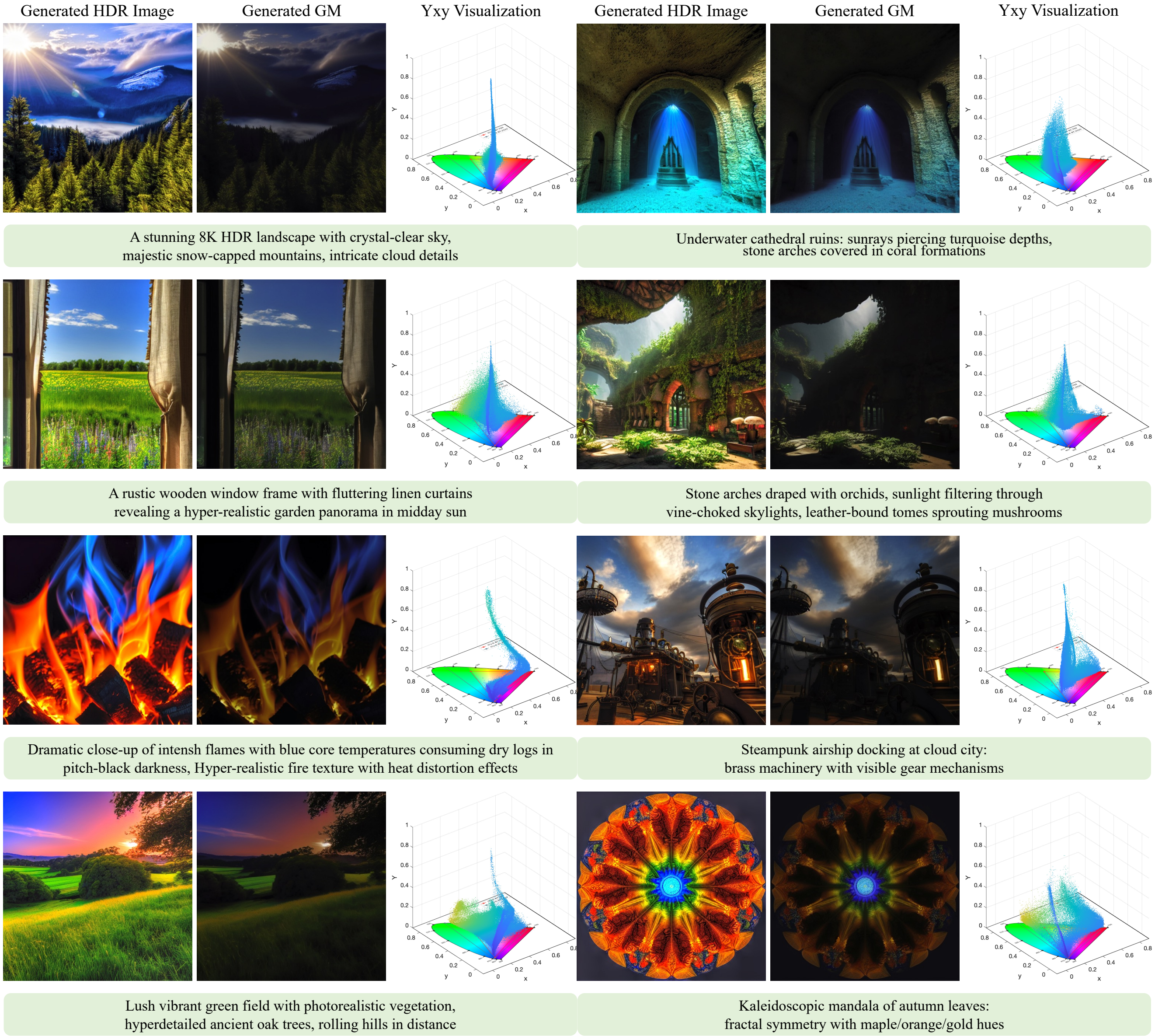

SDR Image
GM
SDR Image
GM
A red bus navigating a bustling urban road.
A living room photo featuring a window with green view outside.
Several green apples neatly arranged in a bowl.
A tennis competition in action, a man actively playing.
A man skating, balanced on a skateboard.
A skier soaring high in the sky.
A motorbike parked in a backyard.
An outdoor picnic, foods spread on a table under a red canvas.

Figure 2. "Text-SDR-GM" Samples in Our Dataset. Our unsupervised dataset construction pipeline enables the decomposed HDR image generation with GM.



A stunning 8K HDR landscape with crystal-clear sky,
majestic snow-capped mountains, intricate cloud details

Underwater cathedral ruins: sunrays piercing turquoise depths,
stone arches covered in coral formations

A rustic wooden window frame with fluttering linen curtains
revealing a hyper-realistic garden panorama in midday sun

Stone arches draped with orchids, sunlight filtering through
vine-choked skylights, leather-bound tomes sprouting mushrooms

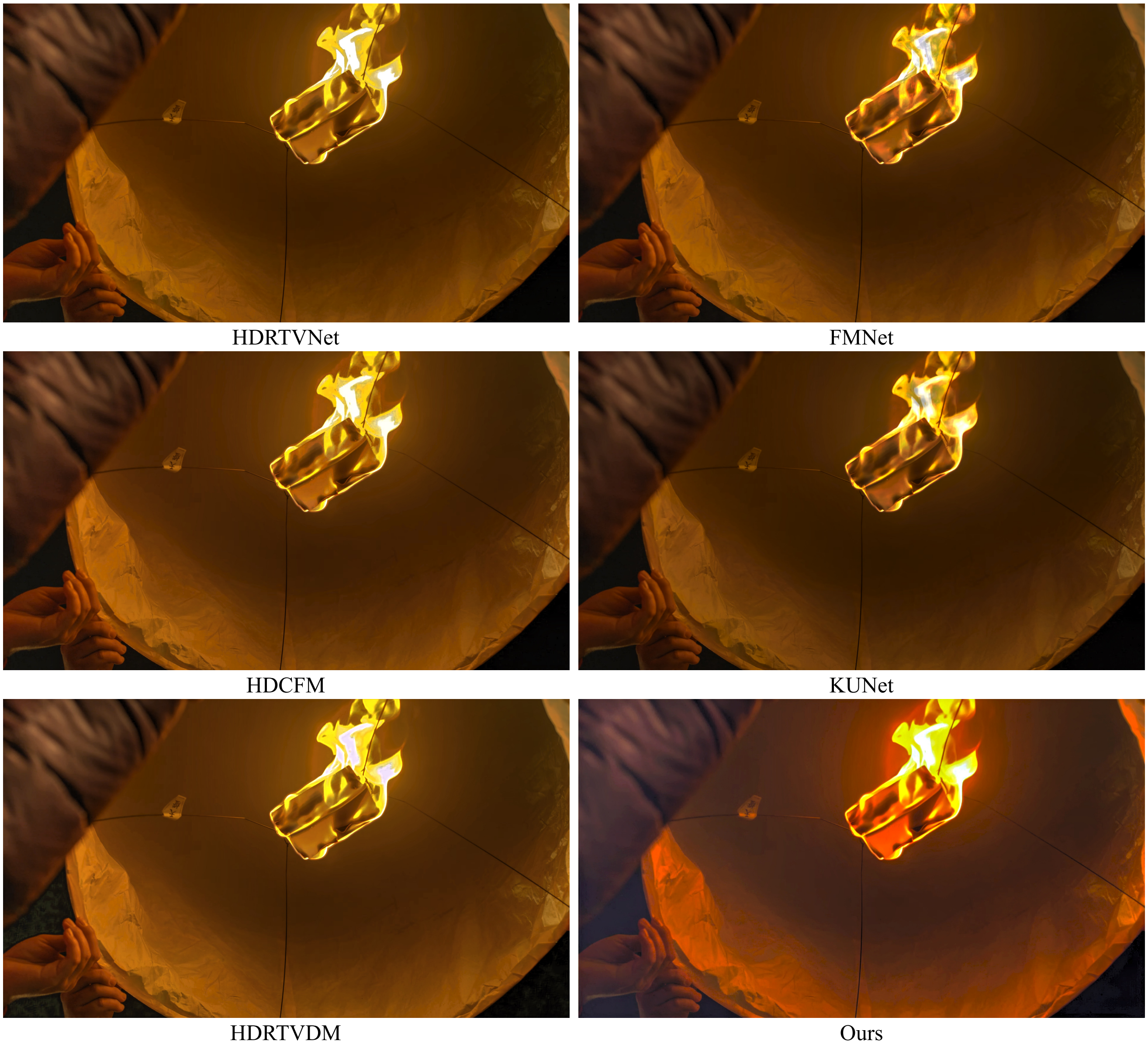
Dramatic close-up of intensh flames with blue core temperatures consuming dry logs in
pitch-black darkness, Hyper-realistic fire texture with heat distortion effects

Steampunk airship docking at cloud city:
brass machinery with visible gear mechanisms

Lush vibrant green field with photorealistic vegetation,
hyperdetailed ancient oak trees, rolling hills in distance

Kaleidoscopic mandala of autumn leaves:
fractal symmetry with maple/orange/gold hues

Figure 3. Text-to-HDR Image Generation Results. All the presented results are of 512 × 512 resolution.



HDRTVNet

FMNet

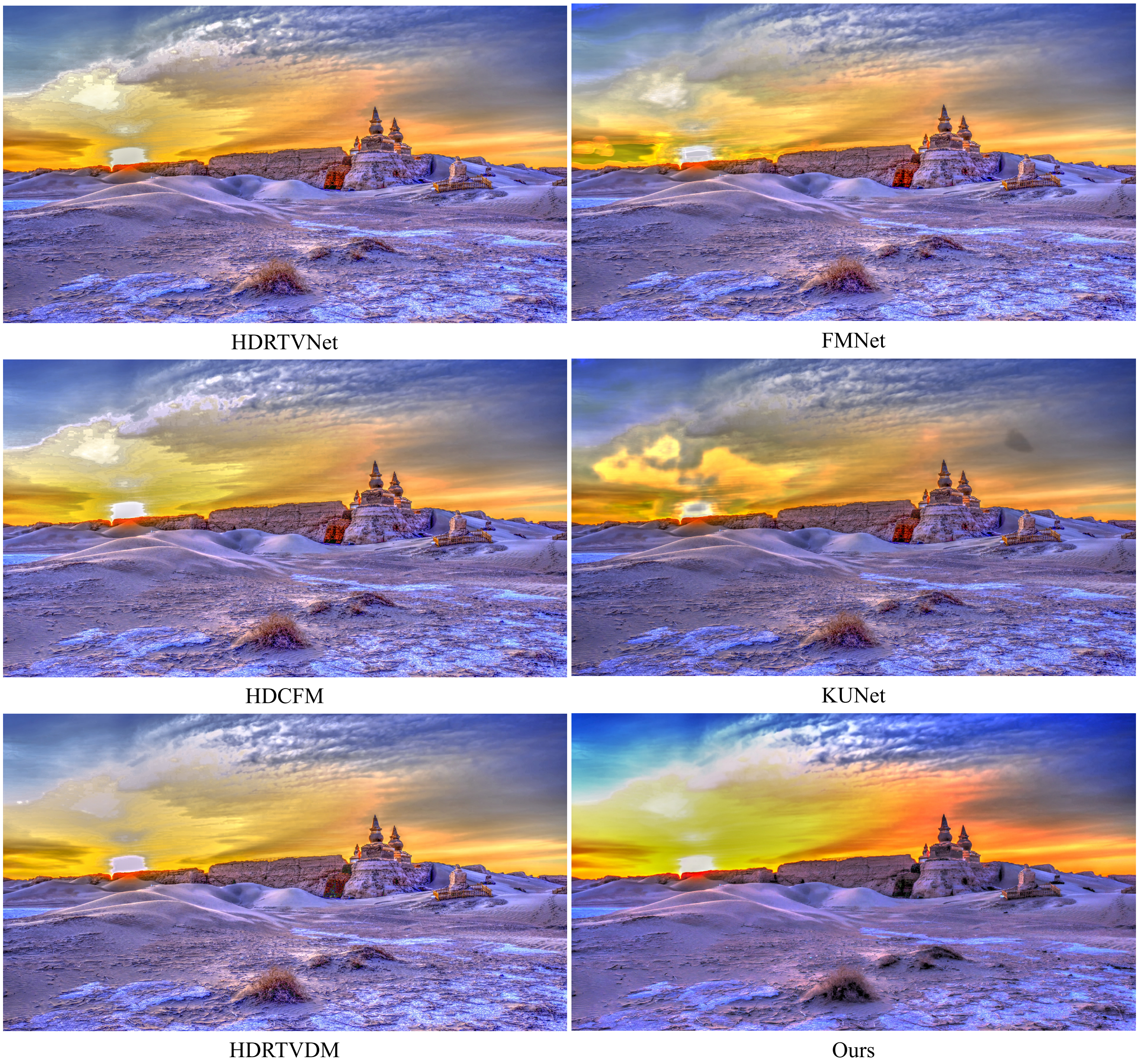
HDCFM

KUNet

HDRTVDM

Ours

Figure 4. Visual Comparison on HDRTV1K [3] Test Set. Our framework produces artifact-free results, especially in the highlighted regions.



HDRTVNet

FMNet

HDCFM

KUNet

HDRTVDM

Ours

Figure 5. Visual Comparison on HDRTV1K [3] Test Set. Our framework produces artifact-free results with vivid colors.