Han WANG, Bolun ZHENG, Quan CHEN, Qianyu ZHANG, Tao ZHANG, Jiyong ZHANG, Xiang TIAN, 2026. Unsupervised single-image high dynamic range rendering via multi-exposure priors. *Engineering Information Technology & Electronic Engineering*, 27(6):250116.
https://doi.org/10.1631/ENG.ITEE.2025.0116

Unsupervised single-image high dynamic range rendering via multi-exposure priors

Key words: High dynamic range (HDR); HDR reconstruction; Single-image HDR; Unsupervised learning; Multi-exposure prior

Corresponding author: Quan CHEN
E-mail: chenquan@alu.hdu.edu.cn

ORCID: https://orcid.org/0000-0003-2858-6771

Motivation

- Existing supervised HDR methods require real HDR images for training, while unsupervised methods usually rely on multi-exposure inputs, increasing data costs and limiting flexibility. USME-HDR learns multi-exposure priors to reconstruct high-quality HDR images from a single low dynamic range (LDR) input without ground-truth HDR supervision, while reducing ghosting artifacts caused by multi-exposure fusion.

Contributions

1. We pioneer an unsupervised HDR rendering method tailored for single-image input. By learning multi-exposure priors, our method eliminates the dependencies on multi-exposure LDR image sequences while suppressing ghosting artifacts in dynamic scenes.
2. We propose a multi-exposure image generation strategy guided by luminance and exposure time. Specifically, the estimated light map and light feature are incorporated to enhance the brightness and texture details of the synthesized images, while exposure time embeddings guide the model to learn accurate luminance variations.
3. Experimental results demonstrate that the proposed method achieves competitive HDR rendering quality from a single LDR input, rivaling multi-exposure methods. USME-HDR suppresses ghosting artifacts in dynamic scenes, exhibiting significant practical utility.

Overview of the USME-HDR

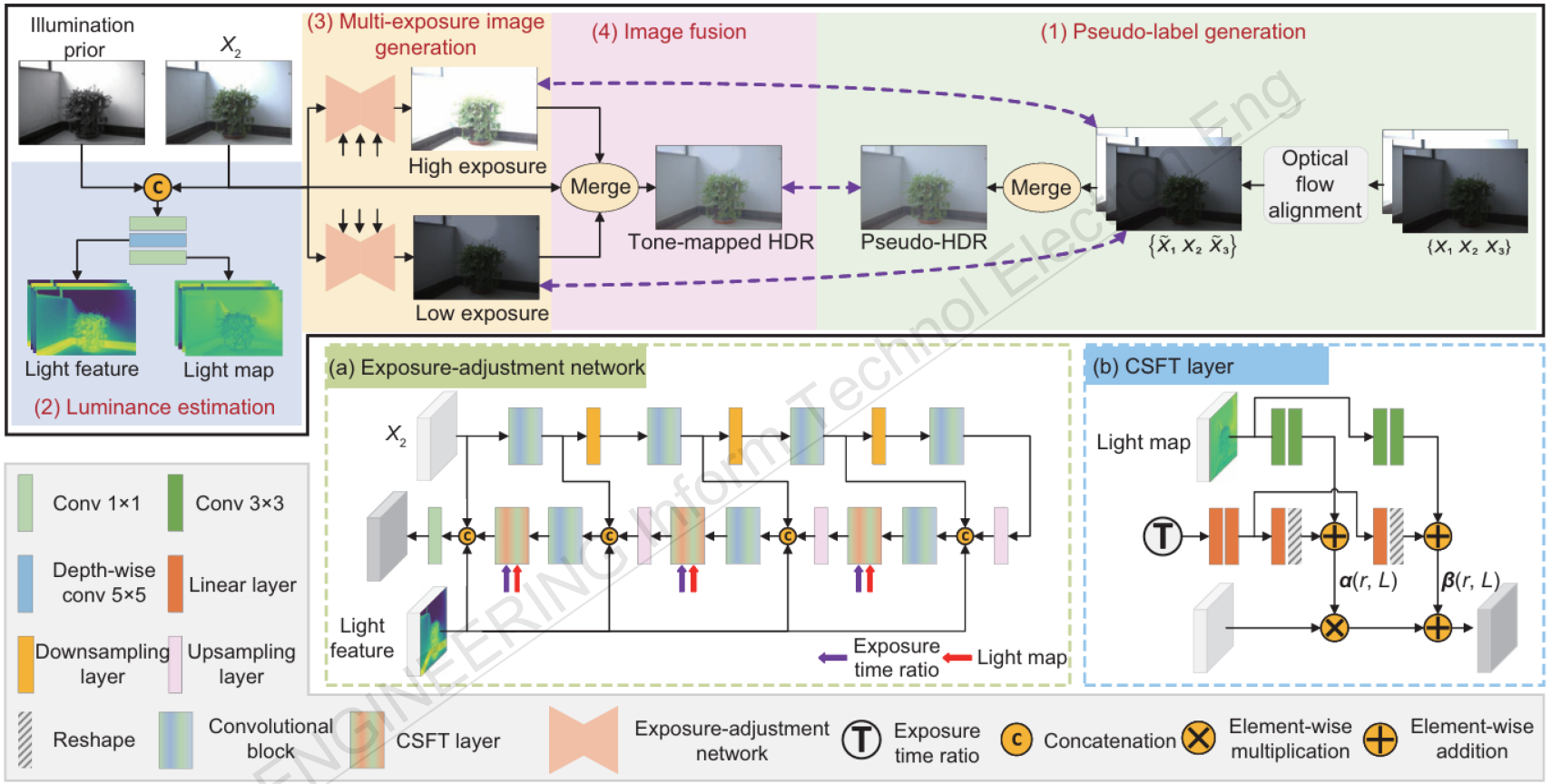
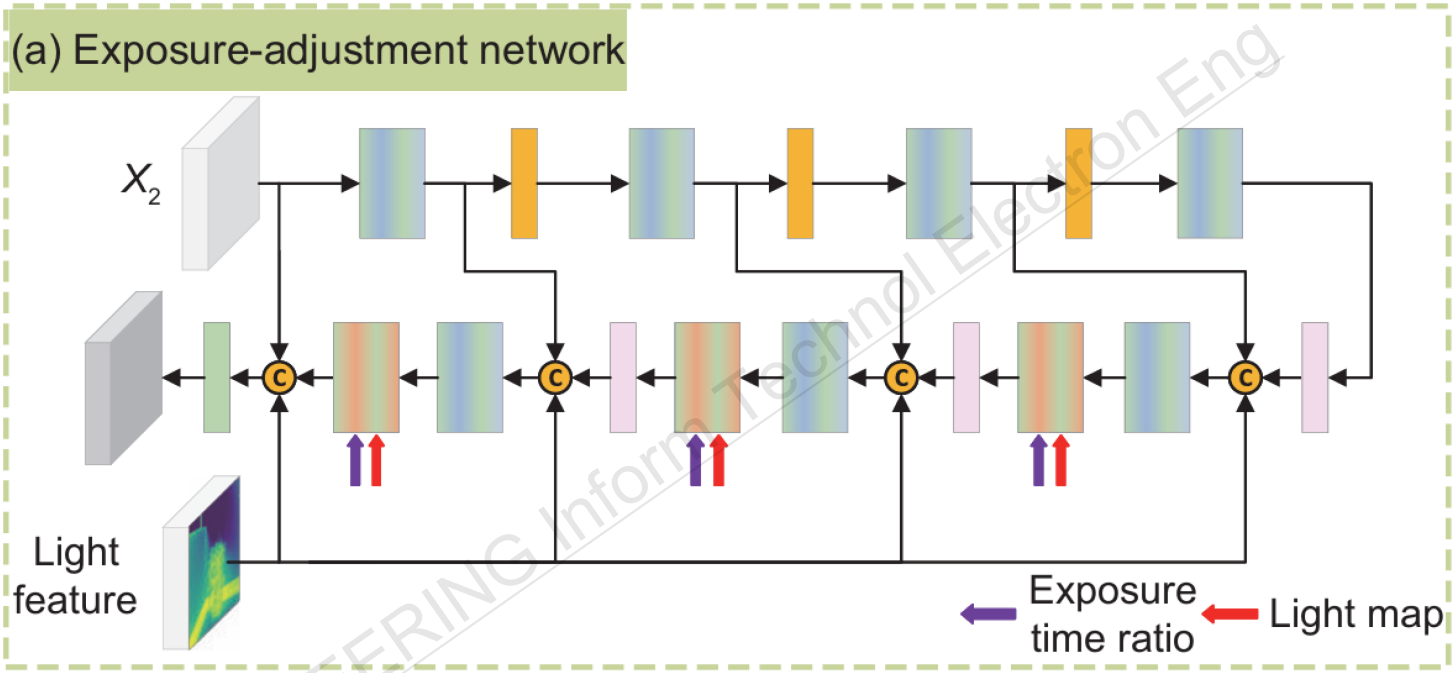


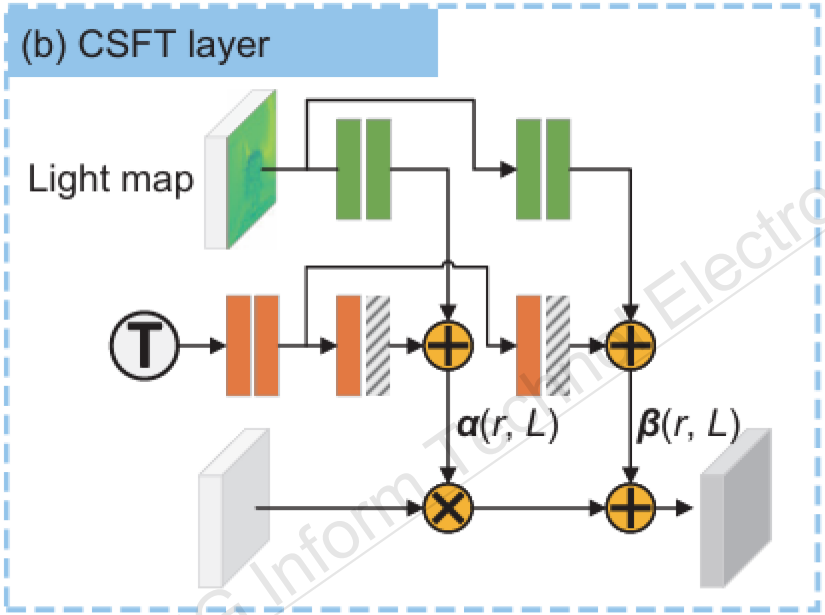
Fig. 2 Overview of the USME-HDR. The input X_2 is decomposed into a light map and a light feature, which are used to guide the subsequent multi-exposure generation process. (a) Architecture of the EAN. The network takes the input image together with luminance-related representations to generate low- and high-exposure images under different exposure levels. The light feature is concatenated with intermediate features at each decoder block, while the light map and exposure time ratio are jointly used as modulation conditions. (b) Structure of the CSFT layer. The light map and exposure time ratio are used to predict the spatially adaptive modulation parameters, which are applied to intermediate features through an affine transformation for luminance-aware exposure modulation. $\tilde{X}_1$ and $\tilde{X}_3$ denote X_1 and X_3 after being aligned with reference to X_2

Exposure-adjustment network



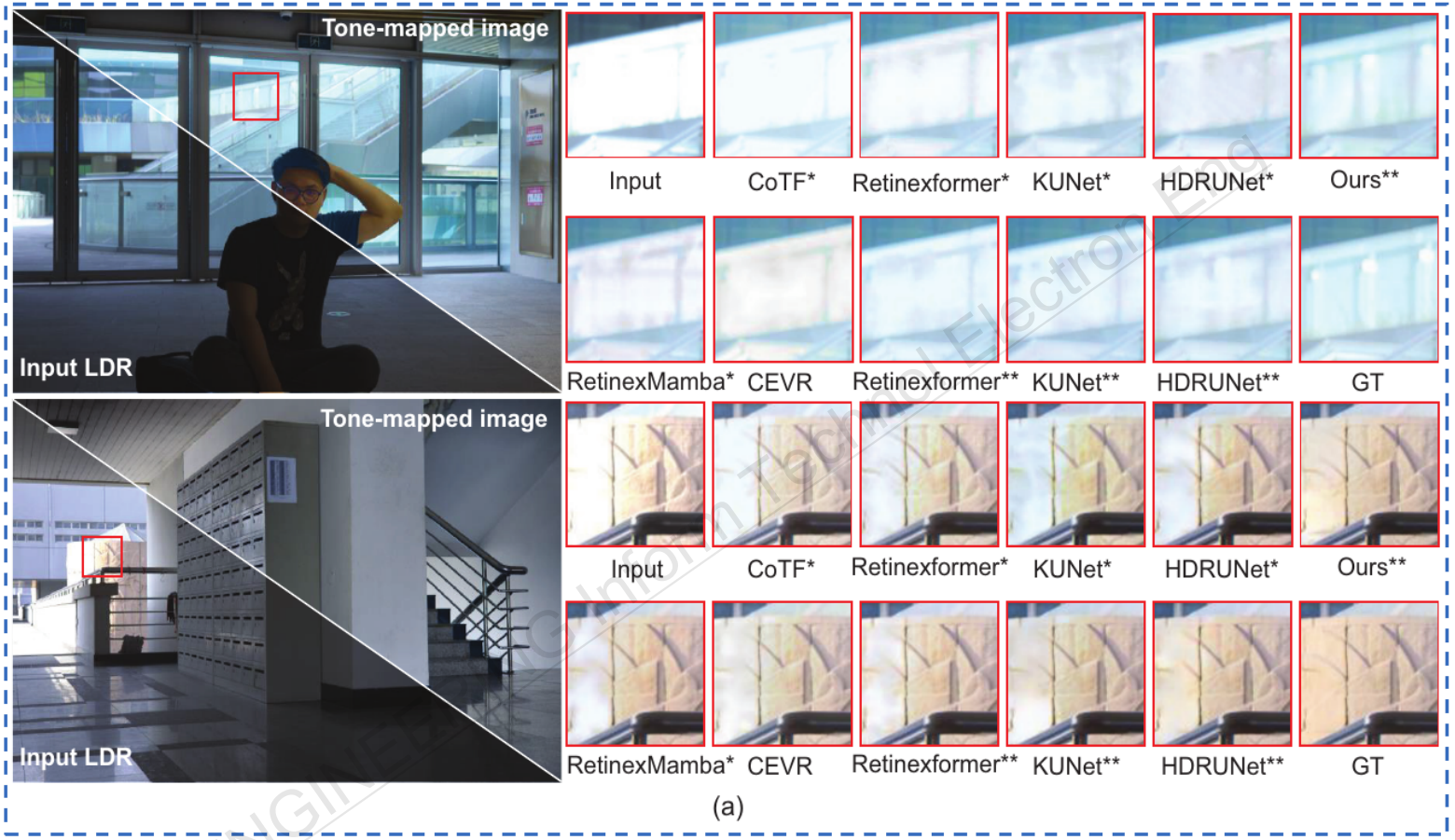
Architecture of the exposure-adjustment network (EAN). The network takes the input image together with luminance-related representations to generate low- and high-exposure images under different exposure levels. The light feature is concatenated with intermediate features at each decoder block, while the light map and exposure time ratio are jointly used as modulation conditions.

CSFT layer



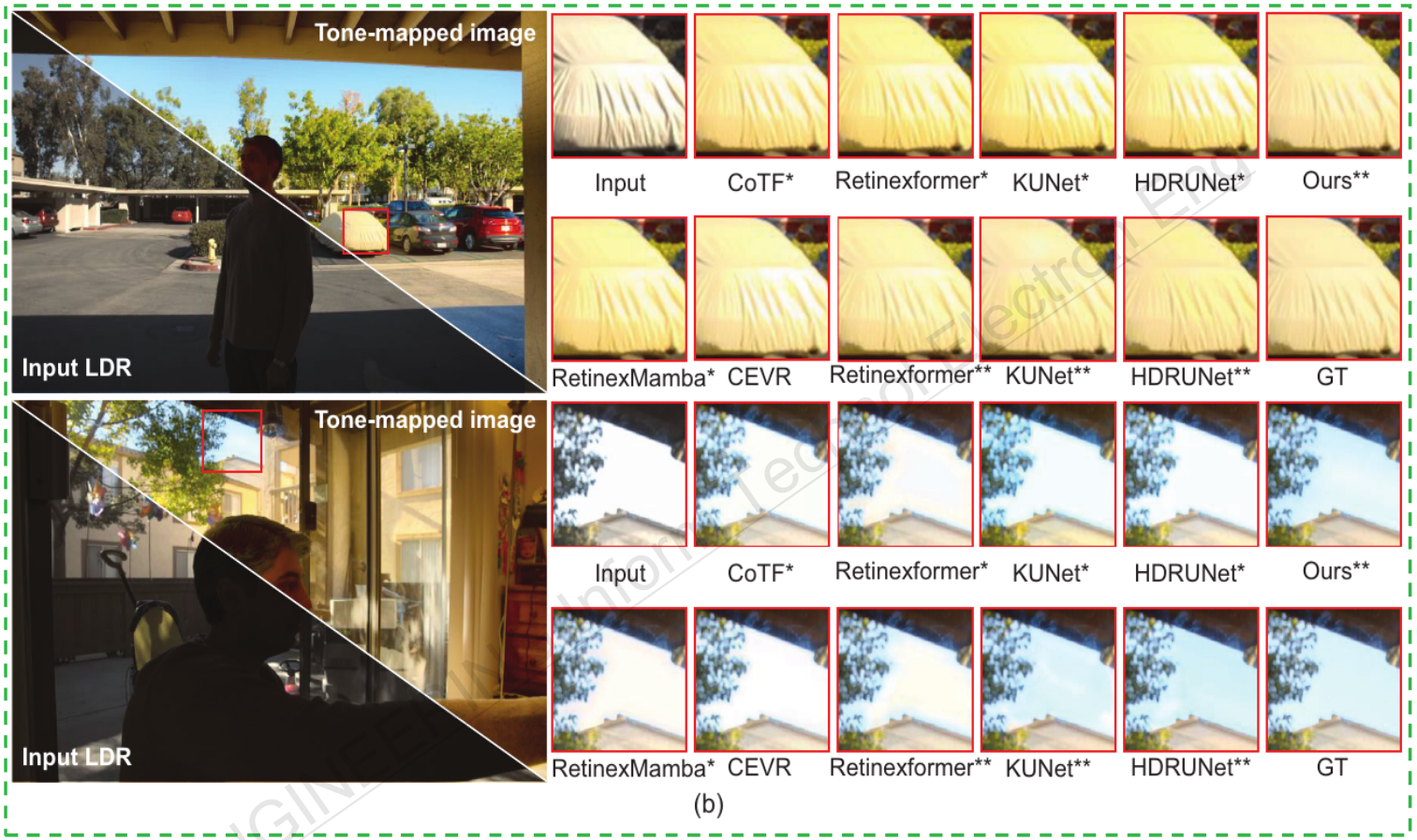
Structure of the composite spatial feature transform (CSFT) layer. The light map and exposure time ratio are used to predict the spatially adaptive modulation parameters, which are applied to intermediate features through an affine transformation for luminance-aware exposure modulation.

Visual Comparison



Visual comparison on RealHDRV dataset. Compared with other methods, our approach better reconstructs details in over-exposed regions while preserving more natural and consistent colors.

Visual Comparison



Visual comparison on Kalantari's dataset. Compared with other methods, our approach better reconstructs details in over-exposed regions while preserving more natural and consistent colors.

Quantitative Comparison

Table 1 Quantitative comparison results on Kalantari's and RealHDRV datasets

Method	Kalantari's dataset					RealHDRV dataset					Computational cost		
	PSNR-μ	PSNR-L	SSIM-μ	SSIM-L	HDR-VDP-2	PSNR-μ	PSNR-L	SSIM-μ	SSIM-L	HDR-VDP-2	N_{MACs} ($\times 10^9$)	N_{params} ($\times 10^6$)	Time (s)
HDRUNet	33.38	32.72	0.9728	0.9455	61.39	29.67	29.18	0.9476	0.9544	60.93	532.70	1.651	0.18
KUNet	30.05	32.42	0.9534	0.9475	61.76	28.97	28.73	0.9214	0.9483	61.44	964.42	1.137	0.215
CoTF	29.64	32.01	0.9723	0.9424	53.99	28.83	28.53	0.9384	0.9558	53.54	2.47	0.31	0.04
Retinexformer	33.93	34.08	0.9758	0.9645	61.64	31.04	29.58	0.9517	0.9508	61.33	389.52	1.606	0.435
RetinexMamba	29.23	33.41	0.9529	0.9418	60.96	31.56	29.73	0.9545	0.9531	62.52	869.35	3.588	1.66
HDRUNet*	36.04	33.63	0.9829	0.9735	60.45	32.32	29.32	0.9727	0.9753	61.48	1065.00	3.190	0.405
KUNet*	34.77	33.49	0.9813	0.9726	59.91	31.04	28.68	0.9735	0.9720	61.22	1929.00	2.273	0.44
CEVR	40.60	36.79	0.9818	0.9747	61.13	37.30	31.32	0.9561	0.9834	61.72	1819.00	6.586	0.329
CoTF*	32.94	33.26	0.9813	0.9640	59.89	34.01	29.14	0.9743	0.9763	60.72	4.97	0.63	0.09
Retinexformer*	36.93	34.20	0.9853	0.9723	61.76	33.85	29.09	0.9714	0.9765	61.04	781.04	3.212	0.92
RetinexMamba*	36.40	33.46	0.9839	0.9709	61.52	34.45	29.91	0.9742	0.9761	61.80	1739.00	7.176	3.21
USME-HDR*	40.92	37.89	0.9872	0.9792	62.47	40.62	32.60	**0.9834**	0.9879	63.07	487.59	3.928	0.16
HDRUNet**	41.12	37.46	0.9871	0.9753	62.47	40.17	31.86	0.9832	0.9851	60.83	1076.00	3.196	0.37
KUNet**	41.18	37.41	0.9873	0.9768	61.99	40.19	32.02	0.9833	0.9854	60.71	1934.00	2.277	0.425
Retinexformer**	41.22	37.46	0.9874	0.9780	61.74	40.52	32.35	0.9829	0.9868	62.70	788.94	3.228	0.98
SelfHDR (multi-exposure)	**43.68**	**41.09**	**0.9901**	**0.9873**	**64.57**	**41.91**	**37.65**	0.9751	**0.9911**	**63.97**	1871.00	1.242	0.353
SelfHDR (unified-input)	42.07	37.50	0.9888	0.9798	61.72	39.68	32.83	0.9748	0.9882	62.77	1871.00	1.242	0.351
USME-HDR**	41.36	38.12	0.9879	0.9794	62.52	40.78	33.01	**0.9834**	0.9882	63.23	494.36	3.932	0.162

Methods marked with $*$ denote indirect reconstruction, where low- and high-exposure images are generated from the middle-exposure input and then fused into an HDR image. Methods marked with $**$ indicate indirect reconstruction with the 6-channel input X_2. Please refer to Section 4.2 for details. For SelfHDR, results under both the original multi-exposure setting and the unified-input setting are reported. The reported time corresponds to processing an image of resolution 1500×1000 on a single RTX 3090 GPU, and N_{MACs} denotes the computational cost under the same configuration. The best and second-best results are highlighted in bold and underlined, respectively

Conclusions

This paper presents USME-HDR, a framework for single-image HDR reconstruction without ground-truth HDR supervision. The proposed method requires only a single LDR image as input at inference time. USME-HDR incorporates an EAN to learn multi-exposure priors, enabling the generation of complementary exposure information from a single input image. Inspired by the Retinex theory, we introduce a light map and a light feature to provide luminance-aware guidance for exposure generation, while exposure time ratio guidance further improves luminance fidelity. Finally, the input LDR image is fused with the generated multi-exposure images, and the overall framework is optimized using synthetic pseudo-labels. Experimental results demonstrate that USME-HDR can reconstruct high-quality HDR images from only a single LDR input, without relying on real HDR images or real multi-exposure LDR inputs, highlighting its practical value.

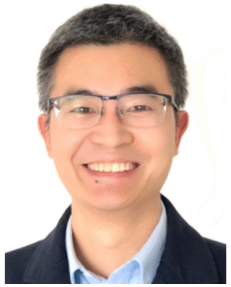
Han WANG received the B.S. degree in electrical engineering and automation from NingboTech University, Ningbo, China. He is currently pursuing the M.Eng. degree with Hangzhou Dianzi University, Hangzhou, China. His research interests include high-dynamic range imaging and image processing.

Bolun ZHENG received the B.S. and Ph.D. degrees in electronic information technology and instrument from Zhejiang University in 2014 and 2019, respectively. He is currently an Associate Professor with Hangzhou Dianzi University. His research interests include computer vision, pattern recognition, image processing, and embedded parallel computing.

Quan CHEN received the B.S. degree from Hangzhou Dianzi University, Zhejiang, China, in 2020, where he is currently pursuing the Ph.D. degree. His research interests include image retrieval, object detection, and image restoration.

Qianyu ZHANG received the B.S. degree from Hangzhou Dianzi University, Hangzhou, China, in 2022, where she is currently pursuing the Ph.D. degree with the School of Automation. Her research interests include image and video processing.

Tao ZHANG received the B.S. and Ph.D. degrees from the School of Computer Science and Technology, Beijing Institute of Technology, China, in 2017 and 2023, respectively. He is currently an associate professor at Hangzhou Dianzi University. His research interests include intelligent information processing, computational photography, and image processing.



Jiyong ZHANG (Member, IEEE) received the B.S. and M.S. degrees in computer science from Tsinghua University in 1999 and 2001, respectively, and the Ph.D. degree in computer science from the École Polytechnique Fédérale de Lausanne (EPFL) in 2008. He is currently a Distinguished Professor with Hangzhou Dianzi University. His research interests include artificial intelligence, machine learning, data mining, and image processing.

Xiang TIAN received the B.Sc. and Ph.D. degrees from Zhejiang University, Hangzhou, China, in 2001 and 2007, respectively. He is currently a Professor at Zhejiang University. His research interests include signal processing and video coding.