LRHDR: Learning Representation-enhanced HDR Video Reconstruction

Supplementary Material

1. Further Analysis

1.1. Further Analysis on ACCR

Distribution alignment effect of RM. We first analyze how RM reduces the discrepancy between features extracted from different exposure inputs. For each sequence, we collect features at the EIC stage (before RM) and after RM from both exposure streams and treat them as two empirical distributions in the feature domain. We then measure their discrepancy using feature-domain metrics, Fréchet distance (FD) [5] and maximum mean discrepancy (MMD) [7]. As shown in Tab. 1, after RM, both FD and MMD are consistently reduced across the evaluated sequences, indicating that the two feature distributions become closer in the feature domain.

Table 1. Distribution alignment effect of RM. We report FD and MMD between low- and high-exposure feature distributions on DeepHDRVideo-D [1] and Cinematic Video [6] under 2-exposure setting. Lower values indicate better alignment, and **bold** marks the best performance.

Model	DeepHDRVideo-D		Cinematic Video	
	FD	MMD	FD	MMD
before RM	2.10	0.11	2.08	0.11
after RM	**0.05**	**0.07**	**0.03**	**0.01**

We further visualize the feature distributions of different exposure features using t-distributed stochastic neighbor embedding (t-SNE) [12] on randomly sampled pixels at the EIC stage (before RM) and after RM. As shown in Fig. 1, before RM, the embeddings from the two exposures tend to form separated clusters, reflecting an exposure-dependent shift in the feature domain. After RM, the two groups become much more interleaved and lie on a tighter common manifold, with a clear increase in overlap between the clusters.

The above quantitative statistics and t-SNE visualizations jointly demonstrate that RM effectively reduces the discrepancy in the feature domain. By pulling the different exposure features toward a common manifold, RM provides ACCR with a more unified representation, which in turn facilitates robust fusion across alternating exposures.

1.2. Further Analysis on APSWF

Ablation on the activation in APSWF. APSWF predicts pixel-wise fusion weights to combine multiple HDR candidates, and the activation function determines how these weights are distributed across inputs. We therefore compare sigmoid, softmax, and entmax with different α values while keeping the rest of the architecture unchanged. We report PSNR_T and SSIM_T on DeepHDRVideo-D and Cinematic Video in Tab. 2. Among these choices, entmax with $\alpha = 1.75$ yields the best overall performance. This result suggests that a moderately sparse activation is more suitable for APSWF, as it suppresses unreliable candidates while still allowing multiple informative inputs to contribute to the final HDR estimate.

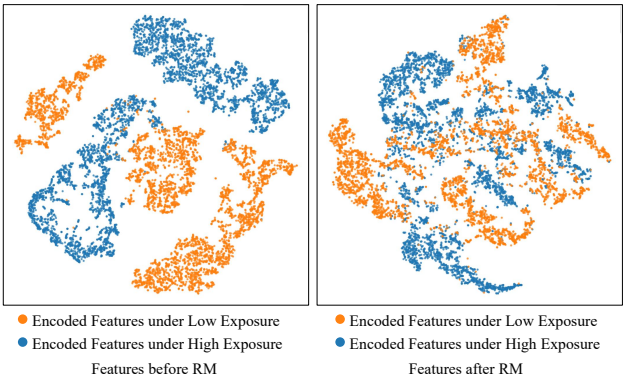

Figure 1. t-SNE visualization of features on Cinematic Video dataset [6].

Visualization of fusion masks. We further compare sigmoid and entmax-1.75 by visualizing their fusion masks and reconstruction results, as shown in Fig. 2. Rows (a) and (b) show the color-coded masks of each linear HDR candidate. Row (a) corresponds to masks produced with entmax, while row (b) shows masks produced with sigmoid. Specifically, from left to right in (a) and (b), they represent the reference frame, the ACCR reference candidate, the interpolated frame, the ACCR interpolated candidate, and two

Table 2. Ablation on the activation in APSWF. We compare sigmoid, softmax, and entmax with different α for predicting fusion masks on DeepHDRVideo-D [1] and Cinematic Video [6] under the 2-exposure setting. **Bold** marks the best result.

Model	DeepHDRVideo-D		Cinematic Video	
	PSNR_T	SSIM_T	PSNR_T	SSIM_T
sigmoid†	45.61	0.9729	40.90	0.9264
softmax††	45.52	0.9733	40.97	0.9246
entmax-1.5	45.51	0.9731	41.17	0.9263
entmax-2	45.80	0.9734	**41.31**	0.9271
entmax-1.75	**45.89**	**0.9753**	41.11	**0.9274**

† Sigmoid is used to predict fusion masks following [3, 13], and all other components of APSWF remain unchanged.
†† Softmax corresponds to entmax-1.

Figure 2. Qualitative comparison of fusion masks and HDR results with different activations in APSWF. First row: fusion masks of each linear HDR candidate for entmax-1.75. Second row: fusion masks of each linear HDR candidate for sigmoid. Bottom: reconstructed HDR crops.

adjacent frames, respectively. Sigmoid produces diffuse masks where multiple candidates are simultaneously active over large regions, which leads to ambiguous fusion near motion boundaries. In contrast, entmax-1.75 yields much sparser and more selective masks, where fewer candidates are activated for each pixel. The bottom row shows the fused HDR crops, and with sigmoid, the moving ping-pong paddle exhibits noticeable ghosting, while entmax-1.75 better preserves sharp edges and agrees more closely with the ground truth.

2. More Comparisons with Previous Methods

2.1. Two-exposure Setting

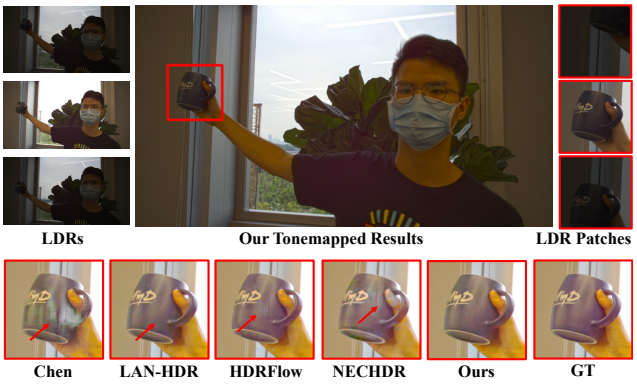

Figure 3. Qualitative comparisons on the DeepHDRVideo [6] dataset under 2-exposure setting. We brightened the results for easier comparison.

We provide additional visual comparisons with representative HDR video reconstruction methods, including Chen [1], LAN-HDR [2], HDRFlow [13], and NECHDR [3]. As shown in Fig. 3, in a scene with large inter-frame motion, competing methods exhibit more visible ghosting or mismatch artifacts, while LRHDR preserves the main structure more clearly in this example. Fig. 4 shows another challenging case where large motion occurs near high-contrast regions. In this scene, competing methods produce more noticeable distortions around motion boundaries, whereas LRHDR yields a cleaner reconstruction with more stable local structure. These examples suggest that LRHDR handles challenging dynamic scenes with strong exposure changes more robustly.

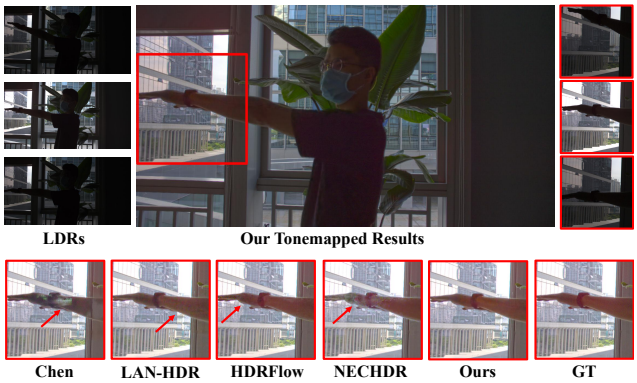

Figure 4. Qualitative comparisons on the DeepHDRVideo [6] dataset under 2-exposure setting. We brightened the results for easier comparison.

2.2. Three-exposure Setting

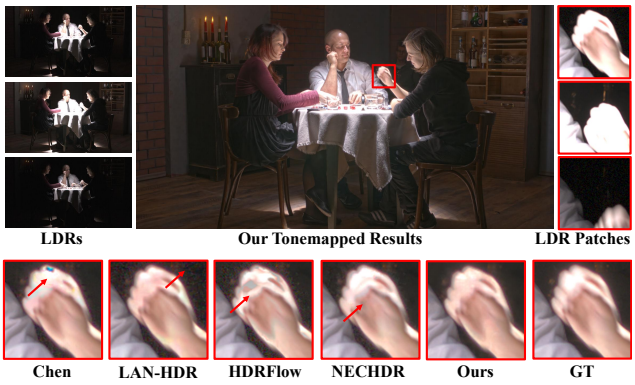

Figure 5. Qualitative comparisons on the Cinematic Video [6] dataset under 3-exposure setting, which is a scene with saturation and motion.

We also provide additional visual comparisons on the Cinematic Video dataset under the 3-exposure setting. As shown in Fig. 5, when motion occurs in saturated regions, competing methods exhibit more visible ghosting or instability, whereas LRHDR recovers more stable intensities in this example. Fig. 6 shows a scene with complex lighting and moving objects. In this case, competing methods produce more noticeable brightness fluctuation, while LRHDR yields visually more consistent reconstruction re-

Figure 6. Qualitative comparisons on the Cinematic Video [6] dataset under 3-exposure setting, which is a scene with complex lighting and motion.

sults. These observations suggest that LRHDR generalizes well to the 3-exposure setting and remains robust in challenging dynamic conditions.

These observations indicate that LRHDR generalizes well to 3-exposure videos and remains robust in challenging dynamic conditions.

3. Details for LRHDR under the Three-Exposure Setting

In the main paper, LRHDR is developed for videos captured with two alternating exposures. Here we describe how it is extended to sequences with three exposures.

Review of the two-exposure model. For the two-exposure setting, the exposure cycle alternates between e_0 and e_1, such as EV+3 and EV+0. Given three frames $\{L_{t-1}^{e_0}, L_t^{e_1}, L_{t+1}^{e_0}\}$, LRHDR reconstructs the HDR output of the reference frame $L_t^{e_1}$. First, the frames $L_{t-1}^{e_0}$ and $L_{t+1}^{e_0}$ are sent to the interpolator to produce a non reference exposure intermediate frame $\hat{L}_t^{e_0}$. Then $L_t^{e_1}$ and $\hat{L}_t^{e_0}$ are processed by two independent EIC encoders to obtain exposure aware features. RM maps these features into a unified representation. A pair of decoders with shared weights then converts this representation to the linear HDR domain. APSWF finally fuses all LDR frames together with their linear HDR candidates and outputs a high quality and ghost free HDR image at time t.

LRHDR for sequences with three exposures. For the three-exposure setting, the exposure cycle alternates between e_0, e_1 and e_2, such as EV-2, EV+0 and EV+2. LRHDR now uses five frames $\{L_{t-2}^{e_0}, L_{t-1}^{e_1}, L_t^{e_2}, L_{t+1}^{e_0}, L_{t+2}^{e_1}\}$ to reconstruct the HDR output of the reference frame $L_t^{e_2}$. First, the pairs $\{L_{t-2}^{e_0}, L_{t+1}^{e_0}\}$ and $\{L_{t-1}^{e_1}, L_{t+2}^{e_1}\}$ are fed to the interpolator to obtain the non reference exposure intermediates $\hat{L}_t^{e_0}$ and $\hat{L}_t^{e_1}$. Then the three frames $\hat{L}_t^{e_0}$, $L_t^{e_2}$ and $\hat{L}_t^{e_1}$ are encoded by three independent EIC encoders to produce $\hat{F}_t^{e_0}$, $F_t^{e_2}$ and $\hat{F}_t^{e_1}$. Two RM modules are used to map these features into the unified representation domain. RM_1 takes $\hat{F}_t^{e_0}$ and $F_t^{e_2}$, and its pair of weight sharing decoders generates linear HDR candidates for exposures e_0 and e_2. RM_2 takes $\hat{F}_t^{e_1}$ and $F_t^{e_2}$, and its decoders generate linear HDR candidates for exposures e_1 and e_2. APSWF then aggregates all input LDR frames and all linear HDR candidates and produces the final HDR reconstruction at time t under the three-exposure setting.

Table 3. Performance of different interpolators.

Model	Params(M)	Time(ms)	PSNR_T
Chen [1]	6.1	522	35.65
LANHDR [2]	7.3	461	38.22
HDRFlow [13]	3.3	55	39.30
NECHDR [3]	8.8	136	40.59
w/ FiLM [10]	41.1	656	41.11
w/ RIFE [8]	16.5	306	41.07
w/ UPR-Net [9]	8.4	371	41.12
w/ OpenCV [4]	6.7	–	40.30
w/ RAFT [11]	11.9	548	40.40

4. Efficiency and Computational Cost

Our main architectural additions are the RM module and sparse fusion, both of which introduce limited overhead. RM is lightweight, and sparse fusion provides a practical speed–quality trade-off by predicting fusion weights at a lower resolution. As shown in Tab. 3 and Fig. 7, replacing the frozen interpolator with UPRNet improves the efficiency–quality trade-off, this variant slightly surpasses NECHDR while using fewer parameters. In addition, the EIC encoders can be executed in parallel, which leaves fur-

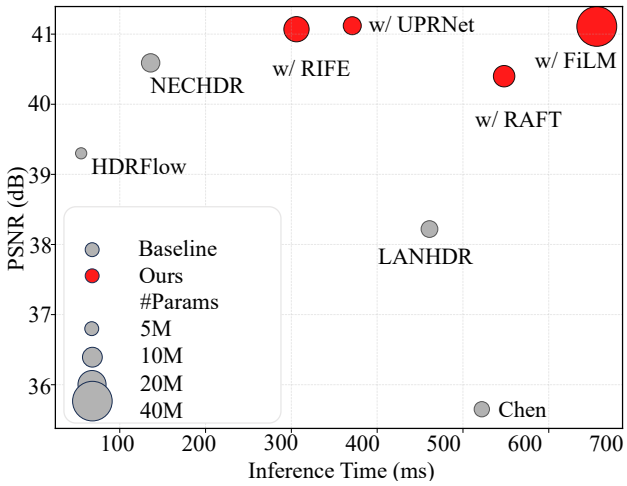

Figure 7. Efficiency-quality visualization of HDR models.

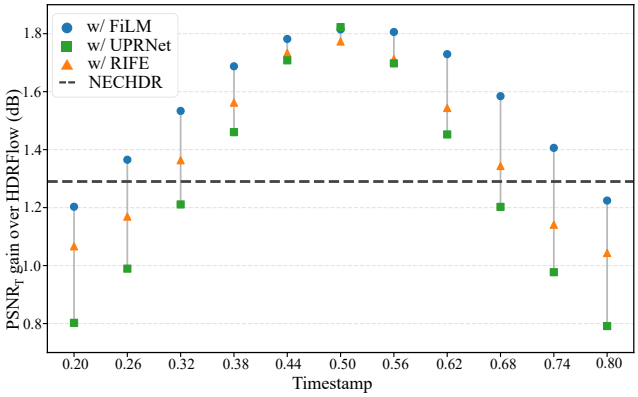

Figure 8. Robustness analysis of timestamps.

ther room for implementation-level speed optimization.

5. Robustness to Temporal Mismatch and Challenging Regions

We further evaluate the robustness of LRHDR to temporal mismatch in the interpolated pseudo-observation. As shown in Fig. 8, sweeping the interpolation timestamp around the center leads to only minor performance fluctuations, and LRHDR consistently maintains the best performance. This result indicates that the proposed framework is robust to small temporal deviations in the interpolated frame.

We also provide quantitative analysis on challenging regions affected by saturation and large motion. While full-frame averages may dilute localized improvements, LRHDR shows clearer gains in these difficult areas. As shown in Tab. 4, using identical 256×256 crops collected from saturated and large-motion regions, LRHDR achieves 45.84 dB PSNR$_T$ and 0.9899 SSIM$_T$, outperforming HDRFlow and NECHDR. These results show that the gains of LRHDR are concentrated in precisely those regions where alternating-exposure HDR reconstruction is most difficult.

The core idea of LRHDR is to decouple same-exposure motion handling from cross-exposure reconstruction. Instead of forcing cross-exposure pixel correspondences with a potentially degraded reference frame, LRHDR first obtains a motion-consistent pseudo-observation with a frozen same-exposure interpolator, and then performs cross-exposure complementarity through an explicit exposure-consistent representation. Building on this aligned representation, APSWF formulates reconstruction as sparse candidate selection rather than dense averaging, which suppresses unreliable candidates and reduces ghosting while preserving fine details.

Table 4. Region-level quantitative comparison. All methods use identical 256×256 crops across the test dataset from saturated, large-motion regions.

Method	HDRFlow	NECHDR	Ours
PSNR$_T$	42.08	44.63	**45.84**
SSIM$_T$	0.9802	0.9868	**0.9899**

6. Hyperparameter Clarifications

The loss weights are tuned on a held-out validation set and then fixed for all experiments. We observe low sensitivity to these hyperparameters: perturbing λ_1, λ_2, and λ_3 by $\pm 20\%$ around the selected values changes PSNR$_T$ by no more than 0.1 dB on average. These results indicate that the training objective is stable under moderate hyperparameter variations.

6.1. Failure Case under EXTREME NOISE

Figure 9 shows a representative failure case of LRHDR under EXTREME NOISE. In this scene, the input frames are contaminated by EXTREME NOISE, making both cross-exposure cue estimation and candidate fusion more difficult. Compared with NECHDR, LRHDR preserves the main hand structure more faithfully and avoids severe artifacts or color distortion. Nevertheless, residual noise and slight brightness instability still remain in very low-SNR regions, indicating the remaining limitation of LRHDR under extremely noisy conditions.

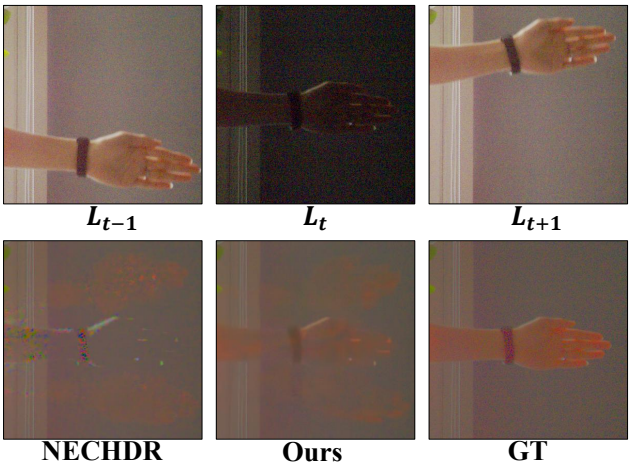

Figure 9. Failure case under EXTREME NOISE. For clarity, the input LDR frames in the top row are brightened by 75% for visualization only.

References

[1] Guanying Chen, Chaofeng Chen, Shi Guo, Zhetong Liang, Kwan-Yee K. Wong, and Lei Zhang. HDR video reconstruction: A coarse-to-fine network and a real-world benchmark dataset. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021. 1, 2, 3

[2] Haesoo Chung and Nam Ik Cho. Lan-hdr: Luminance-based alignment network for high dynamic range video reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12760–12769, 2023. 2, 3

[3] Jiahao Cui, Wei Jiang, Zhan Peng, Zhiyu Pan, and Zhiguo Cao. Exposure completing for temporally consistent neural high dynamic range video rendering. In *Proceedings of the 32nd ACM International Conference on Multimedia*, page 10027–10035, New York, NY, USA, 2024. Association for Computing Machinery. 1, 2, 3

[4] Gunnar Farnebäck. Two-frame motion estimation based on polynomial expansion. In *Proceedings of the 13th Scandinavian Conference on Image Analysis (SCIA)*, pages 363–370, 2003. 3

[5] Maurice Fréchet. Sur quelques points du calcul fonctionnel. *Rendiconti del Circolo Matematico di Palermo (1884-1940)*, 22:1–72, 1906. 1

[6] Jan Fröhlich, Stefan Grandinetti, Bernd Eberhardt, Simon Walter, Andreas Schilling, and Harald Brendel. Creating cinematic wide gamut hdr-video for the evaluation of tone mapping operators and hdr-displays. In *Digital Photography X, part of the IS&T-SPIE Electronic Imaging Symposium, San Francisco, California, USA, February 2, 2014, Proceedings*, page 90230X. SPIE/IS&T, 2014. 1, 2, 3

[7] Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *J. Mach. Learn. Res.*, 13(null):723–773, 2012. 1

[8] Zhewei Huang, Tianyuan Zhang, Wen Heng, Boxin Shi, and Shuchang Zhou. Real-time intermediate flow estimation for video frame interpolation. In *Proceedings of the European Conference on Computer Vision*, 2022. 3

[9] Xin Jin, Longhai Wu, Jie Chen, Youxin Chen, Jayoon Koo, and Cheul hee Hahm. A unified pyramid recurrent network for video frame interpolation, 2023. 3

[10] Fitsum Reda, Janne Kontkanen, Eric Tabellion, Deqing Sun, Caroline Pantofaru, and Brian Curless. Film: Frame interpolation for large motion. In *Proceedings of the European Conference on Computer Vision*, 2022. 3

[11] Zachary Teed and Jia Deng. Raft: Recurrent all-pairs field transforms for optical flow. In *Computer Vision – ECCV 2020*, pages 402–419, Cham, 2020. Springer International Publishing. 3

[12] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9 (86):2579–2605, 2008. 1

[13] Gangwei Xu, Yujin Wang, Jinwei Gu, Tianfan Xue, and Xin Yang. Hdrflow: Real-time hdr video reconstruction with large motions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24851–24860, 2024. 1, 2, 3