

Exposure Completing for Temporally Consistent Neural High Dynamic Range Video Rendering Supplementary Materials

Anonymous Authors

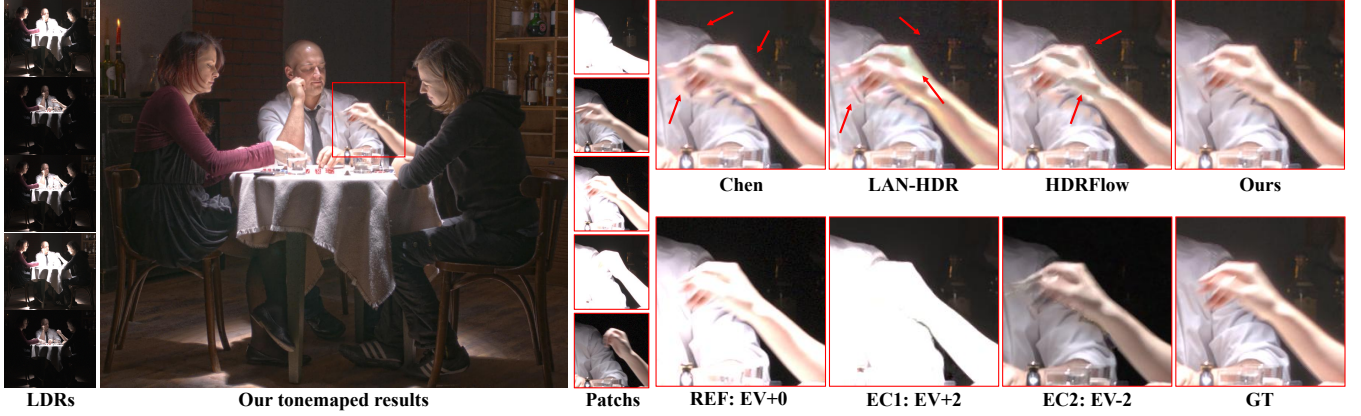


Figure 1: Qualitative comparisons of the three-exposure setting on the Cinematic Video dataset [3]. We compare our proposed method with the current state-of-the-art methods: Chen [1], LAN-HDR [2], HDRFlow [6]. REF refers to reference frame. “EV” refers to the exposure value. “EC1” and “EC2” mean the first and second completed frames with the missing exposure.

1 VIDEO RESULTS

We provide a 3-minute video comparing our HDR videos with results by other methods as an attachment in this supplementary material. One can download and play for a better view of our method.

2 NETWORK ARCHITECTURE

We provide details of the NECHDR network we proposed, mainly including the following sub-networks: feature encoder \mathcal{E} , exposure completing decoder \mathcal{D}_I , k -levels HDR rendering decoders $\{\mathcal{D}_R^k | k = 1, 2, 3, 4\}$ and blending network. Both The feature encoder \mathcal{E} and exposure-completing decoder \mathcal{D}_I are weight-shared in the two-exposure setting or the three-exposure setting. We take input LDR frames with a size of 256×256 as example and illustrate in figures for above sub-networks. In Fig. 2, 3, 4, 5, 6, 7, the parameters for “conv” and “deconv”, listed from left to right, are input channels, output channels, kernel size, stride, and padding. In our network, each “Conv” is followed by a PReLU [4], whereas there is no activation subsequent to each ‘Deconv’ layer.

In the two-exposure setting, the parameter shared feature encoder \mathcal{E} in Fig. 2 outputs pyramid features for the three input LDR frames $\{l_{t-1}, l_t, l_{t+1}\}$, while for the five input LDR frames $\{l_{t-2}, l_{t-1}, l_t, l_{t+1}, l_{t+2}\}$ in the three-exposure setting. As for the exposure completing decoder \mathcal{D}_I in the Fig. 3, we adopt the same structure and parameters in both settings. Particularly, in the three-exposure setting, the parameter shared decoder \mathcal{D}_I is utilized twice, producing two interpolated frames $\{\hat{l}_t^{\epsilon_1}, \hat{l}_t^{\epsilon_2}\}$ and theirs corresponding features $\widehat{\Phi}_{t,1}^k = \{\widehat{\phi}_{t,1}^k | k = 1, 2, 3\}$, $\widehat{\Phi}_{t,2}^k = \{\widehat{\phi}_{t,2}^k | k = 1, 2, 3\}$. The details of HDR render decoder \mathcal{D}_I^k in k levels, where $k = 1, 2, 3, 4$, are shown in Fig. 7, 6, 5, 4, respectively. In each of the Fig. 7, 6, 5, 4, the left illustrates the k -th level \mathcal{D}_I^k under two-exposure setting,

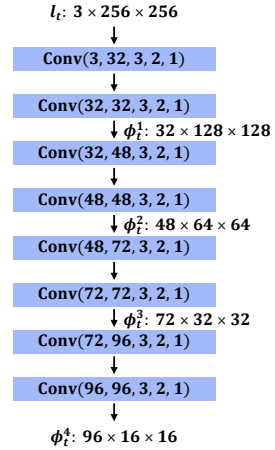
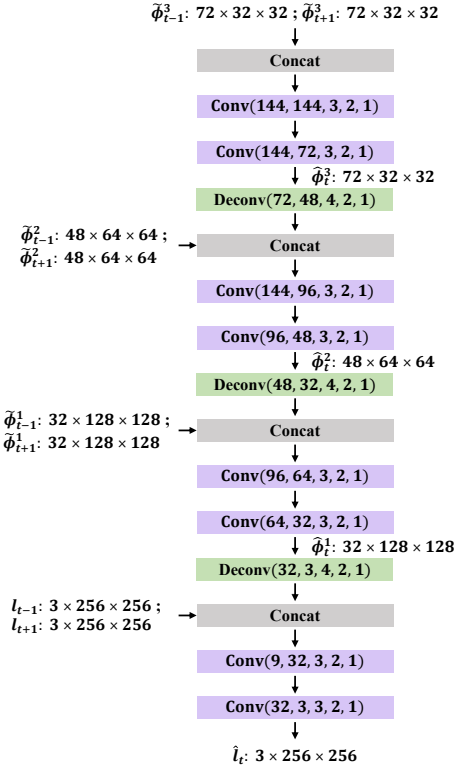


Figure 2: Details of the feature encoder \mathcal{E}

while the right is under three-exposure setting. The details regarding the blending network can be found in [6].

3 MORE QUALITATIVE COMPARISONS

Under the three-exposure setting, we provide more qualitative comparison results in Fig. 1 and Fig. 8 between our method and the current state-of-the-art approaches: Chen [1], LAN-HDR [2], HDRFlow [6]. As shown in Fig. 1, we obtain the completed frames “EC1” and “EC2”. The “EC1” provides high-exposure information with high signal-to-noise ratio to assist in removing noise from the dark background of the reference frame, while “EC2” offers low-exposure foreground object details to restore details missing due to overexposure in the reference frame. Therefore, our proposed method

Figure 3: Details of the exposure completing decoder \mathcal{D}_I

achieves the HDR rendering result closest to ground truth. We also provide two scenarios in Fig. 8, where motions occur with saturation and noise respectively. In the left image of Fig. 8, although “EC1” is not well completed due to the first and fourth input frames with large noise and motion, the well-completed “EC2” provides information about the saturated regions for HDR rendering, resulting in ghost-free HDR result. In the right image of Fig. 8, the completed frame “EC2” provides detailed information with low noise, thus achieving the lowest noise level in these methods. The above visualization results elucidate the performance enhancement of our method under the three-exposure setting, further proving the robustness of our approach across different exposure settings.

4 A LIGHTWEIGHT VERSION OF NECHDR

In both settings, we offer lightweight versions of the NECHDR network by reducing the number of feature channels. In the lightweight versions, feature channels from the first to the fourth levels in feature encoder \mathcal{E} is set to 24, 36, 54, 72, respectively. The channels of both the exposure completing decoder \mathcal{D}_I and HDR rendering decoder \mathcal{D}_R are adjusted accordingly. Furthermore, the number of feature channels in the third and fifth convolution layers of HDR rendering decoder \mathcal{D}_R will be adjusted to 24. As shown in Table 1, reducing the number of feature channels significantly decreases the parameter and inference time of our model, with no noticeable change in performance.

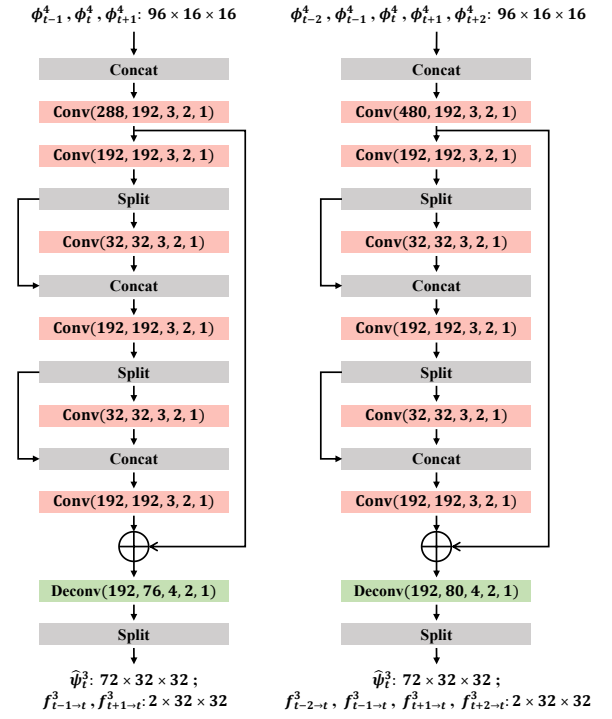
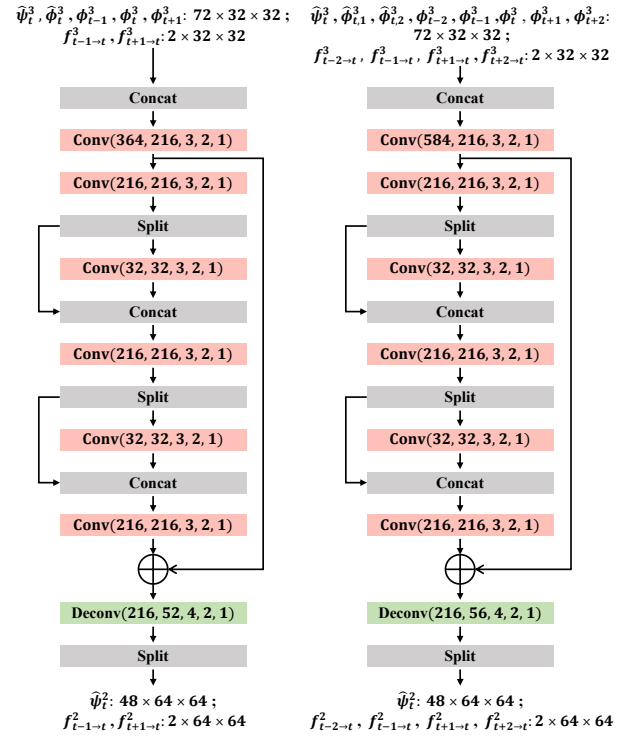
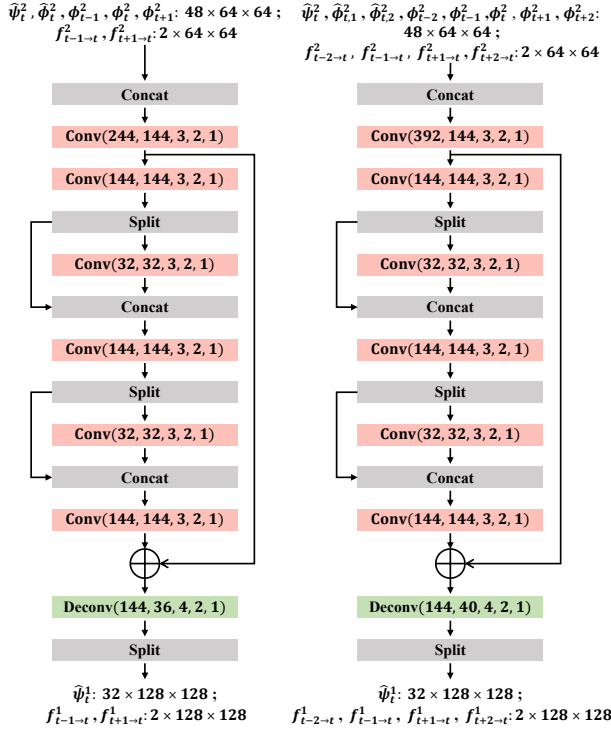
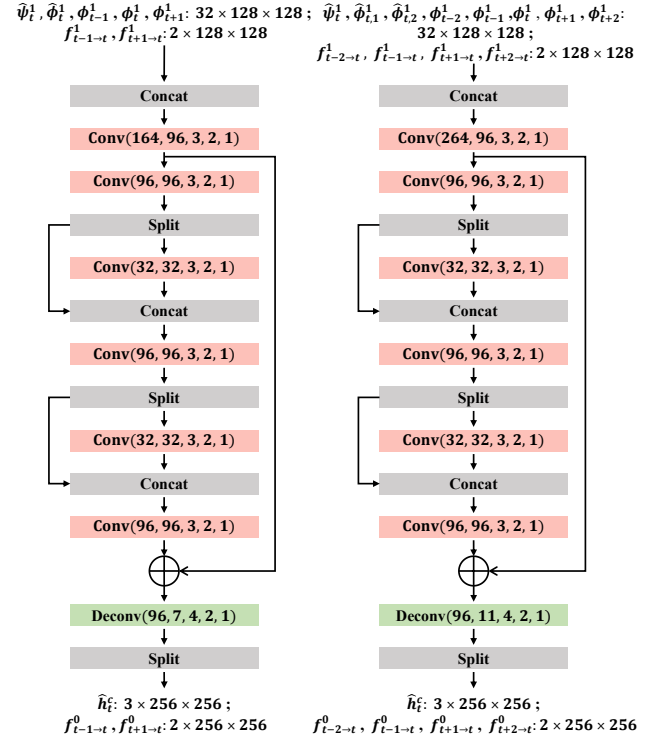
Figure 4: Details of the HDR rendering decoder \mathcal{D}_R^4 . Left: two-exposure; right: three-exposure.Figure 5: Details of the HDR rendering decoder \mathcal{D}_R^3 . Left: two-exposure; right: three-exposure.

Table 1: Quantitative comparisons on the CinematicVideo dataset [3] about parameter, inference time and performance between our “NECHDR” and the lightweight version “NECHDR-lw”.

	two-Exposures				three-Exposures			
	Para. (M)	Time (s)	PSNR _T	SSIM _T	Para. (M)	Time (s)	PSNR _T	SSIM _T
NECHDR	8.8	0.178	40.59	0.9241	9.9	0.163	37.24	0.9102
NECHDR-lw	4.2 (↓ 52%)	0.138 (↓ 23%)	40.33 (↓ 0.7%)	0.9217 (↓ 0.3%)	4.8 (↓ 52%)	0.130 (↓ 20%)	37.11 (↓ 0.3%)	0.9053 (↓ 0.5%)

**Figure 6: Details of the HDR rendering decoder \mathcal{D}_R^2 . Left: two-exposure; right: three-exposure.****Figure 7: Details of the HDR rendering decoder \mathcal{D}_R^1 . Left: two-exposure; right: three-exposure.**

REFERENCES

[1] Guanying Chen, Chaofeng Chen, Shi Guo, Zhetong Liang, Kwan-Yee K Wong, and Lei Zhang. 2021. HDR video reconstruction: A coarse-to-fine network and a real-world benchmark dataset. In *Proc. Int. Conf. Comput. Vis.* 2502–2511.

[2] Haesoo Chung and Nam Ik Cho. 2023. LAN-HDR: Luminance-based Alignment Network for High Dynamic Range Video Reconstruction. In *Proc. Int. Conf. Comput. Vis.* 12760–12769.

[3] Jan Froehlich, Stefan Grandinetti, Bernd Eberhardt, Simon Walter, Andreas Schilling, and Harald Brendel. 2014. Creating cinematic wide gamut HDR-video

for the evaluation of tone mapping operators and HDR-displays. In *Digital photography X*, Vol. 9023. SPIE, 279–288.

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proc. Int. Conf. Comput. Vis.* 1026–1034.

[5] Nima Khademi Kalantari, Eli Shechtman, Connelly Barnes, Soheil Darabi, Dan B Goldman, and Pradeep Sen. 2013. Patch-based high dynamic range video. *ACM Trans. Graph.* 32, 6 (2013), 202–1.

[6] Gangwei Xu, Yujin Wang, Jinwei Gu, Tianfan Xue, and Xin Yang. 2024. HDRFlow: Real-Time HDR Video Reconstruction with Large Motions. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*

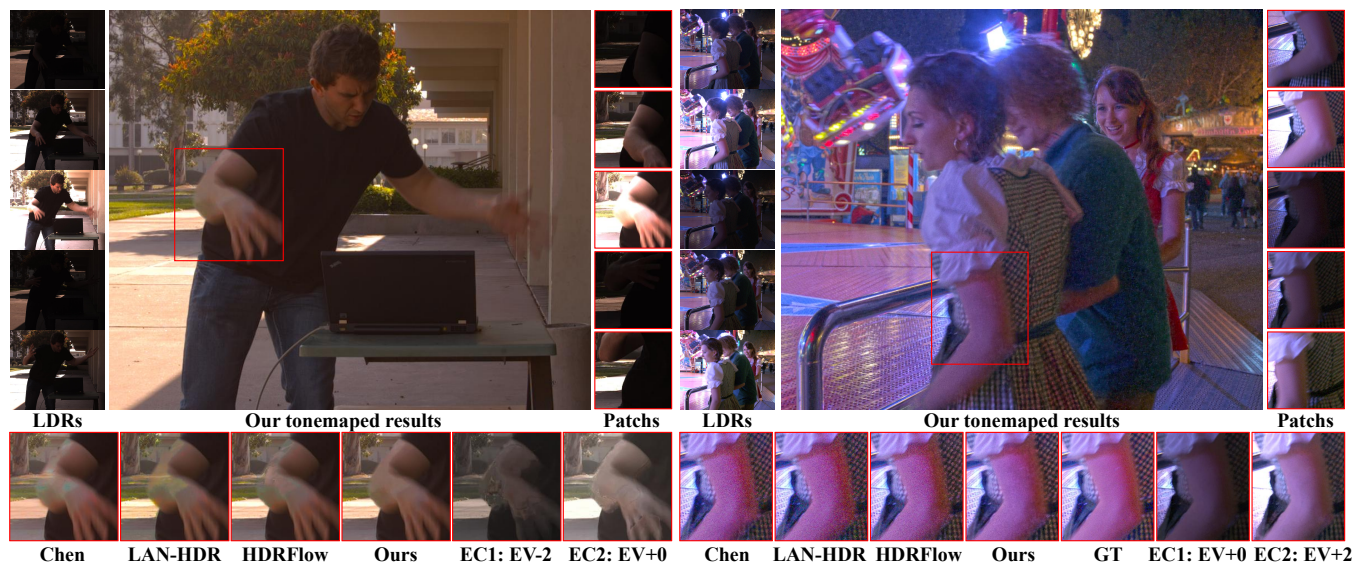


Figure 8: Qualitative comparisons of the three-exposure setting on the Cinematic Video dataset [3] and the HDRVideo dataset [5]. EC1 and EC2 mean the first and second completed frames with the missing exposure.