# Exposure Completing for Temporally Consistent Neural High Dynamic Range Video Rendering

Anonymous Authors



(a) Indoor scene with saturation and motions

(b) Outdoor scene with noise and motions.

Figure 1: Qualitative comparison between the proposed method and HDRFlow [40] on two different scenes. The input low dynamic range (LDR) videos consist of frames with alternate exposures, meaning that the exposure changes and part of exposure information is missing at every time stamp (1-st row). Prior arts, *e.g.*, the HDRFlow, struggle to achieve temporally consistent high dynamic range (HDR) results due to the exposure change and the motion that causes information loss (2-nd row). In contrast, the proposed method achieves temporally consistent HDR reconstruction results (4-th row) by completing the absent exposure frames (3-rd row) at each time stamp. "EV": the exposure value for each input LDR frame. "EC": our exposure completing results for the absent exposure corresponding to the input LDR frames.

# ABSTRACT

High dynamic range (HDR) video rendering from low dynamic range (LDR) videos where frames are of alternate exposure encounters significant challenges, due to the exposure change and absence at each time stamp. The exposure change and absence make existing methods generate flickering HDR results. In this paper, we propose a novel paradigm to render HDR frames via completing the absent exposure information, hence the exposure information

**Unpublished working draft. Not for distribution.** for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission. is complete and consistent. Our approach involves interpolating neighbor LDR frames in the time dimension to reconstruct LDR frames for the absent exposures. Combining the interpolated and given LDR frames, the complete set of exposure information is available at each time stamp. This benefits the fusing process for HDR results, reducing noise and ghosting artifacts therefore improving temporal consistency. Extensive experimental evaluations on standard benchmarks demonstrate that our method achieves state-of-the-art performance, highlighting the importance of absent exposure completing in HDR video rendering. The code will be made publicly available upon the acceptance of this paper.

# **CCS CONCEPTS**

 $\bullet$  Computing methodologies  $\rightarrow$  Artificial intelligence; Computer vision; Image and video acquisition; Computational photography.

and/or a fee. Request permissions from permissions@acm.org.

ACM MM, 2024, Melbourne, Australia

<sup>© 2024</sup> Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM

https://doi.org/10.1145/nnnnnnn.nnnnnn

# KEYWORDS

117

118

119

120

121

123

124

125

126

127

128

129

130

131

132

133

HDR video, Temporal consistency, Exposure completing

#### 1 INTRODUCTION

Compared with the high dynamic range (HDR) of natural scenes, the standard digital camera can only capture low dynamic range (LDR) information. To achieve the goal of "what you see is what you get" in photography, delicate optical systems [22, 33, 38] are designed to simultaneously capture LDR images with different exposures which cover the whole dynamic range to reconstruct HDR results. However, these systems are expensive and un-portable, making HDR contents not easily accessible for general users [18]. To address this limitation, computational methods [17, 18, 29] are designed to render HDR videos from LDR videos whose frames are exposed alternately with different exposures (as shown in 1-st row in Fig. 1).

In this setting, the incomplete exposure information varies along 134 135 the time dimension. Existing neural HDR rendering methods [3, 6, 15, 40] align neighbor LDR frames according to the LDR frame 136 of current time stamp (called reference frame in this paper) and 137 138 fuse the aligned neighbor LDR frames into the reference frame to 139 supplement the missing exposure information. The exposure of reference frame changes at every time stamp, which means that the 140 reference frames of different exposures may have different defects: 141 saturation for highly exposed LDR frames, and noise for lowly 142 exposed ones. Due to that the methods mentioned above heavily 143 depend on reference frames, HDR results from these methods may 144 inherit defects of the reference frames, suffering from the artifacts, 145 i.e., the ghosting (2-nd row in Fig. 1(a)) and noise (2-nd row in 146 Fig. 1(b)), thereby causing temporal inconsistency. 147

To tackle this problem, we revisit early traditional methods [17, 148 149 18] for HDR rendering and are motivated to complete the missing exposure information for each time stamp [17, 18]. This idea fol-150 151 lows the philosophy of optical HDR systems [38] that are designed 152 to obtain different exposed LDR frames at exact the same time. In this way, defects of different exposed frames can be mutually 153 compensated. In this paper, we propose the Neural Exposure Com-154 155 pleting HDR (NECHDR) framework to reconstruct the LDR frames with missing exposure information. In this way, full exposure in-156 formation at each time stamp can be covered by the reference and 157 completed LDR frames, therefore the exposure information is com-158 159 plete and consistent along the time dimension, which benefits the temporal consistency of the rendered HDR videos. 160

In the proposed NECHDR framework, pyramid features of input 161 LDR frames are extracted by the feature encoder, then fed into the 162 exposure completing decoder and the HDR rendering decoder. The 163 exposure completing decoder interpolates the features of neighbor 164 LDR frames at every level of the feature pyramid. The interpolated 165 LDR features are combined with the features from input LDR im-166 ages as the inputs to HDR rendering decoder. The HDR rendering 167 168 decoder estimates coarse HDR results and the optical flows at every level. The flows can facilitate feature interpolating in exposure com-169 pleting decoder. At the end of NECHDR, a simple blending network 170 is used to integrate interpolated LDR frames, input frames, and 171 172 coarse HDR frames, which can achieve high-quality and temporally 173 consistent HDR reconstruction results.

Anonymous Authors

175

176

177

178

Extensive experimental results on multiple public benchmarks demonstrate the superiority of the proposed NECHDR: by completing the missing exposure information (3-rd row in Fig. 1), our method mitigates ghosting resulting from large motions for the time stamps with highly exposed reference frames (4-th row in Fig. 1(a)), and reduces the noise level for the lowly exposed reference frames (4-th row in Fig. 1(b)), hence achieves better temporal consistency. Our work sheds light again on the exposure completing for HDR video rendering. The contributions can be summarized as follows:

• Our work firstly implements the idea of exposure completing for neural HDR rendering.

• Our work proposes a novel HDR video rendering framework, *a.k.a.*, the NECHDR, which completes the missing exposure information by interpolating LDR frames.

• Our NECHDR achieves new state-of-the-art performance on current benchmarks.

#### 2 RELATED WORK

#### 2.1 HDR Image Rendering

HDR imaging technology aims to extend the dynamic range of images by merging LDR images with different exposures. Both the HDR image and video tasks encounter the misalignment between LDR images. Pioneer studies for HDR image rendering employ image alignment techniques to address this issue. Early works [9, 12, 19, 24, 35, 47] often align LDR images globally, then detect and discard unaligned regions. Directly discarding these regions leads to significant information loss, posing challenges for HDR image rendering. To address these issues, optical-flow-based [1, 14, 49] and patch-based methods [13, 28, 37] are proposed. However, large motion still makes these methods fail to produce artifactfree HDR results. With the rapid development of deep learning, many studies for HDR image rendering switch their focus to neural networks [2, 5, 16, 25, 27, 34, 39, 42-44], achieving promising results. Kalantari et al. [16] propose using neural networks to align input LDR images with predicted optical flow. Wu et al. [39] directly map LDR images to HDR images, thereby avoiding alignment errors. Yan et al. [43, 45] use spatial attention to implicitly align LDR images, achieving further improvement. Liu [27] et al.introduce a ghost-free imaging model based on swin-transformer [26]. Yan et al. [42] utilize patch-based and pixel-based fusion to search for information to complement the reference frame from the other frames. However, these HDR image rendering methods assume a fixed exposure for reference frames, typically medium exposure, therefore is not suitable for HDR video rendering where reference frames have alternating exposures.

#### 2.2 HDR Video Rendering

HDR video can be obtained by delicate optical systems, such as scanline exposure/ISO [4, 11], internal/external beam splitter [23, 31, 38], modulo camera [48] and neuromorphic camera [10]. However, these expensive systems are hardly accessible to general users. Therefore, rendering HDR videos from LDR videos is investigated. Kang *et al.* [18] utilize both global and local alignment for reference frames with neighbor frames. Kalantari *et al.* [17] propose a patch-based optimization algorithm to reconstruct HDR video by synthesizing missing exposures for each frame. These traditional methods offer

230

231

232



Figure 2: Pipline of the proposed NECHDR network. Our network mainly consists of three processes: feature encoding for LDRs, exposure completing for LDR frames with missing exposures, and HDR rendering. We extract pyramid features from the input LDR frames using a parameter shared feature encoder. Then, the optical flows predicted by the lower-level HDR rendering decoder is used to warp neighbor frames to the reference frame. Subsequently, the exposure completing decoder performs exposure completing based on the warped neighbor frame features. The input features and the completed feature are fed into the HDR rendering decoder to render the coarse HDR frame. Finally, the original frames, the completed frame, and the coarse HDR frame are fused together through a simple blending network to produce a high-quality HDR result.

a fundamental idea for completing missing exposures, but they require iterative optimization and are prone to producing artifacts. Recently, Kalantari et al. [15] introduce an end-to-end neural net-work consisting of an optical flow prediction model and a fusion network. Chen et al. [3] propose a coarse-to-fine alignment frame-work, which utilizes optical flow for coarse alignment and then employs deformable convolution [7] for fine alignment. Chung et al. [6] employ spatial attention instead of optical flow to achieve alignment from adjacent frames to the reference frame. Xu et al. [40] propose the HDR-domain loss, which uses optical flow supervised by HDR frames for LDR alignment, yielding favorable results. How-ever, when complex situations such as large motion, saturation, and others occur simultaneously, these methods struggle to achieve precise alignment, resulting in artifacts in the final outputs. Besides, as the exposures of reference frames alternate, the reconstructed HDR frames tend to inherit the defects of reference frames, leading to flicker in the output videos. The fundamental issue lies in the lack of exposure information at every time stamp. Different from 

previous methods, we propose a renaissance work which focuses on completing missing exposures for the reference frame in neural HDR video rendering. Through compensating for the defects of the alternatively exposed reference frames by utilizing the completed results, our approach avoids artifacts of ghost and noise and obtains a temporally consistent HDR video result.

# **3 PROPOSED METHODS**

#### 3.1 Overview

Each frame  $l_i$  in the LDR video  $L = \{l_i | i = 1, ..., n\}$  is expected to have multiple exposures  $\mathbf{E} = \{\epsilon_i | i = 1, ..., z\}$  to cover the whole dynamic range of HDR video  $H = \{h_i | i = 1, ..., z\}$  to cover the whole length of video, and z is the number of exposures. However, in our task, only one exposure can be obtained in one certain time stamp, which creates an exposure sequence  $E = \{e_i | i = 1, ..., n\}$  for the LDR video, where  $e_i = \epsilon_{(i \oslash z)+1}$ .  $a \oslash b$  indicates the remainder of a divided by b.

Following previous work [40], we try to render a high-quality 349 HDR video with LDR video under two settings of different alternate 350 351 exposures: two alternate exposures (z = 2) with three frames as input, and three alternate exposures (z = 3) with five frames. We 352 353 use the two-exposure setting as an example to demonstrate how to achieve HDR rendering. In Sec. 3.5, we introduce how to extend our 354 approach to the three-exposure setting. The frame  $l_t$  corresponding 355 to the current time stamp t, is treated as the reference frame.  $l_t$  is 356 combined with its neighbors as input  $\{l_{t-1}, l_t, l_{t+1}\}$ . To render HDR 357 358 frames, we explicitly complete frames  $\hat{l}_t$  with absent exposure  $\epsilon_i$ , 359 where  $i \neq (t \otimes z)+1$ . Therefore, with complete exposure information, 360 we can render a high-quality HDR video.

361 Our method consists of two processes: exposure completing for the frame with the absent exposure at the time stamp t and 362 363 coarse-to-fine HDR rendering. These two processes are coupled 364 together through an encoder-decoder architecture, as illustrated in 365 the Fig. 2. The feature encoder  $\mathcal{E}$  takes LDR frames  $\{l_{t-1}, l_t, l_{t+1}\}$ 366 as input and generates pyramid features for each frame. As the 367 spatial dimensions decrease and the feature channels increase, the 368 encoder outputs feature sets  $\{\Phi_{t-1}, \Phi_t, \Phi_{t+1}\}$  for input frames, each 369 of which is a set of *k*-level pyramid features  $\Phi = \{\phi^k | k = 1, 2, 3, 4\}$ . 370 Considering the distinct characteristics of the two processes, we 371 devise two dedicated decoders. Initially, an exposure completing de-372 coder  $\mathcal{D}_I$  is employed to interpolate frame  $\hat{l}_t$  and its corresponding 373 features  $\widehat{\Phi}_t^k = \{ \widehat{\phi}_t^k | k = 1, 2, 3 \}$ . Subsequently, a coarse-to-fine HDR 374 rendering decoder  $\mathcal{D}_R$  takes { $\Phi_{t-1}, \Phi_t, \Phi_{t+1}$ } and  $\widehat{\Phi}_t$  as input, and outputs optical flows  $F_{t-1 \to t} = \{f_{t-1 \to t}^k | k = 0, 1, 2, 3\}, F_{t+1 \to t} =$ 375 376  $\{f_{t+1\to t}^k | k = 0, 1, 2, 3\}$ , coarse HDR features  $\widehat{\Psi}_t = \{\widehat{\psi}_t^k | k = 1, 2, 3\}$ 377 and coarse HDR frame  $\hat{h}_{t}^{c}$ . The optical flows here are utilized to warp 378  $\{\Phi_{t-1}, \Phi_{t+1}\}$  or  $\{l_{t-1}, l_{t+1}\}$  to time stamp t, yielding warped neigh-379 bor features  $\widetilde{\Phi}_{t-1} = \{ \widetilde{\phi}_{t-1}^k | k = 1, 2, 3 \}, \widetilde{\Phi}_{t+1} = \{ \widetilde{\phi}_{t+1}^k | k = 1, 2, 3 \}$ 380 381 or frames  $\{\tilde{l}_{t-1}, \tilde{l}_{t+1}\}$ . The warped neighbor features  $\{\tilde{\Phi}_{t-1}, \tilde{\Phi}_{t+1}\}$ 382 and frames  $\{l_{t-1}, l_{t+1}\}$  are fed into  $\mathcal{D}_I$ . Subsequently, the exposure 383 completing features  $\widehat{\Phi}_t$  from  $\mathcal{D}_I$  are fed into the HDR rendering 384 decoder  $\mathcal{D}_R$ , assisting it in better restoring lost details due to satu-385 ration and noise when decoding coarse HDR features  $\widehat{\Psi}_t$  and coarse 386 HDR frame  $\hat{h}_t^c$ . This design coupling two decoders enables mutual 387 enhancement of the two processes, thus obtaining more accurate 388 exposure completing frame  $\hat{l}_t$  and better coarse HDR result  $\hat{h}_t^c$ . To 389 obtain the final HDR rendering result, we map  $\{l_{t-1}, l_t, l_{t+1}\}$  and  $\hat{l}_t$ 390 to the linear HDR domain. The function defining the mapping of 391 LDR frames to the linear HDR domain is as follows: 392

$$\kappa_t = l_t^\gamma / e_t, \tag{1}$$

where  $\gamma$  is a hyperparameter and  $e_t$  is the exposure time of  $l_t$ . Finally, the ultimate high-quality HDR rendering result  $\hat{h}_t$  is obtained by fusing the input and completed frames with coarse HDR frame in linear HDR domain using a simple blending network. The details of each process will be presented in the following sections.

# 3.2 Exposure Completing

393

394

395

396

397

398

399

400

401

Under the setting that input LDR frames are with alternating expo sures, completing the absent exposure information is essential for
 rendering HDR frames. However, existing neural HDR rendering
 methods overlook this essential problem, making it difficult for

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

460

461

462

463

464

these methods to accurately recover detailed information when motion occurs with saturation or noise simultaneously, potentially leading to artifacts. Furthermore, due to these methods heavily depend on reference frames, the alternate appearance of noise and saturation caused by the alternate exposures in reference frames affects the temporal consistency of the video. Therefore, to better address the aforementioned issues, we explicitly tackle the fundamental problem of HDR rendering by completing frames with absent exposure corresponding to reference frames.

Given the LDR frame input { $l_{t-1}$ ,  $l_t$ ,  $l_{t+1}$ } with alternate exposures { $e_{t-1}$ ,  $e_t$ ,  $e_{t+1}$ }, where the reference frame and neighbor frames have two different exposures ( $e_t = \epsilon_1$  and  $e_{t-1} = e_{t+1} = \epsilon_2$ ), our task is to complete the frame  $\hat{l}_t$  with absent exposure  $\epsilon_2$  corresponding to reference frame through interpolation.

Our exposure completing process takes the neighbor features  $\{\Phi_{t-1}, \Phi_{t+1}\}\$  and the optical flows  $\{F_{t-1 \rightarrow t}, F_{t+1 \rightarrow t}\}\$  as input. During this process for the features with absent exposure information, the *k*-th level neighbor features  $\{\phi_{t-1}^k, \phi_{t+1}^k\}\$  and optical flows  $\{f_{t-1 \rightarrow t}^k, f_{t+1 \rightarrow t}^k\}\$  are processed to obtain completed feature  $\widehat{\phi}_t^k$ :

$$\widetilde{\phi}_{t-1}^k, \widetilde{\phi}_{t+1}^k = \mathcal{W}(\phi_{t-1}^k, f_{t-1 \to t}^k), \mathcal{W}(\phi_{t+1}^k, f_{t+1 \to t}^k), \qquad (2)$$

$$\widehat{\phi}_t^3 = \mathcal{D}_I^3 \left( \left[ \widetilde{\phi}_{t-1}^3, \widetilde{\phi}_{t+1}^3 \right] \right), \tag{3}$$

$$\widehat{\phi}_{t}^{k} = \mathcal{D}_{I}^{k} \left( \left[ \widetilde{\phi}_{t-1}^{k}, \widetilde{\phi}_{t+1}^{k}, \mathcal{U}_{2}(\widehat{\phi}_{t}^{k+1}) \right] \right), \tag{4}$$

where  $\mathcal{W}(\cdot, \cdot)$  denotes using optical flow to warp neighbor features to the reference feature,  $\mathcal{U}_2$  represents the bilinear upsampling operation with scale factor 2,  $\mathcal{D}_I^k$  (k = 1, 2) represents the middle levels of the exposure completing decoder  $\mathcal{D}_I$ , and  $[\cdot]$  indicates concatenation operation. In the final step of the exposure completing process, the completed frame  $\hat{l}_t$  is explicitly interpolated with neighbor frames  $l_{t-1}, l_{t+1}$  and optical flow  $f_{t-1 \to t}^0, f_{t+1 \to t}^0$  as input:

$$\widetilde{l}_{t-1}, \widetilde{l}_{t+1} = \mathcal{W}(l_{t-1}, f^0_{t-1 \to t}), \mathcal{W}(l_{t+1}, f^0_{t+1 \to t}),$$
(5)

$$\widehat{U}_{t} = \mathcal{D}_{I}^{0}\left(\left|\widetilde{l}_{t-1}, \widetilde{l}_{t+1}, \mathcal{U}_{2}(\widehat{\phi}_{t}^{1})\right|\right), \tag{6}$$

where  $\mathcal{D}_{I}^{0}$  represents the highest level of the exposure completing decoder  $\mathcal{D}_{I}$ .

As illustrated in the Fig. 1, with the above design, we have obtained realistic exposure completing result for the challenging areas with motion, saturation and noise, providing accurate and complete exposure information for subsequent HDR rendering.

# 3.3 Coarse-to-fine HDR Rendering

Building upon the results of above process, we designed a coarse-tofine HDR rendering network based on completed features  $\widehat{\Phi}_t$  and frame  $\widehat{l}_t$ . Taking the original features  $\{\Phi_{t-1}, \Phi_t, \Phi_{t+1}\}$ , completed features  $\widehat{\Phi}_t$  and warped neighbor features  $\{\widetilde{\Phi}_{t-1}, \widetilde{\Phi}_{t+1}\}$  as input, the HDR rendering decoder  $\mathcal{D}_R$  reconstructs HDR features  $\widehat{\Psi}_t$  and optical flows  $\{F_{t-1 \rightarrow t}, F_{t+1 \rightarrow t}\}$  in a coarse-to-fine manner, and Exposure Completing for Temporally Consistent Neural High Dynamic Range Video Rendering

ACM MM, 2024, Melbourne, Australia

finally obtains the coarse HDR frame  $\hat{h}_t^c$ :

$$\widehat{\psi}_{t}^{3}, f_{t-1 \to t}^{3}, f_{t+1 \to t}^{3} = \mathcal{D}_{R}^{4} \left( \left[ \phi_{t-1}^{4}, \phi_{t}^{4}, \phi_{t+1}^{4} \right] \right), \tag{7}$$

$$\psi_t^{\kappa-1}, f_{t-1 \to t}^{\kappa-1}, f_{t+1 \to t}^{\kappa-1} = \sum_{k \in \mathcal{K}} \left( \left[ \frac{k}{k} \hat{k}_k \hat{k}_k \hat{k}_k f_k f_k f_k f_k \hat{k}_k \hat{k}_k \hat{k}_k \hat{k}_k \right] \right)$$

$$\mathcal{D}^{\kappa}\left(\left[\phi_{t}^{\kappa},\phi_{t}^{\kappa},\psi_{t}^{\kappa},f_{t-1\to t}^{\kappa},f_{t+1\to t}^{\kappa},\phi_{t-1}^{\kappa},\phi_{t+1}^{\kappa}\right]\right),$$
(8)  
$$\widehat{h}_{t}^{c},f_{t-1\to t}^{0},f_{t+1\to t}^{0} =$$

$$h_t^c, f_{t-1 \rightarrow t}^0, f_t^0$$

$$\mathcal{D}^{1}\left(\left[\phi_{t}^{1}, \widehat{\phi}_{t}^{1}, \widehat{\psi}_{t}^{1}, f_{t-1 \to t}^{1}, f_{t+1 \to t}^{1}, \widetilde{\phi}_{t-1}^{1}, \widetilde{\phi}_{t+1}^{1}\right]\right),$$
(9)

where  $\mathcal{D}_{R}^{k}$  (k = 2, 3) stand for the *k*-th level of the HDR rendering decoder  $\mathcal{D}_R$ . The warped neighbor features  $\{\widetilde{\phi}_{t-1}^k, \widetilde{\phi}_{t+1}^k\}$  help  $\mathcal{D}_R$ to identify regions with poor alignment and subsequently obtain the refined optical flow  $\{f_{t-1 \to t}^{k-1}, f_{t+1 \to t}^{k-1}\}$ .

Based on the completed frame  $\hat{l}_t$  and the coarse HDR frame  $\hat{h}_t^c$ , we combine them with original LDR frames to obtain the final HDR rendering result  $h_t$ . We adopted a blending network with U-net architecture from Xu et al. [40]. Thus, original and completed LDR frames  $\{l_{t-1}, l_t, l_{t+1}, \hat{l}_t\}$ , along with their corresponding frames  $\{x_{t-1}, x_t, x_{t+1}, \hat{x}_t\}$  in linear HDR domain, and coarse HDR frame  $\hat{h}_t^c$  are fed into the blending network. The blending network calculates fusion weights  $W = \{w_i | i = 0, 1, 2, 3, 4\}$  for the input five HDR domain frames  $\{x_{t-1}, x_t, x_{t+1}, \hat{x}_t, \hat{h}_t^c\}$  and obtains the final HDR rendering result  $\hat{h}_t$  through a weighted average based on the computed weights W:

$$\widehat{h}_{t} = \frac{w_{0}\widehat{h}_{t}^{c} + w_{1}\widehat{x}_{t} + w_{2}x_{t} + w_{3}x_{t-1} + w_{4}x_{t+1}}{\sum_{i=0}^{4} w_{j}}.$$
 (10)

Following Xu et al. [40], we also fuse with the neighbor frames  $\{x_{t-1}, x_{t+1}\}$  to provide information about static regions. As depicted in Fig. 1, when encountering complex scenes involving both motion and saturation or noise, the fused exposure completion frames effectively provide the missing exposure information, thereby obtaining ghost-free and low-noise HDR results. In summary, based on completing the frame with missing exposure information, we finally achieve a high-quality HDR rendering process.

# 3.4 Training Loss

We calculate the losses for the completed LDR features  $\widehat{\Phi}_t$  and frame  $\hat{l}_t$ , rendered HDR features  $\hat{\Psi}_t$  and frame  $\hat{h}_t$ , and optical flows  $\{F_{t-1 \rightarrow t}, F_{t+1 \rightarrow t}\}$ :

The losses for images. The widely adopted  $\mathcal{L}_1$  loss is used to supervise the completed LDR frame  $\hat{l}_t$  and rendered HDR frame  $\hat{h}_t$ , and is defined as follows:

$$\mathcal{L}_{I}^{com} = \left\| \widehat{l}_{t} - \overline{l}_{t} \right\|_{1}, \tag{11}$$

$$\mathcal{L}_{I}^{ren} = \left\| \mathcal{T}(\widehat{h}_{t}) - \mathcal{T}(\overline{h}_{t}) \right\|_{1}, \qquad (12)$$

$$\mathcal{L}_I = \mathcal{L}_I^{com} + \mathcal{L}_I^{ren},\tag{13}$$

where  $\bar{l}_t$  and  $\bar{h}_t$  are the corresponding ground truth for the completed LDR frame  $\hat{l}_t$  and rendered HDR frame  $\hat{h}_t$ , respectively. The  $\mathcal{T}(\cdot)$  is a widely used function to map HDR frame to the tonemapped HDR domain, since HDR images are typically displayed

after tone mapping. This simple differentiable  $\mu$ -law function  $\mathcal{T}(\cdot)$ is defined as follows:

$$\mathcal{T}(h) = \frac{\log(1+\mu h)}{\log(1+h)},\tag{14}$$

where  $\mu$  is a hyperparameter.

The losses for features. Feature space geometry loss proposed in [21] is employed to ensure that the intermediate features obtained from frame interpolation and HDR reconstruction can be more effectively refined to conform to geometrically structured features. The parameter shared encoder  $\mathcal{E}$  is used to obtain corresponding pyramid features  $\overline{\Phi}_t = \{\overline{\phi}_t^k | k = 1, 2, 3\}$  and  $\overline{\Psi}_t = \{\overline{\psi}_t^k | k = 1, 2, 3\}$  from the ground truth of completed LDR and rendered HDR frames. Then, the loss function for supervising the features of completed LDR and rendered HDR frames can be written as:

$$\mathcal{L}_{G}^{com} = \sum_{k=1}^{3} \mathcal{L}_{cen}(\widehat{\phi}_{t}^{k}, \overline{\phi}_{t}^{k}), \qquad (15)$$

$$\mathcal{L}_{G}^{ren} = \sum_{k=1}^{3} \mathcal{L}_{cen}(\widehat{\psi}_{t}^{k}, \overline{\psi}_{t}^{k}), \qquad (16)$$

$$\mathcal{L}_G = \mathcal{L}_G^{com} + \mathcal{L}_G^{ren},\tag{17}$$

where the  $\mathcal{L}_{cen}$  is census loss [32] and computed using the soft Hamming distance between census-transformed [46] feature maps, considering 3×3 patches in a channel-by-channel manner.

The loss for optical flows. The HDR-domain alignment loss [40] is used hierarchically to supervise the learning process of optical flow, which is defined as follows:

$$\mathcal{L}_{F}^{t-1 \to t} = \sum_{k=0}^{3} \|\mathcal{W}(\mathcal{T}(\overline{h}_{t-1}), \mathcal{U}_{2^{k}}(f_{t-1 \to t}^{k})) - \mathcal{T}(\overline{h}_{t})\|_{1}, \quad (18)$$

$$\mathcal{L}_{F}^{t+1\to t} = \sum_{k=0}^{3} \|\mathcal{W}(\mathcal{T}(\overline{h}_{t+1}), \mathcal{U}_{2^{k}}(f_{t+1\to t}^{k})) - \mathcal{T}(\overline{h}_{t})\|_{1}, \quad (19)$$

$$\mathcal{L}_F = (1 - m_t) \odot \left( \mathcal{L}_F^{t-1 \to t} + \mathcal{L}_F^{t+1 \to t} \right), \tag{20}$$

where  $\mathcal{W}(\cdot, \cdot)$  denotes using optical flow to warp neighbor frames to the reference frame,  $\mathcal{U}_s$  represents the bilinear upsampling operation with scale factor s. The mask  $m_t$  indicates well-exposed regions in reference frame. First,  $l_t$  is converted to YCbCr color space to extract luminance y. Then,  $m_t$  is defined as  $\delta_{\text{low}} < y < \delta_{\text{high}}$ , where  $\delta_{\rm low}$  and  $\delta_{\rm high}$  are the low and high luminance thresholds, respectively. This allows the optical flow computation to focus more on regions with poor exposure in the reference frame.

Total loss. Our total loss can be summarized as follows:

$$\mathcal{L}_{total} = \mathcal{L}_I + \alpha * \mathcal{L}_G + \beta * \mathcal{L}_F.$$
<sup>(21)</sup>

#### **Extension to Three Exposures** 3.5

In the alternate exposure setting with three exposures, five LDR frames  $\{l_{t-2}, l_{t-1}, l_t, l_{t+1}, l_{t+2}\}$  are used as input to the network for HDR  $\hat{h}_t$  rendering. The five LDR frames are with corresponding exposure sequences  $\{e_{t-2}, e_{t-1}, e_t, e_{t+1}, e_{t+2}\}$ . In this way, the reference frame has only one certain exposure  $e_t = \epsilon_3$ . Specifically, the first frame  $l_{t-2}$  and the fourth frame  $l_{t+1}$  have the same exposure  $e_{t-2} = e_{t+1} = \epsilon_1$ , while the second frame  $l_{t-1}$  and the fifth frame  $l_{t+2}$  also share the same exposure  $e_{t-1} = e_{t+2} = \epsilon_2$ . Therefore, the

Table 1: Quantitative comparisons of our method with other state-of-the-art methods on the Cinematic Video dataset [8]. The best and the second best results are highlighted in red and blue, respectively.

Methods		2-Exposure			3-Exposure			
		PSNR T	SSIM T	HDR-VDP-2	PSNR T	SSIM $_T$	HDR-VDP-2	
Kalantari13 [17]	2013-TOG	37.51	0.9016	60.16	30.36	0.8133	57.68	
Kalantari19 [15]	2019-CGF	37.06	0.9053	70.82	33.21	0.8402	62.44	
Yan19 [43]	2019-CVPR	31.65	0.8757	69.05	34.22	0.8604	66.18	
Prabhakar [36]	2021-CVPR	34.72	0.8761	68.82	34.02	0.8633	65.00	
Chen [3]	2021-ICCV	35.65	0.8949	72.09	34.15	0.8847	66.81	
LAN-HDR [6]	2023-ICCV	38.22	0.9100	69.15	35.07	0.8695	65.42	
HDRFlow [40]	2024-CVPR	39.20	0.9154	70.98	36.55	0.9039	65.89	
HDRFlow [40](+Sintel)	2024-CVPR	39.30	0.9156	71.05	36.65	0.9055	66.02	
Ours		40.59	0.9241	73.31	37.24	0.9102	68.36	

Table 2: Quantitative comparisons of our method with other state-of-the-art methods on the DeepHDRVideo dataset [3]. The result is the weighted average of all results from both dynamic scenes and static scenes in this dataset. The best and the second best results are highlighted in red and blue, respectively.

		2-Exposure			3-Exposure		
Methods		PSNR T	SSIM $_T$	HDR-VDP-2	PSNR $T$	3-Expos SSIM T 0.9489 0.9331 0.9531 0.9453 0.9569 0.9504 0.9530 0.9535 0.9550	HDR-VDP-2
Kalantari13 [17]	2013-TOG	40.33	0.9409	66.11	38.45	0.9489	57.31
Kalantari19 [15]	2019-CGF	39.91	0.9329	71.11	38.78	0.9331	65.73
Yan19 [43]	2019-CVPR	40.54	0.9452	69.67	40.20	0.9531	68.23
Prabhakar [36]	2021-CVPR	40.21	0.9414	70.27	39.48	0.9453	65.93
Chen [3]	2021-ICCV	42.48	0.9620	74.80	39.44	0.9569	67.76
LAN-HDR [6]	2023-ICCV	41.59	0.9472	71.34	40.48	0.9504	68.61
HDRFlow [40]	2024-CVPR	43.18	0.9510	77.11	40.45	0.9530	72.30
HDRFlow [40](+Sintel)	2024-CVPR	43.25	0.9520	77.29	40.56	0.9535	72.42
Ours		43.44	0.9558	79.20	40.13	0.9550	76.98

exposure completion process takes  $\{l_{t-2}, l_t, l_{t+1}\}$  and  $\{l_{t-1}, l_t, l_{t+2}\}$  as inputs, generates exposure completing results  $\{\hat{l}_t^{e_1}, \hat{l}_t^{e_2}\}$  through the flow-guided completing process, and obtains coarse HDR results  $\hat{h}_t^c$  through the coarse-to-fine HDR rendering process. A total of fifteen images, including the exposure completing frames and multiple original LDR frames, along with their corresponding frames in linear HDR domain, and the coarse HDR result, are fed into the blending network. This blending network calculates seven weights to fuse the seven HDR domain images in a weighted average manner and obtain the final HDR rendering result. More details are provided in supplementary materials.

# 4 EXPERIMENTS

#### 4.1 Experimental Setup

**Datasets.** We utilize synthetic training data generated from the Vimeo-90K dataset [41]. To adapt the Vimeo90K dataset for HDR video reconstruction, we follow prior research [18] to convert the original data into LDR sequences with alternate exposures. To create the ground truth of completed LDR frames, we also generated LDR sequences with missing exposures in the same way. Our framework is tested on the Cinematic Video dataset [8] and DeepHDRVideo dataset [3]. The Cinematic Video dataset has two synthetic videos from indoor and outdoor scenes. The DeepHDRVideo dataset [3] contains both real-world dynamic scenes and static scenes with random global motion augmentation. The HDRVideo dataset [17] is employed solely for qualitative evaluation, as it lacks ground truth. **Implementation details.** We implement our approach using Py-Torch and conduct experiments on an NVIDIA RTX3090 GPU. We employ AdamW optimizer [20] with  $\beta_1 = 0.9$  and  $\beta_2 = 0.999$ . The learning rate is set to  $10^{-4}$ . In our experiments, we set  $\gamma$  in Eq. 1 to 2.2 and  $\mu$  in Eq. 14 to 5000. Following HDRFlow [40], we set  $\delta_{\text{low}}$  to 0.2 and  $\delta_{\text{high}}$  to 0.8. The weighting hyperparameters  $\alpha$  and  $\beta$  for the loss function in Eq. 21 are set to 0.01.

**Evaluation metrics.**  $PSNR_T$ ,  $SSIM_T$  and HDR-VDP-2 [30] are adopted as the evaluation metrics.  $PSNR_T$  and  $SSIM_T$  are computed on the tone-mapped images. HDR-VDP-2 is computed with the number of pixels per visual degree set to 30, which means the angular resolution of the image.

#### 4.2 Comparisons with State-of-the-art

**Quantitative comparisons** between our and other state-of-theart methods on the Cinematic Video [8] and DeepHDRVideo [3] datasets are shown in Table 1 and Table 2, respectively. Compared to Exposure Completing for Temporally Consistent Neural High Dynamic Range Video Rendering

#### ACM MM, 2024, Melbourne, Australia



Figure 3: Qualitative comparison in scenes with over-saturation and motion. Left: 2-Exposure scene from the Cinematic Video dataset [8]. Right: 2-Exposure scene from the HDRVideo dataset [17]."EC" refers to our exposure completion results.



Figure 4: Qualitative comparison in scenes with noise and motion. Left: 2-Exposure scene from the Cinematic Video dataset [8]. Right: 2-Exposure scene from the DeepHDRVideo dataset [3]. "EC" refers to our exposure completion results.

state-of-the-art methods, our approach consistently achieves superior or comparable performance. Especially, our approach achieves state-of-the-art performance on Cinematic Video [8] dataset, outperforming the second-best method by 1.29dB and 0.59dB in terms of  $PSNR_T$  for the 2-exposure and 3-exposure settings, respectively. Qualitative comparisons are shown in Fig. 3 and Fig. 4. we compare our NECHDR with the previous methods: Chen [3], LAN-HDR [6] and HDRFlow [40]. Fig. 3 illustrates the results from scenes with saturation and motion on Cinematic Video dataset [8] and HDRVideo dataset [17] under the 2-exposure setting. Apart from our method, other methods tend to exhibit severe artifacts or detail loss when saturation and motion occur simultaneously. In such challenging scenarios, we achieve accurate and artifact-free exposure completion results for saturated regions by leveraging frame interpolation from neighboring frames. This enables us to fuse high-quality HDR results. And in Fig. 4, we also show the

results encountering noise and motion from the Cinematic Video dataset [8] and DeepHDRVideo dataset [3]. The scene on the left side of Fig. 4 is captured in low-light conditions, resulting in very low signal-to-noise ratio in the low-exposure frames. This can lead to very high noise levels in rendered HDR frames after tone mapping, which makes this scene particularly challenging. Specifically, methods based on optical flow [3, 40] tend to produce noisy artifacts, while attention-based method [6] exhibit more pronounced noise. Our method explicitly reconstructs the absent high-exposure frame with low noise at the current time stamp, achieving the best noise suppression, even with some regions having noise intensity lower than ground truth. The visualization results above explain why we achieve a significant performance improvement compared with other state-of-the-art methods. More qualitative comparisons in the 3-exposure setting are provided in the supplementary materials.

Table 3: Ablation study of NECHDR on the DeepH-DRVideo [3] and CinematicVideo dataset [8]. "EC" refers to exposure completing. "HR" refers to HDR rendering. "Uncoupled" means this two processes work independently.

Model	DeepHDRVideo PSNR <sub>T</sub> SSIM <sub>7</sub>				
EC Baseline	41.09	0.9479	39.13	0.9174	
HR Baseline	42.64	0.9524	39.98	0.9193	
Uncoupled EC, HR	43.17	0.9535	40.27	0.9202	
Coupled EC, HR (Ours)	43.44	0.9558	40.59	0.9241	



LDRs Uncoupled EC, HR Coupled EC, HR (Ours)

Figure 5: Qualitative comparison of the models corresponding to the ablation study on a dynamic scene of DeepH-DRVideo [3] dataset.

# 4.3 Analysis

Ablation Study. We conduct ablation experiments under the 2exposure setting on Cinematic Video dataset [8] and DeepHDRVideo dataset [3], and the quantitative and qualitative results are shown in Table 3 and Fig. 5, respectively. We devised a baseline that employs IFRNet [21] to utilize the neighbor frames to complete the middle time frame that with missing exposure information and achieve HDR results by simply fusing the completed LDR frame with the original LDR frames. The performance of this baseline is shown as "EC Baseline" in the first row in Table 3. Another baseline directly renders HDR results based on IFRNet, which is the "HR Baseline" in second row in Table 3. However, relying solely on either exposure completing or HDR rendering results is with limited performance. By adding the exposure completing decoder, we get a network (third row in Table 3) that contains both decoupled exposure completing and HDR rendering processes. Finally, through feeding the completed features and frames from exposure completing decoder into the process of HDR rendering, we achieve our NECHDR framework in the forth row in Table 3. Based on the qualitative and quantitative results, we can see: (a) coupled exposure completing and HDR rendering processes benefit the quality



Figure 6: Visual comparisons of temporal consistency.

of final HDR results; (b) exposure completing decoder is necessary; (d) the completed results with missing exposure information help the rendering process of HDR reconstruction.

**Temporal consistency.** We show the visual comparisons of temporal consistency in Fig. 6. In Fig. 6, we record a two-pixel-height line traversing all frames of a scene in the Cinematic Video dataset [8] over time and lay them out sequentially to form temporal profiles. Base on the illustration of temporal profiles, we can observe that the horizontal stripes exist in the temporal profiles of other methods. The horizontal stripes comes from the differences between adjacent frames, which represents the temporal inconsistency. In contrast, the horizontal stripes can hardly be observed in our temporal profiles, which means that our proposed method achieves better temporal consistency. Additional visual comparisons regarding temporal consistency can be found in the supplementary materials.

# 5 CONCLUSION

In this paper, we implement the idea of exposure completing for neural HDR video rendering and propose the Neural Exposure Completing HDR (NECHDR) framework. The NECHDR leverages interpolation of neighbor LDR frames to complete missing exposures, providing a complete set of exposure information for each time stamp. This process of exposure completing creates a novel neural HDR video rendering pipeline, which can generate results of less noise and ghosting artifacts, thereby enhancing temporal consistency. Experimental results on multiple public benchmarks demonstrate the superiority of our NECHDR, which may shift the focus of researchers in this area to the exposure completing. Exposure Completing for Temporally Consistent Neural High Dynamic Range Video Rendering

#### 929 **REFERENCES**

930

931

932

933

934

935

936

937

938

939

940

941

942

943

944

945

946

947

948

949

950

951

952

953

954

955

956

957

958

959

960

961

962

963

964

965

966

967

968

969

970

971

972

973

974

975

976

977

978

979

980

981

982

983

984

985

986

- Luca Bogoni. 2000. Extending dynamic range of monochrome and color images through fusion. In *Int. Conf. Pattern Recog.*, Vol. 3. IEEE, 7–12.
- [2] Sibi Catley-Chandar, Thomas Tanay, Lucas Vandroux, Aleš Leonardis, Gregory Slabaugh, and Eduardo Pérez-Pellitero. 2022. Flexhdr: Modeling alignment and exposure uncertainties for flexible hdr imaging. *IEEE Trans. Image Process.* 31 (2022), 5923–5935.
- [3] 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.
- [4] Inchang Choi, Seung-Hwan Baek, and Min H Kim. 2017. Reconstructing interlaced high-dynamic-range video using joint learning. *IEEE Trans. Image Process.* 26, 11 (2017), 5353–5366.
- [5] Haesoo Chung and Nam Ik Cho. 2022. High dynamic range imaging of dynamic scenes with saturation compensation but without explicit motion compensation. In Proc. IEEE Winter Conf. Appl. Comput. Vis. 2951–2961.
- [6] 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.
- [7] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. 2017. Deformable convolutional networks. In Proc. Int. Conf. Comput. Vis. 764-773.
- [8] 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.
- [9] Thorsten Grosch et al. 2006. Fast and robust high dynamic range image generation with camera and object movement. Vision, Modeling and Visualization, RWTH Aachen 277284, 3 (2006), 2.
- [10] Jin Han, Chu Zhou, Peiqi Duan, Yehui Tang, Chang Xu, Chao Xu, Tiejun Huang, and Boxin Shi. 2020. Neuromorphic camera guided high dynamic range imaging. In Proc. IEEE Conf. Comput. Vis. Pattern Recog. 1730-1739.
- [11] Felix Heide, Markus Steinberger, Yun-Ta Tsai, Mushfiqur Rouf, Dawid Pajak, Dikpal Reddy, Orazio Gallo, Jing Liu, Wolfgang Heidrich, Karen Egiazarian, et al. 2014. Flexisp: A flexible camera image processing framework. ACM Trans. Graph. 33, 6 (2014), 1–13.
- [12] Yong Seok Heo, Kyoung Mu Lee, Sang Uk Lee, Youngsu Moon, and Joonhyuk Cha. 2010. Ghost-free high dynamic range imaging. In ACCV. Springer, 486–500.
- [13] Jun Hu, Orazio Gallo, Kari Pulli, and Xiaobai Sun. 2013. HDR deghosting: How to deal with saturation?. In Proc. IEEE Conf. Comput. Vis. Pattern Recog. 1163–1170.
- [14] Katrien Jacobs, Celine Loscos, and Greg Ward. 2008. Automatic high-dynamic range image generation for dynamic scenes. *IEEE Computer Graphics and Applications* 28, 2 (2008), 84–93.
- [15] Nima Khademi Kalantari and Ravi Ramamoorthi. 2019. Deep HDR video from sequences with alternating exposures. In *Comput. Graph. Forum*, Vol. 38. Wiley Online Library, 193–205.
- [16] Nima Khademi Kalantari, Ravi Ramamoorthi, et al. 2017. Deep high dynamic range imaging of dynamic scenes. ACM Trans. Graph. 36, 4 (2017), 144–1.
- [17] 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.
- [18] Sing Bing Kang, Matthew Uyttendaele, Simon Winder, and Richard Szeliski. 2003.
   High dynamic range video. ACM Trans. Graph. 22, 3 (2003), 319–325.
- [19] Erum Arif Khan, Ahmet Oguz Akyuz, and Erik Reinhard. 2006. Ghost removal in high dynamic range images. In IEEE Int. Conf. Image Process. IEEE, 2005–2008.
- [20] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. Proc. Int. Conf. Learn. Represent. abs/1412.6980 (2015). https: //api.semanticscholar.org/CorpusID:6628106
- [21] Lingtong Kong, Boyuan Jiang, Donghao Luo, Wenqing Chu, Xiaoming Huang, Ying Tai, Chengjie Wang, and Jie Yang. 2022. Ifrnet: Intermediate feature refine network for efficient frame interpolation. In Proc. IEEE Conf. Comput. Vis. Pattern Recog, 1969–1978.
- [22] Joel Kronander, Stefan Gustavson, Gerhard Bonnet, and Jonas Unger. 2013. Unified HDR reconstruction from raw CFA data. In IEEE Int. Conf. Comput. Photo. IEEE, 1–9.
- [23] Joel Kronander, Stefan Gustavson, Gerhard Bonnet, Anders Ynnerman, and Jonas Unger. 2014. A unified framework for multi-sensor HDR video reconstruction. Signal Process. Image Commun. 29, 2 (2014), 203–215.
- [24] Chul Lee, Yuelong Li, and Vishal Monga. 2014. Ghost-free high dynamic range imaging via rank minimization. IEEE Sign. Process. Letters 21, 9 (2014), 1045–1049.
- [25] Shuaizheng Liu, Xindong Zhang, Lingchen Sun, Zhetong Liang, Hui Zeng, and Lei Zhang. 2023. Joint hdr denoising and fusion: A real-world mobile hdr image dataset. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.* 13966–13975.
- [26] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In Proc. Int. Conf. Comput. Vis. 10012–10022.

- [27] Zhen Liu, Yinglong Wang, Bing Zeng, and Shuaicheng Liu. 2022. Ghost-free high dynamic range imaging with context-aware transformer. In Proc. Eur. Conf. Comput. Vis. Springer, 344–360.
- [28] Kede Ma, Hui Li, Hongwei Yong, Zhou Wang, Deyu Meng, and Lei Zhang. 2017. Robust multi-exposure image fusion: a structural patch decomposition approach. *IEEE Trans. Image Process.* 26, 5 (2017), 2519–2532.
- [29] Stephen Mangiat and Jerry Gibson. 2011. Spatially adaptive filtering for registration artifact removal in HDR video. In *IEEE Int. Conf. Image Process*. IEEE, 1317–1320.
- [30] Rafał Mantiuk, Kil Joong Kim, Allan G Rempel, and Wolfgang Heidrich. 2011. HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. ACM Trans. Graph. 30, 4 (2011), 1–14.
- [31] Morgan McGuire, Wojciech Matusik, Hanspeter Pfister, Billy Chen, John F Hughes, and Shree K Nayar. 2007. Optical splitting trees for high-precision monocular imaging. *IEEE Comput. Graph. Appl.* 27, 2 (2007), 32–42.
- [32] Simon Meister, Junhwa Hur, and Stefan Roth. 2018. Unflow: Unsupervised learning of optical flow with a bidirectional census loss. In Proc. of AAAI Conf. Artif. Intell., Vol. 32.
- [33] Shree K Nayar and Tomoo Mitsunaga. 2000. High dynamic range imaging: Spatially varying pixel exposures. In Proc. IEEE Conf. Comput. Vis. Pattern Recog., Vol. 1. 472–479.
- [34] Yuzhen Niu, Jianbin Wu, Wenxi Liu, Wenzhong Guo, and Rynson WH Lau. 2021. HDR-GAN: HDR image reconstruction from multi-exposed LDR images with large motions. *IEEE Trans. Image Process.* 30 (2021), 3885–3896.
- [35] Tae-Hyun Oh, Joon-Young Lee, Yu-Wing Tai, and In So Kweon. 2014. Robust high dynamic range imaging by rank minimization. 37, 6 (2014), 1219–1232.
- [36] K Ram Prabhakar, Gowtham Senthil, Susmit Agrawal, R Venkatesh Babu, and Rama Krishna Sai S Gorthi. 2021. Labeled from unlabeled: Exploiting unlabeled data for few-shot deep hdr deghosting. In Proc. IEEE Conf. Comput. Vis. Pattern Recog. 4875–4885.
- [37] Pradeep Sen, Nima Khademi Kalantari, Maziar Yaesoubi, Soheil Darabi, Dan B Goldman, and Eli Shechtman. 2012. Robust patch-based hdr reconstruction of dynamic scenes. ACM Trans. Graph. 31, 6 (2012), 203–1.
- [38] Michael D Tocci, Chris Kiser, Nora Tocci, and Pradeep Sen. 2011. A versatile HDR video production system. ACM Trans. Graph. 30, 4 (2011), 1–10.
- [39] Shangzhe Wu, Jiarui Xu, Yu-Wing Tai, and Chi-Keung Tang. 2018. Deep high dynamic range imaging with large foreground motions. In Proc. Eur. Conf. Comput. Vis. 117–132.
- [40] 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.
- [41] Tianfan Xue, Baian Chen, Jiajun Wu, Donglai Wei, and William T Freeman. 2019. Video enhancement with task-oriented flow. Int. J. Comput. Vis. 127 (2019), 1106–1125.
- [42] Qingsen Yan, Weiye Chen, Song Zhang, Yu Zhu, Jinqiu Sun, and Yanning Zhang. 2023. A unified HDR imaging method with pixel and patch level. In Proc. IEEE Conf. Comput. Vis. Pattern Recog. 22211–22220.
- [43] Qingsen Yan, Dong Gong, Qinfeng Shi, Anton van den Hengel, Chunhua Shen, Ian Reid, and Yanning Zhang. 2019. Attention-guided network for ghost-free high dynamic range imaging. In Proc. IEEE Conf. Comput. Vis. Pattern Recog. 1751–1760.
- [44] Qingsen Yan, Dong Gong, Pingping Zhang, Qinfeng Shi, Jinqiu Sun, Ian Reid, and Yanning Zhang. 2019. Multi-scale dense networks for deep high dynamic range imaging. In Proc. IEEE Winter Conf. Appl. Comput. Vis. IEEE, 41–50.
- [45] Qingsen Yan, Lei Zhang, Yu Liu, Yu Zhu, Jinqiu Sun, Qinfeng Shi, and Yanning Zhang. 2020. Deep HDR imaging via a non-local network. *IEEE Trans. Image Process.* 29 (2020), 4308–4322.
- [46] Ramin Zabih and John Woodfill. 1994. Non-parametric local transforms for computing visual correspondence. In Proc. Eur. Conf. Comput. Vis. Springer, 151– 158.
- [47] Wei Zhang and Wai-Kuen Cham. 2011. Gradient-directed multiexposure composition. IEEE Trans. Image Process. 21, 4 (2011), 2318–2323.
- [48] Hang Zhao, Boxin Shi, Christy Fernandez-Cull, Sai-Kit Yeung, and Ramesh Raskar. 2015. Unbounded high dynamic range photography using a modulo camera. In 2015 IEEE International Conference on Computational Photography (ICCP). IEEE, 1–10.
- [49] Henning Zimmer, Andrés Bruhn, and Joachim Weickert. 2011. Freehand HDR imaging of moving scenes with simultaneous resolution enhancement. In *Comput. Graph. Forum*, Vol. 30. Wiley Online Library, 405–414.

987

988

- 1042 1043
- 1044