A Unified Framework for High-Frame-Rate HDR Video Synthesis

Hue Nguyen*
huent@yorku.ca
Trevor D. Canham*
tcanham@yorku.ca
Michael S. Brown
mbrown@yorku.ca

York University,
Toronto, Ontario

Abstract

Capturing high dynamic range (HDR) video at high frame rates is critical for applications such as cinematic production and autonomous systems, yet it remains challenging with conventional cameras. We present the first unified framework that jointly performs HDR reconstruction and temporal interpolation from sequences with alternating exposures. Unlike prior work that reconstructs only middle frames or uses heavy off-the-shelf interpolators, our lightweight network synthesizes HDR video at arbitrary timesteps in real time on mid-range GPUs. To support this task, we introduce a new dataset of exposure-bracketed video sequences with real-world motion. To reduce reliance on ground-truth HDR data, we also propose a novel self-supervised training scheme that delivers competitive results. Experiments show our approach outperforms existing baselines in efficiency while achieving comparable or better visual quality, establishing a new benchmark for practical HDR video synthesis. Our code is available at https://github.com/huent189/unified-HFR-HDR.

1 Introduction

High dynamic range (HDR) video is crucial for applications such as cinematic production, autonomous navigation, and surveillance, as it enables cameras to capture scenes with extreme variations in light intensity—conditions that conventional systems struggle to handle. However, producing high-frame-rate HDR video, especially in dynamic environments, remains challenging due to the difficulty of recovering fine details in overexposed or underexposed regions while ensuring smooth, artifact-free motion between frames.

Modern cameras capture high-quality signals across a dynamic range of approximately 8 to 14 f-stops (in $\log_2$ units), but real-world scenes often exhibit intensity variations exceeding 17 f-stops [23]. Light falling outside this capture range is either too intense—causing pixel saturation—or too dim to register a clean signal above the sensor's noise floor. In both cases, this leads to a loss of texture and color fidelity in the recorded image.

The effective capture range can also be constrained by components of the imaging pipeline, both before and after sensing. For instance, optical flare can further obscure dark details,

* denotes equal contribution

while post-sensing processes such as quantization, color encoding, and compression can reduce the number of usable intensity levels, introducing visible banding artifacts when displayed at high dynamic ranges.

Devices that capture the widest dynamic range typically rely on high-end hardware and innovative designs. For example, cinema cameras and specialized sensors—such as dual-exposure systems [6] or event-based cameras [3]—can produce high-quality HDR video, but these solutions require custom components. Their high cost, bulky form factors, and high power demands limit their suitability for integration into consumer devices.

A more accessible alternative is exposure bracketing, which involves alternating between under- and over-exposed frames on standard cameras so that their combined dynamic range can be integrated. Previous works [4, 5, 11, 17, 26, 28] have proposed HDR reconstruction methods for video based on this technique, but they are limited in their ability to represent motion at high frame rates.

Another work [18] frames the task as video interpolation but uses an off-the-shelf network that ignores correlations between under- and over-exposed frames, leading to high computational cost. More broadly, existing exposure-bracketed HDR datasets are either limited due to high production costs [9, 18], or fail to capture real scene motion, relying instead on computer graphics or stop-motion sequences [4, 5, 26, 28].

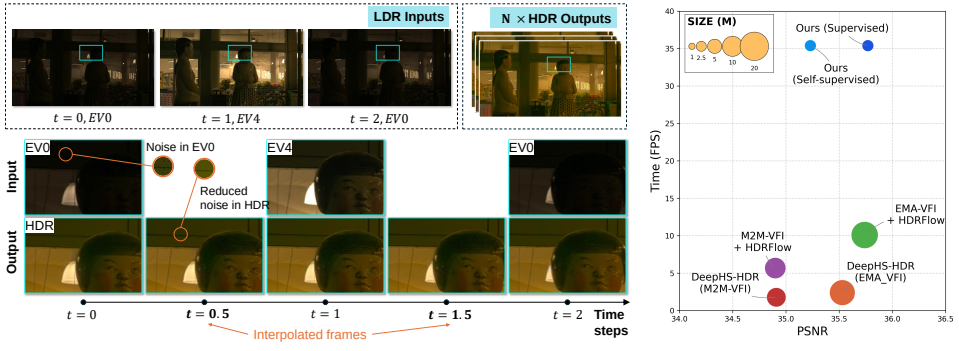
Contributions. Addressing the limitations above, we make the following contributions:

- We introduce a large-scale dataset for training and evaluating the joint HDR video reconstruction and interpolation task. Using a handheld cinema camera with a wide native dynamic range, we captured high-frame-rate (HFR) videos of real motion in diverse scenes, from which exposure-bracketed frames were derived in post-processing.
- We propose a lightweight neural architecture that jointly performs HDR reconstruction and interpolation. This joint formulation reduces computational and memory overhead, and can be integrated into existing HDR reconstruction pipelines to enable interpolation capabilities.
- We present a novel self-supervised learning framework that enables fine-tuning of any pre-trained HDR video reconstruction network for the new interpolation task.
- We conduct comprehensive experiments showing that our method achieves comparable performance in both supervised and self-supervised settings, surpassing state-of-the-art methods while requiring fewer computational resources and less memory.

In the following sections, we review recent work regarding HDR image and video reconstruction, then introduce our proposed model architecture, training framework, and dataset. Finally, we evaluate the effectiveness of these contributions through ablation studies and comparisons with existing baselines.

2 Related Work

Early examples of HDR imaging via multiple exposures date back to the advent of photography [20], but Debevec and Malik were the first to fuse digital exposures for HDR reconstruction [7]. While effective for static scenes, this approach suffers from haloing and ghosting due to temporal offsets between captures—a key challenge for video, which requires high-frequency sampling in dynamic scenes. Subsequent works addressed these artifacts using



(a) (b)

Figure 1: We propose a method for joint high dynamic range video reconstruction and interpolation at **arbitrary timesteps** from a sequence of alternating exposures between a base exposure (EV0) and a higher exposure (e.g., EV4). (a) Input and output examples. Input frames often exhibit noise in dark regions (EV0) and loss of detail in under- and over-exposed areas (EV0/EV4). Our method generates **clean shadows and detailed highlights** with temporal consistency, while increasing the frame rate through interpolation. (b) Comparison of computational cost and reconstruction quality. Our method is more efficient while maintaining comparable or better quality than state-of-the-art baselines.

alignment strategies across exposures [11, 17]. While effective in some cases, these heuristics struggle with large motion or significant exposure differences, and frame rate limitations and motion artifacts persist in multi-exposure HDR video.

Learning-Based HDR Image and Video Reconstruction. Later works apply deep learning to handle these edge cases, but often rely on supervised training with synthetic data, limiting adaptability to real-world scenes and camera sensors [4, 5, 26]. Other methods use stop-motion exposure-bracketed stills [28], which fail to capture natural scene motion. The most closely related work is SelfHDR [31], a self-supervised HDR reconstruction method using multi-exposure inputs. SelfHDR assumes the target image has mid-level exposure, which is reasonable for static images but problematic in video. In HDR video, the target frame may be a high-exposure input with highlight clipping, causing structural detail loss. As a result, applying SelfHDR to video is unreliable.

Video Interpolation. Video interpolation has a long history, notably in compression standards like MPEG [19] and HEVC [25], where intermediate frames are predicted using block-based motion vectors to avoid redundant encoding of static regions. However, these methods are prone to artifacts and struggle with complex motion. Recent advances include interpolation based on the Fourier Shift Theorem [8], and deep learning methods that learn transformations in the latent space of convolutional networks [22] or generative models [27]. Despite their success, these approaches target standard dynamic range (SDR) video and do not exploit exposure redundancy in HDR settings.

HFR-HDR Video Reconstruction. Existing HFR-HDR methods often rely on specialized hardware. Chang *et al*. [3] combined a high-speed spike camera with a low-frame-rate RGB sensor using alternating exposures to achieve 1000 FPS HDR. Similarly, Çoğalan *et al*. [6] used dual-exposure sensors to capture lower and higher exposures simultaneously. While effective, such systems are costly and impractical for consumer use.

To avoid hardware dependence, recent work has focused on software-based HFR-HDR

using alternating exposures. Khan *et al.* [18] proposed DeepHS-HDRVideo, which uses a pre-trained interpolator to align same-exposure frames before merging them into HDR, enabling recursive temporal upsampling on standard cameras.

However, their multi-stage pipeline—using a separate interpolator before HDR merging—introduces significant computational overhead. In contrast, our method retains the hardware-free advantage of DeepHS-HDRVideo but unifies HDR reconstruction and interpolation in a single-stage framework, better leveraging exposure redundancy and enabling real-time HFR-HDR synthesis on mid-range GPUs.

3 Methodology

3.1 Pipeline Overview

Given a low dynamic range (LDR) video sequence $\{\tilde{L}_i \mid i = 1,\ldots,n\}$ captured at a low frame rate (e.g., 30 FPS) using alternating exposure values, our goal is to reconstruct a high-frame-rate (e.g., 60 FPS) high dynamic range (HDR) video, as illustrated in Figure 1. Specifically, our model takes three consecutive LDR frames F_0, F_1, F_2 and a target intermediate time $t \in (0,2)$ as input, and outputs the corresponding HDR frame in linear domain at time t. Following prior work [4, 16, 21], we assume the camera response function (CRF) [10] of the LDR frames is known. Then, to simplify processing and ensure consistency across different cameras and settings, we replace the original CRF with a fixed gamma curve using $L_i = \left(\mathcal{F}^{-1}(\tilde{L}_i)\right)^{1/\gamma}$ with $\gamma = 2.2$ [4].

An overview of our pipeline is shown in Figure 2. We extend existing HDR video reconstruction frameworks by introducing a novel interpolation module for synthesizing HDR frames at any intermediate time. The entire pipeline can be trained end-to-end with ground-truth data or via our self-supervised framework, reducing the need for paired supervision.

3.2 Network architecture

Conventional HDR video reconstruction frameworks [4, 15, 28] typically consist of a flow estimation module followed by a fusion network. We extend this architecture with a novel interpolation module, named *InterpNet*, which synthesizes temporally aligned HDR frames by interpolating optical flow as an implicit quadratic function of time.

Given three consecutive alternating-exposure frames F_0, F_1, F_2, we first compute bidirectional optical flows $F_{1\to0}$ and $F_{1\to2}$ using a frozen pre-trained FlowNet. These flows capture motion between the central frame and its adjacent frames. To estimate the optical flow at an intermediate time $t \in (0,2)$, we feed these flows along with the scalar t into InterpNet.

The core design of InterpNet is motivated by the observation that pixel trajectories across frames often follow a quadratic path in time [12, 29], due to smooth object or camera motion. Specifically, the flow from a given input frame to an intermediate frame at time t can be modeled as a function of t, t^2, and the known flows between neighboring frames. To exploit this prior, we transform the original flows into a higher-dimensional, temporally-aware latent space using the following quadratic basis:

$$\phi(t; f_{1\to0}, f_{1\to2}) = \left[f_{1\to0},\ f_{1\to0}\cdot t,\ f_{1\to0}\cdot t^2,\ f_{1\to2},\ f_{1\to2}\cdot t,\ f_{1\to2}\cdot t^2\right]. \tag{1}$$

This representation encodes temporal dependencies and structures the input space for interpolation. To improve expressiveness, we condition it on flow features from the original

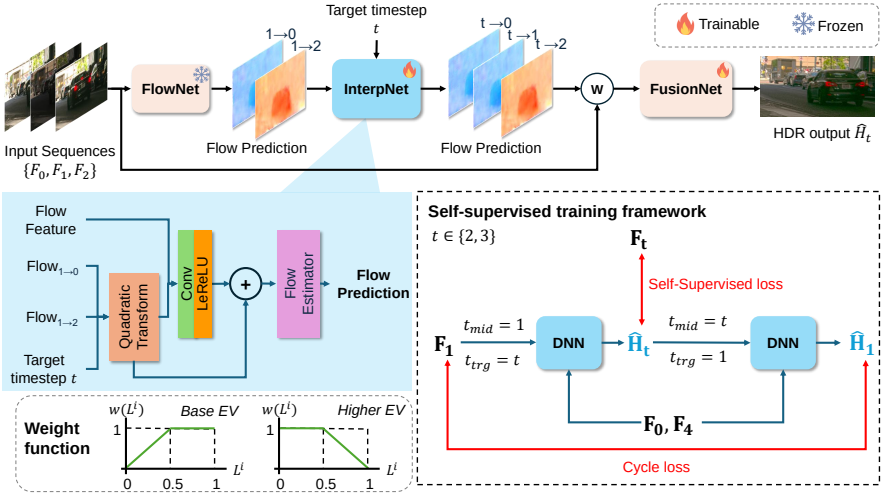


Figure 2: We introduce a novel network architecture and a self-supervised training framework designed to recover both high dynamic range and temporal information from a sequence of alternating exposures. Unlike conventional HDR video reconstruction models, which are typically limited to recovering only the middle frame, our method is capable of *reconstructing an arbitrary frame* at any given timestep t.

FlowNet, which help capture non-linear motion, occlusions, and scene complexity beyond what temporal scaling alone can resolve. The resulting feature volume is passed through a lightweight convolutional sub-network to estimate target flows $F_{t\to0}, F_{t\to1}, F_{t\to2}$, which are then used to warp input frames F_0, F_1, F_2 toward time t. Additional architectural details are provided in the supplementary material

The warped frames, along with their original LDR counterparts, are passed to a fusion network (FusionNet), which merges them to reconstruct the final linearly encoded HDR frame $\hat{H}_t$. Following the design of HDRFlow [28], FusionNet is a U-Net-like architecture that predicts fusion weights for combining the linear input frames. In our implementation, we retain the original FlowNet and FusionNet modules from HDRFlow, incorporating only our InterpNet as a plug-and-play component. This modular design enables easy adoption with minimal computational overhead.

3.3 Training Framework

We introduce two training paradigms for our network: a **supervised** approach that leverages high-frame-rate HDR ground truth when available, and a **self-supervised** alternative that relies solely on alternating-exposure LDR videos.

3.3.1 Supervised Training

Given HDR ground truth H_t^{gt} at time t, we apply a reconstruction loss on both the predicted HDR frame $\hat{H}_t$ and the temporally aligned ground truth frames:

$$\mathcal{L}_{sup} = \mathcal{L}_{rec}(\mathcal{T}(\hat{H}_t), \mathcal{T}(H_t^{gt})) + \sum_{i=0}^{2} \mathcal{L}_{rec}(\mathcal{T}(\mathcal{W}(H_i^{gt}, F_{t\to i})), \mathcal{T}(H_t^{gt})). \tag{2}$$

Here, $\mathcal{T}(\cdot)$ is the differentiable μ-law tone mapping (Eq. 3), and $\mathcal{W}(I,F)$ warps image I using flow F. Following prior work [18, 26, 28], the reconstruction loss $\mathcal{L}_{rec}$ is computed in the tone-mapped domain to better reflect perceptual differences. This can be expressed as:

$$\mathcal{T}(H) = \frac{\log(1+\mu H)}{\log(1+\mu)}, \tag{3}$$

with $\mu = 5000$, the reconstruction loss is defined as a linear combination of the L1 loss and a multi-scale Laplacian pyramid loss (with 5 levels) [24]. The use of the Laplacian pyramid encourages accurate reconstruction of high-frequency details and textures.

3.3.2 Self-Supervised Training

To eliminate the need for HDR ground truth, we propose a self-supervised framework using only alternating-exposure input. Our approach is inspired by classical HDR fusion, which combines multi-exposure LDR images L^i (with exposures e_i) into an HDR image H^{fusion} using linear irradiance $l^i = \text{CRF}^{-1}(L^i)/e_i$ and pixel-wise exposure-aware weights $w(L^i)$:

$$H^{\text{fusion}} = \frac{\sum_i w(L^i)\cdot l^i}{\sum_i w(L^i)}. \tag{4}$$

This equation represents how an HDR image would ideally be formed from multiple exposures. If the full exposure set were available at timestep t, we could train a network to minimize the reconstruction error against H_t^{fusion} at t over a training set D as follows:

$$\mathbb{E}_D\left[\left\|\hat{H}_t - H_t^{\text{fusion}}\right\|_1\right]. \tag{5}$$

By applying triangle inequalities and imposing a constraint on the weight function such that $\sum_i w(L^i) \geq 1$, we can derive an upper bound for this ideal loss:

$$\left\|\hat{H}_t - H_t^{\text{fusion}}\right\|_1 \leq \sum_i w(L_t^i)\cdot\left\|\hat{H}_t - l_t^i\right\|_1. \tag{6}$$

Since we only have a single LDR frame L_t (with corresponding linear irradiance l_t) per timestep in our self-supervised setting, we propose to train our network by minimizing a *partial* term of the upper bound based on this available frame:

$$\mathbb{E}_D\left[w(L_t)\cdot\left\|\hat{H}_t - l_t\right\|_1\right]. \tag{7}$$

Despite relying on a single LDR frame per timestep, the loss remains effective due to the alternating exposure pattern. Sampling the target time t during training exposes the network to both under- and over-exposed frames, enabling the partial loss to approximate gradients effectively.

A potential degenerate solution to minimizing Eq. 7 is for the network to mimic the input's alternating exposure pattern. Our design prevents this: the fusion network, which generates the final HDR output, has no access to the target time t or exposure metadata. Although t is used in flow estimation, the fusion stage lacks temporal cues and cannot directly reproduce the exposure sequence.

In our implementation, we adopt the per-pixel exposure-aware weighting function $w(\cdot)$ from [14], visualized in Figure 2, and apply Eq. 7 in the tone-mapped domain to better reflect perceptual differences. This results in the following self-supervised reconstruction loss:

$$\mathcal{L}_{self\_recon} = w(L_t)\cdot\left\|\mathcal{T}(\hat{H}_t) - \mathcal{T}(l_t)\right\|_1 \tag{8}$$

Model	PU-tone mapping		μ-tone mapping		$Q^*_{LPIPS}\downarrow$	Time
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑		(ms)
HDRFlow [28] (Trivial Copy)	24.60	0.7630	30.62	0.8463	0.1522	24.14
M2M-VFI [13]+ HDRFlow	28.90	0.8726	34.90	0.9190	0.1163	183.08
EMA-VFI [30] + HDRFlow	29.79	<u>0.8934</u>	<u>35.74</u>	<u>0.9334</u>	0.1164	<u>99.08</u>
DeepHS-HDR [18] (EMA-VFI)	<u>29.83</u>	**0.9081**	35.53	**0.9414**	0.1266	417.61
DeepHS-HDR [18] (M2M-VFI)	29.16	0.8844	34.91	0.9259	0.1274	583.81
Ours	**29.88**	0.8927	**35.77**	0.9326	**0.1010**	**28.25**
Ours (Self-supervise)	29.29	0.8835	35.23	0.9266	<u>0.1104</u>	**28.25**

Table 1: Results on our testing set with 2× interpolation. The best score in each column is **bolded**, while the second best is <u>underlined</u>

Finally, we introduce a cycle consistency loss to enhance temporal coherence in the absence of supervision. Given a predicted HDR frame $\hat{H}_t$, we use it along with t_0 and t_2 to re-predict the central frame at t_1, enforcing:

$$\mathcal{L}_{self} = \mathcal{L}_{self\_recon} + \lambda_{cycle}\mathcal{L}_{cycle}, \tag{9}$$

where $\mathcal{L}_{cycle}$ is the $\mathcal{L}_{self\_recon}$ loss applied to the cyclically reconstructed frame, and λ_{cycle} balances its contribution.

4 Dataset

Existing datasets are insufficient for training and evaluating HDR video reconstruction and interpolation under alternating exposures. To address this, we introduce a dataset of 95 raw HDR videos (32,427 frames) captured at 60 FPS in 12-bit ProRes RAW using a Sony FX6 camera. The videos cover diverse scenes, lighting, and motion for robust model development.

Ground Truth. We decode RAW footage into linear HDR radiance frames using Adobe Premiere Pro. For low-light scenes, we apply video denoising selectively to improve signal quality while preserving details.

Input Generation. From each video, we synthesize two exposure levels: a base exposure (EV0) and a higher exposure (e.g., EV4), selected to span the effective captured dynamic range (approximately ∼12 stops). These frames are interleaved to simulate alternating exposures. To maintain realistic noise characteristics, only the higher-exposure frames are denoised. The resulting sequences are saved in 8-bit sRGB format at 30 FPS by dropping every other frame. This processed data serves as both the training and evaluation set for our model. See supp. materials for additional details about the dataset and processing pipeline.

5 Experiments

Implementation details. We trained our network on an NVIDIA L40 GPU using the Adam optimizer for a total of 40 epochs. The learning rate was set to 0.001 with a 3-epoch warm-up, then decreased to 0.0001 and halved at epochs 20 and 30. We used a batch size of 16 for supervised training and 8 for self-supervised training. During training, LDR frame sequences were randomly cropped to 512×512; for self-supervised training, the input crops

Model	PU-tone mapping		μ-tone mapping		$Q^*_{LPIPS}\downarrow$	Time
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑		(ms)
HDRFlow [28] (Trivial Copy)	23.51	0.7244	29.58	0.8184	0.16714	24.14
M2M-VFI [13]+ HDRFlow	25.45	0.7944	31.56	0.8663	0.1591	245.23
EMA-VFI [30] + HDRFlow	26.64	0.8334	32.71	0.8938	0.1590	246.96
DeepHS-HDR [18] (EMA-VFI)	26.89	**0.8545**	32.80	**0.9069**	0.1702	1145
DeepHS-HDR [18] (M2M-VFI)	25.76	0.8133	31.72	0.8798	0.1734	1142
Ours	**27.18**	0.8385	**33.21**	0.8979	**0.1386**	**66.14**
Ours (Self-supervise)	26.32	0.8176	32.40	0.8839	0.1520	**66.14**

Table 2: Results on our testing set with 4× interpolation.

Method	$\text{PSNR}_{\text{PU}}\uparrow$	$\text{SSIM}_{\text{PU}}\uparrow$	$\text{PSNR}_{\mu}\uparrow$	$\text{SSIM}_{\mu}\uparrow$	$Q^*_{LPIPS}\downarrow$
Ours (Self-supervise)	**29.29**	**0.8835**	**35.23**	**0.9266**	**0.1104**
W/o cycle loss	28.36	0.8656	34.35	0.9152	0.1218
W/o FlowNet features	29.05	0.8793	35.00	0.9237	0.1108
Using implicit function of t	28.54	0.8701	34.52	0.9183	0.1206

Table 3: Ablation study on the impact of loss design and interpolation architecture choices.

were augmented with random rotations and flips. In self-supervised training, λ_{cycle} was set to 0.1.

5.1 Comparison with Baselines

We compare our method against two groups of baselines. The first combines state-of-the-art video interpolation models (M2M-VFI [13] and EMA-VFI [30]) with an HDR video reconstruction method (HDRFlow [28]): the target frame at time t is first interpolated, then used with its neighbors as input to the HDR reconstruction network. The second group follows the DeepHS-HDR pipeline [18], combined with various interpolation models (M2M-VFI [13] and EMA-VFI [30]). As a lower bound, we include a trivial copy baseline that simply uses the nearest HDR output to approximate the frame at time t. All baselines were fine-tuned on our dataset using the authors' recommended settings. For DeepHS-HDR, we re-implemented the method based on the paper, as no official code was available.

Quantitative results. Tables 1 and 2 report results on our test set for 2× (30→60 fps) and 4× (15→60 fps) video interpolation. We evaluate using full-reference metrics: PSNR and SSIM computed on tone-mapped outputs using both perceptually uniform (PU)[1] and μ-law mappings[15], the learned perceptual metric Q^*_{LPIPS} [2], and runtime on an RTX 4070 Super at 1280×720 resolution. Our supervised model matches or outperforms baselines in quality while achieving a 2.5× speedup. Notably, our self-supervised variant also surpasses several supervised baselines, demonstrating strong performance without HDR ground truth.

Qualitative results. Figure 3 presents visual comparisons between our method and existing baselines on the test set. Our method consistently produces sharper images with greater detail, particularly in regions with large or fast motion. We attribute this improvement to our single-stage pipeline, which jointly performs HDR reconstruction and interpolation in a single pass. In contrast, two-stage methods typically rely on same-exposure frame pairs (e.g., EV0–EV0 or EV4–EV4) for motion interpolation, which can result in inaccurate or blurred estimates. As a result, these methods struggle to handle fast motion across frames with varying exposures.

Figure 3: Visual comparison of HDR frame interpolation results. We abbreviate M2M-VFI + HDRFlow as M2M-HDRF and DeepHS-HDR + EMA-VFI as DeepHS-EMA. The "Input Overlay" shows the average of input frames in the linear irradiance domain. All linear outputs are tone-mapped using Photoshop's HDR tone mapping for display purposes.

5.2 Ablation Study

We conduct an ablation study to evaluate the contributions of key components in our framework. The results are detailed in Table 3.

Loss function. Using only the self-supervised reconstruction loss (i.e., without the cycle-consistency term) already yields good results, highlighting the effectiveness of our weighted log loss. Adding the cycle-consistency loss further enhances performance across all metrics—particularly in perceptual quality—by enforcing temporal coherence and aiding in the constraining of poorly exposed regions.

Architecture Design. Removing the reuse of FlowNet features leads to drops in accuracy, emphasizing its role in refining latent flow representation within InterpNet. We also evaluate an alternative model where the target time t is represented implicitly by concatenating it to the input. This variant performs noticeably worse, demonstrating the benefit of an explicit, structured interpolation mechanism—especially important in self-supervised settings where no ground truth labels are available to anchor the temporal mapping.

6 Conclusion

We presented a novel unified framework for high-frame-rate HDR video synthesis from exposure-bracketed video, enabling both reconstruction and interpolation in a single end-to-end trainable model. Our method exploits the temporal redundancy and exposure variation across frames to produce temporally consistent HDR video at arbitrary timesteps. By introducing a lightweight interpolation module grounded in quadratic motion modeling and a self-supervised training framework, we reduce the reliance on paired HDR video data while maintaining competitive reconstruction quality. We further contribute a high-quality dataset tailored to this underexplored task. Extensive experiments demonstrate that our approach achieves state-of-the-art accuracy with significantly lower computational cost, making real-time HDR video synthesis feasible on mid-range GPUs. We hope this work and dataset inspire further research in HDR video interpolation and reconstruction for commodity camera systems.

Acknowledgments

This study was funded in part by the Canada First Research Excellence Fund for the Vision: Science to Applications (VISTA) programme and an NSERC Discovery Grant. The authors would also like to acknowledge the benefactors and guides who helped capture the dataset: Jacquie, Chuck and Dylan Canham, Sai Tedla, Christina Das, Justine Giroux, Matthew Kowal, Basile Van Hoorik, Piet Brodowski, Javier Vazquez-Corral, Dana Xue, Maria Pilligua and David Serrano-Lozano, Chris Vidal, Christopher Trombley, Mason Cross, Max Gustafson and Caroline Hess.

References

[1] Maryam Azimi and Rafal K. Mantiuk. PU21: A novel perceptually uniform encoding for adapting existing quality metrics for HDR. In *IEEE Picture Coding Symposium*, 2021.

[2] Peibei Cao, Rafal K. Mantiuk, and Kede Ma. Perceptual assessment and optimization of HDR image rendering. In *CVPR*, 2024.

[3] Yakun Chang, Chu Zhou, Yuchen Hong, Liwen Hu, Chao Xu, Tiejun Huang, and Boxin Shi. 1000 FPS HDR video with a spike-rgb hybrid camera. In *CVPR*, 2023.

[4] 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. In *ICCV*, 2021.

[5] Haesoo Chung and Nam Ik Cho. LAN-HDR: luminance-based alignment network for high dynamic range video reconstruction. In *ICCV*, 2023.

[6] Ugur Çogalan, Mojtaba Bemana, Hans-Peter Seidel, and Karol Myszkowski. Video frame interpolation for high dynamic range sequences captured with dual-exposure sensors. *Computer Graphics Forum*, 2023.

[7] Paul E. Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH*, 1997.

[8] Pierre-Étienne H. Fiquet and Eero P Simoncelli. A polar prediction model for learning to represent visual transforms. In *NeurIPS*, 2023.

[9] Jan Froehlich, 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 *SPIE Digital Photography X*, 2014.

[10] Michael D Grossberg and Shree K Nayar. What is the space of camera response functions? In *CVPR*, 2003.

[11] Jun Hu, Orazio Gallo, Kari Pulli, and Xiaobai Sun. HDR deghosting: How to deal with saturation? In *CVPR*, 2013.

[12] Mengshun Hu, Kui Jiang, Zhihang Zhong, Zheng Wang, and Yinqiang Zheng. IQ-VFI: implicit quadratic motion estimation for video frame interpolation. In *CVPR*, 2024.

[13] Ping Hu, Simon Niklaus, Stan Sclaroff, and Kate Saenko. Many-to-many splatting for efficient video frame interpolation. In *CVPR*, 2022.

[14] Nima Khademi Kalantari and Ravi Ramamoorthi. Deep high dynamic range imaging of dynamic scenes. *ACM ToG*, 2017.

[15] Nima Khademi Kalantari and Ravi Ramamoorthi. Deep HDR video from sequences with alternating exposures. In *Computer Graphics Forum*. Wiley Online Library, 2019.

[16] Nima Khademi Kalantari, Eli Shechtman, Connelly Barnes, Soheil Darabi, Dan B Goldman, and Pradeep Sen. Patch-based high dynamic range video. *ACM ToG*, 2013.

[17] Sing Bing Kang, Matthew Uyttendaele, Simon Winder, and Richard Szeliski. High dynamic range video. *ACM ToG*, 22(3), 2003.

[18] Zeeshan Khan, Parth Shettiwar, Mukul Khanna, and Shanmuganathan Raman. DeepHS-HDRVideo: Deep high speed high dynamic range video reconstruction. In *ICPR*, 2022.

[19] Didier Le Gall. MPEG: A video compression standard for multimedia applications. *Digital Multimedia Systems*, 34(4), 1991.

[20] Gustav Le Gray. Brig on the water, 1856. URL www.metmuseum.org/art/collection/search/261937.

[21] Yuelong Li, Chul Lee, and Vishal Monga. A maximum a posteriori estimation framework for robust high dynamic range video synthesis. *IEEE TIP*, 2016.

[22] Yu-Lun Liu, Yi-Tung Liao, Yen-Yu Lin, and Yung-Yu Chuang. Deep video frame interpolation using cyclic frame generation. In *AAAI*, 2019.

[23] Karol Myszkowski, Rafal Mantiuk, and Grzegorz Krawczyk. *High dynamic range video*, volume 5. Morgan & Claypool Publishers, 2008.

[24] Simon Niklaus and Feng Liu. Context-aware synthesis for video frame interpolation. In *CVPR*, 2018.

[25] Masha T. Pourazad, Colin Doutre, Maryam Azimi, and Panos Nasiopoulos. HEVC: The new gold standard for video compression. *IEEE Consumer Electronics Magazine*, 2012.

[26] Yong Shu, Liquan Shen, Xiangyu Hu, Mengyao Li, and Zihao Zhou. Towards real-world HDR video reconstruction: A large-scale benchmark dataset and A two-stage alignment network. In *CVPR*, 2024.

[27] Quang Nhat Tran and Shih-Hsuan Yang. Efficient video frame interpolation using generative adversarial networks. *Applied Sciences*, 10(18), 2020.

[28] Gangwei Xu, Yujin Wang, Jinwei Gu, Tianfan Xue, and Xin Yang. HDRFlow: Real-time HDR video reconstruction with large motions. In *CVPR*, 2024.

[29] Xiangyu Xu, Li Siyao, Wenxiu Sun, Qian Yin, and Ming-Hsuan Yang. Quadratic video interpolation. *NeurIPS*, 32, 2019.

[30] Guozhen Zhang, Yuhan Zhu, Haonan Wang, Youxin Chen, Gangshan Wu, and Limin Wang. Extracting motion and appearance via inter-frame attention for efficient video frame interpolation. In *CVPR*, 2023.

[31] Zhilu Zhang, Haoyu Wang, Shuai Liu, Xiaotao Wang, Lei Lei, and Wangmeng Zuo. Self-supervised high dynamic range imaging with multi-exposure images in dynamic scenes. In *ICLR*, 2024.