# Ev-iCRF: Self-supervised Event-guided iCRF Estimation for HDR Image Reconstruction

**Xucheng Guo[1], Bing Li[1], Ling Wang[2], Yiran Shen[1]*,**

[1]School of Software, Shandong University.
[2]School of Electrical and Electronic Engineering, Nanyang Technological University.
{xucheng.guo, bing.li}@mail.sdu.edu.cn; linwang@ntu.edu.sg; yiran.shen@sdu.edu.cn

## Abstract

In this paper, we present **Ev-iCRF**, a novel self-supervised pipeline for high dynamic range (HDR) image reconstruction from a single-exposure low dynamic range (LDR) image, guided by asynchronous event streams generated by a bio-inspired event camera. The highlight of **Ev-iCRF** lies in its formulation of the inverse camera response function (iCRF) based on Event-LDR Correspondence. By leveraging the HDR properties of event data, the method enables direct iCRF estimation, offering a new perspective for event-guided HDR imaging. The pipeline is trained in a self-supervised manner using formulation-driven iCRF estimation loss and refinement loss, without the need for synchronized HDR supervision. **Ev-iCRF** adopts a two-stage coarse-to-fine reconstruction pipeline, allowing effective fusion of features from both LDR image and event data. The event information is used to optimize the iCRF, enabling accurate HDR reconstruction from LDR inputs. We evaluate **Ev-iCRF** on real-world datasets, and results show that it outperforms state-of-the-art methods in HDR reconstruction accuracy. Moreover, the reconstructed images demonstrate improved texture fidelity and structural detail.

## Introduction

High dynamic range (HDR) imaging plays a critical role in applications such as autonomous driving (Paul and Chung 2018; Li, Qiao, and Ruichek 2015) and film production (Hasinoff et al. 2016; Reinhard et al. 2015; Tocci et al. 2011), as it provides a wider exposure range than traditional imaging techniques, enabling clearer scene details under challenging lighting conditions. A common approach to reconstruct the HDR image from the single-exposure low dynamic range (LDR) input is inverse tone mapping (iTMO) (Gabriel 2017; Lee, An, and Kang 2018; Khan, Khanna, and Raman 2019; Ning et al. 2018), which aims to extend the dynamic range by inferring texture details in underexposed regions. However, a fundamental limitation remains: recovering details that fail to be captured by the RGB camera due to inherent limited dynamic range. Therefore, LDR image often lose critical information in both highlight and shadow regions as a result of saturation. This loss poses a significant challenge to high-quality HDR reconstruction, making full dynamic range recovery an ongoing problem in the field.

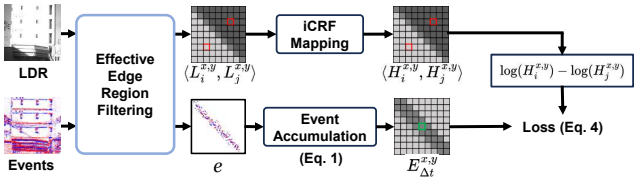


Figure 1: **Illustration of self-supervised iCRF estimation guided by event data.** LDR edge pairs are mapped to HDR using the iCRF, and event data provide irradiance changes. We assume that events (e.g., green box) occurring over a short interval $\Delta t \to 0$ are triggered by the log-intensity difference between HDR edge pairs (e.g., red boxes), which enables loss-based optimization of the iCRF.

Event cameras are emerging bio-inspired sensors that show great potential in addressing the aforementioned problem for HDR reconstruction from a single-exposure LDR image (Han et al. 2020). Unlike traditional CCD/CMOS cameras, event cameras asynchronously generate streams of events in response to per-pixel changes in brightness, offering significantly higher dynamic range (iniVation 2025). This property makes them a promising complementary modality for overcoming the limitations of conventional imaging systems. Existing HDR imaging approaches using event cameras can be broadly categorized into pure event-based and event-guided methods. Event-based methods reconstruct HDR image directly from event streams (Rebecq et al. 2019; Zhu et al. 2022; Weng, Zhang, and Xiong 2021). However, since event cameras primarily capture sparse edge information and lack rich semantic content, the resulting reconstructions often exhibit reduced detail and perceptual quality. In contrast, event-guided methods aim to enhance HDR reconstruction by fusing single-exposure LDR image with synchronized events (Han et al. 2023), leveraging the complementary strengths of both modalities.

However, existing event-guided methods face two critical limitations. First, most rely on end-to-end learning pipelines that lack explicit physical interpretability, which limits their transparency and generalizability across diverse scenarios. Second, they typically require pixel-aligned

*Corresponding author.

HDR–LDR–Event triples for supervised training. Such dataset is extremely scarce, as obtaining HDR ground truth for dynamic scenes demands complex multi-camera beam-splitting systems (Zou et al. 2024), making large-scale collection impractical.

To address the aforementioned limitations, we propose **Ev-iCRF**, a two-stage self-supervised approach for event-guided HDR reconstruction. The key insight underlying **Ev-iCRF** is that event data inherently encode brightness changes, providing rich self-supervision signals in regions where information is lost in LDR image. Fig. 1 illustrates why event data can effectively guide the estimation of the iCRF. **Ev-iCRF** formulates the iCRF based on Event-LDR Correspondence (**Event-Guided iCRF Estimation**) and constructs self-supervised iCRF estimation loss (**Event-guided iCRF Initialization**) and refinement loss (**Complementary Mask-driven iCRF Optimization Module**), using a coarse-to-fine pipeline to directly guide iCRF estimation with the HDR characteristics of event data. Specifically, in the first stage, the method coarsely estimates an iCRF by integrating event data with LDR image. We introduce a feature fusion component for cross-modality fusion and an event-guided iCRF estimation component that maps the LDR image to HDR using imaging models, in which event data are associated with the LDR image to enable self-supervision, as illustrated in Fig. 1. In the second stage, we implement complementary mask-driven optimization, which enhances parameter estimation accuracy through information-orthogonal sampling space construction. This spatial sampling strategy maximizes mutual information independence across distinct regions.

The main contributions can be summarized as follows,

- We propose **Ev-iCRF**, a novel self-supervised event-guided iCRF estimation approach for HDR reconstruction from a single-exposure LDR image, without the need for HDR image as ground truth.
- To accurately estimate the iCRF, we design a two-stage coarse-to-fine approach that leverages the encoding of brightness changes by events to supervise the iCRF estimation process for accurate HDR reconstruction.
- Extensive evaluations on datasets demonstrate that **Ev-iCRF** not only achieves high HDR reconstruction accuracy from a single-exposure LDR image—with enhanced texture fidelity and structural detail—but also outperforms sota methods that rely on ground-truth supervision.

## Related Work

**Event-Guided HDR Image Reconstruction.** Recent advancements in event-based HDR reconstruction have made significant strides through multimodal learning approaches. (Han et al. 2020) enhanced HDR imaging by integrating intensity maps from event cameras with LDR images, later expanding their framework (Han et al. 2023) with a recurrent chroma compensation network to improve color consistency, alongside an intensity upsampling module for high-resolution reconstruction. (Wang et al. 2020) developed a unified sparsity-aware network that simultaneously processes events and LDR inputs, enabling denoising, deblurring, and super-resolution. Several multimodal frameworks have emerged to address key challenges in event-guided HDR imaging. (Yang et al. 2023) introduced HDRev-Net, which tackles modality alignment and flicker suppression through dedicated feature fusion modules. (Messikommer et al. 2022) combined exposure bracketing with event data to mitigate saturation and motion artifacts while preserving color fidelity. (Shaw et al. 2022) utilized attention mechanisms and feature distillation for effective cross-modality integration. Recently, (Guo et al. 2024) introduced a diffusion-based fusion module that achieves 12-stop HDR imaging with enhanced artifact suppression in high-contrast regions. (Weng, Li, and Huang 2024) proposed a joint framework that enhances images with detailed textures in high dynamic range scenes by combining frames with locally underexposed regions and event streams.

By contrast, self-supervised methods are rare. (Xiaopeng et al. 2024) proposed a framework that primarily focused on reconstructing HDR images from dynamically blurred LDR images. It introduces an exposure synthesis and decomposition module, linking the reconstructed HDR image to the LDR image. This method primarily leverages event data to remove motion blur from LDR images in high-speed scenes, with a focus different from that of **Ev-iCRF**. Specifically, it constructs a self-supervised loss based on the relationship between blurred and sharp LDR images, rather than utilizing event data itself in a self-supervised manner within the loss formulation, as done in **Ev-iCRF**.

**Event-to-HDR Image Reconstruction.**

Recent advancements in event-to-HDR reconstruction have focused on improving resolution and handling extreme lighting. (Wang et al. 2019) pioneered event-to-HDR conversion via conditional GANs, while later works (Mostafavi, Wang, and Yoon 2021; Wang, Kim, and Yoon 2020, 2021) tackled resolution limits through supervised and unsupervised super-resolution. To enhance robustness under challenging illumination, (Zhang et al. 2021) proposed an SNN-CNN hybrid for synthetic aperture imaging that preserves fine scene details.

Recent efforts also emphasize architectural efficiency and real-world applicability. (Wang et al. 2021; Wang, Chae, and Yoon 2021; Zheng et al. 2023) introduced knowledge distillation frameworks for joint HDR reconstruction and task-specific learning in non-ideal conditions. Meanwhile, (Zou et al. 2024) presented a keyframe-guided recursive CNN that mitigates event sparsity in high-speed HDR video and released a real-world dataset to address data scarcity in event-based HDR research.

## Methodology

In this section, we first introduce the formulation of the event-guided modeling of the inverse camera response function (iCRF), denoted as $f^{-1}$, which serves as the foundation for the design of **Ev-iCRF**. We then describe the detailed architecture of **Ev-iCRF**. As illustrated in Fig. 2, **Ev-iCRF** adopts a two-stage, coarse-to-fine pipeline, consisting of two primary modules highlighted in different colors. The first stage includes the **Event-Guided iCRF Estimation** module, which produces a coarse estimate of the $f^{-1}$ based on

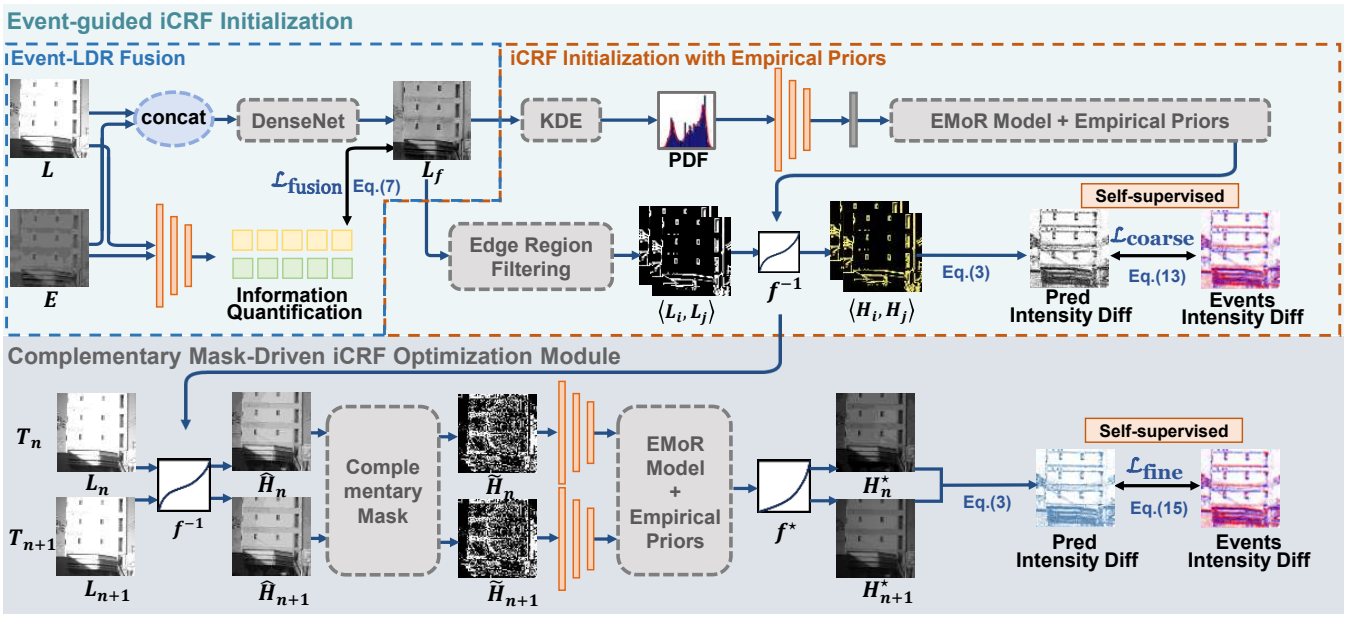


Figure 2: Overview of the Ev-iCRF pipeline. Ev-iCRF adopts a coarse-to-fine two-stage optimization network. First, the event frame and LDR image are processed by a fusion network to supplement missing texture details in the LDR image. Subsequently, an event-guided iCRF estimation network generates an initial coarse iCRF. Finally, a complementary mask-driven refinement network further optimizes iCRF to achieve high precision.

the correspondence between the LDR input image $L$ and its associated event data. The second stage consists of the **Complementary Mask-Driven iCRF Optimization** module, which refines the $f^{-1}$ using additional guidance from event data and adjacent LDR frames , thereby yielding a refined mapping $f^{\star}$ and enabling reliable event-guided HDR reconstruction $H^{\star}$. This two-stage pipeline ensures robust self-supervised estimation of the iCRF for accurate HDR reconstruction. (A summary of notations is provided in the appendix for clarity.)

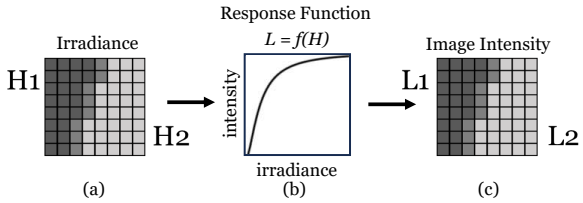


Figure 3: In the camera imaging model, the camera response function is used to map irradiance to intensity.

## Event-Guided iCRF Estimation

Event data inherently encode HDR information, as they continuously capture changes in scene irradiance. This characteristic enables event data to guide the mapping from LDR-to-HDR reconstruction without relying on ground-truth HDR supervision. In other words, they facilitate the design of a self-supervised reconstruction approach. In this section, we leverage the fundamental imaging principles of event cameras to derive an event-guided modeling of iCRF, which serves as the basis for reconstructing HDR image from LDR input.

Specifically, consider an image composed of two adjacent regions with distinct and uniform colors, The corresponding irradiance values can be denoted as $\langle H_1, H_2 \rangle$, as shown in Fig. 3(a). During the imaging process, these irradiance values undergo a transformation governed by the camera response function (CRF), expressed as, $L = f(H)$ which maps physical irradiance to observed image intensity. This transformation, illustrated in Fig. 3(b), characterizes the nonlinear relationship between irradiance and pixel intensity. As a result, the final image contains intensity values $\langle L_1, L_2 \rangle$ corresponding to the original irradiance pair, as depicted in Fig. 3(c).

If we can obtain a set of intensity pairs $\langle L_1, L_2 \rangle$ corresponding to uniformly colored regions in the LDR image, along with their associated irradiance pairs $\langle H_1, H_2 \rangle$, we can estimate the iCRF, $f^{-1}$ (Lin et al. 2004; Lin and Zhang 2005). Once the iCRF is estimated, the LDR image can be transformed into its corresponding HDR representation using the relation $H = f^{-1}(L)$.

In practical applications, obtaining the irradiance pairs $\langle H_1, H_2 \rangle$ typically relies on multi-exposure image synthesis, which suffers from several limitations, including increased acquisition complexity and sensitivity to motion artifacts. In contrast, event cameras—with their exceptionally high dynamic range (up to 140 dB (Posch, Matolin, and Wohlgenannt 2011)), can directly record irradiance changes in the scene, making them well-suited for HDR reconstruction.

For an LDR image $LDR_n$ captured at time $t_n$, we define an **effective edge region** as a region that satisfies two conditions: (1) it consists solely of an edge and its two adjacent uniform grayscale regions, and (2) the grayscale difference between the two regions exceeds a predefined threshold $\theta$. From each image $LDR_n$, we extract effective edge region pairs $\langle L_1, L_2 \rangle$ . The examples of the effective edge region pairs and detailed algorithm can be found in appendix. For the corresponding event data $e(t_1, t_2)$, captured during the time interval $t_n \in (t_1, t_2)$, each event is represented as a quadruple $(t, x, y, p)$, where $t$ denotes the timestamp of the event, $(x, y)$ are the spatial coordinates, and $p \in \{+1, -1\}$ indicates the polarity of the brightness change.

We establish a relationship between the event data and changes in scene irradiance. Specifically, we first accumulate the events occurring within the time interval $(t_1, t_2)$ to construct an event intensity difference at pixel $(x, y)$ is $E^{(x,y)}_{(t_1,t_2)}$. According to the principle of event camera (Lichtsteiner, Posch, and Delbruck 2008), the event intensity difference reflects the logarithmic change in irradiance between two time points at pixel $(x, y)$ , it can also use the accumulation of all event polarities and thresholds during $(t_1, t_2)$ at pixel $(x, y)$ to obtain, and it can be expressed as:

$$E^{(x,y)}_{(t_1,t_2)} = \sum_{(\tau,p)\in(t_1,t_2)} p\theta = \log H^{x,y}_{t_1} - \log H^{x,y}_{t_2} \quad (1)$$

where $\theta$ is the contrast threshold of the event camera.

For the effective edge pair sets $\langle L_1, L_2 \rangle$ extracted from the LDR image $LDR_n$ , we map each pair to the HDR domain using $f^{-1}$, yielding the corresponding irradiance set $\langle H_1, H_2 \rangle$. Let $\Delta t = t_2 - t_1$ denote the time interval of the event data and under the assumption that $\Delta t \to 0$, we establish the following relationship:

$$H_{t_1} - H_{t_2} = H_i - H_j \quad (2)$$

This implies that the irradiance change in the effective edge region over the interval $\Delta t$ corresponds to the intensity difference between the two adjacent regions in the HDR domain. From this, we can approximate the relationship between the event frame and scene irradiance as:

$$E(t_1, t_2) \approx \log(H_i) - \log(H_j). \quad (3)$$

Finally, we define the loss function as:

$$\mathcal{L} = \sum_{\langle i,j \rangle} \Big\| \log f^{-1}(L_i) - \log f^{-1}(L_j) , E^{(i,j)}(t_1, t_2) \Big\|^2_2 . \quad (4)$$

By minimizing $\mathcal{L}$, we leverage event data as a self-supervised signal to guide the estimation of the iCRF from LDR image. This formulation enables effective learning of the iCRF without requiring ground-truth HDR supervision.

## Event-guided iCRF Initialization

Based on the formulation derived above, we design an Event-Guided iCRF Initialization Module to perform coarse estimation of the iCRF. The module begins by fusing information from the event data and the LDR image, using the event data to compensate for missing textures and saturated regions in the LDR input. It then estimates the iCRF by integrating this fused representation with empirical prior knowledge of typical iCRF behaviors, enabling a physically meaningful and data-driven initialization for HDR reconstruction.

**Event-LDR Fusion.** LDR image often lose texture information in overexposed or underexposed regions due to their limited dynamic range, whereas event data can capture complementary details in such areas. To exploit this advantage, we adopt the fusion strategy proposed in (Xu et al. 2022), as illustrated in the top-left quadrant of Fig.2. Below, we briefly describe the Event-LDR fusion process; the detailed design can be found in(Xu et al. 2022).

First, the event data is converted into an intensity map via event-to-video reconstruction (Cadena et al. 2021) and concatenated with the LDR image along the channel dimension. A DenseNet (Huang et al. 2017) extracts joint features, producing a fused image that integrates textures from both modalities.

Then the fusion loss $\mathcal{L}_{\text{fusion}}$ is defined to measure how well the fused output $L_f$ preserves information from the inputs $E$ and $L$. Feature maps are extracted from five convolutional layers of DenseNet, which define as $\theta_C(E)$.

Information content is quantified via the average Frobenius norm of the Laplacian of each feature map:

$$g_E = \frac{1}{5} \sum_{j=1}^{5} \frac{1}{H_j W_j D_j} \sum_{k=1}^{D_j} \|\nabla \theta_{C^k_j}(E)\|^2_F, \quad (5)$$

with similar definitions for $g_L$ and $g_F$ (from $L$ and $L_f$). To emphasize preservation, we adopt adaptive weights $\omega_E$ and $\omega_F$, computed as:

$$[\omega_E, \omega_F] = \text{softmax}\left( \left[ \frac{g_E}{c}, \frac{g_F}{c} \right] \right), \quad (6)$$

where $c$ is a scaling constant. The final fusion loss is defined as:

$$\mathcal{L}_{\text{fusion}} = \omega_E \cdot (1 - \text{S}_{E,F}) + \omega_F \cdot (1 - \text{S}_{L,F}), \quad (7)$$

where $\text{S}_{E,F}$ and $\text{S}_{L,F}$ are SSIM values measuring structural similarity between the fused image and the respective inputs. This formulation encourages the fused output to retain complementary information from both modalities.

**iCRF Initialization with Empirical Priors.** After obtaining the fused image $L_f$, we use it to initialize the iCRF by incorporating empirical prior knowledge derived from the distribution of typical iCRFs.

The iCRF is parameterized as a 1024-dimensional vector, with elements uniformly sampled within the interval [0, 1]. To impose prior structure, we adopt the Empirical Response Model (EMoR) (Grossberg and Nayar 2003) as a statistical prior and assuming that any valid iCRF can be approximated by a linear combination of $k$ principal component basis vectors from EMoR. Following (Liu et al. 2020), we set $k = 11$. To predict the iCRF, we extract features from the LDR image and estimate the combination weights corresponding to the 11 PCA components. The iCRF is then formulated as:

$$f^{-1}(L) = \text{EMoR}(\mathbf{r}), \quad (8)$$

where $\mathbf{r} \in \mathbb{R}^k$ represents the coefficients of the EMoR basis vectors.

To obtain the principal component basis vector $\mathbf{r}$, we leverage prior knowledge indicating that the grayscale histogram of an LDR image is informative for predicting the iCRF. Specifically, we extract the set of pixel values $\{L_i\}_{i=1}^n$ from the edge pairs $\langle L_1, L_2 \rangle$ and use kernel density estimation to derive the probability density function $\hat{f}(L)$:

$$\hat{f}(L) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{L - L_i}{h}\right), \tag{9}$$

where $\hat{f}(L)$ is the estimated density at intensity value $L$, $n$ is the number of samples, $h$ is the bandwidth parameter that controls the smoothness of the estimation, and $K$ is the Gaussian kernel function.

The resulting probability density function $\hat{f}$ is sampled into a 512-dimensional vector. We employ a ResNet1D (He et al. 2016) followed by a fully connected layer for feature extraction. The probability distribution $\hat{f}$ is passed through this feature extraction network, denoted as $\mathcal{N}(\mathbf{x})$, to produce the PCA coefficient vector $\mathbf{r}$. This process can be expressed as, $\mathbf{r} = \mathcal{N}(\hat{f}(L))$. Thus, the estimated iCRF is formulated as:

$$f^{-1}(L) = \text{EMoR}(\mathcal{N}(\hat{f}(L))). \tag{10}$$

Since our method does not rely on HDR image for supervision, the estimated iCRF derived from the PCA basis vector $\mathbf{r}$ may deviate from typical empirical iCRFs observed in real-world data. To address this, we incorporate prior knowledge from the parametric iCRF model proposed by (Eilertsen et al. 2017), defined as:

$$f^{-1}(x) = \frac{(1+\beta)x^\gamma}{\beta + x^\gamma}, \tag{11}$$

where the parameters $\beta$ and $\gamma$ follow a Gaussian distribution with mean $\mu$ and variance $\sigma^2$. To align our estimated iCRF with this prior, we uniformly sample the function $f^{-1}$ to produce a 1024-dimensional vector $\mathbf{v} \in \mathbb{R}^{1024}$, and then apply a parametric fitting procedure. This results in a prior-informed iCRF expressed as:

$$f^{-1}(L) = \frac{(1+\beta)\mathbf{v}^\gamma}{\beta + \mathbf{v}^\gamma}. \tag{12}$$

After initializing the iCRF with empirical priors, we proceed to self-supervised optimize it by integrating it with the iCRF modeling process based on Event-LDR correspondence. As outlined in the previous modeling process, we extract the effective edge regions from the LDR image and establish a relationship between these regions and event intensity differences using Eq.3. Finally, we construct the loss function using Eq.4. The loss function for this module is defined as follows:

$$\mathcal{L}_{\text{coarse}} = \sum_{\langle i,j \rangle} \Big\| \log f^{-1}(L_i) - \log f^{-1}(L_j) , E^{(i,j)}(t_1, t_2) \Big\|_2^2 . \tag{13}$$

By jointly optimizing the fusion loss $\mathcal{L}_{\text{fusion}}$ and coarse iCRF estimation loss $\mathcal{L}_{\text{coarse}}$, we obtain an initial estimation of the iCRF $f^{-1}$. This coarse iCRF is then applied to the LDR image to generate the corresponding coarse HDR image.

## Complementary Mask-driven iCRF Optimization Module

In the first stage, the use of filtered effective edge regions may introduce biases in the estimation of the iCRF, primarily due to the limited sampling density in the LDR image. To mitigate this issue, the second stage introduces a Complementary Mask-Driven iCRF Optimization Module (illustrated in the lower half of Fig. 2). This module refines the iCRF estimation by leveraging adjacent LDR image and their corresponding event data, significantly improving spatial coverage. As a result, the sampling coverage of the scene increases to over 90%, enabling more accurate and robust HDR reconstruction.

Given two consecutive LDR images $L_n$ and $L_{n+1}$ captured at times $T_n$ and $T_{n+1}$, we first apply $f^{-1}$ to obtain initial HDR estimates $\hat{H}_n$ and $\hat{H}_{n+1}$. A refined mapping $f^\star$ is then introduced to optimize these estimates, yielding $H^\star = f^\star(\hat{H})$. To ensure that this refinement complements the initial stage, we define a complementary mask and extract the masked HDR region as $\tilde{H} = M^{\complement} \cdot \hat{H}$.

$$M^{\complement} = \{(x,y) \in \hat{H}\} \setminus \{(x,y) \in L_i\}_{i=1}^n, \tag{14}$$

We initialize $f^\star$ using the same empirical prior $\mathbf{m} \in \mathbb{R}^{1024}$ as in the first stage (see Eq. 10) and optimize it via the parametric prior in Eq. 12. For self-supervised learning, we adopt the Event–LDR-based iCRF modeling (Eq. 3) and define a temporal consistency loss over irradiance differences:

$$\mathcal{L}_{\text{fine}} = \sum_{(x,y)\in\tilde{H}} \Big\| \log f^\star(\tilde{H}_{n+1}) - \log f^\star(\tilde{H}_n) , E(T_{n+1}, T_n) \Big\|_2 . \tag{15}$$

Minimizing $\mathcal{L}_{\text{fine}}$ yields the refined function $f^\star$, and applying it to the masked region gives the final HDR output: $H^\star = f^\star(\tilde{H})$.

# Experiments

In this section, we evaluate **Ev-iCRF** for HDR reconstruction through comparisons with a range of competing methods to demonstrate its superiority. In addition, we conduct comprehensive ablation studies to validate the effectiveness of our key design components.

## Datasets and Metrics

The evaluations are conducted on the EventHDR dataset (Zou et al. 2024), which was collected using a novel optical system capable of capturing paired high-speed HDR videos and event streams. As the real-world dataset to provide HDR-Event pairs, EventHDR contains over 70,000 HDR images. For our experiment, we selected 6,242

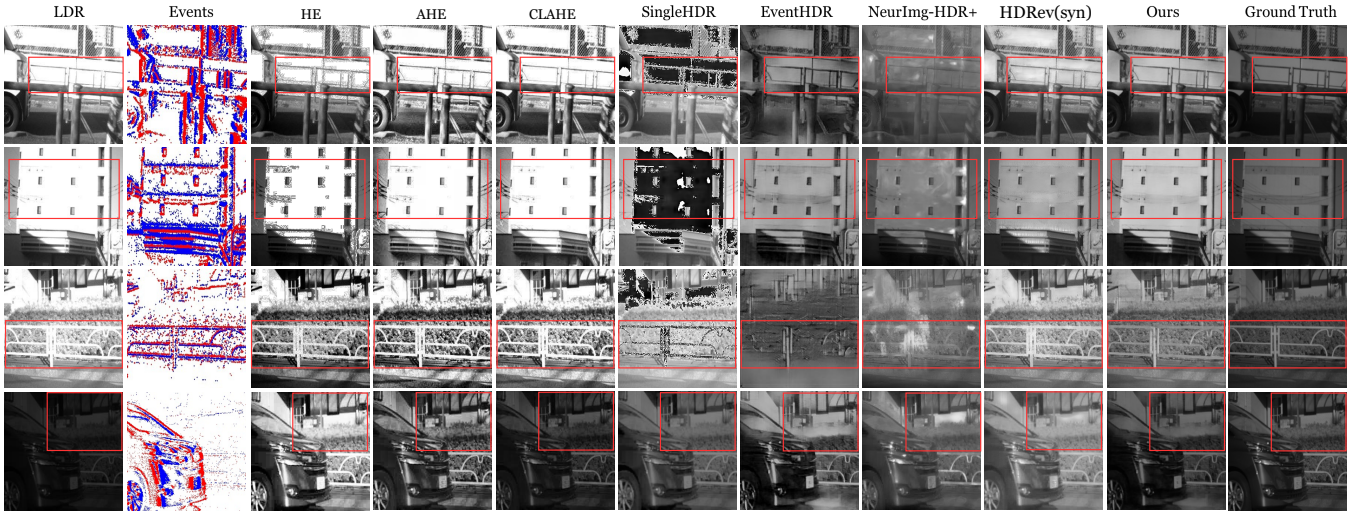


Figure 4: Qualitative comparison of different methods on the EventHDR-real dataset. Our method exhibits richer texture and structural information compared to the other methods.

HDR images and generated corresponding LDR images by multiplying with a series of exposure coefficients. Among these images, 4,242 were used for training, and 2,000 images were randomly selected from the EventHDR test set for evaluation.

We adopt three widely used evaluation metrics, conducting all assessments in the $\mu$-law tone-mapped domain. Specifically, we use $\mu$-law-PSNR, $\mu$-law-SSIM (Wang et al. 2004), and $\mu$-law-LPIPS (Zhang et al. 2018) to measure the similarity between the reconstructed HDR image and the ground truth HDR image.

## Comparison with Other Methods

**Competing Methods.** We first evaluate the performance of **Ev-iCRF** on HDR reconstruction by comparing it against a variety of state-of-the-art methods. These include traditional non-learning-based techniques such as histogram equalization (HE) (Gonzalez and Woods 2008), adaptive histogram equalization (AHE) (Zuiderveld 1994), and contrast-limited adaptive histogram equalization (CLAHE) (Zuiderveld 1994). We also compare against SingleHDR(Liu et al. 2020), a frame-only supervised HDR reconstruction method; EventHDR (Zou et al. 2024), a supervised event-only approach; and two supervised event-guided methods, HDRev (Yang et al. 2023) and NeurImg-HDR+ (Han et al. 2023). We also train our method, **Ev-iCRF**(syn) and HDRev(syn) with a synthetic dataset used in the original work of HDRev, to the generalization capability **Ev-iCRF**. It is important to note that, to the best of our knowledge, there are currently no existing self-supervised event-guided HDR reconstruction methods. **Ev-iCRF** is the first to explore this direction, filling a key gap in the HDR reconstruction landscape.

**Evaluation Results.** We first conduct quantitative evaluations of all HDR reconstruction methods using the test set from the EventHDR dataset. The evaluation metrics include SSIM, PSNR, and LPIPS, as reported in Table 1. The results show that **Ev-iCRF** consistently outperforms all competing methods, achieving the highest scores across all metrics by a significant margin. Specifically, compared to the best-performing existing method, **Ev-iCRF** improves SSIM by 18% (from 0.67 to 0.79), increases PSNR by 11.4% (from 18.47 to 20.58), and reduces LPIPS by 26.3 % (from 0.19 to 0.14).

| Methods | $\mu$-SSIM ↑ | $\mu$-PSNR ↑ | $\mu$-LPIPS ↓ | Supervision |
|---|---|---|---|---|
| HE | 0.44 | 8.09 | 0.41 | Unsupervised |
| AHE | 0.45 | 12.20 | 0.34 | Unsupervised |
| CLAHE | 0.67 | <u>18.47</u> | 0.19 | Unsupervised |
| SingleHDR | 0.55 | 13.59 | 0.32 | Supervised |
| EventHDR | 0.54 | 13.54 | 0.25 | Supervised |
| NeurImg | 0.61 | 14.92 | 0.24 | Supervised |
| HDRev(syn) | 0.45 | 9.62 | 0.33 | Supervised |
| HDRev | 0.26 | 3.40 | 0.71 | Supervised |
| **Ev-iCRF**(syn) | <u>0.70</u> | 15.83 | <u>0.18</u> | Unsupervised |
| **Ev-iCRF** | **0.79** | **20.58** | **0.14** | Unsupervised |

Table 1: Quantitative comparison on the EventHDR dataset. The best results are highlighted in **bold**, and the second-best results are <u>underlined</u>.

Notably, the variant **Ev-iCRF**(syn), trained solely on synthetic data, ranks second in SSIM and LPIPS and third in PSNR when evaluated on the real-world dataset, demonstrating the strong generalization ability of our framework across different data domains.

When compared with other event-guided LDR-to-HDR reconstruction methods, specifically HDRev and NeurImg-HDR+, both of which are trained with supervised HDR ground truth, **Ev-iCRF** achieves notable improvements. It outperforms HDRev and NeurImg-HDR+ by 0.53 and 0.18 in SSIM, achieves gains of 17.18 and 5.66 in PSNR, and reduces LPIPS by 0.57 and 0.1, respectively.

The key advantage of **Ev-iCRF** lies in its iCRF-based re-

construction, which enables accurate mapping of LDR image to the HDR domain while preserving pixel-wise relationships within the same modality. This formulation avoids the direct propagation of noise from the event stream into the reconstructed image. In contrast, methods such as HDRev and NeurImg-HDR+, which rely primarily on feature-level fusion of LDR and event data, are more vulnerable to artifacts and noise introduced by the events.

We also conduct qualitative evaluations and present several reconstructed examples in Fig. 4, with results from different methods arranged in columns for comparison. Overall, **Ev-iCRF** produces the most visually faithful results, closely resembling the ground truth. It effectively reconstructs both underexposed and overexposed regions in the LDR image, where competing methods often fail. For example, SingleHDR, a supervised LDR-only method that also reconstructs HDR images by learning iCRF, fails to reconstruct overexposed regions, often resulting in black pixels. In contrast, our method effectively preserves fine textures and structural details, even under challenging lighting conditions. (Additional qualitative examples under more challenging lighting conditions are provided in the Appendix). Moreover, compared to traditional non-learning-based methods such as HE, AHE, and CLAHE, which enhance contrast by manipulating image histograms, **Ev-iCRF** delivers significantly better visual quality by not only recovering residual details from the LDR image but also enhancing overall texture and structural consistency. Furthermore, in comparison to EventHDR, a method that relies solely on event data, our approach demonstrates clear advantages in preserving low-frequency information and mitigating the noise commonly introduced by event-based representations.

## Ablation Studies

We then conduct comprehensive ablation studies to demonstrate the effectiveness of the key components of **Ev-iCRF**.

| Methods | $\mu$-SSIM ↑ | $\mu$-PSNR ↑ | $\mu$-LPIPS ↓ |
|---|---|---|---|
| iCRF Initialization only | 0.76 | 19.73 | 0.20 |
| iCRF Optimization only | 0.63 | 14.13 | 0.45 |
| **Ev-iCRF** | **0.79** | **20.58** | **0.14** |

Table 2: Ablation results on the EventHDR dataset.

**Effectiveness of the Coarse-to-Fine Strategy.** To evaluate the effectiveness of the proposed coarse-to-fine iCRF estimation strategy, we conduct an ablation study by independently applying the two core modules, Event-Guided iCRF Initialization and Complementary Mask-Driven iCRF Optimization, to the LDR-to-HDR reconstruction task. These two variants are referred to as *iCRF Initialization Only* and *iCRF Optimization Only* in the results table. We compare their performance against the complete **Ev-iCRF** framework using standard quantitative metrics.

As shown in Table 2, the iCRF Initialization Only variant outperforms the iCRF Optimization Only variant, with a 0.13 increase in SSIM, a 5.6 improvement in PSNR, and a 0.25 reduction in LPIPS. This indicates that focusing on effective edge regions enables the network to capture more informative signal content and suppress noise. Furthermore, the complete two-stage framework (**Ev-iCRF**) achieves superior performance across all three metrics, demonstrating that the iCRF optimization module effectively complements the coarse initialization by refining unaddressed regions, thus enhancing overall reconstruction quality.

**Supervised v.s. Self-Supervised.** A key hypothesis proposed in this work is that, within a very short time interval, the intensity value of the event data can be effectively approximated by the irradiance difference in effective edge regions. Based on this assumption, the event data can serve as self-supervision for estimating the iCRF, eliminating the need for ground-truth HDR image.

To validate this hypothesis, we implement a supervised version of **Ev-iCRF** using an iCRF estimation module. In the supervised setup, the effective edge region sets $\langle L_i, L_j \rangle$ in the LDR domain are paired with corresponding ground-truth edge region sets $\langle H_i, H_j \rangle$ in the HDR domain, which are used as direct supervision for estimating the iCRF. Notably, in this supervised variant, event data are only used during feature fusion, and are not involved in the supervision process.

By comparing the iCRFs generated by the self-supervised and supervised versions of **Ev-iCRF** on the same LDR inputs, we observe that both approaches produce highly similar iCRFs, exhibiting strong consistency in both qualitative and quantitative results (see Appendix for detailed visualizations). This consistency further supports the validity of using event data as effective self-supervision.

Additionally, the quantitative evaluation results, summarized in Table 3, show that our proposed self-supervised **Ev-iCRF** achieves comparable reconstruction accuracy to its supervised counterpart, despite not requiring ground-truth HDR labels.

| Methods | $\mu$-SSIM ↑ | $\mu$-PSNR ↑ | $\mu$-LPIPS ↓ |
|---|---|---|---|
| **Ev-iCRF**(Supervised) | 0.75 | **22.58** | **0.135** |
| **Ev-iCRF**(Self-Supervised) | **0.79** | 20.58 | 0.12 |

Table 3: Comparison of supervised and self-supervised.

## Conclusion

In this paper, we introduced **Ev-iCRF**, a novel framework for self-supervised, event-guided iCRF estimation for HDR image reconstruction. By leveraging single-exposure LDR image and their synchronized event data as self-supervision, **Ev-iCRF** estimates the iCRF of an RGB camera to reconstruct HDR image. Compared to conventional end-to-end methods, **Ev-iCRF** provides stronger interpretability through its physically grounded formulation. Extensive experiments on real-world data demonstrate its superior performance, highlighting its effectiveness and practical value in HDR reconstruction tasks. Our self-supervised **Ev-iCRF** achieves reconstruction accuracy that closely approaches that of supervised **Ev-iCRF**, demonstrating its practicality for HDR imaging without requiring costly ground-truth.

## Acknowledgments

This work is partially supported by The Natural Science Foundation of Shandong Province (Major Basic Research) Grant No. ZR2024ZD12 and Open Project Program of State Key Laboratory of Virtual Reality Technology and Systems, Beihang University(No.VRLAB2025C04). This work is supported by the MOE AcRF Tier 1 SSHR-TG Incubator Grant FY24 under Grant No. RSTG7/24.

## References

Cadena, P. R. G.; Qian, Y.; Wang, C.; and Yang, M. 2021. SPADE-E2VID: Spatially-Adaptive Denormalization for Event-Based Video Reconstruction. *IEEE Transactions on Image Processing*, 30: 2488–2500.

Eilertsen, G.; Kronander, J.; Denes, G.; Mantiuk, R. K.; and Unger, J. 2017. HDR image reconstruction from a single exposure using deep CNNs. *ACM transactions on graphics (TOG)*, 36(6): 1–15.

Gabriel, E. 2017. HDR image reconstruction from a single exposure using deep CNNs. *ACM Trans. Graph.*, 36: 178–1.

Gonzalez, R. C.; and Woods, R. E. 2008. *Digital Image Processing*. Pearson Prentice Hall.

Grossberg, M. D.; and Nayar, S. K. 2003. What is the space of camera response functions? In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, volume 2, II–602. IEEE.

Guo, S.; Chen, Z.; Zhang, Z.; Chen, Y.; Xu, G.; and Xue, T. 2024. Event-assisted 12-stop HDR Imaging of Dynamic Scene. *arXiv preprint arXiv:2412.14705*.

Han, J.; Yang, Y.; Duan, P.; Zhou, C.; Ma, L.; Xu, C.; Huang, T.; Sato, I.; and Shi, B. 2023. Hybrid high dynamic range imaging fusing neuromorphic and conventional images. *IEEE Transactions on pattern analysis and machine intelligence*, 45(7): 8553–8565.

Han, J.; Zhou, C.; Duan, P.; Tang, Y.; Xu, C.; Xu, C.; Huang, T.; and Shi, B. 2020. Neuromorphic camera guided high dynamic range imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1730–1739.

Hasinoff, S. W.; Sharlet, D.; Geiss, R.; Adams, A.; Barron, J. T.; Kainz, F.; Chen, J.; and Levoy, M. 2016. Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM Transactions on Graphics (ToG)*, 35(6): 1–12.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778.

Huang, G.; Liu, Z.; Van Der Maaten, L.; and Weinberger, K. Q. 2017. Densely Connected Convolutional Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2261–2269.

iniVation. 2025. DVS346 Dynamic Vision Sensor.

Khan, Z.; Khanna, M.; and Raman, S. 2019. Fhdr: Hdr image reconstruction from a single ldr image using feedback network. In *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 1–5. IEEE.

Lee, S.; An, G. H.; and Kang, S.-J. 2018. Deep recursive hdri: Inverse tone mapping using generative adversarial networks. In *proceedings of the European Conference on Computer Vision (ECCV)*, 596–611.

Li, Y.; Qiao, Y.; and Ruichek, Y. 2015. Multiframe-based high dynamic range monocular vision system for advanced driver assistance systems. *IEEE Sensors Journal*, 15(10): 5433–5441.

Lichtsteiner, P.; Posch, C.; and Delbruck, T. 2008. A 128× 128 120 dB 15 $\mu$s Latency Asynchronous Temporal Contrast Vision Sensor. *IEEE Journal of Solid-State Circuits*, 43(2): 566–576.

Lin, S.; Gu, J.; Yamazaki, S.; and Shum, H.-Y. 2004. Radiometric calibration from a single image. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 2, II–II. IEEE.

Lin, S.; and Zhang, L. 2005. Determining the radiometric response function from a single grayscale image. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, 66–73. IEEE.

Liu, Y.-L.; Lai, W.-S.; Chen, Y.-S.; Kao, Y.-L.; Yang, M.-H.; Chuang, Y.-Y.; and Huang, J.-B. 2020. Single-image HDR reconstruction by learning to reverse the camera pipeline. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1651–1660.

Messikommer, N.; Georgoulis, S.; Gehrig, D.; Tulyakov, S.; Erbach, J.; Bochicchio, A.; Li, Y.; and Scaramuzza, D. 2022. Multi-bracket high dynamic range imaging with event cameras. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 547–557.

Mostafavi, M.; Wang, L.; and Yoon, K.-J. 2021. Learning to reconstruct hdr images from events, with applications to depth and flow prediction. *International Journal of Computer Vision*, 129(4): 900–920.

Ning, S.; Xu, H.; Song, L.; Xie, R.; and Zhang, W. 2018. Learning an inverse tone mapping network with a generative adversarial regularizer. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1383–1387. IEEE.

Paul, N.; and Chung, C. 2018. Application of HDR algorithms to solve direct sunlight problems when autonomous vehicles using machine vision systems are driving into sun. *Computers in Industry*, 98: 192–196.

Posch, C.; Matolin, D.; and Wohlgenannt, R. 2011. A QVGA 143 dB Dynamic Range Frame-Free PWM Image Sensor With Lossless Pixel-Level Video Compression and Time-Domain CDS. *IEEE Journal of Solid-State Circuits*, 46(1): 259–275.

Rebecq, H.; Ranftl, R.; Koltun, V.; and Scaramuzza, D. 2019. High speed and high dynamic range video with an event camera. *IEEE transactions on pattern analysis and machine intelligence*, 43(6): 1964–1980.

Reinhard, E.; Francois, E.; Boitard, R.; Chamaret, C.; Serre, C.; and Pouli, T. 2015. High dynamic range video production, delivery and rendering. *SMPTE Motion Imaging Journal*, 124(4): 1–8.

Shaw, R.; Catley-Chandar, S.; Leonardis, A.; and Perez-Pellitero, E. 2022. Hdr reconstruction from bracketed exposures and events. *arXiv preprint arXiv:2203.14825*.

Tocci, M. D.; Kiser, C.; Tocci, N.; and Sen, P. 2011. A versatile HDR video production system. *ACM Transactions on Graphics (TOG)*, 30(4): 1–10.

Wang, B.; He, J.; Yu, L.; Xia, G.-S.; and Yang, W. 2020. Event enhanced high-quality image recovery. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII 16*, 155–171. Springer.

Wang, L.; Chae, Y.; and Yoon, K.-J. 2021. Dual transfer learning for event-based end-task prediction via pluggable event to image translation. In *Proceedings of the IEEE/CVF international conference on computer vision*, 2135–2145.

Wang, L.; Chae, Y.; Yoon, S.-H.; Kim, T.-K.; and Yoon, K.-J. 2021. Evdistill: Asynchronous events to end-task learning via bidirectional reconstruction-guided cross-modal knowledge distillation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 608–619.

Wang, L.; Ho, Y.-S.; Yoon, K.-J.; et al. 2019. Event-based high dynamic range image and very high frame rate video generation using conditional generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10081–10090.

Wang, L.; Kim, T.-K.; and Yoon, K.-J. 2020. Eventsr: From asynchronous events to image reconstruction, restoration, and super-resolution via end-to-end adversarial learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 8315–8325.

Wang, L.; Kim, T.-K.; and Yoon, K.-J. 2021. Joint framework for single image reconstruction and super-resolution with an event camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11): 7657–7673.

Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4): 600–612.

Weng, J.; Li, B.; and Huang, K. 2024. Event-Based Image Enhancement Under High Dynamic Range Scenarios. In *Proceedings of the Asian Conference on Computer Vision*, 2456–2470.

Weng, W.; Zhang, Y.; and Xiong, Z. 2021. Event-based video reconstruction using transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2563–2572.

Xiaopeng, L.; Zhaoyuan, Z.; Cien, F.; Chen, Z.; Lei, D.; and Lei, Y. 2024. Hdr imaging for dynamic scenes with events. *arXiv preprint arXiv:2404.03210*.

Xu, H.; Ma, J.; Jiang, J.; Guo, X.; and Ling, H. 2022. U2Fusion: A Unified Unsupervised Image Fusion Network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(1): 502–518.

Yang, Y.; Han, J.; Liang, J.; Sato, I.; and Shi, B. 2023. Learning event guided high dynamic range video reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 13924–13934.

Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 586–595.

Zhang, X.; Liao, W.; Yu, L.; Yang, W.; and Xia, G.-S. 2021. Event-based synthetic aperture imaging with a hybrid network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 14235–14244.

Zheng, X.; Liu, Y.; Lu, Y.; Hua, T.; Pan, T.; Zhang, W.; Tao, D.; and Wang, L. 2023. Deep learning for event-based vision: A comprehensive survey and benchmarks. *arXiv preprint arXiv:2302.08890*.

Zhu, L.; Wang, X.; Chang, Y.; Li, J.; Huang, T.; and Tian, Y. 2022. Event-based video reconstruction via potential-assisted spiking neural network. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 3594–3604.

Zou, Y.; Fu, Y.; Takatani, T.; and Zheng, Y. 2024. EventHDR: From Event to High-Speed HDR Videos and Beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Zuiderveld, K. 1994. Contrast limited adaptive histogram equalization. In *Graphics Gems IV*, 474–485. Academic Press.