Mitigating Motion Blur in Neural Radiance Fields with Events and Frames

Marco Cannici and Davide Scaramuzza
Robotics and Perception Group, University of Zurich, Switzerland

Abstract

Neural Radiance Fields (NeRFs) have shown great potential in novel view synthesis. However, they struggle to render sharp images when the data used for training is affected by motion blur. On the other hand, event cameras excel in dynamic scenes as they measure brightness changes with microsecond resolution and are thus only marginally affected by blur. Recent methods attempt to enhance NeRF reconstructions under camera motion by fusing frames and events. However, they face challenges in recovering accurate color content or constrain the NeRF to a set of predefined camera poses, harming reconstruction quality in challenging conditions. This paper proposes a novel formulation addressing these issues by leveraging both model- and learning-based modules. We explicitly model the blur formation process, exploiting the event double integral as an additional model-based prior. Additionally, we model the event-pixel response using an end-to-end learnable response function, allowing our method to adapt to non-idealities in the real event-camera sensor. We show, on synthetic and real data, that the proposed approach outperforms existing deblur NeRFs that use only frames as well as those that combine frames and events by $+6.13dB$ *and* $+2.48dB$*, respectively.*

Multimedial Material: For videos, datasets and more visit https://github.com/uzh-rpg/evdeblurnerf.

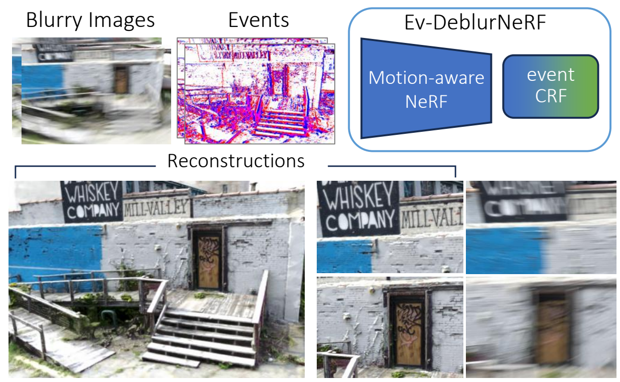

Figure 1. Ev-DeblurNeRF combines blurry images and events to recover sharp NeRFs. A motion-aware NeRF recovers camera motion and a learnable event camera response function models real camera's non-idealities, enabling high-quality reconstructions.

1. Introduction

Neural Radiance Fields (NeRFs) [27] have completely revolutionized the field of 3D reconstruction and novel view synthesis, achieving unprecedented levels of details [2, 3, 43]. As a result, they have quickly found applications in many subfields of computer vision and robotics, such as pose estimation and navigation [36, 53, 59], image processing [12, 24, 28, 47], scene understanding [17, 22, 51], surface reconstruction [1, 48, 54], and many others.

Leveraging multi-view consistency from calibrated images, NeRF exploits supervision from multiple view-points, enabling generalization to novel camera poses and the ability to render view-dependent color effects [43]. However, akin to other methods relying on photometric consistency, NeRF can only deliver high-quality reconstructions when the images used for training are perfectly captured and free from any artifact. Unfortunately, perfect conditions are seldom met in the real world.

For example, in robotics, camera motion is prevalent when capturing images, often resulting in motion blur. Under such conditions, NeRFs are unable to reconstruct sharp radiance fields, thereby impeding their practical use in real-world scenes. Although recent works [6, 18, 24, 49] have shown promising results in reconstructing radiance fields from motion-blurred images by learning to infer the camera motion during the exposure time, the task of recovering motion-deblurred NeRFs still remains significantly ill-posed. Existing image-based approaches typically fail when training images exhibit similar and consistent motion [24], and they are inherently limited by the presence of motion ambiguities and loss of texture details that cannot be recovered from blurry images alone.

In this regard, recent works have shown that event-based cameras can substantially aid the task of deblurring images captured with standard cameras [33, 37, 45, 56]. These sensors measure brightness changes at microseconds resolution and are practically unaffected by motion blur [11], thus directly addressing the aforementioned ambiguities. Motivated by these advantages, the literature has recently looked into the possibility of recovering NeRFs from events [4, 13, 16, 30, 32]. While most of the literature [4, 13, 32] focuses on event-only NeRFs, only two prior works [16, 30] investigate fusing motion-blurred images with events. E-NeRF [16] decouples sharpness and color recovery but struggles at recovering accurate color content, as the rendered images still exhibit blurred colors around sharp edges. E^2NeRF [30], on the other hand, proposes to model the camera motion by combining structure from motion with an event-aided model-based deblurring process. While effective, event supervision is only applied during the exposure time, thus potentially limiting performance under challenging motion conditions.

In this work, depicted in Fig. 1, we propose Ev-DeblurNeRF, a novel event-based deblur NeRF formulation combining learning and model-based components. Inspired by E-NeRF [16], it exploits continuous event-by-event supervision to recover sharp radiance fields. But it departs from E-NeRF in that it models the blur formation process explicitely, exploiting the direct relationship between events triggered during the exposure time and the resulting blurred frames, i.e., the so-called Event Double Integral (EDI) [29]. Unlike E^2NeRF [30], our approach employs this relation as additional training supervision, adding an end-to-end learnable camera response function that enables the NeRF to diverge from the model-based solution whenever inaccurate, resulting in higher-quality reconstructions.

We validate Ev-DeblurNeRF on a novel event-based version of the Deblur-NeRF [24] synthetic dataset, as well as on a new dataset we collected using a Color DAVIS event-based camera [19]. We show that Ev-DeblurNeRF recovers radiance fields that are $+6.13$dB more accurate than image-only baselines, and $+2.48$dB more accurate than NeRFs exploiting both images and events on real data. To summarize, our contributions are:

- A novel approach for recovering a sharp NeRF in the presence of motion blur, incorporating both model-based priors and novel learning-based modules.
- A NeRF formulation that is $+2.48$dB more accurate and $6.9\times$ faster to train than previous event-based deblur NeRF methods.
- Two new datasets, one simulated and one collected using a Color-DAVIS346 [19] event camera, featuring precise ground truth poses for accurate quality assessment.

2. Related Works

Neural Radiance Fields (NeRFs) NeRFs [27] have gained widespread attention in the research community due to their impressive performance in generating high-quality images from novel viewpoints [8, 39]. As a result, ongoing research is constantly broadening NeRFs range of capabilities, extending their use even under unideal settings. Among these, recent works have tackled the problem of recovering sharp neural radiance fields from blurry images. Deblur-NeRF [24] proposes to simultaneously learn the latent sharp radiance field and a view-dependent blurring kernel, using only blurry images as input. PDRF [6] further extends the approach by employing a coarse-to-fine architecture that exploits additional scene features to guide the blur estimation and speed up convergence, while DP-NeRF [18] improves the motion estimation by imposing rigid motion constraints on all pixels. An alternative approach is BAD-NeRF [49], which directly recovers the camera trajectory within the exposure time, taking inspiration from bundle-adjusted NeRF [21]. Despite impressive results, these methods often fail in the presence of severe camera motion or when the training views share similar motion trajectories, challenging their use with in-the-wild recordings. Our approach has a similar backbone architecture but, crucially, it additionally leverages the advantages of event-based cameras to help the reconstruction of sharp NeRFs. This allows us to recover texture and fine-grained details, resulting in improved performance and higher-quality reconstructions, even in the presence of challenging motion.

Event-based image deblurring In recent years, event-based cameras have become increasingly popular in the field of computational photography [9, 25, 40, 41, 50] due to their high dynamic range and temporal resolution. Several methods have been proposed to exploit the unique characteristics of event cameras for image deblurring, starting from model-based methods, such as the event-based double integral (EDI), which explicitly model the relationship between events triggered during the exposure time and the resulting blurry frame [29, 29]. Subsequent works build on these approaches by refining predictions with learning-based modules [14, 46] or directly learning to deblur the image by fusing events and frames [10, 37, 38, 52, 57]. These networks often pair the deblurring task with that of frame interpolation [10, 38], or make use of attention-based modules to further improve quality [37].

Recently, event-based cameras have also been used to recover sharp images from a fast-moving camera by leveraging an implicit NeRF model of the scene. Ev-NeRF [13], later improved in Robust e-NeRF [23], exploits the event generation model [7] to recover the underlying scene brightness, while EventNeRF [32] extends this approach by incorporating color event-cameras. Recent methods [16, 30] have also explored combining event-based cameras with motion-blurred images. E-NeRF [16] shows that incorporating an event supervision loss can enhance the recovery of sharp edges, but it struggles to restore sharp colors due to the lack of explicit blur modeling. Similar to ours, E^2NeRF [30] follows Deblur-NeRF [24] by modeling the camera motion during the exposure time. Notably, in our approach, we exploit continuous event-by-event supervision and employ a novel learnable camera response function that better adapts to real data, resulting in improved reconstruction under fast motion.

3. Method

The proposed Ev-DeblurNeRF aims to recover a latent sharp representation of the scene given a sequence of timestamped blurry colored images $\{(\mathbf{C}_i^{\text{blur}}, t_i)\}_{i=1}^{N_I}$ and events $\mathcal{E} = \{\mathbf{e}_j = (\mathbf{u}_j, t_j, p_j)\}_{j=1}^{N_E}$, specifying that either an increase or decrease in brightness (as indicated by the polarity $p_j \in \{-1, 1\}$) has been detected at a certain time instant t_j and pixel $\mathbf{u}_j = (u_j, v_j)$. Our method employs recent NeRF-based deblurring modules [6, 18] for fast convergence and adapts them to effectively exploit event-based information. Events in our approach serve a threefold purpose: (i) as sharp brightness supervision obtained through a single integral loss [7, 13, 16, 32], (ii) as prior, in the form of the Double Integral (EDI) [29], and lastly (iii) as a learnable event-based camera response function (CRF) that enables adapting to real event-based data. Fig. 2 provides an overview of our proposed method. In the following section, we introduce the basics of event integrals, while in Sec. 3.2 we describe the building blocks of our network.

3.1. Preliminaries

Event-based Single Integral. Let's denote the instantaneous intensity at a monochrome pixel $\mathbf{u}$ on a given time t as $I(\mathbf{u}, t)$. An event $\mathbf{e}_j$ indicates that at time t_j, the logarithmic brightness measured at the pixel location has changed by $p_j \cdot \Theta_{p_j}$ from the last time t_{j-1} an event has been generated from the same pixel location. The quantity $\Theta_{p_j} \in \mathbb{R}^+$ is a predefined threshold that controls the sensitivity to brightness changes. It follows that:

$$log(I(\mathbf{u}, t_j)) - log(I(\mathbf{u}, t_{j-1})) = \Delta L(t_{j-1}, t_j) = p_j \cdot \Theta_{p_j}. \tag{1}$$

Considering the events collected in a time period Δt and denoting as $L(\mathbf{u}, t) = log(I(\mathbf{u}, t))$ the logarithmic intensity, the following relation, here called Event-based Single Integral (ESI), holds:

$$L(t + \Delta t) - L(t) = \Theta \cdot \mathbf{E}(t) = \Theta \int_t^{t+\Delta t} p\delta(\tau) \mathrm{d}\tau, \tag{2}$$

where we dropped the dependency from the pixel location and the polarity p_j in the threshold Θ for readability, with $\delta(\tau)$ an impulse function with unit integral. Besides providing a relation between the difference in instantaneous brightness perceived at two instants and the events captured in between, Equation (4), rewritten as $I(t + \Delta t) = I(t) \cdot \exp(\Theta \cdot \mathbf{E}(t))$, also introduces a way of warping the instantaneous brightness forward or backward in time using the accumulated brightness $\Theta \cdot \mathbf{E}(t)$ measured by the event camera. This relation is utilized in the following.

Event-based Double Integral. Let's now recall that the physical image formation process of a standard frame-based camera can be mathematically represented as integrating a sequence of latent sharp images acquired during a fixed exposure time τ:

$$\mathbf{I}^{\text{blur}}(\mathbf{u}, t) = \frac{1}{\tau} \int_{t-\tau/2}^{t+\tau/2} I(\mathbf{u}, h)\, \mathrm{dh}, \tag{3}$$

where $\mathbf{I}^{\text{blur}}$ is the captured image, which we consider affected by motion blur.

Following [29], by combining Equation (4) with (3), we can finally draw a connection between the blurred image observed at time t, the events recorded during the exposure interval $\Delta T = [t - \tau/2, t + \tau/2]$ and the underlying latent sharp image $I(\mathbf{u}, t)$ at time t:

$$\mathbf{I}^{\text{blur}}(\mathbf{u}, t) = \frac{I(\mathbf{u}, t)}{\tau} \int_{t-\tau/2}^{t+\tau/2} \exp\left(\Theta \mathbf{E}(h)\right) \mathrm{dh}. \tag{4}$$

Solving for $I(\mathbf{u}, t)$, we obtain a model-based deblur of $\mathbf{I}^{\text{blur}}$, guided by the events. In the following, we use this quantity as a prior to supervise our network during training.

3.2. Event-Aided Deblur-NeRF

Our architecture takes inspiration from prior works [6, 18, 24], and is depicted in Figure 2. We aim to recover the scene as a radiance field, implemented by an MLP F_Ω, blindly, by directly modeling the blur formation process at each exposure. Analogous to Equation (3), a blurry color observation generated by the ray $\mathbf{r}(\mathbf{u}, t_i)$ cast by pixel $\mathbf{u}$ during its exposure can be described as the integral of the sharp colors observed by the ray in a time interval $\Delta T_i = [t_i - \tau/2, t_i + \tau/2]$.

Similarly to [18], we learn to estimate the motion of each ray implicitly using a neural module $G_{\mathbf{\Phi}}$. We discretize motion in a finite set of M observations and learn an $SE(3)$ field that rigidly warps pixel rays to each position q:

$$(\mathbf{e^r}_q, \mathbf{t}_q, w_q) = G_{\mathbf{\Phi}}(\mathcal{R}(\boldsymbol{l}_i); \mathcal{T}(\boldsymbol{l}_i); \mathcal{W}(\boldsymbol{l}_i)), \tag{5}$$

where $\boldsymbol{l} \in \mathbb{R}^E$ is a shared learned image embedding, and $\mathcal{R}$, $\mathcal{T}$ and $\mathcal{W}$ are independent MLPs that predict, respectively, a set of rotation matrices $\mathbf{e}_q^r \in SO(3)$, translation vectors $\mathbf{t}_q \in \mathcal{R}^3$, and view weights $w_q \in \mathcal{R}$, one for each discrete position q. The warped rays can thus be finally obtained as $\hat{\mathbf{r}}_q = \mathbf{e^r}_q \mathbf{r}(\mathbf{u}, t_i) + \mathbf{t}_q$.

Following NeRF [8], we render the color at each ray by first sampling a set of 3D points along each ray, and then query a pair of MLPs, one coarse- and one fine-grained, F_Ω^c

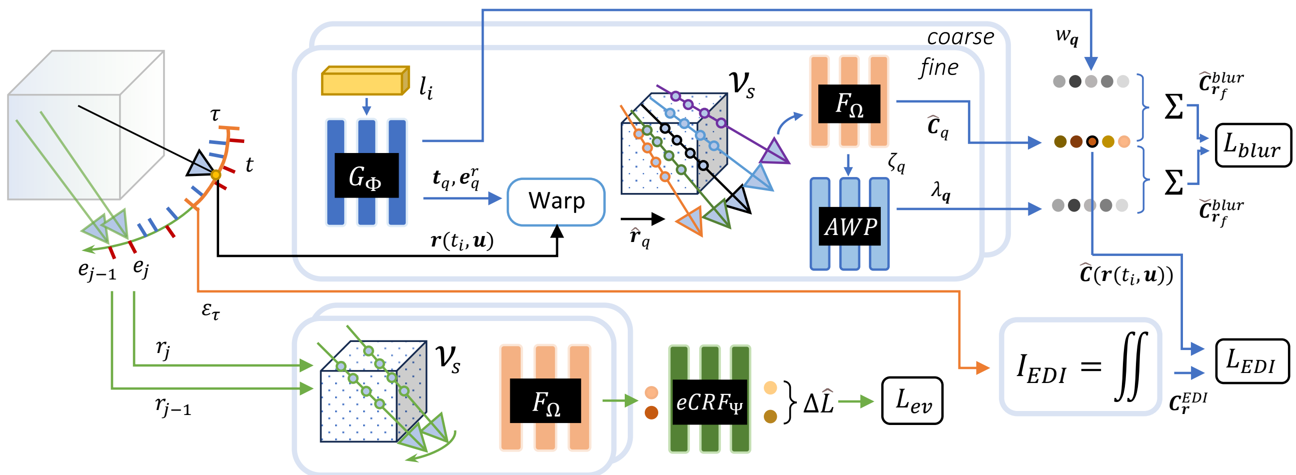

Figure 2. Architecture of the proposed Ev-DeblurNeRF model. For each given ray $\mathbf{r}(\mathbf{u}, t)$, placed at the center of the exposure time τ, we estimate a set of warped rays $\mathbf{r}_q$ using $G_{\mathbf{\Phi}}$. We then sample features from an explicit volume $\mathcal{V}$ and fed these features to F_Ω to compute blurry colors through weighted averaging with L_{blur}. We supervise the color at the center of the exposure time through $\mathcal{L}_{EDI}$ by recovering a prior-based sharp color using the event double integral, considering all events in the exposure time. Finally, we sample a pair of two consecutive events, and supervise their brightness difference, modulated by eCRF, using the observed polarity value via $\mathcal{L}_{ev}$.

and F_Ω^f, to obtain colors and density at each location. Volumetric rendering is then finally used to estimate colors $\hat{\mathbf{C}}_q$ at the predicted camera positions, which are finally fused into a blurry observation

$$\hat{\mathbf{C}}^{\text{blur}}(\mathbf{r}(\mathbf{u}, t_i)) = g\left(\sum_{q=1}^{M-1} w_q \hat{\mathbf{C}}_q\right), \tag{6}$$

where $g(\cdot)$ is a gamma correction function. Inspired by [18], we further refine the composite weights using an adaptive weight proposal network $\lambda_q = \mathcal{AWP}(\zeta_q, \boldsymbol{l}_i, \mathbf{d}_q)$, which takes the ray's samples features ζ_q, directions $\mathbf{d}_q$ and image embedding $\mathbf{l}_i$ to produce refined weights. We use these refined weights in Equation (6) in place of w_q to obtain refined colors $\tilde{\mathbf{C}}^{\text{blur}}$.

The thus rendered synthetic blurry pixel is finally supervised with a ground truth observation $\mathbf{C}_{\text{gt}}$ through:

$$E^b(\mathbf{C}^{\text{blur}}_{\mathbf{r}}) = \left\|\mathbf{C}^{\text{blur}}_{\mathbf{r}} - \mathbf{C}^{\text{blur}}_{\text{gt}}(\mathbf{r})\right\|_2^2 \tag{7}$$

$$\mathcal{L}_{\text{blur}} = \frac{1}{|\mathcal{R}_b|}\sum_{\mathbf{r}\in\mathcal{R}_b} E^b(\hat{\mathbf{C}}^{\text{blur}}_{\mathbf{r}_c}) + E^b(\hat{\mathbf{C}}^{\text{blur}}_{\mathbf{r}_f}) + E^b(\tilde{\mathbf{C}}^{\text{blur}}_{\mathbf{r}_f}), \tag{8}$$

where we consider a batch of pixels $\mathcal{R}_b$, and rewrite $\mathbf{C}^{\text{blur}}_{\mathbf{r}} = \mathbf{C}^{\text{blur}}(\mathbf{r})$. Subscripts c and f indicate if values are obtained through F^c_Ω or F^f_Ω, while $\tilde{}$ if adaptive weights are used.

Event-based supervision via learned event-CRF. When the scene is also captured by an event-based camera, as in our case, blur-free microsecond-level measurements can be exploited to further assist the reconstruction of a sharp radiance field, leveraging the relation in Equation (1) between brightness and generated events. We do so by synthesizing the left-hand side of (1), i.e., the log brightness difference perceived by the event pixel, through volumetric rendering while we take the right-hand side as a ground truth supervision, given recorded event pairs.

In particular, we estimate the log-brightness at each event $\mathbf{e}_j$, produced by the event pixel $\mathbf{u}$ at time t_j, as:

$$\hat{L}(t_j, \mathbf{u}) = log(h(eCRF_\Psi(\hat{\mathbf{C}}(\mathbf{r}_j), p_j))), \tag{9}$$

where we obtain $\hat{\mathbf{C}}(\mathbf{r}_j)$ via volumetric rendering [27] by rendering the ray $\mathbf{r}_j = \mathbf{r}(\mathbf{u}, t_j)$ cast from the camera pose $\mathbf{T}(t_j) \in SE(3)$, approximated via spherical linear interpolation [34] of the available known camera poses. Here, $eCRF_\Psi$ is an MLP that produces a modulated signal $\hat{\mathbf{C}}_e \in \mathbb{R}^3$ from the rendered color $\hat{\mathbf{C}}$ and the polarity p_j, while $h(\cdot)$ is a luma conversion function, implemented following the BT.601 [42] standard.

Given a pair of consecutive events at time t_{j-1} and t_j, we first estimate the log-brightness difference as $\Delta\hat{L}(\mathbf{u}, t_j) = \hat{L}(\mathbf{u}, t_j) - \hat{L}(\mathbf{u}, t_{j-1})$ and then compare it with that observed by the event camera, ΔL, as follows:

$$E^e(\Delta\hat{L}^t_{\mathbf{u}}) = \left\|\Delta\hat{L}^t_{\mathbf{u}} - \Delta L^t_{\mathbf{u}}\right\|_2^2 \tag{10}$$

$$\mathcal{L}_{\text{ev}} = \frac{1}{|\mathcal{U}_e|}\sum_{(t,\mathbf{u})\in\mathcal{U}_e} E^e(\Delta\hat{L}^t_{\mathbf{u}_c}) + E^e(\Delta\hat{L}^t_{\mathbf{u}_f}) + E^e(\Delta\tilde{L}^t_{\mathbf{u}_f}), \tag{11}$$

where we use the compact form $\hat{L}^t_{\mathbf{u}}$ for $\hat{L}(t, \mathbf{u})$, and apply the supervision on fine and coarse levels, as well as on adaptively refined colors. $\mathcal{U}_e$ selects pairs of pixels $\mathbf{u}$ and timestamps t corresponding to received events. Our experiments reveal that applying $\mathcal{L}_{\text{ev}}$ not only during image exposures but also between frames, similar to [16], helps in viewpoints with scarce RGB coverage, as common with fast motion.

In Equation (11), we assume the ideal event generation model of (2). However, real event pixels deviate from the ideal case [11]. Our proposed event CRF function $eCRF_{\Psi}$ learns to compensate for potential mismatches between the ideal model and that of the camera at hand, filling the gap between the rendered color space and the brightness change perceived by the event sensor. Note that, when a color event camera is used, as the one in [19], pixels record color intensity changes following a Bayer pattern. We remove the luma conversion $h(\cdot)$ function in Equation (9), and directly apply the previous loss to the color channel each pixel is responsible for. We refer to this version of the loss as $\mathcal{L}_{\text{ev-color}}$.

Double integral supervision. The eCRF just introduced provides an effective way of handling unmodeled event pixel behaviors. However, blindly recovering the event camera response to colors is not trivial since the only direct source of color supervision comes from Equation (8). In practice, the optimization problem in (11) is under constrained, as the loss, acting on the event CRF, is free to enhance texture details in the radiance field as long as they correctly render once blurred through Equation (3). Inspired by recent works [20], which suggest facilitating NeRF optimization through priors, we propose here to exploit the relationship in (4) to further constrain the NeRF training.

In particular, we consider every original ray $\mathbf{r} \in \mathcal{R}_b$ sampled when optimizing Eq. (8) originating from the mid-exposure pose of image I_i^{blur}, i.e., the rays rendering the latent sharp pixels $\mathbf{C}(\mathbf{r})$. If we simplify and assume these pixels are monochrome, they correspond to $I(\mathbf{u}, t_i)$ in Equation (4). Given this observation, we first rewrite (4) by solving for $I(\mathbf{u}, t_i)$, and then evaluate it channel-wise for the given image at time t_i and ray $\mathbf{r}$, using the observed blurry color $\mathbf{C}^{\text{blur}}$ and the events received at pixel $\mathbf{u}$. We finally collect channels into $\mathbf{C}_{\mathbf{r}}^{EDI} = \left[I^R(\mathbf{u}, t_i), I^G(\mathbf{u}, t_i), I^B(\mathbf{u}, t_i)\right]$, obtaining a model-based sharp latent color. We use this color as a prior in:

$$E^{\text{EDI}}(\hat{\mathbf{C}}_r) = \left\| \hat{\mathbf{C}}_{\mathbf{r}} - \mathbf{C}_{\mathbf{r}}^{\text{EDI}} \right\|_2^2 \tag{12}$$

$$\mathcal{L}_{\text{EDI}} = \frac{1}{|\mathcal{R}_b|} \sum_{\mathbf{r} \in \mathcal{R}_b} E^{\text{EDI}}(\hat{\mathbf{C}}_{\mathbf{r}_c}) + E^{\text{EDI}}(\hat{\mathbf{C}}_{\mathbf{r}_f}) \tag{13}$$

Fast NeRF via explicit features. The additional event-based supervision introduced in Equation (11), while enabling the reconstruction of a high-fidelity sharp NeRF, does come with a notable effect on the training time. Indeed, on top of the rays $\mathcal{R}_b$, needed for optimizing Equations (8) and (13), we also consider an additional pair of rays in $\mathcal{U}_e$ which we employ to render brightness changes across time. We overcome this aspect by taking inspiration from previous works [5, 6] showing that additional explicit features can ease convergence, making the training faster.

Inspired by the hybrid design in [6], we enhance the capabilities of F_{Ω}^c and F_{Ω}^f by incorporating dedicated TensoRF [5] volumes, which we employ as additional input feature spaces for the MLPs. In particular, given a ray $\mathbf{r_u}$ and a set of coarse and fine points $\{\mathbf{x}_k^c\}_{k=1}^S$ and $\{\mathbf{x}_k^f\}_{k=1}^S$ along the ray, we first sample feature volumes:

$$\begin{aligned} f_{s}{}_k^c = \mathcal{V}_s(\mathbf{x}_k^c),\ f_{s}{}_k^f = \mathcal{V}_s(\mathbf{x}_k^f), \\ f_{l}{}_k^c = \mathcal{V}_l(\mathbf{x}_k^c),\ f_{l}{}_k^f = \mathcal{V}_l(\mathbf{x}_k^f), \end{aligned} \tag{14}$$

with $\mathcal{V}_s$ and $\mathcal{V}_l$, respectively, a small and a large TensoRF [5] volume. We use $f_{s}{}_k^c$ as additional features in F_{Ω}^c, while we employ all the features as input to the fine-grained MLP F_{Θ}^f. The structure of F_{Ω}^c and F_{Ω}^f is analogous to that of the original NeRF [27], with the only difference that the MLP predicting σ also takes these extra features as input.

4. Experiments

4.1. Implementation Details.

Training. We build our event-based architecture starting from the Pytorch implementation of DP-NeRF [18]. We use a batch size of 1024 for rays $\mathcal{R}_b$ and 2048 for rays $\mathcal{U}_e$, and sample 64 coarse and additional 64 fine points along each ray. Following [6], we set the number of motion locations to $M = 9$. We use Adam [15] to optimize the multi-objective loss $\mathcal{L} = \lambda_b \mathcal{L}_{blur} + \lambda_e \mathcal{L}_{event} + \lambda_{EDI} \mathcal{L}_{EDI}$, where we set $\lambda_b = \mathcal{L}_{EDI} = 1$, and $\lambda_e = 0.1$. We train the model for a total of $30,000$ iterations, using an initial learning rate of $5 \cdot 10^{-3}$, which we decrease exponentially to $5 \cdot 10^{-6}$ over the course of the training. Further details on the network architectures are provided in the supplementary material.

Ev-DeblurBlender dataset. We evaluate our method on four synthetic scenes derived from the original DeblurNeRF [24] work, namely, *factory*, *pool*, *tanabata*, and *trolley*. We exclude *cozy room* from our conversion as the Blender rendering for this scene relies on an image denoising post-processing step. This step causes the rendered images to show temporally inconsistent artifacts when rendered at high FPS, thereby causing unrealistic event simulation. Differently from [24], where blurry images are obtained by randomly moving the camera at each pose, we use a single fast continuous motion, derived from DeblurNeRF's original poses, lasting around 1s. We simulate a 40ms exposure time by averaging together, in linear RGB space, images rendered at 1000 FPS. We then use the same set of images to generate synthetic events using event simulation [31], making use of a balanced $\Theta = 0.2$ event threshold and monochrome events.

Ev-DeblurCDAVIS dataset. Given the lack of real-world datasets for event-based NeRF deblur that incorpo-

Table 1. Quantitative comparison on the synthetic Ev-DeblurBlender dataset. Best results are reported in bold.

	Factory			Pool			Tanabata			Trolley			Average		
	PSNR↑	LPIPS ↓	SSIM↑	PSNR↑	LPIPS ↓	SSIM↑	PSNR↑	LPIPS ↓	SSIM↑	PSNR↑	LPIPS ↓	SSIM↑	PSNR↑	LPIPS ↓	SSIM↑
DeblurNeRF [24]	24.52	0.25	0.79	26.02	0.34	0.69	21.38	0.28	0.71	23.58	0.22	0.79	23.87	0.27	0.74
BAD-NeRF [49]	21.20	0.22	0.64	27.13	0.23	0.70	20.89	0.25	0.65	22.76	0.18	0.73	22.99	0.22	0.68
PDRF [6]	27.34	0.17	0.87	27.46	0.32	0.72	24.27	0.20	0.81	26.09	0.15	0.86	26.29	0.21	0.81
DP-NeRF [18]	26.77	0.20	0.85	29.58	0.24	0.79	27.32	0.11	0.85	27.04	0.14	0.87	27.68	0.17	0.84
MPRNet [55] + NeRF	19.09	0.37	0.56	25.49	0.39	0.64	17.79	0.42	0.51	19.82	0.31	0.62	20.55	0.37	0.58
PVDNet [35] + NeRF	22.50	0.29	0.71	23.89	0.43	0.52	20.26	0.33	0.64	22.49	0.25	0.74	22.28	0.32	0.65
EFNet [37] + NeRF	20.91	0.32	0.63	27.03	0.31	0.73	20.68	0.31	0.64	21.69	0.25	0.69	22.58	0.30	0.67
EFNet* [37] + NeRF	29.01	0.14	0.87	29.77	0.18	0.80	27.76	0.11	0.87	29.40	0.94	0.89	28.99	0.34	0.86
ENeRF [16]	22.46	0.19	0.79	25.51	0.28	0.72	22.97	0.16	0.83	21.07	0.20	0.80	23.00	0.21	0.79
E^{2}NeRF [30]	24.90	0.17	0.78	29.57	0.18	0.78	23.06	0.19	0.74	26.49	0.10	0.85	26.00	0.16	0.78
(Ours) Ev-DeblurNeRF- -	**32.84**	**0.05**	**0.94**	31.45	**0.14**	**0.84**	**29.20**	**0.06**	**0.92**	**30.60**	**0.06**	**0.93**	**31.02**	**0.08**	**0.91**
(Ours) Ev-DeblurNeRF	31.79	0.06	0.93	**31.51**	**0.14**	**0.84**	28.67	0.08	0.90	29.72	0.07	0.92	30.42	**0.08**	0.90

Table 2. Quantitative comparison on the real-world Ev-DeblurCDAVIS dataset. Best results are reported in bold.

	Batteries			Power supplies			Lab Equipment			Drones			Figures			Average		
	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓	SSIM↑
DP-NeRF [18] + TensoRF [5]	26.64	0.27	0.81	25.74	0.32	0.77	27.49	0.31	0.80	26.52	0.30	0.81	27.76	0.34	0.77	26.83	0.31	0.79
EDI [29] + NeRF	28.66	0.12	0.87	28.16	0.09	0.88	31.45	0.13	0.89	29.37	0.10	0.88	31.44	0.12	0.88	29.82	0.11	0.88
E^{2}NeRF	30.57	0.12	0.88	29.98	0.11	0.87	30.41	0.16	0.86	30.41	0.14	0.87	31.03	0.14	0.85	30.48	0.13	0.87
(Ours) Ev-DeblurNeRF	**33.17**	**0.05**	**0.92**	**32.35**	**0.06**	**0.91**	**33.01**	**0.08**	**0.91**	**32.89**	**0.05**	**0.92**	**33.39**	**0.07**	**0.90**	**32.96**	**0.06**	**0.91**

rate ground truth sharp reference images for quantitative assessment, we introduce a novel dataset composed of 5 real-world scenes. We use the Color-DAVIS346 [19] camera for recording, which captures both color events and standard frames at 346×260 pixel resolution using a RGBG Bayer pattern. We mount the camera on a motor-controlled linear slider to capture frontal-facing scenes and use the motor encoder to obtain poses at 100 Hz. We configure the Color-DAVIS346 with a 100ms exposure time and collect ground truth still images first, followed by a fast motion. Scenes feature 11 to 18 blur training views and 5 ground truth sharp poses with both seen and unseen views.

Baselines. We evaluate our method against frame-only methods as well as methods fusing both images and events. For the first category we follow previous works [18, 18, 24], and select Deblur-NeRF [24], BAD-NeRF [49], DP-NeRF [18] and PDRF [6] as the most recent NeRF-based baselines, as well as single-image and video deblurring methods, namely MPRNet [55] and PVDNet [35], followed by NeRF [27]. Similarly, for the second category, we select E-NeRF [16] and E^2-NeRF as event-based deblur NeRF architectures, and also combine frames deblurred via the events+frames EFNet [37] network with NeRF [27]. We run all baselines with default hyperparameters using the official codebases. We utilize Blender poses in Ev-DeblurBlender and motor encoder poses in Ev-DeblurCDAVIS for all baselines, including E^{2}NeRF, where we compute exposure poses via spherical linear interpolation of the available ones.

4.2. Experimental Validation

Results on Ev-DeblurBlender. We start the evaluation on the synthetic Ev-DeblurBlender dataset to first assess the performance of our method on an ideal case, i.e., where camera poses are accurate and the event generation model is close to the ideal case. Results are reported in Table 1. We test two versions of our network. The first, which we call Ev-DeblurNeRF- -, does not make use of the proposed eCRF module and EDI supervision, while the second, Ev-DeblurNeRF, incorporates the complete architecture presented in Section 3. We found Ev-DeblurNeRF- - to exhibit slightly superior performance on average on this data. As discussed in Section 3, indeed, we designed the eCRF specifically to handle possible variations between RGB and events' response functions, as well as to compensate for mismatches on the event generation model. These issues are not predominant in simulated data, explaining why adding a learnable response function does not improve performance.

Despite this, both versions largely outperform all other baselines, both event-based and frame-based. Compared to DP-NeRF [18], which uses a similar backbone architecture, our method achieves on average a $+3.34$dB higher PSNR, a 52.9% lower LPIPS [58] and 7.14% higher SSIM, highlighting the improvement gained by effectively integrating event-based supervision. This is also evident when considering baselines utilizing an image-deblurring stage prior to NeRF training, which also achieve better performance when events are used. This is the case of EFNet [37], and its variant, which we name EFNet*, that we finetune on the other 3 scenes before deblurring images of a given scene. Despite the high accuracy, these methods fail to produce scene-level consistent deblurring, causing the NeRF to reconstruct floaters and thus decreasing novel-view synthesis performance. Finally, our approach also surpasses both previous event-based deblurring NeRF methods with an average increase of $+5$dB in PSNR, a 50% reduction in LPIPS, and a 16.7% increase in SSIM. Notably, ENeRF, which does

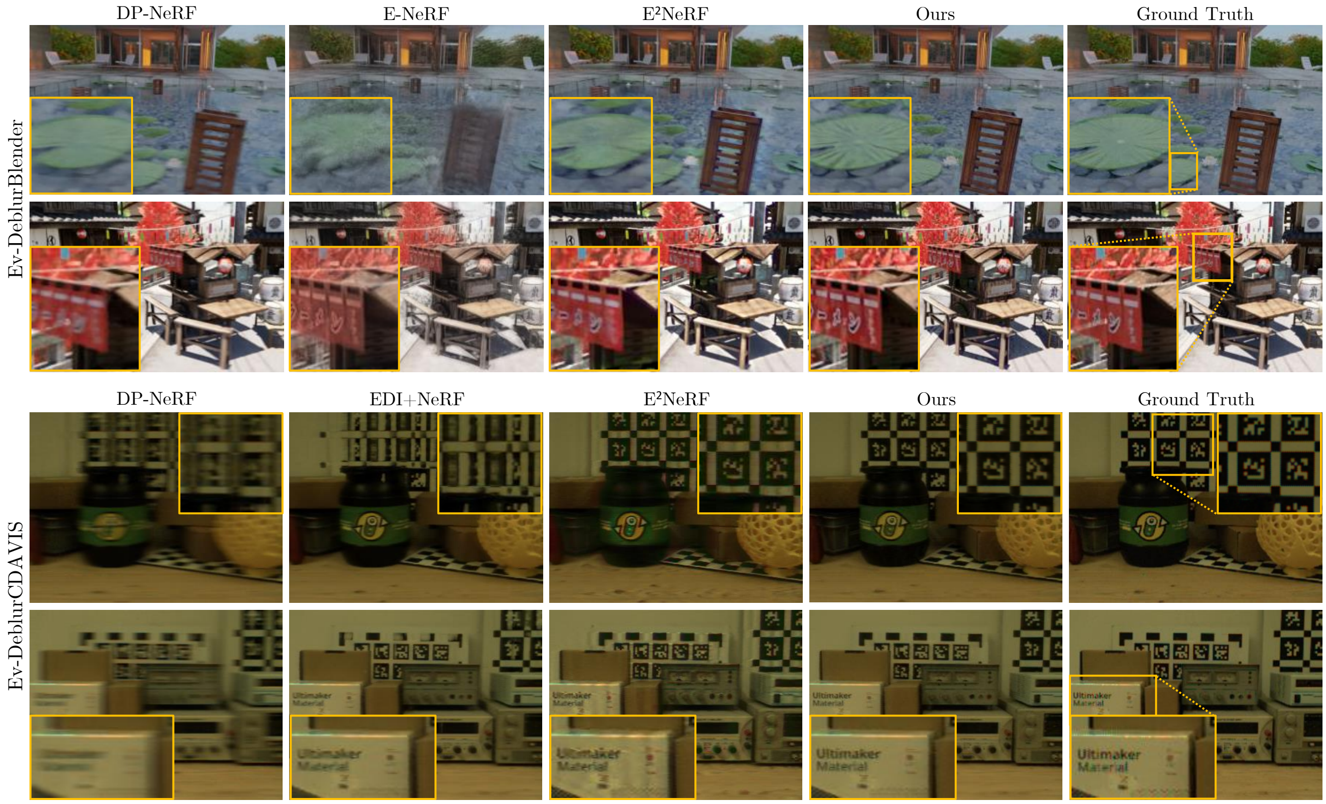

Figure 3. Qualitative comparison on synthetic (top) and real-world camera motion blur (bottom). Ev-DeblurNeRF recovers sharp and fine details, such as the letters in the last example, as well as accurate colors, outperforming other event-based and image-only methods.

not explicitly model the blur formation process, struggles to recover sharp color information, while E^{2}NeRF, exclusively employing event supervision during the exposure time, fails at fully exploiting event-based data. Our method, on the contrary, overcomes both limitations, showcasing the effectiveness of the proposed approach.

Results on Ev-DeblurCDAVIS. In Table 2, we report results obtained on data collected with a real Color-DAVIS346 camera. We select the top-performing NeRF models from the previous evaluation, namely E^{2}NeRF [30] and DP-NeRF [18], which we modify here by integrating the TensoRF modules discussed in Section 3 for a better comparison. Additionally, we include the performance metrics obtained by initially deblurring images using the model-based EDI deblur method, followed by NeRF. An extended analysis including all other baselines is provided in the supplementary materials. Once again, our proposed approach significantly outperforms all baselines, exhibiting an improvement of $+2.5$dB in PSNR and a 4.6% increase in SSIM. A qualitative comparison, depicted in Figure 3, illustrates the capability of the proposed Ev-DeblurNeRF network in reconstructing textures and details, ultimately resulting in a higher-quality novel view synthesis.

Synthesis from sparse blurry views. Utilizing the same setup used for collecting the Ev-DeblurCDAVIS dataset, we study here the robustness of the proposed approach to sparse supervision to highlight the advantage of using events not only within exposure but also in between frames. We collect an additional, longer, sequence with a back-and-forth motion and train the proposed approach with an increasing number of frames $N_f \in \{5, 9, 17, 33\}$, such that each set is a subset of the next and making sure that test poses are within training views but as furthest away as possible from them. Results are reported in Figure 4. Remarkably, Ev-DeblurNeRF attains the highest performance of all methods we tested, with its performance only decreasing by 3.46dB in PSNR when passing from 33 to just 5 views. In contrast, E^{2}NeRF and EDI+NeRF experience a decrease of 13.71dB and 15dB, respectively. These methods struggle to correctly reconstruct the radiance field from viewpoints that are only weakly supervised by blurred images. Our approach, instead, is only marginally affected. More details are provided in the supplementary material.

Robustness to motion blur. In Figure 4, we analyze how the performance of the proposed approach changes as we vary the motion blur intensity. We follow the same setup as before but this time vary the slider speed from 0.1m/s to

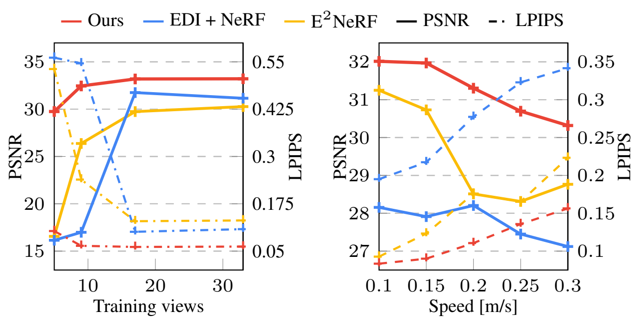

Figure 4. Analysis of the robustness to sparse training views (left) and motion blur intensity (right) on samples from the Ev-DeblurCDAVIS data.

Table 3. Ablation study on Ev-DeblurCDAVIS.

$\mathcal{V}_{c,f}$	$\mathcal{L}_{ev}$	$\mathcal{L}_{ev-color}$	$\mathcal{L}_{EDI}$	eCRF	eCRF w/ p	PSNR↑	LPIPS↓	SSIM↑
✓						27.55	0.26	0.80
✓	✓					28.24	0.14	0.85
✓	✓	✓				29.28	0.12	0.85
✓	✓	✓	✓			32.43	0.10	0.91
✓	✓	✓		✓		30.77	0.11	0.86
✓	✓	✓	✓	✓		32.90	0.07	0.91
✓	✓	✓	✓	✓	✓	**33.17**	**0.07**	**0.91**
	✓	✓	✓	✓	✓	33.03	0.08	0.91

0.3m/s in increments of 0.05m/s. Notably, Ev-DeblurNeRF demonstrates superior robustness, achieving a PSNR of 32.01dB at the highest speed. In contrast, E^{2}NeRF and EDI-NeRF achieve PSNR values of 28.77dB and 27.12dB, respectively. We attribute the higher performance to our choice of decoupling event supervision (Eq. (11)) from blur estimation (Eq. (8)). In contrast to E^{2}NeRF, which fixes the poses used to render blurry images, we leave the NeRF free of optimizing the best camera views to consider for blur estimation as well as their contribution, thus achieving better robustness to different degrees of motion.

Ablations. We conclude the evaluation by studying, in Table 3, the contribution of all the modules introduced in Section 3, using a scene derived from the *Figures* sample of Ev-DeblurCDAVIS. Adding event supervision from Equation (11) improves PSNR by $+0.69$dB, which is further increased by $+1.04$dB when the events' color channel is considered. Similarly, adding $\mathcal{L}_{EDI}$ in Equation 13 as well as the proposed eCRF module, with and without additional polarity features, also results in increased performance. Next, we study the contribution of adding the $\mathcal{L}_{EDI}$ in Equation 13 and the eCRF module. Performance increases in both cases, with a $+3.15$dB increase when adding $\mathcal{L}_{EDI}$ and a $+1.49$dB when adding the eCRF. The highest performance is achieved when both are combined and when the eCRF also utilizes polarity as input, with an increase of $+0.74$dB, and an overall improvement of $+5.62$dB in PSNR with respect to only using images. We finally validate the use $\mathcal{V}_{c,f}$ on the full configuration. Using explicit features guarantees faster training times without sacrificing performance. We obtain a slight boost in PSNR and LPIPS but, most notably, a $\times 10.8$ speedup in training convergence. This model only takes around 3 hours and 30 minutes for training on an NVIDIA A100 GPU, while the same network without $\mathcal{V}_{c,f}$ takes around 38 hours on the same hardware, as it requires more iterations at a lower learning rate. Moreover, in comparison to E^{2}NeRF, which takes around 24 hours to train, our model is 6.9 times faster.

Limitations. We structure the proposed Ev-DeblurNeRF assuming that events and frames can be recorded from the same image sensor. While this is possible with the suggested hardware, namely a ColorDAVIS camera, not all event cameras feature both modalities. While the proposed $\mathcal{L}_{EDI}$ loss requires pixel alignment to work effectively, we believe the proposed method could still be applied in more advanced stereo setups, such as the ones in [25, 41], especially exploiting the proposed eCRF to compensate for different sensor responses. Moreover, our method, similar to [16], estimates event camera poses via interpolation of available ones. This could lead to a performance decrease in case estimated poses are far from actual ones or they are provided at a low frequency. However, we believe refinement of camera poses through event-based methods [26, 44], or a modified approach that only computes $\mathcal{L}_{event}$ at known camera views, could help in mitigating this issue.

5. Conclusions

We present Ev-DeblurNeRF, a novel deblur NeRF architecture that fuses blurry frames with events for sharp NeRF recovery. Our method, exploiting explicit features for fast training convergence, integrates a learnable event-based camera response function and ad-hoc event-based supervision that facilitates fine-grained details recovery. Ev-DeblurNeRF, despite being supervised by model-based priors, can adapt to non-idealities in the camera response, potentially departing from the model-based solution. We validate our method on both synthetic and real data, achieving an increase of $+4.42$dB and $+2.48$dB in PSNR, respectively, when compared to the previous best-performing event-based baseline, and an increase of $+2.74$dB and $+6.13$dB when compared to the top-performing image-only baseline.

6. Acknowledgements.

This work was supported by the National Centre of Competence in Research (NCCR) Robotics (grant agreement No. 51NF40-185543) through the Swiss National Science Foundation (SNSF), and the European Research Council (ERC) under grant agreement No. 864042 (AGILEFLIGHT).

References

[1] Dejan Azinovic, Ricardo Martin-Brualla, Dan B Goldman, Matthias Niebner, and Justus Thies. Neural RGB-d surface reconstruction. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6290–6301. IEEE, 2022. 1

[2] Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. Mip-NeRF: A multiscale representation for anti-aliasing neural radiance fields. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5855–5864. IEEE, 2021. 1

[3] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-NeRF 360: Unbounded anti-aliased neural radiance fields. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5470–5479. IEEE, 2022. 1

[4] Anish Bhattacharya, Ratnesh Madaan, Fernando Cladera, Sai Vemprala, Rogerio Bonatti, Kostas Daniilidis, Ashish Kapoor, Vijay Kumar, Nikolai Matni, and Jayesh K Gupta. Evdnerf: Reconstructing event data with dynamic neural radiance fields. *arXiv preprint arXiv:2310.02437*, 2023. 2

[5] Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. TensoRF: Tensorial radiance fields. In *Lecture Notes in Computer Science*, pages 333–350. Springer Nature Switzerland, 2022. 5, 6

[6] Peng Cheng and Chellappa Rama. Pdrf: Progressively deblurring radiance field for fast and robust scene reconstruction from blurry images, 2023. 1, 2, 3, 5, 6

[7] Guillermo Gallego, Tobi Delbruck, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew J. Davison, Jorg Conradt, Kostas Daniilidis, and Davide Scaramuzza. Event-based vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(1):154–180, 2022. 2, 3

[8] Kyle Gao, Yina Gao, Hongjie He, Denning Lu, Linlin Xu, and Jonathan Li. Nerf: Neural radiance field in 3d vision, a comprehensive review. *arXiv preprint arXiv:2210.00379*, 2022. 2, 3

[9] Jin Han, Chu Zhou, Peiqi Duan, Yehui Tang, Chang Xu, Chao Xu, Tiejun Huang, and Boxin Shi. Neuromorphic camera guided high dynamic range imaging. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1730–1739. IEEE, 2020. 2

[10] Chen Haoyu, Teng Minggui, Shi Boxin, Wang YIzhou, and Huang Tiejun. Learning to deblur and generate high frame rate video with an event camera. *arXiv preprint arXiv:2003.00847*, 2020. 2

[11] Yuhuang Hu, Shih-Chii Liu, and Tobi Delbruck. v2e: From video frames to realistic DVS events. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1312–1321. IEEE, 2021. 1, 5

[12] Xin Huang, Qi Zhang, Ying Feng, Hongdong Li, Xuan Wang, and Qing Wang. HDR-NeRF: High dynamic range neural radiance fields. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 18398–18408. IEEE, 2022. 1

[13] Inwoo Hwang, Junho Kim, and Young Min Kim. Ev-NeRF: Event based neural radiance field. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 837–847. IEEE, 2023. 2, 3

[14] Zhe Jiang, Yu Zhang, Dongqing Zou, Jimmy Ren, Jiancheng Lv, and Yebin Liu. Learning event-based motion deblurring. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3320–3329. IEEE, 2020. 2

[15] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5

[16] Simon Klenk, Lukas Koestler, Davide Scaramuzza, and Daniel Cremers. E-NeRF: Neural radiance fields from a moving event camera. *IEEE Robotics and Automation Letters*, 8(3):1587–1594, 2023. 2, 3, 5, 6, 8

[17] Abhijit Kundu, Kyle Genova, Xiaoqi Yin, Alireza Fathi, Caroline Pantofaru, Leonidas Guibas, Andrea Tagliasacchi, Frank Dellaert, and Thomas Funkhouser. Panoptic neural fields: A semantic object-aware neural scene representation. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12871–12881. IEEE, 2022. 1

[18] Dogyoon Lee, Minhyeok Lee, Chajin Shin, and Sangyoun Lee. DP-NeRF: Deblurred neural radiance field with physical scene priors. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12386–12396. IEEE, 2023. 1, 2, 3, 4, 5, 6, 7

[19] Chenghan Li, Christian Brandli, Raphael Berner, Hongjie Liu, Minhao Yang, Shih-Chii Liu, and Tobi Delbruck. Design of an RGBW color VGA rolling and global shutter dynamic and active-pixel vision sensor. In *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 718–721. IEEE, IEEE, 2015. 2, 5, 6

[20] Zhengqi Li, Simon Niklaus, Noah Snavely, and Oliver Wang. Neural scene flow fields for space-time view synthesis of dynamic scenes. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6498–6508. IEEE, 2021. 5

[21] Chen-Hsuan Lin, Wei-Chiu Ma, Antonio Torralba, and Simon Lucey. BARF: Bundle-adjusting neural radiance fields. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5741–5751. IEEE, 2021. 2

[22] Zhizheng Liu, Francesco Milano, Jonas Frey, Roland Siegwart, Hermann Blum, and Cesar Cadena. Unsupervised continual semantic adaptation through neural rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3031–3040, 2023. 1

[23] Weng Fei Low and Gim Hee Lee. Robust e-nerf: Nerf from sparse & noisy events under non-uniform motion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 18335–18346, 2023. 2

[24] Li Ma, Xiaoyu Li, Jing Liao, Qi Zhang, Xuan Wang, Jue Wang, and Pedro V Sander. Deblur-NeRF: Neural radiance fields from blurry images. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12861–12870. IEEE, 2022. 1, 2, 3, 5, 6

[25] Nico Messikommer, Stamatios Georgoulis, Daniel Gehrig, Stepan Tulyakov, Julius Erbach, Alfredo Bochicchio,

Yuanyou Li, and Davide Scaramuzza. Multi-bracket high dynamic range imaging with event cameras. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 547–557. IEEE, 2022. 2, 8

[26] Nico Messikommer, Carter Fang, Mathias Gehrig, and Davide Scaramuzza. Data-driven feature tracking for event cameras. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5642–5651. IEEE, 2023. 8

[27] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020. 1, 2, 4, 5, 6

[28] Ben Mildenhall, Peter Hedman, Ricardo Martin-Brualla, Pratul P Srinivasan, and Jonathan T Barron. NeRF in the dark: High dynamic range view synthesis from noisy raw images. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16190–16199. IEEE, 2022. 1

[29] Liyuan Pan, Cedric Scheerlinck, Xin Yu, Richard Hartley, Miaomiao Liu, and Yuchao Dai. Bringing a blurry frame alive at high frame-rate with an event camera. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6820–6829. IEEE, 2019. 2, 3, 6

[30] Yunshan Qi, Lin Zhu, Yu Zhang, and Jia Li. E2nerf: Event enhanced neural radiance fields from blurry images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13254–13264, 2023. 2, 6, 7

[31] Henri Rebecq, Daniel Gehrig, and Davide Scaramuzza. Esim: an open event camera simulator. In *2nd Annual Conference on Robot Learning, CoRL 2018, Zürich, Switzerland, 29-31 October 2018, Proceedings*, pages 969–982. PMLR, 2018. 5

[32] Viktor Rudnev, Mohamed Elgharib, Christian Theobalt, and Vladislav Golyanik. Eventnerf: Neural radiance fields from a single colour event camera. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2023. 2, 3

[33] Wei Shang, Dongwei Ren, Dongqing Zou, Jimmy S Ren, Ping Luo, and Wangmeng Zuo. Bringing events into video deblurring with non-consecutively blurry frames. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4531–4540. IEEE, 2021. 1

[34] Ken Shoemake. Animating rotation with quaternion curves. *ACM SIGGRAPH Computer Graphics*, 19(3):245–254, 1985. 4

[35] Hyeongseok Son, Junyong Lee, Jonghyeop Lee, Sunghyun Cho, and Seungyong Lee. Recurrent video deblurring with blur-invariant motion estimation and pixel volumes. *ACM Transactions on Graphics*, 40(5):1–18, 2021. 6

[36] Edgar Sucar, Shikun Liu, Joseph Ortiz, and Andrew J. Davison. *iMAP: Implicit Mapping and Positioning in Real-Time*, pages 6229–6238. IEEE, 2021. 1

[37] Lei Sun, Christos Sakaridis, Jingyun Liang, Qi Jiang, Kailun Yang, Peng Sun, Yaozu Ye, Kaiwei Wang, and Luc Van Gool. Event-based fusion for motion deblurring with cross-modal attention. In *Lecture Notes in Computer Science*, pages 412–428. Springer, Springer Nature Switzerland, 2022. 1, 2, 6

[38] Lei Sun, Christos Sakaridis, Jingyun Liang, Peng Sun, Jiezhang Cao, Kai Zhang, Qi Jiang, Kaiwei Wang, and Luc Van Gool. Event-based frame interpolation with ad-hoc deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18043–18052, 2023. 2

[39] Ayush Tewari, Justus Thies, Ben Mildenhall, Pratul Srinivasan, Edgar Tretschk, Wang Yifan, Christoph Lassner, Vincent Sitzmann, Ricardo Martin-Brualla, Stephen Lombardi, et al. Advances in neural rendering. In *Computer Graphics Forum*, pages 703–735. Wiley Online Library, 2022. 2

[40] Stepan Tulyakov, Daniel Gehrig, Stamatios Georgoulis, Julius Erbach, Mathias Gehrig, Yuanyou Li, and Davide Scaramuzza. Time lens: Event-based video frame interpolation. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16155–16164. IEEE, 2021. 2

[41] Stepan Tulyakov, Alfredo Bochicchio, Daniel Gehrig, Stamatios Georgoulis, Yuanyou Li, and Davide Scaramuzza. Time lens++: Event-based frame interpolation with parametric nonlinear flow and multi-scale fusion. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17755–17764. IEEE, 2022. 2, 8

[42] International Telecomunication Union. Studio encoding parameters of digital television for standard 4: 3 and widescreen 16: 9 aspect ratios. bt.709. *Technical Report, International Telecomunication Union*, 2011. 4

[43] Dor Verbin, Peter Hedman, Ben Mildenhall, Todd Zickler, Jonathan T Barron, and Pratul P Srinivasan. Ref-NeRF: Structured view-dependent appearance for neural radiance fields. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5481–5490. IEEE, IEEE, 2022. 1

[44] Antoni Rosinol Vidal, Henri Rebecq, Timo Horstschaefer, and Davide Scaramuzza. Ultimate slam? combining events, images, and imu for robust visual slam in hdr and high-speed scenarios. *IEEE Robotics and Automation Letters*, 3(2):994–1001, 2018. 8

[45] Patricia Vitoria, Stamatios Georgoulis, Stepan Tulyakov, Alfredo Bochicchio, Julius Erbach, and Yuanyou Li. Event-based image deblurring with dynamic motion awareness. In *Eur. Conf. Comput. Vis.* Springer Nature Switzerland, 2022. 1

[46] Bishan Wang, Jingwei He, Lei Yu, Gui-Song Xia, and Wen Yang. Event enhanced high-quality image recovery. In *Computer Vision – ECCV 2020*, pages 155–171. Springer International Publishing, 2020. 2

[47] Chen Wang, Xian Wu, Yuan-Chen Guo, Song-Hai Zhang, Yu-Wing Tai, and Shi-Min Hu. NeRF-SR: High quality neural radiance fields using supersampling. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 6445–6454. ACM, 2022. 1

[48] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *NeurIPS*, 2021. 1

[49] Peng Wang, Lingzhe Zhao, Ruijie Ma, and Peidong Liu. Bad-nerf: Bundle adjusted deblur neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4170–4179, 2023. 1, 2, 6

[50] Song Wu, Kaichao You, Weihua He, Chen Yang, Yang Tian, Yaoyuan Wang, Ziyang Zhang, and Jianxing Liao. Video interpolation by event-driven anisotropic adjustment of optical flow. In *Lecture Notes in Computer Science*, pages 267–283. Springer Nature Switzerland, 2022. 2

[51] Christopher Xie, Keunhong Park, Ricardo Martin-Brualla, and Matthew Brown. FiG-NeRF: Figure-ground neural radiance fields for 3d object category modelling. In *2021 International Conference on 3D Vision (3DV)*, pages 962–971. IEEE, IEEE, 2021. 1

[52] Fang Xu, Lei Yu, Bishan Wang, Wen Yang, Gui-Song Xia, Xu Jia, Zhendong Qiao, and Jianzhuang Liu. Motion deblurring with real events. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2583–2592. IEEE, 2021. 2

[53] Lin Yen-Chen, Pete Florence, Jonathan T Barron, Alberto Rodriguez, Phillip Isola, and Tsung-Yi Lin. iNeRF: Inverting neural radiance fields for pose estimation. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1323–1330. IEEE, IEEE, 2021. 1

[54] Zehao Yu, Songyou Peng, Michael Niemeyer, Torsten Sattler, and Andreas Geiger. Monosdf: Exploring monocular geometric cues for neural implicit surface reconstruction. *Advances in Neural Information Processing Systems (NeurIPS)*, 2022. 1

[55] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14821–14831. IEEE, 2021. 6

[56] Hongguang Zhang, Limeng Zhang, Yuchao Dai, Hongdong Li, and Piotr Koniusz. Event-guided multi-patch network with self-supervision for non-uniform motion deblurring. *International Journal of Computer Vision*, 131(2):453–470, 2022. 1

[57] Limeng Zhang, Hongguang Zhang, Jihua Chen, and Lei Wang. Hybrid deblur net: Deep non-uniform deblurring with event camera. *IEEE Access*, 8:148075–148083, 2020. 2

[58] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 586–595. IEEE, 2018. 6

[59] Zihan Zhu, Songyou Peng, Viktor Larsson, Weiwei Xu, Hujun Bao, Zhaopeng Cui, Martin R. Oswald, and Marc Pollefeys. NICE-SLAM: Neural implicit scalable encoding for SLAM. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12–786. IEEE, 2022. 1