# Reconstructing Dense Depth of Dark Scenes with Sparse LiDAR, Noisy Events, and Blurry RGB

Jianbo Cao[1,4], Yuqi Han[2], Siming Zheng[3], Bo Wang[1], Tong Guo[1], and Jinli Suo[1⋆]

[1] Department of Automation, Tsinghua University, Beijing, China
{cjb23, bo-wang23, gt23}@mails.tsinghua.edu.cn, jlsuo@tsinghua.edu.cn
[2] Key Laboratory of Social Computing and Cognitive Intelligence, School of Computer Science and Technology, Dalian University of Technology, Dalian, China
yqhanscst@dlut.edu.cn
[3] Vivo Mobile Communication Co., Ltd., Hangzhou, China
zhengsiming@vivo.com
[4] School of Computer Science, Peking University, Beijing, China

**Abstract.** Due to the sparsity of LiDAR measurements, RGB-assisted dense depth reconstruction is widely adopted to provide structural priors for autonomous driving. However, the inherent sensitivity of RGB imaging to illumination conditions makes dense depth reconstruction under low-light scenarios still a critical challenge. Specifically, under long-exposure imaging, motion blur in low-light RGB frames significantly degrades the accuracy of depth reconstruction. To address this issue, we exploit the high-temporal-resolution motion cues captured by event cameras and propose Event-guided Restoration and Upsampling Network (ERU-Net), a unified framework that tightly couples event-guided feature restoration with depth completion. The Event-guided Feature Restoration (EFR) module combines implicit neural representation (INR) and self-recursive optimization to remove motion blur from long-exposure inputs and recover artifact-free features. These features serve as structural priors for the Restoration-Aware Depth Upsampling (RDU) module, enabling accurate completion of sparse LiDAR measurements with fine geometric details. Extensive experiments on challenging synthetic and real-world captured datasets demonstrate that ERU-Net significantly outperforms state-of-the-art approaches, producing accurate and temporally consistent dense depth sequences in severe low-light conditions. Our code will be available at https://github.com/qzm777/Night-RIDE-ECCV2026.

**Keywords:** Depth Completion · Event Camera · Low-light Vision · Autonomous Driving

## 1 Introduction

Reliable depth perception is crucial for safe autonomous driving, as accurate geometric information underpins downstream tasks such as obstacle avoidance

⋆ Yuqi Han and Jinli Suo are co-corresponding authors.

and path planning [8]. LiDAR sensors provide highly precise geometric measurements, making them an indispensable source for depth estimation; however, their inherently sparse sampling limits the completeness of the reconstructed scene [40]. To overcome this limitation, a common strategy is to fuse RGB imagery with sparse LiDAR data: the RGB image, with its rich structural and textural information, serves as guidance that enables the dense reconstruction of the environment [20, 35, 43].

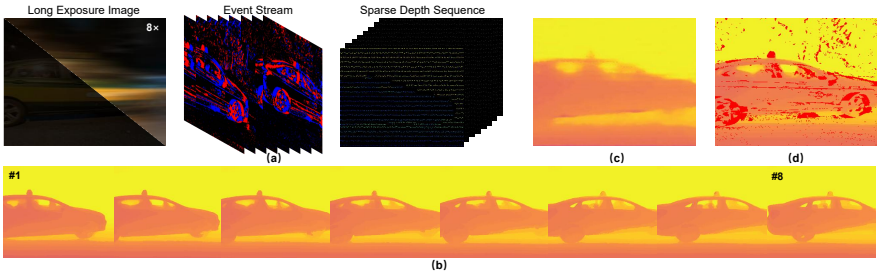


**Fig. 1:** Overview of ERU-Net performance. (a) Input. (b) Sequential outputs of ERU-Net. (c) Reconstruction from sparse LiDAR with blurry RGB input. (d) ERU-Net output with overlaid events illustrating motion estimation accuracy.

This dense depth reconstruction becomes even more challenging under low-illumination conditions, where conventional RGB sensing is severely degraded [15, 16]. In such environments, RGB cameras are constrained to prolonged exposure to accumulate sufficient light, which inevitably induces severe motion blur from either the ego-vehicle or surrounding dynamic objects [14, 25]. This blur is more than a simple visual artifact: it manifests as a complex, non-uniform distortion of structural information, significantly impairing both geometric and semantic perception tasks [4, 52].

By capturing asynchronous changes in pixel intensity with microsecond temporal resolution, event cameras can faithfully record fast motion and low-light dynamics, offering unique advantages in dynamic scenes that are particularly challenging for conventional cameras [6, 18, 26]. The high-frequency motion information embedded in event streams can be leveraged to decouple temporal variations of true depth, enabling the extraction of 3D motion cues and the alignment of scene geometry across different time instances [49], making event cameras powerful for dense depth reconstruction in low-light scenarios.

In this paper, we propose the Event-guided Restoration and Upsampling Network (ERU-Net) for dense depth reconstruction under low-illumination conditions, where long-exposure RGB images provide unreliable motion cues. ERU-Net takes sparse LiDAR measurements, event signals, and long-exposure RGB images as inputs (Fig. 1(a)) and reconstructs sequential dense depth maps that capture the underlying scene geometry (Fig. 1(b)). Specifically, sparse LiDAR provides reliable but spatially sparse depth measurements, serving as the geo-

metric foundation for reconstruction. Event streams capture high-frequency motion information in dynamic scenes, enabling accurate estimation of motion cues even under severe motion blur. Meanwhile, long-exposure RGB images preserve the rich semantic and structural context of the scene, facilitating cross-modal alignment, allowing complementary information from LiDAR and events to be coherently integrated for depth reconstruction.

To this end, we design the Event-Guided Feature Restoration (EFR) module, which integrates an implicit neural representation (INR) with a self-recursive refinement strategy, to recover structured latent representations from blurred long-exposure images captured under extremely low-light conditions. By leveraging event-guided temporal cues, EFR reconstructs geometrically consistent latent features that provide reliable structural priors for subsequent depth inference. Building upon these restored representations, the Restoration-Aware Depth Upsampling (RDU) module introduces a hybrid-guidance confidence module (HGCM) that jointly exploits the restored visual features and dynamic event signals to guide the stable densification of sparse LiDAR measurements.

ERU-Net is trained using a principled multi-stage strategy that culminates in a joint fine-tuning phase. This synergistic training process enables the RDU's depth completion objective to further refine the high-quality features produced by the EFR module, yielding a task-specific representation that is optimally aligned with the goal of accurate depth estimation. Consequently, our framework achieves state-of-the-art performance on the dense depth completion benchmark.

Our main contributions are summarized as:

- We address the challenging task of dense depth reconstruction in low-light scenarios and present a novel event-guided multimodal framework that effectively integrates sparse LiDAR, asynchronous event signals, and motion-blurred RGB imagery.
- By restoring reliable visual representations from motion-degraded images and integrating them with event-derived motion cues, the ERU-Net effectively bridges the complementary characteristics of different modalities, enabling accurate and robust densification of sparse depth observations.
- We build and will release a new, large-scale dataset for low-light 3D perception, featuring synchronized blurred RGB, events, and LiDAR in real-world nighttime scenarios. We demonstrate that ERU-Net achieves state-of-the-art performance on this dataset and synthetic benchmarks.

## 2 Related Work

### 2.1 RGB-Guided Depth Completion

Sparse-to-dense depth completion aims to generate a dense depth map from sparse measurements (e.g., LiDAR) by leveraging a co-registered guidance image (e.g., RGB) [20]. The prevailing paradigm adopts deep encoder–decoder networks (ED-Net) to fuse these multi-modal inputs [12, 24]. However, naive fusion often leads to boundary artifacts and conflicts between heterogeneous features.

To address this, research focuses on enhancing feature propagation, employing methods such as Spatial Propagation Networks (SPN) [2,23] to refine depth predictions through learned correlation, thereby sharpening object boundaries. The dynamic or deformable convolutions [46,47] are introduced to achieve alignment between modalities, particularly in regions with sparse geometry. Other approaches leverage multi-branch or multi-scale architectures to enable more effective feature fusion [5,19,29,50].

Research focusing on fusion architectures has observed that standard ED-Nets often create information bottlenecks. Wang et al. [43] proposed the Depth Feature Upsampling (DFU) network, which identifies a dense-to-sparse bottleneck and introduces a dedicated upsampling stream guided by internal dense features [35,42]. This line of work also encompasses methods such as LRRU [41], Guideformer [27], and CompletionFormer [51].

However, all of these methods, from SPNs [2] to DFU [43], fundamentally assume the availability of a high-quality, sharp, and structurally reliable RGB guidance image—a condition that is completely violated in low-light or dynamic scenarios. Long-exposure sensing in low-light environments is severely affected by motion blur, which introduces misleading structural cues and results in propagated artifacts and distorted geometry.

### 2.2 Event-Guided Image and Video Restoration

As the neuromorphic sensors, Event camera [6] and spike camera [1,48] asynchronously capture pixel-level brightness changes with microsecond-level temporal resolution, enabling robust recording of fast motions under extreme low-light and high-dynamic-range (HDR) conditions where RGB imagery degrade [26,28].

Based on the high temporal sampling rate, a prominent direction is event-guided video deblurring, where blurry video sequences are sharpened using co-registered events [30,33], effectively restoring high-frequency details. These methods assume a blurry video as input, which differs fundamentally from our task of recovering from a single long-exposure frame. Another line targets event-based video frame interpolation (VFI) [3,22], leveraging events to synthesize intermediate frames [7,10,36,37].

A more relevant line of work tackles reconstruction from a single long-exposure blurry image. Pan et al. [21] pioneered this by demonstrating how to "bring a blurry frame alive" at high frame rates by optimizing a latent sharp video. Subsequent works further improved reconstruction fidelity using different network architectures. For example, some methods employed deep generative models to enhance perceptual quality and handled real-world artifacts [53,54], while others explored different event representations or transformer-based backbones to better model long-range temporal dependencies [44]. Others combined deblurring and interpolation [17,32,45] or focused on sparse to dense reconstruction [39].

The above event-guided motion reconstruction methods [21,44,53] are typically evaluated under moderate motion or sufficient illumination. The performance degrades substantially [9,11,49] under extreme motion blur and low photon counts, as encountered in nighttime driving. Event-guided methods offer

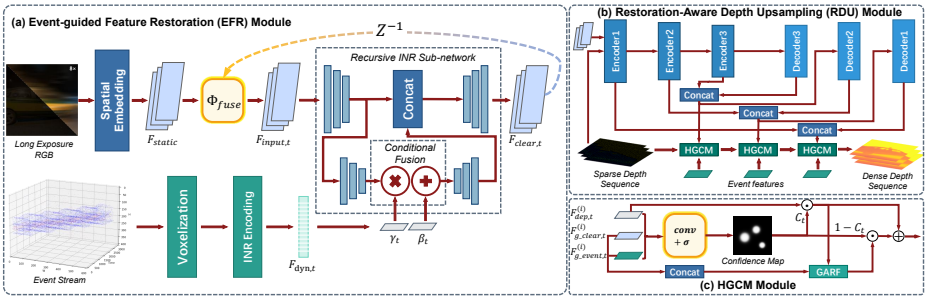


**Fig. 2:** The whole framework of ERU-Net consists of two main components: **(a) The Event-Guided Feature Restoration (EFR) module** implicitly embeds event features into the spatial features via **Conditional Fusion**, integrated with a self-recursive model to generate the latent feature. **(b) The Restoration-Aware Depth Upsampling (RDU) module** introduces the **(c) Hybrid Guidance Compensation Module (HGCM)**, which leverages the restored latent feature, sparse point clouds, and multi-scale features as guidance to reconstruct the dense depth progressively **via a confidence-aware routing mechanism (GARF)**.

strong potential for enhancing imagery in dynamic scenes; however, effectively leveraging the temporal correlations of event streams to guide long-exposure image decoding remains challenging.

## 3 Methods

### 3.1 Overview

We propose ERU-Net (Event-guided Restoration and Upsampling Network) to reconstruct a temporally-consistent, dense depth sequence. Given the single long-exposure blurred RGB image $I_{\text{blur}} \in \mathbb{R}^{H\times W\times 3}$, a corresponding event stream $\mathcal{E}$, and a sparse LiDAR point cloud stream $\{D_{\text{sparse},t}\}_{t=1}^{T}$, the goal of ERU-Net is to reconstruct a dense depth map sequence $\{D_{\text{dense},t}\}_{t=1}^{T}$ over consecutive $T$ time step. ERU-Net consists of two deeply coupled modules: (1) an **Event-Guided Feature Restoration (EFR)** module, which leverages an implicit and self-recursive architecture to reconstruct high-fidelity latent sequencial features $\{F_{\text{clear},t}\}_{t=1}^{T}$; and (2) a **Restoration-Aware Depth Upsampling (RDU)** module, which employs a hybrid confidence-aware mechanism, integrating $\{F_{\text{clear},t}\}$ and $\mathcal{E}$ for dense depth maps $\{D_{\text{sparse},t}\}$ over $t$ time steps.

### 3.2 EFR Module

The EFR module is designed to solve the guidance degradation challenge because the long-exposure blurred RGB generates fallacious structural cues. The EFR module restores a geometrically-accurate latent feature from $I_{\text{blur}}$ and $\mathcal{E}$, creating a reliable guide for the subsequent RDU module.

We first partition the raw event stream $\mathcal{E}$, corresponding to the exposure period of $I_{\text{blur}}$, into $T$ temporal bins, each converted into an event voxel grid $\{V_t\}_{t=1}^{T}$. The self-recursive INR in EFR first employs a spatial encoder $\Phi_{\text{spatial}}$ to extract a static, multi-scale feature pyramid $F_{\text{static}}$ from the blurred image $I_{\text{blur}}$, providing rich yet degraded contextual information of the scene. Concurrently, a temporal encoder $\Phi_{\text{temporal}}$ (corresponding to the "Voxelization and INR Encoding" modules in Fig. 2) processes the event voxel sequence $\{V_t\}_{t=1}^{T}$ to extract a sequence of dynamic motion features $\{F_{\text{dyn},t}\}_{t=1}^{T}$, mapped from the high-dimensional, sparse spatiotemporal dimension into a compact, continuous latent space. The feature vector $F_{\text{dyn},t}$ serves as a learned implicit embedding for the subsequent conditional fusion. In the EFR module, the INR employs a compact implicit representation to model complex spatial motions, since scene motion is generally non-linear and continuously modeled by scale and shift vectors.

The recursive INR sub-network in EFR, denoted as $\Psi_{\text{inr}}$, synthesizes the latent feature sequence $F_{\text{clear}} = \{F_{\text{clear},t}\}_{t=1}^{T}$ in an auto-regressive manner. The $\Psi_{\text{inr}}$ module is implemented as a full encoder–decoder subnetwork that extracts multi-scale representations and propagates high-frequency spatial details through skip connections. Specifically, it consists of two encoding stages $(\Phi_1, \Phi_2)$ and two symmetric decoding stages $(\Psi_1, \Psi_2)$. For each time step $t$, the input features $F_{\text{input},t}$ are first processed through the encoding stages to obtain hierarchical representations, i.e.,

$$E_t^{(1)} = \Phi_1(F_{\text{input},t}); \quad E_t^{(2)} = \Phi_2(E_t^{(1)}). \tag{1}$$

The dynamic feature $F_{\text{dyn},t}$ is further processed to generate $\gamma_t$ and $\beta_t$. Acting as the scale and bias vectors, $\gamma_t$ and $\beta_t$ are integrated into the encoded spatial feature via the conditional fusion, denoted as

$$F_{\text{st},t} = \gamma_t \odot E_t^{(2)} + \beta_t, \tag{2}$$

where $\odot$ denotes element-wise multiplication. Scale and shift vectors enable adaptive modulation of $F_{\text{st},t}$, allowing motion-aware feature transformation beyond rigid displacement.

The spatiotemporal feature $F_{\text{st},t}$ is then upsampled by the decoder stages, which integrate the skip connections from the encoder, denoted as

$$D_t^{(1)} = \Psi_1(F_{\text{st},t}), \quad D_t^{(2)} = \Psi_2(\text{Concat}(D_t^{(1)}, E_t^{(1)})), \tag{3}$$

where the final latent feature is noted as $F_{\text{clear},t} = D_t^{(2)}$.

The model generates features at each time step in a self-recursive manner, producing the dense depth sequence auto-regressively. For the first frame ($t = 1$), $F_{\text{input},1} = F_{\text{static}}$. For subsequent frames ($t > 1$), the output feature $F_{\text{clear},t-1}$ is fed back and fused with $F_{\text{static}}$ (via a fusion block $\Phi_{\text{fuse}}$) to provide a strong temporal prior, denoted as

$$F_{\text{input},t} = \Phi_{\text{fuse}}(F_{\text{static}}, F_{\text{clear},t-1}). \tag{4}$$

### 3.3 Restoration-Aware Depth Upsampling (RDU)

The RDU module performs depth completion by leveraging the restored latent features $\{F_{\text{clear},t}\}$ from the EFR module. The guidance encoder processes $F_{\text{clear},t}$ to produce a pyramid of multi-scale static guidance features $\{F^{(i)}_{\text{guide\_clear},t}\}$, where $i$ indexes the scale level. In parallel, a multi-scale spatial event encoder processes the event voxels $\{V_t\}$ to generate a pyramid of dynamic motion features $\{F^{(i)}_{\text{guide\_event},t}\}$.The depth upsampling stream takes the sparse LiDAR input $D_{\text{sparse},t}$ and encodes it into a low-resolution feature. This feature is then progressively upsampled through a depth decoder. At each upsampling stage, the multi-scale features extracted from the input are regarded as guidance to ensure spatial details and temporal consistency.

The hybrid-guidance confidence module (HGCM) is applied at each upsampling stage $i$ of the depth decoder. Given the upsampled depth feature $F^{(i)}_{\text{dep},t}$, the restored static feature $F^{(i)}_{\text{guide\_clear},t}$, and the dynamic event feature $F^{(i)}_{\text{guide\_event},t}$ at stage $i$, a hybrid confidence map $C^{(i)}_t$ is obtained by simultaneously integrating the features, denoted as

$$C^{(i)}_t = \sigma(\text{Conv}_{3\times3}(\text{Concat}(F^{(i)}_{\text{dep},t}, F^{(i)}_{\text{guide\_clear},t}, F^{(i)}_{\text{guide\_event},t}))) \tag{5}$$

where $\sigma(\cdot)$ is the sigmoid function.

This map assigns low confidence values (close to 0) to regions in $F^{(i)}_{\text{dep},t}$ that are unreliable or sparsely populated. Subsequently, the Guidance with Adaptive Receptive Fields (GARF) [43] leverages a hybrid guidance feature (formed by concatenating $F^{(i)}_{\text{guide\_clear},t}$ and $F^{(i)}_{\text{guide\_event},t}$) to enhance the depth feature, focusing on the low-confidence regions, i.e.,

$$F^{(i)}_{\text{ref},t} = \text{GARF}(C^{(i)}_t \odot F^{(i)}_{\text{dep},t}, \text{Concat}(F^{(i)}_{\text{guide\_clear},t}, F^{(i)}_{\text{guide\_event},t})). \tag{6}$$

GARF is a widely-used dynamic filter that uses larger receptive fields for sparse global context and smaller ones locally to preserve edges and suppress artifacts.

Finally, the RDU updates the depth feature using the confidence map as a gate, denoted as

$$F^{(i+1)}_{\text{dep},t} = (C^{(i)}_t \odot F^{(i)}_{\text{dep},t}) + (1 - C^{(i)}_t) \odot F^{(i)}_{\text{ref},t}. \tag{7}$$

This process enables the network to rely on existing depth features in high-confidence regions, while leveraging the hybrid guidance—derived from both restored static and dynamic event features—to fill gaps in low-confidence (sparse or unreliable) areas. Finally, we derive the high-resolution dense depth $D_{\text{dense},t}$.

### 3.4 Multi-Stage Training Strategy

We adopt a three-stage training strategy that progressively optimizes event-guided restoration and geometry-aware depth reconstruction. Direct end-to-end

training leads to unstable convergence. Pre-training first establishes reliable restoration and geometric propagation, while the final joint fine-tuning aligns restored features with depth-oriented geometric representations under depth supervision.

**Stage 1: EFR Module Pre-training.** We first pre-train the EFR module as an independent video restoration network. Given sharp image sequence, the module is trained on a synthetic dataset using a composite restoration loss $\mathcal{L}_{\text{EFR}}$, which consists of three terms: a main reconstruction loss $\mathcal{L}_{\text{main}}$, a re-blur consistency loss $\mathcal{L}_{\text{reblur}}$, and an event-gradient regularization loss $\mathcal{L}_{\text{grad}}$. The main loss $\mathcal{L}_{\text{main}}$ ensures pixel-level fidelity and structural consistency, defined as

$$\mathcal{L}_{\text{main}} = \lambda_{\text{char}}\mathcal{L}_{\text{char}} + \lambda_{\text{ssim}}\mathcal{L}_{\text{SSIM}} + \lambda_{\text{edge}}\mathcal{L}_{\text{edge}}, \tag{8}$$

where $\mathcal{L}_{\text{char}}$, $\mathcal{L}_{\text{SSIM}}$, $\mathcal{L}_{\text{edge}}$ represent the Charbonnier loss, structural similarity loss, and edge-preserving loss, respectively. All loss terms are computed between the reconstructed sequence $\{I_{\text{recon},t}\}_{t=1}^{T}$ and the corresponding ground-truth sharp sequence $\{I_{\text{gt},t}\}_{t=1}^{T}$.

Next, we define the re-blur loss $\mathcal{L}_{\text{reblur}}$ to enforce consistency between the original blurry RGB $I_{\text{blur}}$ and the re-synthesized blurred version, i.e.,

$$\mathcal{L}_{\text{reblur}} = \|\text{Avg}(I_{\text{recon},t}) - I_{\text{blur}}\|_1, \tag{9}$$

and the event-gradient loss $\mathcal{L}_{\text{grad}}$ is noted as

$$\mathcal{L}_{\text{grad}} = \sum_{t=1}^{T} \|\nabla I_{\text{recon},t}^{\text{gray}} - \nabla I_{\text{event},t}^{\text{gray}}\|_1, \tag{10}$$

which enforces the spatial gradients of the reconstructed grayscale images to align with the motion cues encoded in the event stream, preserving consistent dynamics. The total EFR loss is denoted as

$$\mathcal{L}_{\text{EFR}} = \mathcal{L}_{\text{main}} + \lambda_{\text{reblur}}\mathcal{L}_{\text{reblur}} + \lambda_{\text{grad}}\mathcal{L}_{\text{grad}}. \tag{11}$$

**Stage 2: RDU Module Pre-training.** We freeze the pre-trained EFR and train the RDU module in stage 2, $I_{\text{blur}}$ and $\mathcal{E}$ are passed through EFR to obtain guidance features $\{F_{\text{clear},t}\}_{t=1}^{T}$, which the RDU fuses with sparse depth $\{D_{\text{sparse},t}\}_{t=1}^{T}$ to predict dense depth $\{D_{\text{dense},t}\}_{t=1}^{T}$. The RDU loss $\mathcal{L}_{\text{RDU}}$ emphasizes supervision at reliable sparse LiDAR locations, defined by a binary mask $M_t$ with $M_t = 1$ at valid sparse points and 0 elsewhere. The RDU loss is defined as a composite of dense and sparse supervision

$$\mathcal{L}_{\text{RDU}} = \mathcal{L}_{\text{dense\_depth}} + \lambda_{\text{sparse}} \left(\mathcal{L}_{\text{sparse\_depth}} + \mathcal{L}_{\text{sparse\_grad}}\right). \tag{12}$$

In this equation, the dense depth loss $\mathcal{L}_{\text{dense\_depth}}$ measures L1 error over the full dense ground-truth depth $D_{\text{gt},t}$, i.e.,

$$\mathcal{L}_{\text{dense\_depth}} = \sum_{t=1}^{T} \|D_{\text{dense},t} - D_{\text{gt},t}\|_1; \tag{13}$$

the sparse depth loss $\mathcal{L}_{\text{sparse\_depth}}$ is computed only at valid sparse locations indicated by the mask $M_t$

$$\mathcal{L}_{\text{sparse\_depth}} = \sum_{t=1}^{T} \|M_t \odot (D_{\text{dense},t} - D_{\text{gt},t})\|_1; \tag{14}$$

the sparse gradient loss $\mathcal{L}_{\text{sparse\_grad}}$ enforces gradient consistency at sparse locations

$$\mathcal{L}_{\text{sparse\_grad}} = \sum_{t=1}^{T} \|M_t \odot (\nabla D_{\text{dense},t} - \nabla D_{\text{gt},t})\|_1. \tag{15}$$

The formulation ensures accurate LiDAR depth while propagating geometric information to unknown regions.

**Stage 3: End-to-End Joint Fine-tuning.** In the final and most critical stage, all network weights (EFR and RDU) are unfrozen, and the auxiliary restoration loss $\mathcal{L}_{\text{EFR}}$ is removed. The entire ERU-Net is then fine-tuned end-to-end using only the depth loss $\mathcal{L}_{\text{RDU}}$ with a reduced learning rate. During this stage, gradients from $\mathcal{L}_{\text{RDU}}$ propagate through the RDU into the EFR module, forcing the INR and recursive weights in EFR to adapt. As a result, $F_{\text{clear},t}$ is refined to be not only visually accurate but also optimally suited for the downstream geometric depth completion task.

### 3.5 Implementation Details

The proposed ERU-Net is implemented in PyTorch and trained using the Adam optimizer on a workstation with an AMD EPYC 7H12 CPU and an NVIDIA RTX 3090 GPU, employing a data loader with 2 parallel workers. Training follows a cosine annealing schedule, with the EFR module's learning rate decaying from $5 \times 10^{-4}$ to $1 \times 10^{-6}$ and the RDU module's from $1 \times 10^{-3}$ to $1 \times 10^{-5}$.

The EFR module is pre-trained for 500 epochs with a batch size of 8. The main loss components consist of the Charbonnier loss ($\lambda_{\text{char}} = 1.0$), SSIM loss ($\lambda_{\text{ssim}} = 0.05$), and edge loss ($\lambda_{\text{edge}} = 0.05$), collectively weighted by $\lambda_{\text{main}} = 1$ in the total loss function. In addition, the reblur loss and event gradient loss are incorporated with weights $\lambda_{\text{reblur}} = 0.2$ and $\lambda_{\text{grad}} = 0.05$, respectively, to capture the physical characteristics of the multi-modal input.

The RDU module is pre-trained for 200 epochs. The composite loss $\mathcal{L}_{\text{RDU}}$ comprises the primary dense loss $\mathcal{L}_{\text{dense\_depth}}$ and the sparse supervision components, including $\mathcal{L}_{\text{sparse\_depth}}$ and $\mathcal{L}_{\text{sparse\_grad}}$. The sparse supervision is incorporated with $\lambda_{\text{sparse}} = 0.5$ to ensure high fidelity at LiDAR-sampled points.

The entire ERU-Net is jointly fine-tuned for 100 epochs. In this stage, the learning rate is further reduced, decaying from $1 \times 10^{-5}$ to $1 \times 10^{-7}$. During fine-tuning, only the depth loss $\mathcal{L}_{\text{RDU}}$ is employed to optimize the EFR features exclusively for the geometric completion task.

**Table 1:** The performance comparison on the KITTI and NYUv2 Datasets

| | KITTI Dataset | | | | | NYUv2 Dataset | | | | | Complexity | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **RMSE** | **MAE** | **iRMSE** | **REL** | $\delta_1$ | **RMSE** | **MAE** | **iRMSE** | **REL** | $\delta_1$ | **Params.** | **FLOPs** |
| CSPN | 1324.59 | 397.66 | 3.93 | 0.022 | 99.3 | 6.915 | 3.104 | 7.22 | 0.068 | 92.1 | 52.1 | **146.7** |
| Mari-DC | 1093.50 | 364.40 | 3.44 | 0.084 | 93.9 | 3.902 | 1.685 | 1.45 | 0.035 | 95.8 | 877.1 | 200M |
| DFU | 1908.00 | 451.70 | 3.50 | 0.019 | 98.6 | 5.568 | 2.034 | 2.76 | 0.045 | 94.5 | 39.4 | 624.8 |
| BP-Net | 1613.62 | 381.26 | 3.83 | 0.019 | 98.9 | 5.842 | 2.417 | 2.88 | 0.051 | 93.7 | 107.0 | 263.1 |
| ERU-Net | **888.20** | **209.70** | **2.00** | **0.010** | **99.8** | **2.603** | **1.039** | **0.50** | **0.012** | **99.5** | **29.8** | 346.9 |

# 4 Experiments

## 4.1 Datasets

**KITTI Depth Completion Dataset [8]** is the standard benchmark for outdoor autonomous driving, providing sparse 64-line Velodyne LiDAR synchronized and RGB video. From the video sequences, we synthesize long-exposure blurred inputs with corresponding event streams.
**NYUv2 Dataset [31]** provides RGB images and corresponding dense depth maps captured by a Microsoft Kinect sensor. As it provides dense ground truth, we follow standard evaluation protocols [20, 43] and simulate the sparse LiDAR input by sampling 500 points.
**Night-RIDE Dataset** was collected and will be released to validate ERU-Net on real-world, multi-modal data captured in our target scenario. This dataset was captured using a FLIR industrial RGB camera, an EVK4 event camera, and a 64-line Hesai PandarQT LiDAR, with all sensors rigorously hardware-synchronized. The 3 inputs (RGB, Events, and LiDAR) were aligned and cropped to $256 \times 256$. The dataset was strictly partitioned into training, validation, and testing sets with a ratio of 7:1:2, ensuring no test data is exposed during the training phase.

## 4.2 Evaluation Metrics and Baselines

**Evaluation Metrics.** Following the standard KITTI protocol [8], dense depth is evaluated using RMSE, MAE, iRMSE, REL, and $\delta_1$, where lower is better for RMSE, MAE, iRMSE, and REL (↓) and higher is better for $\delta_1$ (↑).
**Baselines.** We compare our ERU-Net against state-of-the-art (SOTA) RGB-guided depth completion methods:

- **CSPN** [2]: A classic and strong baseline that utilizes convolutional spatial propagation networks to refine depth.
- **BP-Net** [34]: A recent SOTA method (CVPR 2024) based on bilateral propagation.
- **DFU** [43]: A SOTA method (CVPR 2024) introducing a dedicated upsampling guided by internal dense features.
- **Marigold-DC (Mari-DC)** [38]: A state-of-the-art (ICCV 2025) zero-shot diffusion model for depth completion.

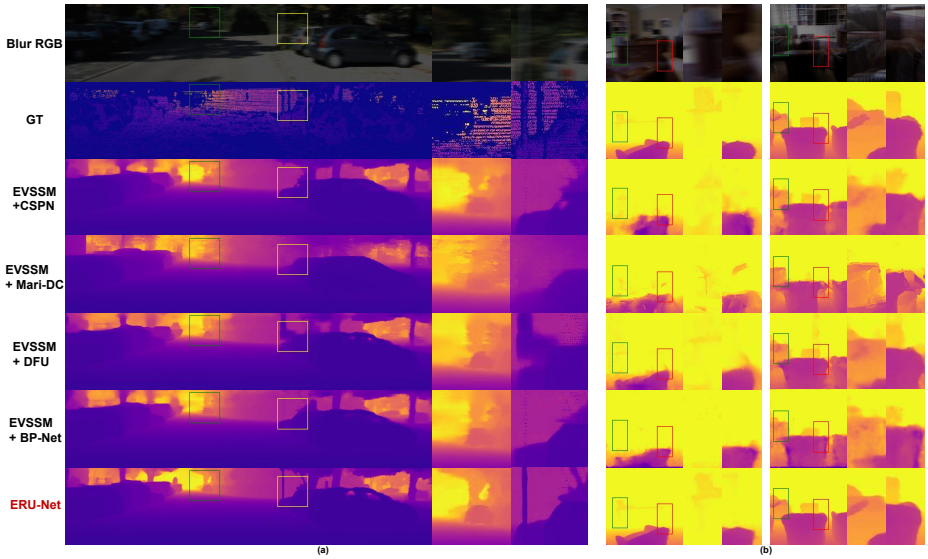


**Fig. 3:** Qualitative performance comparison on simulated datasets, visualizing the results on a challenging outdoor scene from the KITTI dataset (**a**) and two complex indoor scenes from the NYUv2 dataset (**b**). The 4th frame (out of 8) from the reconstructed depth sequence is displayed, with zoomed-in regions for detailed exhibition.

- **EVSSM** [13]: A SOTA (CVPR 2024) image deblurring network based on an efficient visual state space model, serving as the pre-processing for all baselines.
- **EFNet** [33]: A event-based motion deblurring baseline (ECCV 2022) that utilizes a symmetric cumulative event representation and cross-modal attention.
- **REFID** [32]: SOTA method (TPAMI 2025) introducing a unified bidirectional recurrent network for event-based frame interpolation and deblurring.

**Fairness Comparison Illustration.** To ensure a strictly fair comparison, all depth completion baselines (CSPN, BP-Net, DFU, Mari-DC) are re-trained on the deblurred RGB images using official code and optimal hyperparameters. All baselines follow a two-stage "Restore–then–Complete" pipeline: EVSSM [13] pre-processes the long-exposure image ($I_{\text{blur}}$) for guidance, whereas our ERU-Net operates directly on raw inputs.

### 4.3 Performance on Simulated Data

**Quantitative Comparison.** Tab. 1 summarizes the primary quantitative results on the KITTI and NYUv2 benchmarks. Despite applying the state-of-the-art deblurring model EVSSM as a pre-processing step, all baseline pipelines still exhibit substantial errors—for instance, EVSSM+CSPN reaches an RMSE

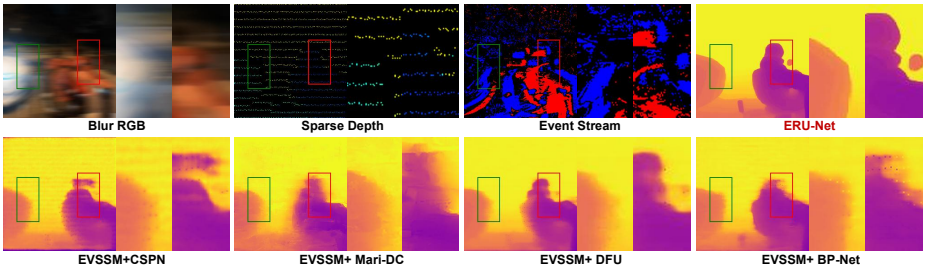


**Fig. 4:** Qualitative results on a real-world dataset. The upper row shows the raw inputs (long-exposure blurred RGB, sparse point cloud from Lidar, and corresponding event stream), and one examplar frame (4th) from the reconstructed 8-frame depth sequence produced by ERU-Net. The lower row shows the baseline results. Zoom-in boxes highlight magnified details.

of 1324.59 and EVSSM+DFU reaches 1908.00 on KITTI. Compared to CSPN, DFU and BP-Net rely on pre-deblurring RGB guidance. Under severe motion blur, deblurring networks introduce artifacts, leading to error propagation. In contrast, CSPN is more robust due to its conservative local propagation, while Mari-DC benefits from strong generative priors. These results indicate that the perceptual-quality–oriented deblurrer fails to recover geometrically faithful guidance features from the severely blurred long-exposure inputs, ultimately leading to degraded depth completion performance.

In contrast, ERU-Net establishes new state-of-the-art performance across all five metrics on both benchmarks. On KITTI, it attains an RMSE of 888.20, marking an 18.8% improvement over the competing baseline, EVSSM+Mari-DC (1093.50). EFR module first recovers a clean and high-fidelity latent feature sequence, which is subsequently fused with the spatial event stream via the proposed HGCM module to effectively guide depth reconstruction. The entire 3-stage training tailors the restored features to geometric consistency.

**Qualitative Comparison.** Fig. 3 presents a clear visual comparison on the two simulated datasets. (Additional experimental results and video demos are provided in the supplementary material.) In both KITTI (a) and NYUv2 (b), the two-stage EVSSM+Baseline pipelines produce artifacts, blur dynamic objects, and fail to recover clean geometric structures in overlapping regions. In contrast, ERU-Net reconstructs sharp and geometrically accurate depth boundaries that closely align with the ground truth. The observations underscore that the deblurrer's restored image offers insufficient geometric fidelity to serve as a reliable guide for high-quality depth completion.

### 4.4 Performance on Real-World Data

ERU-Net is further validated on a challenging real-world nighttime dataset. Fig. 4 presents a qualitative comparison under an extremely low-light scenario. (Additional experimental results and video demos are provided in the supplemen-

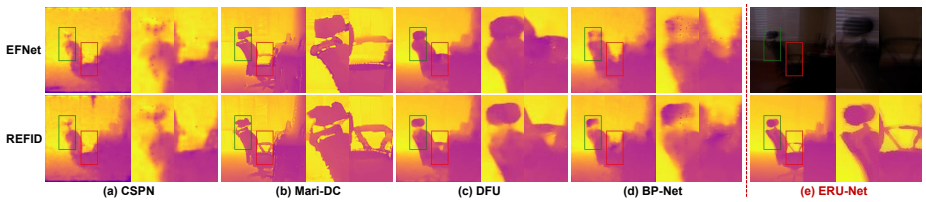


**Fig. 5: Visual Comparison with Event-Deblurring + Depth Reconstruction Pipelines.** Top/Bottom rows use EFNet and REFID pre-processing, respectively. (a)-(d) Baselines (CSPN, Mari-DC, DFU, BP-net). (e) ERU-Net and raw blurred input. ERU-Net recovers fine structures (e.g., chair legs) where baselines fail.

**Table 2:** Comparison with Event-Deblurring + Depth Completion Methods.

| **Methods** | | **RMSE**↓ | **MAE**↓ | **iRMSE**↓ | **REL**↓ | $\delta_1$↑ |
|---|---|---|---|---|---|---|
| EFNet + | CSPN | 6.090 | 2.825 | 6.59 | 0.056 | 93.9 |
| | Mari-DC | 3.352 | 1.385 | 1.15 | 0.024 | 97.5 |
| | DFU | 4.751 | 1.721 | 2.40 | 0.032 | 96.1 |
| | BP-Net | 5.056 | 2.291 | 2.43 | 0.041 | 95.5 |
| REFID + | CSPN | 5.782 | 2.751 | 6.03 | 0.053 | 94.7 |
| | Mari-DC | 3.145 | 1.258 | 0.95 | 0.019 | 98.2 |
| | DFU | 4.371 | 1.532 | 2.30 | 0.028 | 96.9 |
| | BP-Net | 4.455 | 1.926 | 2.06 | 0.034 | 96.6 |
| **ERU-Net (Ours)** | | **2.603** | **1.039** | **0.50** | **0.012** | **99.5** |

tary material.) The inputs include a severely blurred RGB image with strong motion and flare, and a sparse depth map containing very few points on the moving car and pedestrian. In contrast, the event stream clearly captures the underlying motion patterns.

The EVSSM+Baseline pipelines fail to reconstruct the primary foreground objects from the sparse depth and the poorly restored RGB input. As highlighted in the boxed regions, they either produce large holes or fuse the pedestrian into the background. By directly leveraging the raw event stream through its HGCM module, ERU-Net accurately localizes the motion boundaries of both the car and the pedestrian, yielding a clean and well-defined depth reconstruction, confirming the practical applicability of ERU-Net in real-world conditions.

### 4.5 Complexity Analysis

Tab. 1 summarizes model complexity, where baselines reflect the full two-stage "EVSSM+Baseline" pipeline. ERU-Net is the most compact (29.8M params, 346.9G FLOPs), 1.3× smaller and 1.8× more efficient than DFU (39.4M, 624.8G), while CSPN and BP-Net are lighter but less accurate (RMSE 1324.59 and 1613.62). ERU-Net achieves a superior accuracy–efficiency trade-off.

**Table 3:** Quantitative Ablation Study with different modules.

| Variant | RMSE↓ | MAE↓ | iRMSE↓ | REL↓ | $\delta_1$↑ |
|---|---|---|---|---|---|
| w/o EFR | 4.215 | 1.682 | 1.85 | 0.048 | 92.5 |
| w/o HGCM | 2.345 | 0.892 | 0.92 | 0.021 | 98.1 |
| **Full ERU-Net** | **1.505** | **0.502** | **0.40** | **0.009** | **99.6** |

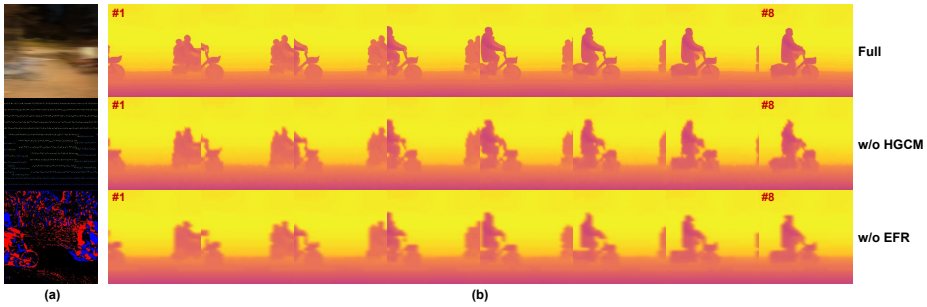


**Fig. 6:** Visual results of the ablation study on a typical real-world dark scene. **(a)** The low-quality multi-modal inputs, including the blurred RGB, sparse depth, and event stream. **(b)** Visual comparison of the reconstructed 8-frame depth map sequences.

### 4.6 Comparison with Event-based Cascaded Pipelines

To rigorously evaluate the effectiveness of the pre-processing paradigm, we construct stronger cascaded baselines by replacing the standard deblurring module with state-of-the-art event-based methods, including EFNet [33] and REFID [32]. As illustrated in Tab. 2 and Fig. 5, these advanced Restore-then-Complete pipelines consistently underperform ERU-Net.

This performance gap arises from two key factors. First, existing event-based deblurring methods are primarily optimized for perceptual restoration and struggle to recover artifact-free structures under severe motion and extremely low illumination, whereas our EFR module restores geometrically consistent latent features that serve as reliable structural priors. Second, the decoupled Image-then-Depth paradigm breaks the geometric consistency required for depth estimation, allowing restoration artifacts to propagate into the completion stage. In contrast, ERU-Net directly integrates the restored latent representation into the RDU module, enabling end-to-end optimization for geometrically consistent depth reconstruction while preserving fine structural details.

### 4.7 Ablation Studies

To validate our two main contributions, we conducted a comprehensive ablation study on a real-world dataset. We compared the full ERU-Net with two degraded variants: (1) w/o EFR, where the EFR module is replaced by a state-of-the-art

deblurrer; (2) w/o HGCM, where the hybrid-guidance stream is removed from the RDU, forcing HGCM blocks to rely solely on the restored latent feature.

As shown in Tab. 3, ablating the EFR module leads to a substantial performance degradation, highlighting that supervision on blurred RGB alone is insufficient and that INR-based restoration is essential. Ablating the HGCM similarly increases depth errors, underscoring the importance of explicit event guidance for reliable upsampling.

Fig. 6(b) visually confirms the superiority of the full ERU-Net pipeline. The w/o EFR variant yields blurred depth with ghosting and scanline artifacts, whereas w/o HGCM recovers shapes with jagged, unclear contours. This ablation demonstrates the synergistic value of our components: EFR, jointly fine-tuned with the depth network, provides task-optimized latent guidance that deblur methods fail to provide. HGCM leverages the spatial event stream to refine motion boundaries, enabling the sharp, detailed reconstructions.

## 5 Conclusion

This paper proposes a unified framework for dense depth acquisition in dark dynamic scenarios where long-exposure RGB images are corrupted and struggle to compensate for the sparse point clouds from Lidar. The framework deeply couples two modules: (1) the Event-Guided Feature Restoration module, which employs an implicit neural representation and a self-recursive strategy to recover geometrically accurate features from blurred images and event streams; (2) the Restoration-Aware Depth Upsampling module, which incorporates a Hybrid-Guidance Confidence Module to fuse restored features, spatial event streams, and sparse LiDAR data. Extensive experiments demonstrate our superior performance compared to state-of-the-arts, providing a robust paradigm for 3D perception under challenging autonomous driving conditions. This approach demonstrates excellent performance, and we plan to develop a miniaturized version of the hybrid prototype to create a compact trimodal setup for nighttime imaging.

## Acknowledgements

This work is supported by the National Natural Science Foundation of China [Grant numbers 62502069], Natural Science Foundation of Liaoning Province [Grant numbers 2025-MS-007], and the Liaoning Young Elite Scientists Sponsorship Program.

## References

1. Chen, K., Chen, S., xing Zhang, J., can Zhang, B., Zheng, Y., zhu Huang, T., han Yu, Z.: Spikereveal: Unlocking temporal sequences from real blurry inputs with spike streams. pp. 62673–62696 (2024)
2. Cheng, X., Wang, P., Yang, R.: Learning depth with convolutional spatial propagation network. IEEE TPAMI **42**(10), 2361–2379 (2020)

3. Dong, J., Ota, K., Dong, M.: Video frame interpolation: A comprehensive survey **19**(2s), 1–31 (2023)
4. Dong, K., Guo, Y., Yang, R., Cheng, Y., Suo, J., Dai, Q.: Retrieving object motions from coded shutter snapshot in dark environment. IEEE TIP **32**, 3281–3294 (2023)
5. Fu, C., Dong, C., Mertz, C., Dolan, J.M.: Depth completion via inductive fusion of planar lidar and monocular camera. pp. 10843–10848 (2020)
6. Gallego, G., Delbrück, T., Orchard, G., Bartolozzi, C., Taba, B., Censi, A., Leutenegger, S., Davison, A.J., Conradt, J., Daniilidis, K., et al.: Event-based vision: A survey. IEEE TPAMI **44**(1), 154–180 (2020)
7. Gao, Y., Li, S., Li, Y., Guo, Y., Dai, Q.: Superfast: 200x video frame interpolation via event camera. IEEE TPAMI **45**(6), 7764–7780 (2022)
8. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? the kitti vision benchmark suite. In: CVPR. pp. 3354–3361 (2012)
9. Gu, C., Learned-Miller, E., Sheldon, D., Gallego, G., Bideau, P.: The spatio-temporal poisson point process: A simple model for the alignment of event camera data. In: ICCV. pp. 13495–13504 (2021)
10. He, W., You, K., Qiao, Z., Jia, X., Zhang, Z., ming Wang, W., liu Lu, H., kun Wang, Y., jun Liao, J.: Timereplayer: Unlocking the potential of event cameras for video interpolation. In: CVPR. pp. 17804–17813 (2022)
11. Huang, X., Zhang, Y., Xiong, Z.X.: Progressive spatio-temporal alignment for efficient event-based motion estimation. In: CVPR. pp. 1537–1546 (2023)
12. Imran, S., Long, Y., Liu, X., Morris, D.: Depth coefficients for depth completion. In: CVPR. pp. 12438–12447 (2019)
13. Kong, L., Dong, J., Yang, M.H., Pan, J.: Efficient visual state space model for image deblurring. In: CVPR. pp. 12710–12719 (2024)
14. Kurimo, E., Lepistö, L., Nikkanen, J., Grén, J., Kunttu, I., Laaksonen, J.: The effect of motion blur and signal noise on image quality in low light imaging. pp. 81–90 (2009)
15. Li, C., Guo, C., Han, L., Jiang, J., Cheng, M.M., Gu, J., Loy, C.C.: Low-light image and video enhancement using deep learning: A survey. IEEE TPAMI **44**(12), 9396–9416 (2021)
16. Liang, J., Yang, Y., Li, B., Duan, P., Xu, Y., Shi, B.: Coherent event guided low-light video enhancement. In: ICCV. pp. 10615–10625 (2023)
17. Lin, S., Zhang, J., Pan, J., Jiang, Z., Zou, D., Wang, Y., Chen, J., Ren, J.: Learning event-driven video deblurring and interpolation. In: ECCV. pp. 695–710 (2020)
18. Liu, H., Peng, S., Zhu, L., Chang, Y., Zhou, H., Yan, L.: Seeing motion at nighttime with an event camera. In: CVPR. pp. 25648–25658 (2024)
19. Liu, L., Song, X., Lyu, X., Diao, J., Wang, M., Liu, Y., Zhang, L.: Fcfr-net: Feature fusion based coarse-to-fine residual learning for depth completion. In: AAAI (2021)
20. Ma, F., Cavalheiro, G.V., Karaman, S.: Sparse-to-dense: Depth prediction from sparse depth samples and a single image. pp. 4776–4783 (2018)
21. Pan, L., Scheerlinck, C., Yu, X., Hartley, R., Liu, M., Dai, Y.: Bringing a blurry frame alive at high frame-rate with an event camera. In: CVPR. pp. 6820–6829 (2019)
22. Parihar, A.S., Varshney, D., Pandya, K., Aggarwal, A.: A comprehensive survey on video frame interpolation techniques. The Vis. Comput. **38**(1), 295–319 (2022)
23. Park, J., Joo, K., Hu, Z., Liu, C.K., Kweon, I.S.: Non-local spatial propagation network for depth completion. In: ECCV (2020)
24. Qiu, J., Cui, Z., Zhang, Y., Zhang, X., Liu, S., Zeng, B., Pollefeys, M.: Deeplidar: Deep surface normal guided depth prediction for outdoor scene from sparse lidar data and single color image. In: CVPR. pp. 3308–3317 (2019)

25. Raskar, R., Agrawal, A., Tumblin, J.: Coded exposure photography: Motion deblurring using fluttered shutter. In: ACM TOG. pp. 795–804 (2006)
26. Rebecq, H., Ranftl, R., Koltun, V., Scaramuzza, D.: High speed and high dynamic range video with an event camera. IEEE TPAMI **43**(6), 1964–1980 (2021)
27. Rho, K., Ha, J., Kim, Y.: Guideformer: Transformers for image guided depth completion. In: CVPR. pp. 6240–6249 (2022)
28. Scheerlinck, C., Rebecq, H., Gehrig, D., Barnes, N., Mahony, R., Scaramuzza, D.: Fast image reconstruction with an event camera. pp. 156–163 (2020)
29. Schuster, R., Wasenmuller, O., Unger, C., Stricker, D.: Ssgp: Sparse spatial guided propagation for robust and generic interpolation. pp. 197–206 (2021)
30. Shang, W., Ren, D., Zou, D.N., Ren, J.S., Luo, P., Zuo, W.M.: Bringing events into video deblurring with non-consecutively blurry frames. In: ICCV. pp. 4531–4540 (2021)
31. Silberman, N., Hoiem, D., Kohli, P., Fergus, R.: Indoor segmentation and support inference from RGBD images. In: ECCV (2012)
32. Sun, L., Gehrig, D., Sakaridis, C., Gehrig, M., Liang, J., Sun, P., Xu, Z., Wang, K., Gool, L.V., Scaramuzza, D.: A unified framework for event-based frame interpolation with ad-hoc deblurring in the wild. IEEE TPAMI **47**(4), 2265–2279 (2025)
33. Sun, L., Sakaridis, C., Liang, J., Jiang, Q., Yang, K.B., Sun, P., Ye, Y., Wang, K., Gool, L.V.: Event-based fusion for motion deblurring with cross-modal attention. In: ECCV. pp. 412–428 (2022)
34. Tang, J., Tian, F.P., An, B., Li, J., Tan, P.: Bilateral propagation network for depth completion. In: CVPR. pp. 9763–9772 (2024)
35. Tang, J., Tian, F.P., Feng, W., Li, J., Tan, P.: Learning guided convolutional network for depth completion. IEEE TIP **29**, 4531–4544 (2020)
36. Tulyakov, S., Bochicchio, A., Gehrig, D., Georgoulis, S., Li, Y., Scaramuzza, D.: Time lens++: Event-based frame interpolation with parametric non-linear flow and multi-scale fusion. In: CVPR. pp. 17755–17764 (2022)
37. Tulyakov, S., Gehrig, D., Georgoulis, S., Erbach, J., Gehrig, M., Li, Y., Scaramuzza, D.: Time lens: Event-based video frame interpolation. In: CVPR. pp. 16155–16164 (2021)
38. Viola, M., Qu, K., Metzger, N., Ke, B., Becker, A., Schindler, K., Obukhov, A.: Marigold-DC: Zero-shot monocular depth completion with guided diffusion. ArXiv **abs/2412.13389** (2024)
39. Wang, B., xing He, J., Yu, L., Xia, G.S., ming Yang, W.: Event enhanced high-quality image recovery. In: ECCV. pp. 155–171 (2020)
40. Wang, Y., Dai, Y., Liu, Q., Yang, P., Sun, J., Li, B.: Cu-net: Lidar depth-only completion with coupled u-net **7**(2), 4413–4420 (2022)
41. Wang, Y., Li, B., Zhang, G., Liu, Q., Gao, T., Dai, Y.: Lrru: Long-short range recurrent updating networks for depth completion. In: ICCV. pp. 9388–9398 (2023)
42. Wang, Y., Mao, Y., Liu, Q., Dai, Y.: Decomposed guided dynamic filters for efficient rgb-guided depth completion. IEEE TCSVT **34**, 1186–1198 (2023)
43. Wang, Y., Zhang, G., Wang, S., Li, B., Liu, Q., Hui, L., Dai, Y.: Improving depth completion via depth feature upsampling. In: CVPR. pp. 21104–21113 (2024)
44. Weng, W.M., Zhang, Y., Xiong, Z.X.: Event-based video reconstruction using transformer. In: ICCV. pp. 2563–2570 (2021)
45. Weng, W.M., Zhang, Y., Xiong, Z.X.: Event-based blurry frame interpolation under blind exposure. In: CVPR. pp. 1588–1598 (2023)
46. Xu, Z., Yin, H., Yao, J.: Deformable spatial propagation networks for depth completion. In: ICIP. pp. 913–917 (2020)

47. Yan, Z., Wang, K., Li, X., Zhang, Z., Xu, B., Li, J., Yang, J.: Rignet: Repetitive image guided network for depth completion. In: ECCV (2022)
48. xing Zhang, J., shan Chen, S., xuan Zheng, Y., han Yu, Z., zhu Huang, T.: Spike-guided motion deblurring with unknown modal spatiotemporal alignment. In: CVPR. pp. 25047–25057 (2024)
49. Zhang, J., Wang, Y., Liu, W., Li, M., Bai, J., Yin, B., Yang, X.: Frame-event alignment and fusion network for high frame rate tracking. In: CVPR. pp. 9781–9790 (2023)
50. Zhang, Y., Wei, P., Li, H., Zheng, N.: Multiscale adaptation fusion networks for depth completion. pp. 1–7 (2020)
51. Zhang, Y., Guo, X., Poggi, M., Zhu, Z., Huang, G., Mattoccia, S.: Completionformer: Depth completion with convolutions and vision transformers. In: CVPR. pp. 18527–18536 (2023)
52. Zhang, Y., Wang, C., Maybank, S.J., Tao, D.P.: Exposure trajectory recovery from motion blur. IEEE TPAMI **44**(11), 7490–7504 (2022)
53. Zhong, Z.Q., Sun, X.Y., Wu, Z.H., Zheng, Y.L., Lin, S., Sato, I.: Animation from blur: Multi-modal blur decomposition with motion guidance. In: ECCV. pp. 599–615 (2022)
54. Zhu, A.Z., Wang, Z., Khant, K.A., Daniilidis, K.: EventGAN: Leveraging large scale image datasets for event cameras. pp. 1–11 (2021)