This CVPR Workshop paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

NTIRE 2025 Challenge on Event-Based Image Deblurring: Methods and Results

Lei Sun* Andrea Alfarano* Peiqi Duan* Shaolin Su* Kaiwei Wang*
Boxin Shi* Radu Timofte* Danda Pani Paudel* Luc Van Gool* Qinglin Liu
Wei Yu Xiaoqian Lv Lu Yang Shuigen Wang Shengping Zhang Xiangyang Ji
Long Bao Yuqiang Yang Jinao Song Ziyi Wang Shuang Wen Heng Sun
Kean Liu Mingchen Zhong Senyan Xu Zhijing Sun Jiaying Zhu Chengjie Ge
Xingbo Wang Yidi Liu Xin Lu Xueyang Fu Zheng-Jun Zha Dawei Fan
Dafeng Zhang Yong Yang Siru Zhang Qinghua Yang Hao Kang Huiyuan Fu
Heng Zhang Hongyuan Yu Zhijuan Huang Shuoyan Wei Feng Li
Runmin Cong Weiqi Luo Mingyun Lin Chenxu Jiang Hongyi Liu Lei Yu
Weilun Li Jiajun Zhai Tingting Lin Shuang Ma Sai Zhou Zhanwen Liu
Yang Wang Eiffel Chong Nuwan Bandara Thivya Kandappu Archan Misra
Yihang Chen Zhan Li Weijun Yuan Wenzhuo Wang Boyang Yao
Zhanglu Chen Yijing Sun Tianjiao Wan Zijian Gao Qisheng Xu Kele Xu
Yukun Zhang Yu He Xiaoyan Xie Tao Fu Yashu Gautamkumar Patel
Vihar Ramesh Jain Divesh Basina Rishik Ashili Manish Kumar Manjhi
Sourav Kumar Prinon Benny Himanshu Ghunawat B Sri Sairam Gautam
Anett Varghese Abhishek Yadav

Abstract

This paper presents an overview of NTIRE 2025, the First Challenge on Event-Based Image Deblurring, detailing the proposed methodologies and corresponding results. The primary goal of the challenge is to design an event-based method that achieves high-quality image deblurring, with performance quantitatively assessed using Peak Signal-to-Noise Ratio (PSNR). Notably, there are no restrictions on computational complexity or model size. The task focuses on leveraging both events and images as inputs for single-image deblurring. A total of 199 participants registered, among whom 15 teams successfully submitted valid results, offering valuable insights into the current state of event-based image deblurring. We anticipate that this challenge will drive further advancements in event-based vision research.

* L. Sun (lei.sun@insait.ai, INSAIT, Sofia University "St. Kliment Ohridski"), A. Alfarano, P. Duan, S. Su, K. Wang, B. Shi, R. Timofte, D. P. Paudel, and L. Van Gool were the challenge organizers, while the other authors participated in the challenge.
Appendix A contains the authors' teams and affiliations.
NTIRE 2025 webpage: https://cvlai.net/ntire/2025/.
Code: https://github.com/AHupuJR/NTIRE2025_EventDeblur_challenge.

1. Introduction

Traditional camera output frames with relatively long exposure time in a fixed framerate. In contrast, event cameras, a kind of neuromorphic sensor, asynchronously capture pixel-wise intensity changes with high temporal resolution [12], and have been applied in various fields such as computational imaging [32, 39–41, 43], human pose estimation [2], depth estimation [30, 34], image segmentation [1, 56], *etc.* .

In recent years, significant efforts have been dedicated to event-based image restoration. Among various tasks, event-based image deblurring has gained the most attention, as the high temporal resolution of event cameras provides valuable priors for motion deblurring [39–41]. Notably, these methods operate under the assumption that input images and events are spatially aligned—a condition that applies to all approaches discussed in this paper.

In conjunction with the NTIRE 2025 Workshop on New Trends in Image Restoration and Enhancement, the Event-Based Image Deblurring Challenge was organized. The objective is to develop a network architecture or solution that effectively integrates events and images to enhance image deblurring performance. We hope that this challenge will serve as a starting point for promoting event-based image

enhancement on a broader stage and contribute to the thriving development of the event-based vision community.

This challenge is one of the NTIRE 2025 Workshop associated challenges on: ambient lighting normalization [47], reflection removal in the wild [51], shadow removal [46], event-based image deblurring [42], image denoising [44], XGC quality assessment [27], UGC video enhancement [37], night photography rendering [10], image super-resolution (x4) [4], real-world face restoration [5], efficient super-resolution [36], HR depth estimation [53], efficient burst HDR and restoration [19], cross-domain few-shot object detection [11], short-form UGC video quality assessment and enhancement [22, 23], text to image generation model quality assessment [13], day and night raindrop removal for dual-focused images [21], video quality assessment for video conferencing [16], low light image enhancement [28], light field super-resolution [48], restore any image model (RAIM) in the wild [25], raw restoration and super-resolution [7] and raw reconstruction from RGB on smartphones [8].

2. NTIRE 2025 Event-Based Image Deblurring Challenge

The goals of this challenge include: (1) promoting research in the area of event-based image deblurring, (2) facilitating comparisons between various methods, and (3) providing a platform for academic and industrial participants to engage, discuss, and potentially establish collaborations. This section delves into the specifics of the challenge, including the dataset, challenge phases and evaluation criteria.

2.1. Dataset

The HighREV dataset [40] is used for both training and evaluation in this challenge. It consists of 1,771 sets of blurry images, corresponding events, and sharp images for training. Additionally, 421 sets are provided as validation data during the development phase, ensuring a comprehensive benchmark for assessing model performance.

2.2. Tracks and Competition

The aim is to obtain a network design capable to produce high-quality results with the best performance measured by PSNR for event-based image deblurring.

Challenge phases Participants were given access to training images from the HighREV dataset. During the validation phase, they could use 421 images from the validation set for model tuning. In the test phase, evaluation was performed on 271 images from the test set. To ensure a fair assessment, the ground-truth images for the test phase remained hidden from participants throughout the challenge.

https://www.cvlai.net/ntire/2025/

Team	Rank	PSNR (primary)	SSIM
IVISLAB	1	42.79	0.9196
MiVideoDeblur	2	42.70	0.9281
404NotFound	3	42.09	0.9300
Give_it_a_try	4	40.37	0.9234
BUPTMM	5	40.21	0.9179
WEI	6	39.46	0.9171
DVS-WHU	7	39.26	0.9101
PixelRevive	8	39.12	0.9112
CHD	9	38.56	0.9055
SMU	10	38.30	0.9047
JNU620	11	37.63	0.9019
colab	12	36.84	0.8962
CMSL	13	31.81	0.8900
KUnet	14	29.42	0.8600
Group10	15	25.93	0.8200

Table 1. Results of NTIRE 2025 Event-Based Image Deblurring Challenge. PSNR and SSIM scores are measured on the 271 test images from HighREV dataset. Team rankings are based primarily on PSNR.

Evaluation protocol Since the aim of this challenge is to foster the development of accurate event-based image deblurring networks, PSNR and SSIM on the 271 testing images are used as the quantitative evaluation metrics. A code example for calculating these metrics is available at https://github.com/AHupuJR/NTIRE2025_EventDeblur_challenge. The code of the submitted solutions and the pretrained weights are also available in this repository.

3. Challenge Results

Table 1 shows the final rankings and test results of the participated teams. The implementation details of each team can be found in Sec.4, while team member information can be found in Appendix A. IVISLAB achieved the first place in terms of PSNR, followed by MiVideoDeblur and 404NotFound as the second and third place, respectively.

3.1. Participants

The challenge attracted 199 registered participants, with 15 teams successfully submitting valid results.

3.2. Main Ideas and Architectures

Throughout the challenge, participants explored various innovative techniques to improve deblurring performance. Below, we summarize some of the key strategies employed by the top-performing teams.

1. **Hybrid architectures demonstrated strong performance,** with all top-3 teams utilizing a combination of transformers and convolutional networks. This approach leverages global features extracted by transformers alongside local features captured by convolutional

layers, both of which contribute to effective event-based image deblurring. Besides, both spatial and channel attention mechanisms play a crucial role in enhancing overall performance.

2. **Pretrained weights matters.** The winning team, IVISLAB, leveraged a backbone model initialized with pretrained weights from ImageNet, demonstrating the advantages of transfer learning in event-based image deblurring.
3. **Cross-modal fusion proves beneficial.** Several teams adopted EFNet [39] and REFID [40, 41] as a baseline model to fuse features from the event and image branches.
4. **Effective training strategies.** Both the second and third-place teams employed progressive learning techniques during training. Additionally, the winning team utilized a large patch size (512×512), which contributed to improved performance.
5. **Incorporating a novel Mamba-based architecture**. Integrating features from both image and event modalities is crucial for enhancing the reconstruction quality of event-based deblurring methods. Team DVS-WHU introduced an innovative Mamba-based architecture to achieve more effective fusion.

3.3. Fairness

To maintain fairness in the event-based image deblurring challenge, specific rules were implemented, primarily regarding the datasets used for training. Participants were permitted to use external datasets for training. However, incorporating the HighREV validation set, whether sharp or blurry images, was strictly prohibited, as this set served to evaluate the overall performance and generalizability of the models. Additionally, the use of HighREV test blurry images for training was not allowed. On the other hand, employing advanced data augmentation techniques during training was considered an acceptable practice.

4. Challenge Methods and Teams

4.1. IVISLAB

To achieve image deblurring, team IVISLAB introduces the Triple Event-stream Image Deblurring Network (TEIDNet). As depicted in Figure 1, TEIDNet converts consecutive events into event voxels at three temporal scales to perceive motion information from blur images and capture fine edges for reconstructing clear images. Furthermore, TEIDNet integrates Shift Window Attention and Channel-Wise Attention blocks to capture local and global contexts, thereby enhancing deblurring accuracy.

4.1.1. Network Architecture

TEIDNet adopts an encoder-decoder architecture to process images and triple-stream event voxels, aiming to estimate the deblurred image. Specifically, when deblurring the image at frame t, TEIDNet considers that the long-term event stream surrounding frame t can aid in motion perception. Therefore, it voxelizes the event data from frame $t - T_l$ to frame $t + T_l$ into a b-bin event voxel $V_{l,t}$. Simultaneously, since the short-term event stream around frame t can help reconstruct high-frequency textures, TEIDNet voxelizes the event data from frame $t - T_s$ to frame $t + T_s$ into a b-bin event voxel $V_{s,t}$. Furthermore, to mitigate color artifacts by leveraging higher-resolution motion information near the current frame, TEIDNet voxelizes the event data from frame $t - T_m$ to frame $t + T_m$ into a b-bin event voxel $V_{m,t}$. Subsequently, the event voxels $V_{l,t}$, $V_{s,t}$, and $V_{m,t}$, along with the blur image I_b, are concatenated and fed into the network. To effectively fuse the features from the image and event voxels, TEIDNet employs convolutional layers to generate fused feature representations. The network then utilizes a dual-branch encoder. The first, a complex branch extracts high-level semantic information from the fused features by leveraging shift window attention to capture local context and channel-wise attention blocks to capture global context. The second, a simple branch utilizes convolutional layers to capture fine-grained details from the fused features. Next, TEIDNet's decoder integrates multiple shift window attention blocks to fuse and upsample the features extracted by the dual-branch encoder. Finally, convolutional layers are employed to predict the deblurred image I_t.

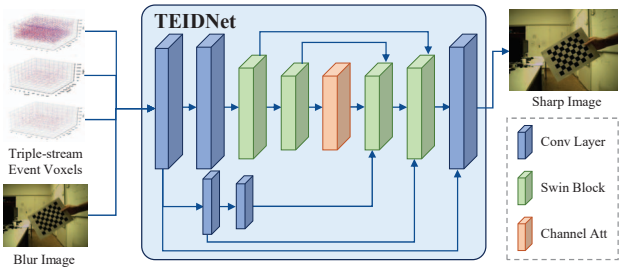


Figure 1. The model architecture of TEIDNet, proposed by Team IVISLAB.

4.1.2. Loss Function

To train TEIDNet, they define a reconstruction loss $\mathcal{L}_r$ for the estimated deblurred image I_t as follows:

$$\mathcal{L}_r = \lambda_1 \mathrm{L}_1(I_t, I_t^{gt}) + \lambda_2 \mathrm{L}_2(I_t, I_t^{gt}) \tag{1}$$

Here, λ_1 and λ_2 are coefficients that balance the loss terms. The function $\mathrm{L}_1(\cdot,\cdot)$ represents the mean absolute error, while $\mathrm{L}_2(\cdot,\cdot)$ denotes the mean squared error. The term I_t^{gt} refers to the ground truth image at frame t.

4.1.3. Implementation Details

TEIDNet is implemented using PyTorch on four Nvidia L20 GPUs. During training, a batch size of 16 is utilized, with

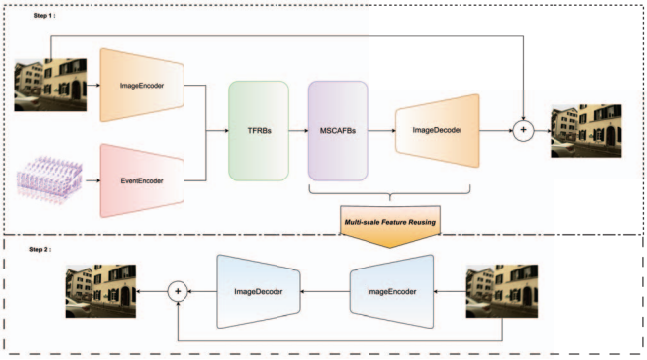


Figure 2. The framework of DASTF-Net, proposed by Team MiVideoDeblur.

input data dimensions of 512×512 pixels. The network weights are optimized over 1000 epochs using the AdamW optimizer, with an initial learning rate set to 2×10^{-5}. A cosine annealing scheduler is employed to decay the learning rate progressively. In addition, they take the checkpoint with good performance and perform a second finetune. To mitigate overfitting, data augmentation techniques such as random flipping and rotation are applied. They also initialize the backbone network parameters using weights pretrained on ImageNet. The specific coefficients and parameters are defined as follows: number of bins $b = 7$, long-term temporal window $T_l = 5$, medium-term temporal window $T_m = 1$, short-term temporal window $T_s = 0$, and loss function weights $\lambda_1 = 1$, $\lambda_2 = 1$.

4.2. MiVideoDeblur

Introduction. As illustrated in Fig. 2, their team proposed the Dual Attention Spatio-Temporal Fusion Network(DASTF-Net). Motivated by EFNet [39], their model employs a two-stage encoder-decoder architecture. Initially, two encoders separately extract multi-scale features from both the image and event data. Based on the EGACA module [40] and the FAF module [45], they have designed the Temporal Fusion Residual Block (TFRB) and Multi-Scale Cross-Attention Fusion Block (MSCAFB), which perform feature fusion in the temporal and spatial dimensions, respectively. By incorporating a dual-attention mechanism, these modules effectively enhance the model's performance. Following feature fusion, the fused features are fed into a Restormer [55], which further leverages the feature information to improve the model's performance.

Training strategy. They employed a four-stage training strategy. In the first stage, the network was trained for 160k iterations using the PSNRLoss function. AdamW Optimizer was used, with an initial learning rate of 2e-4 and a cosine annealing learning rate schedule for updates. Subsequently, in the second stage, data augmentation techniques were introduced, which included adding random Gaussian noise and applying random scaling to the input data. Building upon the model from the first stage, the training continued for 80k iterations with an initial learning rate of 1e-4. For the third and fourth stages, the patch size was progressively increased from 256 to 320 and then to 480. The network was trained for 40k iterations in the third stage and 45k iterations in the fourth stage.

4.3. 404NotFound

Their team proposes EV-Deblurformer[26], a framework consisting of two complementary models designed to fully leverage the temporal dynamics of video sequences and the rich texture details present in single images. The framework includes two distinct components: Video-SFHformer, developed for video-based deblurring, and EFSformer, tailored for single-image deblurring. In Video-SFHformer, they introduce STFBlock to enhance the model's capacity for long-range temporal modeling. In EFSformer, they incorporate STEFusionBlock, which fuses event features from the frequency domain to improve spatial detail restoration. To achieve optimal performance, as shown in Section 4.3.3, a sequence-level ensemble strategy is employed to merge the outputs of both models. A progressive training scheme is also adopted to enhance robustness and effectiveness.

4.3.1. Overall Pipeline

Figure 3 illustrates the overall architecture of their proposed method, EV-Deblurformer. This approach, built upon the two models: Video SFHformer and EFSformer, fully exploits the rich temporal dynamics and sharp edge information provided by event data. For the video deblurring model, they propose the Video-SFHformer based on SFHformer. For the single-image motion deblurring model, they propose the EFSformer built on EFNet[39].

4.3.2. Implementation Details

They implement their proposed network via the PyTorch 2.1.2 platform. Adam optimizer with parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$ is adopted to optimize their network. Motivated by [55] they introduce the progressive training strategy. The training phase of their network could be divided into two stages:

(1) **Initial training of EV-Deblurformer**. They use a progressive training strategy at first. For the video-based motion deblurring model, they start training with patch size 152×152 with batch size of 16 for 250K iterations. The patch size and batch size pairs are updated to $[(192^2, 12), (256^2, 8), (304^2, 8)]$ at iterations [250K, 200K, 150K]. The initial learning rate is 2×10^{-4} and remains unchanged when patch size is 192. Later, the learning rate is set to 1×10^{-4} and 7×10^{-5} for patch and batch size pairs of $(256^2,\ 8)$ and $(304^2,\ 8)$, respectively. They employ a

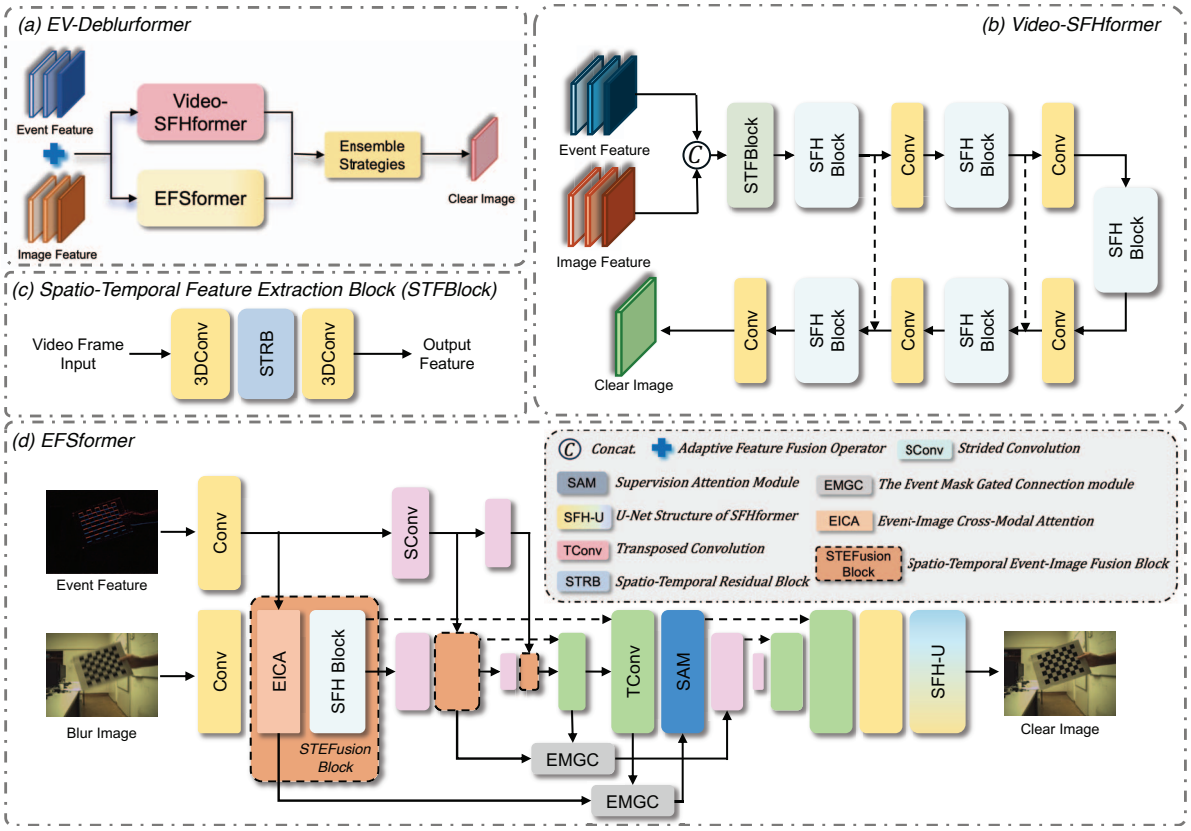


Figure 3. The architecture diagram of EV-Deblurformer, proposed by Team 404NotFound, is designed for event-guided motion deblurring.

cosine annealing learning rate decay strategy, gradually reducing the learning rate. For the single-image-based motion deblurring model, They begin training with a patch size of 192×192 and a batch size of 12 for 250K iterations. During training, patch size and batch size pairs are progressively updated to $(256^2,\ 10)$, $(288^2,\ 8)$, and $(320^2,\ 8)$ at 36K, 24K, and 24K iterations, respectively. The initial learning rate is set to 5×10^{-4}, and later adjusted to 1×10^{-4}, 7×10^{-5}, and 5×10^{-5} corresponding to the updated patch and batch size configurations. A cosine annealing schedule is employed to gradually decay the learning rate throughout the training process. The first stage is performed on the NVIDIA RTX 4090 GPU. They obtain the best model at this stage as the initialization of the second stage.

(2) **Fine-tuning EV-Deblurformer**. For the video-based motion deblurring model, they start training with a patch size of 320×320 and a batch size of 4 for 150K iterations. The initial learning rate is set to 1×10^{-5} and is adjusted to 1×10^{-7} using a cosine annealing schedule, over a total of 150K iterations. They use the entire training data from the challenge without applying any data augmentation techniques. The exponential moving average (EMA) is employed for the dynamic adjustment of the model parameters. For the single-image-based motion deblurring model, they adopt the same training strategy as used in the video-based motion deblurring model. The second training stage is conducted on an NVIDIA RTX 4090 GPU.

(3)**Evaluation Metrics** They utilize two widely adopted reference-based evaluation metrics—Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM)[49]—to evaluate the effectiveness of their method, following prior works[3, 24, 54, 55]. Higher PSNR and SSIM values generally reflect better performance in image restoration tasks.

4.3.3. Ensemble Strategies

Ensemble learning has been proven to be an effective technique in image restoration. Its most basic application involves integrating the outputs of multiple models and applying a fusion strategy to achieve results with better generalization and greater stability in restoration quality.

The HighREV-test dataset consists of four sequences. Among them, one is an outdoor scene, which differs markedly from the other three in terms of object diversity, texture richness, and color composition. Based on this observation, they explore a sequence-level ensemble strategy that selectively exchanges outputs between Video-SFHformer and EFSformer.

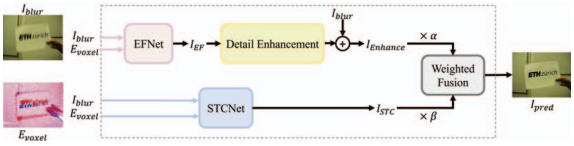


Figure 4. An overview of the method proposed by Team BUPTMM: They set the weights for the fusion, with α set to 0.6 and β to 0.4.

Specifically, they start with the best-performing Video-SFHformer model and replace the output of the outdoor sequence in the HighREV-test set with the corresponding result generated by EFSformer. The results in Table 1 show that their approach yields the best performance, achieving the highest SSIM score and ranking third overall in the NTIRE Event-Based Image Deblurring Challenge.

4.4. Give_it_a_try

4.4.1. General method

This submission is mainly based on the public code of another team. Models used in this submission are EFNet att track fusion and EFNet att track fusion new, which can be found at archs or archs/tested. They change the training strategy, finetune the models and combine two best models to push the limits of scoring.

- How event modality is utilized in the deblurring process: They used the given SCER format event voxels in training, validating and training. The usage is as same as original EFNet [39] since new networks retain the encoder module of the baseline.

4.4.2. Implementation details

- Training:
  In the first stage of training, all models are trained for 2×10^5 iterations with a batch size of 16 by PSNR loss function with AdamW optimizer. In each training batch, each paired images and event voxel are randomly cropped to 256×256 and augmented by random flipping and rotation. The learning rate is initialized as 3×10^{-4}, and a cosine annealing scheduler is used to drop the final learning rate as 10^{-7}. They finetuned the models obtained from the first stage with a patch size of 512×512. At this stage, all models are trained for another 2×10^5 iterations with a batch size of 4 and the learning rate drop from 2×10^{-5} to 10^{-6}. Models are validated for every 10^4 iterations. Other settings remain unchanged.
- Validating and Testing:
  They chose the highest validated models for each network during the fine-tuning stage and average two models' output as final result to improve robustness.

4.5. BUPTMM

4.5.1. Architecture

Our solution is built on EFNet[39] and STCNet[52]. Inspired by [50], they introduce a detail enhancement module that follows the EFNet prediction stage. The whole pipeline is illustrated in Fig. 4. The detail enhancement module adopts a simple U-Net structure.

4.5.2. Implementation Details

Both EFNet and STCNet are initialized with pre-trained GoPro checkpoints. They fine-tune them separately using the NTIRE official training dataset without additional data, aside from the pre-trained GoPro weights. The patch size is set to 1024×1024, and they employ the CosineAnnealingLR scheduler to adjust the learning rate.

The key differences in the training strategies for EFNet and STCNet are as follows:

For EFNet, they train EFNet for 100k iterations with a batch size of 4 using 4 NVIDIA H800 GPUs. The optimizer is AdamW with an initial learning rate of 2e-4. They generate the event voxel grid following the official script, setting the bin size to 24. Due to differences in the event encoder's channel size, they extended the pre-trained GoPro checkpoint weights from 6 to 24 bins. The loss function consists of the L1 loss, the Charbonnier loss, and the Sobel loss, with respective weights of 1.0, 0.5, and 0.5. Unlike the official EFNet implementation, they do not apply a mask between the two stages.

ForNet, they train STCNet for 1000 epochs with a batch size of 8 using 4 NVIDIA H800 GPUs. The optimizer is Adam with an initial learning rate of 2e-4. They use the official event voxel grid with a bin size of 6. The loss function is the Charbonnier loss.

4.6. WEI

Since REFID [40] is an excellent method of event-based blurry video frame interpolation (VFI), considering the differences in modeling image deblurring and VFI problems, they adapt the REFID structure to fit the image deblurring challenge. As shown in Fig. 5, they develop a Bi-directional Gathered Recurrent Network (BGRN) for event-based image deblurring.

4.6.1. Network Architecture

Following REFID [40], the events within the exposure time ($t - \Delta t \rightarrow t + \Delta t$) are represented as a voxel grid $V_{t-\Delta t \rightarrow t+\Delta t} \in \mathbb{R}^{(M+1)\times H \times W}$, where M is set to 9. Furthermore, they divide the voxel $V_{t-\Delta t \rightarrow t+\Delta t}$ into two segments $V_{t-\Delta t \rightarrow t}$ and $V_{t+\Delta t \rightarrow t}$ to perform forward and backward iterations, respectively.

The BGRN consists of image and event branches. Only a blurry image B_t is fed into the image branch, and the

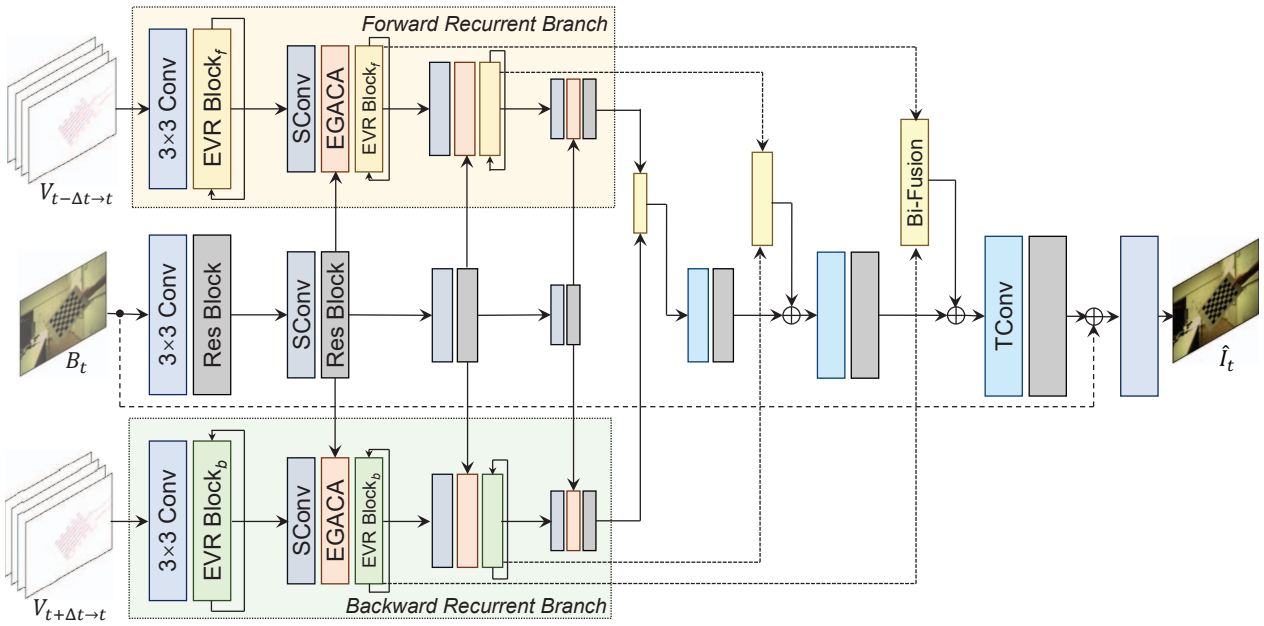


Figure 5. The architecture of the Bi-directional Gathered Recurrent Network (BGRN), proposed by Team Wei, is designed for event-based image deblurring and serves as an enhanced reconfiguration network for REFID. [40]. "EVR Block": event recurrent block [40], "EGACA": event-guided adaptive channel attention [40], "SConv": strided convolution, "TConv": transposed convolution, "Bi-Fusion": bidirectional fusion.

network output is the corresponding sharp image $\hat{I}_t$. Besides, they split the original event branch into a forward recurrent branch and a backward recurrent branch, which respectively and recurrently consumes sub-voxels of forward event voxel $V_{t-\Delta t\to t}$ and backward event voxel $V_{t+\Delta t\to t}$ in a gathered way. In each recurrent iteration, the sub-voxel $V_{sub} \in \mathbb{R}^{2\times H\times W}$ is fed to the event branch, which encodes the event information for the latent frame. To fuse the features obtained from forward and backward recurrent branching, the outputs of both directions are fed into a channel cascade and 1×1 convolution at each scale ("Bi-Fusion" in Fig. 5). Then, they are added element by element with the features of the corresponding scale of the decoder. In addition, to reduce redundancy, they removed the recurrent structure of the decoder section and replaced it with residual blocks. Finally, to make the network learn high-frequency information, the output of the last residual block and the initial features of the blurred image are added element by element, and then the sharp image $\hat{I}_t$ is obtained through a 3×3 convolution.

4.6.2. Implementation details

Training strategy. They train BGRN with the HighREV training dataset specified by the organizer with a batch size of 4 for 200k iterations on an NVIDIA GeForce RTX 3090 GPU. They crop the input images and event voxels to 256×256 for training and use horizontal and vertical flips for data enhancement. AdamW [29] with an initial learning rate of 2×10^{-4} and a cosine learning rate annealing strategy with 1×10^{-7} as the minimum learning rate are adopted for optimization. They use a PSNR loss [39] as supervision.

Ensemble strategy. During testing, they found that images prefixed with "zigzag" showed a large difference in brightness compared to other normal images. To adapt to this sudden change in brightness, they select images with the prefix "sternwatz_window" similar to this scene from the training set. Then, they double their brightness to fine-tune the pre-trained BGRN model for 5k iterations with an initial learning rate of 2×10^{-5}. Therefore, the ensemble strategy is applied when testing, *i.e.*, the abnormally bright images (prefixed with "zigzag") are processed with the fine-tuned model, and the others are processed with the initial pre-trained model.

4.7. DVS-WHU

4.7.1. Network Architecture

Positioned at Fig. 6, the proposed Dual Channel Cross-modal Mamba (DCCM) architecture comprises three primary components: two Shallow Feature Extraction (SFE) modules, a series of N dual channel blocks (with $N = 20$ in their experimental configuration), each containing two Residual Dense Blocks (RDB) [57] and two Cross Modal Mamba (CMM) [14] blocks, and a Global Feature Fusion (GFF) module. Initially, both blur image and events (represented in 24-bin voxel grids) are processed through the SFE module for preliminary feature extraction. Subsequently,

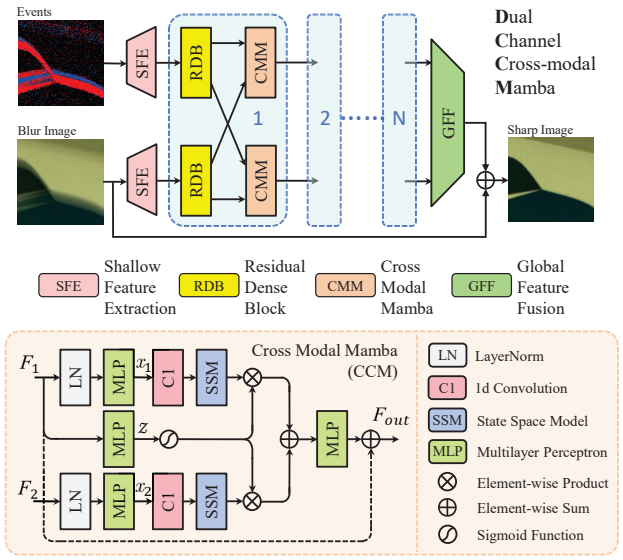


Figure 6. Architecture of DCCM, proposed by Team DVS-WHU.

the dual channel blocks facilitate in-depth feature extraction and cross-modal interaction. Finally, the GFF module synthesizes the ultimate latent sharp image.

The core concept of their network is to establish a mutual compensatory relationship between the features derived from event data and those from blurred images through a dual-channel framework. Specifically, while event data are often characterized by significant noise, images typically exhibit lower noise levels. The CMM block is employed to incorporate image features into the event data, thereby mitigating the noise present in the events. Conversely, event data are rich in sharp edge information, and the CMM block also facilitates the integration of event features into blurred images, ultimately contributing to the deblurred result.

4.7.2. Implementation Details

The network is created with PyTorch and trained on two NVIDIA GeForce RTX 3090 GPUs for 150 epochs with ground-truth-guided L1 norm loss. The training process is composed of two phases. During the first phase, they follow the strategy of Cheng et al.[6] and pretrain their DCCM on the mixed dataset including synthetic REDS dataset[35] and semi-syhthetic HQF dataset[38] with a learning rate fixed at 1×10^{-4} for 50 epochs. In the second phase, the network is fine-tuned on the HighREV dataset[40] where the images are randomly cropped into 256×256 patches with horizontal flipping for data augmentation and the learning rate linearly decays to 1×10^{-5} until the 150th epoch.

4.8. PixelRevive

The model they used was the same as the EFNet[39]. The key to the improved performance of their model lied in the utilization of additional datasets during training and the adoption of larger image sizes in the final fine-tuning phase. They employed a two-stage training strategy. First, they used an Events Simulator called V2E[15] to generate Events from REDS dataset. To generate the dataset, they used timestamp resolution as 0.001, dvs exposure duration as 0.001. The remaining parameters were configured identical to those specified in the V2E paper. They get over 20,000 pairs of events, blur images and sharp images. They trained the model on REDS for 250,000 iters, with gt_size 256, patch size 8. When training on simulated datasets with the HighREV validation set, they observed a paradoxical divergence: while the training PSNR consistently improved, the validation PSNR exhibited a decline. This counterintuitive phenomenon may stem from distributional discrepancies between synthetic data and HighREV characteristics across multiple feature dimensions.

Then, they finetuned it on HighREV train dataset for 200,000 iters, with gt_size 512, patch size 8. The TrueCosineAnnealingLR scheduler was employed in both training phases, configured with a period matching the total training iterations and a minimum learning rate value of 1e-7. After experiments, they found that larger gt_size can improve the PSNR by about 0.5. Experiments showed performance decreases when gt_size exceeds 512 (tested range: 256-608), making 512 the optimal size. Other strategy is same as the EFNet.

4.9. CHD

As illustrated in Fig. 7, team CHD develops an efficient Event-Image Deblurformer Network (EIDFNet) based on the Restormer architecture [55]. To address the computational bottleneck encountered when restoring high-resolution blurry images using event data, they incorporate key design elements from EFNet [39].

4.9.1. Network Architecture

Considering the speed of model training, they still used the official 6-channel voxel grid event representation to achieve a balance between efficiency and precision. They input the blurred image and the event representation with consistent spatial resolution into the network and employ the modified Transformer Block to fuse the cross-modal feature. Firstly, they modify the transformer block in Restormer [55] as a fusion module to achieve full interaction between different feature channels by setting the number of input and output dims in the GDFN and adding 1×1 convolution in the residual connections. Additionally, they build a mutually enhanced fusion encoder based on the Event-Image Cross-Modal Attention Fusion Module (EICA) proposed in EFNet [39]. The enhanced image features are obtained using K and V derived from event embeddings, while Q is sourced from image embeddings. Conversely, the enhanced event features are generated with K and V originating from image embeddings, with Q being drawn from event embeddings.

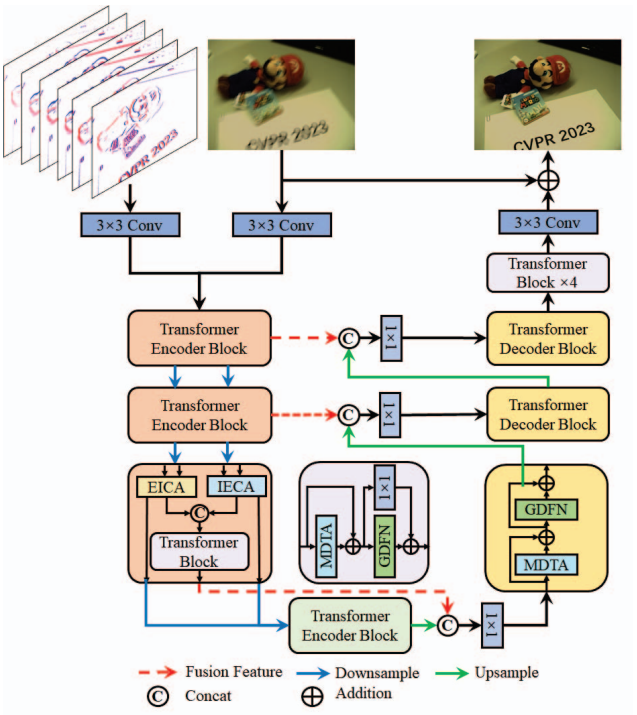


Figure 7. The framework of Event-Image Deblurformer Network (EIDFNet), proposed by Team CHD.

In order to achieve comprehensive integration of event and image features, the enhanced image features and enhanced event features are concatenated along the channel dimension. Subsequently, these concatenated features are fused using a Modified Transformer Block. Ultimately, each encoder produces enhanced image features, enhanced event features, and fused features. The enhanced event and image features undergo downsampling before being input into the subsequent encoder. The fusion feature is directly linked to the corresponding decoding feature through a skip connection.

4.9.2. Training Strategy

They perform progressive learning stategy fllow the settings in Restormer [55] and trained the model on a A100 GPU with L1 loss. The network is trained on smaller image patches in the early epochs and on gradually larger patches in the later training epochs. During the training process, the batch sizes are [4,3,2,2,1,1], and the patch sizes are [128,160,192,256,320,384] with the iterations are [92000,64000,48000,36000,36000,24000]. They employ the AdamW optimizer with an initial learning rate 3e-4 that follows a CosineAnnealingRestartCyclicLR decay strategy.

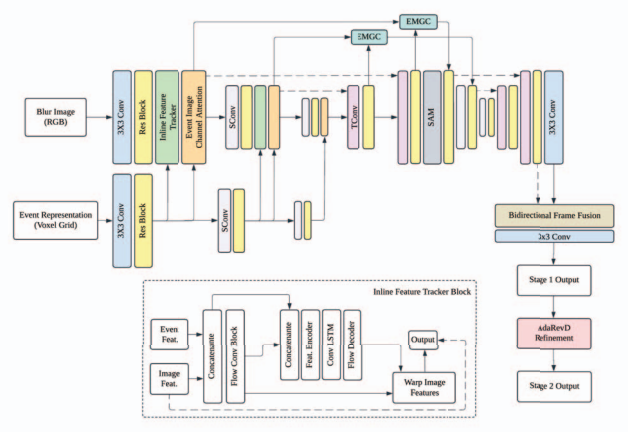


Figure 8. Overview of the proposed pipeline by Team SMU.

4.10. SMU

4.10.1. Motivation

Inspired by recent successes in cross-knowledge sharing between events and RGB frames [39], hierarchical temporal and frequency modelling [18, 40] and stage-wise fine-fusion [20] for the task of event-based RGB deblurring, they propose to modify the base EFNet model [39] such that the the modified model serves as a unified framework which (1) iteratively fine-tunes the coarser deblurred images through two stages of extensive fine-fusion to combat the insufficiencies of the existing decoding techniques while (2) can optionally be made to be specifically aware of propagated frequency information in latent representations to locally and globally filter the blur features in the RGB images through leveraging event features in the frequency domain.

In addition, to the best knowledge, none of the existing methods for event-based RGB deblurring recognizes the importance of feature tracking in this task which can be beneficial especially in challenging conditions such as high contrast (i.e. very bright or dark surroundings) and fast motion (i.e., large pixel displacements within an accumulated event volume) scenarios [33] towards robust performance. To address this limitation, they explicitly employ a data-driven feature tracking module in the pipeline, an inline feature tracker block, such that event feature tracks corresponding to different points in the reference RGB frame are intuitively incorporated in the learning process specifically in the initial stages of the unified framework.

4.10.2. Network Architecture

As depicted in Fig. 8, they propose three main modifications: the inline feature tracker module, bidirectional frame fusion and AdaRevD refinement, to the original EFNet, backed by the motivation as described in section 4.10.1 and validated through the experiments. To this end, they design the inline feature tracker such that the latent RGB and event

features are merged and learned through a flow autoencoder block in combination with a Conv-LSTM block to retrieve the temporal alignment of features. Furthermore, it is to be noted that they place the tracker at an initial stage of the pipeline to ensure that the tracker has the access to the high-level features of each modality, rather than the deeper low-level features, since high-level features, which are close to the input data, are more promising to contain information on temporal propagation, which is critical for co-aligned feature tracking.

Inspired by [20], they design the first stage of refinement using a bidirectional frame fusion block, specifically targeting the spatiotemporal information flow between adjacent coarse frames while in the second stage of refinement, they further refine the output from the first refinement stage with an objective to identify the still remaining degradation patterns in the RGB space and tackle them using an adaptive patch exiting reversible decoder module [31]. Optionally, to implement the frequency-based filtering of blur features, they follow the cross-modal frequency (CMF) module proposed by [18] such that latent representations at each level of the first U-Net are passed through CMF modules, and concatenated in the decoder levels, in a hierarchical fashion to enhance the latent feature representations with frequency-aware characteristics.

4.10.3. Implementation Details

They train the models using one NVIDIA 3090 GPU machine in two stages: (1) primary event-RGB fusion pipeline including the proposed frequency-aware module, explicit feature tracking and the first iteration of refinement based on the bidirectional frame fusion block and (2) second iteration of refinement based on AdaRevD framework [31].

By following the baseline implementation [39], they train the models on the HighREV dataset, in both stages, with an initial learning rate of 2×10^{-4} for a total of 2×10^{4} iterations. The utilized optimizer is AdamW [29] and the learning objective is set to be PSNR loss [39].

4.11. JNU620

As shown in Fig. 9, their framework adopts EFNet [39] as the baseline architecture. To enchance frequency-aware feature processing, a selection frequency block (SF Block) [9] is integrated following each decoder. The architecture introduces two key components: 1) A multi-branch dynamic selection frequency (MDSF) module that adaptively decouples feature mappings into distinct frequency components through dynamic convolution operations; 2) A multi-branch compact selection frequency (MCSF) module specifically designed to expand the receptive field for processing degraded blurry images. Multiple data augmentation strategies were employed, including horizontal and vertical shiftings. For data preparation, they implemented multiple augmentation strategies including horizontal and vertical spatial shifts. The model was trained for 120,000 iterations on an NVIDIA GeForce RTX 3090 GPU with a batch size of 4. The models were optimized by the Adam method with β_1 = 0.9 and β_2 = 0.99 and the weight decay was set to 10^{-4}. The initial learning rate was set to 2×10^{-4}, gradually decreased following a cosine annealing schedule. In inference phase, each test image undergoes augmentation through horizontal and vertival flips before input into the model. The final restored image is generated by averaging all augmented outputs.

4.12. colab

Our team proposes an improved method based on EFNet, named DEFNet (Dynamic Enhanced Fusion Network). This method incorporates three key enhancements. First, we introduce a multi-scale dynamic fusion module, which fuses event and image features at multiple spatial resolutions, significantly improving the restoration of fine details in blurred areas[17]. Second, we enhance the original EICA module by integrating a bidirectional attention mechanism, enabling more effective mutual guidance and interaction between image and event features. Third, for processing event data, we adopt a weighted interpolation strategy[40] that models the dynamic weighting of event sequences more accurately, thereby enriching the temporal details provided to the image restoration process.

4.12.1. Network

Fig. 10 presents the architecture of DEFNet, which is built upon EFNet and incorporates the newly introduced modules: the multi-scale dynamic fusion module and the enhanced EICA module with a bidirectional attention mechanism. These components work collaboratively to optimize the motion deblurring process by improving feature representation and fusion between the image and event data.

During the deblurring process, event streams are used to provide fine-grained temporal variation information that guides the restoration of motion blur in image frames. Specifically, the Symmetric Cumulative Event Representation (SCER) encodes the temporal distribution of events while the enhanced Event-Image Cross-modal Attention Fusion (EICA) module leverages bidirectional attention to facilitate deeper interaction between modalities. Additionally, the integration of weighted interpolation improves the temporal alignment and accuracy of event feature extraction.Together, these components enable DEFNet to more effectively restore motion-blurred images by enhancing edge sharpness, preserving texture, and capturing motion dynamics with higher fidelity.

4.12.2. Implementation Details

We use the AdamW optimizer with an initial learning rate of 2e-4, weight decay of 1e-4, and betas set to [0.9, 0.99].

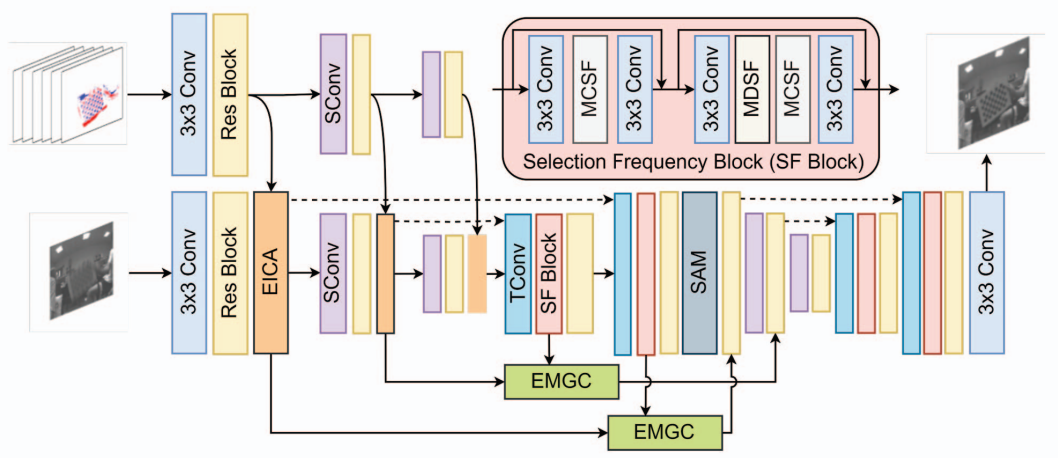


Figure 9. The model framework proposed by Team JNU620.

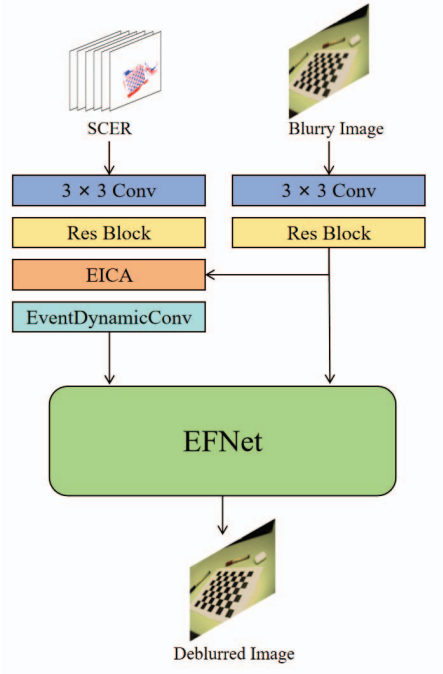


Figure 10. DEFNet architecture, proposed by Team colab.

To dynamically adjust the learning rate, we used the TrueCosineAnnealingLR scheduler with a maximum iteration count of T_max = 200000 and a minimum learning rate of 1e-7. During training, the batch size was set to 4, and 3 worker threads were used per GPU. The total number of training iterations was set to 40000.This method was trained and validated on the HighREV dataset. The model achieved significant improvements on both the training and validation sets, with PSNR and SSIM used as evaluation metrics during training. Validation was performed every 10,000 iterations, and the model was regularly saved.

4.13. CMSL

The Cascade Event Debluring Model With Event Edge Loss was build based on EFNet [39]. An motion edge loss and a cascade framework were introduced to enhance the performance of EFNet.

The EFNet backbone was adopted and two improvements were proposed. Firstly, the event data were organized and represented as voxel [39]. Then, two frame of the event voxels that were most close to the center of the exposure time were multiplied to produce a motion edge frame. The motion edge frame contains the edge of the moving objects in the current frame as shown in fig. 11, fig. 12 is the corresponding edge of the groud truth image (sharp image). As shown in fig. 11 and fig. 12, the motion edge contains clear lines that were consistent with the true edges and could served as a guiding information for image debluring. The edge of the deblured image output by the module should be similar to the motion edge. Therefore, a motion edge loss were proposed as follow:

$$\ell_{edge} = \mathrm{mse}(edge(\widehat{x}) \cdot m, e)$$

$$m_{i,j} = 1 \quad \text{if} \quad e_{i,j} > \tau, \quad \text{else} \quad 0$$

where mse(A,B) is the mean squared error between each element in matrix A and B, $\widehat{x}$ is the output deblured image, e is the motion edge frame, m is the motion edge mask, τ is the threthold parameter.

Secondly, a cascade frame work were proposed that two EFNet was connected in cascade to further enhance the image deblurring ability. The first EFNet took four frames of the event voxels that were relatively remote to the center of the exposure time while the second EFNet took two frames of the event voxels that were relatively close to the center of the exposure time. The two EFNet form a coarse-fine paradigm that gradually remove the motion delur.

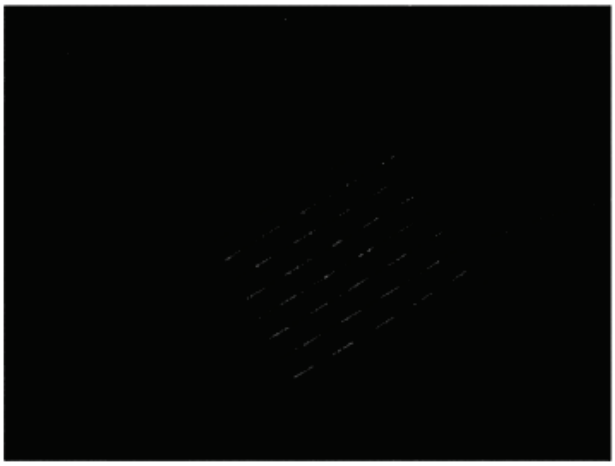


Figure 11. The visualization of the motion edges.



Figure 12. The edges in the ground truth frame

4.14. KUnet

4.14.1. Architecture

Their solution is built upon a custom KUnet backbone tailored for event-based image deblurring. The model employs a dual-encoder strategy that separately processes RGB images and voxelized event data, each through a dedicated encoder branch. At the bottleneck, the features are fused via channel-wise concatenation and passed through a transformer module.

A key novelty in the design is the use of **KANLinear** layers within the transformer block. These layers, based on spline-interpolated kernels, improve attention expressiveness without adding significant computational overhead. This fusion architecture leverages the temporal sharpness of events with the spatial-semantic richness of RGB images to produce high-fidelity deblurred outputs.

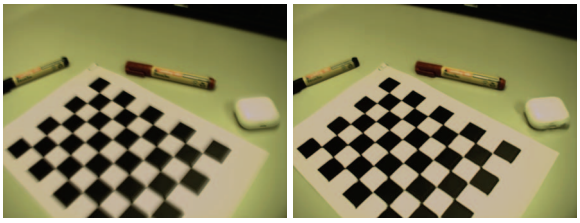


Figure 13. Left: Input blurry frame. Right: output of KUnet, with detailed texture.

4.14.2. Implementation Details

They train the model from scratch on the official NTIRE 2025 HighREV dataset without any external data or pretrained weights. The voxelized events are represented using 6 temporal bins, generating a 6-channel input tensor for the event encoder.

Training was conducted using 2 NVIDIA A100 GPUs with a batch size of 8 and a patch size of 256×256. They trained the network for 150k iterations using the AdamW optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.99$, weight decay = 1e-4) and a CosineAnnealingLR scheduler . Data augmentations included random horizontal flips and rotations.

The loss function includes a PSNR loss weighted at 0.5. Their final checkpoint achieved a peak PSNR of **29.42** on the NTIRE 2025 validation phase.

Inference was performed using a sliding window approach with a max minibatch size of 8. They observed an inference time of $\sim$ 0.15 seconds per frame on an A100 GPU, and a memory footprint of approximately 16 GB during training.

Model Complexity:

- Parameters: 11M
- FLOPs: Not computed
- GPU Memory Usage: 16 GB (training)
- Inference Time: 0.15s/frame

Code and Resources:

- GitHub: https://github.com/Splendor73/NTIRE2025_EventDeblur_challenge_asu
- Pretrained: https://www.dropbox.com/scl/fi/l9td2xtbzxed2bg8tc9w0/17_KUnet.zip
- Results: https://www.dropbox.com/scl/fi/yrky29x2mdwt3k8e40yol/Results.zip

4.15. Group10

The solution is built upon a custom adaptation of the EFNet deblurring framework[39]. The method strategically harnesses both conventional image data and event-based information to mitigate motion blur effectively. Key components of the approach include:

Dual-Stream Network Architecture: The model consists of parallel convolutional streams. One stream processes the blurry input image, while the other processes event data,

which is converted into a voxel grid representation. A cross-modal attention module subsequently fuses the features extracted from both modalities, enhancing the network's ability to recover fine details in dynamic scenes.

Event Data Representation: The raw event data – comprising spatial coordinates, timestamps, and polarity – is transformed into a voxel grid. This process involves temporal normalization and spatial mapping, enabling the network to capture the dynamic nature of motion events with high precision.

Training Strategy: Utilizing mixed precision training to maximize GPU efficiency and accelerate the convergence process. Gradient accumulation is employed to effectively simulate a larger batch size, which is critical for stable training on high-resolution data. The training loss is computed using the Mean Squared Error (MSE) criterion, guiding the network to produce high-quality deblurred images.

Data Pipeline: Custom PyTorch Dataset classes handle the loading and preprocessing of both image and event data. The pipeline includes resizing, normalization, and careful synchronization between blurry images and their corresponding event data, ensuring data consistency across modalities.

Performance Evaluation: The evaluation strategy employs widely accepted metrics such as PSNR and SSIM to quantify restoration quality. Test outputs are resized to their original dimensions and saved as lossless PNG images to preserve the fidelity of the results.

Additional details include:

Parameter Count: The EnhancedEFNet model consists of convolutional layers, CrossModalAttention blocks, and skip connections, leading to a parameter count in the range of millions.

CrossModalAttention layers: These layers introduce additional tensor operations and memory usage. No external pre-trained models were directly used in training.The architecture was trained from scratch on the provided dataset.

GPU Memory Usage: Memory usage is influenced by Batch Size, Default batch size of 4 per GPU, and Voxel Grid Representation, Uses 6 event bins, increasing input size.

CrossModalAttention: Inspired by self-attention mechanisms in Transformer models. Hybrid Loss Function: Combines MSE and L1 loss for better generalization.CosineAnnealingLR Scheduler: Used to dynamically adjust learning rates during training.

Use of Additional Training Data: Only NTIRE Dataset Used:The training was restricted to the HighREV dataset provided by NTIRE. No additional synthetic or external event-based datasets were incorporated. Potential Future Enhancements: Using real-world event datasets (e.g., DSEC, MVSEC) could improve generalization. Fine-tuning with pre-trained image restoration models (like DeblurGAN) could be explored.

Quantitative and Qualitative Improvements Quantitative Improvements (Metrics & Performance): Peak Signal-to-Noise Ratio (PSNR): Achieved PSNR: 25.93.Improved compared to baseline event fusion models. Structural Similarity Index (SSIM): Achieved SSIM: 0.82. Indicates better perceptual quality in restored images. Qualitative Improvements (Visual Results & Generalization): Better Detail Recovery: The attention-based fusion of events and images leads to sharper edges and better contrast in reconstructed images. Works well in low-light or high-motion blur scenarios.

Comparison with Baseline Models: Standard CNN-based deblurring struggles with fine-grained event details, but EnhancedEFNet effectively fuses event features to improve deblurring accuracy. CrossModalAttention aids in spatial alignment of events and images, reducing artifacts. Failure Cases & Future Improvements: Highly blurred images with saturated event data can still cause artifacts. More robust fusion mechanisms (e.g., transformer-based approaches) could further enhance performance.

Acknowledgments

This work was partially supported by the Humboldt Foundation, the Ministry of Education and Science of Bulgaria (support for INSAIT, part of the Bulgarian National Roadmap for Research Infrastructure). Shaolin Su was supported by the HORIZON MSCA Postdoctoral Fellowships funded by the European Union (project number 101152858). We thank the NTIRE 2025 sponsors: ByteDance, Meituan, Kuaishou, and University of Wurzburg (Computer Vision Lab).

A. Teams and affiliations

NTIRE 2025 team

Title: NTIRE 2025 Event-Based Image Deblurring Challenge

Members:

Lei Sun[1] (leo_sun@zju.edu.cn),
Andrea Alfarano[1] (andrea.alfarano@insait.ai),
Peiqi Duan[2] (duanqi0001@pku.edu.cn),
Shaolin Su[3] (shaolin@cvc.uab.cat),
Kaiwei Wang[4] (wangkaiwei@zju.edu.cn),
Boxin Shi[2] (shiboxin@pku.edu.cn),
Radu Timofte[5] (radu.timofte@uni-wuerzburg.de)
Danda Pani Paudel[1] (danda.paudel@insait.ai),
Luc Van Gool[1] (vangool@vision.ee.ethz.ch),

Affiliations:

[1] INSAIT, Sofia University "St. Kliment Ohridski", Bulgaria
[2] Peking University, China
[3] Computer Vision Center, Spain

[4] Zhejiang University, China
[5] University of Würzburg, Germany

IVISLAB

Title: Triple Event-stream Image Deblurring Network
Members:
Qinglin Liu[1] (qlliu@hit.edu.cn), Wei Yu[2], Xiaoqian Lv[1], Lu Yang[3], Shuigen Wang[3], Shengping Zhang[1], Xiangyang Ji[2]
Affiliations:
[1] Harbin Institute of Technology, Weihai
[2] Tsinghua University
[3] Raytron Technology Co., Ltd.

MiVideoDeblur

Title: Event-Based Image Deblurring from Team MiVideoDeblur
Members:
Long Bao[1] (baolong@xiaomi.com), Yuqiang Yang[1], Jinao Song[1], Ziyi Wang[1], Shuang Wen[1], Heng Sun[1]
Affiliations:
[1] Xiaomi Inc., China

404NotFound

Title: Event-Conditioned Dual-Modal Fusion for Motion Deblurring
Members:
Kean Liu[1] (rickyliu@mail.ustc.edu.cn), Mingchen Zhong[1], Senyan Xu[1], Zhijing Sun[1], Jiaying Zhu[1], Chengjie Ge[1], Xingbo Wang[1], Yidi Liu[1], Xin Lu[1], Xueyang Fu[1], Zheng-Jun Zha[1]
Affiliations:
[1] University of Science and Technology of China

Give_it_a_try

Title: Event-Based Image Deblurring from Team Give_it_a_try
Members:
Dawei Fan[1] (dawei.fan@partner.samsung.com), Dafeng Zhang[1], Yong Yang[1]
Affiliations:
[1] Samsung Research China- Beijing (SRC-B)

BUPTMM

Title: Weighted Fusion for Event-based Image Deblurring
Members:
Siru Zhang[1] (zhangsr@bupt.edu.cn), Qinghua Yang[1], Hao Kang[1], Huiyuan Fu[1], Heng Zhang[2], Hongyuan Yu[2], Zhijuan Huang[2]
Affiliations:
[1] Beijing University of Posts and Telecommunications, Beijing, China.
[2] Xiaomi Inc., China.

WEI

Title: Bi-directional Gathered Recurrent Network for Event-based Image Deblurring
Members:
Shuoyan Wei[1] (shuoyan.wei@bjtu.edu.cn),
Feng Li[2], Runmin Cong[3]
Affiliations:
[1] Institute of Information Science, Beijing Jiaotong University
[2] School of Computer Science and Engineering, Hefei University of Technology
[3] School of Control Science and Engineering, Shandong University

DVS-WHU

Title: Dual Channel Cross-modal Mamba for Event-based Motion Deblurring
Members:
Weiqi Luo[1] (wikyluo@whu.edu.cn),
Mingyun Lin[1], Chenxu Jiang[1], Hongyi Liu[1], Lei Yu[2]
Affiliations:
[1] School of Electronic Information, Wuhan University
[2] School of Artificial Intelligence, Wuhan University

PixelRevive

Title: Event-Based Image Deblurring from Team PixelRevive
Members:
Weilun Li[1] (xyj961011@163.com), Jiajun Zhai[1], Tingting Lin[1]
Affiliations:
[1] College of Optical Science and Engineering, Zhejiang University

CHD

Title: Event-Image Deblurformer Network
Members:
Shuang Ma[1] (3125508679@qq.com), Sai Zhou[2], Zhanwen Liu[3], Yang Wang[4]
Affiliations:
[1] Chang'an University, Xi'an, China

SMU

Title: Explicit Feature Tracking and Iterative Refinement for Enhancing Event-based Image Deblurring
Members:
Eiffel Chong[1], Nuwan Bandara[1], Thivya Kandappu[1] (thivyak@smu.edu.sg), Archan Misra[1]
Affiliations:
[1] Singapore Management University

JNU620

Title: Event-Based Image Deblurring from Team JNU620
Members:
Yihang Chen[1] (Ehang@stu.jnu.edu.cn),
Zhan Li[1], Weijun Yuan[1], Wenzhuo Wang[1], Boyang Yao[1], Zhanglu Chen[1]
Affiliations:
[1] Department of Computer Science, Jinan University, Guangzhou, China

colab

Title: Dynamic Enhanced Fusion Network for Event-based Image Deblurring
Members:
Yijing Sun[1] (syj3508852939@163.com), Tianjiao Wan[1], Zijian Gao[1], Qisheng Xu[1], Kele Xu[1]
Affiliations:
[1] National University of Defense Technology

CMSL

Title: Cascade Event Debluring Model With Event Edge Loss
Members:
Yukun Zhang[1] (zhangyukun@cmhi.chinamobile.com), Yu He[1], Xiaoyan Xie[1], Tao Fu[1]
Affiliations:
[1] China Mobile (Hangzhou) Information Technology Co., Ltd, Hangzhou, China

KUnet

Title KUnet
Members:
Yashu Gautamkumar Patel[1] (ypatel37@asu.edu), Vihar Ramesh Jain[1], Divesh Basina[1]
Affiliations:
[1] Arizona State University

Group10

Title: Event-Based Image Deblurring from Team Group10
Members:
Rishik Ashili[1] (rishik67_soe@jnu.ac.in), Manish Kumar Manjhi[1], Sourav Kumar[1], Prinon Benny[1], Himanshu Ghunawat[1], B Sri Sairam Gautam[1], Anett Varghese[1], Abhishek Yadav[1]
Affiliations:
[1] Jawaharlal Nehru University, New Delhi, India

References

[1] Inigo Alonso and Ana C Murillo. Ev-segnet: Semantic segmentation for event-based cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019. 1

[2] Jiaan Chen, Hao Shi, Yaozu Ye, Kailun Yang, Lei Sun, and Kaiwei Wang. Efficient human pose estimation via 3d event point cloud. In *2022 International Conference on 3D Vision (3DV)*, pages 1–10. IEEE, 2022. 1

[3] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *European Conference on Computer Vision*, pages 17–33. Springer, 2022. 5

[4] Zheng Chen, Kai Liu, Jue Gong, Jingkai Wang, Lei Sun, Zongwei Wu, Radu Timofte, Yulun Zhang, et al. NTIRE 2025 challenge on image super-resolution (×4): Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[5] Zheng Chen, Jingkai Wang, Kai Liu, Jue Gong, Lei Sun, Zongwei Wu, Radu Timofte, Yulun Zhang, et al. NTIRE 2025 challenge on real-world face restoration: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[6] Zhangyi Cheng, Xiang Zhang, Lei Yu, Jianzhuang Liu, Wen Yang, and Gui-Song Xia. Recovering continuous scene dynamics from a single blurry image with events. *arXiv preprint arXiv:2304.02695*, 2023. 8

[7] Marcos Conde, Radu Timofte, et al. NTIRE 2025 challenge on raw image restoration and super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[8] Marcos Conde, Radu Timofte, et al. Raw image reconstruction from RGB on smartphones. NTIRE 2025 challenge report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[9] Yuning Cui, Yi Tao, Zhenshan Bing, Wenqi Ren, Xinwei Gao, Xiaochun Cao, Kai Huang, and Alois Knoll. Selective frequency network for image restoration. In *International Conference on Learning Representations*, 2023. 10

[10] Egor Ershov, Sergey Korchagin, Alexei Khalin, Artyom Panshin, Arseniy Terekhin, Ekaterina Zaychenkova, Georgiy

Lobarev, Vsevolod Plokhotnyuk, Denis Abramov, Elisey Zhdanov, Sofia Dorogova, Yasin Mamedov, Nikola Banic, Georgii Perevozchikov, Radu Timofte, et al. NTIRE 2025 challenge on night photography rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[11] Yuqian Fu, Xingyu Qiu, Bin Ren Yanwei Fu, Radu Timofte, Nicu Sebe, Ming-Hsuan Yang, Luc Van Gool, et al. NTIRE 2025 challenge on cross-domain few-shot object detection: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[12] Guillermo Gallego, Tobi Delbruck, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew Davison, Jörg Conradt, Kostas Daniilidis, and Davide Scaramuzza. Event-based vision: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(1):154–180, 2022. 1

[13] Shuhao Han, Haotian Fan, Fangyuan Kong, Wenjie Liao, Chunle Guo, Chongyi Li, Radu Timofte, et al. NTIRE 2025 challenge on text to image generation model quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[14] Xuanhua He, Ke Cao, Jie Zhang, Keyu Yan, Yingying Wang, Rui Li, Chengjun Xie, Danfeng Hong, and Man Zhou. Pan-mamba: Effective pan-sharpening with state space model. *Information Fusion*, 115:102779, 2025. 7

[15] Yuhuang Hu, Shih-Chii Liu, and Tobi Delbruck. v2e: From video frames to realistic dvs events. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1312–1321, 2021. 8

[16] Varun Jain, Zongwei Wu, Quan Zou, Louis Florentin, Henrik Turbell, Sandeep Siddhartha, Radu Timofte, et al. NTIRE 2025 challenge on video quality enhancement for video conferencing: Datasets, methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[17] J Kim, D K Ghosh, and Y J Jung. Event-based video deblurring based on image and event feature fusion. *Expert Systems with Applications*, 223:119917, 2023. 10

[18] Taewoo Kim, Hoonhee Cho, and Kuk-Jin Yoon. Frequency-aware event-based video deblurring for real-world motion blur. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24966–24976, 2024. 9, 10

[19] Sangmin Lee, Eunpil Park, Angel Canelo, Hyunhee Park, Youngjo Kim, Hyungju Chun, Xin Jin, Chongyi Li, Chun-Le Guo, Radu Timofte, et al. NTIRE 2025 challenge on efficient burst hdr and restoration: Datasets, methods, and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[20] Huan Li, Hailong Shi, and Xingyu Gao. A coarse-to-fine fusion network for event-based image deblurring. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 974–982, 2024. 9, 10

[21] Xin Li, Yeying Jin, Xin Jin, Zongwei Wu, Bingchen Li, Yufei Wang, Wenhan Yang, Yu Li, Zhibo Chen, Bihan Wen, Robby Tan, Radu Timofte, et al. NTIRE 2025 challenge on day and night raindrop removal for dual-focused images: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[22] Xin Li, Xijun Wang, Bingchen Li, Kun Yuan, Yizhen Shao, Suhang Yao, Ming Sun, Chao Zhou, Radu Timofte, and Zhibo Chen. NTIRE 2025 challenge on short-form ugc video quality assessment and enhancement: Kwaisr dataset and study. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[23] Xin Li, Kun Yuan, Bingchen Li, Fengbin Guan, Yizhen Shao, Zihao Yu, Xijun Wang, Yiting Lu, Wei Luo, Suhang Yao, Ming Sun, Chao Zhou, Zhibo Chen, Radu Timofte, et al. NTIRE 2025 challenge on short-form ugc video quality assessment and enhancement: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[24] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1833–1844, 2021. 5

[25] Jie Liang, Radu Timofte, Qiaosi Yi, Zhengqiang Zhang, Shuaizheng Liu, Lingchen Sun, Rongyuan Wu, Xindong Zhang, Hui Zeng, Lei Zhang, et al. NTIRE 2025 the 2nd restore any image model (RAIM) in the wild challenge. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[26] Kean Liu, Mingchen Zhong, Senyan Xu, Zhijing Sun, Jiaying Zhu, Chengjie Ge, Xin Lu, Xingbo Wang, Xueyang Fu, and Zheng-Jun Zha. Event-conditioned dual-modal fusion for motion deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 4

[27] Xiaohong Liu, Xiongkuo Min, Qiang Hu, Xiaoyun Zhang, Jie Guo, et al. NTIRE 2025 XGC quality assessment challenge: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[28] Xiaoning Liu, Zongwei Wu, Florin-Alexandru Vasluianu, Hailong Yan, Bin Ren, Yulun Zhang, Shuhang Gu, Le Zhang, Ce Zhu, Radu Timofte, et al. NTIRE 2025 challenge on low light image enhancement: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[29] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. 7, 10

[30] Xingyu Lu, Lei Sun, Diyang Gu, and Kaiwei Wang. Sge: structured light system based on gray code with an event camera. *Optics Express*, 32(26):46044–46061, 2024. 1

[31] Xintian Mao, Qingli Li, and Yan Wang. Adarevd: Adaptive patch exiting reversible decoder pushes the limit of image deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25681–25690, 2024. 10

[32] 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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 547–557, 2022. 1

[33] Nico Messikommer, Carter Fang, Mathias Gehrig, and Davide Scaramuzza. Data-driven feature tracking for event cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5642–5651, 2023. 9

[34] Manasi Muglikar, Guillermo Gallego, and Davide Scaramuzza. Esl: Event-based structured light. In *2021 International Conference on 3D Vision (3DV)*, pages 1165–1174. IEEE, 2021. 1

[35] Seungjun Nah, Sungyong Baik, Seokil Hong, Gyeongsik Moon, Sanghyun Son, Radu Timofte, and Kyoung Mu Lee. Ntire 2019 challenge on video deblurring and super-resolution: Dataset and study. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition workshops*, pages 1996–2005, 2019. 8

[36] Bin Ren, Hang Guo, Lei Sun, Zongwei Wu, Radu Timofte, Yawei Li, et al. The tenth NTIRE 2025 efficient super-resolution challenge report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[37] Nickolay Safonov, Alexey Bryntsev, Andrey Moskalenko, Dmitry Kulikov, Dmitriy Vatolin, Radu Timofte, et al. NTIRE 2025 challenge on UGC video enhancement: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[38] Timo Stoffregen, Cedric Scheerlinck, Davide Scaramuzza, Tom Drummond, Nick Barnes, Lindsay Kleeman, and Robert Mahony. Reducing the sim-to-real gap for event cameras. In *European Conference on Computer Vision*, pages 534–549, 2020. 8

[39] 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 *European Conference on Computer Vision*, pages 412–428. Springer, 2022. 1, 3, 4, 6, 7, 8, 9, 10, 11, 12

[40] 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, 3, 4, 6, 7, 8, 9, 10

[41] Lei Sun, Daniel Gehrig, Christos Sakaridis, Mathias Gehrig, Jingyun Liang, Peng Sun, Zhijie Xu, Kaiwei Wang, Luc Van Gool, and Davide Scaramuzza. A unified framework for event-based frame interpolation with ad-hoc deblurring in the wild. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 1, 3

[42] Lei Sun, Andrea Alfarano, Peiqi Duan, Shaolin Su, Kaiwei Wang, Boxin Shi, Radu Timofte, Danda Pani Paudel, Luc Van Gool, et al. NTIRE 2025 challenge on event-based image deblurring: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[43] Lei Sun, Yuhan Bao, Jiajun Zhai, Jingyun Liang, Yulun Zhang, Kaiwei Wang, Danda Pani Paudel, and Luc Van Gool. Low-light image enhancement using event-based illumination estimation. *arXiv preprint arXiv:2504.09379*, 2025. 1

[44] Lei Sun, Hang Guo, Bin Ren, Luc Van Gool, Radu Timofte, Yawei Li, et al. The tenth ntire 2025 image denoising challenge report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[45] Zhijing Sun, Xueyang Fu, Longzhuo Huang, Aiping Liu, and Zheng-Jun Zha. Motion aware event representation-driven image deblurring. In *European Conference on Computer Vision*, pages 418–435. Springer, 2024. 4

[46] Florin-Alexandru Vasluianu, Tim Seizinger, Zhuyun Zhou, Cailian Chen, Zongwei Wu, Radu Timofte, et al. NTIRE 2025 image shadow removal challenge report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[47] Florin-Alexandru Vasluianu, Tim Seizinger, Zhuyun Zhou, Zongwei Wu, Radu Timofte, et al. NTIRE 2025 ambient lighting normalization challenge. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[48] Yingqian Wang, Zhengyu Liang, Fengyuan Zhang, Lvli Tian, Longguang Wang, Juncheng Li, Jungang Yang, Radu Timofte, Yulan Guo, et al. NTIRE 2025 challenge on light field image super-resolution: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[49] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 5

[50] Wenming Weng, Yueyi Zhang, and Zhiwei Xiong. Event-based blurry frame interpolation under blind exposure. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1588–1598, 2023. 6

[51] Kangning Yang, Jie Cai, Ling Ouyang, Florin-Alexandru Vasluianu, Radu Timofte, Jiaming Ding, Huiming Sun, Lan Fu, Jinlong Li, Chiu Man Ho, Zibo Meng, et al. NTIRE 2025 challenge on single image reflection removal in the wild: Datasets, methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[52] Wen Yang, Jinjian Wu, Jupo Ma, Leida Li, and Guangming Shi. Motion deblurring via spatial-temporal collaboration of frames and events. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 6531–6539, 2024. 6

[53] Pierluigi Zama Ramirez, Fabio Tosi, Luigi Di Stefano, Radu Timofte, Alex Costanzino, Matteo Poggi, Samuele Salti, Stefano Mattoccia, et al. NTIRE 2025 challenge on hr depth from images of specular and transparent surfaces. In *Pro-*

ceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2025. 2

[54] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for real image restoration and enhancement. In *European Conference on Computer Vision*, pages 492–511. Springer, 2020. 5

[55] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5728–5739, 2022. 4, 5, 8, 9

[56] Shaobo Zhang, Lei Sun, and Kaiwei Wang. A multi-scale recurrent framework for motion segmentation with event camera. *IEEE Access*, 11:80105–80114, 2023. 1

[57] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2472–2481, 2018. 7