NTIRE 2025 Event-Based Image Deblurring Challenge Factsheet
- NTIRE2025 Submission -

Anett Varghese
Jawaharlal Nehru University
New Delhi, India
anettv51_soe@jnu.ac.in

B Sri Sairam Gautam
Jawaharlal Nehru University
New Delhi, India
bsrisa59_soe@jnu.ac.in

Rishik Ashili
Jawaharlal Nehru University
New Delhi, India
rishik67_soe@jnu.ac.in

1. Introduction

Motion blur remains a pervasive challenge in computer vision, particularly within dynamic environments such as autonomous vehicles and robotic systems. Traditional deblurring techniques that rely exclusively on frame-based images often exhibit limitations under rapidly changing conditions. In contrast, event-based cameras capture asynchronous brightness changes at the pixel level, providing a rich temporal stream of data that can significantly enhance image restoration.

Our submission builds on the robust EFNet framework [1, 2] by integrating a sophisticated cross-modal attention mechanism. This mechanism effectively fuses the complementary information contained in conventional blurry frames and event-based data, resulting in superior deblurring performance. The proposed solution is implemented in PyTorch with an optimized training strategy that incorporates mixed precision and gradient accumulation, ensuring efficient utilization of GPU resources while maintaining high fidelity in the restored images.

2. Team Details

- Group10.
- Team Leader: Rishik Ashili, +91 8595330068, rishik67_soe@jnu.ac.in.
- Team Members: Manish Kumar Manjhi, Sourav Kumar, Prinon Benny, Himanshu Ghunawat, B Sri Sairam Gautam, Anett Varghese, Abhishek Yadav .
- Jawaharlal Nehru University
- User names:- anettv51_soe
- Best score : 25.93 (PSNR)
- Repository link: Group10 GitHub Repository

3. Method Details

Our solution is built upon a custom adaptation of the EFNet deblurring framework[1]. The method strategically harnesses both conventional image data and event-based information to mitigate motion blur effectively. Key components of our approach include:

- **Dual-Stream Network Architecture:** Our 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:** Our 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.You already track GPU memory with torch.cuda.memory_allocated() and torch.cuda.memory_reserved(). 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: XX dB (you need to replace with actual results from validation).Improved compared to baseline event fusion models.Structural Similarity Index (SSIM):Achieved SSIM: XX (depends on results).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

References

[1] 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

[2] 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. 1