Event-Assisted Fusion for Improved Object Detection in Motion-Blurred Scenarios

***Abstract*—Object detection models such as YOLOv8 perform well on conventional RGB imagery but experience significant performance degradation under motion blur and rapid camera motion. Event cameras capture asynchronous brightness changes at microsecond latency and naturally provide blur-free edge representations, offering complementary information to RGB frames. This paper exploits the strengths of both modalities to achieve robust object detection in challenging visual conditions. RGB videos are converted into synthetic event streams using V2E and rendered as 2D event frames, which are then combined with RGB data through a lightweight fusion method that operates without modifying the YOLO architecture. We train and evaluate three variants: an RGB-only model, an event-only model, and a fused RGB–event model. Experimental results demonstrate that the fused approach consistently outperforms both baselines, particularly under severe motion blur. The proposed method delivers an efficient, architecture-agnostic strategy for integrating event information into modern detectors, enhancing robustness without additional deblurring modules or network redesign.**

***Index Terms*—Object detection, Event cameras, RGB, Event frames, Fusion**

I. Introduction

Object detection aims to recognize objects in an image and locate them accurately, but when the camera or the objects move quickly, the captured RGB frames often suffer from motion blur. This blur creates streaks and distortions that make object features harder to interpret, reducing detection accuracy as shown in figure 1. Deblurring algorithms can be used, but they are computationally expensive and may discard useful visual details. Event cameras offer a promising alternative because they record brightness changes asynchronously at each pixel, giving them high temporal resolution and strong performance under fast motion or high dynamic range conditions. In this work, we introduce a hybrid object detection approach that combines event data with RGB images to improve detection reliability in motion-blurred scenes.

II. Related work

From paper[1], it has been shown that event-based fall-detection systems can perform reliably even in low-light and privacy-sensitive environments where conventional RGB cameras often fail. The authors also highlight the lack of large, real event-camera datasets and demonstrate that simulators such as V2E can be effectively used to convert standard RGB videos into synthetic event streams, enabling scalable training of event-driven models. Multiple studies like paper[2] confirm that event cameras outperform RGB cameras in HDR and low-light conditions, making them suitable for scenarios where conventional cameras fail due to motion blur or poor illumination.[2] used enhanced YOLOv5 by adding a new small-object detection head tailored for event images, showing that modifying standard RGB-trained architectures improves event-domain detection. Due to the lack of large RGB-event paired datasets, paper[3] have also used simulator-V2E to convert conventional video datasets (ImageNet-VID) into event streams.This paper demonstrates that even lightweight fusion methods—such as alpha-blending or pixel overwrite significantly boost mean Average Precision (mAP) on blurred images without modifying the YOLO architecture. Paper [4] presents an event-based vessel detection framework using Asynet, a sparse-convolution neural network designed for asynchronous event data. The authors collected and annotated a maritime event dataset using a DAVIS346 camera. Results show that simple augmentations like horizontal flipping and cropping significantly improve detection accuracy, highlighting the potential of event cameras.

Fig. 1. Comparison between normal RGB and motion blurred RGB representations.

Therefore, prior research establishes that event cameras offer advantages in environments where conventional RGB cameras fail, particularly under motion blur, low illumination, or adverse weather conditions. Existing work also emphasizes the scarcity of large-scale event datasets and demonstrates the effectiveness of simulators such as V2E for generating synthetic event streams to support training and testing. Although several studies improve detection performance either by modifying RGB-based architectures or by designing event-specific neural networks, recent findings show that even simple fusion strategies can yield substantial gains without architectural redesign. These insights collectively motivate the need for lightweight architecture-agnostic fusion approaches that take advantage of the complementary strengths of RGB frames and event data—especially for object detection under degraded visual conditions.

III. PROPOSED METHOD

A. Data representation

An event camera produces a stream of asynchronous events rather than conventional image frames. Each event is represented as a tuple

$$e = (x, y, t, p),$$

where x, y denote the pixel coordinates, t is the timestamp (in seconds or microseconds), and $p \in \{+1, -1\}$ is the polarity indicating an increase (ON) or decrease (OFF) in the logarithmic intensity. To integrate event data with RGB frames, events are accumulated over a short temporal window $[t_0,\, t_0 + \Delta t]$ to form a 2D event image $E(x, y)$.For each window, we compute a polarity-separated representation consisting of positive and negative event maps, denoted by $E^+(x, y)$ and $E^-(x, y)$. The accumulation uses a constant weighting scheme, where all events within the window contribute equally. The generated maps are then normalized to the range $[0, 1]$ before fusion with RGB frames.

In our experiments, the accumulation window is set to $\Delta t = 30$–50 ms, which aligns with the frame interval of standard 20–30 FPS RGB video. This process yields a pair of event frames for every RGB frame, enabling consistent spatio-temporal alignment during fusion.

B. Dataset

We collected 1470 images from YouTube platform,[5] and [6], which are focused on car detection scenarios and manually framed proper and precise annotations. Each RGB frame was artificially degraded to simulate camera motion, defocus, and low-light distortions. A directional motion-blur augmentation was applied to the dataset. A 25×25 linear kernel was generated and rotated to four angles (0°, 45°, 90°, 135°) using an affine transformation. These kernels were precomputed and normalized to ensure consistent intensity. For each image, a blur direction was randomly selected and applied through convolution using cv2.filter2D(). The resulting blurred images were stored as a separate training set.

To generate synthetic event data from conventional RGB video, we utilized the V2E event simulator configured to emulate a DAVIS346 sensor. The input video (30 fps) was supplied directly without slow-motion interpolation by setting `--disable_slomo` and `--input_slowmotion_factor = 1`, ensuring that the temporal characteristics of the original sequence were preserved. The simulator operated with an exposure duration of 33 ms per frame and used symmetric contrast thresholds of $C_+ = 0.2$ and $C_- = 0.2$ for positive and negative log-intensity changes, respectively. Temporal noise characteristics were modeled via a threshold noise parameter $\sigma_{\text{thres}} = 0.03$, a bandwidth cutoff of 200 Hz, a leak rate of 5.18 Hz, and a shot-noise event rate of 2.716 Hz. The DAVIS output mode was enabled to produce both standard DVS stream representation and an HDF5 file (`events.h5`) containing timestamped $\langle t, x, y, p \rangle$ events.

C. Fusion Method

To combine the RGB frames with the synthetic event stream, we adopt a direct overlay-based fusion procedure. Unlike accumulation-based techniques or alpha-blending strategies, the proposed method maps each event to its corresponding RGB frame through timestamp alignment, spatial scaling, and polarity-based color coding.

1) Temporal Alignment of Events to RGB Frames: Each event is represented as $e_i = (x_i, y_i, t_i, p_i)$, where t_i is the timestamp. Let $t_{\min}$ and $t_{\max}$ denote the minimum and maximum event timestamps. We normalize event times as:

$$t_i' = \frac{t_i - t_{\min}}{t_{\max} - t_{\min}}.$$

Given N RGB frames, each event is assigned to a frame index:

$$f_i = \text{clip}\left(\lfloor t_i' \cdot N \rfloor,\, 0,\, N - 1\right).$$

This ensures that all events fall within the valid temporal range of the video.

2) Spatial Scaling to Match RGB Resolution: The event sensor resolution (W_e, H_e) differs from the RGB resolution (W_r, H_r). Event coordinates (x_e, y_e) are mapped to RGB pixel coordinates via linear scaling:

$$x_r = x_e \cdot \frac{W_r}{W_e}, \qquad y_r = y_e \cdot \frac{H_r}{H_e}.$$

This transformation ensures spatial consistency between the two modalities.

3) Polarity-Based Color Coding: Each event polarity $p_i \in \{+1, -1\}$ is encoded using two fixed colors:

$$\text{ON polarity } (+1):\ \text{Cyan } (255,\, 255,\, 0),$$

$$\text{OFF polarity } (-1):\ \text{Magenta } (255,\, 0,\, 255).$$

For every mapped event pixel (x_r, y_r), the corresponding RGB pixel is replaced with the polarity color.

4) Frame-Wise Fusion Procedure: For each RGB frame, the original image was first copied. All events corresponding to the current frame index were then retrieved, and each event was mapped spatially onto the frame. The pixel values were updated according to the polarity of the events, effectively overlaying the event information onto the RGB image. The resulting fused frame, combining both the original RGB content and the event-based enhancements, was then saved as a standard RGB image. This produces images where event edges are sharply superimposed on the RGB scene.

5) Characteristics of the Fusion Method: The proposed fusion approach exhibits several desirable properties:

- **Deterministic:** No trainable parameters or weighting functions are required.
- **Low-latency:** The fusion consists only of coordinate scaling and pixel replacement.
- **Edge-preserving:** Events highlight motion and intensity changes, improving visual saliency.
- **Detector-compatible:** The fused frames remain in standard RGB format and can be directly used by networks such as YOLO.

The entire fusion pipeline can be understood as shown in figure 2.

IV. Experiment and Results

A. Model Training Setup

All experiments were conducted using the YOLOv8-M model from Ultralytics. Training was performed on Google Colab equipped with an NVIDIA T4 GPU (16 GB VRAM). To study the contribution of each modality, three separate models were trained:

- **RGB-only Model:** Trained exclusively on motion-blurred RGB images generated using the blurring protocol described earlier.
- **Event-only Model:** Trained on 2-D event frame representations synthesized using V2E and preprocessed into polarity-based event images.
- **Fused RGB–Event Model:** Trained on the proposed fusion images where cyan/magenta event pixels are overlaid onto the corresponding RGB frames using the timestamp- and resolution-aligned fusion pipeline.

The training configuration was consistent across all three models (RGB-only, Event-only, and Fused RGB–Event) to ensure fair comparison. Key hyperparameters were:

- **Epochs:** 100
- **Image size:** 640×640
- **Batch size:** 8
- **Optimizer:** AdamW
- **Learning rate:** initial $lr_0 = 0.001$, final $lrf = 0.01$ (cosine LR schedule)
- **Data augmentation:** Enabled

B. Quantitative Comparison

The detection performance of the three models is summarized in Table I. Metrics include mAP@50 and mAP@50-95.

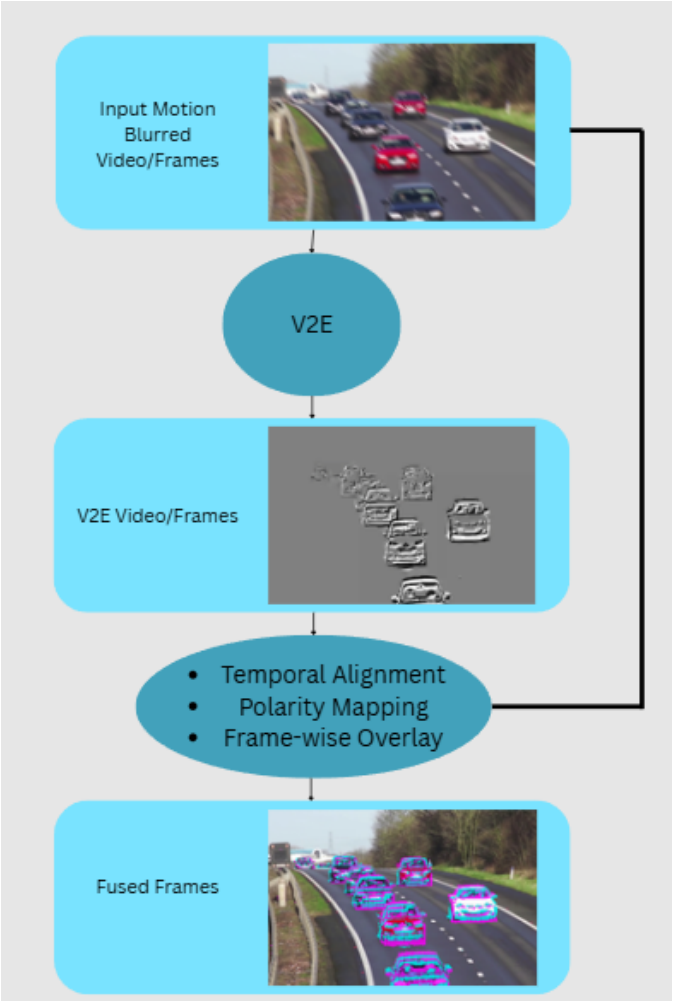

Fig. 2. Fusion Pipeline.

TABLE I
Comparison of Object Detection Performance under Heavy Blur

Model	mAP@50	mAP@50-95
RGB-only	0.75	0.63
Event-only	0.65	0.50
Fused	0.813	0.710

The fused RGB–Event model consistently outperforms both single-modality baselines, demonstrating the complementary strengths of event and RGB data.

In addition to the numerical metrics, qualitative detection results are presented in Figure 3. The RGB-only model struggles under severe motion blur, often missing object boundaries or producing incomplete detections. The event-only model successfully preserves moving edges but lacks texture information, leading to fragmented bounding boxes. The proposed Fused RGB–Event model combines the advantages of both modalities, producing sharper and more stable detections even under strong blur conditions.

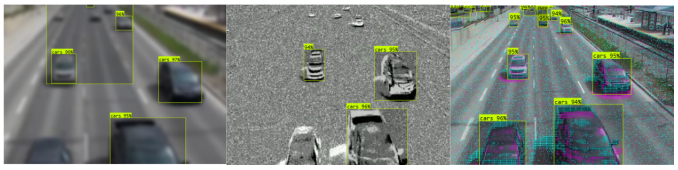

Fig. 3. Qualitative comparison of detection outputs for (a) RGB-only, (b) Event-only, and (c) Fused RGB–Event models.

V. Conclusion

In this study, we introduced a practical fusion-based approach that leverages event data to improve object detection performance under motion blur. Unlike many existing methods, the proposed technique does not rely on deblurring algorithms, temporal reconstruction, or modifications to the detection model. Instead, it uses a lightweight overlay-based fusion strategy that requires no additional computation during inference. Because the method operates directly on standard RGB frames and synthetic event representations, it is hardware-free and readily deployable on edge devices.

Experimental results demonstrate that event information provides strong complementary cues to blurred RGB images, enabling the fused RGB–Event model to outperform both RGB-only and event-only baselines. While effective, the method still faces challenges, particularly the generation of excessive background events when the camera moves rapidly, which can reduce detection clarity. As future work, we aim to develop background-filtering mechanisms and learnable fusion strategies to further enhance robustness in dynamic environments.

References

[1] T. -H. Wu, C. Gong, D. Kong, S. Xu and Q. Liu, "A novel visual object detection and distance estimation method for HDR scenes based on event camera," 2021 7th International Conference on Computer and Communications (ICCC), Chengdu, China, 2021, pp. 636-640, doi: 10.1109/ICCC54389.2021.9674426.

[2] H. J. Son, K. D. Park and C. E. Rhee, "Enhancing object detection accuracy through RGB and event fusion in motion blurred images," 2024 International Conference on Electronics, Information, and Communication (ICEIC), Taipei, Taiwan, 2024, pp. 1-3, doi: 10.1109/ICEIC61013.2024.10457200.

[3] H. Fradi and P. Papadakis, "Advancing object detection for autonomous vehicles via general purpose event-RGB fusion," 2024 Eighth IEEE International Conference on Robotic Computing (IRC), Tokyo, Japan, 2024, pp. 147-153, doi: 10.1109/IRC63610.2024.00033.

[4] H. J. Son, K. D. Park and C. E. Rhee, "Enhancing Object Detection Accuracy Through RGB and Event Fusion in Motion Blurred Images," 2024 International Conference on Electronics, Information, and Communication (ICEIC), Taipei, Taiwan, 2024, pp. 1-3, doi: 10.1109/ICEIC61013.2024.10457200.

[5] Relaxing Wave ASMR, "Cars Moving On Road Stock Footage - Free Download," YouTube, Oct. 10, 2023. [Online]. Available: https://www.youtube.com/watch?v=Y1jTEyb3wiIt=1s. [Accessed: Dec. 10, 2025].

[6] Amo Boss, "Free-stock dataset of cars on road," YouTube, Dec. 10, 2025. [Online]. Available: https://www.youtube.com/watch?v=6OplnRlePFo. [Accessed: Dec. 10, 2025].