

# Time-Specialized Event-Image Alignment for Blur-to-Video Decomposition

## Supplementary Material

In the supplementary material, we provide more details of the architecture of the Event Guide Gating Fusion (EGGF) module and the Conv-Swin Block in the proposed TSANet. We also show more details about our collected EBD dataset and more visual comparison results on test images.

### 1. Architecture details of TSANet

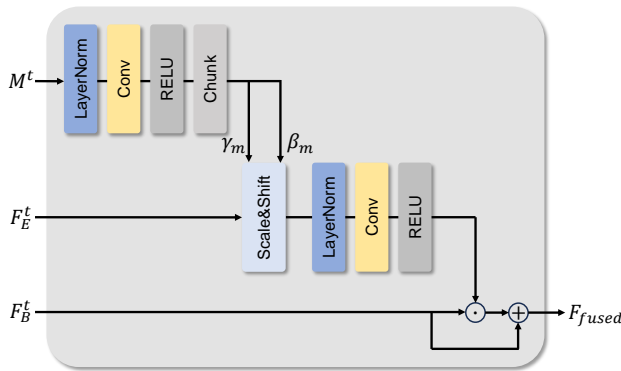


Figure A1. Detail illustration of the architecture of the Event Guide Gating Fusion (EGGF) module.

In our proposed TSANet, we use the Event Guide Gating Fusion (EGGF) module to fuse motion features and image features. Since both have undergone temporal specialized alignment via the Relative Time-Encoded Attention (RTEA) and Timesurface Dynamic Warping (TDW) modules, we do not need to adopt complex fusion methods. As shown in the Figure A1, after supplementing the  $F_E^t$  with more comprehensive motion information at time  $t$  through  $M_t$ , gating parameters for the  $F_B^t$  are generated to enhance the structural details of blurred regions in the  $F_B^t$  and modulate the  $F_B^t$  features.

In the event branch of the TSANet, we use Conv-Swin Block to extract the global spatial-temporal motion features. The detailed architecture is shown in Figure A2. With the help of alternately arranged ConvBlock and SwinTrans Block, we can focus on both local and global motion features simultaneously.

### 2. EBDdataset

To validate the performance of our proposed method on real-event data, we collected a dataset using the DVSync event camera, termed EBDdataset. The DVSync event camera is capable of capturing event data and high-quality RGB

frames simultaneously. Compared with the DAVIS346 Color event camera, it offers higher spatial resolution for both event and RGB data, along with superior RGB imaging quality. In contrast to the common dual-camera (one event camera + one RGB camera) setup with a beam splitter, the DVSync event camera boasts more accurate sensor-level spatial alignment and a wider imaging angle. Leveraging these advantages, we collected 29 color video sequences, encompassing 25,608 sharp frames and their corresponding events, at resolution of  $640 \times 1120$ . These sequences cover diverse scenes (indoor, playground, road, architecture) and motion types (camera motion and object motion). Several samples of EBDdataset are presented in Figures B3 to B5.

EBDdataset is not only applicable to motion decomposition tasks but also suitable for various event-based low-level vision tasks. Blur frames are generated by averaging  $2m+1$  sharp frames, which are then paired with the  $m+1$ -th sharp frame to evaluate the performance of event-based deblurring networks. Additionally, by selecting data at intervals, this dataset can be utilized to explore the performance of event-based frame interpolation networks under real-event scenarios.

### 3. More Visual Comparison Results

We provide more visual comparisons in Figures C6 and C7 on test datasets to show the effect of our method. We have compared our TSANet with other SOTA methods, including image-based methods (BiT and Demfi) and event-based methods (REFID and EvEnhancer). The Qualitative comparisons show that with the time-specialized image feature and event feature we can decompose a blurry frame into a more pleasing and sharper sequence, especially in high dynamic scenes.

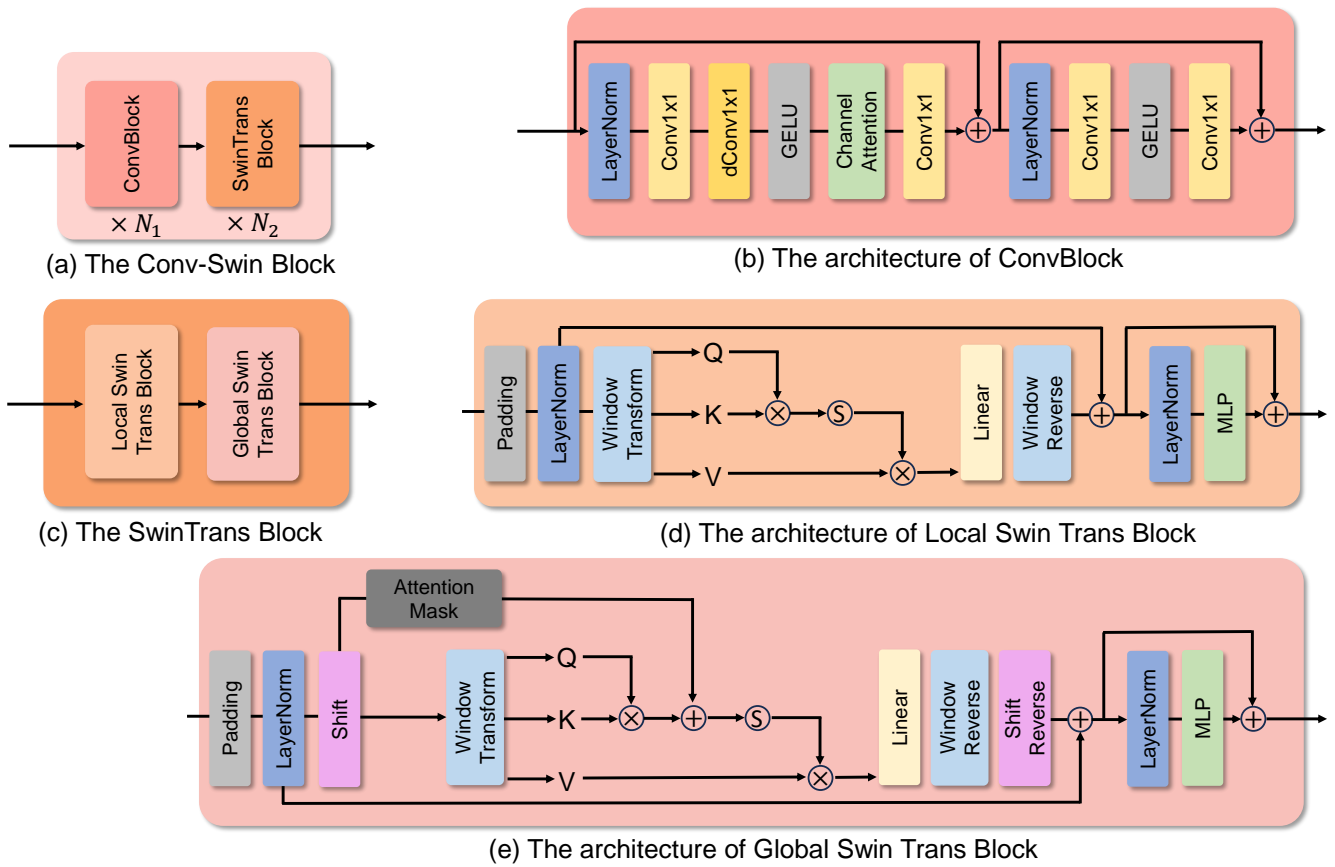
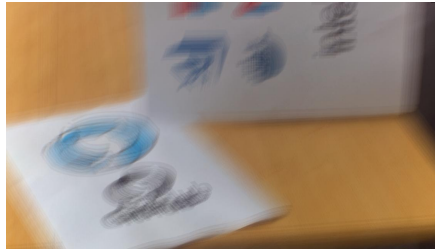
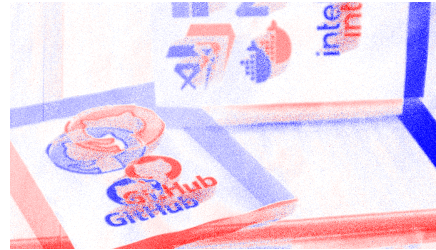


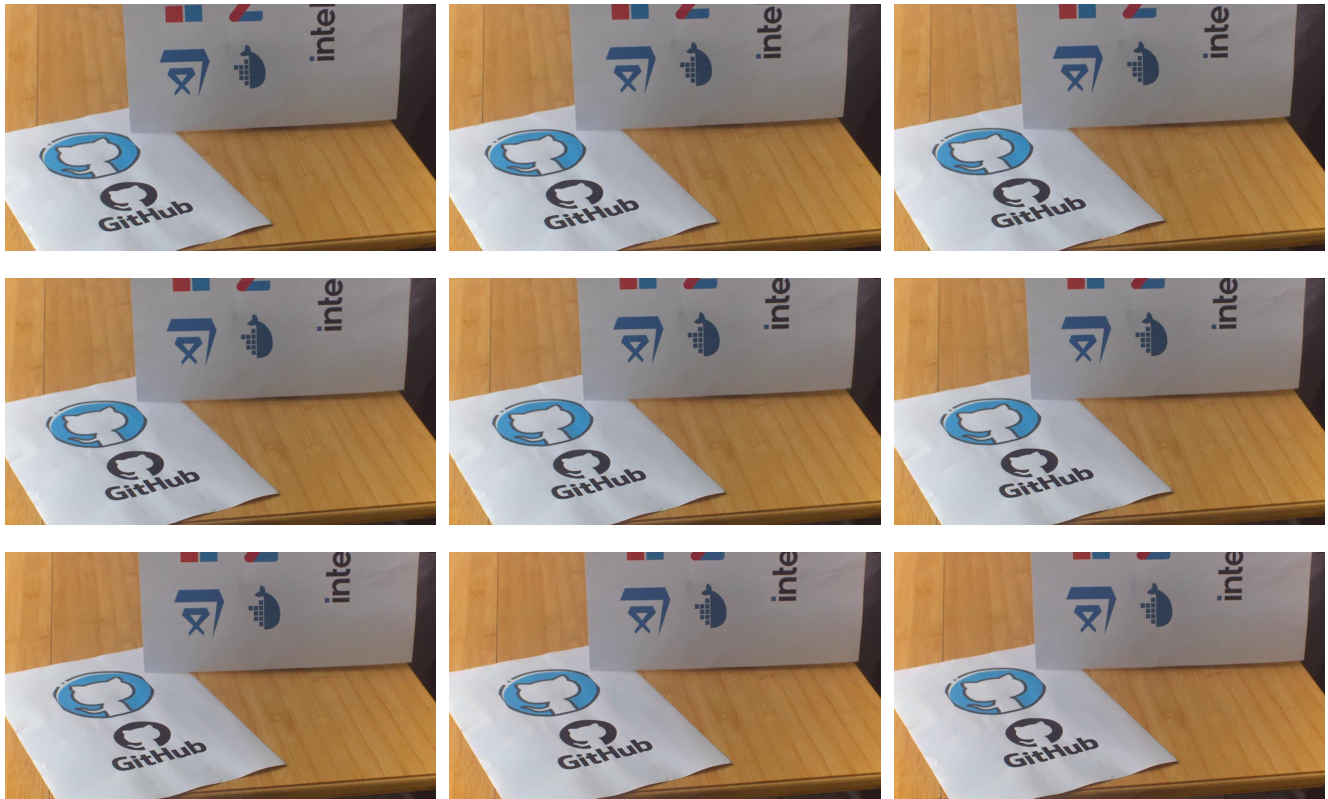
Figure A2. Detail illustration of the architecture of the Conv-Swin Block.



(a) Blurry Frame

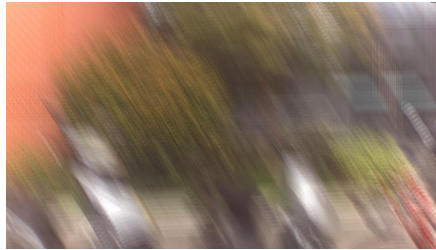


(b) Events

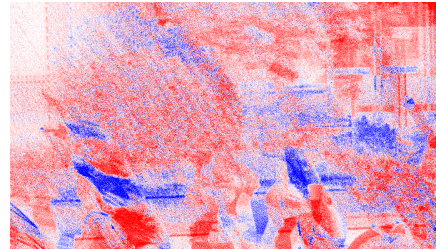


(c) Sharp Frames

Figure B3. Sample of blurry frame, events and sharp frames in EBDdataset.



(a) Blurry Frame



(b) Events



(c) Sharp Frames

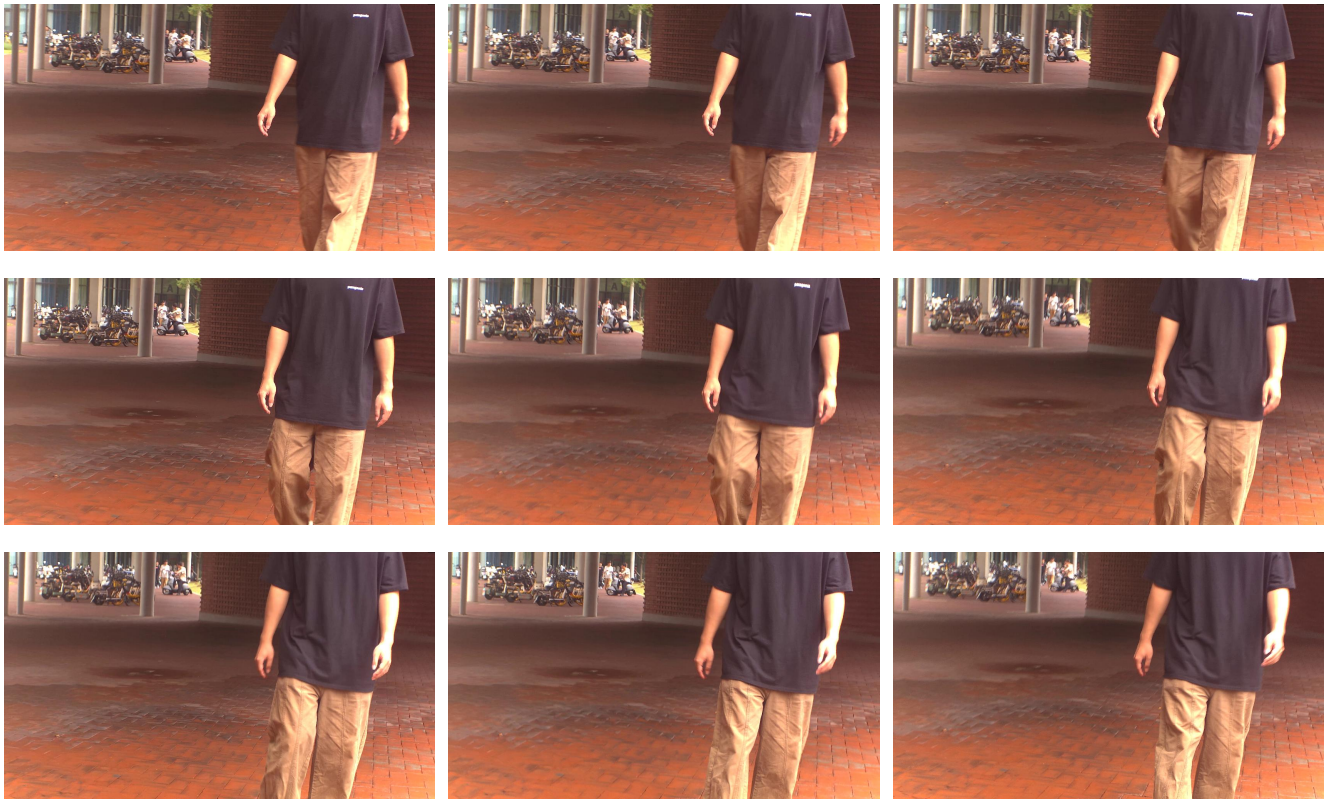
Figure B4. Sample of blurry frame, events and sharp frames in EBDdataset.



(a) Blurry Frame



(b) Events



(c) Sharp Frames

Figure B5. Sample of blurry frame, events and sharp frames in EBDdataset.

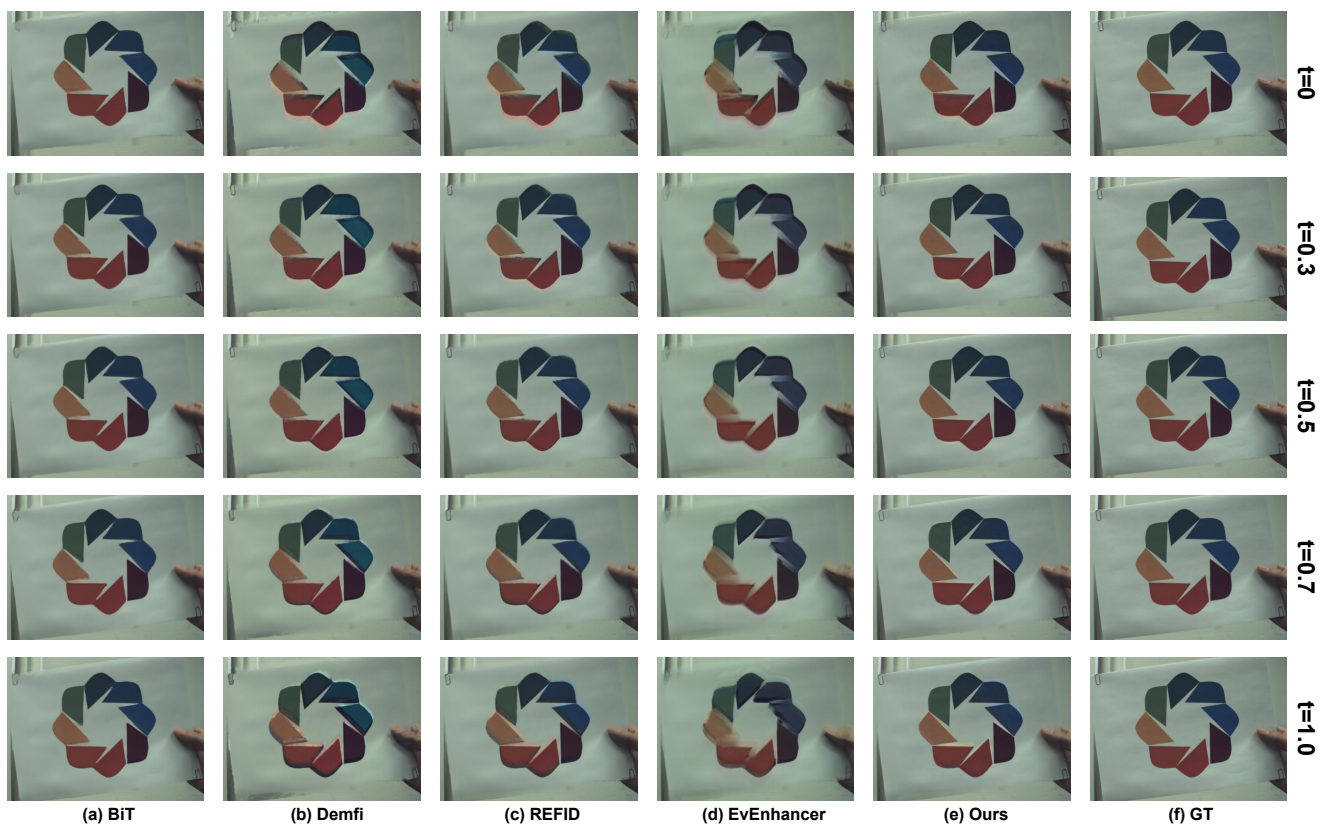


Figure C6. Qualitative comparisons on the HighREV dataset. Zoom in for better view.

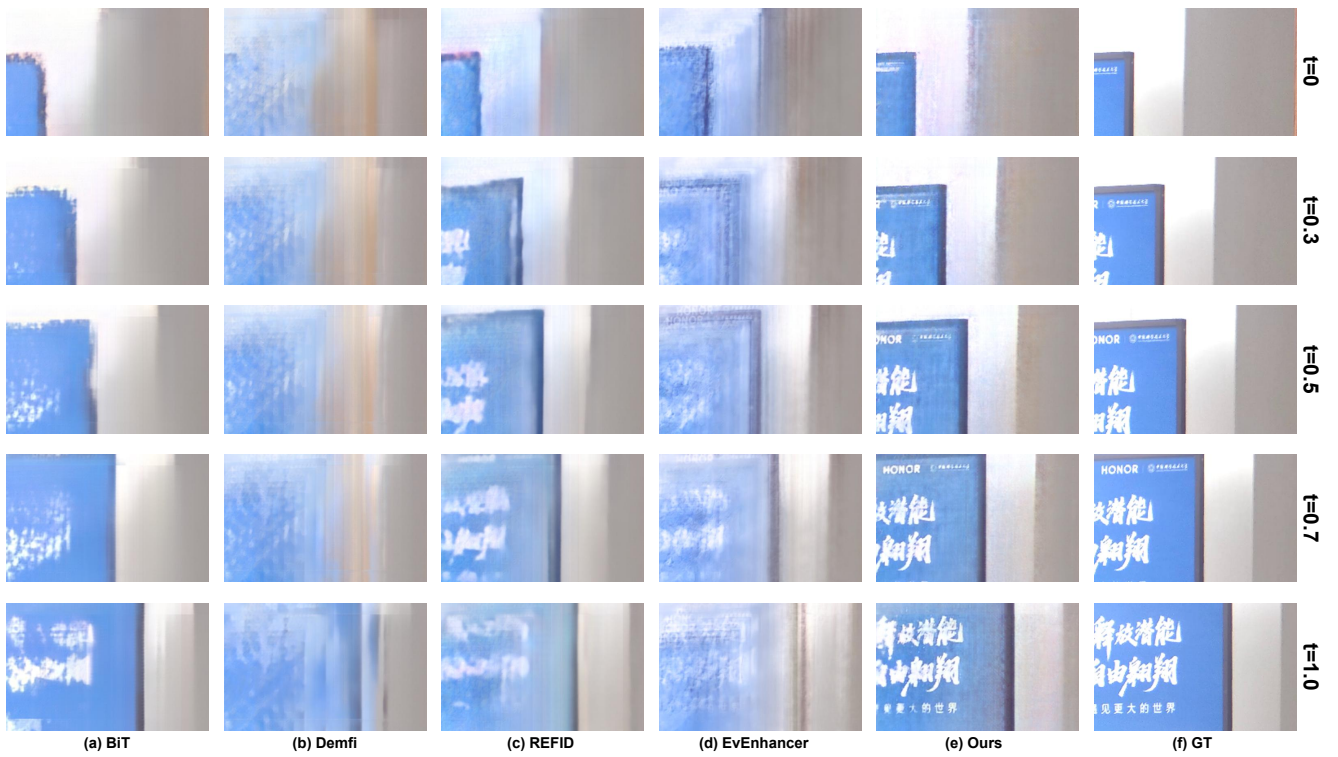


Figure C7. Qualitative comparisons on the EBDdataset. Zoom in for better view.