SUPPLEMENTARY MATERIALS FOR EVENT DENOISING BASED ON ITERATIVE TREE-STRUCTURED INFORMATION AGGREGATION

1. SCHEMATIC DIAGRAM OF THE ITERATIVE INFERENCE PROCESS

The iterative inference process can be divided into two steps which is illustrated in Figure 6. The detailed description is as follows:

(1) *Establish connections.* For the current event $e_{new} = \{x_{new}, y_{new}, t_{new}, p_{new}\}$, identify the pixel set satisfying $x \in [x_{new} - W, x_{new} + W]$ and $y \in [y_{new} - H, y_{new} + H]$ (excluding (x_{new}, y_{new})) and find events with timestamps greater than $t_{new} - T$ as child nodes from the time register. If the number of satisfying pixels exceeds the tree degree K_d, use nearest neighbor pruning (NNP) to select the nearest points. The results of the above process correspond completely to those in Section 3.1.

(2) *Feature aggregation.* After determining the child events of the latest event, the x and y coordinates along with the timestamp of the child events form the 0^{th}-order feature set. Applying the first-order convolution from Section 3.2 to the child nodes' 0^{th}-order features yields the latest event's 1^{st}-order features. Next, reset and write operations are performed: select the register group with the smallest value in the timestamp register at the pixel of the latest event, update it with the latest timestamp t_{new}, set the feature registers from 1st to $D-1$ to 0, and write the new event's first-order features into the reset feature registers. If the event has no child nodes, it is still necessary to reset and write to the time register.

The higher-order convolution of the algorithm is the same as the 1^{st}-order convolution, and the classification module is consistent with the one introduced in Section 3.2.

2. ABLATION STUDY

The ablation experiments investigate the effects of the attention branch in high-order convolution modules, the depth D of the Relation Tree, and the degree K_d of the Relation Tree across three key aspects, as summarized in Table 3.

In the detailed experimental setup, the role of the attention branch in high-order convolution modules is analyzed by removing the branch and retraining the model for evaluation. For the depth of the Relation Tree, configurations of $D = 1$, $D = 2$, and $D = 4$ are explored, while for the degree of the Relation Tree, values of $K_d = 8$, $K_d = 16$, and $K_d = 48$ are tested. To ensure the independence of variables, other modules remain unchanged during these experiments. The accuracy is evaluated across 16 scenes from the DVSNOISE20 dataset, and the average performance is reported.

Table 3. Ablation test of our algorithm

Attention Tests	Removed		
AUC	0.813		
Depth Tests	$D = 1$	$D = 2$	$D = 4$
AUC	0.688	0.792	0.870
Degree Tests	$K_d = 8$	$K_d = 16$	$K_d = 48$
AUC	0.783	0.845	0.869
Original	$AUC = 0.867$		

From the ablation experiments, we observe that the attention mechanism effectively aggregates useful information, thereby enhancing denoising accuracy. Increasing the depth and degree of the Relation Tree significantly improves denoising performance; however, this improvement becomes marginal when the tree depth reaches $D = 4$ and the degree reaches $K_d = 48$. This suggests that $D = 3$ and $K_d = 32$ are sufficient for capturing spatiotemporal correlations in event streams. Further increasing D to 4 or K_d to 48 results in higher computational costs during the inference phase, reducing inference speed without providing substantial performance gains.

3. VISUALIZATION

To visually demonstrate the denoising effects, we visualize selected scenes from the DVSNOISE20 dataset. In this process, the event data over a specific time period is converted into frames for display. Positive polarity events are represented in red, while negative polarity events are shown in blue. The results are presented in Figure 7. From the figure, it can be seen that our method effectively removes noise while preserving useful event information compared to other methods.

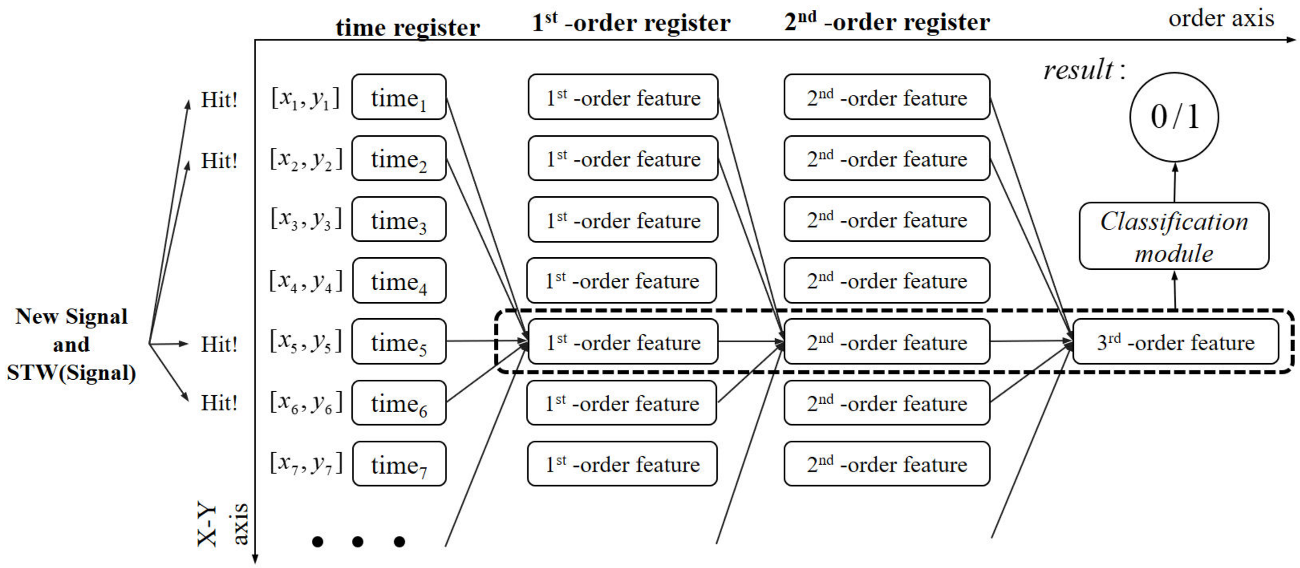

Fig. 6. Illustration of iteration inference process. In the iterative inference process, the sub-node selection is performed first, followed by sequential convolution. For clarity, we set $K_p = 1$, which indicates that only one register group is depicted for each pixel in the figure.

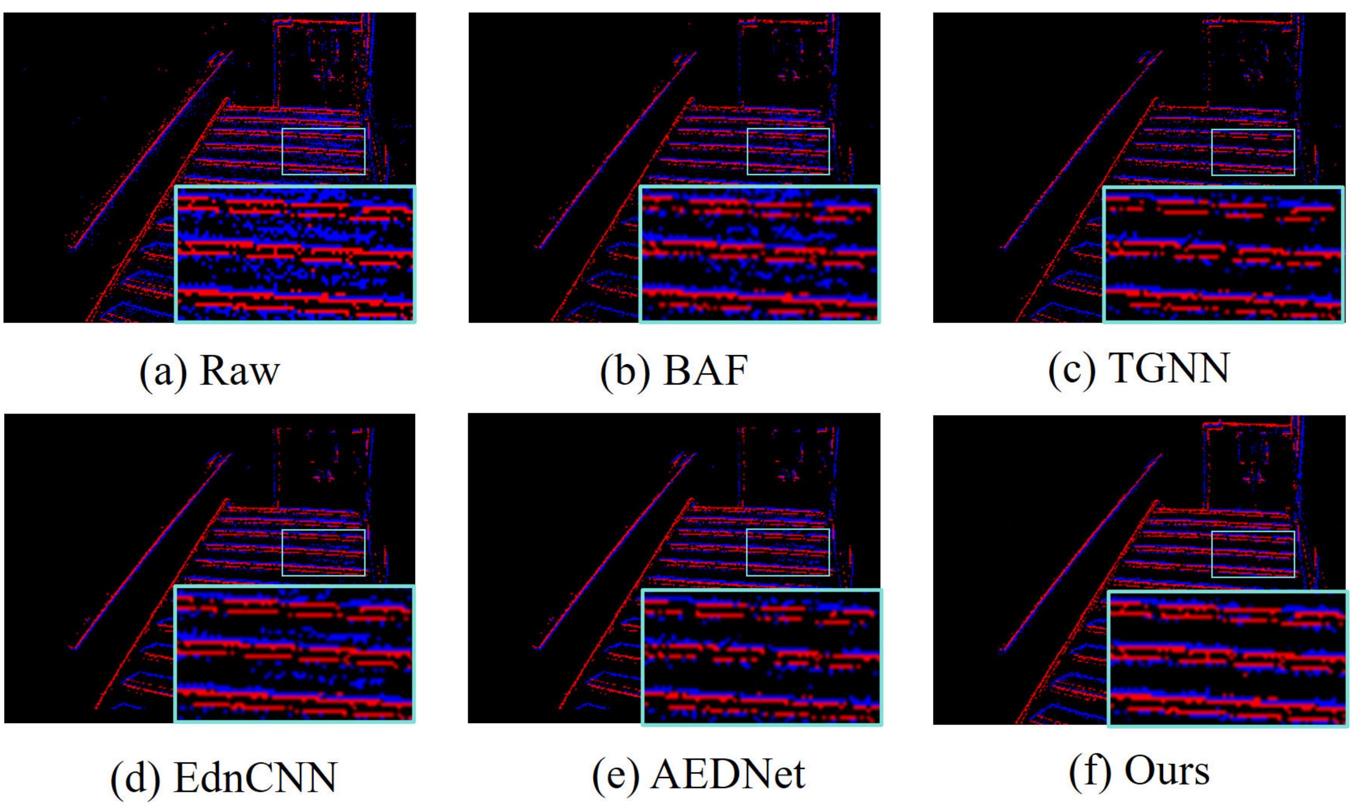

Fig. 7. Visualization of event denoising frames. This figure uses the stairs scene from the DVSNOISE20 dataset to illustrate the denoising effects.