Learning Graph-embedded Key-event Back-tracing for Object Tracking in Event Clouds
(*Supplementary Materials*)

Zhiyu ZHU, Junhui HOU,[*] Xianqiang LYU
Department of Computer Science, City University of Hong Kong
`zhiyuzhu2-c@my.cityu.edu.hk`, `jh.hou@cityu.edu.hk`,
`xianqialv2-c@my.cityu.edu.hk`

1 Details of the Network Architecture

In this section, we provide the detailed network architectures of different modules involved in our framework. We want to notice that the source code and pre-trained model are also included in the supplementary file.

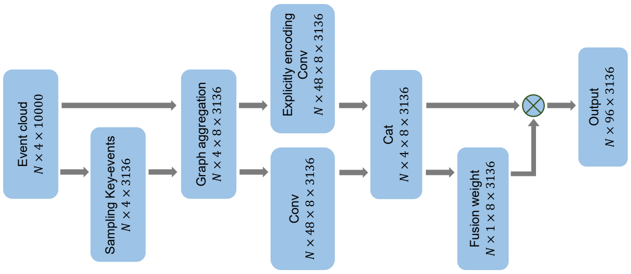

Figure 1: Illustration of the detailed network architecture of the proposed graph-based key-event embedding module.

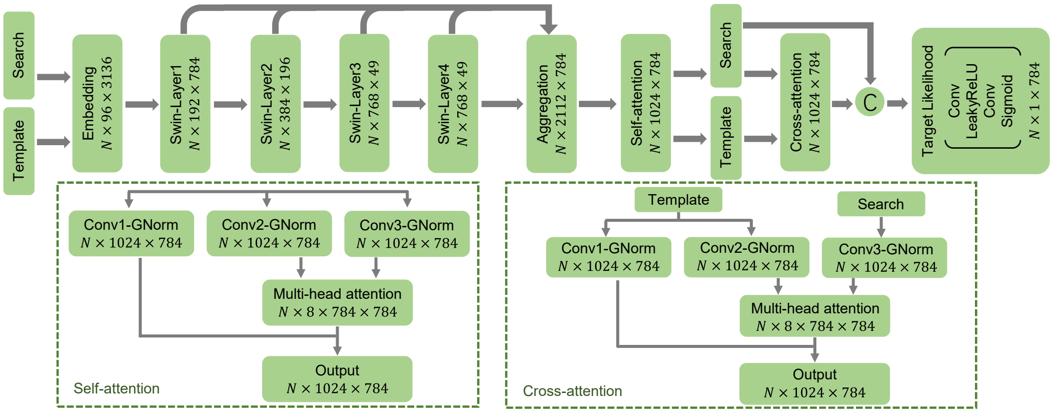

Figure 2: Illustration of the detailed network architecture of the proposed semantic-driven Siamese-matching module for target likelihood prediction.

---

[*] Corresponding author

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

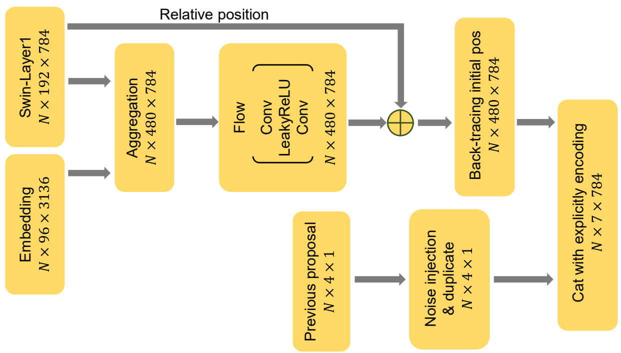

Figure 3: Illustration of the detailed network architecture of the proposed motion-aware key-event back-tracking module for target likelihood prediction.

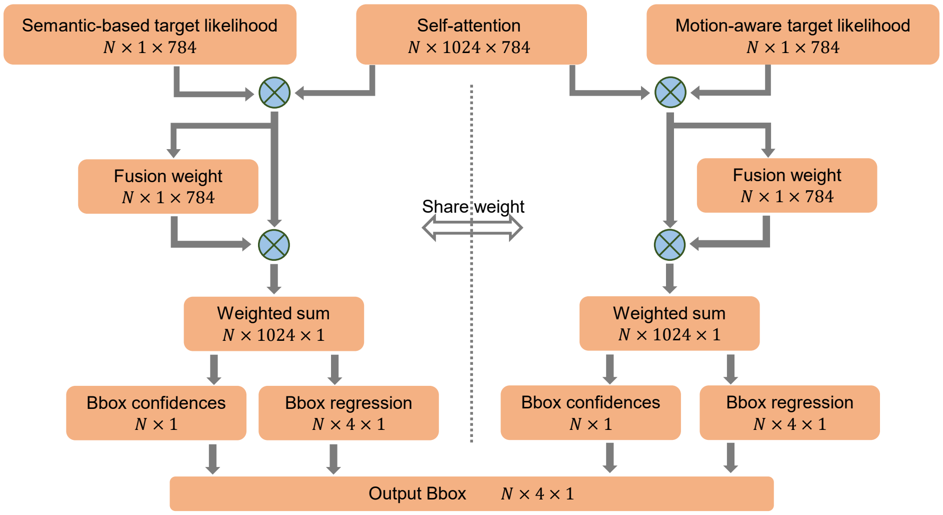

Figure 4: Illustration of the detailed network architecture of the proposed confidence-based object proposal module.

1.1 Graph-based Key-event Embedding

Fig. 1 shows the detailed network architecture of this module. “Graph aggregation" means searching the 8 nearest neighbours of each key-event.

1.2 Semantic-driven Siamese-matching

Fig. 2 shows the detailed network architecture of this module. Note that the spatially down-sampling operation contained in the swin-layers is to resample key-events further using the proposed down-sampling strategy, and feed it with its 3 nearest neighbours to the MLP for dimension reduction.

1.3 Motion-aware Key-event Back-tracing

Fig. 3 shows the detailed network architecture of this module.

1.4 Confidence-based Object Proposal

Fig. 4 shows the detailed network architecture of this module. The semantic-driven likelihood, self-attention, and motion-aware target likelihood are derived from the corresponding modules.

2 Details of Evaluation Metrics

Denote by $\mathbf{B} \in \mathbb{R}^{N\times 4}$ the predicted bounding boxes of event sequences with N event clouds, and $\mathbf{B}_{gt} \in \mathbb{R}^{N\times 4}$ the ground-truth bounding boxes in the form of $[cx,\ cy,\ w,\ h]$ with cx and cy being the spatial position of the central of a bounding-box, and w and h being the width and height. representation. The used metrics for quantitative evaluation are defined as follows:

$$\mathrm{RSR} = \frac{\sum_{i=1}^{N} \mathtt{IOU}\left(\mathbf{B}^i, \mathbf{B}_{gt}^i\right)}{N}, \tag{1}$$

$$\mathrm{OP}_{0.5} = \frac{\sum_{i=1}^{N} \mathbb{I}\left(\mathtt{IOU}\left(\mathbf{B}^i, \mathbf{B}_{gt}^i\right) > 0.5\right)}{N} \text{ with } \mathbb{I}(True) = 1, \mathbb{I}(False) = 0, \tag{2}$$

$$\mathrm{OP}_{0.75} = \frac{\sum_{i=1}^{N} \mathbb{I}\left(\mathtt{IOU}\left(\mathbf{B}^i, \mathbf{B}_{gt}^i\right) > 0.75\right)}{N}, \tag{3}$$

$$\mathrm{RPR} = \frac{\sum_{i=1}^{N} \mathbb{I}\left(\left\|\mathbf{B}^i[1:2] - \mathbf{B}_{gt}^i[1:2]\right\|_2 < 20\right)}{N}, \tag{4}$$

$$\mathrm{RPR}_{0.075} = \frac{\sum_{i=1}^{N} \mathbb{I}\left(\left\|\mathbf{B}^i[1:2] - \mathbf{B}_{gt}^i[1:2]\right\|_2 < 0.075\right)}{N}, \tag{5}$$

where $\mathtt{IOU}(\cdot,\cdot)$ computes the intersection over union of two bounding boxes.