Cross-modal Orthogonal High-rank Augmentation for RGB-Event Transformer-trackers (Supplementary Materials)

Zhiyu Zhu, Junhui Hou, and Dapeng Oliver Wu
Department of Computer Science, City University of Hong Kong
zhiyuzhu2-c@my.cityu.edu.hk; jh.hou@cityu.edu.hk; dapengwu@cityu.edu.hk

In this document, we provide the supplementary information for our submission titled "Cross-modal Orthogonal High-rank Augmentation for RGB-Event Transformer-trackers," including the detailed flowcharts of the employed two-stream and one-stream tracking frameworks in Sec. 1, success and accuracy plots of sequences with different attributes from the COESOT datset in Sec. 2, detailed relation between orthogonality and high-rank property in Sec. 3. Besides, we also refer the reviewers to the attached **source code** and **video demo** for more information.

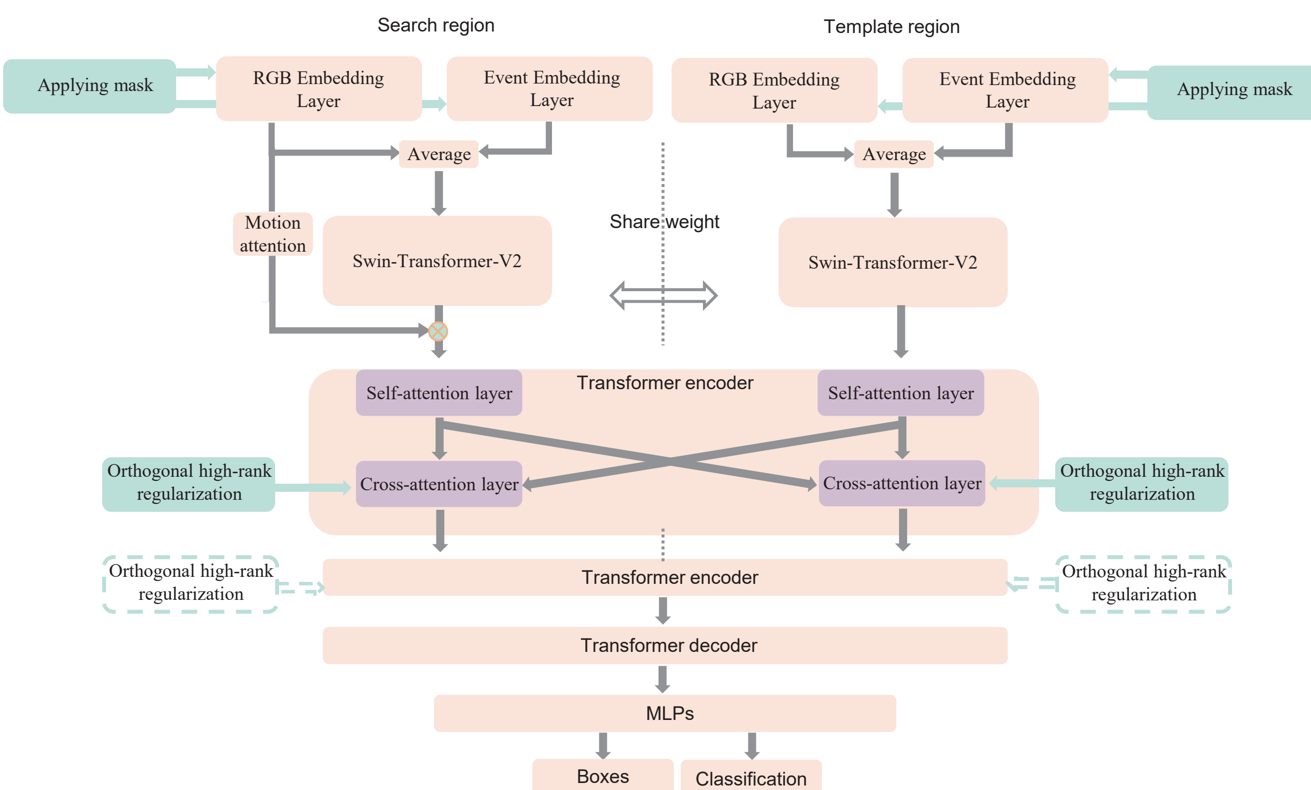

Figure S1. Illustrations of the flowchart of the employed two-stream tracker and the target layers where the proposed plug-and-play training augmentations are applied.

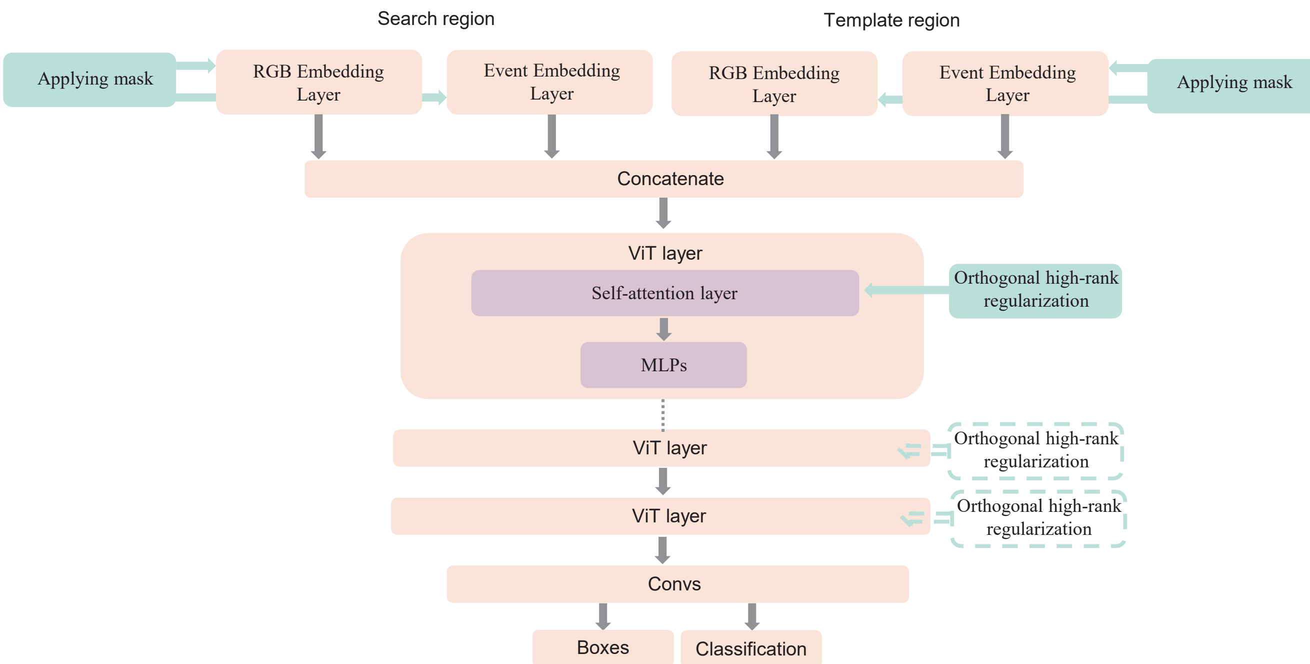

Figure S2. Illustrations of the flowchart of the employed one-stream tracker and the target layers where the proposed plug-and-play training augmentations are applied.

1. Detailed Flowcharts of the Trackers

We show the detailed flowcharts of the employed two-stream and one-stream tracking methods in Figs. S1 and S2, respectively.

2. Success and Accuracy Plots

We show the success and accuracy plots of different attributes on the COESOT dataset in Figs. S3 and S4, demonstrating that the proposed training augmentation techniques can, in most cases, enhance the baseline model.

3. Relation between orthogonality and high-rank

Row-/column-wise orthogonal matrix should be high-rank. Here, we demonstrate the theoretical plausibility of the proposed regularization term. Given a desired attention matrix $M \in \mathbb{R}^{n\times m}(n > m)$ with non-negative elements (for matrix M with $n < m$, we can analyze its transpose $M^\top$), as stated in our manuscript (Lines 437 - 441), ideally, assuming that if tokens are well-embedded and with highly discriminative features, each token will form a unique correlation with its identical counterpart, resulting in each row or column being orthogonal to the others, i.e., $\mathbf{m}_i^\top \cdot \mathbf{m}_j = 0$ and $\|\mathbf{m}_i\|_2^2 > 0$ for $i, j \in [1,\ m]$ and $i \neq j$, where $\mathbf{m}_i \in \mathbb{R}^{n\times}$ is the i-th column of M. Thus, we have

$$M^\top \cdot M = \mathtt{Diag}(\|\mathbf{m}_1\|_2^2, \|\mathbf{m}_2\|_2^2, ..., \|\mathbf{m}_m\|_2^2), \tag{S1}$$

where $\mathtt{Diag}(\cdot)$ denotes a diagonal matrix with inputs as main diagonal elements. Consequently, the matrix M is high-rank, i.e., $\mathtt{rank}(M) = m$. Thus, regularizing an attention matrix or block to be high-rank is feasible to achieve the anticipated orthogonality.

Fig. 7 of the manuscript basically validates the orthogonality of the proposed method. To further directly visualize the variations of matrix orthogonality after applying the proposed augmentation, we visualized the matrix $M^\top \times M$ in Figs. **??**

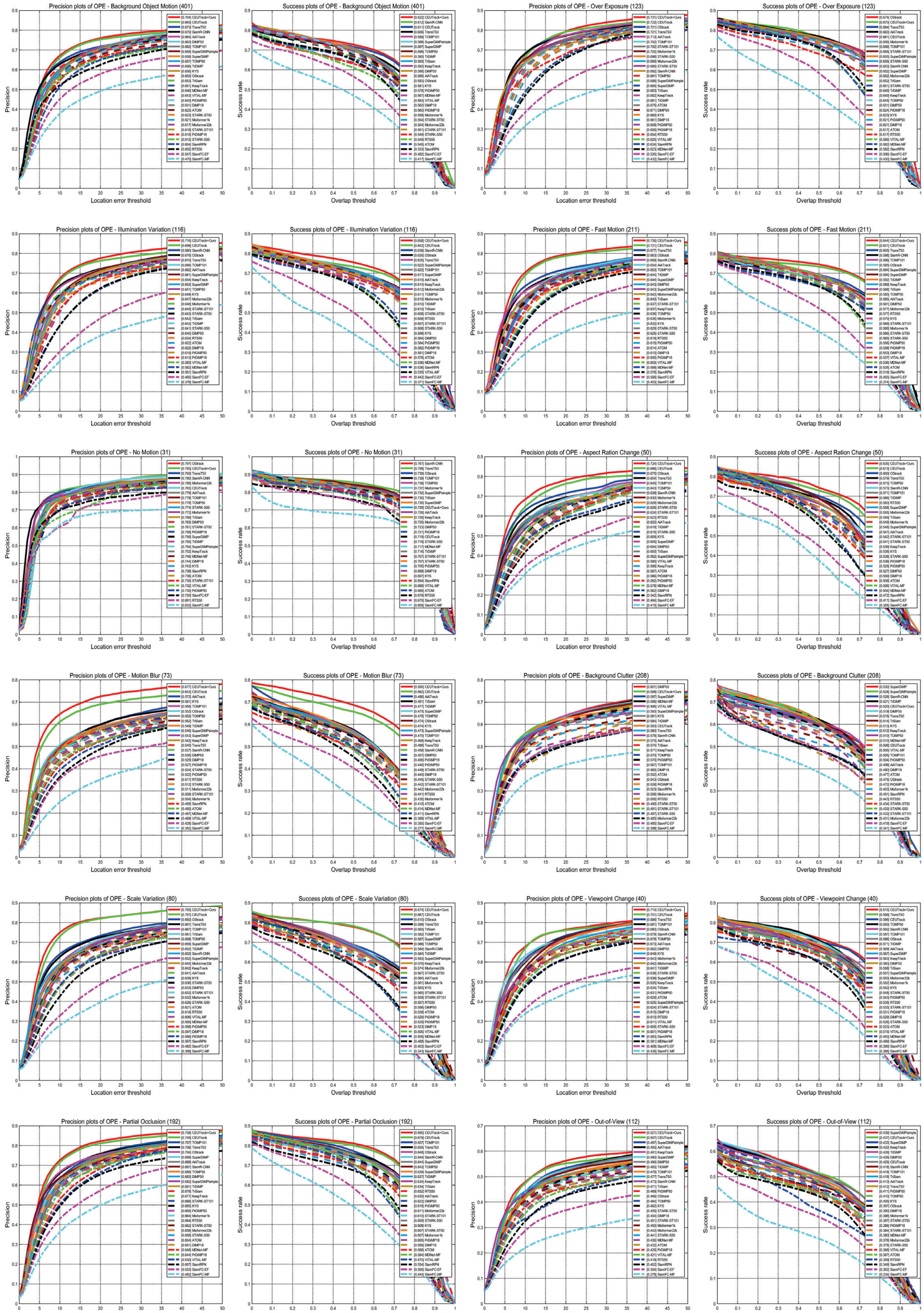

Figure S3. Success and precision plots of different methods under different attributes on the COESOT dataset (*part 1*).

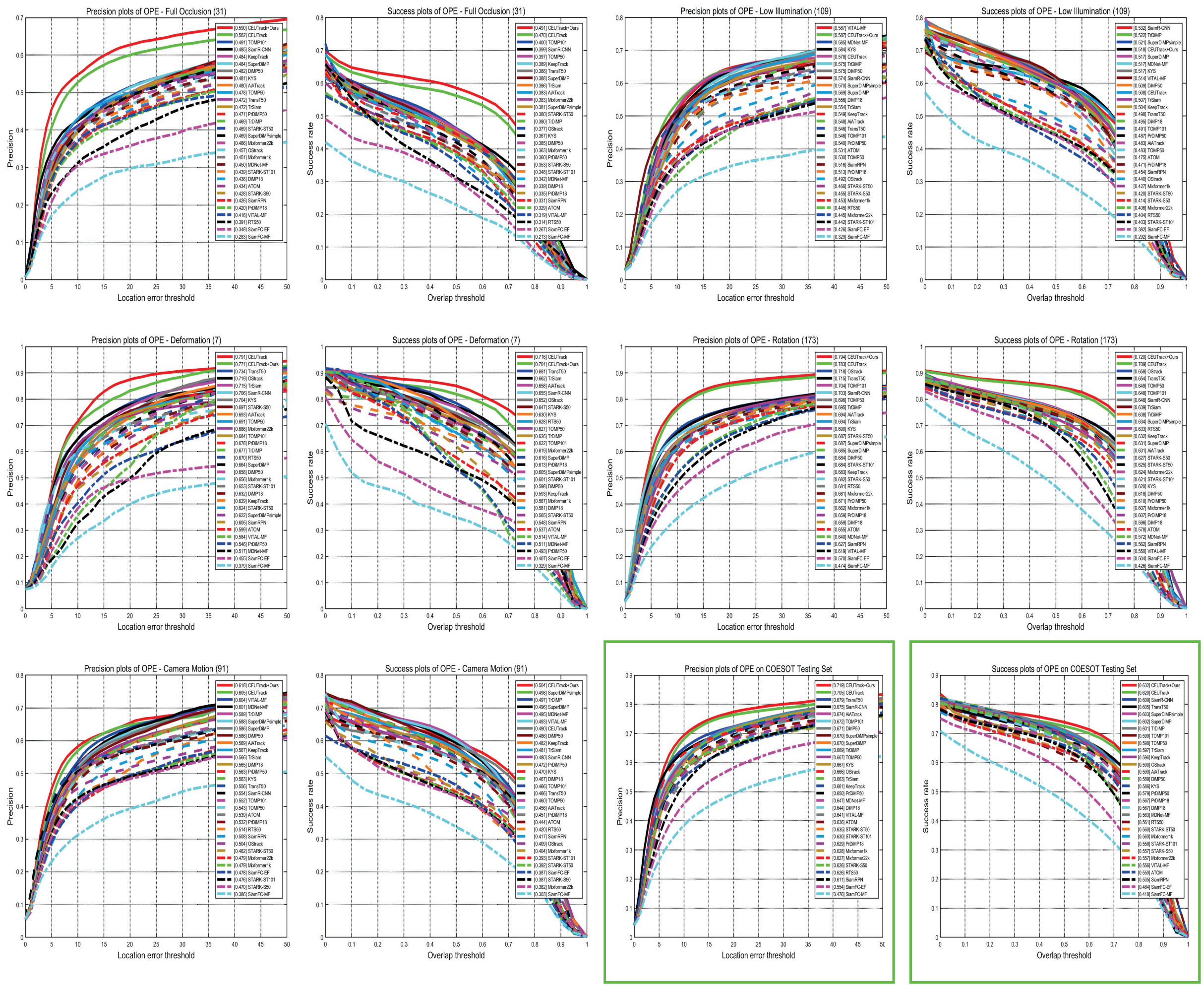

Figure S4. Success and precision plots of different methods under different attributes on the COESOT dataset (*part 2*). The plots of the whole testing test are highlighted in green frames.

Table S1. Results of the ablative study on orthogonalization methods."Dig" denotes directly minimizing the difference between $M^{\mathsf{T}} \cdot M$ and a diagonal matrix.

Methods	SR	PR	NPR	BOC
CEUTrack	62.0	70.5	69.0	20.77
CEUTrack+Dig	62.4	70.9	69.3	21.24
CEUTrack+Ours	63.2	71.9	70.2	21.58

and **??**, convincingly demonstrating the effectiveness of our regularization for enhancing a matrix's row-/column-wise orthogonality.

Other ways to promote orthogonalization. Notably, the presented high-rank regularization is not the only way to accomplish the desired orthogonalization in the attention matrix. we can also accomplish regularization via pushing Eq. (S1) towards a diagonal matrix, and the corresponding results under this regularization are listed in the following Table S1.