AiATrack: Attention in Attention for Transformer Visual Tracking

Shenyuan Gao, Chunluan Zhou, Chao Ma, Xinggang Wang, Junsong Yuan

ECCV TEL AVIV 2022

Limitation of Conventional Attention

$$\text{ConvenAttn}(\mathbf{Q},\mathbf{K},\mathbf{V}) = (\text{Softmax}\left(\frac{\bar{\mathbf{Q}}\bar{\mathbf{K}}^{\mathbf{T}}}{\sqrt{C}}\right)\bar{\mathbf{V}})\mathbf{W_o}$$

The correlation of each key-query pair is calculated independently, which ignores the correlations of other key-query pairs.

→ Result in imperfect correlations, which inhibits the power of Transformer trackers.

Motivation and Insight

If a key has a high correlation with a query, its neighboring keys would also have relatively high correlations with that query (spatial relevance of images).

→ Adaptively seek global consensus among raw correlations.

Methodology

Refine raw correlation map using another attention, which has dynamic weights and a global receptive field.

$$\text{AttninAttn}(\mathbf{Q},\mathbf{K},\mathbf{V}) = (\text{Softmax}(\mathbf{M} + \text{InnerAttn}(\mathbf{M}))\bar{\mathbf{V}})\mathbf{W_o}$$

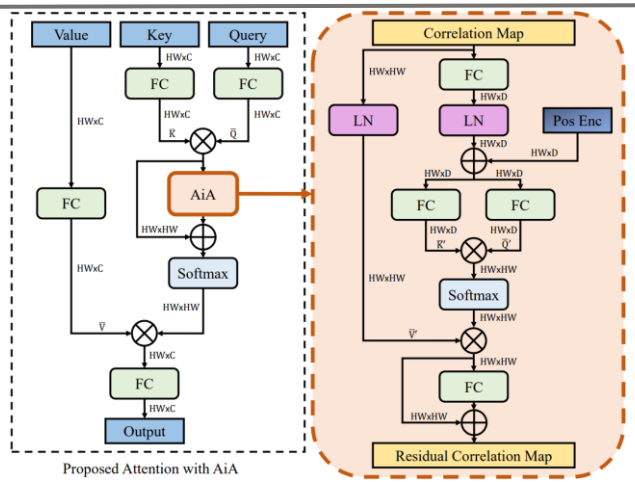
→ Enhance appropriate correlations and suppress unreliable correlations.

→ Can be applied to both self-attention and cross-attention blocks in a typical Transformer tracker.

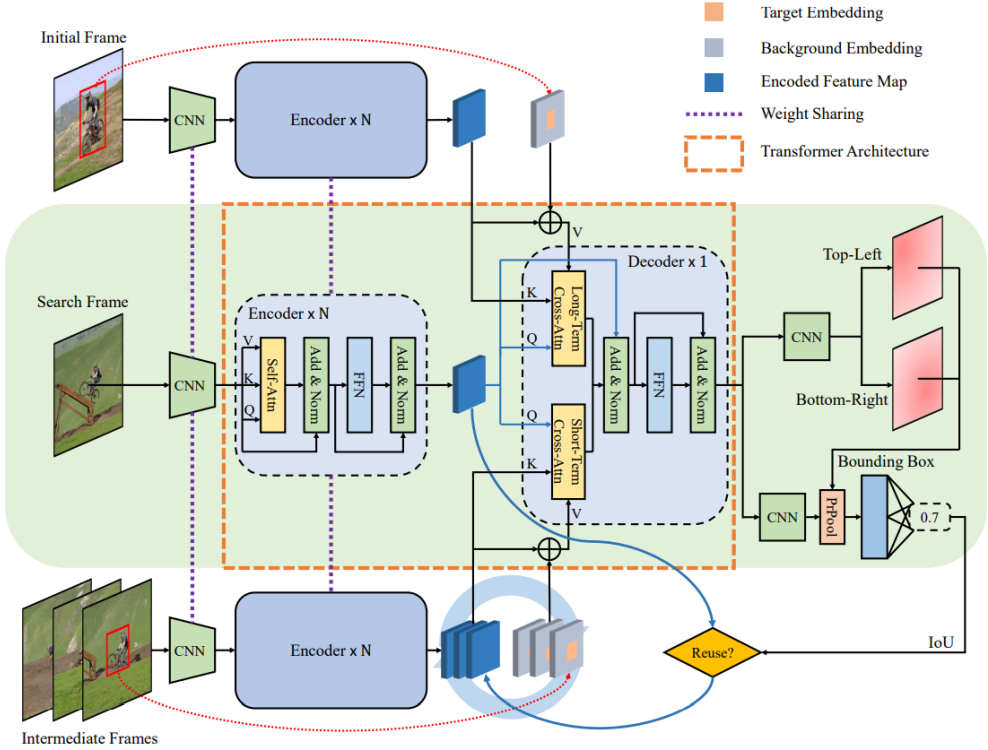
Proposed AiA:

2.5% performance gain on LaSOT.

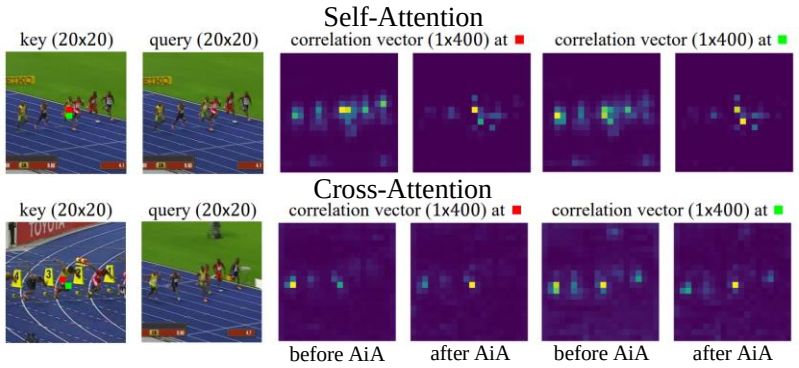
0.8% increase in model parameters.



A streamlined framework to utilize multiple references:



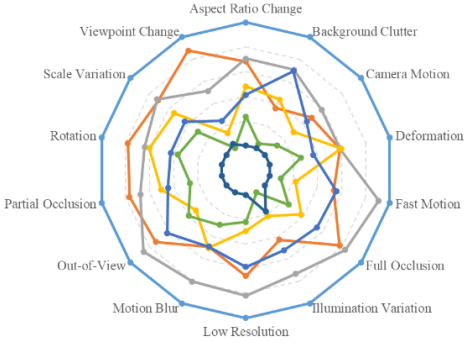
Visualization Results



State-of-the-Art Comparison

Tracker	Source	LaSOT [17] AUC	LaSOT P_{Norm}	LaSOT P	TrackingNet [45] AUC	TrackingNet P_{Norm}	TrackingNet P	GOT-10k [25] AO	GOT-10k $SR_{0.75}$	GOT-10k $SR_{0.5}$
AiATrack	Ours	69.0	79.4	73.8	82.7	87.8	80.4	69.6	63.2	80.0
STARK-ST50 [58]	ICCV2021	66.4	76.3	71.2	81.3	86.1	78.1	68.0	62.3	77.7
KeepTrack [41]	ICCV2021	67.1	77.2	70.2	-	-	-	-	-	-
DTT [61]	ICCV2021	60.1	-	-	79.6	85.0	78.9	63.4	51.4	74.9
TransT [8]	CVPR2021	64.9	73.8	69.0	81.4	86.7	80.3	67.1	60.9	76.8
TrDiMP [53]	CVPR2021	63.9	-	61.4	78.4	83.3	73.1	67.1	58.3	77.7
TrSiam [53]	CVPR2021	62.4	-	60.0	78.1	82.9	72.7	66.0	57.1	76.6
KYS [4]	ECCV2020	55.4	63.3	-	74.0	80.0	68.8	63.6	51.5	75.1
Ocean-online [67]	ECCV2020	56.0	65.1	56.6	-	-	-	61.1	47.3	72.1
Ocean-offline [67]	ECCV2020	52.6	-	52.6	-	-	-	59.2	-	69.5
PrDiMP50 [12]	CVPR2020	59.8	68.8	60.8	75.8	81.6	70.4	63.4	54.3	73.8
SiamAttn [62]	CVPR2020	56.0	64.8	-	75.2	81.7	-	-	-	-
DiMP50 [3]	ICCV2019	56.9	65.0	56.7	74.0	80.1	68.7	61.1	49.2	71.7
SiamRPN++ [34]	CVPR2019	49.6	56.9	49.1	73.3	80.0	69.4	51.7	32.5	61.6

Attribute-Based Evaluation



code
model