

TransT-M: Multi-Template Transformer Tracking

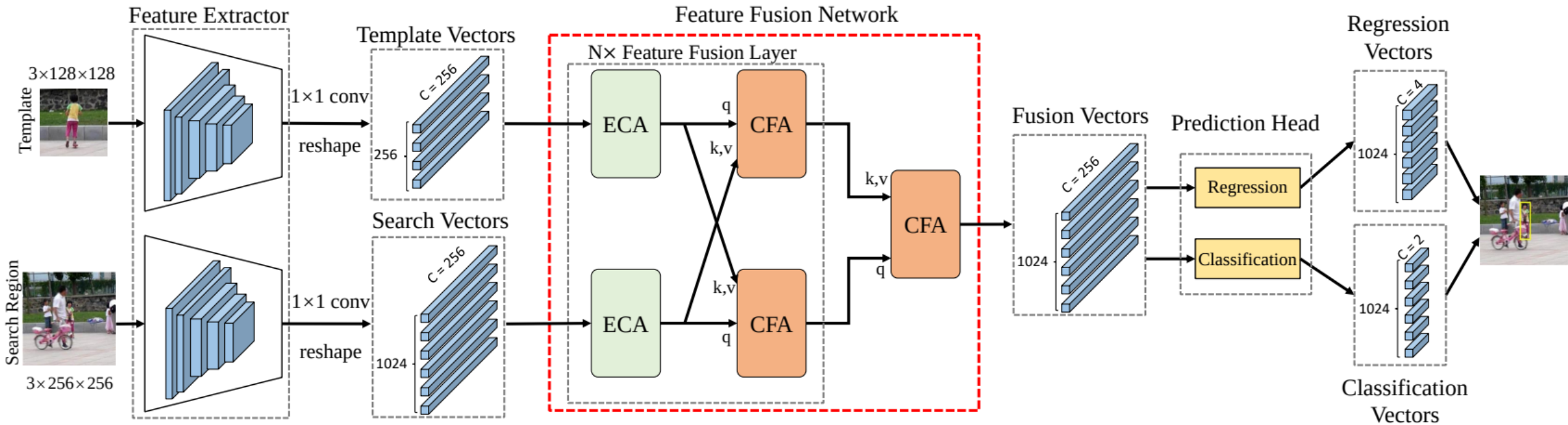
Xin Chen¹, Jiawen Zhu¹, Bin Yan¹, Dong Wang¹, Huchuan Lu^{1,2}, Xiaoyun Yang³

¹Dalian University of Technology

²Peng Cheng Laboratory ³Remark Holdings

Multi-Template Transformer Tracking

Transformer Tracking (TransT)



Xin Chen, Bin Yan, Jiawen Zhu, Dong Wang, Xiaoyun yang, Huchuan Lu. Transformer Tracking. CVPR, 2021.

➤ Code: <https://github.com/chenxin-dlut/TransT>

<https://github.com/chenxin-dlut/TransT>

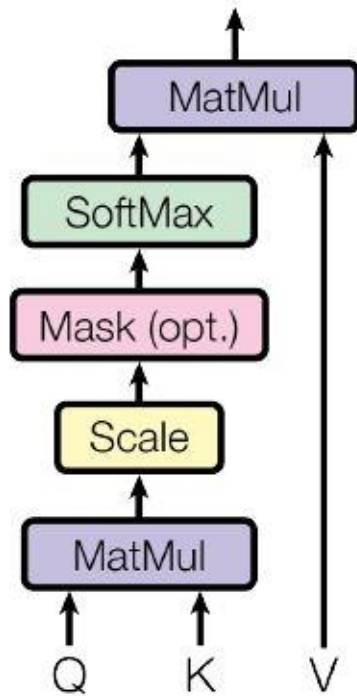
DUT, IIAU-LAB

Multi-Template Transformer Tracking

➤ Attention to Replace “Correlation”

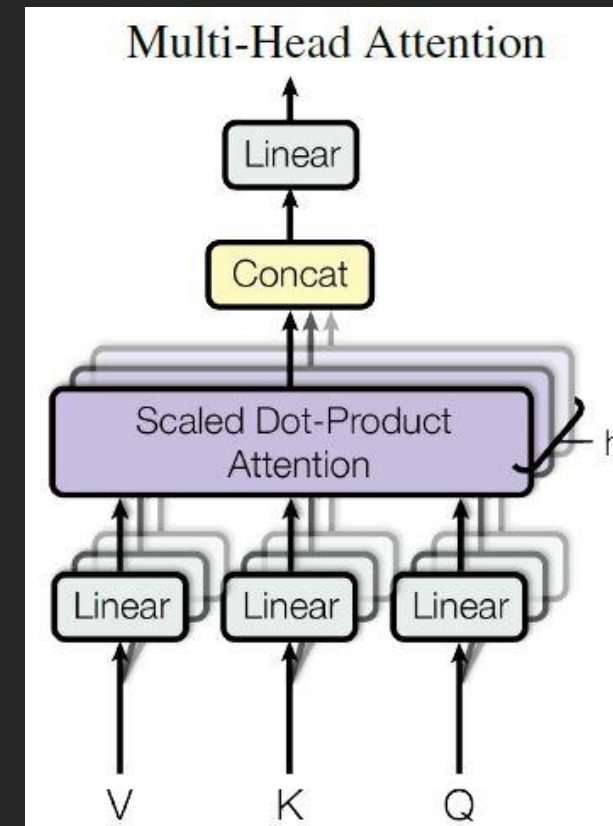
$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}}\right)\mathbf{V}$$

Scaled Dot-Product Attention



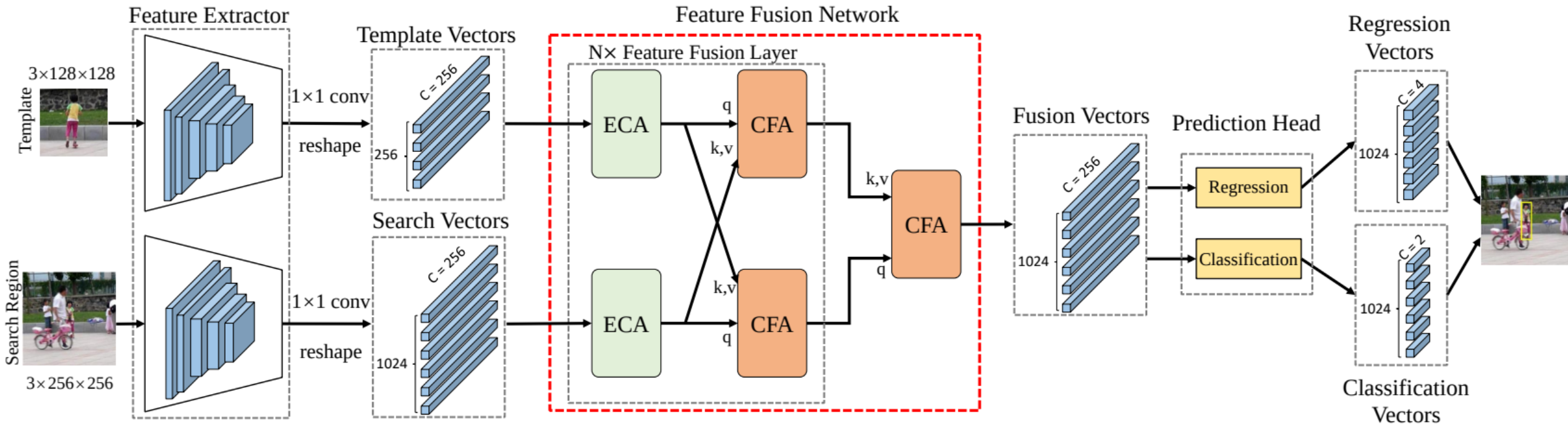
$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\mathbf{H}_1, \dots, \mathbf{H}_{n_h})\mathbf{W}^O$$

$$\mathbf{H}_i = \text{Attention}(\mathbf{Q}\mathbf{W}_i^Q, \mathbf{K}\mathbf{W}_i^K, \mathbf{V}\mathbf{W}_i^V)$$



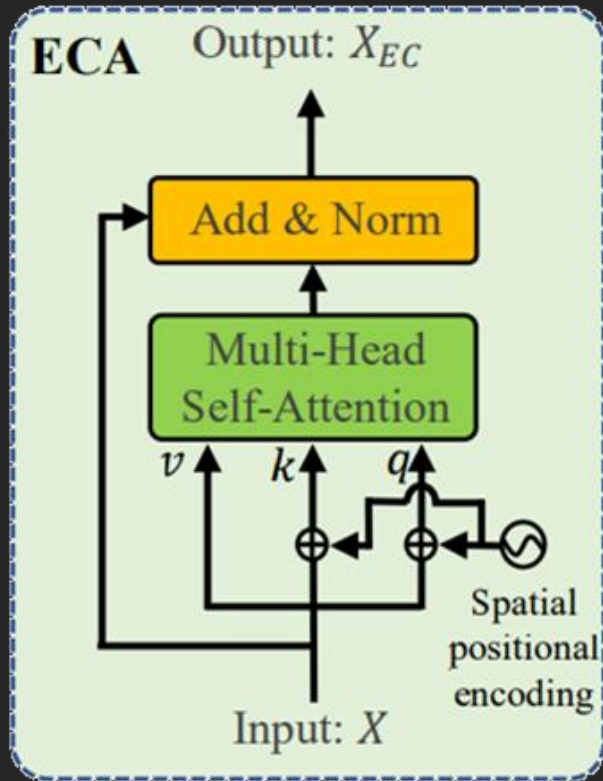
Multi-Template Transformer Tracking

TransT Framework

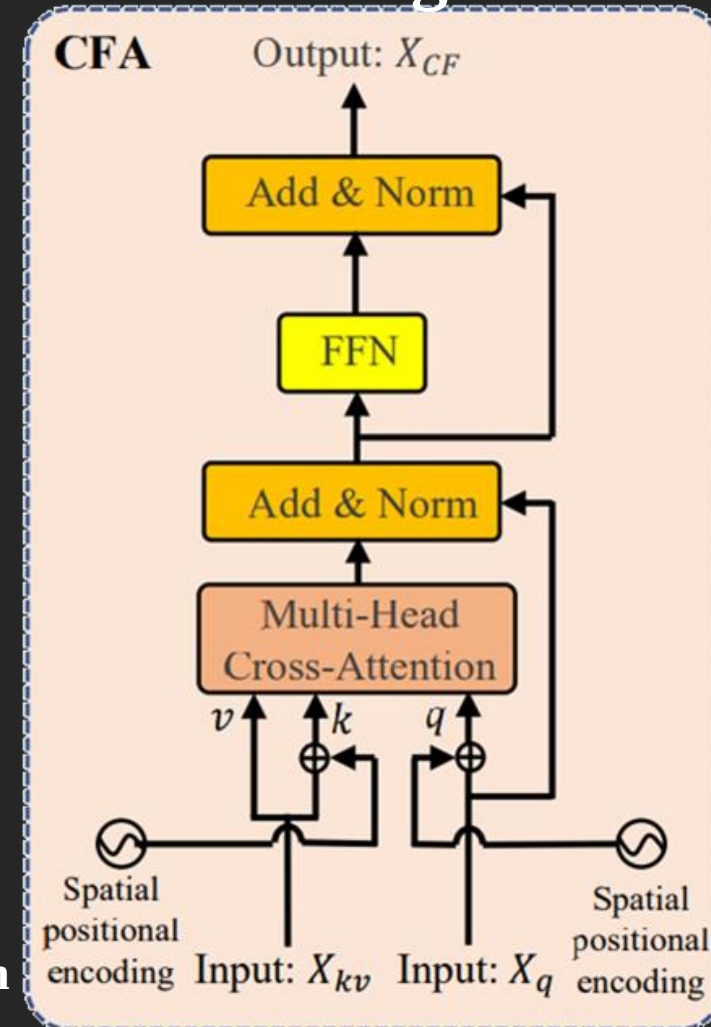


Multi-Template Transformer Tracking

Ego-Context Augment Module

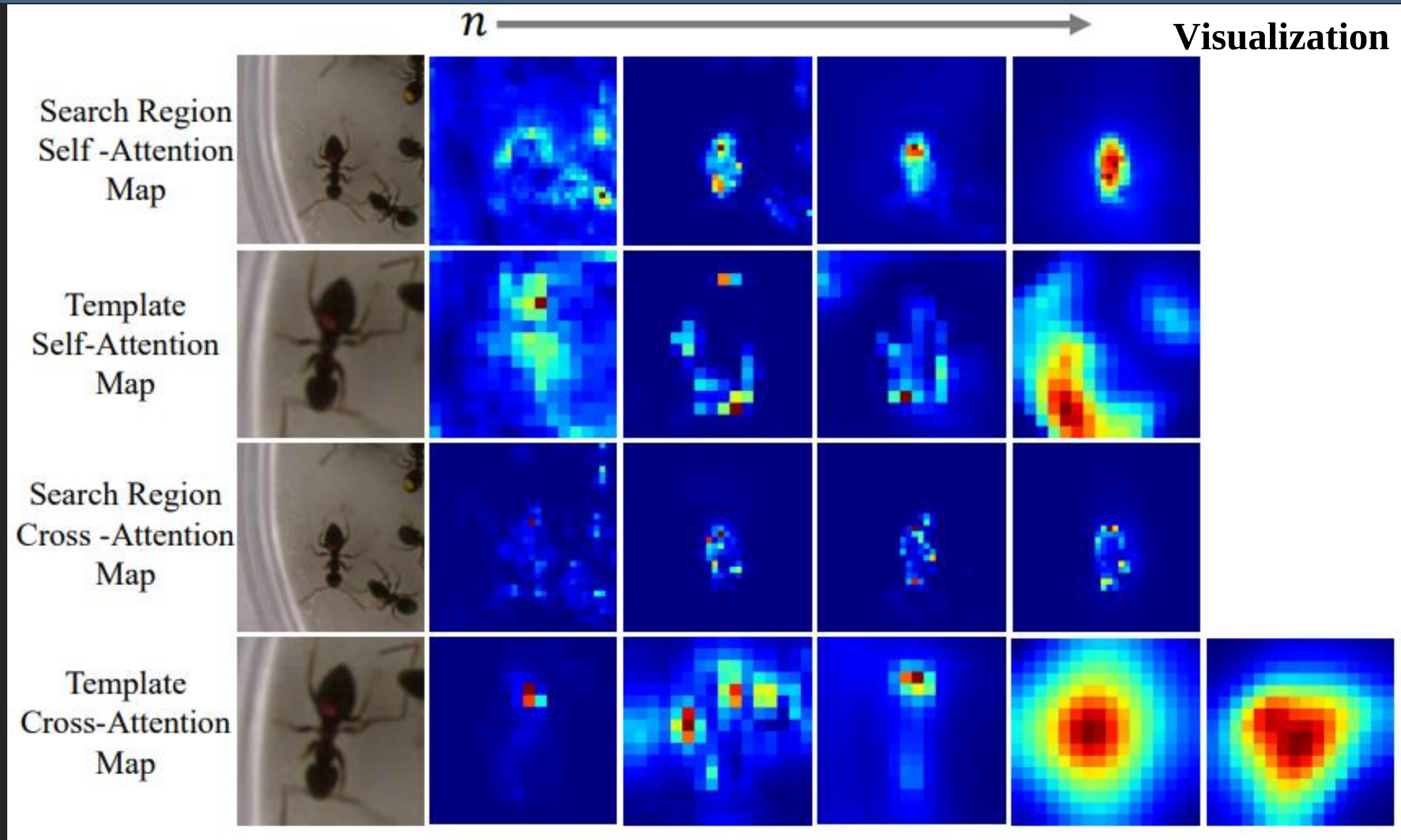


Cross-Feature Augment Module



- ✓ ECA based on self-attention and CFA based on cross-attention
- ✓ CFA performs feature fusion, retaining rich semantic information
- ✓ ECA and CFA establish dependence between long distance features and aggregate global information

Multi-Template Transformer Tracking



Multi-Template Transformer Tracking



- Ground-Truth - Ours - Ocean

Multi-Template Transformer Tracking

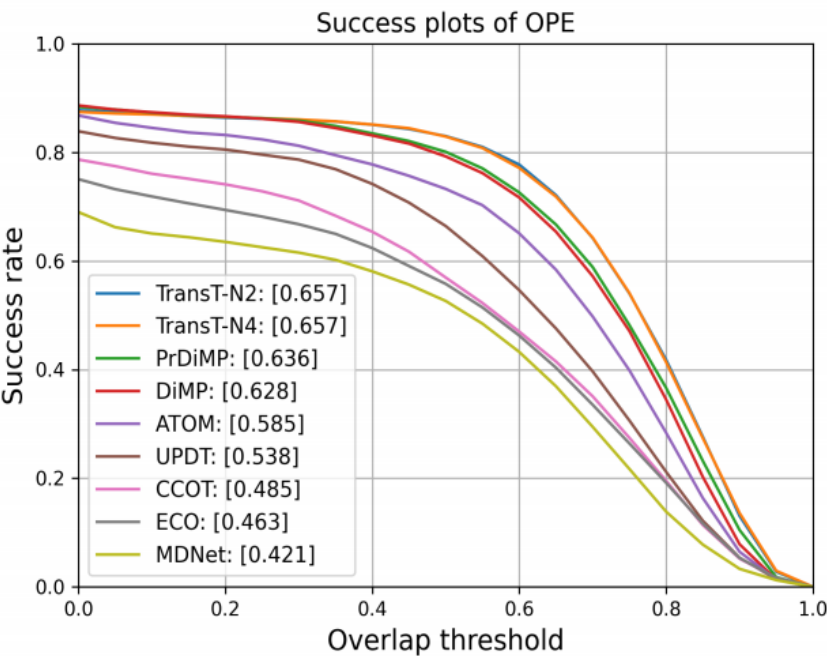
Large-scale Benchmark Results

Method	Source	LaSOT [14]			TrackingNet [30]			GOT-10k [19]		
		AUC	P_{Norm}	P	AUC	P_{Norm}	P	AO	SR _{0.5}	SR _{0.75}
TransT	Ours	64.9	73.8	69.0	81.4	86.7	80.3	72.3	82.4	68.2
TransT-N2	Ours	64.2	73.5	68.2	80.9	86.4	79.2	69.9	80.1	65.9
TransT-GOT	Ours	-	-	-	-	-	-	67.1	76.8	60.9
SiamR-CNN [39]	CVPR2020	64.8	72.2	-	81.2	85.4	80.0	64.9	72.8	59.7
Ocean [48]	ECCV2020	56.0	65.1	56.6	-	-	-	61.1	72.1	47.3
KYS [3]	ECCV2020	55.4	63.3	-	74.0	80.0	68.8	63.6	75.1	51.5
DCFST [49]	ECCV2020	-	-	-	75.2	80.9	70.0	63.8	75.3	49.8
SiamFC++ [44]	AAAI2020	54.4	62.3	54.7	75.4	80.0	70.5	59.5	69.5	47.9
PrDiMP [10]	CVPR2020	59.8	68.8	60.8	75.8	81.6	70.4	63.4	73.8	54.3
CGACD [13]	CVPR2020	51.8	62.6	-	71.1	80.0	69.3	-	-	-
SiamAttn [46]	CVPR2020	56.0	64.8	-	75.2	81.7	-	-	-	-
MAML [40]	CVPR2020	52.3	-	-	75.7	82.2	72.5	-	-	-
D3S [26]	CVPR2020	-	-	-	72.8	76.8	66.4	59.7	67.6	46.2
SiamCAR [16]	CVPR2020	50.7	60.0	51.0	-	-	-	56.9	67.0	41.5
SiamBAN [5]	CVPR2020	51.4	59.8	52.1	-	-	-	-	-	-
DiMP [2]	ICCV2019	56.9	65.0	56.7	74.0	80.1	68.7	61.1	71.7	49.2
SiamPRN++ [21]	CVPR2019	49.6	56.9	49.1	73.3	80.0	69.4	51.7	61.6	32.5
ATOM [9]	CVPR2019	51.5	57.6	50.5	70.3	77.1	64.8	55.6	63.4	40.2
ECO [8]	ICCV2017	32.4	33.8	30.1	55.4	61.8	49.2	31.6	30.9	11.1
MDNet [31]	CVPR2016	39.7	46.0	37.3	60.6	70.5	56.5	29.9	30.3	9.9
SiamFC [1]	ECCVW2016	33.6	42.0	33.9	57.1	66.3	53.3	34.8	35.3	9.8

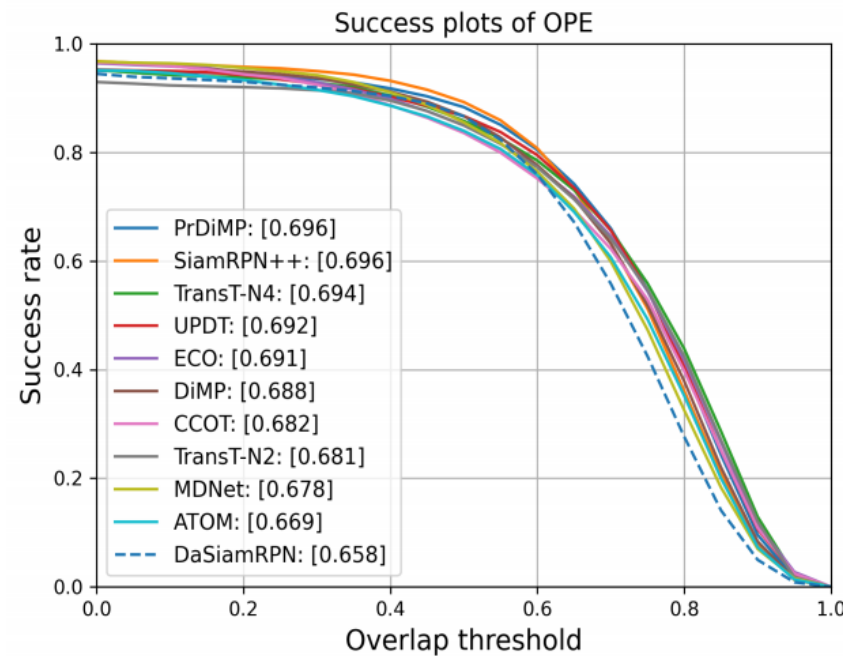
← 46fps
← 66fps

Multi-Template Transformer Tracking

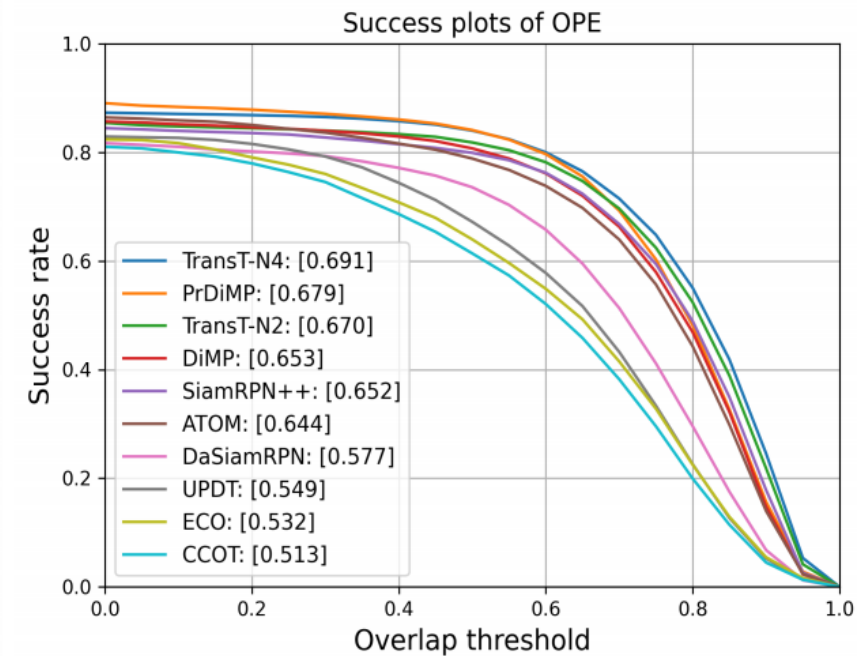
Results on NFS, OTB2015 and UAV123



(a) NFS



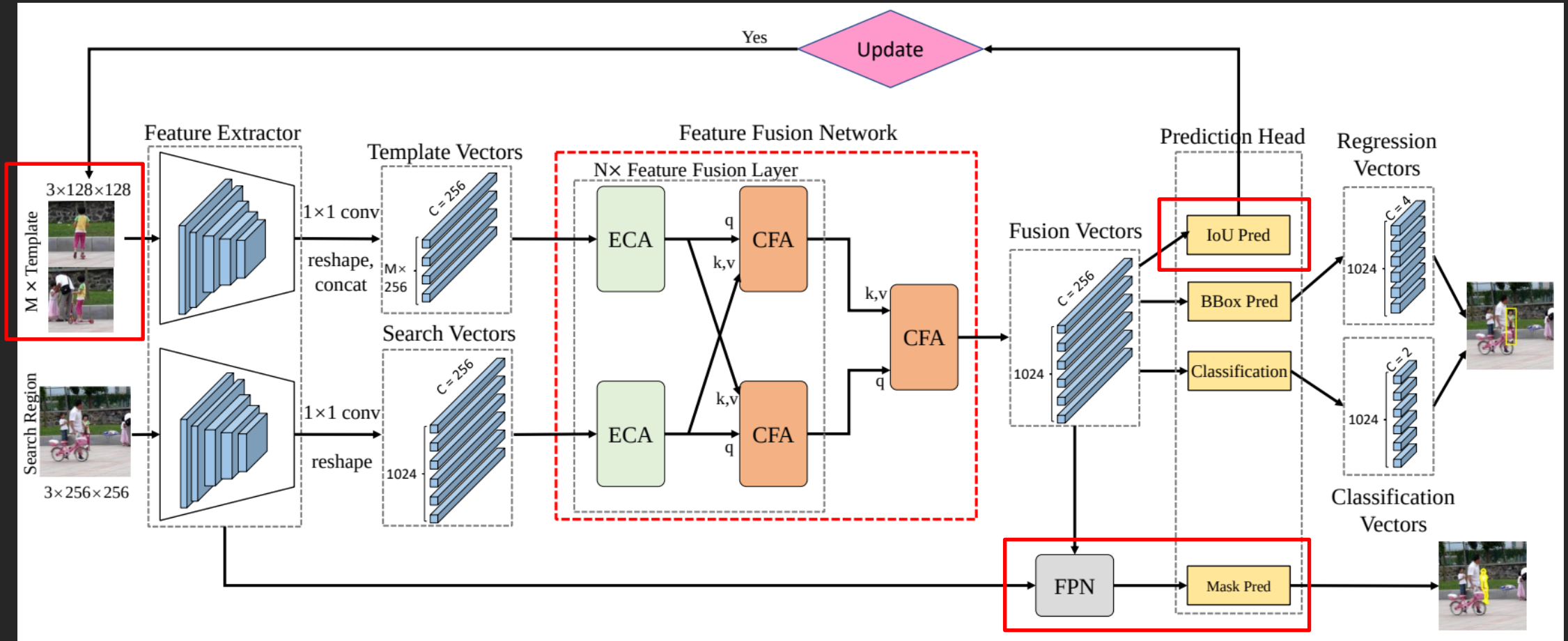
(b) OTB2015



(c) UAV123

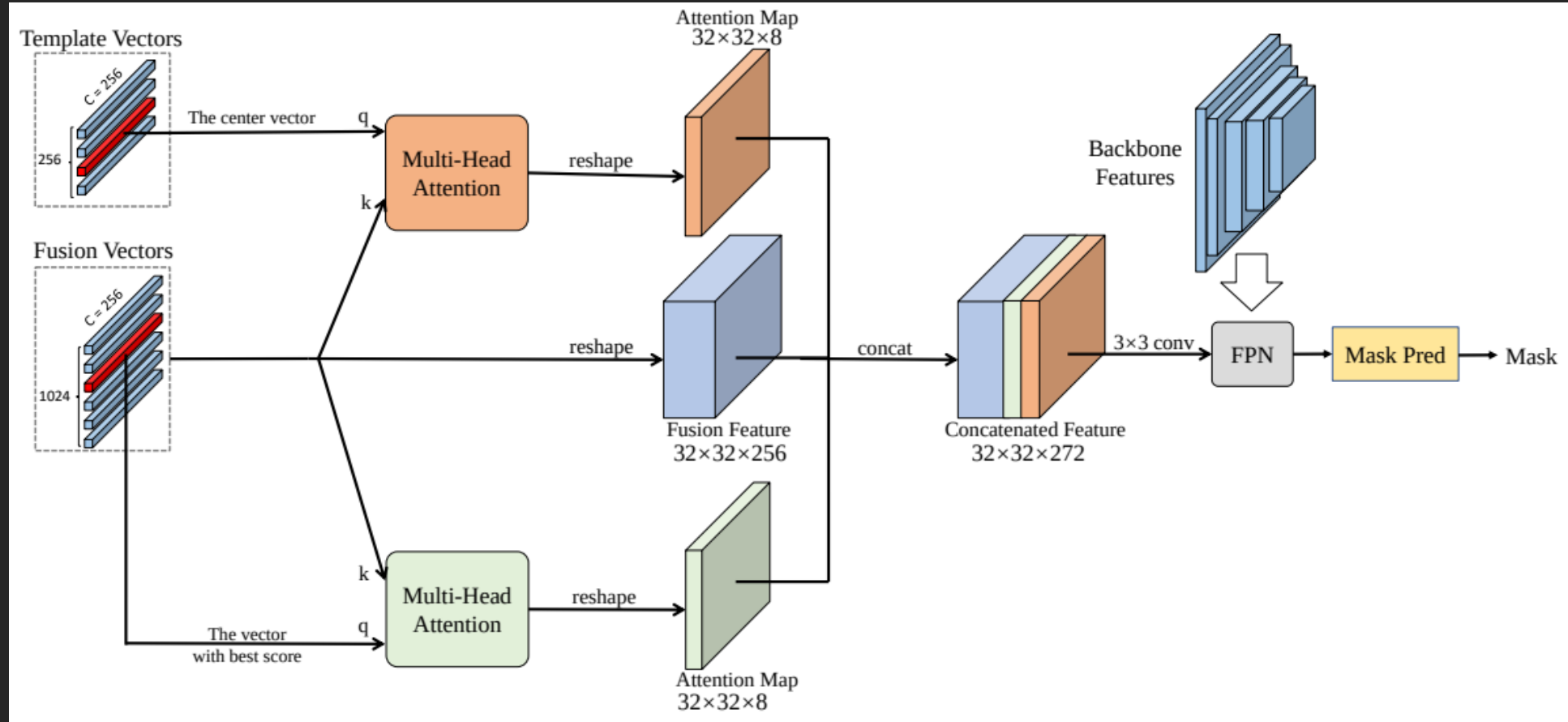
Multi-Template Transformer Tracking

TransT-M Framework



Multi-Template Transformer Tracking

The Segmentation Branch

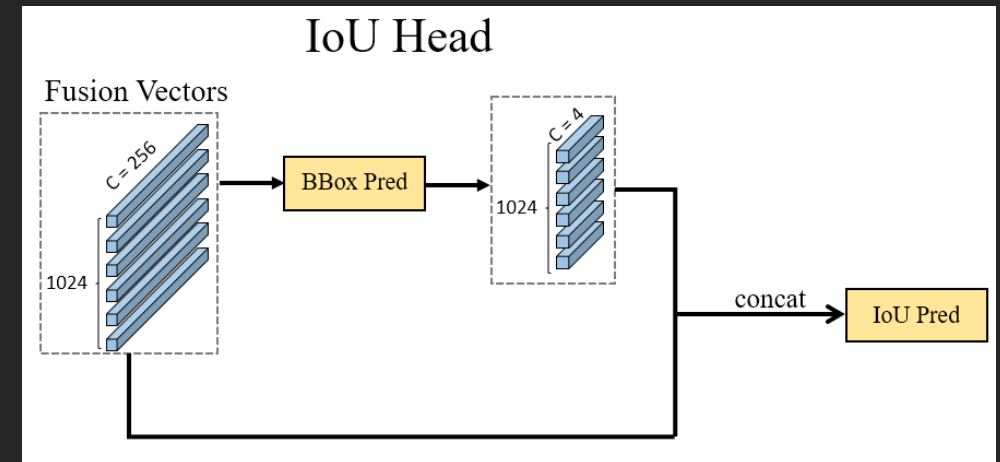


Multi-Template Transformer Tracking

The Multi-template Design

- Concatenate two template in spatial dimension
- Keep the initial template
- Update the other template

The IoU Prediction Head



Multi-Template Transformer Tracking

Experimental Results

Method	seg	EAO	A	R
TransT	-	30.0	47.0	76.8
TransT-SEG	SEG	50.7	74.8	81.7
TransT-M	SEG-na	53.7	73.9	86.2
TransT-M	SEG	55.0	74.2	86.9

Method	Source	LaSOT [37]			TrackingNet [38]			GOT-10k [39]		
		AUC	P_{Norm}	P	AUC	P_{Norm}	P	AO	SR _{0.5}	SR _{0.75}
TransT-M	Ours	65.4	73.9	69.6	82.5	87.3	80.0	74.7	85.5	71.3
TransT	Ours	64.9	73.8	69.0	81.4	86.7	80.3	72.3	82.4	68.2
TransT-GOT	Ours	-	-	-	-	-	-	67.1	76.8	60.9

Multi-Template Transformer Tracking

Thanks!