Transformer Tracking

Xin Chen

chenxin3131@mail.dlut.edu.cn

Mentor: Prof. Dong Wang

Dalian University of Technology

Visual Object Tracking

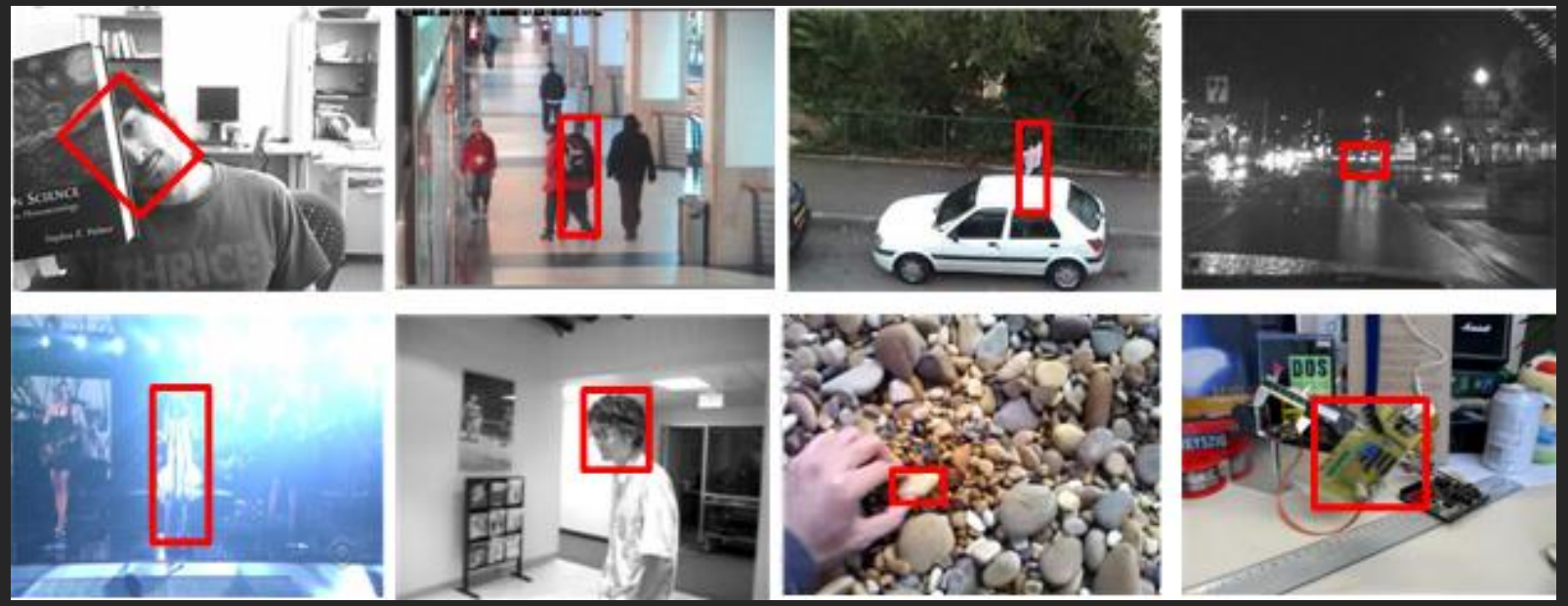
- **Goal**

  Track an arbitrary object in a video given its initial location

  Single-object, **Category-free**

- **Challenges**

  Occlusion, Light Change, Background Clutter, etc.



Papers & Github-repo Recommendation

Papers:

SINT, SiamFC, SiamRPN, SiamDW, SiamRPN++, SiamFC++, Ocean

ECO, ATOM, DiMP, PrDiMP, KYS

TransT, AlphaRefine, LTMU

Github-repo:

[TransT] https://github.com/chenxin-dlut/TransT

[Paper/Code/Results] https://github.com/wangdongdut/Online-Visual-Tracking-SOTA

[Long-term][Paper/Code/Results] https://github.com/wangdongdut/Long-term-Visual-Tracking

[PaperWriting] https://github.com/wangdongdut/PaperWriting

[PyTracking] https://github.com/visionml/pytracking

[PySOT] https://github.com/STVIR/pysot

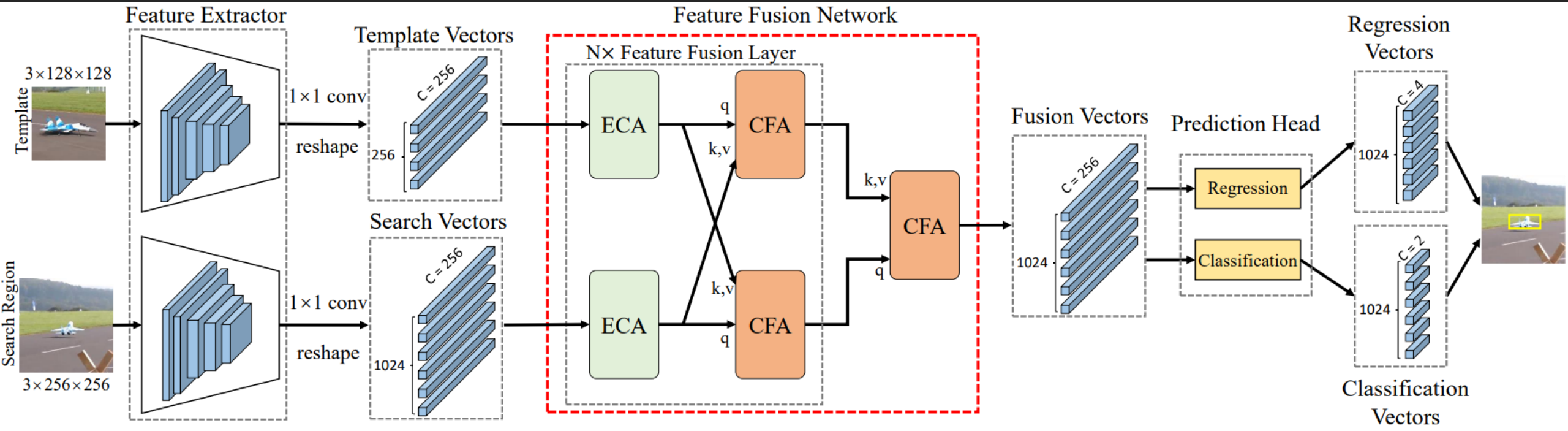
[TracKit] https://github.com/researchmm/TracKit

[OPE-Evaluation] https://github.com/got-10k/toolkit

[VOT] https://github.com/votchallenge

Transformer Tracking

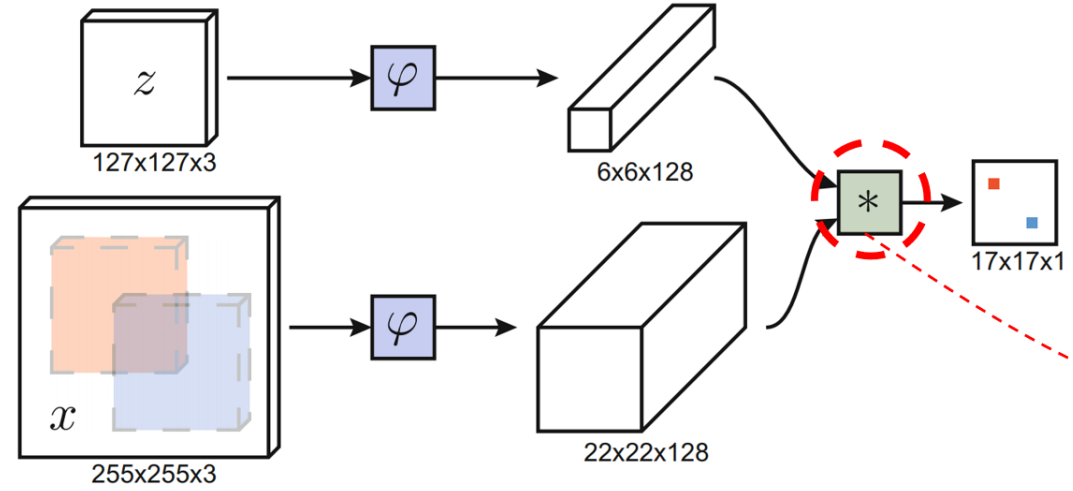
Transformer Tracking (TransT)

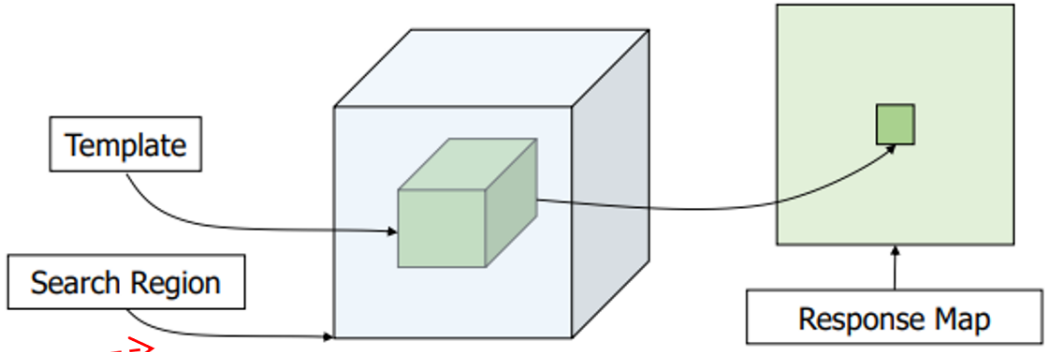


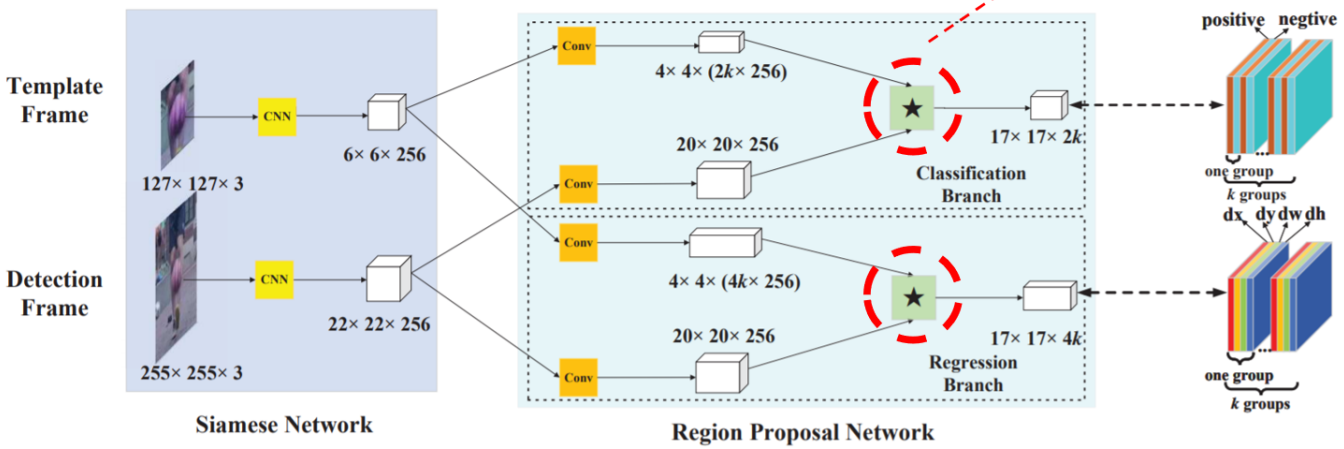
Xin Chen, Bin Yan, Jiawen Zhu, Dong Wang, Xiaoyun yang, Huchuan Lu. Transformer Tracking. CVPR, 2021.

➢ Code: https://github.com/chenxin-dlut/TransT

Transformer Tracking

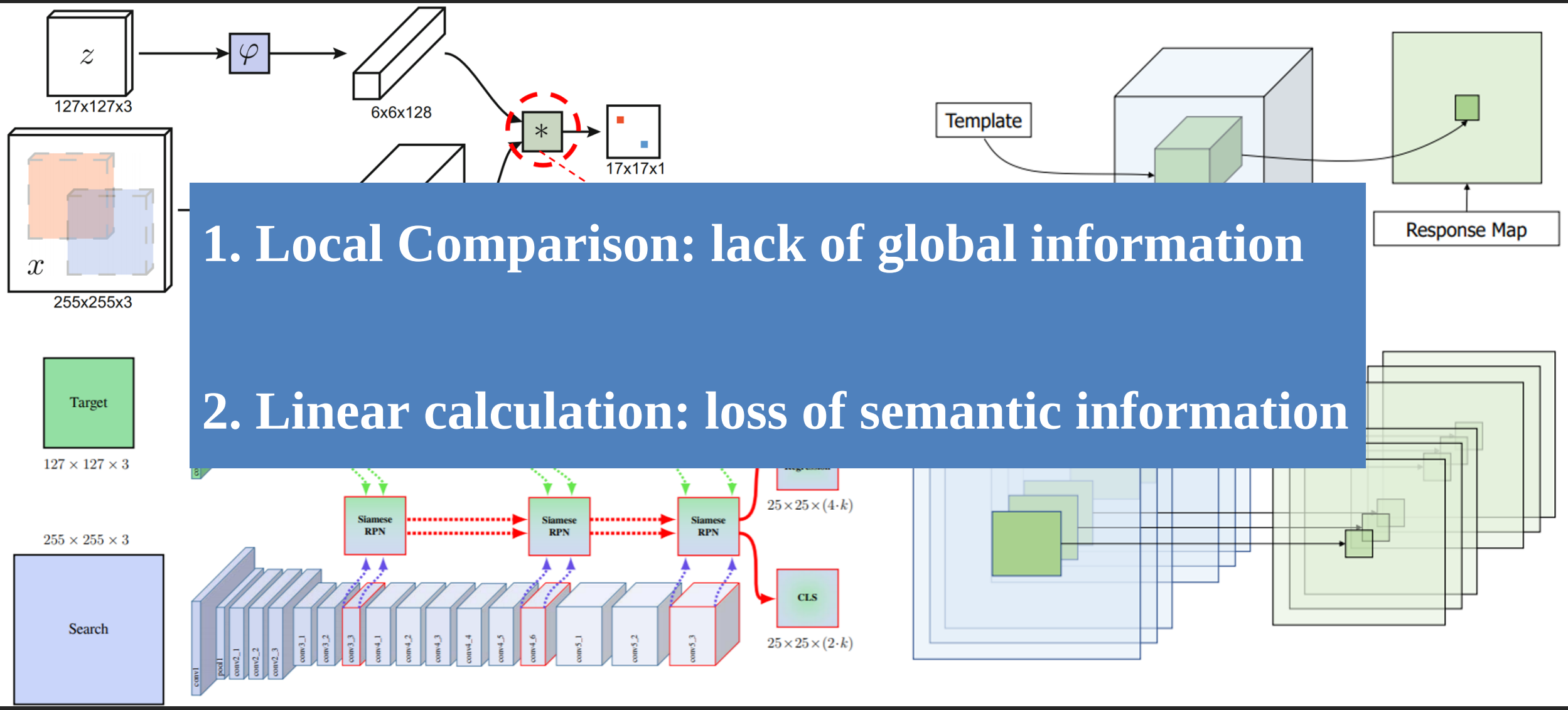






Correlation!

Transformer Tracking

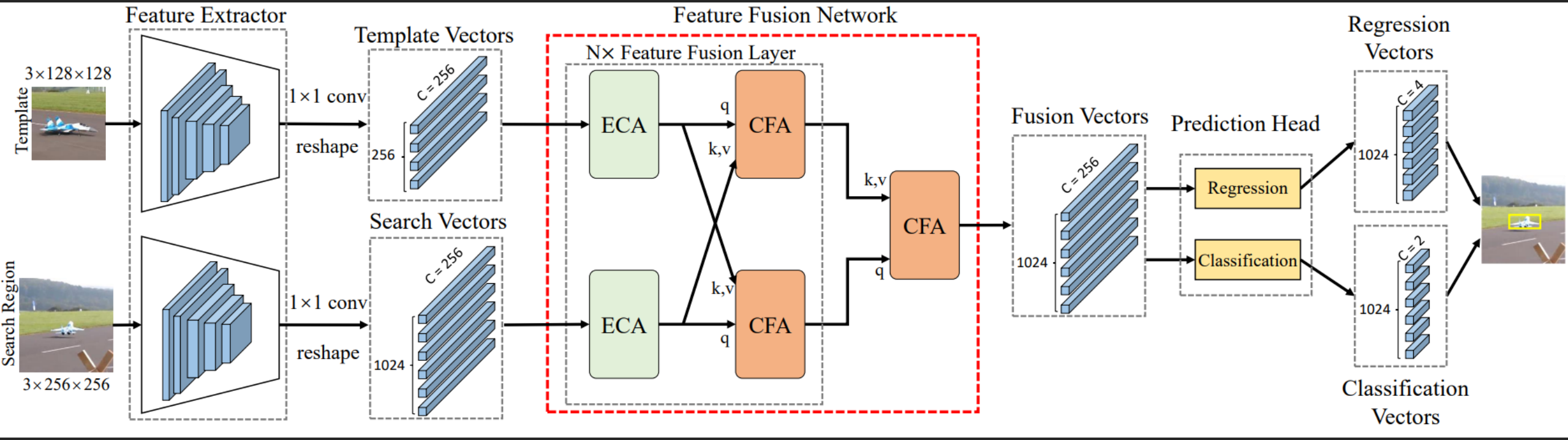


1. Local Comparison: lack of global information

2. Linear calculation: loss of semantic information

Transformer Tracking

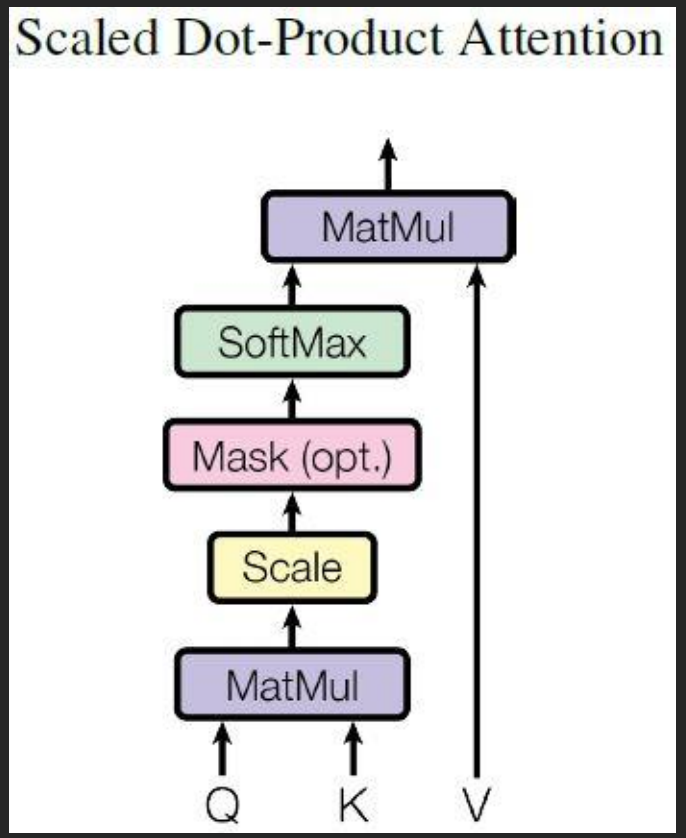
Our TransT Framework



Transformer Tracking

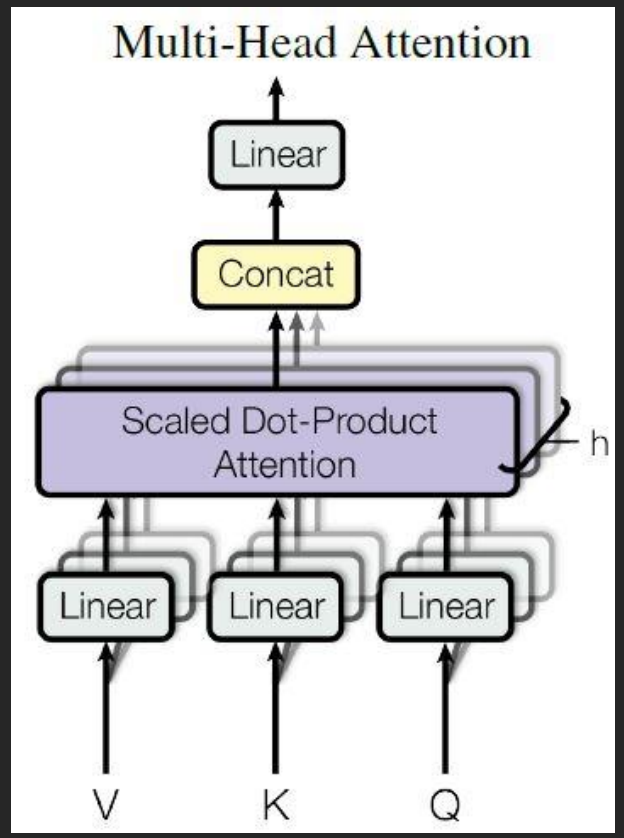
➢ Attention to Replace "Correlation"

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



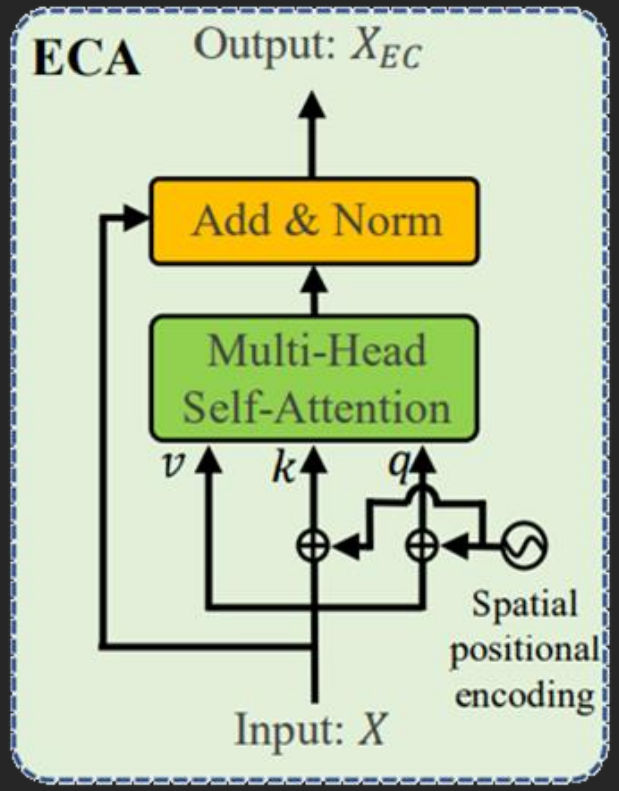
$$\text{MultiHead}(\mathbf{Q},\mathbf{K},\mathbf{V}) = \text{Concat}(\mathbf{H}_1,...,\mathbf{H}_{n_h})\mathbf{W}^O$$

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

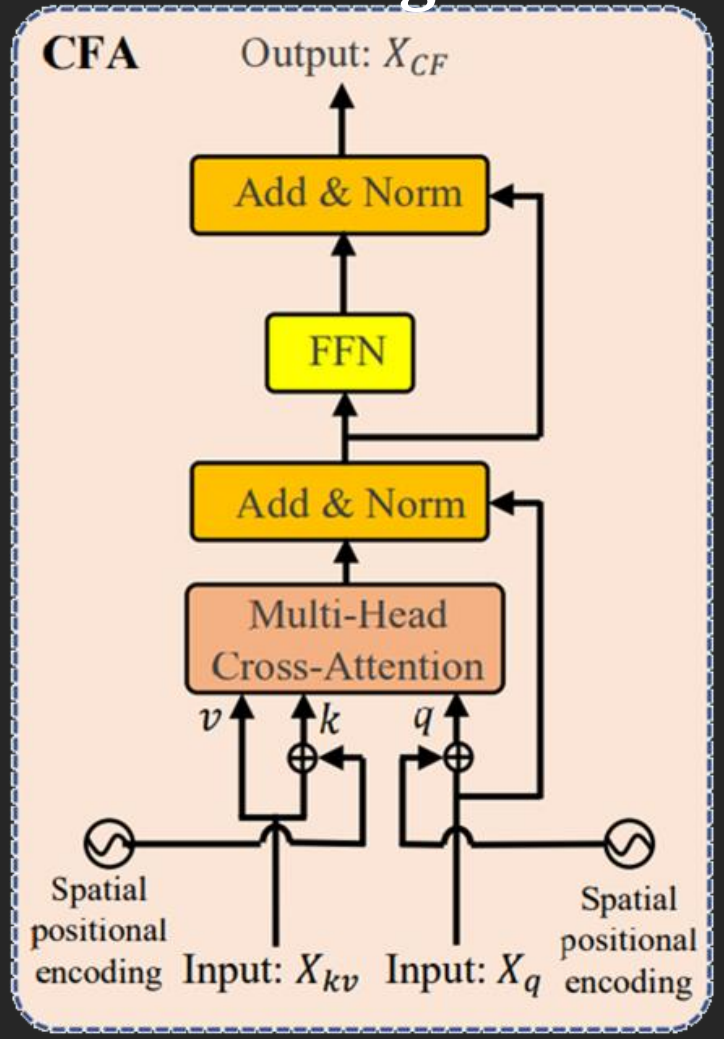


Transformer Tracking

Ego-Context Augment Module



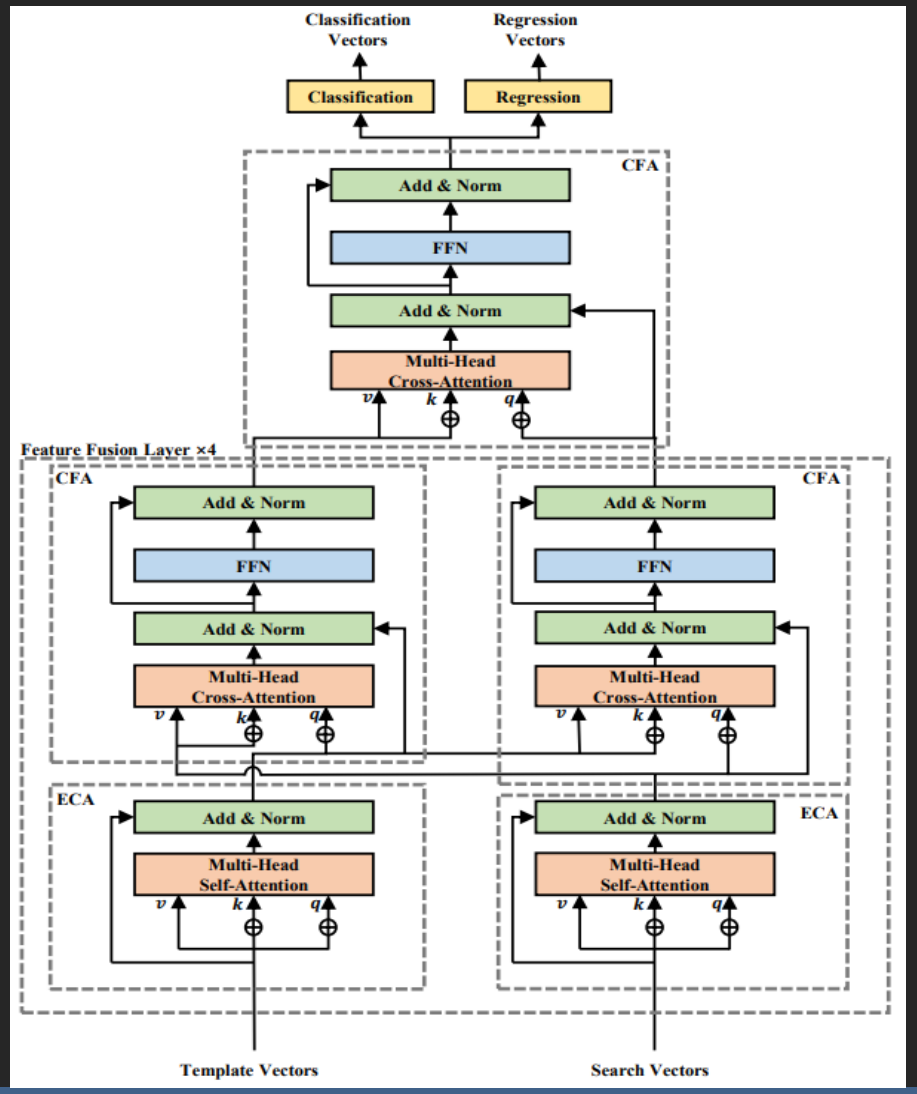
Cross-Feature Augment Module



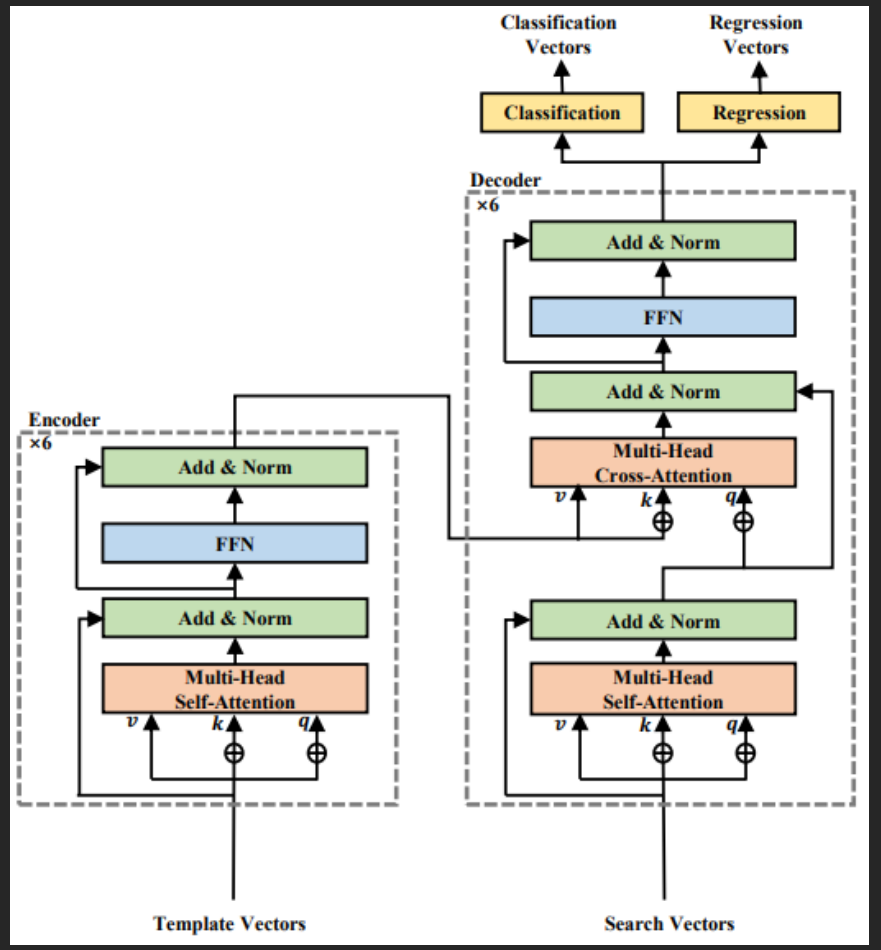
- ✓ ECA based on self-attention and CFA based on cross-attention
- ✓ CFA performs feature fusion, retaining rich semantic information
- ✓ ECA and CFA establish dependence between long distance features and aggregate global information

Transformer Tracking

Our Feature Fusion Network

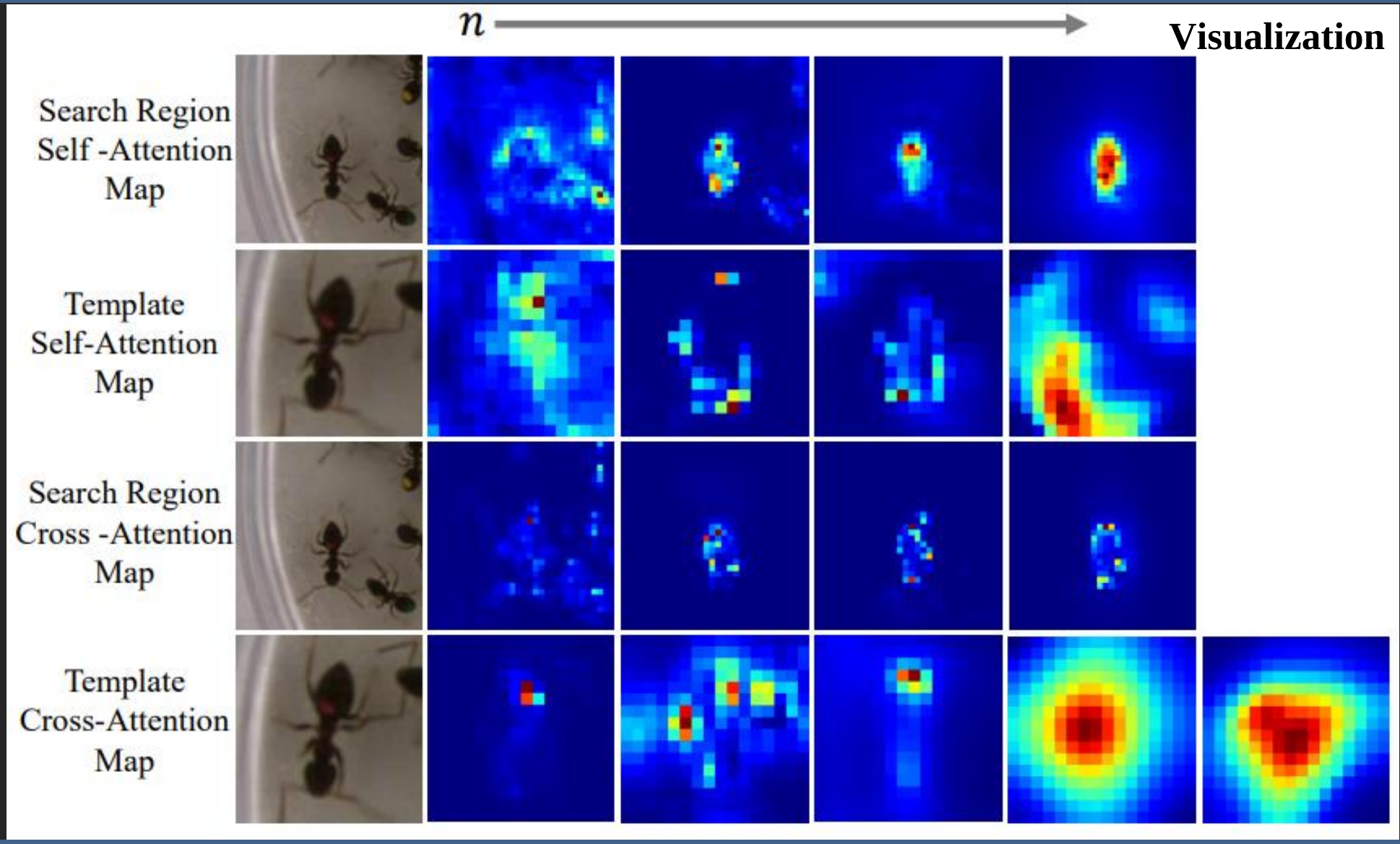


Original Transformer



Similar with DETR

Transformer Tracking



Transformer Tracking

Transformer Tracking

Large-scale Benchmark Results

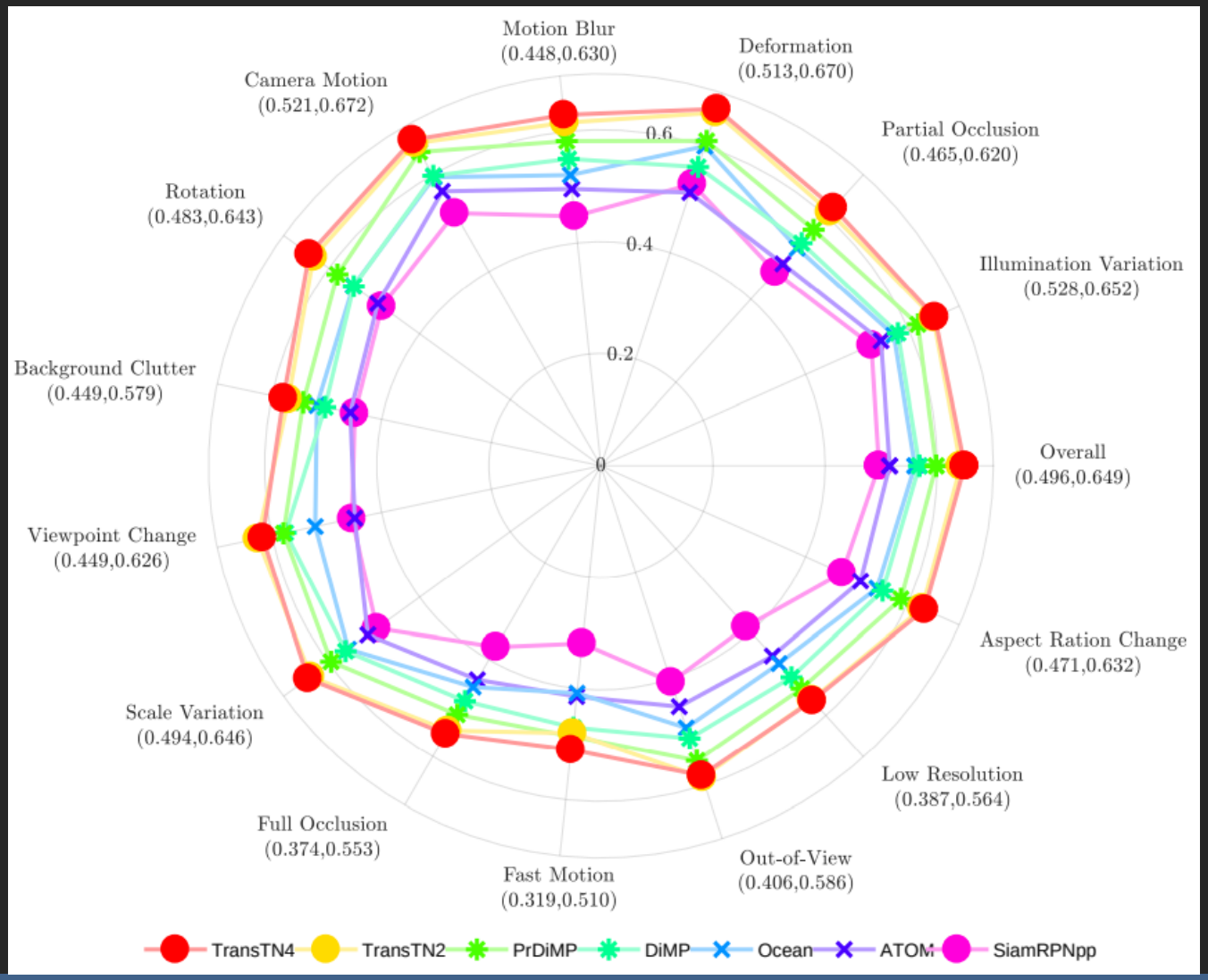
Method	Source	LaSOT [14]			TrackingNet [30]			GOT-10k [19]		
		AUC	P_{Norm}	P	AUC	P_{Norm}	P	AO	$SR_{0.5}$	$SR_{0.75}$
TransT	Ours	**64.9**	**73.8**	**69.0**	**81.4**	**86.7**	**80.3**	**72.3**	**82.4**	**68.2**
TransT-N2	Ours	64.2	**73.5**	68.2	80.9	**86.4**	79.2	**69.9**	**80.1**	**65.9**
TransT-GOT	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

⟸ 46*fps*

⟸ 66*fps*

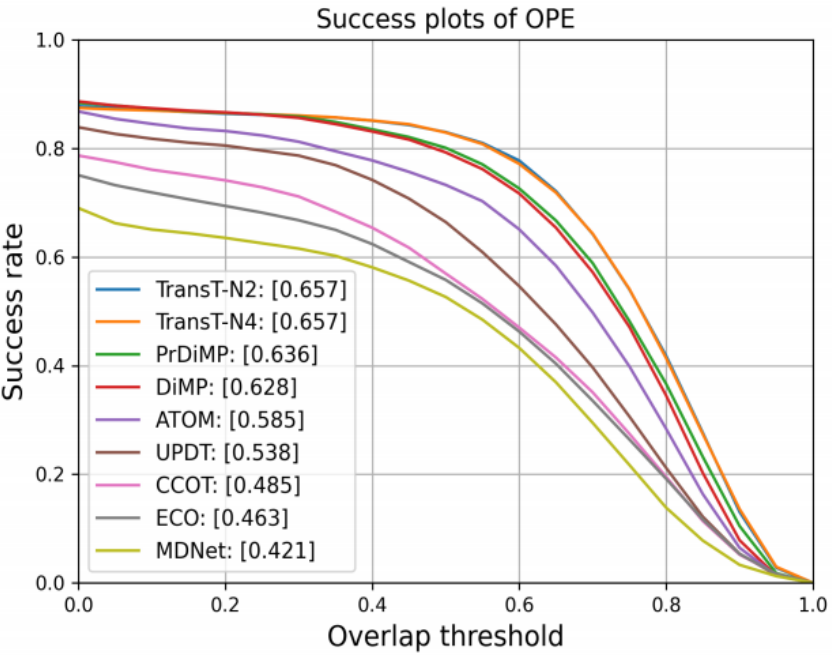
Transformer Tracking

LaSOT Attribute

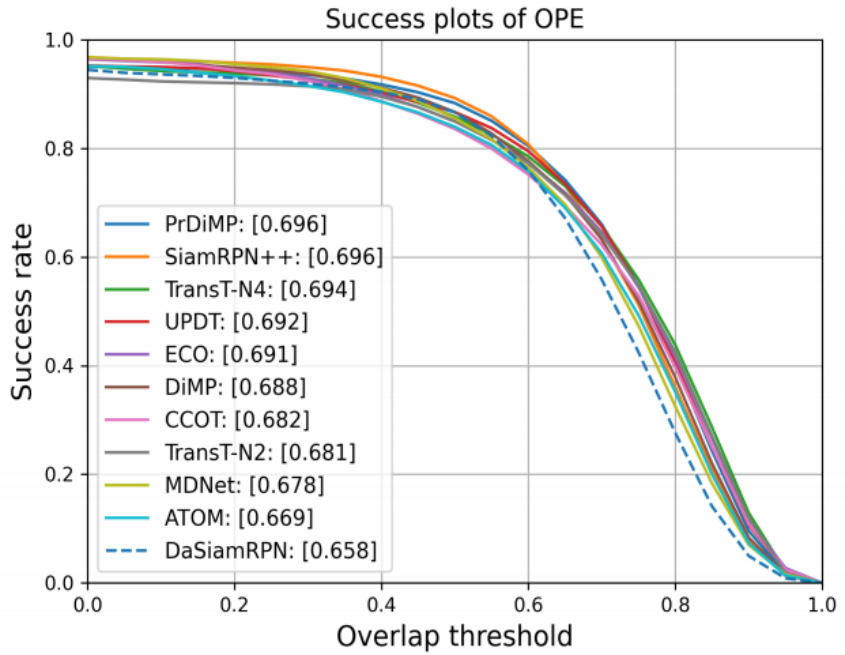


Transformer Tracking

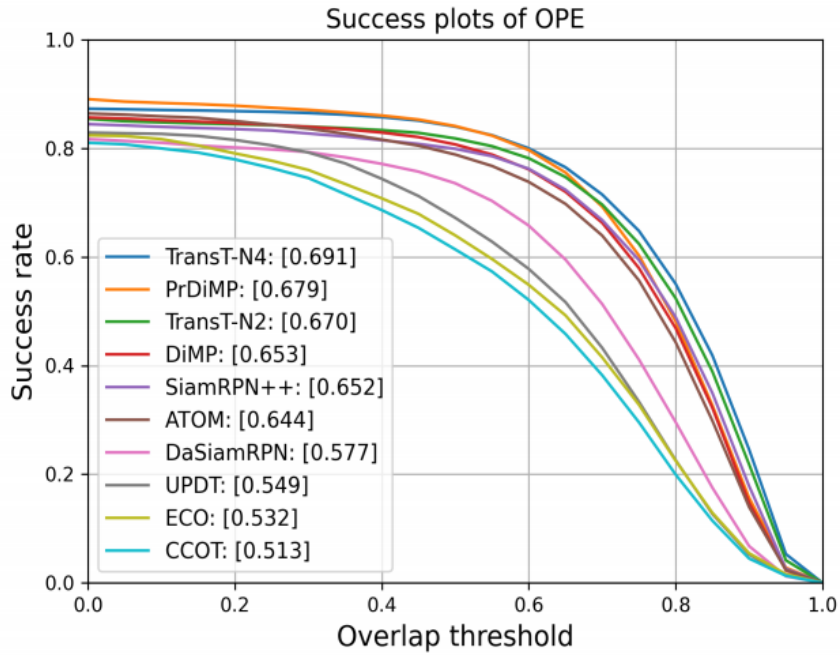
Results on NFS, OTB2015 and UAV123



(a) NFS



(b) OTB2015



(c) UAV123

Transformer Tracking

Results on VOT2020

- TransT-N2 + Alpha-Refine: EAO 48.8
- TransT-N4 + Alpha-Refine: EAO 47.2
- TransT-Mask (add a mask branch on TransT): EAO 49.5

Transformer Tracking

Ablation Study

Method	LaSOT [14]			TrackingNet [30]			GOT-10k [19]		
	AUC	P_{Norm}	P	AUC	P_{Norm}	P	AO	$SR_{0.5}$	$SR_{0.75}$
TransT	**64.9**	**73.8**	**69.0**	**81.4**	**86.7**	**80.3**	**72.3**	**82.4**	**68.2**
TransT-np	62.9	71.5	66.9	81.1	86.4	80.0	71.5	81.5	67.5
TransT(ori)	62.3	71.1	66.2	81.3	86.1	78.9	70.3	80.2	65.8
TransT(ori)-np	60.9	69.4	64.8	80.9	85.6	78.4	68.6	78.2	65.1

Transformer Tracking

Ablation Study

Method	ECA	CFA	Correlation	LaSOT [14]			TrackingNet [30]			GOT-10k [19]		
				AUC	P_{Norm}	P	AUC	P_{Norm}	P	AO	$SR_{0.5}$	$SR_{0.75}$
TransT	√	√		**64.9**	**73.8**	**69.0**	**81.4**	**86.7**	**80.3**	**72.3**	**82.4**	**68.2**
TransT		√		62.9	71.9	66.2	81.1	86.2	79.1	70.6	81.2	65.7
TransT	√		√	57.7	65.4	59.5	77.5	82.2	74.0	62.8	72.2	54.8
TransT			√	47.7	48.6	41.7	68.8	71.4	60.9	50.9	58.0	33.3
TransT-np	√	√		62.9	71.5	66.9	81.1	86.4	80.0	71.5	81.5	67.5
TransT-np		√		61.0	69.6	64.5	80.0	85.0	77.9	68.1	78.3	64.0
TransT-np	√		√	57.3	65.2	58.8	76.2	80.8	72.8	61.4	70.7	53.7
TransT-np			√	35.3	17.9	20.1	46.5	40.3	27.4	38.2	36.8	7.0

Transformer Tracking

Conclusion

➢ A New Transformer-based Tracking Framework

1. A simple and efficient baseline.
2. High performance.
3. Completely offline.
4. Using same hy-parameter and same model for all test-sets.

Unsolved Challenges

- **Distractor**

Occlusion and appearance change when distractors exists.

- **Out of the Search Region**

4 times search region size cannot cover all situations, but a larger search region will aggravate the challenge of distractors.

Transformer Tracking

- Xin Chen, et al. Transformer Tracking.
- Ning Wang, et al. Transformer Meets Tracker: Exploiting Temporal Context for Robust Visual Tracking.
- Bin Yan, et al. Learning Spatio-Temporal Transformer for Visual Tracking.
- Peize Sun, et al. TransTrack: Multiple-Object Tracking with Transformer.
- Tim Meinhardt, et al. TrackFormer: Multi-Object Tracking with Transformers.

Thanks!

(Q & A)