TransT

(Transformer Tracking, 2021)

김동원

목차

Siamese base method

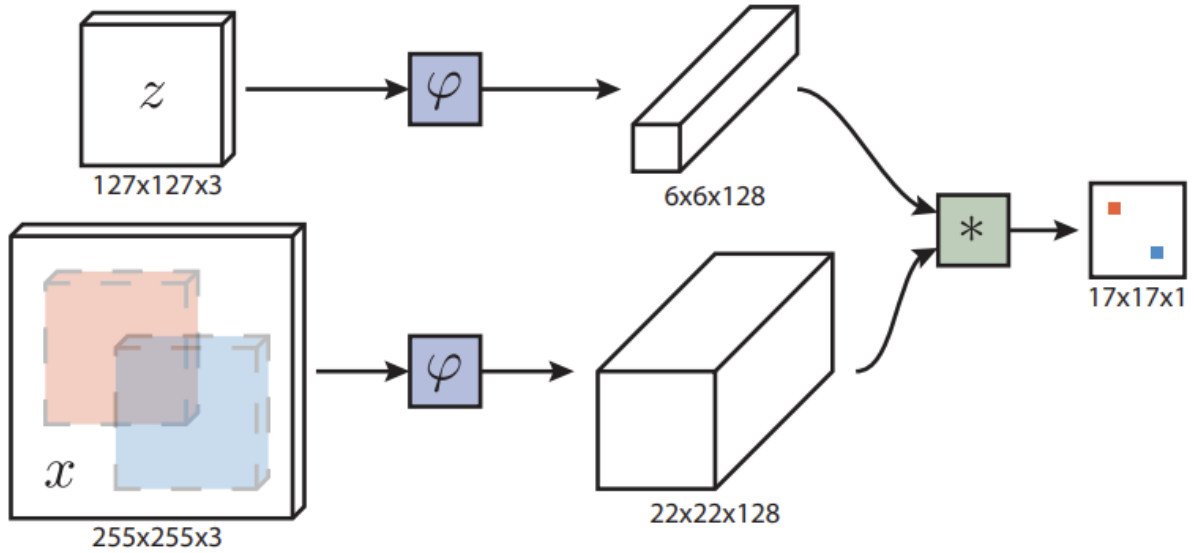
TransT 개요

TransT의 구조

ECA, CFA

Siamese base method

Template 이미지와 search region이미지를 모델을 통해 특징을 추출하여 상관관계를 통해 tracking하는 방법(SiamFC, SiamRPN, ATOM)

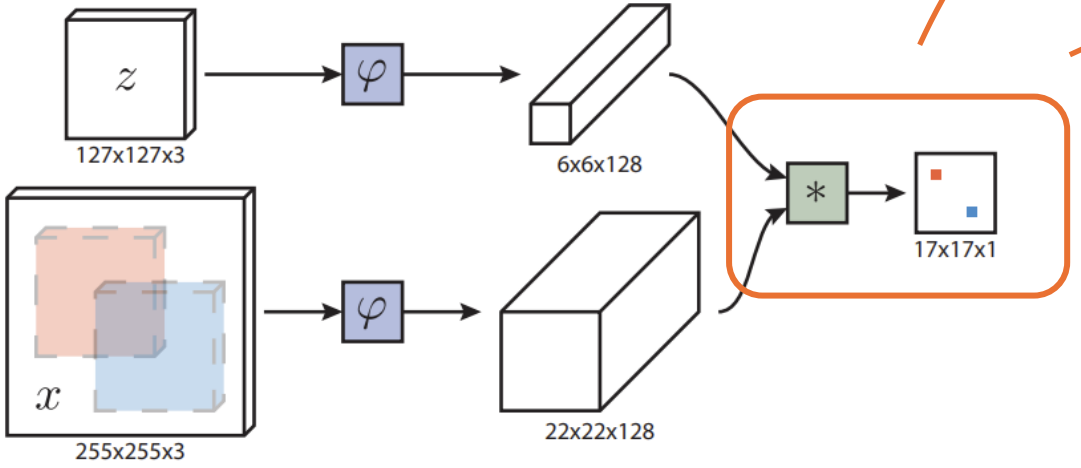


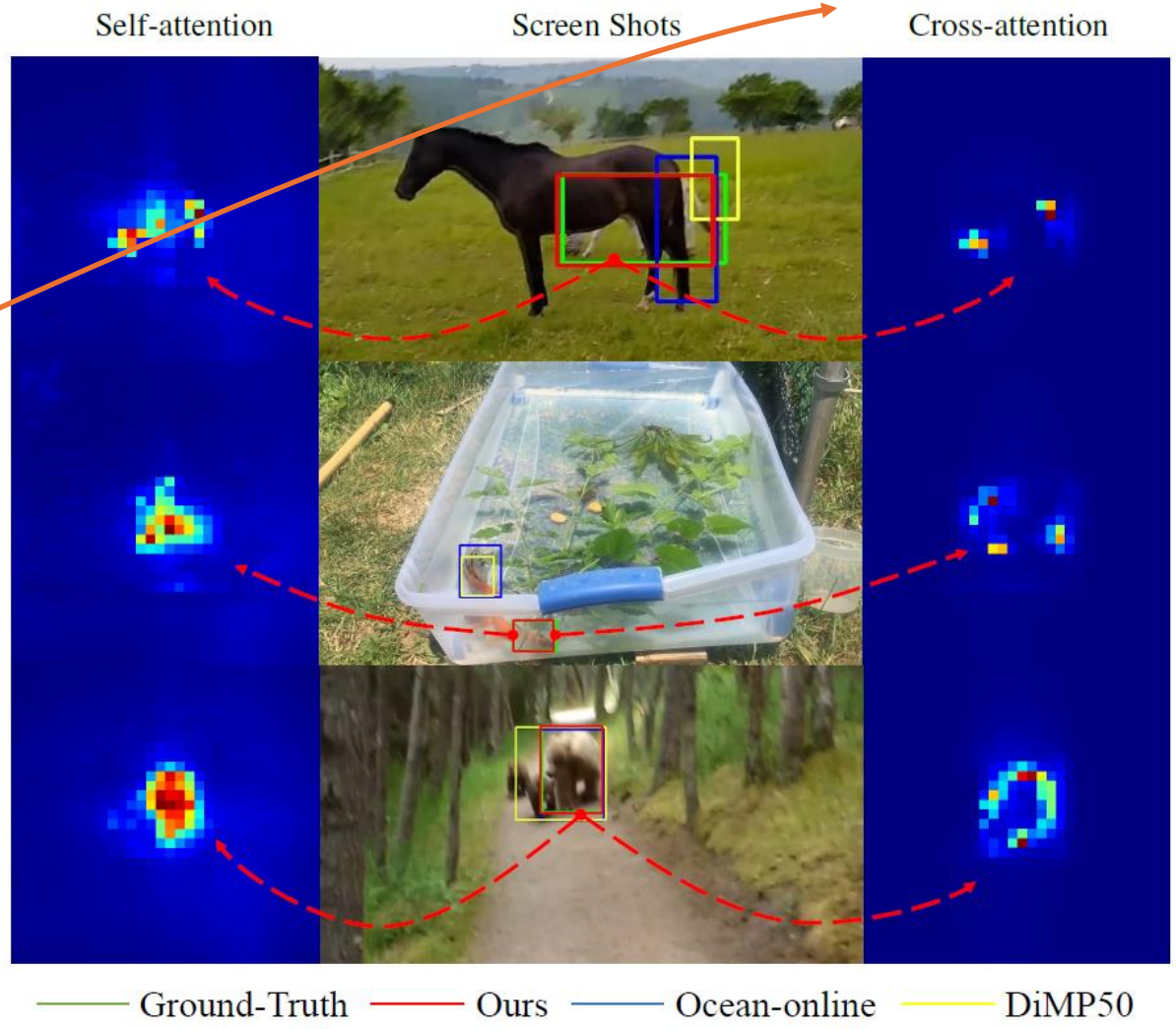
Image(**Fully Convolutional Siamese Networks for Object Tracking, 2021**)

TransT는 〉〉〉 상관관계를 사용하는 방법은 의미적 정보, 복잡한 패턴에 대한 정보가 손실 되므로 transformer의 attention 매커니즘을 적용한 방법

TransT 개요

Siamese network 기반 방법에서 상관관계 대신, transformer의 attention 매커니즘을(ECA, CFA) 적용한 방법이다.



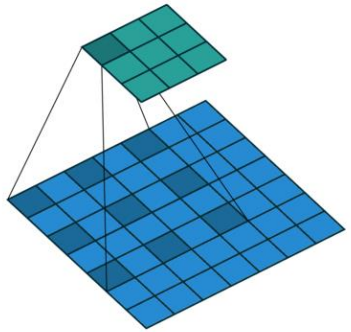


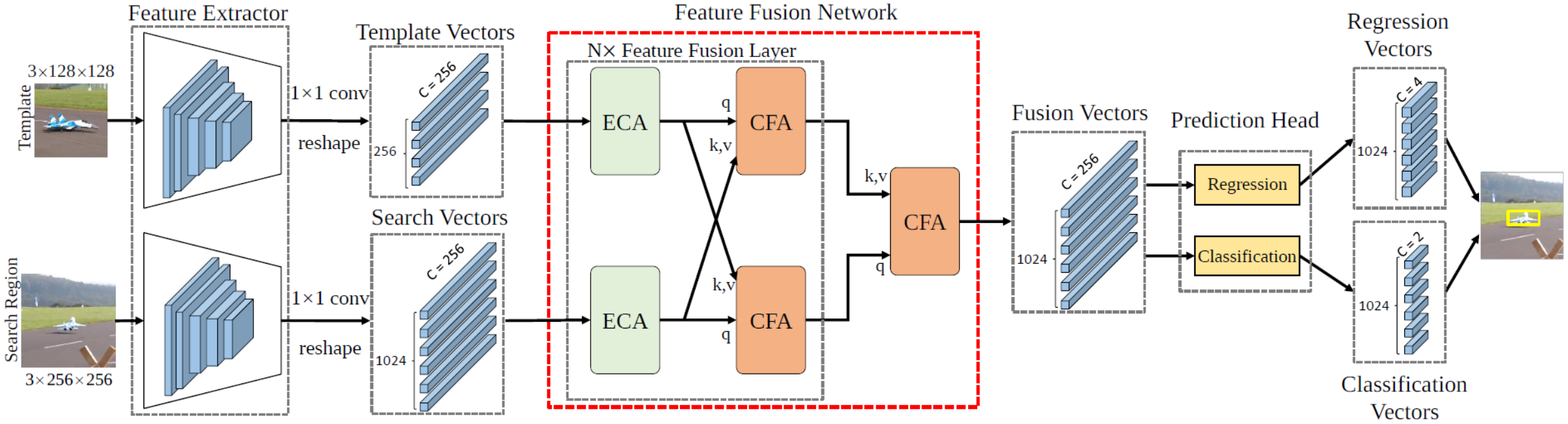
TransT 구조

ResNet 사용

stride 2->1

dilated convolution





T: $3 * H_{z_0} * W_{z_0}$ → $1024 * H_{z_0}/8 * W_{z_0}/8$ → $256 * (H_{z_0}/8 * W_{z_0}/8)$

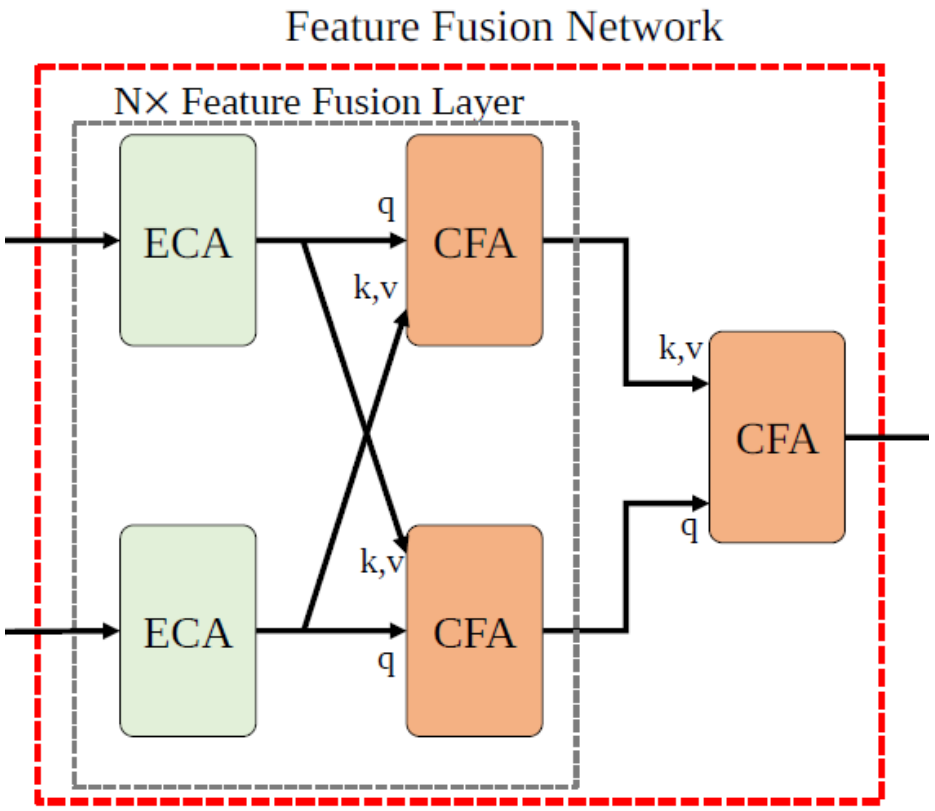
S: $3 * H_{x_0} * W_{x_0}$ → $1024 * H_{x_0}/8 * W_{x_0}/8$ → $256 * (H_{x_0}/8 * W_{x_0}/8)$

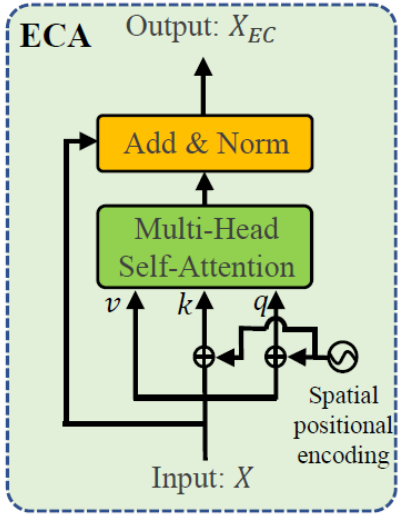
256개의 feature vector

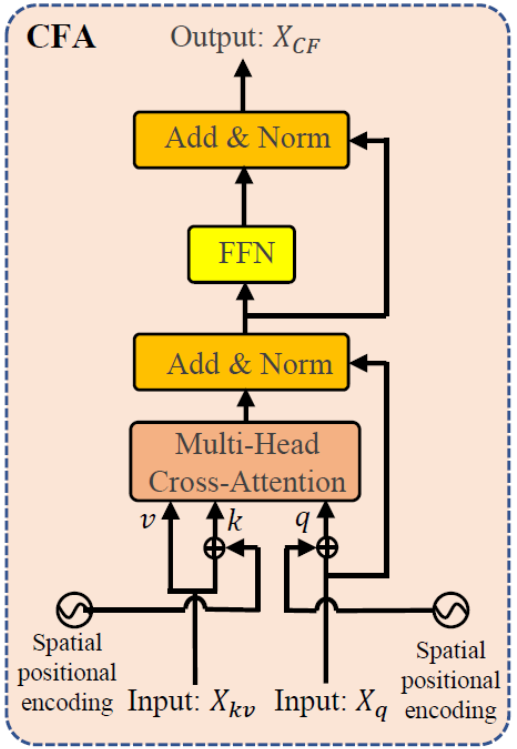
ECA, CFA

Ego-context augment(ECA), Cross-Feature augment(CFA) Modules

ECA: template와 search region을 학습, CFA: template와 search region이 합쳐져 연관성 학습



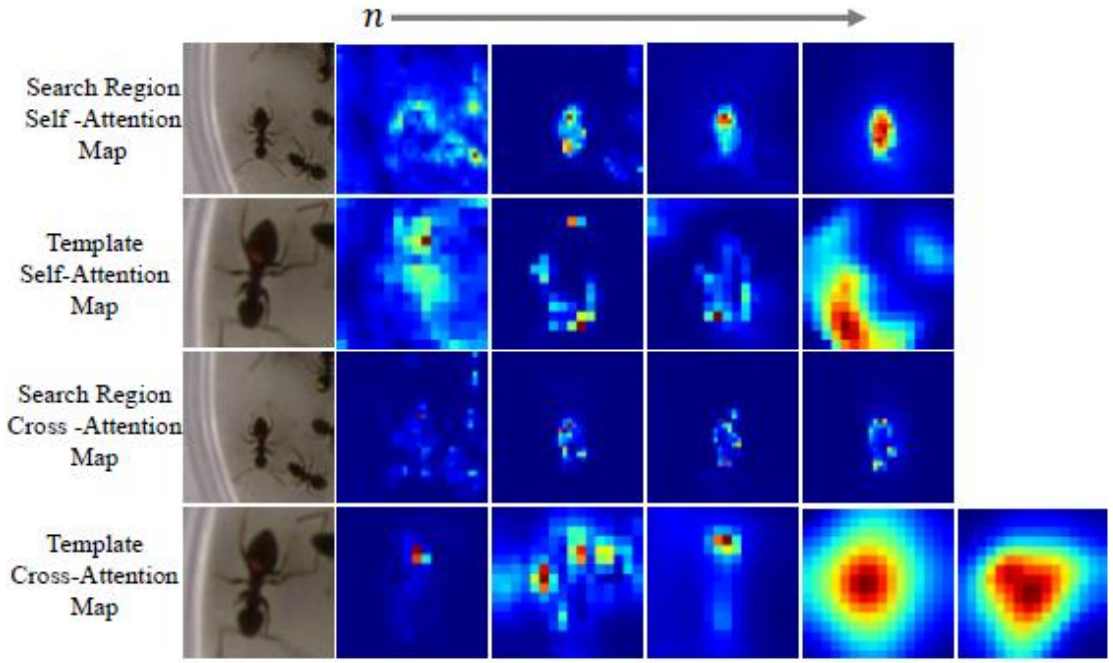




ECA, CFA

Ego-context augment(ECA), Cross-Feature augment(CFA) Modules

ECA: template와 search region을 학습, CFA: template와 search region이 합쳐져 연관성 학습



초반에는 넓은 범위의 정보를 탐색하고 더 깊은 레이어일 수록 template이미지에 집중된다.

Training loss

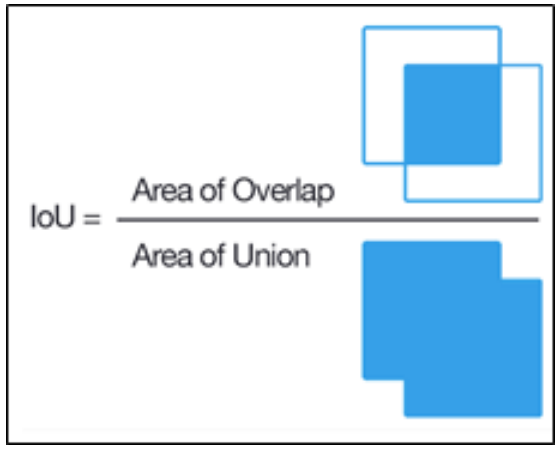
Classification & regression loss 사용

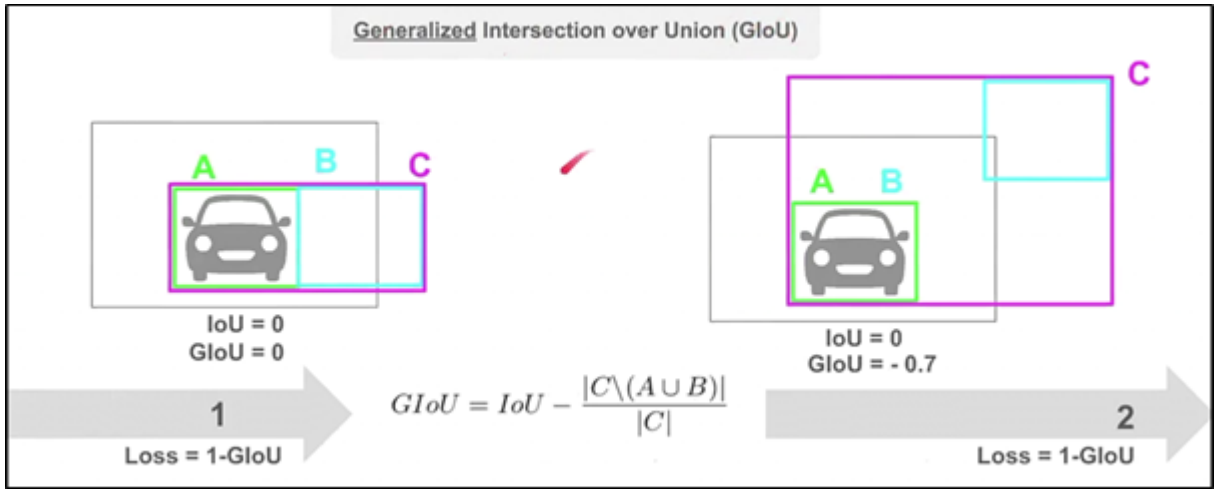
$$\mathcal{L}_{cls} = -\sum_{j}[y_j \log(p_j) + (1 - y_j)\log(1 - p_j)],$$

프레임 내부의 물체 여부 판단

$$\mathcal{L}_{reg} = \sum_{j} \mathbb{1}_{\{y_j=1\}}[\lambda_G \mathcal{L}_{GIoU}(b_j, \hat{b}) + \lambda_1 \mathcal{L}_1(b_j, \hat{b})],$$

물체의 예측과 실제 bounding box의 차이를 계산





Training loss

기본적인 transformer구조를 사용한 모델과 성능 비교 (np: post processing을 적용하지 않은 모델)

Post processing: cosine window penalty를 사용하여 바운딩 박스의 위치와 크기 변화를 자연스럽게 만든다

transT(ori) fusion network를 transformer블록으로 대체한 경우(encoder-template, decoder-search region)

Table 2. Ablation study on TrackingNet, LaSOT, and GOT-10k. The best results are shown in the **red** font.

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

Training loss

Correlation vs attention

Correlation -> depth-wise correlation연산

Table 3. Comparison with correlation on TrackingNet, LaSOT, and GOT-10k. The best results are shown in the **red** font.

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

상관관계로 대체하였을 경우 성능이 크게 하락한다. ECA, CFA가 성능에 중요한 영향을 미친다.

참고자료

Template, search region image link