Reproducibility Study on Adversarial Attacks Against Robust Transformer Trackers

Fatemeh Nourilenjan Nokabadi[1,2,3], Jean-François Lalonde[1,2], Christian Gagné[1,2,3,4]

[1]IID, [2]Université Laval, [3]Mila, [4]Canada CIFAR AI Chair

NEURAL INFORMATION PROCESSING SYSTEMS

Motivations

1- How do transformer-based trackers respond to adversarial attacks?

2- How does the performance of different adversarial attacks vary on tracking datasets as attack parameters are modified?

3- How does the performance of transformer-based trackers compare to other backbone architectures under identical adversarial attack conditions?

	ROMTrack	MixFormerM	TransT	DiMP	PrDiMP	SiamRPN	DaSiamRPN
SPARK(white-box)	N/A	N/A	A	N/A	N/A	A	A
RTAA (white-box)	N/A	N/A	A	N/A	N/A	A	A
IoU (black-box)	A	A	A	A	A	A	A
CSA (black-box)	A	A	A	N/A	N/A	A	A

A. Adversarial Attacks per Tracker Output

Goal: Evaluate the difference before and after the attack when one of the tracker's outputs (bounding box or binary mask) is measured.

TransT-SEG Performance after Attacks

		EAO			Accuracy			Robustness		
Stack	Method	Clean	Attack	Drop	Clean	Attack	Drop	Clean	Attack	Drop
STB	CSA	0.299	0.285	4.68%	0.472	0.477	-1.06%	0.772	0.744	3.63%
	IoU	0.299	0.231	22.74%	0.472	0.495	-4.87%	0.772	0.569	26.29%
	RTAA	0.299	0.058	83.28%	0.472	0.431	8.69%	0.772	0.157	79.66%
	SPARK	0.299	0.012	95.99%	0.472	0.244	48.30%	0.772	0.051	93.39%
STS	CSA	0.500	0.458	8.40%	0.749	0.736	1.73%	0.815	0.779	4.42%
	IoU	0.500	0.334	33.20%	0.749	0.710	5.21%	0.815	0.588	27.85%
	RTAA	0.500	0.067	86.60%	0.749	0.533	28.84%	0.815	0.146	82.08%
	SPARK	0.500	0.011	97.80%	0.749	0.266	64.48%	0.815	0.042	94.84%

MixFormerM Performance after Attacks

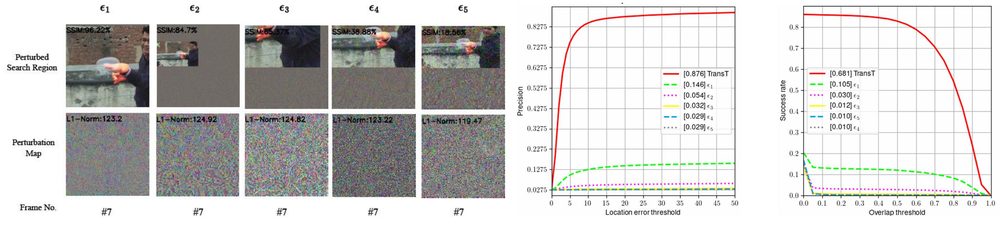
		EAO			Accuracy			Robustness		
Stack	Method	Clean	Attack	Drop	Clean	Attack	Drop	Clean	Attack	Drop
STB	CSA	0.303	0.308	-1.65%	0.479	0.478	0.21%	0.780	0.791	-1.41%
	IoU	0.303	0.246	18.81%	0.479	0.458	4.38%	0.780	0.665	14.74%
STS	CSA	0.589	0.562	4.58%	0.798	0.803	-0.63%	0.880	0.857	2.61%
	IoU	0.589	0.359	39.05%	0.798	0.660	17.30%	0.880	0.677	23.07%

Main Takeaway: The attacks applicable to transformer trackers have more impact on the accuracy of the object mask than the bounding boxes on VOT2022ST dataset.

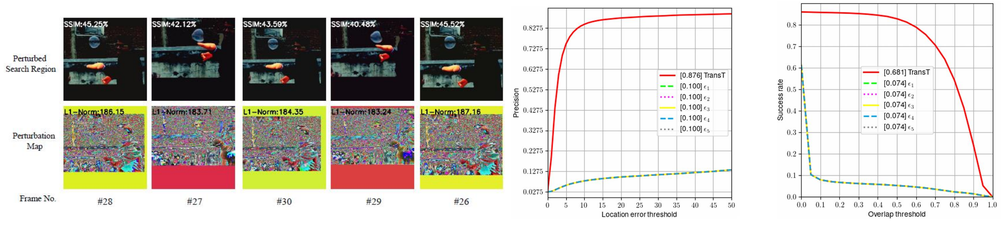
B. Adversarial Attacks per Perturbation Level

Goal: Evaluate the effect of the perturbation level shifts on white-box attacks (SPARK and RTAA) against transformer trackers.

RTAA Performance against TransT Tracker



SPARK Performance against TransT Tracker



Main Takeaways:

1- Increasing the perturbation level on SPARK attack setting results in more super-perturbed regions, i.e. regions with perceptible noise.

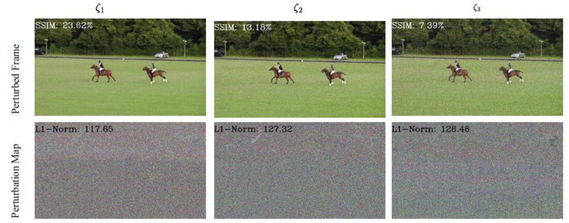
2- Adding the previous perturbations (up to 30 frames) result in more stable performance for SPARK against changes in perturbation levels.

3- For RTAA attack, adding a higher perturbation level generates more perceptible noise for all frames, which damage more the overall tracking performance.

ϵ	No. of frames	SSIM	L1 norm
2.55	7	36.86	176.04
5.1	7	40.96	181.86
10.2	13	41.08	181.33
20.4	13	41.97	182.53
40.8	14	42.53	183.98

C. Adversarial Attack per Upper-Bound

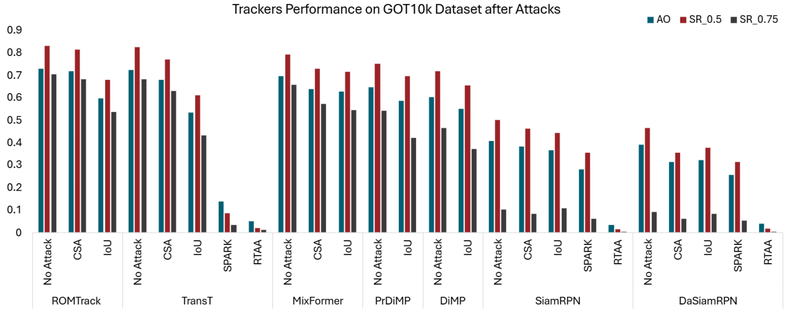
Goal: Evaluate the effect of the upper bound change on black-box attack (IoU) against transformer trackers.



Main Takeaway: The outcome of the IoU attack is sensitive to its initialization. The evaluation process may take a long time due to unsuitable initialization point.

D. Transformer versus Non-transformer Trackers

Goal: Study the adversarial robustness of trackers with different backbones.



Main Takeaways:

1- Despite transformer trackers (ROMTrack, TransT, and MixFormer) showcasing the top-3 performance, their evaluation scores more notably decreased after applying the IoU method.

2- Discriminative trackers also demonstrate a great adversarial robustness and ranked immediately after the transformer trackers on GOT10k dataset.