Adversarial Bounding Boxes Generation (ABBG) Attack against Visual Object Trackers



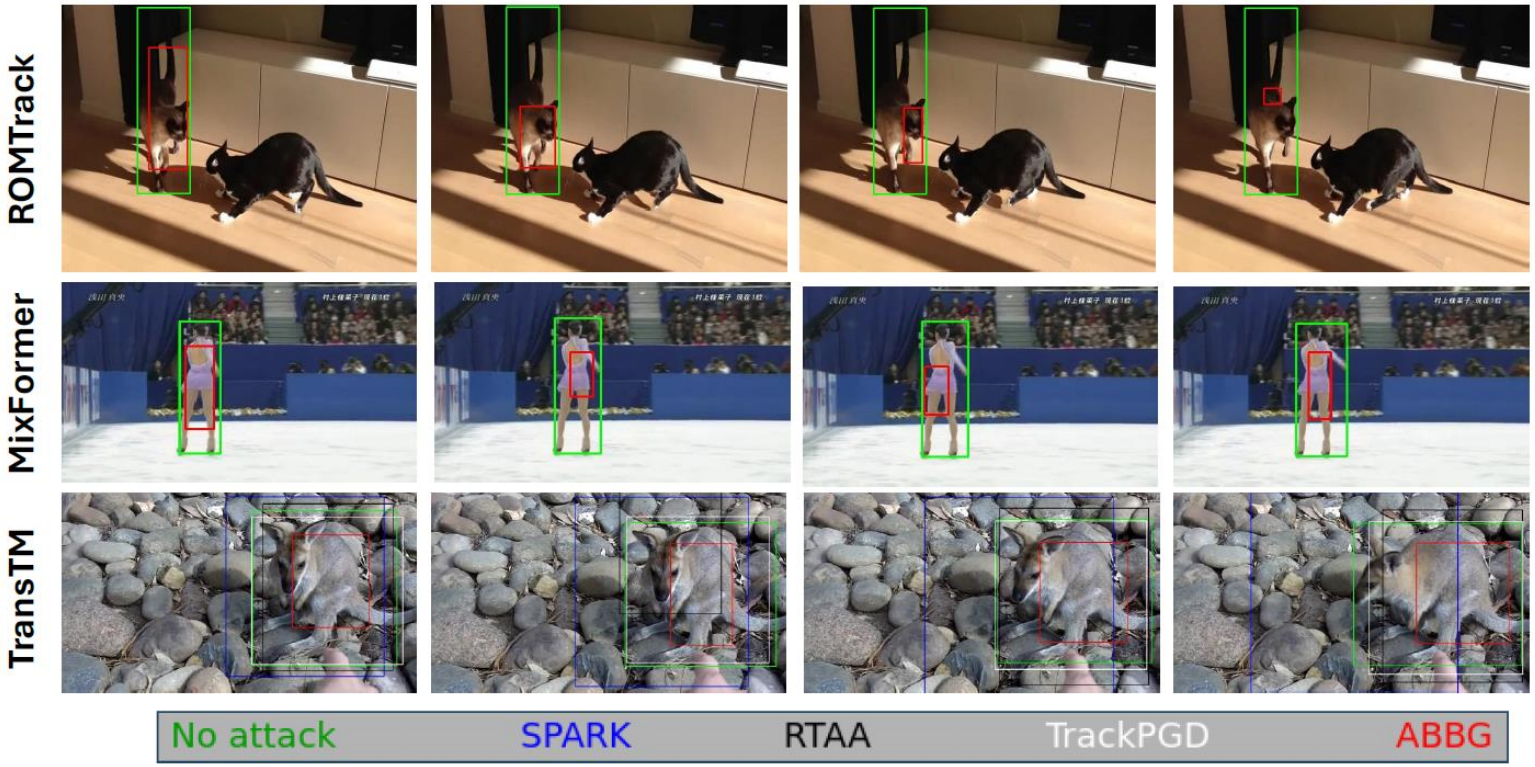
Fatemeh Nourilenjan Nokabadi[1,2,3], Jean-François Lalonde[1,2], Christian Gagné[1,2,3,4]

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

The 3rd New Frontiers in Adversarial Machine Learning
AdvML Frontiers @NeurIPS2024, Vancouver, Canada

Adversarial Attacks against Single Object Trackers

We present a novel white-box approach to attack visual object trackers with transformer backbones using only one bounding box.



Research Contributions

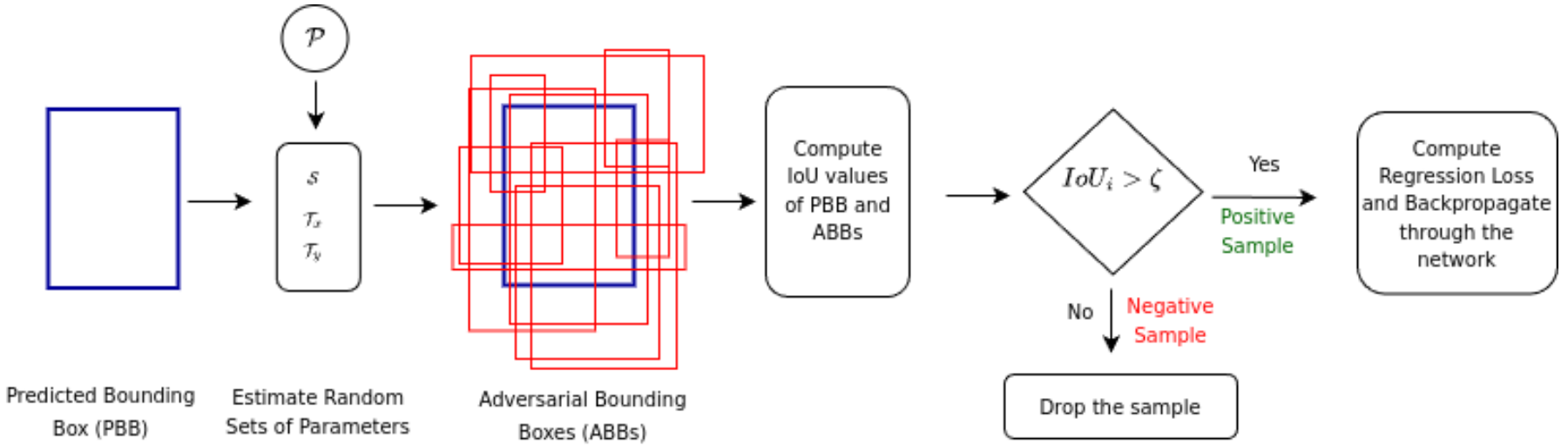
1. AABG attack uses only a single bounding box to challenge the object trackers robustness against adversarial perturbations in a white-box setting.

2. AABG attack is ranked first in several tracking datasets per at least one evaluation metric.

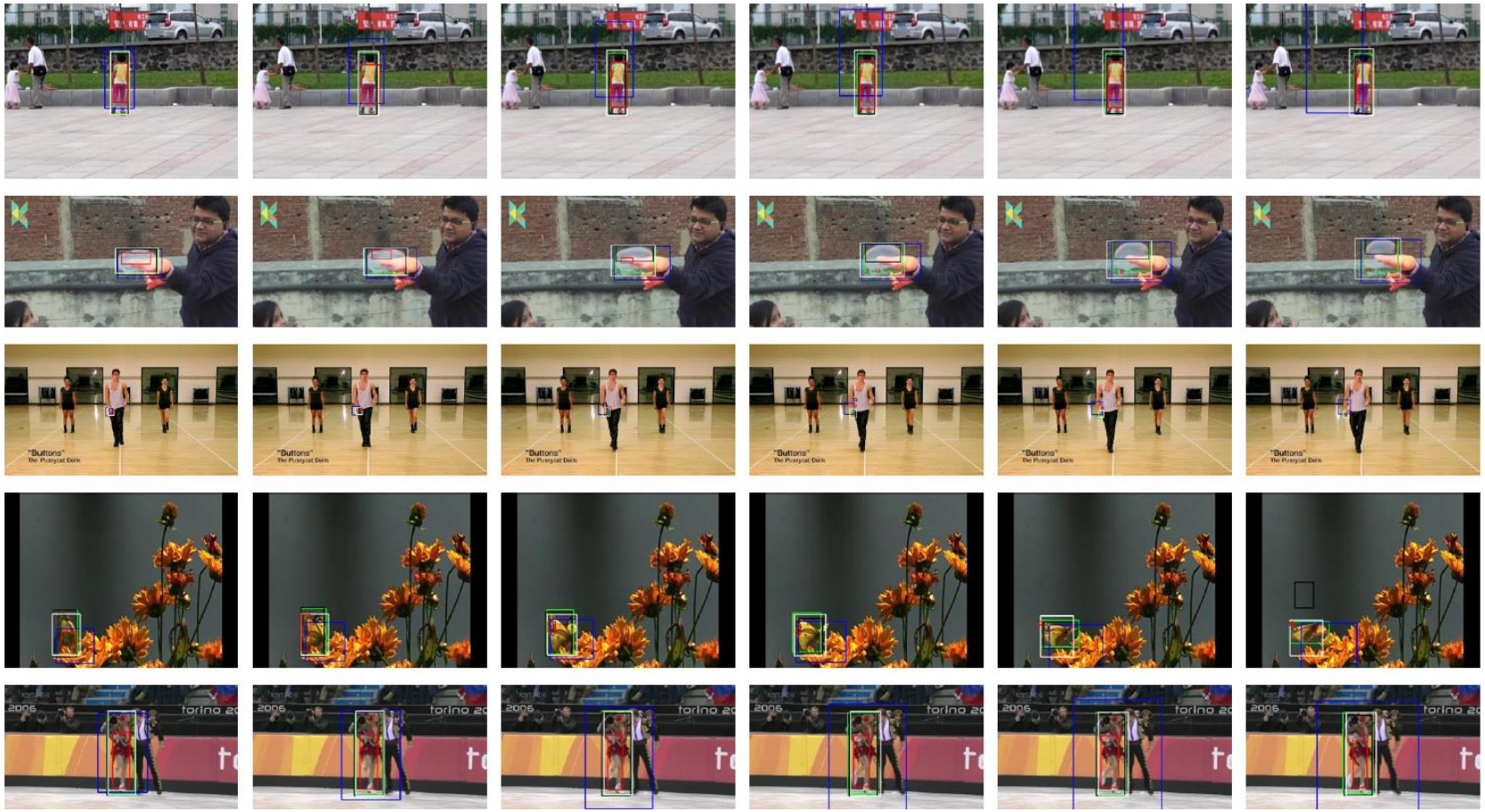
3. Regarding the sparsity and imperceptibility of perturbations, ABBG is ranked 2nd in comparison to other white box attacks.

ABBG Attack

Our goal is to mislead transformer trackers into predicting inaccurate bounding boxes across video frames.

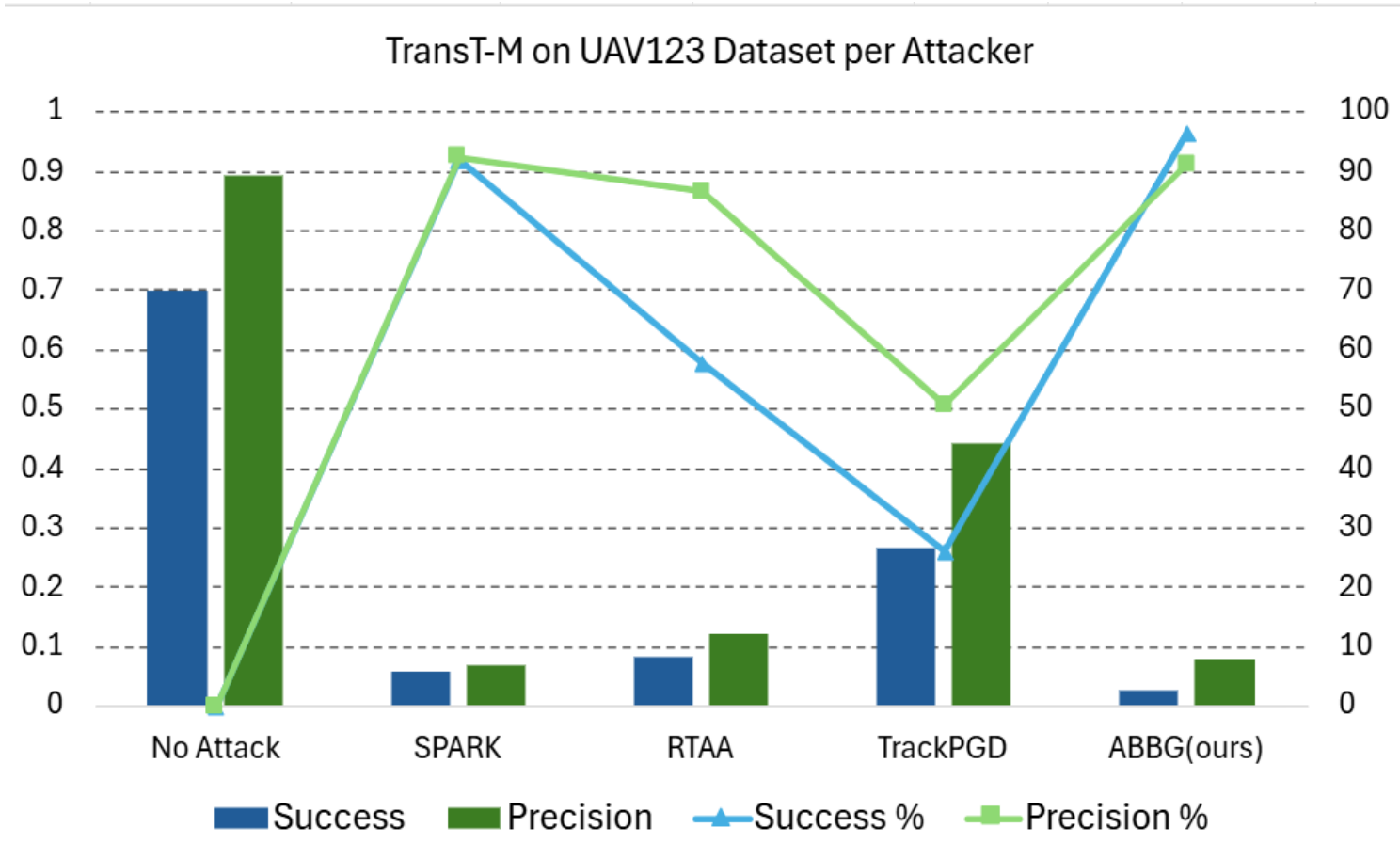


Object Bounding Box Evaluation

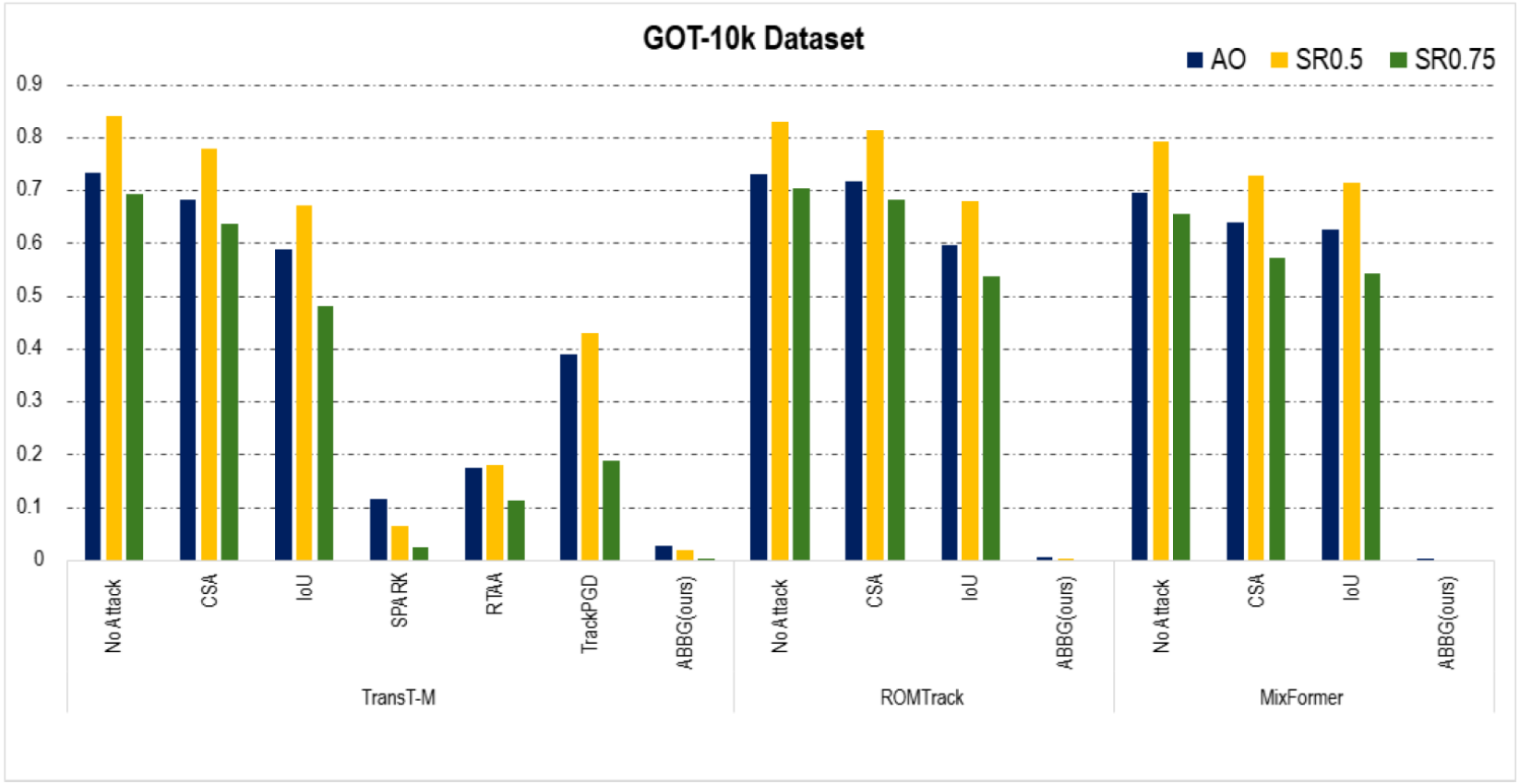


No attack SPARK RTAA TrackPGD ABBG

Object Bounding Box Evaluation



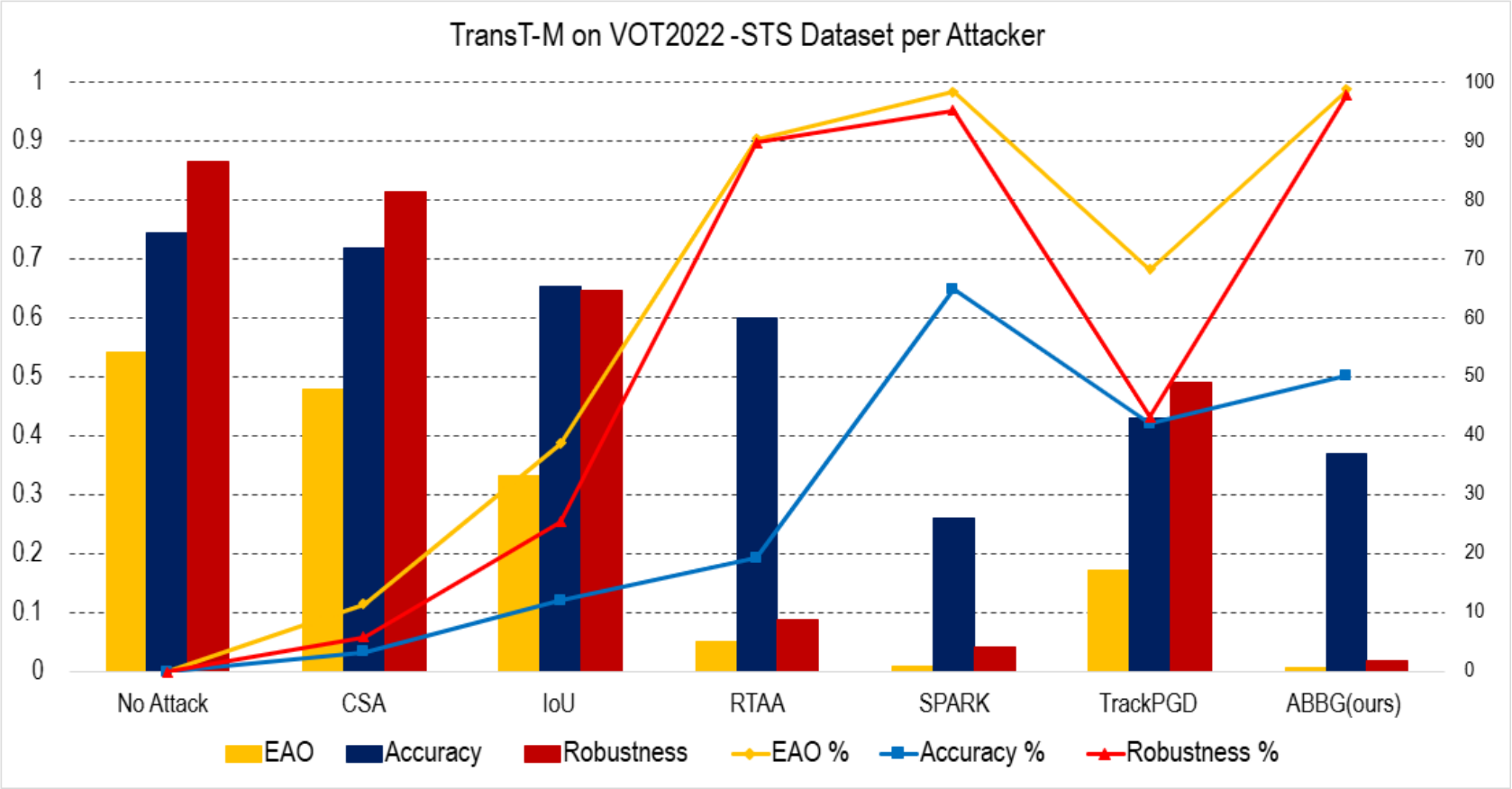
Object Bounding Box Evaluation



Main Takeaway: Among white-box attacks, ABBG is applicable to all three trackers- TransT-M, ROMTrack, and MixFormer. Beyond this versatility, the ABBG attack causes significant drops in all scores.

Binary Mask Evaluation

Binary Mask Evaluation



Main Takeaway: ABBG outperforms other white-box attacks (RTAA, SPARK, and TrackPGD) on the TransT-M tracker, except in accuracy, where it ranks second to SPARK.

Sparsity and Imperceptibility

Tracker	Attacker	L1-Norm ↓	SSIM(%) ↑
TransT-M	SPARK	69.98	94.43
	RTAA	113.48	60.14
	TrackPGD	122.52	64.04
	ABBG (ours)	95.77	89.50

Main Takeaway: ABBG ranks 2nd to SPARK overall but outperforms it in perturbation effectiveness on GOT-10k and VOT2022ST.

Conclusion

- We evaluated the adversarial robustness of three transformer trackers under a novel white-box attack that manipulates target bounding box predictions to generate adversarial perturbations.

- Our attack, ABBG, is broadly applicable across trackers, leveraging the simplicity of a single bounding box as its attack proxy.

- We demonstrated that ABBG consistently outperforms other attack methods, excelling in at least one metric per dataset.

- Notably, ABBG ranks second in both sparsity and imperceptibility among white-box attacks after a fixed number of iterations.