TrackPGD: Efficient Adversarial Attack using Object Binary Masks against Robust Transformer Trackers

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

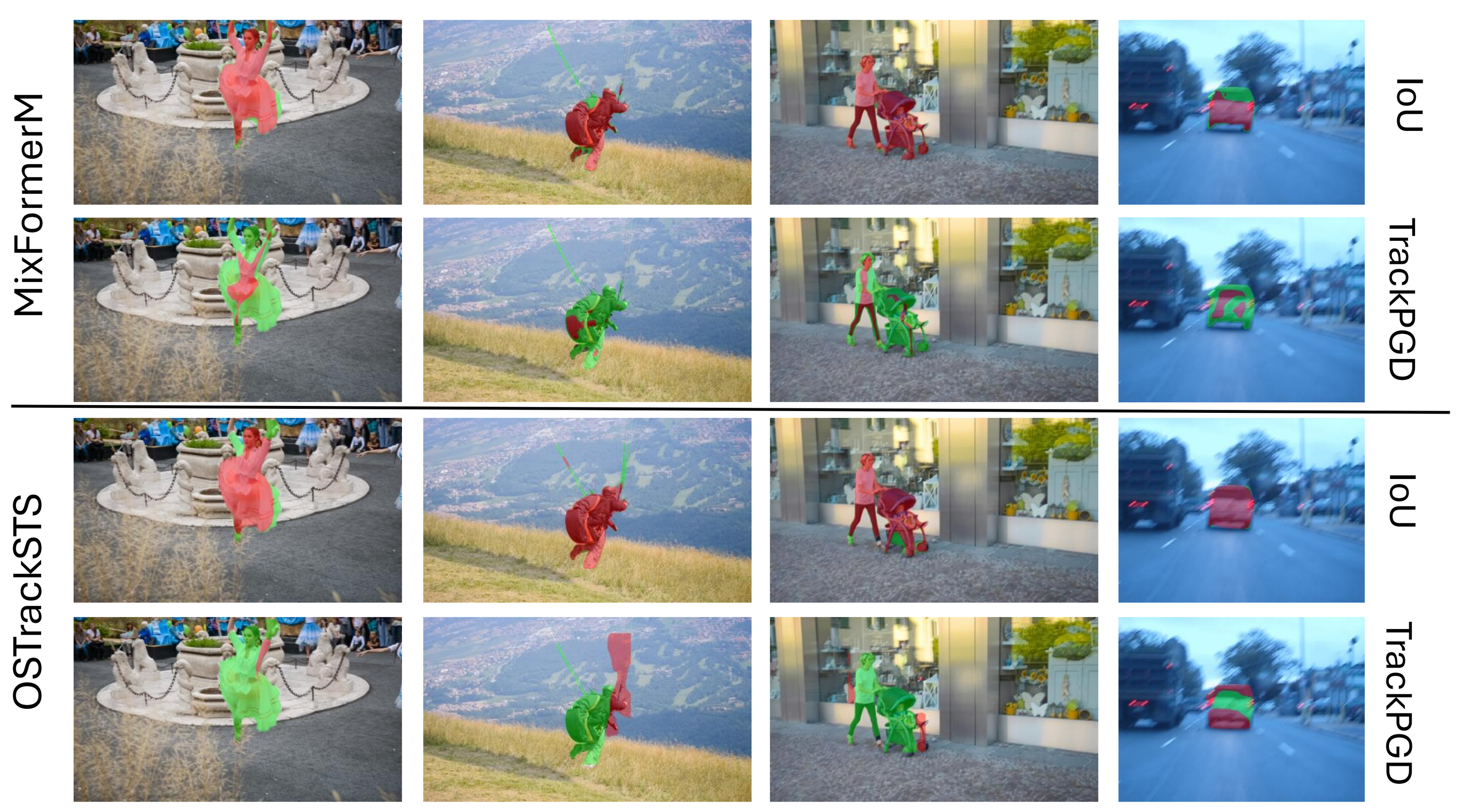
[1]IID, [2]Université Laval, [3]Mila, [4]Canada CIFAR AI Chair

NEURAL INFORMATION PROCESSING SYSTEMS

Contributions

- TrackPGD builds the adversarial noise from the binary mask to attack transformer trackers.
- A new loss in TrackPGD loss is proposed to mislead visual trackers in providing an accurate binary mask.
- Experimental results also demonstrate that the perturbations generated by TrackPGD have a great influence on bounding box predictions in tracking benchmarks.

Attack Setting	Method	Attack Proxy	MixFormerM	OSTrackSTS	TransT-SEG	RTS
Black-box	IoU	Object bbox	✓	✓	✓	✓
	CSA	Object bbox, heat-maps	✓	✓	✓	×
White-box	SPARK	Regression and classification labels	×	×	✓	×
	RTAA	Regression and classification labels	×	×	✓	×
	TrackPGD	**Object binary mask**	✓	✓	✓	✓



Proposed Method

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

Algorithm 1 TrackPGD to attack transformer trackers with segmentation capability

Require: Tracker $\mathcal{F}(\cdot)$, current frame I_τ, previous binary mask $M_{\tau-1}$, perturbation range ϵ, step size α, loss trade-offs λ_1 and λ_2, maximum iteration T, focusing parameter γ, variant of focal loss α_t, probability map p_t

$I^0_{\text{adv}} \leftarrow I_\tau$ ▷ initialization

$G_\tau \leftarrow M_{\tau-1}$ ▷ use last predicted binary mask as ground truth

for $t = 1 \ldots T$ **do**

  $M^t \leftarrow \mathcal{F}(I^{t-1}_{\text{adv}})$ ▷ predict binary mask

  $L_\Delta \leftarrow L_{\text{SegPGD}}(M^t, G_\tau) - L_{\text{SegPGD}}(M^t, 1 - G_\tau)$ ▷ compute difference of SegPGD losses

  $L_{\text{focal}} \leftarrow \alpha_t(1-p_t)^\gamma L_\Delta$ ▷ compute focal loss

  $L_{\text{dice}} \leftarrow 1 - 2\,\text{IoU}(M^t, G_\tau)$ ▷ compute dice loss

  $L \leftarrow \lambda_1 L_{\text{focal}} + \lambda_2 L_{\text{dice}}$ ▷ compute TrackPGD loss

  $I^t_{\text{adv}} \leftarrow I^{t-1}_{\text{adv}} + \alpha\,\text{sign}(\nabla_{I^{t-1}_{\text{adv}}} L)$ ▷ update adversarial example

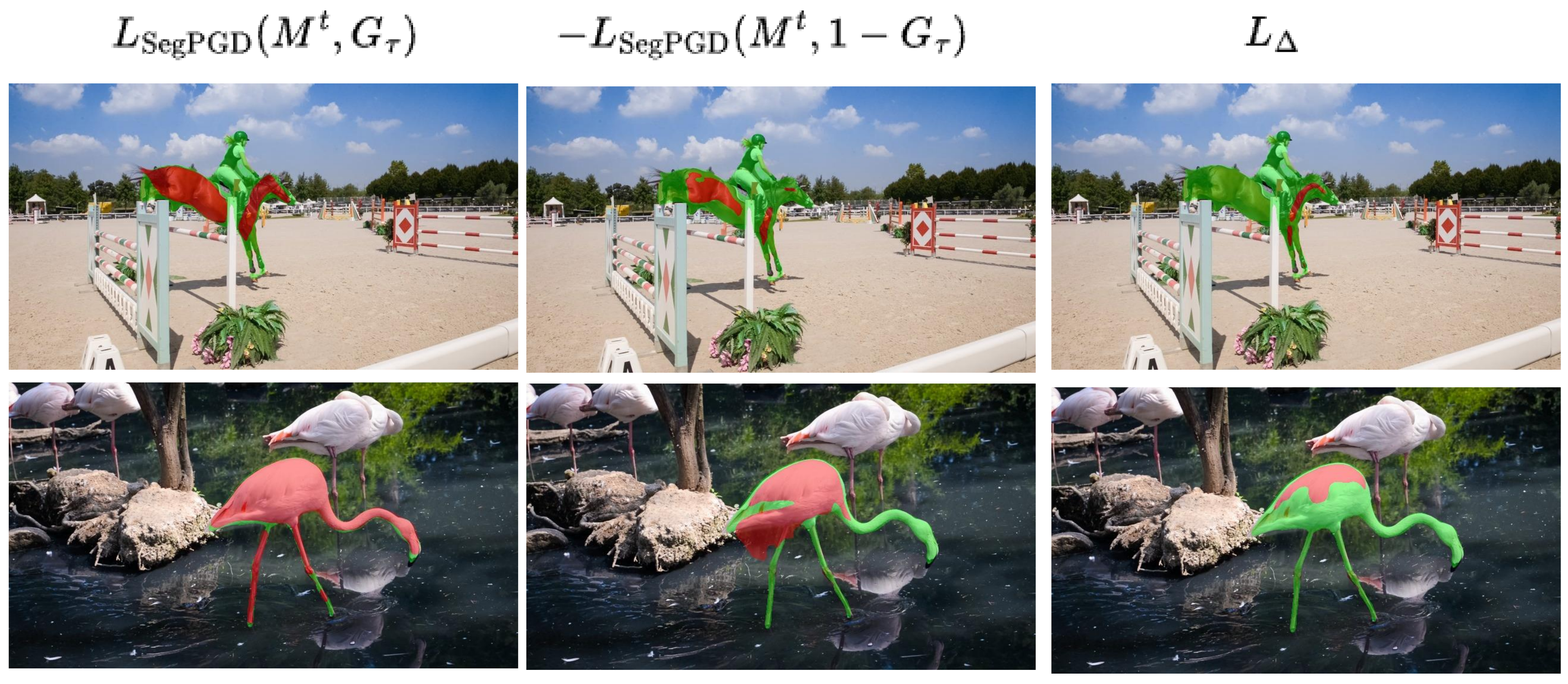
  $I^t_{\text{adv}} \leftarrow \phi^\epsilon\left(I^t_{\text{adv}}\right)$ ▷ clip to the ϵ-ball

end for

Role of Difference Loss: L_Δ

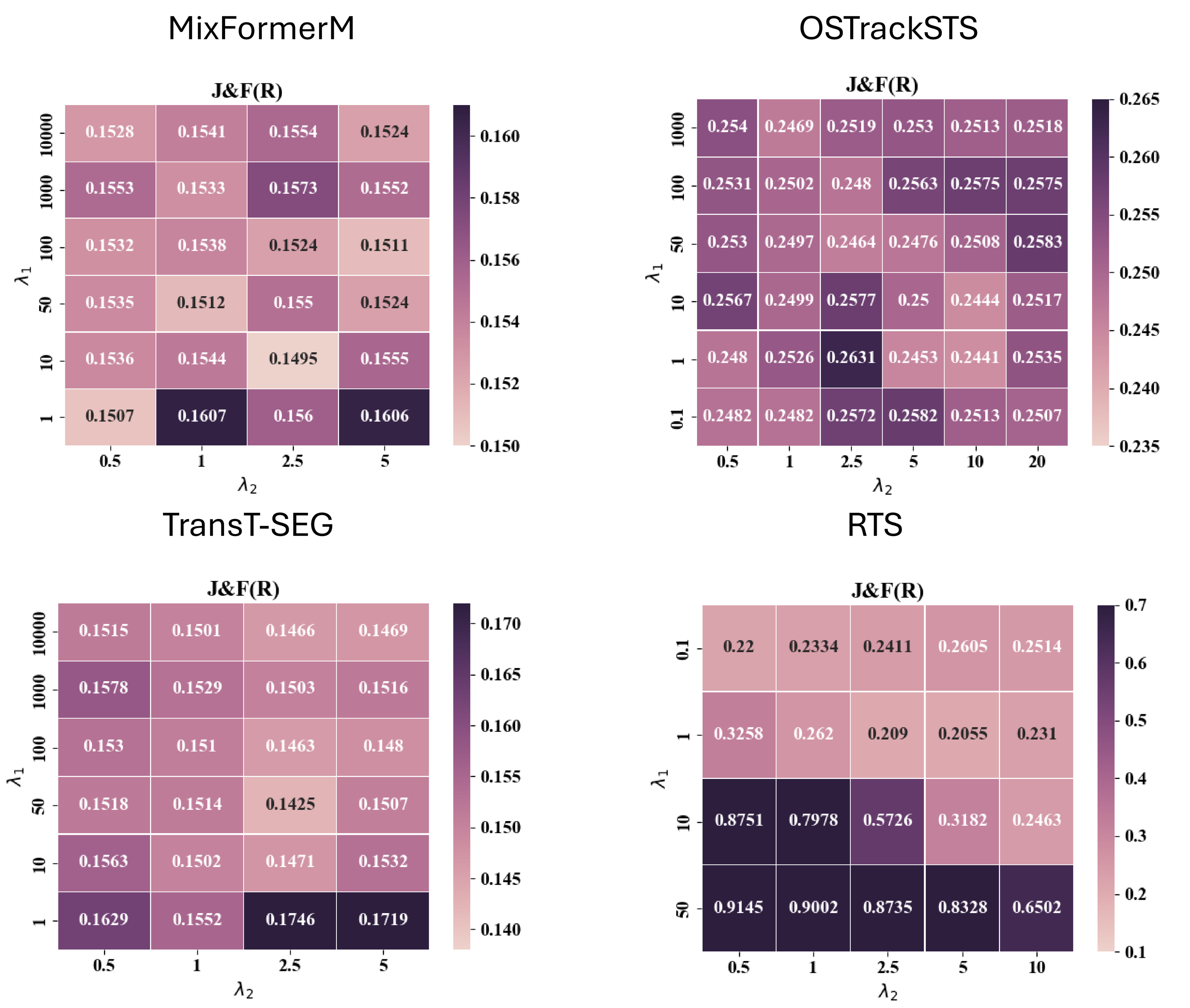
Original	$L_{\text{SegPGD}}(M^t, G_\tau)$	$-L_{\text{SegPGD}}(M^t, 1 - G_\tau)$	L_Δ
85.82	52.86	37.28	**30.30**

$L_{\text{SegPGD}}(M^t, G_\tau)$ $-L_{\text{SegPGD}}(M^t, 1 - G_\tau)$ L_Δ



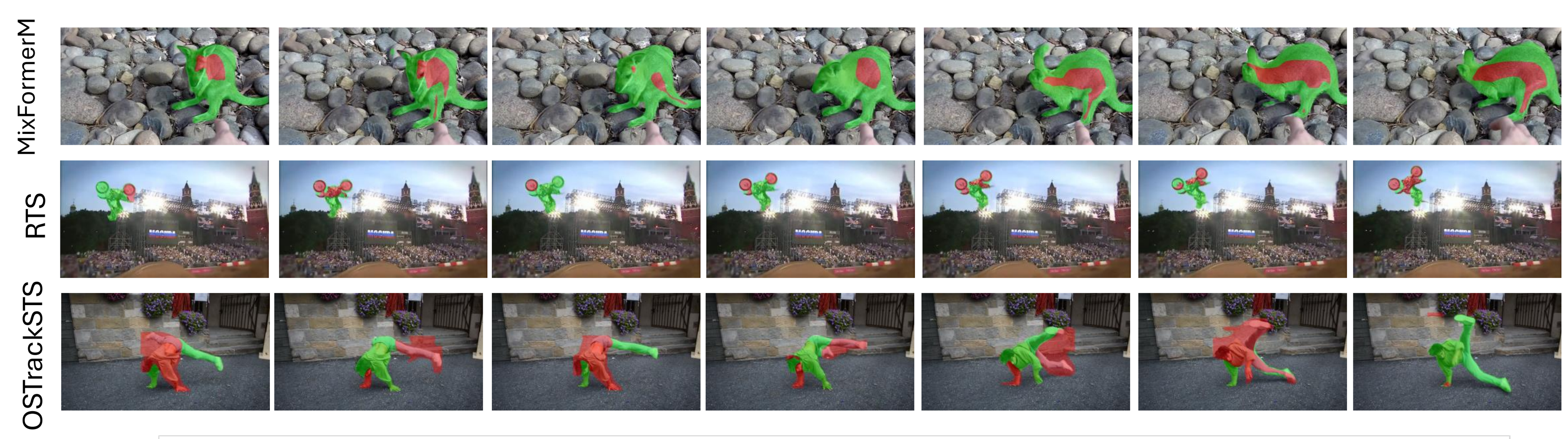
Main Takeaway. Although, the vanilla SegPGD losses also generate inaccurate binary masks, the mask impairment caused by L_Δ is significantly greater.

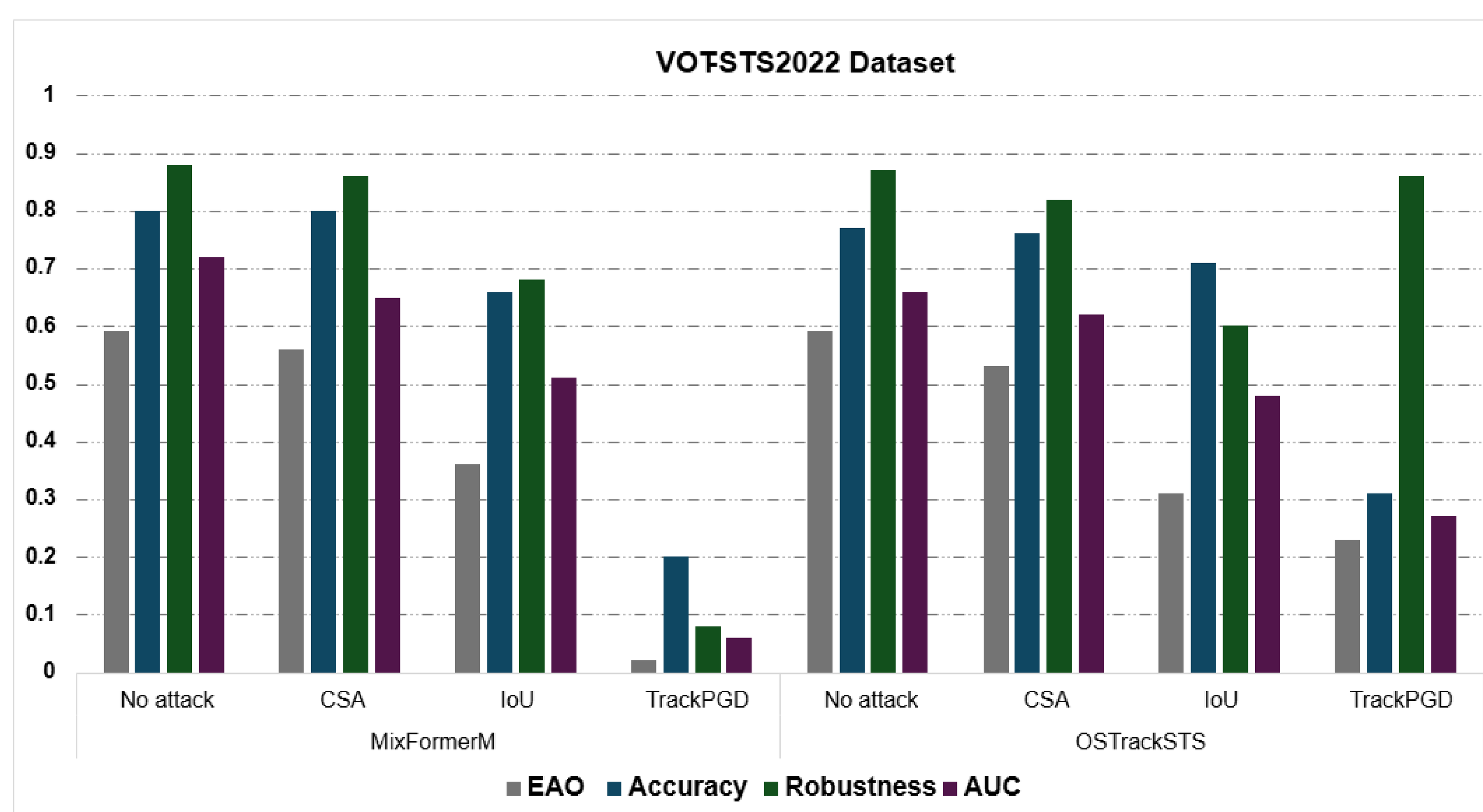
Fine Tuning the hyperparameters



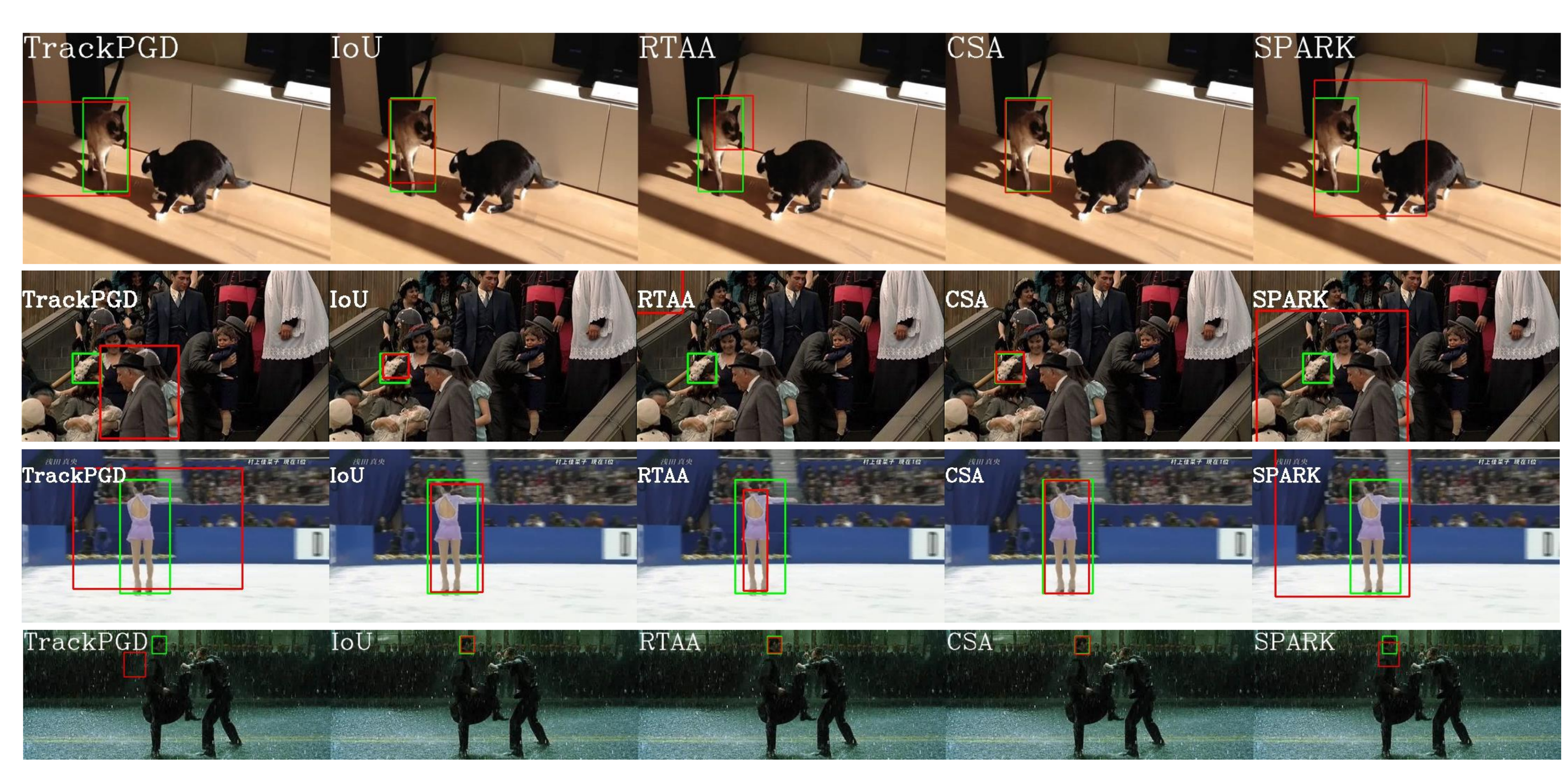
Main Takeaway. The optimal performance occurs when both loss terms are active with intermediate values, showing that neither term alone achieves the best results.

Object Binary Mask Evaluation





Object Bounding Box Evaluation



Main Takeaway. The efficacy of TrackPGD is validated through comprehensive experiments on various transformer and non-transformer networks on popular datasets.

Acknowledgments. This work is supported by the DEEL Project CRDPJ 537462-18 funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Consortium for Research and Innovation in Aerospace in Quebec (CRIAQ), together with its industrial partners Thales Canada inc, Bell Textron Canada Limited, CAE inc and Bombardier inc.

Institut intelligence et données

UNIVERSITÉ LAVAL

Mila

DEEL
DEpendable & Explainable Learning