NEURIPS 2026

Dynamic Physical Adversarial LED Patterns via Reinforcement Learning

Closed-loop control of a wearable LED array to evade person detectors

Konstantin Zyryanov · Nikita Lotts

Lomonosov Moscow State University · HSE University

zyryanov.konstantin.sergeevich@gmail.com · nklotts@edu.hse.ru



MOTIVATION

Physical adversarial attacks today are static

Printed patches
Thys et al., 2019

Adversarial T-shirts
Xu et al., 2020

Camouflage textures
Hu et al., 2022

All are optimized once and printed. Light, distance and viewing angle keep changing — the pattern doesn't.

OUR QUESTION

Can an emissive pattern that adapts to the scene, using only black-box detector feedback, beat a fixed one?

STUDY OVERVIEW

The paper in three lines

Setting	**Closed-loop physical adversarial control** — a policy changes a wearable LED array based on what its cameras see
Approach	**PR-SAC** — Physically Regularized Soft Actor-Critic with latent actions and structured upsampling
Finding	**89.1 / 84.7 / 86.8 % attack success** on YOLOv8 / Faster R-CNN / RetinaNet — **+38–42 pp** over the best static pattern

TESTBED

A wearable array of 9,984 RGB LEDs



39×16^2
LED panels

29,952
RGB control values (d)

Orange Pi 5 Pro
drives the LED matrix

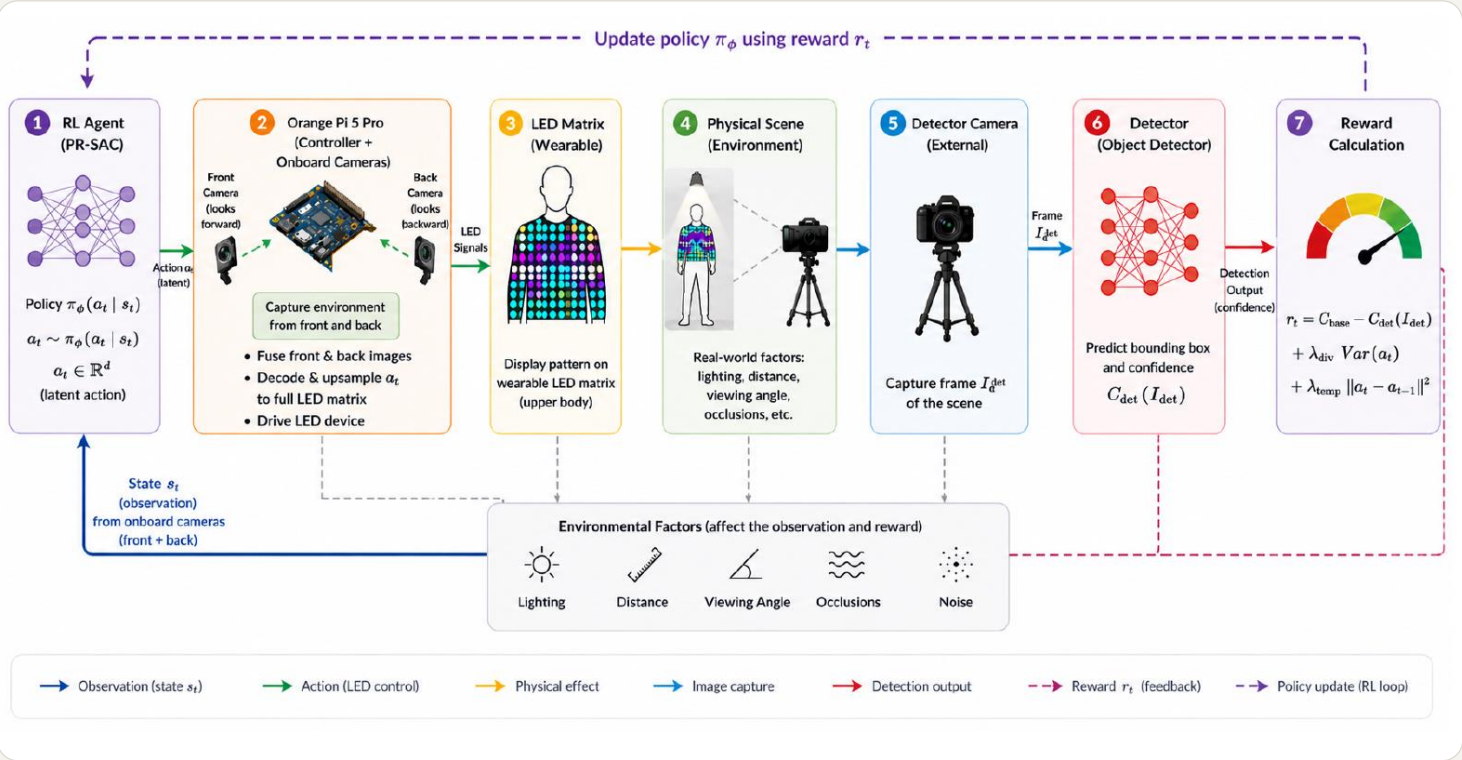
3 GB GPU
training + detector inference

Onboard front + back cameras → what the policy observes

Separate external camera → what the detector is scored on

FORMULATION

The attack as a closed-loop RL problem



Reward

$$r = 1 - C_{det} + \lambda_{div} \cdot Var(a_t) + \lambda_{temp} \cdot \|a_t - a_{t-1}\|^2$$

Threat model

Black-box: no gradients, weights or camera access

Training: detector confidence only

Test time: no detector feedback

CHALLENGE

Why RL is unstable in the physical loop

Noisy reward

Sensor noise and measurement noise in every detector score

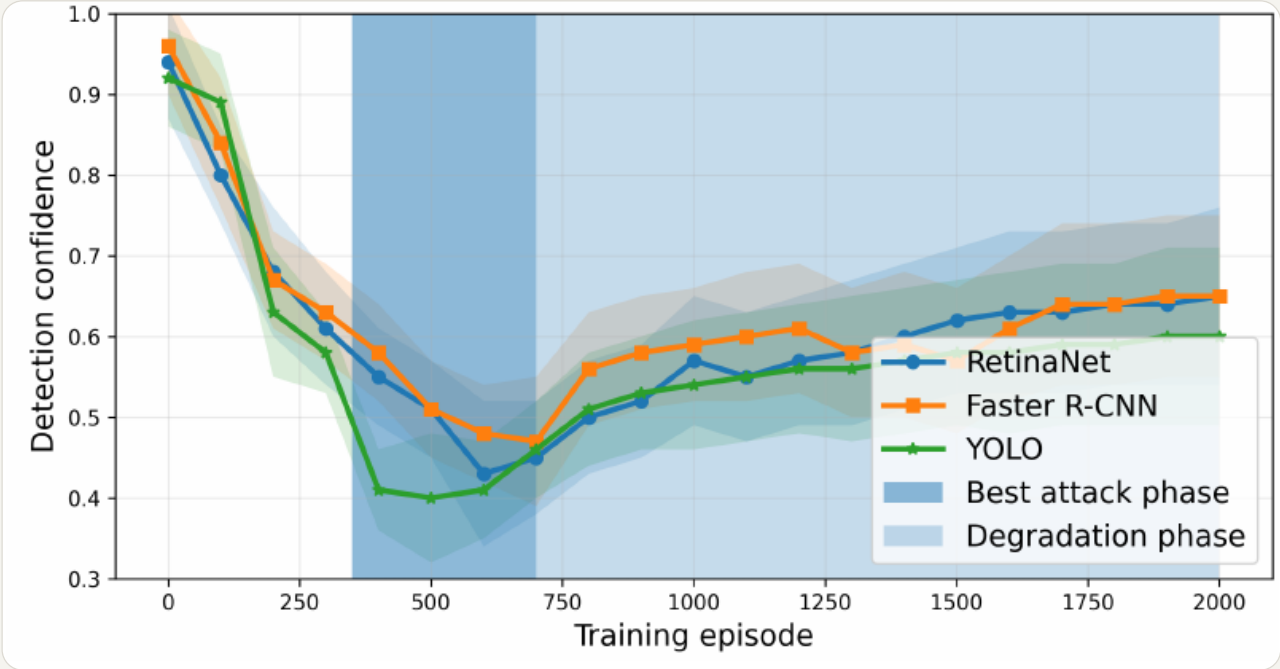
Drifting scene

Pose, light and distance change, so the Bellman operator is time-dependent

Replay mismatch

Old transitions no longer match the current policy or scene

→ Q-drift, entropy collapse, degradation after the peak



Detection confidence over training (mean of 15 runs): best attack around episodes 350–700, then degradation

METHOD

PR-SAC: three fixes on top of SAC

1 · SMOOTH

Regularize the actor

Penalize jumps between neighboring LEDs and between frames

$$+\lambda_s \|\nabla_{LED} a_t\|^2$$
$$+\lambda_t \|a_t - a_{t-1}\|^2$$

2 · TRUST

Down-weight noisy transitions

Critic loss weighted by the estimated reward variance

$$w_i = \frac{1}{\varepsilon + \hat{\sigma}_r^2(s_i, a_i)}$$

3 · RECENT

Recency-weighted replay

Stale samples from an older scene count for less

$$\rho_i = exp(-\beta(T - t_i))$$
$$\tilde{w}_i = \rho_i \cdot w_i$$

Improves stability in practice; no formal convergence guarantee.

METHOD

Latent actions tame 29,952 dimensions

Latent action → Decode → Structured upsampling → **9,984-LED pattern**

Final detection confidence (lower = stronger)

Direct LED control



0.71

Latent action



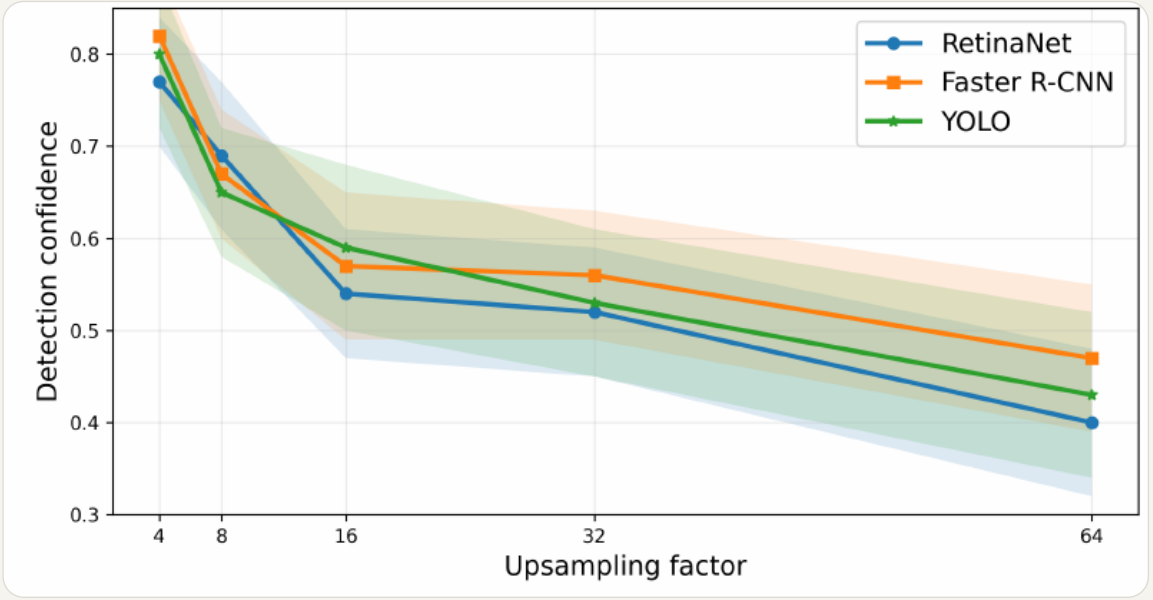
0.60

Latent + structured upsampling



0.43

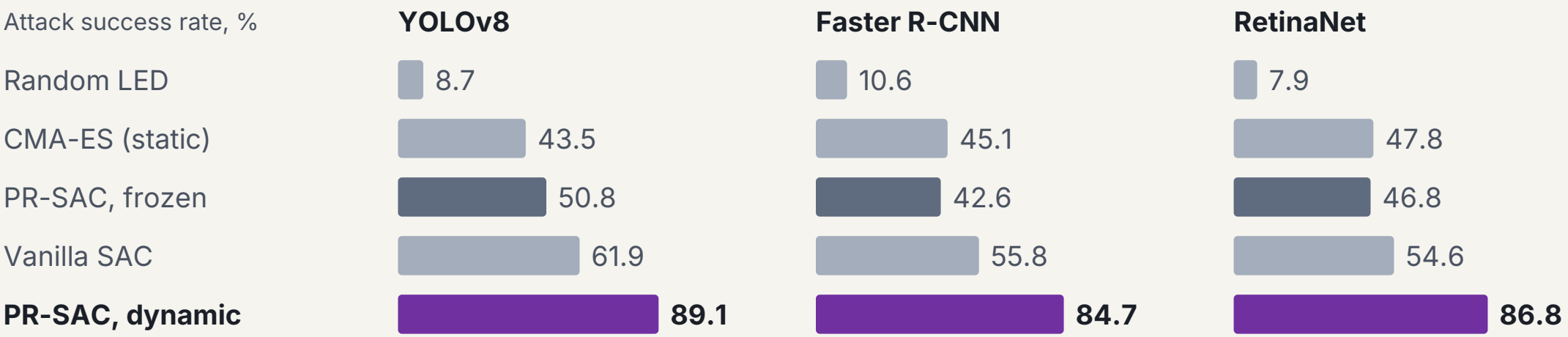
Degradation after peak also drops: 0.19 → 0.18 → 0.13 (15 runs)



Upsampling factor trades detail for stability: low factors are unstable, high ones over-smooth

MAIN RESULT

Adapting beats any fixed pattern



+38 to +42 pp over the **same policy, frozen** to one pattern (+38.3 / +42.1 / +40.0) — the gain comes from adapting, not from the pattern

Mean over 15 independent runs, 300 frames each · Table 2

GENERALIZATION

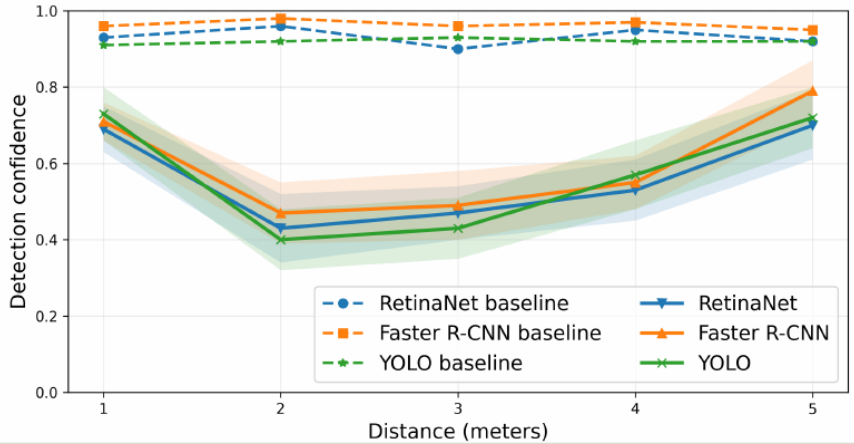
New locations: weaker, but still strong

YOLOv8
89.1 → 77.3%

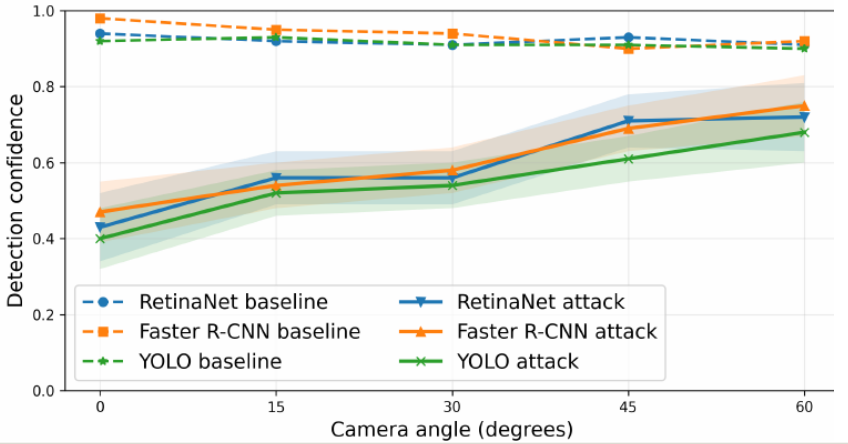
Faster R-CNN
84.7 → 69.8%

RetinaNet
86.8 → 74.5%

Unseen locations, no fine-tuning — still above every static baseline in familiar ones (≤ 50.8%)



Confidence vs. distance (1–5 m)



Confidence vs. viewing angle (0–60°)

Mean over 15 runs · Table 2, Figure 5

TRANSFERABILITY

Patterns partially transfer across detectors

Train ↓ · Test →	YOLO	Faster R-CNN	RetinaNet
YOLO	0.40	0.48	0.51
Faster R-CNN	0.52	0.47	0.56
RetinaNet	0.54	0.55	0.43

Strongest on the training detector

Diagonal: 0.40 / 0.47 / 0.43

Still effective elsewhere

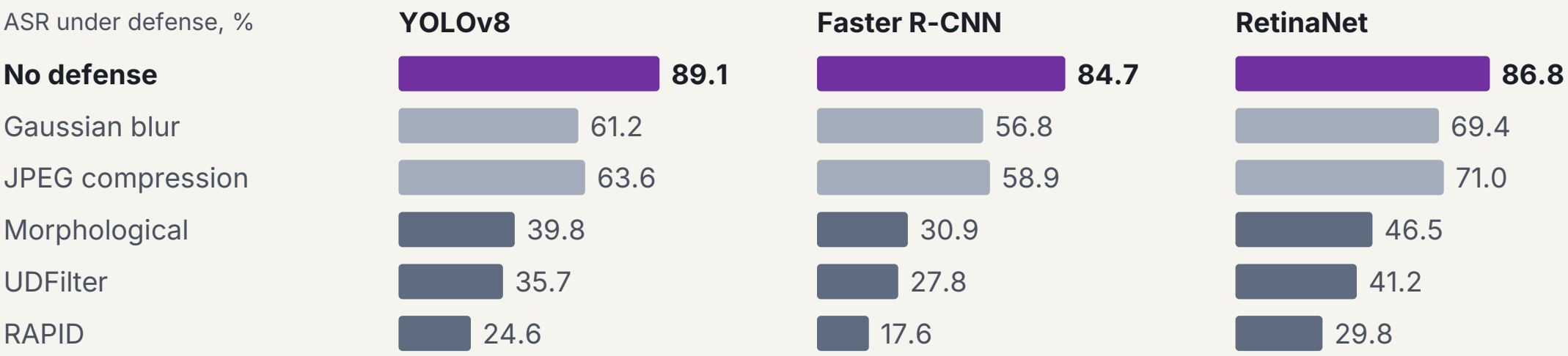
Off-diagonal 0.48–0.56, vs. 0.92–0.96 with no attack

Cells: mean detection confidence; darker = lower = stronger attack

Mean over 15 runs · Tables 3, 5

DEFENSES

Dedicated defenses help; none removes the attack



LIMITATIONS One LED platform, three detectors · adversarial training and certified defenses not tested · training remains unstable (~5–6 h per run)

Same PR-SAC evaluation frames · Table 8

TAKEAWAY

Adaptive physical attacks are much stronger than static ones — robustness evaluations should include them, along with generalization and training stability.

Konstantin Zyryanov · Nikita Lotts

Responsible release: no real-time deployment tools or optimized policies for specific targets are released.

Thank you

BACKUP · FULL RESULTS

Detection confidence and mAP@0.5

Lower = stronger attack

	Detection confidence (Table 5)			mAP@0.5, person (Table 6)		
Method	**YOLOv8**	**F. R-CNN**	**RetinaNet**	**YOLOv8**	**F. R-CNN**	**RetinaNet**
Clean baseline	0.92 ± 0.04	0.96 ± 0.03	0.94 ± 0.03	0.88 ± 0.03	0.85 ± 0.03	0.83 ± 0.03
Random search	0.81 ± 0.04	0.79 ± 0.04	0.82 ± 0.05	0.74 ± 0.04	0.72 ± 0.04	0.73 ± 0.04
CMA-ES	0.71 ± 0.05	0.70 ± 0.05	0.68 ± 0.06	0.64 ± 0.05	0.63 ± 0.05	0.62 ± 0.05
Vanilla SAC	0.59 ± 0.08	0.66 ± 0.07	0.67 ± 0.08	0.54 ± 0.06	0.58 ± 0.06	0.57 ± 0.06
PR-SAC	0.40 ± 0.08	0.47 ± 0.08	0.43 ± 0.09	0.42 ± 0.06	0.47 ± 0.05	0.45 ± 0.05

Mean ± std over 15 independent runs; 95% Student-t confidence intervals in the paper.

BACKUP · ABLATION

Reward design: regularization terms matter

Reward	YOLO	F. R-CNN	RetinaNet
−C	0.62 ± 0.06	0.64 ± 0.08	0.67 ± 0.06
1 − C	0.59 ± 0.07	0.60 ± 0.08	0.63 ± 0.08
−log C	0.56 ± 0.05	0.54 ± 0.06	0.57 ± 0.06
$(1 - C)^2$	0.55 ± 0.06	0.53 ± 0.09	0.61 ± 0.07
+ variance	0.43 ± 0.09	0.51 ± 0.07	0.50 ± 0.09
+ temporal diff	0.40 ± 0.08	0.47 ± 0.08	0.43 ± 0.09

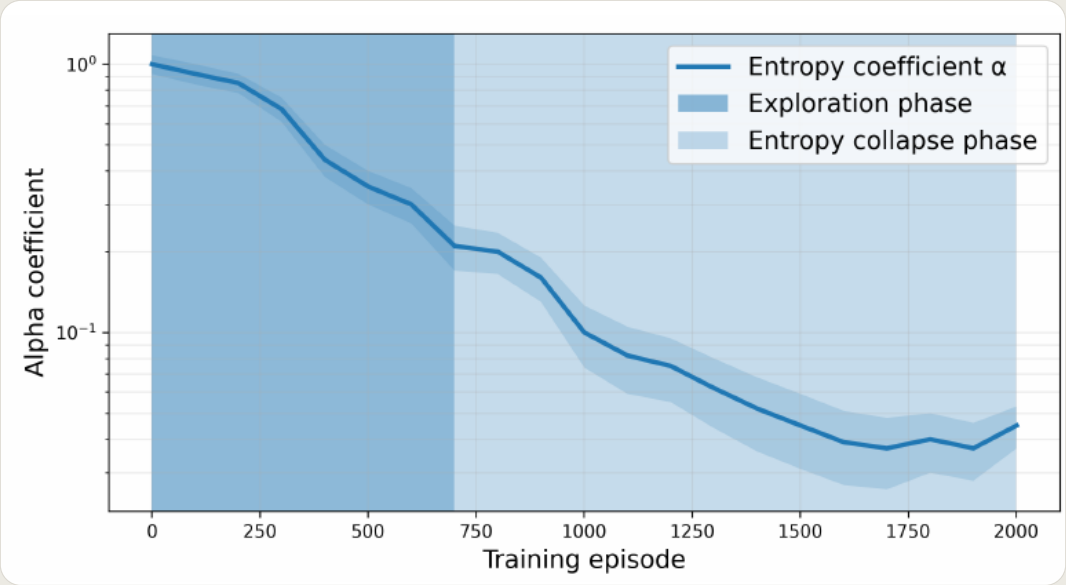
Detection confidence after training; lower = stronger attack.

Default reward: 1 − C plus diversity and temporal terms. Excessive shaping can reduce stability.

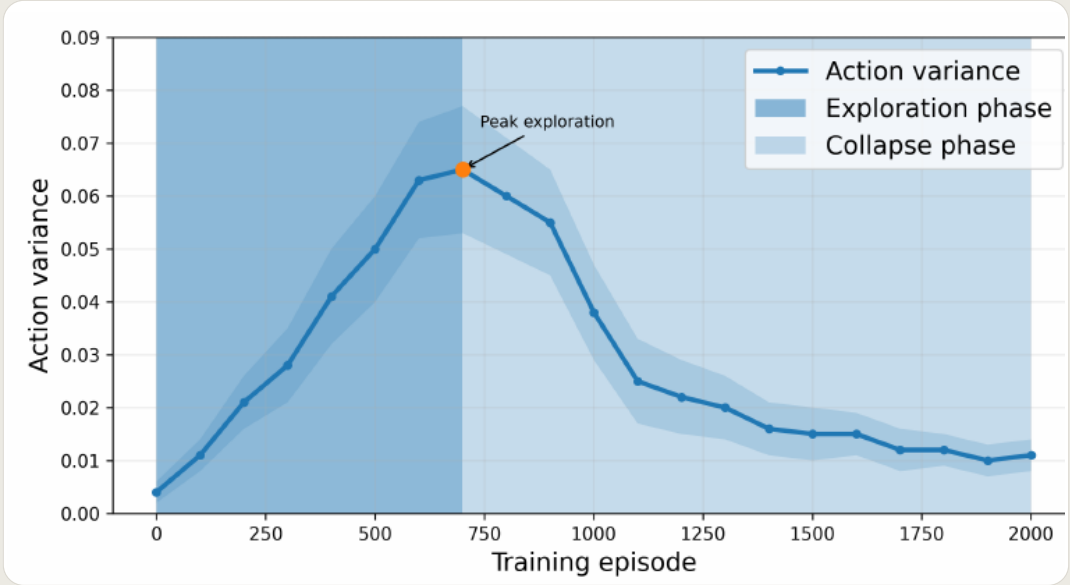
Mean ± std over 15 runs · Table 4

BACKUP · STABILITY

Entropy collapse is delayed, not removed



Entropy coefficient α over training



Action variance over training (peak exploration ≈ episode 700)

Diversity collapse coincides with the degradation phase; checkpoints are chosen on held-out validation confidence.

Mean over 15 runs · Figure 6

BACKUP · PROTOCOL

Evaluation protocol

Attack success rate

% of frames with no person detection matching ground truth at IoU ≥ 0.5 and confidence ≥ 0.6

Training

2,000 episodes of ~10–12 s each, ~5–6 h per run

Baselines

Random LED, vanilla SAC, CMA-ES static pattern in the same latent space with the same reward and query budget

Statistics

15 independent seeds × 300 evaluation frames; run-level mean, std, 95% Student-t CI

Checkpoint selection

Minimum mean confidence on a held-out validation sequence

Scene variation

Operator varies pose, viewpoint, distance, light on a shared schedule; unseen locations without fine-tuning