FlyTrap: Physical Distance-Pulling Attack Towards Camera-based Autonomous Target Tracking Systems

Shaoyuan Xie, Mohamad Habib Fakih, Junchi Lu, Fayzah Alshammari, Ningfei Wang, Takami Sato, Halima Bouzidi, Mohammad Abdullah Al Faruque, Qi Alfred Chen

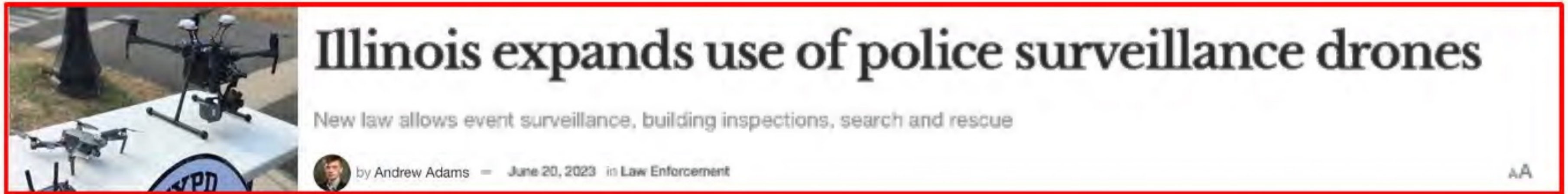
University of California, Irvine

UCI University of California, Irvine AS²Guard

NDSS SYMPOSIUM

Autonomous Target Tracking (ATT)

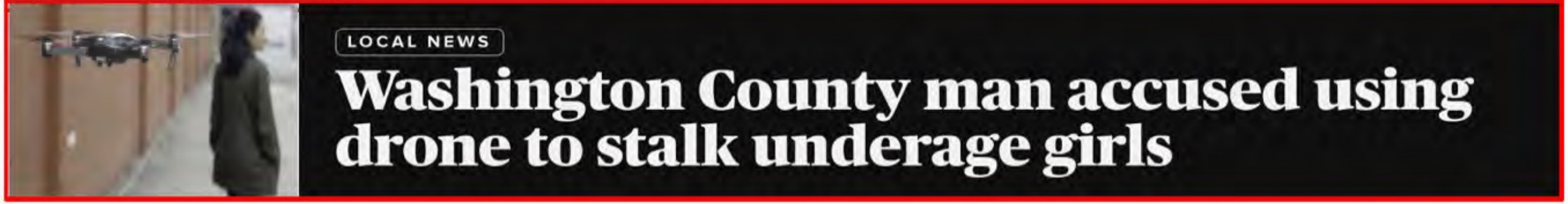
Autonomous Target Tracking Drones are gaining increasing interest to be used in societally critical scenarios in the real world



Illinois expands use of police surveillance drones

New law allows event surveillance, building inspections, search and rescue

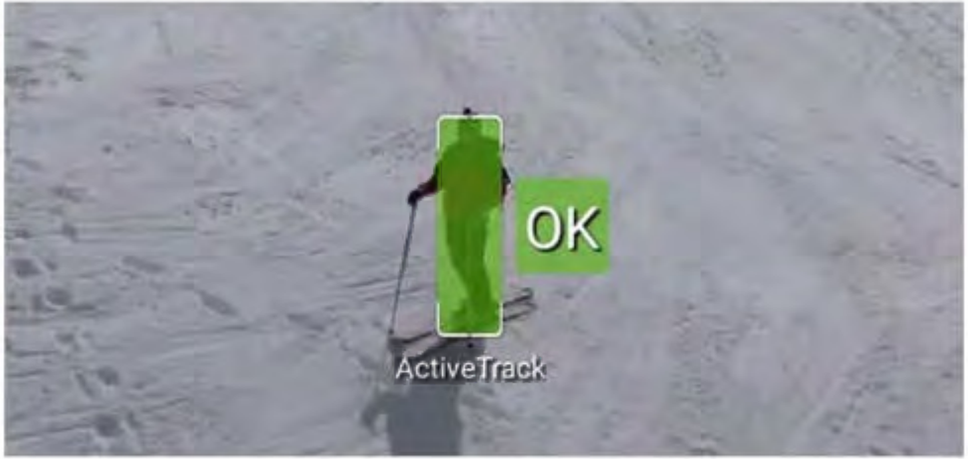
by Andrew Adams – June 20, 2023 in Law Enforcement



LOCAL NEWS

Washington County man accused using drone to stalk underage girls

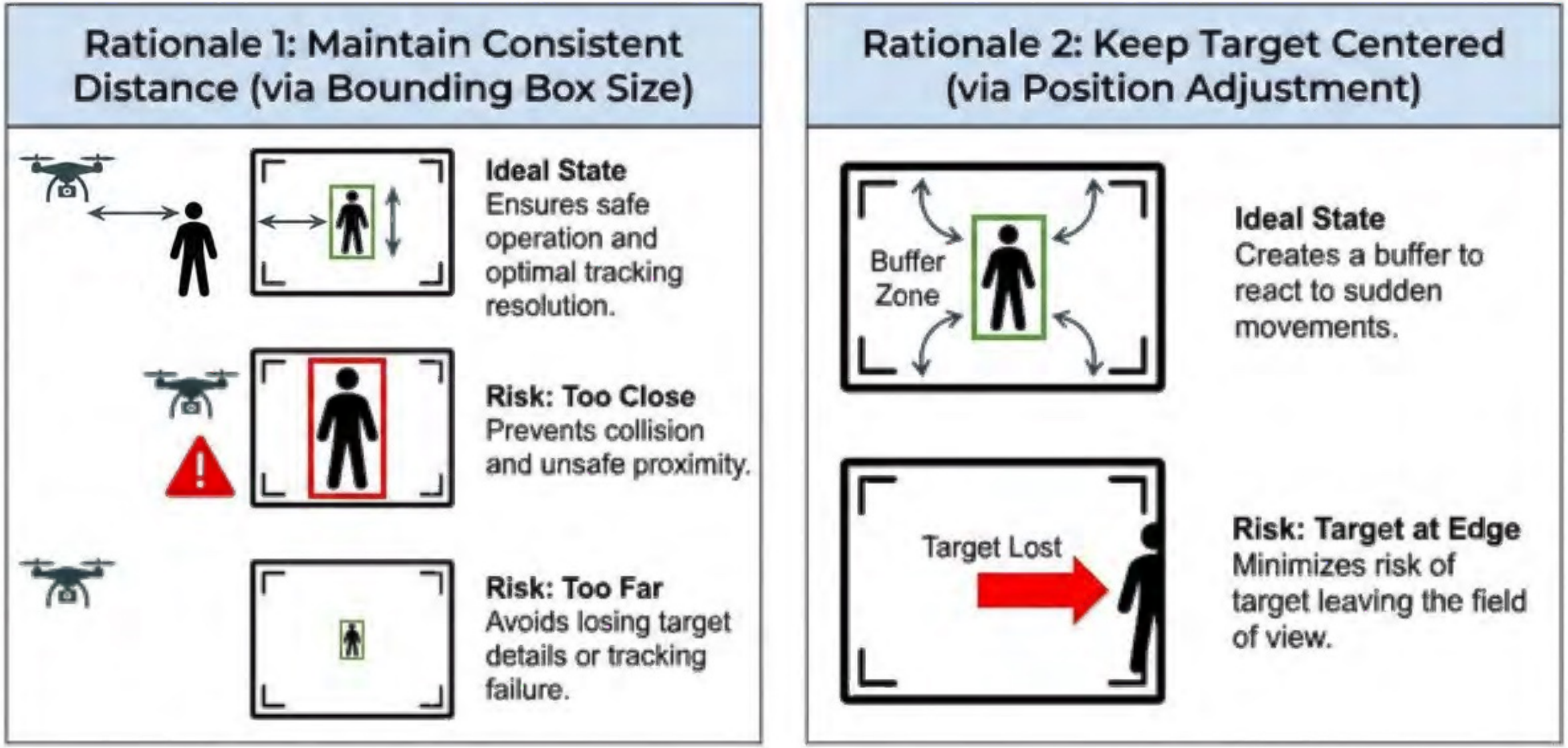
How Autonomous Target Tracking Works



The user selects the target by drawing the bounding box on the controller

Tracking DNN model processes this initial frame and starts to track

How Autonomous Target Tracking Works



Rationale 1: Maintain Consistent Distance (via Bounding Box Size)

Ideal State
Ensures safe operation and optimal tracking resolution.

Risk: Too Close
Prevents collision and unsafe proximity.

Risk: Too Far
Avoids losing target details or tracking failure.

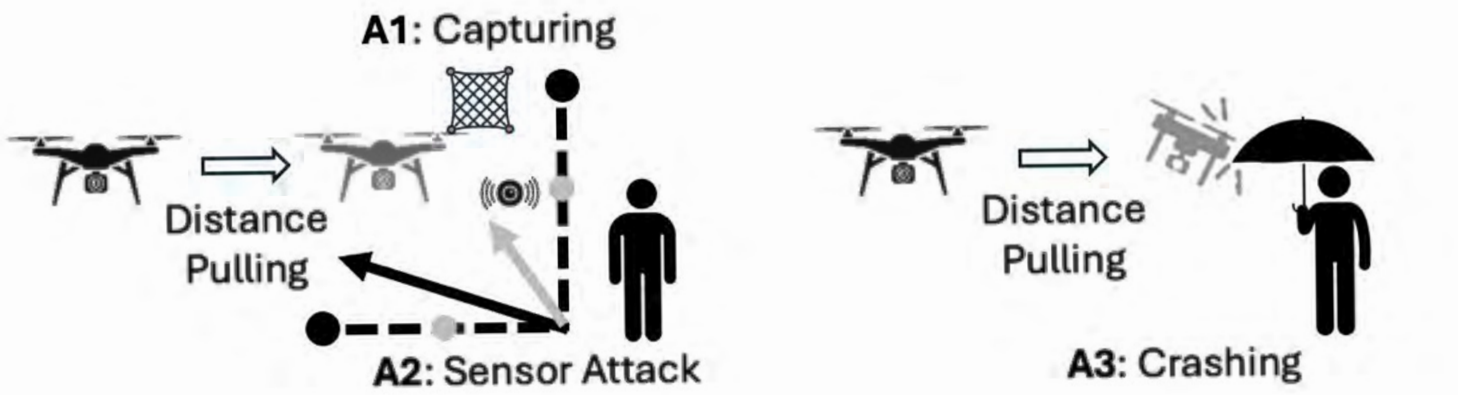
Rationale 2: Keep Target Centered (via Position Adjustment)

Ideal State
Creates a buffer to react to sudden movements.

Risk: Target at Edge
Minimizes risk of target leaving the field of view.

Image generated by Gemini

Attack Exploits



The attacker can exploit the shorter distance for:

1. Capturing: the drone is within the capturing distance of a net gun
2. Sensor Attacks: signal generators are range-limited in nature
3. Crashing: the drone is within the hitting distance

Our Proposed FlyTrap Attack

- **Novel <u>Distance-Pulling Attack</u>** via adversarial perception bounding box shrinkage.
- **<u>Exploits closed-loop control logic</u>**: drone misinterprets spoofed shrinkage as target moving away and accelerates for compensation.
- First systematic security analysis targeting the perception-control logic in **<u>physical ATT systems</u>**.

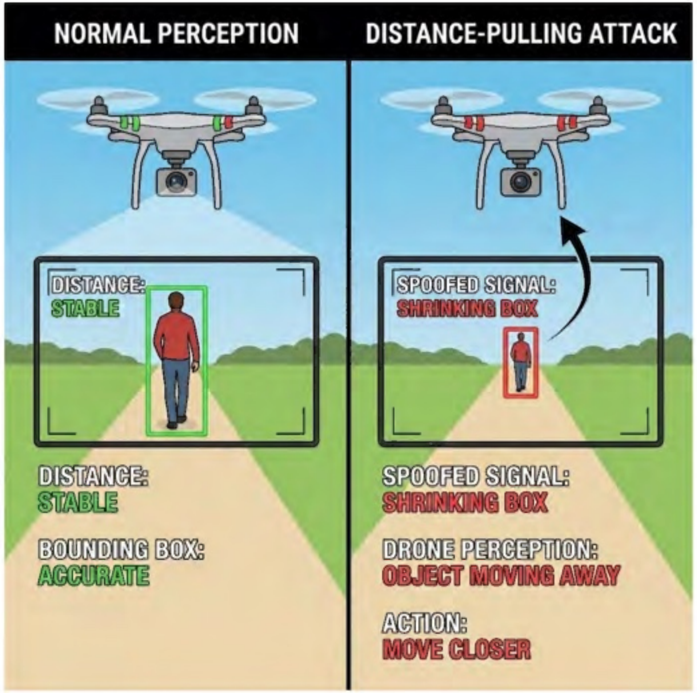
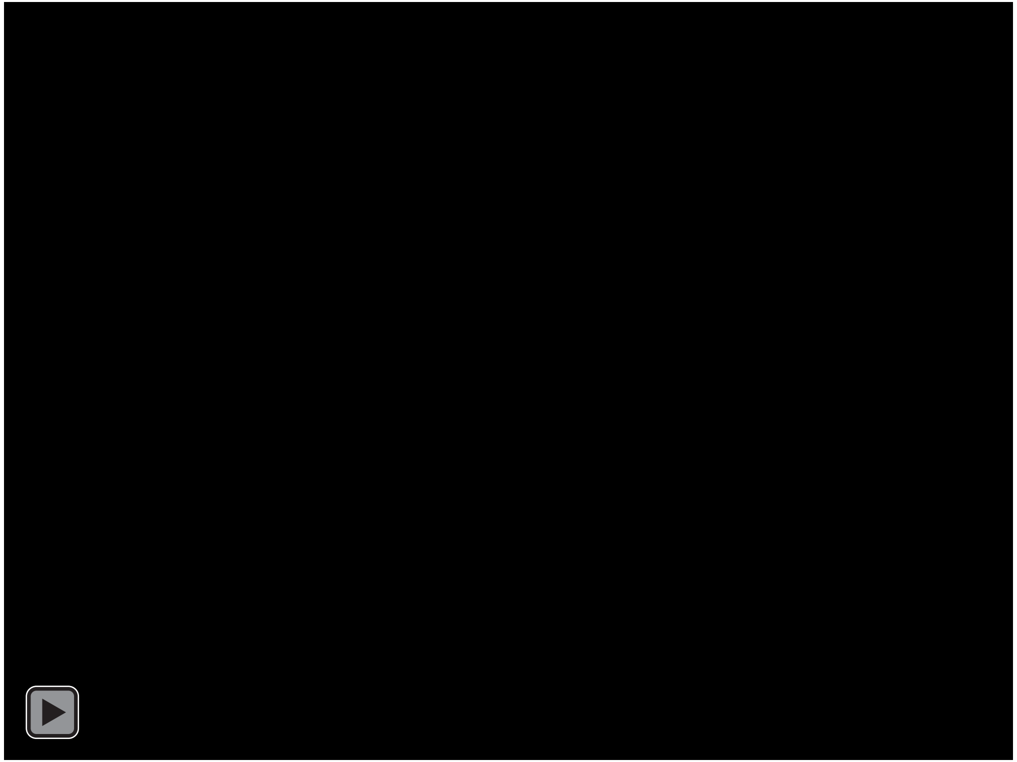


Image generated by Gemini

Threat Model

- We assume the attacker can access the white-box tracking ML models
  - This can be done by purchase the same drone models and conduct reverse engineering
- We also consider black-box threat model, where we use adversarial patterns optimized with one ML model to attack another

Real-World Attack Demonstration

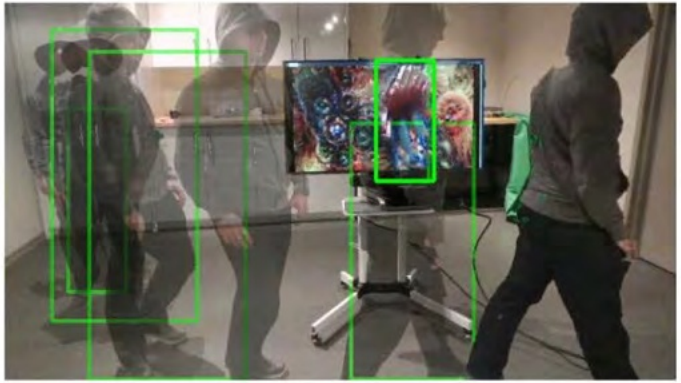


https://sites.google.com/view/av-ioat-sec/flytrap

Design Challenges-1: Physical and Deployable Attack Vectors

Previous attack vectors (e.g., TV screens and projector) are not suitable for outdoor tracking scenarios.

- They are hard to carry and require delicate setups

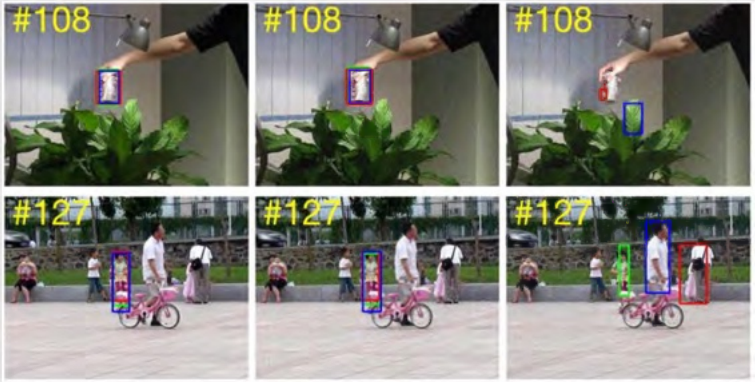


Wiyatno et al. ICCV’19

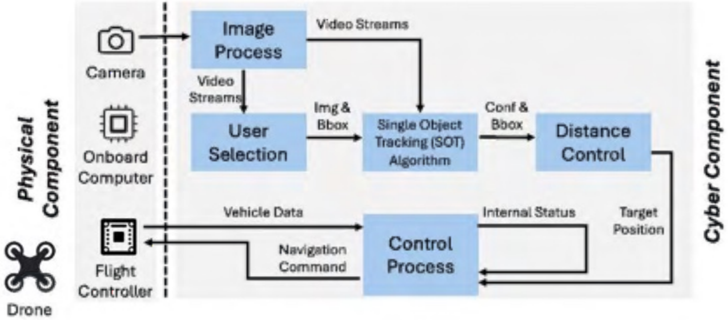
Muller et al. CCS’22

Wiyatno, Rey Reza, and Anqi Xu. "Physical Adversarial Textures That Fool Visual Object Tracking." *ICCV* 2019.
Muller, Raymond, et al. "Physical Hijacking Attacks Against Object Trackers.” *CCS* 2022.

Design Challenges-2: Closed-Loop Effectiveness

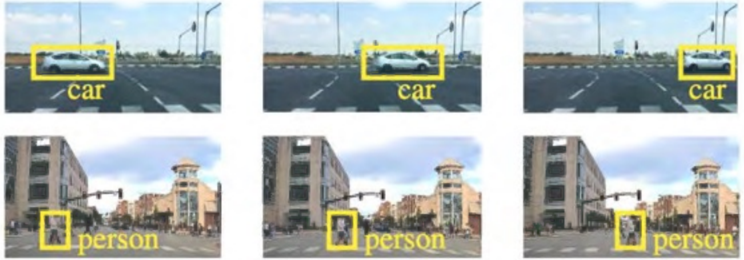


Jia et al. CVPR’21

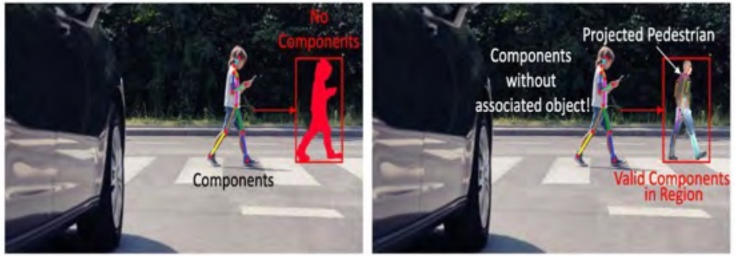


- Previous attacks on object tracking don’t consider the controlling loop
- However, autonomous tracking drones work in closed-loop, the current tracking results influence the future control

Design Challenges-3: Spatial-Temporal Consistency



Man et al. USENIX Security’23


No Components
Components
Components without associated object!
Projected Pedestrian
Valid Components in Region

Muller et al. USENIX Security’24

- Previous defense can already defend against tracking attacks
- The defense approach cross-check other visual features beyond the object detection

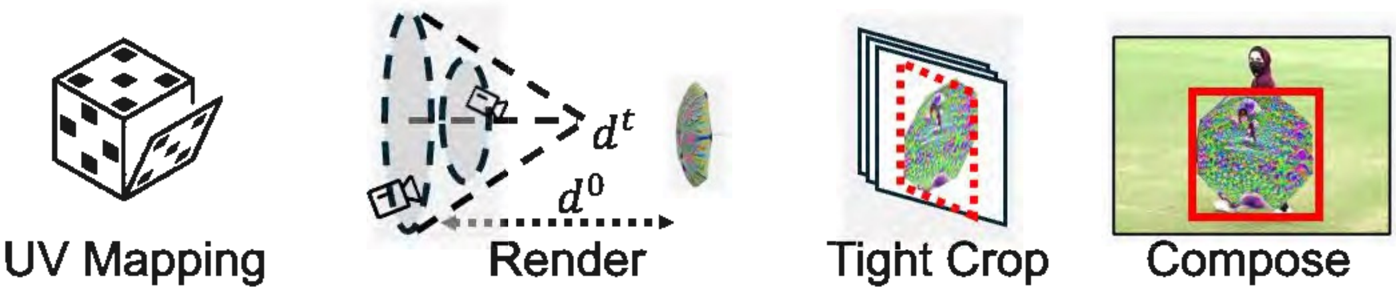
Man, Yanmao, et al. "That Person Moves Like a Car: Misclassification Attack Detection for Autonomous Systems using Spatiotemporal Consistency." *USENIX Security* 2023.
Muller, Raymond, et al. "VOGUES: Validation of Object Guise using Estimated Components." *USENIX Security* 2024.

Attack Vector

We use umbrella as a novel attack vector for printing the adversarial patterns

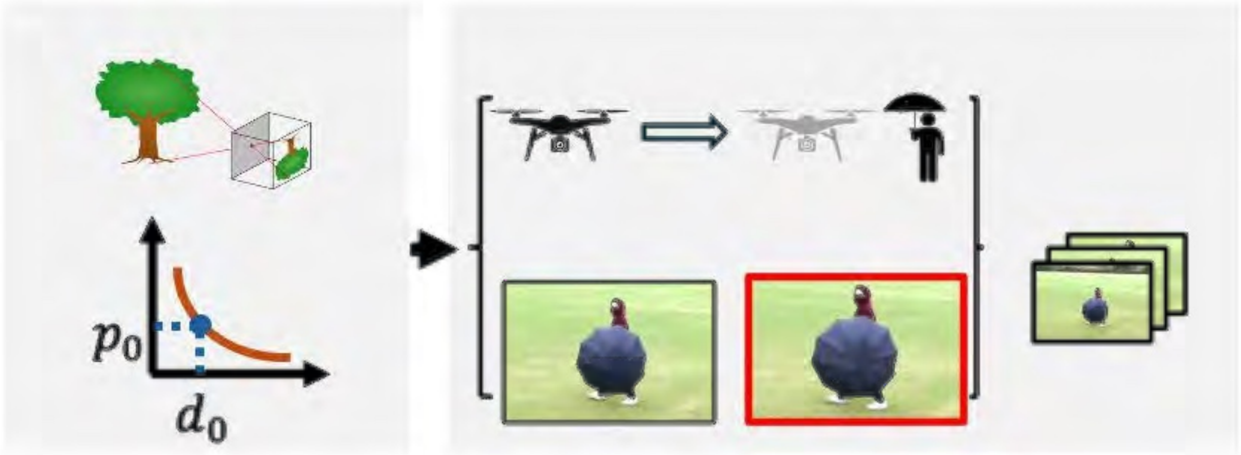
- It offers a large, rigid surface for pattern printing
- It is easy to carry and requires minimal setup to deploy
- It offers fine control, allowing the attackers to maximize the adversarial pattern exposure and conceal themselves

Methodology: Attack Vector Modeling



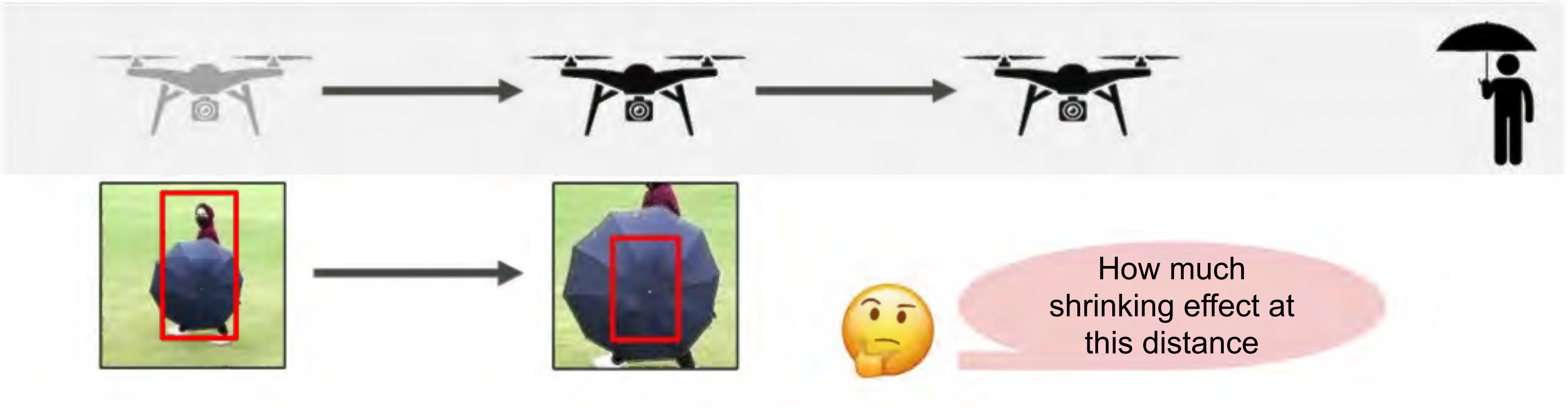
We apply the above pipeline to simulate the umbrella geometry within the optimization process, it includes (1) UV Mapping, (2) Rendering, (3) Tight Crop, and (4) Compose

Methodology: Closed-Loop Simulation



- We first estimate the initial distance based on pinhole camera model
- Then, we simulate the image at a closer distance by zooming-in
  - This simulates the image view during the distance-pulling process

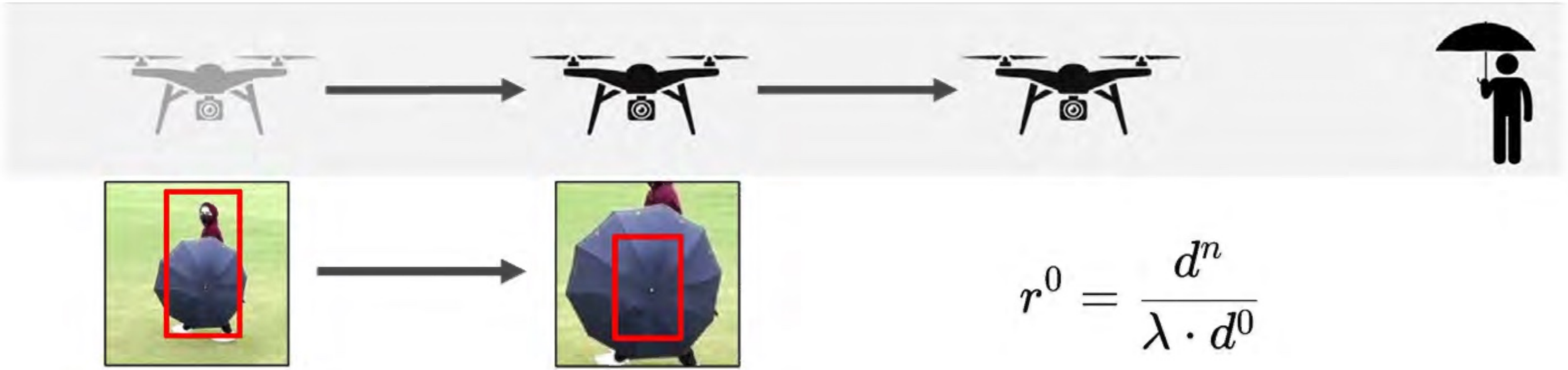
Methodology: Adversarial Objective Derivation



How much shrinking effect at this distance

- We need to know the adversarial objective at each simulated distance
- One possible solution is to set them as small as possible
  - However, this will break the spatial-temporal consistency

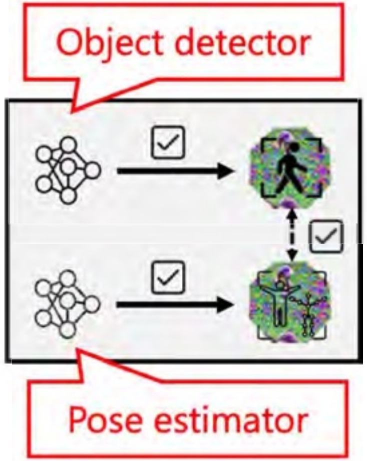
Methodology: Adversarial Objective Derivation

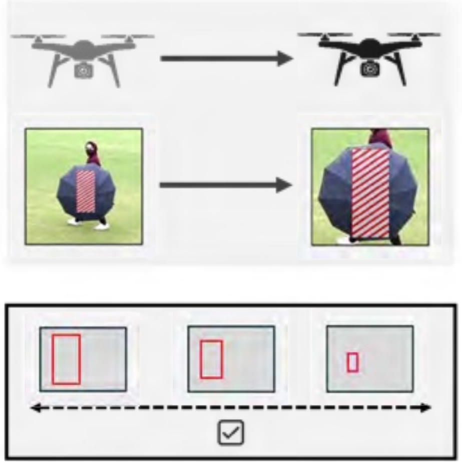


$$r^0 = \frac{d^n}{\lambda \cdot d^0}$$

- We derive that, to pull the drone into the next distance, the upper bound shrink rate must satisfy the above formula
- Therefore, the adversarial objective can only be selected between 0 and the derived upper bound

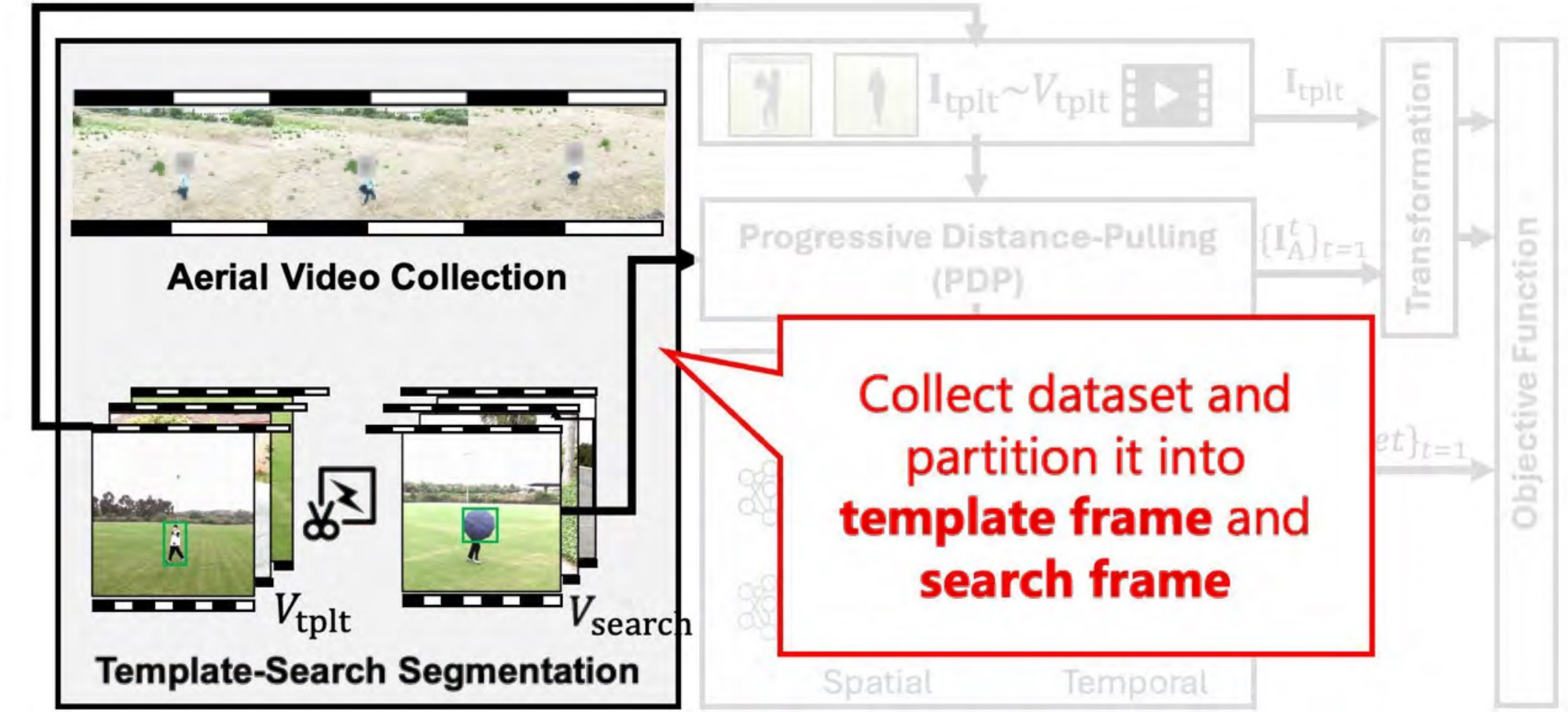
Methodology: Spatial-Temporal Consistency





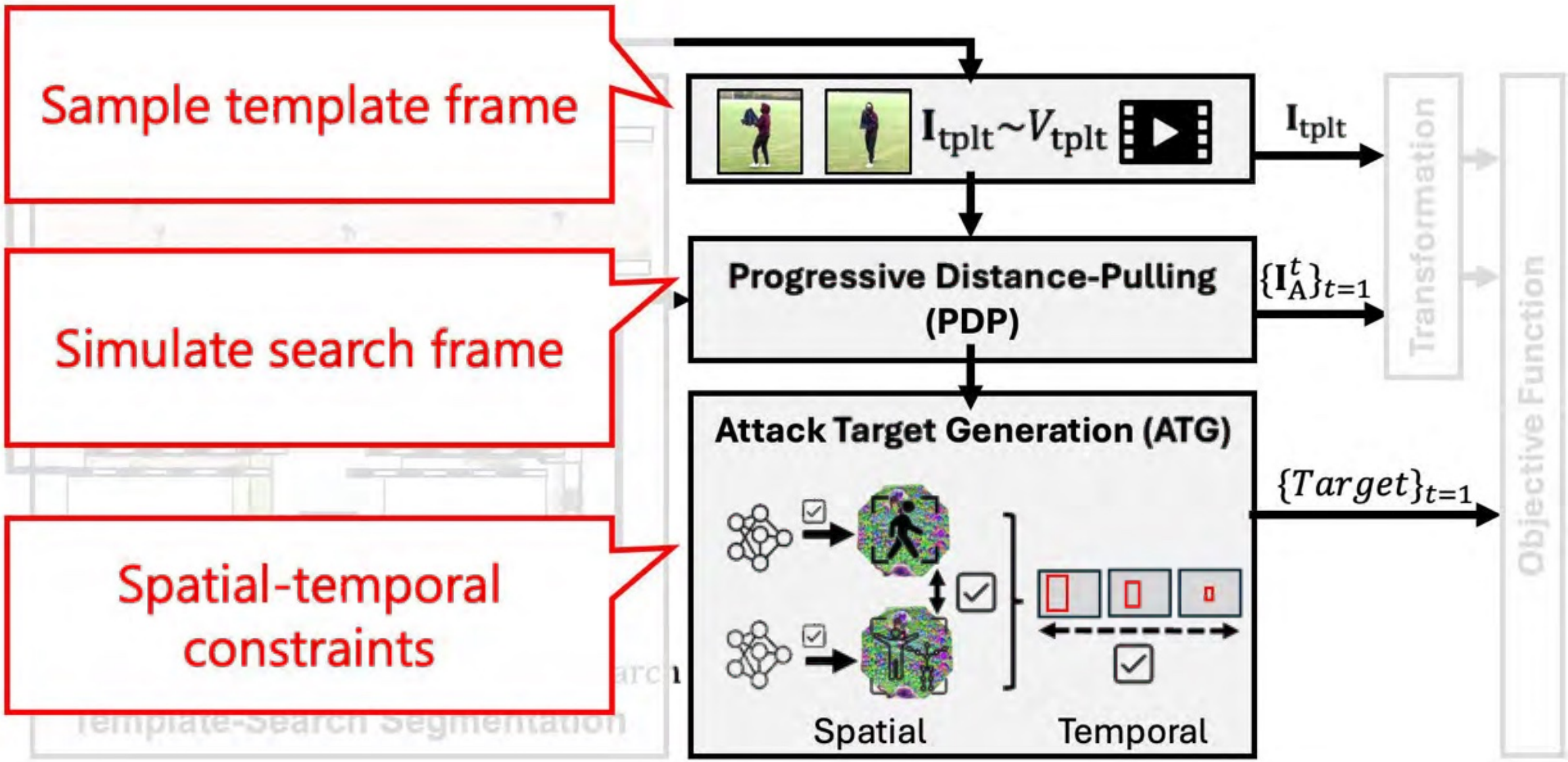
Given the simulated images within the distance-pulling effect, we can control the spatial-temporal feature by adding additional constraint into the adversarial objective

Optimization Pipeline Overview


Aerial Video Collection
V_{tplt}
V_{search}
Template-Search Segmentation
$\mathbf{I}_{\text{tplt}} \sim V_{\text{tplt}}$
$\mathbf{I}_{\text{tplt}}$
Progressive Distance-Pulling (PDP)
$\{\mathbf{I}_{\text{A}}^{t}\}_{t=1}$
$et\}_{t=1}$
Transformation
Objective Function
Spatial
Temporal

Collect dataset and partition it into **template frame** and **search frame**

Optimization Pipeline Overview



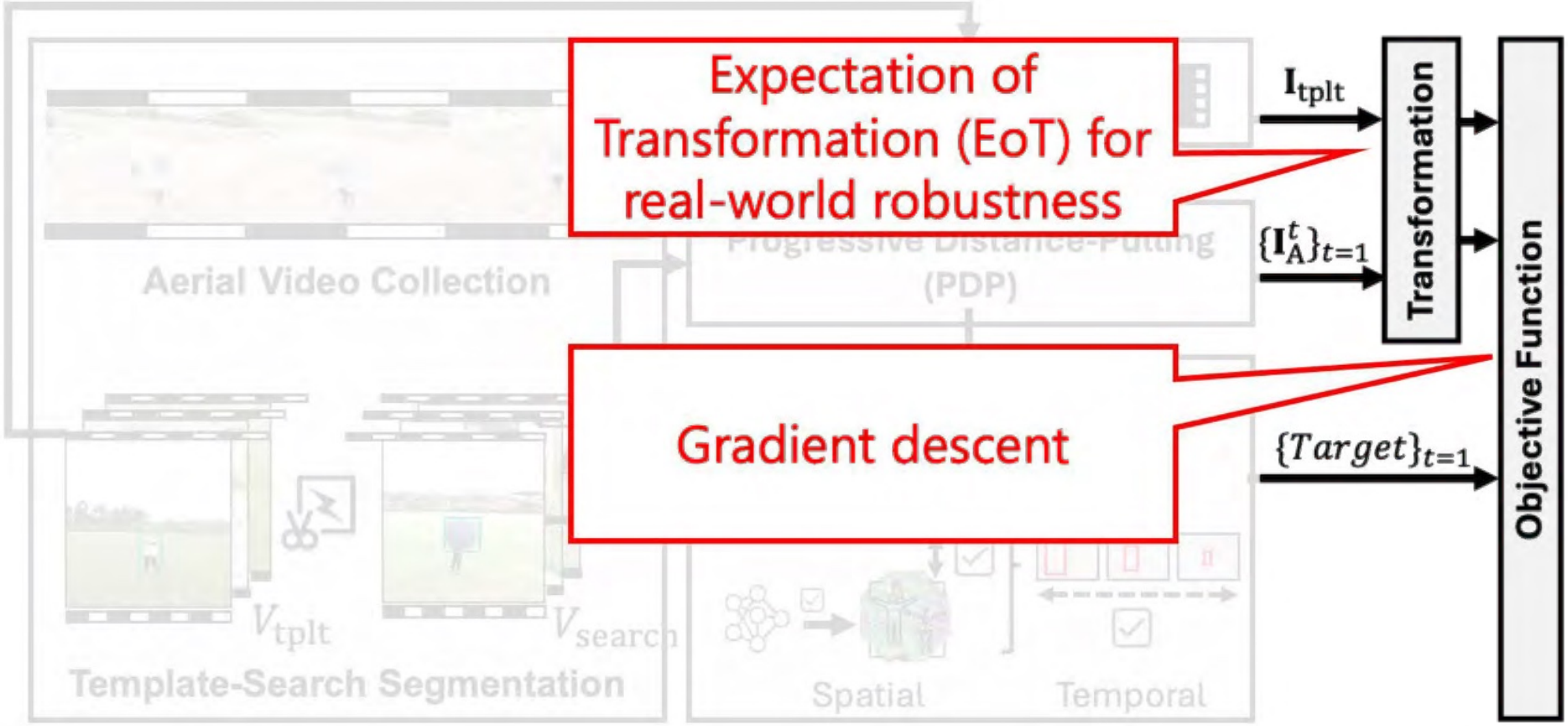
Sample template frame

Simulate search frame

Spatial-temporal constraints

$\mathbf{I}_{\text{tplt}} \sim V_{\text{tplt}}$
$\mathbf{I}_{\text{tplt}}$
Progressive Distance-Pulling (PDP)
$\{\mathbf{I}_{\text{A}}^{t}\}_{t=1}$
Attack Target Generation (ATG)
$\{Target\}_{t=1}$
Spatial
Temporal
Transformation
Objective Function
Template-Search Segmentation

Optimization Pipeline Overview



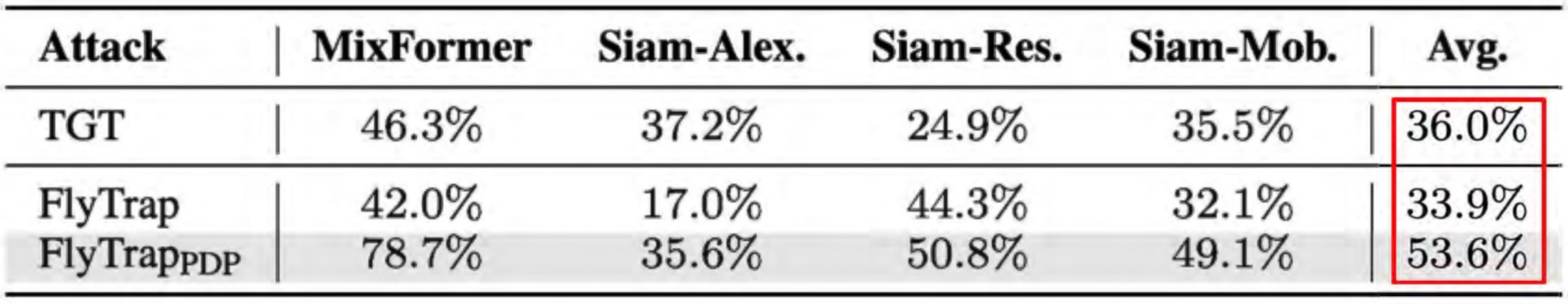
Expectation of Transformation (EoT) for real-world robustness

Gradient descent

Evaluation Setups

- We collect the dataset including 4 different individuals and 4 different background
- We evaluate both CNN-based model and Transformer-based model
- We compare with a straightforward baseline: target photo (TGT)

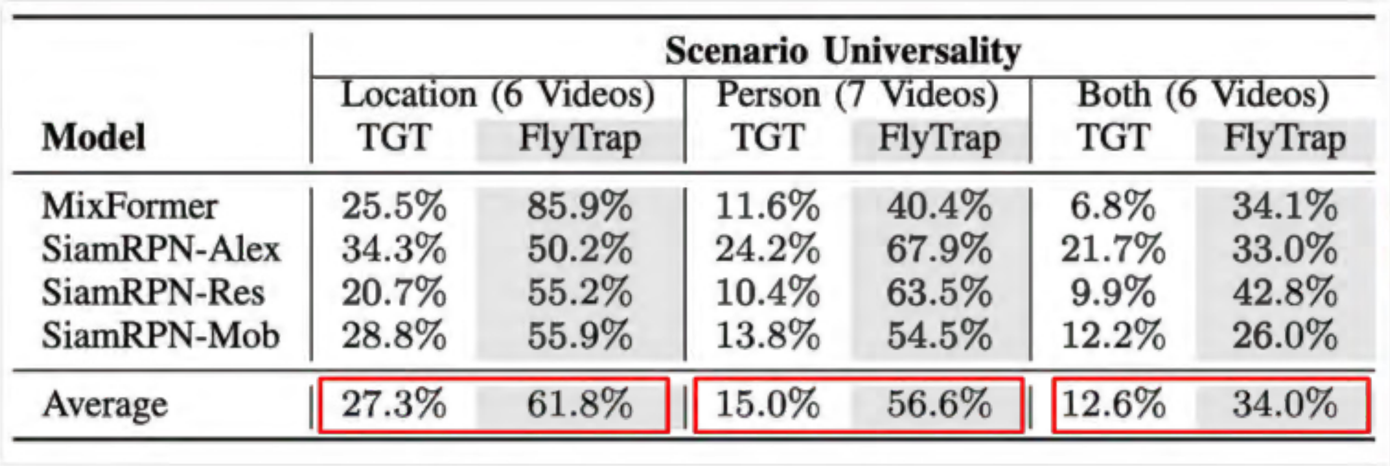
Evaluation: Attack Effectiveness



Attack	MixFormer	Siam-Alex.	Siam-Res.	Siam-Mob.	Avg.
TGT	46.3%	37.2%	24.9%	35.5%	36.0%
FlyTrap	42.0%	17.0%	44.3%	32.1%	33.9%
FlyTrap$_{PDP}$	78.7%	35.6%	50.8%	49.1%	53.6%

- FlyTrap attack can achieve much better effectiveness, surpassing the baseline TGT by **48.9%**
- Our PDP design can improve the ASR by **58.1%**

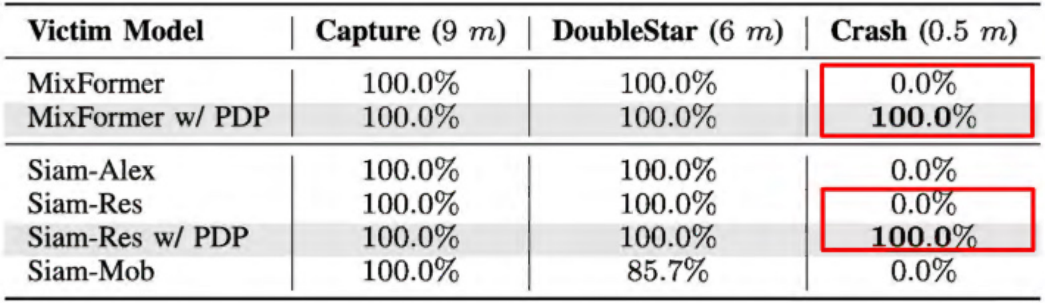
Evaluation: Attack Universality



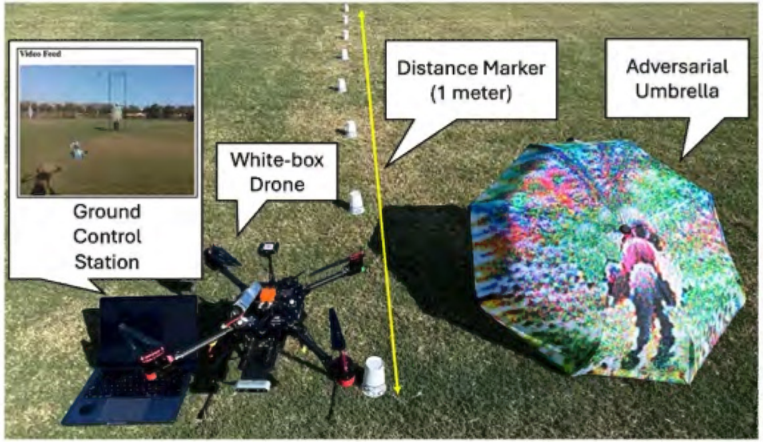
Model	Scenario Universality					
	Location (6 Videos)		Person (7 Videos)		Both (6 Videos)	
	TGT	FlyTrap	TGT	FlyTrap	TGT	FlyTrap
MixFormer	25.5%	85.9%	11.6%	40.4%	6.8%	34.1%
SiamRPN-Alex	34.3%	50.2%	24.2%	67.9%	21.7%	33.0%
SiamRPN-Res	20.7%	55.2%	10.4%	63.5%	9.9%	42.8%
SiamRPN-Mob	28.8%	55.9%	13.8%	54.5%	12.2%	26.0%
Average	27.3%	61.8%	15.0%	56.6%	12.6%	34.0%

- FlyTrap attack can achieve much better universality to unseen background and person identity
- This property can support effective attack across diverse environment and person clothing

Evaluation: Real-World Closed-Loop Attack

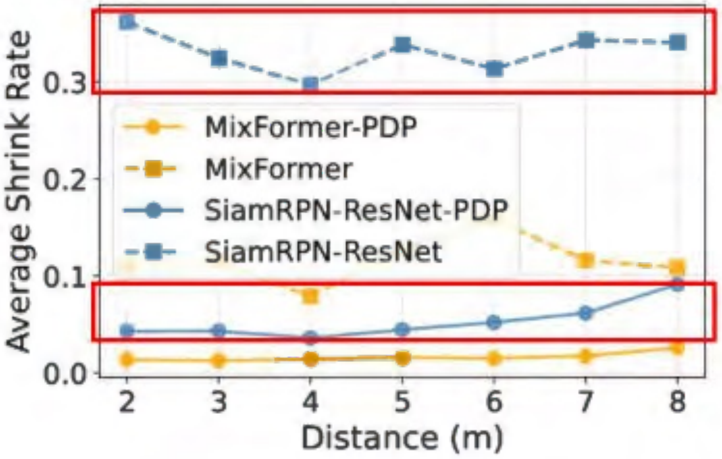


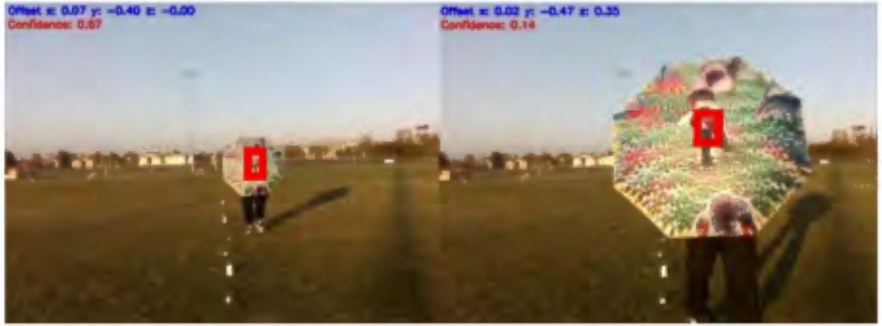
Victim Model	Capture (9 m)	DoubleStar (6 m)	Crash (0.5 m)
MixFormer	100.0%	100.0%	0.0%
MixFormer w/ PDP	100.0%	100.0%	**100.0%**
Siam-Alex	100.0%	100.0%	0.0%
Siam-Res	100.0%	100.0%	0.0%
Siam-Res w/ PDP	100.0%	100.0%	**100.0%**
Siam-Mob	100.0%	85.7%	0.0%



- Real-world closed-loop evaluation illustrates the effectiveness of distance-pulling
- Our design can pull the autonomous tracking drone within 0.5 meters with **100% success rate**

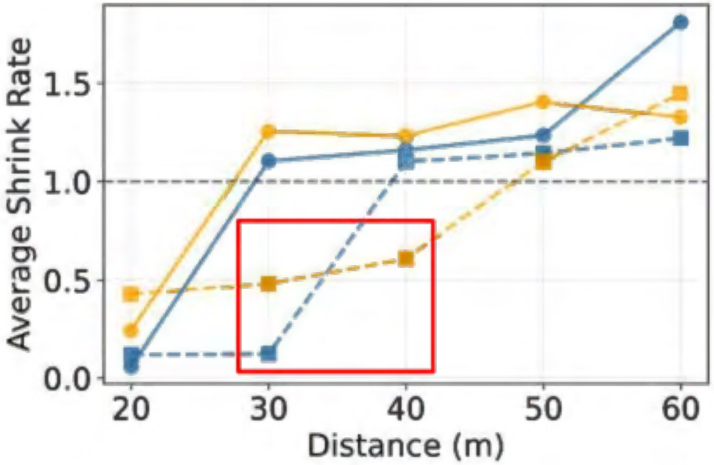
Evaluation: Real-World Attack Distance




Offset x: 0.07 y: -0.40 z: -0.00
Confidence: 0.87
Offset x: 0.02 y: -0.47 z: 0.35
Confidence: 0.14

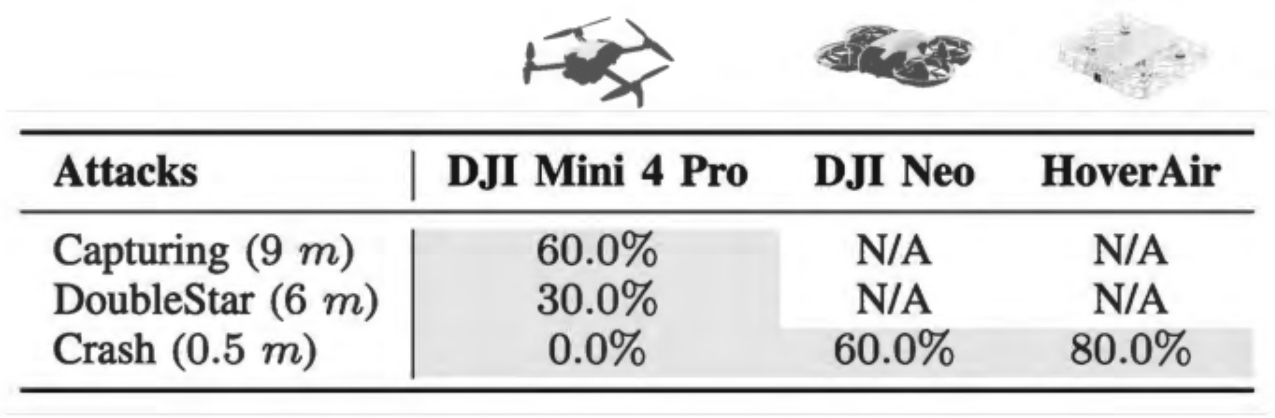
- Our Progressive-Distance Pulling design can consistently achieve lower shrink rate within 8 meters

Evaluation: Real-World Attack Distance



- Our attack can potentially generalize beyond 30 meters (*i.e., shrink rate smaller than 1.0*) even trained on data around 20 meters

Evaluation: Commercial Drones



Attacks	DJI Mini 4 Pro	DJI Neo	HoverAir
Capturing (9 m)	60.0%	N/A	N/A
DoubleStar (6 m)	30.0%	N/A	N/A
Crash (0.5 m)	0.0%	60.0%	80.0%

- Our attack can potentially transfer to commercial drones
- We can pull the DJI Neo and HoverAir-X1 within 0.5 meters over 60% success rate
- We can pull the DJI Mini 4 Pro drone within 6 meters with 30% success rate

Vulnerability Disclosure

We performed responsible vulnerability disclosure to impacted manufacturers, including DJI and HoverAir

Future Directions

- **Commercial Products**: better understanding of commercial ATT systems for real-world vulnerability identification
- **Defense**: develop adversarial training, certified robustness specifically to the single object tracking models and satisfy the real-time efficiency for ATT drones

Summary

- **Problem formulation**
  - We are the first to define distance-pulling attacks of camera-based ATT drones with novel attack vectors
- **Novel designs**
  - We propose **FlyTrap**, including a progressive distance-pulling and a controllable spatial-temporal consistency design
- **Evaluation**
  - We construct a new dataset and define new metrics for comprehensive evaluation. We show the attack effectiveness and universality
- **Physical-world impact**
  - We implement full-stack ATT drones, craft physical adversarial umbrellas, and conduct end-to-end evaluations in real-world setups

Thank you!

Q&A

FlyTrap: Physical Distance-Pulling Attack Towards Camera-based Autonomous Target Tracking Systems

Shaoyuan Xie, Mohamad Habib Fakih, Junchi Lu, Fayzah Alshammari, Ningfei Wang, Takami Sato, Halima Bouzidi, Mohammad Abdullah Al Faruque, Qi Alfred Chen

**All drone data and experiments presented in this work were completed before December 22, 2025*

Project & Demo

shaoyux@uci.edu