Adversarial Attacks On Aerial Vehicle Policies

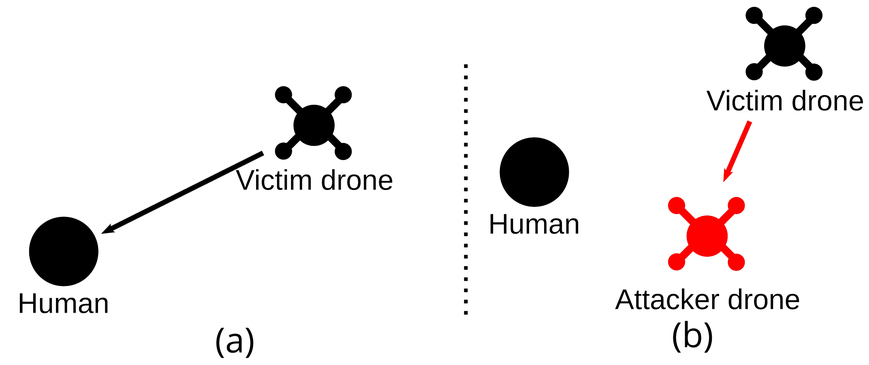
Pia Hanfeld[1,2,3], Wolfgang Hönig[3], Marina M.-C. Höhne[4], Michael Bussmann[1,2]

[1] Center for Advanced Systems Understanding, Görlitz, Germany
[2] Helmholtz-Zentrum Dresden-Rossendorf, Dresden, Germany
[3] IMRC Lab, Technical University Berlin, Germany
[4] UMI Lab, Technical University Berlin, Germany

Motivation

- **Unmanned Aerial Vehicles** (UAVs), like **quadrotors,** are utilized for **industrial** and **civil** applications
- Adversarial attacks could have severe negative impact on safe operation
- State-of-the-art adversarial attacks mainly focus on autonomous vehicle instead of UAV policies
- Adversarial attacks on UAVs
  - should be **physically realizable**,
  - with a special focus on **low-power AI**,
  - and **exploit the hardware properties** of a quadrotor

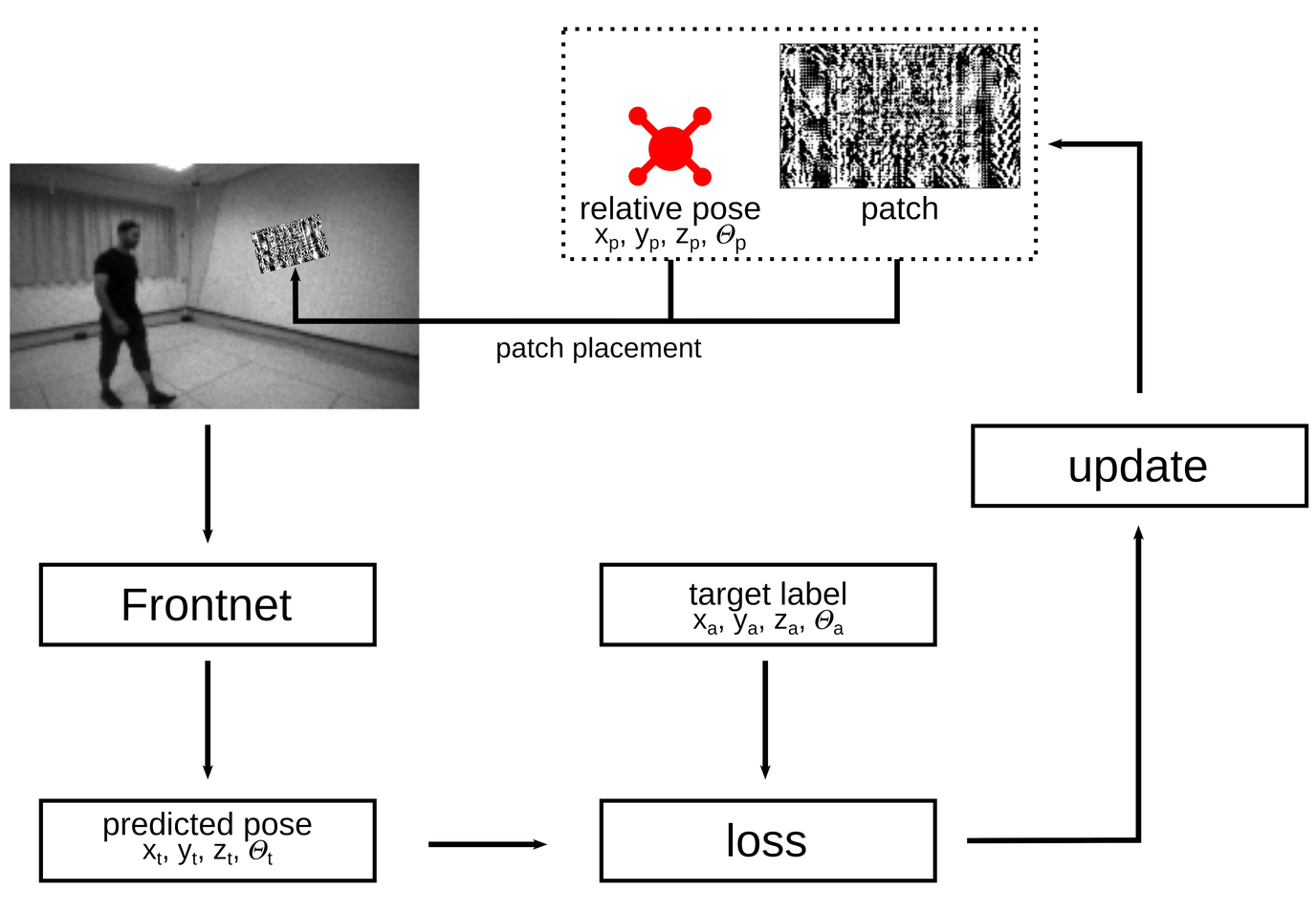
Goal



- Case (a) depicts the **regular scenario**: the victim drone is following a human target
- A neural network is predicting the relative pose of the human from camera images
- The prediction directly influences the control of the victim drone
- Case (b) depicts the **attack scenario**: an attacker drone has **full control over the victim drone**

Adversarial Attack

- We perform a **white-box adversarial attack**: the parameters of the neural network are known
- The adversarial attack will not influence the whole camera image but a small area and thus create an **adversarial patch**
- The fully optimized patch will be printed and attached to the attacker drone
- We both optimize the pixel values of the patch and the **relative pose** of the carrying attacker drone to enable control over it
- Using a patch placement algorithm, we create the artificial, **manipulated input** to the neural network
- We calculate the $\mathbf{l_2}$ **loss** between the **predictions** of the neural network and a predefined adversarial **target label**
- This loss is used to calculate the gradients w.r.t. the patch and relative pose and perform a **gradient decent step to update both**



Future Work

- Simultaneous optimization of multiple patches for full control in x, y, and z direction
- Real-world experiments
- Investigate impact on quantized neural net (low-power AI)

References

[1] Palossi, D. et al. (2022). Fully Onboard AI-Powered Human-Drone Pose Estimation on Ultralow-Power Autonomous Flying Nano-UAVs. IEEE Internet of Things Journal, 9(3), pp. 1913–1929. doi:10.1109/JIOT.2021.3091643

[2] Giernacki, W. et al. (2017). Crazyflie 2.0 quadrotor as a platform for research and education in robotics and control engineering. 22nd International Conference on Methods and Models in Automation and Robotics (MMAR). pp. 37-42. doi: 10.1109/MMAR.2017.8046794

SPONSORED BY THE
Federal Ministry of Education and Research
STAATSMINISTERIUM FÜR WISSENSCHAFT KULTUR UND TOURISMUS
Freistaat SACHSEN
SACHSEN
This action is financed with tax funds on the basis of the budget approved by the Saxon State Parliament.