Bio-Inspired Adversarial Attack

Bowei Xi
Department of Statistics
Purdue University
xbw@purdue.edu

Joint Work with Yujie Chen, Zhan Tu, Fan Fei, Xinyan Deng

Malicious Attacks Against AI

AI successful at complex tasks, such as image classification, object recognition etc.

Forward thinking AI is not secured against potential cyber and cyber physical attacks.

Adversarial attacks can cause traffic signs to be mis-classified.

Adversaries actively transform their objects to avoid detection.

Most of the literature focused on digital attacks – adding minor perturbation to a digital input.

Bio-Inspired Design

We introduce a bio-inspired attack, which is a physical attack using a moving object. It is inspired by biological intelligence.

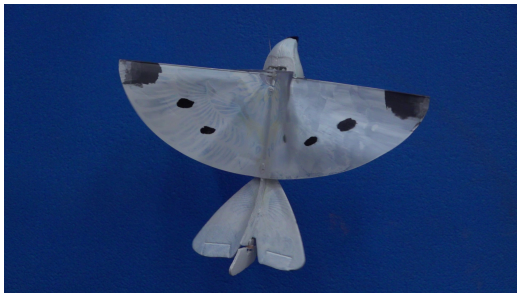
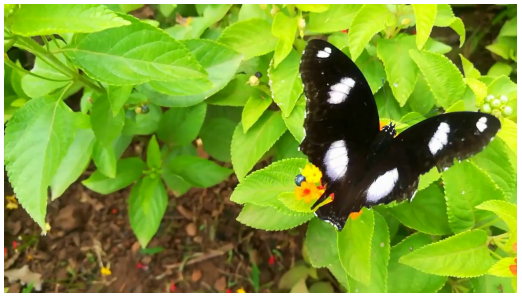
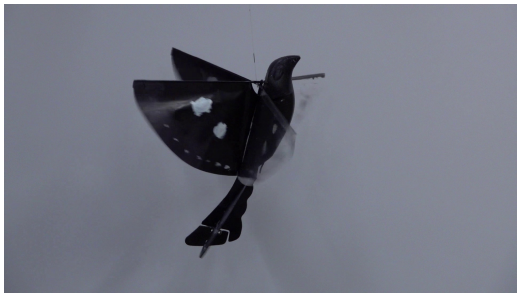
Biological intelligence has mechanisms to make living organisms hidden, and adapt to changing environments.

Praying mantis can remain hidden due to their camouflage coloration. Coloration scheme changes with the surroundings.

They move with a rocking motion mimicking a swaying plant in the wind.

Three Flapping Wing Robots



Three Flapping Wing Robots

The original robot is bird shaped, has two pairs of wings, one with color and the other transparent.

We apply camouflage, creating a superposition of three different patterns – head and body resembling a bird, tail resembling a small aircraft, wings resembling a butterfly.

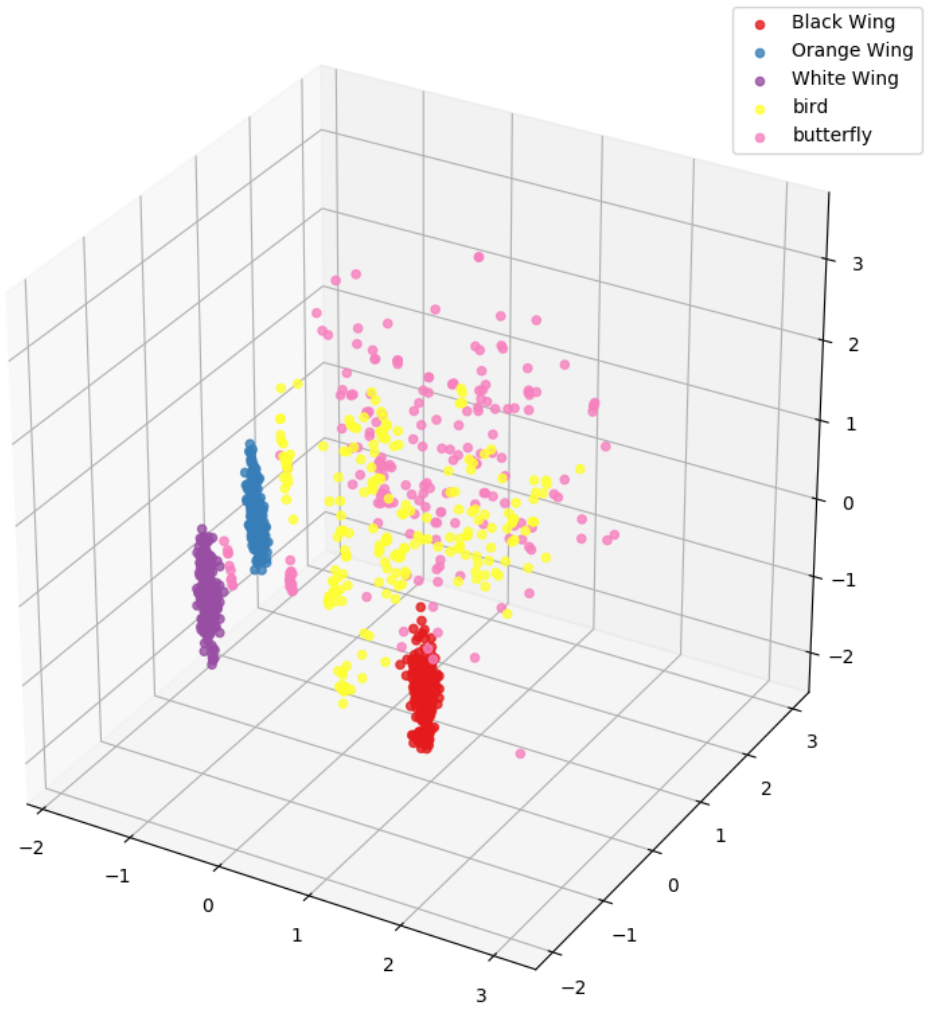
We record them flying. Both the videos and individual frames are analyzed.

Two successful attack strategies: 1) superposition of multiple patterns leading to unpredictable labels; 2) certain motion to make an object detector blind.

Attack by Superposition of Multiple Patterns

We retrain Inception V3 three times with 9 image classes. In each run, the robot class include the frames from two videos. The frames from the third video are the test samples.

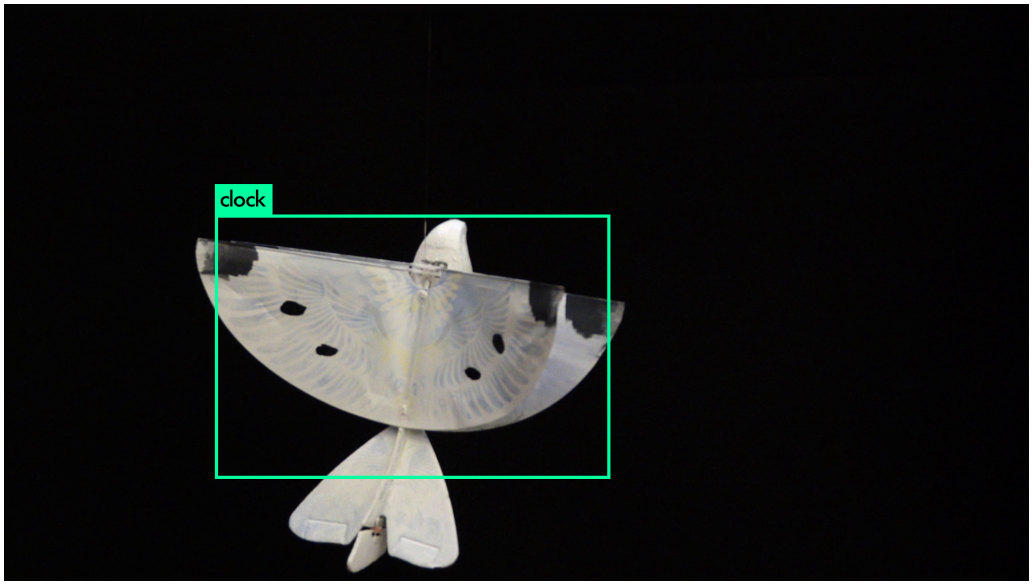
Orange	butterfly(233), robot(7)
White	butterfly(182), robot(82)
Black	butterfly(1), robot(195), bird(4), black cat(3)


Black Wing
Orange Wing
White Wing
bird
butterfly
−2
−1
0
1
2
3

Attack by Rapid Motion

We use YOLOv3. YOLOvs can process both frames and videos.

Although YOLOv3 may identify a (wrong) object from many individual frames, it is blind to the robots when processing the videos (bounding box not present throughout the videos.)



In one frame, YOLOv3 labels it as "clock" with score 0.58.

Discussion

It is possible to design bio-inspired adversarial attack using moving physical objects, which can fool different DNN based systems.

AI needs improvement, beyond simply increasing training data size.

A rapidly moving object makes it difficult to match the frames with a ground truth object, and breaks the dependency among consecutive frames.