APRICOT: A Dataset of Physical Adversarial Attacks on Object Detection (Supplementary Material)

A. Braunegg, Amartya Chakraborty, Michael Krumdick, Nicole Lape, Sara Leary, Keith Manville, Elizabeth Merkhofer, Laura Strickhart, and Matthew Walmer

The MITRE Corporation
{abraunegg, achakraborty, mkrumdick, nflett, sleary, kmanville, emerkhofer, lstrickhart, mwalmer}@mitre.org

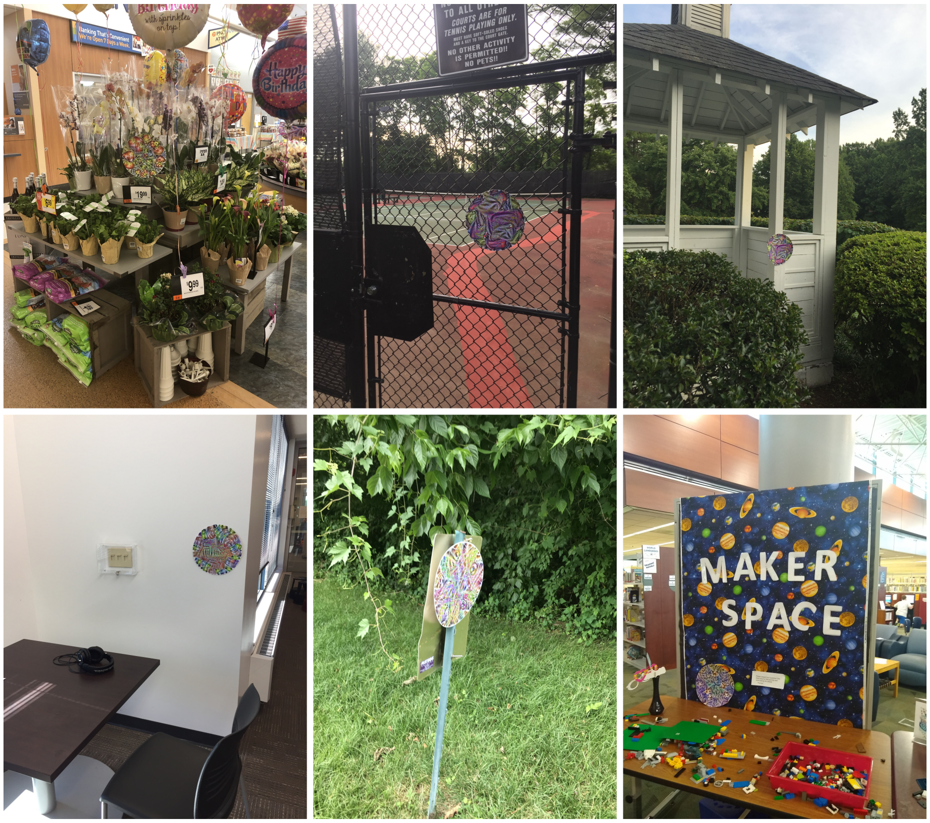

Fig. 1: Sample images from the APRICOT dataset. Images have been downsampled to reduce file size.

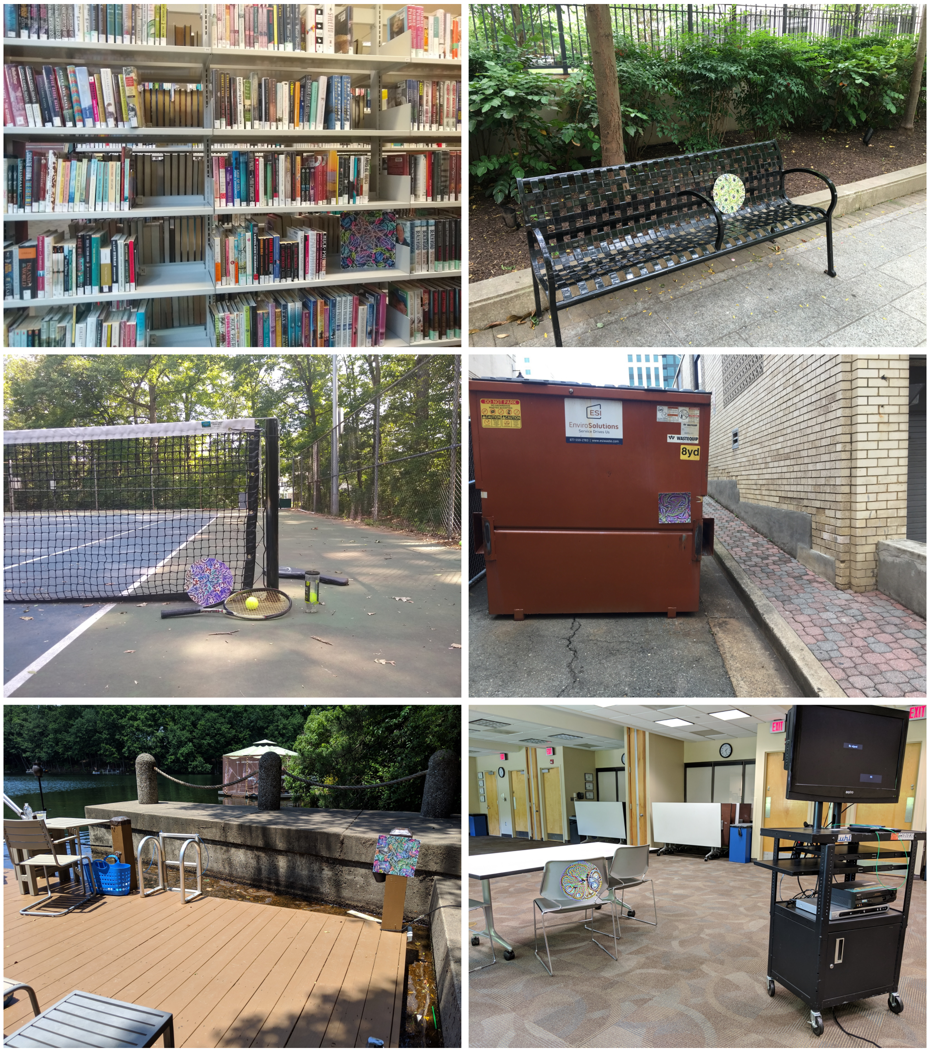

Fig. 2: Sample images from the APRICOT dataset

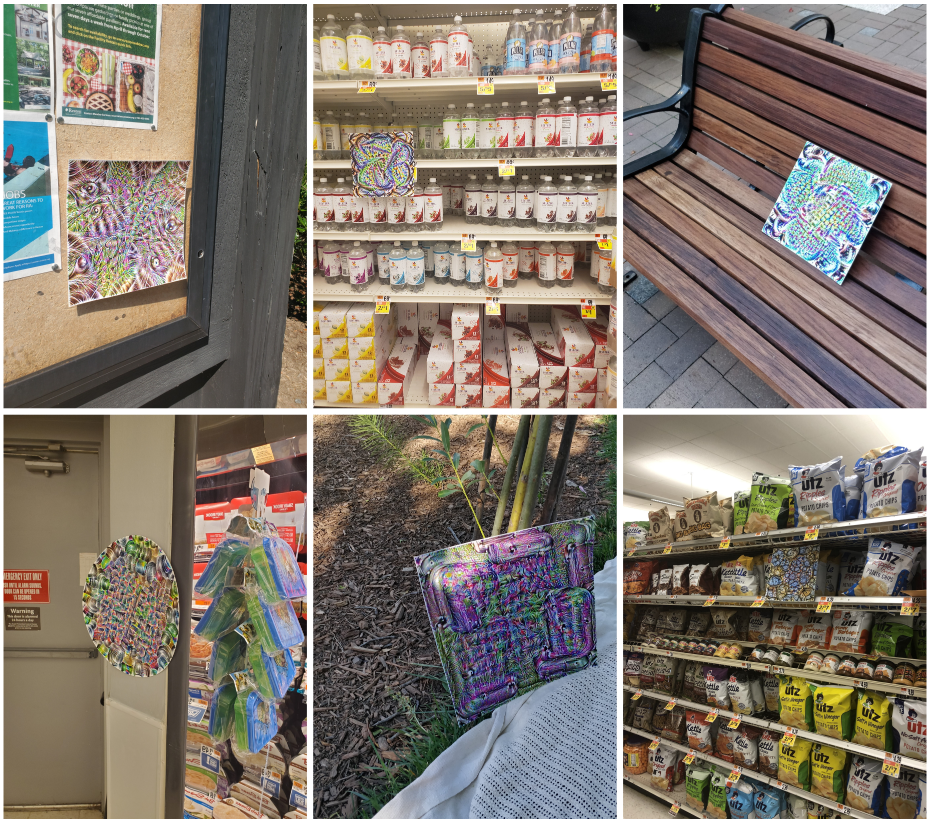

Fig. 3: Sample images from the APRICOT dataset

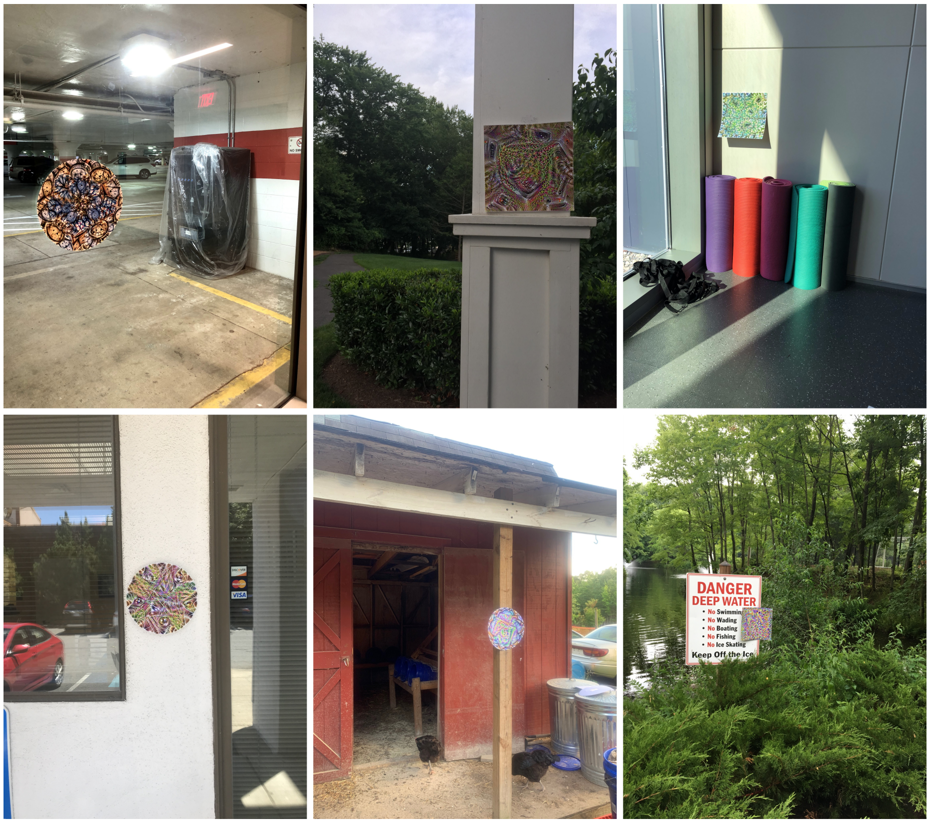

Fig. 4: Sample images from the APRICOT dataset

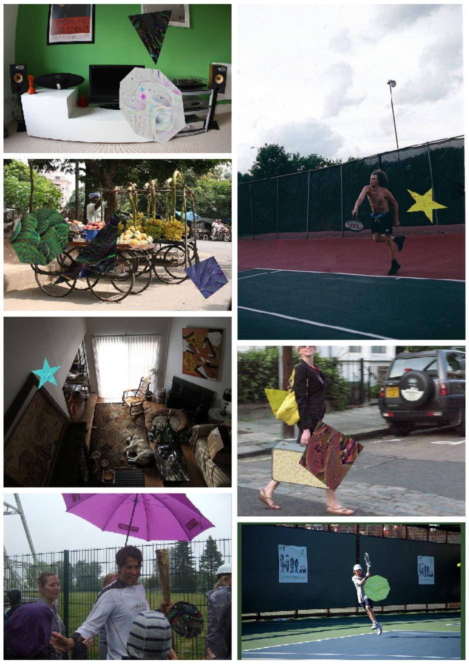

Fig. 5: Sample of "flying patch" images with digital patches superimposed over COCO images, which were used to train adversarial patch detectors

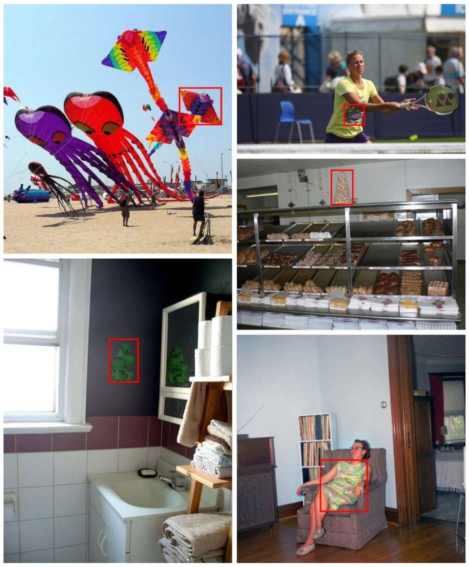

Fig. 6: Sample of high-scoring false positive patch detections in the COCO 2017 test set as produced by the best-performing model trained on flying patch images

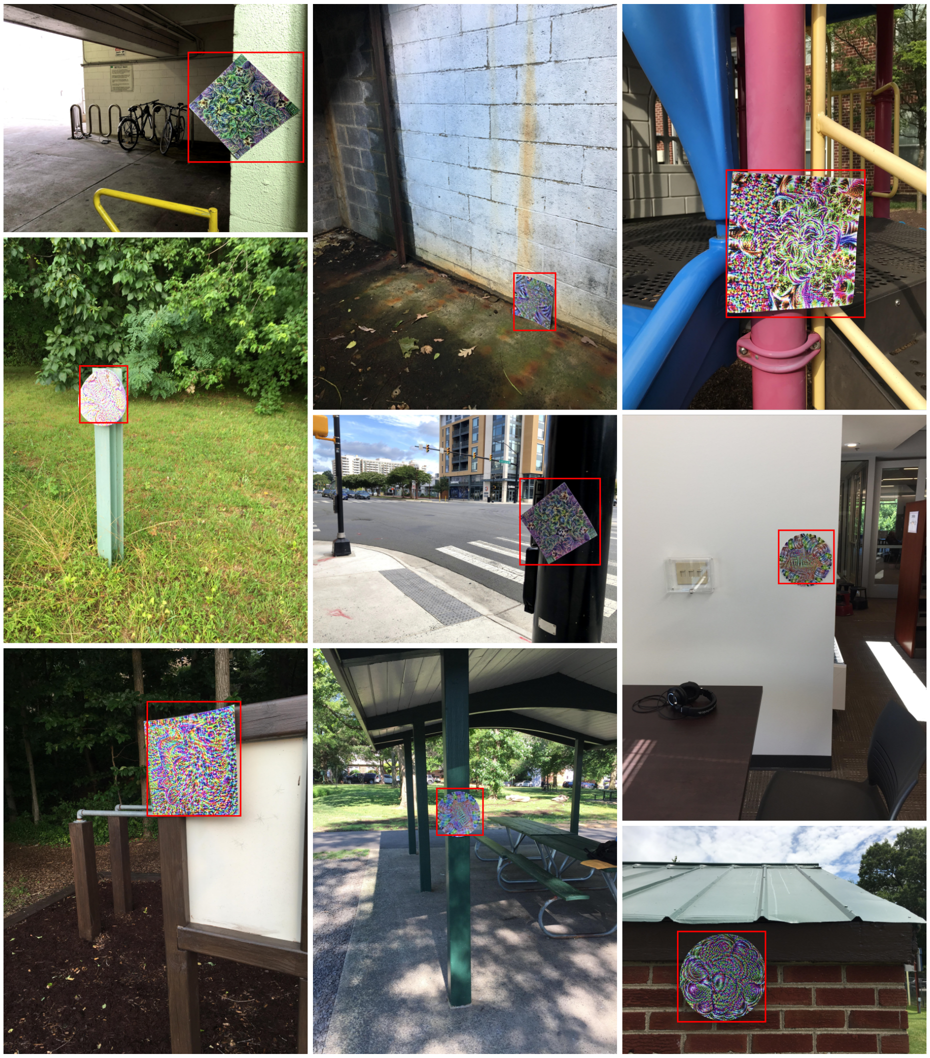

Fig. 7: Sample of APRICOT patches well-detected by the best patch-detector model trained on synthetic flying patches (Joint Adv. High Conf.). Detector outputs marked in red.

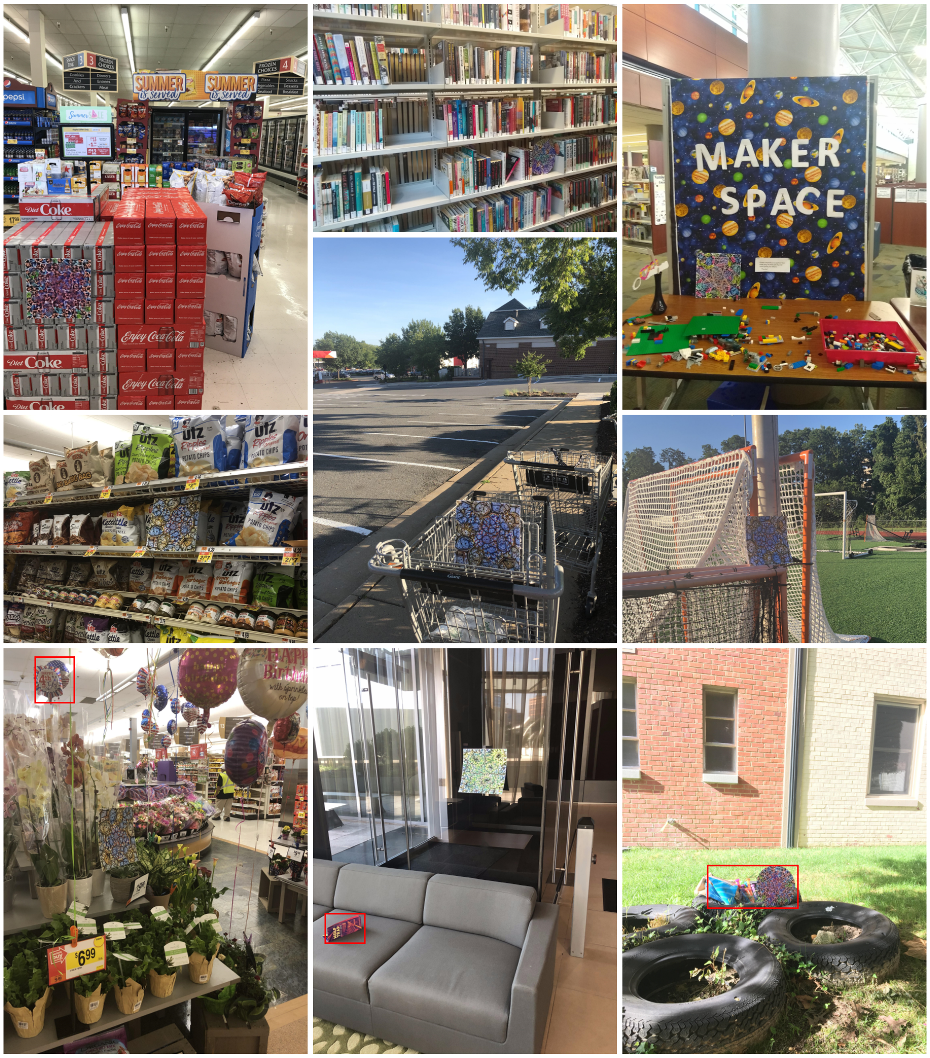

Fig. 8: Failure cases for the best patch-detector model. The top 6 images show patches that were not detected, likely due to the more complex backgrounds. The lower 3 images show benign objects incorrectly flagged as adversarial (network output marked in red). The bottom right image shows benign objects interfering with patch localization. Some images have been cropped for spacing.