ARTIFACT EVALUATED usenix ASSOCIATION AVAILABLE

USENIX'23 Artifact Appendix: Towards Targeted Obfuscation of Adversarial Unsafe Images using Reconstruction and Counterfactual Super Region Attribution Explainability

Mazal Bethany, Andrew Seong, Samuel Henrique Silva,
Nicole Beebe, Nishant Vishwamitra, Peyman Najafirad
The University of Texas at San Antonio

A Artifact Appendix

A.1 Abstract

We release the image reconstruction and explainability-based image obfuscation code that was used in our paper's experiments.

A.2 Description & Requirements

A.2.1 Security, privacy, and ethical concerns

We do not release the datasets used in our paper due to various privacy and ethical reasons.

A.2.2 How to access

Stable Reference: https://github.com/SecureAIAutonomyLab/uGuard/tree/dbd98a38611af486d992b36024f78a96f99d43cc

A.2.3 Hardware dependencies

We ran our experiments on a desktop system with an Nvidia 1080 ti GPU, and 64 GB RAM. CUDA compatible GPU's are required for our project.

A.2.4 Software dependencies

The project was designed to be run in a conda environment using python. An extensive list of software dependencies is contained within the the environment.yml file on the project repository.

A.2.5 Benchmarks

We do not release datasets or model weights, though our code is extendable to other datasets.

A.3 Set-up

A.3.1 Installation

To set up the system, users should first install conda to their system, clone the code repository, navigate to the repository, being building the environment using "conda env create -f environment.yml" and then activate the conda environment using "conda activate uGuard".

A.3.2 Basic Test

To run the code, users can navigate to the scripts directory. Users would need to add their own datasets to the datasets directory and edit the scripts so that they point to the correct datasets, save paths, etc.

To test that all packages are correctly installed, users can simply run the scripts. If the only errors received are related to missing files due to missing folders or model weights, this indicates that the environment is functioning correctly.

A.4 Version

Based on the LaTeX template for Artifact Evaluation V20220926. Submission, reviewing and badging methodology followed for the evaluation of this artifact can be found at https://secartifacts.github.io/usenixsec2023/.