AMLI Team Project

Jasmine DeHart — OUDALab

June 2021

1 ViperLib — Detect and Mitigate

This project seeks to build a library to search, detect, and mitigate visual content (image and video data). Below is a rough schedule and guide exploration and code creation. The team will implement an open source Python Library for *Viperlib*. This team will work along side Jasmine DeHart and Christan Grant from the University of Oklahoma.

Email: dehart.jasmine@ou.edu

Supervised by: Jasmine DeHart

1.1 Project Background

Private visual content exposes intimate information that can be detrimental to your finances, personal life, and reputation. Private visual content can include baby faces, credit cards, phone numbers, social security cards, house keys and others. Consumers may want to share content of themselves but hide all or parts of their visual content.

This research will impact everyday actors of SMNs and in Smart Cities by providing a mechanism to identify and mitigate sensitive information found in visual content posted on SMNs. With the use of *ViperLib*, we can lower the amount of malicious, financial and personal, attacks made on these platforms. The promise of this research endeavor is to increase the general understanding of privacy concerns and help uncover understudied cyber-attack vectors. We have proposed the use of a redaction spectrum to obfuscate visual from actors or machines (Figure 1).

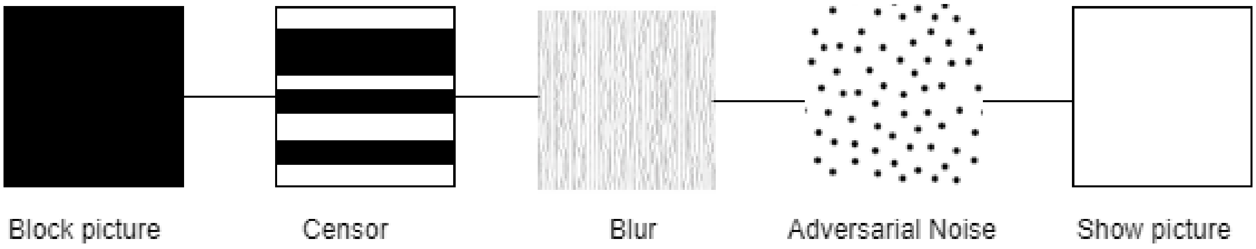

Figure 1: Spectrum of visual content redaction techniques

Once a visual content privacy leak is detected, we can handle these leaks in different ways. The first option is to block the picture, this will remove the content and/or the user's affiliation with the content from SMNs. The second option is use a censor. Censoring is essentially removing a person or object from a visual content and it will insert a blank space where the object once existed. The third option is to blur content. Blurring the content will allow the user to have some control over what is being seen without causing too much distortion. The surrounding objects will still remain visible and the leaked object will be less visible but not removing them. The fourth option is to use adversarial noise. We believe that adversarial noise will be important feature added to visual content to help protect SMNs user from computer attacks. By adding a few pixels, we could (1) hide the visual content from deep learning systems, and (2) still allow the images to be visible to humans.

1.2 Application/End User

This application is designed for end-users to process visual content and mitigate it. The bounding box coordinates will be the boundaries when redaction the content. When implementing their systems, the modeler can upload the package and call a specific method from the package.

1.3 Tasks

Your task is to begin creating ViperLib.

Week 1 Creating/Testing an Object Detection Model

Week 2 Develop Blur/Censor Redaction Techniques

Week 3 Develop Adversarial Noise Redaction Technique

Week 4 Merge Code into Python Library

Additional tasks will include estimating the likelihood an image can be uncovered after redaction, how much the object is showing after redaction, detecting or counting # of objects in an image, and lastly application design.

1.4 Relevant Classroom Knowledge

- Tensorflow
- Image Classification/Image-Video Classification
- Transfer Learning
- Visualization
- Acquiring Data
- Python - OOP