Project Progress Report: “Self-Guiding Sentinels: an Accurate Physical Attack Surveillance System”

Sarah Mundy, *srm2238*

1. Overview

The Self-Guiding Sentinels project aims to develop an innovative computer vision system capable of detecting and responding to physical surveillance equipment. Unlike existing systems that primarily focus on network infrastructure security, this project extends protections to physical surveillance assets by differentiating between legitimately innocuous activities and malicious tampering attempts.

The system employs a hybrid architecture combining Vision Transformers with lightweight diffusion components to process surveillance video in real-time. Wehen attacks are detected, the system triggers appropriate audio responses, including anthropomorphic expressions of “pain”, to deter attackers and alerts security personnel. This approach leverages cutting edge deep learning techniques optimized for surveillance applications while minimizing computational costs.

2. Research Questions

1. How can we effectively distinguish between legitimately innocuous activities such as maintenance and malicious tapering attempts on surveillance cameras using computer vision techniques?
2. What is the optimal hybrid architecture combining Vision Transformers and diffusion models to achieve high accuracy in physical attack detection while maintaining computational efficiency?
3. What pre-attack indicators can be reliably identified to enable preventative measures before serious damage occurs?
4. What is the statistical time between attacks, and how does this impact our view of false negatives versus false positives?
5. How can we develop a system that is robust against sophisticated evasion attempts designed to mimic maintenance activities?
6. How effective are anthropomorphic audio responses in deterring attackers compared to traditional alarm systems? How can we effectively compare these?

3. Value to User Community

Prospective users for this project include security system integrators and manufacturers, facility security personnel and managers, research institutions focused on surveillance technology, critical infrastructure protection teams, and public safety organizations. The Self-Guiding Sentinels system provides several unique benefits including: early detection of physical tampering before critical damage occurs, reduced false alarms through sophisticated event classification, deterrence through contextual audio responses, protection for physical components beyond standard network security measures, and real-time response capabilities with minimal latency. This project will be made available to the user community through an open source GitHub repository containing the model architecture, training code, and evaluation scripts. The repository will also include comprehensive documentation and pre-trained models which can be fine tuned for specific deployment scenarios. It will also include sample datasets for testing and validation.

4. Demo

Elevator Pitch
Self-Guiding Sentinels is a breakthrough in physical security that teaches surveillance cameras to protect themselves. Unlike traditional systems that only detect network intrusions, our solution uses advanced computer vision to distinguish between legitimate maintenance and malicious tampering. When under attack, the system responds with contextual audio cues, including human-like expressions of ‘pain’, deterring attackers and alerting security personnel. This innovative approach reduces false alarms while providing real-time protection for critical surveillance infrastructure.

Demo Plan
Introduction - Overview of the system architecture and objectives
Tampering Detection - Demonstrate system’s abilite to detect various tampering scenarios using pre-recorded footage
Response Mechanism - Show an audio response and display the system threat levels based on severity and how it logs and reports incidents
Performance Metrics - Show F1 score, ROC AUC, and response latency (maybe resource utilization)
Q&A

5. Delivery

The project deliverables will be made available through:

1. A project specific github repository will be created and made available including all code, documentation, and non-proprietary data.
2. Expected project components will be:
    a. Source code for the Vision Transformer and diffusion model architecture
    b. Training and evaluation scripts
    c. Model weights for pre-trained components
    d. Associated documentation:

        i. Installation guide
        ii. Usage examples
        iii. Deployment considerations
    e. Validation test suite

3. The repository will include references to external dependencies rather than copying them. These include:
    a. PyTorch framework
    b. Publically available UHCTD dataset