ECE 535
Professor Anwar
September 28th, 2025

Security Analysis of Gaze Estimation in Mixed Reality Project Proposal

Kelsey Erb, Breanna Gushman, and Jialing Lin

Motivation

Mixed reality (MR) head-mounted devices are becoming increasingly popular for applications like gaming, training, and everyday tasks. A key factor in these devices is gaze estimation, which tracks where the user is looking so virtual content can be aligned accurately with the physical world. In order to do this, these devices are equipped with many sensors which capture personal and sensitive information such as body motion, eye gaze, hand joints, and facial expressions. This data is used to train models and improve user experience. However, pre-trained models can be attacked with backdoors, which causes them to give incorrect outputs, compromise the integrity of the devices, and harm user experience. Due to its growing use, it is worth exploring and testing how secure these models are and what can be done to improve the security of them.

Design Goals

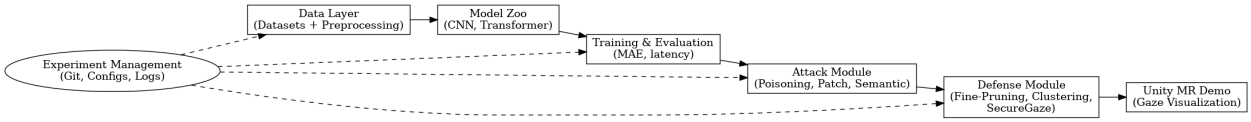
The goal of this project is to better understand the security challenges, specifically regarding backdoor attacks, of gaze estimation in MR. We aim to do the following:

- Find and collect existing gaze estimation datasets and models.
- Add backdoors to the models to see how they can be compromised.
- Analyze the security of the models, including findings from existing research.
- Test ways to remove or defend against backdoors while keeping the model accurate.
- Summarize our results, including what worked, what didn't, and what could be improved upon in future work.

Deliverables

- Spreadsheet/links of existing gaze datasets and gaze estimation models.
- Written report on security analysis on the data based on existing literature.
- GitHub repository of completed work.
- Visualization in Unity that shows anticipated gaze vectors with clean vs. backdoored to demonstrate real-world risk.
- Final project demonstration slides and/or poster.
- Written final project report on research, process, difficulties, findings, and solutions.

System Blocks



- Data Layer
    - Organize datasets into train/validation/test sets.
    - Preprocess images so they are ready for training.
- Model Zoo
    - Start with a simple CNN model
    - Compare with a Transformer-based model (ViT/Swin) to see if it improves accuracy.
- Training and Evaluation
    - Train models on gaze datasets and evaluate performance.
    - Record speed and model size to compare efficiency.
    - Use basic image augmentations to test robustness.
- Attack Module
    - Introduce backdoors by adding triggers to some training images.
    - Measure how well these attacks succeed compared to normal accuracy.
- Defense Module
    - Test defenses that detect and remove backdoors.
- MR Integration (Unity)
    - Build a Unity demo that shows where the model thinks the user is looking.
    - Compare outputs of a clean model, a backdoored model, and a defended model.
- Experiment Management
    - Use Git for version control and keep configuration files organized.
    - Log results to make comparisons easier.

Hardware and Software Requirements

To achieve the goal of understanding and defending against backdoor attacks in estimation for mixed reality, we will require both hardware and software components:

- On the hardware side, the setup includes a monitor that presents gaze targets, and a webcam that captures real-time facial images of the user. This simulates a mixed reality environment where gaze tracking enables interaction. Additionally, physical triggers, such as white tape, will be used to test how the gaze estimation model responds to visual inputs that activate hidden backdoor behaviors.

- On the software side, we will be using Python, which supports trigger injection and defense algorithms through libraries like TensorFlow, and Unity to create the mixed reality environment and visualize gaze interactions. Python and Unity will communicate through an API, allowing real-time exchange of gaze estimation and user input between the model and the interface.
- To simulate backdoor attacks, we will add backdoor triggers to the training data, causing the model to produce incorrect outputs when these triggers are present. To defend against such attacks, we will analyze the behavior of existing gaze estimation models to identify which inputs lead to incorrect predictions. By comparing the performance on clean input and triggered inputs, we can uncover patterns that activate back door behavior, which allows the models to remove the backdoor behaviors using techniques such as reverse-engineering triggers.

Team Member Responsibilities

Kelsey - Data Analysis & Writing

- Focus on finding and organizing gaze estimation datasets, analyzing the data, and interpreting the results. She will also take the lead on writing the project reports and summarizing findings made by the group.

Jialing - Software & Algorithm Design

- Responsible for implementing the software components of the project, such as modifying gaze estimation models. She also will design and test algorithms for adding and removing backdoors from the models.

Breanna - Setup & Research

- Will handle the initial setup of tools, environments, and frameworks needed for the project. She also will do background research, review existing work, and connect findings to existing work.

Project Timeline

September

- Project proposal and initial GitHub submitted.
- Preliminary research to find related work.

October

- Finalized list of relevant resources.
- Find gaze estimation datasets/models and begin security analysis and initial testing of backdoor attacks.
- Report current process, relevant research, and initial findings.
- First meeting with Professor Anwar to discuss progress.

November

- Continue to report on findings as testing is ongoing.
- Finalize security analysis report.
- Finalize implementation of backdoor attacks.
- Work on demonstration in Unity.
- Second meeting with Professor Anwar to discuss progress.

December

- Organize deliverables and do finishing touches.
- Third meeting with Professor Anwar to discuss progress.
- Final project presentation.
- Submit final deliverables to Canvas.

References

L. Du, Y. Liu, J. Jia, and G. Lan, “SecureGaze: Defending Gaze Estimation Against Backdoor Attacks,” in *Proceedings of the 23rd ACM Conference on Embedded Networked Sensor Systems. Association for Computing Machinery*, pp. 102–115, 2025, https://doi.org/10.1145/3715014.3722071

M. Goldblum *et al*., "Dataset Security for Machine Learning: Data Poisoning, Backdoor Attacks, and Defenses," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 2, pp. 1563-1580, 1 Feb. 2023, https://doi.org/10.1109/TPAMI.2022.3162397