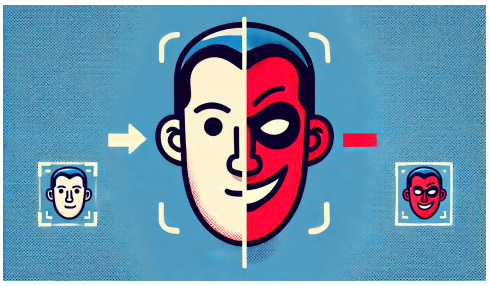
BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF INFORMATION TECHNOLOGY

Evaluating Backdoor Vulnerabilities in Facial Recognition Systems

Samuel Vojtáš (xvojta09), Rastislav Budinský (xbudin05)
Brno University of Technology, Faculty of Information Technology



Introduction:

What happens when facial recognition system is tricked into thinking an impostor is someone else? Could this seemingly uniquely-identifying biometrics system be abused by a subtle change resulting in a massive security flaw? Our project dives into these questions by exploring how backdoor vulnerabilities threaten AI-facial recognition systems. With subtle image changes, we reveal just how easy it might be to fool even the most advanced systems.

Methodology:

- **Dataset:** Images of known identities are used to simulate misclassification. Poisoning images by embedding a simple to understand square pattern.
- **Experiment Design:** Evaluate accuracy for clean & poisoned samples.
- **Tools:** ArcFace pre-trained model and PyTorch for fine-tuning final model.

Visual Representation:

Figure 1: How does it actually look like?

Key Findings:

Scenario	Trigger	
	w/	w/o
Impostor → Impostor	**3**	**7**
Impostor → Victim	**5**	**1**
Total	**8/9**	**8/9**

Scenario	Actual	Total
Victim → Victim	**13**	**13**
Non-Victim/Non-Impostor → Correct Class	**64**	**66**

Conclusion:

What we found was both surprising and alarming: even small image altering can have devastating effects on facial recognition accuracy. This raises the question—are these systems ready for real-world challenges? As we push AI further into our lives, securing it from such vulnerabilities has never been more critical.

Acknowledgments:

This project was completed as part of **BUT FIT's Biometric System** course, inspired by similar research paper[1].

[1] https://publications.idiap.ch/attachments/papers/2024/Unnervik_THESIS_2024.pdf