XIAOJIAN YUAN

+(86) 137-2102-5535 ◇ Hefei, Anhui, P.R.China
xjyuan@mail.ustc.edu.cn ◇ Website ◇ GitHub ◇ LinkedIn ◇ Zhihu

EDUCATION

University of Science and Technology of China (USTC) Sept. 2021 - Jun. 2024 (expected)
Master student in Cyber Science and Technology Advisor: Weiming Zhang
Research Interest: Trustworthy Machine Learning (Robustness, Privacy, Fairness)

China University of Mining and Technology (CUMT) Sept. 2017 - Jun. 2021
B.Eng. in Information Security, School of Computer Science and Technology
GPA: **90.69**/100, Rank: **3rd**/137 (top 2%), Graduated with Honors

PUBLICATIONS

Xiaojian Yuan, Kejiang Chen, Jie Zhang, Weiming Zhang, Nenghai Yu, Yang Zhang. Pseudo Label-Guided Model Inversion Attack via Conditional Generative Adversarial Network. Proceedings of the 37th AAAI Conference on Artificial Intelligence (AAAI), 2023. **(Oral Presentation)**

RESEARCH EXPERIENCE

PLG-MI Attack: Pseudo Label-Guided Model Inversion Attack Apr. 2022 - Aug. 2022
University of Science and Technology of China *Hefei, China*

- Proposed Pseudo Label-Guided MI (PLG-MI) attack, which can make full use of the target model and leverage pseudo-labels to guide the output of the generator during the training process.
- Proposed a simple but effective strategy to provide public data with pseudo-labels.
- Demonstrated the gradient vanishing problem of cross-entropy loss commonly adopted in previous MI attacks and use max-margin loss to mitigate it.
- Achieved the state-of-the-art performance of white-box model inversion attack.

Adversarial Robustness based on Self-Supervised Learning Feb. 2021 - May 2021
China University of Mining and Technology (Undergraduate Thesis) *Xuzhou, China*

- Reproduced a SimCLR-based adversarial defense method using PyTorch.
- Designed a MoCo-based adversarial defense method.
- Implemented a robust traffic sign recognition system.

PROJECTS

Fer2013 - Facial Emotion Recognition
This work is the final project of the computer vision course of USTC which **achieves the highest single-network classification accuracy on FER2013 based on ResNet18**. To my best knowledge, this work achieves state-of-the-art single-network accuracy of 73.70 % on FER2013 without using extra training data, which exceeds the previous work of 73.28%. [Github][Zhihu]

ISASearch: An Article Search Engine Based on Distributed Crawler This work is the final project of the Information Content Security course of CUMT. I first used the Scrapy crawler framework and the NoSQL database Redis to implement a distributed crawler, and crawled technical articles from three online communities; then I chose ElasticSearch to build a search service; finally, I built a visual site through Django, For users to search articles transparently. [Github]

HONORS & AWARDS

- First-Class Scholarship, USTC, 2021.10, 2022.10
- Outstanding Graduate, CUMT, 2021.06

- Outstanding Undergraduate Thesis, CUMT, 2021.06
- **China National Scholarship**, 2020.12
- First-Class Scholarship, CUMT, 2018.12, 2019.12

COMPETITIONS

- CVPR2021 Security AI Challenger PHASE VI Track1: White-box Adversarial Attacks on ML Defense Models, Twelfth Place Award (12/1682, 0.71%)
- CVPR2021 Security AI Challenger PHASE VI Track2: Unrestricted Adversarial Attacks on ImageNet, Seventeenth Place Award (17/1559, 1.09%)

EXTRA-CURRICULAR ACTIVITIES

Actively write blog posts and social media posts (Zhihu). Representative posts (in Chinese):

- Past and Present of Noise Contrastive Estimation: From NCE to InfoNCE **(60k views, 1.1k likes)** [link]
- Understanding of Deep InfoMax (DIM) **(27k views, 141 likes)** [link]
- Fer2013 Facial Emotion Recognition: Make the Course Project SOTA? **(9.8k views, 19 likes)** [link]

Participated in some CTF competitions during my undergraduate years and mainly focused on Web Security.

- National College Student Information Security Contest, Second Prize
- National College Student Software Testing Competition (Web Security Track), Third Prize
- Information Security Contest of CUMT, First Prize

SKILLS

Programming languages:	Python, C, C++
Web Technologies:	HTML, CSS, JavaScript, PHP, Flask
Miscellaneous:	MySQL, Linux, Git, LaTex, Markdown