JAECHUL (Harry) ROH

Final Year Student in Computer Engineering, HKUST

jroh@connect.ust.hk | https://www.jrohs.com | https://github.com/jcroh0508 | Google Scholar

EDUCATION

Hong Kong University of Science and Technology **September 2017 – Present**

B.Eng. in Computer Engineering, School of Engineering Clear Water Bay, Hong Kong

RESEARCH INTERESTS

My research interests rely on **Trustworthy AI** and **Robustness** of deep neural networks. I am especially fascinated by both adversarial attack as well as adversarial training to overcome the vulnerability of various machine learning models in a wide range of domains. I am also interested in exploring other fields of study such as the relationship between adversarial attack and federated learning, backdoor attacks/defense, fairness, and natural language processing. I am currently working on the topic of adversarial attacks to federated learning as my *Final Year Thesis* research under the supervision of Prof. Jun Zhang.

PUBLICATIONS

[1] Robust Smart Home Face Recognition under Starving Federated Data

ROH, Jaechul, Yajun Fang

IEEE International Conference on Universal Village 2022 *(IEEE UV2022)*

[paper][code][slides][video]

[2] MSDT: Masked Language Model Scoring Defense in Text Domain

ROH, Jaechul, Minhao Cheng, Yajun Fang

IEEE International Conference on Universal Village 2022 (IEEE UV2022)

[paper][code][slides][video]

[3] Impact of Adversarial Training on the Robustness of Deep Neural Networks

ROH, Jaechul

2022 IEEE 5th International Conference on Information Systems and Computer Aided Education (ICISCAE)

[paper][code]

RESEARCH / WORK EXPERIENCE

Adversarial Attacks to Federated Learning **August 2022 - Present**

Final Year Thesis, Supervisor: Prof. Jun Zhang Clear Water Bay, Hong Kong

- Investigating various gradient inversion attack methods in the context of federated learning that could be utilized under more flexible settings.

IEEE International Conference on Universal Village 2022 **May 2022 – October 2022**

Student Research Program, Supervisor: Dr. Yajun Fang (CSAIL, MIT) Cambridge, Massachusetts

- Experimented the robustness of federated learning in smart home face recognition system

MSDT: Masked Language Model Scoring Defense in Text Domain **December 2021 – May 2022**

Independent Work Research, Supervisor: Prof. Minhao Cheng Clear Water Bay, Hong Kong

- Proposed a novel improved textual defense method against backdoor attack on pre-trained language models.

Personal Research Project **January 2022 – March 2022**

Topic: "Impact of Adversarial Training on the Robustness of Deep Neural Networks"

- Experimented the effectiveness of various methods of adversarial training on improving the robustness of neural networks against classifying perturbed histopathological images.

Histopathological Scan Cancer Detection **December 2021 - January 2022**
2022 Personal Winter Project, Supervisor: Prof. Mark Vogelsberger (MIT)
- Demonstrated a user-friendly application that aids to classify whether a histopathologic scan contains metastatic cancer using modified Convolutional Neural Network and modified Resnet-18.
- In charge of implementing the neural networks for the classification task.

Super Chain AI (Conard International) **June 2021 – August 2021**
NLP Engineer Intern (Artificial Intelligence Team) Kowloon Bay, Hong Kong
- In charge of topic modeling and semantic analysis based on customer reviews and assigning specific semantics to the topics extracted.
- Competitors' analysis through web-scrapping customer reviews from other drop-shipping websites.

Military Service at Head Quarter of 12th Infantry Division **July 2018 – March 2020**
Sergeant of Republic of Korea Army Injae, Kang Won Do, Republic of Korea
- Officer Administrative Clerk Specialist
- Squad Leader of the Head Quarter

PRESENTATIONS

IEEE International Conference on Universal Village 2022 **October 2022**
Oral Presentations in Session 10B: Learning Algorithm Development, Analysis and Interpretability
- Paper - **Robust Smart Home Face Recognition under Starving Federated Data**
- Paper - **MSDT: Masked Language Model Scoring Defense in Text Domain**

Presentation Project on "Adversarial Attack" **November 2021**
Machine Learning course Final Project, Instructor: Prof. Dit-Yan YEUNG Clear Water Bay, Hong Kong
- 30-minute video presentation on the topic of "Adversarial Attack"
- Papers reviewed: *"Explaining and Harnessing Adversarial Examples" & "Is Bert Really Robust? A Strong Baseline for Natural Language Attack on Text Classification and Entailment"*
[slides][video]

SKILLS / LANGUAGES

Programing Languages: Python, C, C++
Frameworks/Libraries: PyTorch, Hugging Face, NLTK, spaCY, MongoDB, Selenium, Beautifulsoup
Languages: Korean (Native), English (Native), Chinese (Fluent)