Yuning Han

30 Park N. Lane, NJ 07310 | +1(929)684-6644

yh3612@columbia.edu | https://www.linkedin.com/in/yuning-han-827a56288/

EDUCATION

COLUMBIA UNIVERSITY — New York City, NY

M.S. in Electrical Engineering — *Sept 2023 - Present*

GPA: 3.81/4.00

Tesla Scholarship of Columbia University Electrical Engineering Department (Honored)

SHANGHAI JIAOTONG UNIVERSITY (SJTU) — Shanghai, CN

BS in Information Engineering — *Sept 2019 - Jun 2023*

GPA: 3.52/4.0

Eastern China's second prize in the Freescale Cup National College Student Smart Car Competition 2020 (Top 20%)

RESEARCH EXPERIENCE

GRADUATE RESEARCH ASSISTANT

BACKDOOR ATTACK AND DEFENSE ON DIFFUSION MODEL — Boston, MA

Advisor: Prof. Yingjie Lao, EE Department, Tufts University — *April 2024 – Nov 2024*

- Project leader and the first author of the paper '*UIBDiffusion: Universal Imperceptible Backdoor Attack for Diffusion Models*', **CVPR 2025 accepted**.
- Design a universal imperceptible image trigger generator based on Universal Adversarial Perturbation (UAP) with a trainable network and an image classifier network, generate various triggers based on different datasets and classifier networks, all proved to be imperceptible both on image level and on noise level during the diffusion process.
- Use our trigger to backdoor attack diffusion models and achieve an amazingly high Attack Success Rate (ASR) with **100%** in a very low data poisoning rate (**5%**). The model performs well on different generation models and different samplers.
- Evaluate our trigger on Elijah and TERD, two SOTA backdoor attack defense methods. Experiments show that our trigger can fully escape these two defense methods with a **100%** escaping success rate and existing methods can't reverse our trigger.

UNDERGRADUATE RESEARCH ASSISTANT

ADAPTIVE SEMANTIC COMMUNICATION SYSTEM FOR VIDEOS ON MACHINE LEARNING — Shanghai, CN

Advisor: Prof. Zhiyong Chen, Institute Of Wireless Communications, SJTU — *Jul 2022 - June 2023*

- Propose a hierarchical adaptive semantic communication ML system, applying it to the processing of video semantics used for video reconstruction at the users' head receiving compressed key frames with no background knowledge.
- Propose a learnable model to extract and encode video keyframes based on the U-net structure. The receiver can use the compressed key frames to reconstruct the full frames with no background knowledge based on the trained model. Collaborate with the Kinetics dataset to train the model and evaluate its performance.
- Achieved a compression ratio of **99.87%**, video keyframe restoration accuracy is over **95%** under simple noise circumstances.

INTERNET COMMUNICATION SCENARIO SECURITY PROTECTION — Shanghai, CN

Advisor: Prof. Liyao Xiang, John Hopcroft Computer Science Center, SJTU — *Jul 2021 - Feb 2022*

- Build an attacking defense model against typical attacking scenarios based on GMM, GAN, and DQN.
- Propose a model to defend the black-box attack against datasets based on GAN with clean and poisoned datasets, making the model resilient to tag attacks. Train and evaluate over CelebA dataset, and achieve a classification accuracy of over **90%**.
- Design a DQN network to improve the model's performance compared with the original model, and improve accuracy by **15%**.

ACADEMIC PROJECT

GRADUATE: Deep research in the Swin-Transformer model, reimplement and evaluate, achieve a final classification accuracy of **99.64%**, compared to **63.4%** without shifting window. *Advised by Dr. Mehmet Kerem Turkcan, Columbia University, 2023.*

UNDERGRADUATE: Development and integration of A-C RL-based ABR algorithm for video streaming, achieve an error rate lower than **0.02** over online transformation scenario. *Advised by Prof. Roger Zimmermann, School Of Computing, NUS, 2021.*
Vision and electromagnetic guidance solutions for smart cars, won the second prize. *Advised by Prof. Bing Wang, SJTU, 2020.*

SKILLS AND TECHNOLOGIES

- Programming Languages: C/C++, Python (mainly)
- Technologies: PyTorch, TensorFlow, CNN, RNN, Transformer, Diffusion Model, NeRF, Large Language Model, AI Security

PUBLICATION

[1] *UIBDiffusion: Universal Imperceptible Backdoor Attack for Diffusion Models, Yuning Han, Bingyin Zhao, Rui Chu, et al.*