ZHANHAO HU

Soda Hall, University of California Berkeley, CA, 94720
+1 341-333-8522 ◇ zhanhaohu.cs@gmail.com

EDUCATION

Tsinghua University, Beijing *2017 - 2023*
Ph.D., Computer Science and Technology
Advisor: Bo Zhang and Xiaolin Hu
Dessertation: The Practicality of Physical Adversarial Examples for Deep Learning Models.

Tsinghua University, Beijing *2013 - 2017*
B.S., Mathematics and Physics
Advisor: Xiaolin Hu
Dessertation: STDP-based learning for spiking neural networks

RESEARCH EXPERIENCE

University of California, Berkeley January 2024 - Present
Postdoctoral Researcher *Berkeley, California*

- I'm affiliated with the Institute for Data Science (BIDS) and advised by Prof. David Wagner. I led and was involved in multiple projects primarily related to security issues in Large Language Models (LLMs), such as evaluating, detecting, and defending against existing threats to LLMs.

Tsinghua University July 2023 - December 2023
Research Assistant *Beijing, China*

- I worked with Prof. Xiaolin Hu at the Tsinghua Laboratory of Brain and Intelligence (THBI). My research mainly concerned security issues in Computer Vision (CV) models, such as privacy and physical adversarial examples.

PUBLICATIONS

Under review & *Preprint*

Zhanhao Hu, Julien Piet, Geng Zhao, Jiantao Jiao, and David Wagner (preprint). Toxicity Detection for Free. arXiv preprint arXiv:2405.18822, 2024

Qiongxiu Li, Lixia Luo, Agnese Gini, **Zhanhao Hu**, Xiao Li, Chengfang Fang, Xiaolin Hu, Jie Shi. (under review). On the Hardness of Input Reconstruction Attack via Gradient Sharing in Federated Learning: A Cryptographic View.

Xiaopei Zhu, Siyuan Huang, **Zhanhao Hu**, Jianmin Li, Xiaolin Hu. (under review). Physical Adversarial Attack to Person Detectors in Infrared Images based on 3D Modeling.

Published

Xiao Li, Wei Zhang, Yining Liu, **Zhanhao Hu**, Bo Zhang, Xiaolin Hu (2024). Language-Driven Anchors for Zero-Shot Adversarial Robustness. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

Xiao Li, Qiongxiu Li, **Zhanhao Hu**, Xiaolin Hu (2024). On the Privacy Effect of Data Enhancement via the Lens of Memorization. IEEE Transactions on Information Forensics and Security (IEEE TIFS).

Xiaopei Zhu, **Zhanhao Hu**, Siyuan Huang, Jianmin Li, Xiaolin Hu (2023). Hiding from Infrared Detectors in Real World with Adversarial Clothes. Applied Intelligence.

Zhanhao Hu, Wenda Chu, Xiaopei Zhu, Hui Zhang, Bo Zhang, Xiaolin Hu (2023). Physically Realizable Natural-Looking Clothing Textures Evade Person Detectors via 3D Modeling. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

Tong Wang, Xiaohui Kuang, Qianjin Du, **Zhanhao Hu**, Huan Deng, Gang Zhao. (2023). Driving into Danger: Adversarial Patch Attack on End-To-End Autonomous Driving Systems Using Deep Learning. In 2023 IEEE Symposium on Computers and Communications (ISCC)

Zhanhao Hu, Jun Zhu, Bo Zhang, Xiaolin Hu (2022). Amplification trojan network: Attack deep neural networks by amplifying their inherent weakness. Neurocomputing, 505, 142-153.

Zhanhao Hu, Siyuan Huang, Xiaopei Zhu, Fuchun Sun, Bo Zhang, Xiaolin Hu (2022). Adversarial texture for fooling person detectors in the physical world. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR Oral).

Xiaopei Zhu, **Zhanhao Hu**, Siyuan Huang, Jianmin Li, Xiaolin Hu (2022). Infrared invisible clothing: Hiding from infrared detectors at multiple angles in real world. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR Oral).

Zhanhao Hu, Tao Wang, Xiaolin Hu (2017). An stdp-based supervised learning algorithm for spiking neural networks. In Neural Information Processing: 24th International Conference (ICONIP).

TALKS AND SYMPOSIA

Adversarial texture for fooling person detectors in the physical world. Oral Presentation at the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022/06

Adversarial texture for fooling person detectors in the physical world. Shield Laboratory. 2022/06

Adversarial texture for fooling person detectors in the physical world. Beijing RealAI Intelligent Technology Co., Ltd. 2022/05

Adversarial texture for fooling person detectors in the physical world. Beijing Jiangmen Development Venture Capital Center, L.P. 2022/05

Defend Adversarial Examples by Robust Sparse Coding Neural Networks. Poster Presentation at the McGovern Institutes Joint Neuroscience Symposium, MIT. 2019/03

TEACHING EXPERIENCE

Neural and Cognitive Computation (No.80240642), Tsinghua University *2019 Autumn*
Teaching Assistant

Neural and Cognitive Computation (No.80240642), Tsinghua University *2018 Autumn*
Teaching Assistant

PROFESSIONAL SERVICE

Journal Reviewer

- TIP, TPAMI, TNNLS

Conference Reviewer

- CVPR, ECCV, AAAI, ICML, ICIST, ICACI, ICICIP, ISNN

Workshop Reviewer

- ICML2021 adversarial ML workshop