600 W Grove Pkwy, Apt 2039, Tempe, AZ, 85283
☎ 734-2392-561
✉ xiaocw@asu.edu
http://xiaocw11.github.io/
Google Scholar: citations 5000+

Chaowei Xiao

Education

2015.8-2020.8 **University of Michigan, Ann Arbor**.
Ph.D. in Computer Science, EECS. Advisor: Prof. Mingyan Liu.
Dissertation: "Secure Learning in Adversarial Environments".

2011.8-2015.7 **Tsinghua University**.
B.S. in Computer Software.
B.S. in Economics, School of Economic and Management.

Employments

2022.8-now **Assistant Professor**, *School of Computing and Augmented Intelligence*, Arizona State University.

2020.9-now **Research Scientist**, *AI Algorithm Group*, NVIDIA Corporation.

Honors & Awards

2022	**ACM Gordon Bell Special Prize** for HPC-Based COVID-19 Research
2022	**Best Paper Award** in ICML SRML workshop
2021	**Best Paper Award** in ESWN
2019	Exhibition of “Physical Stop Sign” in London Science Museum.
2014	**Best Paper Award** in MobiCom
2014	First Prize in the 32nd Tsinghua Great Challenge Cup
2014	Intel Chinese Outstanding Student Scholarship
2013-2014	National Innovation and Entrepreneurship Training Program
2013	Tencent Chinese Outstanding Student Scholarship
2012-2014	First Class Scholarship for Overall Excellence

Three Representative Publications (* indicates equal contributions)

[3] **Chaowei Xiao***, Zhongzhu Chen*, Kun Jin*, Jiongxiao Wang*, Weili Nie, Mingyan Liu, Anima Anandkumar, Bo Li, Dawn Song. *DensePure: Understanding Diffusion Models towards Adversarial Robustness.* ICLR 2023. **TL;DR**: Based on our previous [31], which, for the first time, discovered the effective diffusion model to defend against unseen adversarial examples without adversarial training, our work theoretically explained why and how diffusion models could enhance adversarial robustness and achieved state-of-the-art L2-certified robustness.

[2] Yulong Cao*, Ningfei Wang*, **Chaowei Xiao***, Dawei Yang*, Jin Fang, Ruigang Yang, Alfred Chen, Mingyan Liu, Bo Li. *Invisible for both Camera and LiDAR: Security of Multi-Sensor Fusion based Perception in Autonomous Driving Under Physical-World Attacks*. IEEE Symposium on Security and Privacy 2021. **TL;DR**: We are the first to show vulnerabilities of the multi-sensor fusion perception framework of real-world autonomous driving systems (e.g., Baidu Apollo) by physically generating adversarial objects.

[1] **Chaowei Xiao***, Jun-Yan Zhu*, Bo Li, Warren He, Mingyan Liu, Dawn Song. *Spatially Transformed Adversarial Examples*. ICLR 2018. **TL;DR**: Our work broadened the traditional Lp-based adversarial examples and opened up a new domain on non-Lp-bounded adversarial examples.

Publications (* indicates equal contributions)

Summary Total Citations: 5497, H-Index: 23 (Google Scholar [link (click)], as of Nov 22, 2022)

27 papers in commonly-recognized top-tier machine learning conferences (NeurIPS, ICML, ICLR, CVPR, ICCV, ECCV, CORL, ICDM, AAMAS, IJCAI)

8 papers in commonly-recognized top-tier security and system conferences (IEEE Security & Privacy, USENIX Security, ACM CCS, ACM MobiCom, IEEE INFOCOM)

1 paper in Top Survey journal (ACM Computing Survey) and 2 papers in Top journals (TMC, TDSC)

Students mentored by me are underlined

[39] **Chaowei Xiao***, Zhongzhu Chen*, Kun Jin*, Jiongxiao Wang*, Weili Nie, Mingyan Liu, Anima Anandkumar, Bo Li, Dawn Song. *DensePure: Understanding Diffusion Models towards Adversarial Robustness.* ICLR 2023.

[38] Shutong Wu, Jiongxiao Wang, Wei Ping, Weili Nie, **Chaowei Xiao**. *Defending against Adversarial Audio via Diffusion Model.* ICLR 2023

[37] Zichao Wang, Weili Nie, Zhuoran Qiao, **Chaowei Xiao** , Richard Baraniuk, Anima Anandkumar . *Retrieval-based Controllable Molecule Generation* ICLR 2023 (spotlight)

[36] Zhiyuan Yu, Yuanhaur Chang , Ning Zhang, **Chaowei Xiao**. *SMACK: Semantically Meaningful Adversarial Audio Attack* USENIX Security 2023

[35] Maxim Zvyagin*, Alexander Brace*, Kyle Hippe*, Yuntian Deng*, Bin Zhang, Cindy Orozco Bohorquez, Austin Clyde, Bharat Kale, Danilo Perez-Rivera, Heng Ma, Carla M. Mann, Michael Irvin, J. Gregory Pauloski, Logan Ward, Valerie Hayot, Murali Emani, Sam Foreman, Zhen Xie, Diangen Lin, Maulik Shukla, Weili Nie, Josh Romero, Christian Dallago, Arash Vahdat, **Chaowei Xiao**, Thomas Gibbs, Ian Foster, James J. Davis, Michael E. Papka, Thomas Brettin, Rick Stevens, Anima Anandkumar, Venkatram Vishwanath, Arvind Ramanathan, GenSLMs: Genome-scale language models reveal SARS-CoV-2 evolutionary dynamics. (ACM Gordon Bell Special Prize)

[34] Manli Shu, Weili Nie, De-An Huang, Zhiding Yu, Tom Goldstein, Anima Anandkumar, **Chaowei Xiao**. *Test-Time Prompt Tuning for Zero-Shot Generalization in Vision-Language Models.* NeurIPS 2022

[33] Boxin Wang, Wei Ping,**Chaowei Xiao**, Peng Xu, Mostofa Patwary, Mohammad Shoeybi, Bo Li, Anima Anandkumar, Bryan Catanzaro. *Exploring the Limits of Domain-Adaptive Training for Detoxifying Large-Scale Language Models* NeurIPS 2022

[32] Yulong Cao, Danfei Xu, Xinshuo Weng, Z. Morley Mao, Anima Anandkuma, **Chaowei Xiao**, Marco Pavone. *Robust Trajectory Prediction against Adversarial Attacks.* CORL 2022 (Oral presentation)

[31] Weili Nie, Brandon Guo, Yujia Huang, **Chaowei Xiao**, Arash Vahdat, Anima Anandkumar. *Diffusion Models for Adversarial Purification.* ICML 2022.

[30] Daquan Zhou, Zhiding Yu, Enze Xie, **Chaowei Xiao**, Anima Anandkumar, Jiashi Feng, Jose M Alvarez. *Understanding the robustness in vision transformers.* ICML 2022

[29] Yulong Cao, **Chaowei Xiao**, Anima Anandkumar, Danfei Xu, Marco Pavone. *AdvDO: Realistic Adversarial Attacks for Trajectory Prediction.* ECCV 2022

[28] Zhuowen Yuan, Fan Wu, Yunhui Long, **Chaowei Xiao**, and Bo Li. *SecretGen: Privacy Recovery on Pre-trained Models*. ECCV 2022

[27] Sina Mohseni, Zhiding Yu, **Chaowei Xiao**, and Jay Yadawa, Haotao Wang and Zhangyang Wang. *Taxonomy of Machine Learning Safety: A Survey and Primer*. ACM Computing Survey 2022. **TL; DR**: a comprehensive survey to discuss the problem and opportunities in machine learning safety.

[26] Xiaojian Ma, Weili Nie, Zhiding Yu, Huaizu Jiang, **Chaowei Xiao**, Yuke Zhu, Song-Chun Zhu, Anima Anandkumar. *RelViT: Concept-guided vision transformer for visual relational reasoning.* ICLR 2022

[25] Jianwie Liu, **Chaowei Xiao**, Kaiyan Cui, Jinsong Han, Xian Xu, Kui Ren. *Behavior Privacy Perserving in RF Sensing.* IEEE Transcations on Dependable and Secure Computing, 2022

[24] Jianwei Liu, Yinghui He, **Chaowei Xiao**, Jinsong Han, Le Cheng, Kui Ren, Physical-World Attack towards WiFi-based Behavior Recognition, IEEE International Conference on Computer Communications (INFOCOM) 2022

[23] Xinlei Pan*, **Chaowei Xiao***, Warren He, Jian Peng, Mingjie Sun, Jinfeng Yi, Mingyan Liu, Bo Li, Dawn Song . *Characterizing Attacks on Deep Reinforcement Learning.* AAMAS 2022

[22] Jiachen Sun, Yulong Cao, Christopher Choy, Zhiding Yu, Anima Anandkumar, Z. Morley Mao, and **Chaowei Xiao**. *Adversarially Robust 3D Point Cloud Recognition Using Self-Supervisions.* NeurIPS 2021

[21] Haotao Wang, **Chaowei Xiao**, Jean Kossaifi, Zhiding Yu, Animashree Anandkumar, and Zhangyang Wang. *AugMax: Adversarial Composition of RandomAugmentations for Robust Training.* NeurIPS 2021

[20] Chen Zhu, Wei Ping, **Chaowei Xiao**, Mohammad Shoeybi, Tom Goldstein, Anima Anandkumar, Bryan Catanzaro. *Efficient Transformers for Language and Vision.* NeurIPS 2021

[19] Mingjie Sun*,**Chaowei Xiao***, Zichao Li*, Haonan Qiu, Mingyan Liu, Bo Li*Can Shape Structure Features Improve Model Robustness under Diverse Adversarial Settings?*. ICCV 2021

[18] Aria Rezaei, **Chaowei Xiao**, Bo Li, Jie Gao. *Application-driven Privacy-preserving Data Publishing with Correlated Attributes.* EWSN 2021 (Best Paper Award)

[17] Yulong Cao*, Ningfei Wang*, **Chaowei Xiao***, Dawei Yang*, Jin Fang, Ruigang Yang, Alfred Chen, Mingyan Liu, Bo Li. *Invisible for both Camera and LiDAR: Security of Multi-Sensor Fusion based Perception in Autonomous Driving Under Physical-World Attacks*. IEEE Symposium on Security and Privacy 2021.

[16] Huan Zhang, Hongge Chen, **Chaowei Xiao**, Bo Li, Mingyan Liu, Duane Boning, Cho-Jui Hsieh. *Robust Deep Reinforcement Learning against Adversarial Perturbations on State Observations*. NeurIPS 2020 (Spotlight)

[15] Haonan Qiu*, **Chaowei Xiao***, Lei Yang*, Xinchen Yan, Honglak Lee, Bo Li. *SemanticAdv: Generating Adversarial Examples via Attribute-conditional Image Editing*. ECCV 2020

[14] Huan Zhang, Hongge Chen, **Chaowei Xiao**, Sven Gowal, Robert Stanforth, Bo Li, Duane Boning, Cho-Jui Hsieh. *Towards Stable and Efficient Training of Verifiably Robust Neural Networks*. ICLR 2020

[13] **Chaowei Xiao***, Dawei Yang*, Bo Li, Jia Deng, Mingyan Liu. *Realistic Adversarial Examples in 3D Meshes*. CVPR 2019 (Oral Presentation)

[12] **Chaowei Xiao**, Ruizhi Deng, Bo Li, Taesung Lee, Benjamin Edwards, Jinfeng Yi, Dawn Song, Mingyan Liu, Ian Molloy. *AdvIT: Characterizing Adversarial Frames in Videos Based on Temporal Information*. ICCV 2019

[11] Yulong Cao, **Chaowei Xiao**, Benjamin Cyr, Yimeng Zhou, Won Park, Sara Rampazzi, Qi Alfred Chen, Kevin Fu, Z. Morley Mao. *Adversarial Sensor Attack on LIDAR-based Perception in Autonomous Driving*. CCS 2019

[10] Liang Tong, Bo Li, Chen Hajaj, **Chaowei Xiao**, Ning Zhang, Yevgeniy Vorobeychik. *Improving Robustness of ML Classifiers against Realizable Evasion Attacks Using Conserved Features*. USENIX Security 2019

[9] Kin Sum Liu, **Chaowei Xiao**, Bo Li, Jie Gao. *Performing Co-Membership Attacks Against Deep Generative Models*. ICDM 2019

[8] **Chaowei Xiao**, Ruizhi Deng, Bo Li, Fisher Yu, Mingyan Liu, Dawn Song. *Characterize Adversarial Examples Based on Spatial Consistency Information for Semantic Segmentation*. ECCV 2018

[7] **Chaowei Xiao***, Jun-Yan Zhu*, Bo Li, Warren He, Mingyan Liu, Dawn Song. *Spatially Transformed Adversarial Examples*. ICLR 2018.

[6] **Chaowei Xiao**, Bo Li, Jun-Yan Zhu, Warren He, Mingyan Liu, Dawn Song. *Generating Adversarial Examples with Adversarial Networks*. IJCAI 2018

[5] **Chaowei Xiao**, Armin Sarabi, Yang Liu, Bo Li, Tudor Dumitra, Mingyan Liu. *From Patching Delays to Infection Symptoms: Using Risk Profiles for an Early Discovery of Vulnerabilities Exploited in the Wild*. Usenix Security 2018

[4] Kevin Eykholt*, Ivan Evtimov*, Earlence Fernandes, Bo Li, Amir Rahmati, **Chaowei Xiao**, Atul Prakash, Tadayoshi Kohno, Dawn Song. *Robust Physical-World Attacks on Deep Learning Visual Classification*. CVPR 2018

[3] Chenshu Wu, Zheng Yang, **Chaowei Xiao**. *Automatic Radio Map Adaptation for Indoor Localization using Smartphones*. TMC 2017

[2] Chenshu Wu, Zheng Yang, **Chaowei Xiao**, Chaofan Yang, Yunhao Liu, Mingyan Liu. *Static Power of Mobile Devices: Self-updating Radio Maps for Wireless Indoor Localization*. INFOCOM 2015

[1] Lei Yang, Yekui Chen, Xiangyang Li, **Chaowei Xiao**, Mo Li and Yunhao Liu. *Tagoram: Real-time Tracking of Mobile RFID Tags to High Precision Using COTS Devices*. MobiCom 2014 (Best Paper Award)

Workshop Papers

[4] **Chaowei Xiao***, Zhongzhu Chen*, Kun Jin*, Jiongxiao Wang*, Weili Nie, Mingyan Liu, Anima Anandkumar, Bo Li, Dawn Song. *DensePure: Understanding Diffusion Models towards Adversarial Robustness* . NeurIPS SRML 2022 (Contributed Talk)

[3] Jiachen Sun, Qingzhao Zhang, Bhavya Kailkhura, Zhiding Yu, **Chaowei Xiao**, Z Morley Mao. *Benchmarking robustness of 3d point cloud recognition against common corruptions*. ICML SRML 2021 (Best Paper Award)

[2] Shengchao Liu, Weili Nie, Chengpeng Wang, Jiarui Lu, Zhuoran Qiao, Ling Liu, Jian Tang, Anima Anandkumar, **Chaowei Xiao**. *MoleculeCLIP: Learning Transferable Molecule Multi-Modality Models via Natural Language*. NeurIPS AI4Science

[1] Boyi Li, Zhiding Yu, De-An Huang, Weili Nie, Linxi Fan, **Chaowei Xiao**,Serge Belongie, Kilian Q. Weinberger, Anima Anandkumar. *Weakly-Supervised Referring Image Segmentation with Multimodal Transformers*. NeurIPS InterNLP 2022

Funding (Total: My share $481,660)

Granted Exploring and mitigating unrestricted adversarial examples. Open Philanthropy. Role: Single PI. My share: $200,000.

Granted A Federated Query Optimizer for Privacy-Preserving Analytics and Machine Learning. Department of Homeland Security 2022 CAOE Research Grants for privacy. Total: $874,998 Role: Co-PI. My share: $281,660

Submitted DARPA ECOLE TA1&TA2. DOD-DARPA: Information Innovation Office (IIO). Total:$5,420,547 Role: Co-PI. My share: $600,951

Submitted DARPA Castle TA1. DOD-DARPA: Information Innovation Office (IIO). Total:$10,835,576 Role: Co-PI. My share: $530,425

Submitted DARPA Castle TA3. DOD-DARPA: Information Innovation Office (IIO) Total:$2,136,688 Role: Sub-contractor. My share: $480,352

Submitted Interdisciplinary Systems-based Training in Precision Nutrition (INTERAICT). NIH grant of advanced training in AI for precision nutrition science research (RFA-OD-22-027).

Selected Media Press

2022 FIERCE Electronics.Nvidia, others work to use LLMs to predict Covid variants

2022 HPCWire. Gordon Bell Special Prize Goes to LLM-Based Covid Variant Prediction

2022 Insightsfy.Speaking the Language of the Genome: Gordon Bell Finalist Applies Large Language Models to Predict New COVID Variants

2022 Scientific Computing World. ACM awards researchers for HPC-Based COVID-19 Research

2022 Tencent News. Speaking the Language of the Genome: Gordon Bell Finalist Applies Large Language Models to Predict New COVID Variants

2019 Analytics. Elon Musk Might Be Right. New Research Exposes Vulnerabilities In LiDAR-based Autonomous Vehicle.

2019 Synced. Researchers Fool LiDAR with 3D-Printed Adversarial Objects.

2019 Popular Science. Self-driving cars still have major perception problems

2019 GCN and Conversation. Autonomous vehicles can be fooled to see nonexistent obstacles

2017 Wired. Security News This Week: A Whole New Way to Confuse Self-Driving Cars.

2017 Fortune. Researchers Show How Simple Stickers Could Trick Self-Driving Cars

2017 Nature. Why deep-learning AIs are so easy to fool.

2017 IEEE SPECTRUM. Slight Street Sign Modifications Can Completely Fool Machine Learning Algorithms.

2017 Yahoo News. Researchers demonstrate the limits of driverless car technology.

2017 Telegraph. Graffiti on stop signs could trick driverless cars into driving dangerously.

Selected Industry Discussions & Responses

Triggered 20+ Autonomous Driving (AD) companies such as Tesla, GM, Daimler, Baidu, TuSimple, Aptiv, Hyundai, Volkswagen, Bosch, Lyft, Nuro, Toyota, Hyundai, Kia, and Volvo to start investigating our newly-discovered security vulnerabilities in AD localization and/or perception algorithms

Talks

Tiny Step towards Socially Responsible Machine Learning.

2022/8 Virtual Seminar series on Challenges and Opportunities for Security & Privacy in Machine Learning

2022/4 ICLR SRML opening remarks

Machine Learning in Adversarial Environment and Beyond.

2022/2 AAAI 2022 1st International Workshop on Practical Deep Learning in the Wild

2022/2 AAAI 2022 workshop on Adversarial Machine Learning and Beyond

2021/10 Hong Kong Baptist University

2021/9 University of Science and Technology of China

2021/8 Tsinghua University

2021/8 Zhejiang University

2021/7 ICML SRML opening remarks

Machine Learning in Adversarial Environment.

2020/11 Waterloo ML + Security + Verification Workshop

2020/3 Google Brain

2020/3 Facebook AI Research

2020/3 Nvidia Research

2020/3 Uber ATG Research

2020/3 Amazon AWS

2020/2 Visa Research

2020/2 Ant Finance

2020/2 ByteDance

Machine Learning: the Good, the Bad, and the Ugly.

2019/9 Microsoft Research

2019/3 Amazon Graduate Research Symposium

2019/2 University of Michigan, Ann Arbor

2018/6 Baiduxlab

Adversarial Objects for Lidar-Based Autonomous Driving System.

2019/8 Microsoft Security Workshop

2019/6 CVPR workshop on Adversarial Machine Learning in Real-World Computer Vision Systems

Characterizing Adversarial Frames in Videos Based on Temporal Information.

2018/8 IBM Watson Research Lab

Research Experience

2019 Microsoft Research, Redmond, USA. Research Intern at Deep Learning Group

2018 IBM Watson Research Lab, New York, USA. Research Intern at IBM Research AI group

Teaching & Mentoring Experience

2022-2023 Instructor, CSE 598 Machine Learning Security, Privacy and Fairness, ASU, Fall 2022. Design a new course on current topics in machine learning security, privacy and fairness. Course Evaluation: 4.7/5.0

2021-now Jiongxiao Wang (ASU Ph.D.): machine learning security

2021-now Yijin Yang (ASU Ph.D.): machine learning privacy

2018-now Jiachen Sun (UMich Ph.D.): adversarial examples in 3D vision systems

2018-now Yulong Cao (UMich Ph.D.): adversarial attacks in autonomous driving systems

2021 Manli Shu (UMaryland Ph.D.): zere-shot robustness in foundation models

2021 Boxin Wang (UIUC Ph.D.): large language model detoxification

2021 Chen Zhu (UMaryland Ph.D.): efficient and robust transformer for vision and language

2019 Kaizhao Liang (UIUC B.S, now applying Ph.D.): verifiably RNN Robust Models

2019 Max Wolff (Viewpoint school): attacks on Face verification system.

2016-2018 Ruizhi Deng (UMich B.S, SFU M.S and now SFU Ph.D.): Adversarial attacks in semantic segmentation and audios

2018 Mingjie Sun (Tsinghua B.S, now CMU Ph.D.): graph poisoning attacks

Academic Services

Organizer Workshop on Socially Responsible Machine Learning (Founder) in NeurIPS 2022, ICLR 2022, ICML 2021

Workshop on Neural Architectures: Past, Present and Future in ICCV 2021

Workshop on 1st International Workshop on Adversarial Learning for Multimedia in ACM-MM 2021

Workshop on Adversarial Machine Learning in Real-World Computer Vision System in CVPR 2019

PC member CVPR, ICCV, ECCV, ICLR, ICML, NeurIPS, CCS

Area Chair AAAI, CVPR

Panelist NSF 2023, NSERC 2021