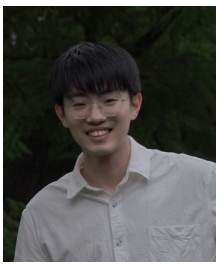
Mengnan Zhao

Citizenship: China Phone: +86 15822843520

dlutzmn@mail.dlut.edu.cn https://dlut-lab-zmn.github.io/



Summary

- I am a Ph.D. candidate at Dalian University of Technology (DUT), advised by Prof. Baocai Yin and Prof. Lihe Zhang. I obtained my master's degree from DUT under the supervision of Prof. Bo Wang. My current research focuses on AI security, specifically on LLM security, adversarial attacks and defenses, scene graph generation, data inference attacks, and watermarking forensics.

Education

2021.09 – Present **Ph.D. candidate**, Dalian University of Technology
Thesis title: *Privacy Protection in Image-Text Transformation Tasks*.

2018.09 – 2021.06 **M.Sc.**, Dalian University of Technology
Thesis title: *Research on attack and defense of adversarial examples*.

2014.09 – 2018.06 **B.S.**, Tianjin University of Technology

Experience

2021.03 – 2021.06 **Intern Engineer**, Institute of Automation, Chinese Academy of Sciences

Published Papers

Journal.

1. **M. n. Zhao**, L. h. Zhang, Y. q. Kong, and B. c. Yin, "Adversarial attacks on scene graph generation," *IEEE Transactions on Information Forensics and Security*,
2. **M. n. Zhao**, L. h. Zhang, Y. q. Kong, and B. c. Yin, "Class correlation correction for unbised scene graph generation," *Pattern Recognition*,
3. **M. n. Zhao**, B. Wang, W. k. Guo, and W. Wang, "Protecting by attacking: A personal information protecting method with cross-modal adversarial examples," *Neurocomputing*, vol. 551, p. 126 481, 2023.
4. **M. n. Zhao**, L. h. Zhang, Y. q. Kong, and B. c. Yin, "Temporal knowledge graph reasoning triggered by memories," *Applied Intelligence*, 2023.
5. **M. n. Zhao**, B. Wang, W. Wang, Y. q. Kong, T. h. Zheng, and K. Ren, "Guided erasable adversarial attack (geaa) toward shared data protection," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 2468–2482, 2022.
6. B. Wang (Tutor), **M. n. Zhao**, W. Wang, X. r. Dai, Y. Li, and Y. q. Guo, "Adversarial analysis for source camera identification," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 11, pp. 4174–4186, 2021.
7. B. Wang (Tutor), **M. n. Zhao**, W. Wang, F. Wei, Z. Qin, and K. Ren, "Are you confident that you have successfully generated adversarial examples?" *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 6, pp. 2089–2099, 2020.

8. **M. n. Zhao**, B. Wang, F. Wei, M. n. Zhu, and X. Sui, "Source camera identification based on coupling coding and adaptive filter," *IEEE Access*, vol. 8, pp. 54 431–54 440, 2019.

Conference.

1. **M. n. Zhao**, L. h. Zhang, Y. q. Kong, and B. c. Yin, "Fast adversarial training with smooth convergence," in *Proceedings of the International Conference on Computer Vision (ICCV)*, 2023.
2. **M. n. Zhao**, X. r. Dai, B. Wang, F. Yu, and F. Wei, "Further understanding towards sparsity adversarial attacks," in *International Conference on Artificial Intelligence and Security*, Springer, 2022, pp. 200–212.

Preprints

Under Review in CCF-A Conf.

1. **M. n. Zhao**, L. h. Zhang, Y. q. Kong, and B. c. Yin, *Advprompte: Enhancing the reliability of erasing concept*.
2. **M. n. Zhao**, L. h. Zhang, Y. q. Kong, and B. c. Yin, *Catastrophic overfitting: A potential blessing in disguise*.
3. **M. n. Zhao**, L. h. Zhang, Y. q. Kong, and B. c. Yin, *Eipformer: Emphasizing instance positions in 3d instance segmentation*.
4. **M. n. Zhao**, L. h. Zhang, Y. q. Kong, and B. c. Yin, *Separable multi-concept erasure from diffusion models*.

Authorized Patents

CN201911050075	**M. n. Zhao**, B. Wang. A Point Attack Method Based on Weighted Spectrum Generation of Adversarial Samples
CN201910871685.X	**M. n. Zhao**, B. Wang. An Image Source Identification Method Based on Adaptive Filtering and Coupled Encoding.
CN112381149A	**M. n. Zhao**, B. Wang. A Reasonable Adversarial Analysis Method for Source Camera Recognition Based on Deep Learning.
CN201911050099.5	B. Wang, **M. n. Zhao**. A Defense Method Based on Deceptive Attackers to Adversarial Examples.
CN113821770A	**M. n. Zhao**, B. Wang. A Targeted Adversarial Poisoning Attack Method for Shared Data Protection.

Professional Activities

Reviewer for

- IEEE Transactions on Neural Networks and Learning Systems
- IEEE Transactions on Multimedia
- IEEE Transactions on Circuits and Systems for Video Technology
- European Conference on Computer Vision
- IEEE Transactions on Network Science and Engineering
- Journal of King Saud University-Computer and Information Sciences
- International Journal of Machine Learning and Cybernetics