Ge YAN

Mobile: +1(858)531-8601 | Email: geyan@ucsd.edu | San Diego, CA | Website: https://windymount.github.io/

RESEARCH INTEREST

My research interest is mainly in **trustworthy machine learning** and **responsible AI**. My goal is to make current machine learning models **more robust and interpretable** and provide a quantification of **uncertainty**.

EDUCATION

Department of Computer Science and Engineering, UC San Diego Ph.D. student advised by Prof. Tsui-wei Weng.	Current
Department of Electrical and Computer Engineering, UC San Diego M.S. in Machine Learning and Data Science (GPA 3.95/4)	Mar 2023
School of Mathematical Sciences, Peking UniversityB.S. in Information and Computing Science (GPA 3.484/4)	Jun 2021

PUBLICATIONS

Provably Robust Conformal Prediction with Improved Efficiency

Ge Yan, Yaniv Romano, Tsui-wei Weng, ICLR 2024.

VLG-CBM: Training Concept Bottleneck Models with Vision-Language Guidance

Divyansh Srivastava*, **Ge Yan***, Tsui-wei Weng, NeurIPS 2024. * Equal contribution

SKILLS

Programming Languages: Python, C. **Proficiency** in Python (**8**+ **years** of experience). **Solid mathematical foundation:** linear algebra, probability, mathematical/real analysis. **Rich experience** in conducting deep learning experiments with Pytorch.

RESEARCH EXPERIENCE

Provably robust conformal prediction with improved efficiency

Advisor: Professor Tsui-wei Weng, UC San Diego

- Studied robust conformal prediction which generates prediction sets that are robust against adversarial examples.
- Provided a theoretical analysis and proposed two methods based on theoretical insight.
- Successfully reduced the inefficiency of the current baseline by up to 48.80% on ImageNet.

INTERNSHIP

Applied Scientist Intern, Amazon, San Diego

- Developed machine-generated text detector with style embeddings.
- Enhanced model robustness to adversarial inputs and accelerated previous model.
- The model is adopted into internal tool as a light-weight and robust detection model.

Data Scientist Intern, DiDi Technology, Beijing.

- Developed and maintained internal causal inference tools to provide guidance for marketing decisions.
- Implemented Bayesian Structural Time Series (BSTS) algorithm for counterfactual prediction.
- The tool is used averagely 100+ per month by internal data analyst teams.

Jun 2023 – Aug. 2023

Jun 2024 - Aug. 2024