

Yiqiu Shen

Ph.D., Statistics

Department of Data Sciences and Operations
University of Southern California
Los Angeles, CA
✉ yiqiushen2025@gmail.com
🌐 <https://yiqiushen.github.io>
in <https://www.linkedin.com/in/yiqiu-shen/>

Education

- Aug 2019 – July 2025 **Ph.D. in Statistics**, *University of Southern California*, Los Angeles, CA
GPA: 3.96
Field of Interest: High Dimensional and Robust Statistics, Statistical Learning Theory
Advisor: Dr. Stanislav Minsker
- Aug 2016 – May 2019 **B.S. in Applied and Computational Mathematics**, *University of Southern California*, Los Angeles, CA
GPA: 3.92, *summa cum laude*

Professional Experience

- May 2023 – Aug 2023 **Data Scientist Intern**, *Delphire Inc.*, Los Angeles, CA
- Utilized **Pandas**, **Numpy**, and **Seaborn** to model and visualize wildfire impact on properties using historical data (1990–2020).
 - Created **spatial indexes** using **PostGIS** to integrate Microsoft Building Footprints and Zillow housing data and estimated monetary damage, bringing the query time down to under **1s** for a dataset of **11 mil geoshapes**.
 - Operated **FlamMap** and **WindNinja** to simulate wildfire perimeters and assess damage reduction using Sentinel units.

Selected Projects

- 2024 **Ultrasound Image Classification Project**
- Designed a deep learning pipeline using **PyTorch** to classify and segment breast ultrasound images.
 - Fine-tuned **ResNet** and **CLIP** models with linear probing and compared against custom CNNs, achieved **83% accuracy**.
- 2020 **Membership Identification in Large Scaled Sparse Networks**
- Implemented SIMPLE in R and Python to analyze S&P1000 company networks.
 - Improved computation efficiency of both experiment and simulation and reduced running time by 90% using **parallel processing**.
- 2020 **High Dimensional Classification via Spiked Eigenvalue Theory**
- Developed and coded a novel classification method using spiked eigenvalue theory to classify viruses and bacteria from codon data.
- 2020 **Pairwise Relationship Identification via Mixed Neural Networks**
- Built a **CNN-RNN hybrid model** in Tensorflow to identify virus-host relationships.
 - Achieved state-of-the-art **87% accuracy** on test sets.

Publications

- 2024 **The Impact of Contamination and Correlated Design on the Lasso**
Authors: Minsker, Stanislav, **Shen, Yiqiu**
Probability and Statistics Letters. To appear
- 2021 **Minimax Supervised Clustering in the Anisotropic Gaussian Mixture Model**
Authors: Minsker, Stanislav, Mohamed Ndaoud, **Shen, Yiqiu**
Submitted to Journal of Machine Learning Research. Minor Revision
- 2024 **Concentration and Moment Inequalities for Heavy-Tailed Random Matrices**
Authors: Jirak, Moritz, Stanislav Minsker, **Shen, Yiqiu**, and Martin Wahl
Submitted to Probability and Related Fields. Minor Revision

Skills

- Languages Python, R, MATLAB, SQL (PostgreSQL), Java
- ML PyTorch, TensorFlow, Keras
- Frameworks
- Software \LaTeX , Microsoft Excel, SAS JMP, Mathematica, Tableau, PostgreSQL, PostGIS, FlamMap
- Languages English (Proficient), Japanese (JLPT N2), Chinese (Native)
- Other Skills Webscraping (BeautifulSoup), Teaching, Technical Communication, AI-assisted Front-end Engineering (Bolt.new), Git