Alan Wang

Open to relocate | 858-539-5564 | zew013@ucsd.edu
Personal Website: zew013.github.io
Linkedin: linkedin.com/in/alanwang-ucsd
GitHub: github.com/zew013

Education (2020-2025 Continuing MSCS at UCSD)

The University of California San Diego | GPA:3.985/4.0 | Major GPA: 4.0/4.0

B.S. | Primary Major - Data Science | Double Major - Joint Major in Math and Economics

Data Science Courses: Data Structures & Algorithms, Web & Data Mining, Visualization, Machine Learning, Deep Learning, Computer Vision, Image Processing, Recommender System, Network Science, Graph Theory, Database, Scalable Data Systems & Analytics, Spatial Data Science.

Math & Econ Courses: Calculus, Linear Algebra, Numerical Analysis, Optimization, Probability, Statistics, Micro & Macroeconomics, Econometrics (Hypothesis Testing, Regressions, Time Series, etc.), Operations Research.

Research Experience

DS-ML Capstone | Prof. Alex Cloninger & Prof. Rayan Saab | Oct. 2023 – Present

From Theory to Pixels: Diffusion Models for Image and Data Generation | *GitHub, Webpage*

- Reproduced the Denoising Diffusion Probabilistic Models and the Denoising Diffusion Implicit Models.
- Using probing classifiers to elucidate whether/how the Latent Diffusion Model scene representations such as depth are encoded in the diffusion process.

DS-ML Research Assistant | Rappel Laboratory | Feb. 2023 – Present

Image Segmentation and Propagation Analysis Pipeline for cAMP Waves in Cell Aggregation Stage | *GitHub*

- Experimented with computer vision methods and deep learning models to segment videos of cAMP waves.
- Developed a two-stage Python workflow. First, utilized image and signal processing techniques to segment videos and extract the cell signals. Then, creatively grouped waves using the unsupervised clustering algorithm DBSCAN and calculated velocity vector fields by performing least squares on activation maps through 3-by-3 kernels.
- Successfully presented my work during lab meetings. Optimized and parallelized the code, reducing average processing time from 50 minutes to 4 minutes. Trained other lab members to use my program.

DS Research Assistant | Prof. Richard Carson & Prof. Dale Squires | Dec. 2021 – Dec. 2022

Data-Driven Analysis of Ethical Preferences in UN Membership Policies

- Conducted an extensive review of United Nations documents focusing on membership assessment policies.
- Developed an ETL pipeline using Python and AWS to prepare assessment data spanning 70 years from the UN Digital Library. Sharply improved processing speed and accuracy, especially for handwritten records dating back to the 1950s.
- Applied a Fixed Effect OLS on the processed panel dataset, and deduced the ethical preferences which could serve as a standard to guide and facilitate multilateral cooperation by reducing conflicts and information costs.

Variations in Conditional Logit Assumptions via Monte Carlo Simulation

- Designed experiments to evaluate the properties of Conditional Logit under different assumptions by performing Monte Carlo simulation and controlling the error term under Minimum and Maximum Gumbel Distributions.
- Coded an automated program to analyze statistical metrics and create academic graphs in Stata.

Professional & Teaching Experience

Data Science Intern | Grant Street Group | May. 2023 – Aug. 2023

- Proposed an initiative for the ML-powered transaction monitoring system. Experimented with multiple ML/DL models (Random Forest, ARIMA, Prophet, Isolation Forest, Temporal Fusion Transformer, etc.) to build an unsupervised anomaly detection solution.
- Led a dynamic team of four and orchestrated our different skillsets. Our new system utilized the Prophet time series model and was tested in the Exasol database, enhancing fraud and anomaly detection by 40% and 70% respectively compared with the previous static threshold-based system.
- Presented our work in poster presentations and presented during Summer Company Meeting.

Instructional Assistant | Halıcıoğlu Data Science Institute | Mar. 2023 – Present

Theoretical Foundations of DS – Prof. Janine Tiefenbruck, The Practice and Application of DS – Prof. Sam Lau

- Automated the grading process by developing test cases and grading systems on Python and Jupyter Notebook.
- Assisted over 400 students by conducting office hours, leading labs and discussions, answering questions on the forum, as well as creating and grading assignments and exams.

CSE-PACE Program Designer | UC San Diego CSE Department | May. 2022 – Sep. 2022

NSF-supported Project: Inclusive longitudinal peer mentoring for community building and retention. | *Webpage*

- Addressed issues that disproportionately affect students from historically marginalized groups by crafting courses that prioritized communication and individual-peer relationships over sheer knowledge acquisition.
- Designed 7 courses in the CSE Cohort Program by drawing topics in computing from blogs, news articles, and videos.
- The program has been officially integrated as part of the CSE curriculum by the end of Summer 2022.

Selected Leadership & Project Experience

Data Analyst, Tech VP | Lumnus Consulting (Student Enterprise) | Nov. 2021 – Feb. 2023

Webpage

- Built machine learning models and created visualizations for various team projects.
- Built connections with local businesses. Invited alumni in industry and upperclassmen majoring in related majors on a regular basis to talk about their academic and career experiences. Organized events such as data analysis projects, intern presentation, and case-study competitions throughout the quarter.
- Developed and maintained our website using React.js and Heroku.

Deep Learning Projects

eBay 2023 ML Competition | May. 2023 – June. 2023

- Achieved a weighted F1 score of 0.86 on the Named Entity Recognition task by finetuning on a pretrained DeBERTa model. Gained a deeper understanding of text data processing, tokenization, and BERT downstream tasks.

Amazon Massive Intent Dataset Classification with BERT & Web Deployment | Dec. 2022 – Jan. 2023 | GitHub, Webpage

- Used BERT as encoder and a Neural Network as the decoder to classify text intentions.
- Self-studied Knowledge Distillation. Used the trained BERT-NN as the teacher model and BiLSTM as the student model to compress the model size from over 439MB to 70MB while maintaining a similar accuracy level.
- Deployed the compressed model on my own webpage using Flask-RESTful, Amazon S3, and Heroku.

CNNs and LSTMs with PyTorch for Image Captioning on COCO Dataset | Nov. 2022 – Dec. 2022 | GitHub, Report

Neural Network from Scratch | Sep. 2022 – Oct. 2022 | GitHub, Report

- Implemented a neural network in Python and coded backpropagation, mini-batch gradient descent, and cross-validation using NumPy from scratch. Added early stopping, momentum, and L1 & L2 regularization to enhance the model. Conducted experiments with sigmoid, tanh, ReLU, and softmax as activation functions.

Data Analysis Projects

Analysis of Power Outage Status in the Continental U.S.

- Went through the full process of questioning, data gathering, data mining, explorative data analysis, missingness assessment, hypothesis test, baseline model, final scikit-learn ML pipelines, fairness analysis, and visualization.

Professional Skills

Programming:

Python, R, Stata, SQL, Java, HTML, JavaScript, CSS, LaTeX, MATLAB

Libraries & Frameworks:

NumPy, SciPy, Pandas, sklearn, Matplotlib, Plotly, PyTorch, OpenCV, bs4, D3.js, Dask, GeoPandas, Spark, REST API, etc.

Technology:

Microsoft Office Suite, Git & GitHub, AWS, Basic Shell scripting, Tableau, Heroku, ImageJ, Exasol, Jupyter.