Shuyun **TANG**

928 San Benito Rd, Berkeley, CA, 94707

(+1) 904-325-5896 | shuyuntang@Berkeley.edu | 44Shu | shuyuntang

Natural Languages: *English, Mandarin Chinese, Japanese*
Programing Languages: *Python, R, C++, Java, Julia, SAS.*
ML Libraries: *PyTorch, Keras, Tensorflow, Scikit-Learn*

Education

University of California, Berkeley

Master of Art in Statistics — *Aug 2021 - Jun. 2022 (Expected)*

- Incoming graduate student.
- Courses: Advanced Probability and Statistics, Software Engineering, Optimization, Reinforcement Learning, Graph Theory.

University of California, Santa Barbara

Bachelor of Science in Statistics and Data Science — *Aug. 2019 - May 2021*

- GPA: 3.95/4.0
- Graduated with the highest honor (top 1% students).
- Courses: Vector Calculus, ODE, Linear Algebra, Numerical Analysis, Regression Analysis, Object oriented programming, Data Structures, Machine Learning (graduate level), Natural Language Processing, Computational Vision.

Research Experience

Multi-modal Emotion Recognition with Graph Neural Networks — *Dr. Zhaojie Luo*

Frontier Intelligent System Research Laboratory, Osaka University, working paper (first author) — *Jan. 2021 - Now*

- Applied a novel hierarchical multi-modal feature fusion to the graph level.
- Used attention-based encoder with Graph Convolutional Networks. Trained using Valence-Arousal learning pipeline.
- Achieved new state-of-the-art results in the following data: IEMOCAP, MELD, AVEC. Library used: NetworkX, Pytorch Geometric.
- Supervised and granted funding from Prof. Hiroshi Ishiguro.

Joint Fovea and Optic Disc Segmentation for Degenerated Retina — *Prof. Michael Beyeler*

Bionic Vision Lab, UCSB, Paper submitted to OMIA 2021 (first author) — *November 2020 - July 2021*

- Proposed a new U-Net based attention model with multiple local bottleneck structures.
- Conducted extensive experiments and ablation studies to prove it in segmenting the landmarks in disease fundus images (Age-Related Macular Degeneration, Glaucoma), which are usually hard to segment with traditional methods.
- Achieved new state-of-the-art results in the following data: REFUGEE, Baidu ADMD, Messidor.

Joint Modeling of EEG, fMRI and Structural MRI — *Prof. Ambuj K. Singh*

Dynamo Lab, UCSB, Intend to submit to AAAI 2021 (co-fisrt author) — *Mar. 2021 - Now*

- Combined high temporal resolution EEG and high spacial resolution fMRI for better brain activity modeling.
- Used spatial-temporal graph neural networks with adaptive adjacency matrix to solve the time series prediction.
- Interpreted the signal source localization among various brain tasks.

Academic Projects & Experience

Kaggle Competition Expert

Ranked top 500, out of 160,000. Profile can be found ***here***. Open-sourced repo can be found ***here*** — *June. 2020 - Now*

- **Cornell Bird Identification, Silver Medal**: Bird audio prediction, used ResNexts with Librosa/SED, and post-process.
- **Google Landmark Recognition, Silver medal**: Landmark recognition over 2 million images, used ensembled EfficientNetB5-B7 with refined GeM pooling, trained with custom learning pipeline.
- **Halite By Two Sigma, Silver medal:** AI game agent development based on Reinforcement Learning (DQN) and heuristic algorithm.
- **Global Wheat Detection, Bronze medal:** Wheat ears detection and feature extraction based on YoloV3 and ResNets with pseudo labeling and stacking.

Real Time Machine Learning Based Chat Client

UCSB Data Science Fellowship Capstone Project. Codes and report paper can be found ***here*** *Jan. 2020 - May 2021*

- This year-long project was in partnership with Invoca, an NLP-based company that creates conversation intelligence platforms. We built a real-time chat client to react to customers' text, classify to sub-categories, and recommend useful links.
- Responsible to build text cleaning pipelines, train the traditional ML models and transformer models. This made extensive use of PyTorch (for building models) and Scikit-Learn (for text manipulating).
- Assisted to build database and retrieve text input using MongoDB.

SSH Port 22 Hacker Passwords Pattern

UCSB Data Science Club Project. Codes can be found ***here*** *Sep. 2019 - Mar. 2020*

- Organized the team's workflow, used K-mean clustering, PCA, Hamming distance, and custom data preprocess pipeline to analyze more than 40,000 password string patterns. Provided useful insights and interpretation about some vulnerable password patterns.
- Selected to present in the Annual UCSB Data Science Showcase.
- Worked as the project manager and mentor in the club.

Fake Job Posting Prediction and Analysis with Machine Learning

UCSB PSTAT Department Undergraduate Research Project. Codes and report paper can be found ***here*** *Dec. 2020 - Mar. 2021*

- Developed and explored data-driven strategies to identify fraudulent job posts and advertisements. Mentor: Prof. Trevor Ruiz.
- Experimented a variety of ML models including Regression models, Tree models, LSTM, and BERT with statistical sampling methods.
- Data used: University of the Aegean Public 18k Job Posts data.

Undergraduate Learning Assistant

UCSB PSTAT Department *Sep. 2020 - Jan. 2021*

- Assisted Dr. Dawn Holmes in FALL 2020: PSTAT10 - Principles of Data Science.
- Responsible for grading, preparing curriculum, the teaching of open labs, and office hours. Topics include data analysis in R and MySQL.

Publications

- **Tang, S.**, Qi, Z., Granley, J., Beyeler, M., *U-Net with Hierarchical Bottleneck Attention for Landmark Detection in Fundus Images of the Degenerated Retina.* Under submission. [Paper Link]

Honors & Activities

2019-2021 Deans Honor Awardee. *UCSB*

2019-2021 L&S Honor Student. *UCSB*

2019 National Honor Society North FL Chapter Scholarship. *NHS*

2018-2020 Volunteer in Disaster Action Team, North FL Chapter. *Red Cross*

2010-Now Competitive Yo-Yo player, world rank 8th, national rank 3rd, state champion. *USA*

2017-Now Established Yo-Yo brand Original Throw to support and sponsor players to innovate techniques. *China*