Joonyoung (Aaron) Bae

Email: joonyoungbae.aaron@gmail.com | **Github:** github.com/n99joon | **[Google Scholar]**

Education

University of Southern California (USC) Los Angeles, CA

Master of Science in Computer Science (General Track) (GPA: 4.0 / 4.0) *Jan 2024 – Dec 2025*

- Thesis : Clustering and Structural Analysis of High Dimensional Data and Manifold | Advisor: Jiapeng Zhang

University of Hong Kong (HKU) Hong Kong, HK

Bachelor of Engineering (Computer Engineering & Minor in Mathematics) (GPA: 3.44 / 4.3) *Sep 2017 – Jun 2023*

Publications

CoreSPECT: Enhancing Clustering Algorithms via an Interplay of Density and Geometry

Chandra Sekhar Mukherjee, Joonyoung Bae*, Jiapeng Zhang (* Equal Contribution)* *Arxiv (2025)*

- Designed a plug-and-play clustering enhancement framework, based on density-geomtery correlation observations, boosting existing algorithms (e.g. Kmeans & HDBSCAN) to near-SOTA performance with ∼50x faster runtime
- Developed the initial model assumptions and established theoretical guarantees for our framework

Concepts and Models of Environment of Self-Adaptive Systems:A Systematic Literature Review

Yong-Jun Shin, Joon-Young Bae, Doo-Hwan Bae *APSEC (2021)*

Teaching and Research Experience

Graduate Research Assistant USC

Advisor: Jiapeng Zhang *Aug 2024 – Present*

- Focused on improving clustering and visualization methods mainly for single-cell and bulk RNA sequencing biology datasets. Main direction is to observe, formularize, and leverage latent intrinsic structures of the data.
- Currently working on a new UMAP-based visualization method for biology datasets that enhances robustness across runs and enables hierarchical display of datapoints sorted by our novel centrality/clusterability measure

Course Grader of CSCI567: Machine Learning USC

Instructor: Haipeng Luo *Jan 2025 – Jun 2025*

- Ranked 1st out of 250 (Fall 2024) and graded on the topics of Supervised/Unsupervised Learning, SVM, Kernel Methods, Neural Networks, Reinforcement Learning, Multi-armed Bandits, Learning in Games, etc

Full-Time Research Assistant HKU

Advisor: Eric Schuldenfrei, Faculty of Architecture *Jun 2023 – Sep 2023*

- Developed 3D-Multilateration program using Gradient Descent to locate a device only using noisy signal strength
- Implemented object detection on a microprocessor, ESP32-CAM, with limited computation power using decision tree based algorithms (XGBoost and RandomForest) in C++, for smart building applications

Undergraduate Research Assistant KAIST

Advisor: Doo-Hwan Bae *Jul 2020 – Sep 2020*

- Mainly conducted a Systematic Literature Review on Self-Adaptive System and its environment

Honors and Awards

Credit Suisse Global Coding Challenge 2021 (7th in World, 2nd in Asia) 2021

- Algorithm problems scored on accuracy and runtime among $+20,000$ participants (Coded in C & C++)

Republic of Korea Army Torchlight Award 2020

- Top distinction in performance in my cohort of Cybersecurity Division (∼20 selected individuals)

HKU Worldwide Undergraduate Student Exchange Scholarship 2019-2020

HKU Dean's Honours List 2017-2018, 2019-2020

Experience

USC CS Theory Group USC

Member *Jan 2024 – Present*

- Presented "Clustering and Structural Analysis of High Dimensional Data and Manifold" at the weekly gathering
- Attended weekly talks and gained insights on wide variety of theoretical computer science topics

Cybersecurity Division Republic of Korea Army

Mandatory Military Service *Oct 2020 – Apr 2022*

- Selected after interviews & exams and executed cyber-emergency response, real-time monitoring, packet analysis, and general maintenance of the Korean Army network systems, using tools such as SIEM, Firewall, and IPS

Undergraduate Summer Internship Hong Kong

Big Data Architect Ltd. *Jul 2019 – Aug 2019*

- Refactored legacy SQL code in traditional database systems and migrated them to new data warehouse structures

Projects

cplearn | *Python Package (published on PyPI)* Jun 2025 – Present

- Co-authored a novel clustering and visualization python toolkit for data with core-periphery-like structures
- **CoreSPECT:** a clustering enhancement framework that identifies most-to-least separable layers in the data regarding clusterability and clusters using the novel two-fold method, details shown in the aforementioned paper
- **CoreMAP:** a visualization tool that enables hierarchical display of most-to-least separable points by incorporating a novel anchoring idea and clustering results on the core nodes to the attraction-repulsion dynamics of UMAP

Co-occurrence Embedding based Knowledge Base Compression in RAG Sep 2025 – Present

- Developing Retrieval Augmented Generation(RAG) with memory efficient document retrieval by Word2Vec-like embedding of documents and human-feedback guided multi-document summarization, without loss of accuracy

Regularization on Various Stages of Convolutional Neural Network(CNN) Jan 2024 – Jun 2024

- Demonstrated that adding various regularization methods on input, internal layers, and labels on CNN can boost performance of 450k parameters baseline model to achieve similar accuracy as AlexNet with 15x more parameters.

Quantitative Finance Projects Dec 2023

- Worked on improving arbitrage cycle detection using variations of Bellman-Ford algorithm
- Applied offline reinforcement learning to predict optimal threshold for stop-loss orders given historical signal data

Age and Gender Prediction Jun 2022 – Jul 2022

- Performed Data Analysis, Data Preprocessing (Feature Engineering, PCA, Mutual Information Score), Modeling (XGBoost, Regression, etc), and Performance Analysis (Feature Importances, Ensemble/Stacking) of a model predicting gender and age from health data

Certifications

Cisco Certified Network Associate Routing and Switching (CCNA R&S) 2020

- Passed the exam on network and IP fundamentals, security fundamentals, and automation and programmability

Audited Coursework

Below is the list of audited courses that I fully attended at USC, but not listed in the transcript

MATH435: Introduction to Differential Geometry

MATH505B: Applied Probability

EE546: Mathematics of High-dimensional Data

EE588: Optimization for the Information and Data Sciences