

JAE HONG KIM

1200 E California Blvd, MC 356-48, Pasadena, CA 91125 || jaehongkim.xyz || jkim9@caltech.edu

Education

California Institute of Technology, Pasadena, CA

2013 - 2019 (anticipated)

Ph.D. in Physics with Subject Minor in Computer Science. GPA 3.

Stony Brook University, Stony Brook, NY

2009 - 2013

B.S with Honors in Physics and Mathematics (double major). GPA 3.94/4.00

Work Experience

Second Spectrum, Los Angeles, CA

Summer 2017

Data / Software Engineer Intern

Projects

Selected Works for Doctoral Thesis:

Database Adapter

- Traditionally, high energy physicists store data that they customize for a particular analysis across many files that are susceptible to misplacement and corruption. I recognized that RDMS provide more robust solutions and created an automated way to convert existing files and insert them into tables in a PostgreSQL database.
- The PostgreSQL adapter written in C++ allows any user to insert their files into the database by passing the list of columns and files to a python wrapper. The final version of the package had runtime improvements over the first working version of more than 100x by choosing a more suitable data structure and optimizing Postgres bulk insertion performance.

Unsupervised Domain Adaptation

- The methods of most measurements in particle physics are developed by using simulated data. The analysis procedures are fixed prior to being applied to the detector data.
- The differences between the simulation and the real data can affect the estimated quantities. This is typically quantified as an addition source of certainty (systematic uncertainty).
- I implemented unsupervised domain adaptation algorithms such as DANN and PixelDA using Tensorflow.
- Learned a common representation of the two datasets while minimizing the loss in classification performance (4% lower).

Selected Works from the Internship:

Measuring Physical Load on NBA Players

- Using player location data extracted by computer vision, we can estimate the physical toll on a player.
- Trainers and coaches can use this information to determine if a player is fatigued or over-practiced, allowing players to be in their peak physical shape for games.
- The load is extracted from the time-series data is compared with the baseline value from the limited data from accelerometers attached to players. The agreement between the two values were within 10%.

Programming Languages and Tools

- Python, SQL, git, and C++.
- NumPY, Scikit-learn, and Tensorflow.