Peter Pao-Huang

+1 608 622 2672 | peter.paohuang@gmail.com | LinkedIn | GitHub |

Education

University of Illinois at Urbana-Champaign — Champaign, Illinois
B.S. in Computer Science; ***GPA: 3.96/4.00*** — *Sep 2020 – May 2024*

Relevant coursework: Discrete Mathematics, Data Structures, Computer Architecture, System Programming, Numerical Methods, Probability & Statistics, Algorithms, Machine Learning, Programming Languages & Compilers, Real Analysis, Molecular Dynamics, Probability Theory, Optimization

Publications

Pao-Huang, P., Thrush, K., Montemayor, D., Levine, M., 2023. Unpaired Single Cell Dataset Alignment with Wavelet Optimal Transport. *In Submission.*

Pao-Huang, P., Jing, B., Berger, B., 2023. Scalable Multimer Structure Prediction using Diffusion Models. *NeurIPS 2023 Generative AI and Biology Workshop.*

Jing, B., Erives, E., **Pao-Huang, P.**, Corso, G., Berger, B. and Jaakkola, T., 2023. EigenFold: Generative Protein Structure Prediction with Diffusion Models. *ICLR MLDD.*

Allabadi, G., Lucic, A., **Pao-Huang, P.**, Wang, Y. X., & Adve, V. (2023). Semi-Supervised Object Detection in the Open World. *arXiv preprint arXiv:2307.15710*

Zhao, Y., Sharif, H., **Pao-Huang, P.**, Shah, V., Sivakumar, A. N., Valverde Gasparino, M., ... & Adve, V. (2023). ApproxCaliper: A Programmable Framework for Application-aware Neural Network Optimization. *Proceedings of Machine Learning and Systems, 5*

Research Experience

Altos Labs — San Diego, California, United States
Research Intern, Advised by Professor Morgan Levine — *Jun 2023 – Oct 2023*

- Worked with the Morgan Levine group on generative modeling efforts for single-cell omics and imaging in cellular rejuvenation.
- Created a novel optimal transport method ("Wavelet Optimal Transport") for unpaired modality alignment using spectral graph wavelets.

Computation & Biology Group (MIT CSAIL) — Cambridge, Massachusetts
Research Assistant, Advised by Professor Bonnie Berger — *Oct 2022 – Jun 2023*

- Researched generative modeling and physics-informed deep learning applied to drug discovery.
- Developed a novel diffusion-based generative model for protein conformation sampling and folding.
- Created a new diffusion SE(3)-equivariant model to predict large multimeric structures that is more scalable than previous methods.

LLVM Research Group — Urbana, Champaign
Research Assistant, Advised by Professor Vikram Adve — *Sept 2020 – Sept 2022*

- Co-authored paper published in MLSys for neural network optimization (structured pruning, low-rank factorization, and quantization) and approximate computing on edge computing devices.
- Developed PyTorch-based federated semi-supervised learning framework for distributed, computer vision-based systems.

Work Experience

Two Sigma Investments — New York, New York
Software Engineering Intern — *Jun 2022 – Aug 2022*

- Integrated Bayesian statistical modeling to critical classes of forecast models.
- Built framework for evaluating probabilistic programming languages (PPLs) like PyMC and Stan, used by quantitative researchers to select the optimal language to express Bayesian models.

Avant Chicago, Illinois

Machine Learning Intern *May 2021 – Aug 2021*

- Developed a Python-based machine learning library utilizing distributed computing to enable model training and scoring on large-scale datasets.
- Deployed library through Docker on production Kubernetes clusters used for in-house model training and evaluation tasks.
- Created parallelized versions for Scikit-learn models, transforms, & hyperparameter optimization

Skills

Programming Languages: Python, Swift, Java, C++, C, Javascript, Verilog, Stan

Libraries & Frameworks: Flask, Tensorflow, PyTorch, Keras, CUDA, SQL, GIT, HTML, CSS, AWS, Docker, Digital Ocean, ONNX, Kubernetes, Dask, Stan, PyMC

Projects

Epius.ai: Platform for Automated Machine Learning

- Created an automated machine learning platform targeting small businesses that can utilize the technology in daily business operations.
- Developed evolutionary algorithm that produces an optimal machine learning pipeline (algorithm, hyperparameters) for the client's datasets made available to frontend through a REST API.

TAPIR: Thermomechanical Advanced Polymer Informatics & Resource

- Developed object-oriented Python framework for integrating quantum mechanical simulation data, experimental data, and cheminformatics into a predictive engine for macromolecular design.
- Utilized Python ML/DS libraries including Scikit-learn, Keras, Pandas, Matplotlib, and TPOT.