

Haotian Teng

E-mail: haotiant@cs.cmu.edu

Personal Website: <https://haotianteng.github.io/>

EDUCATION

Carnegie Mellon University (CMU), Pittsburgh, PA, United States

Doctor of Philosophy, Computational Biology, School of Computer Science (SCS)

Advisor: Ziv Bar-Joseph, Machine Learning and Computational Biology department, SCS, CMU.

2019 - present

Co-Advisor: Carl Kingsford, Computational Biology department, SCS, CMU

Core Courses: Machine Learning 10-701, Deep Reinforcement Learning 10-703, Probabilistic Graphical Models 10-708, Convex Optimization 10-725, Advanced Natural Language Processing 11-711, Computational Genomics 02-710

University of Queensland, Brisbane, Queensland, Australia

Master of Bioinformatics, Queensland Brain Institute

2016 - 2018

Peking University, Beijing, China

Bachelor of Science, Physics

2011 - 2015

RESEARCH EXPERIENCE

Ph.D. Candidate, Carnegie Mellon University, Pittsburgh, PA

Target-based drug generation using E(3) equivalent graph neural network and diffusion model

Dec 2022 - Present

- Implemented an E(3)-equivalent graph neural network combined with a probabilistic diffusion model to generate small molecules with both high binding affinity to target protein pockets and enhanced druggability.

Clustering single-cell spatial transcriptomics data using hybrid model.

Aug 2019 - Feb 2022

- Introduced FICT, a Gaussian-Multinomial probabilistic graphical model, to cluster cells from single-molecule spatial transcriptomics data, discovering six novel neuron development-related genes based on spatially-aware clusters.
- Developed a general gene embedding algorithm using a denoising autoencoder, achieving a 17.3X speed increase compared to UMAP.
- Created a simulation pipeline based on Metropolis-Hastings sampling to validate spatial transcriptomics clustering results.

Detecting N6-methyladenosine (m6A) modification in RNA using long read sequencing

Jan 2021 – Jan 2023

- Developed a hybrid PyTorch model that combines Non-Homogeneous Hidden Markov Model (NHMM) and Convolutional-Recurrent Neural Network (CRNN) for semi-supervised training on time-series signals to detect m6A modifications in RNA sequencing.
- Achieved state-of-art accuracy in RNA methylation detection, obtain a 22% and 7% relative improvement on yeast and human HEK293T cell lines.

Neural architecture search using deep reinforcement learning

Jan 2021 – Jul 2021

- Identified 3 novel convolutional neural layer structures using neural architecture search and achieved better performance than the usual residual connection layer on several common tasks in the OpenAI gym library with 20% fewer parameters.

WORK EXPERIENCE

Applied Scientist @ Amzon - Alexa

May 2023 - Sep 2023

- Teach Large Language Models (GPT-J and Llama2) to use tools through self-supervised learning.

Algorithm Engineer @ Alibaba - Cloud

Feb 2019 – March 2019

- Improving LRU-based cache through pattern similarity prediction.

Bioinformatics Engineer @ Novogene Europe

Sep 2018 - Feb 2019

- Optimized the human resequencing and laboratory automation pipeline cooperatively in a group with 7 people.
- Designed and developed the long-read sequencing platform as group leader with 3 group members.

SELECTED PUBLICATIONS

- **Teng, H.**, Stoiber, M., Bar-Joseph, Z., & Kingsford, C. (2024). Detecting m6A RNA modification from nanopore sequencing using a semi-supervised learning framework. *bioRxiv*, 2024, <https://doi.org/10.1101/2024.01.06.574484>
- **Teng, H.**, et al. (2022). Clustering Spatial Transcriptomics Data. *Bioinformatics*, 38(4), 997-1004. **(Citation 19)**
- **Teng, H.**, Cao, M. D., Hall, M. B., Duarte, T., Wang, S., & Coin, L. J. (2018). Chiron: translating nanopore raw signal directly into nucleotide sequence using deep learning. *GigaScience*, 7(5), giy037. **(Citation 137)**
- Avitan, L., Pujic, Z., Mölter, J., Van De Poll, M., Sun, B., **Teng, H.**, Amor, R., Scott, E.K. and Goodhill, G.J., 2017. Spontaneous activity in the zebrafish tectum reorganizes over development and is influenced by visual experience. *Current Biology*, 27(16), pp.2407-2419. **(Citation 73)**

SKILLS

- Programming: Python, C++, Bash, MATLAB, R
- Packages&Platforms: Tensorflow, PyTorch, OpenGL, BWA, SAMtools, Velvet.
- Deep Learning: CNN, RNN, LSTM, GNN, Diffusion model, GAN, VAE, BERT, Transformer, deep Q-learning