**Haotian Teng**

E-mail: [haotiant](mailto:haotian.teng@uqconnect.edu.au)@andrew.cmu.edu

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

**EDUCATION**

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

*Doctor of Philosophy, School of Computer Science (SCS)*

Advisor: Ziv Bar-Joseph, Machine Learning and Computational Biology department, SCS, CMU. 2019 - present

**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. Research Assistant, Advised by Prof. Ziv Bar-Joseph, Carnegie Mellon University**

**Target-based drug generation using graph neural networks with probabilistic diffusion model.** Nov 2022 – present

* Adapted an E(3) equivalent graph neural network with a probabilistic diffusion model to generate small molecules with high binding affinity to a given target protein pocket and high druggability.

**Neural architecture search using deep reinforcement learning** Jan2021 – 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.

**Text mining and data annotation for named entity recognition** Mar 2022 – Jul 2022

* Designed a fully automated pipeline to collect text from topic-specified online academic articles.
* Conducted data cleaning and formatting using customized regular expressions and labeled the data using label-studio.
* Finetuned the SciBERT model with adapt layer on our labeled dataset and with the CoNLLpp training dataset and achieved similar performance compared to the state-of-the-art model on the CoNLLpp test dataset.

**Clustering the single-cell spatial transcriptomics data with novel probabilistic graphical model.** Aug2019 - Feb 2022

* Developed a robust probabilistic graphical model [FICT](https://academic.oup.com/bioinformatics/article/38/4/997/6384569) that clusters the cell for spatial transcriptomics data. By considering the spatial distribution in addition to the gene expression, the model can give cell clusters variated in spatial domain.
* Conducted simulation to get desired cell distribution with Metropolis-Hastings swap algorithm.
* Constructed regularized denoise auto-encoder that achieve dimensional reduction faster than UMAP and applicable to any dimensions. It disentangles complicated clusters failed by t-SNE which does not consider nonlinear transformations.

**PROFESSIONAL EXPERIENCE**

**Bioinformatics Engineer @ Novogen, Beijing, China** Sep2018 - Feb 2019

* Optimized the human resequencing and laboratory automation pipeline cooperatively in a medium size group with 7 people, achieved 50% faster processing speed and estimated 13% cost saving.
* Initiated the first long-read sequencing cloud platform in China, achieved real-time online basecalling and data analysis.

**SELECTED PUBLICATIONS**

* **Teng, H.**, Yuan, Y., & Bar-Joseph, Z. (2021). Clustering Spatial Transcriptomics Data. *Bioinformatics*, *38*(4), 997-1004. (**Citation 21)**
* **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 135)**
* 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 70)**
* **Teng, H.** “A neuron-muscle circuit model of C.elegans’s locomotion.” *Bachelor of Science Thesis: Peking University, 2015*

**SKILLS**

* Programming: Python, C++, Bash, MATLAB, R, PHP, CSS
* Selected Packages&Platforms: Tensorflow, PyTorch, Linux, OpenGL, BWA, SAMtools, Velvet.
* General Machine Learning: Linear Regression, Logistic Regression, SVM, Decision Tree, Random Forest, Probabilistic Graphical Models, RBM, GMM, KNN, K-Means, PCA, t-SNE, UMAP, Bagging, AdaBoost, Gradient Boosting.
* Deep Learning: CNN, RNN, LSTM, GMM, Diffusion model, GAN, VAE, BERT, Transformer, deep Q-learning