# HAIDONG YI

haidyi@cs.unc.edu ◇ Google Scholar

## EDUCATION

**University of North Carolina at Chapel Hill** *Sept. 2019 - present*
Ph.D. student of Computer Science (expected graduate date: 05/2024)
Advisor: Junier B. Oliva

**Nankai University** *Sept. 2016 - Jun. 2019*
M.E. in Computer Science and Technology
Advisor: Han Zhang

**Nankai University** *Sept. 2012 - Jun. 2016*
B.E. in Intelligent Science and Technology[1]
B.S. in Mathematics and Applied Mathematics[2] (Co-Major)
Advisor: Han Zhang[1] & Changliang Zou[2]

## PUBLICATIONS

**Peer reviewed publications (* denotes equal contribution)**

- Yang Li*, **Haidong Yi***, Christopher M Bender, Siyuan Shan, Junier B Oliva, "Exchangeable Neural ODE for Set Modeling", Advances in Neural Information Processing Systems (NeurIPS). Virtual, 2020.
- Ausland, Catherine, Jinfang Zheng, **Haidong Yi**, Bowen Yang, Tang Li, Xuehuan Feng, Bo Zheng, and Yanbin Yin. "dbCAN-PUL: a database of experimentally characterized CAZyme gene clusters and their substrates." Nucleic Acids Research (2020).
- **Haidong Yi**, Le Huang, Bowen Yang, Javi Gomez, Han Zhang, and Yanbin Yin. "AcrFinder: genome mining anti-CRISPR operons in prokaryotes and their viruses." Nucleic Acids Research (2020).
- Wang, Xiao, **Haidong Yi**, Jia Wang, Zhandong Liu, Yanbin Yin, and Han Zhang. "GDASC: A GPU parallel based web server for detecting hidden batch factors." Bioinformatics (2020).
- **Haidong Yi***, Ayush T. Raman*, Han Zhang, Genevera I. Allen, Zhandong Liu, "Detecting hidden batch factors through data-adaptive adjustment for biological effects", Bioinformatics, 2018, 34(7):1141-1147.
- Huang Le, Han Zhang, Peizhi Wu, Entwistle Sarah, Li Xueqiong, Yohe Tanner, **Haidong Yi**, Yang Zhenglu, Yanbin Yin, "dbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation", Nucleic Acids Research, 2018, 46(Database issue):D516-D521.
- Liu Wei, Ma Shunjian, Sun Mingwei, **Haidong Yi**, Wang Zenghui, Zengqiang Chen, "Sequential quadratic programming-based fast path planning algorithm subject to no-fly zone constraints", Engineering Optimization, 2015, 48(8):1401-1418.

## PROJECTS

**Develop a fast implementation of convex biclustering** *July 2020 - Oct. 2020*
*Project Leader, advised by Eric C. Chi*

- I developed a C/C++ program called COBRAC to accelerate the full path solution of convex biclustering.
- COBRAC achieves 2.5-4x computing acceleration by iteratively compressing the problem size (code).
- We developed a web server to run COBRAC and generate the clustering dendrogram (website).

**Develop deep learning method for set modeling** *Mar. 2020 - Jun 2020*
*Project Participant, advised by Junier B. Oliva*

- We proposed ExNODE, an exchangeable module for set modeling, which explicitly captures the intradependencies among set elements (Accepted by Neurips 2020).
- ExNODE represents a type of invertible flow transformation on which the invariant set likelihood can be achieved.

- I developed a ExNODE-based model for set classification, which achieves state-of-the-art performance on point cloud classification task with much less parameters.
- I developed a VAE-based model for time variant set modeling, which can be used to generate temporal sets.

**Design Constraint Conditions of Arc No-fly Zones** *Sept. 2013 - Sept. 2014*
*Project Leader*

- I modeled an constraint of arc no-fly zone using line segment approximation.
- I expanded a C++ software to solve the constrained optimization using sequential quadratic programming.

**Detect Batch Factors in High Dimensional Gene Expression Dataset** *Sept. 2016 - Sept. 2017*
*Project Leader, advised by Zhandong Liu*

- I modeled a data-adaptive method to estimate biological effects using convex optimization.
- I detected and classified batch factors using semi-nonnegative matrix factorization method.
- I developed an R package, DASC, to identify and classify batch factors.

## SKILLS

| | |
|---|---|
| **Development Languages:** | Python, C/C++, R |
| **Development Framework:** | Pytorch, OpenMP/MPI, OpenBlas |
| **Software** | Docker, Git |

## TEACHING

| | |
|---|---|
| **COMP411: Computer Organization** | Teaching Assistant, Fall 2019 |
| **COMP488: Data Science in the Business World** | Teaching Assistant, Spring 2020 |
| **COMP550: Algorithm and Analysis** | Teaching Assistant, Fall 2020 |

## AWARDS

- College Top 5 Undergraduate Thesis Award, Nankai University, 2016
- Meritorious Winner of Interdisciplinary Contest In Modeling Certificate of Achievement (ICM), 2015.
- College Top 1 Award of "Hundreds of Undergraduate Students' Projects", Nankai University, 2014

## SOFTWARES

**DASC** Link to Github Repository

- An R package used for identifying batches and classifying samples into different batches in high dimensional gene expression dataset.

**Acrfinder (website)** Link to Github Repository

- A python code used for identifying anti-CRISPR (Acr) and Acr-associated (Aca) genomic loci in gene sequence data.