

# Daniel (Chaofan) Tao

423 Towerview Dr., Durham, North Carolina

☎ (+1) 919-884-1590 | ✉ chaofan.tao@duke.edu | 🌐 www.danieltao.me | 📷 danieltao | 📱 chaofan daniel tao

## Education

### Duke University

B.S OF COMPUTER SCIENCE (CS)

Durham, NC

August, 2017-December, 2020

- GPA: 3.9/4.0
- Named to the Dean's List (top third in the school) every semester
- **Courses:** Machine Learning (graduate level), Database, Operating Systems, Computer Architecture, Internet, Data Structure, Algorithms

## Experience

### InterSystems

Cambridge, MA

DATA PLATFORM DEVELOPMENT INTERN

May, 2019-August, 2019

- Worked in the Core Development team on a project about **PMML** using ObjectScript
- Compared the IRIS PMML with standard JPMML with Java and Python, fixed over a dozen existing bugs in IRIS PMML
- Added functionality to support NN, SVM, Kmeans, Naive Bayes, and Random Forest in IRIS PMML with ObjectScripts
- Improved the runtime performance of IRIS PMML by up to 8 times by optimizing data input methods
- Worked on a industrial codebase that would influence real customers, wrote unit tests to ensure quality
- Used Swarm and Perforce for source control

### Duke Prediction Analysis Lab

Durham, NC

UNDERGRADUATE RESEARCH ASSISTANT

October, 2017-Present

- Researched the application of interpretable NN in medical area with the help of prof. **Cynthia Rudin**
- Published ***This Looks Like That: Deep Learning for Interpretable Image Recognition*** on NeurIPS 2019
- Applied PPNet to give a self-explained diagnosis of breast cancer mammograms, achieved **state-of-the-art** AUC
- Used existing attention NN and CNN models such as VGG, ResNet, RA-CNN, R-CNN, U-Net, etc.
- Used Tensorflow, Keras, Scikit-learn and Pytorch for the dataset of Breast Cancer
- Used GPU computing and data parallelism to speed up training

### Duke Information Initiative

Durham, NC

SOFTWARE DEVELOPER INTERN

May, 2018-August, 2018

- Built an **interactive web app** with Plotly to demonstrate and analyze datasets of single cell sequencing
- Built a pipeline to reduce dimension, cluster, and visualize single cell sequencing data
- Used TSNE, PCA, autoencoder, KMeans, and other deep learning clustering methods
- Achieved cluster accuracy of **76%** on testing dataset

## Publications

### This Looks Like That: Deep Learning for Interpretable Image Recognition

Durham, NC

NEURIPS 2019

2019

- Proposed a new neural network architecture that uses prototypes to produce interpretable classification results
- Got accepted as a **spotlight (2 %)** paper

## Skills

### Programming

**Fluent:** Python, Java

**Proficient:** JavaScript, C, Cpp, Assembly, Latex

### Tools

Git, Perforce, Maven

### Web Dev.

HTML, CSS, Bootstrap, React, Django

### Database

MySQL, PostgreSQL, MongoDB, XQuery

## Projects & Clubs

### <DAML>

Durham, NC

CO-FOUNDER

September, 2019-Present

- Started Duke Applied Machine Learning club that applies machine learning in real life problems
- Current customers include US Air Force, Duke Brain Center, Duke Immerse VR center, etc.
- Lead data scientist team consists of P.hDs, undergrads, and masters

### <Router Simulator>

Durham, NC

C

April, 2019

- Built a virtual router using Routing Information Protocol (RIP).
- The router could perform tasks like ping, traceroute, etc. dynamically