

THANH NGOC (NATALIE) PHAM

874 Flemington St., Pittsburgh, PA, 15217

npham.cmu@gmail.com | (412) 785-4349 | [Github](#) | [LinkedIn](#)

EDUCATION

Carnegie Mellon University (CMU), Pittsburgh, PA

May 2023

Master of Science, Data Analytics for Science, GPA: 4.12/4.33

Chinese University of Hong Kong (CUHK), Hong Kong

Jul 2022

Bachelor of Science, Mathematics – Computational Big Data Analytics Stream, GPA: 3.05/4.00

University of Pittsburgh, Pittsburgh, PA

Apr 2022

Undergraduate semester exchange program, GPA: 4.00/4.00

RESEARCH EXPERIENCE

Machine Learning Department, CMU

Pittsburgh, PA

Research Assistant

April 2022 – Now

- Study the S4 model and its variants for processing long-sequences and time series data.
- Analyze the performance of online learning algorithms such as Follow The Leader and online gradient flow in the continuous time setting.

Mathematics Department, CUHK

Hong Kong

Participant of Undergraduate Research Opportunities Program

Jan 2022 – Jul 2022

- Studied data-driven reduced-order modeling for time-dependent problems.
- Used reduced-basis methods and data-driven approaches (such as Gaussian Process or Neural Networks) to approximate numerical solution to partial differential equations such as Navier-Stoke equations.

Shanghai Jiao Tong University

Hong Kong - Virtual

Summer Research Intern

Jul 2021 – Aug 2021

- Investigated 3D object detection algorithms based on point cloud and semi-supervised learning, and analyzed their advantages and disadvantages.
- Implemented some of the algorithms (such as PV-RCNN, Point-RCNN) and validated them using experimental results with KITTI 3D dataset.

PROJECTS

Streamline Protein Image Processing, CMU

Jan 2023 – May 2023

- Collaborated with a team of data scientists from Bristol Myers Squibb as part of master's capstone project.
- Generated pseudo-labels for 4000+ unlabeled protein crystallization images and HS-AFM data.
- Developed computational strategies using deep learning methods (rVAE, CNNs) to automate image processing and analysis workflows on multiple datasets.

Analyze Data from Public Reports of Asian giant hornets in Washington, CUHK

Jan 2021 - Feb 2021

- Constructed a model for predicting the existence of Asian giant hornets using public reports and reasonably allocate workforce to remove the hornets' nest for the Mathematical Contest in Modeling (MCM) 2021
- Applied the Metropolis Hasting algorithm to predict the spread of the hornets based on 14 positive public reports to update possible existence in a spatial distribution.
- Trained an image classifier with convolutional neural networks on Google Colab with 85% accuracy to evaluate the likelihood of false reports submitted by the public.

WORK EXPERIENCE

Machine Learning Department, CMU

Pittsburgh, PA

Graduate Teaching Assistant

Sep 2023 – Now

- Assist the lecturer of the course 10-708: Probabilistic Graphical Models for designing assignments and grading.
- Hold office hours and recitation sessions to clarify students' understanding of the course materials, including topics such as Monte Carlo Markov Chain, Variational Autoencoders, Causal Inference.

LnData Co. Ltd

Hong Kong - Virtual

NLP Data Science Intern

May 2021 – Aug 2021

- Performed data analysis for a dataset of 40,000+ posts collected from social media.
- Built an algorithm using fastText and Scikit-learn libraries in Python to extract keywords from the dataset provided and assign them to clusters.

SKILLS

Technical: MATLAB, Python, R, C++, Latex, Microsoft Office

Language: Vietnamese, English, Mandarin, Cantonese

AWARDS

Honours At Entrance, **CUHK**

Sep 2018

Faculty Admission Scholarship for Science Students, **Science Faculty, CUHK**

Sep 2018