Roberto Halpin Gregorio

M.S. Student, Computer Science
Cornell University

rgh224@cornell.edu • 🖥 roberto-hg.github.io

Education

Cornell University
M.S. in Computer Science,   *2020 – Spring 2023 (expected)*
Ithaca, NY

Cornell University
B.S. in Computer Science,   *2016 – 2020*
Ithaca, NY

Experience

PRINCIPAL MACHINE LEARNING (ML) RESEARCHER   **Cornell University**
Self-Supervised Learning (M.S. Thesis Research)   *Oct 2020 - Current*

- Analyzing the effect of augmenting datasets with GAN generated images in self-supervised learning.
- Achieved a significant accuracy improvement using SimCLR model under the CIFAR-10 dataset.
- Extending these experiments to semi-supervised, supervised, and transfer learning tasks.
- Advised by Bharath Hariharan and Madeleine Udell.

ML RESEARCHER   **Cornell University**
Missing Indicator Method (MIM)   *Spring 2022 - Current*

- Tackled the important issue of missing data in real-world datasets.
- Developed Selective MIM (SMIM), a novel method that adds missing indicators only for features that have an informative missing pattern.
- Performed extensive empirical justification in a variety of settings and across a wide range of supervised learning models.
- Analyzed and visualized MIM features using an attention-based transformer model.

ML RESEARCHER   **Cornell University**
Autonomous Vehicles - Amodal Segmentation   *Aug 2018 - Dec 2020*

- Developed amodal segmentation algorithms for road identification with self-driving cars.
- Designed a labeling tool used in-house and setup in Amazon MTurk for an amodal segmentation dataset project.
- Evaluated state-of-the-art 3D trackers and object detectors on full sensor datasets - KITTI, NuScenes, Lyft, Waymo, Argo.
- Collected and built a new synthetic dataset for amodal segmentation of road images.
- Worked under Wei-Lun Chao, Kilian Weinberger, Bharath Hariharan, and Mark Campbell.

ML RESEARCHER   **Cornell University**
Distributed Machine Learning   *Oct 2019 - May 2020*

- Developed a new asynchronous centralized distributed machine learning architecture.
- Created a novel asynchronous SGD update scheme and utilized the RDMA network protocol in the parameter server setting.
- Implemented multiple baselines using a distributed parameter server architecture on TensorFlow.
- Worked under Chris De Sa and Ken Birman.

ML RESEARCH ASSISTANT   **Cornell University**
Computer Vision with Deep Learning   *Summer 2018*

- Analyzed modern deep stereo regression research and codebases.
- Coded from-scratch implementations of many deep learning computer vision white papers.
- Participated and presented in the weekly research lab reading group.

RESEARCH ASSISTANT   **U.C. Davis**
Parallel Sparse Matrix Multiplication Optimization through GPUs   *Summer 2017*

- Implemented an efficient sparse vector-matrix multiplication algorithm from scratch using CUDA and C.
- Analyzed state-of-the-art Parallel Sparse Matrix Multiplication research.
- Learned parallel programming in C with CUDA.

Projects

Contrastive Representation Learning   **Cornell University**
Research Project   *Spring 2022*

- Developed and coded a new visual representation learning method called MIM-CLR.
- MIM-CLR combines contrastive and masked image modeling approaches.
- Our proposed method achieves significant accuracy improvements over previous work (>10%) on the STL-10 dataset.

Representation Learning Theory   **Cornell University**
Research Project   *Fall 2021*

- Analyzed how to bound the best case loss and risk for learned representations.
- Improved the generality of previous bounds in literature.
- Extended bounds to include a k-layer neural network classifier.

Pancreas Tumor Segmentation   **Cornell University**
Research Project   *Fall 2018*

- Addressed the problem of the small amount of available medical data by using 2D slices of 3D voxels.
- Created and coded new models that utilized our new generated dataset to segment pancreas tumors.

Teaching and Courses

CS 4670: Computer Vision
Teaching Assistant   *Spring 2022*

CS 6787: Advanced Machine Learning Systems
Teaching Assistant   *Fall 2021*

CS 4787: Principles of Large-Scale Machine Learning
Teaching Assistant   *Spring 2019, 2020, 2021*

CS 4700: Foundations of Artificial Intelligence
Teaching Assistant   *Fall 2020*

CS 4780: Machine Learning for Intelligent Systems
Teaching Assistant   *Fall 2019*

Relevant Coursework
Probability, Linear Algebra, Multivariable Calculus, Artificial Intelligence, Machine Learning for Data Science, Machine Learning for Intelligent Systems.

PhD Level Courses:
Computer Vision, Advanced Machine Learning Systems, Advanced Topics in Machine Learning, Machine Learning Theory, Deep Generative Models, Mathematics of Data Science, Bayesian Statistics and Data Analysis, Statistical Principles, Matrix Computations, Algorithmic Ideas for Speeding Up Optimization.

Skills

Languages: Python, R, Julia, Java, Javascript, C, C++, MATLAB

Other: TensorFlow, PyTorch, scikit-learn, OpenCV, Git, Slurm, CUDA, LaTeX, AWS, Amazon MTurk