Roberto Halpin Gregorio

rgh224@cornell.edu • roberto-hg.github.io • github.com/roberto-hg

Education

Cornell University **Ithaca, NY**
M.S. in Computer Science, Current. GPA: 3.98 *2016 – Spring 2023 (expected)*
B.S. in Computer Science, 2020. Major GPA: 3.76

Experience

Image Augmentation via Generative Models (Thesis Research) **Cornell University**
Machine Learning (ML) Researcher *Oct 2020 - Current*

- Analyzing the effect of augmenting datasets with GAN generated images in self-supervised learning
- Method improves SimCLR's self-supervised representations under the CIFAR-10 dataset
- Implemented training and evaluation codebase in Python using PyTorch, TensorFlow, Pillow, and OpenCV

Handling Real-World Missing Data Research **Cornell University**
ML Researcher *Apr 2022 - Current*

- Developed Selective MIM (SMIM), a novel method that addresses missing tabular data
- Performed extensive empirical justification in a variety of settings and across a wide range of supervised learning models using PyTorch, scikit-learn, OpenML and real-world clinical data
- Underwent the academic peer review process at NeurIPS, AISTATS, and KDD.

ML, Computer Vision, AI Teaching Assistant **Cornell University**
Teaching Assistant *Jan 2019 - May 2022*

- Held weekly office hours; created, tested, and graded programming, problem set, and exam questions and solutions
- Managed and mentored 10+ undergraduate teaching assistants

Autonomous Vehicles - Amodal/Panoptic Segmentation Research **Cornell University**
ML Researcher *May 2018 - Dec 2020*

- Developed amodal segmentation algorithms for road identification with self-driving cars
- Designed a custom JavaScript web labeling tool used in-house and on Amazon MTurk with AWS
- Evaluated state-of-the-art 3D trackers and object detectors in PyTorch and TensorFlow on full sensor datasets - KITTI, NuScenes, Lyft, Waymo, Argo
- Collected and built a new synthetic dataset for amodal segmentation of road images

Distributed Deep Learning Research **Cornell University**
ML Researcher *Oct 2019 - May 2020*

- Created novel asynchronous SGD optimization scheme using RDMA network protocol in the centralized distributed setting
- Implemented multiple baselines using a distributed parameter server architecture in TensorFlow (Python)

Parallel GPU Sparse Matrix Multiplication Optimization Research **U.C. Davis**
Research Assistant *Summer 2017*

- Implemented an efficient sparse vector-matrix multiplication algorithm from scratch using CUDA and C
- Brainstormed ideas for optimizations in Sparse Matrix Multiplication using parallel algorithms

Projects

Contrastive Representation Learning PDF *Spring 2022*

- Primary contributor on a novel visual representation learning method that achieves >10% accuracy improvements over previous work on the STL-10 dataset
- Implemented in Python using PyTorch

Representation Learning Theory PDF *Fall 2021*

- Discovered new bounds on the performance of downstream classifiers based on feature representation properties

Pancreas Tumor Segmentation PDF *Fall 2018*

- Improved mean IoU by a factor >15% on 3D CT scan images of pancreas by employing transfer learning to segment tumors
- Implemented in Python using PyTorch

Skills

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

Other: PyTorch, TensorFlow, NumPy, scikit-learn, Pillow, OpenCV, CUDA, AWS