Roberto Halpin-Gregorio

rgh224@cornell.edu • (650) 417-3894 • Portfolio: roberto-hg.github.io • LinkedIn: roberto-hg

Education

Cornell University — **Ithaca, NY**
M.S. in Computer Science, GPA: 3.98 — Present
B.S. in Computer Science, Major GPA: 3.76 — May 2020

Experience

Bharath Hariharan Lab — October 2020 - Present
Research Assistant — Data Augmentation via Deep Generative Models `PDF`

- Built, trained, and tested deep generative models to augment image datasets, improving training speed and accuracy by over 8% in multiple machine learning applications
- Augmented vision datasets with more than 10,000 data-efficient Generative Adversarial Network (GAN) generated images, improving prediction in machine learning tasks using custom PyTorch + TensorFlow modules and Pillow + OpenCV image processing
- Configured, monitored, and logged custom distributed training and inference runs on remote clusters with over 300 users, 500 compute nodes and 200 GPUs

Madeleine Udell Lab — April 2022 - June 2023
Research Assistant — Missing Real-world Data `KDD23 Publication`

- Developed novel methods to handle missing tabular data in real-world datasets by incorporating informative missing patterns, improving RMSE, AUC, and wall-clock times over standard techniques
- Performed missing data experiments on real-world data: MIMIC-III clinical data and several OpenML datasets with linear + logistic regression, XGboost, multilayer perceptron (MLP), and tabular transformer models using PyTorch and scikit-learn

Kilian Q. Weinberger Lab — May 2018 - May 2021
Research Assistant — Autonomous Vehicles

- Led the development of an end-to-end data, training, and inference pipeline for occluded road segmentation using state-of-the-art Convolutional Neural Network (CNN) architectures in PyTorch and the Berkeley Deep Drive (BDD), Cityscapes, and in-house curated datasets
- Constructed a robust Amazon MTurk labeling pipeline with a custom JavaScript web labeling tool for road segmentation labeling using Amazon Web Services (AWS) tools
- Implemented and tested multiple state-of-the-art 3D trackers and object detectors in PyTorch and TensorFlow on full sensor datasets – KITTI, NuScenes, Lyft, Waymo, Argo

Cornell Systems Lab — October 2019 - May 2020
Research Assistant — Distributed Machine Learning `PDF`

- Developed a novel asynchronous distributed machine learning scheme for multi-GPU clusters, improving training time by 13% while maintaining statistical performance compared to other distributed methods
- Implemented deep learning models: ResNet, VGG, and random Fourier feature logistic regression with SGD and SVRG in a distributed GPU parameter server setting using C++ BLAS, and TensorFlow and PyTorch distributed

John Owens Lab — U.C. Davis — June 2017 - August 2017
Research Assistant — Sparse GPU Optimization

- Designed and implemented custom NVIDIA GPU kernels using CUDA and efficient sparse algorithms in C to improve sparse vector-matrix and matrix-matrix multiplication over standard implementations

Cornell Computer Science Department January 2019 - May 2022
Teaching Assistant

○ Provided real-time assistance to thousands of students during weekly office hours and online course Q&A forums

○ Created, tested, and graded assignments/exams; organized and oversaw grading sessions

○ Managed and mentored 10+ undergraduate teaching assistants

○ Relevant courses taught: Machine Learning, Computer Vision, Principles of Large-Scale Machine Learning, Advanced Machine Learning Systems, Artificial Intelligence

Skills

Languages: Python, C++, Javascript, C

ML Areas: Computer Vision, Deep Learning, Data Augmentation, Training and Inference Pipelining, Distributed Training, Generative Modeling, Self-supervised Learning, Data Cleaning and Preparation, Object Detection, Semantic Segmentation, Image Processing, Tabular Data, Missing and Sparse Data, Tracking

ML Models: CNNs, GANs, MLPs, Transformers, XGBoost, Classic ML/Statistical Models

Libraries/Tools: PyTorch, TensorFlow, NumPy, scikit-learn, Pillow, OpenCV, CUDA, AWS, Docker, Amazon MTurk

Other Projects

Visual Self-supervised Learning **PDF** :octocat: Spring 2022
○ Designed and implemented a novel masked-image vision transformer and contrastive self-supervised learning method that achieves >10% accuracy improvements over previous work on the STL-10 dataset using PyTorch

Representation Learning Theory **PDF** Fall 2021
○ Discovered new theoretical bounds on the performance of classifiers based on feature representation properties

○ Improved generalization of previous bounds in literature and extended bounds to a k-layer neural network

Pancreas Tumor Segmentation **PDF** :octocat: Fall 2018
○ Designed a novel transfer learning strategy in pancreas tumor segmentation using CNNs implemented in PyTorch, improving mean Intersection over Union by 15% on Memorial Sloan Kettering Cancer Center's dataset

Relevant Coursework

ML/AI: Computer Vision, Deep Generative Models, Advanced Machine Learning Systems, Advanced Topics in Machine Learning, Machine Learning Theory, Machine Learning, Artificial Intelligence

Stats: Bayesian Statistics and Data Analysis, Statistical Principles

Other: Mathematics of Data Science, Matrix Computations, Algorithmic Ideas for Speeding Up Optimization