Roberto Halpin Gregorio

Email: rgh224@cornell.edu • Homepage: roberto-hg.github.io

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   Oct 2020 - Present
Research Assistant — Data Augmentation with Deep Generative Models
○ Augmented vision datasets with data-efficient GAN generated images improving low data regime prediction in supervised, semi- and self-supervised learning tasks using PyTorch, TensorFlow, Pillow, and OpenCV
○ Achieved improved linear probing accuracy with SimCLR's representations trained using Adam and LARS on a GAN augmented CIFAR-10 dataset

Madeleine Udell Lab   Apr 2022 - Present
Research Assistant — Missing Real-world Data
○ Developed a novel selective method that improves handling of missing tabular data over standard imputation
○ Performed missing data experiments on real-world data: MIMIC-III clinical data and several OpenML datasets with linear regression, XGboost, MLP, and tabular transformer models using PyTorch and scikit-learn
○ Analyzed explainability of our method's features by visualizing the attention weights of a trained transformer

Kilian Q. Weinberger Lab   May 2018 - May 2021
Research Assistant — Autonomous Vehicles
○ Collected, cleaned, and built an synthetic dataset of over 8000 road scenes for amodal segmentation based on the Berkeley Deep Drive (BDD) and Cityscapes datasets
○ Developed novel Dilated ResNet based architectures that incorporate modal information and a smoothing loss to improve the amodal segmentation task in road identification on several amodal road scene datasets using PyTorch
○ Designed JavaScript web labeling tool for amodal road segmentation labeling used in-house and on Amazon MTurk
○ Constructed a Amazon MTurk labeling pipeline using AWS cloud storage with specific time and budget constraints
○ Tested multiple configurations of 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   Oct 2019 - May 2020
Research Assistant — Distributed Machine Learning
○ Designed a novel asynchronous HOGWILD!-esque SGD optimization scheme using RDMA network protocol over multiple GPU machines improving training time and statistical performance over other distributed methods
○ Implemented ResNet, VGG, and random Fourier feature logistic regression with SGD and SVRG in a distributed GPU parameter server setting using TensorFlow and PyTorch distributed to serve as baselines

Cornell Computer Science Department   Jan 2019 - May 2022
Teaching Assistant
○ Held weekly office hours; created, tested, and graded programming, problem set, and exam questions and solutions
○ Managed and mentored 10+ undergraduate teaching assistants; organized and oversaw grading sessions
○ Relevant courses taught: Machine Learning for Intelligent Systems, Computer Vision, Principles of Large-Scale Machine Learning, Advanced Machine Learning Systems, Foundations of Artificial Intelligence

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

- Designed a custom NVIDIA GPU kernel using CUDA and an efficient sparsity-minded algorithm in C to improve sparse vector-matrix multiplication over a standard implementation

Publications

Mike Van Ness, Tomas M. Bosschieter, **Roberto Halpin-Gregorio**, Madeleine Udell. 2023. The Missing Indicator Method: From Low to High Dimensions. To appear in *KDD 2023 Applied Data Science Track*

Projects

Visual Self-supervised Learning   PDF ⬡   Spring 2022
- Designed and implemented a novel masked-image and contrastive visual 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 the generality of previous bounds in literature and extended bounds to a k-layer neural network classifier

Pancreas Tumor Segmentation   PDF ⬡   Fall 2018
- Designed novel 2D-to-3D transfer learning strategy in pancreas tumor segmentation using dilated ResNets implemented in PyTorch improving mean Intersection over Union by 15% over the 2D baseline on Memorial Sloan Kettering Cancer Center's pancreas tumor dataset

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, Javascript, C
Other: PyTorch, TensorFlow, NumPy, scikit-learn, Pillow, OpenCV, CUDA, AWS, Amazon MTurk