Abhay Singh

contact
as2626@cornell.edu
(404) 353-0477
linkedin.com/in/as2626
github.com/as2626

education

Cornell University, Ithaca, NY
B.S. in Computer Science, GPA: 4.15/4.30 — Aug 2018 – May 2021
M.S. in Computer Science — Aug 2021 – May 2023

coursework
∗ = in progress
† = teaching assistant

CS 6241: Data Science Numerics∗	ECE 6970: Statistical Distances∗	CS 6670: Computer Vision
ORIE 6510: Probability∗	CS 4780: Machine Learning†	CS 4820: Algorithms†
MATH 4130: Analysis I (Honors)	MATH 4315: Linear Algebra	CS 4850: Math Foundations
CS 4120: Compilers	CS 4410: Operating Systems	CS 2110: Data Structures & OOP

publications

Combining Label Propagation and Simple Models Out-performs GNNs (ICLR 2021)
Qian Huang, Horace He, Abhay Singh, Ser-Nam Lim, and *Austin Benson*

Better Set Representations For Relational Reasoning (NeurIPS 2020)
Qian Huang, Horace He, Abhay Singh, Yan Zhang, Ser-Nam Lim, and *Austin Benson*

experience

Yext, New York, NY
Software Engineering Intern — May 2020 – Aug 2020
- Designed and integrated static code analysis tool used firm-wide on over 80% of codebase to scan vulnerable Java code at compile-time
- Wrote multi-threaded Golang script to determine unprotected customer apps that downloads and parses terabytes of API log data on-the-fly via AWS S3, and makes remote-procedure calls to fetch app data by API key; improved performance by 4x relative to previous solution
- Integrated webhooks to automate modification of company repository permissions using Github's REST API, notifying teams automatically via Slack and email

Morgan Stanley, New York, NY
Technology Summer Analyst — June 2019 – Aug 2019
- Architected and implemented end-to-end data pipeline to process and analyze over 800,000,000 entries of financial data daily with highly optimized, parallelized Python scripts, using NumPy and Pandas
- Reduced mainframe consumption by 90%, from 5000 to 500 CPU-seconds, saving tens of millions of dollars in annual costs
- Created and deployed firm-wide DevOps web tool to analyze large text-based datasets

Cornell Unmanned Air Systems, Ithaca, NY
Computer Vision and DevOps Lead — Oct 2018 – present
- Designed and implemented custom object detection and classification model (Mask R-CNN variant with multi-head output) in multi-task learning setting on collected aerial imagery dataset, in PyTorch
- Lead all computer vision tasks on team, with individual efforts directly increasing classification task accuracy by 32% and object detection mAP IoU by over 80%

projects

Few-Shot Clustering Instance Segmentation (FS-CIS) Net
- Designed novel neural network architecture to perform proposal-free few-shot instance segmentation, showcasing results in graduate-level course, CS 6670: Computer Vision
- Validated approach on PASCAL-5i dataset, showing comparable performance to few-shot Mask R-CNN inspired methods with significant speedups at inference time

Continual Learning with Lottery Tickets
- Proposed and demonstrated effectiveness of novel training scheme to resist catastrophic forgetting, a phenomena in which a model overfits to the most recently seen data in a multi-task learning setting

Xi Compiler
- Wrote optimized compiler in Scala for language Xi, in team of 4; approximately 10,000 lines of code
- Includes lexing, parsing, type-checking, intermediate code generation, various optimizations including dataflow analysis, and emitting assembly instructions with non-trivial register allocation

languages & technologies
Python, Java, OCaml, Scala, Go, Julia, C/C++, Bash, JavaScript, SQL
PyTorch, Keras/TensorFlow, Git, Jupyter, Docker, Bazel, Gradle, Terraform