Yash Sharma

Curriculum Vitae

210 Lake Street, Ithaca
NY, 14850
+1 845 290 4694
yash.sharma200999@gmail.com
yash-s20.github.io
20 Sep 1999

MS CS student at Cornell University,
B.Tech in CSE from Indian Institute Technology Bombay

Education

2022– **Master of Science in Computer Science**, *(ongoing)*, from **Cornell University NY**
GPA: 4/4

2017–2021 **Bachelor of Technology in Computer Science & Engineering with Honors**,
Minor in AI & Data Science, from the **Indian Institute of Technology (IIT) Bombay**,
Department Rank 10
CGPA: Major 9.68/10, Honors 10/10, Minor 9.40/10

Work Experience

Sep 2021– Aug 2022 **Software Engineer**, *Samsung Electronics*, Suwon, South Korea,
Advanced S/W Lab, Network Business Unit
Part of the **P**hysical **U**plink **S**hared **CH**annel team of Samsung's 5G vRAN Development Lab, working on high performance virtualized physical layer for next-generation wireless communication, working on core-cycle optimization. Utilized intel intrinsic AVX instructions for efficient parallel processing of data. Reduced bottlenecks in uplink signal processing pipeline to achieve upto 20% speedup

Publications

Oct 2020 **Improving Low Resource Code-switched ASR using Augmented Code-switched TTS**,
Interspeech 2020, Shanghai, China (online)
Y.Sharma, B.Abraham, K.Taneja, P.Jyothi

Feb 2020 **WeLineation: crowdsourcing delineations for reliable ground truth estimation**,
SPIE 2020, Houston TX, USA
S. Goel, Y. Sharma (joint first author), ML. Jauer, TM. Deserno

Feb 2020 **STAPLE performance assessed on crowdsourced sclera segmentations**,
SPIE 2020, Houston TX, USA
ML. Jauer, S. Goel, Y. Sharma, TM. Deserno, M. Gijs, T. Berendshot, C. Bertens, R. Nuijts

Research Projects

Aug 2020 – June 2021 **Improving code-switched Automatic Speech Recognition**,
Undergraduate Thesis, IIT Bombay & Microsoft IDC,
Guides - Prof. Preethi Jyothi and Basil Abraham
Focused on improving performance of end-to-end ASR models on Gujarati-English speech by conditioning transformer layers on language ID of text in a per-layer supervised method. Proposed two methods of introducing language specific parameters and explainability in the multi-head attention mechanism, and implemented a Temporal Loss that helps maintain continuity in input alignment.

Fall 2020 **GCC Translation Validation**,
RnD Project, IIT Bombay, **Guide** - Prof. Amitabha Sanyal
Design of an automated debugger for a GCC plugin that formally detects bugs in or validates the translation of C programs to machine code. We shifted the plugin's support from G++ to GCC, and created a custom Intermediate Representation inspired from GIMPLE to better analyze translation of programs

Dec 2019 – June 2020 **Improving Low Resource Code-switched ASR using Augmented Code-switched TTS**, *RnD Project*, IIT Bombay & Microsoft IDC, **Guides** - Prof. Preethi Jyothi and Basil Abraham

Developed end-to-end based Automatic Speech Recognition models trained on Hindi and English monolingual data. Proposed techniques to leverage code-switched Text to Speech (TTS) to improve performance in low-resource settings. Ideated a new loss function to target underlying distributions of languages in the data. Used augmentation and encoder freezing to avoid over-fitting on synthetic artefacts.

Internship Experience

Summer 2020 **Network Automation & Kubernetes Service Load testing**, *Samsung Electronics, Korea*, Remote Intern

Involved understanding of **kubernetes** cluster (deployment, pods and services) and load balancing in detail. Built an automated testing framework using **Locust**, supporting various **Layer 4** & **Layer 7** protocols to evaluate performance of Samsung's load balancing tools.

Summer 2019 **WeLineation - Crowdsourcing and Consolidation of Medical Segmentation**, *Techniche Universität Braunschweig*, Research Intern, Guide - Prof. Thomas Deserno

Implemented a variant of **STAPLE**, an expectation maximization algorithm, with a custom **Markov Random Field** (MRF) prior to delineate ground-truth like segmentations from crowdsourced data. Performed a controlled study to test the system and rank user performance. Presented WeLineation in the 2020 SPIE Medical Imaging Conference at Houston, TX

Teaching Experience

Fall 2022 **TA**, *CS 3410 - Computer System Organization & Programming*, Cornell University, Course Instructor - Prof. Anne Bracy

Involves grading exams, spear-heading assignments, holding office hours and leading lab discussions

Fall 2020, Fall 2019 **TA**, *CS 251 - Software Systems Lab*, IIT Bombay, Course Instructor - Prof. Amitabha Sanyal

Orchestrating smooth running and preparing assignments for the lab course of "SSL" for second year students of CSE department, a fully online semester in 2020, and a in-person course in 2019. Granted honourable mention in 2019 and **best TA award** in 2020

Fall 2018 **TA**, *MA 105 - Calculus*, IIT Bombay, Course Instructors- Prof. Shripad Garge, Prof. Sourav Pal, Prof. Saurav Bhaumik

Took weekly sessions of 50 freshmen students from various departments, evaluated exam papers and volunteered to help beyond class hours

Technical Course Projects

Completed at Cornell University

Fall 2022 **Few-shot Compositional Action Recognition**,

Course - Advanced Topics in ML, Instructor - Prof. Kilian Weinberger

Ongoing project on leveraging decomposition of action into nouns and verbs in the few-shot setting

Completed at IIT Bombay

Spring 2021 **Open-Ended Reinforcement Learning**,

Course - Advances in Intelligent and Learning Agents, Instructor - Prof. Shivaram Kalyanakrishnan

Used Uber's POET algorithm to apply **evolutionary strategies** to solve increasingly complex mazes

Spring 2021 **Analysis of Negative Interference in Multilingual Models**,

Course - Advanced Machine Learning, Instructor - Prof. Sunita Sarawagi

Analysed negative interference and improved performance in various NLP tasks on the GLUECoS dataset using the proposed meta learning approach in [Wang et al, EMNLP 2020]

Spring 2021 **Low Resource Morphological Inflection**,

Course - Deep Learning in NLP, Instructor - Prof. Pushpak Bhattacharya

Implemented Low-Resource Morphological Inflection model in PyTorch, following [Anastasopoulos et al, ACL 2019]

Spring 2021 **n-thread Lamport algorithm on NuSMV**,
Course - Analysis of Concurrent Programs, Instructors - Prof. Ashutosh Gupta & Prof. Krishna S
Implemented the lamport algorithm on NuSMV using wraparound queues. Wrote a python script to generalize this NuSMV program to arbitrary number of threads

Spring 2021 **Parallel and Concurrent Programming in Haskell**,
Course - Design & Implementation of Functional Languages, Instructor - Prof. Amitabha Sanyal
Understanding the use of Parallel and Concurrent programming monadic interfaces provided in Haskell

Spring 2020 **Self Load-Balancing Server**,
Course - Virtualization and Cloud Computing, Instructor - Prof. Mythili Vutukuru
Made a server-manager using the libvirt API to manage multiple connections and failure due to timeouts

Fall 2019 **VQA - Inferring and executing programs**,
Course - AIML, Instructor - Prof. Ganesh Ramakrishnan
Used parallel forward propagation and hard parameter sharing to optimize existing architectures for Visual Question Answering without loss in performance

Fall 2019 **What's NE(x)T - a content-based music recommendation system**,
Course - Automatic Speech Recognition, Instructor - Prof. Preethi Jyothi
Implemented a recommendation system based on audio signals, feedback of user likings and bag-of-word lyrics

Fall 2018 **Secure Personal Cloud**,
Course - Software Systems Lab, Instructor - Prof. Soumen Chakrabarti
Constructed a 'zero-knowledge' cloud server and client with end-to-end encryption using AES, Triple DES and RC4 encryption techniques, following industrial standards, with keys stored locally.

Technical Skills

ML tools pyTorch, kaldi, keras, tensorflow

Languages C, C++, python, bash, Haskell, Racket, HTML/CSS, Javascript, Android, LaTeX, SQL, Java, Prolog, Answer Set Programming

Softwares Perforce & Swarm, Jira, Jenkins, GNU/Linux, Docker, Git, MATLAB, Android Studio, QTSpim

Courses Undertaken

Computer Science Topics in Computational Sustainability, Advanced Topics in ML, Theoretical ML, Automatic Speech Recognition, Concurrent Programming, Functional Programming, Deep Learning for NLP, Operating Systems, Computer Architecture, Computer Networks, Software Systems Lab

Mathematics Discrete Structures, Calculus, Linear Algebra, Differential Equations, Systems and Control

Others Environment Sciences, Psychology, Quantum Physics, Biology, Chemistry, Economics

Language Proficiency

English **Proficient**
Hindi **Proficient**
French **Intermediate** *8 years of curriculum learning*
Korean **Amateur** *Passed TOPIK-I Level 2*

Extracurricular

Winter 2018 Qualified for the final round of Microsoft AI Challenge

Spring 2018 Hosted speaker sessions and organized shows as an organizer in E-Summit, a two-day business event conducted by Entrepreneurship Cell, IIT Bombay

2015 Stood first in a Shakespearan themed Inter-School Dramatics competition

Oct 2013 Participated in the first Junior Model United Nations conference in Indus International School