Ketaki Joshi

ketaki.joshi@yale.edu
+1 203-906-4461
https://joshi-ketaki.github.io/

Education

- **Yale University** New Haven, USA
  Ph.D Computer Science August 2019 - Present

  M.Phil in Computer Science August 2019 - 2021
  – **Thesis:** "Single Source Code, Hardware Agnostic Heterogeneous Systems."

  Masters in Computer Science August 2019 - 2020
  – **Thesis:** "Detecting Computational Clones in Brain Models."

- **University of Pune** Pune, India
  Bachelor's in Computer Engineering 2011 - 2015
  Institute Rank : 1/200, University Rank : 5/9000
  – **Thesis:** "OptGen : A Custom Compiler Optimization Generator."

Professional and Research Experience

- **Yale University** New Haven, USA
  Ph.D candidate(Mentor: Guilherme Cox) August 2019 - Present

  – Mitigating Catastrophic Forgetting in Online Machine Learning using Brain-Inspired Techniques.

  * Developed a technique inspired by cognitive science to reduce catastrophic forgetting in online machine learning applications.
  * Quantified the impact of this strategy on recently developed machine learning models for memory access prefetching, a significant issue in computer system optimization.
  * Achieved a speedup of 1.6 times compared to existing regularization techniques to mitigate catastrophic forgetting. Also, reduced external storage by a factor of 16 compared to existing replay techniques in initial prototype evaluations.

  – Prefetching using Principles of Cognitive Systems.

  * Explored the application of brain-inspired, machine learning techniques to enhance memory management in tiered memory systems.
  * Identified the presence of replay in the GPU memory management model. This can help address fundamental issues in online learning, such as catastrophic forgetting, without the need for external implementation.

  – Leveraging Computer Systems to Better Understand Cognitive Systems.

  * Developed a code clone detection tool to identify parts of cognitive models that are computationally similar.
  * This enables neuroscientists to reuse, construct, and comprehend cognitive models.

  Teaching Fellow for CPSC 323: Introduction to Systems Programming Spring '21, Spring '22, Fall '22

  – Designed and deployed a toy compiler to introduce students to compiler optimizations and assembly code generation, using a custom-designed toy programming language and a subset of x86 ISA.

Nvidia Santa Clara, USA
Unified Virtual Memory Intern(Mentor: Guilherme Cox) June 2022 - August 2022

- Identified the root cause, proposed and delivered a new data eviction policy to improve memory management in tiered memories.
- The improved policy delivers two orders of magnitude performance improvement and enables efficient execution of emerging applications.

Architecture Research Intern(Mentor: Daniel Lustig) June 2021 - October 2021

- Identified and developed a prototype to execute traditional GPU applications on CPU SIMD units to reduce data transfer costs across tiered memory systems without compromising programmability.
- Achieved speedup by a factor of 1.5 compared to pure CPU-SIMD execution and a sppedup by a factor of 1.8 compared to a pure GPU execution.

Nvidia Pune, India

GPU Compiler Developer January 2017 - August 2019

- Delivered compiler frontend and backend interface design, development and code generation for the deep learning matrix operations. The instructions were exposed in CUDA 10.0 and CUDA 10.1. (MMA etc.)
- Delivered the entire assembly code generation and decoding for Turing architecture.
- Led analysis of key benchmarks to identify opportunities within the compiler for a newly introduced architecture feature of uniform register files.
- Designed and deployed a framework to auto-generate assembly and decoding of assembly instructions for the compiler.

Tools Developer June 2015 - July 2016

- Identified the root cause and developed a no-reference image analysis tool to identify artifacts in images rendered across a different of GPU architectures.
- Achieved an accuracy of 98% and eliminated the time-consuming manual analysis.
- The work was submitted to Nvidia internal technical conference and an Invention Submission Form was filed.

Shoreline IoT Pune, India

Member of Technical Staff September 2016 - January 2017

- Led the prototype development of a remote monitoring device for industrial boilers.
- Collaborated with both the design and technical teams to ensure the delivery of a reliable, effective solution which significantly improved the efficiency and safety of boiler operations, leading to substantial cost savings and decreased downtime.

Indian Institute of Technology, Bombay Mumbai, India

Undergraduate Thesis Intern May 2014 - June 2015

- Developed a sophisticated tool that, when provided with data flow equations, can generate optimization passes. These optimization passes are then seamlessly integrated into the compiler. This offers programmability and simplicity, making it easier to experiment with and develop new compiler optimizations.

Indian Institute of Tropical Meteorology Pune, India

Junior Year Intern February 2014 - June 2014

- Deployed a sophisticated N-SAT solver to be utilized in weather prediction analysis modules. This solver plays a critical role in enhancing the accuracy and efficiency of these analysis and decision-making modules, improving the overall performance of the weather prediction system.

Skills, Awards, and Service

- **Domain Tools and Skills:**
  - **Programming Languages, Frameworks and Source Control:** C, Python, x86 Assembly, Nvidia PTX Assembly, CUDA, C++, MATLAB, Octave, Pytorch, Raytune, Git, PerForce.
  - **Writing and Presentation Tools:** LaTeX, MS Office, Libre Office
  - **Writing and Presentation Courses:** Academic Writing (Yale Course: CPSC 992), Great Presentations (Yale Course: CPSC 993).

- **Awards :**
  - **Ackerman Faculty Award Committee 2022:** Selected to be the sole representative of the entire Computer Science Graduate Community.
  - **Travel Grants:** ISCA 2023, ISCA 2022, MICRO 2020, CRA Women Graduates Meet 2020.
  - **Professional Recognition:** Recognized as a "Strong Contributor" for work done during full-time employment at Nvidia.
  - **Persistent Systems Gold Medal:** Gold Medal for being the 2014 department topper in junior year.
- **Service:**
  - **Yale Society for Women Engineers Board:** Currently serving as the Communications Chair.
  - **Committee Member on 2022 Ackerman Faculty Award:** Participated as the committee member in evaluating, debating, and finalizing the winning candidate.
  - **Nvidia Women in Technology Board Member:** Served on the board in the India chapter.
- **Mentorship:**
  - Mentored two women Yale sophomores for navigating internships, career, and academic challenges under the Women in Sciences at Yale (WISAY) mentorship initiative.

Publications and Presentations:

- **Ketaki Joshi**, Raghavendra Pradyumna Pothukuchi, Andre Wibisono, Abhishek Bhattacharjee "Mitigating Catastrophic Forgetting in Long Short-Term Memory Networks.", arXiv:2305.17244 [cs.LG].
- Wu Michael, **Joshi Ketaki**, Sheinberg Andrew Cox Guilherme, Khandelwal Anurag, Pothukuchi Raghavendra Pradyumna, Bhattacharjee Abhishek, "Prefetching Using Principles of Hippocampal-Neocortical Interaction.", HOTOS '23.
- **Ketaki Joshi**, Guilherme Cox, Jan Vesely "Access Guided Eviction for Unified Virtual Memory.", Summer'22 Internship Talk at Nvidia.
- **Ketaki Joshi**, "Single Source, Hardware Agnostic Heterogenous Systems.", Fall'21 Candidacy Exam Talk at Yale University.
- **Ketaki Joshi**, Daniel Lustig, Oreste Villa, "CUDA Task launcher for GPU and CPU SIMD units.", Fall'21 Internship Talk at Nvidia Research.
- J. Veselý, R. P. Pothukuchi, **K. Joshi**, S. Gupta, J. D. Cohen and A. Bhattacharjee, "Distill: Domain-Specific Compilation for Cognitive Models.", 2022 IEEE/ACM International Symposium on Code Generation and Optimization (CGO), 2022, pp. 301-312, doi: 10.1109/CGO53902.2022.9741278.
- J. Veselý, R. P. Pothukuchi, **K. Joshi**, S. Gupta, J. D. Cohen, and A. Bhattacharjee, "Cognac: Domain-Specific Compilation for Cognitive Models."
- **Ketaki Joshi**, Uday Khedker "A Custom Compiler Optimization Pass Generator.", Fall' 15 guest speaker at ACM-W chapter of Cummins College of Engineering for Women, Pune University.

References

Available on request.