Ketaki Joshi

ketaki.joshi@yale.edu
+1 203-906-4461
Personal Website

Education

- **Yale University** — New Haven, USA
  Ph.D Computer Science (Advisor: Abhishek Bhattacharjee) — 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."

Research and Teaching Experience

- **Yale University** — New Haven, USA
  Ph.D candidate(Advisor: Abhishek Bhattacharjee) — August 2019 - Present

  – Mitigating catastrophic forgetting using principles of Episodic Memory (Current).
    * Developed a cognitively inspired technique to mitigate catastrophic forgetting continual learning applications.
    * Quantifying impact on recently developed LSTM networks for memory access prefetching - an important problem in computer systems optimization.
    * Achieved speedup by factor of 1.6 compared to existing regularization technique and reduced external storage by factor of 16 than existing replay implementation in initial prototype evaluations.
  – Prefetching using principles of Complementary Learning System (CLS).
    * Investigated use of CLS-inspired techniques to improve GPU memory management.
    * Identified and investigated the existence of natural replay in GPUs. This would potentially remove the need to implement explicit replay techniques to avoid catastrophic forgetting.
  – Enhance cognitive systems research using algorithmic principles.
    * Developed a similarity detection technique to identify computational similarity between cognitive models.
    * This tool helps neuroscientists reuse, build and understand cognitive models.
  – Developed asynchronous system calls in CertiKOS (a certified Operating System) which were two orders of magnitude faster than existing synchronous system calls.

  Teaching Fellow for CPSC 323: Introduction to Systems Programming — Spring '21, Spring '22, Fall '22
  – Responsibilities include designing instructional materials, and assignments, proctoring, and grading exams.
  – Designed a toy compiler to introduce students to local, global, and peephole optimizations, register allocation, and assembly code generation.

- **Indian Institute of Technology, Bombay** — Mumbai, India
  Undergraduate Thesis Intern (Advisor: Uday Khedker) — May 2014 - June 2015
  – Developed a tool that given data flow equations, generates optimization passes, and integrates them in the compiler.

- **Indian Institute of Tropical Meteorology** Pune, India
  Undergraduate Research Intern (Mentor: Narendra Karamarkar) February 2014 - June 2014
  - Deployed a N-SAT solver as a tool for use in weather prediction analysis modules.

Professional Experience

- **Nvidia** Santa Clara, USA
  Unified Virtual Memory Intern(Mentor: Guilherme Cox) June 2022 - August 2022
  - Deployed an access-aware eviction algorithm that enabled support for irregular access patterns along with traditional streaming patterns.
  - Achieved a performance improvement of two orders of magnitude in the best case and same as the existing algorithm in the worst case.

  Architecture Research Intern(Mentor: Daniel Lustig) June 2021 - October 2021
  - Delivered a driver shim to enable execution of traditionally GPU-targeted applications on any available hardware substrate. This was done transparently without compromising on the programmability of the application developer. Demonstrated this by enabling CUDA programs to execute on CPU SIMD units along with GPUs.
  - Achieved speedup of a factor of 1.5 compared to pure CPU-SIMD execution and 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, support and assembly 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 generation and decoding for Turing architecture.
  - Led analysis of key benchmarks to identify opportunities for using newly introduced uniform register file within the compiler.
  - Led the design and implementation of a framework to auto generate assembly and decoding of assembly instructions for the compiler.

  Tools Developer June 2015 - July 2016
  - Developed a no-reference image analysis tool to detect artefacts in images rendered across different GPU architectures. Eradicated existing manual analysis and achieved 98% accuracy.
    Submitted to Nvidia internal conference and filed an Invention Submission Form

- **Shoreline IoT** Pune, India
  Member of Technical Staff September 2016 - January 2017
  - Led prototype development of a device for remote maintenance of IoT systems.

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 Tools:** Latex, Word.
  - **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 at Nvidia for Turing Architecture Compiler Development in August 2019.
  - **Persistent Systems Gold Medal:** Gold Medal for being the department topper in junior year in 2014.
- **Service:**
  - **Yale Society for Women Engineers Board Member:** 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

- Abhishek Bhattacharjee [abhishek@cs.yale.edu]: Professor of Computer Science, Yale University.
- Guilherme Cox [gcox@nvidia.com]: Software Engineer - Unified Virtual Memory Group, Nvidia