Sukriti Paul

https://www.sukritipaul.in

Email: sukriti5@umd.edu
Phone: 240-789-7157

Education

- **University of Maryland** — College Park, USA
  PhD Student in Computer Science — *Jan 2025 – Present*
- **University of Maryland** — College Park, USA
  Master of Science in Computer Science; GPA: 4.0 — *Aug 2023 – May 2025*

Publications

- V. Singla, K. Yue, **S. Paul**, R. Shirkavand, M. Jayawardhana, A. Ganjdanesh, H. Huang, A. Bhatele, G. Somepalli, T. Goldstein; From Pixels to Prose: A Large Dataset of Dense Image Captions *arXiv:2406.10328, 2024*
- **S. Paul**, H. Devi, C. Seelamantula, VR. Mujeeb, AS. Prasad; Fully-automated semantic segmentation of wireless capsule endoscopy abnormalities *IEEE 17th International Symposium on Biomedical Imaging (ISBI), 2020*

Experience

- **University of Maryland, College Park** — College Park, MD
  Research Graduate Assistant - Advisor: Prof. Tom Goldstein — *Jun 2024 - Aug 2024*
  - **PixelProse Dataset**: Core contributor in building a 16M+ image dataset with high-quality synthetic captions for training Idefics3. Curated 50k subset for diffusion alignment.
  - **Autoregressive Image Generation**: Benchmarked SOTA image tokenizers (COSMOS, EMU3, Janus, OpenMagViT2) across diverse dataset categories using quality and throughput metrics. Scaled PixelProse tokenization with efficient latent caching and trained via LLamaGen.
  - **Diffusion Models**: Led fine-tuning of 'SD3 Medium' on PixelProse, achieving 52% throughput increase via optimized caching, data sharding, and model parallelism.
- **NonExomics** — Boston, MA (Remote)
  Founding Genome Data Scientist- Advisor and CEO: Prof. Sudhakaran Prabakaran — *Jun 2021 - Jun 2023*
  - **Protein Structure Prediction**: Predicted structures of 250K novel proteins using five state-of-the-art algorithms. Developed proprietary prediction algorithm with optimized inference pipeline.
  - **ML for Genomics**: Developed ML techniques to study mutations on novel proteins from the 'Dark Genome' and protein evolution. Built a GNN to predict disease-protein-drug interactions.
  - **Impact**: Helped shortlist 99 drug targets and establish partnerships with Illumina Accelerator, New York Genome Center, and AWS Life Sciences.
- **American Express** — Bengaluru, India
  Business Analyst-2, Merchant Recommender System Team — *Dec 2019 - Oct 2020*
  - **Feature Engineering**: Engineered 120 new features and rationalized 543 model features, boosting customer engagement by 6.3% and general spend by 4.5%.
  - **Recommender System**: Optimized hybrid Collaborative Filtering model with XGBoost, increasing monthly engagement by 3.1% across 9 industries.
  - **Large-Scale Analysis**: Conducted analyses for 2M+ card users, implementing 'Central Biller' Logic and adjusting merchant suppression for pandemic trends.
- **Indian Institute of Science, Spectrum Lab** — Bengaluru, India
  Research Assistant - PI: Prof. Chandra Sekhar Seelamantula — *Oct 2018 - Nov 2019*
  - **Medical Imaging**: Designed an encoder-decoder network for semantic segmentation of 9 Wireless Capsule Endoscopy lesions with collaboration with Command Hospital Air Force, Bangalore. Work published in ISBI conference.
  - **AI Diagnostic System**: Developed a prototype AI-powered diagnostic web application for real-time detection of WCE abnormalities, reducing screening time from 4 hours to minutes. Project was awarded a grant by the **Bill & Melinda Gates Foundation** via the Global Grand Challenges 2020.

Projects

- **Post Training Quantization of Image Tokenizers**: Investigated post-training quantization of tokenizers using logarithmic and per-tensor techniques. Discovered asymmetric resilience where decoders function effectively at lower bit precision while encoders require higher bits.
- **Steerable Fast Bilateral Edge Detectors**: Novel noise-robust algorithm for color images, reducing runtime by 3x for real-time sidewalk detection in a government-funded project.

Skills & Awards

- **Skills**: Deep Learning, Computer Vision, PyTorch, TensorFlow, Python, C++, AWS, Docker
- **Awards**: **Google CSRMP Scholar** (2023), ACM Women Best Officer Award (2018), Google APAC Women Techmakers Scholar (2017)