Curriculum Vitae - Aditya Sarkar

PhD Student · University of Maryland

+1 (858) 518-3066 asarkar6@umd.edu kingston-aditya kingston-aditya aditya-sarkar

Research Interests

Video Understanding Action recognition

Multimodal learning Visual (images, videos) · Textual (natural language) · Audio (speech, music)

Education

University of Maryland College Park — College Park, MD, US

PhD in Computer Engineering — 2024 – 2028

- **Specialization**: Multimodal Learning
- Joined UMIACS in 2024, conducting research in computer vision
- **Advisor**: Prof. Ang Li, **GPA**: 4.00/4

Indian Institute of Technology (IIT) Mandi — India

BTech in Electrical Engineering (First Class Honors) — 2019 – 2023

- **Specialization**: Computer Networks and Protocols
- President of India Gold Medal Awardee in 2023
- **Thesis**: Analytical study of IEEE 802.11bd protocol
- **Advisor**: Sreelakshmi Manjunath, **Major GPA**: 3.98/4

Experience

Institute for Advanced Computer Studies, University of Maryland — College Park, MD

Graduate Research Assistant — Jan. '25 – Present

- **Collaborators**: David Jacobs, Shlok Mishra (Meta)
- Explored various multimodal models to identify the root causes of their issues and addressed them using a range of targeted approaches.

Statistical Visual Computing Lab, University of California — San Diego, CA

Student Researcher — Sept. '23 – Nov. '24

- **PI**: Nuno Vasconcelos
- Proposed a training-free pipeline to detect misalignment between image and text. Used it for a variety of selective predictions tasks based on taxons for captioning, classification and image-text matching.

Machine Learning and Genomics Lab, University of California — Los Angeles, CA

Student Researcher — July '20 – May '22

- **PI**: Serghei Mangul, Eleazar Eskin.
- Trained a simple yet effective attention based pipeline for deletions detection in SV Callers trained on the UCLA BIG genome dataset, outperforming SOTA algorithms. Published 2 papers in Genome Biology (IF 17.9) and Briefings in Bioinformatics (IF 13.99).

Publications

Memory Augmented Plug-and-Play Selective Prediction — Sep '24 - Jan '25

Aditya Sarkar, Yi Li, Jiacheng Cheng, Shlok Kumar, Nuno Vasconcelos — *Under Review in ICLR '26*

- Our hypothesis was that similar samples to query image-text pair can help calibrate and reduce variance of the CLIP score. If true, this score can be used as an efficient selective predictor.
- Proved the hypothesis by proposing a training-free selective predictor that relies on external retrieval dataset to detect misalignment in predictions of large multimodal and representation models.
- **Applications.** Improved performance for not only closed set but also open set selective prediction.

The Spectrum of Temporal Understanding — Aug '25 - Oct '25

Aditya Sarkar, Y. Li, J. Cheng, S. Mishra, D. Jacobs, N. Vasconcelos — *Under review in CVPR '26*

- We hypothesized that temporal understanding exists as a layered spectrum *ie.* each layer represents an independent factor that determines complexity of a video-question pair and they exist as a spectrum.
- We demonstrated this by creating a dataset of synthetic videos with controlled variations in specific factors. Evaluated Video-LMMs on these variants to show variance in performance - it performs well on some variants but poorly on others. Showed this for real videos using similar approaches.
- **Applications.** Used it for tagging existing benchmarks and videoLMMs with their complexity and specialization, thereby reducing computation in benchmarking.

Image Generation From Interleaved Image-Text Prompt Jan '25 - Present
Aditya Sarkar, X. Pan, S.N. Sai, S. Ghosh, S. Mishra, A. Singh, D. Jacobs *To be submitted in ICML '26*

- Developed a pipeline to curate a TI2I benchmark dataset from the standard T2I-CompBench dataset. The objects in their text prompts were aligned to images from classification datasets to create interleaved image-text prompts.
- Proposed a novel reward function based on object detection and matching them with DINOv2 score, and used it to finetune the Qwen-Image Edit model with Flow-GRPO.

Analyzing cross-modal miscalibration of Multimodal Models Oct '25 - Present
Aditya Sarkar, S. He, S. Mishra, D. Jacobs, A. Li *Ongoing*

- We hypothesize that a significant portion of CLIP's architecture is dedicated to object recognition, while only a small subset is responsible for understanding relationships between objects.
- To test this, we aim to identify the minimal set of modules involved in relational understanding using interpretability techniques such as attention patching and attribution, and then fine-tune only those layers using DPO to improve CLIP's performance on WinoGround and What'sUp benchmarks.

Honors & Awards

- **Graduate fellowship** at University of Maryland in 2024.
- **Walmart Predoc Fellowship**, Department of CSA, IISc Bangalore in 2023.
- **Graduate scholarship**, Department of CS, UCLA in 2022.
- **National Talent Search Scholarship**, Government of India in 2017.

Technical Skills

Programming Languages	Python, MATLAB, C++, HTML/CSS, Bash/Unix, Git.
Deep Learning	PyTorch, Tensorflow, Accelerate, Distributed Training.
Languages	English, Japanese

Academic Service

Peer reviewer	ICCV '25
Teaching Assistant	ENEE222, ENEE290, CMSC415