Arjun Somayazulu

Austin, TX | ars7452@my.utexas.edu | arjunrs1.github.io | linkedin.com/arjunsomayazulu

Education

The University of Texas at Austin, Ph.D. in Computer Science (Advisor: Prof. Kristen Grauman) 2022 – Present

- **Area:** Computer Vision, Multi-modal Video Understanding (Vision, Language, Audio, 3D Pose)
- **Coursework:** Natural Language Processing, Visual Recognition, Robot Learning, Spoken Language Technologies, Distributed Computing, Algorithms
- **Activities:** Board member – Graduate Representative Association of Computer Sciences (GRACS)

Johns Hopkins University, BS in Biomedical Engineering, Computer Science 2018 – 2022

- **Honors:** Richard J. Johns Award, Tau Beta Pi (Engineering Honor Society), Dean's List (every semester)

Experience

Research Intern, Meta FAIR – New York, NY May 2024 – Jan 2025

- Worked on project developing a generalizable method to improve performance of video understanding models on video with severe action and/or object occlusions, with Profs. Kristen Grauman and Lorenzo Torresani.

Publications

Learning Activity View-invariance Under Extreme Viewpoint Changes via Curriculum Knowledge Distillation 2025
Arjun Somayazulu, Effrosyni Mavroudi, Changan Chen, Lorenzo Torresani, Kristen Grauman
arXiv 2025 (Under submission)

– Proposed a method to make vision-language models (VLMs) robust to severe object and action occlusions present in activity video captured from challenging viewpoints. Outperforms SOTA temporal sentence grounding methods by 8% on video with severe occlusions.

ActiveRIR: Active Audio-Visual Exploration for Acoustic Environment Modeling 2024
Arjun Somayazulu, Sagnik Majumder, Changan Chen, Kristen Grauman
IROS 2024 (Oral)

– Introduced an embodied active navigation method that efficiently explores an indoor environment and actively collects samples to build a high-quality spatial-acoustic map within real world time and sampling constraints. Achieve higher quality spatial-acoustic maps than SOTA methods using >70% fewer agent steps.

Ego-Exo4D: Understanding Skilled Human Activity from First- and Third-Person Perspectives 2024
Kristen Grauman *et al.*, including **Arjun Somayazulu**
CVPR 2024 (Oral)

– Introduced a large-scale skill understanding dataset with ego- and exocentric views. Led development of baseline models for the novel 'Demonstrator Proficiency' skill estimation task and its corresponding benchmark. Paper cited 260+ times to date.

Self-Supervised Visual Acoustic Matching 2023
Arjun Somayazulu, Changan Chen, Kristen Grauman
NeurIPS 2023

– Proposed a model that edits speech acoustics to match a target visual scene, learned from uncurated web-video without supervision. Achieve >26% reduction in acoustic error (RTE) compared to SOTA supervised approaches relying on smaller, curated datasets.

Teaching/Service

- **Reviewer:** CVPR 2024, CVPR 2025, IROS 2024, NeurIPS 2025, CVIU 2025, IJCV 2025, CVPR 2026
- **Teaching Assistant:** EN.601.226 Data Structures (Spring 2021), EN.601.675 Machine Learning (Fall 2021)