

SOUNAK MONDAL

📧 sounakcs.github.io ✉ somondal@cs.stonybrook.edu 🔗 LinkedIn 🎓 Google Scholar

Research Interests

I am interested in research on **multimodal learning**, particularly **vision-language modeling**. My PhD thesis focuses on using **vision-language representation learning** and **multimodal foundation models (e.g., multimodal LLMs)** for modeling human visual attention (eye gaze).

Education

Stony Brook University, Stony Brook, NY <i>Doctor of Philosophy, Computer Science</i> Advised by Minh Hoai Nguyen, Dimitris Samaras, Gregory Zelinsky	August, 2020 – Present GPA: 4.0/4.0
Jadavpur University, Kolkata, India <i>Bachelor of Engineering, Computer Science & Engineering</i>	August, 2013 – May, 2017 GPA(Absolute Grading): 8.65/10.00, Rank: 3/54

Selected Publications

- Generative Gaze Decoding using Multimodal LLMs**
Sounak Mondal, D. Samaras, G. Zelinsky, M. Hoai *NeurIPS 2025 (under submission)*
- Gaze-Language Alignment for Zero-Shot Prediction of Visual Search Targets from Human Gaze Scanpaths**
Sounak Mondal, N. Sendhilnathan, T. Zhang, Y. Liu, M. Proulx, M.L. Iuzzolino, C. Qin, T.R. Jonker *ICCV 2025*
- Few-shot Personalized Scanpath Prediction**
R. Xue, J. Xu, Sounak Mondal, H. Le, G. Zelinsky, M. Hoai, D. Samaras *CVPR 2025*
- Look Hear: Gaze Prediction for Speech-directed Human Attention**
Sounak Mondal, S. Ahn, Z. Yang, N. Balasubramanian, D. Samaras, G. Zelinsky, M. Hoai *ECCV 2024*
- Diffusion-Refined VQA Annotations for Semi-Supervised Gaze Following**
Q. Miao, A. Graikos, J. Zhang, Sounak Mondal, M. Hoai, D. Samaras *ECCV 2024*
- Unifying Top-down and Bottom-up Scanpath Prediction using Transformers**
Z. Yang, Sounak Mondal, S. Ahn, R. Xue, G. Zelinsky, M. Hoai, D. Samaras *CVPR 2024*
- Gazeformer: Scalable, Effective and Fast Prediction of Goal-Directed Human Attention**
Sounak Mondal, Z. Yang, S. Ahn, D. Samaras, G. Zelinsky, M. Hoai *CVPR 2023*
- Target-absent Human Attention**
Z. Yang, Sounak Mondal, S. Ahn, G. Zelinsky, M. Hoai, D. Samaras *ECCV 2022*
- Characterizing Target-absent Human Attention**
Y. Chen, Z. Yang, S. Chakraborty, Sounak Mondal, S. Ahn, D. Samaras, M. Hoai, G. Zelinsky *CVPRW 2022*

Experience

Research Scientist Intern Meta Reality Labs Research, Burlingame, CA Self-Supervised Learning for Vision and Sensor Data via Multimodal LLMs	June, 2025 – Present
Research Scientist Intern Meta Reality Labs Research, Redmond, WA Vision-Language Modeling of Eye Gaze Behavior (paper accepted at ICCV'25); Part-Time Student Researcher from 10/2024	June, 2024 – December, 2024
Research Intern UII America, Cambridge, MA Vision-Language (Multimodal) Modeling - Scene Graph Generation from captions using Large Language Models (LLMs)	May, 2023 – August, 2023
Graduate Researcher CV Lab, Stony Brook University, NY Vision-Language Modeling for gaze prediction (CVPR'23, ECCV'24), and gaze decoding (ICCV'25, NeurIPS'25 submission). Also worked on gaze prediction models for visual search (ECCV'22, CVPR'24, CVPR'25), and gaze estimation (ECCV'24).	November, 2020 – Present
NLP Engineer Samsung Research Institute, Bangalore Enhancements of Bixby digital assistant: (1) Low resource intent classification via transfer learning, (2) Sequence labeling for Named Entity Recognition and speech end-point detection, (3) Lightweight and fast text classification architecture (ICSC'20)	June, 2017 – August, 2020
Summer Intern Samsung Research Institute, Bangalore Context awareness in SVoice platform for Natural Language Processing	May, 2016 – July, 2016
Undergraduate Researcher Indian Statistical Institute, Kolkata Video Action Recognition/Detection (ICAPR'17, ICVGIP Workshop'16) advised by Sanjoy Kumar Saha, Bhabatosh Chanda	July, 2015 – June, 2017

Academic & Technical Details

Graduate Courses: Computer Vision, Natural Language Processing, Robotics, Machine Learning, Database Systems
Languages & Frameworks: Python, C++, C, Java, PyTorch, TensorFlow, Hadoop (familiar), Spark (familiar)
Voluntary Service: Reviewer for CVPR, ICCV, NeurIPS, TPAMI