Sounak Mondal

sounakcs.github.io | somondal@cs.stonybrook.edu | LinkedIn | Google Scholar

Research Interests

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

Education

Stony Brook University, *Stony Brook, NY*
Doctor of Philosophy, Computer Science ***(GPA: 4.0/4.0)*** *Aug 2020 – Present*

Advisors: Minh Hoai Nguyen, Dimitris Samaras, Gregory Zelinsky

Jadavpur University, *Kolkata, India*
Bachelor of Engineering, Computer Science & Engineering ***(GPA: 8.65/10, Rank: 3/54)*** *Aug 2013 – May 2017*

Experience

Meta

Research Scientist Intern, *Reality Labs Research, Burlingame, CA* *Jun 2025 – Dec 2025*
- Modeling Vision and Sensor Data via Multimodal LLMs.

Research Scientist Intern, *Reality Labs Research, Redmond, WA* *Jun 2024 – Dec 2024*
- Vision-Language Modeling of Eye Gaze Behavior (published at *ICCV 2025*, patent filed)

CV Lab, Stony Brook University

Graduate Researcher, *Stony Brook, NY* *Nov 2020 – Present*
- Vision-Language Modeling for gaze prediction (CVPR'23, ECCV'24), and gaze decoding (ICCV'25 accepted + 1 under review)
- Gaze prediction modeling for search (ECCV'22, CVPR'24, CVPR'25); multimodal modeling for gaze estimation (ECCV'24)
- Modeling attentional states, e.g., mind-wandering vs. sustained attention, from eye gaze during online learning (ongoing)

Samsung

NLP Engineer, *Natural Language Processing Team, Bangalore, India* *Jun 2017 – Aug 2020*
- Enhancements of Bixby virtual assistant.
- Low resource intent classification via transfer learning
- Sequence labeling models for Named Entity Recognition and speech end-point detection
- Lightweight and fast text classification architecture (ICSC'20)

Summer Intern, *SVoice Team, Bangalore, India* *May 2016 – Jul 2016*
- Context awareness in SVoice platform for Natural Language Processing

UII America

Research Intern, *Computer Vision, Cambridge, MA* *May 2023 – Aug 2023*
- Vision-Language Modeling: Scene Graph Generation from captions using Large Language Models (LLMs).

Indian Statistical Institute

Undergraduate Researcher, *Kolkata, India* *Jul 2015 – Jun 2017*
- Video Action Recognition/Detection (ICAPR'17, ICVGIP Workshop'16).

Selected Publications

Generative Gaze Decoding via Multimodal LLMs **Under review**
Sounak Mondal, D. Samaras, G. Zelinsky, M. Hoai

Gaze-Language Alignment for Zero-Shot Prediction of Visual Search Targets from Human Gaze Scanpaths **ICCV 2025**
Sounak Mondal, N. Sendhilnathan, T. Zhang, Y. Liu, M. Proulx, M.L. Iuzzolino, C. Qin, T.R. Jonker

Look Hear: Gaze Prediction for Speech-directed Human Attention **ECCV 2024**
Sounak Mondal, S. Ahn, Z. Yang, N. Balasubramanian, D. Samaras, G. Zelinsky, M. Hoai

Gazeformer: Scalable, Effective and Fast Prediction of Goal-Directed Human Attention **CVPR 2023**
Sounak Mondal, Z. Yang, S. Ahn, D. Samaras, G. Zelinsky, M. Hoai

Few-shot Personalized Scanpath Prediction **CVPR 2025**
R. Xue, J. Xu, **Sounak Mondal**, H. Le, G. Zelinsky, M. Hoai, D. Samaras

Unifying Top-down and Bottom-up Scanpath Prediction using Transformers **CVPR 2024**
Z. Yang, **Sounak Mondal**, S. Ahn, R. Xue, G. Zelinsky, M. Hoai, D. Samaras

Target-absent Human Attention **ECCV 2022**
Z. Yang, **Sounak Mondal**, S. Ahn, G. Zelinsky, M. Hoai, D. Samaras

Diffusion-Refined VQA Annotations for Semi-Supervised Gaze Following **ECCV 2024**
Q. Miao, A. Graikos, J. Zhang, **Sounak Mondal**, M. Hoai, D. Samaras

Characterizing Target-absent Human Attention **CVPRW 2022**
Y. Chen, Z. Yang, S. Chakraborty, **Sounak Mondal**, S. Ahn, D. Samaras, M. Hoai, G. Zelinsky

Selected Awards

- ICCV 2025 Doctoral Consortium
- CVPR 2023 Travel Grant Award
- ACM Best B.E. Thesis Award, 2017

Academic & Technical Details

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