





SOUNAK MONDAL

 [sounakcs.github.io](https://github.com/sounakcs)  somondal@cs.stonybrook.edu  LinkedIn  Google Scholar

Education

Stony Brook University, Stony Brook, NY

August, 2020 – Present

Doctor of Philosophy, Computer Science

GPA: 4.0/4.0

Advised by Minh Hoai Nguyen, Dimitris Samaras, Gregory Zelinsky. I also collaborate with Niranjan Balasubramanian.

Jadavpur University, Kolkata, India

August, 2013 – May, 2017

Bachelor of Engineering, Computer Science & Engineering

GPA (Absolute Grading): 8.65/10.00, Rank: 3/54

Publications

- Where to Look: Gaze Prediction for Multimodal Visual Search** *Under Review*
Sounak Mondal, S. Ahn, Z. Yang, N. Balasubramanian, D. Samaras, G. Zelinsky, M. Hoai
- Diffusion-Refined VQA Annotations for Semi-Supervised Gaze Following** *Under Review*
Q. Miao, A. Graikos, J. Zhang, Sounak Mondal, 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
- 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
- Target-absent Human Attention** *ECCV 2022*
Z. Yang, Sounak Mondal, S. Ahn, G. Zelinsky, M. Hoai, D. Samaras
- Characterizing Target-absent Human Attention** *CVPR Workshop 2022*
Y. Chen, Z. Yang, S. Chakraborty, Sounak Mondal, S. Ahn, D. Samaras, M. Hoai, G. Zelinsky
- ICAN: Introspective Convolutional Attention Network for Semantic Text Classification** *ICSC 2020*
Sounak Mondal, S. Modi, S. Garg, D. Das, S. Mukherjee
- Violent/Non-Violent Video Classification based on Deep Neural Network** *ICAPR 2017*
Sounak Mondal, S. Pal, S.K. Saha, B. Chanda
- A Beta Distribution Based Novel Scheme for Detection of Changes in Crowd Motion** *ICVGIP(WS) 2016*
S. Pal, Sounak Mondal, S.K. Saha, B. Chanda

Experience

Research Intern | UII America, Cambridge, MA

May, 2023 – August, 2023

Vision-Language (Multimodal) Modeling - Scene Graph Generation from captions using Large Language Models (LLMs)

Graduate Researcher | CV Lab, Stony Brook University, NY

November, 2020 – Present

- Generative Modeling of multimodal perception using Multimodal LLMs and diffusion models (ongoing)
- Human Attention (Eye Gaze) Prediction for multimodal (vision-language) tasks (CVPR'23, one paper under submission)
- Human Attention (Eye Gaze) Prediction for visual search and free-viewing (ECCV'22, CVPRW'22, CVPR'24)
- Human Gaze Estimation using VQA and diffusion models (one paper under submission)

NLP Engineer | Samsung Research Institute, Bangalore

June, 2017 – August, 2020

- Deep learning-based enhancements for Bixby digital assistant, such as (1) low resource Intent Classification via transfer learning, (2) Sequence Labeling for Named Entity Recognition and speech end-point detection
- Proposed a lightweight and fast text classification architecture (ICSC'20)

Summer Intern | Samsung Research Institute, Bangalore

May, 2016 – July, 2016

Context awareness in SVoice platform for Natural Language Processing

Undergraduate Researcher | Indian Statistical Institute, Kolkata

July, 2015 – June, 2017

Video Action Recognition/Detection (ICAPR'17, ICVGIP Workshop'16) advised by Sanjoy Kumar Saha, Bhabatosh Chanda

Academic & Technical Details

Research Experience: Computer Vision, Deep Learning, Multimodal Learning, Natural Language Processing

Graduate Courses: Computer Vision, Natural Language Processing, Robotics, Machine Learning, Database Systems

Languages & Frameworks: Python, C++, C, Java, L^AT_EX, PyTorch, TensorFlow, Hugging Face, OpenCV, NLTK, spaCy, Hadoop (familiar), Spark (familiar)