Sounak Mondal

sounakcs.github.io 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

1. **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
2. **Diffusion-Refined VQA Annotations for Semi-Supervised Gaze Following** *ECCV 2024*
   Q. Miao, A. Graikos, J. Zhang, **Sounak Mondal**, M. Hoai, D. Samaras
3. **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
4. **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
5. **Target-absent Human Attention** *ECCV 2022*
   Z. Yang, **Sounak Mondal**, S. Ahn, G. Zelinsky, M. Hoai, D. Samaras
6. **Characterizing Target-absent Human Attention** *CVPR Workshop 2022*
   Y. Chen, Z. Yang, S. Chakraborty, **Sounak Mondal**, S. Ahn, D. Samaras, M. Hoai, G. Zelinsky
7. **ICAN: Introspective Convolutional Attention Network for Semantic Text Classification** *ICSC 2020*
   Sounak Mondal, S. Modi, S. Garg, D. Das, S. Mukherjee
8. **Violent/Non-Violent Video Classification based on Deep Neural Network** *ICAPR 2017*
   Sounak Mondal, S. Pal, S.K. Saha, B. Chanda
9. **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 Scientist Intern | Meta Reality Labs Research, Redmond, WA June, 2024 – Present

Vision-language modeling of human eye gaze behavior in AR/MR.

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, ECCV'24)
- 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 (ECCV'24)

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 classfication 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, LaTeX, PyTorch, TensorFlow, Hugging Face, OpenCV, NLTK, spaCy, Hadoop (familiar), Spark (familiar)