Md Mohaiminul Islam

✉ mmiemon@cs.unc.edu

🌐 Webpage　　　in LinkedIn　　　🎓 Google Scholar　　　⭘ GitHub

Research Interest

Machine Learning	🔖 Deep Learning, Multi-Modal Learning, Self-Supervised Learning, Unsupervised Clustering, Large Language Modeling, Natural Language Processing.
Computer Vision	🔖 Video Understanding, Vision-Language Modeling, Long-Range Video Modeling, Egocentric Vision, Human Activity Understanding, Procedural Planning.

Education

2021 – Present	🔖 **Ph.D. in Computer Science**, 5*th* Year, *UNC Chapel Hill* Advisor: Gedas Bertasius
2021 – 2023	🔖 **Masters in Computer Science**, *UNC Chapel Hill* Advisor: Gedas Bertasius
2014 – 2018	🔖 **B.Sc. in Computer Science and Engineering** *Bangladesh University of Engineering and Technology*, CGPA 3.86/4.00 Advisor: Shamsuzzoha Bayzid

Experience

May 24 – Aug 24	🔖 Research Scientist Intern, *Facebook AI Research*, **Meta AI**, New York, NY, USA Advisors: Lorenzo Torresani, Tushar Nagarajan, Huiyu Wang Topic: Multimodal Large Language Models, Efficient Long Video Understanding Publication: BIMBA (CVPR 2025)
May 23 – Aug 23	🔖 Research Scientist Intern, *Facebook AI Research*, **Meta AI**, Menlo Park, CA, USA Advisors: Xitong Yang, Kris Kitani Topic: Video Agents, Procedural Learning Publication: VidAssist (ECCV 2024, Oral)
May 22 – Aug 22	🔖 Machine Learning Research Intern, **Comcast AI**, Washington DC, USA Advisors: Mahmudul Hasan, Tony Braskich Topic: Scene Detection, Efficient Long-Range Video Models Publication: TranS4mer (CVPR 2023)
Apr 19 – Dec 20	🔖 Lecturer, **University of Asia Pacific**, Dhaka, Bangladesh
Nov 18 – Mar 19	🔖 Software Engineer, **Samsung Research**, Dhaka, Bangladesh

Selected Publications

10 BIMBA: Selective-Scan Compression for Long-Range Video Question Answering
M. Islam, T. Nagarajan, H. Wang, G. Bertasius, L. Torresani
In *Computer Vision and Pattern Recognition (**CVPR**)* 2025.

9 ReVisionLLM: Recursive Vision-Language Model for Temporal Grounding in Hour-Long Videos
T. Hannan, **M. Islam**, J. Gu, T. Seidl, G. Bertasius
In *Computer Vision and Pattern Recognition (**CVPR**)* 2025. [arxiv]

8 Video ReCap: Recursive Captioning of Hour-Long Videos
M. Islam, N. Ho, X. Yang, T. Nagarajan, L. Torresani, G. Bertasius
In *Computer Vision and Pattern Recognition (**CVPR**)* 2024. [arxiv]

7 Ego-Exo4D: Understanding Skilled Human Activity from First- and Third-Person Perspectives
K. Grauman, **M. Islam**, et al.
In *Computer Vision and Pattern Recognition (**CVPR**)* 2024. [arxiv]

6 Propose, Assess, Search: Harnessing LLMs for Goal-Oriented Planning in Instructional Videos
M. Islam, T. Nagarajan, H. Wang, F. Chu, K. Kitani, G. Bertasius, X. Yang
In *European Conference on Computer Vision (**ECCV**)* 2024. (Oral) [arxiv]

5 RGNet: A Unified Clip Retrieval and Grounding Network for Long Videos
T. Hannan, **M. Islam**, T. Seidl, G. Bertasius
In *European Conference on Computer Vision (**ECCV**)* 2024. [arxiv]

4 A Simple LLM Framework for Long-Range Video Question-Answering
C. Zhang, T. Lu, **M. Islam**, Z. Wang, S. Yu, M. Bansal, G. Bertasius
In *Empirical Methods in Natural Language Processing (**EMNLP**)* 2024. [arxiv]

3 Efficient Movie Scene Detection using State-Space Transformers
M. Islam, M. Hasan, K. Athrey, T. Braskich, G. Bertasius
In *Computer Vision and Pattern Recognition (**CVPR**)* 2023. [arxiv]

2 Long Movie Clip Classification with State-Space Video Models
M. Islam, G. Bertasius
In *European Conference on Computer Vision (**ECCV**)* 2022. [arxiv]

1 COVID-DenseNet: A Deep Learning Architecture to Detect COVID-19 from Chest Radiology Images
M. Islam, T. Hannan, L. Sarker, Z. Ahmed
In *International Conference on Data Science and Applications (**ICDSA**)* 2022. [arxiv]

Selected Projects

EgoOSCC   A transformer architecture using the divided space-time attention mechanism that can detect object state changes in egocentric videos.

Audio2Video   A conditional GAN approach to generate plausible videos from natural sounds. Proposed a 3D CNN model for video and CNN-RNN model for audio.

ProteinSeq   A CNN-LSTM model with an attention mechanism to predict the three-dimensional structure of a protein from its amino acid sequence.

BenglaAI   A CNN-based model for Bengali Handwritten Digit Recognition.

UniShare   An academic resource-sharing platform developed using PHP, CodeIgniter framework, MySql, JavaScript, HTML, CSS, and XAMPP server.

PoliceBox   An interactive server-client system developed using Java Swing, JavaFX, MySql, and Socket Programming to efficiently manage various services provided by the police.

Skills

Coding   Python, Java, C, C++, Matlab, HTML, CSS, JavaScript, Android.

ML tools   PyTorch, Keras, TensorFlow, Jupyter, Distributed training.

Database   Oracle, MySQL, PostgreSQL, SQLite.

Misc.   Shell Script, LaTeX, Git, Dockers, Arduino.

Miscellaneous Experience

Invited Talks

02/2025	🔖 Multimodal Webinar, **TwelveLabs**. [Video]
03/2024	🔖 Research Summit for Egocentric Perception with Project Aria, **Meta**, Redmond, WA.[Video]
12/2024	🔖 Ego4D Winter Symposium, **Meta**. [Video]

Organizer and Reviewer

2024	🔖 **Organizer**, *T4V: Transformers for Vision* Workshop, **CVPR** 2024.
	🔖 **Organizer**, *T4V: Transformers for Vision* Workshop, **CVPR** 2023.
	🔖 **Organizer**, *Workshop on Video-Language Models* Workshop, **NeurIPS** 2024.
2022-23	🔖 **Reviewer**, CVPR, ECCV, ACL, IJCV.

Awards and Honors

2022	🔖 **2nd place**, *Ego4D: Object State Change Clasification Challenge*, Ego4d Workshop, **CVPR** 2022.
2018	🔖 **Champion Student Poster Award**, International Conference on Networking, Systems and Security, **NSysS** 2017.
	🔖 **9th place**. *Bengali Handwritten Digit Recognition Challenge*, **Kaggle** Competition 2018.
2016-18	🔖 **Deans List Award**, Bangladesh University of Engineering and Technology.
	🔖 **University Merit List**, Bangladesh University of Engineering and Technology.

Leadership

2014-18	🔖 **Class Captain** Bangladesh University of Engineering and Technology.