Md Mohaiminul Islam

mmiemon@cs.unc.edu

Webpage 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**, *4th* 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 23 – Aug 23 Research Scientist Intern, *Facebook AI Research*, ***Meta AI***, Menlo Park, CA, USA
Advisors: Xitong Yang, Kris Kitani

May 22 – Aug 22 Machine Learning Research Intern, ***Comcast AI***, Washington DC, USA
Advisors: Mahmudul Hasan, Tony Braskich

Apr 19 – Dec 20 Lecturer, **University of Asia Pacific**, Dhaka, Bangladesh

Nov 18 – Mar 19 Software Engineer, **Samsung Research**, Dhaka, Bangladesh

Selected Publications

Peer-Reviewed Publications

5 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. [pdf]

4 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. [pdf]

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. [pdf]

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

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. [pdf]

ArXiv Preprint

2 RGNet: A Unified Clip Retrieval and Grounding Network for Long Videos
T. Hannan, **M. Islam**, T. Seidl, G. Bertasius
In *ArXiv Preprint* 2023. [pdf]

1 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 *ArXiv Preprint* 2023. [pdf]

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.

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

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.

Organizer and Reviewer

2023 — **Organizer**, *T4V: Transformers for Vision* Workshop, **CVPR** 2023.

2022-23 — **Reviewer**, Conference: CVPR, ECCV, Journal: International Journal of Computer Vision.

Invited Talks

03/2024 — Research Summit for Egocentric Perception with Project Aria, **Meta**, Redmond, WA.