Md Mohaiminul Islam

✉ mmiemon@cs.unc.edu
🌐 Webpage   in LinkedIn   🎓 Google Scholar   ⎇ GitHub

Research Interest

Computer Vision	🔖	Video Understanding, Vision-Language Modeling, Efficient Long-Range Video Models, Video LLMs, Egocentric Vision.
Machine Learning	🔖	Natural Language Processing, Large Language Models, Multi-Modal Learning, Efficient Sequence Modeling, LLM Agents.

Education

2021 – Present   🔖 **Ph.D. in Computer Science**, $5th$ Year
The University of North Carolina at Chapel Hill
Advisor: Gedas Bertasius

2021 – 2023   🔖 **M.Sc. in Computer Science**
The University of North Carolina at Chapel Hill
Advisor: Gedas Bertasius

2014 – 2018   🔖 **B.Sc. in Computer Science and Engineering**, CGPA 3.86/4.00
Bangladesh University of Engineering and Technology
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: Video Large Language Models, Long-Range Video Question Answering
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: Multimodal LLM Agents, Procedural Planning, Test-Time Scaling
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 Segmentation, 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

Awards and Honors

2025   🔖 **Doctoral Consortium Award**, **CVPR** 2025.

🔖 **Egocentric Vision (EgoVis) 2023/2024 Distinguished Paper Award**, **CVPR** 2025. [arxiv]

🔖 **1st place**, *EgoSchema Challenge*, EgoVis Workshop, **CVPR** 2025. [arxiv]

2022   🔖 **2nd place**, *Object State Change Clasification Challenge*, Ego4D Workshop, **CVPR** 2022. [arxiv]

2017   🔖 **Champion Student Poster Award**, **NSysS** 2017.

2016-18   🔖 **Deans List Award**, Bangladesh University of Engineering and Technology.

Peer-Reviewed 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. [arxiv]

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]

Skills

Coding	🔖	Python, Java, C, C++, Matlab, HTML, CSS, JavaScript, Android.
ML tools	🔖	PyTorch, Keras, TensorFlow, Jupyter, Distributed training.
Database	🔖	Oracle, MySQL, PostgreSQL, SQLite.
Web Dev	🔖	Html, CSS, JavaScript, XML, Android.
Hardware	🔖	AVR Micro-controller, Arduino.
Misc.	🔖	Shell Script, LaTeX, Git, Dockers, Arduino.

Miscellaneous Experiences

Invited Talks

02/2025	Long-Range Video Question-Answering and Video-Language Reasoning [Video] Multimodal Webinar, **TwelveLabs**.
12/2024	Ego4D Winter Symposium, **Meta**. [Video]
03/2024	Research Summit for Egocentric Perception with Project Aria, **Meta**, Redmond, WA. [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, NeurIPS, ACL, ACMMM.

Leadership

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