Jehanzeb Mirza

Boston, USA. · +18572258760 · jmirza@csail.mit.edu
LinkedIn: linkedin.com/in/jmiemirza · Portfolio: jmiemirza.github.io

Research Interests: Online Learning, Vision-Language Understanding, Unsupervised Representation Learning

I am particularly interested in multi-modal learning for vision-language models, focusing on understanding their limitations and taking steps toward mitigating these limitations. Specifically, I am interested in improving the fine-grained reasoning capabilities of these models.

Education

Graz University of Technology *Jan 2021 - April 2024*
Doctor of Philosophy in Computer Science (Grade: 1.0, with distinction)
Supervisor: Prof. Dr. Horst Bischof | Examiner: Prof. Dr. Serge Belongie Graz, Austria.

Karlsruhe Institute for Technology *Oct 2017 - Aug 2020*
Master of Science in Electrical Engineering and Information Technology
Supervisor: Prof. Dr. Jürgen Becker Karlsruhe, Germany.

National University of Science and Technology *Sep 2013 - June 2017*
Bachelor of Science in Electrical Engineering Islamabad, Pakistan.

Work Experience

Postdoctoral Researcher *Nov 2024 – Present*
Massachusetts Institute of Technology Boston, USA
- Multi-modal learning with speech, vision, and language.

Computer Vision Project Assistant *Jan 2021 – Oct 2024*
Graz University of Technology Graz, Austria
- Mainly focused on employing self-supervised and unsupervised representation learning techniques for making deep neural networks robust to distribution shifts on the fly at test time.
- Along with the main research focus on online learning, I also worked extensively with large language models (LLMs) particularly focusing on multi-modal (vision-language) models.

Research Scientist (Intern) *May 2024 – August 2024*
Sony AI Tokyo, Japan
- Worked on multi-modal learning with vision, language, and audio data.

Master Thesis *Jan 2020 – Jul 2020*
Intel Labs Karlsruhe , Germany
- Evaluated the robustness of state-of-the-art deep neural networks in degrading weather conditions in the context of Autonomous Driving. The study included evaluation of both 2D and 3D object detectors.

C++ developer (Intern) *Oct 2019 – Dec 2019*
Intel Labs Karlsruhe , Germany
- Worked on state estimation of objects detected in the environment using the Unscented Kalman Filter. The development was done by extensively using C++ and OpenCV.

Platform Application Engineer (Intern) *Mar 2019 – Aug 2019*
Intel Karlsruhe , Germany
- Designed a framework to help the team to automate some important internal tasks. The framework involved designing a PCB to interface the boards with a micro controller.

Technical Skills

Languages: Python, C++
Deep Learning Frameworks: PyTorch
Data Modalities: Images, Point Clouds, Videos, Natural Language

Selected Publications (Lead Author)

ConMe: Rethinking Evaluation of Compositional Reasoning for Modern VLMs | Neural Information Processing Systems (NeurIPS) *2024*

Meta-Prompting for Automating Zero-shot Visual Recognition with LLMs | Proceedings of the European Conference for Computer Vision (ECCV) *2024*

LaFTer: Label-Free Tuning of Zero-shot Classifier using Language and Unlabeled Image Collections | Neural Information Processing Systems (NeurIPS) *2023*

MATE: Masked Autoencoders are Online 3D Test-Time Learners | Proceedings of the IEEE/CVF International Computer Vision Conference (ICCV) *2023*

ActMAD: Activation Matching to Align Distributions for Test-Time-Training | Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) *2023*

Video Test-Time Adaptation for Action Recognition | Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) *2023*

The Norm Must Go On: Dynamic Unsupervised Domain Adaptation by Normalization | Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) *2022*

An Efficient Domain-Incremental Learning Approach to Drive in All Weather Conditions | Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPR) *2022*

Invited Talks

CSAIL, MIT. (May, 2024)

Center for Robotics, Paris Tech. (September, 2023)

VIS Lab, University of Amsterdam. (October, 2023)

Cohere. (October, 2023)

Academic Services

Reviewer: CVPR, ICCV, ECCV, NeurIPS, ICLR, TPAMI

Program Chair: What's next in Multi-Modal Foundation Model Workshop at CVPR 2024/2025

Student Supervised Projects (Selected)

Master Thesis: Online Test-Time Training for 3D point clouds with Masked Autoencoders (Completed)

Master Thesis: How Much are Data Augmentations Worth for 3D Representation Learning? (Completed)

Master Thesis: Online Perception System Deployment in Real-Time Scenarios (Ongoing)

Bachelor Thesis: Online Domain Incremental Learning for Driving in Adverse Weather Conditions (Completed)

Bachelor Thesis: Online Test-Time Training for Vision-Language Models (Ongoing)

References

Ph.D. Supervisor (TU Graz): Prof. Dr. Horst Bischof (bischof@icg.tugraz.at)

Ph.D. Advisor (TU Graz): Dr. Horst Possegger (possegger@icg.tugraz.at)

Collaborator (MIT-IBM): Dr. Rogerio Feris (rsferis@us.ibm.com)

Collaborator (MIT-IBM): Dr. Leonid Karlinsky (leonidka@ibm.com)