Manish kumar Govind

+1 203-673-7006 mgovind@charlotte.edu linkedin.com/in/manishgovind portfolio

Research Interests

Robot Learning, Computer Vision, Multimodal Representation Learning

Education

The University of North Carolina at Charlotte **Jan 2025 – Dec 2027**
Ph.D in Computer Science – GPA 4.0/4.0
Advisor: Dr. Srijan Das *Charlotte, NC*

The University of North Carolina at Charlotte **Sep 2022 – Dec 2024**
Master of Science in Computer Science – GPA 4.0/4.0 *Charlotte, NC*

Malviya National Institute of Technology, Jaipur **Jul 2015 – May 2019**
Bachelor of Technology in Computer Science – GPA 7.2/10.0 *Jaipur, India*

Work Experience

Charlotte Machine Learning Lab, UNC Charlotte **Sep 2023 – Current**
Graduate Research Assistant *Charlotte, NC*

- Developing multi-modal **World Action Models (WAM)** for robot manipulation by integrating complementary visual and sensory modalities [e.g., Geometry, motion, and tactile] to leverage rich spatial and physical cues to enhance policy.
- Investigated **continual learning** approaches for WAMs, addressing catastrophic forgetting across sequential robot manipulation tasks through **replay-based methods**.
- Designed **multi-modal latent action pretraining** technique to scale internet videos for robot manipulation using complementary visual modalities such as **Depth Maps**, **Optical Flow**, and **Point Tracks**.
- Gained hands-on experience with **Vision-Language-Action (VLA)** models including **OpenVLA**, **Pi0**, and **LAPA**, focusing on cross-modal policy learning and latent action representation.
- Domain Adaptation of Vision-Language Models(VLMs) to **new target domains such as robot control** using visual probes, while avoiding catastrophic forgetting of source-domain of exocentric videos.
- Enhanced exocentric video understanding of VLMs by integrating egocentric information through cross-view knowledge transfer.
- Video understanding of daily living activities by enhancing (VLMs) with object and pose cues using **curriculum learning.**
- Proposed a **novel sparse attention mechanism** for Vision Transformers (ViTs), demonstrating its effectiveness on image classification, action recognition, and robot policy learning with **95% sparsity**.

Lowe's Tech **Jan 2025 – Dec 2026**
Student Researcher *Charlotte, NC*

- Developed a **multi-camera person re-identification system** and improved model performance for production-level deployment and scalability.
- Designed a **data curation pipeline** to process real-time video streams and created a ReID dataset capturing challenging real-world scenarios.
- Applied **knowledge distillation** and **model quantization** techniques to deploy lightweight ReID models for real-time inference in retail stores.

Mindstaq (Collaboration with UNC Charlotte) **May 2024 – Aug 2024**
Summer Intern *Remote*

- Designed the architecture for **Wingmate**, an AI-driven project management assistant for generating automated meeting summaries and reports.
- Transformed project data into **graph-structured representations** and built a pipeline for automated reports using LangChain, OpenAI, and Flask.
- Applied extensive **prompt engineering** to optimize LLM responses for graph-based reasoning and automated agenda/report generation.

Unitx **Oct 2020 – Jul 2022**
Software Engineer *Thuwal, Saudi Arabia*

- Led development of **Vizard**, a real-time edge-based computer vision platform for drones and CCTV, enabling deployment in bandwidth-constrained robotic settings.
- Optimized deployment pipelines and APIs, achieving **40–50ms latency** and **90% faster** CI/CD workflows.

Droom **Jun 2019 – Aug 2020**
Associate Software Developer *Gurgaon, India*

- Developed scalable CMS and filter APIs serving **100K+ product listings**, doubling content team productivity.
- Optimized data pipelines and vehicle history reports, reducing processing time by **50%**.

Publications

- Manish Kumar Govind, Dominick Reilly, Smit Patel, Hieu Le and Srijan Das. "World Action Models Enable Continual Imitation Learning with Recurrent Generative Replays" *Under review*, 2026. [arXiv] [Project Page]
- Manish Kumar Govind, Dominick Reilly, Pu Wang, and Srijan Das. "**UniLACT**: Depth-Aware RGB Latent Action Learning for Vision-Language-Action Models," *Under review*, 2026. [arXiv] [Project Page]
- Dominick Reilly, Manish Kumar Govind, Le Xue, and Srijan Das."**VisCoP**: Visual Probing for Video Domain Adaptation of Vision-Language Models," *In proceedings of European Conference on Computer Vision (ECCV)*, 2026. [arXiv]
- Dominick Reilly, Manish Kumar Govind, Le Xue, and Srijan Das. "**From My View to Yours**: Ego-to-Exo Transfer in VLMs for Understanding Activities of Daily Living," *In proceedings of European Conference on Computer Vision (ECCV)*, 2026. [arXiv]
- Dominick Reilly, Rajatsubhra Chakraborty, Arkaprava Sinha, Manish Kumar Govind, Pu Wang, Francois Bremond, and Srijan Das. "**LLAVIDAL**: Benchmarking Large Language Vision Models for Daily Activities of Living," In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025. [arXiv] [Code]
- Ali Khaleghi Rahimian, Manish Kumar Govind, Subhajit Maity, Dominick Reilly, Christian Kümmerle, Srijan Das, and Aritra Dutta. "**Fibottention**: Inceptive Visual Representation Learning with Diverse Attention Across Heads," 2024. [arXiv] [Code]
- Trupti Bavalatti, Osama Ahmed, Dhaval Potdar, Rakeen Rouf, Faraz Jawed, Manish Kumar Govind, and Siddharth Krishnan. "A Systematic Review of Open Datasets Used in Text-to-Image (T2I) Generative AI Model Safety." *IEEE Access*, 2024.

Academic Projects

Video Summarization using State Space Models **Dec 2024**

- Introduced **SUMBA** - A video summarization framework using SSMs achieving a **20% improvement over SOTA** baselines on standard video summarization datasets TVSum and SumME.

Report Generation For Project Management **May 2024**

- Developed a LangChain application to manage project data and generate daily meeting reports, enhancing team productivity and project tracking.

CassiNav: SLAM and Object Detection for on campus Vehicle **Dec 2023**

- Employed ORB-SLAM3 for real-time mapping and localization of the UNC Charlotte campus autonomous vehicle Cassi, alongside YOLOv9 to attain a mean IOU score of **0.89** in real-time object detection with an inference time of approximately **2.5** sec.

Technical Skills

Programming: Python, ROS, C/C++, CUDA, Bash
Frameworks & Tools: Pytorch , Hugging face, LeRobot, NumPy, Pandas, Matplotlib, Scikit-Learn , Git, Docker, MLflow
Robots & Simulators: xARM7 Manipulator, UR5e Manipulator, TurtleBot, Gazebo, PyBullet, OpenAI Gym

Awards & Achievements

CVPR Graduate Student Travel Award, 2025
Graduate Student Program Travel Award, UNC Charlotte, 2024
National Merit Scholarship Award, *MHRD, Government of India*, 2016