

Mosam Dabhi

PH.D. STUDENT

Carnegie Mellon University, Pittsburgh, PA, USA

✉ mosam@cmu.edu | 🏠 mosamdabhi.github.io

Research Interests

I focus on advancing AGI through scalable multi-modal AI, 3D vision, and neuro-symbolic reasoning. My research centers on developing geometric foundation models that enable cross-domain generalization for scene understanding and human-computer interaction. These models tackle challenges in out-of-distribution (OOD) reasoning and offer scalable AI solutions for complex real-world applications.

| | |
|-------------------------|---|
| Computer vision | 3D Vision, Text2Video, Multi-view geometry [1–3] |
| Machine Learning | Optimization, SSL, Generative modeling, Graph [2–4] |
| Graphics | Implicit Functions, Differentiable Rendering [5] |
| Robotics | SLAM, Control-theory [6–10] |

Education

Carnegie Mellon University

PH.D. IN ROBOTICS

- **Advisor:** Simon Lucey and Laszlo Attila Jeni

Pittsburgh, USA

Aug. '21 - Present

Carnegie Mellon University

M.S. IN ROBOTICS

- **Advisor:** Simon Lucey
- Thesis: Multi-view NRSfM: Affordable Setup for High-Fidelity 3D Reconstruction

Pittsburgh, USA

Aug. '19 - May '21

National Institute of Technology

B.TECH. IN ELECTRONICS AND COMMUNICATION ENGINEERING

Surat, India

Aug. '13 - May '17

Research Experiences

Carnegie Mellon University

PH.D. STUDENT

- **Advisors:** Simon Lucey, Laszlo Attila Jeni
- Developed Geometric Foundation Models that advance AGI by integrating 3D vision and neuro-symbolic reasoning, enabling cross-domain generalization and scalable solutions for complex, real-world tasks.
- Developed scalable models like MV-NRSfM [1], MBW [2], 3D-LFM [3], RAT4D [4], enabling efficient 3D label generation and ground truth in natural settings, advancing AGI through generalizable, scalable AI.
- MBW: Scalable auto-labeling with minimal supervision, using spatiotemporal priors for landmark labeling.
- 3D-LFM: A scalable model that lifts 2D signals into 3D representations using graph-based transformers, enabling generalization across diverse object categories with minimal supervision, supporting OpenAI's goal of AGI through efficiency.
- RAT4D: Developed a system that generates dense, animatable 3D models of deforming objects from RGB videos without pre-defined templates, facilitating AGI-related applications such as generative AI video corrections and multimodal learning.

Pittsburgh, USA

Aug. '19 - Present

Apple Inc.

RESEARCH SCIENTIST INTERN (AI)

- LLMs, LVMs, and 3D (Vision Pro team).

Cupertino, US

May '24 - Present

Apple Inc.

RESEARCH SCIENTIST INTERN (AI)

- Multimodal time-series modeling, shaping the foundation for lightweight AI foundation models (Vision Pro team).

Cupertino, US

June '23 - Aug. '23

Apple Inc.

RESEARCH SCIENTIST INTERN

- Few-shot learning and Out-Of-Distribution adaptation algorithms (Vision Pro team).
- Auto-labeling in the wild, driving substantial financial savings by auto-generating 3D labels for computer vision applications.

Cupertino, US

May '22 - Aug. '22

Apple Inc.

RESEARCH INTERN

Cupertino, US

May '21 - Aug. '21

- Active learning and self-supervised learning strategies (*Vision Pro team*).

Apple Inc.

RESEARCH INTERN

Cupertino, US

May '20 - Aug. '20

- Foundational work on ML based multi-view 3D geometry, paving the way for economical setups to generate 3D groundtruth labels for 3D vision applications (*Vision Pro team*).

Carnegie Mellon University

RESEARCH ASSISTANT

Pittsburgh, US

May '17 - May. '19

- **Advisor:** Nathan Michael
- Robotic exploration and mapping in real-time for search and rescue operations, enabling superior robot-to-robot communication on extra-terrestrial and sub-terrestrial surfaces [10].
- Achieved flights in challenging, GPS-denied terrains, hitting accelerations over 12 m/s^2 [9]
- Experience-driven Model Predictive Control (EPC) tailored for platforms with computational constraints [6, 7].
- Motion and path planning in cluttered environments through mixed-integer programming [8].

Indian Institute of Science

RESEARCH INTERN

Bangalore, India

May '15 - Jul. '15

- **Advisor:** Prasanta Kumar Ghosh
- Home automation via speech, harnessing MFCC features and speaking rate analysis.

Publications

- [1] **Mosam Dabhi**, Chaoyang Wang, Kunal Saluja, László A Jeni, Ian Fasel, and Simon Lucey. High Fidelity 3D Reconstructions with Limited Physical Views. In *2021 International Conference on 3D Vision (3DV)*. IEEE, 2021.
- [2] **Mosam Dabhi**, Chaoyang Wang, Tim Clifford, László A Jeni, Ian Fasel, and Simon Lucey. Multi-view Bootstrapping in the Wild. In *Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*. NeurIPS, 2022.
- [3] **Mosam Dabhi**, László A Jeni, and Simon Lucey. 3d-lfm: Lifting foundation model. *2024 Computer Vision and Pattern Recognition (CVPR)*, 2024.
- [4] **Mosam Dabhi**, Simon Lucey, and László A Jeni. Rat4d: Rig and animate any object without templates in 4d. *Under submission*, 2024.
- [5] **Mosam Dabhi**, Simon Lucey, and Laszlo A Jeni. Template-free deformable 3d models from videos, 2024. Manuscript in submission.
- [6] **Mosam Dabhi**, Vishnu R Desaraju, and Nathan Michael. Evaluation of Explicit Experience-driven Predictive Control on a Computationally Constrained Platform. Technical report, Carnegie Mellon University, Pittsburgh, PA, June 2017.
- [7] **Mosam Dabhi**, Alexander Spitzer, and Nathan Michael. Aggressive Flight Performance using Robust Experience-driven Predictive Control Strategies: Experimentation and Analysis. Technical Report CMU-RI-TR-19-08, Carnegie Mellon University, Pittsburgh, PA, June 2019.
- [8] **Mosam Dabhi**, Vishnu Desaraju, and Nathan Michael. Planning Aggressive, Dynamically Feasible and Optimal Trajectories for Autonomous Vehicles in Cluttered Environments using Mixed Integer Programming. Technical report, Carnegie Mellon University, Pittsburgh, PA, 2016.
- [9] Alex Spitzer, Xuning Yang, John Yao, Aditya Dhawale, Kshitij Goel, **Mosam Dabhi**, Matt Collins, Curtis Boirum, and Nathan Michael. Fast and agile vision-based flight with teleoperation and collision avoidance on a multirotor. In *International Symposium on Experimental Robotics*, pages 524–535. Springer, 2018.
- [10] Wennie Tabib, Kshitij Goel, John Yao, **Mosam Dabhi**, Curtis Boirum, and Nathan Michael. Real-Time Information-Theoretic Exploration with Gaussian Mixture Model Maps. In *Robotics: Science and Systems*, 2019.

Honors & Awards

- '19 - '23 **Apple Research Grant**, Apple Inc.
- '17 **Research Scholarship**, FICCI, India
- '16 - '17 **Summer Scholar**, Robotics Institute Summer Scholar
- '16 **Undergraduate thesis funding**, TEQIP Award, MHRD, Government of India

Academic Services

- '22 - '24 **Conference Paper Reviewer**, NeurIPS; CVPR; ICLR; ICML; AAAI; ICCV; ECCV; WACV
- '21 - '24 **Conference Paper Reviewer**, ICRA, IROS
- '20 - '24 **Conference Paper Reviewer**, International Conference on Humanoid Robots
- '21 - '24 **Admissions Committee**, Master of Science, Robotics, CMU (MSR)
- '23 - '24 **Admissions Committee**, Master of Science, Computer Vision, CMU (MSCV)
- '22 - '23 **M.S. in Robotics Thesis Committee**, Examinee: Heng Yu, Aarush Gupta
- '17 - '19 **Admissions & Admin. Committee**, Robotics Institute Summer Scholars, CMU (RISS)

Teaching Experiences

Carnegie Mellon University

TEACHING ASSISTANT

- **Spring 2022** : Robot Localization and Mapping
- **Fall 2022** : Geometry-Based Methods in Vision
- **Fall 2023** : Advanced Machine Learning with Large Datasets

Relevant coursework

Carnegie Mellon University

LEARNING FOR 3D VISION (**A+**), ADVANCED COMPUTER VISION (**A**), GEOMETRY METHODS IN VISION (**A+**), ADVANCED MACHINE LEARNING (**A**), CONVEX OPTIMIZATION (**A**), ROBOT LOCALIZATION AND MAPPING / SLAM (**A+**), MATH. FUNDAMENTALS FOR ROBOTICS (**A**), KINEMATICS, DYNAMICS, AND CONTROLS (**A+**)

Proficient Skills

Programming languages

PRIMARY: PYTHON, C/C++, \LaTeX , MATLAB

SECONDARY: CUDA, LUA, HTML, JAVASCRIPT

Software libraries

PRIMARY: TRANSFORMERS, PYTORCH, TENSORFLOW, COLMAP, BLENDER, GENERATIVE AI FRAMEWORKS (LIKE DALL-E, SWARM)

SECONDARY: OPENAIGYM, SSL, TORCH, CAFFE, OPENCV, VLFEAT, PTHREAD