Amitangshu Mukherjee

Email | Website | Google Scholar | LinkedIn

Research Overview

Ph.D. candidate specializing in **active vision** architectures for **robust perception**. I develop foveated learning frameworks that mitigate **domain shift** and **long-tail fragility** while significantly reducing sample complexity. My current work adapts these efficiency priors to **vision–language models** and **multimodal LLMs**, enabling token-efficient, high-fidelity inference at scale.

Education

Doctor of Philosophy (PhD) in Computer Engineering — **Summer 2020-Present**
Purdue University — Expected Graduation: Spring 2026

- Advisor: **Dr. Kaushik Roy**
- Research: **Cognitive Inspired AI**, **Scene Understanding**, **Robust Deep Learning**, **Multimodal Models**

MS in Computer Engineering — **Fall 2017 - Spring 2020**
Iowa State University

- Advisors: **Dr. Chinmay Hegde & Dr. Soumik Sarkar**
- Research Area : **Domain Adaptation**, **Autonomous Driving**, **Generative Adversarial Networks**, **Robust Deep Learning**

Experience

Purdue University | Graduate Research Assistant | NRL | C-BRIC

- Foveated Object-Centric Learning (FocL): Designed a multi-viewpoint foveated-vision training framework improving adversarial robustness by **61%** with **56% less labeled data** and **∼65% lower cumulative sample loss** (memorization proxy); integrated **SAM**-based dorsal–ventral pipeline, boosting ImageNet-V2 Top-1 by **+7 pp** and COCO zero-shot mAP by **3–4%**. [1]
- Multimodal Privacy: Extending framework to Self-Supervised Learning (SSL) for ViT pretraining and **VLM grounding**, specifically targeting the reduction of memorization in foundation models.
- Active Vision for Enhanced Robustness: Developed a dorsal–ventral active-vision system with ViT ventral backbones, improving robustness to state-of-the-art transformer-based black-box transfer attacks on ImageNet with **1.75**× higher accuracy, and achieving **1.5–2**× higher accuracy on the **ImageNet-A** hard-sample benchmark. [2, 3]
- Hierarchical Models for Subpopulation Shift: Developed a hierarchical training approach that reduces severe errors by **7.5–8 pp** and improves accuracy by **1.5–2 pp** on long tail, subpopulation-shifted subsets. [4]
- Adaptive Foveated Vision & FastVLM Integration: Developing a token-efficient adaptive-resolution framework using factorized positional embeddings and Patch-n-Pack to process native-resolution object crops, improving small-object Top-1 accuracy by **+7.3 pp** with **74% fewer FLOPs**. Integrating this architecture into a FastVLM-based multimodal object-centric grounding pipeline via dual-branch token injection to reduce small-object hallucinations in cluttered real-world visual scenes.

IOWA STATE UNIVERSITY | Graduate Research Assistant | SCS Lab

- Attribute GAN-Based Semantic Adversarial Attacks: Developed a conditional GAN–based optimization scheme to generate semantic adversarial examples, driving target classifier accuracy down to **1%** on CelebA and outperforming iterative baselines [5].
- GAN-Based Domain Adaptation for Autonomous Driving: Built an attribute-controlled GAN to synthesize adverse driving conditions (e.g., rain, night), improving detector mAP by **3–5%** on BDD and KITTI [6, 7, 8].

Internship

Qualcomm | Video Research Intern | Summer 2023 | San Diego, USA

- Joint Learning of Optical Flow, Depth and Camera Pose via Transformers: Developed a **multi-decoder Transformer architecture** for autonomous-driving videos in the Multimedia R&D team, jointly estimating optical flow, depth, and camera pose.
- Designed a **self-supervised training framework** leveraging **epipolar geometry** and **view synthesis** to learn flow, depth, and pose jointly in a resource-efficient manner for edge deployment.

Selected Publications

[1] **Amitangshu Mukherjee**, D. Ravikumar, and K. Roy, "From clutter to clarity: Visual recognition through foveated object-centric learning (focl)," in *NeurIPS 2025 Workshop: Reliable ML from Unreliable Data*, 2025.

[2] **Amitangshu Mukherjee**, T. Ibrayev, and K. Roy, "On inherent adversarial robustness of active vision systems," *Transactions on Machine Learning Research*, 2025.

[3] T. Ibrayev, M. Nagaraj, **Mukherjee, Amitangshu**, and K. Roy, "Exploring foveation and saccade for improved weakly-supervised localization," in *Proceedings of The 2nd Gaze Meets ML workshop, NeurIPS 2023*, PMLR, 2024.

[4] **Mukherjee, Amitangshu**, I. Garg, and K. Roy, "Encoding hierarchical information in neural networks helps in subpopulation shift," *IEEE Transactions on Artificial Intelligence*, 2023. An earlier version appeared as a poster at the 9th Workshop on Fine-Grained Visual Categorization (FGVC9), CVPR 2022.

[5] A. Joshi, **Amitangshu Mukherjee**, S. Sarkar, and C. Hegde, "Semantic adversarial attacks: Parametric transformations that fool deep classifiers," *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019.

[6] **Amitangshu Mukherjee**, A. Joshi, S. Sarkar, and C. Hegde, "Attribute-controlled traffic data augmentation using conditional generative models," in *CVPR Workshop on Vision for All Seasons (VAS)*, 2019.

[7] **Amitangshu Mukherjee**, A. Joshi, S. Sarkar, and C. Hegde, "Semantic domain adaptation for deep classifiers via gan-based data augmentation," in *Machine Learning for Autonomous Driving, NeurIPS workshop (Poster)*, 2019.

[8] **Amitangshu Mukherjee**, A. Joshi, S. Sarkar, and C. Hegde, "Generative semantic domain adaptation for perception in autonomous driving," in *Journal of Big Data Analytics in Transportation, Nature Springer*, 2022.

Interested Areas

• Domain Adaptation • Scene Understanding • Multimodal Learning • Multimodal Inference • Structure from Motion • Autonomous Driving • Scene Understanding • Perception in Robotics • Generative AI • Embodied AI • Adversarial Robustness • Large Vision Models • Foveated Learning • Cognitive Inspired AI • Vision Language models • AR/VR

Relevant coursework

• Computational Perception • Deep Learning • Generative Adversarial Networks • Machine Learning for Cyber-Physical Systems • Convex Optimization • Random Variables • Linear Algebra • Foundational AI Models • AI Ethics and Society

Skills

Programming & OS: Python, Bash, Linux (RedHat, Ubuntu), Git
ML & Deep Learning: PyTorch, TensorFlow, NumPy, scikit-learn, HuggingFace Transformers
Computer Vision & Multimodal: OpenCV, SAM/SAM 2, vision–language (CLIP/VLM-style) pipelines
Compute & Tools: SLURM-based GPU clusters, Conda, LaTeX
Datasets: ImageNet, COCO, KITTI, BDD100K, CelebA, Part-ImageNet, Ref COCO, ImageNet A, ImageNet V2, LVIS

Reviewer and Responsibilities

- **Computer Vision: CVPR** (2023, 2024, 2026), **ICCV** (2023), **ECCV** (2024)
- **AI: ICLR** (2025, 2026), **ICML** (2025), **NeurIPS** (2025), **AAAI** (2026) **TinyICLR** (2024),
- Transactions on Machine Learning Research (**TMLR**), September 2025 - Present
- **Workshops:** Machine Learning for Autonomous Driving Workshop at **NeurIPS** (2020–present)
- **Career Chair**, Electrical and Computer Engineering Graduate Student Association (**ECEGSA**) at Purdue University, Summer 2022-2023