Pratik Bhowal

Pittsburgh/San Diego | pbhowal@andrew.cmu.edu | +1-412-901-0557 | LinkedIn | Google Scholar
Website

Education

- **MS Robotics | August 2024 - June 2026**
  Carnegie Mellon University
  Research Focus Robot Learning, Manipulation, Computer Vision, 3D Vision
  Courses Learning for 3D, Discrete Differential Geometry, Advanced Computer Vision, Intro to Robot Learning, Mechanics of Manipulation, Robo-Math
- **B.E in Instrumentation & Electronics Engineering | 2018- 2022**
  Jadavpur University | GPA 9.23/10
  Relevant Coursework => Deep Learning for Visual Computing, Artificial Intelligence and Machine Learning, Optimization Techniques for Engineering Design, Reinforcement Learning, Brain Computer Interface, Linear Control Systems, Advanced Process Control, Computer Organization Architecture and Networking, DBMS, Data Structure and Algorithm

Research Experience

- **Carnegie Mellon University (RPaD Lab)** | David Held & Zackery Erickson
  - **3D and 2D goal Conditioned with future Pre-training plans**
    * Initially, I developed a 3D goal prediction algorithm based on the MVT architecture. To handle multimodality, I used GMM.
    * However, going forward, we plan to pretrain the network using open-source datasets such as DROID, and to avoid relying on camera-to-EEF transformations in those datasets, we shifted to predicting 2D goals instead.
    * In the 2D goal prediction setup, I implemented an MVT-based high-level policy to generate 2D goals for the low-level controller. The low-level policy, a diffusion policy style model, takes these 2D goals along with RGB-D inputs to produce the final 3D goal for execution
  - **Articubot** (**Website**) | **RSS2025** [3]
    Developing a point-cloud–based hierarchical policy capable of opening a wide range of articulated objects, including cupboards, refrigerators, and dishwashers.
    * Zero-shot sim-to-real door-opening policy trained on thousands of simulated demonstrations. The system uses a hierarchical setup with a weighted-displacement model as the high-level policy and a diffusion-based controller as the low-level policy.
    * In this project, I explored multiple combinations of high-level and low-level policies. For the high-level module, I evaluated several Transformer-based DP3 variants, experimented with a 3DDA-style diffusion network for our policy, and investigated a modified weighted-displacement approach that integrates diffusion components into the framework. On the low-level side, I focused on improving the policy using a ToolFlowNet-style architecture.
    * To improve model generalization, I developed several handle-augmentation strategies, including shifting and scaling existing handles. I also created a new handle dataset by extracting handles from PartNet objects and designed an algorithm to scale and place these handles onto cupboard doors, enabling the policy to handle broader variations in handle geometry and placement.
    * To enable the policy to operate on multiple doors specified by the user, I designed a point-prompting architecture. While the project was later abandoned, I generated the dataset and trained an initial policy. I recorded all points on relevant links to later sample point prompts and trained the policy using a one-hot encoding for the prompt, combined with the point cloud and gripper points, to indicate which door to open. In simulation, the robot was able to successfully open the door specified by the point prompt on a small dataset in simulation.
  - **In-context Imitation Learning (Website) | Submitted ICRA 2026** [2]
    * Articubot is not fully reliable in real-world scenarios for out-of-distribution objects. To address this, we use human demonstrations to condition the model whenever it fails during real-world execution.

    - In this project, I explored various techniques to condition the policy on human demonstrations, including different FiLM variants and attention-based conditioning approaches.
    - Later, it was observed that the policy often ignored the human demonstrations. I developed policies that applied various observation distortions to encourage the model to give greater importance to the demonstrations.
    - I also developed an algorithm to convert human demonstrations into a gripper four-point representation, enabling more effective conditioning of the model.
    - I designed the real-world system to capture demonstrations, convert them into the gripper four-point representation, and integrate the full policy. I then extensively tested and deployed the policy on a Franka Arm for tabletop tasks and an XArm mounted on a mobile base.
  - **Bimanual Manipulation**
    - I came up with a number of architectural changes that could adapt Articubot High Level Policy to the bimanual changes. The initial changes created issues such as improper coordination between the 2 arms, essential for Bimanual Manipulation tasks like coordinated movements or object transfer. To enable coordination, I designed architectures that use the predictions of one arm to inform the other's actions, implementing variations of this approach in both the GMM based architectures.
- **University of Wateroo (Critical AI Lab)** | Sirisha Rambhatla
  - **Explaining Disentanglement Learning in VAEs** (**Paper**) | **ICML 2024** [1]
    - Exploring the architecture, assumptions, and the loss functions of the VAE to explain its success in Disentanglement Learning from a non-linear perspective.
    - Modeled VAE decoders as compositions of linear and nonlinear transformations and used Taylor expansions to overcome issues inherent in purely linear assumptions.
    - Derived, under standard polarized-regime assumptions for VAEs, that minimizing reconstruction loss enforces orthogonality among the columns of the decoder's linear component.
    - Demonstrated how this emergent orthogonality provides a theoretical explanation for disentanglement in VAEs.
- **Carnegie Mellon University | Research Intern**
  - Developed a meta learning based technique for video anomaly detection. In this project I developed meta learning techniques for Video Anomaly Detection.
  - Developed a novel Network for Saliency Detection. In this project I have developed a novel deep learning architecture for Saliency Detection.
  - Detection of Mitochondria in Tomograms. In this project I developed an object detection module for the detection of mitochondria in tomograms.
- **Research Assistant | Jadavpur University**
  - **Integrating the Dempster Shafer Belief Structure into the Fuzzy Measure Theory (Paper) | IEEE Transactions on Fuzzy Systems** [4]
    - Developed a low-complexity framework for computing fuzzy measures used in Choquet integrals for classifier and deep model fusion.
    - Showed that Dempster–Shafer (DS) belief structures provide partial information about underlying fuzzy measures and leveraged this insight to design a new calculation method. Proved a general theorem describing a family of fuzzy measures consistent with a given DS belief structure, expressed as weighted combinations of the basic assignment function.Demonstrated that this formulation removes the need to explicitly enforce monotonicity constraints during fuzzy measure computation.
    - Designed an efficient procedure for evaluating the basic assignment function using a Monte-Carlo–inspired approach adapted from value-function estimation in Markov decision processes.
    - Applied the proposed method to fuse deep learning models across multiple domains, achieving accuracy improvements over existing fuzzy-measure–based fusion techniques.
  - **Using Shapley Values to compute Fuzzy Measures (Paper) | IEEE Journal of Biomedical and Health Informatics** [5]
    - Developed a Choquet-integral–based fusion framework requiring fuzzy measures over all classifier subsets, and designed a principled approach to estimate these measures. Applied coalition game theory by using the Shapley value to compute each classifier's payoff within the grand coalition, grounding the fuzzy measures in rigorous marginal-contribution analysis.

    - Used information-theoretic quantities—mutual information and conditional mutual information—to evaluate redundancy, independence, and interdependence among classifiers, guiding the computation of marginal contributions. Constructed a scoring function that favors classifiers with high relevance to the true label and low redundancy with already-selected classifiers, while incorporating penalties for overly repetitive high-accuracy models.
    - Computed three separate Choquet integrals based on these measure sets and combined their outputs via majority voting to improve robustness of the final fused decision.
    - Applied the resulting fusion framework to deep learning models for medical imaging–based COVID-19 screening.
- **University of Surrey | Research Intern**
  - Survey of Transformer Architecture. I helped in the comprehensive literature review of the different types of Transformer Architecture with special emphasis on the Fine Grained Task.
- **Indian Institute of Technology, BHU, Varanasi | Summer Research Intern**
  - Realistic And Textured Terrain Data generation using GANs. In this project I first generated Height Maps using GANs. Next Texturing of these Height Maps was done using the satellite images present. This texture generation was also done using GANs.
- **Nagoya University, Nagoya, Japan |Student Exchange Program , Sakura Science Program**
  - In this exchange program I worked on Hyperspectral Imaging and AI analysis of data obtained by Hyperspectral Imaging.

Professional Experience

- **Nvidia**
  IC2 (Deep Learning - Audio (Maxine SDK) Team) | April 2024 - August 2024
  IC1 (Deep Learning - Audio (Maxine SDK) Team) | August 2022 - March 2024
  - **AI-Mic (Studio Voice):** This converts low quality and noisy audio recordings to studio-quality audio.
    - I worked on a transformer- and LSTM-based generative algorithm, exploring multiple network variants. This included implementing causal convolutions and causal transformers, as well as experimenting with different positional embeddings to optimize transformer performance. I designed both waveform and spectral encoders for the network and evaluated various loss function strategies, including embedding, classification, and discriminator losses, to improve training effectiveness.
    - I optimized the model architecture and training/inference pipeline to lower latency without compromising quality, implementing caching for convolutional layers and key-value (KV) caching for transformer architectures.
    - I ported the model to an NVIDIA Jetson to leverage the Deep Learning Accelerator (DLA) and reduce GPU usage. Porting the large model to Jetson was challenging due to latency constraints, and adapting it for the DLA required implementing custom functions for unsupported modules such as LSTMs.
  - **Speaker Focus (Speaker-Separation)**
    - For this project, I worked on developing the architecture, which was a compact LSTM-based model.
    - I also developed the data pipeline and dataloader, incorporating multiple speakers and adding noise on the fly at varying SNR levels. Additionally, I worked with both 16 kHz and 48 kHz audio data.
  - **Super Resolution (Bandwidth Extension)** -
    - Worked on developing Diffusion Based models for Speaker Resolution.
- **Qualcomm | Software Development Intern (Qgenie AI team)** | June 2025 - August 2025
  - Working on Multimodal RAGs for LLM models.
  - Working on Finetuning techniques for LLM models
- **Nvidia** | **System Software Development Intern (GeForce Now Team)**| January 2022 - June 2022
  - Worked with state-of-the-art methods for language translation and developed REST API and "Progressive Web Apps" for it.

Publications

[1] Why do variational autoencoders really promote disentanglement? ; **Pratik Bhowal**, Achint Soni, Sirisha Rambhatla ; Proceedings of the 41st International Conference on Machine Learning ; **(ICML -2024)**

[2] In-context Imitation Learning for Articulated Object Manipulation; **Pratik Bhowal*, Yufei Wang***, Jacinto Suner, Zackory Erickson, David Held ; **(ICRA 2026)**
* : Joint First Author

[3] ArticuBot: Learning Universal Articulated Object Manipulation Policy via Large Scale Simulation ; Yufei Wang*, Ziyu Wang*, **Mino Nakura**[&], **Pratik Bhowal**[&], **Chia-Liang Kuo**[&], Yi-Ting Chen, Zackory Erickson[+], David Held[+] ; **(RSS 2025)**
* : Joint First Author, [&] : Joint Second Author, [+] : Equal Advising

[4] Evaluation of fuzzy measures using Dempster–Shafer belief structure: A classifier fusion framework ; **P Bhowal**, S Sen, JH Yoon, ZW Geem, R Sarkar ; **IEEE Transactions on Fuzzy Systems**

[5] Choquet integral and coalition game-based ensemble of deep learning models for COVID-19 screening from chest X-ray images ; **Pratik Bhowal**, Subhankar Sen, Jin Hee Yoon, Zong Woo Geem, Ram Sarkar; **IEEE Journal of Biomedical and Health Informatics**

Skills

- **Subjects of Interest:** Robot Learning, 3D Vision, Theoretical Machine Learning, Vision
- **Languages:** Python, C++, C
- **Frameworks:** Pytorch, Deoxys, IsaacGym, ManiSkill, Pybullet, CoppeliaSim

Academic Extracurriculars

- **IEEE Jadavpur University Student Branch:**
  - Conducted a PCL Board Workshop
  - Conducted two AI Workshops
  - Conducted multiple Women in Tech Seminers
- **Hyperspectral Imaging:**
  - Conducted a joint conference on Hyper Spectral Imaging and AI for hyper spectral imaging between Japan Science and Technology Agency (JST) and Jadavpur University.

References

- **Prof. David Held** Associate Professor, Robotics Institute, Carnegie Mellon University, dheld@andrew.cmu.edu
- **Prof. Zackery Erickson**, Assistant Professor, Robotics Institute, Carnegie Mellon University, zackory@cmu.edu
- **Prof. Sirisha Rambhatla**, Assistant Professor, David R. Cheriton School of Computer Science, University of Waterloo sirisha.rambhatla@uwaterloo.ca
- **Ram Sarkar**, Computer Science Department, Jadavpur University ram.sarkar@jadavpuruniversity.in
- **Ambrish Dantrey**, Senior Manager, NVIDIA, adantrey@nvidia.com