Debottam Dutta

Room 261, Coordinated Science Laboratory, Urbana, IL, 61801

dd24@illinois.edu debottam-dutta7.github.io Google Scholar

RESEARCH SUMMARY

PhD researcher on **generative and multimodal models** (diffusion, VAEs, EBMs) spanning **vision, audio, and music**. Focused on controllable image and audio generation, efficient inference and source separation. Built **sampling-time controllers** that improve compositional fidelity in diffusion models, and multi-source **music generation** systems using latent diffusion.

EDUCATION

Ph.D. in Electrical and Computer Engineering
University of Illinois Urbana-Champaign — Aug, 2022–Present
Advisor: Prof. Romit Roy Choudhury (SiNRG)

M.Tech in Signal Processing
Indian Institute of Science, Bangalore — July, 2021
Advisor: Prof. Sriram Ganapathy (LEAP Lab)

B.Tech in Electronics and Communication Engineering
National Institute of Technology, Silchar — May, 2018

RESEARCH INTERESTS

Generative Models (Diffusion, VAEs, EBMs); Compositional and Controllable Generation; Audio and Music Generation; Multi-channel Source Separation; Efficient and Robust Inference.

RESEARCH EXPERIENCE

Graduate Research Assistant *Signals & Inference Research Group (SiNRG), UIUC*
Advisor: Prof. Romit Roy Choudhury

- **Compositional generative modeling:** Designed sampling-time controllers improving multi-object balance in diffusion/VLM backbones (no finetuning), achieving zero-shot compositional gains.
- **Music Generation:** Designed multi-instrument controllable music generation system based on latent instrument codes and diffusion priors.
- **Speech Enhancement:** Worked on multi-channel speech separation and enhancement for wearable smart glasses.

Graduate Student Researcher *Indian Institute of Science (IISc)*
Advisor: Prof. Sriram Ganapathy

- Interpretable representation learning for acoustic signals; contributed to large-scale health acoustics project targeting respiratory illness detection from crowd-sourced acoustic signals.

WORK EXPERIENCE

Research Fellow LEAP Lab, IISc — May 2021 – June 2022

- **Coswara:** Built pipelines and models for COVID-19 screening from cough/breath/speech as part of a public web diagnostic tool; contributed to a large, symptom-rich dataset release.
- **Speech Enhancement:** Built algorithms for improving ASR and listening quality of far-field reverberated speech.

Teaching Assistant, UIUC — Fall 2025, Spring 2024

- *Deep Generative Models*; *Real-World Algorithms for IoT and Data Science*: office hours, assignment/exam design and grading.

Summer Research Intern IIT Madras — May–July, 2017

- Time-series error analysis and modeling of IMU sensor data with autoregressive methods.

ACADEMIC HONORS & AWARDS

- **AR Buck "Knight" Fellowship** (ECE, UIUC).
- **AICTE-PG Scholarship**.
- **Ishan-Uday Scholarship**, **Ishān Bikās Scholarship** (Govt. of India).
- **Ananda Ram Borooah Award** (Govt. of Assam, India).

SELECTED PROJECTS

CO3: Contrasting Concepts Compose Better (Paper) Nov 2024 – Present

- Inference-time controller for diffusion/VLMs enabling balanced composition of objects/attributes without retraining.
- Lightweight, gradient-free, and compatible with common vision foundation backbones; evaluated via CLIPScore/ImageReward/BLIP-VQA.

Learning Energy-based Variational Latent Prior for VAEs (Paper) Dec 2023 – Sep 2024

- Introduced an energy-based latent prior in SOTA VAEs, improving sample quality and inference robustness across modalities.
- Demonstrated scalability to hierarchical VAEs for high-resolution sample generation.

Multi-Source Music Generation with Latent Diffusion (Paper) (Demo) Jan – Sep 2024

- Latent diffusion prior over instrument codes for controllable instrument-conditioned music generation; evaluated with FAD and subjective metrics.

Coswara Dataset (Paper) (Dataset) (Web App) May 2021 – June 2022

- Led data collection and curation for a large respiratory sound dataset; built a deep learning-based diagnostic tool for COVID-19 detection from cough/breath/speech.

SKILLS

Generative Modeling: Diffusion (image/audio), VAEs, EBMs, Latent Priors
Audio/Music: torchaudio, librosa, ESPNet
Frameworks: PyTorch, HuggingFace Diffusers/Transformers, CUDA
Tooling: Slurm, Weights&Biases, Git
Programming: Python, C/C++, MATLAB

COURSES TAKEN

Deep Learning, Speech Processing, Computer Vision, Optimization, Information Theory, Random Processes, Matrix Theory.

CURRENT PREPRINTS

[1] Debottam Dutta, Jianchong Chen, Rajalaxmi Rajagopalan, Yu-Lin Wei, Romit Roy Choudhury, **"CO3: Contrasting Concepts Compose Better"**, https://arxiv.org/pdf/2509.25940

[2] Debottam Dutta, Chaitanya Amballa, Zhongweiyang Xu, Yu-Lin Wei, Romit Roy Choudhury, **"Learning Energy-based Variational Latent Prior for VAEs"**, https://arxiv.org/pdf/2510.00260

SELECTED PUBLICATIONS

[3] Zhongweiyang Xu, Debottam Dutta, Yu-Lin Wei, Romit Roy Choudhury, **"Multi-Source Music Generation with Latent Diffusion"**, NeurIPS 2024 Workshop on Audio Imagination.

[4] Sattwik Basu, Debottam Dutta, Yu-Lin Wei, Romit Roy Choudhury, **"Estimating Multi-chirp Parameters using Curvature-guided Langevin Monte Carlo"**, ICASSP 2025, Hyderabad, India, pp. 1–5.

[5] Anurenjan Purushothaman, Debottam Dutta, Rohit Kumar, Sriram Ganapathy, **"Speech Dereverberation With Frequency Domain Autoregressive Modeling"**, IEEE/ACM TASLP, vol. 32, pp. 29–38, 2024.

[6] Debottam Dutta, Debarpan Bhattacharya, Sriram Ganapathy, Amir H. Poorjam, Deepak Mittal, Maneesh Singh, **"Acoustic Representation Learning on Breathing and Speech Signals for COVID-19 Detection"**, Proc. Interspeech 2022, pp. 2863–2867.

[7] Debarpan Bhattacharya, Debottam Dutta, Neeraj Kumar Sharma, Srikanth Raj Chetupalli, Pravin Mote, Sriram Ganapathy, Sahiti Nori, Sadhana Gonuguntla, Murali Alagesan, **"Analyzing the Impact of SARS-CoV-2 Variants on Respiratory Sound Signals"**, Proc. Interspeech 2022, pp. 2473–2477.

[8] Debottam Dutta, Purvi Agrawal, Sriram Ganapathy, **"A Multi-head Relevance Weighting Framework for Learning Raw Waveform Audio Representations"**, WASPAA 2021, pp. 191–195.

Full list on Google Scholar.