

Mustafa Chasmai

 mustchasmai |  website |  mustafa-chasmai |  mustafa1728

Education

Ph.D. in Computer Science

Supervised by Prof. Subhransu Maji | GPA: 4 / 4 | Expected Graduation May 2027

University of Massachusetts Amherst

2023 - current

Bachelor of Technology (B.Tech)

Computer Science and Engineering | GPA: 9.35 / 10

Indian Institute of Technology Delhi

2019 - 2023

Publications

Bioacoustic Geolocation: Species Sounds as Geographic Signals

Mustafa Chasmai, Wuao Liu, Subhransu Maji, Grant Van Horn

ICML '26

[arxiv]

- First work geolocating audio at the global scale. Unlike images, geographic signals ambiguous in audio.
- Show the potential of species sounds as geolocation signals with strong oracles and proof of concepts.
- Multimodal (audio, image, text) geolocation with cross-modal contrast. Debunk movie fake bird sounds.

The iNaturalist Sounds Dataset

Mustafa Chasmai, Alexander Shepard, Subhransu Maji, Grant Van Horn

NeurIPS '24

[paper]

- New dataset for acoustic recognition of ~5000 challenging (even for experts) fine-grained species.
- Benchmark popular models, demonstrating benefits of geographic priors and visual cues for species.
- Demonstrate utility as pretraining data for a variety of downstream bioacousting monitoring tasks.

Robust Prototypical Few-Shot Segmentation with Regularized Neural-ODEs

Prashant Pandey*, Mustafa Chasmai*, Tanuj Sur, Brejesh Lall

MICCAI '22, TMI '23

[paper], [paper]

From Forks to Forceps: A New Framework for Segmentation of Surgical Instruments

Britty Baby, Daksh Thapar, Mustafa Chasmai, Tamajit Banerjee, Subhashis Banerjee, Chetan Arora

WACV '23

[paper]

A View Independent Classification Framework for Yoga Postures

Mustafa Chasmai, Nirjhar Das, Aman Bhardwaj, Rahul Garg

SNCS '22

[paper]

Internships

Metadata co-signals for Bioacoustics Modeling at scale

Student Researcher | Bioacoustics | Ecology | Multimodal Learning | ICML '26

Google, Montreal

May - Dec 2025

- Use large scale citizen science data to build models for novel bioacoustics tasks like species identification.
- Use auxiliary multimodal metadata like location, text descriptions and images to get robust audio features.

Long Form Video Question Answering

PhD Research Intern | Multimodal AI | Vision Language Models | CVPR '25 workshop

Dolby, San Francisco

May - Aug 2024

- Factored approach of first retrieving relevant moments in long video and then answering questions.
- Benchmarked approach with 4 large VLMs (proprietary and open) on 4 long form video QA datasets.

Explainable AI (XAI) API Development

Deep Learning Research Intern | Responsible AI | Attention Mechanisms

Sony, Japan

May - July 2022

- Developed reusable and modular APIs for 3 XAI methods with rigorous tests and integration in nnabla.
- Implemented SOTA models of DenseNet, ResNext with attention branch on NNC, Sony's GUI for DNNs.

Unsupervised Domain Adaptation for Action Recognition

Prof. Haiping Lu, CS Sheffield | Transfer Learning | Adversarial Training

University of Sheffield, UK

May - Aug 2021

Miscellaneous

- **Technical Skills** | Python, C++ | Pytorch, Huggingface, Sklearn, Gradio | Google Colab, slurm, LLM API
- **Fellowships and Awards** | IITD Top 7% (3 sems) | INPhO gold, OCSC for IPHO | KVPY | AIR 91 JEE '19 |
- **Extra Curricular** | Student Mentor | AI-ML club | Photography | Physics & Astro | Rubik Cube (20×20) |