

Research Interests

Multimodal Machine Learning, Computer Vision, Affective computing, Machine learning for Healthcare

Education

- 2018–2024 **University of Southern California(USC)**.
 - Ph.D in Electrical and Computer Engineering; **GPA: 3.88/4.00**
 - **Advisor:** Prof. Shrikanth Narayanan
- 2014–2016 **Indian Institute of Technology Bombay**.
 - M.Tech in Electrical Engineering; **CPI: 9.51/10**
 - **Specialization:** Control and Computing (Rank: 2/16)
 - **Advisor:** Prof. Subhasis Chaudhuri
- 2010-2014 **Jadavpur University**.
 - B.E. in Electronics and Telecommunication Engineering; **CPI: 9.34/10**
 - **Specialization rank:** 2/46

Selected Publications

Please see [Google Scholar](#) for the complete list of publications

- **AffectEdit: Diagnosing and Fixing Evaluation for Affective Image Editing**
Kilichbek Haydarov, [Digbalay Bose](#), Sayan Nag, Sanjoy Chowdhury, Mohamed Elhoseiny | *Under review*
- **Diagnosing Evidence Utilization in Multimodal Document Question Answering**
Debolena Basak, [Digbalay Bose](#), Koustava Goswami, Maunendra Sankar Desarkar | *KDD Research track 2026*
- **Can Multimodal Foundation Models Help Analyze Child-Inclusive Autism Diagnostic Videos?**
Aditya Kommineni, [Digbalay Bose](#), Tiantian Feng, Shrikanth Narayanan | *Interspeech 2025*
- **Domain Adaptation for Sentiment Analysis Using Robust Internal Representations**
Mohammad Rostami, [Digbalay Bose](#), Shrikanth Narayanan, Aram Galstyan | *Findings of EMNLP 2023* [pdf]
- **MM-AU: Towards Multimodal understanding of advertisement videos**
[Digbalay Bose](#), Rajat Hebbar, Tiantian Feng, Krishna Somandepalli, Anfeng Xu, Shrikanth Narayanan | *ACM Multimedia 2023* [pdf]
- **SEAR: Semantically-grounded Audio Representations**
Rajat Hebbar, [Digbalay Bose](#), Shrikanth Narayanan | *ACM Multimedia 2023* [pdf]
- **Contextually-rich human affect perception using multimodal scene information**
[Digbalay Bose](#), Rajat Hebbar, Krishna Somandepalli, Shrikanth Narayanan | *IEEE International Conference on Acoustics, Speech and Signal Processing 2023* [pdf]
- **FedMultimodal: A Benchmark for Multimodal Federated Learning**
Tiantian Feng, [Digbalay Bose](#), Tuo Zhang, Rajat Hebbar, Anil Ramakrishna, Rahul Gupta, Mi Zhang, Salman Avestimehr, Shrikanth Narayanan | *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2023* [pdf]
- **A dataset for Audio-Visual Sound Event Detection in Movies**
Rajat Hebbar, [Digbalay Bose](#), Krishna Somandepalli, Veena Vijai, Shrikanth Narayanan | *IEEE International Conference on Acoustics, Speech and Signal Processing 2023* [pdf]
- **MovieCLIP: Visual Scene Understanding in Movies**
[Digbalay Bose](#), Rajat Hebbar, Krishna Somandepalli, Haoyang Zhang, Yin Cui, Kree-Cole Mclaughlin, Huisheng Wang, Shrikanth Narayanan | *Winter Conference on Applications of Computer Vision (WACV) 2023* [pdf] [project page]

Patents

- **Visually Guided Query Processing**
Ashok Pon Kumar Sree Prakash, Ayushi Dalmia, Amith Singhee, [Digbalay Bose](#), Sumanta Mukherjee, Raghavendra Singh, Vikas C. Raykar | *US Patent (US10878291B2), 2020* [pdf]

Work Experience

- January 2025-Present **Adobe Research, Research Scientist.**
- o Architected multi-agent systems that synthesize collections of documents into holistic visualizations, generating cohesive charts and infographics.
 - o Designed a multi-agent pipeline for controllable advertisement video generation, transforming source documents and websites into targeted, on-brand video content.
 - o Built a scalable VLM-as-a-Judge framework enabling multimodal, agentic evaluation of image and infographic inclusion in AI-generated presentations.
 - o Co-developed LLM-as-a-Judge evaluation metrics for document-to-podcast generation, shipped as part of Adobe Acrobat.
- May 2023-August 2023 **NVIDIA Corporation, Computer Vision & Graphics Intern.**
- o Built deep learning models for controllable portrait video animation in NVIDIA Maxine.
- May 2022-August 2022 **NVIDIA Corporation, Software Engineering Intern.**
- o Developed visual and audio-visual deep learning models for high-fidelity facial animation of avatars in Maxine ARSDK.
- July 2016-June 2018 **IBM Research, India, Research Software Engineer.**
- o Built an end-to-end soil-moisture extraction system from satellite imagery using image interpolation techniques, shipped as part of IBM Geospatial Analytics.
 - o Designed explainable deep learning models for image classification and visual search in the Cognitive Fashion retail effort.
- May 2013 - July 2013 **Indian Statistical Institute, Kolkata, Research Intern.**
- July 2013 **Advisor: Prof. Subhamoy Maitra, Applied Statistics Unit**
- o Developed a key recovery scheme based on the properties of Slid Pairs for stream cipher Salsa20.

Summer schools

- July 2020 - August 2020 **Oxford Machine Learning Summer School.**
- o Acceptance rate: 15% [[Certificate](#)]

Research Experience

- o **Research Assistant, Signal Analysis and Interpretation Laboratory, USC** 2018 – Dec 2024
Advisor: Prof. Shrikanth Narayanan
PhD focused on multimodal understanding of media content — advertisements, movies, and affect — with publications at KDD, ACM MM, WACV, and ICASSP.
 - **Advertisement video understanding [ACM MM 2023]**: Introduced a large-scale advertisement benchmark and multimodal models for semantic video understanding.
 - **MovieCLIP: Visual scene understanding [WACV 2023, w/ Google Research]**: Released a large-scale weakly-labeled dataset of movie shots and trained deep models for scene and genre classification on HVU and MovieScope.
 - **Multimodal federated learning [KDD 2023]**: Co-developed multimodal benchmark tasks and baseline models for federated learning.
 - **Context-driven affect perception [ICASSP 2023]**: Built a multimodal context fusion module for apparent emotion recognition on EMOTIC and CAER-S.
 - **Computational analysis of facial paralysis [Facial Plast. Surg. Aesthet. Med.]**: Built a web app for recording clinical sessions and a landmark-based pipeline with novel asymmetry measures for standardized score prediction (with Dr. Kochhar & Dr. Voelker, PNI).
 - **Selected applied collaborations**: Gender representation in media with the Geena Davis Institute (See Jane) [US '22, India '23].
- o **Research Scholar, Vision and Image Processing Lab, IIT Bombay** 2015 – 2016
Advisor: Prof. Subhasis Chaudhuri
 - **Master's thesis — Sparsity and metric learning for classification**: Developed a hierarchical fine-grained image classification scheme combining a self-tuning variant of spectral clustering with large-margin nearest neighbor.

Skills

- o **Languages**: Python, C/C++, JavaScript, R, Bash, SQL
- o **Deep Learning**: PyTorch, JAX, TensorFlow, HuggingFace (Transformers, Diffusers)
- o **LLM & Agent Frameworks**: LangChain, LangGraph, LlamaIndex, AutoGen, DSPy
- o **Fine-tuning & Serving**: PEFT, LoRA/QLoRA, vLLM, Ollama
- o **RAG & Evaluation**: FAISS, Chroma, Pinecone, LangSmith, Ragas, Weights & Biases
- o **Computer Vision**: OpenCV, Scikit-Image, PIL, Scikit-learn
- o **Creative & 3D**: Blender, Maya, ComfyUI, VTK

Selected Academic Projects

- **StyleIT: Style Guided Image Captioning** (CSCI 699, USC) [Report]
- **Future sales prediction using ensemble models (CSCI 567, USC)** [Report] [Code]
 - Obtained a world rank of 80 among 8292 teams in the Kaggle future sales competition
- **Visual Question Answering : Attention and Fusion based approaches (CSCI 599, USC)** [Report] [Code]
 - Awarded the best project by the poster session sponsors, Neudesic.
- **Multimodal Emotion Recognition from speech utterances (EE 599, USC)** [Slides]

Professional Service

- **Conference Review:**
 - 2026: CVPR, ECCV, ICLR, KDD
 - 2025: CVPR, BMVC, EMNLP, WACV
 - 2024: CVPR, ACM MM, BMVC, ICASSP
 - 2023: ACL, EMNLP, WACV, CoNLL, NeurIPS Phys, ACII
 - 2022: ACM MM, COLING
 - 2021: ICME
 - 2020: ICASSP

Invited Talk

- Invited talk on **Visual Scene Recognition in movies** at Netflix Research, 2023
- Invited talk on **Understanding context in movies: Taxonomies, Benchmarks and Challenges** at the [3rd Media Understanding Workshop on Context and Environment](#) by [Google Research](#) and [Center for Computational Media Intelligence, USC](#).

Selected Awards and Honors:

- **2025:** Recognized as an outstanding reviewer at CVPR 2025.
- **2021:** Phase 1 Finalist in **OpenCV AI Competition** as part of USC SAIL Team (top 200 teams globally).
- **2018:** Awarded **Annenberg Fellowship** by **University of Southern California** .
- **2016:** **Managers choice award** by **IBM Research, India** for research contributions in data-driven soil moisture modeling.
- **2016:** **Academic excellence award** by **IIT Bombay** for ranking among top 1% students of Masters in Electrical Engineering.
- **2014:** Secured All India Rank **251** out of 216367 candidates in **Graduate Aptitude Test in Engineering (GATE) 2014**

Selected Coursework

- **USC:** Advanced Computer Vision, Grounding Natural Language, Machine Learning, Deep Learning and its Applications, Affective Computing, Random Processes.
- **IIT Bombay:** Computer Vision, High-Performance Scientific Computing, Matrix Computations.

Teaching Experience

- **Fall 2020:** Teaching Assistant, EE 599: Deep Learning Systems (USC)

Mentoring

- **exploreCSR:** Mentored 3 senior and 1 freshman student for [exploreCSR](#) workshop.
- **USC Viterbi:**
 - Mentored 5 students spanning undergraduate (freshman through sophomore) and Master's levels across Computer Science and Electrical & Computer Engineering as part of CURVE project.