Rohit Gupta

Citizen of India, Currently present in US on F1 visa status

🛛 (+1) 4079905985 | 🕿 rohitgupta.hpf@gmail.com | 🌴 rohit-gupta.github.io | 🖸 rohit-gupta | 🖇 0WukQpMAAAAJ | 🖬 rohitguptahpf

Education	
2019 - Ongoing	PhD. in Computer Science, University of Central Florida, Orlando
2015 - 2017	M.Tech. in Computer Science and Engineering, IIT Kanpur, Kanpur, India
2010 - 2014	B.Tech. in Electrical Engineering, IIT Kanpur, Kanpur, India

Selected Publications

ECCV 2024	Open Vocabulary Multi-Label Video Classification
	R Gupta , MN Rizve, A Tawari, J Unnikrishnan, S Tran, M Shah
	(Link: https://www.ecva.net/papers/eccv_2024/papers_ECCV/papers/05599.pdf)
CVPR 2023	Class Prototypes based Contrastive Learning for Classifying Multi-Label and Fine-Grained Educational Videos
Citations: 15	R Gupta , A Roy, S Kim, C Christensen, T Grindal, S Gerard, M Cincebeaux, A Divakaran, M Shah
	Link: https://openaccess.thecvf.com/content/CVPR2023/html/Gupta_Class_Prototypes
AAAI 2023	Contrastive Self-Supervised Learning Leads to Higher Adversarial Susceptibility
Citations: 10	R Gupta, N Akhtar, A Mian, M Shah
	Link: https://ojs.aaai.org/index.php/AAAI/article/view/26733
ICPR 2020	RescueNet: Joint building segmentation and damage assessment from satellite imagery
Citations: 112	R Gupta, M Shah
	Link: https://ieeexplore.ieee.org/document/9412295
CVIU. Jun '22	TCLR: Temporal Contrastive Learning for Video Representation
Citations: 203	I Dave, R Gupta , M N Rizve, M Shah
	Link: https://www.sciencedirect.com/science/article/pii/S1077314222000376
MediaEval 2018	Linear Models for Video Memorability Prediction Using Visual and Semantic Features
Citations: 18	R Gupta, K Motwani
	Link: http://ceur-ws.org/Vol-2283/MediaEval_18_paper_31.pdf
	Cassandra: Detecting Trojaned Networks from Adversarial Perturbations
IEEE Access, Jul '21	X Zhang, R Gupta , A Mian, N Rahnavard, M Shah
Citations. 20	Link: https://ieeexplore.ieee.org/document/9502110
	DeepSAR: Vessel Detection in SAR Imagery With Noisy Labels
ICIP 2022	M Pillai, T Baweja, A Bhattacharya, R Gupta , M Shah
	Link: https://ieeexplore.ieee.org/document/9898020
IEEE J. Photovolt Citations: 62	Automated Defect Detection and Localization in Photovoltaic Cells Using Semantic Segmentation
	of Electroluminescence Images
	J Fioresi, D Colvin, R Frota, R Gupta , M Li, H P Seigneur, S Vyas, S Oliveira, M Shah, K O Davis
	Link: https://ieeexplore.ieee.org/document/9650542

Pre-Prints _____

Under Deview	ViLLaGE: Video LLM for Generative and Embedding Tasks
Under Review	R Gupta , J Unnikrishnan, S Tran, R Hamid, M Shah
Under Review	BrailleVision: Text Instruction Tuning of LLMs to Improve Visual Skills
	R Gupta , MN Rizve, P Tirupattur, M Shah
Under Review	StretchySnake: Flexible VideoMamba for Short and Long-Form Action Recognition
	N Siddiqui, R Gupta , Sirnam S, M Shah
	OpenReview: https://openreview.net/forum?id=SMIVEeoSyI
Under Review	Self-Supervision is Not All You Need: In Defense of Semi-Supervised Learning.
	R Gupta , MN Rizve, Sirnam S, N Kardan, M Shah
Under Deview	SB-Bench: Stereotype Bias Benchmark for Large Multimodal Models.
Under Review	V Narnaware, A Vayani, R Gupta , Sirnam S, M Shah
	arXiv: https://arxiv.org/abs/2502.08779)
Under Deview	DEDD-SEQ: Dual Encoder-Denoising Decoder for Video to GPS Sequence Prediction
Under Review	PP Kulkarni, R Gupta , M Shah
	(Private Pre-Print (Google Drive): https://drive.google.com/file/d/18XUXTc6fVDL1iAi6zqfDkMSyzChkWYSx)
Linder Deview	GAEA: A Tourism and Geo-localization Assistant
Under Keview	R Campos, A Vayani, PP Kulkarni, R Gupta , A Dutta M Shah
	(Private Pre-Print (Google Drive): https://drive.google.com/file/d/1memo4cmwpFby93JihlG527eQ7AeahpBw)
Under Review	TimeLogic: A Temporal Logic Benchmark for Video QA
	Sirnam S, R Gupta , H Kuehne, M Shah
	(Private Pre-Print (Google Drive): https://drive.google.com/file/d/1bRTCiP_J2MzEJhr5VqZgrHLgJMWi2bCr)

Applied Scientist Intern, Amazon Nile, Rufus Multi-Modal Team

- Developed a single model capable of solving a wide spectrum of video understanding tasks: Question Answering, Video Retrieval, Localization etc
- Built an unified Video LLM with the capabilities of outputting text as well as embeddings in order to solve all the tasks.
- Carried out medium scale distributed post-training on 10 million+ videos.

Applied Scientist Intern, Amazon Search Science and AI, M5 Team

- · Worked on multi-label open vocabulary video classification; recognizing objects and actions in videos not present in training data.
- Leveraged LLMs to assist CLIP Text Encoder by providing class attributes and designed Temporal Modeling adapters for the CLIP Vision Encoder
- Achieved competitive results across multiple video classification benchmarks in zero-shot open vocabulary setting. • Work resulted in a successful publication at ECCV 2024 main conference.

Research Intern, SRI International

- Was tasked with solving the problem of classifying youtube education videos into fine-grained topic categories within subjects.
- Developed multi-label, multi-modal prototype contrastive learning to solve fine-grained video content understanding.
- Work resulted in a successful poster publication at CVPR 2023 main conference, and was invited as an Oral presentation at a workshop. Sep. 2017 - Jul. 2019,

Research Engineer, Conduent Labs (erstwhile Xerox Research)

 Contributed to a variety of projects in Computer Vision: Video memorability prediction, Analyzing multi-modal data for smart-city applications, Instance recognition and image classification for augmented reality (AR) and appearance based re-identification of cars for traffic flow analysis.

Data Scientist, Fuzzy Logix

• Developer on DB Lytix[™] suite of machine learning, statistical and financial algorithms embedded into data warehouses like Teradata[™]& Netezza[™]

Achievements ____

2023	Oral Presentation , Workshop on Large Scale Holistic Video Understanding, CVPR
2021	1st Place and Jury Prize, VI-Priors Action Recognition Challenge, ICCV
2019	Fellowship, ORCGS Doctoral Fellowship, UCF
2018	1st Place, MediaEval 2018: Predicting Media Memorability Task
2010	National Rank 433 (Top 0.1%), Joint Entrance Exam, Indian Institutes of Technology

Service _

2025	Organizer, Workshop on Video Large Language Models, CVPR	
2022-present	Reviewer, CVPR, ECCV, ICCV, AAAI, ICLR, IEEE Journals (TIP, TNNLS, TCSVT)	
2020, 2022	Mentor, NSF Research Experiences for Undergraduates, UCF-CRCV REU Site	
2015-16	Teaching Assistant, Courses: Fundamentals of Computing, Machine Learning	IIT Kanpur

Research Overview

Joint Generative and Embedding Video LLM	Amazon
Internship and Graduate Research Project	2024
• Developed a multi-modal large language model (LLM) capable of generating both embeddings and text to solve video understanding	tasks.
Text Instruction Tuning of LLMs for Better Visual Understanding	UCF
Graduate Research as Lead Researcher	2024
• Designed a lightweight, specialized text instruction tuning stage for LLMs to improve their downstream multi-modal understanding p	performance.
Multi-Label Open Vocabulary Video Classification	Amazon
Internship and Graduate Research Project	2023
• Developed a video classification model capable of recognizing object and action classes not seen during training.	
Multi-Label Contrastive Learning for Fine-Grained Educational Video Classification	SRI & UCF
Internship and Graduate Research Project	2022
• Achieved state of the art results on a novel dataset of education videos and two prior benchmark datasets (YouTube-8M and COIN)	
Robustness of Contrastive Self-Supervised Representations	UCF
Graduate Research as Lead Researcher	2021
• Identified root causes of the adversarial vulnerability of contrastive self-supervised models and boosted the robustness of SSL by ab	out 5%
Conversational Tourism and Geolocalization Assistant	UCF
Graduate Research as Collaborator	2025
• Helped design a dataset and finetune a multi-modal LLM to create a geolocalization expert conversational assistant for tourism.	
Benchmarking Multi-Modal LLMs	UCF
Graduate Research as Collaborator	2025
 SB-Bench: Benchmark to identify if models end up relying on stereotypes in situations with limited information. TLQA: Benchmark to comprehensively test the video understanding capabilities of Video LLMs based on logical operators. 	
Temporal Contrastive Learning of Video Representations	UCF

GRADUATE RESEARCH AS COLLABORATOR

• Devised a SotA self-supervised contrastive learning approach for video classification by enforcing temporal distinctiveness in feature space.

May-Nov 2024, Seattle

May-Nov 2023, Palo Alto

May-Aug 2022, Menlo Park (remote)

Bangalore, India

Bangalore, India

Jul. 2014 - Jun. 2015,