Raja Kumar

PhD Student in Computer Science
University of Southern California

Los Angeles, CA
kumarraj@usc.edu
raja-kumar-58971010a
raja-kumar
Google Scholar

Summary

Research Interest Computer Vision, Vision Language Model (VLM), Multimodal Reasoning, Fine-grained Image and Video Understanding, Representation Learning, 3D vision

Summary At USC, my research focuses on fine-grained visual understanding of vision language foundation model. Specifically, I am exploring post-training methods to leverage foundation models for domain specific image and video understanding tasks under zero-shot/few-shot settings. During my MS at UCSC, I worked on 3D face reconstruction from few view images. Earlier, at samsung research I worked on Deep learning model quantization for efficient inference on resource constrained devices.

Education

Aug. 2024–April 2028 (expected) **University of Southern California**, *PhD Computer Science*, Advisor: **Prof. Ram Nevatia**

Sep. 2022–March 2024 **University of California Santa Cruz**, *MS in Computer Science*, Advisor: **Prof. James Davis**

July. 2015–May 2019 **Indian Institute of Technology (IIT) Kharagpur**, *B.Tech in Electrical Engineering*, Advisor: **Prof. Pranab K Dutta**

Publications

ICLR 2026, [Paper] [Code] DiVE-k: Differential visual reasoning for fine-grained image recognition. **Raja Kumar**, Arka Sadhu, Ram Nevatia

ECCVW 2024 [Paper] [Code] Few-shot Novel View Synthesis using Depth Aware 3D Gaussian Splatting. **Raja Kumar**, Vanshika Vats

ICCVW 2023, [Paper] Disjoint Pose and Shape for 3D Face Reconstruction. **Raja Kumar**, Jiahao Luo, Alex Pang, James Davis

ArXiv 2023, [Paper] 3D Face Reconstruction: Is Model-Based Classic Passive Stereo Competitive. Jiahao Luo, **Raja Kumar**, Alex Pang, James Davis

ArXiv 2023, [Paper] [Code] Task Oriented Conversational Modelling With Subjective Knowledge. **Raja Kumar**

IJCNN 2021, [Paper] [Code] Hybrid and Non-Uniform Quantization Methods Using Retro Synthesis Data for Efficient Inference.
Raja Kumar, Tejpratap GVSL, Pradeep NS

Patents

US Patent, 2024 Method and Electronic Device for Efficiently Reducing Dimensions of Image Frame.
Tejpratap GVSL, **Raja Kumar**, Dewashish Avinash Dharkar.
Patent No: US20230252602A1

US Patent, 2022 Method and Electronic Device for Quantizing DNN Model.
Tejpratap GVSL, Arun Abraham, **Raja Kumar**, ..., Prasen Kumar Sharma.
Patent No: US20230068381A1

Skills

Languages Python, C++, Assembly, Javascript, SQL
Libraries PyTorch, Pytorch3D, Tensorflow, Caffe
Tools git, AWS, OpenCV, OpenGL

Work Experience

Aug. 2024–Present **Graduate Student Researcher**, *IRIS Lab, USC*
- **Fine-Grained VL Process Reward Models (Ongoing):** Developing a vision-language PRM that verifies step-level reasoning on tasks requiring pixel-level visual grounding, using synthesized process-supervised data from hierarchical fine-grained attributes.
- **DiVE-k (ICLR 2026):** Proposed a framework using an LVLM's top-k generations as MCQ options to elicit differential visual reasoning via RL (GRPO), achieving state-of-the-art on five fine-grained benchmarks with up to 10% gains in base-to-novel generalization.

April 2024–July 2024 **Applied Science Intern**, *Flawless AI, Santa Monica*
- **3D Multomodal Facial Synthetic Data Generation:** Optimized 3D facial synthetic data generation pipeline and Advanced multi-modal understanding by augmenting synthetic data generation to integrate depth maps and positional encoding, creating a richer dataset for enhanced geometric cues

Sep. 2022–Mar. 2024 **Graduate Student Researcher**, *VIS Group, UCSC*
- **Few shot 3D Face Reconstruction (ICCV-W):** Proposed an end-to-end pipeline that disjointly solves for pose and shape using a 3DMM face prior with stereo matching, achieving topological consistency and improved accuracy over SoTA on few-view 3D face reconstruction.
- **Model-Based Passive Stereo for 3D Face Reconstruction:** Showed that classic two-view passive stereo regularized by a FLAME morphable model surprisingly outperforms SoTA deep learning methods for casual 3D face reconstruction, with comprehensive analysis of resolution, viewpoint, and compression factors.
- **Depth-Aware 3D Gaussian Splatting (ECCV-W):** Introduced monocular depth priors with a scale-invariant depth loss to constrain 3D Gaussian splatting under sparse views, achieving 10.5% PSNR, 6% SSIM, and 14.1% LPIPS improvements over vanilla 3DGS.

Jun 2019–Feb 2022 **Machine Learning Engineer**, *Samsung Research, Bangalore*

- Worked on DL model's quantization techniques for faster and accurate inference of models on edge devices like mobile phones.
- Proposed novel data-free quantization methods which improved the inference by 10-20% and fixed the accuracy issue for various in-house models.
- Reduced the dependency on training data for quantization.

Academic Experience

Reviewer Expert Systems With Applications, IJCNN, BMVC, ICLR Workshop

Teaching **USC**: Advanced Computer Vision, Introduction to AI, Database Systems, Introduction to Programming Systems Design.
UCSC: Analysis of algorithms, Computer Architecture, Computer Systems and Assembly Language.

Relevant Courses

Advanced Computer Vision	Deep Learning and its application	Computer Vision
ML for NLP	Adversarial and Trustworthy Foundation Models	Neural Computation