

VEDANT JOSHI

+1-(858)-305-4782

vejoshi@ucsd.edu

[linkedin.com/in/vedant-joshi](https://www.linkedin.com/in/vedant-joshi)

[vedrocks15.github.io](https://github.com/vedrocks15)

EDUCATION

University Of California San Diego

Master of Science in Computer Science; GPA : 4.0/4.0

Sep. 2023 – Mar. 2025

La Jolla, California

Indian Institute Of Information Technology Kottayam

Bachelor of Technology (Hons.) in Computer Science; GPA : 9.82/10.0 - Gold Medalist

Aug. 2017 – April 2021

Valavoor, Kerala

TECHNICAL SKILLS

Languages: Python, Java, C, C++, SQL, JavaScript, Golang

Developer Tools: PyTorch, Tensorflow, JAX, OpenCV, SciKit, TensorRT, Pandas, NumPy, Matplotlib, Onnx, Git, Docker, Open3D

Cloud Technologies: AWS, Microsoft Azure cognitive services, Google Cloud Platform

EXPERIENCE

Apple

June 2024 – September 2024

Video Engineering Intern

San Diego, California

- Improved the grounding/localisation performance of vision-language models (**V-LMs**) on benchmarks by **2.5%** through longer region descriptions using **Qwen2** LLM & novel visual adapters that supported **4x** higher resolution images.
- Encoded improved **3D** information in machine generated instruction tuning data by using **normalised point clouds & region graphs** to create stronger evaluation benchmarks for 3-D aware, **LoRA** fine-tuned VLMs.
- Worked on building image to text token co-relation metrics in V-LM pipelines such as **GROMA**, **Llava** & **CLIP**, to identify hallucinations & incorrect localisation cases which lead to a sizeable reduction in incorrect image descriptions.

UCSD Graduate Student Researcher

November 2023 – June 2024

Dr. Mark Christopher

UCSD, California

- Worked with domain experts to generate a **50% larger** eye disease pre-training dataset using text conditioned **diffusion pipeline** with **geometric constraints** imposed by 2D multi-view eye-scans to achieve highly variant training samples.
- Researched on glaucoma classification problem using a combination of multi-modal fusion, **Vision Transformers** and hard negative mining approaches to attain a **10%** AUC score improvement over SOTA.

Tonbo Imaging

February 2023 – July 2023

Computer Vision & Imaging Engineer - I

Bangalore, Karnataka

- Achieved a **20%** improvement in detection performance on thermal images by re-engineering the base layers & detection scales of **YOLO v5** which significantly improved the robustness of **Tonbo's** autonomous driving software stack for heads up displays.
- Reduced the error rate by **7%** in real time depth map generation from monocular videos using **PoseNet** & self-supervised view synthesis for **Nvidia Xavier NX** devices that improved the spatial awareness of self-driving systems.
- Leveraged Generative A.I. based **diffusion models** coupled with **neural style transfer** losses & **sub-pixel convolutions** to generate context specific thermal images from RGB videos to solve the data shortage problem for training perception models.

Vedantu Innovations

July 2021 – December 2022

EdTech Data Scientist - I

Bangalore, Karnataka

- Architected a real time graph based image search engine by repurposing joint embedding models such as **BYOL** & **SimCLR**, through [domain specific augmentations](#) which lead to a **72%** reduction in redundant elements in the search space.
- Productionized a novel solution for [profanity detection](#) in real time tuition classes that achieved a **10%** improvement in recall over regular expressions by using contrastive learning to perform **zero shot learning** on LSTMs for low resource languages.
- Improved the quality of matches returned by elastic search engine by **40%**, through creation of a text cleaning pipeline that used n-gram SimHashing & **Levenstein** distances to remove results that had no syntactic similarity with a particular text cluster.

TCS Rapid Labs

September 2020 – March 2021

Research Intern (Won Best Paper at IEEE 8th ICSCC)

Online

- Generated a **25%** character error rate on the task of single word lip reading from [videos](#) by re-modelling the **LipNet** model from word to character level along with an efficient **CTC loss** implementation that helped patients suffering from hearing deformities.
- Proposed [FYE0](#), an **attention** based LipNet model, which reduced the character error rate by **2.5%** through improved context signal generation & provide model transparency through heat maps that showcased predicted character & time frame alignment.

PROJECTS

Attention Splats : 3D Scene Editing | Prof. Manmohan Chandrekar

April 2024

- For interactive scene editing, boosted the text to vision alignment of 3D feature field inside the **Gaussian Splatting** framework using a novel approach of local and global **transformers**, along with distillation from **2D-foundational models** such as SAM.
- Proposed an innovative shift encoding **MLP** along with robust voxelization technique to handle **single GPU** memory constraints.

Edge device object detection | B.tech Hons. Project

January 2020

- Researched on mixed precision, **layer fusion** & quantization aware training for **SSD & YOLO** to achieve a **22 FPS** detection rate on a Nvidia Jetson Nano along with a **0.4** mAP result for camera mounted drones.