

Jaiminkumar Ashokbhai Bhoi

+13864568222 ajaymin28@gmail.com FL, USA - 32765
[in/ajaymin28](https://in.linkedin.com/in/ajaymin28) github.com/ajaymin28 ajaymin28.github.io

Professional Experience

- Graduate Research Assistant** *Center for Research in Computer Vision(CRCV)* **FL, USA** 05/2024 - 01/2025
- Proposed a novel solution for Dynamic Scene Graph Generation (DSGG) with VLMs and MLLMs, giving **10-40% performance improvement using just 5-10% of training data** across varying top-K metrics, while maintaining the recall-precision balance.
 - Efficiently finetuned SOTA Multi-Modal LLMs with Flash Attention using **High Performance Computing(HPC)** on Action Gnome and VidVRD datasets using PyTorch, Python, with customized frame selection using OpenCV and Decord.
 - Created** a novel dataset of **2D/3D shapes and real images** to evaluate reasoning limits in **MLLMs/VLMs** (Gemma3, InternVL3, Qwen2-VL), revealing critical **biases and failure cases** in geometric and in-context understanding of SOTA VLMs.
- Individual Contributor** *University of Central Florida* **FL, USA** 12/2023 - 04/2024
- Built** an automated system for tibia/femur angle measurement using segmentation with **Segment Anything Model** and a **light weight MLP classifier using Pytorch**, reducing manual review time of 3D scans(DICOMS) of the patient by **35%**
- Software Engineer / Research Associate** *Tata Consultancy Services* **Bangalore, India** 06/2018 - 09/2023
- Executed and delivered three projects to production: In-vehicle Infotainment System(IVIT), Computer Vision on Qualcomm RB5 (QCRB5) robotics board, and **Container Image Analytics(CIA)** that contributed to the goal of **saving billions in USD**.
 - Designed and implemented **C++ based video analytics solution** (theft detection, barcode switch on high value products) to **prevent losses in retail self-checkout environments**, addressing an industry-wide annual loss of **\$90B**.
 - Developed and deployed end-to-end **Computer Vision models and algorithms** that **saved \$4M** in container repair and cleaning costs, and **reduced lead time from 12 to 1 day** for 10% of repair volume while ensuring high accuracy and performance.
 - Deployed four end-to-end Computer Vision solutions developed in C++: Dlib **face recognition**, **PosNet**, **YOLO v3** based product detection, and queue counting on the **QCRB5(ARM64)** processor.
 - Quantized** Tensorflow models to **TFLite/ONNX** with **float16,int8** and **int4** precision for efficient on-device inference, **minimizing model size** trading negligible accuracy loss with inference speed.
 - Leveraged Android delegates (DSP, CPU, GPU, NNAPI)** to execute multiple Computer Vision solutions concurrently using **multi-threading and multi-processing** techniques, **increasing throughput** of edge devices.
 - Fine-tuned deep learning models using **Tensorflow** on production image datasets, achieving **over 90% accuracy** on defect and quality inspection using **image classification, object detection, and segmentation** tasks.
 - Built a **Continuous Learning Framework (CLF)** with customized **AzureML Ops**, reducing retraining efforts by **80%** and accelerating iteration cycles with **human-in-the-loop feedback** by developing custom annotation tools.
 - Deployed scalable multi-model APIs with **Flask/RestX** and **Docker** on **Azure Kubernetes**, leveraging auto-scaling to efficiently process **10k+ high-quality images per hour** through optimized **ONNX** hierarchical chained inference.

Education

Master's in Computer Vision (*University of Central Florida*) **FL, USA** 08/2023 - 04/2025
Bachelor's in Computer Engineering (*A. D. Patel Institute of Technology*) **GJ, India** 04/2014 - 03/2018

Skills

Python, C++, PyTorch, TensorFlow, Keras, OpenCV, Image Processing, Computer Vision, Image recognition, Classification, Object Detection, Tracking, Segmentation, NumPy, Machine Learning, NLP, Deep Learning, Transformers, Quantization, LLM, Software Engineering, MLOps, Slurm, Cmake, Docker, containerization, Kubernetes, AzureML, REST, API, Git, DeepSpeed

Patents & Research

- Method and system to detect a text from multimedia content captured at a scene. (**US-12333832-B2**)
- What can Off-the-Shelves Large Multi-Modal Models do for Dynamic Scene Graph Generation? ([ongoing](#))
- An efficient ensemble-based deep learning model for the diagnosis of cervical cancer (ISCAIE-22)

Projects

PaliGemma-Video	08/2025 - current
Can Visual Language Models(VLMs) understand shapes?	02/2025 - 05/2025
Video Understanding (Using Neuro Symbolic AI)	02/2024 - current
Human Activity Recognition(HAR) on Static Images using CLIP	08/2023 - 12/2023
Container Image Analytics (CIA)	02/2021 - 06/2023
Computer Vision on Qualcomm RB5 Robotics Development Board	06/2020 - 02/2021
Image and Video Analytics Toolkit (IVTK)	04/2020 - 06/2020