Jaiminkumar Ashokbhai Bhoi

+13864568222 ajaymin28@gmail.com Oviedo, FL, USA - 32765
in/ajaymin28 github.com/ajaymin28 𝕏 ajaymin28.github.io

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*

Professional Experience

Graduate Research Assistant *Center for Research in Computer Vision(CRCV)* **FL, USA** *05/2024 - 01/2025*

- **Designed and implemented** an MLLM-based method for dynamic video scene graph generation, **improving performance by 10-40%** for different top-Ks and achieving **state-of-the-art performance** (SOTA) on benchmark datasets.

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 classifier**, reducing manual review time of 3D scans(DICOMS) of the patient by **35%**

Software Engineer / Research Associate *Tata Consultancy Services* **Bangalore, India** *06/2018 - 06/2023*

- **Executed and Delivered** three projects to production: **IVI system**, **Computer Vision on QC RB500 board**, and **Container Image Analytics** that directly impacted human lives and **saved billions in USD**.
- **Delivered** 60+ sprints, aligning feature delivery with the product roadmap through close **collaboration with Product Manager**, Scrum Masters, and subject matter experts as well as coss-functional teams.

Skills

Python ,C++ **PyTorch**, TensorFlow, **OpenCV**, **Computer Vision**, **Image recognition**, Video Analytics, Classification, Detection, Segmentation, Tracking, **Classical Machine Learning**, Deep Learning, Transformers, Quantization, LLMs, Software Engineering, Deployment, **MLOps**, **Docker**, **AzureML**, feature extraction, Edge deployment, **REST APIs**, Git, SQL, Web Development.

Projects & Research

Video Understanding (Using neurosymbolic AI approach) *Ongoing Research* *02/2024 - current*

- **Proposed a novel solution** for Dynamic Scene Graph Generation (DSGG) with MLLMs, demonstrating a **10-40% performance improvement using just 5-10% of training data** across varying top-K metrics, while maintaining the recall-precision balance.
- Continuously evaluated and experimented with SOTA models and methods for generating DSGG.
- Efficiently finetuned SOTA MLLMs(Video-LLAVA, LLaVa-OneVision, InternVL2) with Flash Attention using High Performance Computing(HPC) on **Action Gnome** and **VidVRD** datasets.
- **Benchmarked** and analyzed model's performance **demonstrating reduced predicate perplexity** after finetuning as well as maintaining performance for long-trail predicates.

DumbVLMs *(Visual Language Models)* *02/2025 - 05/2025*

- **Created** a novel dataset of **2D/3D shapes and real images** to evaluate reasoning limits in **MLLMs/VLMs** (LLaVA-One-Vision, InternVL3, Qwen2-VL), revealing critical **biases and failure cases** in geometric and in-context understanding of SOTA VLMs.
- **Generated 14k synthetic images** and **50k VQA queries** for robust, scalable evaluation of multimodal models.
- **Collected 200 real images** to support **shape/object matching, odd-one-out, and rotation reasoning** evaluation tasks.

Container Image Analytics (CIA) *02/2021 - 06/2023*

- Developed and deployed **Computer Vision 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.
- Fine-tuned deep learning models on **production image datasets** using **TensorFlow**, achieving **over 90% accuracy** in defect and quality inspection for **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.
- Distilled and quantized for **ARM processors**, and developed a cross-platform MVP in **Flutter** for edge deployment.

Computer Vision on Qualcomm RB5 Development Board *06/2020 - 02/2021*

- Designed and implemented a **video analytics solution** to **prevent losses in retail self-checkout environments**, addressing an industry-wide annual loss of **$90B**.
- **Deployed 4 Computer Vision solutions Dlib face detection**, **PosNet-based theft detection**, **YOLO v3 ticket-switch detection**, and queue counting in **C++** on the **Qualcomm RB5 board**, advancing edge AI capabilities.
- **Compiled and optimized OpenCV and Dlib** with **OpenBLAS** for **ARM**, achieving a **4x performance boost** on edge hardware.
- **Quantized** models to **TFLite** for efficient on-device inference, **minimizing model size** with negligible accuracy loss.
- Developed and **optimized** real-time **computer vision pipelines** for deployment on edge.
- **Leveraged Android delegates** (**DSP**, **CPU**, **GPU**, **NNAPI**) to execute multiple Computer Vision solutions concurrently, **increasing throughput** of edge devices.

Human Activity Recognition on Static Images (HAR) *Github* *08/2023 - 12/2023*

- Addressed the challenge of **static-image action recognition** by fine-tuning a CLIP model, enabling **accurate classification** of human activities and **improving interpretability** using self-attention visualization.
- Applied zero-shot learning technique to **address mutually inclusive human actions**, increasing classification robustness in scenarios with limited labeled data.

Self-Supervised Distillation with No Labels on X-ray Images *Github* *02/2024 - 03/2024*

- Tackled the lack of annotated X-ray data by **fine-tuning a DINO self-supervised model** on chest X-rays for pneumonia classification, **achieving a 95.5% test accuracy** and demonstrating strong generalization.
- Enhanced model deployment by **distilling knowledge into smaller ViT** and EfficientNet architectures, **reducing inference latency and resource usage** for real-world applications.

DinIE (Distillation with No Labels for Image and EEG) *Ongoing Research* *01/2024 - Current*

- Achieved near image-level representation of EEG visual stimuli features by **training** a self-supervised learning model using the **DINO framework**, enabling better alignment with visual representations.
- Used **Domain Invariant Learning** framework to learn subject-invariant features to **promote cross-subject retrieval**.
- Actively working on improving cross-session, cross-subject understanding for EEG data with multi-modal self-supervised learning.