

Jaiminkumar Ashokbhai Bhoi

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

Education

Master's in Computer Science (Specialized in Computer Vision)

FL, USA 08/2023 - 04/2025

[University of Central Florida](#)

Bachelor's in Computer Engineering

GJ, India 04/2014 - 03/2018

[A. D. Patel Institute of Technology](#)

Professional Experience

Research Assistant [Center for Research in Computer Vision\(CRCV\), University of Central Florida](#) **FL, USA** 05/2024 - 01/2025

- Achieved SOTA performance in video scene graph generation by designing and implementing a multimodal LLM framework, yielding a 12% accuracy improvement.

Individual Contributor [College of Medicine, University of Central Florida](#)

FL, USA 12/2023 - 04/2024

- Developed an automated angle measurement system for tibia and femur alignment using segmentation (SAM) and classification head, reducing manual inspection time for 3D patient data by 35%.

Software Engineer/Research Associate [Cognitive AI, Tata Consultancy Services](#)

Bangalore, India 06/2018 - 06/2023

- Successfully delivered three impactful projects: In-vehicle Infotainment System, Computer Vision on the QC RB500 board, and Container Image Analytics, along with several Image processing-based MVPs and produced patented computer vision solutions.
- Practiced and applied Agile Methodologies at a large-scale project delivering 32+ sprints. Collaborated with the product manager and scrum master to identify backlogs and features that served the business needs.
- Led an initiative in an adaptive environment to automate unit and pre-integration testing with simulators in a multi-node environment, as well as a new hardware setup for the physical real-time testing on assets and binaries.
- Managed a team of two junior developers to design, develop, and deploy multiple production-grade deep learning models on the cloud as well as edge devices for quality assurance use cases that helped multiple stakeholders save up to 60% of cost in quality inspection and maintenance.

Skills & Interests

- Programming & Tools:** Python, TensorFlow, PyTorch, OpenCV, Docker, Kubernetes, Flask, RESTful APIs, Git version control, SQL, Javascript, Android Studio, Software Engineering
- AI/ML Techniques:** Computer Vision, Object Detection/Segmentation/Tracking, Pruning, Quantization, Image Processing, Multimodal models, Machine Learning, Deep Learning, Video Scene Graph Generation, Self-Supervised Learning(SSL), algorithm development, Natural language processing, neural networks, transformers, Large Language Models(LLM), Prompt Engineering, Chain of Thoughts, RAG
- Platforms & Frameworks:** Cloud computing , AzureML, NVIDIA Jetson, ARM64, Android, Linux, Edge Computing
- Soft Skills:** Team Leadership, Analytical, Agile Development, Problem solving, Cross-Functional Collaboration

Patents & Publications

- What can Off-the-Shelves Large Multi-Modal Models do for Dynamic Scene Graph Generation?
- An Efficient Ensemble-Based Deep Learning Model for the Diagnosis of Cervical Cancer: (ISCAI-2022)
- Aerial Video Analytics-based dynamic Non-linear distance measurement between on-ground objects.
- Method and system to detect a text from multimedia content captured at a scene.

Projects & Research

Computer Vision on Qualcomm RB500 Development Board

06/2020 - 02/2021

- Spearheaded the design and development of various computer vision solutions including face detection (utilizing Dlib), self-checkout theft detection (employing PosNet), barcode-switching detection (leveraging YoloV3), queue counting, and person tracking heatmap generation, all implemented in C++.
- Engineered Java Native Interfaces (JNI) to facilitate seamless communication between C++ and Java components for Android applications.
- Optimized performance on ARM64 devices by compiling custom Android libraries for OpenCV, Qualcomm Snappy, and Tensorflow, and integrating OpenBLAS for compiling DLIB, resulting in a notable performance enhancement of up to 4x.
- Successfully quantized deep learning models to enable their deployment on Android edge devices with minimal loss of accuracy.
- Demonstrated expertise in leveraging various Android delegates, including DSP, CPU, GPU, and NNAPI to concurrently execute four computer vision solutions on edge devices.

Container Image Analytics (CIA)

02/2021 - 06/2023

- Fine-tuned deep learning models including VGG16, MobileNetV2, and RCNN, achieving over 90% accuracy on production data.
- Designed and implemented a Continuous Learning Framework (CLF) using AzureML, significantly reducing manual training efforts by 80%.
- Developed REST APIs for deep learning models, leveraging AzureML, Docker, Flask/RestX, and Azure Kubernetes. These APIs efficiently handle 10,000+ requests per hour with auto-scaling capabilities, resulting in optimized cloud resource usage and a 60% reduction in cloud service costs.
- Engineered and delivered multiple user interfaces for data collection and validation purposes, utilizing C# Blazor, HTML/CSS, and jQuery-based web pages, reducing customers' efforts and providing AI-aided judgment.
- Quantized and distilled models for ARM-based processors and developed an MVP using a Flutter app for edge inference, utilizing method channels for cross-device code development for speed and efficiency of the models.

Human Activity Recognition on Static Images (HAR) [Github](#)

08/2023 - 12/2023

- Fine-tuned CLIP model for classifying human actions based on a static image, leveraged few-shot capabilities for identifying mutually inclusive actions. Leveraged self-attention maps for model explanation and reasonable AI.

Self Checkout Theft Prevention (RetailEye) [Github](#)

08/2023 - 12/2023

- Designed an innovative solution for tracking customer self-checkout experience to prevent theft. Used pose estimation & classification to track the sequence of activities performed by customers to find anomalies. Ported the solution to NVIDIA Jetson device for on-edge device computation, as well as adhering to user privacy.

Self-Supervised Distillation with No Labels on X-ray Images [Github](#)

02/2024 - 03/2024

- Leveraged off-the-shelf DINO model and fine-tuned on Chest X-ray images for Pneumonia classification, achieving 95.5% on test data. Distilled models to smaller ViTs and EfficientNets for speed and efficiency.

Video Understanding (with the Neurosymbolic approach)

02/2024 - 01/2025

- Generated zero-shot dynamic scene graph using prompt engineering and chain of thoughts utilizing SOTA Video LMMs.
- Fully finetuned Video LMMs as well as LORAs for customized and improved scene graph generation.
- Evaluated large-scale video datasets: Action Gnome and VidVRD.
- Investigated perplexity of predicted scene graph triplets, revealing a significant reduction in predicate perplexity after fine-tuning, offering insights into the model's behavior.

DumbVLMs (Dataset benchmark for 2D/3D shapes understanding, and sub features match)

02/2025 - 05/2025

- Constructed a dataset for various shapes with various combinations to benchmark SOTA VLMs (LLaVA-one-vision, Qwen-VL-2, Paligemma, InternVL3) for reasoning capabilities, yielding weaknesses and biases of different VLMs.