

Jaiminkumar Ashokbhai Bhoi

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

Education

Master's in Computer Science (specialized in computer vision)

FL, USA 08/2023 - 05/2025

[University of Central Florida](#)

Professional Experience

Research Assistant [CRCV lab, University of Central Florida](#)

FL, USA 05/2024 - present

- Designed and implemented a novel video scene graph generation framework using multimodal large language models, achieving a 12% improvement in benchmark accuracy, setting SOTA performance.

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 segmentations and classification, reducing manual inspection time for 3D patient data by 35%.

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

Bangalore, India 06/2018 - 06/2023

- Led the development of a patented crowd anomaly detection system using optical flow to identify sudden velocity and directional changes, enhancing public safety monitoring.
- Delivered a COVID-19 safety compliance solution for face mask detection, deployed across 100+ facilities.
- Designed and developed production-grade computer vision solutions on Azure Cloud as well as Edge, reduced cloud service costs by 60% through optimized scaling.

Skills & Interests

- Programming & Tools:** Python, C++, TensorFlow, PyTorch, OpenCV, Docker, Kubernetes, Flask, REST API
- AI/ML Techniques:** Computer Vision, Multimodal AI, Deep Learning, Video Scene Graph Generation, Self-Supervised Learning
- Platforms & Frameworks:** AzureML, NVIDIA Jetson, ARM64, Android Edge Devices
- Soft Skills:** Team Leadership, Agile Development, Cross-Functional Collaboration

Patents & Publications

- What can Off-the-Shelves Large Multi-Modal Models do for Dynamic Scene Graph Generation? (CVPR 2025, Tier-1 Conference)
- 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 (Patent Filed)
- Method and system to detect a text from multimedia content captured at a scene (Patent Published, 2023)

Projects & Research

EEGVis (Understanding Visually Evoked Potentials of EEG signals) [Github](#)

01/2024 - present

- Researching on human brain signals to decode visual perception capabilities through EEG signals.

Neuro-Symbolic AI based Video understanding

04/2024 - 11/2024

- Designed and implemented a video scene graph generation framework using multimodal large language models (MLLM), achieving state-of-the-art (SOTA) performance. this work is submitted to the prestigious CVPR (Tier-1 conference).

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.

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 models for ARM-based processors and developed a Flutter app for edge inference, utilizing Method channels for cross-device code development for speed and efficiency of the models.

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++.
- 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.
- Engineered Java Native Interfaces (JNI) to facilitate seamless communication between C++ and Java components for Android applications.
- Demonstrated expertise in leveraging various Android delegates including DSP, CPU, GPU, and NNAPI to concurrently execute four computer vision solutions on edge devices, showcasing adept multi-threading capabilities.