AJITH BALAKRISHNAN

AI Research Engineer

Computer Vision — NLP — GenAI

Email: ajithkannampara@gmail.com
Mobile: +91 9400763896
LinkedIn: linkedin.com/ajith2094
Github: github.com/Ajithbalakrishnan
Scholar: scholar/Ajithbalakrishnan

Personal Profile

Aspiring Machine Learning Engineer with 7 years of experience with a knack for developing solutions to challenging problems involving complex data. Critically skilled in analysing data & deducing actionable insights to provide robust solutions using concepts of Machine Learning, Computer Vision, and NLP.

Technical Skills

- **Languages:** Python3, C, C++, Go, Kotlin, SQL.
- **DL Libraries:** Pytorch, Tensorflow, Keras, and Darknet.
- **Tools/Libraries:** ONNX, Flask, Deepstream, TensorRT, FastAPI, Django.
- **Gen-AI Stack:** Huggingface, Langchain, Langgraph, Langsmith, Ollama.
- **Cloud Platforms:** AWS (EC2, S3, Bedrock), Azure (Azure ML Studio).
- **Computer Vision:** OpenCV, TensorFlow Object Detection API, Detectron2, TAO Toolkit, MMDetection.
- **CNN Models:** YOLO, F-RCNN, SSD, U-Net, ResNet, RetinaNet, etc.
- **NLP Libraries:** SpaCy, Hugging Face Transformers, BERT, T5, mBART, etc.
- **LLM Architecture:** Llama-3, Mistral, Phi, Donut, InterVL2, PaliGemma, ColPali, etc.
- **Databases:** MySQL, PostgreSQL, MongoDB, Weaviate.
- **MLOps Tools:** MLflow, Airflow, Tensor Board, Jenkins.
- **Data Engineering:** Apache Spark, Kafka, Airflow.
- **Visualization Tools:** Tableau, Power BI, Plotly, Dash.
- **Version Control:** Git, GitHub, GitLab, Bitbucket, Azure Git.
- **Deployment and DevOps:** Docker, Kubernetes, Jenkins, Ollama.
- **Hardware Skills:** Jetson Nano, TX2, Xavier, OAK, Raspberry Pi, Arduino-Uno, ESP 32.

Experience

Senior Data Scientist **Sept 2024 – Present**

Moder (Archwell) *Bangalore (India)*

Project: DASH (Expanse AI) - Automated Document Processing Platform

- Leading the development of GenAI-based document extraction based on the agnetic **RAG** framework and deployed in **AWS ECS** framework.
- Design and develop multi-model feature extraction for context-aware information retrieval using **locally hosted LLMs** and **Bedrock models**.
- Finetuned Small-Vision-Language models for VQA and document classification for finance and insurance documents based on **InternVL-2**, **Donut** architecture, resulting in an accuracy jump by **+12%.**
- Researching efficient multi-vector retrieval using **layout** & **context-aware** information representation with the help of **MM-LLM** and **Weaviate DB** and achieved **98.6%** extraction accuracy.
- Collaborated effectively within Agile teams to deliver high-quality solutions. Experienced in **Scrum** and **Kanban boards**, utilizing tools like **Slack** and **Jira** for project management.

Associate Research Staff Member **Sep 2019 - Aug 2024**

V-Labs (SLK Softwares) *Bangalore (India)*

Project: Lockbox- Automated Information Retrieval from Scanned Docs

- Led the development and deployment of an Information extraction solution based on a hybrid ensemble approach with a **six-sigma confidence** level (3.4 errors per million).
- Created small VLM using **ViT** encoder and **mBART** decoder as a pipeline architecture to perform information extraction for handwritten documents with OCR task as the pretraining approach.
- Proposed **FasterRCNN**-based **Chargrid** algorithm to improve localization F1-score by 13.45 on documents.
- Developed an advanced **OCR** algorithm based on a **vision-language transformer**, resulting in a **17%** reduction in CER compared to CRNN alternatives.

Project: SentixAi- An Intelligent Finance Advisory Platform

- Developed an LLM fine-tuning framework using **PEFT-LoRA** and **QLoRA**, with supporting trainers like **SFTT**, **DPO**, and **ORPO** methods with **flash attention** for ABSA & IR tasks.

- Developed RAG pipeline using Python & **LangChain** with **GPT4** and various **Llama** models for intelligent document extraction.
- Created solutions based on Instruction fine-tuning and soft prompt-tuning on compound ABSA tasks using LLMs **FlanT5**, **Llama2**, and **Mistral**.

Project: Face Recognition: A Distributed Edge Computation Framework

- Developed a Multi-stream object detection pipeline using Python, **DeepStream** & **Jetson EdgeAI** with **DeepSORT** and **IOU trackers**.
- Finetuned **YOLO-v4** and **SSD** models with **QAT** using **NVIDIA-TAO** and achieved an accuracy of **0.97%+** for Near-field and Far-field detection datasets(Custom).

Electronics Development Engineer **April 2017- Jul 2017**
CINESOFT Research & Development *Kerala (India)*

- Implimented and deployed a **ResNet18**-based deep learning model to automatically identify anomalies in machine welding and robotics assemblies, resulting in a **30%** reduction in quality check time.
- Developed a **VGGNet**-based anomaly detection system to ensure product quality by identifying deviations in dimensions and shapes. This system achieved a **74%** reduction in FP while maintaining a **90%** STP.

R&D Embedded Engineer **Oct 2015 - April 2017**
ARVIN Technologies *Kerala (India)*

- Successfully executed projects such as road line detection using **MLP**, **RPi** & OpenCV, and the creation of a bespoke Bluetooth Keyboard with microcontroller integration.
- Developed Computer Vision algorithm to detect straight lane line markings on the road using **OpenCV** Image Processing, Color Masks, **Canny Edge Detection**, and **Hough Transform**.

Education

Master of Technology in Robotics and Automation (GATE 2017) **2017-2019**
College of Engineering Trivandrum (KTU), Kerala *8.01/10 CGPA*

Bachelor of Technology in Electronics and Communication Engineering **2011-15**
Vidya Academy of Science and Technology, Calicut University, Kerala *6.7/10 CGPA*

Internships

- **Research Intern - CTO ST** | *Wipro LTD, Bangalore (India)* Feb 2019 - Jul 2019
- **Research Intern - Robotics** | *MyRobo Robotics Academy, Kerala (India)* May 2018 - Feb ,2019
- **Industrial Automation by Bosch Rexroth** | *CET Centre of Excellence in Automation Technologies, Kerala, 2019*

Publications

Refine3DNet: Scaling Precision in 3D Object Reconstruction from Multi-View RGB Images using Attention
– Ajith , Dr.Sreeja, Dr.Linu Shine. | ICVGIP, IIIT Bangalore, 2024 | Github | Paper*

Academic Projects

A Novel Hybrid Ensemble Approach For 3D Object Reconstruction from Multi-View Images
Research Intern, Wipro LTD, Bangalore (India)

- Studied different attention mechanisms in latent space and feature diffusion from multiple images. The proposed novel Generator-refiner network with attention layers outperformed existing models by 4%.

Certifications & Licenses

- Oracle Cloud Infrastructure 2024 Generative AI, Oracle.
- Machine Learning Operations (MLOps), Coursera.
- Machine Learning in Production, Deeplearning.Ai
- Mathematics for Machine Learning: Linear Algebra, Coursera.
- Neural Networks and Deep Learning, Coursera, Deeplearning.ai.
- Convolutional Neural Networks for Visual Recognition, Stanford - Spring 2018 - CS231N.
- Getting Started with AI on Jetson Nano, NVIDIA DALI.
- Advanced PG Diploma in Embedded Systems & VLSI Technology from SMEC Labs, Ernakulam.