JAZIB MAHMOOD

8563983918 • mahmood.j@zentreelabs.com • linkedin.com/in/jazib-computer-vision • github.com/jazib-sudo

Seasoned deep learning professional with 7+ years of expertise in NLP and Computer Vision. Specialized in model optimization and hardware-specific kernel development, consistently delivering innovative solutions at the forefront of deep learning and hardware integration.

PROFESSIONAL EXPERIENCE

Zentree Labs, Hyderabad, Telangana: Sr Research & Dev Engineer Jan 2023 – Present

- Sparsification, Pruning and Quantization of deep learning models
- Developing Core for LLM-based applications using frameworks like langchain,

AttainU Pvt Ltd, Bengaluru, Karnataka: Technical Learning Head Aug 2021 – Oct 2022

- Serving as the lead Data structure and Algorithms instructor for over ten batches. (Python, JS)
- Contributed to the development of the startup and oversaw over ten employees of the instructor team.
- Designed the curriculum for the Data Structures and Algorithms module.

Deep Learn Labs Pvt Ltd, Hyderabad, Telangana: Sr R&D engineer Dec 2017 – Dec 2019

- Algorithmic Research & Development
- Explored Cuda kernel programming and implemented various deep-learning operators on specific devices(x86,CUDA,ARM

VIZONCRAFTS SOFTSOLS PRIVATE LIMITED, Hyderabad, Telangana: ML engineer Jan 2014 - Sept 2017

- Algorithmic Research & Development
- responsible for designing and implementing robust models for NLP and Computer Vision applications, ensuring their accuracy and efficiency.

INDUSTRY PROJECTS

ZentreeLabs , Hyderabad, Telangana: R&D dev Engineer July 2023 – Dec 2023

- Developed RAG-based pipeline using langchain for developing a conversational AI agent that generates responses from OpenAI's chatGPT LLM for answering client's queries regarding the financial information of public/private companies, present in the redShift database in AWS.
- Removed the Hallucination problem from LLM's response.
- Finetuned an auxiliary deep-learning model (BERT) to refine and enrich the query provided by the client to the LLM.

ZentreeLabs , Hyderabad, Telangana: R&D Engineer Jan 2023 – June 2023

- Performed Quantization and Pruning of various computer vision deep-learning models, like YOLO, SSD, Resnet and VGGx.
- Benchmarked various quantization strategies.
- Made model ready to be deployed on edge device with significant memory and power consumption reduction.

Deep Learn Labs Pvt Ltd , Hyderabad, Telangana: Sr. R&D Engineer Feb 2019 – Dec 2019

- **Traffic Counts**
- The Main contribution towards this project is Algorithm development and Optimization.
- Literature survey on ML-based similarity measures and gauging the performance on the traffic counts data.
- Implemented Correlation best match selection for tracking, later optimized it for speed using DFTs.
- Optimization of the inference pipeline using TensorRT implementations of Vehicle Detector and respective post-processing functions.

Deep Learn Labs Pvt Ltd , Hyderabad, Telangana: Sr. R&D Engineer Feb 2018 – Oct 2018

- **Harris Corner Detector.**
- Understanding the algorithm.
- Reverse engineering the existing implementation.
- Proposing the architecture and optimization possible to fit on the hardware.

Vizon Crafts Pvt Ltd , Hyderabad, Telangana: ML engineer Jan 2016 – Nov 2017

- **Driver In-Attentiveness Detection**
- Designing the architecture that will alert the driver on time using the concepts of computer vision and Machine Learning.
- Detecting the required features using computer vision-based algorithms and training the systems with machine learning algorithms to detect the driver's inattentiveness.

Vizon Crafts Pvt Ltd , Hyderabad, Telangana: ML engineer trainee Oct 2014 – Sept 2015

- **JEWELUKZ**
- Devised guidelines in data collection required for the JEWELUKZ application.
- Backend processing of the jewellery data – Stitching, Matting and deciding on the resolutions of the jewellery data required for Tab, Web and Kiosk-based solutions.
- Implementing Matting techniques to be followed for realistic blending.

ACADEMIC PROJECTS

Hear Me Out: Fusional approaches for audio augmented temporal action localization 2022

Oral presentation at 17th International Conference on Computer Vision Theory and Applications (VISAPP) 2022

- Designed a State-of-The-Art multimodal fusion approach for temporal activity localization using residual multi-modal attention.

Activity detection model for Denso Corporation, Japan 2019 - 2020

Designed and developed a fully supervised DL model for activity localization and evaluation

TECHNICAL SKILLS

Deep learning Frameworks: Pytorch, Keras, Tensorflow, langchain, numpy, pandas
Deep learning Architectures: CNNs, GCNs, Transformers, BERT, LLMs - Llama7b, PalM, Falcon, T5
Deeplearning Tools: Octave, Netron
Programming: Python, C, C++, JavaScript, Git, MATLAB, Optimal DL operators implementation on specific hardware(x86,ARM,CUDA), Kernel programming in CUDA
Certifications:

- Deep Neural Networks with PyTorch (https://coursera.org/share/591c2d9f35a4c69f4421ed84af52b284) – June 2022,
- Neural Networks and Deep Learning(https://coursera.org/share/be62a3150e2d92e3a450a48532708893) - June 2022

EDUCATION

M.S. by Research, CSE May 2022
IIIT, Hyderabad 7.8 CGPA
Centre for Visual Information Technology, Lab
Relevant coursework: Statistical Methods in AI, Computer Vision, Deep learning

ACHIEVEMENTS

GATE AIR 673 2019
Graduate Aptitude Test in Engineering (GATE) - 2019 rank = 673 All India Rank(99.37 percentile))

VISAPP 2022 2022
Published and paper in VISAPP 2022 and gave Oral presentation at the conference