Jay Lal

linkedIn.com/in/jay-lal | github.com/thejaelal | jaylal3013@gmail.com | +1 7165208003

Available for full time roles starting Jan 2024 - Open to relocation in USA

Skills

Languages : Python, C/C++, Java | **Database Technologies:** SQL, NoSQL, Hadoop, Spark, Hive
ML Toolkit: PyTorch, Tensorflow, Keras, Hugging Face, MLFlow, Scikit-Learn, OpenCV, CUDA, NLTK, Spacy, LangChain, LlamaIndex, Numpy, Pandas, Matplotlib
Deployment:AWS, Google Cloud Platform (GCP), REST, Flask/Django, TensorRT, ONNX, Ray, Docker/Kubernetes,Git,Linux
ML Lifecycle: Problem Mapping, Data Processing, Model Training & Evaluation, Optimization, Deployment

Experience

AI Researcher *(AI Innovation Lab, SUNY Research Foundation, NY, USA)* *March 2022 - Present*

- Implemented RAG-based **LLM** chat on scientific articles, incorporating **multi-modal** data using custom image embeddings tailored for charts and Sci-BERT text embeddings with **LlamaIndex**, for aiding academic research.
- Developed a **Generative AI** based natural language chart image editing solution using LLMs and **Diffusion models**.
- Led the research for robust transformer-based **Action Recognition** in Soccer **videos** using multi-task learning.
- Led the research on **Transformer**-based chart **Instance Segmentation**, achieving a 17% increase in data extraction accuracy for real-world complex multi-line chart images**.**

Machine Learning Engineer *(Karza Technologies, Mumbai, India)* *April 2020 - January 2022*

- Developed system for data-efficient novel class **Object Detection** by augmenting YOLO with cascading classifier.
- Implemented **GAN** based text data generation in Few Shot text image stylization for text recognition on novel styles.
- Improved **OCR** accuracy by **11%** using synthetic feature supervision and **Spatial Transformer Networks.**
- Implemented **Information Extraction** in semi-structured documents using **Graph Neural Networks (GNN)**.
- Contributed towards developing an automated Financial Statement analysis system from PDFs/Images, for extraction and categorization of text data using **Natural Language Processing (NLP)** and **Image Processing**.
- Increased throughput and **reduced latency** by 15% for several **Deep Learning models** on GPU with APIs serving more than **10,000 requests** per day in **production** using **TensorRT** and other pipeline optimizations.

Computer Vision Engineer *(Artivatic.ai, Bengaluru, India)* *August 2019 - March 2020*

- Developed an end to end system for automated extraction of Handwritten Insurance Forms using **Deep Learning** and **Computer Vision**, scaled using **Ray**, resulting in savings of 750+ man-hours per year.
- Optimized these models for on-device inference using model compression techniques like **Knowledge Distillation** and **Quantization**, resulting in 80% model size reduction and **30% faster** inference.

Graduate Analyst *(BARCLAYS, Pune, India)* *July 2018 - August 2019*

- Designed and executed an automated team-performance insight system by aggregating data from heterogeneous sources, processed using **Spark** and **Hive** and visualized team metrics for enhanced decision-making.

Projects

- ***Hallucination detection in Dialogue Agents (Conversational AI):*** Developed a solution to detect and correct fact-hallucination in knowledge grounded dialogue agents using **BERT** and **GPT-2**.
- ***Virtual Home Interior Planner***: Applied **Structure from Motion** (SfM) on user videos for room **3D Reconstruction** and used U-Net for furniture semantic segmentation, emphasizing **3D Geometry** in home design.
- ***Automated Visual Inspection of Road Damage:*** Designed a solution to spot road damage (such as cracks, potholes) from a UAV Camera video feed via **Object Tracking** and Detection using **MobileNet-SSD** and **Kalman filter**.
- ***Open-Domain Retrieval ChatBot***: Developed an information retrieval chat-bot using **Apache Solr**, indexing data from Wikipedia and Reddit, parsing user queries using **NER** and retaining context using BERT.

Selected Publications & Articles

- LineFormer: Rethinking Chart Data Extraction, *International Conference on Doc Analysis & Recognition (ICDAR) 2023*
- Decoding the performance secret of world's most popular Data Science library - Numpy, *Medium, 2019*
- An Offline Handwriting Recognition Approach for Ruled Pages using CNN and LSTM, accepted to *ICSICCS-2018*

Education

University at Buffalo, SUNY, NY, USA *Jan 2022 - Dec 2023*
Masters in Computer Science (GPA 3.83/4)
Coursework: Machine Learning, Computer Vision, NLP, Deep Learning for Biometrics, Information Retrieval.

University of Mumbai, India *June 2018*
Bachelors in Computer Engineering (GPA 8.17/10)
Coursework: Applied Statistics, Data Structures, Algorithms, Operating Systems, Image Processing, Artificial Intelligence

Accomplishments

- OpenCV AI Competition Finalists 2021.
- Core Member of the Data Science team at Karza Technologies, winning the **NASSCOM AI Challengers 2021**
- Presented ***'Demystifying Deep Learning'*** tech-talk to 100+ attendees, including Barclays' VPs & Directors.