Arun Kumar Tiwary

Machine Learning Engineer
@Navikenz India Pvt. Ltd.
+91-9555013247
tiwaryarun084@gmail.com | LinkedIn | GitHub

Google Certified
Professional ML Engineer
Blockchain ID: 761551
Experience: 6+ Years

Profile

Having 6+years of experience in the datascience and specialize in Machine Learning, Deep Learning and MLOPs domain. Having strong academic background, with a M.Phil degree in Applied Physics from IIT(ISM) Dhanbad. Over the course of my career, have worked on a variety of projects and tasks, ranging from object detection(Computer Vision models) to Large language model(NLP (LLMs)) training and inference on GPU/TPUs. Very passionate,about staying up-to-date with the latest technologies, and always looking for new challenges and opportunities to learn and grow.

Skills

Technologies: Machine Learning, Deep Learning, Cloud(AWS/GCP) and MLOPs
Frameworks: Pytorch, TensorFlow, Darknet
Programming: C, C++, Python
Deep Learning: Experienced with model training/inference, training and inference cost optimizations, have knowledge and interests in, distributed data parallel and model parallel training, trasfer-learning, tuning, post-training model optimizations for performance and quality and benchmarking. Model prunning, quantization, iterative prunning, model architecture modification, feature extraction, model architecture design. training models from scratch. Debugging for model operator supports, hyper-parameter tuning, model architecture analysis and search. Model compression and conversions. Make the models deployable on cloud and embedded devices with resource constrained environments. Build light-weight models and have interests in tarnsformer based generative LLMs and LCVMs. Having end-to-end knowledge of deep-learning process, from data collection and preparation to model building, and deployment. Post training model optimiztions, for inference time reduction and model size compression without significantly compromising model performance.
State of The Art DL Algorithms: YOLO, BERT, Transformers, Stable-Diffusion
Computer Vision Tasks: Classification, Regression, Image Classification, Object Detection, Pose-Keypoints Detection, Image Segmentation and Tracking.
NLP(Natural Language Processing) Tasks: Text Classification, Text Summarization, Text Generation.
State of The Art DL Model Archs: CNN, LSTM, YOLO, GANs, Auto-Encoder Models, Stable-Diffusion, Huggingface Transformers, BERT, GPTs, Bloom
Vision Model Frame-Works: NanoDet, OpenPose, Yolo, Faster-Yolo, etc.
Model Training/Inference Accelerators (Software): Nvidia-Cudnn, Nvidia-TensorRT, TVM, Octo-ML, Hummingbird, ONNX, Qualcomm SNPE, AWS-Neuron-SDK
Hardware Accelerators: Nvidia-Cuda based GPU Machines, Google TPUs, AWS-Tranium(Trn1) AWS-Inferentia(Inf1) Instances.
Benchmarking: Model Performance, Quality (MLCommon Standards)
MLOPs: Docker, Docker-compose, Kubernaties, MLCube, MLFlow, AWS, GCP, AWS-Sagemaker, GCP-Vertex-AI, TorchServe and TensorflowServing
DevOps: Git, Kubernates Pods, Docker/Podman, Docker-compose
Model Deployment: Cloud: AWS/AWS(Sagemaker), GCP/ GCP(Vertex-AI)
Edge devices: Qualcomm Neural Network AI-SDK(SNPE), Open-Vino Google Corel Board.
Python Frame-Works: Numpy, Scipy OpenCV, Matplotlib, Pandas Scikit-Learn.
Build System: CMake,Make, and Bazel
Scripting: Shell(Bash)
Operating Systems: Linux(Ubuntu), Windows, WSL
Troubleshooting: OS/Scripts/Model debugging
Machine-Learning Algorithms: Linear Regression, Logistic Regression, Support Vector Machines, Radom-Forest, PCA, KNN, K-Means, Many Others

Reinforcement Learning:
Quantum Machine-Learning:
Languages: English, Hindi

Education

IIT, Delhi — Delhi, India
Quantum Computing: Certification in Quantum Computing & Machine Learning; — *Jan 2023 – Aug 2023 (Expected)*

IIT(ISM), Dhanbad — Jharkhand, India
M.Phil in Applied Physics; — *– 2012*

University Dept. of Physics, Ranchi University, Ranchi — Jharkhand, India
M.Sc. in Physics; — *– 2009*

Jaiprakash University, Chhapra — Bihar, India
B.Sc. in Physics; — *– 2005*

Work Experience

Machine Learning Engineer @Navikenz India Private Ltd. — Noida, India
Natural Language Processing R&D — *Sep 2022 – Present*

- Worked on cost optimization of model training and inference. Did hardware (AWS-Trainium, AWS-Inferentia and Google TPUs) benchmarking for infererence time reduction, model compression, tranfer-learning, cloud cost optimizations, by benchmarking inference/training metrics on diffrent hardware. Worked on common NLP tasks sentiment analysis, text classification, and generative LLMs. Worked on TensorRT, TVM, Hummigbird for post-training, model optimizations for performance and quality benchmarking and MLOPs. Wrote inference scripts for diffrent accelerators.
- **Impact on Business:** Huge cost reduction due to post-training optimizations and benchmarking of inference and training of large models.

Software Engineer @HCL Technologies Ltd. — Noida, India
Deep Learning (Computervision) R&D — *Dec 2018 – Sep 2022*

- Worked on Deep Learning (Computer Vision), Image classification, Video-Classification, Object Detection, Object Segmentation, Image Denoising and Enhancement, Super-Resolution, Auto-annotation, Model conversions and Post-training model optimizations
- **Impact on Business:** Huge operational cost reduction due to auto-annotation and post-training optimizations and tranfer-learning.

Machine Learning Programmer @Data Analytics Services Pvt. Ltd. — Gurugram, India
Machine Learning, Deep Learning(Computer Vision/NLP) R&D — *Sep 2017 – Dec 2018*

- Worked on Machine Learning, Deep-Learning(Computer Vision, NLP), DevOPs tasks, Cloud Basic Software devleopment envionment setup and troubleshooting.
- **Impact on Business:** Huge operational cost reduction due to all troubleshootings and infrastructure setup.

Freelancer, Freelance Teaching and R&D — Delhi, India
Did freelance teaching and R&D on blockchain, LFS and AI — *Jan 2014 – Sep 2017*

- During this period, explored the frontiers of cutting edge technologies, from building "Linux System from Scratch(LFS)" and AI to building bitcoin based custom altcoin with diffrent genesis block, along with freelaunce teaching assignments.

Senior Research Fellow @IIT Kharagpur — Kharagpur, India
SRF(TA) Teaching Assistanceship with R&D — *Dec 2012 – Jan 2014*

- Studied the stability, bifurcations and dynamics of "Optical Dynamical Systems". Did courses in Linear Algebra, Calculus, Probability & Statistics, Quantum mechanics, Non-Linear Dynamics and Dynamical Systems and many other related courses. Solved systems of ODEs using MATLAB and studied the dynamics.

Projects

@Navikenz India Pvt. Ltd

- Working on post training model optimization frameworks: TVM, Octo-ML, ONNX, NVIDIA-TensorRT and HummingBird
- Worked on Google TPU(V3-8) Pytorch-pt1.13/AWS(Trainium/Inferentia instances) inf1.6x.large. Did training and inference for Huggingface BERT model.
- Did model debugging for operator support issues. Debugged issues of model inference on various devices. Resolved the issues in model conversion to various formats. Did model performance and quality benchmarking for diffrent accelerator hardwares.

- Wrote inference scripts for Octoml/Onnx/Pytorch(CPU/GPU)/Tensorflow(CPU/GPU)/Keras/TensorRT for various trained models.
- Worked/configured/used on Nvidia-NGC containers. Used nvidia-docker for Containerised GPU access.
- Worked on Tesla-K80/A100/V100 AWS-Cloud instances for various activites.
- Worked on AWS/GCP cli Sagemaker/Vertex-AI and on MLOPs Workflows for AWS/GCP
- Working on transformer based models (BERT/Bloom/GPT)
- Doing a course in "Quantum-Computing and Machine Learning" from IIT Delhi.
- Worked on AWS-Neuron-SDK.
- Hired Machine Learning, Data-Science, DevOPs and MLOPs professionals, and mentoring interns.
- Interested in working on LLMs/LCV models (Fastfold/Stable-Diffusion/GPT)
- Interested to work on training and inference of Large Models with Billions of Parameters.
- Interested to work on Distributed data/model parellel training and inference
- Pursuing Huggingface Reinforcement Learning course.

@HCL Technologies Pvt. Ltd

- Worked on integration of Darknet based jersey number detection model with tflite based person and pose key-point detection model c++ inference code. Separated inference code, wrote cmake files for automating the build process on Linux(Ubuntu) as well as windows CPU and GPU(Cuda) machines. Automated and merged jersey detection code with TensorFlow-lite pose detection model inference. Wrote single script to build and infer C++ inference both on CPU and GPU on Ubuntu and windows machine.
- Worked on auto annotation project for HCL in-house auto- annotation tool. My role was to write various scripts for auto- generating XML, TXT, and JSON files in COCO, Pascal-VOC and Darknet data formats, from pre-trained, object detection and pose-detection models. Also convert in-house annotation format to Pascal-VOC, COCO and Darknet data formats by parsing in- house data-format for client. Also from in-house annotation format to customer format.
- Worked on Covid-19 compliance project, developed the Mask, No Mask and Flu and and Non Flu classifiers using Posenet- keypoint features.
- Worked for beauty-product recommendation system based on classification of wrinkled face and non-wrinkled face classifier. By using dlib face-detector.
- Worked on an elderly monitoring system, in which task was build a solution for hospitals and places living old people. Monitor their habits and state of residence. By using Infrared image data of Night and RGB data in daylight illumination. Person Detection model was trained with classes, standing, sitting, lying and falling on IR and RGB data.
- Auto-Detect Books from images and crop for a Library Catalog Project.
- Built video classifier by auto-random selection of 30 frames and classify it for having offensive and non-offensive content.
- Also, mentored HCL Tech-Bees enrolled in Study cum Job programs for Deep Learning. Also helped the team members in various trouble-shooting, as and when required and contributed to the team and company by my best efforts.
- Worked on various object-tracking methods using optical-flow, motion-vectors and deep-sort algorithm. For auto-annotation project.
- **Other Tasks @ HCL**
- GPU environment setup and resolve the issues in installation of CUDA and CuDNN libs.
- Built and installed caffe on CPU and GPU machines.
- Prepared data-sets, prepared scripts for auto-annotation for object detection and pose key-points detection model.
- Built flu detection model using pose keypoints model features. Explored model architecture search for flu detection model using Auto-Keras Auto-ML and Keras tuner.
- Have parsed Yolov3-tiny weights for using it in Keras model.
- Converted models from "Checkpoints to Frozen graphs", "Checkpoints to protocol buffer files", "from Protocol Buffers to ONNX", from "ONNX models to snpe-dlc" and "Pytorch models to SNPE-dlc etc.
- Worked on model-size reduction and performance-boosting.
- Built classifiers at the top of object detection and pose-detection models.
- Built machine-learning models for classification tasks using scikit-learn.
- Explored optimal hyper-parameter space for Machine Learning models using grid-search.
- Explored hyper-parameter space of neural-network models using keras tuner and pytorch tuner for small models.

- Did NN model graph analysis and extraction of features from frozen graph, also did graph visualization and drawing, using graphviz and pydot libs.
- Worked on SNPE-DLC conversion, environment setup and conversion of frozen models to dlc.
- Have worked on video-classification task.
- Have worked on one-shot and iterative prunning of pose- detection model and reduced pose-detection model size.
- Built and trained pose key-points detection model, on our own data-set. Worked on data-set preparation. Improving model execution time performance and accuracy.
- Have experimented with Lottery ticket algoritham on pose- detection model.
- Used model prunning techniques and Lottery Ticket algorithm for model performance improvement.
- Did localization Improvement for object detection model following Reinforcement Learning based, hierarchical search method.
- Worked on different image processing and deep learning techniques of object tracking and detection.
- Worked on optical-flow, template matching and deep-sort based object tracking algorithms.
- Resolved the issues in Google-corel board
- Debugged and trouble-shooted the issues in model conversions and setup. Resolved the issues in SNPE/NCNN/Open-vino model conversion.
- Worked on Appium server setup, Appium-inspector setup, Android-Studio Setup and Appium client setup, extraction of page source and parsed the relevant information for auto-annotation of relevent part screenshots of a android app.
- Expored Huggingface Image captioning using ransformers.
- Did model performance metrics calculation.
- Worked on lib-torch model conversion and building with lib-torch.
- Did memory profiling and analysis of different PyTorch and TensorFlow models.
- Explored image enhancement models, calculated performance metrics, did setup of model training environment, for image de-noising, and super-resolution models.

@Data Analytics Services Pvt. Ltd

- **DevOps and Cloud AWS/GCP:** Worked on deployment of auto-form fill OCR Solution, R-Shiny Application for Banking Sector Client. Resolved the issues in model deployment and dockerization. Have deployed models on AWS and google GCP cloud intances. Have written Dockerfiles and docker-compose yml files. Experience with virtual machines, CI/CD pipelines with Jenkins and Git.
- Experiance with big data eco-system setup, (installation and configuration of Hadoop(HDFS), Pyspark and Apache-Spark.
- Setup and configuration of Gitlab Server, Apache, Nginix webservers, LDAP Server(Phpmyladpadmin), RStudio-Server, Shiny Server, RStudio-Connect Server, NFS server, SSHFS Server, SSH-Server, Redmine Server and different others.
- Worked on Tesseract-Ocr and OCR project. Built models for face detection, face recognition, gender classification, age prediction, logo detection, initially using Haar-Cascade and later using deep learning.

@IIT Kharagpur, WB

- Chaos and Bifurcation Studies in Nonlinear optical dynamical systems. Study of onset and regions of instabilities and bifurcation of Pulsed Laser Systems. Worked on study of stability and bifurcation analysis of Non-Linear Optical Dynamical Systems.

@IIT(ISM), Dhanbad

- M. Phil Numerical Study of Nonlinear Dynamical Systems.

@Ranchi University

- M. Sc: A brief Survey of General Theory of relativity.

Awards & Achievements

- IIT Kharagpur, WB, Senior Research Fellowship
- IIT(ISM), Dhanbad, WB, M.Phil Fellowship

Certifications

- *Google Certified Professional Machine Learning Engineer*Nov 2022

Publications

- *Analysis and chaos synchronization of Genesio–Tesi system applying sliding mode control techniques International Journal of Dynamics and Control,* 2022