

Sunil Acharya

☎ 7406893061 ✉ sun1lach@outlook.com
🌐 sun1lach.me [in sun1lach](https://www.linkedin.com/in/sun1lach) [sun1lach](https://github.com/sun1lach)

Education

Liverpool John Moores University <i>Master of Science in Machine Learning & Artificial Intelligence (76%)</i>	Nov. 2020 - Aug. 2022 Liverpool, UK
<ul style="list-style-type: none">• Research Thesis: Evaluate the effectiveness of various knowledge distillation methods on smaller neural network architectures.	
International Institute of Information Technology <i>7th Summer School on Machine Learning</i>	Aug. 2023 - Sept. 2023 Hyderabad, IN
International Institute of Information Technology <i>Post graduate diploma in Machine Learning & Artificial Intelligence (GPA: 3.60 / 4.00)</i>	Nov. 2020 - Nov. 2021 Bengaluru, IN
Visvesvaraya Technological University <i>Bachelor of Engineering (BE), Electronics & Communication Engineering (74.32%)</i>	Aug. 2012 - May. 2016 Bengaluru, IN

Experience

Intel Corporation <i>AI Software Solutions Engineer</i>	June. 2019 – Present Bengaluru, IN
<ul style="list-style-type: none">• LLM Enablement & Distributed Training: Enabled and scaled LLM training on Intel Gaudi 2 & 3 accelerators using DeepSpeed, Megatron-LM, with various distributed parallel strategies; productionized distributed checkpoint conversions for Hugging Face compatibility to external customers.• Optimized Edge AI: Designed and deployed Intel Automated Vision Checkout at retail sites in India, achieving low-latency (70ms) deep learning inference and scalability on Intel's integrated GPU, supporting 100+ daily transactions at the store front.• Federated AI: Contributed to Linux Foundation's 'securefederatedai/openfl' by introducing JAX/FLAX support, federation long and short lived component timeout feature, and interactive examples for secure, efficient federated learning on private medical records.• LLM Fine-Tuning as a Service: Designed a ZenML/cnvrng.io-based fine-tuning framework for internal use as a premium offering.• Performance Profiling: Experience in accelerators compute/memory profiling, device & host trace analysis to optimize distributed training workloads and application level optimization for edge inference.• Data Pipelines: Built an event-driven, cost-optimized AWS data pipeline, reducing costs by 8x and publishing product data in 5 minutes (down from 6 hours), supporting 40M API hits monthly.• Public Speaking: Presented demos and delivered technical talks on cloud cost optimization and product innovations at Intel India Innovation day, Intel ConnectiON and Intel India tech talk series (Cloud Community of Practice) events.	
Western Digital <i>Software Engineer</i>	Apr. 2017 – May. 2019 Bengaluru, IN
<ul style="list-style-type: none">• Experience building cross-platform CLI and GUI tools for performance profiling of SD/uSD, USB, and enterprise SSD devices using open source tools like FIO, VDBench, etc.• Successfully designed, developed, and deployed an E2E Distributed Client Server Application utilized by several teams to execute 350+ mobile devices in parallel to test, validate, and certify uSD cards.• Experience in configuring and chaining 48/24 ports Brocade Network switches & tune power class ports to support 80+ PoE surveillance cameras for automated uSD/SD card testing.	
sketchmyroom.com (Rhythm of space) <i>Software Engineer Trainee</i>	Aug. 2016 – Dec. 2016 Bengaluru, IN
<ul style="list-style-type: none">• Interiors and architecture design-related experience with Full Stack Web Application Development. Primarily responsible for the development of a backend application that exposes in-house portfolios to clients via APIs, giving them access to a variety of Architecture designs and an online customization option.	
Bharat Sanchar Nigam Limited (BSNL) <i>Engineering Trainee/Intern</i>	Jan. 2015 – Dec. 2015 Bengaluru, IN
<ul style="list-style-type: none">• Operational testing of wireless equipment. Configured routers, modems, and mainline distribution frames. Examined optical fiber functioning at various stages.• A hands-on approach to fusion splicing of optical fiber and monitoring various signal parameters using OTDR devices.	

Projects & Certifications

Certification on "Accelerators for Deep Learning" - IIT Roorkee | [certificate link](#)

- Executive certificate on Accelerators for Deep Learning covering deep learning algorithms and computer architecture with an emphasis on AI acceleration on various computing systems, such as FPGAs, mobile/desktop GPUs, smartphones, ASICs, DSPs and CPUs.

Custom & Efficient CNN architectures from Scratch | [project link](#)

- Designed and Implemented 25k, 143k, 340k, 600k & 1M parameter efficient custom CNN architectures.
- Upon knowledge distillation on these custom CNN architectures, model accuracy surpasses ResNet-18/34/152 baselines with 10-20x less model parameters on FMNIST & CIFAR-10.

Intel AI Everywhere Conference | *Hackathon*

- Feature-engineered using VIF, RFE, and PCA. Built a hyperparameter-tuned logistic regression, random forest, and XGBoost model to predict the outcome (pass or fail) of new turn-ins using historical records of turn-ins.
- Selective filtering was done based on the classification outcome of the turn-ins to save execution time in the DevOps pipeline in the context of HW design validation.

GradCam Visualization of CIFAR-10 dataset with Albumentations. | [project link](#)

- Focuses on building and training a ResNet-18 model on the CIFAR-10 dataset.
- Implemented data augmentation using the Albumentations library, a custom dataset loader, plotting train and test loss curves, GradCam visualization of randomly sampled misclassified images, and visualization of misclassified images with labels and appropriate legends.

Style Transfer using Generative Adversarial Network (GAN) | *Project*

- Built a Generative adversarial model(modified U-Net Architecture) which can generate artificial T1 to T2 and vice-versa MRI images of different contrast levels from existing MRI scans.

Technical Skills

Languages: Python, C++, Java

Libraries/Frameworks: PyTorch, TF/Keras, JAX/FLAX, LitGPT, OpenFL, Transformers, Deepspeed, Megatron-LM, ZenML (MLOps), Llama-Factory, Vue.js, FastAPI, perfetto, fio

Technologies, Platform & Practices: Clean Architecture (Domain Driven Design), HuggingFace, Git/Gerrit (code reviews), AWS, microservices, containers, Jenkins, VS Code.