Investigation and Deployment of optimized Machine learning models on embedded systems

Abstract

Machine Learning (ML) models have become a powerful tool for solving complex problems in areas like image recognition, anomaly detection, speech processing, and system control. Once trained, these models can make accurate predictions and decisions based on new data, which has led to their wide adoption in industries ranging from healthcare to manufacturing.
However, most ML models are typically developed and tested on high-performance computers with powerful processors and large memory resources. Deploying these models in real-world scenarios — especially in devices like sensors, monitoring units, and small-scale controllers — often requires running them on low-cost embedded platforms with very limited computing power and storage.
This project focuses on bridging that gap by studying how trained Neural Network (NN) models can be optimized and deployed on affordable embedded hardware such as microcontrollers and single-board computers. Techniques like model compression, quantization, and hardware-specific optimization are explored to ensure the models can perform fast and accurate predictions in real-time, even with limited resources. The project aims to demonstrate how intelligent decision-making can be brought directly to field devices, enabling smarter, more autonomous systems in areas like environmental monitoring, predictive maintenance, and IoT applications. The student has to train time series models and then deploy them on low platforms such as ESP-32/ R-Pi after performing optimizations in terms of memory and response.

References:

[1] Han, S., Mao, H., & Dally, W. J. (2015), “Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding.”, https://arxiv.org/abs/1510.00149
[2] https://www.tensorflow.org/lite/microcontrollers
[3] Sze, V., Chen, Y. H., Yang, T. J., & Emer, J. S. (2017).“Efficient Processing of Deep Neural Networks: A Tutorial and Survey”, Proceedings of the IEEE, 105(12), 2295-2329., https://doi.

Academic Project Requirements:

1) Required No. of student(s) for academic project: 1

2) Name of course with branch/discipline: M.E./M.Tech Electronics and Instrumentation Engineering

3) Academic Project duration:

(a) Total academic project duration: 48 Weeks
(b) Student's presence at IPR for academic project work: 4 Full working Days per week

Email to: himanshu.tyagi@iterindia.in[Guide's e-mail address] and
project_ece@ipr.res.in [Academic Project Coordinator's e-mail address]

Phone Number: 079 -07923962224 [Guide's phone number]