MS THESIS IN EMBEDDED AI / TINYML

At R&D Global Technology, Digital and Sustainability (TDS) we are looking for a Ms Thesis Intern who will work in a team responsible for the development of smart functionalities for home-appliances. The increasing demand for Artificial Intelligence (AI) applications has led to a surge in the development of AI models that can run on ultra-low power, often a milliwatt or less, microcontrollers with constrained storage. These devices are used in various applications such as IoT devices, wearables, and autonomous systems where energy efficiency is crucial. The white goods industry is keeping up, and has exciting data-driven use-cases to integrate into its' embedded systems.
The position involves working in an international, multi-competence and cross product line environment.

MS THESIS PROJECT

Project title: "Optimized Code Generation for Ultra-Low Power Microcontrollers: Integration into a consolidated platform for Efficient AI model deployment"

Problem definition: The deployment of AI models as part of a larger system on the edge platforms poses significant challenges due to their limited computational resources and memory constraints. By edge platform we refer to low-power CPUs, tiny memory, and low-power display technologies.To address this challenge, researchers have developed optimized libraries such as CMSIS-NN that provide a reference implementation for neural network inference on Cortex-M microcontrollers. This library is built upon and improved by other frameworks/libraries, but there is evidence to room for modifications to achieve optimal performance, as it is a highly researched area [1,2].
Librararies as such are a low-level software/library designed to efficiently perform the computation required for machine learning inferences.

Expected output:

Literature review: Review papers and open-source frameworks (and their white papers) related to optimized inference scheduling / inference kernel libraries, that encapsulate innovative strategies for leveraging hardware capabilities. Leveraging hardware capabilities in non-ML HW can be through strategies such as reordering operator execution, improvements in memory scheduling and where possible trading computation for memory efficiency.

Benchmarking: Benchmark short-listed third-party strategies on custom and public deep neural networks used in signal processing and computer vision. The metrics for comparison are CPU and peak RAM usage. Suggested platforms to test on are S32K1 from NXP, STM32F Discovery.

Strategy development: Identify the best strategy for optimizing AI model deployment based on the benchmarking results.

Code-development: Develop/enrich existing third-party solutions that are suitable to be incorporated into the current in-house Model-Based Design toolbox, a consolidated internal platform for Code Generation in real-time control context, developed with Mathworks product.

Verification: Verify the performance of the optimized platform on various AI models and benchmarking datasets.

The project will require regular meetings with the team to discuss progress, share findings, and receive feedback. The student will also be responsible for presenting the final report and demonstrating the consolidated platform to the team.

YOU

- Team player;
- Good listener and communicator;
- Proactive. You are self-driven, results-oriented with a positive outlook. You are assessing and investigating feasibility of new technologies and ideas;
- Adaptive. You like challenges and you are flexible to adapt to new situations and contexts;
- Creative. You love to explore new ways of thinking and you are oriented towards innovation;

EDUCATION & EXPERIENCE

- Experience in Embedded Systems and one or more of the following areas:Neural Networks, Deep Learning, Signal Processing, Computer Vision
- Good knowledge in MATLAB/Simulink and C/C++or programming language like Python, Java
- Knowledge of machine learning and image processing frameworks: TensorFlow, PyTorch, Keras
- Good in English, written and spoken

REFERENCES

1. TinyML for Ultra-Low Power AI and Large Scale IoT Deployments: A Systematic Review, N. Schizas, A. Karras et al., 2022
2. Tiny Machine Learning: Progress and Futures, J. Lin, L. Zhu, 2024