PhD Forum. DCIS 2022. November 16 – 18. Pamplona, Spain

Deep Learning Algorithms on Egde Devices

Daniel Enériz Group of Electronic Design (I3A-UZ)
Dr. Nicolás Medrano and Dra. Belén Calvo

Instituto Universitario de Investigación de Ingeniería de Aragón
Universidad Zaragoza

BACKGROUND & MOTIVATION OF THE THESIS

Cloud computing is the default choice for neural networks, but

- Requires Internet connection
- Reduces bandwidth
- Introduces an extra latency
- More energy consumption
- Privacy at risk

Edge computing is more convenient in real-time and portable applications

Food quality detection [1], EEG activity detection for BCIs [2], computer-aided heart murmur diagnostic with PCG analysis [3], …

removing the Internet-related issues, but

Requires optimized implementations since edge devices have less computer power than datacenter servers

INNOVATION

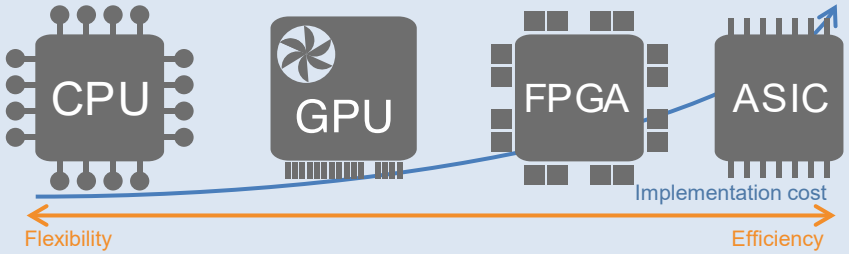
Low-spec and **low-cost** devices are targeted, requiring further **optimization**

Usage of **low-resolution datatypes** to reduce the computation impact with **performance losses** due to model quantization

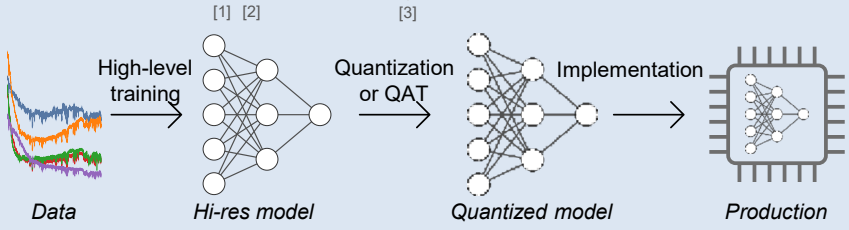
Quantization-aware training (QAT) enabling **null accuracy loss** at the cost of more expensive training processes

OBJECTIVES AND WORK PLAN

The main objective of this PhD thesis is to implement DL models in low-spec CPUs, GPUs and FPGAs, comparing the workflows required and the performance of the models in each platform.



Time	2021	2022	2023	2024
HW	FPGAs		CPUs	GPUs
SW	HLS		TFLite/C	CUDA
Data	Public			Experimental



Current Technology Readiness Level (TRL) 3

- *Empirical proof of concept of the implementation process in all HW platforms*

METHODOLOGY

Implementation process

- **Data** formatting and preprocessing
- **High-level training** of the model using Keras/PyTorch
- Model translation to the **low-level description** (C/C++) with low-resolution datatypes
- **Optimization** of the model considering the targeted device
- Model **deployment** in HW
- Performance **evaluation**

RESULTS AND IMPACT

- ***Expected results include successful implementations in each HW platform***
- ***Different applications in biochemical sensor fusion will be tested***

[1] D. Enériz, N. Medrano, and B. Calvo, 'An FPGA-Based Machine Learning Tool for In-Situ Food Quality Tracking Using Sensor Fusion', *Biosensors*, vol. 11, no. 10, p. 366, Oct. 2021

[2] A. C. Hernandez-Ruiz, D. Enériz, N. Medrano, and B. Calvo, 'Motor-Imagery EEGNet-Based Processing on a Low-Spec SoC Hardware', in *2021 IEEE Sensors Conference*, Oct. 2021

[3] D. Enériz *et al.*, 'Implementation of a Heart Sound Segmentation Deep Model on a Low-Cost FPGA', submitted to the 25th Euromicro Conf. on Digital System Design (DSD2022)

This work is funded by the Gobierno de Aragón through the PhD grant BOA20201210014 and the Ministero de Ciencia e Innovación, through the PID2019-106570RB-I00 project.

More info & contact

Daniel Enériz eneriz@unizar.es
Facultad de Ciencias, Universidad de Zaragoza, Pedro Cerbuna,12, Zaragoza, 50009