Internship on TinyML (Machine Learning on IoT devices)

Master 2 – 5 months – 02-04/2023

Guillaume MULLER & Mihaela JUGANARU

November 18, 2022

1 Administrative Context

Mines Saint-Etienne (MSE), one of the graduate schools of Institut Mines-Télécom, the first group of graduate schools of engineering and management in France under the supervision of the Ministry of the Economy, Industry and Digital Technology, is assigned missions of education, research and innovation, transfer to industry and scientific, technological and industrial culture.

MSE consists of 2,400 graduate and postgraduate students, 400 staff, a consolidated budget of €46M, three sites on the Saint-Etienne campus (Auvergne Rhone-Alpes region, Lyon Saint-Etienne metropolitan area), a campus in Gardanne (SUD region, Aix Marseille metropolitan area), a site in Lyon within the digital campus of Auvergne Rhone-Alpes Region, six research units, five teaching and research centres and one of the leading French science community centres (La Rotonde €1M budget and +40,000 visitors per year). The Times Higher Education World University Ranking ranked us for 2022 in the 251-300 range for Engineering and Technology. Our work environment is characterised by high Faculty-to-Student, Staff-to-Faculty and PhD-to-Faculty ratios, as well as comprehensive state-of-the-art experimental and computational facilities for research, teaching and transfer to industry.

The Henri Fayol Institute, one of the school's 5 training and research centers, brings together professors in industrial engineering, applied mathematics, computer science, environment and management around the theme of overall business performance. The Henri Fayol Institute is strongly involved in flagship projects of the Industry of the Future and the City of the Future.

2 Scientific Context

In recent years, Artificial Intelligence, in particular **Neural Networks (NN)**, has shown **impressive results** in many applications, often beating humans in many domains, from Games (AlphaGo...) to Health Care (skin & eye cancer detection...). However, training such models requires **large amounts of computing power**, thus of energy; sometimes more than a small city over a year (e.g. GPT-3). As energy is the main source of release of CO_2 in the atmosphere, such technological progress unfortunately goes along with the destruction of our planet. This goes in the opposite direction of UN's Sustainable Development Goals, that we need to achieve quickly to ensure our survival as a whole society.

3 Topic: TinyML

The field of **TinyML** seeks to find ways of implementing Machine Learning (ML) models (particularly NN) on small devices, with **limited CPU power, RAM capacity, Network bandwidth and Battery life**. Techniques developed in this domain could provide elements for a global solution, thus allowing to continue producing **positive social impacts** with AI/ML/NN (better health care, optimized transportation...), **without destroying our planet**.

This internship proposes to explore state of the art techniques for reducing both the size and the training time of a NN, using small devices to impose strict energy consumption constraints.

Keywords: *Artificial Intelligence, Neural Network, Deep Learning, IoT, TinyML, Quantization, Pruning, Distillation, Training, Gradient Descent, Back-Propagation.*

4 Organization

The internship will take place at Espace Fauriel in Saint-Etienne, in the ISI department of Institut Fayol.

The internship will follow a 3 steps plan:

1. The student will start with trying to reproduce the toy (but realistic) application which consists in designing a glove/bracelet that can recognize the characters drawn in the air by a person [Fre21]. Through this example the student will learn about techniques like **Quantization**, **Pruning** and **Distillation**. These techniques allow reducing the size of a Big NN that was previously learned on a standard computer. This solves the problem of the energy consumption at *inference time, but not at training time.*

2. Then, the student will explore state of the art techniques for **training a NN directly on a small device**, based on researches like [Lin+22].

3. Based on these experiments, the student will be able to explore more realistic scenarios adapted to **Industry 4.0** (e.g. the "Augmented Technician") or **Health Care** (e.g. "Smart Orthosis"), where we need *both inference and training* to be executed *on-device*, in order to detect custom gestures that can change over time.

5 Job requirements

The student should have prior following skills:

- Solid background in Machine Learning, in particular Deep Learning
- Strong coding in Python skills
- Minimal background in IoT/Arduino
- Curiosity of anything technological/scientific & Motivation for Sustainable Development
- Master 2 or last year engineering school

6 Application

To apply, please send your CV, cover letter, and any other useful information before **January, 15 2023** to guillaume.muller@emse.fr

7 References

References

[Fre21] Zack Freedman. *AI Data Glove: Somatic*. 2021. URL: `https://www.youtube.com/watch?v=6raRftH9yxM`.

[Lin+22] Ji Lin et al. "On-Device Training Under 256KB Memory". In: *arXiv preprint arXiv:2206.15472* (2022). URL: `https://tinyml.mit.edu/`.