Homework #2

Updated Literature Description

Author:	Dereck Alpizar
Matriculation No.:	7023782
Course of Studies:	Machine Learning
First examiner:	Prof. Dr. Elmar Wings
Submission date:	November 1, 2022

University of Applied Sciences Emden/Leer · Faculty of Technology · Mechanical Engineering Department
Constantiaplatz 4 · 26723 Emden · http://www.hs-emden-leer.de

1 Elements of TinyML on Constrained Resource Hardware

Keywords: IoT, TinyML, Neural Networks, Deep Learning Algotithms, Arduino Nano 33 BLE Sense, Quantization Training.

The article presents three TinyML applications by using the Arduino Nano 33 BLE Sense that can provide trustworthy data from IoT to the cloud for Neural Network data processing algorithms. Implementations given are all related to temperature sensing and alerting under given conditions: Temperature Classification, Anomaly Detection and Weather Forecasting. [TGK22]

The article gives usage examples with result by using TinymL (Arduino Nano 3 BLE Sense) with little results, and few details on the SW implementation. It clearly specifies the algorithms implemented. Therefore, it is good to get an idea of possible projects that can be developed under this technology. However, if you are looking for solutions to a specific HW or SW issue, this may lack information.

2 A Practical View on Training Neural Networks in the Edge

Keywords: Embedded Systems, Resource-Constrained Environments, Embedded AI, Edge AI, Neural Networks, Efficient Training.

In this paper, authors summarize techniques that make training on embedded devices feasible. They present 4 alternatives for this matter, which include:

- Reduce the size of the neural networks itself.
- Accelerate the training process of the neural networks.
- Reduce the memory consumption in neural network training.
- Don't need all data at the same time to train neural networks.

Then the authors also expose existing techniques that address training in resource-constrained environments as well as techniques that are also suitable for training on embedded devices, such as incremental learning. [RS22]

The paper is interesting in the sense that it offers a wide motivation on the challenges that training machine learning models on Edge Devices implicitly has. It approaches 4 different solutions to help the reader get an idea on how to approach a more elaborated solution. One thing I consider relevant to mention for this paper is that proposed solutions are mostly targeted for MCUs, FPGAs, single-board computers and smartphones. It leaves behind ASICs; therefor, more complex and specific application might feel some lacking information. Also, the information provided doesn't get too practical, so don't expect to find any further solution, than flow diagram charts.

Bibliography

[RS22] M. Rüb and A. Sikora, "A practical view on training neural networks in the edge," *IFAC-PapersOnLine*, vol. 55, no. 4, pp. 272–279, Jul. 2022, ISSN: 2405-8963. DOI: 10.1016/j.ifacol.2022.06.045. [Online]. Available: https://doi.org/10.1016/j.ifacol.2022.06.045.

[TGK22] V. Tsoukas, A. Gkogkidis, and A. Kakarountas, "Elements of tinyml on constrained resource hardware," in *Advances in Computing and Data Sciences*, M. Singh, V. Tyagi, P. K. Gupta, J. Flusser, and T. Ören, Eds., Cham, Switzerland: Springer International Publishing, 2022, pp. 316–331, ISBN: 978-3-031-12641-3. DOI: 10.1007/978-3-031-12641-3_26. [Online]. Available: https://doi.org/10.1007/978-3-031-12641-3_26.