Arrhythmia Classifier Using Convolutional Neural Network with Adaptive Loss-aware Multi-bit Networks Quantization

1st Hanshi Sun
School of Electronic Science and Engineering
Southeast University
Nanjing, Jiangsu, China
preminstrel@gmail.com

2nd Ao Wang
School of Electronic Science and Engineering
Southeast University
Nanjing, Jiangsu, China
Ao0323@outlook.com

3rd Ninghao Pu
School of Electronic Science and Engineering
Southeast University
Nanjing, Jiangsu, China
1742237531@qq.com

4th Zhiqing Li
School of Electronic Science and Engineering
Southeast University
Nanjing, Jiangsu, China
lizhiqing@seu.edu.cn

5th Junguang Huang
School of Electronic Science and Engineering
Southeast University
Nanjing, Jiangsu, China
huangjunguang@seu.edu.cn

6th Hao Liu†
School of Electronic Science and Engineering
Southeast University
Nanjing, Jiangsu, China
nicky_lh@seu.edu.cn

***Abstract*—Cardiovascular disease (CVDs) is one of the universal deadly diseases, and the detection of it in the early stage is a challenging task to tackle. Recently, deep learning and convolutional neural networks have been employed widely for the classification of objects. Moreover, it is promising that lots of networks can be deployed on wearable devices. An increasing number of methods can be used to realize ECG signal classification for the sake of arrhythmia detection. However, the existing neural networks proposed for arrhythmia detection are not hardware-friendly enough due to a remarkable quantity of parameters resulting in memory and power consumption. In this paper, we present a 1-D adaptive loss-aware quantization, achieving a high compression rate that reduces memory consumption by 23.36 times. In order to adapt to our compression method, we need a smaller and simpler network. We propose a 17 layer end-to-end neural network classifier to classify 17 different rhythm classes trained on the MIT-BIH dataset, realizing a classification accuracy of 93.5%, which is higher than most existing methods. Due to the adaptive bitwidth method making important layers get more attention and offered a chance to prune useless parameters, the proposed quantization method avoids accuracy degradation. It even improves the accuracy rate, which is 95.84%, 2.34% higher than before. Our study achieves a 1-D convolutional neural network with high performance and low resources consumption, which is hardware-friendly and illustrates the possibility of deployment on wearable devices to realize a real-time arrhythmia diagnosis.**

***Index Terms*—arrhythmia, neural network, quantization, multi-bit networks, multi-class classification, machine learning, resource-constrained devices**

I. Introduction

In many healthcare scenarios, patients are diagnosed with a remarkable variety of diseases, including Cardiovascular disease (CVDs), a universal deadly disease [22]. The electrocardiogram (ECG) depicts the human heart's electrical activity and is significant for accurate diagnoses. However, in the early stage, with unobvious symptoms and short duration, some arrhythmias may be challenging to recognize [11], resulting in severe consequences. Therefore, real-time heart rate detection deployed on low-power devices have come under the spotlight.

Neural networks establish a mapping from low-level signals to high-level semantics by simulating a hierarchical structure similar to the human brain to achieve the hierarchical feature expression of data, which has powerful information processing capabilities, promoting the development of algorithms and models for ECG classification methods [13]. Although the detection and classification accuracy of the neural network model seems considerable [25], its huge trainable network parameters consume a large amount of memory and require more time for complex computation, which makes it difficult to deploy on low-power hardware platforms.

To tackle this issue, we consider both the design of network structure and the adaptation of quantitative compression method, which can reduce the accuracy degradation from typical quantization methods, even improve the accuracy in that model error is optimized by cited adaptive bitwidth quantization method. The contribution of this paper has three aspects:

- An adaption of cited adaptive loss-aware quantization(ALQ) is proposed to lower the memory and power consumption of a 1-D convolutional neural network while maintaining or even improving the classification accuracy.
- Based on our novel compression method, a 17 layer convolutional neural network (CNN) architecture for cardiac arrhythmia (17 classes) detection based on long-duration ECG fragment analysis is proposed, and it realizes an