Benchmark Report: DEEPCRAFT™ Ready Model for Baby Cry Detection

Introduction

The DEEPCRAFT™ Ready Model for Baby Cry Detection is a commercial solution for consumer electronics manufacturers. We compare its performance against cutting-edge models and solutions to ensure real-world effectiveness. We evaluate our model against the most recent research paper with the highest accuracy and an industrial solution integrated into common mobile phones.

Evaluation

We evaluate our model against the following models:

1. CNN 9.6M[1]: a recently published research paper that adopts a 2D-CNN with specialized nonsymmetric kernels
2. a software feature integrated into common mobiles phones

The comparisons were performed in the following ways:

- Model hardware performance
    - We compared the memory footprint and inference time of the Ready Model against the CNN 9.6M model
- Model accuracy performance
    - We compared the recall and the outlier accuracy against the model found in common mobile phones

We deployed our model, generated using DEEPCRAFT™ Studio code, onto Infineon's PSOC™ 6 with an IoT sense expansion kit (CY8CKIT-028-SENSE).

The CNN 9.6M model does not provide a pre-trained model, so we cannot evaluate the model accuracy. The model found in common mobile phones is not publicly available, so it's not included in the hardware evaluation.

Model Hardware Performance

In Figure 1, we present the performance of our model in terms of inference time and memory footprint. The DEEPCRAFT™ Ready Model requires 500 times less

[1] https://link.springer.com/chapter/10.1007/978-3-030-31764-5_7

memory than the CNN 9.6M. Moreover, the DEEPCRAFT™ model is 40 times faster. Note that the actual inference time of our model may differ when deployed in the microcontroller. This is a relative comparison in a CPU, given that the CNN 9.6M cannot be deployed in a PSOC™ 6 microcontroller.

.

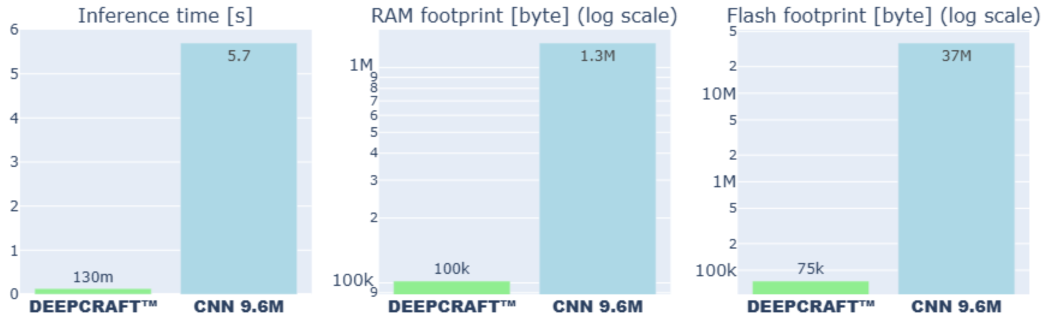

Figure 1 Evaluation of our model against state-of-the-art research model regarding inference time and memory footprint.

Model Accuracy Performance

Figure 2 compares our model's performance with a model found in common mobile phones. Our model outperforms the other in recall, but the other model is better at identifying outlier sounds. However, these rare sounds don't significantly impact our model's performance.

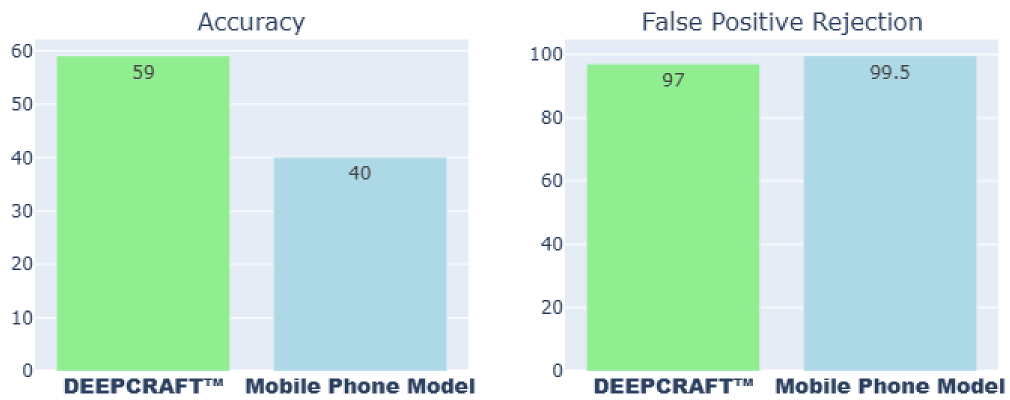

Figure 2 Evaluation of our model against the industrial model from common mobile phones.

Accuracy is the number of correct detections (true positives) divided by the total number of events occurring (true positives). The higher the number, the better. In this case, the event is a baby cry sound.

$$Accuracy = \frac{True\ Positives}{Total\ Positives}$$

False Positive Rejection is the number of negative sounds correctly rejected or not triggered divided by the total number of negative sounds. The higher the number, the better. In this case, the negative sounds are any sound that is a baby crying, i.e., talking, laughing, etc.

$$False\ Positive\ Rejection = \frac{True\ Negatives}{Total\ Negatives}$$