imagimob
An Infineon Technologies Company

Benchmark Report: DEEPCRAFT™ Ready Model for Siren Detection

Introduction

The DEEPCRAFT™ Ready Model for Siren 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. 1D-CNN[1]: a recently published research paper that adopts a 1D-CNN on Fourier transformations of the audio signals
2. a software feature integrated into common mobile phones

The comparisons performed on the models were in terms of the following:

- Model hardware performance
    - We compared the memory footprint and inference time of the Ready Model against the 1D-CNN 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 generation, onto Infineon's PSOC™ 6 with an IoT sense expansion kit (CY8CKIT-028-SENSE).

The model from the research paper does not provide a pre-trained model, so we cannot evaluate the model accuracy. The model found in common mobiles phones is not publicly available, so it's not included in the hardware evaluation.

[1] https://www.sciencedirect.com/science/article/pii/S0263224123013489

Model Hardware Performance

In Figure 1, we present the performance of our model in terms of inference time and memory footprint. The research model requires significant memory, both RAM and flash, to be deployed in production. Therefore, that model is unsuitable for hardware with memory limitations, such as a PSOC™ 6 microcontroller. The DEEPCRAFT™ model requires 300 times less memory than the research model. Moreover, the DEEPCRAFT™ model is >12000 times faster than the research model in terms of inference time. Note that the actual inference time is hardware dependent. This is a relative comparison in a CPU, given that the research model cannot be deployed in a PSOC™ 6 microcontroller.

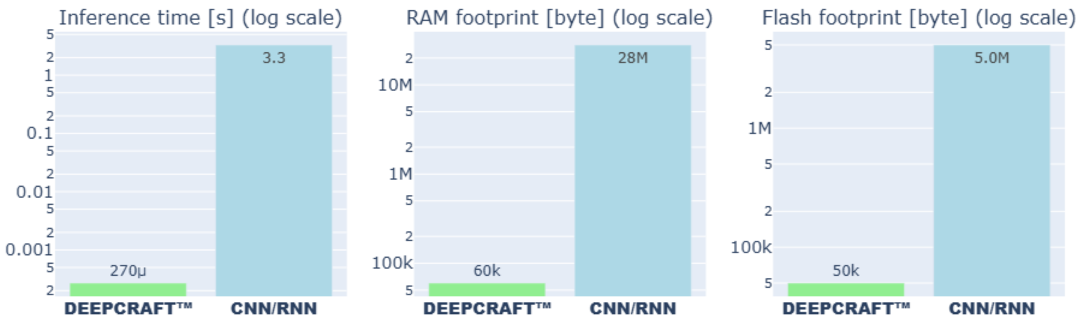

Figure 1 Evaluation of our model against state-of-the-art research model regarding inference time and memory footprint.

Model Accuracy Performance

Figure 2 compares our model's performance with a model found in common mobile phones. The other model shows better recall at 70 dB siren volume, while our model outperforms at 80 dB and above, which is typical for emergency vehicle sirens. The other model is better at identifying outlier sounds, but these sounds are rare and don't significantly impact the DEEPCRAFT™ model's overall performance.

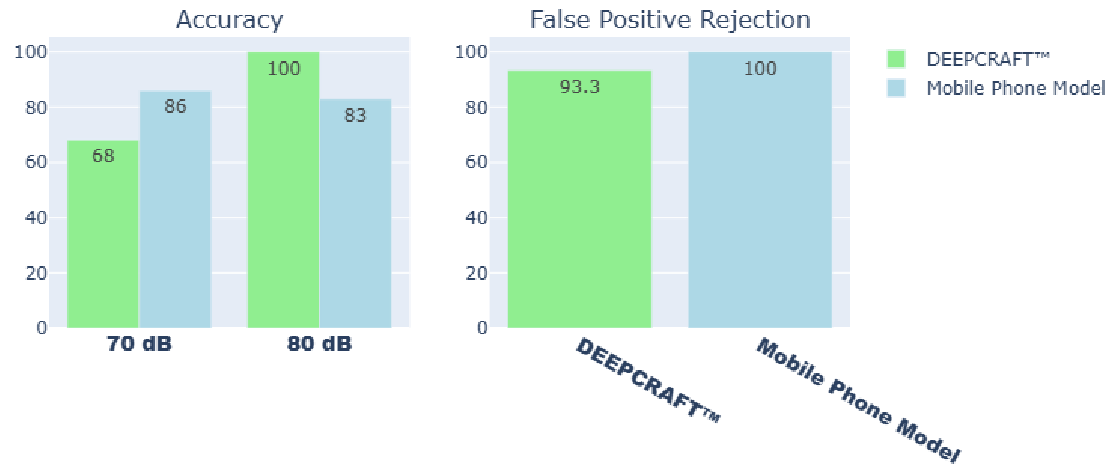

Figure 2 Evaluation of our model against the industrial model found in common mobile phones.

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 siren 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, not a siren, i.e., talking, laughing, etc.

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

Learn more at imagimob.com