

Abstract

Real-time Classification of Bird Vocalizations with CNN on a System-on-Chip Applying Hardware-Software-Co-Design Techniques and High-Level Synthesis

Knut Pröpper, Anastasiia Purtova, Lutz Leutelt
Hamburg University of Applied Sciences

This research aimed to evaluate the use of high-level synthesis (HLS) within the hardware-software codesign paradigm with application to efficient implementation of convolutional neural networks (CNN) for birds vocalization classification.

The classification model was trained using supervised learning with selected and preprocessed bird vocalizations from the Xeno-Canto database of wildlife. The preprocessing and feature extraction of the audio data involves noise reduction, silence removal and calculation of psycho-acoustically weighted MEL spectrograms which were used for CNN model training. With the applied Sequential CNN classifiers an accuracy of approx. 94% was achieved.

For implementing on an AMD Zynq UltraScale+ MPSoC, the Python solution of the classifier was analyzed module by module with a focus on the design principles of hardware software co-design and HLS. Depending on the requirements of the module, preprocessing and feature extraction modules are implemented either as software, hardware description language (HDL) or HLS-based hardware acceleration, while CNN inference is performed on the processing system using the PYNQ framework. The different variants are compared in terms of latency, resource usage and development effort.

The evaluation showed that HLS is highly effective for abstract, dataflow-oriented signal processing tasks, such as spectrogram generation. It can significantly reduce development time without compromising performance. However, traditional HDLs remain superior for low-level interfaces such as I2S audio input due to their precise control over timing and protocol details. The CNN implementation using PYNQ has shown that high-level frameworks can effectively integrate with custom hardware modules to create a cost-efficient, real-time audio classification system on an SoC.