VUB VRIJE UNIVERSITEIT BRUSSEL

The faculty of Engineering of the Vrije Universiteit Brussel invites you to attend the public defense leading to the degree of

DOCTOR OF ENGINEERING TECHNOLOGY

of   **Jurgen Vandendriessche**

The public defense will take place on **Friday 29th March 2024 at 5:00 pm** in room **I.0.01** (Building I, VUB Main Campus)

To join the digital defense, please click here
Meeting ID: 372 589 449 35
Wachtwoord: 4F4UAw

More information can be found on
https://jurgenvandendriessche.github.io/PhdJurgen/

TOWARDS SMART ACOUSTIC CAMERAS FOR SIMULTANEOUS SOUND LOCALIZATION AND RECOGNITION

BOARD OF EXAMINERS

Prof. dr. ir. Omar Hegazy

Prof. dr. ir. Dimitrios Angelis

Prof. dr. An Braeken

Prof. dr. ir. Hichem Sahli

Prof. dr. ir. Kris Steenhaut

Prof. dr. ir. Alberto Izquierdo

Prof. dr. ir. Michael Hubner

PROMOTORS

Prof. dr. ir. Abdellah Touhafi

Prof. dr. ir. Bruno Tiago da Silva Gomes

Abstract of the PhD research

Acoustic cameras are devices that visualize sound by utilizing an array of microphones. The signal from each microphone is combined using a beamforming algorithm to generate an acoustic heatmap or acoustic image. These beamforming algorithms tend to have a high computational cost, which increases with the number of microphones. The combination of a high number of Input/Output (I/O) requirements for the microphones combined with the high amount of parallel computations makes Field Programmable Gate Arrays (FPGAs) very suitable for processing the signals from these microphone arrays. FPGAs have a low power consumption, which makes them especially viable when targeting battery powered devices such as handheld acoustic cameras or nodes in a sensor network. Despite the high computational power per watt of FPGAs, satisfying real time scenarios still present a challenge, especially when targeting acoustic images with a higher resolution. To overcome this challenge, a multi-mode acoustic camera has been developed. The camera supports multiple modes depending on the task at hand. To satisfy the real time requirement for each mode, the resolution of the acoustic heatmap can be adapted. A second limitation of the existing acoustic cameras is the identification of the type of sound, which commonly requires human expertise to recognize and profile the sound.

In recent years, deep learning, a form of Artificial intelligence (AI), has shown promising results towards the task of sound recognition by using Convolutional Neural Networks (CNNs). However, most of the research focuses on improving the accuracy of such models without considering the limitations one encounters when deploying such a model on resource constrained devices such as FPGAs. FPGAs are used nowadays for embedding deep learning inference, mainly using two architectures. One type of architecture uses a general-purpose soft-core inside the Programmable Logic (PL) of the FPGA. On the other hand, there are also dataflow-based architectures that translate each layer in a CNN to a functional block in the PL. Embedding these CNNs for inference on FPGAs is not a trivial task and come with trade-offs in terms of resource consumption, accuracy, supported layers,... These two architectures are compared against other embedded solutions such as Google's edge Tensor Processing Unit (TPU) and a Raspberry Pi (RPi) to find the best fit for acoustic cameras. Acoustic cameras are targeted in this instance because they can identify the location of a sound source, which is not possible when using one microphone. Furthermore, existing beamforming techniques such as delay-and-sum reconstruct audio signals, which can be used for audio classification tasks.