Hardware Realization of Hybrid AI Models Targeting SoC and FPGA Devices

Summary

The research focuses on the hardware realization of hybrid AI models, specifically targeting System on Chip (SoC) and Field-Programmable Gate Array (FPGA) devices. The primary objective is to design a scalable and energy-efficient AI accelerator that supports hybrid deep learning models, including CNN-Transformers and complex models operating in the frequency domain. This will be collaborative work with other students who develop a framework that facilitates the design of lightweight (hardware-aware), secure, and reliable hybrid AI models. These models will be deployed into the designed accelerator and validated through FPGA demonstrators, ensuring their energy efficiency and performance on diverse IoT and edge hardware platforms.

Research field:	Information and communication technology
Supervisors:	Jaan Raik Masoud Daneshtalab
Availability:	This position is available.
Offered by:	School of Information Technologies Department of Computer Systems
Application deadline:	Applications are accepted between October 01, 2024 00:00 and October 25, 2024 23:59 (Europe/Zurich)

Description

The research initiative is centered on the hardware implementation of hybrid Artificial Intelligence (AI) models, with a specific emphasis on System on Chip (SoC) and Field-Programmable Gate Array (FPGA) devices. The primary goal of this project is to design a scalable and energy-efficient AI accelerator capable of supporting sophisticated hybrid deep learning models. These models include Convolutional Neural Networks (CNNs) integrated with Transformer architectures, as well as more complex models that operate within the frequency domain.

To achieve this, our approach will involve a collaborative effort with other students who are dedicated to developing a comprehensive framework. This framework will facilitate the design of lightweight, hardware-aware hybrid AI models that are not only efficient but also secure and reliable. The focus will be on ensuring that these models can be seamlessly integrated into the AI accelerator.

The proposed AI accelerator will undergo rigorous testing and validation using FPGA demonstrators. This process will be critical in evaluating the energy efficiency and overall performance of the hybrid AI models when deployed on various Internet of Things (IoT) and edge hardware platforms. By leveraging the reconfigurable nature of FPGAs, we aim to demonstrate the practical applicability and advantages of our designs in real-world scenarios.

Furthermore, the project will explore innovative techniques for optimizing the interplay between hardware and AI models. This includes fine-tuning the computational load and minimizing power consumption without compromising the performance and accuracy of the AI models. Through this collaborative and iterative approach, we aim to push the boundaries of what is possible with hybrid AI models on modern SoC and FPGA platforms, paving the way for advanced applications in edge computing and IoT environments.networks emerging around experiments?

Responsibilities and (foreseen) tasks:

- Design the AI Accelerator: Develop a scalable and energy-efficient AI accelerator that supports hybrid deep learning models, including CNN-Transformers and models operating in the frequency domain.
- Collaborate on Framework Development: Work with other students to create a framework that facilitates the design of lightweight (hardware-aware), secure, and reliable hybrid AI models.
- Deploy Models on the Accelerator: Implement the hybrid AI models into the designed AI accelerator.
- Validate Performance and Efficiency: Test and validate the energy efficiency and performance of the deployed models using FPGA demonstrators across diverse IoT and edge hardware platforms.

Applicants should fulfil the following requirements:

- a master's degree in computer science, computer engineering, electronics or relevant areas
- a clear interest in the topic of the position
- excellent command of English
- strong and demonstrable writing and analytical skills
- capacity to work both as an independent researcher and as part of an international team
- capacity and willingness to provide assistance in organizational tasks relevant to the project

The following experience is beneficial:

- Skills: VHDL/Verilog, openCL, HLS tools, Python/pytorch, be familiar with FPGA boards.
- Merit: Proven experience designing digital components with AXI protocols and deploying neural networks on embedded systems

The candidate should submit their CV along with a research plan for the topic, including the overall research goals.

We offer:

- 4-year PhD position in one of the leading nanoelectronics research centers in Europe with a large portfolio of ongoing pan-European and national research and innovation projects
- The chance to do high-level research in one of the most dynamic digital government contexts globally
- Opportunities for conference visits, research stays and networking with globally leading universities and research centers

About the Department

Department of Computer Systems focuses on the areas of design of dependable computing systems, including reliability, verification and testing of nanoelectronic systems, intelligent and control systems, virtual reality, and biorobotics. The department is highly engaged in international cooperation in research, and teaching provided through all study levels – bachelor's, master's and doctoral studies.

About Taltech

TalTech - Tallinn University of Technology (https://www.taltech.ee/en/) is an internationally recognized research university located in Tallinn, Estonia. As the flagship of Estonian engineering and technical education, TalTech offers high-quality education in various fields such as technological, natural, exact, social, and health sciences. With a mission to develop the economy and industry of the Baltic Sea region, TalTech provides an inspiring environment for students with exciting cultural and sporting activities.

About Estonia:

Estonia is a true digital society. In just 20 years, Estonia has become one of the most technologically advanced societies in the world. Some of the fastest broadband speeds in the world are widely available across the country. But more importantly, so is the wireless Internet which covers everything.

In Estonia you are never more than a 30-minute drive away from a forest or a lake. The living environment is very clean, relaxed and safe. According to the World Health Organization, Estonia has the best overall air quality in the entire world.

For further information, please contact Prof Masoud Daneshtalab masoud.daneshtalab@taltech.ee and Prof. Jaan Raik jaan.raik@taltech.ee or visit https://taltech.ee/en/department-of-computer-systems

To get more information or to apply online, visit https://taltech.glowbase.com/positions/859 or scan the the code on the left with your smartphone.