Politecnico di Torino

Dipartimento di Automatica e Informatica

AVVISO DI SEMINARIO

Martedì, 21 Maggio, ore 11.30

DAUIN, Aula Ciminiera (5° piano)

The Quest for Open-Source tinyML Heterogeneous Hardware Acceleration: A 10+ Year PULP Journey"

Prof. Francesco Conti
Università di Bologna

Abstract:
In the last few years, our perception of what constitutes a "tinyML device" has shifted from simple microcontrollers to complex heterogeneous SoCs suited to execute DNNs directly at the extreme edge in real time and at minimal power cost. These devices provide ultra-low latency and high energy efficiency necessary to meet the constraints of advanced use cases that cannot be satisfied by cloud solutions. However, how can tinyML hardware keep up with the evolution of the AI landscape, continuously pushing towards much larger and more complex models? The costs to develop new accelerators and NPUs for each evolutive step in AI are hard to sustain. A possible way forward is given by the open-source model for digital hardware, popularized by RISC-V: multiple actors - both academic and industrial - collaborate on the development of digital technology that can benefit all parties. In this presentation, I discuss a 10+-year "quest" to push the performance and energy efficiency of tinyML further by exploiting a fully open-source model based on the PULP Platform initiative. I show how the open-source cooperative model makes it possible to combine different ideas and contributions in a technologically portable way, acting as an innovation catalyst and enabling the fast pace of evolution required to keep up with new ideas in AI within a tiny power budget.

Bio:
Francesco Conti holds the position of Tenure-Track Assistant Professor in the Department of Electrical, Electronic, and Information Engineering at the University of Bologna, Italy. He completed his Ph.D. in electronic engineering at the same university in 2016 and worked as a Post-Doctoral Researcher at ETH Zürich between 2016 and 2020. His research is centered on hardware acceleration in ultra-low power and highly energy efficient platforms, with a particular focus on System-on-Chips for Artificial Intelligence applications. He is a senior contributor to the open-source PULP Platform project initiative,

and has focused also on technology transfer, most notably as a consultant for the development of the GAP9 System-on-Chip with GreenWaves Technologies.

Over his career, he has contributed to over 90 international conference presentations and journal articles, earning him multiple accolades, such as the 2020 IEEE Transactions on Circuits and Systems Darlington Best Paper Award and the 2018 ESWEEK CODES+ISSS Best Paper Award. He is a member of the IEEE Circuits and Systems Society, Solid-State Circuits Society, and Council for Electronic Design Automation, and serves as Associate Editor for the IEEE Transactions on Computer-Aided Design of Circuits and Systems.