Energy-Efficient AI Chips for Atrial Fibrillation Detection

A system for the reliable detection of atrial fibrillation that is as energy-efficient as possible – that was the task of the pilot innovation competition of the German Federal Ministry of Education and Research (BMBF).

AI systems can improve healthcare, increase patients' chances of recovery and support doctors in their diagnoses. But they may also require high amounts of energy. Together with the Fraunhofer Institute for Integrated Circuits IIS, a working group from our High Performance Computing department was awarded first prize. Our project HALF (Holistic AutoML for FPGAs) is a holistic approach to the optimization of artificial neural networks and FPGA (Field Programmable Gate Arrays) architectures. For the competition, the neural network not only had to consider performance, but also had to be supplemented by the factor of energy efficiency, so that the dangerous atrial fibrillation is detected with maximum efficiency to allow longer runtimes for mobile ECG devices.


© istockphoto/Nastasic

AI model decides on the energy consumption of the hardware

But how exactly do you find the networks that meet the defined requirements and specifications? "There are various search strategies here, and we use an evolutionary approach. We start with ten different randomly selected networks, train and test them. Then we select the two best networks and mutate them, creating ten new network variants," explains Dr. Jens Krüger. "We repeat this process with a unique approach for the selection and mutation until we find the best network. This is a process of automated machine learning – or AutoML."

In search of the ideal network

The researchers are extending this process to include a holistic approach that looks not only at the neural network but also at the hardware, since the AI model affects the hardware's energy consumption. To do this, they use FPGAs as a form of a programmable chip. An FPGA can be reprogrammed any number of times – very helpful in finding the optimal neural network mapping.

This has now resulted in a new unifying methodology that is energy efficient, reduces development time for optimal neural network topologies, and enables appropriate FPGA implementations. These software tools are suitable not only for FPGAs, but also for a wide variety of chips and system architectures, and enable the analysis of patient data even on mobile devices. This potential has been recognized and awarded by the BMBF a resulted in the Neural Architecture Search Engine (NASE) Tool.

Further information at www.itwm.fraunhofer.de/HALF_en

Contact

Dr. Jens Krüger
Department "High Performance Computing"
Phone +49 631 31600-4541
jens.krueger@itwm.fraunhofer.de