Evolutionary Hardware-Aware Neural Architecture Search

To reduce human effort, neural architecture search (NAS) methods have been developed to automate the design process of deep neural networks (DNNs). Later, hardware-aware NAS methods were introduced to optimize DNN architecture (and weights) for a given hardware platform. The current trend is to co-optimize the hardware accelerator as a part of NAS and utilize the principles of approximate computing to optimize the resulting accelerator further. In this talk, we will survey the critical elements of NAS methods that – to various extents – consider hardware implementation of the resulting DNNs and employ an evolutionary algorithm as the main search engine. As most research in this area deals with NAS for image classification using convolutional neural networks (CNN), our case studies will be devoted to this application. We will present the EvoApproxNAS method capable of an automated design of CNN topology with an automated selection of suitable approximate multipliers for each convolutional layer to find the best tradeoff between the CNN's accuracy and energy.