FPGA Machine Learning Datapaths

Invited Paper/Abstract

By Martin Langhammer, Intel Corporation

Of the many different implementation targets for machine learning, FPGAs are unique in their ability to provide both computational density and computational flexibility. In addition to the large numbers of embedded integer and floating point blocks, the traditional soft logic resources can be configured to create extensive arrays dense low precision integer structures. The floating point blocks modes of the embedded resources are typically used for training, although the larger numbers of integer operators in these blocks can also be combined with soft logic implemented structures to create additional floating point functions. Inference can use these embedded fixed point operators, smaller integer operators extracted from the larger integer blocks, or the soft logic datapaths. Complementing the arithmetic resources are memory structures of various granularity, with bandwidth enabled by configurable connectively able to source and sink each embedded arithmetic block independently.

In this short paper we will review all of the features of modern FPGAs that can be used to implement machine learning applications. We will then examine high density tensor constructions in soft logic, and show that the ability to tailor the precision and representations of the arithmetic generate a high performance, low power solutions for a wide variety of workloads.