Contribution ID: **95** Type: **Lightning 5 min talk + poster**

Leveraging Intel FPGA AI Suite and AI Tensor Blocks for Real-Time, Energy-Efficient Deep Learning Inference

In the presentation, the introduction of the Intel FPGA AI Suite alongside the revolutionary AI Tensor Blocks recently incorporated into the latest FPGA device families by Intel for deep learning inference is showcased. These innovative FPGA components bring real-time, low-latency, and energy-efficient processing to the forefront. They are supported by the inherent advantages of Intel FPGAs, which include I/O flexibility, dynamic reconfiguration, and long-term support. This presentation will explore the Intel FPGA AI Suite, demonstrating its flexibility in achieving scalable performance and seamless integration with industry-leading frameworks such as TensorFlow and PyTorch, facilitated by Quartus Prime Software. Additionally, the pivotal role of AI Tensor Blocks in boosting deep learning inference performance is emphasized.

Focus areas

Author: AHMAD, Jahanzeb

Presenter: AHMAD, Jahanzeb

Session Classification: Lighting talks