ikatod / Freepik

TOP 6 REASONS

to choose FPGA for hardware acceleration of DNN models

Deep Neural Network

ignitarium
Spark On.

ABOUT THIS CHECKLIST

Hardware acceleration is defined as a process in which an application will offload a high computational task into specialised hardware for achieving high efficiency when compared to software implementation in CPU alone. To achieve accurate results in real-time, better models operating on a larger dataset are required. Also, time taken for decision making is an important factor. As new Deep Learning models evolve, the model structure becomes more complex. Thus, a huge number of operations and parameters, as well as more computing resources are needed. Three options for Hardware Accelerators are GPUs, ASICs and FPGAs.

Read through this e-book to understand why engineers are increasingly choosing FPGAs over GPUs and ASICs for hardware acceleration of Deep Neural Network models.

TOP 6 REASONS

1. FPGA offers high performance per watt when compared to GPU, making it a strong candidate for DNN computations and inference.

2. Architecture is customizable and flexible so that the required resources can be used.

3. Provide high throughput with massive parallelism at low latency.

4. FPGA has block RAM which allows faster data transfer compared to off-chip memory.

5. FPGAs are reconfigurable according to application. This enables a reduction in time to market. As the new machine learning algorithm evolves, less development time and reconfigurability make them a better option when compared to ASIC.

6. Apart from power efficiency and throughput, the speed of a DNN deployed on an FPGA can be further increased when the inferred algorithm uses low numeric precision in the calculation. For example, the quantization process converts a 32-bit or 64-bit floating-point network model to a fixed point which reduces computations by maintaining reasonable accuracy.

CONCLUSION

While FPGAs are being increasingly used, engineers still hesitate to adopt FPGA due to difficulty in programming.

To reduce complexity, tools like High-Level Synthesis (HLS) that synthesize high level languages to HDL codes exist. There are different hardware frameworks developed by FPGA vendors and other third-party companies to implement inference on FPGA.

Composite teams with in-depth understanding of AI-ML frameworks and FPGA architectures are able to leverage these tools in order to accomplish high computing goals for their products.

ignitarium
Spark On.

LOOKING TO ACCELERATE YOUR HARDWARE FOR DNN MODELS?

As new Deep Learning models evolve, the model structure becomes more complex. Thus, a huge number of operations and parameters, as well as more computing resources are needed. That's exactly why product teams start looking at options for hardware acceleration.

We understand this.

Our VLSI engineers have executed several projects for clients which include implementation of quality inspection machine vision cameras for the assembly line and real time inferences for endoscopy images on popular FPGA platforms by Xilinx, Lattice and Intel.

Schedule a complimentary consultation with our experts to see how we can help you with using FPGAs for high computing tasks in your product. Together, we can chart your path forward.

Here's what you will walk away with:

1. Clarity on why FPGAs for DNNs
2. Frameworks available
3. Process flow for implementation
4. Verification and benchmarking methods available

SCHEDULE A COMPLIMENTARY CONSULTATION

partnership with

XILINX intel. LATTICE SEMICONDUCTOR