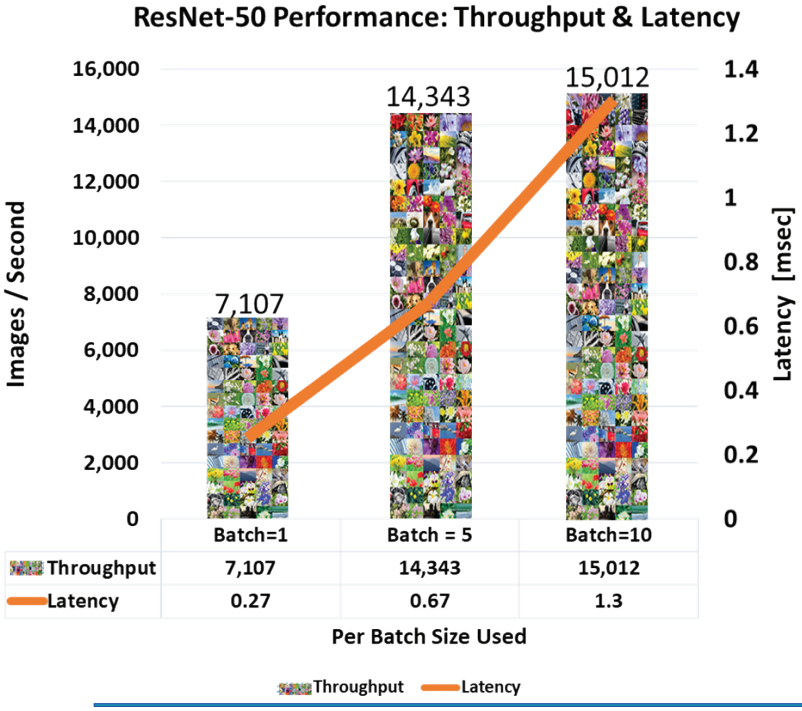
habana
Goya

Habana Labs Goya™ is the industry's first commercially available deep learning inference processor product-line designed specifically to deliver superior performance, power efficiency and cost savings.

Habana Labs HL-10x PCIe cards incorporate a single Goya HL-1000 Processor and are designed to accelerate various AI inferencing workloads, such as image recognition, neural machine translation, sentiment analysis, recommender systems and many others.

Goya HL-1000 processor is a ground up design for neural network processing that incorporates a fully programmable TPC™ (Tensor Processing Core) along with its SynapseAI™ software stack.



SPECIFICATIONS

Processor Technology	**Goya HL-1000**
System Interface	**PCIE Gen 4.0 x 16** **8-pin auxillary power connector** **full length, full height**
Form Factors & Product SKUs	**HL-100: dual-slot, passive cooling** **HL-101: single-slot, passive cooling** **HL-102: dual-slot, active cooling**
Memory	**4/8/16GB DDR4 memory** **ECC protected**
Thermal Design Power	**TDP - 200W**
Scenario Power & Efficiency	**ResNet-50: 15,012 images/second @100W (=150 images/sec/watt)**

habana
Goya

Fast Deployment with SynapseAI:

Habana Labs' SynapseAI is a comprehensive inference software toolkit that simplifies the development and deployment of deep learning models. SynapseAI provides inference network model compilation and runtime, eliminating the need of low level programing.

The Goya platform is training-platform-agnostic: the deep neural network can be trained on any hardware platform. The SynapseAI then compiles the trained model for use on the Goya processor and the result is an optimized execution code in terms of accuracy, latency, throughput and efficiency for inferencing usage.

Goya Software Stack