Handle Up to 2.94x the Frames Per Second for ResNet50 Image Classification with AWS M6i Instances Featuring 3rd Gen Intel® Xeon® Scalable Processors

ResNet50

Classify 2.94x the frames per second on 96 vCPU AWS M6i instances with 3rd Gen Intel Xeon Scalable processors

vs. M6a instances with AMD EPYC processors

Classify 2.49x the frames per second on 16 vCPU AWS M6i instances with 3rd Gen Intel Xeon Scalable processors

vs. M6a instances with AMD EPYC processors

For Image Classification Workloads, M6i Instances Outperformed M6a Instances with AMD EPYC Processors at Multiple Instance Sizes

For organizations selecting AWS cloud instances to run image classification workloads, knowing which instance type provides the performance to reduce time to insights can be a challenge. Test results show for image classification, AWS M6i instances featuring 3rd Gen Intel® Xeon® Scalable processors can meet demanding deep learning needs.

The latest 3rd Gen Intel Xeon Scalable processors offer features that focus on boosting deep learning workloads, including Intel Advanced Vector Extensions (Intel AVX-512) instructions as well as Intel Deep Learning Boost Vector Neural Network Instructions (VNNI). Principled Technologies tested two AWS instance types using int8 precision and batch size of 128 to compare ResNet50 image classification performance across instance sizes.

- AWS M6i instances with 3rd Gen Intel Xeon Scalable processors
- AWS M6a instances with 3rd Gen AMD EPYC processors

In these tests, AWS M6i instances with 3rd Gen Intel Xeon Scalable processors classified up to 2.94x the frames per second compared to M6a instances with AMD EPYC processors. This significant level of increased performance indicates that organizations could purchase fewer cloud instances to handle their deep learning needs by selecting AWS M6i instances.

Making Sense of Data Faster on Large Instances

On large, 96 vCPU instances, AWS M6i instances with 3rd Gen Intel Xeon Scalable processors offered a significant performance advantage over the M6a instances with AMD EPYC processors. As Figure 1 shows, AWS M6i instances classified 2.94x the frames per second of M6a instances.

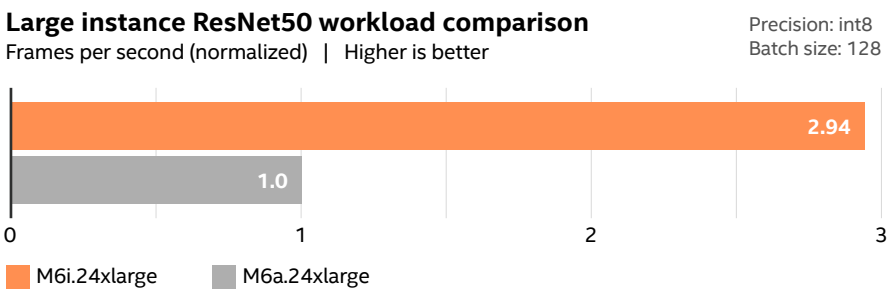

Figure 1. The relative rate of frames per second for large M6i and M6a instances (96 vCPUs) completing image classification using the ResNet50 benchmark. Higher numbers are better.

Making Sense of Data Faster on Small Instances

On small, 16 vCPU instances, AWS M6i instances with 3rd Gen Intel® Xeon® Scalable processors offered a significant performance advantage over the M6a instances with AMD EPYC processors, classifying 2.49x—more than twice—the frames per second on the ResNet50 benchmark (see Figure 2).

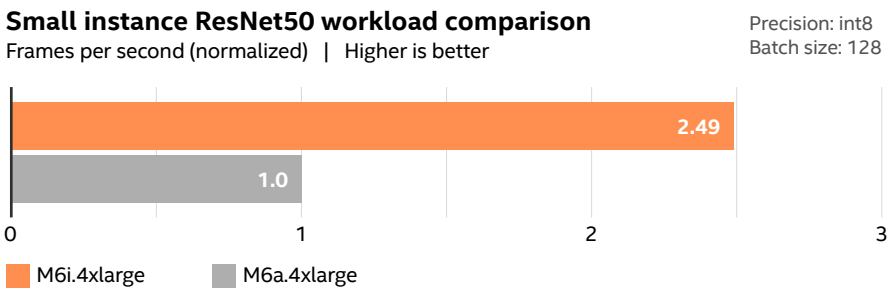

Figure 2. The relative rate of frames per second for small M6i and M6a instances (16 vCPUs) completing image classification using the ResNet50 benchmark. Higher numbers are better.

Conclusion

Whether you run image classification workloads to improve self-driving cars or to assess patients for medical diagnoses, classifying more frames per second can lead to faster insights. Tests show that opting for AWS M6i instances with 3rd Gen Intel Xeon Scalable processors over M6a instances with 3rd Gen AMD EPYC processors can accelerate image classification by as much as 2.94 times.

Learn More

To begin running your image classification workloads on AWS M6i instances, visit https://aws.amazon.com/ec2/instance-types/m6i/.

For more details about these third-party test results, visit https://facts.pt/DNLGDAd.

Printed in USA 0322/JO/PT/PDF US001 ♻ Please Recycle