



TAKE AI PERFORMANCE TO THE NEXT LEVEL

With Intel® Deep Learning Boost



Available on 2nd Generation Intel® Xeon® Scalable Processors

EMBEDDED AI ACCELERATION

Intel® AVX-512 with Intel® DL Boost (VNNI)

1 Instruction for int8 Convolutions



Intel® AVX-512 without Intel® DL Boost

3 Separate Instructions for int8 Convolutions



DEEP LEARNING USE CASES

Image Recognition

Object Detection

Recommendation Systems

Speech Recognition



OPTIMIZED FRAMEWORKS AND DATA TYPES**



OPTIMIZED LIBRARIES

INTEL® MKL-DNN

**This list and more frameworks are optimized. Intel® Optimization for Caffe Resnet-50 performance does not necessarily represent other framework performance.



Intel® Xeon® 9200 Platinum Processors

up to **30x FASTER** with Intel® Deep Learning Boost¹



2nd Generation Intel® Xeon® Scalable Processors

up to **14x FASTER** with Intel® Deep Learning Boost²

ACCELERATE INFERENCE PERFORMANCE

Intel® Deep Learning Boost



Learn More: ai.intel.com intel.com/yourdata

All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest Intel product specifications and roadmaps.

Performance results are based on testing as of 7/26/2018(2x and 3X) and may not reflect all publicly available security update. See configuration disclosure for details. No product can be absolutely secure. Optimization Notice: Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit: <http://www.intel.com/performance>

1) Up to 30X AI performance with Intel® DL Boost compared to Intel® Xeon® Platinum 8180 processor (July 2017). Tested by Intel as of 2/26/2019. Platform: Dragon rock 2 socket Intel® Xeon® Platinum 9282(56 cores per socket), HT ON, turbo ON, Total Memory 768 GB (24 slots/ 32 GB/ 2933 MHz), BIOS: SE5C620.86B.0D.01.0271.120720180605 (ucode: 0x200004d), Ubuntu 18.04.1 LTS, kernel4.15.0-45-generic, SSD 1x sda INTEL SSDSC2BA80 SSD 745.2GB,nvme1n1 INTEL SSDPE2KX040T7 SSD 3.7TB, Deep Learning Framework: Intel® Optimization for Caffe version: 1.1.3(commit hash: 7010334f159da247db3fe3a9d96a3116ca06b09a), ICC version 18.0.1, MKL DNN version: v0.17 (commit hash: 830a10059a018cd2634d94195140cf2d8790a75a,model: https://github.com/intel/caffe/blob/master/models/intel_optimized_models/int8/resnet50_int8_full_conv.prototxt, BS=64, DummyData, 4 instance/2 socket, Datatype: INT8vs Tested by Intel as of July 11th2017: 25 Intel® Xeon® Platinum 8180 CPU @ 2.50GHz (28 cores), HT disabled, turbo disabled, scaling governor set to "performance" via intel_pstate driver, 384GB DDR4-2666 ECC RAM, CentOS Linux release 7.3.1611 (Core), Linux kernel 3.10.0-514.10.2.el7.x86_64, SSD: Intel® SSD DC S3700 Series (800GB, 2.5in SATA 6Gb/s, 25nm, MLC),Performance measured with: Environment variables: KMP_AFFINITY=granularity=fine, compact, OMP_NUM_THREADS=56, CPU Freqset with cpupowerfrequency-set -d 2.5G -u 3.8G -g performance. Caffe: (http://github.com/intel/caffe/), revision f96b759f71b2281835f690af267158b82b150b5c. Inference measured with "caffetime --forward_only" command, training measured with "caffetime" command. For "ConvNet" topologies, dummy dataset was used. For other topologies, data was stored on local storage and cached in memory before training. Topology specs from https://github.com/intel/caffe/tree/master/models/intel_optimized_models (ResNet-50), Intel C++ compiler ver. 17.0.2 20170213, Intel MKL small libraries version 2018.0.20170425. Cafferun with "numactl-l".

2) Up to 14X AI Performance Improvement with Intel® DL Boost compared to Intel® Xeon® Platinum 8180 Processor (July 2017). Tested by Intel as of 2/20/2019. 2 socket Intel® Xeon® Platinum 8280 Processor, 28 cores HT On Turbo ON Total Memory 384 GB (12 slots/ 32GB/ 2933 MHz), BIOS: SE5C620.86B.0D.01.0271.120720180605 (ucode: 0x200004d), Ubuntu 18.04.1 LTS, kernel4.15.0-45-generic, SSD 1x sda INTEL SSDSC2BA80 SSD 745.2GB,nvme1n1 INTEL SSDPE2KX040T7 SSD 3.7TB, Deep Learning Framework: Intel® Optimization for Caffe version: 1.1.3(commit hash: 7010334f159da247db3fe3a9d96a3116ca06b09a), ICC version 18.0.1, MKL DNN version: v0.17 (commit hash: 830a10059a018cd2634d94195140cf2d8790a75a,model: https://github.com/intel/caffe/blob/master/models/intel_optimized_models/int8/resnet50_int8_full_conv.prototxt, BS=64, DummyData, 4 instance/2 socket, Datatype: INT8vs Tested by Intel as of July 11th2017: 25 Intel® Xeon® Platinum 8180 CPU @ 2.50GHz (28 cores), HT disabled, turbo disabled, scaling governor set to "performance" via intel_pstate driver, 384GB DDR4-2666 ECC RAM, CentOS Linux release 7.3.1611 (Core), Linux kernel 3.10.0-514.10.2.el7.x86_64, SSD: Intel® SSD DC S3700 Series (800GB, 2.5in SATA 6Gb/s, 25nm, MLC),Performance measured with: Environment variables: KMP_AFFINITY=granularity=fine, compact, OMP_NUM_THREADS=56, CPU Freq set with cpupowerfrequency-set -d 2.5G -u 3.8G -g performance. Caffe: (http://github.com/intel/caffe/), revision f96b759f71b2281835f690af267158b82b150b5c. Inference measured with "caffetime --forward_only" command, training measured with "caffetime" command. For "ConvNet" topologies, dummy dataset was used. For other topologies, data was stored on local storage and cached in memory before training. Topology specs from https://github.com/intel/caffe/tree/master/models/intel_optimized_models(ResNet-50), Intel C++ compiler ver. 17.0.2 20170213, Intel MKL small libraries version 2018.0.20170425. Cafferun with "numactl-l".

Intel, the Intel logo, Xeon, and Xeon Scalable logos are trademarks of Intel Corporation in the U.S. and/or other countries. *Other names and brands may be claimed as the property of others. ©Intel Corporation.