



Intel® Software Development Tools Optimize Deep Learning Performance for Healthcare Imaging

The Deep Learning Deployment Toolkit from Intel helps deliver optimized inferencing on Intel® architecture, helping bring the power of AI to clinical diagnostic scanning and other healthcare workflows



OPTIMIZED MODEL
Exceeds GE Inferencing Target

14x FASTER	5.9x ABOVE TARGET
----------------------	--------------------------------

Executive Summary

Deep learning and other forms of artificial intelligence (AI) offer exciting potential to streamline medical imaging workflows, enhance image quality, and increase the research value of imaging data. To deliver on the promise of AI-enhanced medical imaging, developers and other innovators must deploy their AI solutions on flexible platforms that provide high performance and scalability for deep learning innovations without driving up costs.

As part of its company-wide commitment to the AI revolution, Intel offers processors, tools, and frameworks for high-performance deep learning on cost-effective, general-purpose Intel® architecture. In addition to Intel® Xeon® Scalable processors and Intel® Solid State Drives, key technologies include Intel's Deep Learning Deployment Toolkit and the Intel® Math Kernel Library for Deep Neural Networks (Intel® MKL-DNN). These technologies provide an easy way for innovators to deploy and integrate their deep learning models in optimal fashion on various Intel architectures from a variety of frameworks and training platforms. Using Intel architecture, developers can deliver their innovations without adding costs and complexity to the deployment environment.

Intel and GE Healthcare explored the Intel® technologies with one of GE's deep learning image-classification solutions. They found that optimizing the trained GE solution with Intel's Deep Learning Deployment Toolkit and Intel MKL-DNN improved throughput an average of 14 times over a baseline version of the solution and exceeded GE's throughput goals by almost six times. These findings show a path to practical AI deployment for next-generation diagnostic scanners and a new era of smarter medical imaging.

Table of Contents

- Executive Summary 1
- Business Challenge:
High-Performance Inferencing for
Practical Deep Learning 2
- Solution Overview: Deep Learning
Deployment Toolkit from Intel 2
- Putting Intel's Inferencing
Performance to the Test 3
- Flexible Deployment..... 3
- Conclusion..... 4



Business Challenge: High-Performance Inferencing for Practical Deep Learning

Medical images are valuable in diagnosing a wide range of health issues, as well as planning treatment and assessing its results. Aggregated with other sources of healthcare and demographic information, imaging data can also lead to novel insights that inspire next-generation treatment breakthroughs.

Deep learning and other forms of artificial intelligence (AI) offer exciting potential to improve medical imaging workflows and enhance medical imaging quality. Deep learning is a branch of AI and machine learning in which developers create mathematical and neural network models, and use vast amounts of data to “train” them to perform tasks such as recognizing and classifying medical images. Once a trained model is sufficiently accurate, it can be deployed with other algorithms as an “inference engine,” to evaluate and categorize real-world inputs from digital cameras and sensors.

Deep learning solutions with Intel® technologies are helping to deliver high-throughput inferencing in a variety of medical imaging use cases. For example, a development team from Zhejiang University’s School of Mathematics Sciences* and Zhejiang DE Image Solutions Co., Ltd.* developed a deep learning solution that is helping to improve thyroid cancer screening in China. The Intel® technology-based solution examines ultrasound images to identify and classify thyroid lesions.

China’s government is using an Intel technology-based deep learning solution to expand the capacity of its health system to detect two common causes of preventable blindness. Aier Eye Hospital Group*, China’s leading hospital network in eye care, worked with Medimaging Integrated Solutions* (MiiS*) to create the cloud-based solution, which runs on Intel® Xeon® Scalable processors and integrates deep learning inferencing capabilities with a handheld ophthalmoscope from MiiS.

For medical imaging AI solutions, the deployment architecture must deliver high inferencing throughput that keeps pace with busy radiology workflows, but doesn’t restrict flexibility or add needless complexity and costs to the deployment environment. Training infrastructure is

significantly different from inference infrastructure, and choosing a highly optimized inference infrastructure can offer significant speed-up in inference throughput.

Solution Overview: Deep Learning Deployment Toolkit from Intel

In addition to its powerful processors and solid state drives (SSDs), Intel offers tools for optimized inferencing on flexible, cost-effective Intel® architecture.

The Deep Learning Deployment Toolkit from Intel is a free set of tools that lets users optimize deep learning models for faster execution on Intel® processors. The toolkit imports trained models from Caffe*, Apache MXNet*, and TensorFlow*, regardless of the hardware platforms used to train the models. Developers can quickly integrate various trained neural network models with application logic using a unified application programming interface (API). The toolkit maximizes inference performance by reducing the solution’s overall footprint and optimizing performance for the chosen Intel® architecture-based hardware.

The toolkit also enhances models for improved execution, storage, and transmission and enables seamless integration with application logic. Developers can also test and validate their models on the target hardware environment to confirm accuracy and performance. The result is an embedded-friendly inferencing solution with a small footprint and excellent performance for high-throughput deployment on industry-standard technologies.

Intel’s Deep Learning Deployment Toolkit includes the **Intel® Math Kernel Library for Deep Neural Networks (Intel® MKL-DNN)**, a high-performance library designed to accelerate neural network primitives, increase application performance, and reduce development time on multiple generations of Intel processors. Intel MKL-DNN is also available as a standalone, open source package. The Deep Learning Deployment Toolkit is a component of the Intel® Computer Vision Software Development Kit.

Intel® Deep Learning Deployment Toolkit for AI-Enhanced CT Imaging at GE Healthcare

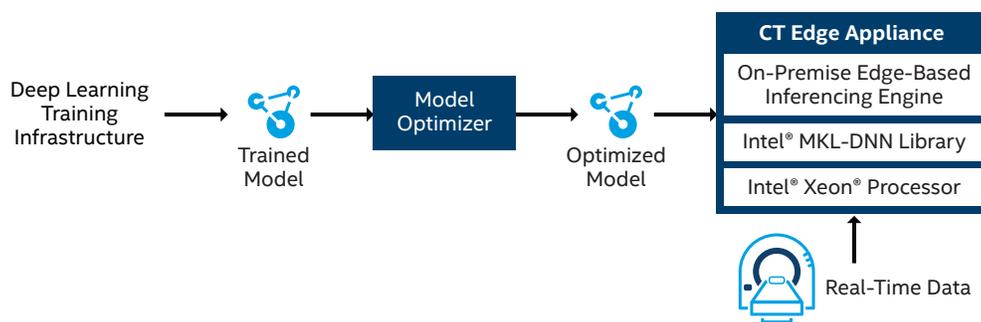


Figure 1. Optimized models use the Deep Learning Deployment Toolkit from Intel and the Intel® Math Kernel Library for Deep Neural Networks (Intel® MKL-DNN) to deliver outstanding inferencing performance for practical deployment of AI solutions at the “edge” of the enterprise in clinical and research settings.

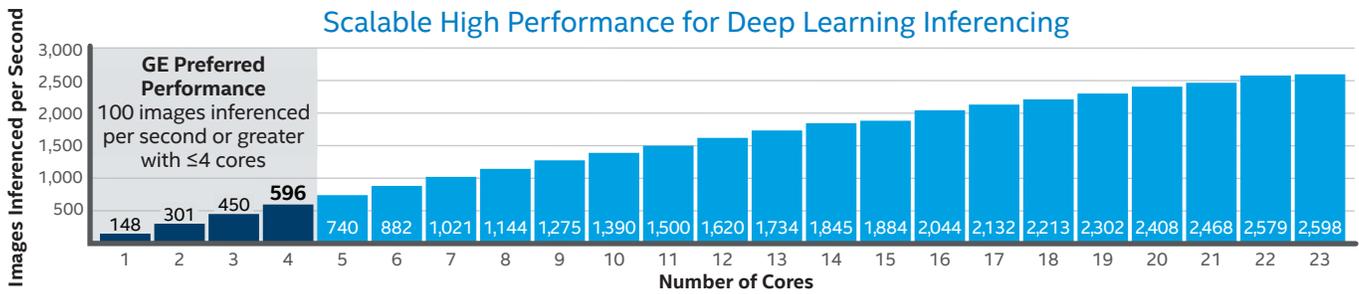


Figure 2. The Intel® Xeon® processor E5-2650 exceeded GE’s inferencing performance target by nearly six times on four cores, and demonstrated steady, scalable performance for solutions needing more cores.

Putting Intel's Inferencing Performance to the Test

GE Healthcare, a leading provider of medical imaging equipment and other healthcare technologies, worked with Intel to test inferencing performance for one of GE’s deep learning solutions. Developed by a team from GE Healthcare’s Computed Tomography (CT) scan division, the solution classifies and tags scanned image slices, making it easier to find relevant images and use them for research or clinical comparison. GE presented a paper on its AI anatomy classifier at the prestigious SPIE Medical Imaging 2018 Conference*. SPIE* is an international society advancing an interdisciplinary approach to the science and application of light.

GE’s CT imaging specialists had developed and trained their new AI-based model using the Python* programming language and open source software, including the TensorFlow and Keras* libraries for deep learning. Collaborating with some of Intel’s AI-optimization experts, they used Intel’s Deep Learning Deployment Toolkit to set up and optimize the solution for the Intel technology environment and test the solution’s performance. Reflecting the need for cost-effective inference performance, the CT experts set a target of classifying 100 images per second using no more than four dedicated cores of an Intel® Xeon® processor.

The solution ran on a system powered by the Intel® Xeon® processor E5-2650 v4 at 2.20GHz, and configured with 264 GB of memory, an Intel® Solid State Drive Data Center 480 GB, and CentOS Linux* 7.4.1708. The system ran at Intel’s Datacenter Health and Life Sciences Lab in Hillsboro, Oregon, and used a privacy-protected data set of 8,834 CT scan images.

The results were impressive. The optimized codes produced by Intel’s Deep Learning Development Toolkit and Intel MKL-DNN improved inferencing throughput an average of 14 times over the baseline TensorFlow model running on the same system, and easily met GE’s performance target. In fact, a single core of the Intel Xeon processor E5-2650 ran nearly 150 percent faster than GE’s performance target, and four cores of the processor exceeded GE Healthcare’s goal by nearly six times (see Figure 2). Intel benchmarks show the new Intel® Xeon® Platinum 8180 processor delivering up to 2.4 times higher performance for a range of AI workloads compared to the previous generation.

“We want to keep deployment costs down for our customers, so we need the performance and flexibility to run a whole range of AI and imaging processing tasks in a variety of clinical environments,” said David Chevalier, principal engineer, GE Healthcare. “We think using general-purpose processors, tools, and frameworks from Intel can offer a cost-effective way to leverage AI in medical imaging in new and meaningful ways.”

Flexible Deployment

By delivering excellent inferencing performance on just a handful of CPU cores, Intel Xeon processors and Deep Learning Deployment Toolkit can enable innovators such as GE Healthcare to run one or more AI solutions at high performance on the same cost-effective edge-server infrastructure that handles local image-processing tasks such as reconstruction, registration, segmentation, and noise reduction. AI developers can take advantage of Intel’s performance-tuned frameworks and tools to deploy their inferencing solutions on Intel architecture regardless of the development and training environment they used.

CPU-based inferencing on Intel processors also offers innovators such as GE the flexibility to change where inferencing occurs. While medical imaging AI deployments currently run in inferencing appliances at the on-premise server, next-generation inferencing can move to an on-premise or off-premise cloud or into the scanner itself (Figure 3).

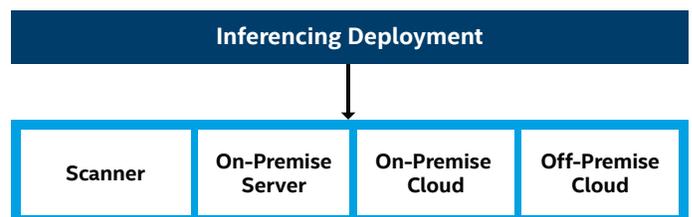


Figure 3. Running on Intel® processors gives AI innovators the flexibility to deploy inferencing capabilities in diagnostic imaging equipment, on-premises infrastructure, or secure external clouds.

Conclusion

With high inferencing throughput on flexible Intel architecture, innovators such as GE Healthcare move closer to delivering practical medical imaging solutions that use deep learning and other forms of AI to improve image quality, diagnostic capabilities, and clinical workflows.

Find the solution that is right for your organization. Contact your Intel representative or visit intel.com/healthcare.

Learn More

You may find the following resources useful:

- [Download the Deep Learning Deployment Toolkit from Intel](#)
- [Download the Intel® Math Kernel Library for Deep Neural Networks](#)
- [Read the White Paper: Inferencing Solution Simplifies AI Adoption for Medical Imaging](#)
- [Read the Case Study: Aier Eye Hospital Brings Deep Learning to Blindness Prevention](#)
- [Read the White Paper: Use AI and Imaging Data to Unlock Insights and Improve Healthcare](#)

Read about relevant Intel products:

- [Intel® Xeon® processor Scalable family](#)
- [Deep Learning Deployment Toolkit from Intel](#)
- [Intel® Solid State Drive Data Center Family](#)
- [Intel® Math Kernel Library for Deep Neural Networks](#)

Solution Provided By:



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 information go to intel.com/benchmarks. Benchmark results were obtained prior to implementation of recent software patches and firmware updates intended to address exploits referred to as "Spectre" and "Meltdown". Implementation of these updates may make these results inapplicable to your device or system.

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software, or service activation. Performance varies depending on system configuration. No computer system can be absolutely secure. Check with your system manufacturer or retailer, or learn more at intel.com.

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. Notice revision #20110804

Intel does not control or audit the design or implementation of third-party benchmark data or Web sites referenced in this document. Intel encourages all of its customers to visit the referenced Web sites or others where similar performance benchmark data are reported and confirm whether the referenced benchmark data are accurate and reflect performance of systems available for purchase.

Intel the Intel logo, and Xeon are trademarks of Intel Corporation in the U.S. and other countries.

GE, GE Healthcare, and the GE logo are trademarks of General Electric Company

*Other names and brands may be claimed as the property of others. © Intel Corporation

0318/JSTAR/KC/PDF

336993-001US