CASE STUDY

Health & Life Sciences

GE Healthcare: Speeding Cardiac Ultrasound with AI Technologies

Intel and GE Healthcare used deep learning tools to improve the efficiency of ultrasound measurements

As part of an automatic tool called "Auto VTI" for measurement of cardiac echo, GE Healthcare created a module that uses deep learning to automatically position the measurement point on the image. GE Healthcare's inference engine was hand-coded in C++, though, which made it difficult to develop and to maintain. Intel showed GE Healthcare an easier way: Using the Intel® Distribution of OpenVINO™ toolkit (open visual inference and neural network optimization), trained models could be optimized and run with 3.63x[1] the performance of the bespoke code[2].

Challenge

- The Venue family of point-of-care ultrasound systems from GE Healthcare, including Venue and Venue Go, was based on bespoke C++ code for deep learning inference.
- The C++ code wasn't compatible with future deep learning models GE might want to use, and was difficult to code and maintain.
- GE had to perform hand-coded optimizations to take advantage of Intel® Advanced Vector Extensions 2 (Intel® AVX2).

Solution

- The Intel Distribution of OpenVINO toolkit enables GE Healthcare to optimize its TensorFlow-based Keras model for deep learning.
- The inference engine in the toolkit is itself optimized to take advantage of Intel® architecture features, including Intel AVX2.

Results

- Using the Intel Distribution of OpenVINO toolkit, GE Healthcare was able to achieve performance 3.63x faster than its original code[1].
- Intel helped GE Healthcare integrate the toolkit in the software for GE Healthcare's Venue family of artificial intelligence (AI)-enabled ultrasound systems.

Using AI to Automate Heart Measurements

In busy hospitals where quality of care is a priority, it's important for doctors to be able to get test results quickly. Using portable ultrasound equipment, doctors can measure blood flow in the heart at the patient's bed, without requiring the support of specialist ultrasound operators. The measurement taken is called the Left Ventricle Outflow Tract Velocity Time Integral (LVOT VTI). It plays an important role in treating heart conditions including cardiogenic or other types of shock, where organs do not get enough blood or oxygen.

[2] For more complete information about performance and benchmark results, visit www.intel.com/benchmarks. For more complete information about security notices and configurations see Endnote 1.

The measurement process is difficult, though: it involves positioning the probe on the patient's body and keeping it stable while positioning the measurement point (the doppler gate) on the console touchscreen. The signal is embedded in white noise, so the physician needs to understand and trace the signal on the touchscreen. Some doctors performed this measurement less often because the process is so cumbersome.

GE Healthcare provides medical diagnostic and monitoring technologies, including the Venue family of point-of-care ultrasound systems which includes Venue and Venue Go. The company spotted an opportunity to use deep learning during algorithm development (specifically, convolutional neural networks) to create a solution that helps doctors position the doppler gate automatically, making measurements easier to take. For its Venue family of AI-enabled scanners, GE Healthcare trained a deep learning model using the Keras neural network library using a set of ultrasound images with data for the correct location of the doppler gate. (The algorithms are developed, compiled, and sealed in GE's lab where they undergo full testing. The validated algorithms do not use machine learning or otherwise change when in use by doctors.)

GE Healthcare coded its own inference engine to achieve the performance its software required. Because the scanning equipment is portable and battery-powered, GPUs could not be used due to power limitations. The systems were designed to allocate workloads to specific processor cores. Since only one core was available for the deep learning workload, GE Healthcare needed to optimize the single-core performance of its inference engine. Functions had been hand-coded to take advantage of the Intel AVX2 instructions available on the Intel® Core™ i7 processor used in the ultrasound device. Intel AVX2 provides instructions that can process multiple data items at the same time, accelerating performance. GE also used Intel® Integrated Performance Primitives, a library of ready-to-use functions that are highly optimized for Intel architecture, which speeds up the software development cycle.

Nevertheless, developing and maintaining a bespoke inference engine is a difficult undertaking. "Our existing C++ inference engine is limited in terms of its abilities," said Dani Pinkovich, who leads the algorithms team at GE Healthcare POCUS. "It doesn't cover all the deep learning models, and was created by two people in a short period of time." Integration of new models would require extensive software development work.

The company wanted to see whether there was a better way, and turned to Intel for help.

Accelerating Inference by 3.63x[1]

Intel proposed a solution based on the Intel Distribution of OpenVINO toolkit. The Intel Distribution of OpenVINO toolkit is a free software kit that helps developers and data scientists speed up computer vision workloads, streamline deep learning inference and deployments, and enable easy, heterogeneous execution across Intel platforms from edge to cloud. This toolkit can be used to optimize an existing deep learning model based on frameworks including TensorFlow, Caffe and MXNet. The toolkit carries out optimizations on the model and converts it into an intermediate representation (IR) file which can then run on the toolkit's inference engine. The inference engine is itself optimized to take advantage of processor features available where the workload runs, such as Intel AVX2 on Intel Core i7 processors.

The latest release, 2019 R3, of the toolkit includes network loading optimizations in order to improve performance and speed up time-to-inference by reducing model loading times. This is especially useful when shape size changes between inferences, such as in cases for medical imaging using a Keras model where the region of interest is pre-defined and/or input shape is defined by the previous network in the pipeline.

Testing showed that using the Intel Distribution of OpenVINO toolkit, single core inference performance was increased by 3.63x[1] (see Figure 1)[3].

"We entered into this research aiming to have an engine that worked as well as our bespoke solution but that also enabled us to use different deep learning models," said Pinkovich. "Using the Intel Distribution of OpenVINO toolkit running on the Intel Core i7 processor, we've not only opened up the opportunity to use a wider range of models, but we've already achieved an excellent performance improvement. Improved performance enables us to use larger models and achieve more accurate doppler gate placement results."

"In many other tools, we have to work with configurations of the system where CPU loading is up to 80 percent," said Pinkovich. "Each percentage of loading on the CPU affects

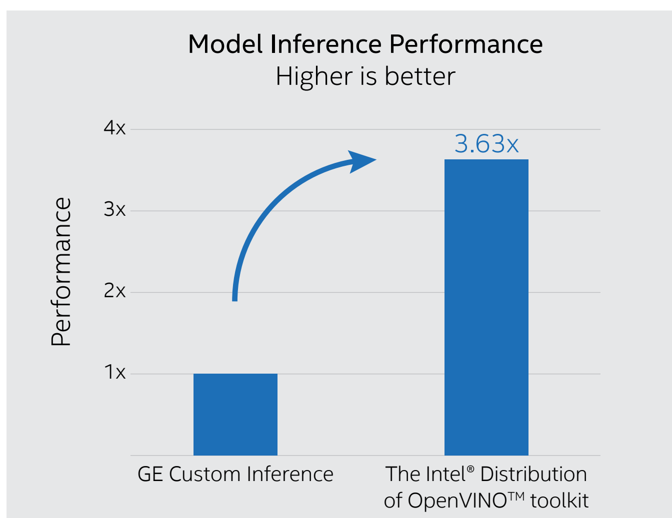

Figure 1. The Intel® Distribution of OpenVINO™ toolkit was 3.63x faster than GE's custom inference engine, coded in C++[1].

[3] For more complete information about performance and benchmark results, visit www.intel.com/benchmarks. For more complete information about security notices and configurations see Endnote 1.

Technical Components of the Solution

- Intel® Distribution of OpenVINO™ toolkit. This toolkit accelerates the development of computer vision solutions, enabling deep learning inference on edge devices. Based on convolutional neural networks, the toolkit optimizes an existing trained model and provides an inference engine and model optimizer that maximizes performance on Intel architecture, such as CPU, integrated GPU, VPU and FPGA.
- Intel® Core™ i7 processor. The Intel® Core™ i7-7567U processor used by GE has a base frequency of 3.50GHz and a max turbo frequency of 4.00 GHz, with a cache of 4MB. It has two cores and supports four threads. It supports a maximum memory size of 32GB and incorporates Intel® Iris® Plus processor graphics.

the time resolution, where, for example, we can only process one in five or one in ten frames. Using the toolkit, we will not have to compromise and will have a better resolution. What's more, we won't have to spend time creating our proprietary inference engine in the future."

Intel Helps Enable Transformation

Intel worked closely with GE Healthcare to run the proof of concept based on the Intel Distribution of OpenVINO toolkit, using trained models that GE provided to Intel's software engineers. "It was an intensive and fast collaboration," said Pinkovich. "Intel created an environment based on the OpenVINO toolkit running our models, and talked us through the results. Then we built the same environment on our main computer in-house, with the help of the team from Intel. We saw the same performance, so the proof of concept was successful. Intel provided us with strong support as we integrated the inference engine into our system."

Using the new solution, GE Healthcare will be able to make it much easier for doctors to measure blood flow in the heart, potentially helping to improve the clinical outcomes for patients.

Spotlight on GE Healthcare

As a leading global medical technology and life sciences company, GE Healthcare provides a broad portfolio of products, solutions and services used in the diagnosis, treatment and monitoring of patients and in the development and manufacture of biopharmaceuticals. GE Healthcare helps to improve outcomes for healthcare providers and for therapy innovators around the world. This means increased capacity, improved productivity and better patient outcomes. GE also aims to support its customers in the pursuit of precision health: health care that is integrated, highly personalized to each patient's needs and that reduces waste and inefficiency.

www.gehealthcare.com

Learn More

- Intel® Distribution of OpenVINO™ toolkit
- Intel® Integrated Performance Primitives
- Intel® Core™ i7-7567U processor
- Latest Performance Benchmarks

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

[1] 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 www.intel.com/benchmarks

Configurations: Testing carried out by Intel 1st September 2019. Intel® Core™ i7-7567U CPU @ 3.50GHz; TensorFlow and Keras; Intel® Distribution of OpenVINO™ toolkit, Microsoft Windows, Custom CNN topology

Performance results are based on testing as of the date set forth in the Configurations and may not reflect all publicly available security updates. See configuration disclosure for details. No product or component 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.

Notice Revision #20110804

Intel does not control or audit third-party data. You should review this content, consult other sources, and confirm whether referenced data are accurate.