(intel®)

Mapping the cortex

Intel® Distribution of OpenVINO™ Toolkit accelerates the analysis of 3D brain scans

The Intel® Distribution of OpenVINO™ Toolkit, an open visual inference and neural network optimization toolkit, is a free software 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.

Research is an indispensable basis for science

Researchers at the Max Planck Institute for Human Cognitive and Brain Sciences* in Leipzig are working on an exciting project that seeks a deeper understanding of the human brain. The task presents a non-trivial challenge. Today, researchers are conducting detailed studies of post-mortem brain tissue, the results of which will be used in the future to obtain insights directly from non-invasive MRI imaging of living brain tissue.

Today's MRI technology, however, has limitations in resolution and cannot provide the necessary detail to represent the cerebral cortex's structure accurately. Therefore, the team plans to learn from high-resolution post-mortem data and correlate that information later with lower-resolution MRI scans of living tissue.

Making a deeper understanding of the brain possible

The human cerebral cortex is composed of complex tissue with a rich inner structure, consisting of multiple cell types with fibrous connections organized in six layers. The spatial arrangement of these layers varies within the cortex depending on the functional organization of different brain regions. Obtaining a deeper understanding of the organization involves the study of the cortex's microstructure at the cellular level.

To date, scientific knowledge of the brain structure derives from the analysis of two-dimensional cross-sections. This approach helped identify the distribution of the brain's functional areas. However, comprehension of the entire three-dimensional microstructure of the brain represents a much more significant challenge. Now, a combination of 3D imaging techniques and advanced image analysis brings scientists closer to understanding the brain in its entirety.

Trained neural networks must process the large 3D images of cell structures produced by high-resolution microscopy. The challenge is that the 3D image data is larger than the previously used 2D images by a factor of 1,000. Even high performance computing (HPC) systems struggle to process the data volumes of this size.

To overcome this hurdle, scientists from the Max Planck Institute for Human Cognitive and Brain Sciences in Leipzig turned to colleagues from the Max Planck Computing and Data Facility* (MPCDF) in Garching. The MPCDF offers the technical expertise and HPC infrastructure needed for scientific simulations.

In search of a universal approach

In early 2018 the scientists determined that a traditional approach, involving splitting the analysis workload into several parallel tasks, was possible. Unfortunately, no available applications could accomplish the task. The team decided to develop an application, which addressed not only their immediate research needs but could also be used in similar future projects.

The biggest challenge lies with the volume of data contained in the 3D images. When working with 2D images of individual brain layers, expert scientists could evaluate each image individually and identify the relevant elements. However, with 3D imaging, thousands of layers would require individual evaluation. A task of that magnitude would take scientists months or even years to complete.

Using machine learning, though, the timeframe for image evaluation can be reduced substantially. By training the algorithm to recognize important aspects of each image, the software can then break down the image into meaningful segments.
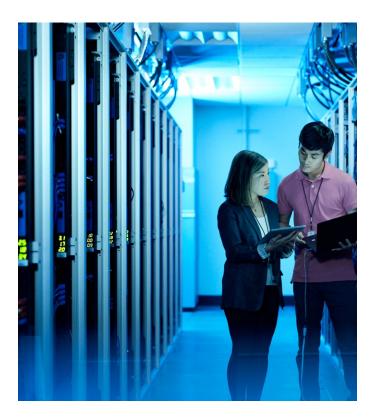
Extremely high hardware requirements

Colleagues at the MPCDF shared their expertise in neural network training techniques and offered an HPC system to help complete the task. A major difficulty consisted in running inference on a full 3D dataset. Using a non-parallelized approach, this effort would require a compute node with 24 TB of RAM. Even the HPC environment at the MPCDF in Garching could not accommodate this extreme requirement on one compute node.

Performing 3D convolutions that are particularly compute-intensive adds another difficulty to the process. A conventional approach using parallelization – distributing the workload homogenously among several nodes using an application developed by the MPCDF – offered a possible ad-hoc approach. The MPCDF, however, wanted to use a non-parallelized application, which offered faster and easier applicability to similar problems without the need to rewrite the application each time.

To meet this need, the MPCDF contacted Intel and the OpenVino team. OpenVINO™ (Open Visual Inference and Neural Network Optimization) toolkit is free software for developers and data analysts that facilitates image processing, optimizes inference applications, and helps run workloads across heterogeneous Intel resources from edge devices to the cloud.

The first and most important achievement of the OpenVino team at Intel was helping the Max Planck Society reduce the 24 TB memory requirement by a factor of 15[1]. As a result, processing each image required only 1.5 TB of RAM.

Unlike standard deep learning frameworks, the OpenVINO toolkit allows reuse of memory as it moves through the layers during inference. Typical frameworks cater to both training and inference, and they retain the values for each layer in memory. Since this is not necessary for inference-only applications, the OpenVINO toolkit discards previous values to re-use memory efficiently through the inference process.

"As a result of the project with the Max Planck Computing and Data Facility (MPCDF), we further developed and optimized the Intel OpenVINO Toolkit. By adopting a flexible approach, we can now use the OpenVINO toolkit for future MPCDF projects and 3D medical imaging in general."

- Yury Gorbachev, Principal Engineer and Lead Architect of OpenVINO™ Toolkit, Intel Corporation

"The 3D data volume is at least a 1,000 times larger than the previous 2D data volume, making the analysis and evaluation of individual layers by human experts impossible. By contrast, with the OpenVINO toolkit processing times of one 3D image are now under an hour."

- Andreas Marek, Senior High Performance Computing Expert, and Lead of the Data Analytics Group, Max Planck Computing and Data Facility (MPCDF)

Extensive customization of the Intel OpenVINO Toolkit

Because the toolkit did not initially offer the necessary 3D convolution capability, Intel developers adapted and extended the software. Numerous other findings and requirements posed by the Max Planck project led to further streamlining of the program code, including those related to the input format ingestion.

Working together, teams from MPCDF and Intel pursued the goal of maximizing the speed of image processing. Initial tests required 24 hours of execution time for each image. This result was already a dramatic improvement compared to the time required for a human to perform the same task. The team did not stop there: their further efforts first reduced image processing time down to eight hours and finally the software is now able to complete the process in less than an hour.

Part of the underlying technical solution used by the MPCDF and Intel included 1.5 TB of Intel® Optane™ DC Persistent Memory in the Intel® Xeon® Gold 6248 processor-based nodes, enabling both the performance characteristics and the price-performance ratio needed for this application. It also provided an excellent substitute for large quantities of DDR4 RAM.

Conclusion

By adopting the Intel® Distribution of OpenVINO™ Toolkit, the MPCDF and Intel developers unleashed the power of neural networks, enabling faster insights from 3D image data while minimizing hardware requirements. Intel Optane DC Persistent Memory played a significant role in the solution. Because the project-specific workload was a perfect match for the performance characteristics of Intel persistent memory architecture, the team achieved excellent results in a cost-efficient manner. Another advantage of the OpenVINO solution is its flexibility. Since the application does not require parallelization, it can be adapted easily for future MPCDF applications or other use cases.

Working in tandem with Intel® Xeon® Scalable processors, Intel® Optane™ DC Persistent Memory can increase system performance by placing workload data closer to the processor, thereby reducing latency. When configured in Memory Mode, Intel Optane DC Persistent Memory behaves like an extended memory pool that can address up to 6 TB in a two-socket system. In App Direct Mode, Optane functions more like RAM with the supplemental benefit of non-volatile memory.

The Max Planck Computing and Data Facility (MPCDF, formerly known as RZG) is a cross-institutional competence center of the Max Planck Society to support computational and data sciences. It originated as the computing center of the Max Planck Institute for Plasma Physics (IPP) which was founded 1960 by Werner Heisenberg and the Max Planck Society (MPS). In close collaboration with domain scientists from different Max Planck Institutes, the MPCDF is engaged in the development and optimization of algorithms and applications for high performance computing as well as in the design and implementation of solutions for data-intensive projects. The MPCDF operates a state-of-the-art supercomputer, several mid-range compute systems and data repositories for various Max Planck institutes, and provides an up-to-date infrastructure for data management including long-term archival.

Summary

Evaluating three-dimensional data helps us develop a better understanding of the human brain. However, analyzing the vast data sets derived from the imaging process creates an enormous Big Data challenge.

Using the Intel® Distribution of OpenVINO™ Toolkit, and Intel® Optane™ DC Persistent Memory the MPCDF team reduced their HPC system memory footprint by a factor of 15, and with it, image analysis time dropped to a mere hour per image.