

Intel, Winning Health and AMAX jointly optimize AI-aided medical system to empower smart healthcare

As one of main application fields, the medical and healthcare industry has a soaring demand for artificial intelligence (AI) technology and is increasing investment in it. To promote “healthcare + AI” and help to build a smart healthcare ecosystem, Winning Health Technology Group Co., Ltd. (hereinafter referred to as “Winning Health”), AMAX and Intel have made extensive cooperation in AI application. The three parties have successfully innovated and optimized the application of AI-assisted devices in healthcare scenarios, further driving AI application in the medical industry.

“AI + medical imaging” is one of the latest collaboration outcomes of Winning Health, AMAX, and Intel. Based on AMAX Deep Learning All-In-One, which carries 2nd Gen Intel® Xeon® Scalable processor and OpenVINO™ toolkit, Winning Health launched a medical imaging AI solution to assist with the diagnosis of 20+ common abnormal images. These diagnostic results can be presented as preliminary medical reports to reduce doctors’ repeated manual labor. At the radiologist’s workstation and in clinician’s report reading system, they can also promptly remind doctors of critical patients’ conditions to avoid treatment delay.

Challenge: How to empower medical imaging diagnosis with AI

With the gradual evolution of medical imaging technology and continuous introduction of medical imaging devices, medical institutions are witnessing an explosive growth of medical imaging data. This trend has brought a heavy workload of medical image reading. Fortunately, AI can help solve this problem. With AI, medical institutions can effectively improve diagnostic accuracy and efficiency, shorten waiting time, and reduce treatment cost.

With regard to AI-assisted medical imaging solution, in addition to the development of deep learning algorithm, the construction of infrastructure platform is also an important factor worth considering.

First, among the options of deep learning infrastructure, traditional deep learning applications often choose GPU for training. Although GPU can deliver the required model inference speed, medical institutions have to purchase additional dedicated-purpose GPU servers. This results in extra costs. Therefore, how to harness existing computing resources for deep learning deployment is an important subject.

Second, although the threshold for deep learning has lowered to some extent, specialists are still indispensable for establishing and maintaining deep learning environment. AI applications, such as AI-aided diagnostic imaging of pulmonary nodules, involve many optional brands and models of software, basic hardware, and dependent packages. It is difficult to choose from them as different tasks require different resources. In addition, the deployment of deep learning system architecture is relatively complicated. It is not easy for the architecture to adapt to various frameworks and models, and its operation and maintenance are also complicated.

Solution: Winning Health AI Medical Imaging solution based on AMAX Deep Learning All-In-One

To help medical institutions promote AI technology in medical imaging applications and deliver AI capability for scenarios like AI-aided diagnostic imaging of pulmonary nodules, Winning Health has built the infrastructure platform with AMAX Deep Learning All-In-One, which is equipped with 2nd Gen Intel Xeon Scalable processor and OpenVINO toolkit.



Figure 1: AMAX Deep Learning All-In-One

AMAX Deep Learning All-In-One employs the concept of software and hardware integration, and deeply integrates machine learning platform with traditional hardware. Leveraging the holistic scheduling of this deep learning platform, AMAX Deep Learning All-In-One features fine-grained authority management and control, safe and efficient data management, and comprehensive and delicate monitoring management capabilities, which can help users maximize resource utilization.

The 2nd Gen Intel Xeon Scalable processor carried by AMAX Deep Learning All-In-One not only delivers powerful general-purpose computing capability, but features Intel® DL Boost technology with vector neural network instructions (VNNI). It greatly improves the performance of AI inference, and accelerates AI inference workloads (such as image recognition, object detection, and image segmentation) in data centers, enterprises, and intelligent edge computing environments. In addition, with outstanding flexibility and scalability, this processor can help medical institutions better meet the needs of multiple workloads and reduce the total cost of ownership (TCO).

The core software of AMAX Deep Learning All-In-One is the AI Max platform that employs OpenVINO toolkit. For deep learning users, this platform provides functions such as rapid resource scheduling, resource usage monitoring, and distributed interactive development. It also supports highly customizable all-in-one hardware based on 2nd Gen Intel Xeon Scalable processor to meet users' computing demands. Meanwhile, AMAX Deep Learning All-In-One integrates and optimizes existing deep learning framework, encapsulates a large number of underlying operations, and completes function debugging before delivery to enable out-of-the-box services, greatly reducing the technical threshold for deep learning practitioners.

OpenVINO toolkit provides highly optimized neural network computing capability, which is delivered by Intel® Deep Learning Deployment Toolkit (Intel® DLDT), its main inference optimization module. Intel DLDT contains two modules, Model Optimizer (MO) and Inference Engine (IE), which help to realize best execution capability on target terminal devices. In addition, OpenVINO toolkit provides extended channels to retrain and deploy models such as license plate recognition, face detection, and facial expression recognition. It also supports generating widely applicable classification network models based on Pytorch.

Paired with 2nd Gen Intel Xeon Scalable processor, OpenVINO toolkit is able to significantly improve deep learning inference capability. It also contains a large number of pre-trained models, which have been tuned and encapsulated by Intel to facilitate users' secondary development. In addition, users can call OpenVINO model converter to convert models into standard Intermediate Representation (IR) format and optimize them, greatly improving deep learning performance on the Intel-based platform.

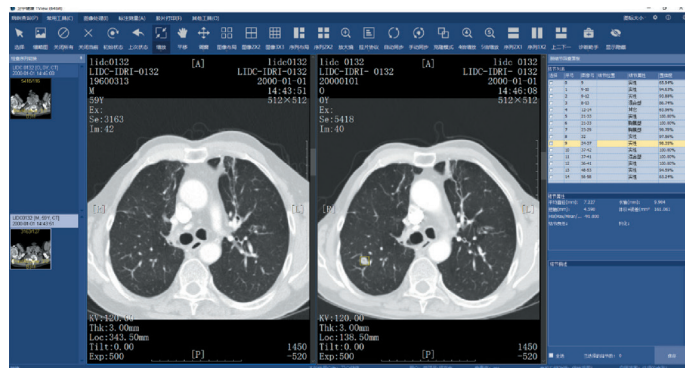


Figure 2: Winning Health Medical Imaging AI application - TView

The 2nd Gen Intel Xeon Scalable processor also significantly improves INT8-based inference performance. For accuracy conversion of pre-trained models, users can employ the accuracy calibration tool of OpenVINO toolkit to reduce the accuracy loss during the conversion from FP32/16 to INT8. This tool can also measure the model performance (throughput and latency), obtain the performance index of each layer and the whole model, reduce the accuracy loss, and ensure accurate, complete, and timely inference.

Verification: Inference speed for AI-aided diagnostic imaging of pulmonary nodules significantly boosted

Continuous efforts of experts and technical teams from the three parties have paid off. As a result of joint optimization and technological innovation in both software and hardware, Winning Health Medical Imaging AI solution based on AMAX Deep Learning All-In-One has achieved significant improvement in efficiency. Taking AI-aided diagnostic imaging of pulmonary nodules as an example, the overall model inference time has sharply shortened from more than 0.5h to less than 2m.

To verify the performance of this solution, Winning Health ran some tests using AMAX Deep Learning All-In-One, which is equipped with 2nd Gen Intel Xeon Scalable processor. Three configurations were tested and compared - Linux Pytorch Default: numa off, Linux Pytorch Optimized (numa mapped, 2 tasks, each assigning 36 sub-tasks to 36 logical cores), and OpenVINO toolkit (software and hardware configurations are shown in Table 1).

CPU	Intel Xeon Gold Processor 6240*2
Memory	12 slots * 32 GB , 2666MT/S
Storage	Intel SSD D3-S4510
Software Version	OpenVINO=2019.3.376, DNN library, Python=3.5, Pytorch=1.3.0, Linux versions=OS: CentOS Linux 8 (Core), Kernel: 5.6.4-1.el8.elrepo.x86_64

Table 1: Software and hardware configurations for the tests

Test data (as shown in Figure 3) shows that in the three tasks of segmentation, detection, and false-positive removal, OpenVINO toolkit increases the inference speed by 10-30 times.

In addition to AI image inference for pulmonary nodules, Winning Health Medical Imaging AI solution based on AMAX Deep Learning All-In-One has been successfully applied in other medical imaging fields such as bone age detection. This proves that the solution can help medical institutions boost inference speed and effectively control cost, without sacrificing the inference accuracy.

Outlook: Jointly build “AI + healthcare” ecosystem to empower smart healthcare

The cooperation between Winning Health, AMAX, and Intel has effectively improved the performance, cost effectiveness, and flexibility of medical imaging AI applications. For instance, Winning Health AI-aided diagnostic imaging system for pulmonary nodules has improved the model inference speed from more than 0.5h to less than 2 minutes, matching the total diagnostic time measured with the graphics card without sacrificing the inference accuracy. This also saves the cost of server procurement.

Intel is a leading provider of a wide range of products and technologies in deep learning and AI. To promote widespread application of AI in the medical and healthcare field, Intel is working with Winning Health, AMAX, and other partners to build a more dynamic “AI + healthcare” ecosystem, helping users apply medical imaging AI in different segments, widening the application of AI in the medical industry, and promoting healthcare development.

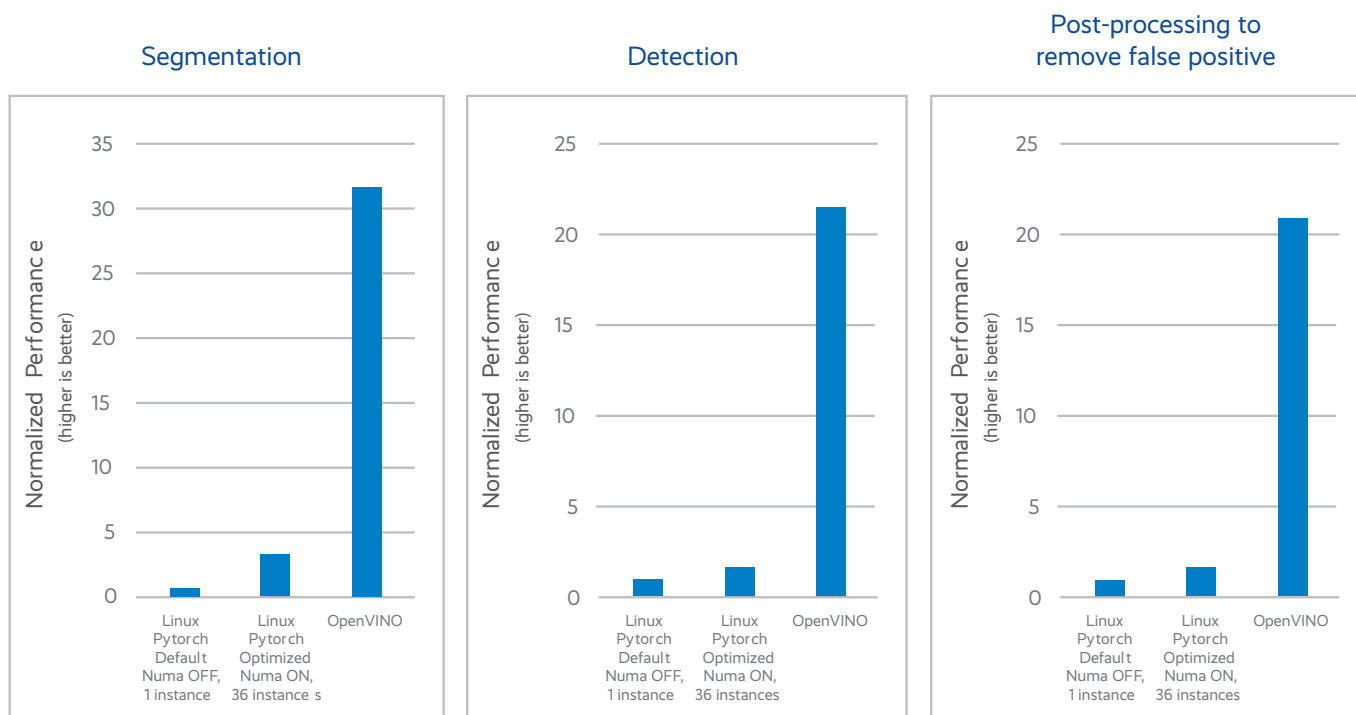


Figure 3: Performance comparison of 3 configurations in 3 test tasks

Note 1: Tested by Winning as of 4/17/2020. 2 socket Intel Xeon Gold 6240 Processor, 18 cores, HT ON, Total Memory 384 GB (12 slots/ 32 GB/ 2666 MHz), BIOS: SE5C620.86B.02.01.0010.010620200716 (ucode: 0x400002C), CentOS 8, 5.6.4-1.el8.elrepo.x86_64, Deep Learning Framework: Pytorch, Compiler: gcc 7.3, MKL DNN version: v0.20.5, precision: FP32, dataset: 357x4x3x96x512x512, Customized 3D Unet. Configuration 1: Linux Pytorch Default Numa OFF, 1 instance; Configuration 2: Linux Pytorch Optimized Numa ON, 36 instances; Configuration 3: OpenVINO.

Note 2: As more tests are conducted, the performance benchmark results may be modified. The results depend on the specific platform configuration and workload. The results do not necessarily represent other performance benchmarks, and other performance benchmark results may be more or less inhibited. Performance results are based on tests as of April 20, 2020 and may not represent all published security updates. Please check the configuration statement for details.

About Winning Health

Established in 1994, Winning Health specializes in information technology for medical and healthcare. They are committed to providing excellent products and services for medical and healthcare institutions, and continuously improving people's hospital experience and health conditions. Learn more about [Winning Health](#).

About Intel

Intel Makes Possible the Most Amazing Experiences of the Future. You may know us for our processors. But we do so much more. Intel invents at the boundaries of technology to make amazing experiences possible for business and society, and for every person on Earth.

Harnessing the capability of the cloud, the ubiquity of the Internet of Things, the latest advances in memory and programmable solutions, and the promise of always-on 5G connectivity, Intel is disrupting industries and solving global challenges. Leading on policy, diversity, inclusion, education and sustainability, we create value for our stockholders, customers, and society.

About AMAX

AMAX is an award-winning global leader in data center, open-architecture platforms, HPC, Deep Learning, application-tailored cloud and OEM server manufacturing solutions designed towards the highest efficiency and optimal performance. Whether you are a Fortune 1000 company seeking significant cost savings through better efficiency for your global data centers or a software startup seeking an experienced manufacturing partner to design and launch your flagship product, AMAX is your trusted solutions provider, delivering the results you need to meet your specific metrics for success. For more information about AMAX, visit <http://www.amax.com>.