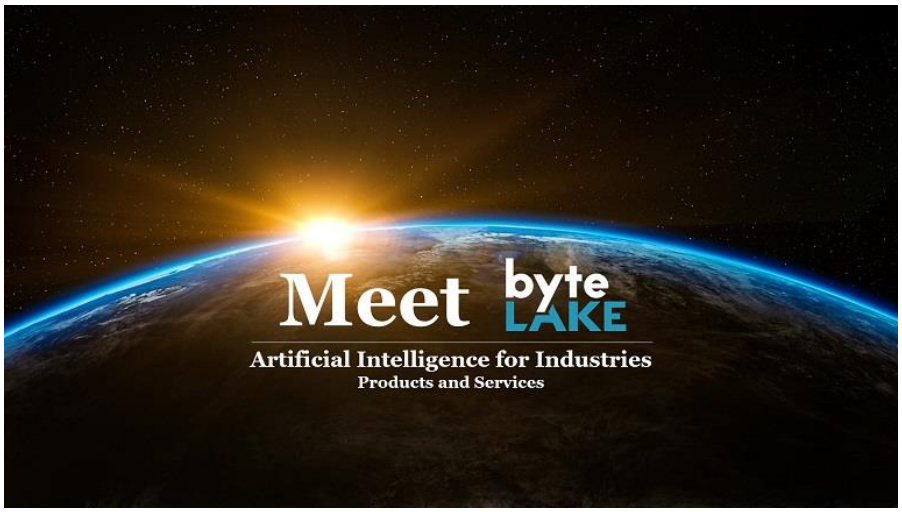


Cognitive Services accelerated with OpenVINO

AI FOR INDUSTRY 4.0

byteLAKE and Intel work together to increase performance of data analytics and AI-assisted visual inspection in Industry 4.0 scenarios. Document presents benchmarks for byteLAKE's Cognitive Services before and after optimizations with OpenVINO.

Artificial Intelligence

Machine Learning

Deep Learning

Computer Vision

Edge AI

Cognitive Automation

RPA

HPC

FPGA / GPU

byte LAKE

byteLAKE

Europe & USA

+48 508 091 885
+48 505 322 282
+1 650 735 2063

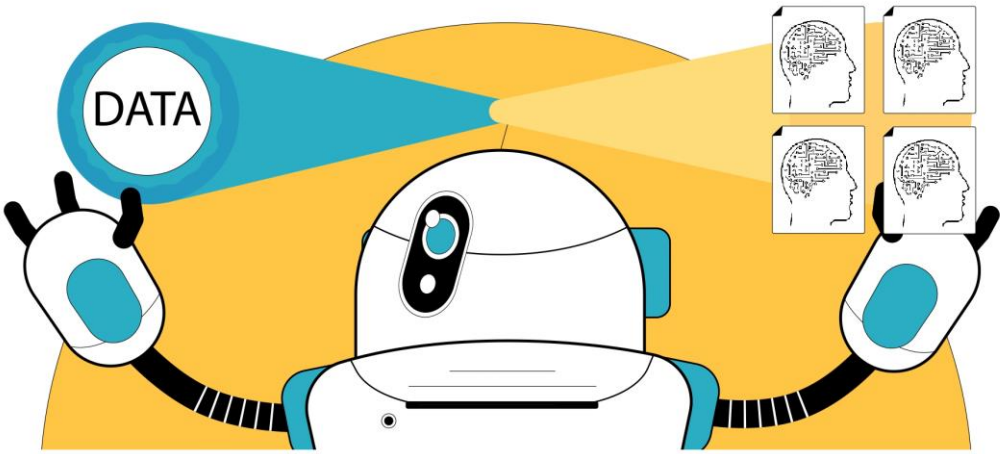
byteLAKE's Cognitive Services: Artificial Intelligence for Industry 4.0

byteLAKE's Cognitive Services is a collection of Artificial Intelligence (AI) models designed to address Industry 4.0 needs. Each AI model has been designed and trained to be razor-focused on specific industrial jobs, therefore ensuring maximum accuracy.

byteLAKE have made a strategic decision to work with various industry leaders and talented researchers in efforts to combine human knowledge, industry expertise, and know-how with the best AI algorithms and technologies. Humans and machines both make mistakes. Therefore, byteLAKE's Cognitive Services have been designed to effectively bring out the best of both worlds.

In essence, byteLAKE's Cognitive Services focus primarily on the following two areas:

- **AI-assisted visual inspection** for efficient quality of products & process monitoring.
- **AI-powered Big Data / IoT sensors data analytics** to find trends, enable predictive maintenance, answer questions like why something happens, what will likely happen and to find the collective meaning of the data extracted from many sources.



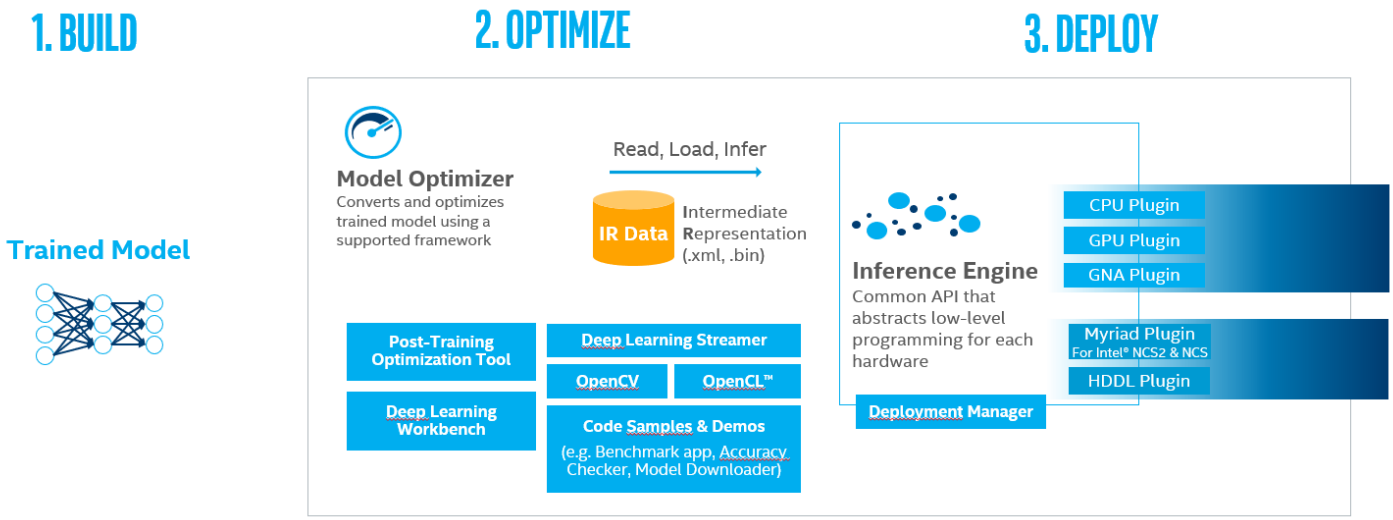
Intel® Distribution of OpenVINO™ toolkit

A toolkit for developing applications and solutions that use deep learning. The toolkit extends workloads across Intel® hardware (including accelerators) and maximizes performance.

- Enables deep learning inference from edge to cloud
- Accelerates AI workloads, including computer vision, audio, speech, language, and recommendation systems
- Supports heterogeneous execution across Intel® architecture and AI accelerators—CPU, iGPU, Intel® Movidius™ Vision Processing Unit (VPU) and Intel® Gaussian & Neural Accelerator (Intel® GNA)—using a common API
- Speeds up time to market via a library of functions and preoptimized kernels
- Includes optimized calls for OpenCV, OpenCL™ kernels, and other industry tools and libraries.

Learn more: https://software.intel.com/content/www/us/en/develop/tools/openvino-toolkit.html

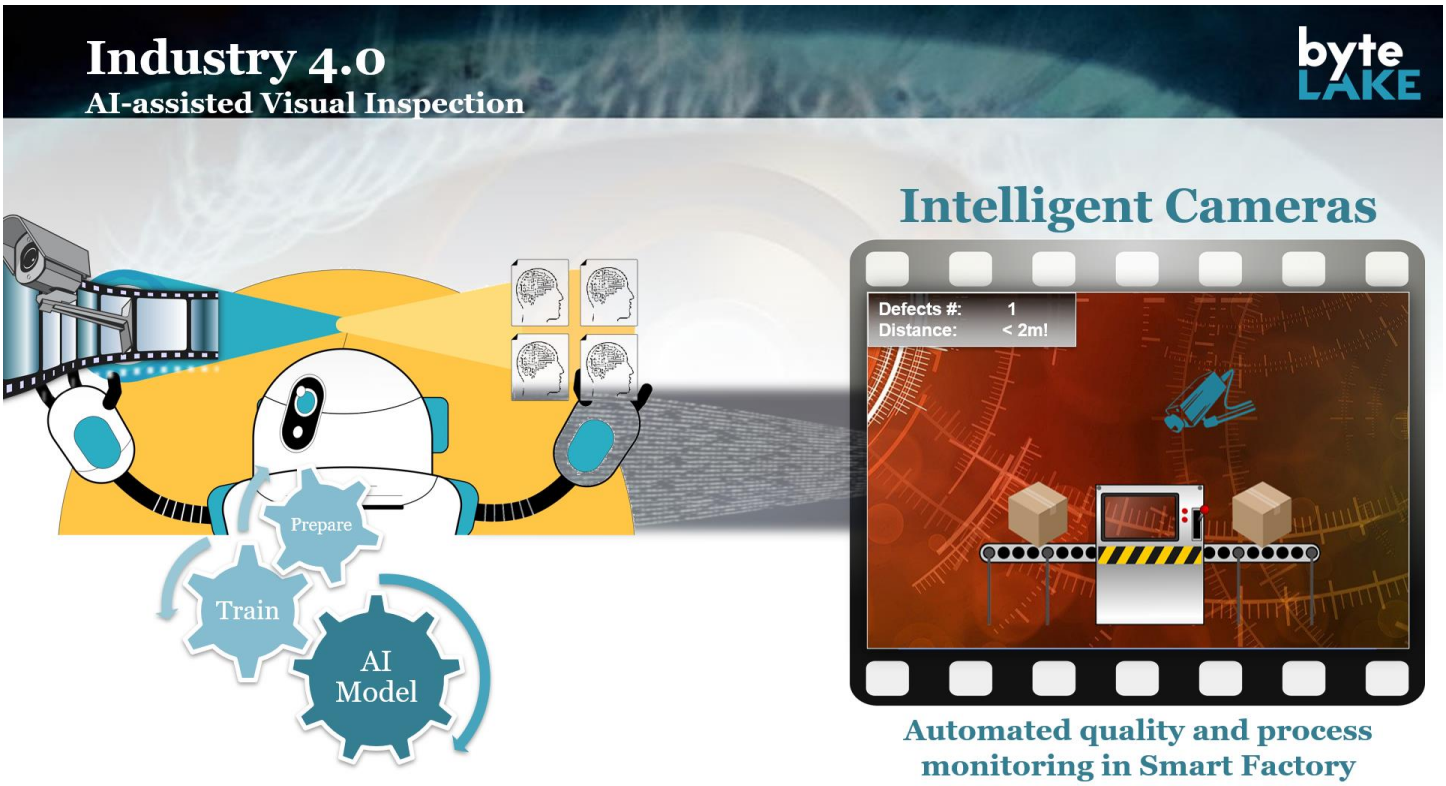
You might also want to read another byteLAKE's report about how we leveraged OpenVINO to optimize byteLAKE's Cognitive Services, our AI product for Industry 4.0 (AI-assisted Visual Inspection, AI-powered Big Data Analytics). Find out more at: www.byteLAKE.com/en/CognitiveServices.



AI-assisted visual inspection in Industry 4.0 – performance is important

Artificial Intelligence (AI) brings value across industries, giving machines the abilities once reserved for humans. On one hand, its fast adoption is driven by a constantly growing number of software frameworks, availability of a specialized hardware, and Big Data. On the other hand, effective human-robot collaboration translates into increased efficiency, decreased number of defects, routine jobs automation, faster results, and numerous costs optimizations.

AI-assisted visual inspection is a scenario where AI is leveraged to transform computers into intelligent machines to identify objects, analyze scenes and activities in real-life visual environments. For instance, a camera is used to take pictures of products or production lines and AI algorithms provide analysis related to products quality, quantity or even analyze and help monitor production processes.



Performance of such systems is critical. AI-assisted visual inspection is as useful as its timely and accurate responses. Therefore, the underlying software, or the AI algorithms to be precise, need to be optimized and generate results without latencies or even in real-time. These days the designers have access to plethora of optimization techniques to deliver optimal results, including programming tips and tricks, well documented lessons learned, software frameworks and hardware options.

However, once every reasonable improvement has been incorporated, can there be any way to improve the performance even further and at the relatively low cost? To answer that question, byteLAKE and Intel worked together to benchmark how byteLAKE's Cognitive Services performance could possibly be improved through its integration with Intel's OpenVINO.

AI for Paper Industry, benchmarked use case

Paper production is a multi-phase process during which a natural phenomenon called Wet Line (sometimes Dry Line) is observed. AI-assisted visual inspection of the process can help efficiently detect and monitor that phenomenon.

byteLAKE's Cognitive Services includes a dedicated AI model which has been designed and trained specifically for this task ("Wet Line Detector"). It can work in real-time and continuously analyze frames received from industrial cameras. AI algorithms inspect the surface where the fabric is formed and detect the so-called Wet Line. In addition, the algorithms measure and estimate the position of the Wet Line as well as its width. This information is then presented to the paper machine operator who can react accordingly and i.e., apply required settings. More about the solution can be found at: https://www.bytelake.com/en/ai-for-paper-industry/.

Benchmark results

byteLAKE's "Wet Line Detector" (part of byteLAKE's Cognitive Services) has been used for the benchmarking purposes and the results are presented below. Based on these, byteLAKE's Cognitive Services has been updated accordingly.

General configuration of the byteLAKE's Cognitive Services

- "Wet Line Detector" trained to detect Wet Line and provide its measurements
- Software highlights: DarkNet C++ OpenMP/CUDA framework, YOLO, Python
- Supported hardware infrastructure: cross-platform
- Optimization: OpenVINO YOLO/Python
- Neural network / Wet Line: 23 CNN layers, 5 pooling layers, single reorg layer, single region layer
- Hardware used for inferencing (Edge AI):
    - CPU: Intel(R) Core(TM) i5-8500T CPU @ 2.10GHz
    - GPU: NVIDIA Quadro P1000

Inferencing sequence (benchmarking procedure):

1. Load 100 images to RAM
2. *<starting the timer>*
3. Inferencing (all cores used): images analysis, detecting a wet line, drawing the results in RAM
4. *<stopping the timer>*
5. Saving the results

I/O related latencies have been eliminated and only the time of inferencing has been measured. Below results are a median for 5 consecutive trials for each configuration.

CPU inferencing results **without OpenVINO** (baseline):

Detection time for all images [seconds]: 75.0

Average analysis time per image [seconds]: 0.75

FPS: 1.33

CPU inferencing results **with OpenVINO**:

Detection time for all images [seconds]: 7.99

Average analysis time per image [seconds]: 0.08

FPS: 12.52

ACCELERATION: 9.4x

Integration with OpenVINO let us significantly increase the performance of the images' analysis and Wet Line detection / measurements by 9.4 times. It must be noted that the acceleration is the result of software optimizations only and the hardware has not been changed.

Furthermore, OpenVINO allows us to change the data types from FP32 to FP16. In our configuration, however, this did not have any impact on performance results. It reduced the size of the model though (by 50%). Explanation: "As CPU now supports FP16 (while internally upscaling to FP32 anyway) and because this is the best precision for a GPU target, you may want to always convert models to FP16." (source).

Further study is required to explore the potential benefits with changing the data types. In the past we saw performance benefits while changing the precision of the arithmetic (case studies).

For the comparison purposes, we also ran the baseline version on GPU.

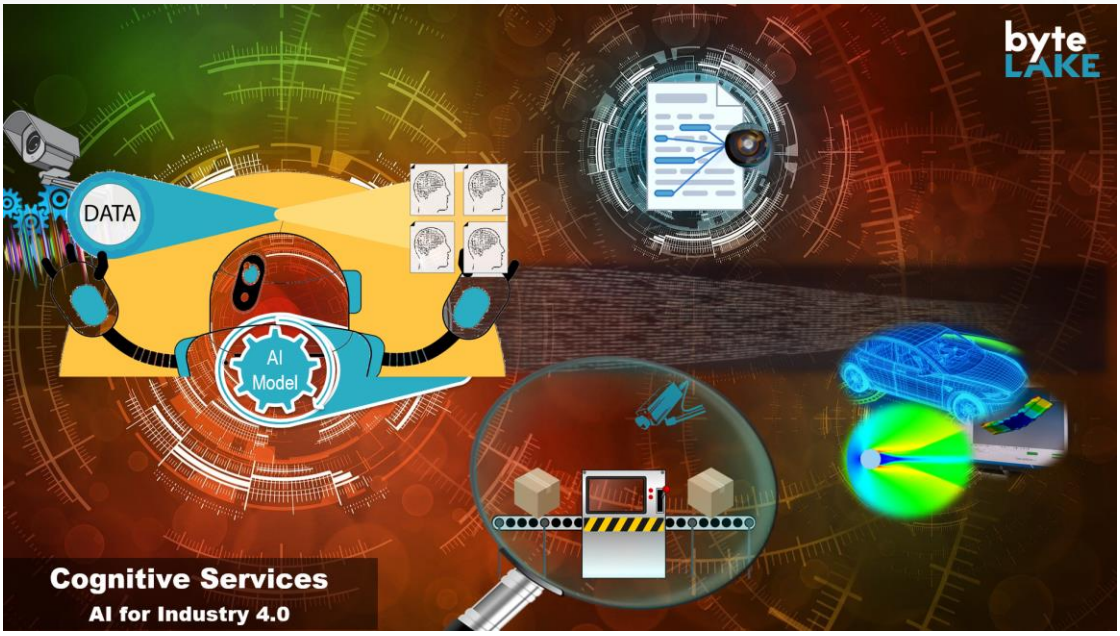
GPU inferencing with CUDNN:

Detection time for all images [seconds]: 8.07

Average analysis time per image [seconds]: 0.08

FPS: 12.39

As one can see, GPU led us to almost as good results as those achieved with OpenVINO. Therefore, from performance perspective we can say that in our example tandem CPU/OpenVINO was able to generate comparable results to GPU/CUDA.



Conclusions

Integration with OpenVINO is a straightforward process and the effort to do so is relatively small. The results, however, are stunning. We have not observed any decrease in accuracy. When it comes to the performance, as presented above, **byteLAKE's Cognitive Services gained almost 10x improvement without any hardware upgrades**. Then depending on accuracy needed and case by case this number can soar up, leading us to faster inferencing and in turn generating timely results.

"We have been using various Intel technologies at byteLAKE for many years and always been impressed by their performance and scalability offered. When designing the Cognitive Services for Industry 4.0, one of the key elements on our roadmap was to ensure maximum performance while leveraging our clients' existing infrastructure, which in many cases is based on Intel processors. Today I am excited to announce that having Intel's OpenVINO integrated into byteLAKE's products for Industry 4.0 including AI-assisted visual inspection, we not only meet the highest performance requirements but also ensure the availability of our product across various hardware configurations", said Marcin Rojek, co-founder of byteLAKE.

Key takeaways

- **byteLAKE's Cognitive Services automate visual inspection and Big Data processing across industries.** Each AI model has been designed and trained to be razor-focused on specific industrial jobs, therefore ensuring maximum accuracy.
- **Can be re-trained to handle a variety of scenarios** related to visual inspection/quality monitoring automation, products counting, objects recognition and historic data analysis to find hidden answers in the data (i.e. trends, information about why something happened or what will likely happen and when).
- **New AI models are constantly added by byteLAKE** which gradually increases the number of scenarios that can be handled off-the-shelf. To do so byteLAKE collaborates with a growing number of industry/manufacturing leaders.
- **Cognitive Services is an add-on to existing tools/software in factories and its integration is a straightforward process (compatibility).**
- **byteLAKE as single source for all components** of the solution (sensors/cameras, edge devices, servers, data acquisition/processing, deployment, post-delivery customer care etc.)
- **Globally available through growing network of integrators.**
- **Optimized for Intel technologies** (OpenVINO) ensuring compatibility and maximum performance across various hardware configurations.

- Link to recording: https://youtu.be/skM77hdPCjw
- Learn more: www.byteLAKE.com/en/CognitiveServices
- Contact us: CognitiveServices@byteLAKE.com

byteLAKE

Artificial Intelligence for Industries. Products and Services.

www.byteLAKE.com

About byteLAKE

byteLAKE is a bespoke AI & HPC software company developing AI-powered solutions for enterprises. The company offers both products and services, enabling innovative, AI-powered automation and data-driven, proactive operations across various industries i.e. AI-assisted Visual Inspection and Big Data analytics for manufacturing, AI-accelerated Computational Fluid Dynamics, AI for Industry 4.0, workflow and document processing automation etc. To learn more about byteLAKE's innovations, go to byteLAKE.com.