intel®

Vismarty Accelerates Detection and Extraction of Regulated Imagery with Intel® AI Technologies

Intel optimizations of Vismarty IDFS algorithms running on Intel Arria® 10 FPGAs and Intel Distribution of OpenVINO™ toolkit boost performance by 3X[1]

INTEL®
AI BUILDERS
MEMBER

VISMARTY 睿视智觉

Government regulations that restrict Internet-distributed imagery to help protect citizens require content providers to carefully monitor their data to stay compliant. This is clearly a challenge, considering how quickly and easily creators can produce and publish disallowed content and how fast they can distribute it through video, online sales, and social platforms.

Many countries restrict explicit and disturbing imagery that promotes illegal and unacceptable activities. With hundreds of millions of people contributing to social networks and self-publishing platforms, service providers cannot manually keep up with monitoring and extracting disallowed content. They depend on advanced technologies that offer effective and fast detection and prevention of illegal images from distribution.

Vismarty Technology Co. Ltd. combines their artificial intelligence (AI) algorithms and FPGA optimizations with Intel® technologies to accelerate detection and blocking of disallowed imagery across multiple distribution technologies in real time. Service providers can now effectively and efficiently monitor and prevent unwanted content on their networks.

Automate and Accelerate Image Compliance in Real Time

The development of advanced mobile communication technologies and Internet applications allows nearly instant distribution of images and video through streaming platforms, online sales services, and social networks. The heterogeneous nature of the data makes it difficult for service providers to remain compliant with regulations using manual methods of content monitoring and extraction.

With the adoption of AI technologies, computer vision is gradually replacing manual compliance reviews. However, given the massive volumes of image data, mainstream system solutions in the market today are challenged to fully unleash system intelligence due to bottlenecks in computing power.

Vismarty, an AI solution developer, focuses on computer vision algorithms and FPGA acceleration technology, delivers high-performance image analytics services that leverage computer vision algorithms powered by FPGA technology. Relying on self-training systems and acceleration architectures, Vismarty delivers a host of services and solutions deployable in the cloud and on-premise to enterprises and public entities.

The Vismarty image detection and filtering system (IDFS) meets market demands for real-time response requirements with simultaneous processing of multiple sources. Vismarty IDFS allows enterprises and regulatory entities to monitor compliance of content published through live broadcast, short video, social networking, online sales, and other media platforms. Vismarty IDFS is used in a host of applications, including:

- **Live video –** Live video has many concurrent broadcast streams. Vismarty IDFS performance and scalability enable it to monitor the content of all video feeds, identifying suspicious imagery and blocking it in real time.

- **Online Shopping –** A wide range of suspicious content is distributed through online malls. Vismarty IDFS monitors and detects uploaded images, including non-compliant or illegal content, such as disturbing pictures, images promoting terrorism, and other non-compliant content.

- **Website BBS –** User Generated Content (UGC) provides a large volume of Internet imagery. Vismarty IDFS identifies and blocks suspicious and non-compliant images to help customers reduce the risk of business violations.

Vismarty IDFS can be deployed easily across cloud services, such as AWS, Lenovo, KS cloud, TE, HUPU, and others, or as an on-premise solution.

526 Frames Per Second: Accelerating Image Detection by Over 3X[1,2]

While GPUs can be used for computer vision applications, Vismarty chose to optimize their algorithms on Intel architecture using Intel AI technologies. Optimizing their IDFS solution for Intel Programmable Acceleration Cards (PAC) with Intel Arria® 10 GX FPGAs, Vismarty was able to achieve 526 frames per second throughput with low latency on image detection, which is 322% faster performance than running on a GPU (see Figure 1).[1]

The key benefits of the Vismarty IDFS solution running on Intel architecture include:

- **Low latency –** processing time for a single image is less than 2 milliseconds.

- **High throughput –** the Intel PAC with Intel Arria 10 GX FPGA processes more than 3.5 million pictures per hour (with the Intel Stratix® 10 Programmable Acceleration Card, throughput reached 10 million pictures per hour).

- **High power efficiency –** the solution processes 50,000 pictures per watt (for FPGA only).

- **High accuracy –** greater than 99 percent recognition rate.

- **Flexibilit –** easily supports different image recognition applications.

- **Easy deployment –** an API interface between the customer's sources and the solution enables fast deployment.

With high performance and accuracy at low power, the solution helps customers to reduce investment in hardware infrastructure and lower operating costs in order to meet regulatory compliance requirements.
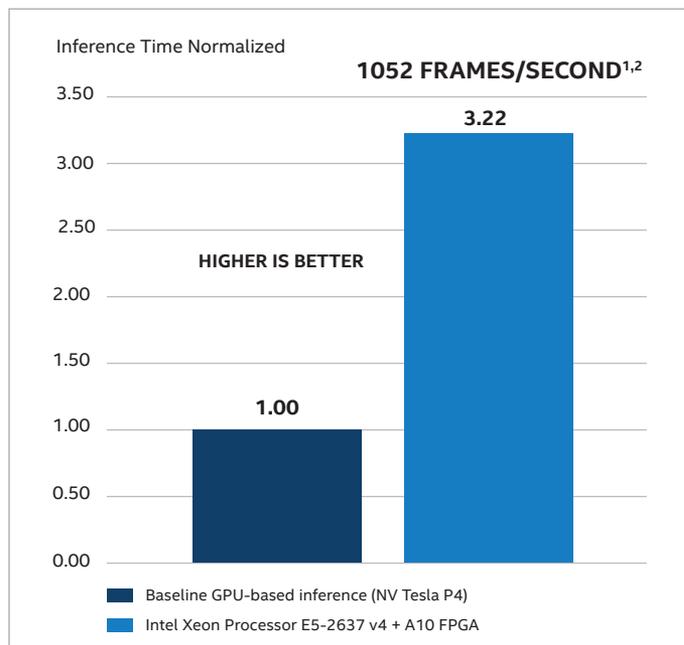

Figure 1. Optimizations for Vizmarty's IDFS solution on Intel PACs achieve 3.22X faster performance vs GPUs.[1]

The Vismarty IDFS solution was developed on the TensorFlow framework for image detection with the Intel Distribution of OpenVINO™ toolkit 2018 R4 software release for computer vision using the Mobilenet v2 topology.

Intel AI technologies and accelerating hardware allowed Vismarty to optimize their algorithms for greater image processing performance.

- Intel Optimizations for TensorFlow framework include Intel Deep Neural Network Library (Intel DNNL) optimizations that allow Vismarty algorithms to take advantage of advanced capabilities in the architecture of the Intel PAC with Intel Arria 10 GX FPGA.

- Intel Distribution for OpenVINO toolkit provides deep learning and computer vision optimizations that help accelerate image inferencing on Intel architecture, including CPU, iGPU, VPU and FPGA, and accelerators with VPU and FPGA.

The Vismarty IDFS solution uses the Intel PAC with Intel Arria 10 GX FPGA for neural network compute along with their IDFS software development kit (SDK) to provide an API service connector to the customer based on Vismarty's specific neural network models for the target applications. With the API connector, the IDFS solution can be directly deployed into datacenters with the Intel PAC with Intel Arria 10 GX FPGA card installed, easily integrating it into an existing image and video data processing workflow.

Conclusion

The Vismarty IDFS solution, optimized for the Intel PAC with Intel Arria 10 GX FPGA, accelerates image detection and extraction for specific types of content disallowed by regulating bodies. The Vismarty algorithms running on Intel hardware with Intel AI technology optimizations increased image throughput to 526 fps—322 percent faster[1,2] than running the same algorithms on a GPU—with high efficiency and accuracy.

The IDFS solution can be deployed across cloud infrastructures or on-premise in an existing datacenter. A Vismarty IDFS API allows fast integration into an existing IT environment along with Intel PAC with Intel Arria 10 GX FPGA accelerators.

With high performance and accuracy at low power, the solution helps customers to reduce investment in hardware infrastructure and lower operating costs in order to meet regulatory compliance requirements.

For more information about Vismarty, visit **http://www.vismarty.com/**.

Learn more about the Intel AI Builders program at **https://builders.intel.com/ai/membership**.

INTEL® AI BUILDERS MEMBER

Vismarty is a member of the **Intel AI Builders Program**, an ecosystem of industry-leading independent software vendors (ISVs), system integrators (SIs), original equipment manufacturers (OEMs), and enterprise end users, which have a shared mission to accelerate the adoption of artificial intelligence across Intel platforms.

---

1. Baseline: 164fps @ 6.116ms per image running Mobilenet v2 model on NV Tesla P4 GPU

   Optimized: 526fps @1.902ms per image running on Intel Xeon processor

   Testing conducted on Intel® Xeon® E5-2637 v4 processor with Intel PAC with Intel Arria 10 GX FPGA accelerator comparing baseline config with NV Tesla P4 GPU (Batch size:1)on the same trained dataset.

   Configuration : Intel® Xeon® CPU E5-2637 v4 @ 3.50GHz, 8 physical cores, 112 GB physical memory; Network Mobilenet-v2 | Network Type CNN | Batch Size 2| Precision: Vismarty's Lossless INT| Dataset Synthetic

   GPU benchmark: Performed in Huawei Cloud with following instance: GPU Inference Acceleration Pi1

   GPU : Nvidia P4 8GB GDDR5

   SW env: CentOS 7.4 TensorFlow 2.0+CUDA 10.0+cuDNN7.5

2. 526 FPS was tested using one kernel in the Intel FPGA. Vismarty can run two kernels per Intel FPGA card, delivering up to double the tested performance, while maintaining 1.928 ms of latency.

Performance results are based on testing as of dates shown in configuration and may not reflect all publicly available security updates. No product or component can be absolutely secure. See configuration disclosure for details.

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. 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: http://www.intel.com/performance