Solution Brief

Intel® Network Builders
Multi-access Edge Computing
Retail

Brick-and-Mortar Customer Analytics at the Intelligent Edge

Kibernetika Audience Analytics applies deep learning models to video streams from cameras placed in retail stores to analyze customer foot traffic and optimize store operations. The solution is powered by ingredients that include Intel® Smart Edge Open and the Intel® Distribution of OpenVINO™ toolkit, running on Intel® Xeon® Scalable processors and Intel® Movidius™ Myriad™ X vision processing units (VPUs).

KIBERNETIKA.AI

The retail industry's transformation in the past two decades is far more than a story of e commerce out-maneuvering brick-and-mortar locations with the inherent advantages of greater convenience and virtual storefronts that can reach anywhere. Despite the ongoing transition, brick-and-mortar retail continues to dwarf e commerce in terms of revenue, although its dominance is gradually shrinking. As of the second quarter of 2021, the U.S. Department of Commerce reports that e-commerce represents just 13.3 percent of total sales by the retail industry as a whole, but with that percentage growing as a long-term trend.[1]

Going forward, an important strategy for brick-and-mortar retail is to take advantage of the same types of rich customer data that online retailers use to assess customer behavior while shopping. Similar to approaches taken by e-commerce companies, operators of physical stores can use this information to help comprehensively understand customer presence and fine-tune the shopping experience to optimize sales volume and profit. Whereas the means to gather that information are built into the clickstream for online retail, specialized measures are required for brick-and-mortar stores.

Deep Learning-Based Shopper Observation and Insight

The Kibernetika Audience Analytics solution helps retailers capture and describe data about shoppers and their behavior, then applies deep-learning algorithms to that data to generate actionable insights. The solution primarily uses data from in-store video feeds and point-of-sale terminal transactions. Analytics generate a variety of visualizations that assist with interpretation, including heatmaps as shown in Figure 1, which show how much time customers spend at various physical locations within a store.

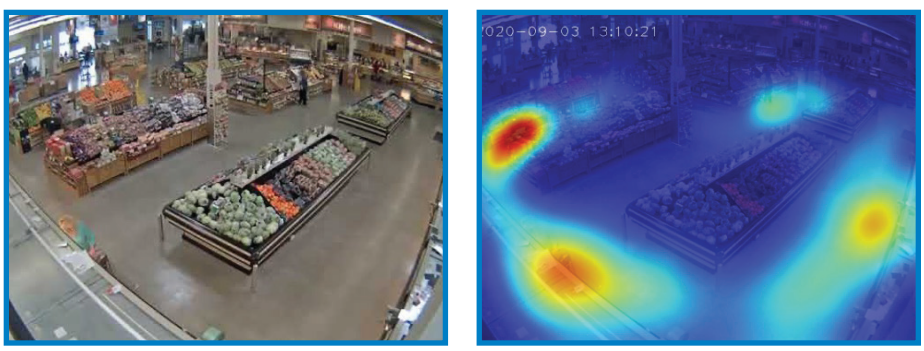

Figure 1. Video images (left) transformed into a shopper-presence heatmap (right).

Heatmaps reveal the amount of attention that customers pay to particular in-store displays, which can be used to gauge the effectiveness of product placement and optimize overall store layouts. Improving in-store design can direct customers to the products that best support sales goals and fine-tune promotions, helping increase targeted linger time and ultimately boost revenue. The solution also generates tracking maps, another type of data visualization that works alongside heatmaps by showing the paths of foot traffic through a store.

Adding further to sophisticated behavioral understanding, the Kibernetika solution provides demographic details such as the distribution of customers according to age and gender, with the variations in behavior among them. This analysis can reveal patterns of attention by different types of people as they move through the store. Based on that information, the solution can help retailers target sub-groups of shoppers with products and promotions tailored to each audience.

Kibernetika Audience Analytics can serve many other functions as well, such as keeping a running total number of customers in the store, watching the length of checkout lines to manage customer wait time, and monitoring stock levels of items on the shelves. Visual intelligence generated by the solution can even help detect shoplifting in progress.

Analytics on Multi-access Edge Computing (MEC) Infrastructure

The solution performs cloud-based analytics at the network edge—either in-store or nearby—so video stream data does not need to be transmitted over long distances. This critical bandwidth efficiency increases overall scalability and cost-effectiveness, especially for large retail chains with many locations. Edge computing also allows the solution to reduce latency, enabling real-time or near-real-time usages. To maximize flexibility, Kibernetika Audience Analytics can be deployed using any combination of cloud infrastructure, including customer-owned on-premise environments, MEC platforms provided by service providers, and public clouds.

The solution architecture is illustrated at a high level in Figure 2. The Audience Analytics service consists of three pods: the Audience UI, API, and database. The Audience UI allows human operators to interact with the environment, including control of stream configurations, which can include sources such as cameras, local or internet-based files, and manual streaming using FFmpeg. The Audience API provides programmatic control plane and data plane access, while the Audience database (together with attached storage) holds both raw video and analytics outcomes as needed.

The Heatmap Streaming service reads incoming streams as specified by the Audience Analytics configuration. It works with the solution's Heatmap logic to provide analytics and visualizations based on those inputs, assisted by the machine learning engine. The Heatmap Streaming service periodically sends picture samples, heatmap labeling information, and video chunks back to the Audience Analytics service by means of the Audience API. Various elements of this solution architecture can be scaled independently to meet the needs of specific functionalities and implementations.

Intel Enablement with Hardware and Software Building Blocks

The value of heatmaps in audience analytics is directly proportional to heatmap density, which corresponds to the amount of audience data that can be represented in a given heatmap. Specifically, the heatmap provides analysts with a richer view of audience behavior when the system can render frames more quickly on a larger number of simultaneous video streams in real time. Figure 3 illustrates the increased heatmap density that is possible using the 3rd Gen Intel® Xeon® processor compared to its predecessor.[2]

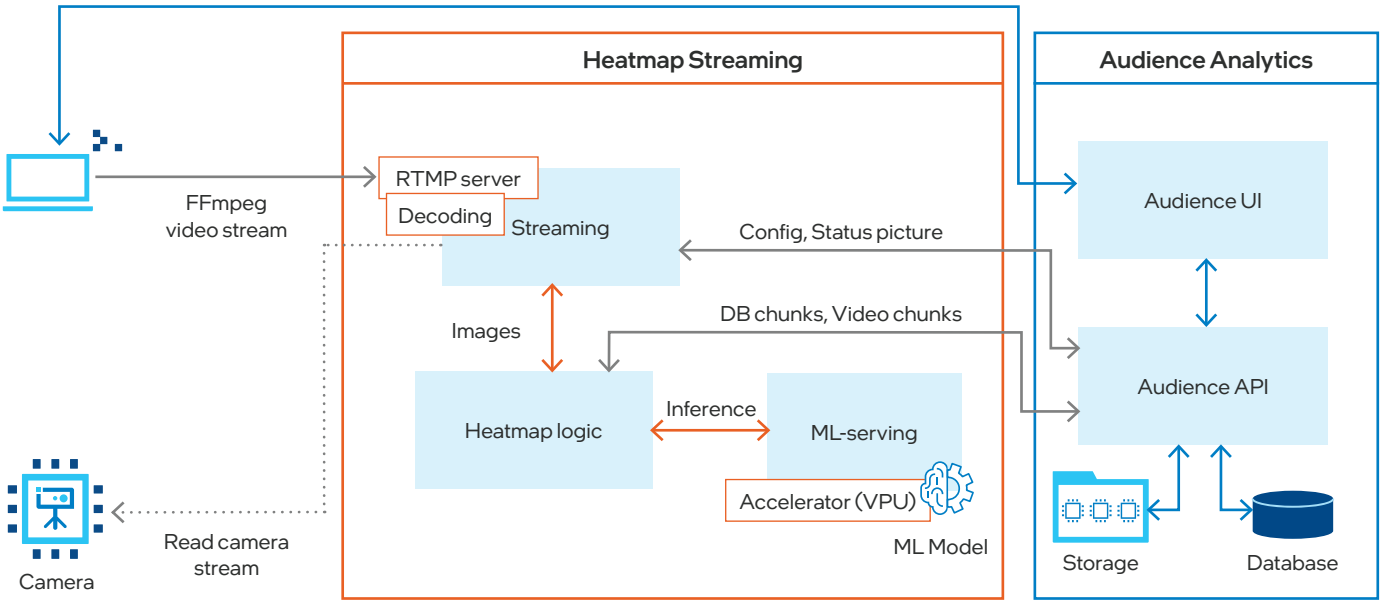

Figure 2. Kibernetika Audience Analytics solution architecture.

This data shows that the 3rd Gen Intel Xeon Scalable processor can render a given number of frames in approximately 80 to 85 percent of the time required by the 2nd Gen Intel Xeon Scalable processor. That increased performance is relatively constant as the number of streams being processed in real time increases from 40 to 320, demonstrating the scalability of the benefit as heatmap implementations use more cameras to provide more comprehensive data or coverage over a larger area.

The Kibernetika Audience Analytics solution is built using open source components and optimized to run on Intel architecture. That optimization enables the software to deliver performance that scales smoothly to platforms with varying levels of compute power for different use cases and implementation sizes. Key Intel technologies that are applicable to the solution include the following:

- **Intel® Smart Edge Open** is a royalty-free edge computing software toolkit that enables highly optimized and performant edge platforms to on-board and manage applications and network functions with cloud-like agility across any type of network. It provides software primitives that enable the solution to deliver high performance and scalability across the full range of Intel platforms. Toolkit capabilities used by the Kibernetika solution include enhanced platform awareness (EPA), high-density deep learning (HDDL) card resource allocation, Intel® Visual Compute Acceleration (Intel® VCA) card orchestration, and the Consumer-Producer API.
- **Intel® Distribution of OpenVINO™ toolkit** streamlines the process of building, optimizing, and running deep-learning inference models based on convolutional neural networks (CNNs) at the edge, including a library of pre-optimized kernels and computer functions. The toolkit helps accelerate inference performed by the Kibernetika solution, enabling sophisticated analysis.
- **Intel® Movidius™ Myriad™ X vision processing unit (VPU)** accelerates machine learning-driven visual workloads using a dedicated, on-chip Neural Compute Engine. The VPU processing pipeline is built to be programmed using the Intel Distribution of OpenVINO toolkit. The Intel Movidius Neural Compute SDK streamlines creation and validation of the high-quality deep-learning models used by the Kibernetika solution and performs image analysis.
- **Intel® Xeon® Scalable processors** offer flexibility for edge workloads, with options at a variety of feature levels and core counts. They provide high per-core performance and accelerate deep learning functions using Intel® Deep Learning Boost with Vector Neural Network Instructions (VNNI). The platform's balanced architecture is complemented by Intel components such as accelerators, Ethernet adapters, and persistent memory.

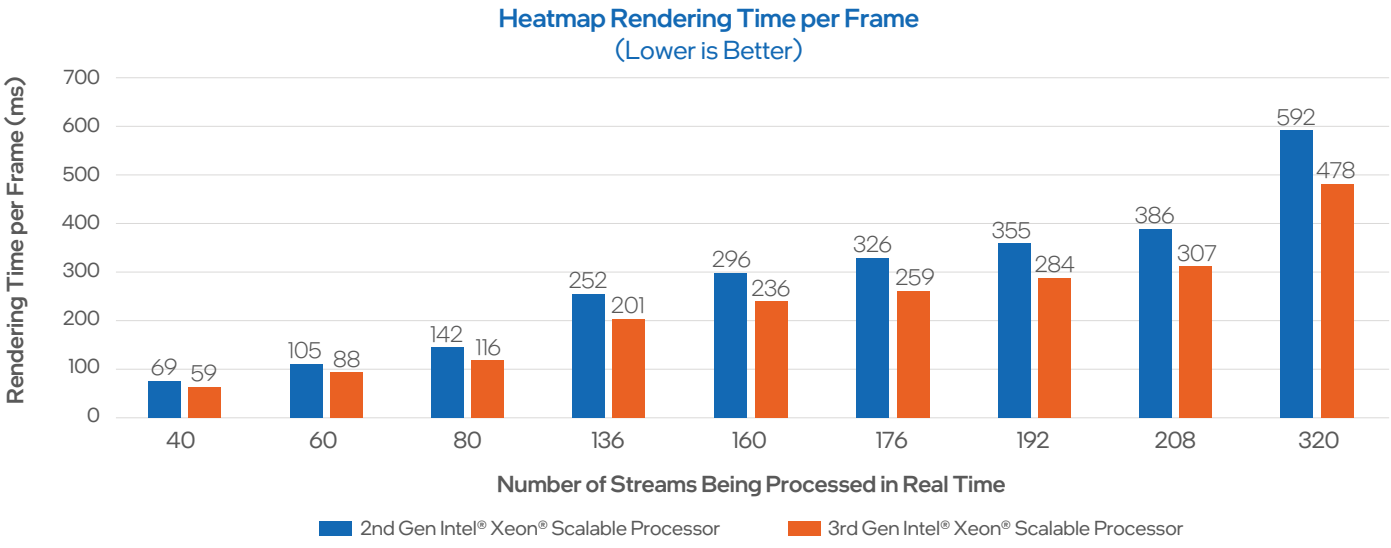

Figure 3. Increased heatmap density with 3rd Gen Intel® Xeon® processors.[2]

Conclusion

Kibernetika is moving the retail industry beyond digitization with visual intelligence that helps understand customer mindsets, motivations, and buying behaviors. Edge-based audience analytics are a critical enabler on the path to giving brick-and-mortar retail businesses parity of customer insight with the e-commerce segment, and the 3rd Gen Intel Xeon Scalable processor helps retailers advance that insight even further compared to predecessors.[2] A deeper understanding of how customers interact with the store environment can help these retailers achieve ROI from the solution by influencing those interactions, improving the customer experience, and increasing sales.

More Information

Intel® Network Builders: networkbuilders.intel.com

Intel Distribution of OpenVINO Toolkit: software.intel.com/content/www/us/en/develop/tools/openvino-toolkit.html

Kibernetika and Intel AI Builders: builders.intel.com/ai/membership/kibernetika

Kibernetika.AI Platform: kibernetika.ai/product

Solution provided by:

intel® + KIBERNETIKA.AI

[1] U.S. Department of Commerce, U.S. Census Bureau News. "Quarterly Retail E-commerce Sales 3rd Quarter 2021." Washington, D.C.: (2021). https://www.census.gov/retail/mrts/www/data/pdf/ec_current.pdf (PDF). Retrieved October 12, 2021.

[2] Testing completed by Kibernetika September 15, 2021.

OLD SYSTEM: 2x Intel® Xeon® Platinum 8260M processor (24 cores/48 threads, 2.4 GHz); Intel® Hyper-Threading Technology enabled; Intel® Turbo Boost Technology disabled; 384 GB RAM; CentOS 7.9, Docker 20.10.2; kuberlab/realtimeheatmap:1.0.4.

NEW SYSTEM: 2x Intel® Xeon® Platinum 6330N processor (28 cores/56 threads, 2.3 GHz); Intel® Hyper-Threading Technology enabled; Intel® Turbo Boost Technology disabled; 384 GB RAM; CentOS 7.9, Docker 20.10.2; kuberlab/realtimeheatmap:1.0.4.

Performance varies by use, configuration, and other factors. Learn more at https://www.intel.com/PerformanceIndex.

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

Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.

Your costs and results may vary.

Intel technologies may require enabled hardware, software, or service activation.

You may not use or facilitate the use of this document in connection with any infringement or other legal analysis concerning Intel products described herein. You agree to grant Intel a nonexclusive, royalty-free license to any patent claim thereafter drafted which includes subject matter disclosed herein.

The products described may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

1221/DO/MESH/346432-001US