Case Study

Intel® Network Builders
Telecommunications

Artificial Intelligence Deep Learning-Based Video Analytics at the 5G Network Edge

DeepSight AI Labs worked in a cross-industry collaboration to deploy real-time artificial intelligence deep learning-based video analytics at the 5G network edge. This work represents one of the world's first tight integrations between a multi-access edge computing (MEC) platform and 5G stand-alone (SA) network core. It enables a range of use cases based on machine learning-based video analytics.

Closed-circuit television cameras (CCTVs) have become ubiquitous across the globe, with millions of units deployed at a pace that continues to accelerate. The global CCTV market is expected to grow at a CAGR of approximately 12.1 percent through 2028, reaching an estimated value of USD 6.68 billion in that year.[1] This proliferation intends to address concerns over public safety, as well as more pedestrian uses in commercial, industrial, and many other settings. Inexpensive cameras attached to increasingly capable and cost-effective civilian unmanned aerial vehicles (UAVs, or drones) in particular enable a massive and open-ended variety of use cases.

Even as CCTVs capture massive amounts of visual data, video analytics are applied to only a small fraction, dramatically reducing its value. Video analytics powered by machine learning lets information captured by CCTV cameras be applied to a range of use cases that range from traditional roles such as security to more forward-looking ones such as smart cities. Communication service providers (CoSPs) are positioned to profit from new service offerings across this spectrum by transforming video data streams into insights.

The present lack of video analytics is due in part to the complexity and expense of backhauling massive data to a central cloud or data center for processing. By placing compute resources closer to where data is produced, multi-access edge computing (MEC) offers ultra-low latency, reduces network congestion, and ensures high bandwidth that can be used for time-critical applications, as shown in Figure 1. Edge computing is widely seen as transformative to enterprise computing; Deloitte estimates that by 2023, 70 percent of enterprises will be using the intelligent edge.[2] DeepSight AI Labs is working with the ecosystem to optimize the performance and cost-effectiveness of these new usages.

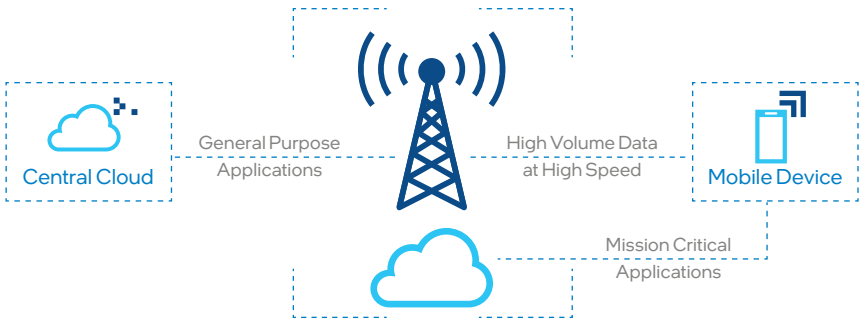

Figure 1. Multi-access edge computing (MEC).

Real-Time Computer Vision with DeepSight Platform

In addition to lacking video analytics to make full use of the information that CCTV cameras capture, traditional implementations depended on wired transmission of signals, which drove up costs, limited scalability, and made installations in many locations impractical. By contrast, DeepSight Platform leverages 5G network infrastructure to build highly scalable and stable video analytics solutions that can process live feeds from thousands of cameras in real time, taking advantage of 5G's advances in high bandwidth, low latency, and enhanced security. DeepSight Platform is built to be retrofitted into existing implementations, and it supports any CCTV camera, anywhere 5G can reach, including remote areas and onboard drones. Cloud-based deep learning enables implementations to scale to thousands of CCTV cameras and beyond, while implementations can span many industries, end devices, and functionalities, as shown in Figure 2.

DeepSight Platform provisions and orchestrates cloud-native video analytics applications based on microservices and containers at the network edge and provides resource orchestration in production. It also provides comprehensive configuration and management of network parameters, as well as streamlined provisioning for CCTV cameras. In addition, the platform integrates with other network elements such as industrial controllers and IoT actuators to provide real-time control of electromechanical systems.

In addition to providing full customizability of analytics, DeepSight includes a range of ready-to-deploy capabilities right out of the box, including detection of intrusions, camera tampering or faults, face masks, and helmets, as well as crowd counting and social distancing monitoring.

The platform can generate a number of different types of alerts based on set parameters, including phone calls or texts, email, WhatsApp, audible alarms, or alerting based on IoT devices. In addition, a built-in dashboard shows alerts based on video analytics and related information in real time, keeping interested parties continuously apprised of situations as they occur. Those user interfaces can also include supplementary information—such as historical data, heatmaps, and other visualizations—to add further to the value and actionability of the insights produced by DeepSight Platform and its associated applications.

Video analytics applications based on DeepSight Platform can be deployed on any MEC infrastructure and any public or private 5G network. CoSPs can offer novel subscription-based services based on the technology, creating new revenue streams and increasing average revenue per user (ARPU). At the same time, subscription-based services reduce complexity and cost of ownership for customers, who do not have to maintain equipment or associated networking. The solution is also compliant with privacy regulations including the European Union's General Data Protection Regulation (GDPR).

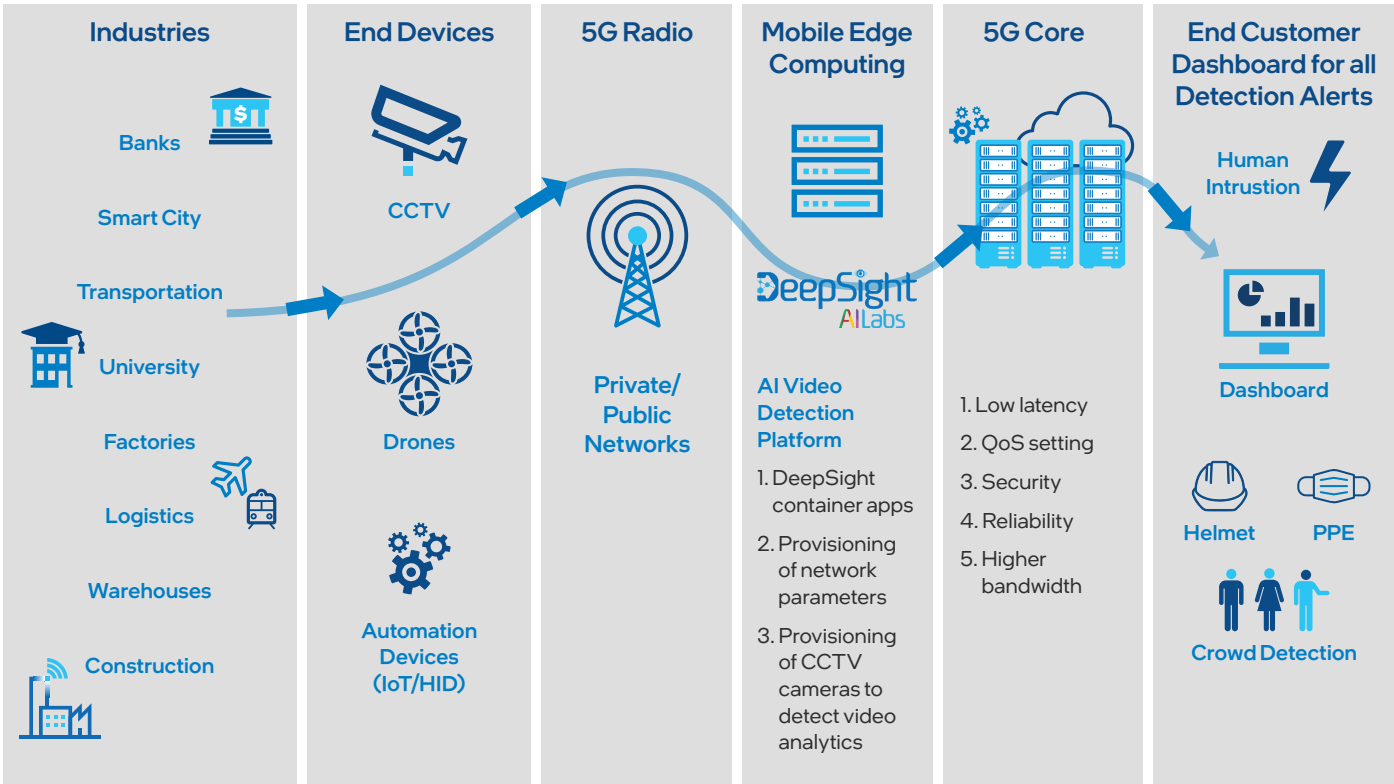

Figure 2. DeepSight Platform implementations on 5G networks.

DeepSight Monitoring Capabilities Based on Video Analytics

Video analytics can derive many types of meaning from video streams, analogous to human understanding of the same scenes but with distinct advantages. Chiefly, analytics can look at hundreds or thousands of video streams at once, without the attention lapses that affect human operators, who can monitor no more than a handful over an extended period. A few examples of the many types of capabilities the DeepSight Platform enables for applications are illustrated in Figure 3.

Figure 3. Example DeepSight Platform monitoring capabilities based on video analytics.

DeepSight Use Cases Across Industry Verticals

Applications based on DeepSight Platform support potential use cases within many industry sectors, some examples of which are illustrated in Figure 4.

Figure 4. Potential DeepSight Platform use cases by sector.

DeepSight Real-World Integration of Edge Video Analytics with a 5G Networks

A recent proof of concept by Capgemini Engineering, Ericsson, Intel, and Telefónica performed a pioneering integration of the intelligent edge with 5G SA network core, as illustrated in Figure 5. DeepSight AI Labs participated in the collaboration to provide use cases based on real-time image processing and video analytics, demonstrating streamlined development potential for third parties of customized applications for the intelligent edge. In particular, the project combines cloud advantages of cost-effective scalability on demand with the performance and convenience of on-premise solutions. It provides reference implementations of 3GPP network components to accelerate customer time to value. This work was conducted at the 5TONIC lab in Madrid, which was built by Telefónica and IMDEA Networks to advance 5G network innovation.

The solution architecture implements Capgemini Engineering's ENSCONCE Edge platform with a DeepSight application server, integrated with Ericsson's 5G SA core on Telefónica's mobile network. Compute infrastructure was based on 3rd Gen Intel® Xeon® Scalable processors, with optimized operation enabled by Intel® Smart Edge Open and Intel® Distribution for OpenVINO™ toolkit. The cloud-native intelligent edge dynamically steers traffic between the mobile network and destination applications.

Data Flow and Operation

In this implementation, application intelligence is deployed at the network border, close to the 5G user plane function (UPF). Direct data access from the CCTV cameras helps provide low-latency connectivity, while control-plane connectivity from cameras to the MEC platform enables discovery of the edge video-processing server. CCTV cameras connect to the network core and DeepSight application server using 5G SIM cards and VPN tunnels. Control connections from the MEC platform to the network exposure function (NEF) within the 5G core support capabilities that include the following:

- Requests for device information (e.g., location)
- Requests for guaranteed quality of service (QoS)
- Modification of charging policies
- Optimization of network routing for edge applications

The test application detected intrusions (the presence of humans in the video frame at specified times of day), multiple people, and helmets. Upon detection, the application sent alerts over the internet to an administrator dashboard.

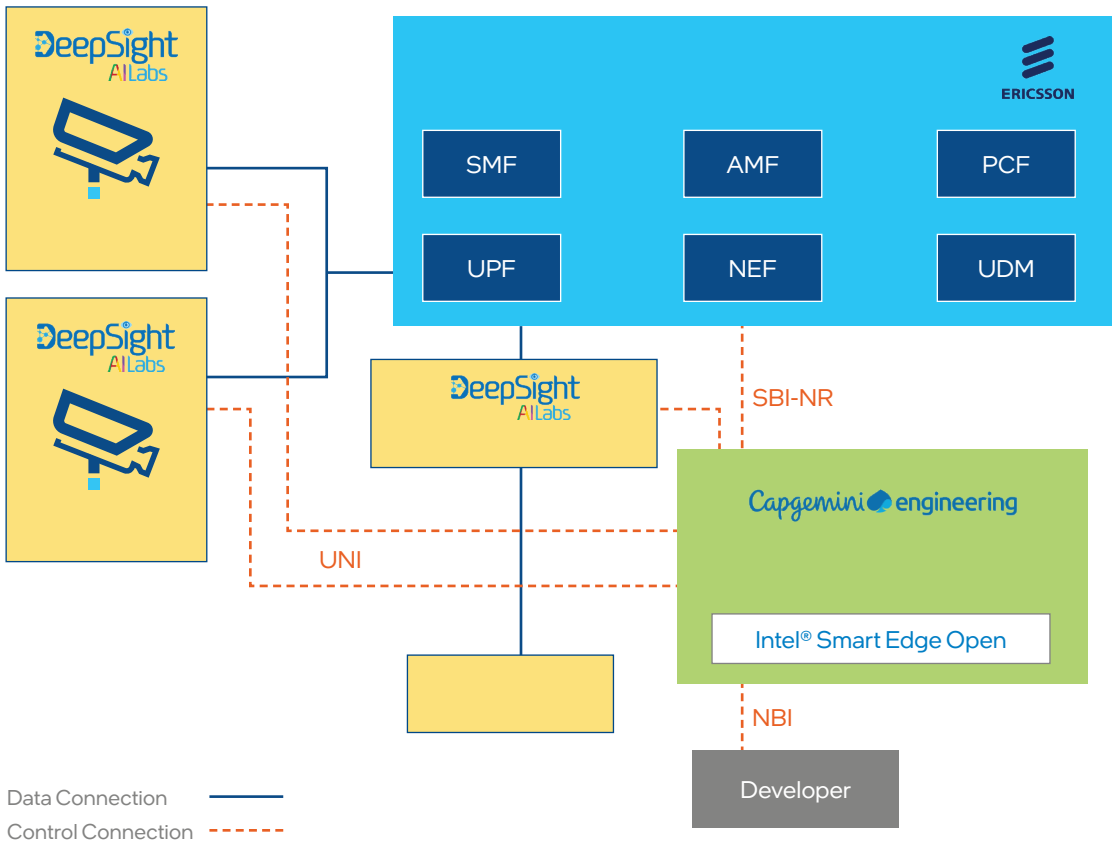

Figure 5. High-level solution architecture.

Results

The proof of concept shows the 5G network-readiness of solutions based on DeepSight Platform. Testing by DeepSight AI Labs confirms that the test application successfully detects objects and conditions from video streams at the intelligent edge in real time for representative use cases and generates dashboard alerts based on that information. Test measurements found latency under 30 milliseconds and throughput between CCTV cameras and the MEC platform of 400Mbps, offering efficient, instantaneous processing. With implementation completed in under three weeks, this work also showcases the advantages of DeepSight Platform in accelerating time to market for new services by CoSPs.

Foundations Built on Intel Technologies

DeepSight Platform is engineered explicitly for Intel® architecture, including use of the latest platform technologies to accelerate compute-intensive training and inference for machine learning, particularly for implementations that include large numbers of CCTV cameras. It builds on open-source software toolkits that are optimized for performance on Intel® processors.

3rd Gen Intel® Xeon® Scalable Processors

Built explicitly for flexibility, 3rd Gen Intel Xeon Scalable processors are engineered to run advanced computations such as machine-learning-driven video analytics on the same physical platforms as mainstream application workloads. They are available in a wide variety of SKUs with a range of options, features, and core counts.

To accelerate artificial intelligence training and inference, the processor includes enhancements to Intel® Deep Learning Boost (Intel® DL Boost) that deliver up to 1.59x higher real-time inference throughput, compared to the previous-generation platform.[3] Enhanced Intel DL Boost introduces new sets of Intel® AVX-512 instructions that speed up computer vision workloads and streamline deep learning inference:

- **Brain Floating Point 16-bit (bfloat) instructions** compute floating-point numbers encoded to occupy 16 bits instead of 32, increasing efficiency for AI workloads with high compute intensity but low precision requirements.
- **Vector Neural Network Instructions (VNNI)** combine three instructions into one, to optimize utilization of compute resources, cache, and on-package I/O for accelerated AI inference.

Intel® Smart Edge Open

Intel® Smart Edge Open (formerly known as OpenNESS) is a royalty-free edge computing software toolkit for building platforms optimized for performance at the edge. The toolkit provides functional building blocks that enable flexible integration into edge solutions for 5G use cases with abstraction of the underlying network complexity. This cloud- native, microservice-based platform includes lightweight resource orchestration capabilities, building on the Kubernetes control plane for provisioning and configuring edge resources. It also provides a range of use case-specific SDKs that enable streamlined convergence of diverse workloads onto edge platforms. The toolkit is built for tight integration with Intel architecture, including processor platforms and accelerators.

Intel® Distribution of OpenVINO™ Toolkit

DeepSight Platform uses capabilities of the Intel Distribution of OpenVINO toolkit to build, run, and optimize its underlying machine-learning models. Designed explicitly with a “write once, run anywhere” model, the toolkit enables applications to be deployed across architectures based on any combination of CPUs, GPUs, and accelerators. It provides development tools, libraries, and pre-optimized kernels to streamline development, enabling high performance from convolutional neural networks (CNNs) deployed for DeepSight inference workloads at the intelligent edge.

Conclusion

DeepSight Platform enables new edge-oriented business models for CoSPs on 5G networks, including subscription-based services such as AI-enabled surveillance, monitoring, and safety applications, in compliance with GDPR and other privacy regulations. It accelerates time to market for edge-based video analytics solutions while also delivering optimized performance on Intel processors, platforms, and accelerators. DeepSight Platform is a critical asset for providers as they seek new ways to monetize their 5G investments and differentiate their services for a competitive advantage.

More Information

Intel® Network Builders: networkbuilders.intel.com

Intel Smart Edge: intel.com/content/www/us/en/collections/technology/smart-edge.html?s=Newest%0D

DeepSight AI Labs: deepsightlabs.com

5TONIC Laboratory 5G and MEC Integration: https://www.5tonic.org/news-20210412-5tonic-laboratory-completes-pioneering-tight-integration-between-5g-sa-network-and-edge-platform/

Solution provided by:

intel + DeepSight AILabs

[1] Data Bridge Market Research, February 2021. "Global Closed Circuit Television (CCTV) Camera Market – Industry Trends and Forecast to 2028." https://www.databridgemarketresearch.com/reports/global-closed-circuit-television-cctv-camera-market

[2] Deloitte, 2020. "Global TMT Predictions 2021." https://www2.deloitte.com/content/dam/insights/articles/US93838_TMT_Predictions_2021/93838_TMT-predictions-2021-infographic.pdf.

[3] Performance varies by use, configuration and other factors. See [122] at intel.com/3gen-xeon-config

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 is committed to respecting human rights and avoiding complicity in human rights abuses. See Intel's Global Human Rights Principles. Intel's products and software are intended only to be used in applications that do not cause or contribute to a violation of an internationally recognized human right.

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.

0322/DO/MESH/346430-001US