HAILO

Top Performing Video Management System

HAILO-8™ POWERFUL VIDEO ANALYTICS AT SCALE

Video Management Systems — a growing market with growing demands

A Video Management System, also known as VMS, serves as a global term for all the software elements that are involved in handling multiple video channels at scale. The system collects inputs from a number of cameras and other sensors and addresses all related aspects of video handling, such as storage, retrieval, analysis and display. Video analytics is increasingly becoming an important part of the VMS, enabling advanced capabilities that are applied to multiple streams concurrently.

VMS systems are used in different contexts. Most commonly, VMS is used for security and surveillance, enhancing personal safety while maintaining individual privacy. Another typical use case is extraction of business intelligence from user behavior analysis for the purpose of customer experience improvement in retail and other industries.

Traditionally, the analysis of live video streams used to be manual, relying on human perception for visual identification of events happening in each feed. This method has many disadvantages as it is not scalable, and prone to errors caused by operator's fatigue, leading to false alarms, missed occurrences, and wasted time and money.

Nowadays, deep learning is enabling the automation of the analytics task, thereby allowing for easier scalability, and improvement in overall performance. This eventually leads to lower total cost of ownership (TCO).

According to a recent market research[1], the video management system market size is expected to reach $31B by 2027, growing at a Compound Annual Growth Rate (CAGR) of 23.1% in the next 5 years. The key drivers for this growth are increasing security concerns in parallel to rapid adoption of IP cameras for surveillance and security applications.

Functionality of aTypical VMS

A typical VMS can include one or more of the following functions:

1
Management

The front end used by the VMS operator, enables diligent control over the video streams. In advanced systems it might be one of the termination points for metadata collected by the analytics entity for the purpose of logging and later on searching and displaying.

2
Streaming

This entity is responsible for the video data transmission including compression, decompression, time synchronization and video content manipulation to guarantee smooth delivery (e.g., regulation buffers; data shaping; frame dropping, etc).

3
Storage

This entity is responsible for indexing, storing, and enabling the retrieval of the data and metadata when and as needed. Efficient and secure handling of the storage fabric is also an inherent part of this entity.

4
Display

The display entity is responsible for serving the various views, analytics and insights to the operator clients / endpoints. Display and visualization could be handled on numerous platforms and in multiple versions that reflect different requirements and viewpoints (e.g., display walls; mobile clients; zoom in/out on specific streams, toggle between different views, etc.).

5
Analytics, Monitoring & Event Triggering

This is where most intelligence is applied in order to extract metadata from the video streams. Not all the components of analytics, monitoring & event triggering are currently implemented in VMS by Artificial Intelligence (AI), but adoption of AI technologies and applications for these functions is on the rise given the potential benefits at the system level. For example, smart indexing and search functionality at the storage entity can be handled more efficiently and automatically using AI.

VMS analytics can be implemented, in either of the following system components:

- At the data source, on smart cameras
- At an aggregation point such as smart gateways, usually referred to as Smart NVR (network video recorder)
- On the cloud, which is sometimes administered as a service. The latter usually referred to as VSaaS (Video Surveillance as a Service)

The Benefits of Edge AI for VMS

To enable real-time and accurate video analytics, VMS systems require high-performance computing. AI is transformational to this market as it allows faster and more accurate event identification at a lower cost.

Hailo provides a range of top-performing AI accelerators, which are specifically designed for deep learning tasks performed in real-time and with minimal power consumption, size and cost. These groundbreaking chips deliver data-center-class performance to video analytics for video management systems, whether analytics take place on cameras, small gateways, or large servers, supported by our ecosystem of global leading hardware partners.

Performing advanced analytics on the edge is game changing as it enables:
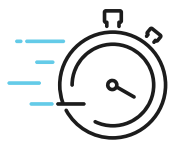
Real Time Insights

Low latency and higher frame rates enable timely video processing for detection and search on video streams. Low latency enables rapid action-reaction cycle with real-time generation of alerts and insights that allow operators and first responders to take immediate action for maximum safety. Higher throughputs deliver better visibility to scenarios of high dynamics nature.

High Accuracy

Improved performance by lowering the chances for false alerts and missed detections, usually requires more compute power. In many cases limited computational capacity forces implementers to sacrifice performance to meet these limitations.

Easy Integration

A comprehensive software suite which supports a wide range of platforms, operating systems, and neural network models.

Cost Efficiency

Unrivalled cost efficiency, with more compute per unit price that is translated into more complexity per application, more applications per stream and more streams per platform.

Improved Privacy

Video analytics on the edge means that personally identifiable information (PII) does not need to be transmitted and stored in the cloud, enabling maximum security without compromising compliance.

High Reliability

Products designed to withstand harsh conditions, and conform with industrial operating conditions, with extended temperature range of -40°c to 85°c; and low power consumption which eliminates the need for active cooling.

INTELLIGENT VIDEO ANALYTICS FOR VIDEO MANAGEMENT SYSTEMS

DETECT / RESPOND / ANALYZE

Detect

The first layer of video analytics in VMS is scene understanding and metadata extraction for the purpose of both real-time response as well as long term storage for future search and analysis. Multiple neural networks can be deployed for this phase, leveraging advanced algorithms to perform deep-learning tasks. These tasks serve as the foundation on which more complex tasks identify pre-defined events and trigger specific response. Some of the more common applications are described in the following pages.
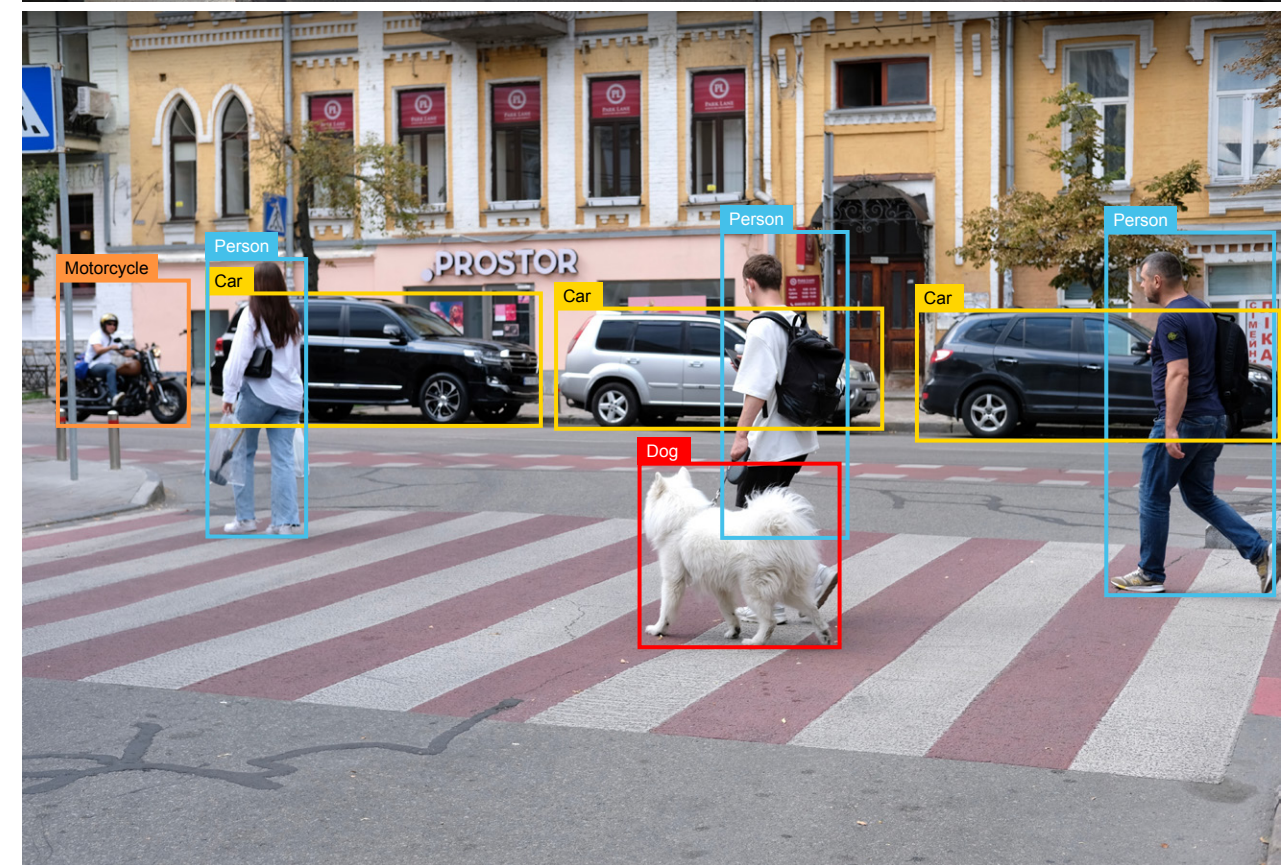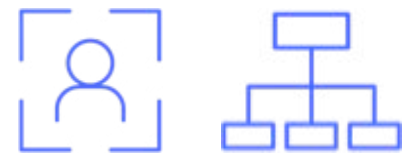
Detect ›
Object Recognition

Object detection is the most common task and was the first task to emphasize the benefits of harnessing artificial intelligence to replace classic machine vision processing. It is usually the first task in the pipeline that precedes further, more complex analytics. Object recognition models are used to identify a specific object or a class of objects within a frame and could be used to distinguish between classes of objects (for example between people to other objects in the background of the frame), or to identify different entities of the same class (for example, identify every person or every car in the frame).

A relatively complex task of object recognition is License Plate Recognition (LPR, also known as ANPR – Automatic Number Plate Recognition), which requires 3 different models to identify a vehicle in the frame, the license plate within the vehicle, and the characters (numbers or letters) within the license plate. The plate number is the unique identifier of the vehicle which serves to identify it for the purpose of billing, tolling, access control, etc.

The relevant neural network tasks commonly used for this kind of applications are **object detection** followed by **classification**.
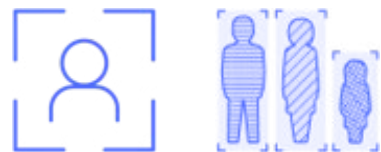
Detect ›
Counting

Counting the number of people in a specific area is a common task, especially in places where occupancy is limited and has safety implications. Counting could be used, for example, to manage or control load in closed areas like an elevator or a concert hall, to optimize energy consumption in a large building, or to verify there is no living soul in a building during a fire or other emergency.

Machine learning is used for this task to achieve more accurate outcome in challenging cases like partially hidden objects, occlusions etc.

Furthermore, AI-driven scene understanding enables a 'scene-aware' counting, i.e., counting in relevant semantic areas within a given frame (e.g., in the illustrated scene - counting people on the walkway and on either side of it).

Common neural network models used for such tasks are usually **object detection or instance segmentation.**

Gate D59
## Estimated Crowd Size: 351

Detect ›
Density Estimation

A bird's eye view of density can be used to identify an evolving situation. Unlike counting, this is relevant in cases in which the exact number of objects is not required but rather an overall density.

Density is usually represented by means of heatmaps that provide clear and intuitive way to observe behavior over time and potentially highlight activity patterns. It can serve as a tool for identifying obstacles; unused areas etc.

The relevant neural network task used for this kind of applications would be **semantic segmentation.**
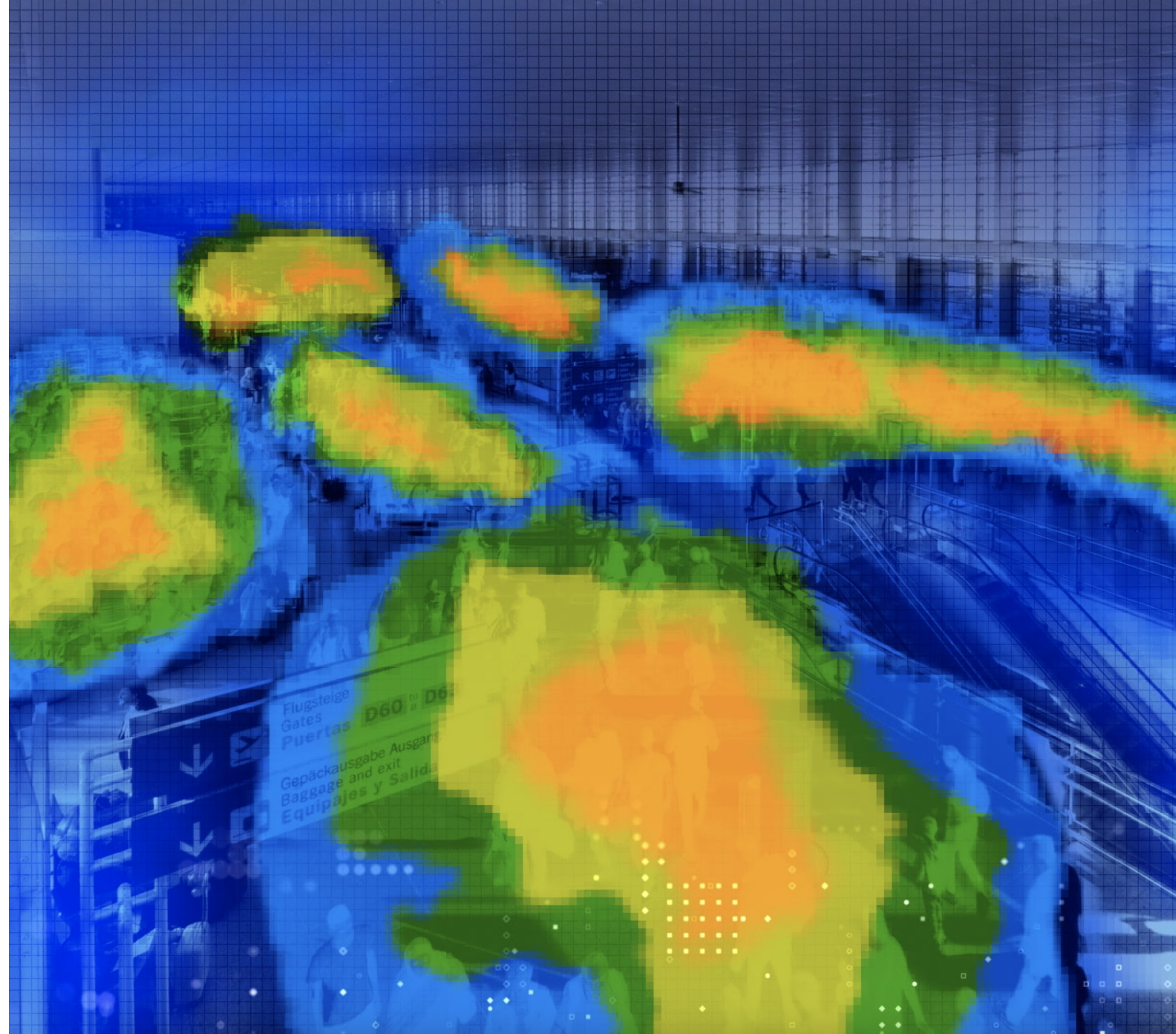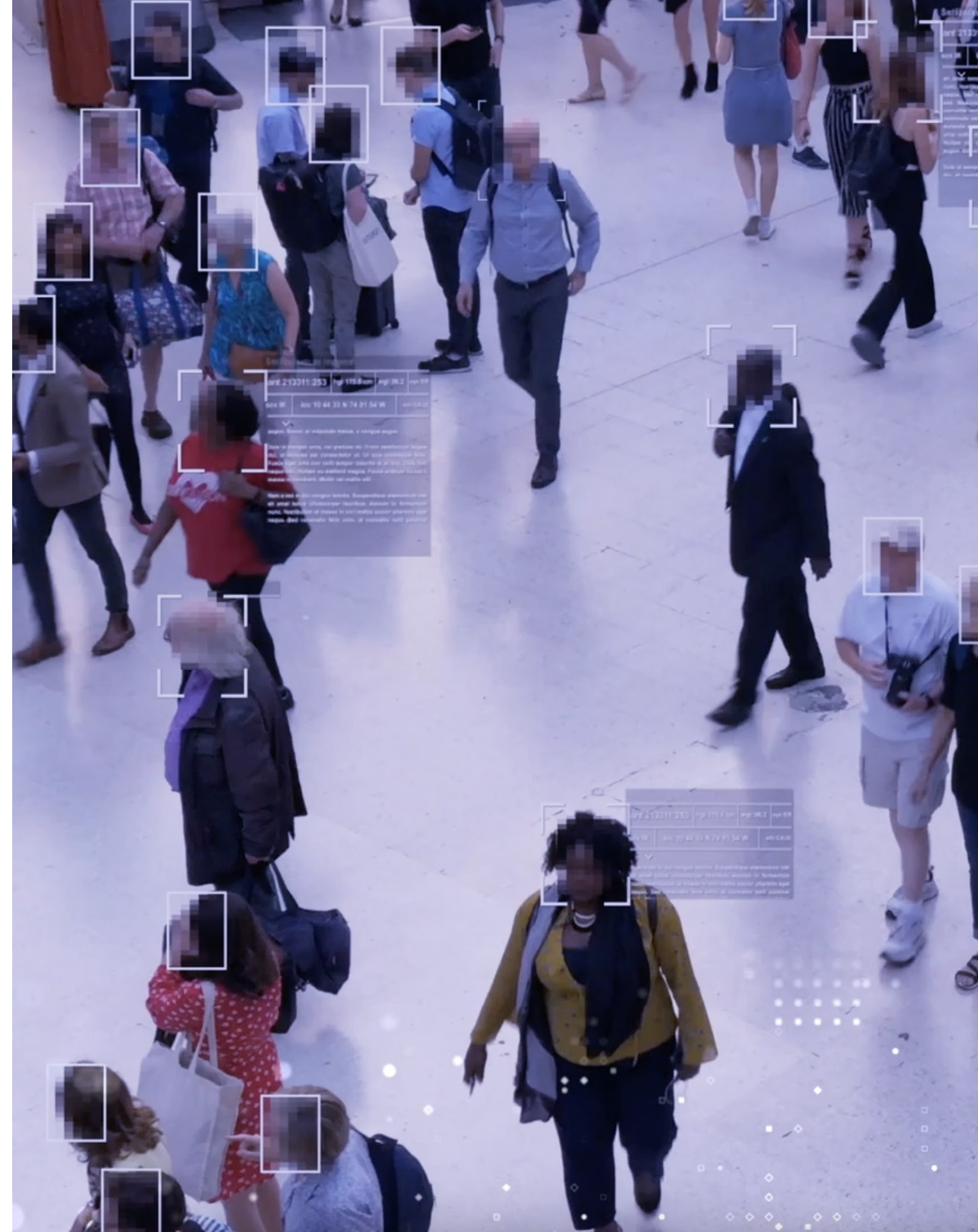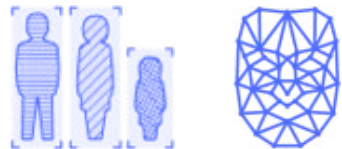
Detect ›
Object Attributes

Identification of specific attributes of people's face, people's body, vehicles or other objects. These attributes can assist in unique identification of a person or an object in a specific scene, or re-identification of the same unique entity in the same scene over time, or in multiple locations. This data can be used to trigger specific events, or over time to analyze behavior and provide statistics.

Instance segmentation is usually required for this task in order to provide both classification and a unique ID per entity, alongside other models specifically designed for **attribute recognition** and **facial landmark** extraction.

A very important aspect of this task, that is very likely to be applied, is the **anonymization** of Personally Identifiable Information (PII) which is carried out to protect private or sensitive information that connects specific individuals to the stored data, thereby preventing disclosure of raw visual reference of the objects or other sensitive data in the scene.
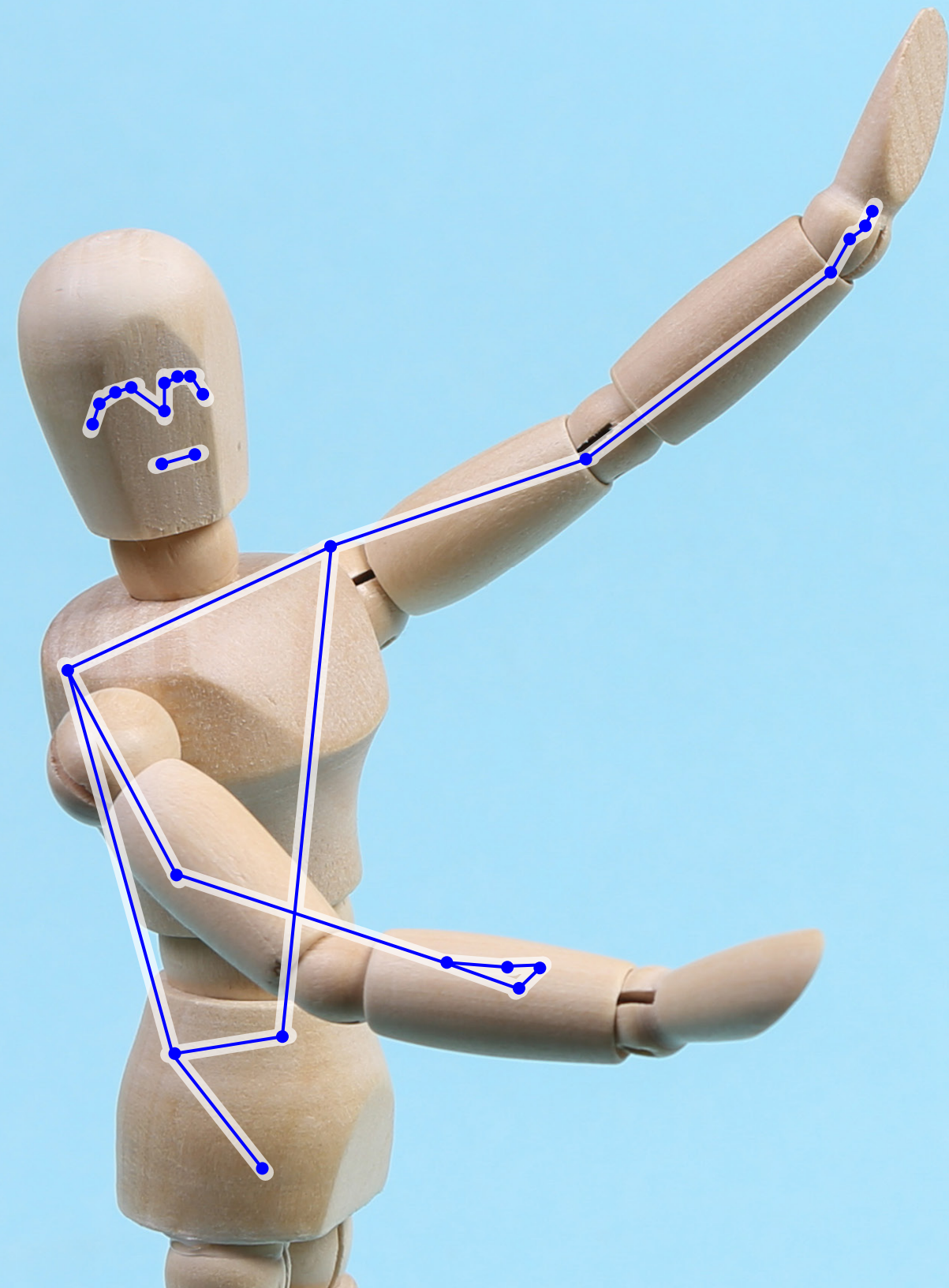
Detect ›
Gesture Estimation

Analyzing the posture of a human body in space, in order to understand gestures of the body, and translate them to a 2D/3D model of the body (also known as a "stick figure"). Analysis of a sequence of such gestures enables interpretation of specific behaviors for the purpose of behavioral analysis.

The neural networks used for this task are of the **Pose estimation** models.

Detect ›
Distance
Measurement

Calculating the distance between two or more people or objects, based on the accurate 3D location of each entity in a defined space. Vast experience in this type of applications was gained from Advanced Driver Assistance Systems (ADAS), used to calculate the distance between adjacent cars to maintain safe driving distance and provide forward collision warning. This application relies mainly on **3D object detection** AI models
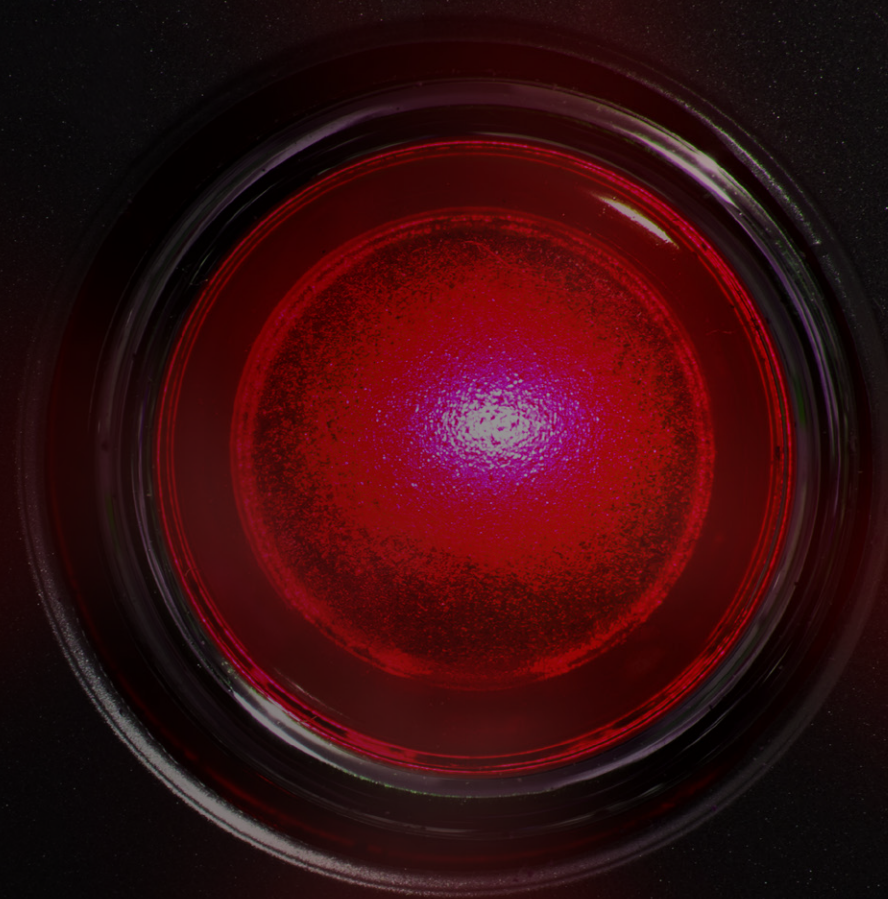
Respond

The second layer of video analytics in VMS is the real-time event triggering based on insights from the Detection phase. In this phase pre-defined events will trigger a specific response, such as setting off an alarm or a call to action for security personnel and/or first responders.

Some examples of applications that use machine learning to trigger a response based on identified events are listed in the next few pages.

Respond ›
Crowd
Management

Based on people and vehicle counting, and density estimation, the accumulation of large number of entities will create an event in a specific location which may trigger a response or call to action. Examples could be load balancing for example, if queue is too long, open another checkout in the supermarket or department store, or another counter in the bank or at the airport. Another example could be the triggering of a more dynamic management of traffic lights to release bottlenecks and alleviate traffic jams.

Respond ›
Perimeter Protection

Face and person attributes as well as gesture estimation and distance measurement, can all serve to secure and protect an area with restricted access or an unsafe zone. For example, entrance permission or denial can be granted based on facial attributes, without the collection and storage of personally identifiable information. A tripwire application can be used to alert security personnel when people or vehicles cross a pre-defined border and enter a restricted or dangerous area. This can be done in a direction sensitive manner so that one-directional or bi-directional restriction can be set. The advantage of using advanced machine learning algorithms is that the accuracy level is high, and no stray cat or dog will set false alarms. Another benefit of AI is the ability to dynamically set the border and adapt to changing physical conditions, rather than manually sketching a border line which does not accurately fit the contour of the restricted or monitored area.

Respond ›
Social
Distancing

2020 was a year to be undoubtedly remembered by the books of history. A global pandemic changed the way we work, travel, shop and communicate. Social distancing became a recommended and sometimes mandatory means for preventing the spread of the virus and was usually enforced through stickers or lines drawn on the floor, marking the recommended space between people in queue or waiting rooms. AI-based distance measurement can be used to detect whether people maintain social distancing in public places, supporting health authorities in their effort to contain the plague.

Respond ›
Behavioral Analysis

Artificial intelligence can be leveraged to recognize distress signals and trigger an alert to call first responders to the scene as soon as possible, and much faster than a call to 911 would. Such SOS applications use surveillance cameras in public areas to run applications of pose and gesture estimation which are programmed to identify specific behaviors based on a sequence of poses. For example, such applications can identify distress signals, a person fainting or having a seizure, and alert an ambulance to the scene.

Other applications are programmed to identify loitering, vandalism or violence, and even to recognize a weapon in the frame, alerting police patrols to the area.

Respond ›
Lost Person / Unattended Object

Another application that can be very helpful for security personnel is searching and tracking lost people based on property indexing and gesture estimation. A quick search for certain keywords can help parents track a lost child in the mall or at the airport, or alternatively can help security locate a suspect in a crowd or large building.

The same rational can be applied to locate unattended objects left behind by mistake or in more extreme cases, with intention to cause harm or damage.

Respond ›
License Plate Recognition (LPR)

LPR can be used for access control and billing in parking lots or garage. In these cases, a camera will usually handle one car at a time. In other cases, such as free-flow tolling, this task needs to be carried out in high-speed and high-resolution, even in poor lighting conditions such as night-time, rain or fog. In this case, a single camera should be able to capture in a matter of seconds the license plate number of multiple vehicles before they move out of the frame. To this end high performance AI-based video analytics are required.

For additional information on LPR and Intelligent Transportation Systems, refer to our ITS eBook.

Analyze

The third layer in video analytics for video management systems is indexing, storage and retrieval of metadata. Artificial intelligence is leveraged to cost-efficiently index and store relevant information from the video streams.

Deep insights can be extracted by exploring patterns over long-term observations and joint analysis of multiple points of view.

This can be employed for multiple purposes as exemplified in the following pages.

Analyze ›
Indexing & Recording

To efficiently manage the data and reduce communication and storage costs, artificial intelligence is used to differentiate between meaningful events and background footage. This enables the system to record and store only meaningful events as defined by the user. The camera captures the scene on a 24/7 basis, but only real events such as person or vehicle entering or exiting the frame, or crossing a pre-defined threshold or border, would necessitate recording at full speed. Other parts of the video can be trimmed.

With relevant metadata attached, this allows indexing for easy and rapid retrieval.

Analyze ›
Summarization

Summarization is another important feature of AI at the interrogation level, creating an edited and concise version of the camera input, which only includes the significant events and insights, while cropping out "dead footage".

For example, an intersection camera is recording vehicles and people crossing throughout the day, but the summary will only present instances in which people and vehicles got too close to each other, based on distance measurement applications, therefor storing only the events of accidents, or near-accidents.

Analyze ›
Data Extraction

Data extraction is the act of Identifying patterns in stored data, based on the metadata extracted from the video stream. The better, and more elaborate this metadata is, the better the results and insights will be. The quality of the metadata relies on the quality of the analytics, and this is where advanced algorithms exhibited by deep learning, come into play.

Security personnel and law enforcement agents can perform a freeform text-based search of the database in order to locate a specific scenario or retrieve an evidence for an alleged crime. Keyword search can be done for classification (person, animal, vehicle, etc.) or for specific attributes (male/female, young/old, hair color, shirt color, etc.). An advanced search can be done based on a combination of metadata, to identify more complex events, e.g., tailgating of two cars at the parking lot exit.

Tracking uniquely identified entities (e.g., a person or a vehicle) can be used for business intelligence. Behavioral analytics can be applied to understand user preferences and improve user experience.

In a supermarket, for example, sales and marketing experts can analyze the trail of customers based on their gender, age, etc. and apply changes to placement of products or promotion of brands to enhance sales and reduce costs.

The underlying task to achieve this purpose is usually termed re-identification.

About Hailo

- Hailo is developing top performance AI processors, specifically designed for edge applications.

- The company is focused on building AI accelerators efficient and compact enough to compute and interpret vast amounts of data in real-time.

- Complicated AI applications that could previously only run on the cloud in datacenters, can now run on the edge, providing high-accuracy, real-time analytics with a fraction of the power consumption.

- Hailo's solutions bridge the gap between existing and future AI technologies and the compute capacity required to power these advanced applications.