Publication date:
May 2024
Author:
Tommy Zhu

The Future of the True AI Camera

Technologies enable advanced image quality, video transmission, security, and video analytics

Brought to you by Informa Tech

Contents

Summary	2
Enhanced video image quality	5
Improved video transmission and security	12
AI-powered video processing	18
Summary and conclusions	26
Appendix	27

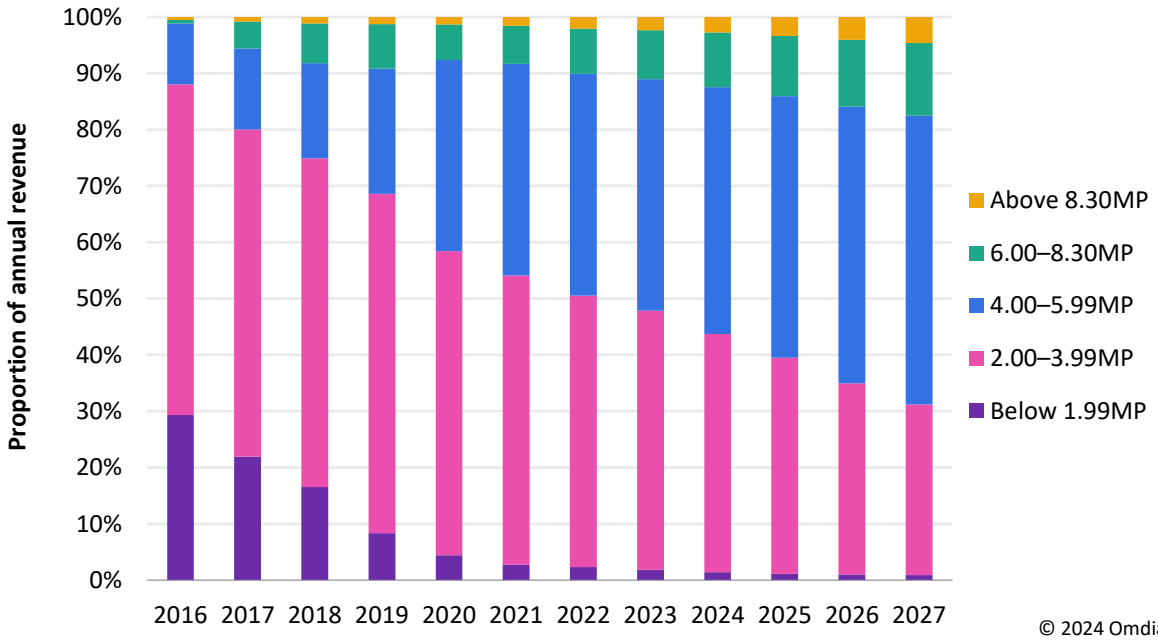
Summary

Security camera market overview

The video surveillance industry has been accelerating to higher resolution, wider adoption of network technology, and more embedded artificial intelligence (AI) functionalities. Omdia estimates that the professional security camera market has been worth $14.8bn in 2023, accounting for 53% of total video surveillance market revenue[1]. The security camera market is forecast to increase to $20.4bn by 2027.

In 2023, 52.1% of network cameras shipped with resolutions higher than 4MP, and the proportion is forecast to exceed 68.8%[1] by 2027 because of the growing demand for improved image quality.

Figure 1: Global network camera unit shipments by resolution, 2016–27



Source: Omdia

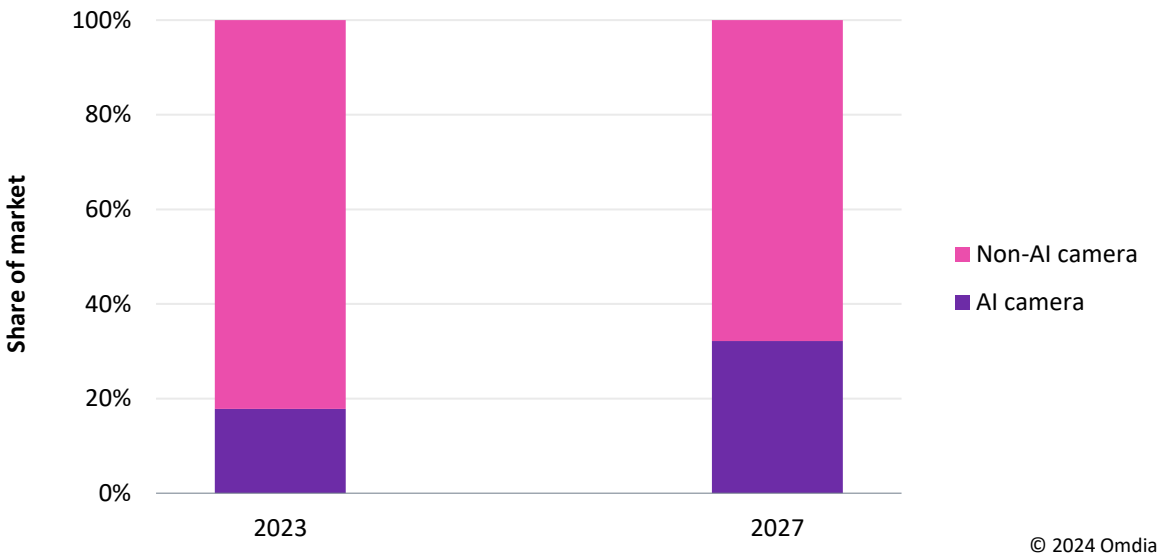
[1] Omdia Video Surveillance and Analytics Report – 2023 (August 2023). The 2023 figures, based on 2022 actuals released in August 2023, may differ from actual figures due to publish in 2024.

AI camera market

With the spread of AI applications into mobile and embedded markets, demand has emerged for hardware that can perform neural network inference at the edge in the context of power-, price-, and area-constrained systems-on-chip (SoCs). Omdia estimates that globally 28 million cameras embedded with AI analytics has been shipped in 2023[1]. With further SoC development, we expect the use of embedded analytics to continue. The adoption of AI cameras has been growing rapidly across various vertical markets including:

- Traffic monitoring: Captures speeding vehicle via License Plate Recognition (LPR) algorithm.
- City surveillance: Safeguards public places by detecting crow gathering and abandoned object.
- Retail: Manages store operations using people counting and heat mapping functionalities.

Figure 2: Global, network cameras' analytic capability



Source: Omdia

[1] Omdia Video Surveillance and Analytics Report – 2023 (August 2023).

Technological trends

Although AI is commonplace in the security industry, the emergence of ChatGPT at the end of 2022 reignited the industry's passion for it. However, there is still a long way to go before foundation model and multimodal are adopted at the edge. Today's video surveillance cameras have in fact been continuously evolving with technological advancements in optics, sensing, codecs, and processing powered by artificial intelligence. AI is now being seen in almost every aspect of surveillance cameras, including image sensors, signal processors, video codecs, and video content analytics.

This white paper discusses the features a modern camera should encompass:

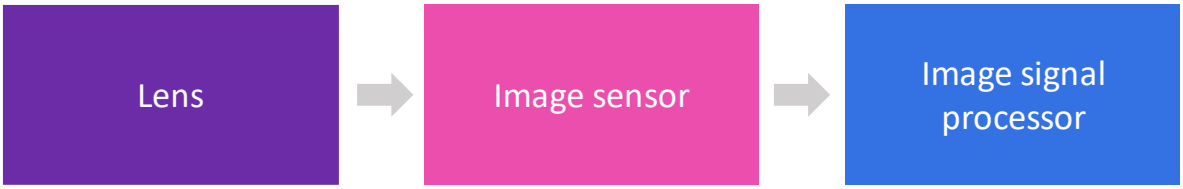
- Enhanced video image quality via lens, image sensor, and AI image signal processor (ISP)
- Improved video transmission efficiency and security via smart codes and security technologies
- Powerful video processing via miscellaneous algorithms and multi-algorithm concurrency

Enhanced video image quality

Challenges and solutions

Increasing number of surveillance cameras deployed on city corners have become a critical tool for public safety. AI cameras have brought city management, business operations, and safety and security to the next level. Although the ability to capture clear color images in daytime is essential for almost all surveillance cameras, video images in the nighttime are not yet ideal. It is evident that most criminal and safety incidents take place at night. A lack of effective video images of such incidents will result in considerable damage. To understand how to achieve clear images in low-light conditions, it is necessary to understand the imaging mechanism of a camera (see **Figure 3**). The light (photons) passes through the lens to the image sensor and is transformed into a digital image signal, which is subsequently processed by the ISP to create a usable image.

Figure 3: Creation of a digital image in a camera



Source: Omdia

Conventionally, a camera's low-light performance is improved by upgrading the elements listed below:

- **Lens:** Coating the lens for less light loss; larger aperture
- **Image sensor:** Larger sensor to gather more light
- **ISP:** Fine parameter tuning for a better image
- **Illuminator:** Increase visible and infrared (IR) spectrum

However, the conventional solution is faced with several challenges:

- Flare and ghost

- Blur of moving objects
- Undesirable performance in varying light conditions
- Light pollution at nighttime brought by flashlights

To overcome these challenges, manufacturers strive for technological breakthroughs in lenses, image sensors, and ISPs.

Latest developments in low-light full-color technologies

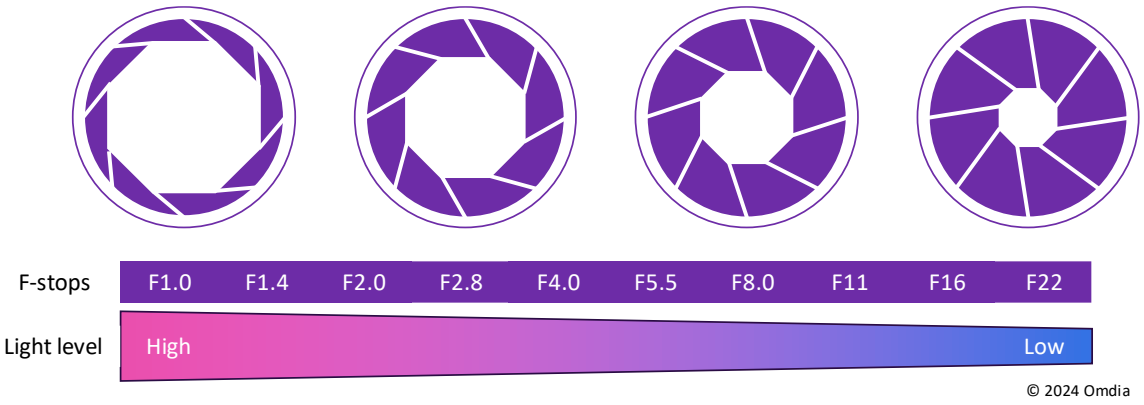
Lens

As the first point of contact of optical path, the lens plays a critical role in the image quality of a camera. The aperture is one of the most important parts of a camera lens. It is a mechanical diaphragm that controls the amount of light that passes through the lens. The opening of a set of circular blades determines the size of aperture. The working mechanism of aperture is like the human eyes—adjusts to light by contracting and expanding the size of the pupil. Apertures can adapt to different lighting conditions by altering the diameter of the opening. Dimmer lighting will need a larger aperture to let more light hit the image sensor. This is why most surveillance camera manufacturers utilize large apertures to acquire full-color imaging in low light conditions.

The size of aperture is expressed by F-stops or F-numbers, representing the focal length of the lens divided by the effective diameter of the aperture. The larger the f-number is, the smaller the aperture is. For example, F1.0 indicates a large aperture while F16 indicates a small aperture. The standard F-numbers commonly seen in the market from larger apertures to small apertures are F1.0, F1.4, F2.0, F2.8, F4.0, F5.6, F8.0, F11, F16, F22**.** Each increment on the standard F-stop scale is the power of the square root of two and halves the amount of light that reaches the image sensor.

The relationship between aperture F-numbers and light level is shown in **Figure 4**.

Figure 4: Aperture and light level



Source: Omdia

Image sensor

Common industry approaches

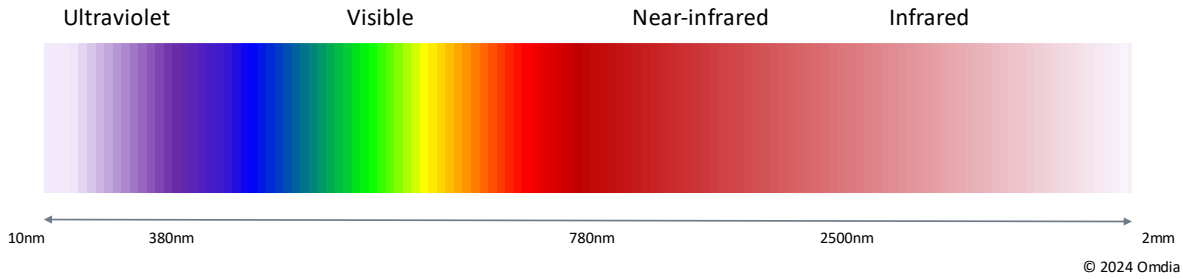
Single sensor solution

The typical approach to enhancing video image quality, especially for low-light scenarios, is to apply a larger image sensor. There are millions of photosites (the parts of a sensor that receive photons) hosted on an image sensor, gathering light and translating into pixels. All else being equal, larger image sensors contain larger photosites and therefore deliver better low-light images than a small sensor. However, as the size image sensor increases, the cost rises.

Apart from using larger image sensors, how does a security camera work in low-light conditions?

This has to do with the mechanism of the human eye. The human eye is sensitive to light of wavelengths between 400nm and 700nm (visible light). There are two types of photoreceptors on the retina: the rod sensing intensity of light for night vision and the cone sensing wavelength of light for color vision. Human eyes lose color vision when the wavelength goes beyond 700nm. The typical image sensor, on the other hand, can capture a maximum wavelength of 1,100nm (light with a wavelength between 700nm and 1,100nm is within near-infrared range). Because of the difference between the human eye and the image sensor, the image captured by an image sensor will present unnatural color to an eye because of the impact of near-infrared (NIR) in low-light conditions.

Figure 5: Part of the electromagnetic spectrum



Source: Omdia

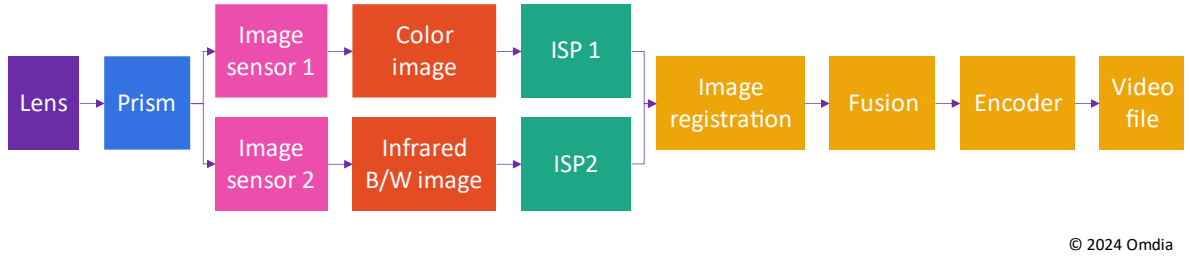
A conventional low-light camera uses an IR cut filter to block infrared light in the daytime and only allows visible light to pass through the camera. Colors will be vividly represented and not distorted by NIR light. In the nighttime, with the help of IR illuminator, the IR cut filter disengages, and NIR will be captured by the sensor to produce a clear monochrome image. Further, to achieve a color image in low-light conditions, manufacturers also employ large sensors supplemented by a white light illuminator for color reproduction. However, this introduces digital noises and loss of detail as well as light pollution.

In recent years, various manufacturers have developed new solutions to acquire full-color images in low-light conditions.

Dual-sensor solution

A typical approach is to apply a dual-sensor solution. A prism splits the light coming through the lens into two light rays: one NIR ray to capture the brightness and sharpness of the image and another visible light ray to reproduce full color. The two light rays hit two sensors and an ISP channel, which process the color image and monochrome image respectively and subsequently converge the two images to create a true full-color image in low-light conditions. However, this solution introduces higher cost and increases the complexity of assembly and calibration.

Figure 6: Dual-sensor solution

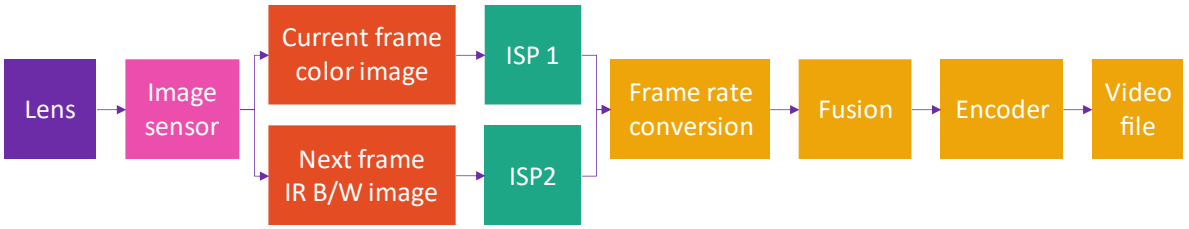


Source: Omdia

Dual-frame exposure solution

Another approach is to apply a dual-frame exposure solution via one image sensor. This solution uses a high-speed switchover in front of an image sensor to capture two successive frames: the infrared frame and the color frame. Like the dual-sensor solution, the infrared frame captures the brightness of the image, and the color frame provides a full-color image. The two frames are subsequently sent to two ISP channels and then the digital signal processor (DSP) integrates the two frames to produce a color image with brightness details. The solution provides a cost-saving option for full-color imaging in low-light conditions because only one image sensor is applied. However, because of the time gap between the two frames, blur will appear in the image if fast-moving objects are captured.

Figure 7: Dual-frame exposure solution

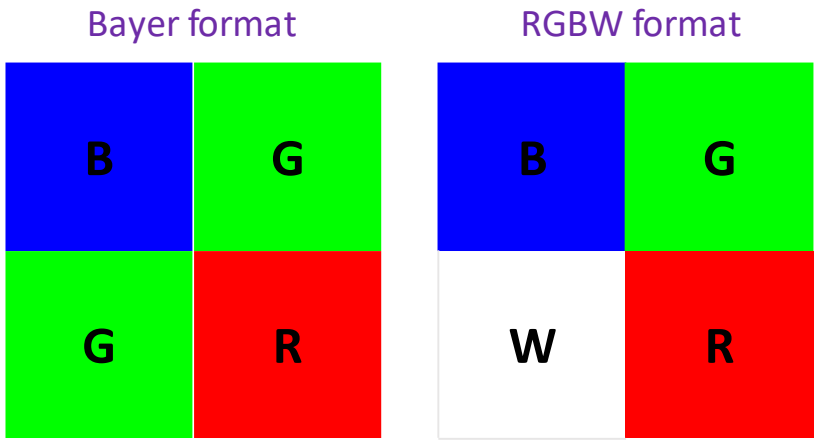


Source: Omdia

Red, green, blue, white image sensor

With advancements in image sensing and signal processing, new image sensors adopted by new generations of smartphones are gaining popularity in surveillance cameras to overcome the challenges in low-light conditions. Most conventional color sensors use a monochrome sensor as a base and apply a color filter array (CFA) to restore color. The most widely adopted CFA in surveillance cameras is the Bayer RGGB (red, green, green, blue) pattern. Inspired by the mechanism of the human eye, the red, green, blue, white (RGBW) CFA pattern replaces a Bayer green pixel with a white pixel to enhance the limited sensitivity of the Bayer CFA under low-light conditions. This is because the white pixel captures all spectrum rather than just RGB spectrum, so the sensor's sensitivity is improved. However, the RGBW sensor suffers from loss of color information caused by elimination of a green pixel, which is the color to which the human eye is most sensitive. This issue is well tackled by developments in color restoration algorithms and the increasing computing power of image signal processors.

Figure 8: Pixel replacement from Bayer format to RGBW format



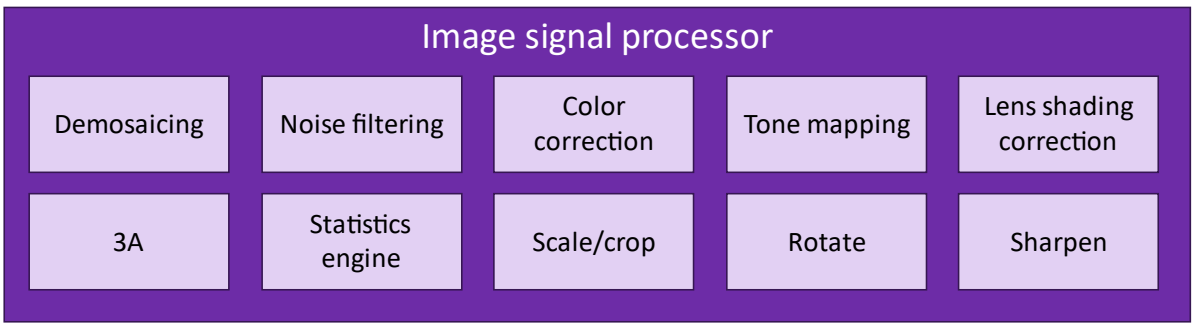
Source: Omdia

Image signal processor

Conventional ISP

Apart from the application of the advanced optical technologies discussed above, capturing clear color images in low-light conditions also requires a powerful ISP. Because of the mechanism of lens and image sensor, the raw image created by the image sensor has different color and brightness than is seen in human vision, resulting in unwanted distortion. The ISP plays a critical role in translating raw sensor data into visually appealing images with a series of functional algorithms including noise reduction, color correction, and brightness and contrast adjustment. A functional block diagram of a typical ISP is shown in **Figure 9**.

Figure 9: ISP high-level functional block diagram



Source: Omdia

Conventional ISP requires complex parameter tuning and considerable effort from engineers to adapt to varying lighting, weather, and application conditions. Even so, after a long development cycle, the image may not present the desired effect in real-world scenarios.

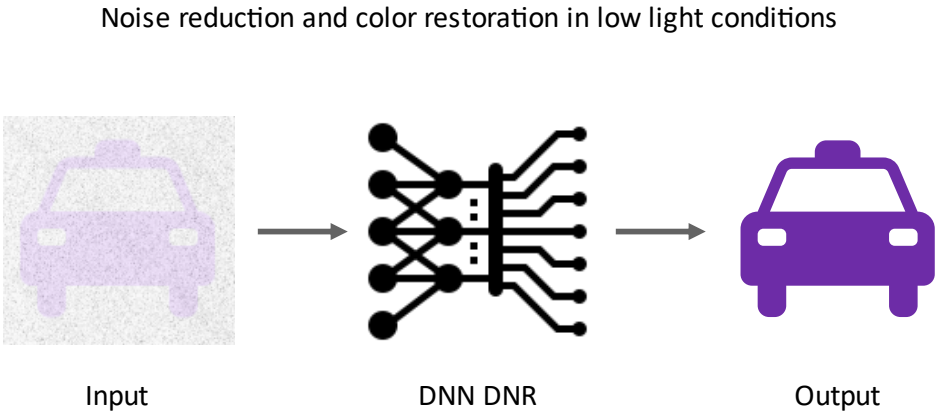
AI image signal processor

With rapid developments in AI, leveraging algorithms and neural processing units (NPUs) to enhance ISP processing capability becomes viable. An AI-ISP can automatically tune complex parameters via a deep neural network (DNN) to yield better images than a conventional ISP.

The most prominent contribution of AI-ISP is to improve digital noise reduction (DNR) in low-light conditions. Intelligent DNR aims to reduce noise on the image for fine details. In addition, it can improve codec efficiency for lower bandwidth consumption. This is because the encoder treats the visible noise as moving objects. Applying DNNs, AI-ISP can learn the distribution of noise and signal via many datasets. As a result, the algorithm can intelligently distinguish between noise and signal and between moving objects and a static scene, avoiding potential motion blur.

In summary, AI-ISP can significantly reduce the amount of time used for parameter tuning while enabling the camera to adapt to varying light conditions. With the help of AI-ISP, cameras can provide full-color images while operating 24/7.

Figure 10: Illustration of DNN DNR



Source: Omdia

Improved video transmission and security

Challenges faced by the industry

Nowadays, video cameras have become common in every area of life. The application of cameras is also going beyond safety and security as the Internet of Things (IoT) and digitization continues to transform every industry. The transmission, storage, and safety of video feeds is becoming a critical concern. Omdia estimates that approximately 114 exabytes (EB)[1] of hard drives used in video surveillance were sold in 2023[2]. This is driven by the following factors:

- **Increasing number of higher-resolution cameras shipped:** Almost 52.1% of network cameras shipped in 2023 had a resolution of more than 4MP[2].
- **Higher frame rate required:** Certain vertical markets such as transportation, sports, and casinos mandate a higher frame rate of 30fps or above.
- **Longer retention times:** Legislation and rules stipulate video retention for more than 30 days; for example, Dubai's Security Industry Regulatory Agency (SIRA) requires the retention of video for a minimum of 31 days[3].

These factors together pose great challenges to the bandwidth, storage capacity, and information security of network and system infrastructure. To address these challenges, the industry has been continuously applying new technologies and measures to improve video-encoding efficiency and network safety.

Video codecs

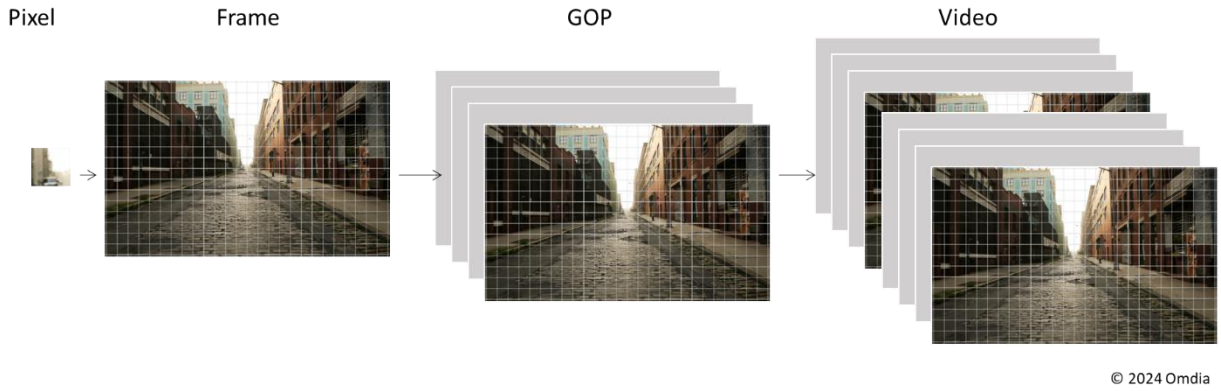
Video codec basics

A pixel is the fundamental unit of a frame/picture. A group of pictures (GOP) comprises several frames to be processed, and several GOPs constitute a video, which is measured in frames per second (fps). Neighboring pixels within a single frame and a large proportion of consecutive frames within a GOP are often similar or even the same. Transmitting every detail for each frame can be

[1] Omdia Enterprise and IP Storage Used in Video Surveillance Report – 2023 (August 2023)
[2] Omdia Video Surveillance and Analytics Report – 2023 (December 2023)
[3] Security Industry Regulatory Agency, Government of Dubai. (2020). Preventative Systems Manual.

wasteful. Compression is designed to remove redundant information from an image frame and GOP and therefore reduce the data file size and bit rate. There is a tradeoff between too much compression, which can reduce image quality, and too little compression, which can result in large file size and bit rate without additional useful information being gathered from the image.

Figure 11: Relationship of pixel, frame, GOP, and video



Source: Omdia

Compression processing is done on a network video surveillance camera before its images are transmitted over the network to a storage device. Nowadays, the industry offering is rapidly transitioning from the H.264 compression algorithm to H.265. The most common compression algorithm used for video surveillance in 2023 was H.265. It can reduce file sizes by around 40–50% compared with H.264. Omdia forecasts that 96.4% of network cameras shipped will adopt H.265, up from less than 10% in 2016[1].

Smart codec

With developments in AI algorithms and powerful edge-computing chips, a smart codec is introduced to video surveillance to further improve compression efficiency, reducing the burden of transmission and storage of video streams.

In a standard codec, three key parameters for encoding are usually set to fixed values:

- **Compression level of an image:** The entire frame is set to the same compression level without distinguishing between static background and dynamic foreground. Usually, users of video surveillance are only interested in moving objects rather than in static background in a video.

[1] Omdia Video Surveillance and Analytics Report – 2023 (August 2023)

- **Interval of GOP:** The full image frames are not sent consecutively with the standard H.265 because of the similarity of neighboring frames. Instead, an I-frame (intracoded frame or reference frame) with full image information is sent, followed by P-frames (predicted frame) with only small updates of the scene from prior frames. Each GOP starts with an instantaneous decoder refresh frame (IDR frame, a special type of I-frame) at the least amount of compression followed by multiple P-frames with high compression. In standard codec, the I-frame interval is fixed. Periodically sending I-frames is wasteful when there is no motion in the scene.

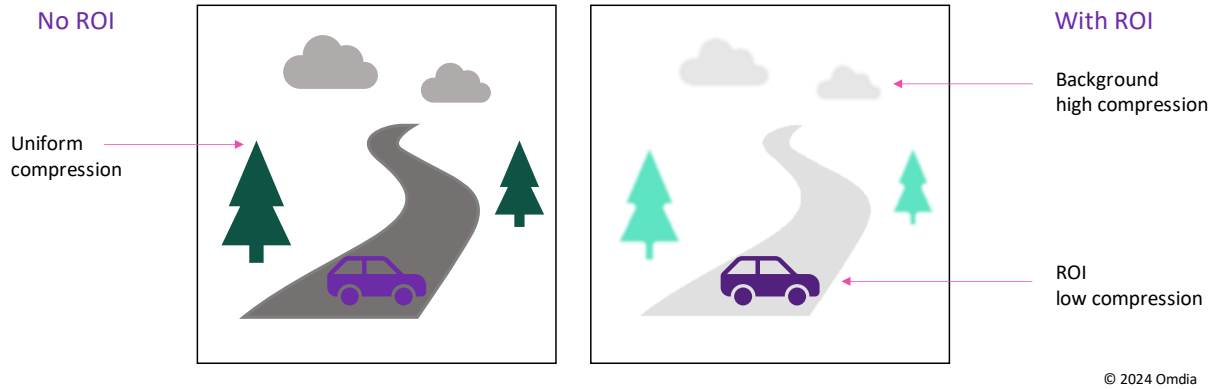
- **Frame rate:** Frame rate is used to measure the number of images within a specified period of a video, using measurement of frames per second. Because of the nature of the human eye, video at more than 10–12fps is considered to be coherent. In a video surveillance context, frame rate is often fixed at 15fps or higher depending on application scenario. Higher fps often translates into faster catching of a scene, for example, a moving vehicle.

Smart codecs intelligently adjust these parameters to achieve a lower transmission bit rate and a smaller video file size by constantly detecting surrounding conditions such as lighting, motion, and weather.

Dynamic region of interest

Users are usually interested in a specified area on video images, known as the region of interest (ROI). The clarity of the image in the key area can be higher than in other areas. Dynamic ROI can ensure effective observation over a specified area in the case of insufficient network bandwidth. Users can specify one or more areas on images as ROIs. The image quality in the ROIs can be different (higher or lower) from that in other areas. That is, the system performs near-lossless compression (high bit rate) on ROIs and lossy compression on non-ROIs (low bit rate), ensuring higher quality for reconstructed images and achieving a higher compression ratio. With the help of AI, ROIs can be automatically identified without human intervention.

Figure 12: Dynamic ROI



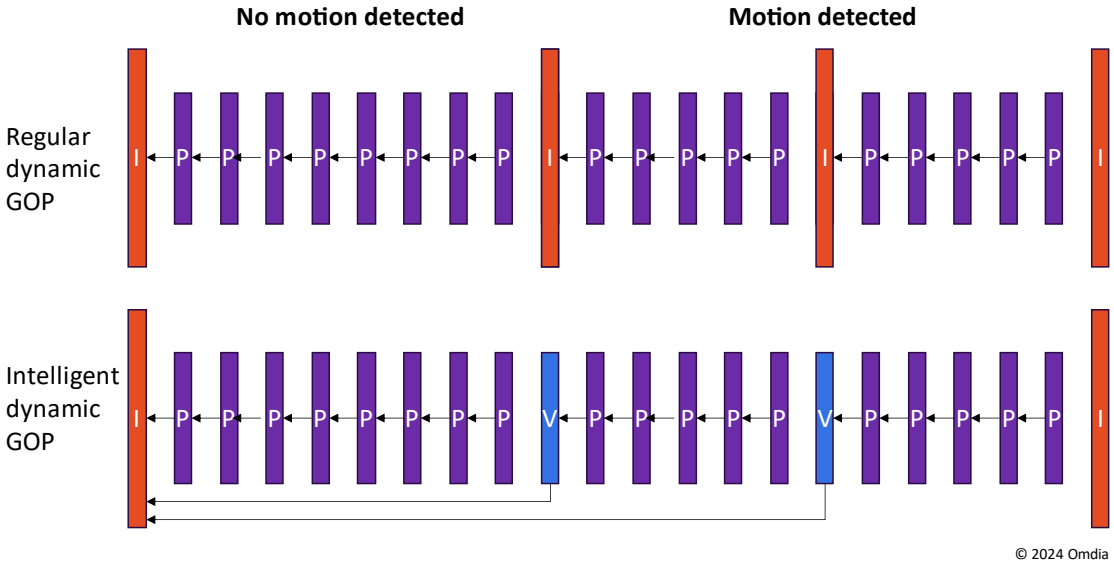
Source: Omdia

Dynamic GOP

The I-frame of a GOP contains a large amount of information and is processed with the least compression, whereas the subsequent P-frames contain very minimal information by referencing the previous frame. A smart codec dynamically adjusts the interval of I-frames based on the amount of motion and complexity in the video. An I-frame is inserted less frequently when the scene is static and more frequently if activity is detected. As a result, encoding and compression efficiency can be improved. However, if an I-frame is lost during transmission or incurs encoding errors, the subsequent P-frames are not able to encode correctly, causing loss of video for a period of time.

Intelligent frame reference and virtual I-frame techniques have been developed to solve this issue. The P-frame references the prior I-frame (long-term reference frame) and a virtual I-frame (short-term reference frame, essentially a P-frame). Therefore, when there is any motion in the scene, a virtual I-frame will be inserted instead of an I-frame, further reducing the bit rate. This approach can also better distinguish static portions of an image and moving objects, improving ROI compression.

Figure 13: Dynamic GOP



Source: Omdia

Adaptive variable bit rate control

Adaptive variable bit rate (VBR) control allows bit rate fluctuation during collection to ensure stable quality of encoded images.

The bit rate control algorithm detects the object status (moving or static) in the current scene, uses a higher bit rate for encoding when an object is moving, and decreases the bit rate when the object becomes static. Bit rate control is performed inside the encoder. The algorithm determines the

scene based on the amount of motion detected during encoding and then adjusts the bit rate control policy.

Unlike common VBR control, adaptive VBR control can effectively decrease the bit rate when an object is static and gradually increase the bit rate when an object is moving while ensuring the image quality. Adaptive VBR control features higher real-time performance, progressive bit rate control, and excellent image quality.

Video security

Risks faced with network cameras

As security solutions converge with IoT networks, the line between physical and cybersecurity continues to blur, and the gap between solutions for traditional security systems and for network infrastructure narrows. With growing numbers of network cameras becoming interconnected, the risks of cyber-physical attack also increase significantly. For example, South Korean police reported a hacker using an automatic program to gain access to more than 400,000 home security cameras and attempting to sell video footage online between August and November 2021.

Cameras are mainly vulnerable to the following risks:

- Hackers access the camera via the port for login authorization or tamper with the software and hardware of the camera.
- Hackers access the camera network to spread malware (including worms, spyware, ransomware, adware, and trojans) and acquire sensitive information.

Security protection measures

With rapid adoption of edge AI and cloud-based surveillance solutions, end users are becoming particularly wary and sensitive to cybersecurity threats, data protection, and privacy concerns. Solutions that emphasize cybersecurity with features such as encrypted surveillance feeds and secure data storage have seen a considerable increase in market demand.

Intrusion detection

The intrusion detection component provides the following functions: promiscuous network adapter detection, key file-tampering detection, unauthorized superuser detection, botnet detection, rootkit detection, and CoinMiner virus detection. This component performs security detection on the networks, files, users, and malicious programs of cameras. An intrusion detection system is installed on a camera to provide multilayer visualized intrusion detection capabilities. This helps detect hacker intrusion behavior in a timely manner and protect camera security.

Media security technology

With media security technology, the backend video security platform generates initial encryption information and sends the encrypted information to cameras. The cameras obtain the final private key based on negotiated encryption rules, encrypt the collected video data and final private key in combination mode, and send the encrypted video data to the security center. After receiving

encrypted streams, the platform decrypts the private key, preventing media information theft and enhancing media transmission security.

Digital watermark technology

During stream data output through video encoding, watermark information (which functions as protection information) related to the stream frame (including the current number of frame bytes, time, device MAC address, and SN) is stored in the stream data packet defined by users. The data packet is stored in disks or transmitted with the compressed stream.

AI-powered video processing

The development of AI and deep learning has continued to be a significant market driver within the video analytics industry, and the impact this technology will have on the analytics market is expected to continue to be a driver of growth and technological development for the video surveillance industry. The ability to convert unstructured video data into meaningful insight will move the video surveillance industry into the artificial intelligence of things (AIoT) world. This is expected to have implications for nearly all vertical markets, especially retail, transportation, city surveillance, and large-scale critical and national infrastructure.

In video analytics, deep learning is primarily used as a translation technology. Deep learning turns frames and pixels from video into metadata, turning an unstructured video into more structured data. Therefore, it helps understand what is in the video or image at very high levels of accuracy at very high frame rates. As such, it is a translation layer. Machine learning (ML) works at the metadata level, performing data analysis or statistical analysis of structured data to help with contextual understanding and situational awareness. Over the last few years, video analytics has increasingly become a productivity tool, and analyzing the contents of a video is nearly a fully solved problem. Omdia expects this to spur evolution in productivity tools, in that ML and data analytics will become increasingly important; all this is built on deep learning and the AI technology underpinning it.

There are mainly two phases involved in deep-learning development: training and inference.

Training is the process by which a DNN learns from a curated dataset and adjusts its parameters to minimize errors. It uses artificial neurons to process abstract data types such as images. It memorizes the rules and patterns of the data by looking at examples and learning from its mistakes.

Inference is making predictions from novel data via a model the DNN has learned from the training process. It bridges the gap between the data it was trained on and ambiguous information in the real world. It uses features and algorithms to score and categorize new data and produce actionable results.

Both training and inference are compute intensive and require high-performance hardware and software solutions. However, inference in general requires less compute power than training because the inference process is less compute intensive. In the video surveillance context, inference is a major concern to end users.

The surge in AI video analytics adoption in the past decade has been closely linked with developments in hardware. The development of general-purpose computing on graphics processing units (GPUs) initially spurred this trend; the trend itself then drove the development of very powerful server GPUs and their integration into high-performance computing systems, largely focused on cloud and data center deployments. With the spread of AI applications into mobile and embedded markets, the demand emerged for hardware that could perform neural network inference at the edge in the context of power-, price-, and area-constrained SoCs. At the same time,

the increasingly advanced SoCs developed for the enormous smartphone market began to penetrate the security market. These devices are general-purpose computing platforms, optimized for constrained, embedded form factors.

Figure 14: A hierarchy of artificial intelligence domains

Artificial intelligence

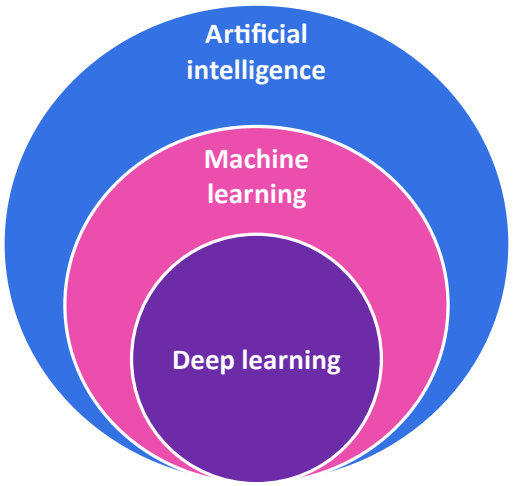
The body of science that studies how to enable machines to perform independent problem -solving, inference, learning, knowledge representation, and decision -making

Machine learning

An application or set of algorithms that allow machines to automatically find and learn patterns by feeding them data without explicit programming.

Deep learning

The use of neural networks with multiple layers to analyze and interpret data for training and inferencing.



Source: Omdia

Embedded AI camera

In an increasingly AIoT world, the ability to process analytics at the edge is fundamental to developing an intelligent video surveillance market. As a result, AI cameras are now becoming standard in the high-end market. As the cost of chipsets and algorithms falls, AI cameras are transitioning to midrange and low-tier markets.

Definition of AI camera

Omdia defines an AI camera as any network camera that is specifically designed to fully process deep-learning analytics on the camera, a deep-learning video analytic that is based on the use of rules-based analytics and deep-learning analytics in conjunction to detect, identify, and classify objects in a video stream. Typically, an AI camera will include some form of advanced image processor (SoC, DLPu) with above 1 TOPS computing power. Deep-learning analytics performances should not be limited in any manner to allow processing on the camera.

Omdia estimates that 18% of network cameras shipped in 2023 were AI cameras[1]. A large proportion of the midrange market is expected to have been penetrated by 2027, with AI cameras accounting for 32% of all network cameras shipped globally.

AI in video analytics

AI technologies have been evolving for years, but recent advancements in foundation models and generative AI have significantly enhanced and showcased AI's enormous potential. The killer applications are those such as ChatGPT, Stable Diffusion, and DALL-E that are reshaping industries and the consumer experience. Although several security vendors have announced generative AI specific to video surveillance, there is still a long way to go for wider adoption of multimodal foundation models across the security industry.

Given their complementary nature, traditional AI based on machine learning and deep leaning will still be required in video surveillance. Generative AI, with its holistic understanding, can contribute to overall scene comprehension, but traditional AI will still play a role in capturing local features and spatial hierarchies.

Types of AI video analytics at the edge

At present, there are a growing number of analytics solutions available in the market because demand from the security industry is very fragmented. Some are particularly deployed in the data center or cloud, while others are particularly effective at the edge camera because of application scenarios. For example, an automatic number plate recognition (ANPR) camera can detect a speeding vehicle, capture the license plate, and identify the type, model, and color of the vehicle. Many tasks and algorithms are involved in this scenario.

In general, the most frequently used basic tasks in video analytics include the following:

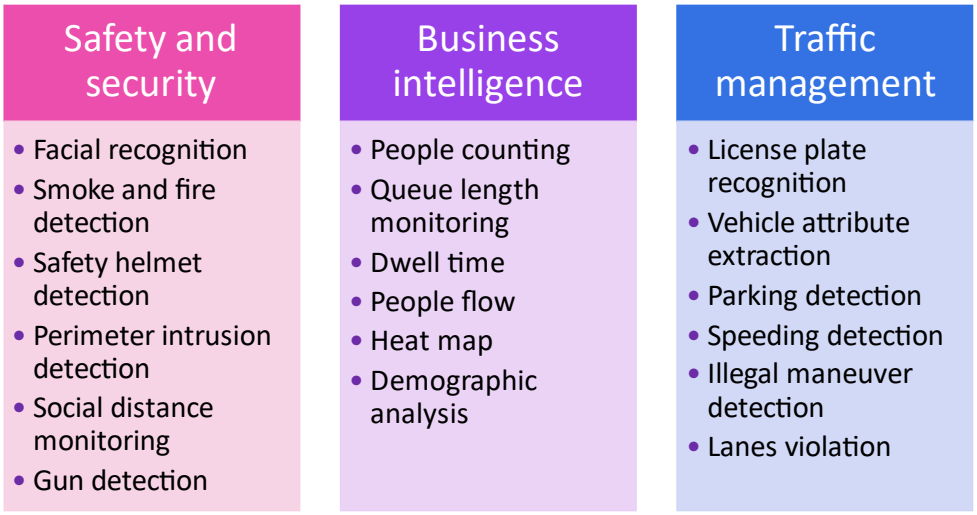
- **Image classification** is the most basic task, which classifies an image input to a set of predefined categories (e.g., car, animal, person, inanimate objects).
- **Object detection** locates and classifies an object in an image (drawing a bounding box tagged with the object classification). It takes into account the location, size, and rectangular shape of the object.
- **Image segmentation** includes instance segmentation and semantic segmentation aiming to outline the shape of an object in the image instead of using a bounding box. Semantic segmentation takes a step further than instance segmentation by detecting at pixel level. Some manufacturers use this technology to detect ROI in a video for a smart codec.

[1] Omdia Video Surveillance and Analytics Report – 2023 (August 2023)

- **Object tracking** tracks the motion and behavior of an object using successive video frames as inputs. This can be used for event and behavior analysis and for situational awareness (e.g., tripwire, loitering, fighting, people counting, heatmap).

AI video analytics algorithms can also further be categorized by application scenarios (**Figure 15)**.

Figure 15: AI video analytics, example use cases



Source: Omdia

Drivers of adoption

As diverse edge video analytics algorithms are developed, AI cameras are seen in more and more use cases. There are also many other reasons why AI cameras are gaining fast adoption in the industry:

- **Cost-efficiency:** Streaming large amounts of video to be processed on server-based GPUs can be significantly more expensive than processing it locally on the camera itself. Processing the analytics on the camera means everything does not have to be uploaded, and expensive infrastructure to process the video is not needed.

- **Latency and system redundancy:** On-camera analytics means data is not sent back to the server, so there is no waiting for a response. This is very important in applications that need quick feedback loops, such as facial-recognition access-control gates. Additionally, even if connection is lost the analysis can still be run.

- **Bandwidth concern:** As the industry increasingly adopts higher-resolution cameras, edge analytics allow users to do analytics on the full-resolution raw image. Processing at higher resolution allows greater accuracy and reliability because objects can be detected at a greater

distance. Typically, in a server-based architecture, this video would need to be uploaded. For that there are two choices: either use a lot of bandwidth and stream the full 4K video at high resolution or compress it, which means losing some resolution and accuracy.

- **Privacy and data sovereignty:** When it is unwise or forbidden by regulation to hold sensitive data centrally, it must be left on the device where it was created. Because analytics are increasingly being used for privacy-related algorithms such as facial recognition, it has been suggested that on-camera processing is more secure.

Deployment of embedded AI

Deep-learning video analytics promises unprecedented performance but also requires significantly more computational power than many traditional video analytics. However, the development of deep-learning models is a complex, time- and resource-consuming process. After training at the data center or cloud, a deep-learning model needs to be deployed on an edge device (e.g., a camera) for applications in the real world.

Model optimization

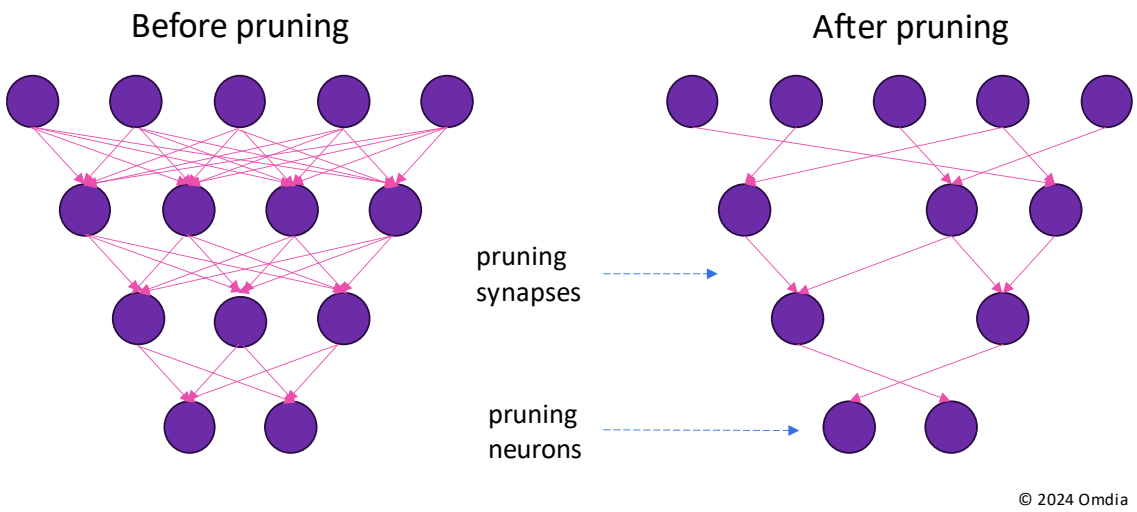
Deep-learning models have many parameters and require a significant number of samples to train. Model training is almost always carried out at the data center or cloud with powerful AI accelerators such as GPUs. However, the large size and high compute consumption of the original model makes it hard to deploy at the edge. To overcome this challenge, model compression techniques have been developed to improve the model's efficiency without reducing accuracy.

Model compression is the process of improving the performance, efficiency, and resource utilization of deep-learning models, especially for resource-constrained devices such as cameras. It can be achieved through various techniques aimed at reducing the model's size, inference time, and memory footprint while maintaining or improving its accuracy. Some of the commonly used techniques for model optimization are listed below.

Pruning

Pruning is the technique of removing some of the weights or neurons of a neural network that are not essential for the model's accuracy. This can reduce the model size, inference speed, and energy consumption.

Figure 16: Pruning improves model efficiency

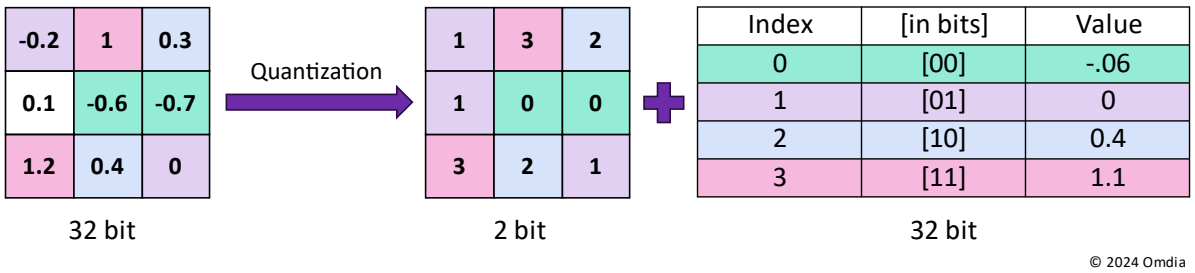


Source: Omdia

Quantization

Quantization is the technique of reducing the precision or bit width of the model's weights and activations from floating-point numbers to integers or binary values. This can also reduce the model size, inference speed, and energy consumption and enable hardware acceleration.

Figure 17: Quantization

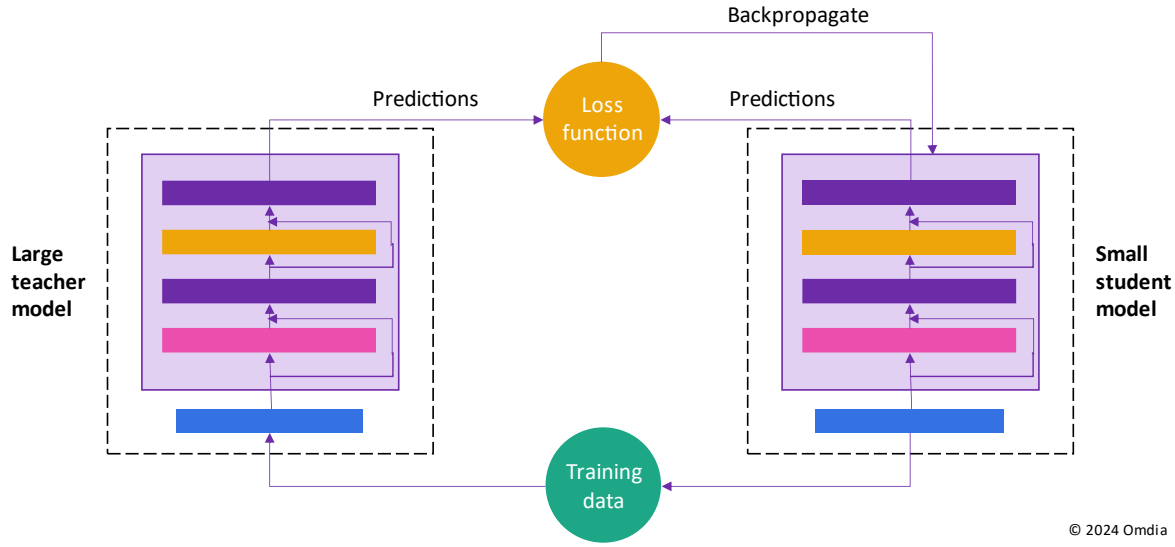


Source: Omdia

Knowledge distillation

Knowledge distillation is the technique of transferring knowledge or information from a large or complex teacher model to a smaller or simpler student model. This can involve using the teacher model's outputs or intermediate representations as additional supervision for the student model.

Figure 18: Knowledge distillation



Source: Omdia

Compressed models often perform similarly to the original but use a fraction of the computational resources. Some of the benefits of model compression are:

- **Conserves storage space:** Model compression can reduce the size of the model, making it more suitable for edge devices with limited memory.
- **Reduces computational demand:** Model compression can reduce the complexity of the model, making it more efficient and faster to execute.
- **Reduces energy consumption:** Model compression can reduce the power consumption of the model, making it more sustainable and scalable.
- **Improves model performance:** Model compression can improve the accuracy and robustness of the model, making it more reliable and stable.

Building an open ecosystem for embedded AI

With the rapid development of AI software and hardware, many deep-learning analytics algorithms have become available for edge AI cameras in the market, each solving a specific problem. A conventional AI camera is normally prepackaged with a few fixed algorithms to meet the requirements of a predefined scenario. For example, an AI camera deployed at a retail store may be preloaded with algorithms for people counting, dwell times, and heatmap for business intelligence

and intrusion detection for security and safety. The functionality of the intelligence is rarely changed once the AI camera is installed because of the fixed software and hardware specifications.

The industry has seen a growing demand for flexibility and diversity of AI applications. For example, the AI algorithms for business intelligence also meet new requirements such as demographic analysis and shelf management. The traditional approach will require either more AI cameras or backend appliances with various algorithms to meet the end users' increasing demands. There is a need for AI cameras to be adaptable to the requirements of diverse scenarios.

A true AI camera embedded with a powerful AI accelerator allows on-demand deployment and online upgrade of algorithms, providing continuous improvements and adaptability to evolving security operation needs. This flexibility enables AI cameras to keep pace with technological advancements and deliver optimal performance. The development of camera-first ecosystems with dedicated device operating systems, applications platforms, and developer community has been increasingly critical to wider adoption of AI cameras.

Open operating system

Having an open standard and operating system specifically for security and IoT devices means different hardware and software vendors can all use the same platform. In theory, this enables easy configuration of best-of-breed components, with more flexibility for changes and tighter cybersecurity controls than standard interoperability protocols allow. This could also mean system configurations and camera applications are able to change more often due to an easier delivery model provided by the ecosystem, not constrained by the current more manual and time-consuming processes.

Developer community

While hardware and operating systems are critical elements, an established developer community that provides a software stack is also essential for a full solution. Particularly within the video analytics space, the ability to give developers the tool sets to optimize performance and accelerate their time to solution is important. Fundamentally, an SDK is meant to help developers and video analytics providers create rich video analytics pipelines. SDK's do a lot of work for the developer in terms of memory management and tying into underlying API's, which trains a neural network in industry standard frameworks, but then optimizes them for the target hardware.

App store for AI camera

Such AI cameras require a marketplace or app platform at their back. This is a concept derived from the smartphone, meaning an AI camera can install different algorithms and applications. The development of edge processing and AI cameras, distributed processing architecture advances, software developments, and the change in sales channels from pre- to post-installation sales have all driven the development of app platforms that enable the interchangeability of analytics depending on specific needs and requirements. These platforms are intended to offer increased flexibility and deployment architecture that suit the end user's needs. At the same time their "openness" is designed to prevent lock-in.

Summary and conclusions

As the AIoT concept gains traction, AI cameras are playing an important role in sensing the real-world environment, enhancing the efficiency of safety, security, and business operations. A true AI camera should not only capture crystal clear images in any lighting conditions but should also stream high-quality video feeds to backends securely without taking much storage and bandwidth. There is also a need for AI cameras to adapt to growing end-user demand and keep up with technological developments.

The latest developments seen in AI cameras include:

- **Enhanced video quality:** An increasing number of AI cameras are now adopting higher resolution, larger aperture, lower noise, advanced image sensors for greater brightness, and AI-enabled ISPs with noise reduction to provide full-color images in low-light conditions.
- **Improved video transmission and security:** AI cameras adopting smart codec can save storage and bandwidth while maintaining image quality. Video image protection measures including intrusion detection, data encryption, and watermarking can improve the safety and robustness of the surveillance network.
- **Advanced video processing powered by AI:** AI cameras should be able to adapt to the growing demand for flexibility and diversity of deep-learning algorithms. This can be achieved with an open application platform that supports on-demand deployment and online upgrade of algorithms.
- **Open ecosystem built surrounding AI camera:** An established ecosystem for AI cameras can be a disruptive force for video surveillance. Once different IoT devices, AI cameras and algorithms are on the same platform and operating system, greater device interconnectivity and convergence can be achieved.

Appendix

Further reading

Enterprise and IP Storage Used in Video Surveillance Report – 2023 (December 2023)

Video Surveillance and Analytics Report – 2023 (August 2023)

Security Industry Regulatory Agency, Government of Dubai. (2020). Preventative Systems Manual

Author

Tommy Zhu
Senior Analyst, Physical Security
customersuccess@omdia.com

Get in touch

www.omdia.com
customesuccess@omdia.com

Omdia consulting

Omdia is a market-leading data, research, and consulting business focused on helping digital service providers, technology companies, and enterprise decision makers thrive in the connected digital economy. Through our global base of analysts, we offer expert analysis and strategic insight across the IT, telecoms, and media industries.

We create business advantage for our customers by providing actionable insight to support business planning, product development, and go-to-market initiatives.

Our unique combination of authoritative data, market analysis, and vertical industry expertise is designed to empower decision-making, helping our clients profit from new technologies and capitalize on evolving business models.

Omdia is part of Informa Tech, a B2B information services business serving the technology, media, and telecoms sector. The Informa group is listed on the London Stock Exchange.

We hope that this analysis will help you make informed and imaginative business decisions. If you have further requirements, Omdia's consulting team may be able to help your company identify future trends and opportunities.

Copyright notice and disclaimer

The Omdia research, data, and information referenced herein (the "Omdia Materials") are the copyrighted property of Informa Tech and its subsidiaries or affiliates (together "Informa Tech") or its third-party data providers and represent data, research, opinions, or viewpoints published by Informa Tech and are not representations of fact.

The Omdia Materials reflect information and opinions from the original publication date and not from the date of this document. The information and opinions expressed in the Omdia Materials are subject to change without notice, and Informa Tech does not have any duty or responsibility to update the Omdia Materials or this publication as a result.

Omdia Materials are delivered on an "as-is" and "as-available" basis. No representation or warranty, express or implied, is made as to the fairness, accuracy, completeness, or correctness of the information, opinions, and conclusions contained in Omdia Materials.

To the maximum extent permitted by law, Informa Tech and its affiliates, officers, directors, employees, agents, and third-party data providers disclaim any liability (including, without limitation, any liability arising from fault or negligence) as to the accuracy or completeness or use of the Omdia Materials. Informa Tech will not, under any circumstance whatsoever, be liable for any trading, investment, commercial, or other decisions based on or made in reliance of the Omdia Materials.