Hanwha Vision

White Paper

Edge AI Cameras

November 2025

Contents

1. **Introduction**
2. **Object Detection Technology**
   - 2.1. Advantage of Deep Learning-Based Architectures
   - 2.2. Strategy for Training and Enhancing Object Detection Models
   - 2.3. Optimization for Edge AI (Model Lightweighting)
3. **Object Classification Technology**
   - 3.1. Advantages of Deep Learning-Based Classification Architectures
   - 3.2. Technical Challenges in Object Classification
   - 3.3. Hanwha Vision's Strategy for Enhancing Attribute Classification Models
4. **BestShot**
   - 4.1. Technical Implementation Mechanism and Utilization
5. **System Integration and AI Application Areas**
6. **Conclusion**

1. Introduction

Conventional video surveillance systems suffered from inherent limitations due to their reliance on human operators, who were required to simultaneously monitor multiple CCTV feeds. This dependency often led to "human error," where critical incidents or subtle situational changes were missed, resulting in low overall operational efficiency.

To overcome these challenges, AI technology has been introduced into the realm of video surveillance. This transition has enabled not only the accurate detection of key objects such as people and vehicles but also the analysis of object information through attribute extraction and Automatic Number Plate Recognition (ANPR). Furthermore, AI technology that accurately classifies objects and selectively presents only meaningful events to operators has drastically improved control room efficiency and realized effective monitoring for large-scale systems.

This white paper aims to provide an in-depth explanation of the core intelligent video analysis technology embedded in Hanwha Vision AI cameras. It covers system architecture, model training and enhancement strategies, and optimization technologies designed to enhance operational efficiency in the field. The goal is to help users gain a deeper understanding of Hanwha Vision’s AI camera capabilities and secure optimal performance in their operations.

2. Object Detection Technology

Object detection is one of the most fundamental and critical technologies in computer vision. **Hanwha Vision AI cameras adopt a deep learning-based architecture** to detect pre-defined objects of interest—such as people, faces, vehicles, and license plates—in real time and visualize them by accurately marking their areas with bounding boxes.

2.1. Advantage of Deep Learning-Based Architectures

In the past, traditional object detection functions were developed based on computer vision algorithms utilizing manually defined features, such as Background Subtraction, Haar Cascades, and HOG (Histogram of Oriented Gradients). While these methods offered relatively fast processing speeds, their accuracy was limited due to high sensitivity to environmental factors, such as lighting changes, complex backgrounds, and partial occlusion. Additionally, a significant drawback was the lack of flexibility, as algorithms had to be redesigned to recognize new object types.

In contrast, deep learning-based object detection utilizes artificial neural networks trained on massive amounts of image data to automatically extract and define the features necessary for recognizing objects. This approach allows the system to learn shapes, textures, and contextual information on its own without human design, enabling more sophisticated and flexible detection.

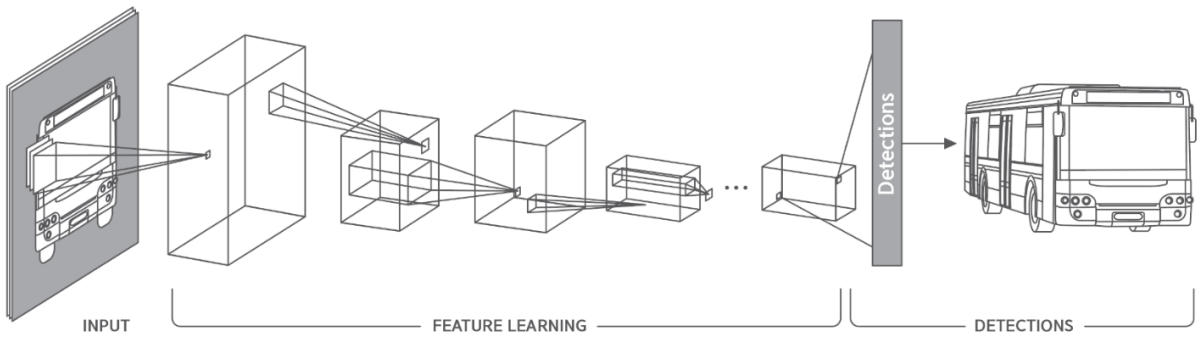

Figure 1: Deep Learning-Based Object Detection Architecture

Hanwha Vision secures a technological edge by adopting this deep learning-based architecture and optimizing it for camera implementation:

- **High Detection Performance:** The CCTV environment encompasses a diverse range of objects, from small, distant targets to large, nearby ones. Hanwha Vision utilizes optimized algorithms to detect objects of varying sizes simultaneously. This ensures accurate detection of small distant objects while

providing sophisticated detection capabilities to separate and recognize overlapping or occluded objects in complex, crowded areas.

- **Low False Alarm Rate:** The architecture minimizes false alarms by ensuring stable detection accuracy even under unpredictable environmental conditions, such as varying lighting, weather, and camera angles.
- **Real-Time Inference Stability:** The lightweight model structure and optimized computational efficiency of Hanwha Vision's AI allow for seamless, ultra-fast analysis directly on the camera. This supports high-reliability real-time video analysis with extremely low latency, enabling immediate alarms and tracking responses when events occur.

2.2. Strategy for Training and Enhancing Object Detection Models

The performance enhancement of Hanwha Vision's AI object detection is achieved through the following four-stage training pipeline.

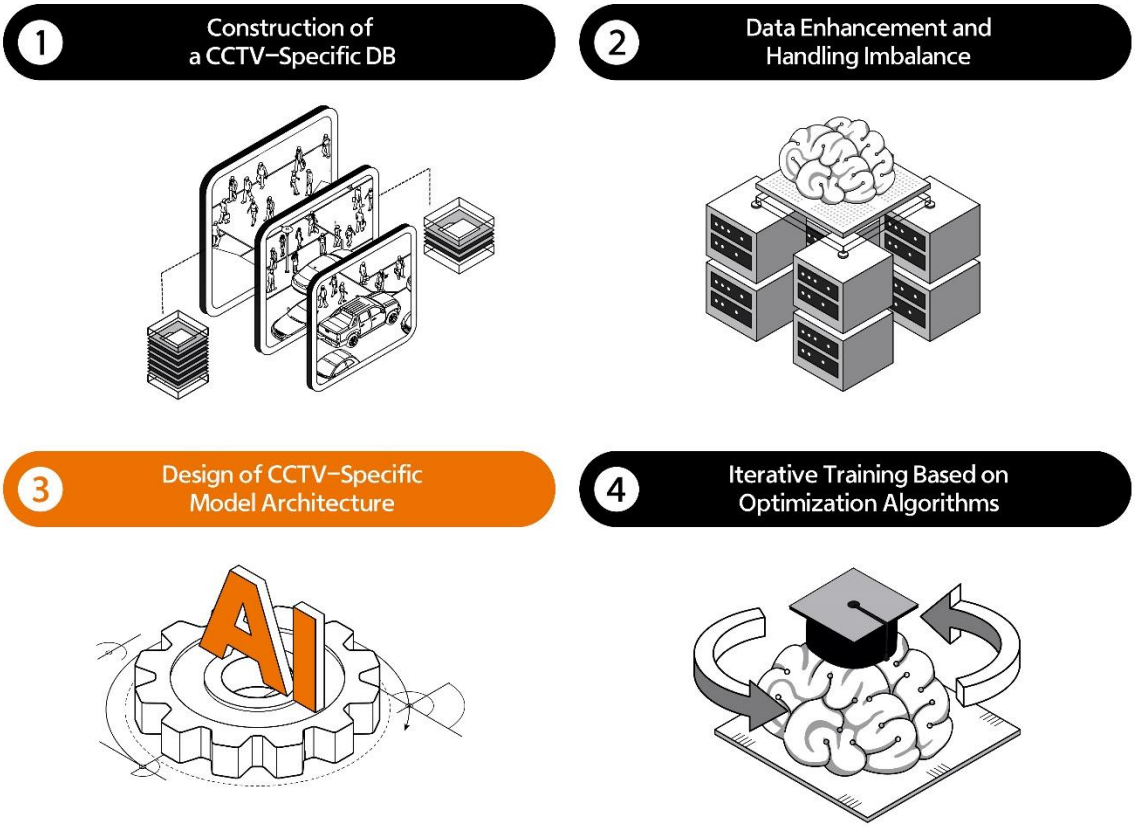

Figure 2: Hanwha Vision Object Detection Model Training Pipeline

- **Stage 1: Construction of a CCTV-Specific Database (DB)**

  The superiority of a deep learning model's performance is directly attributed to the quantity and diversity of its training data. To maximize accuracy, Hanwha Vision collects high-quality CCTV data designed to reflect actual operating environments systematically. This dataset covers all potential extreme conditions found in real-world scenarios, including indoor/outdoor settings, day/night cycles, backlighting, severe weather, and high congestion. The collected footage undergoes an annotation[1] process—specifying bounding boxes and labels—to create a final dataset suitable for AI model training.

- **Stage 2: Data Enhancement and Handling Imbalance**

  To maximize accuracy in CCTV environments, two major data imbalance issues must be overcome:

  - **Class Imbalance**: Bridging the frequency gap between common objects (e.g., cars) and rare objects (e.g., bicycles) to ensure balanced learning for all object classes.
  - **Edge Case**[2] **Accuracy**: Securing sufficient data for low-frequency conditions/states—such as figures in extremely low light or people in crouching positions—unlike standard pedestrian data. This prevents performance degradation in unpredictable situations.

  To solve these issues, Hanwha Vision strengthens the collection of edge cases and refines the training dataset by considering class balance. For edge cases, the weight is intentionally adjusted, or Data Augmentation[3] techniques are applied to maximize learning effectiveness. Customer VoC is analyzed to reflect diverse real-world cases, and data closely aligned with actual usage conditions is selectively collected.

  During this process, sensitive personal information such as faces and license plates is encrypted or blurred to comply with privacy regulations.

- **Stage 3: Design of CCTV-Specific Model Architecture**

  Based on this data, Hanwha Vision has designed a dedicated model architecture that reflects the specific characteristics of the CCTV environment (objects of various sizes and aspect ratios, real-time requirements) to maximize training efficiency and performance. This architecture is optimized to

[1] Annotation is the process of labeling the ground truth information that a model must learn onto the data. This involves drawing bounding boxes around objects of interest in the video, labeling the object's type and attributes, and transforming the data into a format that the model can use for training.

[2] Edge Case refers to boundary conditions that, while low in frequency, can lead to AI model prediction failures.

[3] Augmentation is a technique that increases the quantitative and qualitative diversity of training data by applying transformations—such as geometric deformations and changes to brightness and saturation—to the original images.

simultaneously achieve improved generalization capabilities for complex objects and minimized inference latency.

As a result, it provides balanced detection precision—from small distant license plates to large nearby vehicles—while maintaining high FPS (Frames Per Second).

- **Stage 4: Iterative Training Based on Optimization Algorithms**

  Once the specialized architecture (Stage 3) and the enhanced training dataset (Stages 1 & 2) are ready, the AI model enters the full training phase. The model learns to predict object locations and types based on the high-quality dataset, undergoing hundreds of iterative training sessions to minimize the loss between prediction results and actual ground truths. Through this deep iterative learning, the model acquires the ability to detect objects with consistent high accuracy across various environmental conditions.

2.3. Optimization for Edge AI (Model Lightweighting)

The deep learning-based object detection models described above possess complex structures and high accuracy, which traditionally require significant computation and memory, making them difficult to operate directly on a camera. These constraints often lead to slower inference speeds and long latency, which directly impact system performance in CCTV environments where real-time analysis is essential.

To overcome these fundamental limitations, Hanwha Vision applies an optimal combination of lightweighting technologies such as Quantization (QAT/PTQ)[4], Knowledge Distillation[5], and Pruning[6], tailored for each SoC (System on a Chip). These technologies are designed to reduce model size and computational load while minimizing accuracy loss, enabling the stable operation of original large-scale models in an Edge environment.

[4] Quantization reduces model size and computational load by converting the model's 32-bit floating-point weights into 8-bit integer format. QAT (Quantization-Aware Training) minimizes accuracy loss by incorporating quantization during the training process. PTQ (Post-Training Quantization) quantizes the model after the training phase is complete.

[5] Knowledge Distillation is a process that transfers the predictive knowledge and decision-making expertise from a large, computation-heavy model to a lightweight model. This enables the lightweight model to boost its computational efficiency without compromising performance.

[6] Pruning reduces computational complexity and simplifies the model structure by eliminating unnecessary weight connections within the network.

3. Object Classification Technology

Object classification technology began with the function of identifying objects within a video into pre-defined classes or categories. Recently, with the advancement of AI, this technology has evolved beyond simple classification to a level where complex and detailed Attribute Information can be precisely extracted.

Hanwha Vision AI cameras utilize this technology to extract and classify detailed attribute information for people, faces, and vehicles in real time, as shown in Table 1 below.

Detected Objects	Attribute Category	Supported Attribute Items
Person	Gender	Female, Male
	Upper clothes color	(Up to 2 colors simultaneously)
	Lower clothes color	
	Bag	Carries a bag, Does not carry a bag
	Directions	(8 cardinal and intercardinal directions)
Face	Gender	Female, Male
	Age	Young, Adult, Senior
	Glasses	Wearing glasses, Not wearing glasses
	Face mask	Wearing mask, Not wearing mask
Vehicle	Type	Car, bus, truck, motorcycle, bicycle
	Color	(Up to 2 colors simultaneously)
	Directions	(8 cardinal and intercardinal directions)
License plate	License plate	Presence of a license plate

Table 1. Detailed Attributes Extracted by Hanwha Vision AI Cameras

This precise attribute classification automatically extracts meaningful information from video, contributing to improved operational efficiency in control systems and advancing video surveillance from simple monitoring roles into intelligent systems.

3.1. Advantages of Deep Learning-Based Classification Architectures

Deep learning technology has fundamentally improved the accuracy and expressiveness of object classification, enabling the precise recognition of detailed

object attributes. Previous machine learning methods, based on manually defined features like HOG, Haar, or color histograms, relied on rules to classify simple information like color or shape. These methods were vulnerable to varied environmental conditions (lighting changes, occlusion, low resolution) and struggled to add new attributes or handle complex scenarios.

In contrast, AI-powered object classification effectively resolves these limitations, significantly boosting video analysis accuracy and efficiency. Deep learning models integrate learning across shape, color, texture, lighting, and contextual information to deliver stable attribute classification performance across diverse environments. They secure the following technical advantages:

- **Operational Stability in Varied Environments**: The model delivers high accuracy regardless of diverse conditions such as changing lighting, partial occlusion, or low resolution, ensuring consistent performance in real-world scenarios.
- **Metadata for Search**: Extracted attribute data serves as high-dimensional metadata, enhancing search efficiency in control systems. Operators can quickly search for and track objects matching specific conditions amidst vast amounts of video data.
- **Business Intelligence (BI) Enablement**: Facial attribute information and vehicle type statistics are utilized as valuable data for BI analysis, such as access control and customer demographics, supporting optimized decision-making for operations.

3.2. Technical Challenges in Object Classification

To achieve the high-reliability classification performance required in actual CCTV environments, three major technical challenges that hinder model generalization and fairness must be overcome:

- **Fine-grained Class Classification**

  Classification becomes difficult when the difference between classes is subtle. For example, distinguishing between a white vehicle and a gray vehicle—where the difference is minute saturation and brightness—requires fine-grained classification. This demands a significantly higher level of expressiveness and precise feature learning than general classification.

- **Class Imbalance and Bias**

  Class imbalance occurs due to a disparity in the number of samples between classes in the training dataset. Over-training on majority classes often leads to under-training on minority classes, causing the model’s prediction performance to be biased toward the dominant classes. For example, if 'Female' comprises 95% of the data and 'Male' only 5%, the AI model is highly likely to predict 'Female' more frequently.

- **Prediction Bias**

  Prediction bias extends beyond simple data imbalance, stemming from environmental factors, algorithm design, and social elements. This bias causes the model to over-rely on specific patterns, severely compromising fairness and reliability. If training data is skewed toward specific conditions, the AI will show higher accuracy for those conditions. Conversely, for conditions that are under-represented or represent minority conditions, prediction errors increase. This leads to misclassification bias against specific groups and can result in discriminatory performance.

3.3. Hanwha Vision's Strategy for Enhancing Attribute Classification Models

To overcome the limitations outlined above and ensure the highest level of reliability and fair performance in real-world environments, Hanwha Vision implements a multi-faceted model enhancement strategy.

To ensure generalization across diverse CCTV environments, Hanwha Vision's classification models are built upon a domain-specific data collection and verification strategy that accurately reflects actual operating conditions (indoor/outdoor, day/night, backlit, weather changes, crowd density).

To improve fine-grained classification performance between similar classes (e.g., white ↔ gray, bicycle ↔ motorbike), models are designed to make accurate judgments even when classification boundaries are ambiguous. We also continuously manage learning data quality through precise annotation. Furthermore, we increase responsiveness to edge cases frequently occurring in CCTV environments—such as partial occlusion, mask-wearing, motion blur, and frame boundaries—through custom augmentation techniques[7].

Regarding prediction bias, we strengthen model fairness and reliability by incorporating diverse environmental and social factors into the learning process alongside class balancing. This involves introducing fairness evaluation metrics and utilizing bias detection and correction algorithms to prevent over-reliance on specific conditions or groups.

Through this comprehensive approach, Hanwha Vision builds object classification models that possess the reliability and fairness demanded by actual CCTV operations, moving beyond merely achieving high theoretical accuracy.

[7] Custom Augmentation refers to intentionally simulating and adding scenarios to the training data that are known to specifically and frequently occur in CCTV environments, thereby addressing issues that typically degrade model performance.

4. BestShot

BestShot is a feature that encapsulates Hanwha Vision's AI object detection and tracking technologies. It accurately locates an object of interest (Person, Face, Vehicle, License Plate), continuously tracks the object, and selects the single best representative image for the user. Instead of reviewing entire video footage, users can instantly grasp the object's features through this high-quality BestShot, using it for rapid search and analysis to maximize operational efficiency.

For example, the BestShot function seamlessly tracks an approaching vehicle and extracts a representative image captured at the moment when its unique attributes—such as license plate, color, detailed type, and direction of movement—are clearest and largest. For people, it selectively saves the moment the face is most clearly visible at close range, which can be instantly used for accurate suspect identification and monitoring during an incident.

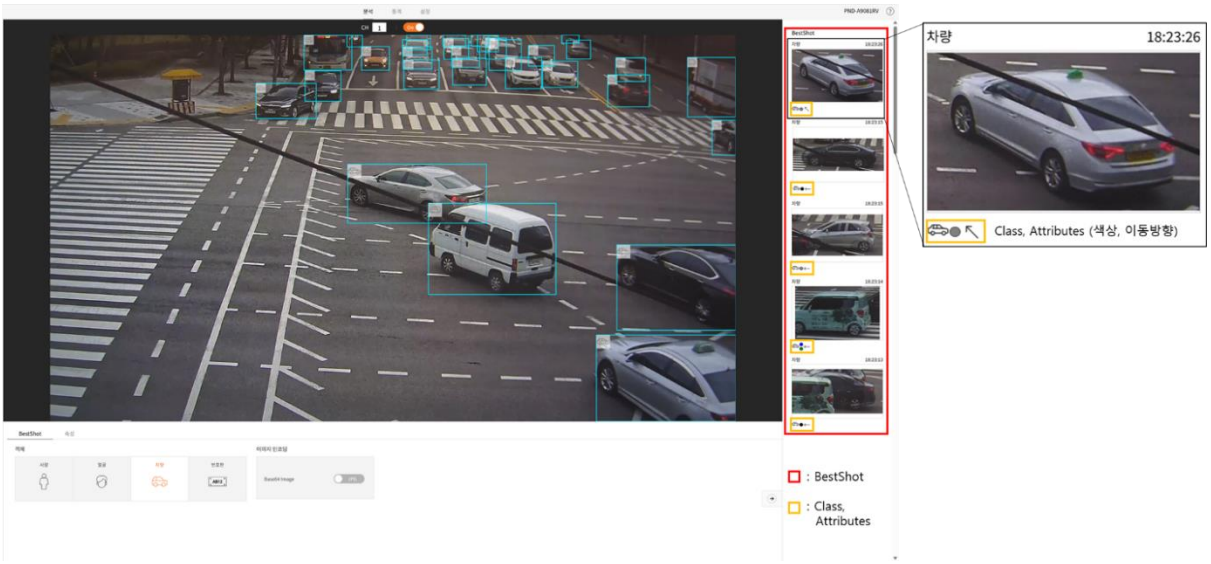

Figure 3. Hanwha Vision AI Camera BestShot Screen

Note that acquiring such high-quality evidence must comply with privacy laws and site operation policies, requiring masking or anonymization. Hanwha Vision's Edge AI technology supports Privacy Masking for this compliance.

Users can configure BestShot options for People, Faces, Vehicles, and License Plates via the [Analysis] > [BestShot & Attribute] settings tab in the camera web viewer.

4.1. Technical Implementation Mechanism and Utilization

- **High-Quality Selection Algorithm**: An optimal image selection algorithm comprehensively evaluates multiple factors, including object size changes, detection accuracy, degree of occlusion, and clarity, to select the largest, clearest, and most distinct BestShot.

- **Ensuring Continuity**: Reliable tracking technology prevents the generation of duplicate BestShots for the same object and minimizes missed objects, ensuring continuity throughout the tracking process.
- **Metadata Integration and Utilization**: Extracted BestShot images are automatically linked with timestamps and detailed attribute metadata at up to 4K resolution. This high-quality data source is then utilized for subsequent analysis and deep searches, such as license plate recognition and similarity searches.

5. System Integration and AI Application Areas

The Hanwha Vision AI-based object detection, classification, and BestShot technologies described so far are integrated into various intelligent operational functions of the camera to maximize user control efficiency.

AI Function	Description
Analytics	Based on AI object detection, this triggers intelligent event alarms for virtual lines and areas, such as Crossing, Intrusion, and Loitering. This supports rapid response by operators during critical situations.
Statistics	Provides statistical data, including People Counting, Queue Management, and Heatmaps, based on detected objects. This is utilized for BI to support operational optimization decisions.
Object Auto Tracking	Automatically and continuously tracks the movement of detected objects (People/Vehicles) within the video, helping operators intensively monitor specific targets without losing their path.
Privacy Masking	The AI-based DPM (Dynamic Privacy Masking) function automatically applies masking to objects in the video, allowing users to enhance video surveillance while complying with privacy regulations.
Video Compression	Hanwha Vision’s WiseStream is a technology where AI efficiently compresses unnecessary data in the video. It maintains high quality for important objects like people or vehicles while applying high compression rates to non-interest areas, effectively reducing data size.

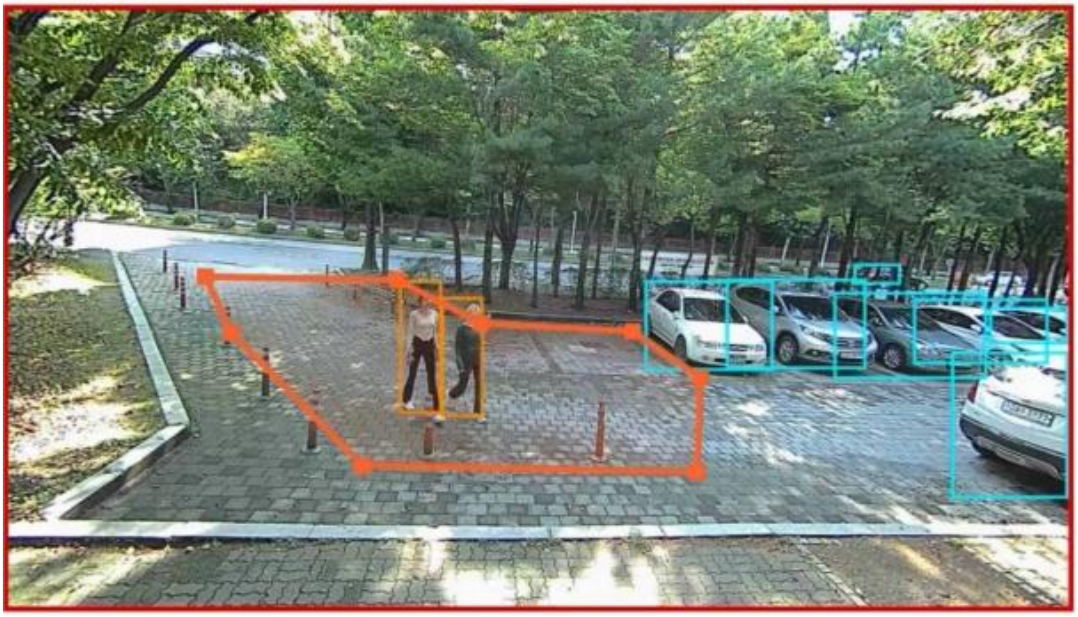

Figure 4. Event Occurrence Screen by Hanwha Vision AI Analysis

6. Conclusion

Hanwha Vision AI cameras represent a comprehensive solution that perfectly integrates a core AI technology stack, ranging from deep learning-based object detection and classification to BestShot. By applying model training strategies tailored to CCTV environments and Edge AI optimization technologies, Hanwha Vision has achieved the highest level of accuracy and real-time performance, solidifying its technological leadership.

Users can therefore significantly enhance their operational efficiency through intelligent video analysis and statistical features, leading to the creation of a more stable and sustainable video analysis ecosystem.

Hanwha Vision
13488 Hanwha Vision R&D Center,
6 Pangyo-ro 319-gil, Bundang-gu, Seongnam-si, Gyeonggi-do, Korea
www.HanwhaVision.com