



Vision AI for a safer, smarter world



Vision AI solution from HPE, NVIDIA®, and IronYun™

IronYun Vaidio® AI Vision Platform provides AI-enabled video search and analytics for security, health, safety, and operations. Powered by HPE ProLiant DL320 Gen11 Server and up to four NVIDIA L4 Tensor Core GPUs with a unique compact design, purpose-built for fast, efficient, and accurate AI video processing.

Computer vision is a broad field of artificial intelligence (AI) and computer science that focuses on interpreting visual data. Today, computer vision is used in a range of applications, from object recognition and autonomous vehicles to navigation and security. AI and machine learning (ML) technologies are crucial in powering computer vision technology. AI enables computer vision to understand, recognize, and analyze all types of visual data. AI models consume, absorb, and learn from the huge amount of labeled and unlabeled visual data.

How Vision AI is automating our spaces

Vision AI is a type of computer vision technology that leverages AI algorithms and deep learning models to analyze images and draw insights based on a set of criteria. Analyzing IoT sensors requires a lot of data input, and in order to manage both the camera's input and the AI workflows there's an important need for robust infrastructure to capture, store and processing the data in timely manner.

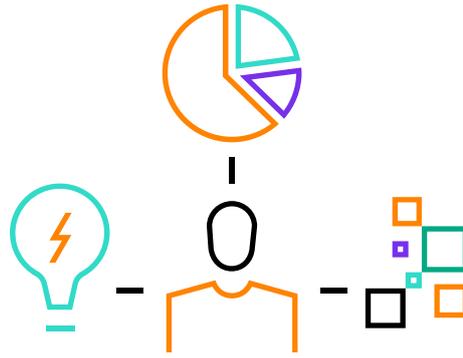
This technology has a tremendous benefits: improved operational efficiency, ability to react instantly to concerns, and provide public and worker safety.

Top use cases for Vision AI

- Worker safety
- Business operations efficiency
- Warehouse logistics
- Smart city
- Transportation analytics
- Retail operations
- Campus security
- Traffic analysis
- Intelligent airports and many more

What are the top benefits of computer vision AI?

- Automate tasks
- Increase accuracy
- Enhance safety
- Immediate response
- Scalability
- Intuitive and user-friendly
- Affordability
- Reliability
- Improve customer experience



Vaidio AI Vision Platform by IronYun

The Vaidio AI Vision Platform from IronYun uses advanced AI to deliver three key functions: active real-time monitoring for accurate alerting; intelligent, accelerated video search for incident investigation; and rich video metadata for business intelligence. Vaidio gives enterprise personnel reliable, real-time situational information, faster response and search times, and valuable operational data. Vaidio can quickly and cost effectively improve the security and safety posture of any enterprise.

Business challenge

- Improve facility security, safety, and access control
- Make corporate camera and security systems more effective
- Accelerate incident response time and incident investigation

Solution

- Vaidio AI Vision Platform by IronYun
- 30+ advanced, AI-enabled video analytics on a single platform
- Monitor 1,000s of cameras in real-time
- Search 1,000s of hours of video in seconds
- Mine the video stream for business intelligence data

Key functions

- Active real-time perimeter monitoring and intrusion detection
- Object recognition, detection and search
- Access control
- Forensic video search for incident investigation
- Works with any new or existing ONVIF IP camera
- Built-in integration with 28 video management systems (VMSs)

Differentiation

- Scalable — from 10s to 1,000s of cameras
- Open — works with existing IP cameras and video management systems
- Comprehensive — 30+ advanced AI video analytics for security, safety, and operations
- Flexible — deploy analytics a la carte as needed and multiple analytics per camera
- Validated — architecture and performance tested and certified by NVIDIA

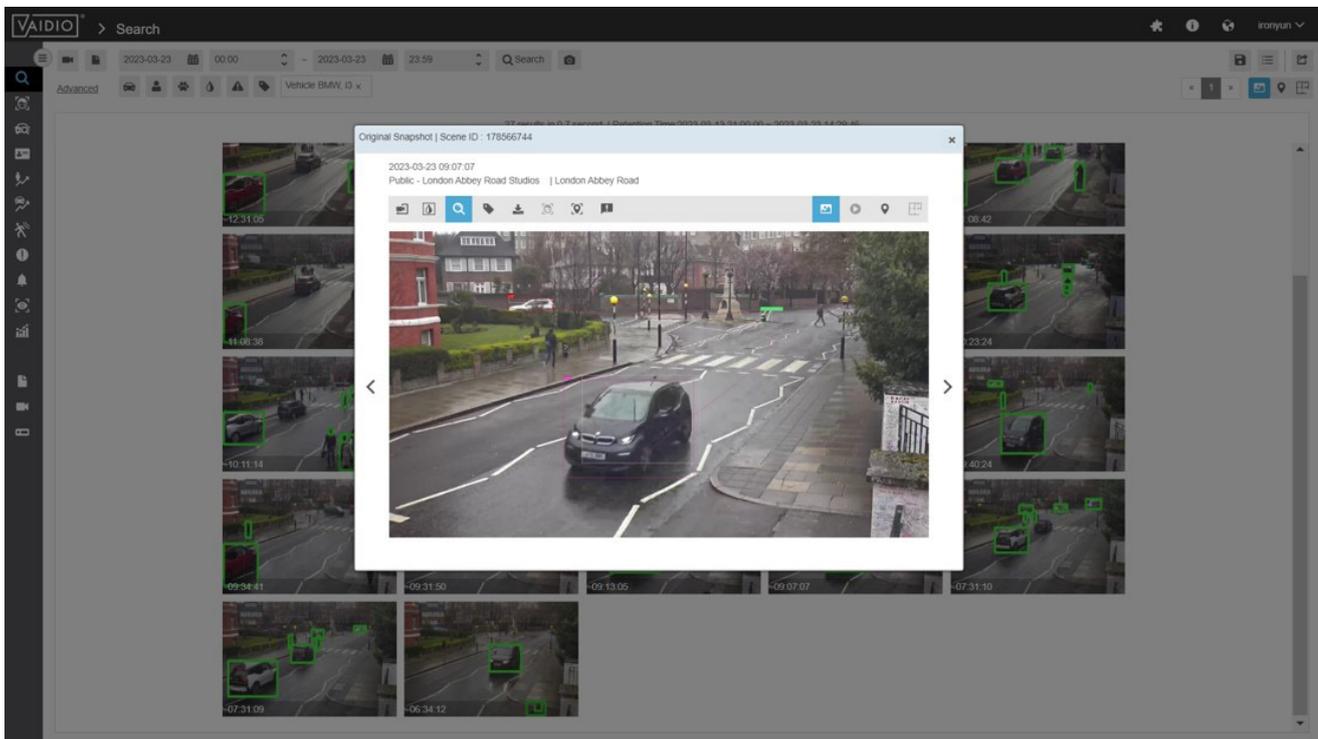


Figure 1. Instantly find an object from the past 24 hours of footage

Vaidio AI Video Analytics applications

Safety and security

- Intrusion detection
- Anomaly detection
- Crowd
- Fall
- Loitering
- Object counting
- Object detection
- Object left/Removed
- Perimeter monitoring
- PPE detection
- Restricted areas
- Scene change
- Sensor integration
- Smoke, Fire

Operations

- Camera health
- Dwell time
- Heat map
- Material inspection
- Privacy protection
- Operational data

Vehicles

- Access control
- Parking management
- Speed detection
- Traffic flow
- Vehicle make/Model
- Vehicle wrong direction

Access control

- Object recognition
- ID verification
- Visitor check-in

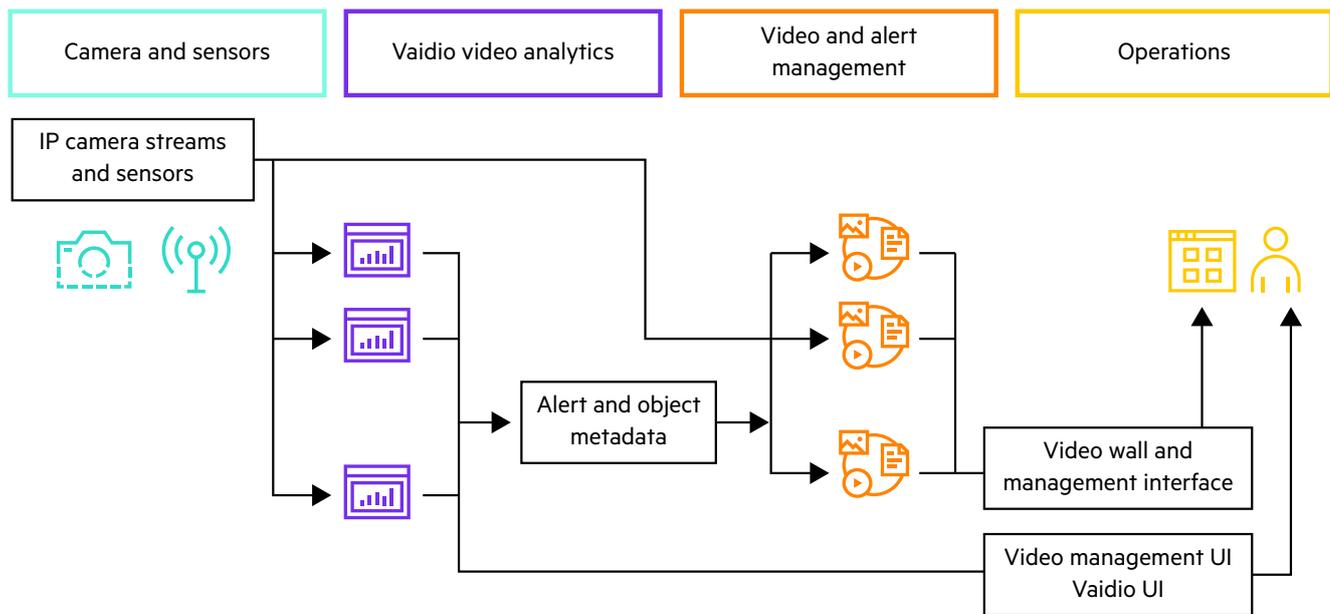
Health screening

- PPE detection
- Distancing
- Occupancy
- Temperature

Investigation

- Video search
- Video evidence

Vaidio AI Vision Platform architecture



Solution components

- New or existing IP cameras
- New or existing network video recorder (NVR) or video management system (VMS)
- HPE ProLiant DL320 Gen11 Server (NVIDIA certified and GPU accelerated)
- Vaidio Platform Software
- Vaidio Analytics Software (per analytic, per camera)
- TD SYNEX GoldSeal Support & Upgrades



Sample deployment

Small deployment 30 IP cameras	Medium deployment 200 IP cameras	Large deployment 1,000 IP cameras
30 Vaidio analytics licenses	200 Vaidio analytics licenses	1,000 Vaidio analytics licenses
1 Server	4 Servers	20 Servers 1 Command Center server

Accelerator-optimized compute — HPE ProLiant DL320 Gen11 Server with NVIDIA L4 Tensor Core GPU

The HPE ProLiant DL320 Gen11 Server is a 1U 1P server with a unique compact design purpose-built for computing at the edge and the data center delivering exceptional performance for computer vision with up to four single-wide (NVIDIA L4 Tensor Core GPU) or two double-wide GPU accelerators. The HPE ProLiant Gen11 advantage has the latest breakthrough advances in performance with fourth-generation CPU, PCIe Gen5, DDR5 memory, and NVMe storage.

The NVIDIA L4 Tensor Core GPU powered by the NVIDIA Ada Lovelace architecture delivers universal, energy-efficient acceleration for video, AI, visual computing, graphics, virtualization, and more. Packaged in a low-profile form factor, L4 is a cost-effective, energy-efficient solution for high throughput and low latency in every server, from the edge to the data center to the cloud.

With fourth-generation Tensor Cores and 1.5x larger GPU memory, NVIDIA L4 GPUs paired with the CV-CUDA library take video-content understanding to a new level. In a recent benchmark, the L4 delivered up to 120x¹ higher AI video performance than CPU-based solutions, letting enterprises gain real-time insights and implement smart-space solutions.

Trusted security by design

HPE ProLiant Gen11 servers are engineered with a fundamental security approach to defend against increasingly complex threats through an uncompromising focus on state-of-the-art security innovations built into our DNA.

- **HPE exclusive immutable digital fingerprint** with the silicon root of trust from HPE validates the lowest level of firmware to BIOS and software to help ensure the system is secure and in a known good state before the server even boots.
- **Protect your infrastructure, workloads, and data from threats to hardware,** and risks from third-party software, with a trusted edge-to-cloud security posture built on an HPE compute core hardened through a proven zero trust approach to security.

Modernize your compute management experience

Seamlessly monitor, manage, and gain visibility of your distributed compute environment.

- **Unified compute management** — streamline compute management operations with a centralized, cloud experience. With real-time access to servers, you can quickly gain visibility to your distributed environment, identify issues, and update servers in a few clicks.
- **Simplify and automate tasks** — save time and money with agile server lifecycle management that reduces manual efforts, drives better efficiency of server deployments and updates, gives visibility into server health status, and alerts you to critical hardware failures.
- **Secure compute operations** — take a zero trust approach incorporating multi-factor authentication, security certificates, and industry-leading security innovation of HPE ProLiant — so you can easily establish governance and compliance controls across the entire environment.

¹ Measured performance: 8x L4 vs. 2S Intel® 8362 CPU server comparison, end-to-end video pipeline with CV-CUDA decode, preprocessing, inference (SegFormer), postprocessing, encode, NVIDIA TensorRT 8.6 vs. CPU-only pipeline using OpenCV 4.7 and PyTorch inference.



NVIDIA Metropolis partner ecosystem

NVIDIA Metropolis, a partner program focused on bringing to market a new generation of vision AI applications, nurtures a rich ecosystem and offers powerful developer tools to supercharge vision AI applications that are designed to make the world's most important spaces and operations safer and more efficient. NVIDIA Metropolis partners, such as IronYun, are paving the way for AI-powered video analytics solutions for frictionless retail, streamlined inventory management, traffic engineering in smart cities, optical inspection on factory floors, patient care in healthcare facilities, and more. Businesses can now take advantage of this cutting-edge technology and the extensive Metropolis developer ecosystem to create, deploy, and scale AI and Internet of Things (IoT) applications from the edge to the cloud.

HPE and NVIDIA solution for IronYun Vaidio AI Video Analytics



HPE ProLiant DL320 Gen11 Server with NVIDIA L4 Tensor Core GPUs

Feature	Specifications
GPU support (front loaded)	Front loaded — up to four NVIDIA L4 Tensor Core GPUs (Up to 4 SW or 2 DW — front loaded)
	
Processor	4th Generation Intel® Xeon® Scalable processors
Memory	Up to 2 TB of DDR5 up to 4800 MT/s
Drive count	Up to 4 SFF NVMe/SAS/SATA or up to 8 EDSFF E3.S 1T NVMe SSD
Boot options	Optional internal RAID 1 M.2 NVMe (hot-pluggable) 2x SATA/NVMe M.2 connector (onboard)
I/O	Embedded 2x1GbE networking ports Up to 2 x16 PCIe Gen5 Up to 1 x16 OCP slot
Storage controller	HPE ProLiant Gen11 controllers (PCIe and OCP) + VROC NVMe/SATA
Management	HPE iLO 6
Chassis depth (GPU front end)	30.4" / 772 mm



Accelerate your AI outcomes with HPE and NVIDIA

Ideal for computer vision and video analytics, the HPE ProLiant DL320 Gen11 Server has a unique compact design, purpose-built for edge computing. It can pack up to four NVIDIA L4 GPUs in a 1U form factor to power smart spaces and loss prevention solutions and deliver insights in near real time. Customers can leverage offerings from the NVIDIA Metropolis ecosystem to deploy solutions targeted for their industry and use case.

Speed time to value with HPE and NVIDIA, together we deliver innovative AI technologies, software, and systems to help you accelerate your AI outcomes.

- [HPE ProLiant AI inference solutions](#)
- [HPE and NVIDIA AI collaboration](#)

Learn more at

[HPE ProLiant Gen11 servers](#)

[NVIDIA Metropolis](#)

[IronYun Vaidio AI Video Analytics](#)

BlueAlly

844-960-3902

[ServerComputeWorks.com](#)

Sales@ServerComputeWorks.com

[→ Request a quote](#)

**Hewlett Packard
Enterprise**

© Copyright 2023 Hewlett Packard Enterprise Development LP. The information contained herein is subject to change without notice. The only warranties for Hewlett Packard Enterprise products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. Hewlett Packard Enterprise shall not be liable for technical or editorial errors or omissions contained herein.

Intel Xeon and Intel are trademarks of Intel Corporation or its subsidiaries in the U.S. and/or other countries. NVIDIA, the NVIDIA logo, CUDA, and TensorRT are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. All third-party marks are property of their respective owners.

a00134038ENW