

The Industry 4.0 Case for Edge AI Accelerators

Edge AI accelerators optimize real-time processing, reduce latency, and improve energy efficiency, making them superior to traditional GPUs for edge deployments.

Current Compute Environment

In the era of Industry 4.0, AI-based computer vision enables machines to interpret visual data—essential functions for automation, quality control, and operational efficiency.

Historically, AI computer vision relied on GPUs to handle these functions. GPUs can handle heavy matrix math computations. GPUs are very good at processing graphics and scientific workloads. Additionally, AI training is suited for GPU technology.



However, GPUs are not practical for most industrial environments. They are primarily designed for desktop or data center use, making them less suitable for industrial applications. Additionally, significant expense is added with higher power and airflow requirements. Off-the-shelf GPU boards, such as the NVIDIA RTX 5090, do not meet industrial temperature standards. Their size and heat dissipation requirements also complicate achieving acceptable ingress protection (IP) ratings in industrial computing systems.

AI accelerators bridge the gap between traditional CPUs and the intensive tensor processing required by AI computer vision workloads. They are a core technology for realizing Industry 4.0 goals.

Benefits of Edge AI Accelerators in Industrial Environments

AI accelerators make AI processing feasible in industrial environments by delivering the necessary performance in a smaller footprint with lower power dissipation. Using AI accelerators reduces system costs, boosts reliability, and lowers total ownership cost (TCO).

MemryX Cascade 100 AI Accelerators

MemryX Cascade 100 AI Accelerators support AI models originally developed for GPUs. Thousands of GPU-targeted models can be compiled to run on Cascade 100 AI Accelerators without modifying the model source code. This compatibility provides a straightforward way to replace a GPU with a MemryX Cascade 100 in typical x86 or ARM-based industrial computing systems.



A stereoscopic camera for 3D imaging measuring time of flight depth sensor in a factory.

Key Use Cases

Quality Control and Inspection

AI computer vision systems automate product inspections directly on production lines, identifying defects with high precision and consistency.

These systems improve product quality, reduce human error, and operate faster than manual checks. In manufacturing, computer vision spots flaws in components within milliseconds on high-speed lines. Quick identification helps maintain high standards and minimizes waste.



Autonomous Navigation and Vehicle Guidance

Autonomous Guided Vehicles (AGVs) use AI vision to transport materials without human intervention. Factories and warehouses benefit from labor cost savings and streamlined logistics by tracking materials at rest and in transit. Reports suggest using robots in warehouses can human walking by up to 80%.



Worker Safety and Compliance

AI computer vision systems monitor worker behavior to ensure compliance with safety protocols. These systems alert supervisors when workers lack protective equipment or enter restricted areas.

In high-risk environments like construction or areas near robots, rapid detection of non-compliance helps maintain safety standards and lowers accident rates.

Process Optimization

AI-powered computer vision revolutionizes process optimization in logistics by providing real-time insights. It detects conveyor jams, misrouted packages, or equipment malfunctions in logistics hubs, prompting immediate corrective action.

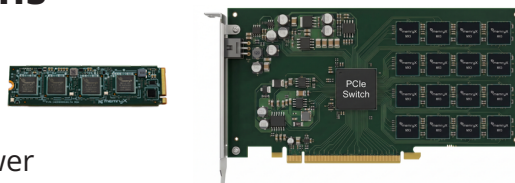
With faster, more accurate visual processing, AI vision enables predictive insights and operational improvements. AI vision is vital for modern logistics and operational management.



The Role of MemryX Cascade 100 Solutions

AI accelerators are purpose-built to address the specific AI computational needs of edge environments like factories. Unlike general-purpose processors, MemryX Cascade 100 boards are optimized for edge AI inference with a focus on power efficiency, low latency, and high throughput. This approach is ideal for real-time industrial applications where energy efficiency and fast processing are crucial for operational success.

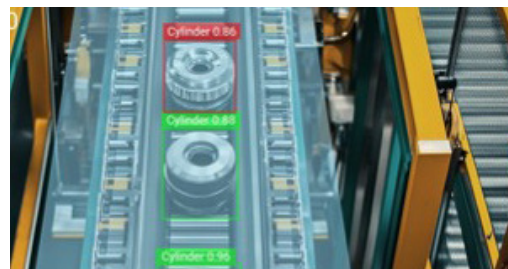
Cascade 100 boards provide excellent AI computer vision performance and accuracy for applications such as autonomous navigation and automated quality inspection. They can be added to industrial computing systems (IPC) via M.2 or PCI Express slots and support x86 or Arm processors under Windows or Linux. Cascade 100 boards are the best way to add AI processing power to industrial systems without compromising size, weight, power (SWP), ruggedness, and reliability. MemryX Cascade 100 is the key to achieving success in computer vision in Industry 4.0.



AI Accelerator Capabilities

Defect Detection

On high-speed production lines, Cascade 100 accelerates AI models used for defect detection and enables real-time analysis. Models are trained to identify specific defects and can be updated in the field as new defect types emerge. This adaptability and longevity are valuable in dynamic production settings. Common models for defect detection include ResNet, VGG, and YOLO, all of which run successfully on the MX3 platform. These AI vision methods outperform prior approaches in speed and accuracy, especially in complex scenarios.



Semantic Segmentation

Semantic segmentation models label each pixel in an image to identify various elements. This output is often visualized as color overlays. Segmentation helps path-planning software identify roads and other areas to navigate or avoid. The YOLO model family offers segmentation-optimized versions. Cascade 100 demonstrates high performance with YOLO, DeepLabV3, MediaPipe, and Unet. Many of these models run three times faster on the MX3 than on GPUs.



Object Detection & Pose Recognition

Cascade 100 solutions quickly identify safety issues such as missing Personal Protective Equipment (PPE). Models trained to detect proper PPE usage generate alerts when they find unsafe environments, increasing workplace safety. Pose recognition combined with object detection also improves factory safety by identifying individuals in distress or in hazardous positions. Models like MediaPipe Holistic compile efficiently for Cascade 100 solutions, enabling these critical applications.



MemryX Cascade 100 vs. GPU in Edge Computing

GPUs have been central to advancing AI by supporting parallel processing for algorithms such as convolutional neural networks (CNNs). They helped overcome significant barriers, enabling computers to sense in new ways.

Despite their importance in AI's growth, GPUs are unsuitable for many industrial environments. Better solutions are required. MemryX enables developers to compile GPU-based AI models and run them on more efficient accelerator architectures tailored for industrial use.

Cascade 100 boards are designed to provide efficient, real-time edge processing with minimal energy consumption. Ideal for power-sensitive Industry 4.0 applications, including autonomous navigation, quality inspection, and safety monitoring. Scalable Cascade 100 solutions based on the efficient MX3 AI accelerator, a cascadable building block IC that integrates AI processing and memory on a single die to achieve high efficiency and performance for a wide variety of applications. Multiple chips can be cascaded for low-cost or high-performance needs.

The Cascade 100 family includes products with PCI Express (Cascade 100P), M.2 (Cascade 100M), USB (Cascade 100U) and also a Raspberry Pi HAT+ form factor (Cascade 100R). These products span a wide range of performance and capability from < 5 Watts to 70 Watts for the 16-chip PCI Express card.

Cascade 100 Product Family



	Cascade 100R	Cascade 100U	Cascade 100M	Cascade 100P
				
Performance	12 TFLOPS	12 TFLOPS	26 TFLOPS	104 TFLOPS
MX3 Devices	2	2	4	16
Power	5 Watt	5 Watt	8W Typical 12W Max	Up to 75 Watt

MemryX Cascade 100 Family Value Propositions

Accurate out of the box

- BFloat16 processing is more accurate than INT8
- Skips the quantize-and-revalidate workflow
- Runs trained models as-is without needing hand-tuning

Efficient and self-contained

- On-chip weights mean low power, low latency, small footprint
- No external DRAM — BOM isn't exposed to DRAM price spikes
- Optimize for low power, high performance or steady throughput

Scalable solutions with a common toolchain

- Same MX3 and SDK across the line — build once, deploy anywhere
- PCI Express, Raspberry Pi HAT+, USB-C, or M.2/embedded
- Large library (600+) of pre-validated computer vision models

Designed for the edge

- Local inference — private, offline, plug-and-play
- No cloud, quota, or recurring cost
- Prototype on the bench, ship in the product

Conclusion

A primary goal of Industry 4.0 is to boost productivity by making production processes more efficient. AI accelerators like the Cascade 100 Family help achieve this by integrating intelligence directly into the workflow. Industry 4.0 also aims to improve flexibility, allowing production systems to adapt quickly and economically to changing needs, including personalized manufacturing. AI accelerators play a significant role in enabling this adaptability via system connectivity and easy model compilation.

The innovative MemryX at-memory architecture makes AI both feasible and cost-effective in industrial environments by resolving the limitations of traditional GPUs. To explore how the MemryX Cascade 100 products can enhance your Industry 4.0 plans or to discuss deployment options, please contact us at memryx.com/contact.

110 Miller Ave. Suite 1. Ann Arbor, MI 48104



MemryX Inc. is an AI semiconductor startup headquartered in Ann Arbor, MI USA with branches in Taipei and Hsinchu, Taiwan. We develop a highly scalable and innovational AI accelerator offering high performance, low power, and customer ease of implementation for embedded Edge AI vision-based applications and real-time processing.

Information contained is subject to change without notice. MemryX Inc logo, MX3, and other designated brands included herein are registered trademarks in the US and other countries.

© 2026 MemryX Inc., All Rights Reserved