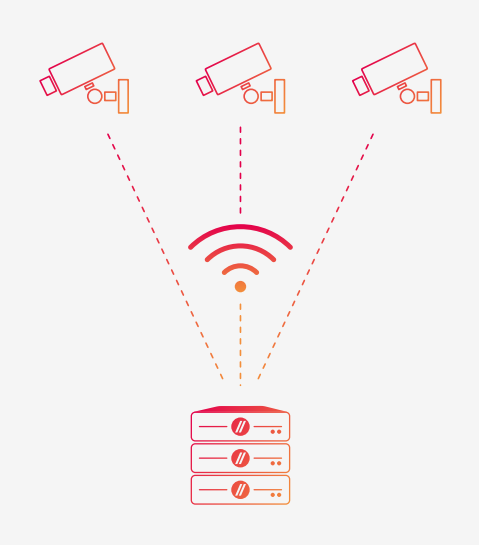
Accelerating Computer Vision with GPU on Scale Computing Platform

HIGHER PERFORMANCE COMPUTING AT THE EDGE

Computer vision (CV) is a field of artificial intelligence that trains computers to interpret and understand the visual world. Using digital images from cameras and videos and deep learning models, machines can accurately identify and classify objects — and then react to what they "see."



The integration of graphics processing unit (GPU) technology with computer vision significantly enhances the performance and efficiency of data processing tasks. Originally developed for graphics tasks, GPUs are now widely used in various computing tasks, especially those involving parallel processing, such as artificial intelligence, machine learning, and computer vision applications.

GPUs, with their parallel processing capabilities, accelerate complex computations required for image recognition, object detection, and other computer vision applications. This leads to faster analysis, real-time processing, and improved accuracy, which are crucial for industries relying on automated visual insights.

COMPUTER VISION OPPORTUNITY

CV market capabilities have grown exponentially, with most AI solutions incorporating CV in some form. This growth is driven by advancements in machine learning algorithms, increased computational power, and the proliferation of data from various sources such as cameras and sensors. Since CV can be deployed virtually anywhere, it has introduced challenging infrastructure and processing demands.

Integrating Computer Vision-Based Solutions

Integrating computer vision and artificial intelligence (AI) can significantly benefit organizations across many industries. These technologies often accompany digital transformation and innovation, providing organizations with the tools to automate processes, enhance operational efficiency, and improve decision-making. However, the implementation of CV solutions can be daunting due to the complexity of managing vast amounts of data, ensuring real-time processing, and maintaining high levels of accuracy and reliability.

Real-Time Actionable Insights

Computer vision environments necessitate real-time, actionable insights to create value for an organization. CV's low latency requirement makes edge computing an ideal deployment model. By processing data closer to the source, edge computing reduces the time it takes to analyze and act on information, which is critical for applications requiring immediate response, such as autonomous vehicles, industrial automation, and real-time surveillance.

Simplifying Data Use and Analysis

Deploying CV at the edge simplifies data use and enables easier data analysis. Edge computing allows organizations to process and filter data locally, reducing the need for large-scale data transfers to central data centers. This not only decreases bandwidth usage and associated costs but also enhances data privacy and security by limiting the exposure of sensitive information.

In its Forecast Analysis: Artificial Intelligence Software, 2023-2027, Worldwide, Gartner projects that global spending on AI software, which includes computer vision applications, will reach $297 billion by 2027. This growth underscores the increasing adoption of AI technologies across various sectors, driven by the need for automation and optimization.

ADVANCED APPLICATIONS OF GPU-ACCELERATED COMPUTER VISION

The integration of GPU technology with computer vision has revolutionized various industries by enabling applications that require substantial computational power beyond the capabilities of traditional CPUs. GPUs accelerate the processing of complex algorithms and large datasets, making real-time data analysis and decision-making possible.

Healthcare

Image Analysis and Diagnostics. In healthcare, computer vision is used to analyze medical images such as MRI, CT scans, and X-rays. GPUs accelerate the processing of high-resolution images, enabling faster and more accurate diagnostics. For example, detecting tumors, fractures, or other anomalies often involves deep learning models that are computationally intensive.

3D Reconstruction. Reconstructing 3D models from 2D images requires substantial computational power, which GPUs provide.

Manufacturing

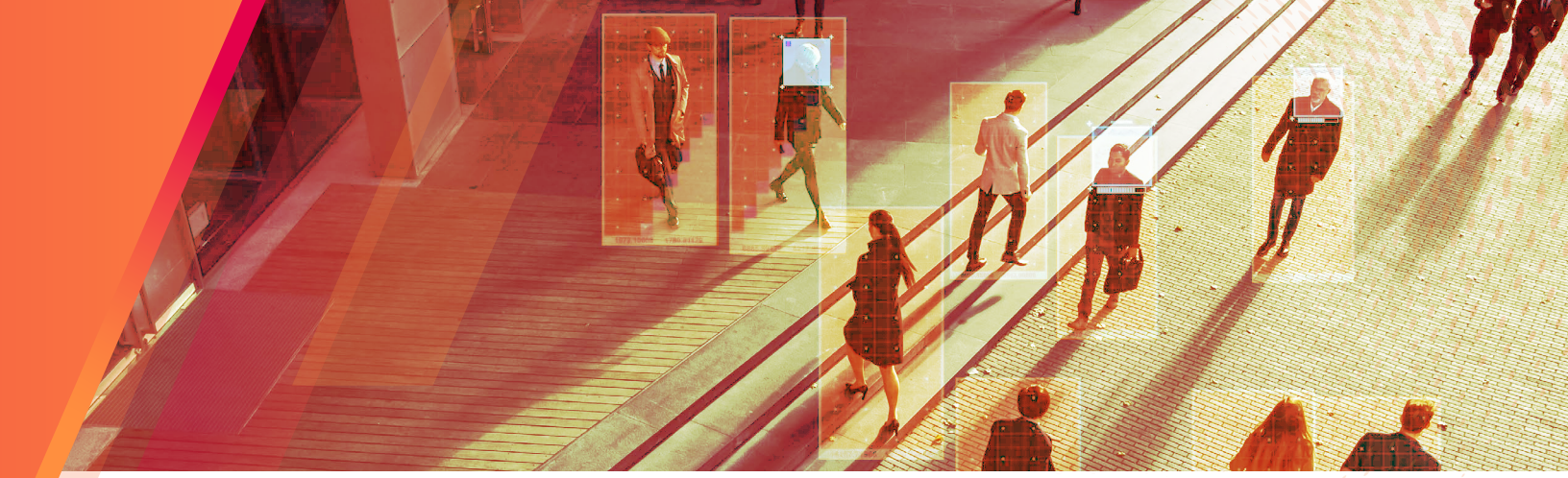
Quality Control. Manufacturing processes use computer vision for quality control by inspecting products for defects in real time. High-speed image processing with GPUs allows for the rapid detection of defects, ensuring that defective products are identified and removed from the production line.

Robotics. Industrial robots use computer vision for tasks such as picking and placing items, assembly, and packaging. These tasks require precise object recognition and spatial analysis, which are computationally demanding and benefit from GPU acceleration.

Security and Surveillance

Facial Recognition. Security systems use computer vision for facial recognition to identify individuals in real time. GPUs process the vast amount of visual data from surveillance cameras, enabling quick and accurate identification.

Behavior Analysis. Analyzing behavior patterns in surveillance footage to detect suspicious activities or safety violations involves complex algorithms that require GPU acceleration to process data efficiently.



Retail

Customer Behavior Analysis. Retailers use computer vision to analyze customer behavior in stores, such as tracking movement patterns, dwell times, and product interactions. GPUs enable the real-time processing of video data to generate actionable insights.

Inventory Management. Automated inventory management systems use computer vision to monitor stock levels and detect out-of-stock situations. This requires high-speed image processing, which GPUs facilitate.

Agriculture

Crop Monitoring. Drones equipped with cameras and computer vision systems monitor crop health by analyzing images for signs of disease, nutrient deficiencies, and pest infestations. GPUs process high-resolution images quickly, providing timely insights for farmers.

Yield Prediction. Predicting crop yields involves analyzing large datasets, including images, weather data, and soil conditions. GPUs help process these datasets rapidly, allowing for accurate yield predictions.

INDUSTRY APPLICATIONS

CV and graphics process unit (GPU) requirements in your edge environment will vary depending on a multitude of variables and on the industry you operate in.

Healthcare	Manufacturing	Retail	Hospitality	Financial Services
• Research • Patient care • Enhanced diagnostics • Supply chain • Safety and security	• Factory automation • Loss detection • Inventory management • Predictive maintenance • Robotic systems • Safety and security	• Facial recognition • Augmented reality • Smart assistants • Smart inventory • Loss detection • Analyzing customer traffic • Dynamic experience • Improved product placement	• Facial recognition • Augmented reality • Loyalty • Enhanced customer experience • Smart assistants • Analyzing customer traffic • Accurate customer insights	• Risk mitigation • Fraud detection • Facial recognition • Safety and security • Enhanced customer experience

ADVANTAGES OF RUNNING EDGE-BASED COMPUTER VISION ON SC//PLATFORM

Running CV applications at the edge, close to the data capture point, offers numerous benefits that address key challenges and enhance operational efficiency:

Data Explosion. Computer vision generates massive amounts of data, which has value only when it can be properly collected and analyzed. But bandwidth isn't free, and transferring all that data to the cloud for processing is both impractical and cost-prohibitive. Edge computing allows all this rich data to be collected and processed locally.

Resiliency. Reliable connectivity is key when applications are running from a centralized location. Whether it's a complete outage, occasional drop or simply high error rates, any interruption is bound to affect the availability and performance of applications relying on that connection. Running applications locally means they can continue to operate as expected, even without a connection to the cloud or data center.

Latency. Information takes time to travel across a network. The longer it takes, the more it impacts end-to-end processing times. Expectations for application response times vary from one application and organization to the next. However, the more an application experience benefits from a real-time response, the more important it is to remove distance as a factor. Edge computing brings applications closer to where they are used, reducing lag time and improving efficiency.

Regulatory Compliance. Complying with data security and privacy regulations for images is a serious business. The risk of interception and potential for regulatory non-compliance increases every time data is moved. By definition, the cloud is a fuzzy place, making it difficult to know exactly where data is and where it has been. The more data can be collected and processed on-site, the simpler maintaining compliance becomes.

Cost Savings. By processing data locally, organizations can reduce the costs associated with bandwidth and cloud storage, optimizing their overall IT budget

---

SCALE COMPUTING PLATFORM

Scale Computing Platform, the fully integrated compute, storage, virtualization, and disaster recovery environment simultaneously runs legacy and modern applications on the same infrastructure. There's no need to manage the complexity of separate hardware and software components to support individual point solutions.

Enterprise infrastructure has become critical for data-intensive industries, whether you are delivering critical patient outcomes, customized and real-time customer experiences, or advancing supply chains for zero-touch delivery and logistics. SC//Platform technologies offer IT teams the ability to deliver high-performance infrastructure and streamlined application experiences for employees and customers across industries.

For workloads where CPU alone is not enough, SC//Platform supports several models throughout the product family with both integrated and discrete GPU options. These GPU resources can be directly allocated to specific workloads, providing the necessary computational power for demanding tasks.

Implementing a strong CV solution is a critical combination of hardware and software. For more information, contact your Scale Computing partner or account manager or visit scalecomputing.com.