

# Solution Brief

Edge AI Inference  
Computer Vision SaaS



## AI Inference Proof of Concepts in Days, Not Months, with Viso Suite



**Viso Suite packages AI inference capabilities into easy-to-deploy SaaS offerings enabled by the Intel® Distribution of OpenVINO™ toolkit**

AI-enabled computer vision and inference are revolutionizing workflows up and down the supply chain for businesses across all industries. The number of applications continues to grow, and the computer vision market is expected to reach a valuation of USD 51.3 billion by 2026 at a compound annual growth rate (CAGR) of 26.3 percent.<sup>1</sup>

### **Challenge: Complexity and edge performance requirements**

Traditionally, computer vision and AI inference deployments have been limited to large tech companies and enterprises with the resources to architect, deploy, and manage complex edge solutions. These efforts include hiring or training developers with AI expertise to build, train, and tune AI inference models. Successful deployments also require high performance in IoT devices and edge servers to process compute- and graphics-intensive AI workloads. These barriers have made it difficult for small- and medium-sized businesses to take advantage of computer vision, putting them at risk for being left behind.

### **Solution: Flexible, scalable computer vision Software as a Service**

viso.ai offers computer vision Software as a Service in a unified, cloud-based platform that businesses of any size can use. Their flagship product, Viso Suite, integrates with the Intel® Distribution of OpenVINO™ toolkit and enables users to select pretrained, optimized AI inference models with a single click. Users can execute AI inference on multiple Intel-enabled devices, with plugins available to support heterogeneous architecture and requests in parallel. Nico Klingler, cofounder of viso.ai, says, "Viso Suite offers a low-code experience for AI experts and enables users with minimal coding experience to benefit from the features and optimization of the OpenVINO toolkit. With Viso Suite, businesses can bring their solution to market faster<sup>3</sup> than traditional, custom development—think days instead of months." Viso Suite dramatically simplifies computer vision and AI inference adoption, granting access of previously exclusive technology to organizations of any size while accelerating time to market with lightning-fast proof of concepts.

*"Viso Suite offers a low-code experience for AI experts and enables users with minimal coding experience to benefit from the features and optimization of the OpenVINO™ toolkit. With Viso Suite, businesses can bring their solution to market faster<sup>3</sup> than traditional, custom development—think days instead of months."*

—Nico Klingler, cofounder of viso.ai

## How it works

Viso Suite provides users with a cloud-based workspace to help accelerate the creation and help automate the management of computer vision projects. Users can quickly build software modules with AI algorithms in an intuitive drag-and-drop interface that includes popular frameworks like TensorFlow, PyTorch, and the OpenVINO toolkit. Users will input their existing computer vision hardware into a device management dashboard that allows Viso Suite to execute chosen workflows on specific nodes or edge devices.

### The Viso Suite workflow

- **Step 1: Build.** Drag and drop software modules into a computer vision workflow. The best tools, databases, and cloud infrastructure are integrated out of the box and include the OpenVINO toolkit.
- **Step 2: Deploy.** Enroll and deploy workflows to edge devices. Viso Suite seamlessly integrates with the latest computer vision hardware, including Intel® processors and VPUs.
- **Step 3: Monitor.** Gather insights from real-time metrics dashboards to guide your decision-making.
- **Step 4: Manage.** Add or remove users, access, locations, security, devices, and data all from one place.
- **Step 5: Browse.** Use the Viso Marketplace to expand platform capabilities with prebuilt extensions and even upload extensions of your own.

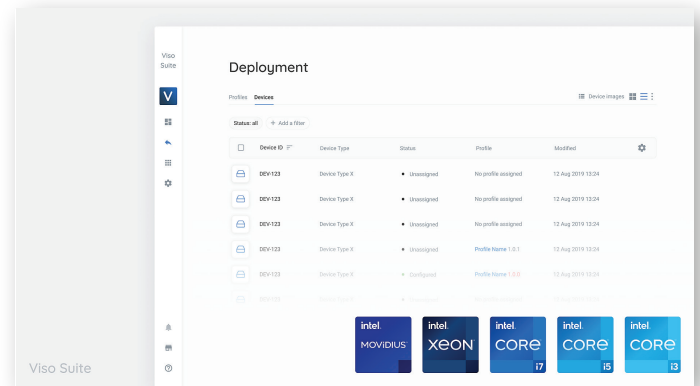
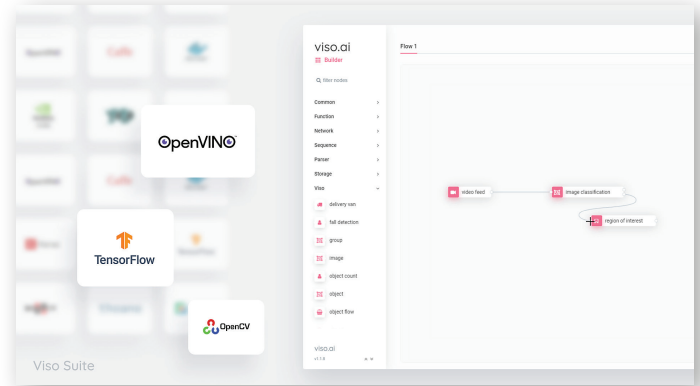


Figure 1: Users construct computer vision workflows via a drag-and-drop interface and then deploy workflows to available devices in the Viso Suite dashboard.

## The OpenVINO™ toolkit accelerates AI inference

### Intel® Distribution of OpenVINO™ toolkit performance

in frames per second (fps) vs. TensorFlow<sup>2</sup>

Up to  
**3.23x**  
higher fps

Intel® Celeron® CPU and iGPU

Up to  
**4.46x**  
higher fps  
Intel® Xeon® Scalable processor

With Viso Suite, users can select pretrained and optimized AI models from the OpenVINO toolkit in just a few clicks. Users can also execute converted and optimized models across diverse Intel® architectures using or enabled by the OpenVINO toolkit, based on their existing hardware configurations. The OpenVINO toolkit includes pretrained AI models that are optimized for higher inference performance on Intel architecture compared to popular AI frameworks like TensorFlow. In benchmark tests comparing ResNet-50 (a simulated, 50-layer convolutional neural network) performance, the OpenVINO toolkit with int8 quantization outpaced TensorFlow executing fp32 data types by a factor of up to 3.23x using an Intel® Celeron® processor with integrated GPU.<sup>2</sup> For Intel® Xeon® Scalable processor-enabled configurations, the factor increases up to 4.46x.<sup>2</sup>

### OpenVINO toolkit capabilities

- **Pretrained, preoptimized inference models:** Developers can use the Open Model Zoo to find and deploy models that are ready for inference now, or they can use their own deep learning models.
- **High-performance deep learning inference:** The OpenVINO toolkit enables developers to convert and optimize AI models to achieve higher performance<sup>2</sup> for deep learning inference applications.
- **Streamlined development:** The toolkit includes inference tools for low-precision optimization, media processing, computer vision libraries, and preoptimized kernels.
- **Write once, deploy anywhere:** Developers can deploy inference models across a diverse combination of Intel® CPUs, VPUs, and FPGAs and benefit from optimized performance on-premises, on device, or in the cloud.

For workloads and configurations, visit [Intel.com/PerformanceIndex](https://www.intel.com/performance/index). Results may vary.



Figure 2: Performance optimization makes it easy to deploy AI object recognition across several use cases. Images were built on Viso Suite.

### Hardware flexibility and avoiding sunk costs

viso.ai engineers further note that OpenVINO toolkit optimization makes it easier to upgrade to next-gen hardware while maintaining high performance over a wide range of use cases. This gives customers the advantage of migration portability; no organization is beholden to a specific configuration or stack. Customers can adapt to evolving market requirements by upgrading their architecture and still benefit from the performance optimization of the OpenVINO toolkit and the low-code environment that Viso Suite has to offer.

### Expanding OpenVINO toolkit capability, scalability

Viso Suite customers can implement plugins to their AI inference workflows by subscribing to new extensions through the integrated Viso Marketplace. This includes the ability to support parallelization across heterogeneous devices. The OpenVINO toolkit simplifies parallel inferring by efficiently processing workloads on all available hardware. Parallelization becomes especially effective when combined with supported architectures that benefit from OpenVINO toolkit optimization. Gaudenz Boesch, cofounder of viso.ai, comments, “Especially combined with the Intel® Movidius™ Myriad™ X VPUs, the OpenVINO toolkit provides a scalable solution to a variety of computer vision problems. The possibility to scale from one to multiple VPUs at low cost and low power consumption makes it ideal for moving from simple prototypes to heavy-load systems in production.”

### Viso Suite summary of benefits



#### OpenVINO™ toolkit optimization

Organizations of all sizes get access to the OpenVINO toolkit, with optimized performance on Intel® CPUs, GPUs, and VPUs.



#### Reduced time

Bring solutions to market fast vs. traditional AI development time. Deliver a full proof of concept (PoC) in days instead of months.<sup>3</sup>



#### Reduced cost

Access Viso Suite with a monthly subscription. Avoid the expensive licensing, infrastructure, and DevOps costs of traditional AI development.



#### Extended expertise

Get started fast with an intuitive interface and familiar tools. Viso Suite is built for AI experts and also IT professionals with less experience.

### Conclusion: An open playing field for computer vision

With Viso Suite, customers combine the streamlined AI development workflow from a cloud-based unified platform with their existing edge devices and infrastructure. By spending less time developing their own AI workflows and computer vision software, businesses can realize key benefits and fast time to market for otherwise complex and cost-prohibitive deployments.

Boesch and the viso.ai engineering team call this the democratization of computer vision: “The value proposition of disruptive low-code technology gives companies of any size the flexibility and agility to adapt to the fast-changing industry reality while accelerating the time to market of new applications to keep up with changing markets.” Fewer barriers to entry allow for a more open playing field and allow for more-automated, more-innovative use cases to emerge.



## Intel Distribution of OpenVINO toolkit

OpenVINO is a complete development toolkit for deploying AI on Intel® hardware. It optimizes and converts deep learning models into high-performance inference engines that can run on any mix of Intel hardware.

Developers can choose between standard support releases or Long-Term Support (LTS). Standard releases provide new versions of the toolkit every quarter, ideal for early-stage projects and developers looking to take advantage of the latest innovations in deep learning. LTS offers long-term maintenance and support, a great choice for later-stage developers focused on leveraging the toolkit's existing features and functionality.

Learn more at [intel.com/openvino](https://intel.com/openvino).

## Learn more

Get started with Viso Suite at [viso.ai](https://viso.ai).

Explore the OpenVINO toolkit at [intel.com/openvino](https://intel.com/openvino).

### About viso.ai

Switzerland-based viso.ai seeks to advance humanity through the power of computer vision. The company works closely with organizations in farming, medicine, transportation, and disaster response to guide product development for their widely accessible AI vision application, Viso Suite.

[viso.ai](https://viso.ai)



1. "The AI in computer vision market is estimated to be valued at USD 15.9 billion in 2021 and reach USD 51.3 billion by 2026, at a CAGR of 26.3% between 2021 and 2026," globenewswire.com, June 2021.
2. "Intel® Distribution of OpenVINO™ toolkit Benchmark Results," OpenVINO.org, date of access: August 2021. See "resnet-50-TF [224x224]" for comparative data. See "Platform Configurations" for full configuration data. [https://docs.openvino.org/latest/openvino\\_docs\\_performance\\_benchmarks\\_openvino.html](https://docs.openvino.org/latest/openvino_docs_performance_benchmarks_openvino.html)
3. Source: Estimation and internal measurements provided by viso.ai.

#### Notices and disclaimers

Intel is committed to respecting human rights and avoiding complicity in human rights abuses. See Intel's [Global Human Rights Principles](#). Intel® products and software are intended only to be used in applications that do not cause or contribute to a violation of an internationally recognized human right.

Performance varies by use, configuration, and other factors. Learn more at [intel.com/PerformanceIndex](https://intel.com/PerformanceIndex).

Performance results are based on testing as of dates shown in configurations and may not reflect all publicly available updates.

Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.

Intel® technologies may require enabled hardware, software, or service activation.

No product or component can be absolutely secure.

Your costs and results may vary.

© Intel Corporation. Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others.

0921/ADS/CMD/PDF