

Al Containers[™]

Version 1.3 | As seen on <u>AI Containers</u>™

Deploying AI Containers™ securely, at scale

In <u>edge inference</u>, the classification and/or decision of a neural network typically happens at the edge (on the camera). While inferring on one single device is relatively straightforward, deploying different <u>neural networks</u> from various <u>frameworks</u>, securely, over the air, to millions of different edge devices in the field—is a lot more complicated.

AnyConnect provides a solution to this problem through an <u>Al Store</u>[™], enabling users to deploy Al Containers[™] securely to the edge, <u>at scale</u>, with <u>access control</u>. The AnyConnect <u>Console</u> monitors all devices on the platform, and provides the ability to centrally view, control and manage deployed containers.

a. Select Al Container(s)[™] from our Al Store[™].

- **b.** Optionally, upload your own AI Container(s)[™] to our AI Store[™].
- **c.** Deploy selected AI Containers[™] to your cameras with AnyConnect's <u>OTA</u> <u>Programming</u> feature.



Choose your AI Container(s)[™] via AnyConnect Console



A Repository of AI Containers[™]



AnyConnect updates your AI Containers[™] for Cloud Inference

AnyConnect's Smarter Al[™] Camera Platform offers a repository of Al Containers[™] in its cloud. The repository holds containers with both free and paid trained neural networks, as well as some software logic.

The repository also supports leading <u>deep learning frameworks</u>, such as <u>TensorFlow</u>, <u>PyTorch</u>, <u>Keras</u>, <u>MXNet</u>, <u>ONNX</u>, and more.



"Copyright©2020" : "All Rights Reserved" "Page" : "3"

Al Container™ at the Edge

AnyConnect's Smarter AI[™] Camera Platform deploys AI Containers[™] to the edge, seamlessly, securely, and <u>over-the-air</u>. The system will deliver the container with the right framework, to the right edge device, and its neural network accelerator automatically. This system supports heterogeneous edge device deployments with different types of edge inference accelerators, like <u>CPUs</u>, <u>GPUs</u>, <u>Intel</u> <u>Movidius</u>, <u>Google Coral Edge TPU</u>.

The AnyConnect console allows you to manage and monitor the deployment of Al Containers[™] to the edge seamlessly. The management system built into the platform will automatically convert the trained neural networks and their associated logic to the right frameworks needed by your edge device infrastructure.

Our platform will also provide statistics on the quality of inferences centrally.



AnyConnect OTA Programming deploys your AI Containers[™] for Edge Inference



Al Containers™ in depth



Al Containers[™] help you to solve plumbing, deployment & security challenges related to the deployment of Al applications to the edge. Although chipmakers have created common Al runtimes, such as Intel <u>OpenVino</u>, <u>Qualcomm Neural Processing SDK</u>, NVIDIA <u>JetPack SDK</u>, etc. to simplify porting Al Models to their chips, many problems remain as deploying an Al Application to the edge on a new product always requires long, complicated and costly engineering efforts. Let's look at what Al Containers[™] look like.

AnyConnect's Al Containers[™] have four main parts – Egresses, Ingresses, Configuration Interface, and the Al Compute, Security & Storage Interface. Al Containers[™] enable system administrators to deploy an Al Application to a new product graphically in minutes. Let's dive into each of those four sections.



"Copyright©2020" : "All Rights Reserved" "Page" : "5"





"Copyright©2020" : "All Rights Reserved" "Page" : "6"

Ingress

All video, audio, and data (e.g., sensor data, positioning, etc.) streams required by an AI Application is explicitly defined in the Ingress part of AI Containers[™]. The Ingress side specifies the format of streams in detail, such as resolution, framerate, and color for video, resolution & number of channels for audio, etc. As an administrator, you'll know before deploying an AI Container[™] Over-the-Air (OTA), if this container is compatible with the camera product. For instance, does this product have a suitable AI Accelerator, does it have the required sensors/imagers/microphones, and finally, does it have enough available compute capacity.

Configuration Interface

As AI Models come from different suppliers and AI Containers[™] can host multiple models, the configuration of these models and the associated software logic is complicated. AI Containers[™] standardize the configuration of AI Models and the embedded software logic, ensuring easy and engineering department-free deployments.

Egress

Al Models' inferences format (what Al Models predict/recognize/classify) are standardized and directly usable to create events and notifications. On top of that, the egress interface provides inference quality metrics. Inference quality metrics are recorded in AnyConnect's cloud to provide an understanding of the Al Model's inference quality per camera and over time.

AI Compute, Security & Storage

Al Containers[™] do not replace or add layers between Al Models and their runtimes; they offer unified, standardized, and monitored access to compute, storage (slow storage such as SD card and fast storage such as <u>NVME SSD</u>) as well as security resources. Security is critical as many Al models come with copy protection, either encryption or a Digital Right Management (<u>DRM</u>). By providing access to Crypto Cores, such as Trust Zone and Trusted Platform Modules (<u>TPM</u>s), Al Containers[™] allow secured Al Models on almost any device.

For more information about AI Containers and how they compare with other container systems such as Docker and Kubernetes, please visit our page <u>AI Containers™ vs.</u> <u>Docker & Kubernetes</u>.



Al at the edge has never been that simple.

Contact Us

