Deploying AI Edge Computing with AAEON Solutions and Intel® distribution of OpenVINO™ Toolkit

Case Study: ComBox and LARGA on-board Bus Passenger Counting System
Product: VPC-3350S with AI Core X

Introduction

As more industries and companies turn to AI and Edge Computing to bring intelligent applications to manage and operate their businesses, the field of suppliers and solutions offering AI Edge platforms is increasing as well. However, as a developer or someone looking to design and deploy their own solutions, there are many things to consider. Most hardware platforms, while sharing many open source resources, still rely on their own proprietary acceleration and optimization software. Whether a developer is new to the field or looking to migrate their models to more efficient hardware platforms, picking the solution with the best software and toolkit support is just as important as the hardware itself.

AAEON is helping make this step easy for clients by offering a range of solutions compatible with the Intel® distribution of OpenVINO™ toolkit. These platforms with the innovative OpenVINO toolkit make it easier than ever for developers to deploy their models to the edge, whether they're new developers entering the market for the first time, or migrating their models from other hardware environments.

To help illustrate the advantages of using the OpenVINO toolkit, this paper will look at the case study from leading embedded Edge AI developers ComBox and LARGA. These two companies partnered together to deploy a passenger counting system for public buses, powered by AAEON's VPC-3350S with AI Core X, and utilizing the OpenVINO toolkit to help optimize and quickly deploy their software models.

Through this case study, the partnered companies looked at three key points; Data Center vs Edge deployment, Intel based solutions vs other comparable platforms, and the advantage and features of OpenVINO.

Part 1: Data Center vs AI at the Edge

One of the first considerations of deploying smart networks is whether to go with a cloud-based system hosted by a data center or to utilize an edge platform. ComBox and LARGA looked at the advantages of each platform to determine the best way to deploy their passenger counting application.
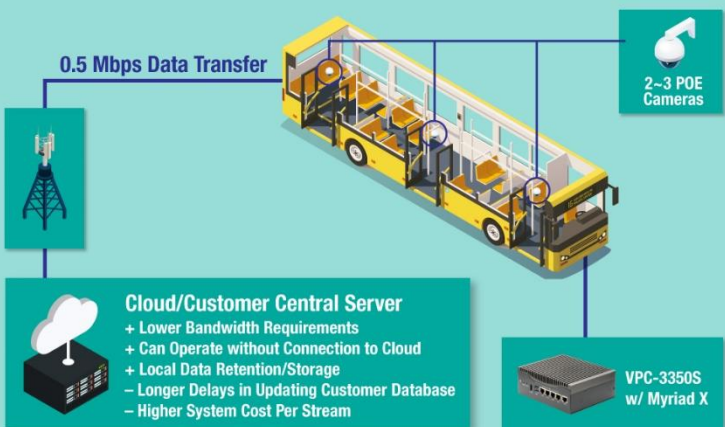
	Data Center Inference	Edge Inference	
Device	ComBox Server with 8 x 8th Gen Intel Core Processors	AAEON VPC-3350S	AAEON VPC-3350S + AI Myriad X
Inference Device	8 x Intel integrated graphics	Intel HD 500	Intel HD 500, Myriad X
Number of Processed Streams	60	2	3
Price in $USD	10,000	850	1,000
Cost of 1 stream over 12 months, $USD	$14 /month	$35 /month	$28 /month
Communication Channel Requirements, Mbps	4 Mbps each RTSP stream	0.5 Mbps data transfer	
Backup communication channel	4 Mbps each RTSP stream (preferably)	Not required	
Impact of Connection Loss	The gap is critical, data loss	The gap is not critical, timing when channel appears	
Delay of getting data in a personal account	10 seconds	Up to 24 hours	

Table 1: Comparing performance and costs of data center and edge operation. Source: ComBox and LARGA.
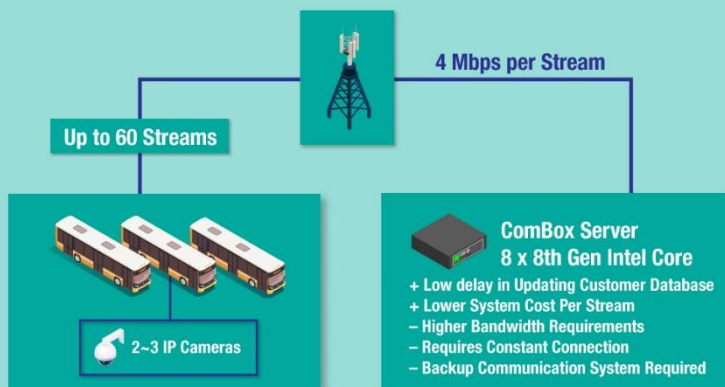
As the table demonstrates, a data center offers pure performance that's greater than an individual edge system, even on a per cost basis. However, there are some critical issues which edge systems can overcome. First, the data center requires the raw video footage to be sent to the server, requiring higher bandwidth connections. Higher bandwidth generally means higher communications provider costs, especially when deploying a cellular based wireless network. Edge systems process the video onboard, only sending data to the central server as needed, reducing the bandwidth required to transfer the data, thus lowering communications costs.

Secondly, a data center-based system requires constant communication. If data transfer is interrupted, then the data center cannot provide the application service. To avoid this, a data center ideally relies on backup communication streams, which adds even more communication costs. Edge systems, on the other hand, do not require constant communication with the central management system in order to function, and should communication service be unreliable or temporarily disconnected, the system can store data onboard and wait until the connection is restored to transfer any necessary data.

Lastly, deploying a data center system still requires installation of communication gateways to connect cameras and sensors to the data center itself. Edge systems can double as gateways, not only providing the computing power needed at the edge, but also being the bridge between edge and the central cloud management system. This helps reduce some of the installation infrastructure required to deploy the system.

Part 2: Intel Advantage

Settling on an Edge-based application, the next step is picking the right hardware solution. There are several key points to look at when determining the right solution, such as costs, I/O configuration, installation requirements, and other points specific to the application in order to provide reliable operation once deployed.

While it is tempting to go with the highest performance dedicated edge systems to power an application, it is usually better to stick with a "good enough" approach. AAEON systems powered by Intel® processors offer a key

advantage in this regard as even low-power CPUs such as the Intel® Atom™ x5 with integrated Intel HD500 graphics offers enough performance to handle two video streams in the passenger counting application developed by ComBox and LARGA. Adding in the AI Core X from AAEON, featuring the Intel® Movidius® Myriad™ X VPU, the VPC-3350S used in this case can process three video streams while also leaving enough processing power on the Intel Atom processor to handle other important tasks, such as communication with the central management system, or collecting vehicle data from the bus.

While dedicated edge systems provide greater performance, they are generally more expensive on a per unit basis or may require modification in order to integrate better into the environment where it is deployed. The VPC-3350S with AI Core X offers four POE ports and is designed for operation within vehicles, whereas a comparable dedicated edge system is much more expensive and may not be suited to the harsh voltage changes that can occur on a vehicle.

	VPC-3350S	VPC-3350S with AI Core X	Comparable Edge Gateway System
Inference Device	Intel HD 500	Intel HD 500 with Myriad X	System on Chip processor
Base processing speed	30 fps	45 fps	40 fps (approx.)
System Cost	$850	$1000	$1200

Table 2: Comparing performance and costs of VPC-3350S with Edge Gateway System.

Part 3: Deploying with OpenVINO and AAEON Solutions

With the hardware advantages known, it's time to look at the software. The Intel® distribution of OpenVINO™ toolkit provides developers with a solution designed to help optimize AI inference performance across a broad range of Intel products, including both Intel processors and/or built-in graphics controllers to dedicated accelerators and VPUs such as the Intel® Movidius® Myriad™ X. Being compatible with a broad range of Intel hardware helps bring flexibility and scalability to AI applications. AAEON helps clients unlock both thanks to a broad range of embedded platforms, but also designing with expandability in mind. The VPC-

3350S, for example, can be deployed as is with the Intel distribution of OpenVINO, or can be scaled up with two Intel Movidius Myriad X VPUs thanks to our AI Core X and AI Core XM modules.

For ComBox and LARGA, there is another significant feature which OpenVINO offered. Originally, these two companies had designed and programmed their inference utilizing another AI software environment. Having trained their models before selecting the hardware to deploy on, migrating from one platform to another may seem like a difficult obstacle to get over. However, OpenVINO toolkit is able to optimize

models trained in other environments to work flawlessly to deploy onto Intel hardware quickly, with little to no training needed.

The Intel distribution of OpenVINO toolkit helps to reduce deployment time and speed up time to market. Compatible with common software inferences such as TensorFlow, and support and sample inferences provided by Intel, it is very easy to jump into the Intel hardware environment whether designing an inference for the first time, or migrating from other hardware systems.

Product Intro

The VPC-3350S Mobile NVR from AAEON provides users with choice, flexibility and customization not offered in other Mobile NVRs. The VPC-3350S offers the Intel® Atom® x5 E3940 processor (formerly Apollo Lake) as standard, with options for Pentium® N4200, Celeron® N3350 and Atom® x7 E3950. The core feature of the VPC-3350S is its four PoE Ports, allowing the system to connect to and power a wide range of devices. The VPC-3350S can also be configured with an integrated AI module featuring Intel® Movidius™ Myriad™ X.

The VPC-3350S utilizes an innovative design providing customers with two configurations to choose from; the compact mobile Industrial system, and the flexible In-Vehicle platform. The Industrial configuration offers an I/O compliment perfect for use in machine vision applications, while the In-Vehicle configuration features an innovative modular design allowing ultimate flexibility for less cost, and greater customization.

The AI Core X mPCIe module features the Intel Movidius Myriad X, a low-power high-performance VPU designed for AI acceleration with Edge Computing. The Intel Movidius Myriad X offers speeds up to 105 fps (80 typical) and over 1 trillion floating point operations as a dedicated neural network accelerator. The AI Core X is compatible with the Intel distribution of OpenVINO Toolkit, and supports TensorFlow and Caffe frameworks.

For developers needing an even more powerful solution, AAEON also offers the VPC-3350AI. Built on the VPC-3350S, AI acceleration performance is pushed even further with a built in AI Core XM module, featuring two Intel Movidius Myriad X VPUs, supporting speeds up to 210 fps (160 typical). With the same features as the VPC-3350S and compatibility with the Intel distribution of OpenVINO Toolkit, the VPC-3350AI is perfect for developers needing more power for their edge inference application.

About ComBox Technology

ComBox Technology (www.combox.io) is a developer and designer of mobile computing centers based on CPU, GPU, VPU and FPGA for solving high-tech tasks, training and execution of neural networks and artificial intelligence. The company has its own unique stack of technologies, successfully used in the implementation of large-scale projects. The team is made of experienced specialists in the field of design and implementation of electronic computing equipment, who have the necessary technical skills for projects of any scale. Specializing in solving complex technical problems since 2005, ComBox Technology has established itself as a reliable partner in both the Russian and international markets.

During its work, the company has participated in more than 200 projects, such as the development and implementation of an impurity control system in water, a system for monitoring the process of reactive ion etching, coordinate control systems, and others, as well as the design and construction of electronic components for various purposes.

About LARGA

LARGA Group of Companies (www.larga.group) was founded in 1998. For over 20 years, LARGA has been successfully working in the field of high-tech, engineering and video analytic solutions. In the field of video services, the company specializes in providing cloud solutions, and is also a supplier and integrator of leading world developers. The LARGA cloud-based video streaming platform (LVS) is software designed for the simultaneous reception, storage and transmission of audio and video streams, as well as their analysis and control. One LVS server supports up to 10,000 streams (cameras). Their own Larga.Videoserver platform allows users to create clusters of media servers to scale to any number of streams (cameras). Technological superiority allowed LARGA to become the exclusive partner of MTS PJSC in the field of video analytics (MTS is part of BIG 3 operators of the Russian Federation, and one of the largest operators in Europe). According to an independent study (https://www.cnews.ru/reviews/rynok_reshenij_dlya_videonablyudeniya), LARGA ranks 11th place among suppliers of Russian integrators providing video services. In addition to video applications, the company provides solutions in the field of RPA, BI, microsensory solutions, as well as automation of production, including the digital twins engineering.

About AAEON

Established in 1992, AAEON has become one of the leading designers and manufacturers of advanced AI, IoT and edge computing platforms. Committed to innovative engineering, AAEON provides Industry 4.0 integrated solutions, hardware and intelligent automated services for premier OEM/ODMs and system integrators worldwide, as well as IoT solution platforms that seamlessly consolidate virtual and physical networks. AAEON designs reliable and high-quality platforms to power AI Edge Computing, Advanced IoT solutions, and the adoption of next generation technology such as 5G based networks. AAEON also offers customized end-to-end services from initial product conceptualization and product development through to volume manufacturing and after-sales service programs. AAEON is also committed to continuously redefining and harmonizing the management and development processes of the industry.

With its constant pursuit of innovation and excellence, AAEON became a member of the ASUS group in 2011, enabling the company to further strengthen its leadership, access advanced technology from ASUS, and leverage resources from within the group. AAEON is poised to offer more diversified embedded products and solutions at higher quality standards to meet world-class design and manufacturing demands in the years to come.

AAEON is an Associate Member of the Intel® Internet of Things Solutions Alliance.