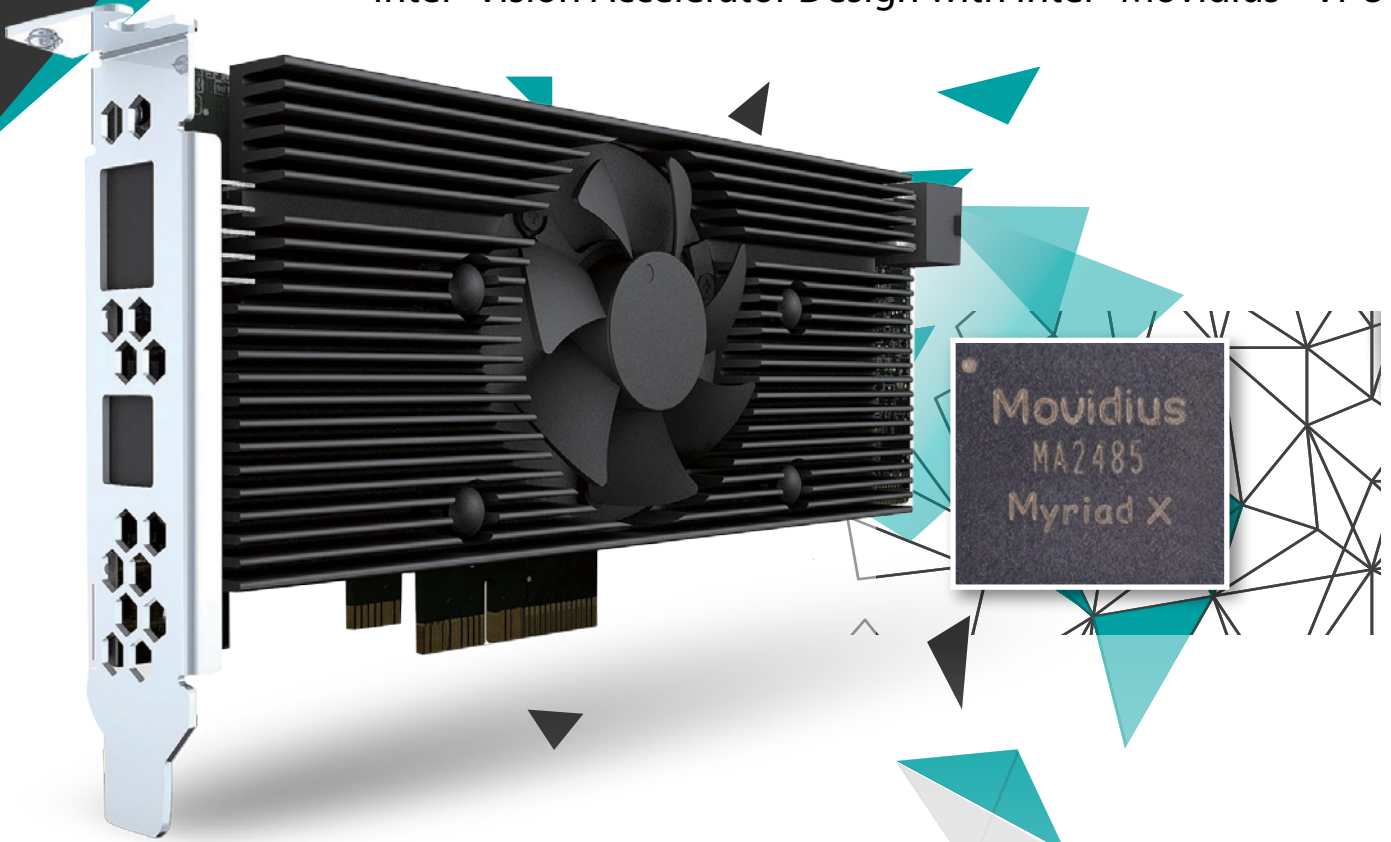




An Intel® Vision Accelerator Design Product

# Mustang-V100-MX8

Intel® Vision Accelerator Design with Intel® Movidius™ VPU

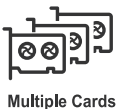


## Accelerate To The Future

Powered by Open Visual Inference & Neural Network Optimization (OpenVINO™) toolkit

- Ubuntu 16.04.3 LTS 64bit, CentOS 7.4 64bit (Support Windows® 10 in the end of 2018 & more OS are coming soon).
- Supports popular frameworks...such as TensorFlow, MxNet, and CAFFE.
- Easily deploy open source deep learning frameworks via Intel® Deep Learning Deployment Toolkit .
- Provides optimized computer vision libraries to quick handle the computer vision tasks.

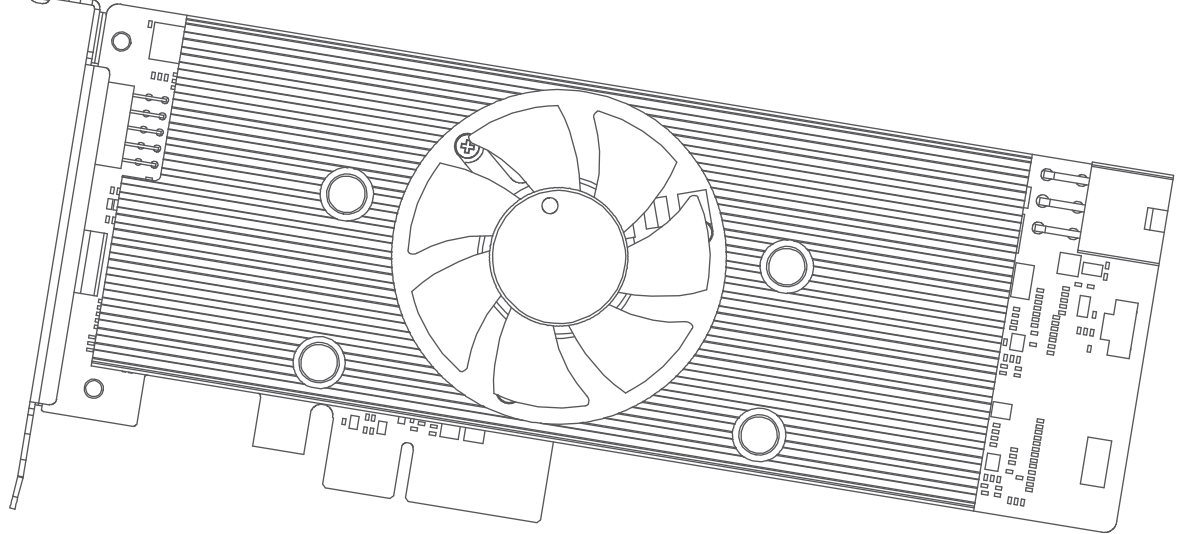
A Perfect Choice for AI Deep Learning Inference workloads



OpenVINO™ toolkit

[www.ieiworld.com](http://www.ieiworld.com)

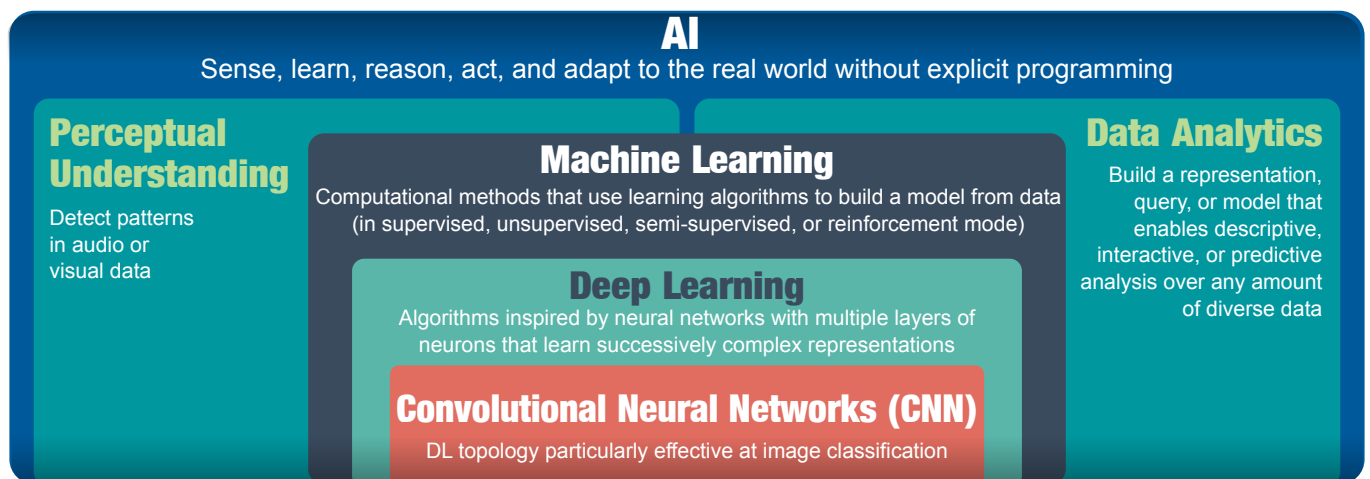




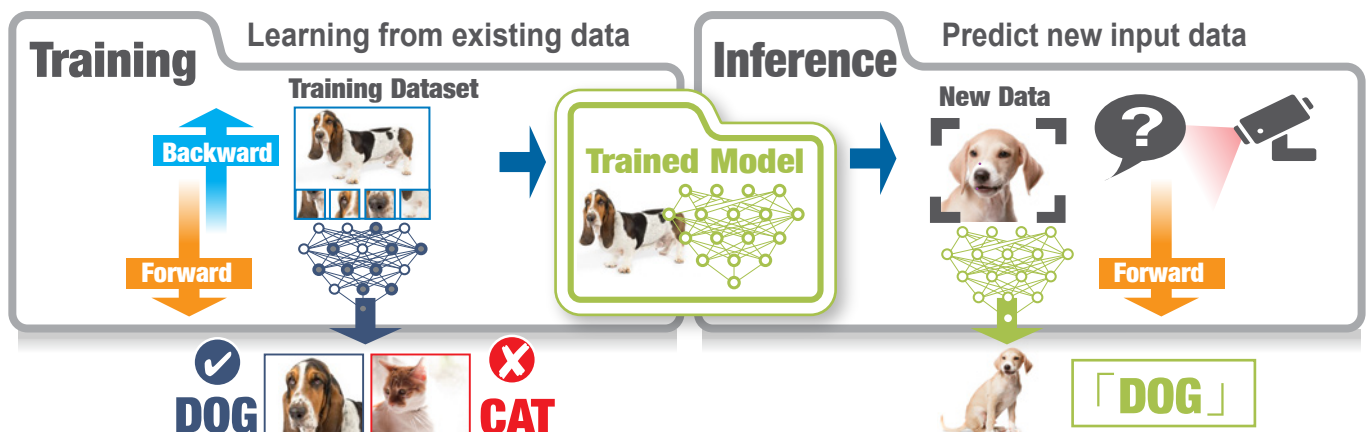
# Mustang-V100-MX8

## Deep learning and inference

Deep learning is part of the machine learning method. It allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. Deep neural network and recurrent neural network architectures have been used in applications such as object recognition, object detection, feature segmentation, text-to-speech, speech-to-text, translation, etc. In some cases the performance of deep learning algorithms can be even more accurate than human judgement.



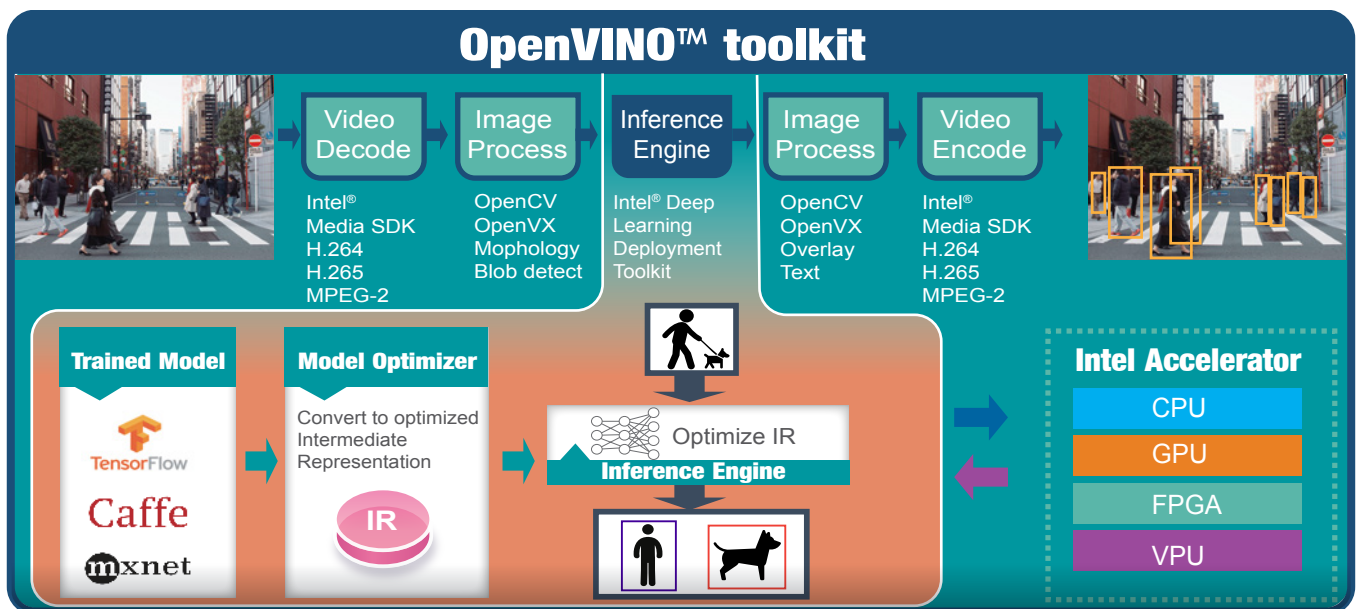
In the past, machine learning required researchers and domain experts knowledge to design filters that extracted the raw data into feature vectors. However, with the contributions of deep learning accelerators and algorithms, trained models can be applied to the raw data, which could be utilized to recognize new input data in inference.



## OpenVINO™ toolkit

Open Visual Inference & Neural Network Optimization (OpenVINO™) toolkit is based on convolutional neural networks (CNN), the toolkit extends workloads across Intel® hardware and maximizes performance.

It can optimize pre-trained deep learning model such as Caffe, MXNET, Tensorflow into IR binary file then execute the inference engine across Intel®-hardware heterogeneously such as CPU, GPU, Intel® Movidius™ Neural Compute Stick, and FPGA.



## IEI Mustang-V100-MX8

In AI applications, training models are just half of the whole story. Designing a real-time edge device is a crucial task for today's deep learning applications.

VPU is short for vision processing unit. It can run AI faster, and is well suited for low power consumption applications such as surveillance, retail, transportation. With the advantage of power efficiency and high performance to dedicate DNN topologies, it is perfect to be implemented in AI edge computing device to reduce total power usage, providing longer duty time for the rechargeable edge computing equipment. AI applications at the edge must be able to make judgements without relying on processing in the cloud due to bandwidth constraints, and data privacy concerns. Therefore, how to resolve AI task locally is getting more important.

In the era of AI explosion, various computations rely on server or device which needs larger space and power budget to install accelerators to ensure enough computing performance.

In the past, solution providers have been upgrading hardware architecture to support modern applications, but this has not addressed the question on minimizing physical space. However, space may still be limited if the task cannot be processed on the edge device.

We are pleased to announce the launch of the Mustang-V100-MX8, a small form factor, low power consumption, and high performance VPU base AI edge computing solution compatible with IEI TANK-870AI compact IPC for those with limited space and power budget.

## Mustang-V100-MX8 AI Application Ready Edge Computing

- **Compact:**

Only half-height ,half-length ,single slot. The conventional accelerator have large form factor which is the drawback for compact edge systems.

- **Low power consumption:**

Compared to CPU or GPU or FPGA, VPU power consumption is extremely efficient (approximate 2.5W for each Intel® Movidius™ Myriad™ X VPU), and this feature is a great advantage in edge computing.

- **Multi-Tasks:**

Eight Intel® Movidius™ Myriad™ X VPU can execute multiple topologies simultaneously.

The graphic features a central image of the Mustang-V100-MX8 AI accelerator card, which is a compact, half-height, half-length, single-slot PCIe card with a black heat sink and a fan. To the left, a stylized brain icon contains the text "AI Application Ready Edge Computing". To the right, three key features are highlighted: "Compact" with a dashed box around a small image of the card; "Low power consumption" with a battery icon and a circular arrow; and "Multi-Tasks" with the Intel Movidius logo and "x8". The background is teal with a network pattern of white dots and lines.

AI Application Ready Edge Computing

Compact

Low power consumption

Multi-Tasks

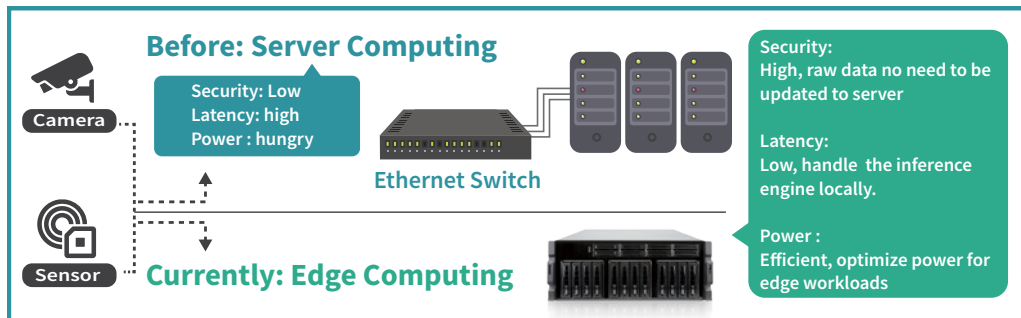
intel Movidius x8

**Mustang-V100-MX8**



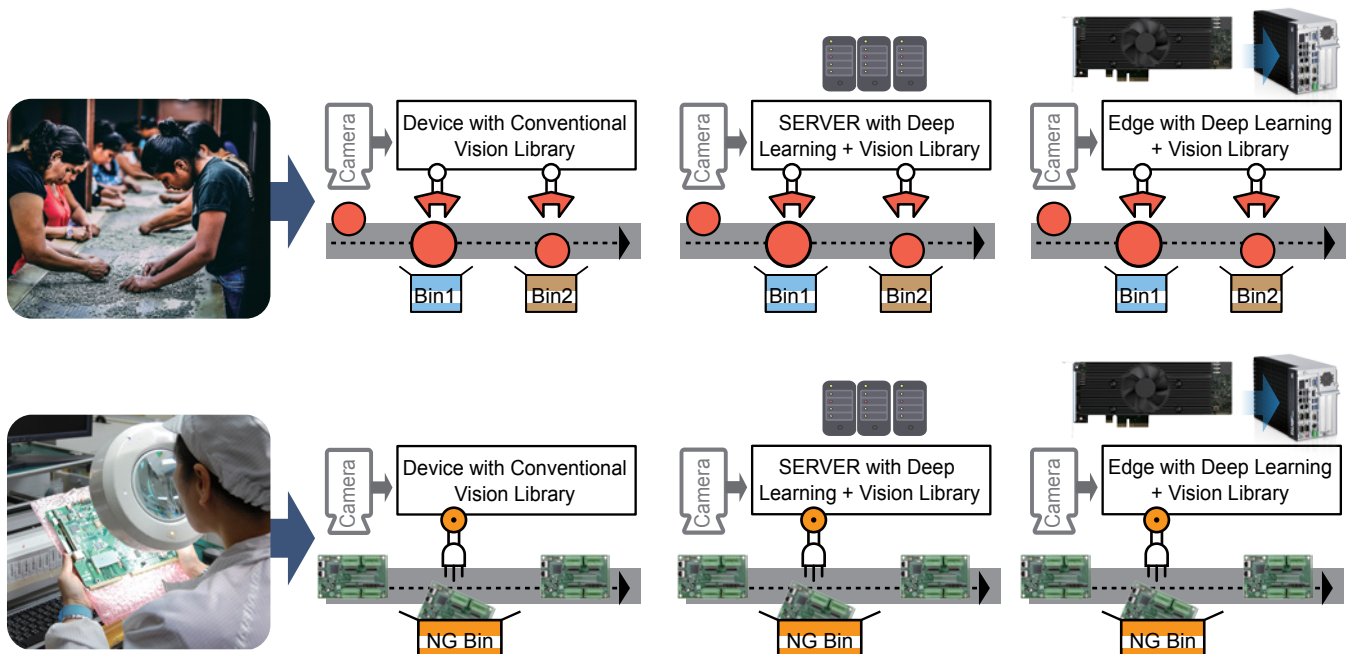
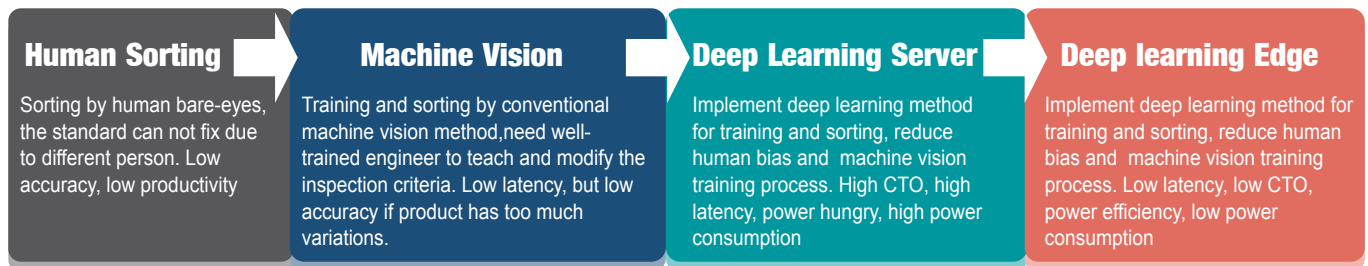
# Edge Computing

Wikipedia defines Edge Computing as “pushing the frontier of computing applications, data, and services away from centralized nodes to the logical extremes of a network.” Today, most of AI technology still rely on the data center to execute the inference, which will increase the risk of real-time application for applications such as traffic monitoring, security CCTV, etc. Therefore, it’s crucial to implement a low-latency real-time edge computing platform.

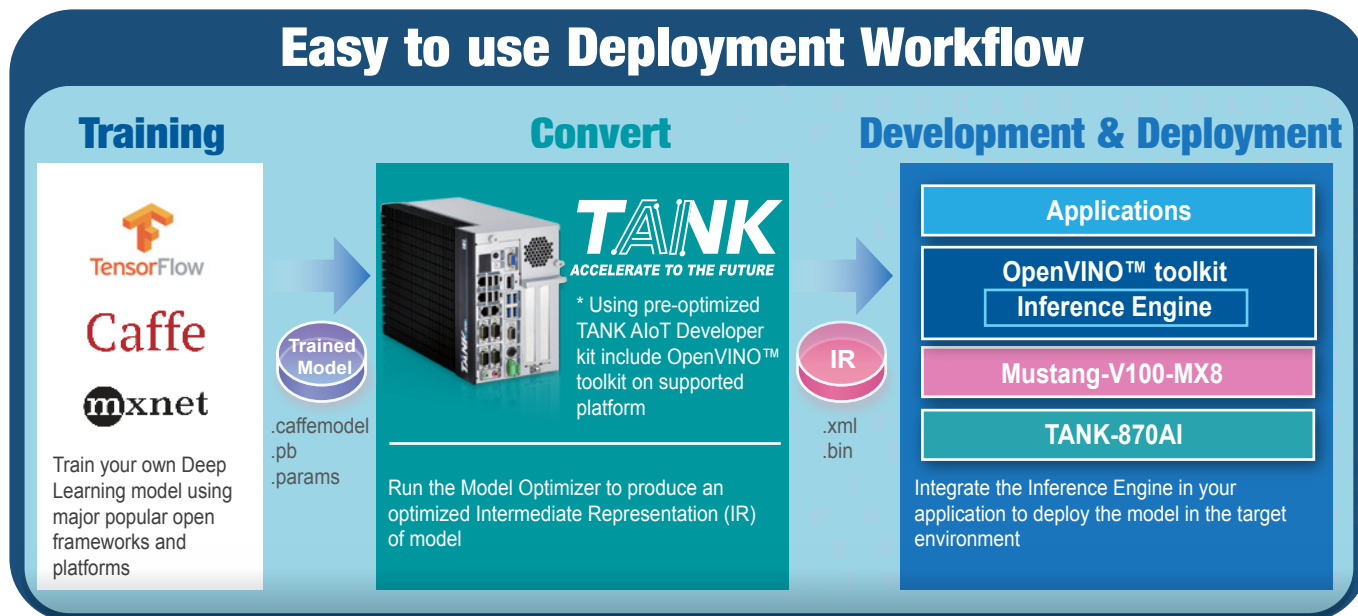


## The advantages of edge computing:

- Reduce data center loading, transmit less data, reduce network traffic bottlenecks.
- Real-time applications, the data is analyzed locally, no need long distant data center.
- Lower costs, no need to implement sever grade machine to achieve non complex applications.



# Software

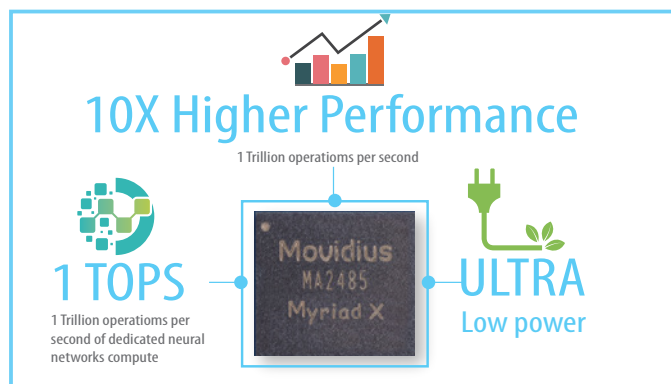


## Mustang-V100-MX8 Software

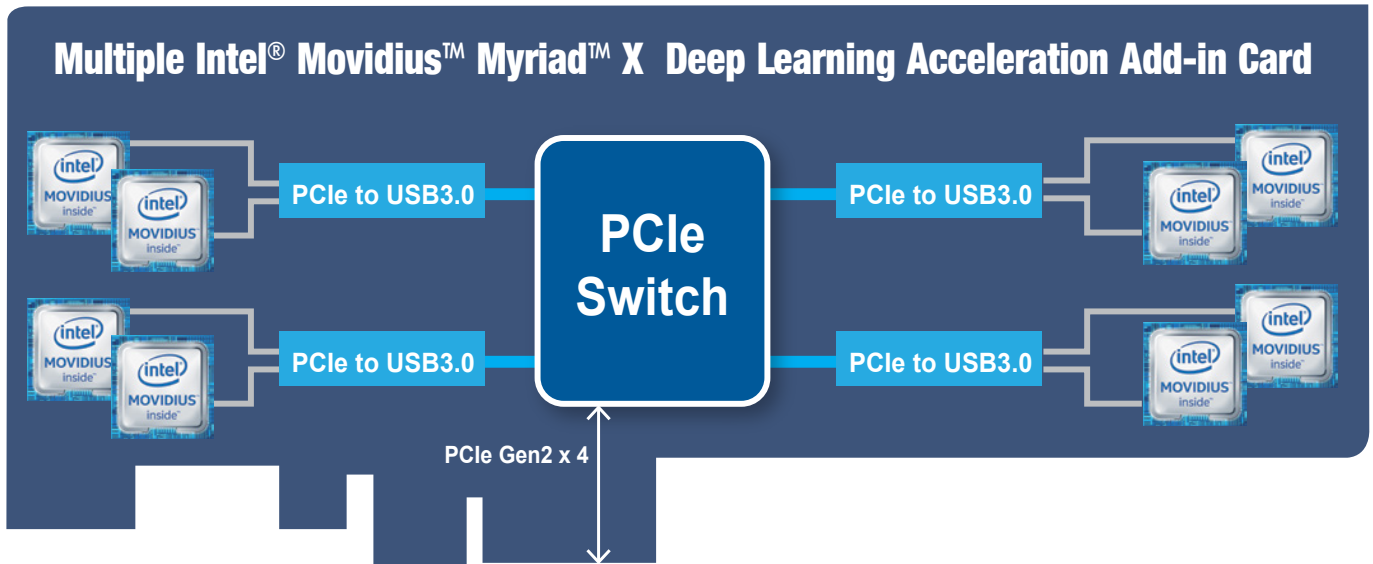
- Operating Systems  
Ubuntu 16.04.3 LTS 64bit, CentOS 7.4 64bit  
(Support Windows 10 in the end of 2018 & more OS are coming soon)
  - OpenVINO™ Toolkit
    - Intel® Deep Learning Deployment Toolkit
      - Model Optimizer
      - Inference Engine
    - Optimized computer vision libraries
    - Intel® Media SDK
    - \*OpenCL™ graphics drivers and runtimes.
    - Current Supported Topologies: AlexNet, GoogleNet V1, Yolo Tiny V1 & V2, Yolo V2, SSD300, ResNet-18, Faster-RCNN. (more variants are coming soon)
  - High flexibility, Mustang-V100-MX8 develop on OpenVINO™ toolkit structure which allows trained data such as Caffe, TensorFlow, and MXNet to execute on it after convert to optimized IR.
- \*OpenCL™ is the trademark of Apple Inc. used by permission by Khronos

### Key Features of Intel® Movidius™ Myriad™ X VPU:

- Native FP16 and fixed point 8 bit support
- Rapidly port and deploy neural networks in Caffe and Tensorflow formats
- End-to-End acceleration for many common deep neural networks
- Industry-leading Inferences/S/Watt performance



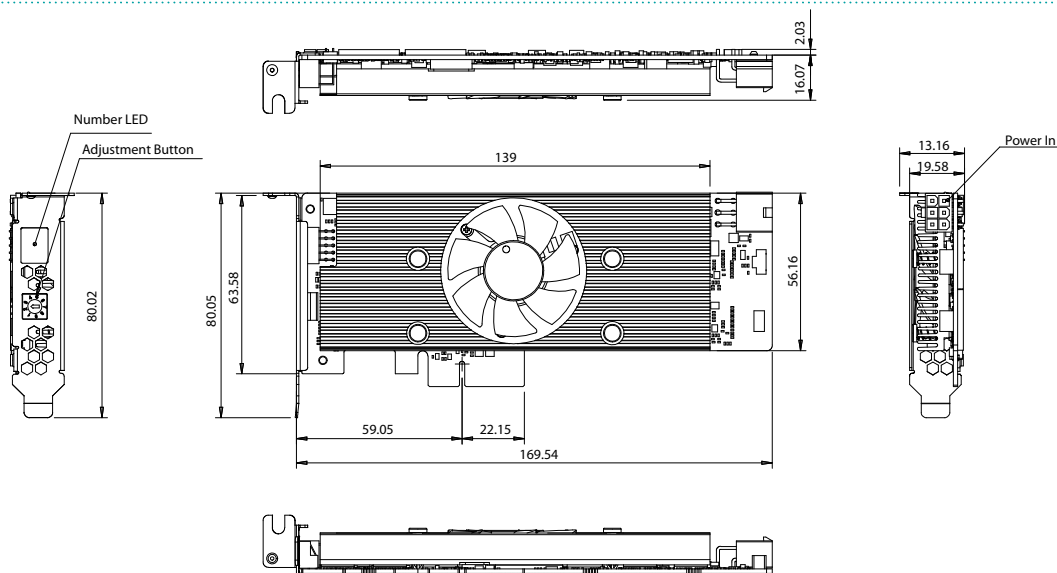
# Hardware



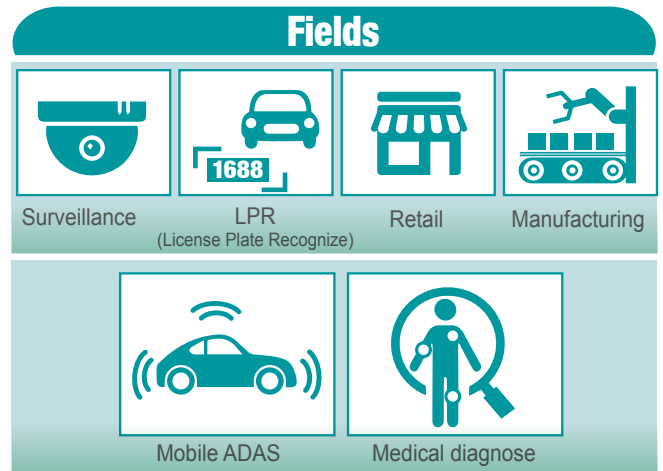
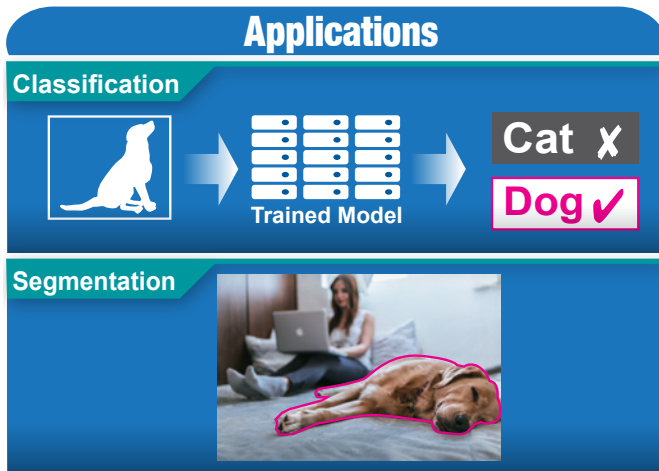
## Mustang-V100-MX8 Block Diagram

- Intel® Movidius™ Myriad™ X VPU delivering up to 1 TOPs of dedicated networks compute
  - Interface: PCIe Gen2 x 4
  - Form Factor: Standard Half-Height, Half-Length, Single-slot
  - Cooling: Active fan.
  - Operation Temperature : 5°C~55°C(ambient temperature)
  - Operation Humidity : 5% to 90% relative humidity
  - Power Consumption: < 30W
  - Power Connector: \*Preserved PCIe 6-pin 12V external power
  - DIP Switch/LED Indicator: Identify card number.
- \*Standard PCIe slot provides 75W power, this feature is preserved for user in case of different system configuration

## Dimensions (Unit: mm)



# Applications



		TS-X77 with GPU	GRANG-C422 with GPU	TANK-870AI with Mustang-F100-A10	TANK-870AI with Mustang-V100-MX8
Applications	Inference Training	○	○		
	Inference Engine	○	○	○	○
	Image Classification	○	○	○	○
	Image Segmentation	○	○	○	
Features	Energy Efficient			○	○
	Low-latency			○	○
	Compact Size			○	○

## Surveillance

### • Traffic

The Mustang-V100-MX8 edge computing device can be utilized to capture data and send traffic to a control center to optimize a traffic light system. It can also perform license plate recognition (LPR) to help law enforcement if vehicles break traffic laws or help parking services identify available parking spaces to assist drivers in congested urban areas.

### • Security

With the algorithms developed using the Mustang-V100-MX8 edge device, trained deep neural networks now have inference capabilities to identify suspicious persons to alert law enforcement or for security departments to early warning scenarios.

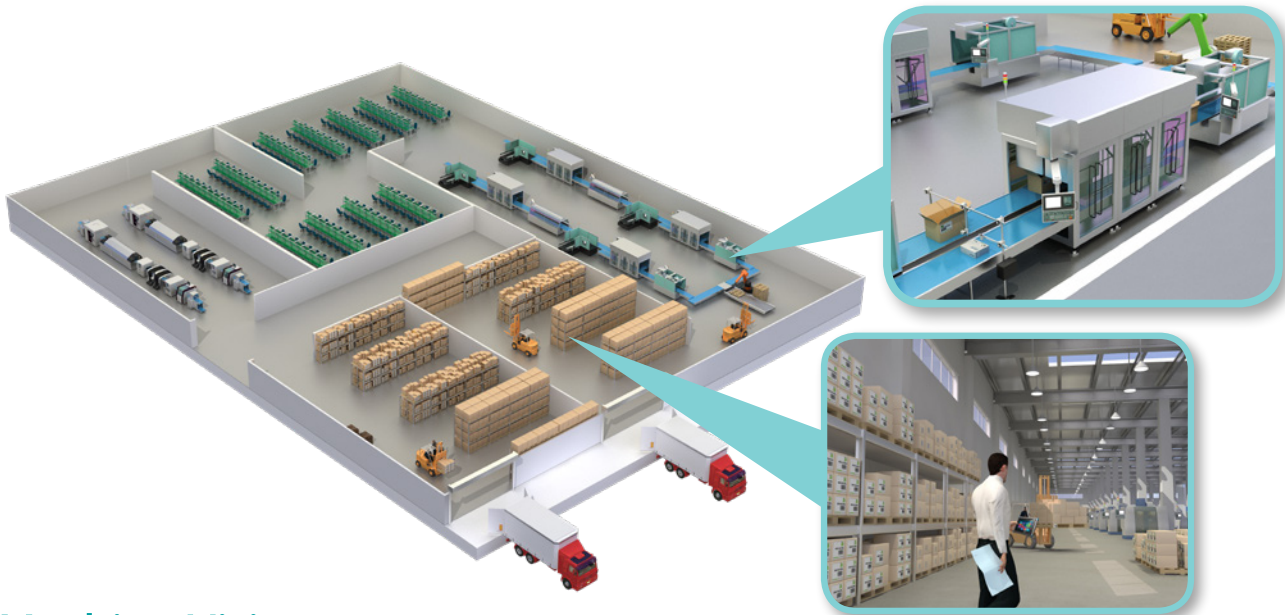




# Industrial Manufacturing

## • Industrial automation

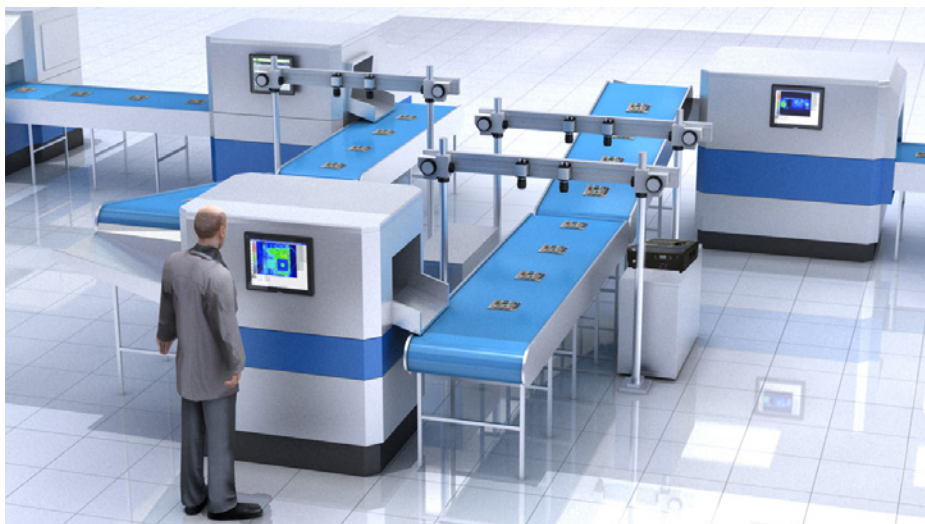
Mustang-V100-MX8 solutions help enable intelligent factories to be more efficient on work order schedule arrangements. In today's production line, sticking to manufacturing schedules is becoming more and more important for business efficiency. From raw material storage to fabrication and complete products, all information from factory such as manufacturing equipment process time and warehouse storage status are essential to achieve production goals. Solutions based on AI technology can produce more detailed, accurate, and meaningful digital models of equipment and processes for product management.



## • Machine Vision

Implementing AI into machine vision makes smart-automation applications easier. Previously, factory AOI needed sophisticated engineers to fine tune inspection parameters such as length, width, diameters and many other specifications that required many adjustments.

The Mustang-V100-MX8 powered using AI technologies supports workloads such as defect detection and quality control to improve production yield.

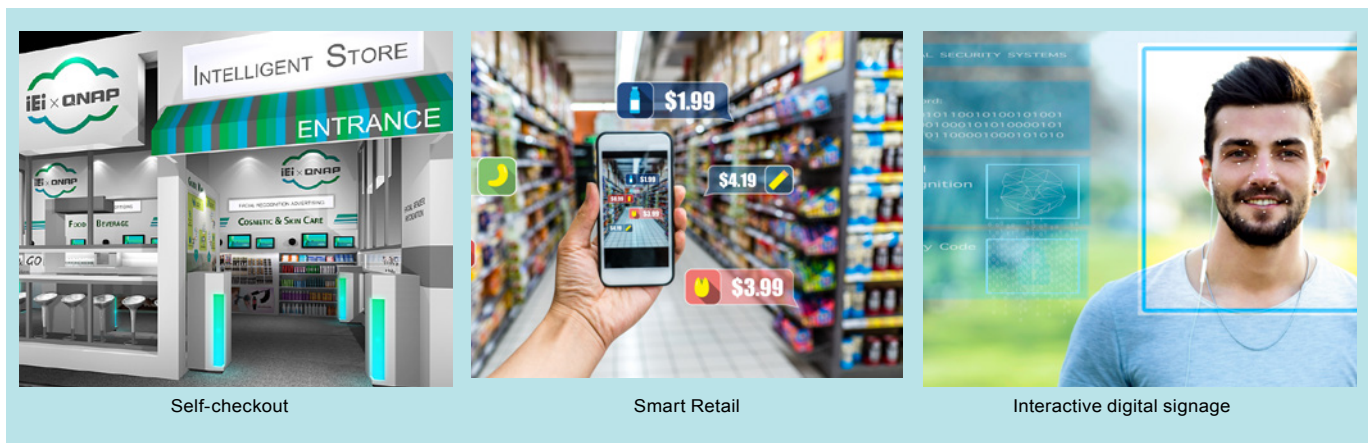


# Retail

## • Smart Retail

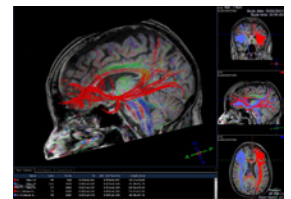
Using the Mustang-V100-MX8 for computer vision solutions at the edge of retail sites can quickly recognize the gender and age of the customers and provide relevant product information through digital signage display to improve product sales and inventory control. Self-checkout can reduce human resource cost so that retail owners can spend more resources on promoting products and understanding business patterns.

In addition, it can help to analyze customer's in-store behavior, and provide customer information based on gender and age to facilitate product positioning. Quickly converting the business intelligence gained and help build better business practices and increase profitability.



## • Medical Diagnostics

With AI based technology, healthcare and medical centers can diagnose, locate and identify suspicious areas such as tumors and other abnormalities more quickly and accurately. Using trained models on the Mustang-V100-MX8 can be used to locate and identify abnormalities with a high degree of accuracy helping doctors and researchers quickly serve the patient.



### Case Study Eye Related Disease (Age-related macular degeneration)

Trained Model .pb      IR .xml .bin

#### Training

The 22K Labeled OCT image data are used to train an image classification model (using Inception v3) to recognize the eye disease.

**GRAND-C422**

#### Convert

The model optimizer is used to convert the trained model to IR file.

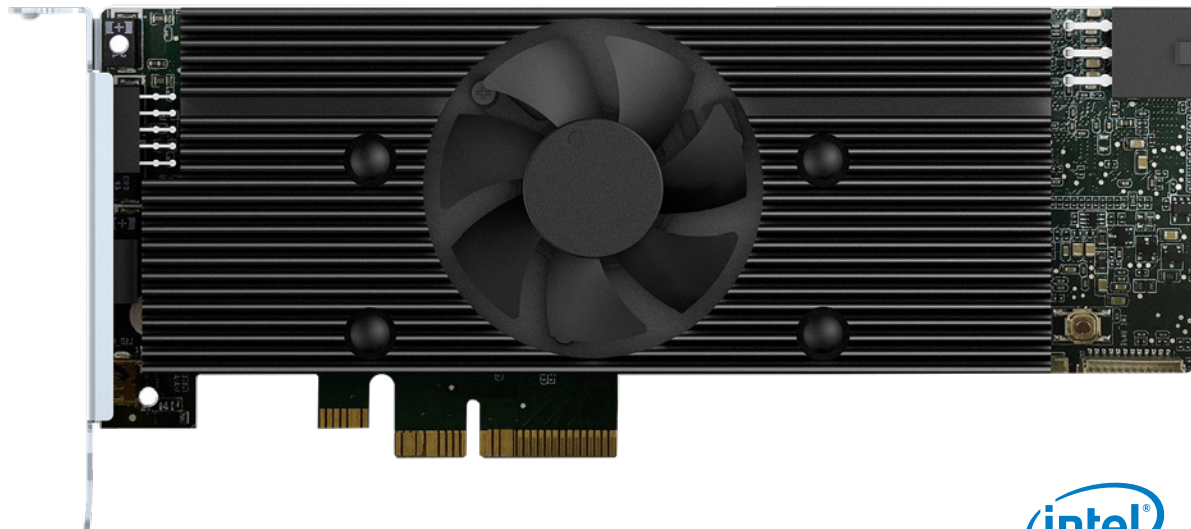
**TANK-870AI**

#### Development & Deployment

An eye disease classification program is developed, and integrated the Inference Engine to gain the great performance and efficiency for age-related macular degeneration classification.

**TANK-870AI** Inference Engine

# Mustang-V100-MX8



## Feature

- Half-Height, Half-Length, Single-slot compact size
- Low power consumption ,approximate 2.5W for each Intel® Movidius™ Myriad™ X VPU.
- Supported OpenVINO™ toolkit, AI edge computing ready device
- Eight Intel® Movidius™ Myriad™ X VPU can execute multiple topologies simultaneously.

## Specifications

Model Name	Mustang-V100-MX8
Main Chip	Eight Intel® Movidius™ Myriad™ X MA2485 VPU
Operating Systems	Ubuntu 16.04.3 LTS 64bit, CentOS 7.4 64bit (Support Windows 10 in the end of 2018 & more OS are coming soon)
Dataplane Interface	PCI Express x4 Compliant with PCI Express Specification V2.0
Power Consumption	<30W
Operating Temperature	5°C~55°C (ambient temperature)
Cooling	Active fan
Dimensions Standard	Half-Height, Half-Length, Single-slot PCIe
Operating Humidity	5% ~ 90%
Power Connector	*Preserved PCIe 6-pin 12V external power
Dip Switch/LED indicator	Identify card number

\*Standard PCIe slot provides 75W power, this feature is preserved for user in case of different system configuration

## Packing List

1 X Full height bracket
1 x External power cable
1 x QIG

## Ordering Information

Part No.	Description
Mustang-V100-MX8-R10	Computing Accelerator Card with 8 x Movidius Myriad X MA2485 VPU, PCIe Gen2 x4 interface, RoHS



# IEI Tank AIoT Development Kit



## Feature

- 6th/7th Gen Intel® Core™ processor platform with Intel® Q170/C236 chipset and DDR4 memory
- Dual independent display with high resolution support
- Rich high-speed I/O interfaces on one side for easy installation
- On-board internal power connector for providing power to add-on cards
- Great flexibility for hardware expansion
- Support Open Visual Inference & Neural Network Optimization (OpenVINO™) toolkit
- Support Ubuntu 16.04 LTS

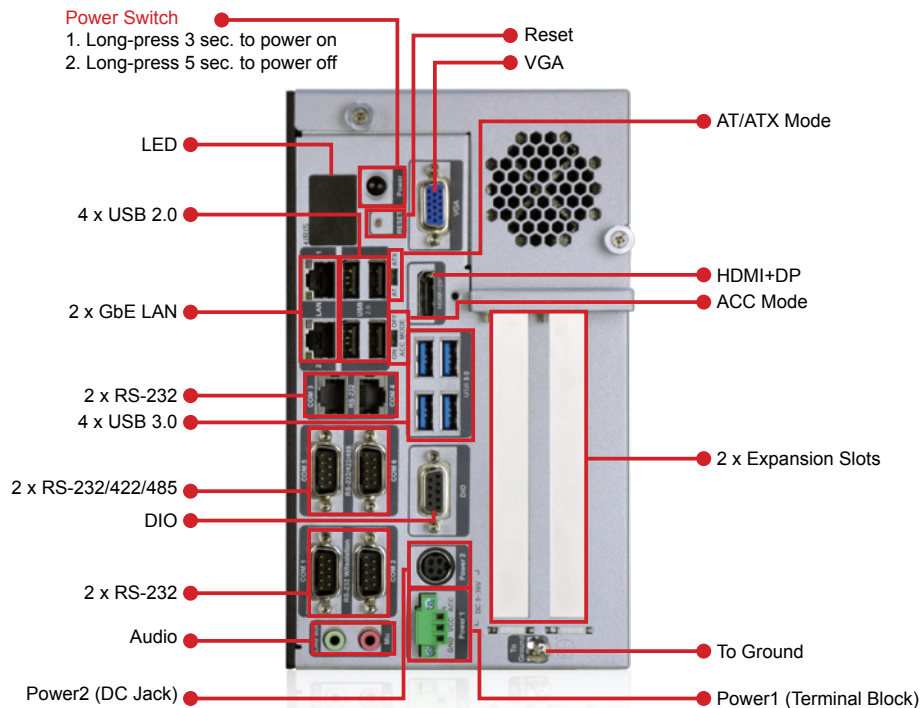


## Specifications

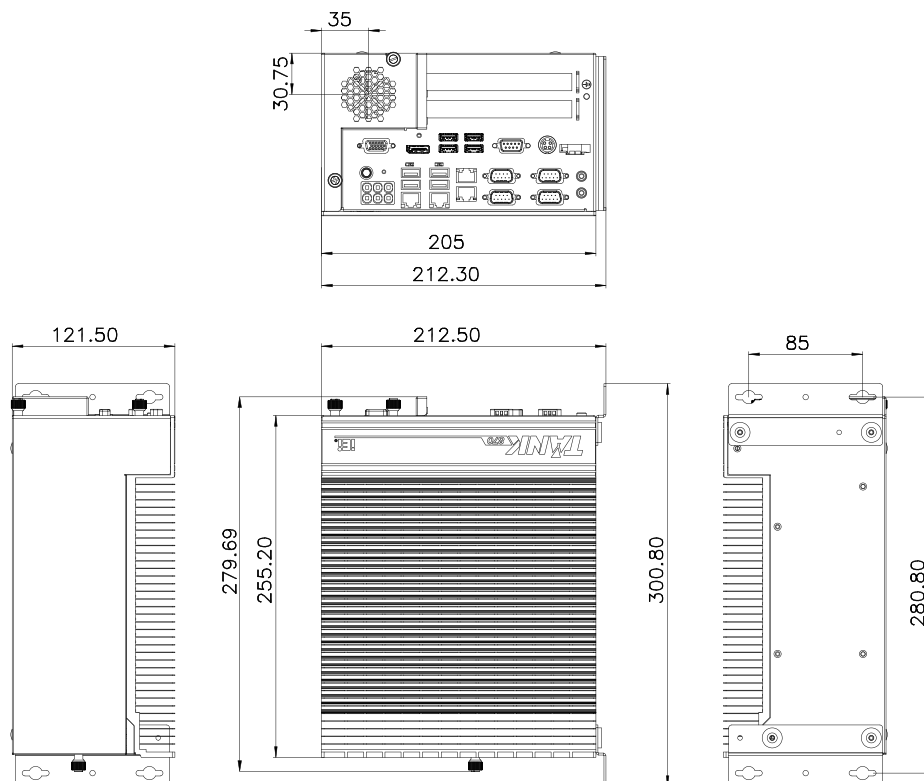
Model Name		TANK AIoT Dev. Kit
Chassis	Color	Black C + Silver
	Dimensions (WxHxD) (mm)	121.5 x 255.2 x 205 mm (4.7" x 10" x 8")
	System Fan	Fan
	Chassis Construction	Extruded aluminum alloys
	Weight (Net/Gross)	4.2 kg (9.26 lbs)/ 6.3 kg (13.89 lbs)
Motherboard	CPU	Intel® Xeon® E3-1268LV5 2.4GHz (up to 3.4 GHz, Quad Core, TDP 35W) Intel® Core™ i7-7700T 2.9GHz (up to 3.8 GHz, Quad Core, TDP 35W) Intel® Core™ i5-7500T 2.7GHz (up to 3.3 GHz, Quad Core, TDP 35W) Intel® Core™ i7-6700TE 2.4 GHz (up to 3.4GHz, quad-core, TDP=35W) Intel® Core™ i5-6500TE 2.3 GHz (up to 3.3GHz, quad-core, TDP=35W)
	Chipset	Intel® Q170/C236 with Xeon® E3 only
	System Memory	2 x 260-pin DDR4 SO-DIMM, 8 GB pre-installed (for i5/i5KBL/i7 sku) 16 GB pre-installed (for i7KBL sku) 32 GB pre-installed (for E3 sku)
Storage	Hard Drive	2 x 2.5" SATA 6Gb/s HDD/SSD bay, RAID 0/1 support (1x 2.5" 1TB HDD pre-installed)
I/O Interfaces	USB 3.0	4
	USB 2.0	4
	Ethernet	2 x RJ-45 LAN1: Intel® I219LM PCIe controller with Intel® vPro™ support LAN2 (iRIS): Intel® I210 PCIe controller
	COM Port	4 x RS-232 (2 x RJ-45, 2 x DB-9 w/2.5KV isolation protection) 2 x RS-232/422/485 (DB-9)
	Digital I/O	8-bit digital I/O, 4-bit input / 4-bit output
	Display	1 x VGA 1 x HDMI/DP 1 x iDP (optional)
	Resolution	VGA: Up to 1920 x 1200@60Hz HDMI/DP: Up to 4096x2304@24Hz / 4096x2304@60Hz
	Audio	1 x Line-out, 1 x Mic-in
	TPM	1x Infineon TPM 2.0 Module
	Expansions	Backplane
PCIe Mini		1 x Half-size PCIe Mini slot 1 x Full-size PCIe Mini slot (supports mSATA, colay with SATA)
Power	Power Input	DC Jack: 9 V~36 V DC Terminal Block: 9 V~36 V DC
	Power Consumption	19 V@3.68 A (Intel® Core™ i7-6700TE with 8 GB memory)
	Internal Power output	5V@3A or 12V@3A
Reliability	Mounting	Wall mount
	Operating Temperature	Xeon® E3 -20°C ~ 60°C with air flow (SSD), 10% ~ 95%, non-condensing i7-7700T -20°C ~ 35°C with air flow (SSD), 10% ~ 95%, non-condensing i5-7500T -20°C ~ 45°C with air flow (SSD), 10% ~ 95%, non-condensing i7-6700TE -20°C ~ 45°C with air flow (SSD), 10% ~ 95%, non-condensing i5-6500TE -20°C ~ 60°C with air flow (SSD), 10% ~ 95%, non-condensing
	Operating Vibration	MIL-STD-810G 514.6 C-1 (with SSD)
	Safety/EMC	CE/FCC/RoHS
OS	Supported OS	Linux Ubuntu 16.04 LTS



# Fully Integrated I/O



# Dimensions (Unit: mm)



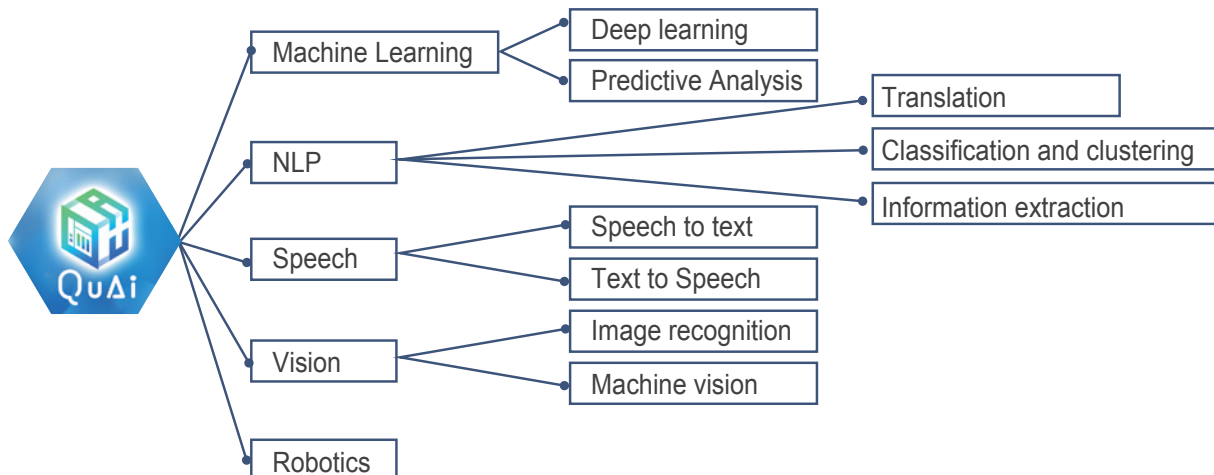
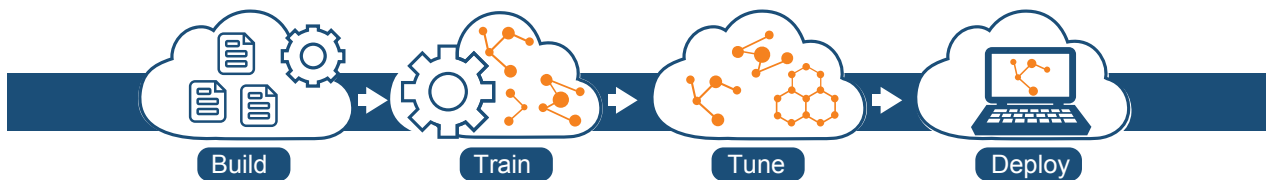
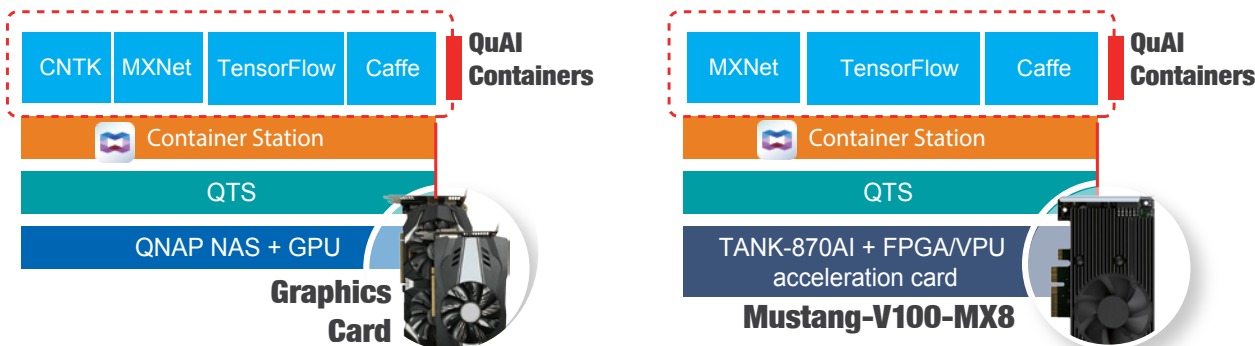
# QNAP QuAI

QNAP Systems, Inc. is a wholly owned subsidiary under IEI Group which specializes in providing networked solutions for file sharing, virtualization, storage management and surveillance applications. QNAP implements deep learning method into its main product Network-Attached Storage (NAS) to enable more AI applications.

QNAP QuAI enables data scientists and developers to quickly build, train, optimize, and deploy machine-learning models with high-performance machine-learning algorithms that come with a wide range of supported AI frameworks.

QuAI is an integrated platform to empower your AI-related computing needs. QNAP NAS now supports graphics cards, Intel FPGA acceleration card, and Intel VPU acceleration card; from training to edge computing, it provides additional computational power and end-to-end solution to help run your tasks more efficiently. On top of that, software enhancements are also provided to help you deploy your solutions faster than ever.

Major frameworks and libraries are supported through Container Station (1.8 and later), such as Caffe, MXNet, TensorFlow, CNTK and NVIDIA CUDA. You can easily migrate existing containerized solutions to the QuAI platform, or start a new one with QuAI, to fully realize benefits of cognitive technologies.



# 2018



**\*Specifications are subject to change without prior notice.**

#### Headquarters

威強電工業電腦 IEI Integration Corp.

No. 29, Zhongxing Rd., Xizhi Dist., New Taipei City 221, Taiwan  
TEL : +886-2-86916798 / +886-2-26902098 FAX : +886-2-66160028  
sales@ieiworld.com www.ieiworld.com

#### America

IEI Technology USA Corp.

138 University Parkway, Pomona, CA 91768  
TEL : +1-909-595-2819 FAX : +1-909-595-2816  
sales@usa.ieiworld.com usa.ieiworld.com

#### China

威強電工業電腦 IEI Integration (Shanghai) Corp.

515, Shen Fu Rd., Xin Zhuang Industrial Develop Zone, Shanghai, 201108, China  
TEL: +86-21-3116-7799 FAX: +86-21-3462-7797  
sales@ieiworld.com.cn www.ieiworld.com.cn