IEI AI Solution and Vision Acceleration Card

Accelerate To The Future

Outline

- AI frameworks and tools
- Intel® OpenVINO™ Toolkit
- AI Roadmap
- AI Accelerate Cards
- AI Accelerate Platforms
- Applications

AI frameworks and tools

Frameworks

Caffe
Caffe2
CNTK
MXNet
Neon
PyTorch
Tensorflow
…

Topology / Model architectures

Image Classification
AlexNet, VGG16, GoogLeNet, ResNet, MobileNet, etc.

Object Detection
SSD, Yolo v1/v2/v3,R-FCN, RCNN, Faster RCNN, etc.

Image Segmentation
SegNet, U-Net, FCN, DeepLab v1/v2, etc.

Face Detection / Recognition
MTCNN, DeepFace, Facenet, etc.

Video Classification
RNN, LSTM, etc.

Speech Recognition
DeepVoice, WaveNet, etc

Training Platform

intel Intel® MKL
NVIDIA® CUDA

Inference Platform

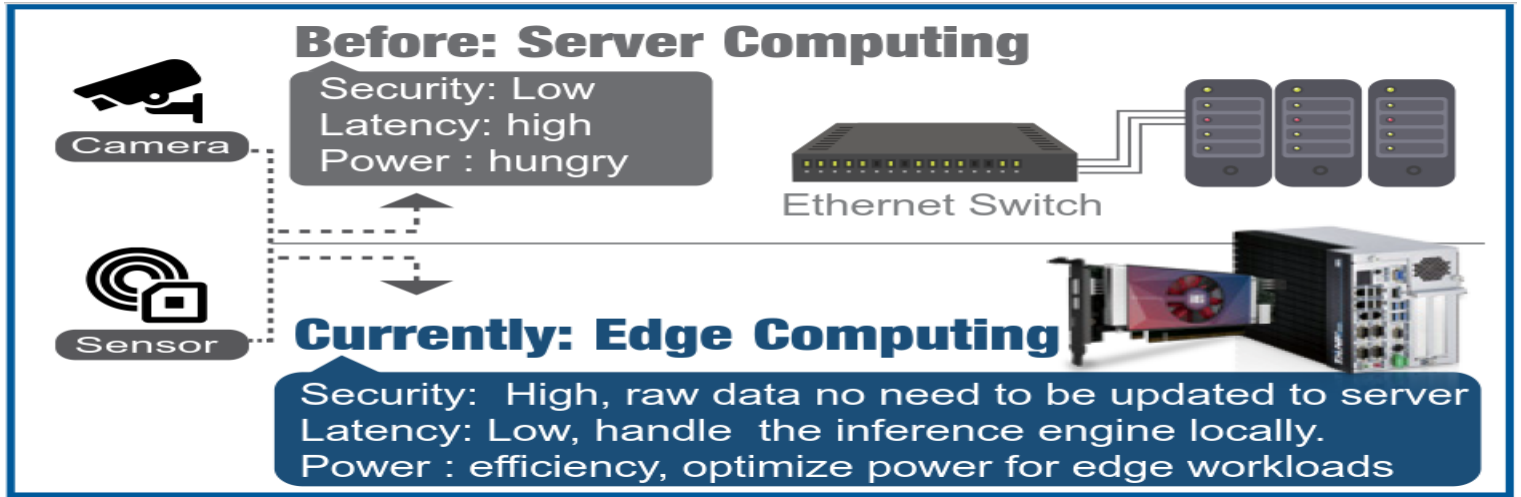
intel Intel OpenVINO
TensorRT

OS

Linux, Windows 10

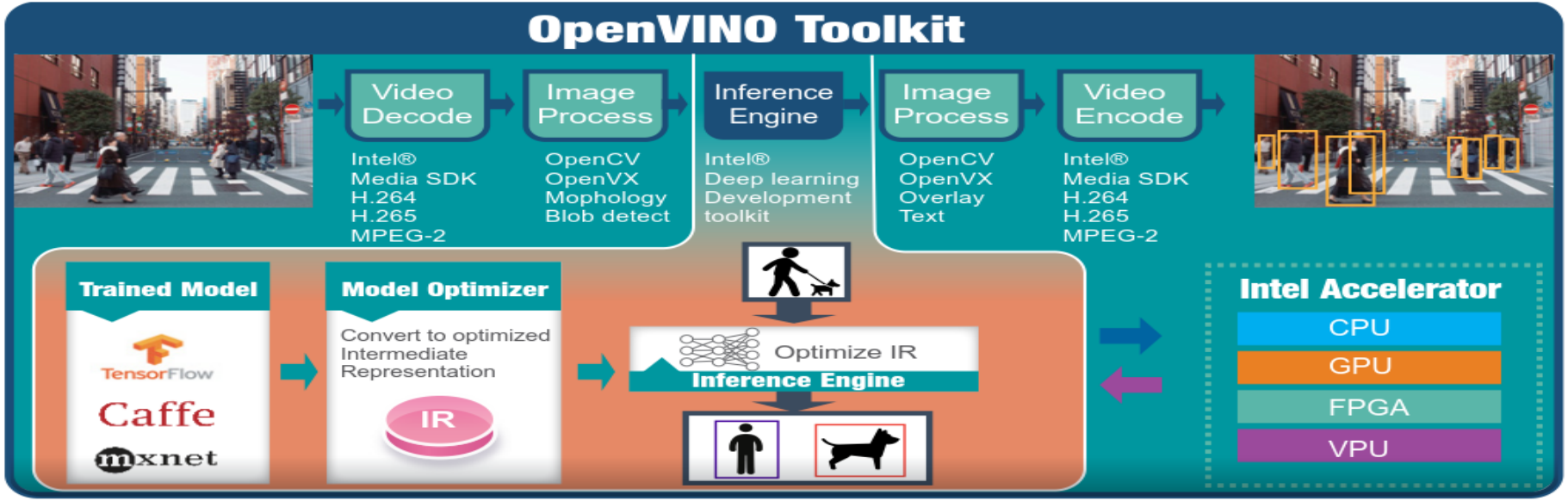
Inference Edge Computing

- 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.



OpenVINO™ Toolkit

- Allows users to easily deploy open source deep learning frameworks for Intel architecture to realize the concept of one SDK for multiple acceleration platforms (CPU, GPU, FPGA, VPU).
- OpenVINO™ toolkit can optimize pre-trained deep learning model such as Caffe, MXNET, Tensorflow into IR binary file then run the inference engine.



Mustang AI Accelerate Selections

Accelerator

CPU

Mustang-200

- Two Intel Core ULT
- 4 DDR4 UDIMM
- 2 NVMe, 2 eMMC
- 10GbE based
- PCIe x4 interface

Intel Kabylake ULT

Accelerator

FPGA

Mustang-F100-A10

- Intel Arria 10 GX 1150 FPGA
- PCIe Gen3 x 8
- Low profile , half size

Intel FPGA

Accelerator

VPU

Mustang-V100-MX8

- Intel Movidious solution
- 8 x Myriad X VPU
- PCIe Gen2 x4
- Low profile , half size

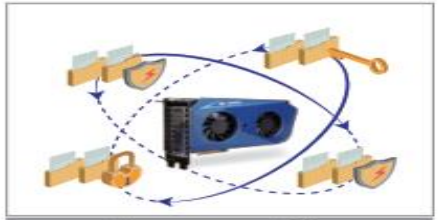
Intel VPU

Mustang-200

Mustang-200

- 10 Gigabit Ethernet Based x86 Computing Nodes support decentralized computing architecture.
- Perfectly Integrated QNAP QTS-Lite provides a flexible and secure developing environment
- Support Virtualization technology, Virtualization Machine (VM) & Docker Container technology
- Fit standard server, compatible with PCI-Express x4, x8, x16
- Increase computing power, support decentralize computing
- Achieve higher densities computing and lower the total cost.

Video Transcoding/ Live Streaming



Numerous File Encryption/Decryption

Numerous Face Analysis

Numerous Car Plate Analysis

Mustang-F100-A10

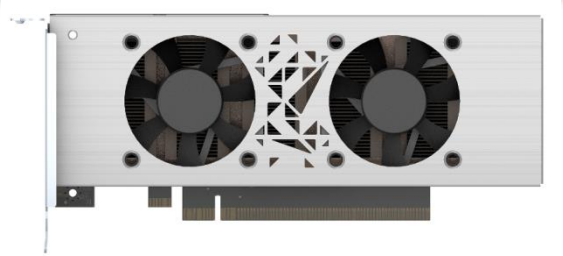
Mustang-F100-A10

Compact Size: Half-height, Half-Length, double slot, compact size.

Low latency: Algorithms implemented into FPGA provide deterministic timing, with latencies one order of magnitude less than GPUs.

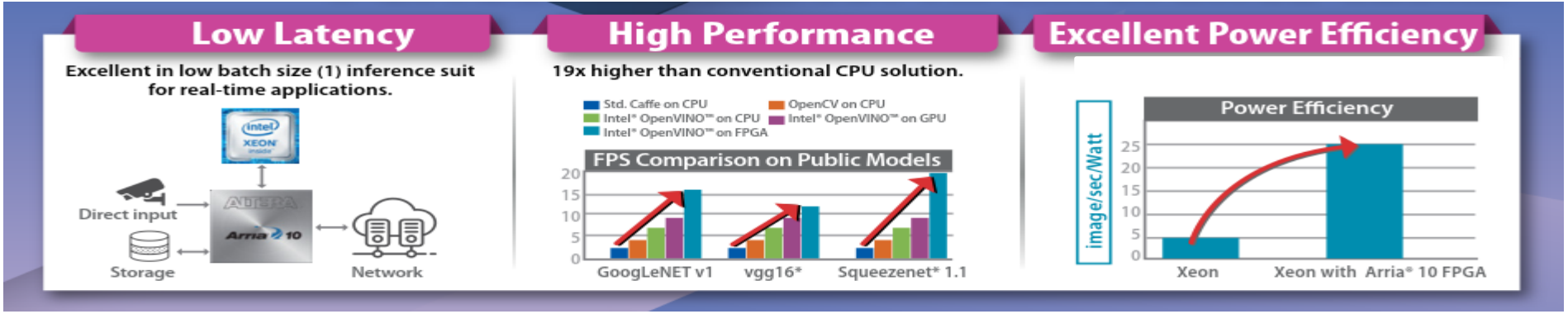
Low power consumption: (25images/Sec/Watt)

Compared to CPU or GPU, FPGA power consumption is extremely efficient.



PCI Express x 8 AI acceleration card, Low profile, Intel® Arria10 FPGA, RoHS

- Power economy
- 1.4 TFLOPS.



Mustang-V100-MX8

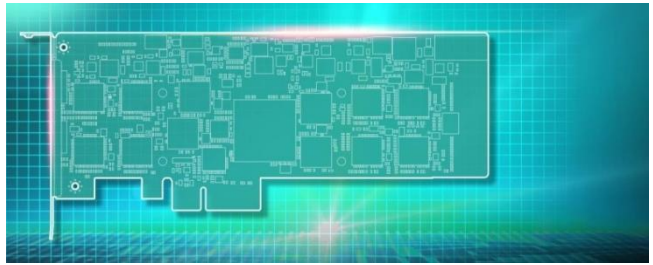
Mustang-V100-MX8

Multi-Tasks: Eight myriad X chips can execute eight topologies simultaneously.

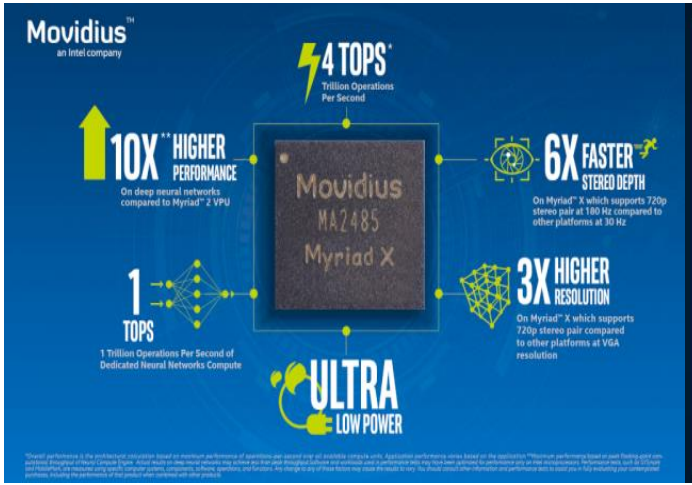
Compact Size: Half-height, half-length , single slot.

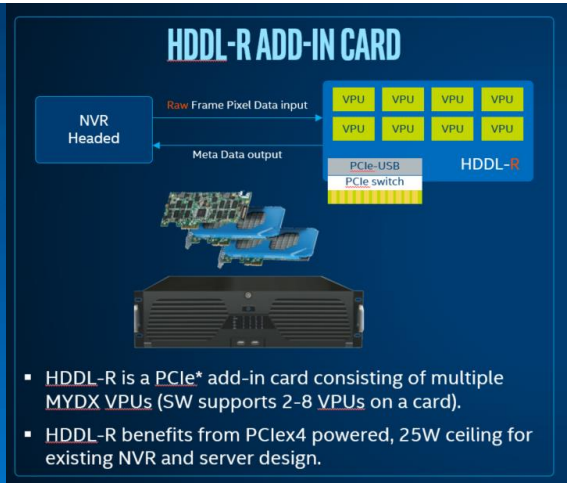
Low power consumption:

approximate 2.5W for each Myriad X Soc.



Mustang-V100-RX8-R10





PCI Express x 4 AI acceleration card, Low profile, 8 Intel® Movidius™ Myriad ™ X VPU, RoHS

- One card for eight inference
- Power economy
- 1 TFLOPS for each chips, 8 TFLOPS for one card.

Mustang-V100-mPCIe, M.2

- **Intel® Movidius™ Myriad™ X mPCIe**
  - 2x MYDX
  - Intel® OpenVINO™ Toolkit
  - Available Q3, 2019
  - Dimension: 30 x 50mm

- **Intel® Movidius™ Myriad™ X M.2 A/E Key**
  - 1x MYDX
  - Intel® OpenVINO™ Toolkit
  - Available Q3, 2019
  - Dimension: 22 x 42mm

- **Intel® Movidius™ Myriad™ X M.2 B/M Key**
  - 2x MYDX
  - Intel® OpenVINO™ Toolkit
  - Available Q3, 2019
  - Dimension: 22 x 60mm

AI Accelerate Card Family

Interface		CPU	FPGA	VPU x8	VPU x4	VPU x2	VPU x1
PCIe	N/A	**Mustang-200** intel Kabylake CPU	**Mustang-F100-A10** Intel® Arria10 FPGA 2018/10/B	**Mustang-V100-MX8** Myriad X 2018/10/E	**Mustang-V100-MX4** Myriad X		
	10G SFP				**Mustang-V100-MX4 -10G2SF** Myriad X + SFPx2		
	10G RJ45				**Mustang-V100-MX4 -10G1T** Myriad X + RJ45x1		
	M.2				**Mustang-V100-MX4 -2P** M.2 M key x2		
MPCIE						**MYX-MPCIE-MX2** Myriad X	
M.2						**MYX-M2BM-MX2** Myriad X, BM Key	**MYX-M2AE-MX2** Myriad X, AE Key

MP	**Developing**

QNAP x IEI offerings - Making AI possible!

Caffe2 Microsoft CNTK

mxnet TensorFlow

Docker

Linux

iEi H/W

Industrial AI Training

Intel Xeon W
GRAND-C422-20D

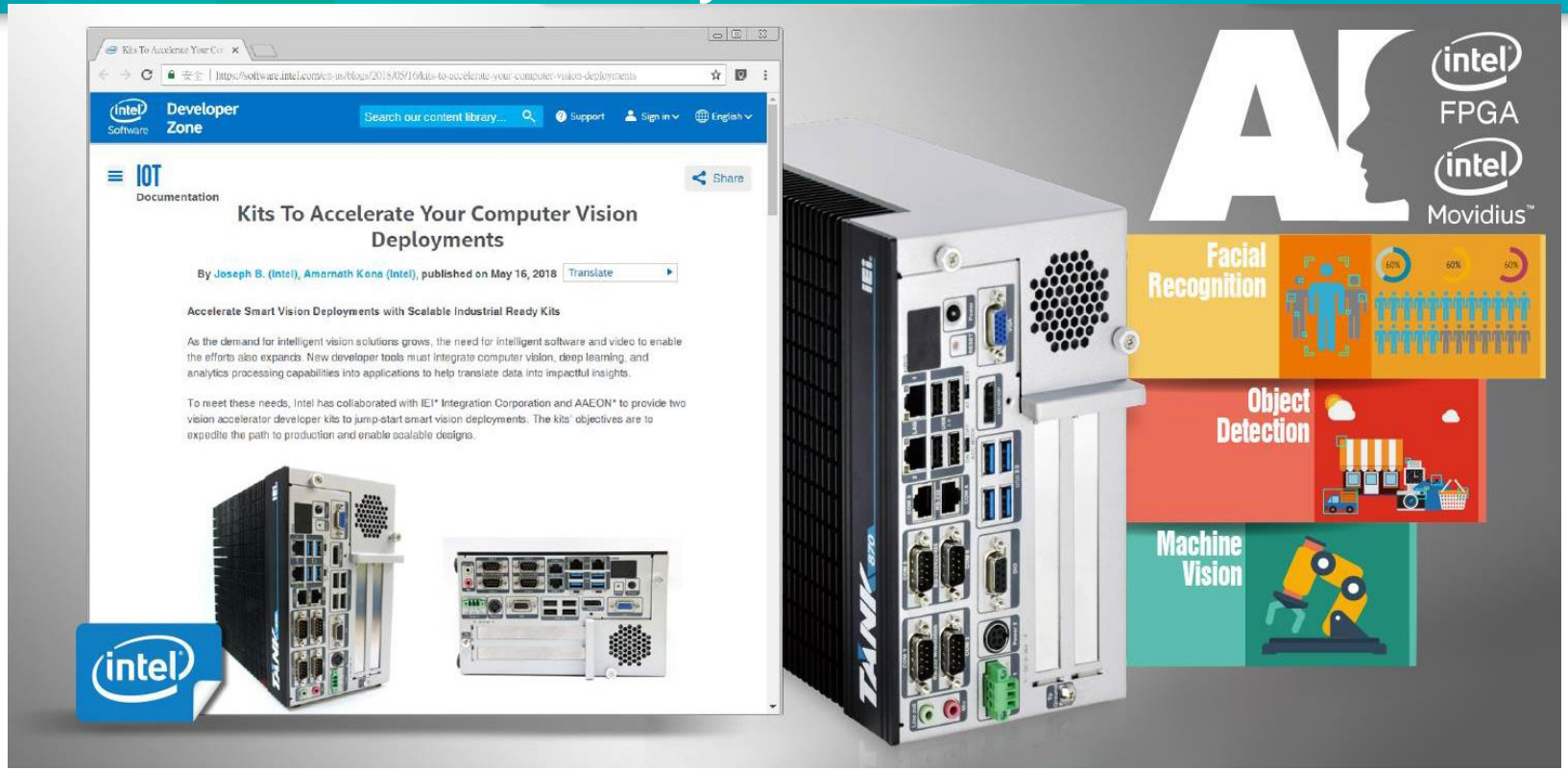
Industrial AI Inference

Intel Skylake
TANK-870AI

Intel Skylake
PAC-400AI

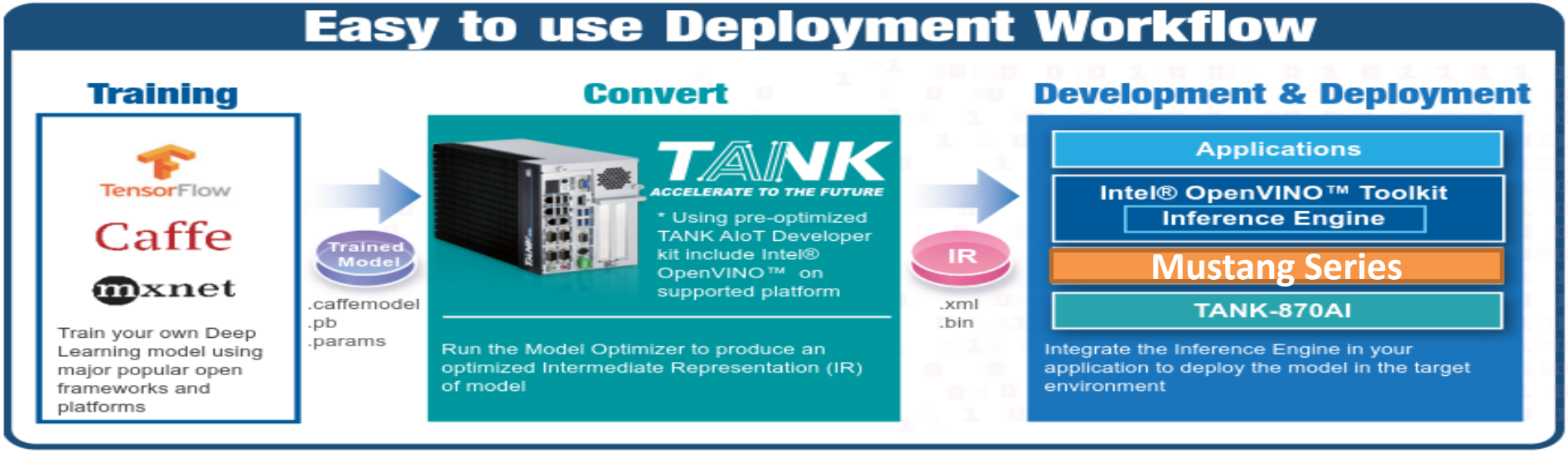
Intel Coffee Lake
PACK-500AI

AI Inference System TANK-870AI

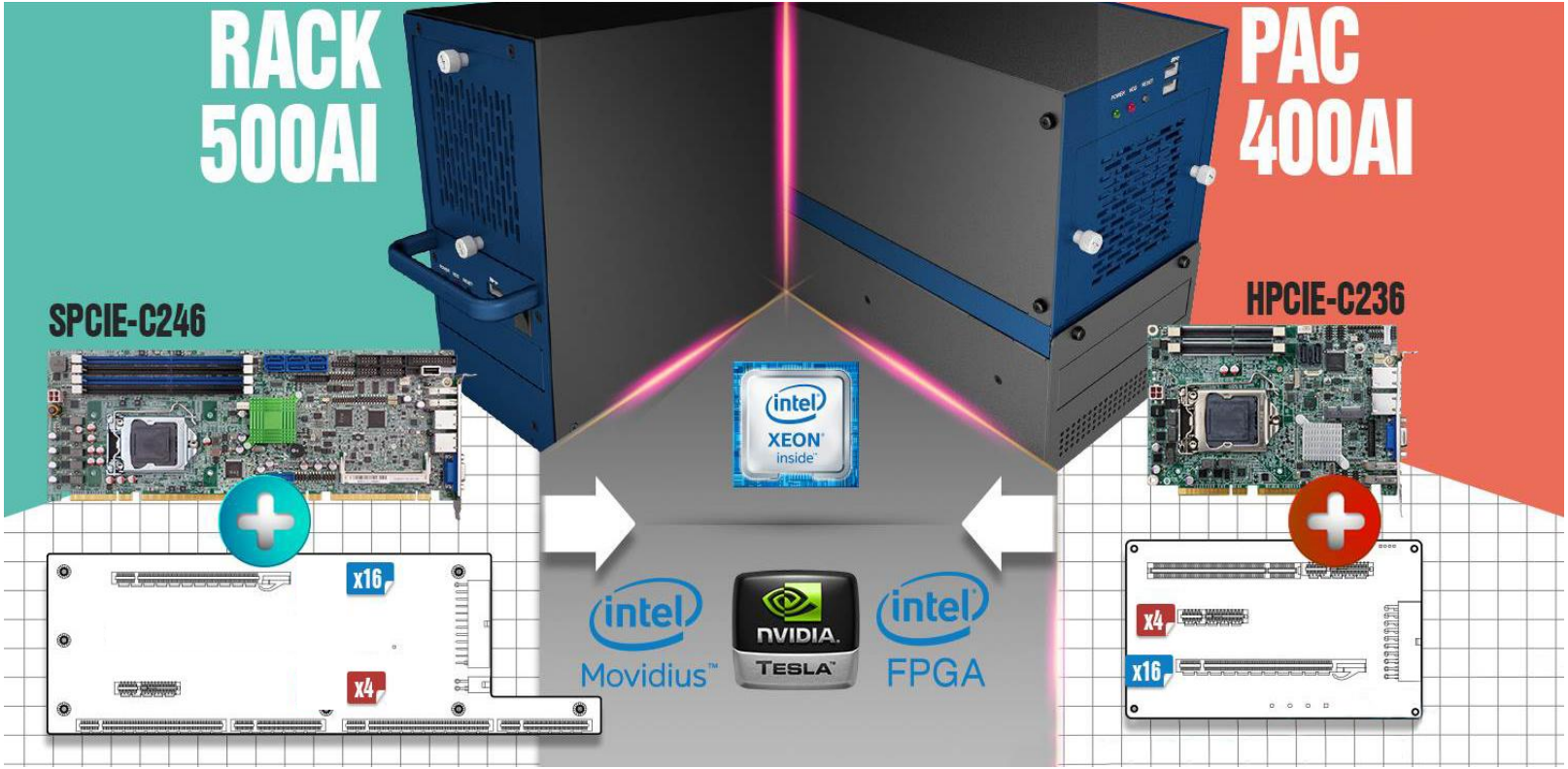


AI Inference System TANK-870AI

- TANK-870AI
  - Deep learning inference ready.
  - CPU VPU FPGA accelerating.



AI Inference System RACK-500AI & PAC-400AI



Applications

- **Retail**
  - Self checkout
  - Interactive digital signage
  - Customers behavior analyze

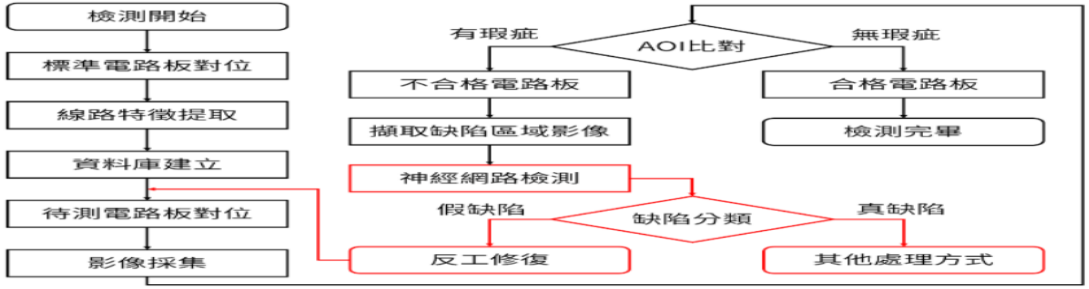
Self-checkout

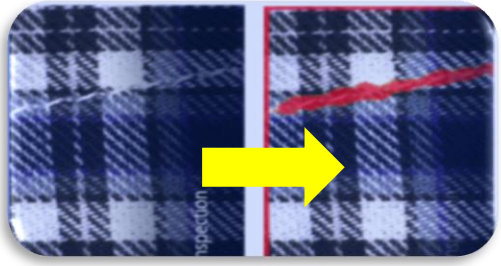
Applications

Machine Vision

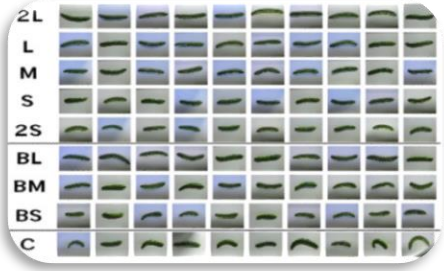
- Assist conventional AOI to double check fail products.
- Food, textile industry which does not have consistence features.

AOI





Textile Defect

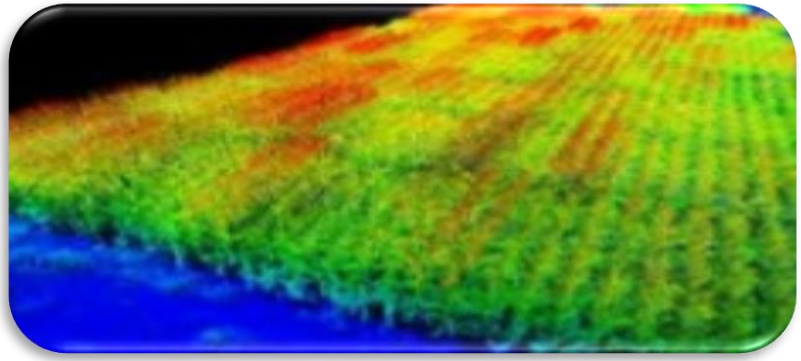


Food Classification

Applications

- **Agriculture**
  - Livestock Monitoring
  - Agriculture Robots
  - Drone Analytics
  - Precision Farming

3D image drone, pesticide spray

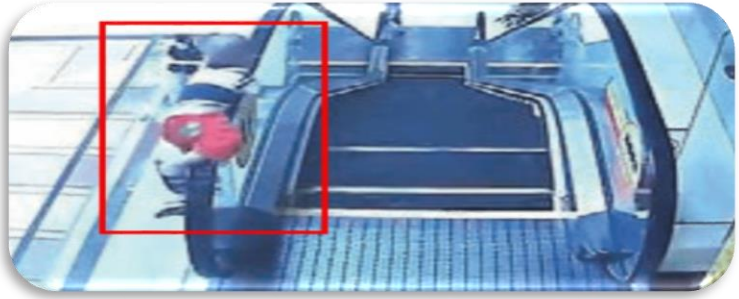


Condition monitoring

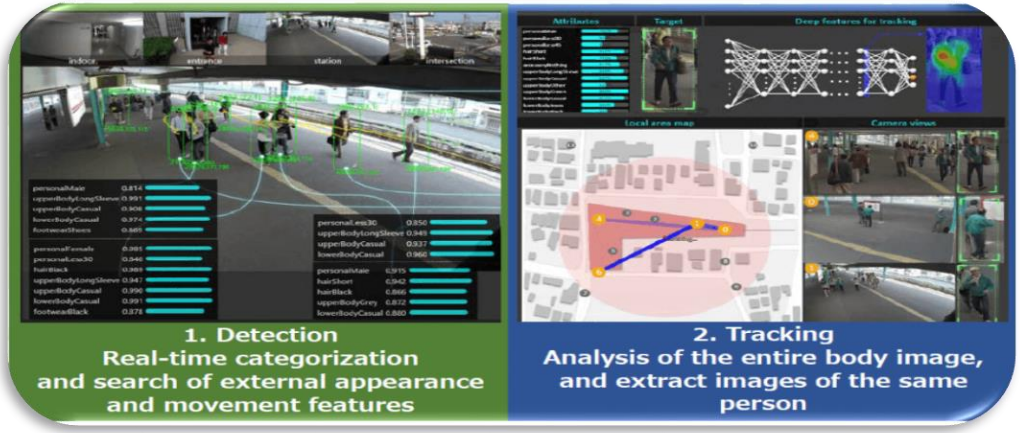
Strawberry Harvesting robot

Applications

- **Surveillance**
  - Behavior monitoring
  - Facial recognition
  - People counting



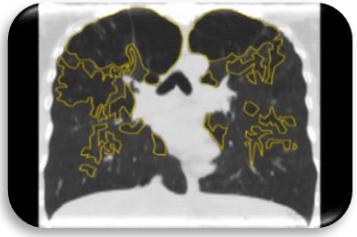
Unexpected behavior monitoring

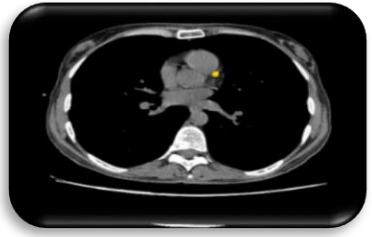


Applications

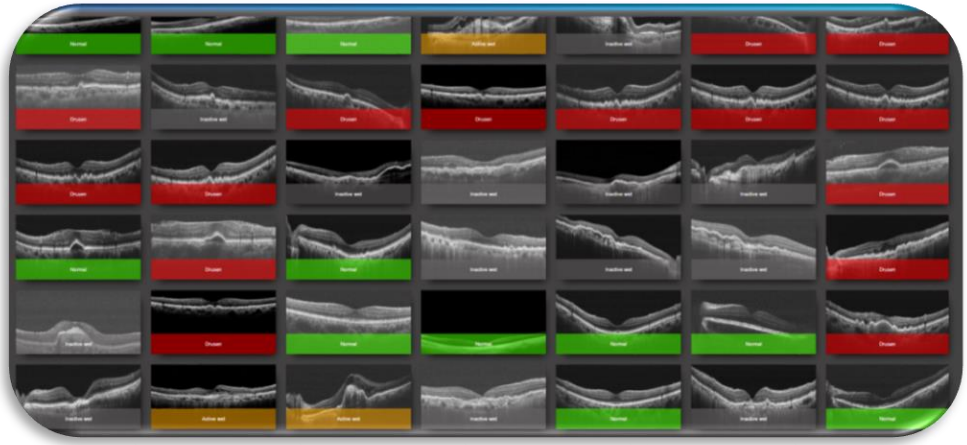
- **Healthcare**
  - Diagnosis assistant
  - Patient monitoring

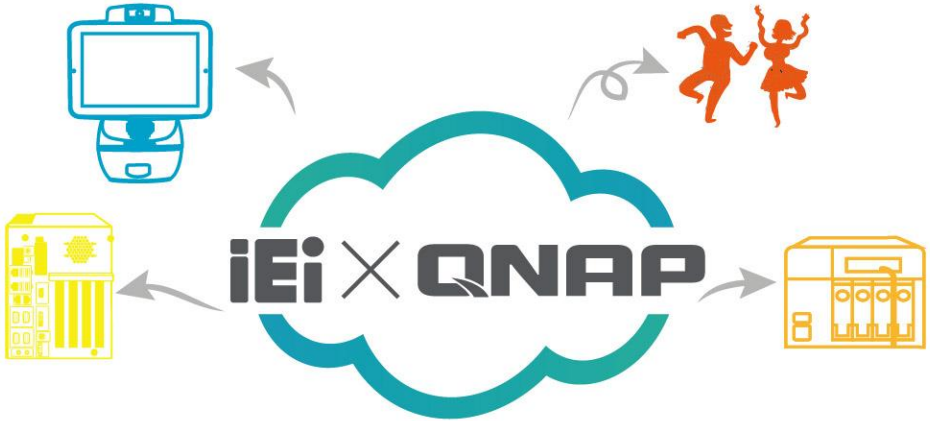
Diagnosis





Baby CAM





Thanks for your attention!