COMBOX X64 MYRIADX (MOVIDIUS) PCIE BOARD FOR NEURAL NETWORKS INFERENCE

COMBOX

TECHNOLOGY

LEGAL NOTICES AND DISCLAIMERS

This document contains information on products, services and/or processes in development. All information provided here is subject to change without notice.

ComBox Technologies features and benefits depend on system configuration and may require enabled hardware, software or service activation. Learn more at **https://combox.io** or from the OEM or retailer. No computer system can be absolutely secure. Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Statements in this document that refer to ComBox Technology plans and expectations for the quarter, the year, and the future, are forward-looking statements that involve a number of risks and uncertainties.

The products described may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.

ComBox Technology does not control or audit third-party benchmark data or the web sites referenced in this document. You should visit the referenced web site and confirm whether referenced data are accurate.

The code names presented in this document are only for use by ComBox Technology to identify products, technologies, or services in development, that have not been made commercially available to the public, i.e., announced, launched or shipped. They are not "commercial" names for products or services and are not intended to function as trademarks.

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

AI MARKET OVERVIEW

TRADITIONAL COMPUTER VISION

DEEP LEARNING

HEURISTIC ALGORITHMS

AI/DEEP LEARNING

HUMAN ENGINEERED

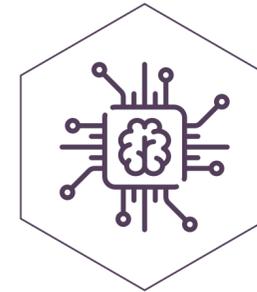REAL-TIME, HIGHER ACCURACY

AI MARKET OVERVIEW

MARKET TRENDS

2 500 petabytes of video surveillance data generated daily in 2019[1]

2025 $40 BILLION FROM $654 MILLION[2]

effective data analysis can not only deliver actionable insights, but significantly improve operations — helping boost profits

TECHNOLOGY TRENDS

BY 2021 82% of all IP traffic will be video[3]
More video data is collected every day, making it critical for businesses to implement robust data analysis strategies

2019 » 2024 (compound annual growth[4])
+67% AI cameras
+31% AI NVRs/recorders

1. Spectra Tech Brief: Meeting the Challenges of Storing IP Video Surveillance Content. July 2017
2. $40 Billion Artificial Intelligence in Marketing Market — Global Forecast to 2025. February 2018
3. Cisco Complete Visual Networking Index (VNI) Forecast, 2017–2022. February 2019
4. OMDIA Video Surveillance & Analytics Intelligence Database July 2020

WE ARE HELPING THESE INDUSTRIES MAKE PEOPLE'S LIVES

SAFER

SMART cities
• Public safety & security
• Traffic, parking and LPR
• Emergency response

MORE EFFICIENT

Financial sector
• People counting
• Reduce customer wait time
• ATM facial recognition

Robotics
• Manufacturing automation
• Industrial

HIGHER QUALITY

Industrial
• Machine vision
• Asset inspection
• Augmented reality

MORE FUN

Casino gaming
• Public safety & security
• Facial recognition

PRODUCTIVE

Transportation
• Public safety
• Traffic & people counting

RELAXING, SAFER

Home & retail
• Security
• Responsive retail advertising
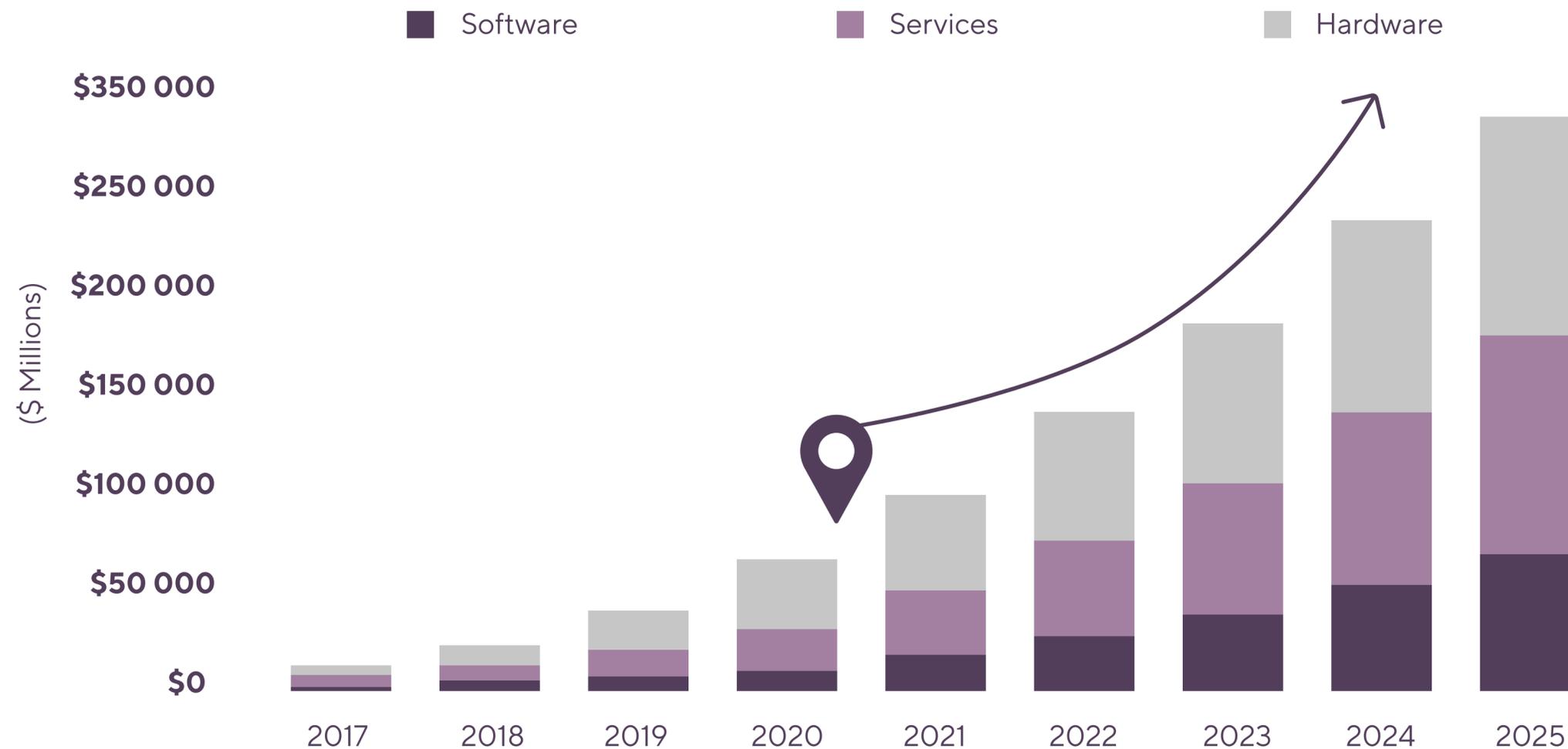• Digital home assistant

HEALTHFUL

Medical imaging
• Segmentation
(MRI scans to detect disease)

THE AI COMPUTING ERA OFFERS GROWING OPPORTUNITY

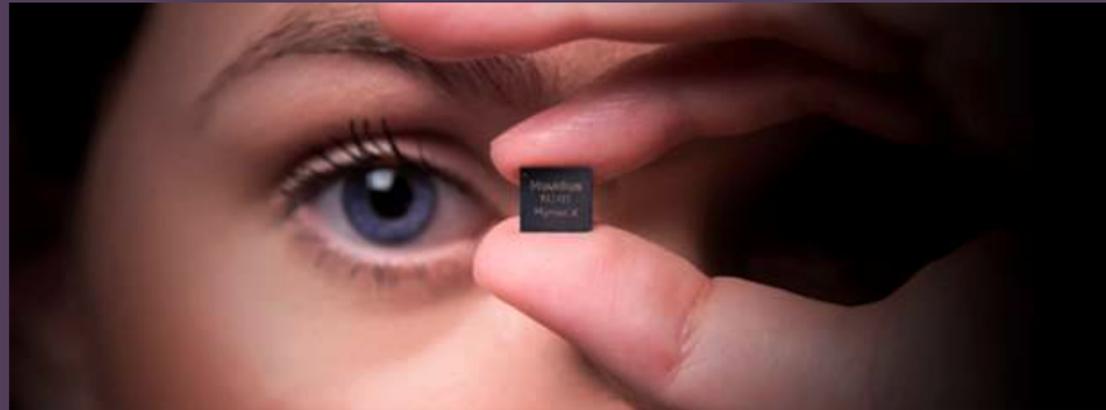## Deep Learning Revenue World Markets

■ Software ■ Services ■ Hardware

- AI to create **$13 TRILLION** additional economic activity **by 2030**[5]

- Machines to impact jobs—in a good way **58M NEW JOBS by 2022**[6]

5. https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-AI-frontier-modeling-the-impact-of-ai-on-the-world-econom
6. https://www.forbes.com/sites/amitchowdhry/2018/09/18/artificial-intelligence-to-create-58-million-new-jobs-by-2022-says-report/#710234644d4b

Chart Source: Tractica

WHAT IS A VISION PROCESSING UNIT (VPU)?

THE INTEL MOVIDIUS VPU IS A CLASS OF PROCESSORS THAT:

⬡ **Processes sensory inputs**

⬡ **Computes vision or AI inference algorithms to extract meaning from multi-modal sensor data**

⬡ **Has a uniquely power-efficient architecture that includes:**

- custom accelerator engines
- programmable processors
- a central scratchpad memory

INTRODUCING COMBOX X64 MOVIDIUS PCIE BOARD FOR NEURAL NETWORKS INFERENCE, A COMPUTE-EFFICIENT PCIE BOARD FOR CLOUD AI
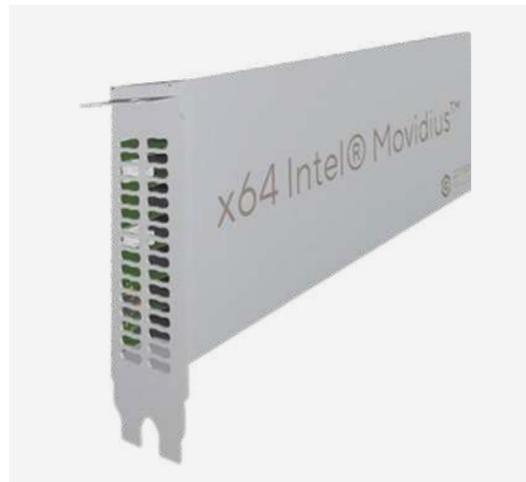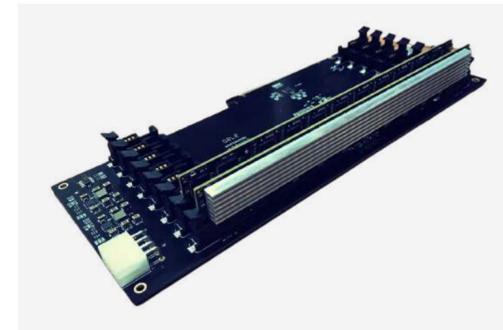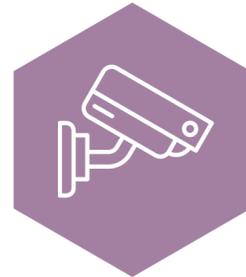

⬡ **High Performance CV & Deep Learning Inference**

⬡ **Power Efficiency with Low Cost. World-class performance per watt per dollar**

⬡ **Simultaneous Video Decode/Analytics/Encode on one board with 64 Movidius. Optimized hardware Codec included with acceleration for CV/DL**

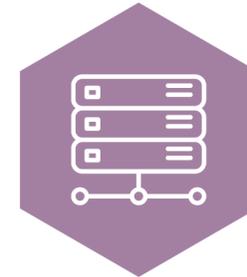# COMBOX NUC SERVER ENABLES A WIDE SET OF NEW APPLICATIONS
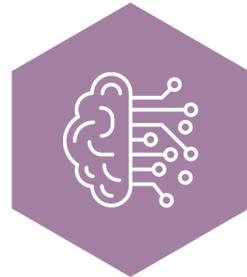
AI appliance

NVR

On premise server

Shopper tracking

Medical imaging
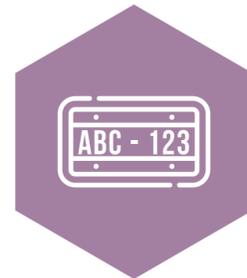
Auto plate recognition

ABC - 123
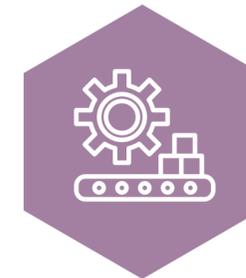
Industrial automation

and many more...

COMBOX X64 MOVIDIUS PCIE PRODUCT SPECIFICATION

Product VPU	**Movidius MyriadX VPU**
Processor number	**MA2485**
Count of VPU onboard	**64 pcs**
Status	**Presale order**
Warranty	**1 year**
Processor base frequency	**700 MHz**
Max. operating temperature	**105C**
Min. operating temperature	**0C**
Operating temperature range	**0-105C**
Board size	**Full-length, full-height, double-width**
PCIe	**PCIe 3.0 x4**
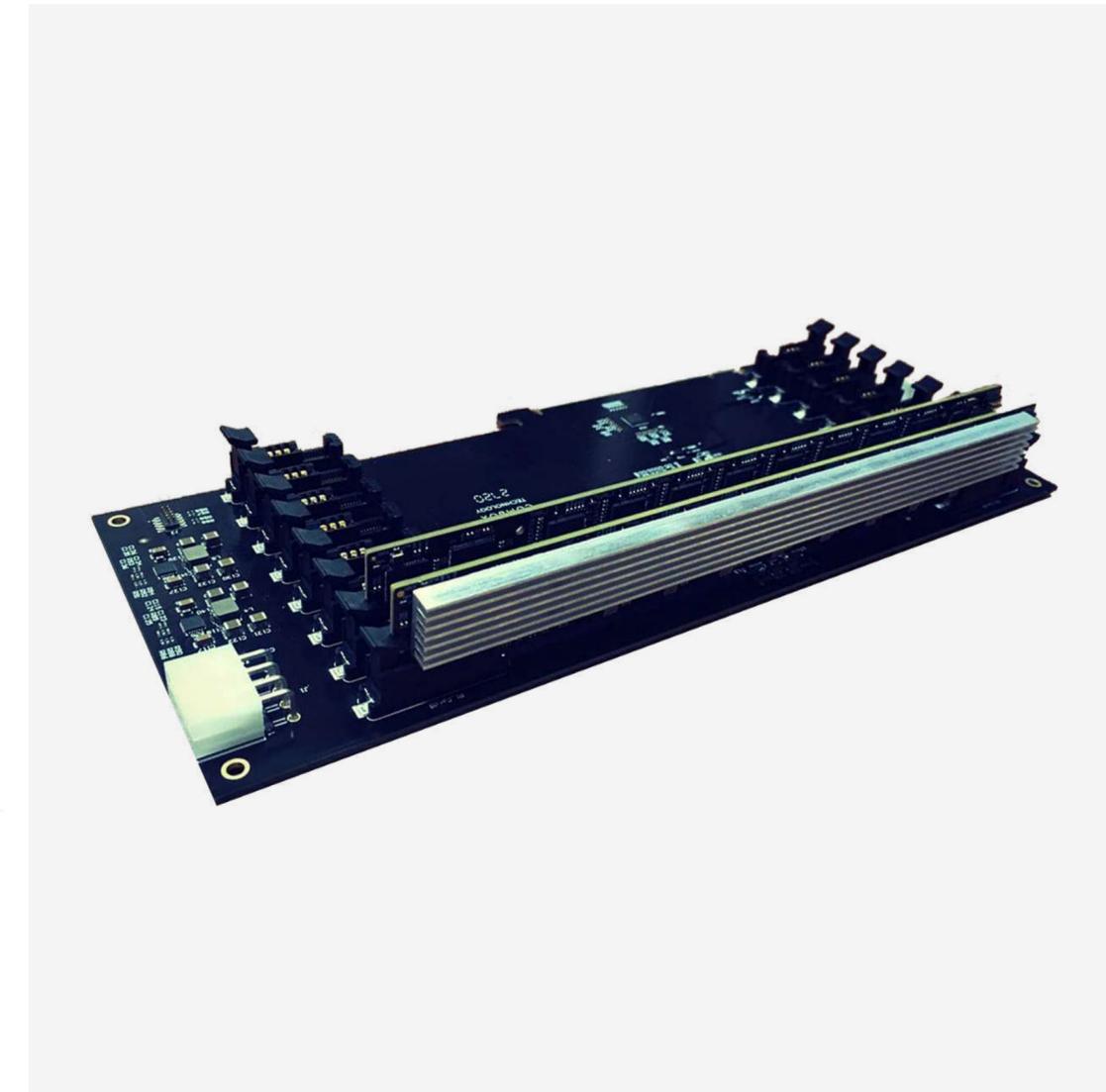
EXAMPLE INFERENCE SYSTEM USING HOST AND 4X PCIE X64 MOVIDIUS BOARDS

SYSTEM CONFIGURATION:

- Platform Supermicro 1029GQ-TRT, 1U

- HDD 1024 Gb SSD M.2 Intel 660p series

- CPU Intel Xeon 4114 Skylake x2 pcs

- RAM Kingston Server Premier DDR4 DIMM 16 Gb PC4-19200 x8 pcs

- ComBox x64 Movidius PCIe board x4 pcs
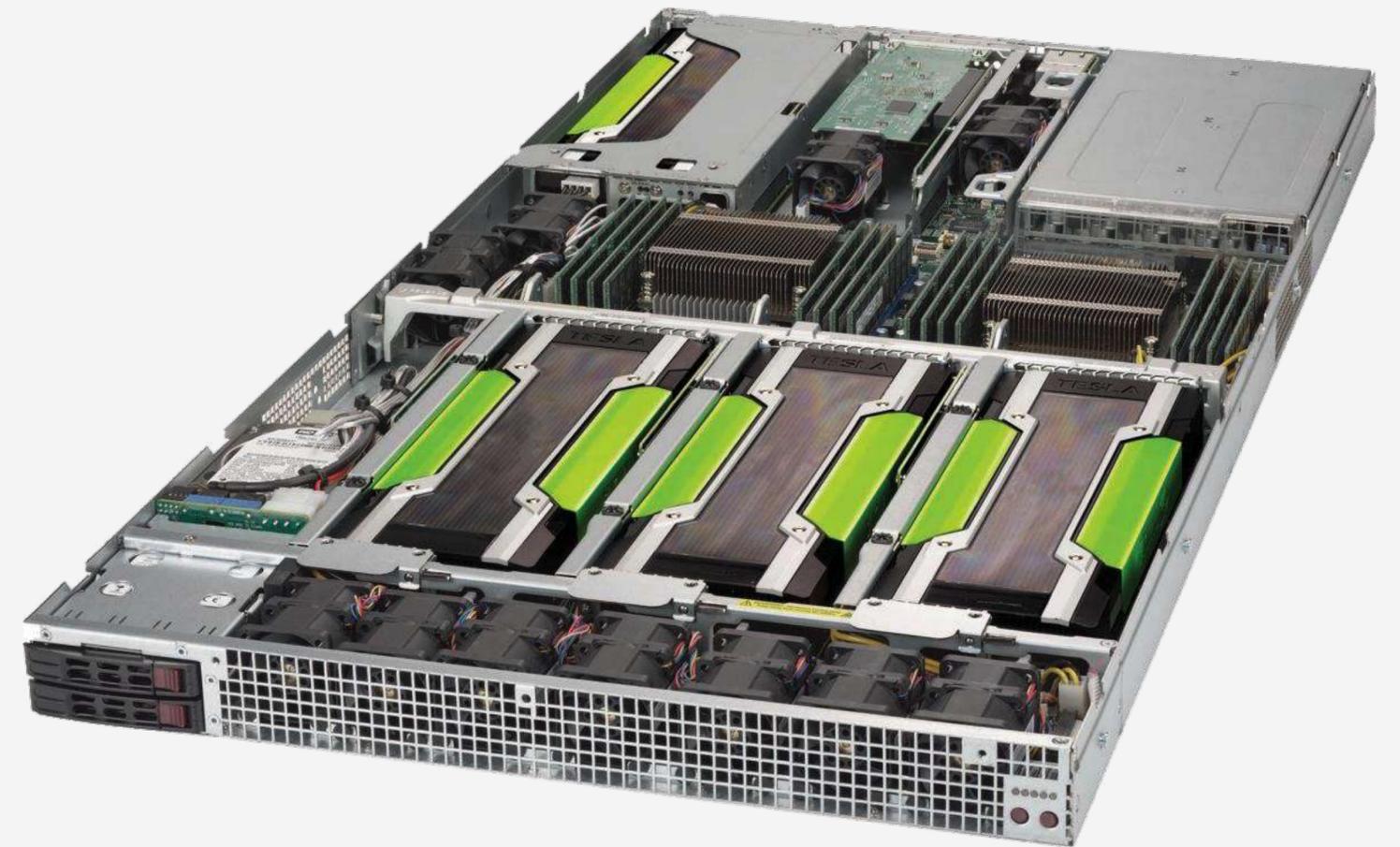
DL INFERENCE PERFORMANCE

- Operating system — Ubuntu 18.04.5.

- Measurements are made using ResNet50 and SSD Mobilenet v.2 model with FP16 weights.

- The benchmark_app utility included in Intel OpenVINO is used to measure performance.

- Inference tested with Intel OpenVINO 2021.1 release

DNN	FPS x64 Movidius board
RESNET50	2240
SSD MOBILENET V.2	2880

Testing technique
Performance measurements are performed with blocks of 1 frame and 8 frames are fed to the input of the neural network (batch_size 1 and 8, respectively). The number of frames per second is counted and recorded.

SOFTWARE ARCHITECTURE AND OVERVIEW

PRODUCTIVITY

- Intel OpenVINO™ toolkit

- Ease-of-use programming model

- Simple and industry standard SDKs and APIs

FLEXIBILITY

- Different programming modes

- Consistent UX/DX across product line

- Backward compatibility

PERFORMANCE

- Expose full capability of VPU

- NCE and Shaves for Inference SDKs

- HW Codec for media encode/decode SDKs

Intel OpenVINO toolkit makes media analytics development easier

Intel OpenVINO toolkit (Open Visual Inference and Neural network Optimization) is a free toolkit facilitating the optimization of a deep learning model from a framework and deployment using an inference engine onto Intel hardware.

The toolkit has two versions: OpenVINO toolkit, which is supported by open source community and Intel Distribution of OpenVINO toolkit, which is supported by Intel.

OpenVINO was developed by Intel. The toolkit is cross-platform and free for use under Apache License version 2.0.

The toolkit enables a write-once, deploy-anywhere approach to deep learning deployments on Intel platforms, including CPU, integrated GPU, Intel Movidius VPU, and FPGAs.

WHAT'S INSIDE THE OPENVINO™ TOOLKIT

INTEL® DEEP LEARNING DEPLOYMENT TOOLKIT

Model Optimizer
Convert and optimize

IR **IR=Intermediate**
Representation format

Inference engine
Optimized inference

- TensorFlow
- Caffe
- mxnet
- **Trained models**

COMPONENT TOOLS

Traditional computer vision for Intel CPU/CPU with integrated graphics — optimized computer vision libraries

OpenCV   ***OpenVX***

Helps increase processor graphics performance — Linux* only

*Intel® Media SDK   ***OpenCL™**
*Intel® Integrated Graphics drivers and runtimes

Linux for FPGA only

*FPGA runtime environment (RTE)
(from Intel® FPGA 50K for openCL™)
*Bitstreams

*Intel® FPGA Deep Learning Acceleration Suite (DLAS)

CPU	GPU	FPGA	VPU

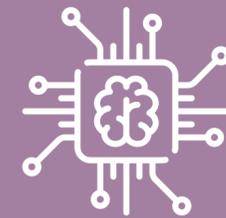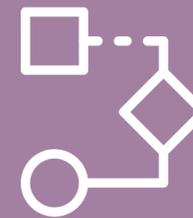# ADVANTAGES OF USING OPENVINO

Write once

Deploy anywhere

Easy to use

Leverage common algorithms

High-performance deep learning inference

TYPICAL PIPELINE

#01

Media input

#02

Encode/decode

#03

Pre/post processing

#04

Deep learning inference

#05

Object tracking

#06

Media output

#07

Data move and management

ADVANTAGES OF USING OPENVINO

Also you can use Gstreamer, an open source multimedia framework that helps you build and connect-up media application pipelines:

Media-handling components support playback, audio/video streaming and processing

Includes Gstreamer Video Analytics (GVA) plugin for inference, processing, debugging

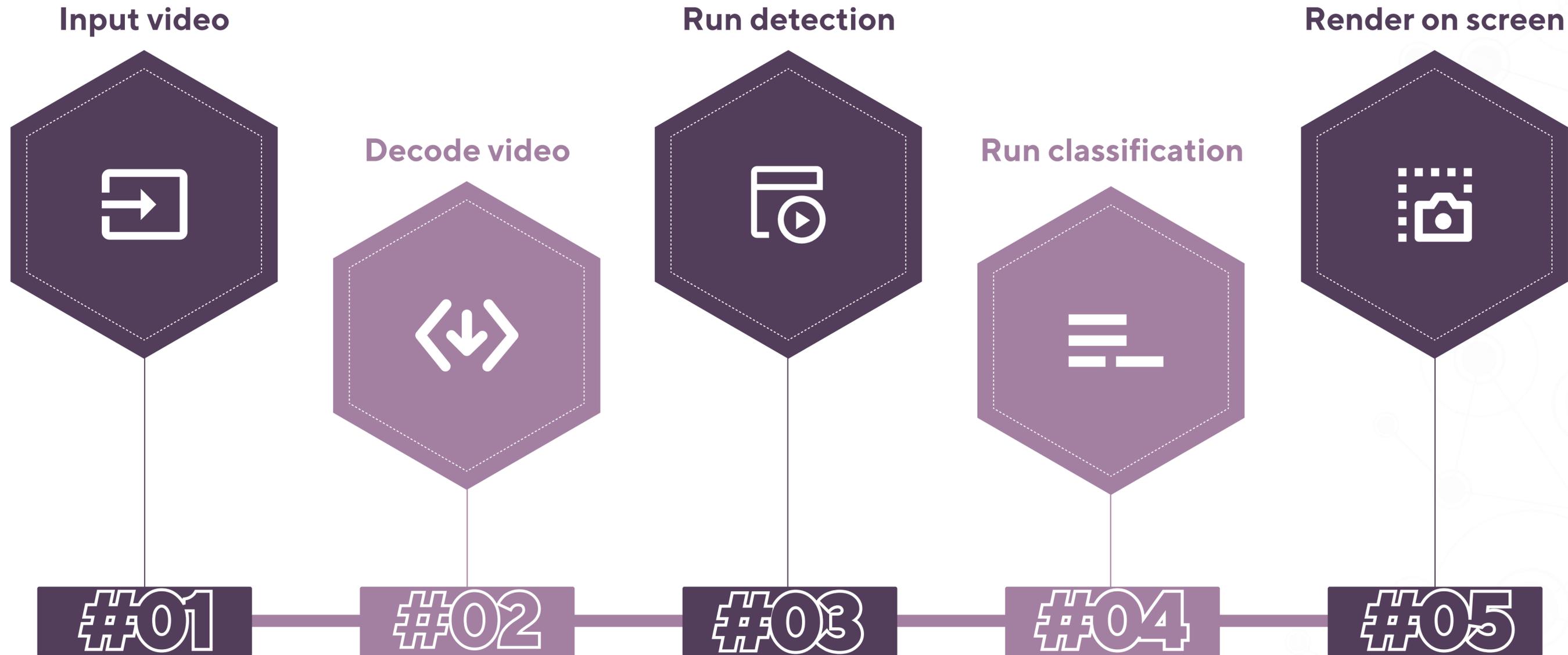**Intel Movidius VPU can be used for inference**

Each step can be configured with extensive set of Gstreamer-level properties

Developers can add new codecs and filters by writing a simple plugin with a generic interface

ADVANTAGES OF USING OPENVINO

Input video

Decode video

Run detection

Run classification

Render on screen

#01

#02

#03

#04

#05

INTEL MOVIDIUS MYRIAD X VPU PARITY NNS

Model name	FWK	Compiles for	Executes on	Accurate on
googlenet-v1	TF	yes	yes	yes
mobilenet-v2	PyTorch	yes	yes	yes
resnet-50-pytorch	PyTorch	yes	yes	yes
squeezenet1.1	PyTorch	yes	yes	yes
yolo-v2-tiny-ava-0001	TF	yes	yes	yes
yolo-v2-ava-0001	TF	yes	yes	yes
googlenet-v3	TF	yes	yes	yes
tiny_yolo_v2	TF	yes	yes	yes
resnet-18	PyTorch	yes		
facenet-20180408-102900	TF	yes	yes	yes
yolo_v3	TF/Keras	yes	yes	yes
resnet-101	CF	yes	yes	yes
resnet-152	CF	yes	yes	yes
ssd512	CF	yes	yes	yes

INTEL MOVIDIUS MYRIAD X VPU PARITY NNS

Model name	FWK	Compiles for	Executes on	Accurate on
googlenet-v4	TF	yes	yes	yes
icnet-camvid-ava-0001	TF	yes	yes	yes
ssd_mobilenet_v1_coco	TF	yes	yes	yes
vgg16	TF	yes	yes	yes
tiny_yolo_v1	TF/Keras	yes	yes	yes
yolo-v2-tiny-vehicle-detection-0001	TF	yes	yes	yes
resnet-50-tf	PyTorch	yes	yes	yes
unet-camvid-onnx-0001	PyTorch	yes	yes	
vehicle-attributes-recognition-barrier-0042	TF	yes	yes	yes
vehicle-license-plate-detection-barrier-0106	TF	yes	yes	yes
mobilenet-v2-1.0-224	CF	yes	yes	yes
vehicle-attributes-recognition-barrier0039	CF	yes	yes	
googlenet-v1	CF	yes	yes	
googlenet-v2	CF	yes	yes	

INTEL MOVIDIUS MYRIAD X VPU PARITY NNS

Model name	FWK	Compiles for	Executes on	Accurate on
googlenet-v4	CF	yes	yes	yes
inception-resnet-v2	TF	yes	yes	
mobilenet-v1-0.25-128	TF	yes	yes	
mobilenet-v1-0.50-160	TF	yes	yes	
mobilenet-v1-1.0-224	TF	yes	yes	
mobilenet-v2-1.4-224	TF	yes	yes	yes
resnet-101	TF	yes	yes	
resnet-152	TF	yes	yes	
resnet-v2-101	TF	yes		
resnet-v2-50	TF	yes		
ssd_mobilenet_v2_coco	TF	yes	yes	
vgg16	CF	yes		
vgg19	CF	yes		
yolo-v3	CF	yes	yes	

INTEL MOVIDIUS MYRIAD X VPU PARITY NNS

Model name	FWK	Compiles for	Executes on	Accurate on
resnet-34-pytorch	CF	yes	yes	
mobilenet-v2	CF	yes	yes	
resnet-50	CF	yes	yes	
mobilenet-v1-1.0-224	CF	yes	yes	
mobilenet-ssd	CF	yes	yes	
age-gender-recognition-retail-0013	CF	yes	yes	yes
emotions-recognition-retail-0003	CF	yes	yes	
face-detection-retail-0004	CF	yes		
person-detection-retail-0013	CF	yes	yes	
head-pose-estimation-adas-0001	CF	yes		
inception-resnet-v2	CF	yes	yes	
squeezenet1.1	CF	yes	yes	
ssd_resnet50_v1_fpn_coco	TF	yes	yes	
ssd_mobilenet_v2_quantized	TF	yes	yes	
ssdlite_mobilenet_edgetpu_coco_quant	TF	yes	yes	
yoloV3	CF	yes	yes	

COMBOX
TECHNOLOGY

START WORKING WITH US

COMBOX TECHNOLOGY

combox.io

info@combox.io

+7(812)748-27-27

Russia, Saint-Petersburg, Bolshaya Pushkarskaya str., 20, of. 5-514

product specification