HAILO

Empowering Intelligence

Hailo-8™: The World's Top Performing AI Processor for Edge Devices

Enables Data-class AI Applications in an Embedded Power Envelope

August 2022

About Hailo

A leading AI chipmaker for edge devices, founded in 2017

1st generation in MP

Headquartered in Israel with offices in USA, Germany, Japan, China, Korea, Taiwan

Patented structure-defined dataflow architecture

190 + employees with extensive experience from leading tech companies

Total $224M funding including Strategic Investors

NEC & ABB

A growing worldwide partner ecosystem

CES 2020 Innovation Awards Honoree

EU Horizon 2020 Recipient

AI Semi Cool Vendor by Gartner

Best Edge AI Processor of 2021

ISO 9001 CERTIFIED

ISO 14001

Intelligence Become a Necessity

Hailo’s **powerful** and **scalable** AI technology enables new capabilities in various markets

Automotive

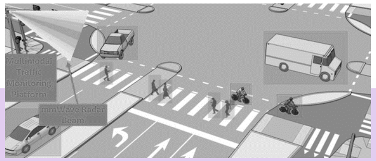
Autonomous
Vehicles, ADAS

Smart City

Public safety & security

Smart Home

Security,
Assisted Living



ITS (Intelligent Transportation System)

Traffic control, Tolling,
Law enforcement

Smart Retail

Cashierless Store, Inventory
Management

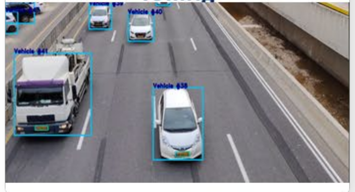
Industry 4.0

Factory Automation

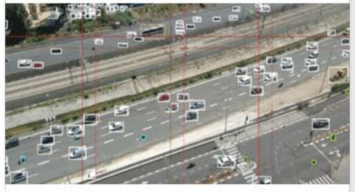
Deep Learning at the Edge with Hailo-8™

Use-case Examples:

Traffic Management & Tolling



Traffic Monitoring



Intersection Safety

Public Health Monitoring

Autonomous Delivery

Quality Inspection

Factory Safety

Retail Automated Checkout

Smart Building

Advanced Driver Assistance (ADAS)

Front Facing Perception

Access Control

Device Examples:

Intelligent Cameras

Intelligent NVR

Industrial Gateways

In-Vehicle

ADAS ECU

Autonomous

Hailo-8™ Highlights

The World's Most Powerful and Efficient Edge AI Processor

High Performance
26 TOPS
Efficient AI architecture

Comprehensive SW Tools
Mature dataflow compiler
Efficient RT library

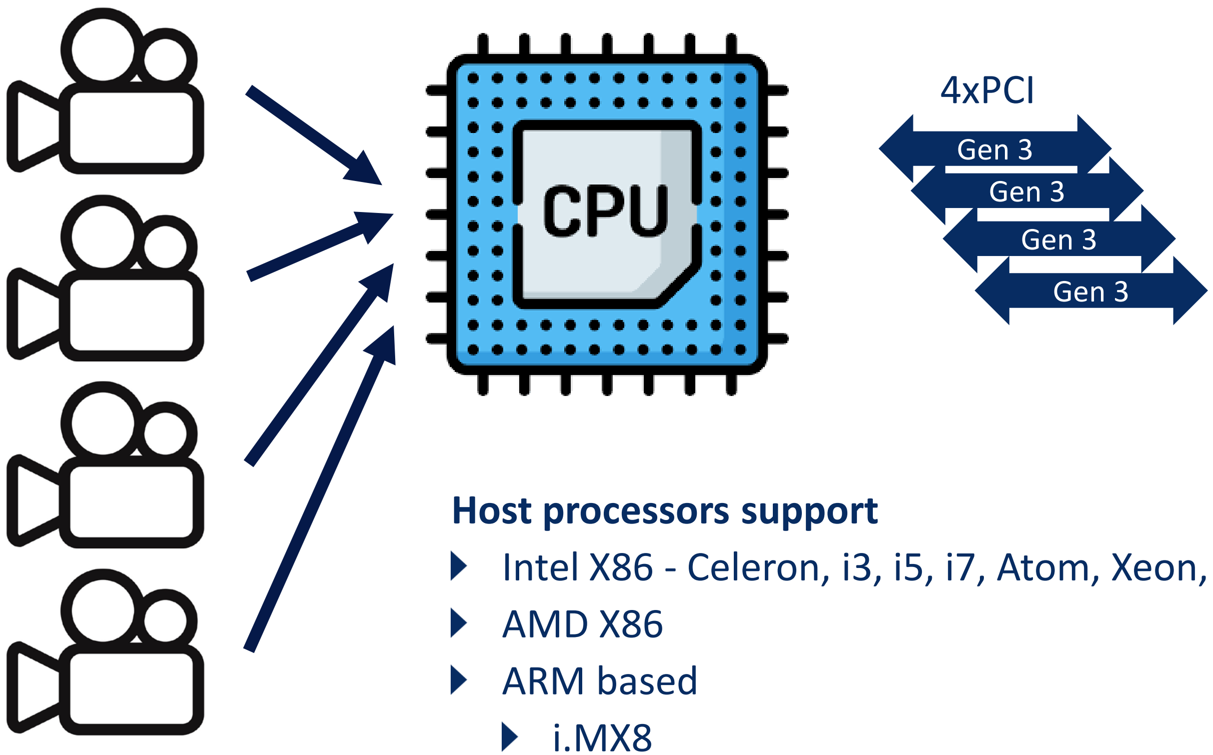
Industrial & Automotive Grades
Industrial: -40°C to 85°C
Automotive: -40°C to 105°C

Power Efficiency
Typical Power Consumption: 2.5W

Single Chip Solution
No External DRAM required

Scalable & Flexible
Multi-streams
Multi-model
Multi-chip

Hailo-8™ System Usage



Host processors support

- Intel X86 - Celeron, i3, i5, i7, Atom, Xeon, …
- AMD X86
- ARM based
  - i.MX8
  - Layerscape (LX2160)
  - S32G
  - Raspberry Pi
  - FPGA SoC – Xilinx Zynq
  - Renesas R-CAR V3H/V4H
  - SocioNext SC2A11

- **Flexibility & Scalability**
  - **Performance scalability** (1x to 12x Hailo-8 → 26 to 312 TOPS)
  - **Host processor type** (X86 & ARM)
  - **Interface w/Host** (PCIe / Ethernet)

Hailo-8™ Product Offering

Hailo-8™ AI Processor

- 26 TOPS
- Industry-leading power efficiency
- 17 x 17 FCBGA

Hailo-8™ M.2 AI Acceleration Module

- PCIe Interface
- M.2 form factor
  - M.2 Key M 2242/2260/2280
  - M.2 Key B+M 2242/2260/2280
  - M.2 Key A+E 2230
- Extended temperature support: -40°C to 85°C

M key
4 lanes

B+M key
2 lanes

A+E key
2 lanes

Hailo-8R™ mPCIe AI Acceleration Module

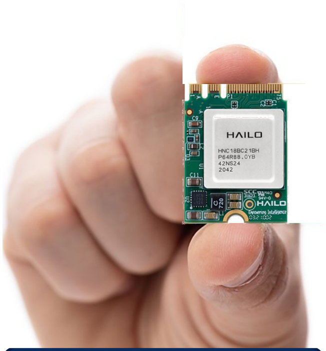
- PCIe Interface
- mPCIe form factor 3050
- Extended temperature support: -40°C to 85°C

PCIe Acceleration Card

- PCIe Interface
- Multi-chip configuration (x4, x5, x6)
- Up to 156 TOPS
- Typical power: 35W

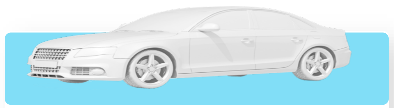
Lanner Falcon-H8

Hailo-8™ M.2 Starter Kit

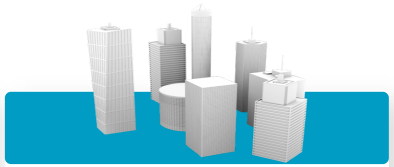


MSRP: $179

- AI accelerator module for developing and prototyping edge AI applications and specifically for video analytics solutions
  - **M.2 module** with Hailo-8™ AI accelerator processor
  - Best-in-class real-time performance utilizing the Hailo-8™ **26 TOPS** compute power
  - Industry-leading power efficiency with typical power consumption of **2.5W**
  - Higher **cost-efficiency** (TOPS/$) compared with existing solutions
- Robust software toolchain supports state-of-the-art NN models and applications out-of-the-box
- Suitable for various applications

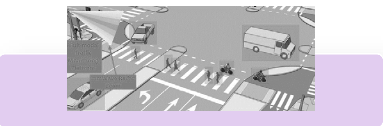


Automotive

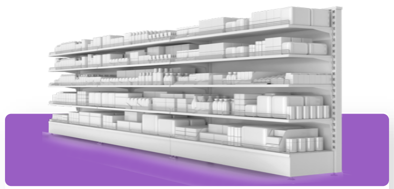


Smart City

Smart Home



ITS



Smart Retail

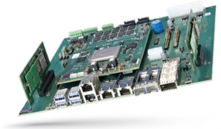
Industry 4.0

Hailo-8™-Powered Edge AI Solutions

FOXCONN

BoxiEdge

MicroSys



AIP-LX2160A

BASLER

prB-IMX8MP

Vecow

VAC-1100

APB-3000AI

Compulab

Fitlet2

Tensor

kontron

KBox A 150-WKL-AI-H8

pITX-iMX8M-AI-H8

DELL

OptiPlex 7080

OptiPlex 3070

Precision 3930

NEXCOM

VTC1021

NISE-51

NISE-52

AAEON

Xtreme i11

UPS Squared Pro

UPS Squared 6000

Lanner

LEC-2290H

LEC-7242H

AXIOMTEK

RSC101

ebox710-521-fl

Variscite

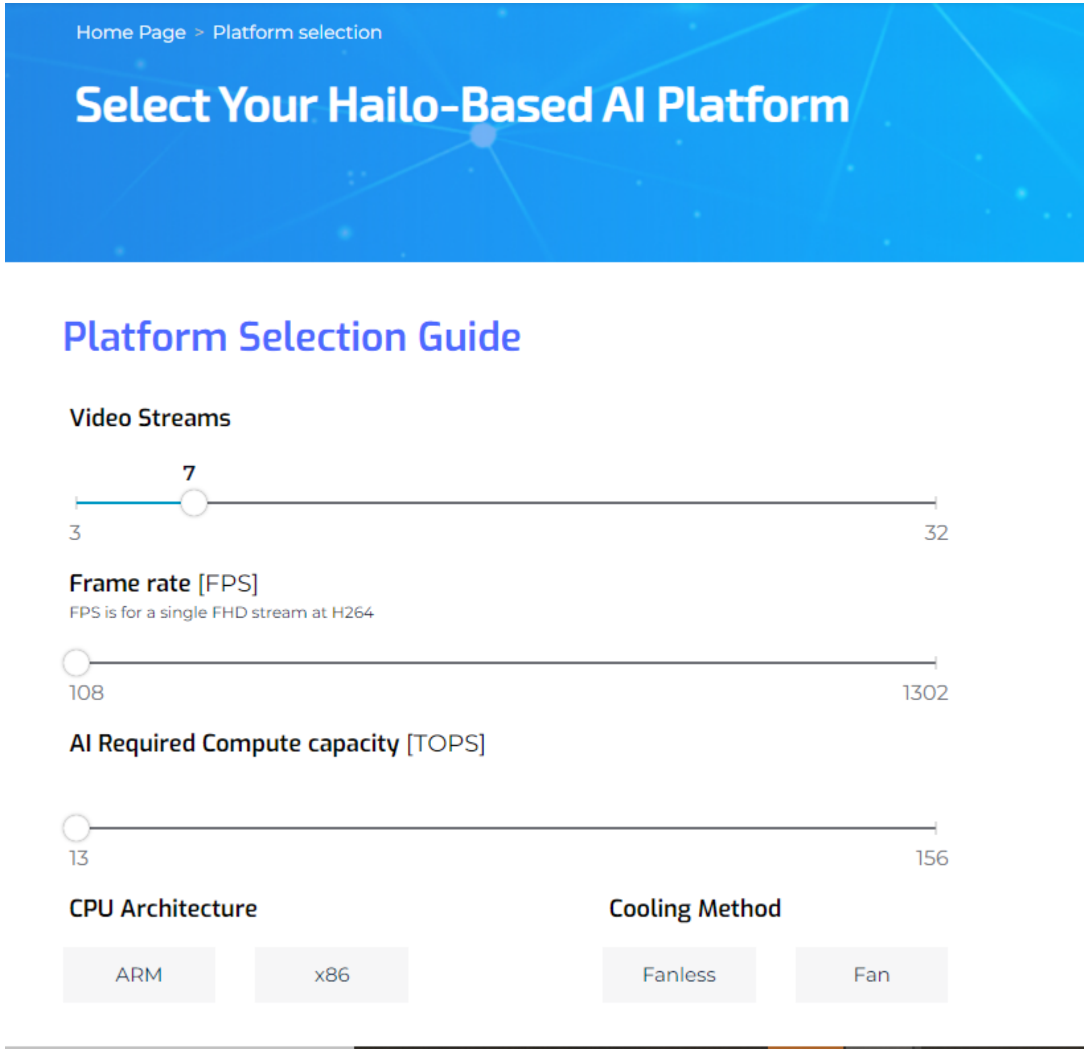
DART-MX8M-PLUS

VAR-SOM-MX8M-PLUS

HAILO

https://hailo.ai/our-partners/

Platform Selection Guide



Home Page > Platform selection

Select Your Hailo-Based AI Platform

Platform Selection Guide

Video Streams

7

3 32

Frame rate [FPS]

FPS is for a single FHD stream at H264

108 1302

AI Required Compute capacity [TOPS]

13 156

CPU Architecture

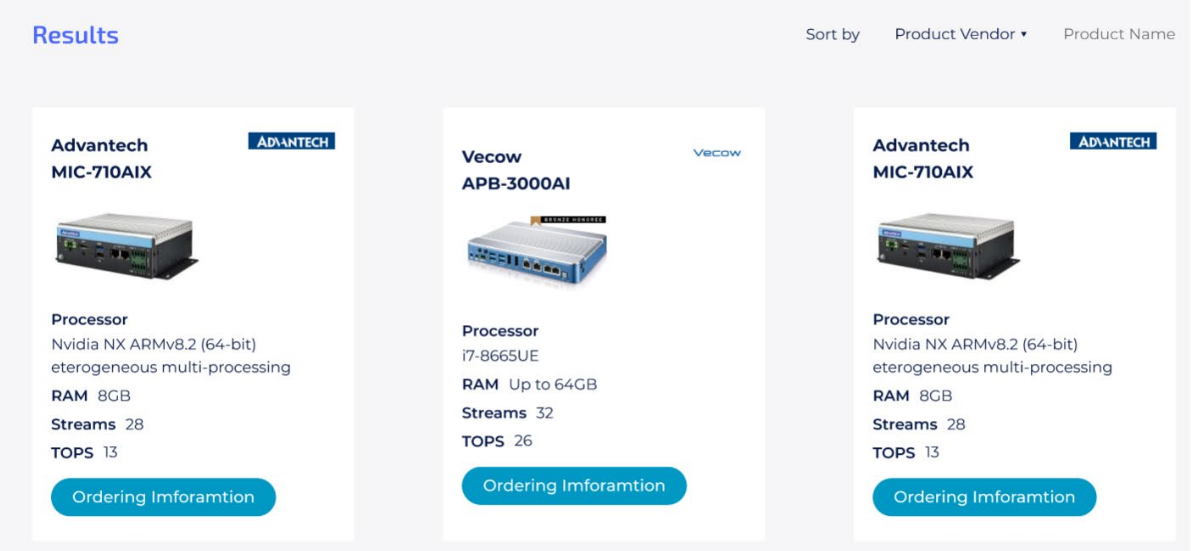
ARM x86

Cooling Method

Fanless Fan

Quickly find a H/W platform with Hailo inside

- Based on the database maintained by BD
- Clear criteria for selection and de-selection



Results

Sort by Product Vendor ▾ Product Name

Advantech
MIC-710AIX

Processor
Nvidia NX ARMv8.2 (64-bit) eterogeneous multi-processing

RAM 8GB

Streams 28

TOPS 13

Ordering Imforamtion

Vecow
APB-3000AI

Processor
i7-8665UE

RAM Up to 64GB

Streams 32

TOPS 26

Ordering Imforamtion

Advantech
MIC-710AIX

Processor
Nvidia NX ARMv8.2 (64-bit) eterogeneous multi-processing

RAM 8GB

Streams 28

TOPS 13

Ordering Imforamtion

https://hailo.ai/product/platform-selection/

Hailo & NXP Joint Offering

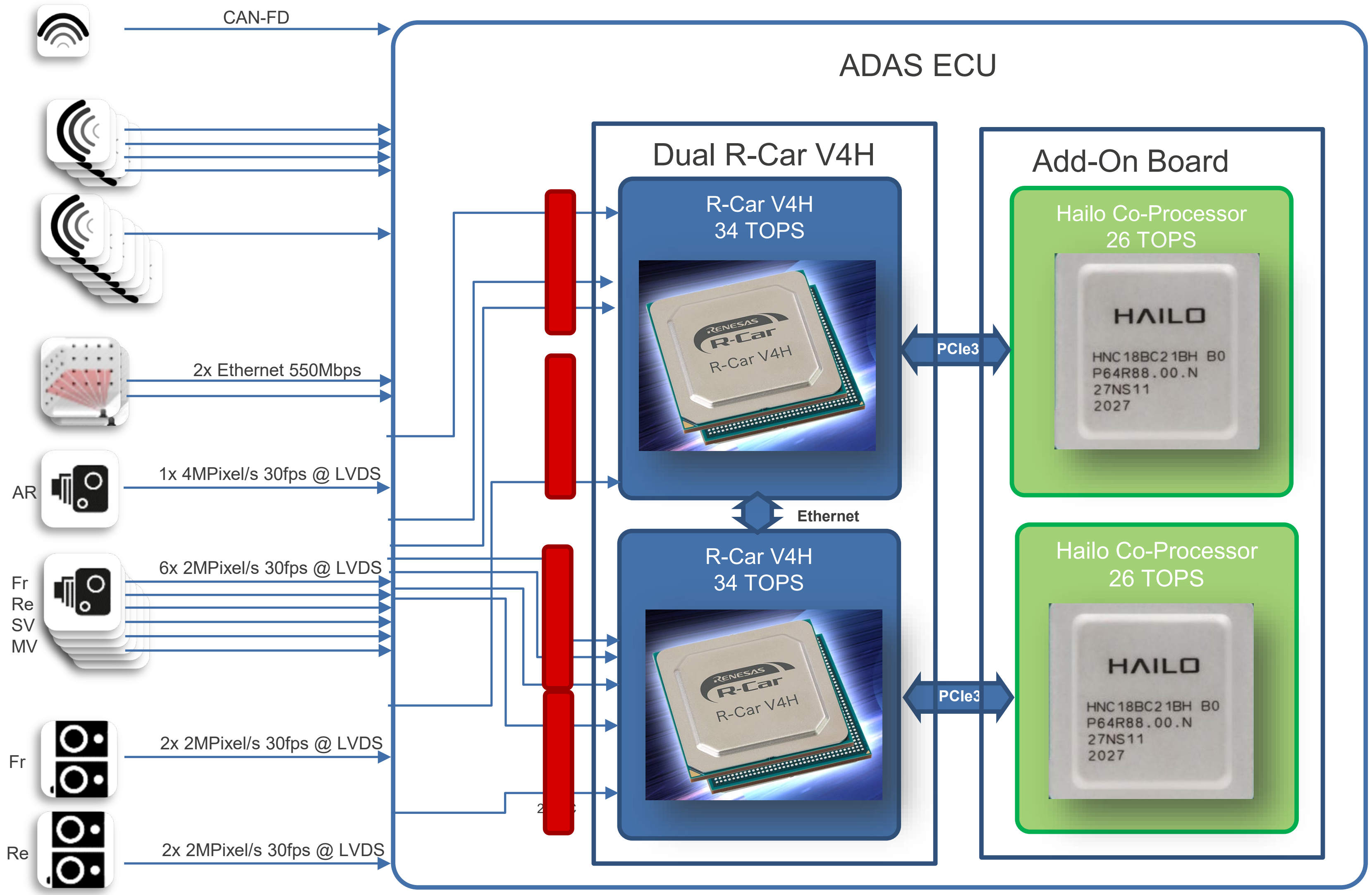
Combining NXP’s Arm®-Based Processors with Hailo-8™ AI Processor for a powerful, scalable and efficient AI offering for embedded products

- Generic edge device designs
  - Hailo-8 combined with Arm® based NXP® i.MX 8 Series delivers a powerful, power-efficient and cost-effective platform for edge devices
  - Application-ready hardware is available from Kontron

REGISTERED PARTNER NXP

- Automotive driven designs
  - Hailo-8™ combined with Arm® based NXP® S32 Automotive and NXP® Layerscape® platforms results in a high-performance, scalable, safe and efficient automotive grade solution
  - Application-ready hardware is available from MicroSys

Hailo & Renesas Joint Autonomous Drive ECU Concept



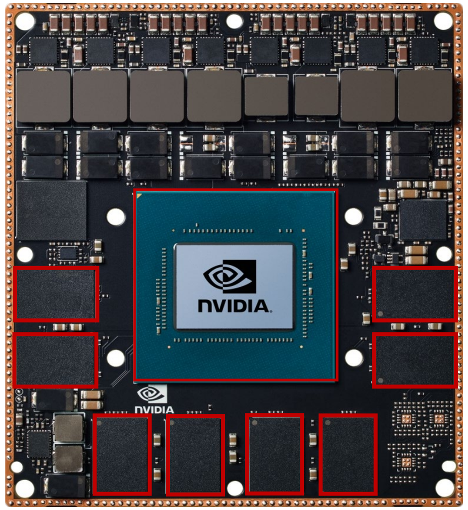
- Independent scalability in AI and compute allowing flexibility for L2-L4 ADAS designs
- Best-in-class power efficiency enabling passively cooled ECUs
- Cost-efficient solution “pay for what you need”
- Pay as you grow with Hailo AI accelerator roadmap
- Open software ecosystem allowing OEMs/Tiers control and innovation

Combining Renesas R-Car V4H with Hailo AI Co-Processor

Unprecedented AI Performance

Comparison on Inference Compute Performance

NVIDIA AGX Xavier



General Purpose GPU
+ External Memory

Hailo-8™
M.2 A+E Key

Dedicated AI Chip
No External Memory

ResNet-50 Benchmark

Device	Total Power [Watt]	Total Power Efficiency [TOPS/W]
Hailo-8™	1.6	3.0
Nvidia Xavier AGX	32	0.14

Conditions:

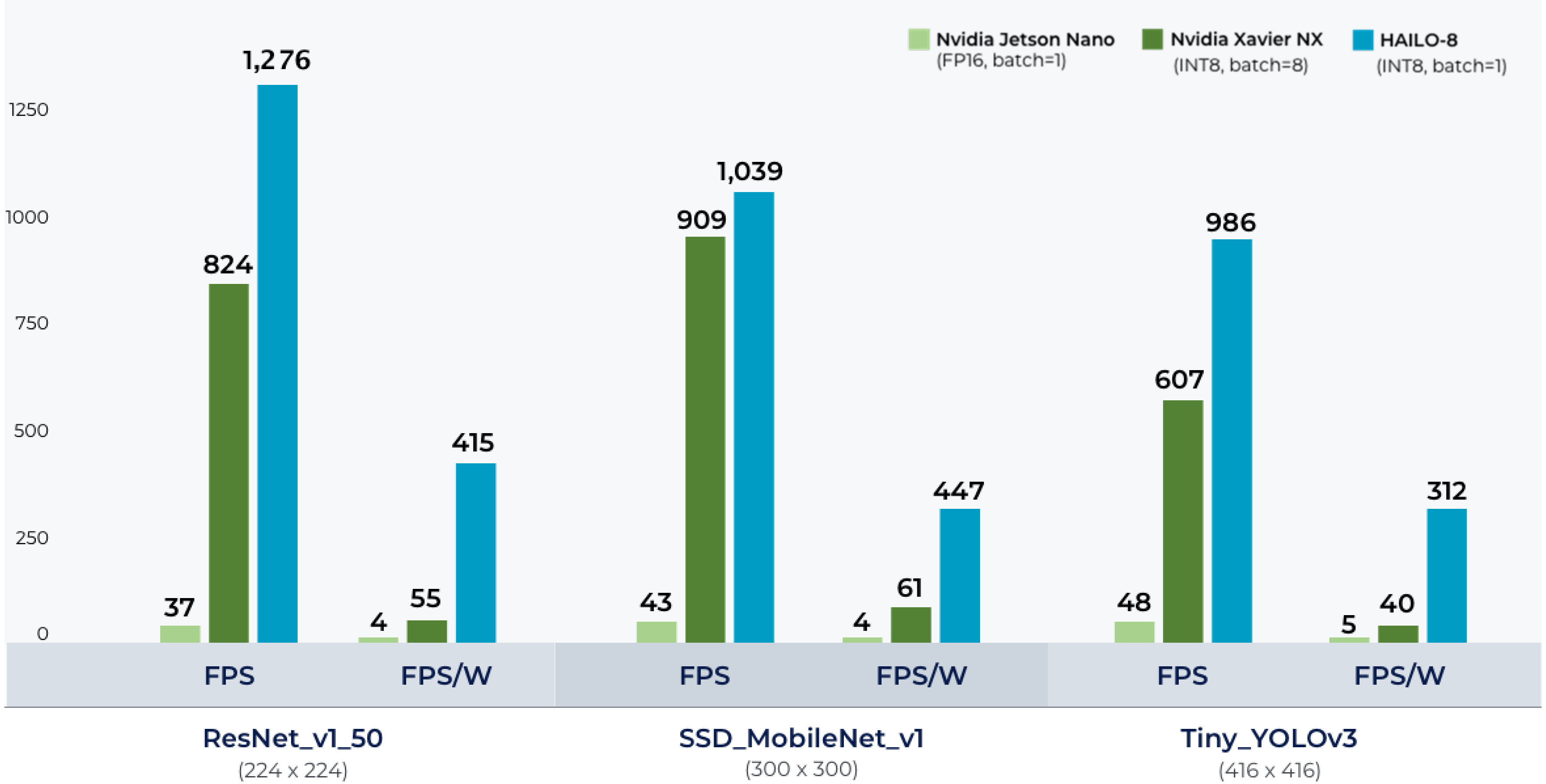
- TOPS (8-bit): Xavier 32 TOPS, Hailo-8 26 TOPS
- 224x224 image resolution feed @ 656 FPS
- 8-bit precision
- Batch size = 1

X15 Better
Area Efficiency

X20 Better
Power Efficiency

Unprecedented Performance at the Edge

Hailo-8 offers higher performance and as much as x8 the power efficiency of Nvidia's best edge device



Remarks

- SDK version 3.9.0 (June 2021), measured at room temp on a single Hailo-8 device through PCIe interface on a Hailo EVB. System host: Intel® Core™ i5-9400 CPU @ 2.90GHz)
- **Xavier NX results are using batch=8** (while Hailo-8 and Jetson Nano are using batch=1) and that **Jetson Nano is limited to FP16** (while Hailo-8 and Xavier NX are INT8). Nvidia results for batch=1 and INT8, respectively, are expected to be lower.
- FPS & power figures for Nvidia Jetson Nano and Xavier NX are sourced from the Nvidia website and Github repo, retrieved 12/07/21

Hailo-8™ Unprecedented AI Performance and Power Efficiency

	Intel Myriad X	Google Edge TPU	Hailo-8™	Hailo-8™ outperforms
Performance FPS	87	385	2,613	x30 vs. **Myriad X** x6 vs. **Edge TPU**
Power Efficiency FPS/W	35	275	1,267	x30 vs. **Myriad X** x4 vs. **Edge TPU**

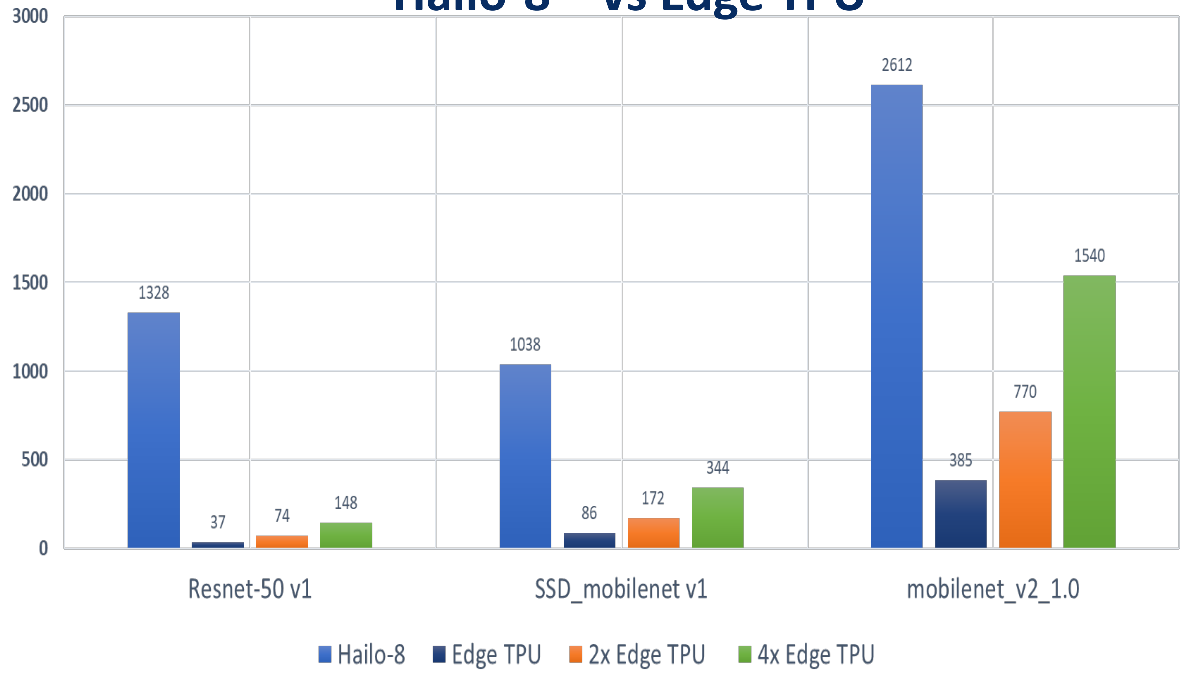
The Hailo-8™ M.2 AI Acceleration module unprecedented AI capabilities

Provides the scalability to run advanced video analytics DL models in high-resolution & high-frame rate

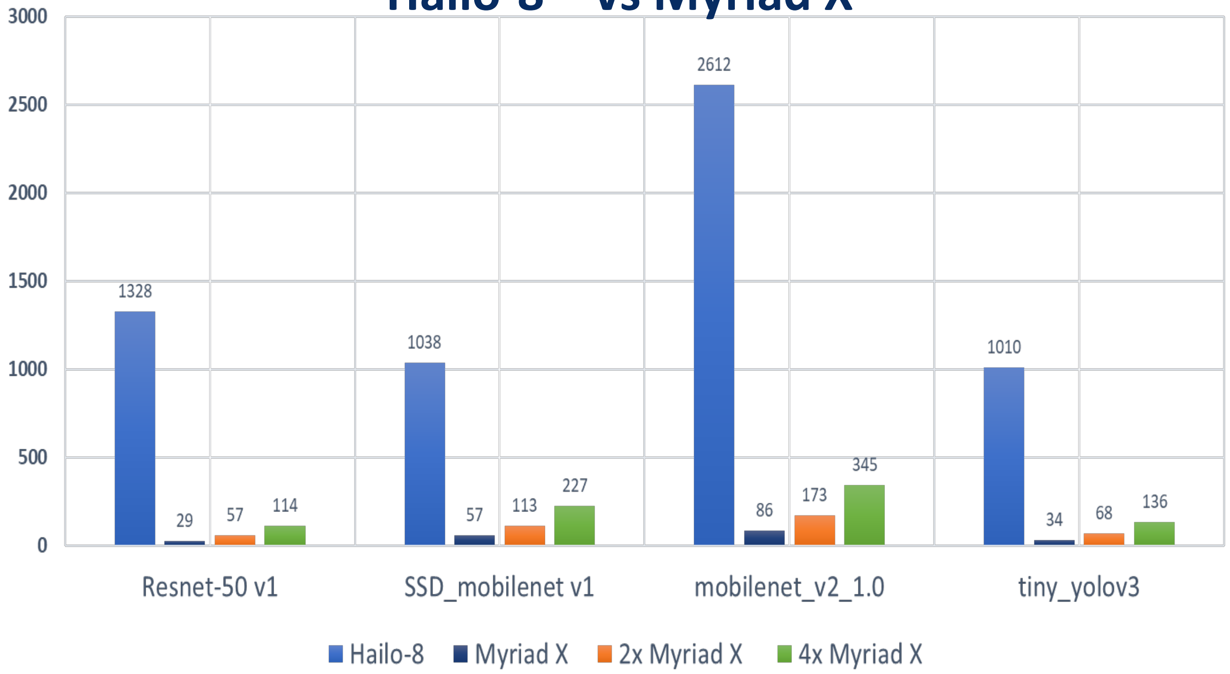
** Based on published performance (intel, Google) on the MobileNet_v2_1.0 BM, input resolution: 224x224*

Hailo-8™ Unprecedented Performance at the Edge

Hailo-8™ vs Edge TPU



Hailo-8™ vs Myriad X



Hailo-8™ **outperforms**

x10 vs. Edge TPU

x2 vs. 4 Edge TPU devices

Hailo-8™ **outperforms**

x26 vs. Myriad X

x6 vs. 4 Myriad X devices

- **Hailo-8 figures are based on SDK Q1 2022 version, measured at room temperature on Hailo-8 device through PCIe interface on a Hailo-8 evaluation board (system host: Intel Core i5-9400 CPU @ 2.90GHz)**
- **Edge TPU figures are for batch=1 and INT8, while Myriad X is batch=1 and FP16**
- **Intel Myriad X figures sourced from: https://docs.openvinotoolkit.org/latest/openvino_docs_performance_benchmarks_openvino.html , retrieved April 2022**
- **Google Edge TPU figures sourced from here and here retrieved April 2022; FPS is converted from latency in ms (1 divided by ms/1000)**

https://hailo.ai/developer-zone/benchmarks/

Hailo-8™ Measured Benchmarks

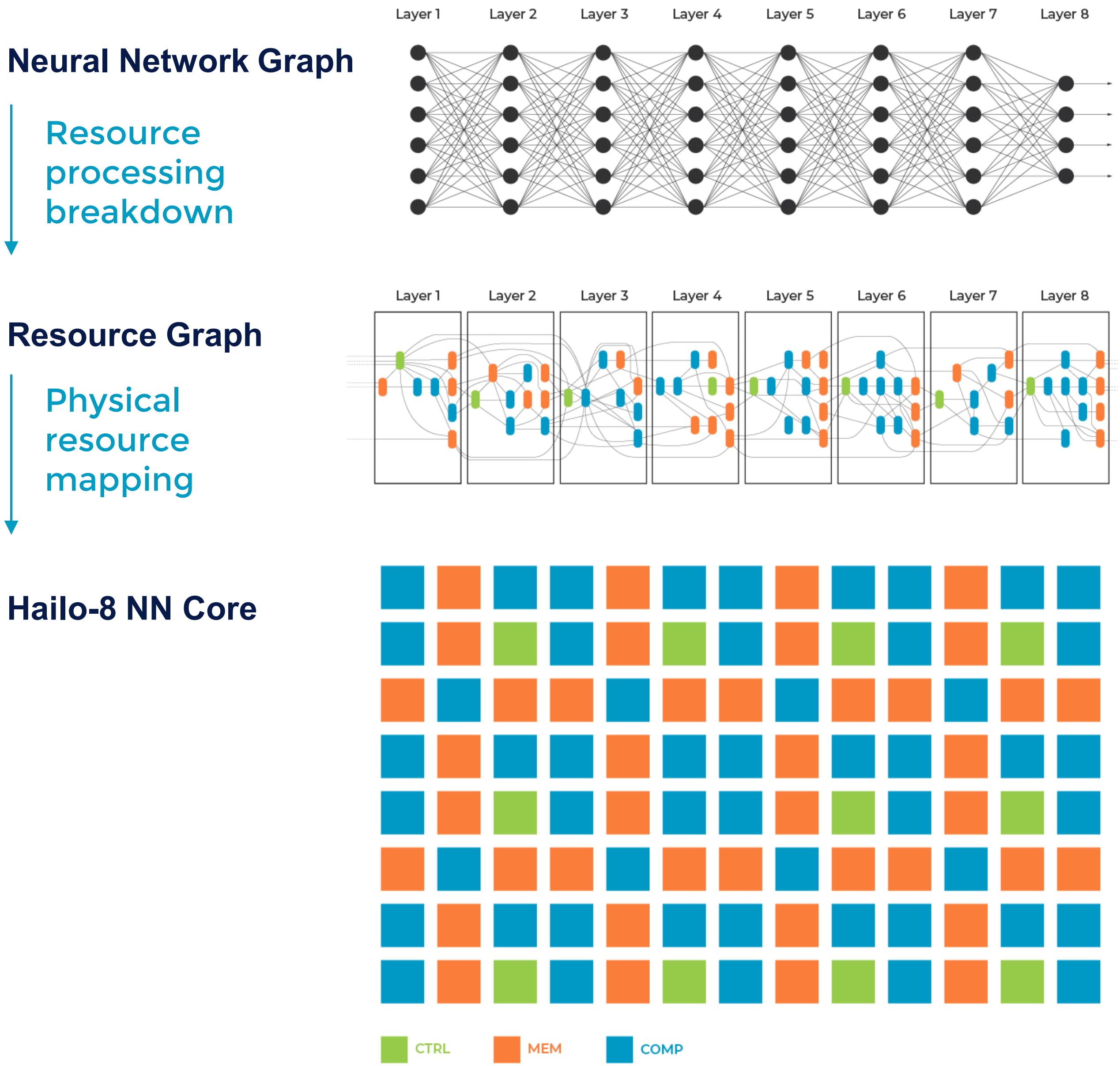
Model	Type	Input Resolution	FPS	Total Power [W]	FPS/W
ResNet-50 v1	Classification	224x224	1,328	3.1	428
MobileNet_v2_1.0	Classification	224x224	2,613	2.1	1,267
MobileNet_v3[4]	Classification	224x224	3,519	1.9	1,852
RegNetx_800mf	Classification	224x224	2,462	2.0	1,254
EfficientNet_M	Classification	240x240	891	3.2	278
SSD_MobileNet_v1	Object Detection	300x300	1,055	2.3	453
Tiny_YOLOv3	Object Detection	416x416	1,010	3.2	315
YOLOv3[5]	Object Detection	608x608	60	4.3	14
YOLOv4[5]	Object Detection	512x512	70	3.04	23
YOLOv5m	Object Detection	640x640	218	4.2	53

Notes:

1. **Based on Dataflow compiler version 3.18.0 (Q2 2022)**
2. **Measurements were taken at room temperature through PCIe interface on Hailo-8 evaluation board**
3. **System host: Intel(R) Core(TM) i5-9400 CPU @ 2.90GHz**
4. **MobileNet_v3 - the benchmarked model flavor is Mobilenet V3 Large Minimalistic**
5. **Performance figures are for processing 8 simultaneous streams**

Hailo-8™ NN Core: Unique, Powerful and Scalable

- **Dataflow** vs. decision making
- **Physically** distributed computation
- **Software abstraction** allows quickly running a variety of NN models
- Smaller elements lead to **Lower power**
- **>20 patents** pending



Neural Network Graph

Resource processing breakdown

Resource Graph

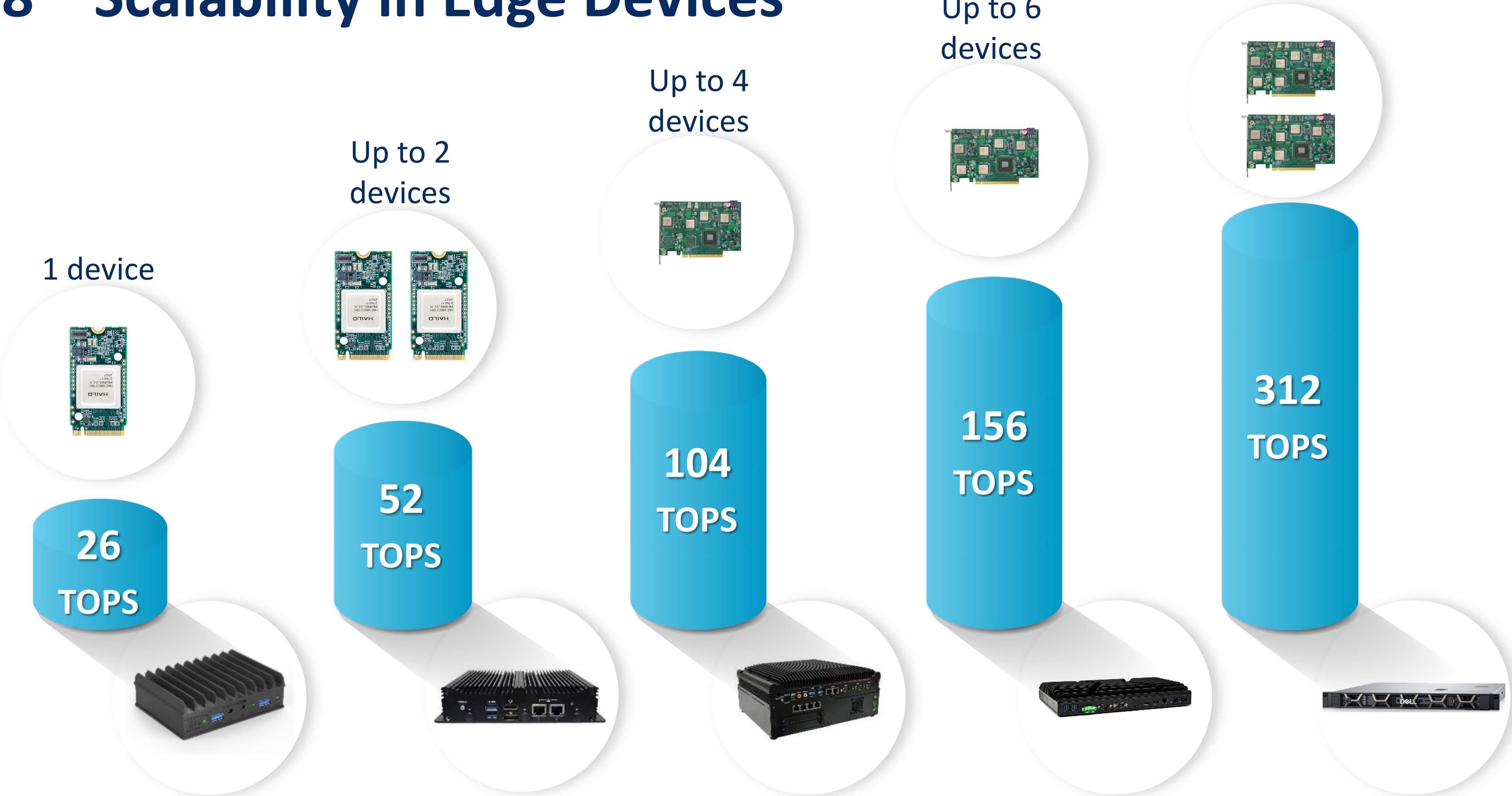
Physical resource mapping

Hailo-8 NN Core

Hailo-8™ Scalability in Edge Devices

1 to **12** devices

26 to **312 TOPS**
of AI processing



1 device

26 TOPS

Up to 2 devices

52 TOPS

Up to 4 devices

104 TOPS

Up to 6 devices

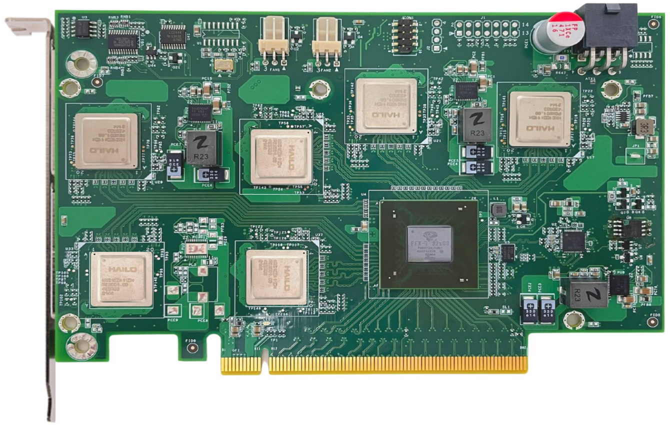
156 TOPS

Up to 12 devices

312 TOPS

Passively cooled | Highly **Scalable** | **Multiple** vendors

Falcon-H8: PCIe Accelerator Card with Multiple Hailo-8™



- Off-the-shelf PCIe for high-performance video analytics systems
  - **PCIe accelerator** card with x4, x5 or x6 Hailo-8™ devices in a standard PCIe single slot form factor provided by Lanner
  - Delivers up to **156 TOPS** for video analytics at a typical power consumption of **35W,** no auxiliary power required
  - Higher **cost-efficiency** (TOPS/$) compared with existing solutions
- Robust software toolchain supports state-of-the-art NN models and applications out-of-the-box
- A powerful platform for edge AI and video analytics:
  - High-performance Edge AI Boxes and video analytics servers for **Smart Retail**, **Smart City**, and more
  - Edge servers, **industrial** PCs and gateways
  - Industrial and commercial **robots**
  - Evaluation and prototyping for **ADAS/AV** sensing

Falcon-H8 Performance, Power and Cost

NVIDIA T4 PCIe

General Purpose GPU

130 TOPS

Power 70W

Falcon-H8 PCIe

Dedicated AI Processors

156 TOPS (w/6 Hailo-8™ devices)

Typical Power 35W

ResNet-50 Benchmark

	Performance [FPS]	Power [Watt]	Power Efficiency [FPS/W]
Falcon-H8 *1 (4x Hailo-8)	5,313	32	166
Falcon-H8 *1 (6x Hailo-8)	7,692	38	202
Nvidia T4 *1	1,109		
Nvidia T4 *2	3,288	70	47
Nvidia T4 *3	4,909	70	70

- 224x224 image resolution
- 8-bit precision
- *1 Batch size = 1
- *2 Batch size = 8
- *3 Batch size = 128
- Source: Nvidia T4 performance

Falcon-H8

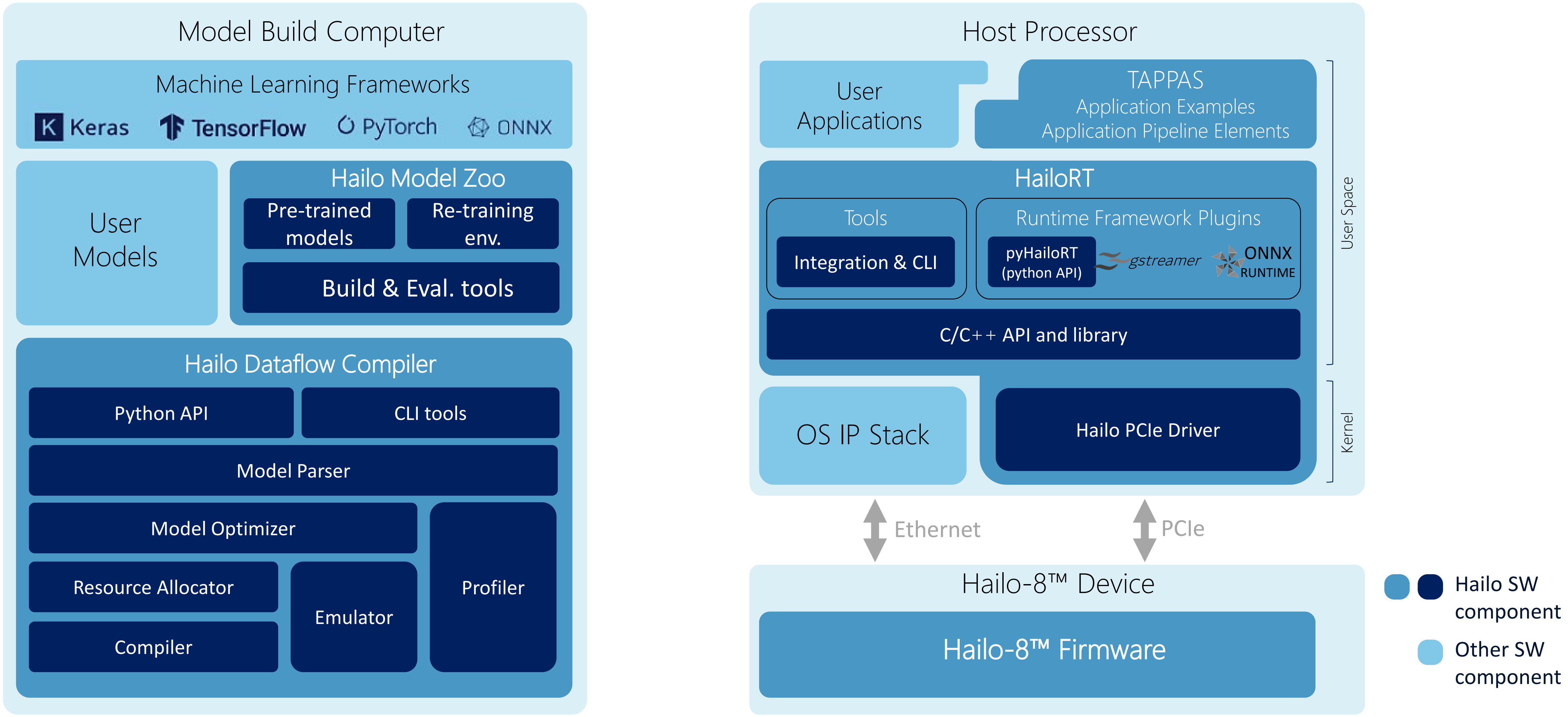
X4.5 Higher
Cost Efficiency

X3 Better
Power Efficiency

Hailo Software Toolchain and Developer Tools

Model Build Environment

Runtime Environment



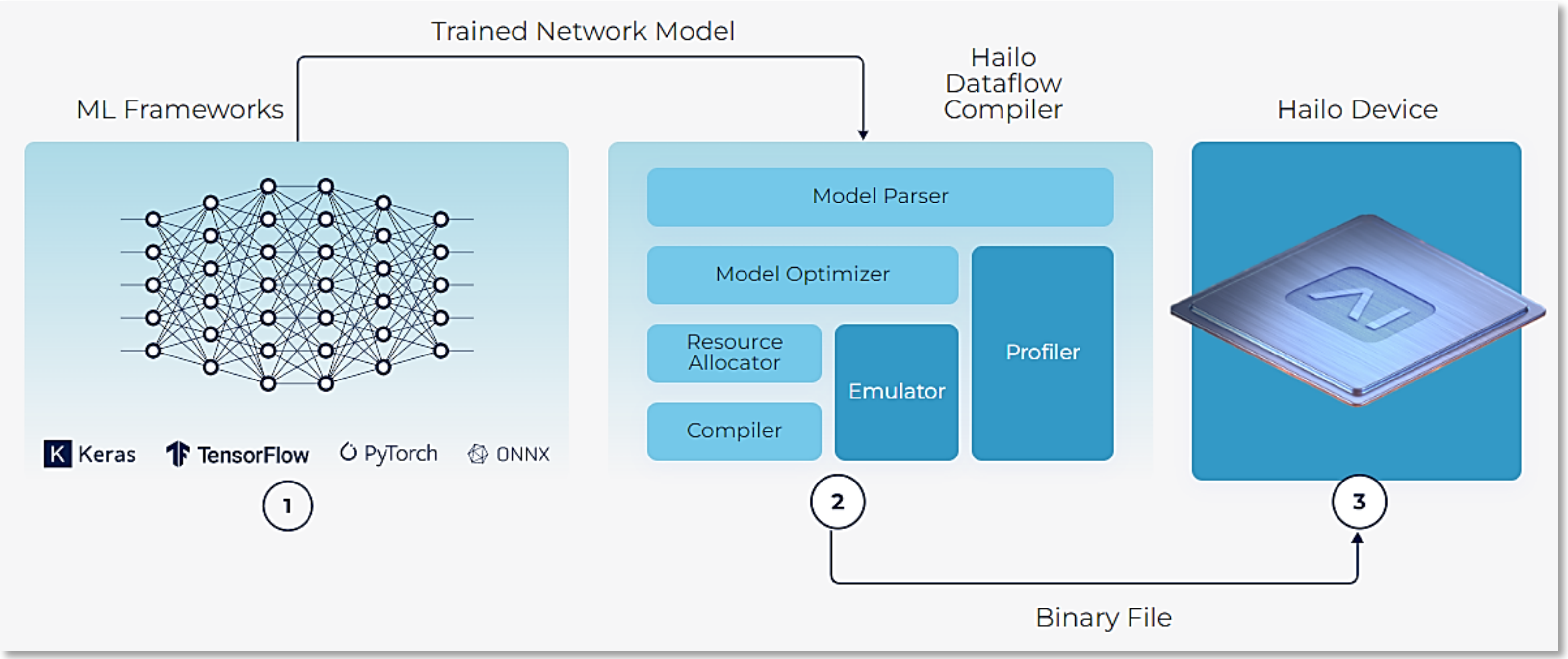
HAILO

Hailo Dataflow Compiler

Automated software toolset

converting trained models to

Hailo's executable format

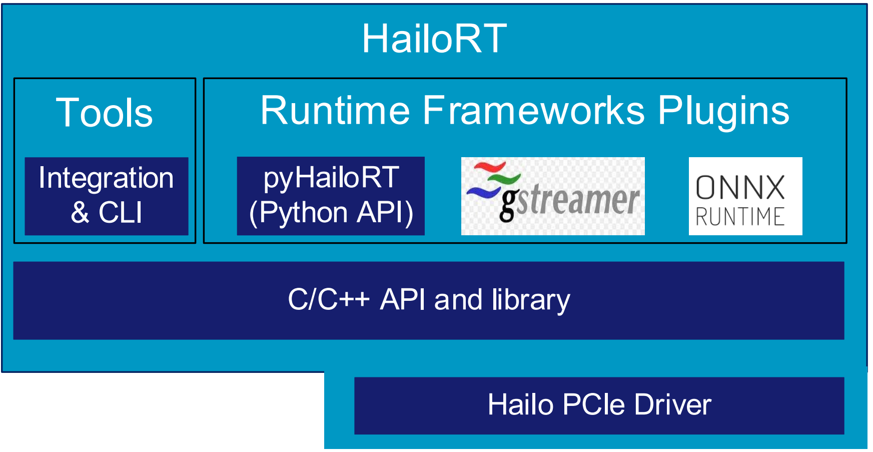


- Efficient quantization scheme allowing flexibility between performance and accuracy
- Automated resource allocation for meeting user's requirements in FPS, latency and power consumption
- Accurate profiling (FPS, power, latency) and bit-exact emulation of expected accuracy
- Supporting multiple Hailo devices and forward compatible

HailoRT Key Software Modules

Production-grade, light, runtime software precompiled for x86 & AArch64 for the host CPU; Open-source in github

- Runtime frameworks Integration
  - pyHailoRT - Python API
  - Standard frameworks support: GStreamer, ONNX runtime
- Integration Tool
  - for verification of the hardware integration of Hailo-8™ M.2 & mPCIe modules
- CLI Tools
- HailoRT Library
  - C/C++ API for control and data transfer to/from Hailo device
- PCIe Driver
  - External kernel module. Can be installed using DKMS framework
- Yocto Layer
  - Enables integration of Hailo's software into Yocto environment
  - Includes recipes for the HailoRT library, pyHailoRT and the PCIe driver



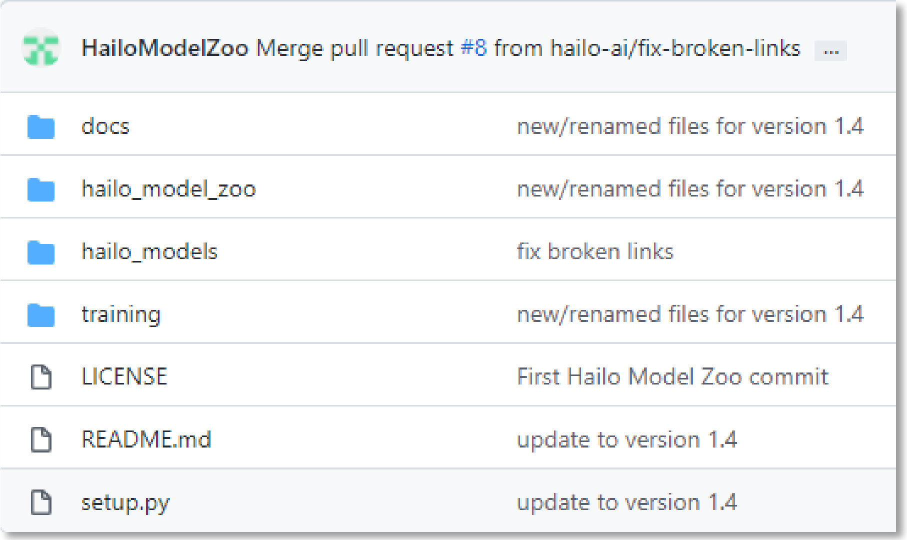
■■ Hailo SW component

□▨ Other SW component

Hailo Model Zoo

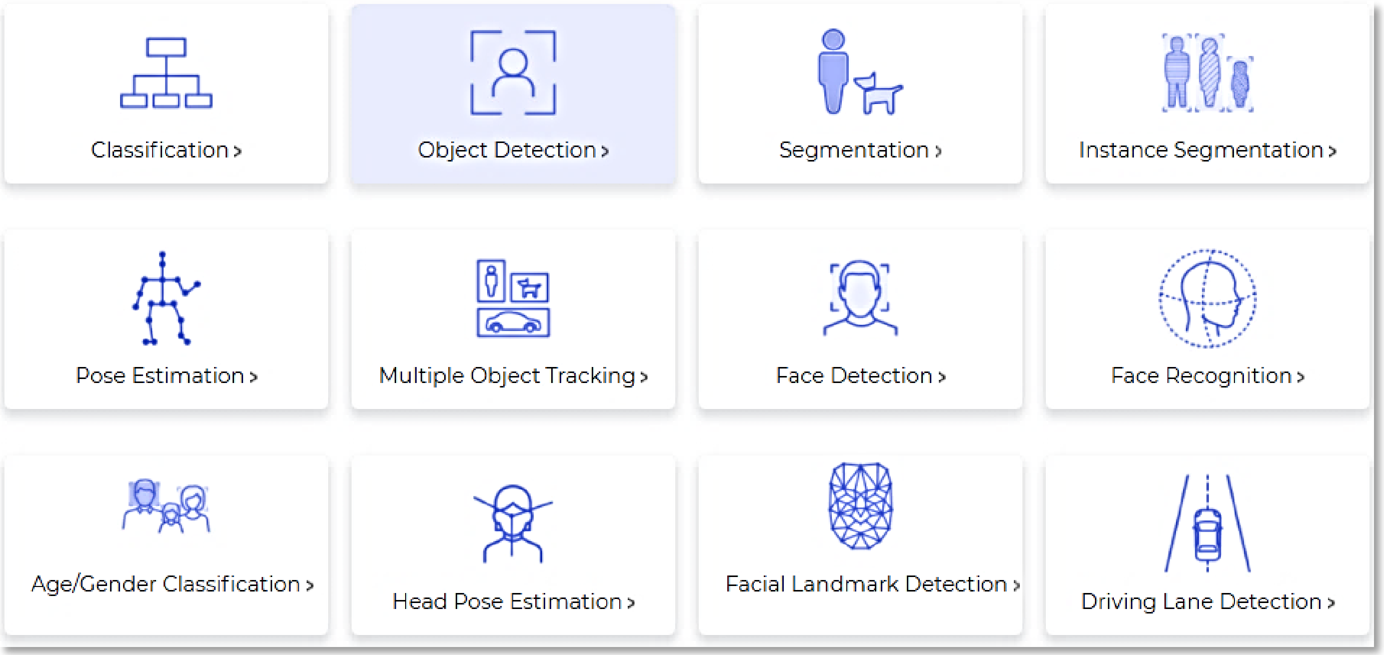
A variety of common and state-of-the-art pre-trained models and tasks in TensorFlow and ONNX

- Opensource repository (available on GitHub)
- Quickly and easily reproduce Hailo-8 performance for evaluation and development
- Models can be re-trained with updated datasets



HailoModelZoo Merge pull request #8 from hailo-ai/fix-broken-links ...

docs	new/renamed files for version 1.4
hailo_model_zoo	new/renamed files for version 1.4
hailo_models	fix broken links
training	new/renamed files for version 1.4
LICENSE	First Hailo Model Zoo commit
README.md	update to version 1.4
setup.py	update to version 1.4



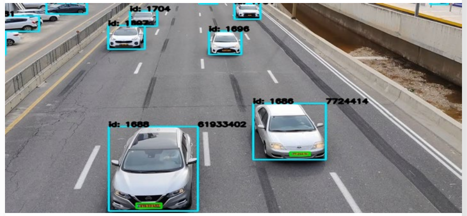
Hailo AI Template APPlications And Solutions (TAPPAS)

Suite of high-performance, pre-trained template AI tasks and applications elements with production-grade pipeline

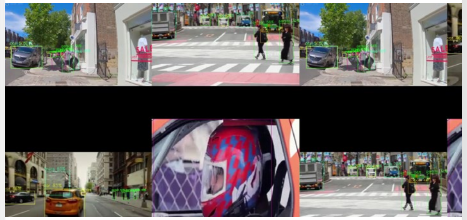
- Suitable for variety of categories and industries
- Useful for demos and can be used as reference designs
  - Accelerate time to market by reducing development and deployment effort
  - Model(s) can be easily replaced

https://hailo.ai/developer-zone/tappas-apps-toolkit/

License Plate Recognition

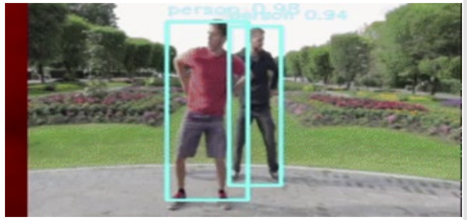


Multi Streams Multi Device Object Detection



Multi Person Multi Camera Tracking

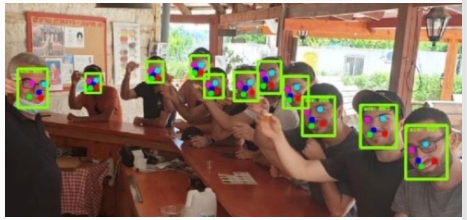
Object Detection and Depth Estimation



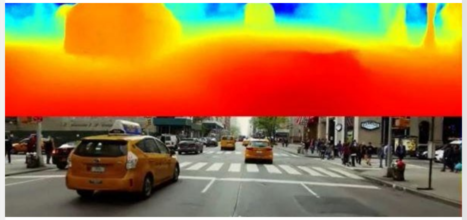
Semantic Segmentation

Pose Estimation

Facial Detection & Recognition



Depth Estimation



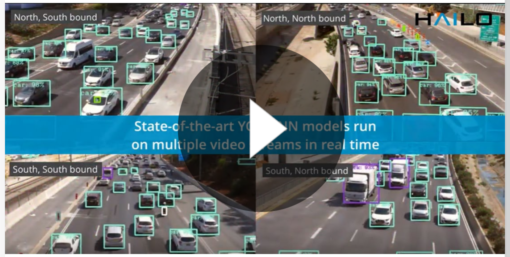
Instance Segmentation

Hailo Demos

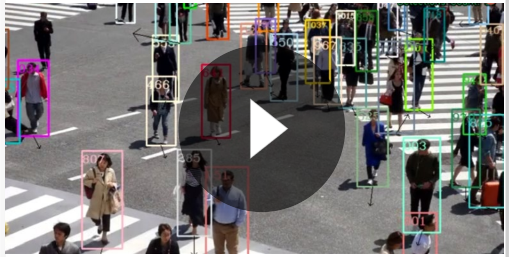
Object Detection
on 15 video streams

Detection with High Power Efficiency

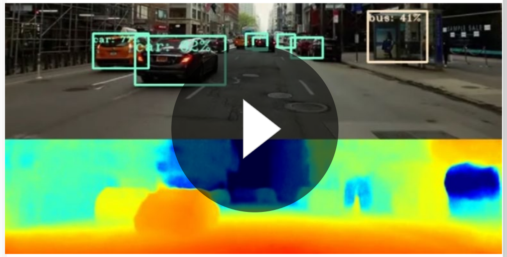
Multi-sensor IVA for Smart City



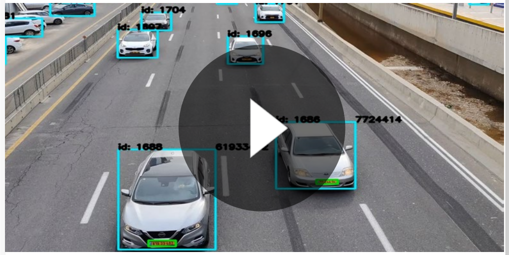
Multiple Object Tracking



Depth Estimation & Object Detection



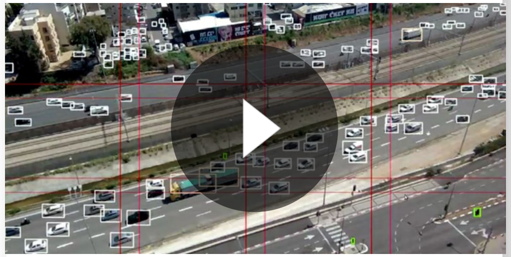
Vehicle License Plate Recognition (LPR)



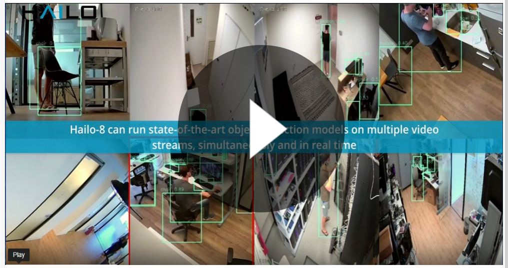
Object Detection w/Yolo V5M

Semantic Segmentation

Improved Object Detection w/ Tiling

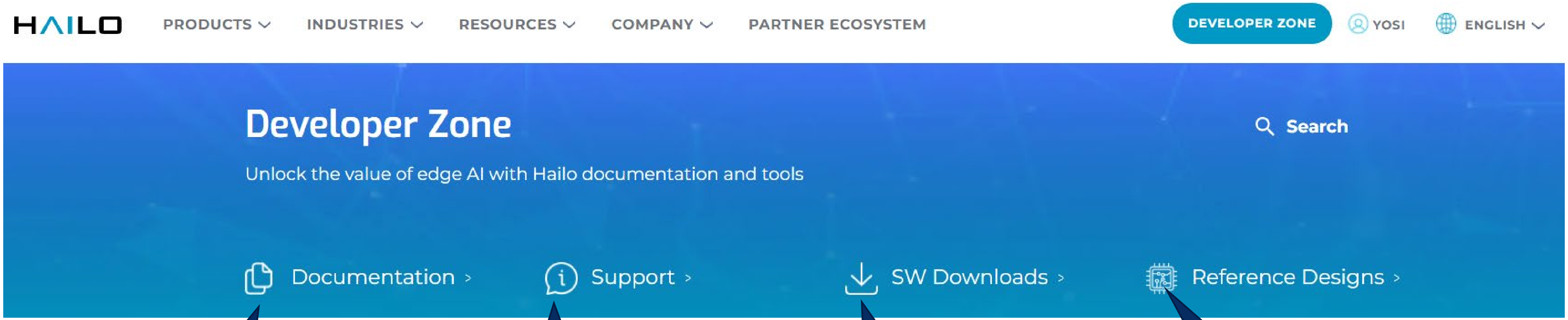


Intelligent NVR Ref Design


HAILO
Hailo-8 can run state-of-the-art obj ction models on multiple video
streams, simultane y and in real time
Play

https://hailo.ai/resources/#demos

Software & Documentation – Developer Zone & Github



HAILO PRODUCTS INDUSTRIES RESOURCES COMPANY PARTNER ECOSYSTEM DEVELOPER ZONE YOSI ENGLISH

Developer Zone

Unlock the value of edge AI with Hailo documentation and tools

Search

Documentation › Support › SW Downloads › Reference Designs ›

Documentation for Hailo products and software suite https://hailo.ai/developer-zone/documentation/

Video tutorials, User guides, App notes https://hailo.ai/developer-zone/customer-support/

Hailo software suite https://hailo.ai/developer-zone/sw-downloads/

Reference Designs https://hailo.ai/developer-zone/reference-design

Open source: https://github.com/hailo-ai/hailo_model_zoo

https://hailo.ai/developer-zone/

HAILO

Empowering Intelligence

THANK YOU

🌐 hailo.ai ✉ contact@hailo.ai