Drive for better vision

WiseEye 2

Himax Ultra Low Power Endpoint AI Processor

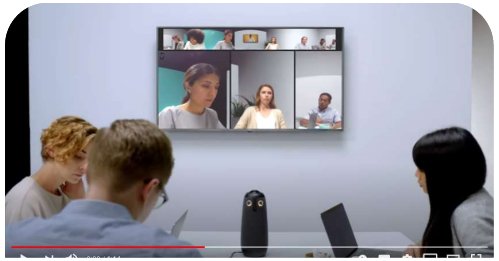
Himax Technologies, Inc.
奇景光電股份有限公司

Always-on Smart Sensing will be Everywhere

Smart Home

Notebook

Smart City



Consumer Appliances

Smart Retail

Wearables

Smart Building

Smart Agriculture

Always-On Sensing Product Roadmap

MP/ES | Developing | Planning

* Planned projects are subject to change without notice

Always-on Sensing ASIC (DL)

- Object detection
- Anomaly detection
- Keyword detection
- Sensor fusion
- Vibration detection

TensorFlow Lite

WE-I Plus 28nm ULP

EM9D, XY	Q3'20

400MHz, Ultra-Low-Power
2 GOPS
2MB SRAM / 2MB Flash
TFLu / MLI Kernels

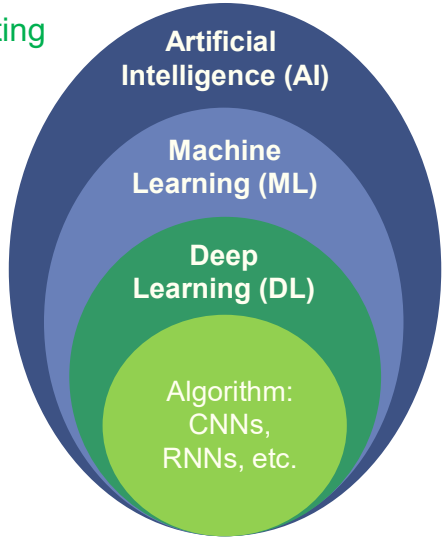
- High FPS tracking & counting
- Object recognition
- Speech recognition
- Biometric awareness
- High FPS Gesture control
- Object detection
- Anomaly detection
- Keyword detection
- Sensor fusion
- Vibration detection

TensorFlow Lite

WE-II 22nm ULL

Dual Cortex-M55 Ethos-U55	Q4'22

400MHz, Extreme-Low-Power
50 GOPS
2.5MB SRAM / 4MB Flash
TFLu / CMSIS-NN Kernels



AI	2020	2021	2022

WE1 AI Processor Overview

System Control

- JTAG
- OSC / PLL
- Clock Generator
- Reset Generator
- DMA
- Timers
- Watchdog Timer

Internal Memory

- 64 KB Boot ROM
- 320 KB Program SRAM
- 320 KB Data SRAM
- 1472 KB System SRAM

DSP

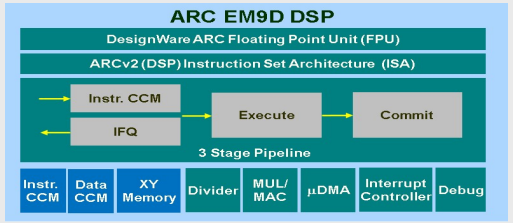


Image Accelerator

- Motion Detection
- 2x2 / 5x5 Filter
- JPEG Codec
- HOG Extraction
- Re-Sampler

Power Management

- LDO
- Power Management Unit

- SPI Master x2
- SPI Slave x1
- I2C Master x3
- I2C Slave x1
- UART x2
- PWM x3
- GPIOs
- I2S Tx
- I2S Rx
- PDM Rx
- 1/4/8-bit CMOS Sensor I/F

ADC

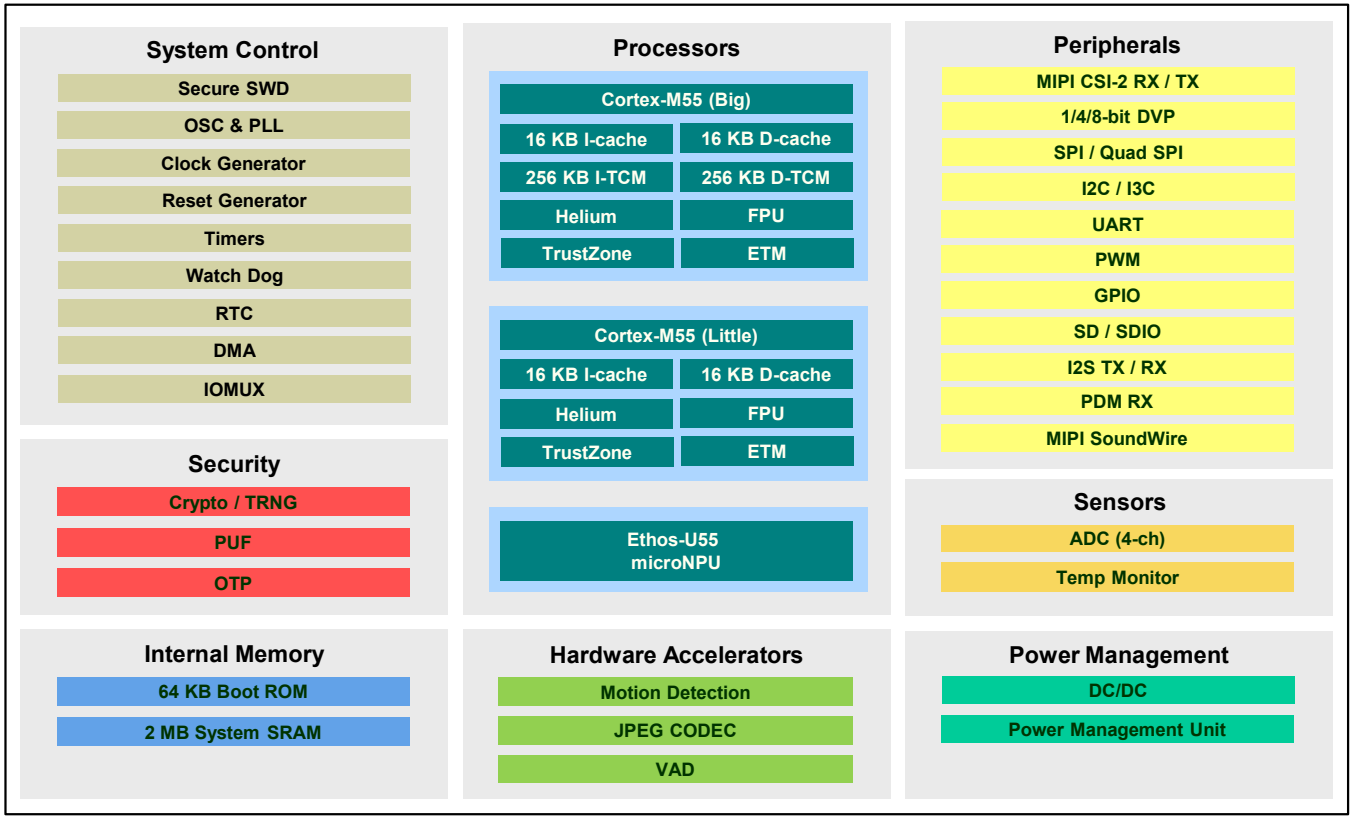
- ADC x1 (4-ch)

Security

- Ciphers and TRNG
- OTP

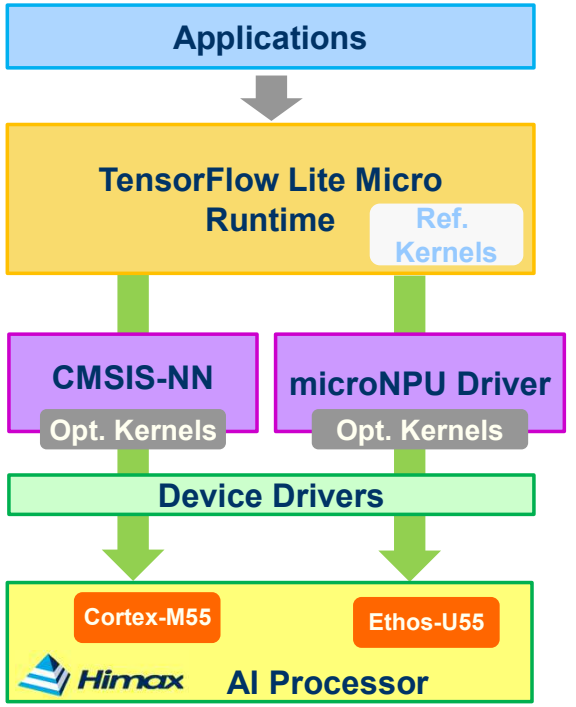
WE2 AI Processor Overview

- Maximum ML computing (50 GOPS)
  - Cortex-M55 (Big), up to 400MHz
  - Cortex-M55 (Little), up to 100MHz
  - Ethos-U55, 64MACs, up to 400MHz
  - 2MB SRAM, 512KB TCM
  - External Flash, up to 32MB
- Optimum energy efficiency
  - DC/DC
  - DVFS power management
  - Internal power islands
- Rich peripherals
  - MIPI CSI-2 RX/TX
  - 1/4/8-bit DVP
  - SPI/I2C/I3C/UART/GPIO
- Security
  - PUF/TRNG
  - Cryptography
  - TrustZone



Open-Source Software Development Platform

- Embedded ML computing engine in AI processor
  - Arm Cortex-M55 MCU with ML acceleration
  - Arm Ethos-U55 microNPU
- Optimized software development flow for Embedded & ML
  - TensorFlow Lite for Microcontrollers
  - Open-source Vela NN optimizer tool for Ethos-U55 microNPU
  - Open-source CMSIS-NN libraries for Cortex-M55 MCU
  - Unified software development flow for MCU and microNPU integration
- Unified flow to accelerate ML performance
  - NN operators are accelerated by microNPU driver by default
  - Fallback to CMSIS-NN, then reference kernels

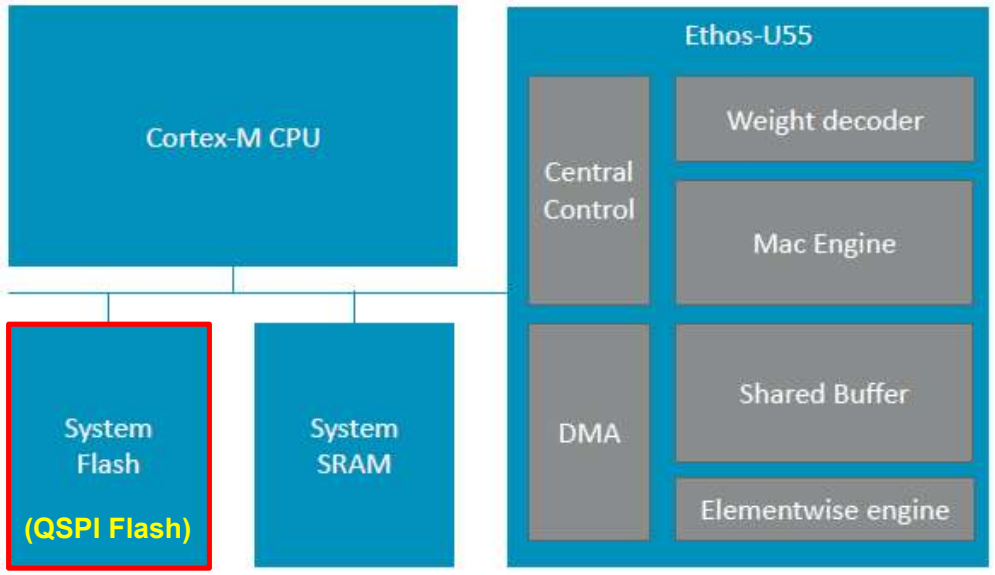


Usecase: TFLu Person Detection – Inference Speed & Energy Efficiency

- Inference time & FPS
  - Cortex-M55 (CPU)
  - Ethos-U55 (NPU)
    - Weights in internal SRAM
    - Weights in external Flash

Processor	Inference (cycles)	Inference (fps)	Ratio
WE-I	14,000,000	29	1.0x
WE-II: Cortex-M55	8,418,498	48	1.7x
WE-II: Ethos-U55 (SRAM)	441,446	906	31.7x
WE-II: Ethos-U55 (Flash)	3,072,112	130	4.6x

***Model: TFLu MobileNet V1 250KB INT8 Person Detection**



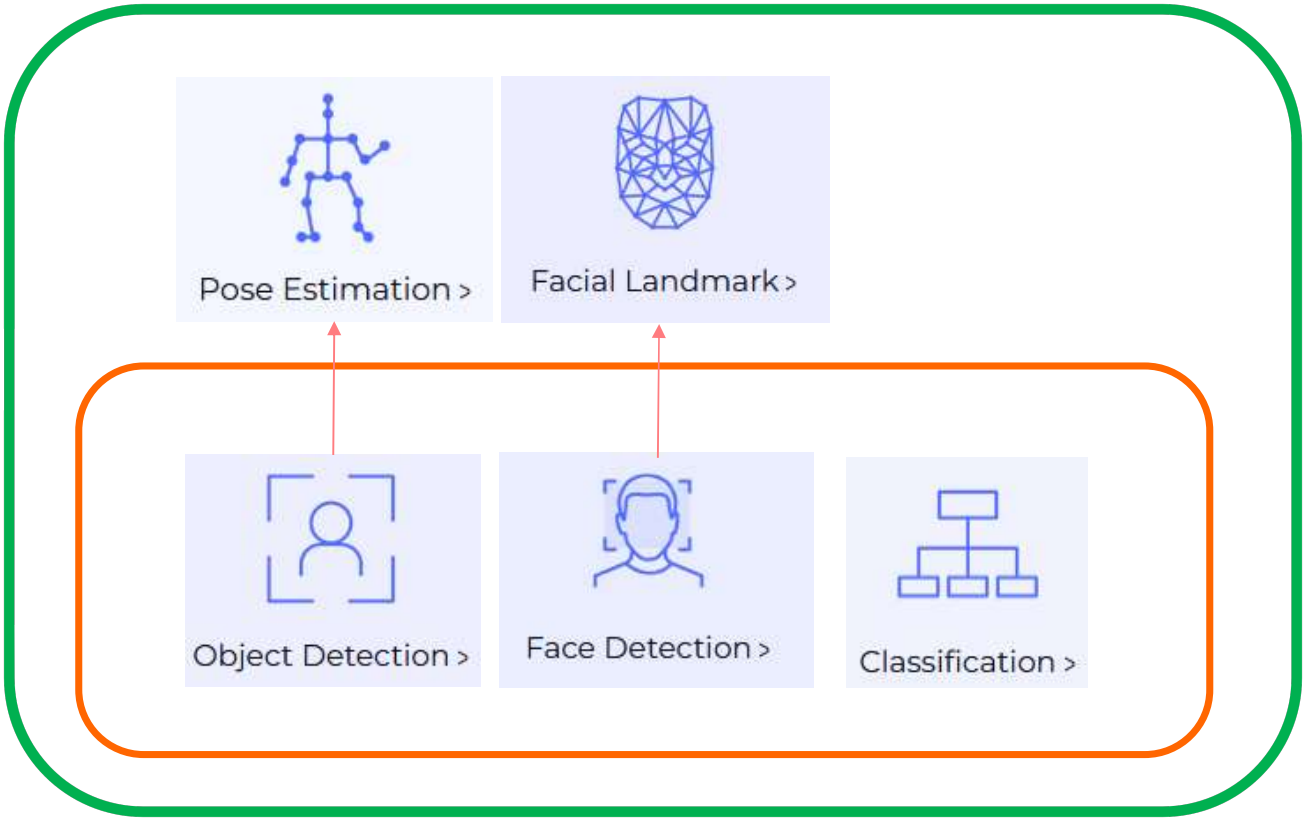
WE2 reference model

- Object Detection
  - Models: EfficientNet-lite0 and Yolo-Fastest-1.1-xl
  - Platform: TFLM and Vela optimizer
  - ML inference: Ethos-U55 NPU only
    - All operators are supported by TFLM and Ethos-U55

Neural Network	Data Type	Weight Size	Image Resolution	Inference Cycle Count (Weight in SRAM)	Inference Cycle Count (Weight in Flash)
EfficientNet-lite0	INT8	4.7 MB	224x224 (RGB)	N/A	61,147,848 (6.5 FPS*)
Yolo-Fastest-1.1-xl	INT8	0.925 MB	256x256 (RGB)	13,082,766 (30.5 FPS*)	25,919,207 (15.5 FPS*)

*Ethos-U55@400MHz

DL Model Apps: Beyond “Bounding Box” -> “2D Key Points” Detection



https://hailo.ai/products/hailo-software-suite/model-zoo/

Key Point Detection in WE2

68 points face landmark detection

Run 4 models for each frame

VGA@12FPS

Key Point Detection in WE2

17 key points of human pose

qVGA@6fps

https://blog.tensorflow.org/2021/05/next-generation-pose-detection-with-movenet-and-tensorflowjs.html

WE2 vs. WE1

32x increase machine learning computing

50x increase energy efficiency

∞ Endpoint AI Imagination

Drive for better vision