2023 embedded VISION SUMMIT

Introducing the i.MX 93: Your "Go-To" Processor for Embedded Vision

Srikanth Jagannathan Senior Product Manager NXP Semiconductors



Outline

- Introduction to NXP application processors
- Introduction to i.MX 93
 - Architecture
 - Features
 - Target applications
- i.MX 93 machine vision features
- Machine vision software enablement flow and demo



Scalable Compute Platforms



i.MX & LAYERSCAPE APP PROCESSORS

FUNCTIONAL INTEGRATION





i.MX 93 Applications Processors Family



i.MX 6

- 12 product families
- Offers software and pin-pin compatibility between families for easy compute scalability
- Arm[®] v7-A processors

i.MX 8 Family Advanced Graphics, Vision & Performance

i.MX 8M Family General Purpose Edge Computing

i.MX 8ULP Family Ultra Low Power with Graphics

i.MX 8X Family Safety certifiable & Efficient Performance

i.MX 7 Family Flexible Efficient Connectivity





Reliable Processors for Long-life Platforms

 NXP's Product Qualification program considers conditions experienced by the processor when running in customer application

Qualification Level	Characteristics	Temperature Range
Commercial or Consumer	SoC is 50% on for 5-year life	Typically: 0°C to +95°C Tj
Industrial (Longest operating life)	SoC is 100% on for 10-year life	Typically: -40°C to +105°C Tj Extended: -40°C to +125°C Tj
Automotive	SoC is 10% on for 15-year life	Typically: -40°C to +125°C Tj

NXP's Product Longevity program goal is to ensure stable supply of products for automotive and industrial designs

	NXP i.MX 93 Series Application Processor Families are part of the NXP 15-year Product Longevity Program																
2	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039



embedded

SUMMI

i.MX 93 Family – Differentiated Features



© 2023 NXP Semiconductors

embedded

VISION

SUMMIT



i.MX 93 Family: Vision Applications Across Segments

- Keys features used in vision applications
- Camera interface: MIPI-CSI
- Other Interfaces: USB, Ethernet, Wi-Fi connectivity and other low-speed interfaces (SPI, UARTs etc)
- Processing: A55, NPU, M33

Industrial Automation	Smart Home	Smart City	Automotive
 Industrial Machine vision 	Smart doorbell	 Smart lighting 	 Driver Monitoring System (DMS)
 Industrial scanning/printing 	Smart lock	Traffic control	 Object Monitoring System (OMS)
	 Smart home hub/ Hue Bridge 		















embedded

SUMMIT

SIO



© 2023 NXP Semiconductors

AL/ML NPU (Neural Processing Unit)



Expanding Edge ML with ARM[®] Ethos™-U65



- High efficiency and small memory footprint
- HW acceleration for high compute NN + Cortex-M for other operations
- Model compression and on-the-fly weight decompression
- Comprehensive software and tools with NXP's eIQ[®] ML Software Development Environment



embedded

SUMMIT



BRINGING MCU-CLASS ML EFFICIENCY TO THE CORTEX-A WORLD

Ethos-U65 Hardware Architecture





- The Ethos-U system includes Cortex-M, Ethos-U65, SRAM and Messaging Unit (MU).
- Ethos-U65 is controlled by Cortex-M. Cortex-A has no direct access with Ethos-U NPU.
- Communication between CPUs is through DDR memory and the MU.
- Ethos-U NPU can access both DDR and SRAM over embedded DMA interface.
- Model weights, bias, input and output feature maps are stored in DDR
- SRAM size is 384 KB



Ethos-U65 Software Architecture



- TensorFlow Lite (TFLite) interpreter or Inference API to run the model on NPU:
 - TFLite interpreter: NPU optimized operations on NPU, all other operations on Cortex-A55.
 - Inference API: all operations not supported by NPU will be fallback to Cortex-M core.
 - Requires precompiling a model to be able to run on the NPU
 - Non-compiled model will run on Cortex-A55 core only.
 - Users can compile models using the **Arm Vela compiler** or **NXP eIQ Portal.**

embedded

SUMMIT

VISIO

The ARM Vela Compiler

- ARM Vela compiler tool converts a quantized TFLite model and exports it to an optimized version that can be run on the Ethos-U65 NPU (Vela model).
- Conversion is one way: TFLite \rightarrow Vela model only
- Converts TFLite operators to custom operators that can run on the NPU
 - Unsupported TFLite operators will run on CPU
- Additionally, tool compresses model to reduce model size (~70%) and SRAM size (~90%)
- Tool is **free to download** and open-source: <u>https://github.com/nxp-imx/ethos-u-vela</u>









Compiling a Model for i.MX 93 SoC



Method 1: Use eIQ Portal's Model Viewer

	🗠 mobilenet_v1_	1.0_224_quant.tflite	- elQ Model Tool 2.7.0				-		×	
	File Edit View	Manage Targets	Profiling Help							
	Open								_	
	Open Recent									
	Export									
	Convert		ONNX (.onnx)							
	Close		Deepview RT (.rtm) Tensorflow Lite (.tflite)	onversion	Options	×				
	Exit	Ente	Tensorflow Lite Vela/iMX93 (.tflite) r Model Name		1 224 224	3				
		Input I	Names ⑦		Output Names ⑦					
		inpu	t + /		MobilenetV1/Prediction	+ Add Output				
2		Labels	file ⑦		Quantization Settings ⑦					
1		Cho	bose File No file chosen		Enable Quantization ?					
1	Vela									
	Vela Configuration File ⑦									
ł	Choose File vela.ini									
	Convert									
				Conv2D						

Method 2: Use the Command Line Tool

\$ git clone https://github.com/nxp-imx/ethos-u-vela.git \$ cd ethos-u-vela \$ git checkout lf-5.15.71_2.2.0 \$ pip3 install . \$ vela mobilenet_v1_1.0_224_pb_int8.tflite

Both methods create a Vela model file that can be used on the i.MX 93's NPU



NPU Performance Increase for Quantized Models



 Measured inference/sec on i.MX 93 with NPU and Cortex A55 running at 1 GHz

embedded

VISIO

SUMMIT



Machine Vision Software Enablement



eIQ ML SW Development Environment





Solutions and Services from NXP and NXP Eco-System Partners

- ML Applications
- Optimized Models
- Optimization Tools and Modules
- Development tools
- • Datasets
- Training
- Sensor solutions



eIQ ML SW Development Environment





eIQ ML SDE Inference Engine Options



NXP eIQ[®] ML Software Development Environment



* Planned

Additional support for devices not listed can be available or requested

Machine Vision Use Case & Demos



Sample Enablement Flow Video and Demo





Detects movement of car driver to estimate alertness to drive

Driver Monitoring System (DMS) on the i.MX 93

When driving, milliseconds can save a life. **i.MX 93's on-board NPU** allows for quick and responsive alerts to keep the driver focused on the road.

Driver OK (0.0%)

Distracted: No Drowsy: No Yawn: No Smoking: No Phone: No





2023 NXP Semiconductors





- NXP just launched the i.MX 93 applications processor family first in next generation i.MX 9 series
- i.MX 93 enables customers to develop cost-optimal solution for low-/mid-end vision applications using following features
 - Hardware: NPU, MIPI-CSI interface
 - Software: eIQ[™] toolkit
 - Enablement: plethora of reference demos targeting machine vision application
- Please email <u>Srikanth.Jagannathan@nxp.com</u> and <u>Ali.Ors@nxp.com</u> for any questions related to i.MX 93 samples/EVKs/eIQ[™]/ML Demos etc.







- i.MX 93 product page: www.nxp.com/imx93
- i.MX 93 evaluation kit (EVK) page: www.nxp.com/imx93evk
- eIQ[™] page <u>www.nxp.com/EIQ</u>
- Please visit NXP booth for more demos and discussion



Thank You

