How to develop edge AI camera applications simpler, faster, and more affordably

Shyam Jagannathan, TI

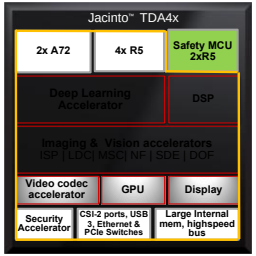
Manisha Agrawal, TI

Agenda

- Recap
- Edge AI SDK
  - Architecture overview
  - Example camera inferencing application dataflow
  - GStreamer plugin designs
  - SDK roadmap and performance
- Demonstration
- Call to action

TI edge AI | revolutionizing applications from factory to home

Processor for practical edge AI



- High-speed image acquisition
- Low-latency, low-power vision and AI processing

ti.com/edgeai

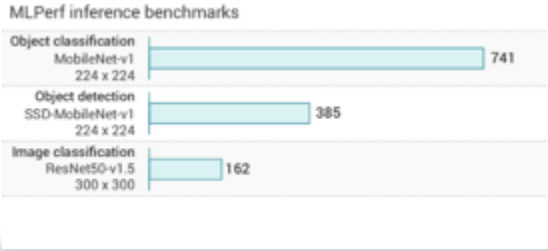
Learn with Free Cloud Tool

ti.com/edgeaicloud

Build with 8 TOPS starter kit

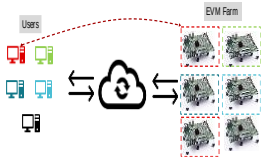
P/N: SK-TDA4VM: <$250

Energy efficient AI architecture



Get started for free

- Example scripts
- TI Model Zoo
- Training videos



Fast Development Cycle

Industry standard APIs

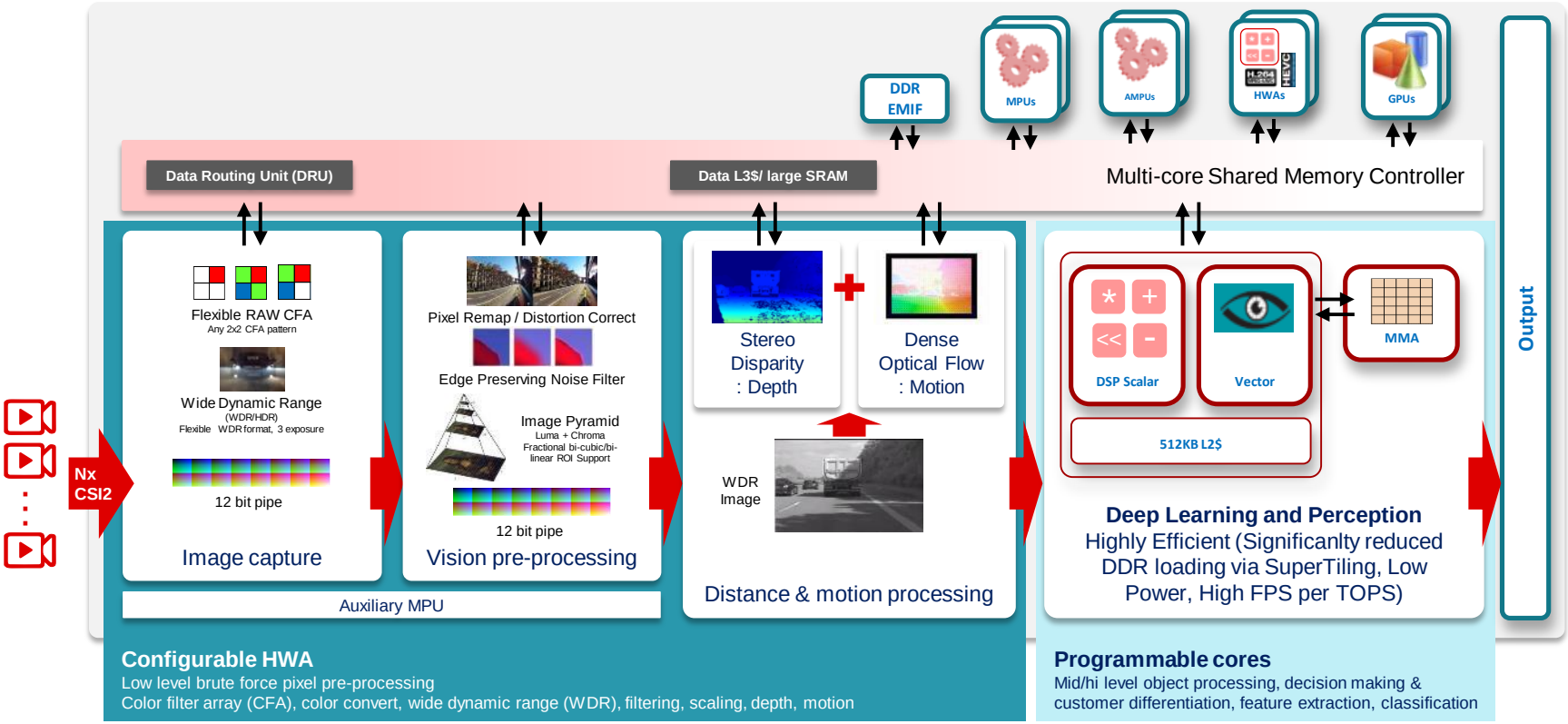
gstreamer, tvm, ROS, docker, TensorFlow Lite, ONNX

3P Eco-system

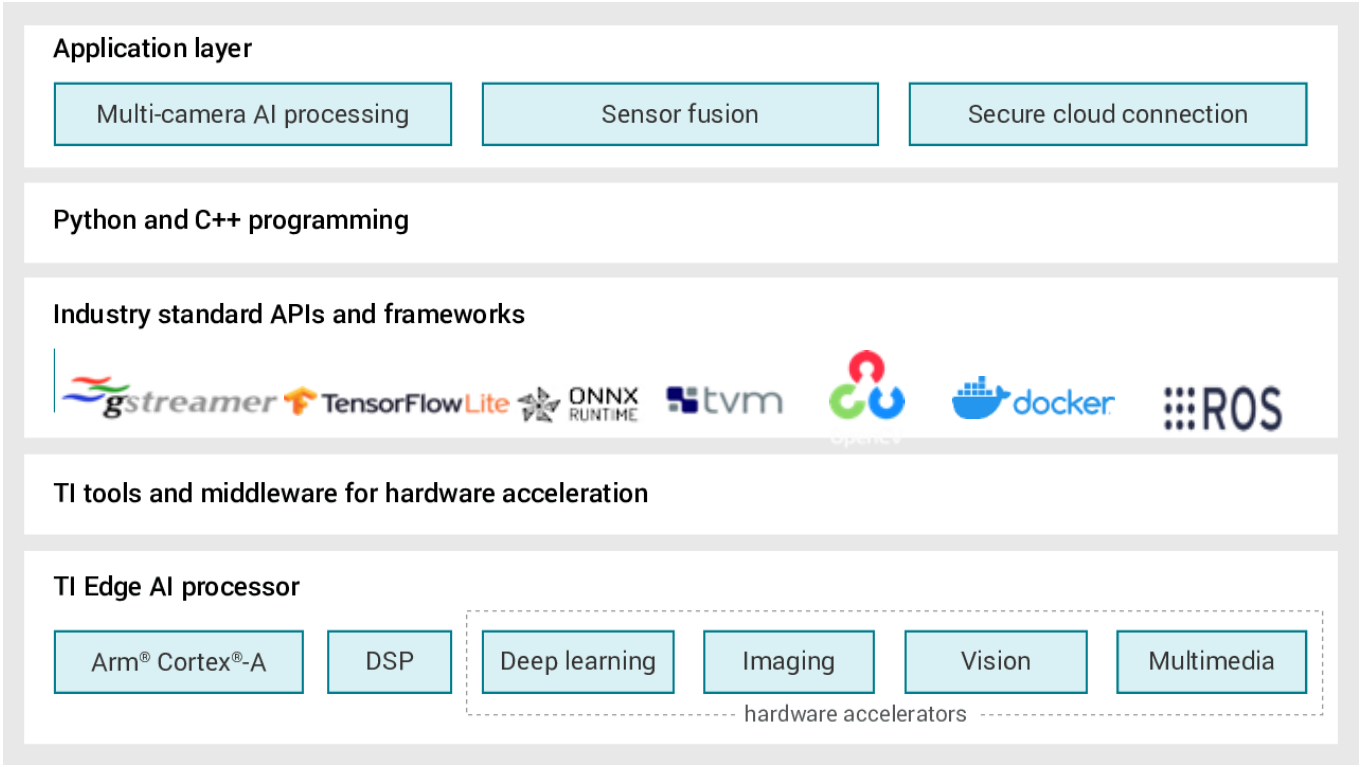
aws, RidgeRun, Allied Vision, PHYTEC, ignitarium, D3 Engineering, kudan

ti.com/edgeai for all the resources you have to get started!

Efficient data movement | more system level performance

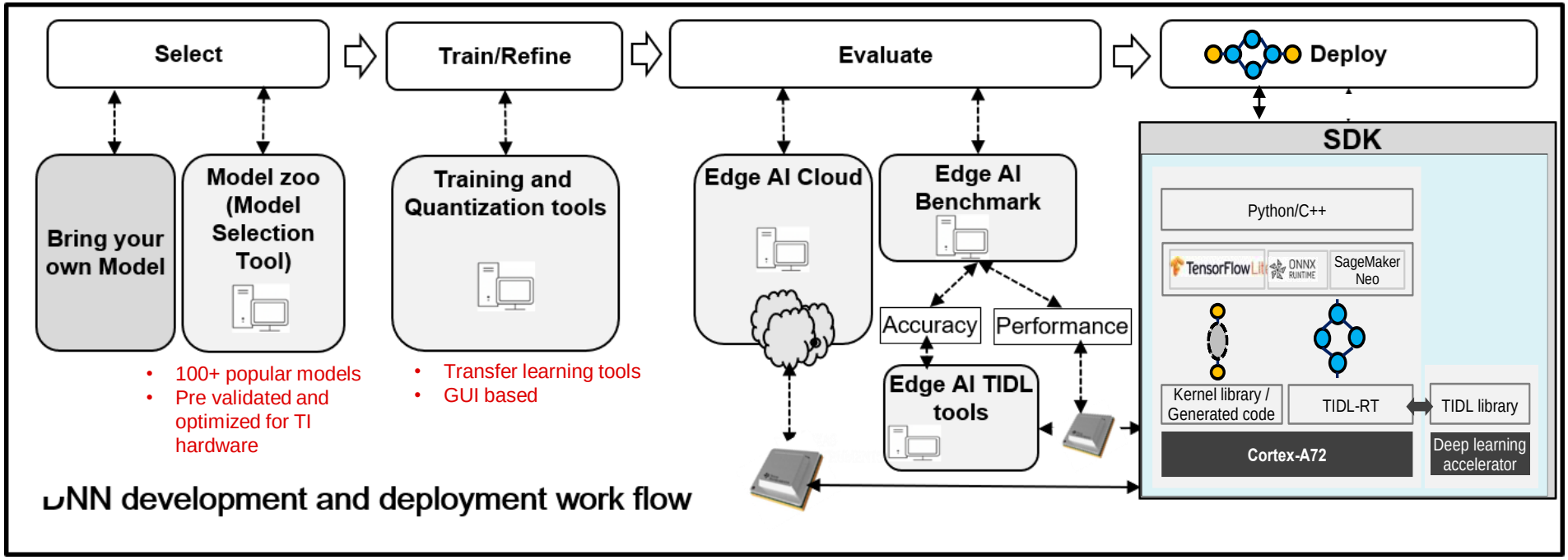


It is easy to develop | With industry standard APIs



Full software from TI makes applications and demo development much easier!

Extensive tools | Faster DL model development & deployment



Edge AI Starter Kit | What you can build

Make factories, cities and home - smart and safe

Learn through Academy

www.ti.com/edgeaiacademy

Fundamentals

Learning
Part 1: Introduction to AI
Learn about AI applications, AI functions
Get started >

Learning
Part 2: Understanding an AI System
Learn about Deep learning system architecture
Get started >

Learning
Part 3: SW Architecture
Learn the details of SW programming environment
Get started >

Step-by-step code development

Hands-on-coding
Part 4: Building "Hello-World" AI application
Build your first AI app on a PC and port it to TDA4VM
Get started >

Hands-on-coding
Part 5: Deep Learning Deployment Demystified
Model compilation without
Get started >

Hands-on-coding
Part 6: Putting it all together: end-to-end AI application development
Video in, Analytics, Video Out
Get started >

Explore, build and contribute to projects

www.ti.com/edgeaiprojects

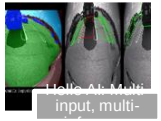
Smart cameras and AI Boxes


Hello AI: Classification

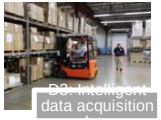

Hello AI: Object detection


Hello AI: Semantic segmentation


Hello AI: Single input, multi-


Hello AI: Multi input, multi-

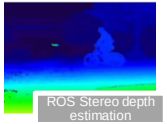

AWS: People counter IoT

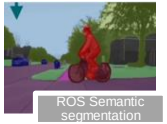

D3: Intelligent data acquisition


Ignitarium: People protection


Ignitarium: Multi-channel AI server

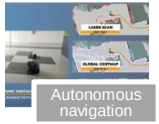
Robotics


ROS Stereo depth estimation


ROS Semantic segmentation


ROS 3D obstacle detection


ROS Visual localization


Autonomous navigation


Kudan SLAM

Customers, Third-party, Community & Hobbyist

Edge AI SDK | Overview

EdgeAI applications

Deep learning Python/C++ demo applications on Yocto/Docker

- TFLite runtime, ONNX runtime, Neo-AI-DLR
- edgeai-tidl-tools – Standalone examples, Jupyter notebooks
- edge_ai_apps – GStreamer based Deep Learning demos
- edgeai-gst-plugins – GStreamer custom TI plugins
- Gstreamer
- OpenCV

Features supported on LINUX

Aarch64 Linux OS

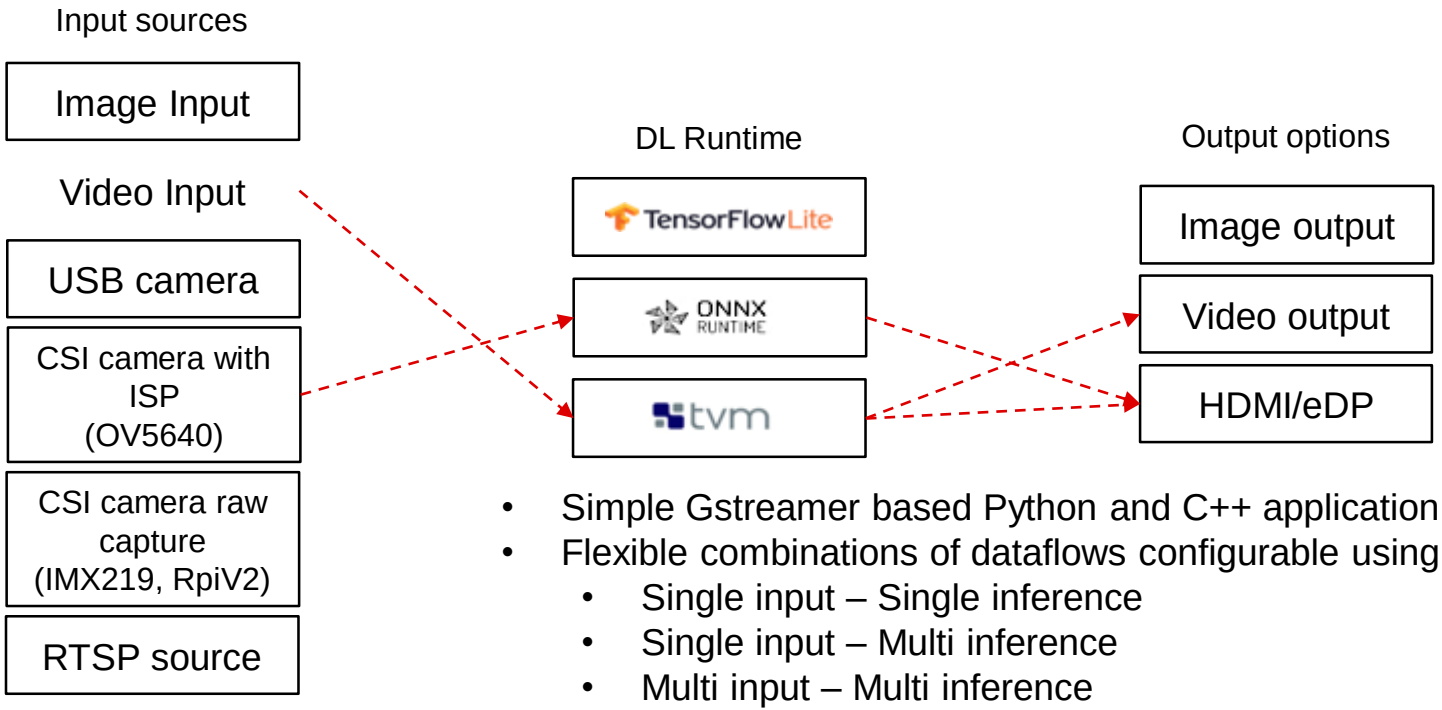
- Bootloader, Linux kernel, Filesystem
- Secure boot, ARM Trusted Firmware
- Foundational tools and drivers
- V4L2 Capture driver (CSI, USB)
- Display driver (DSS)
- HWA Codecs driver (Decoder, Encoder)
- Image sensor tuning tool

Features supported on HWA and RTOS

Remote core software running RTOS

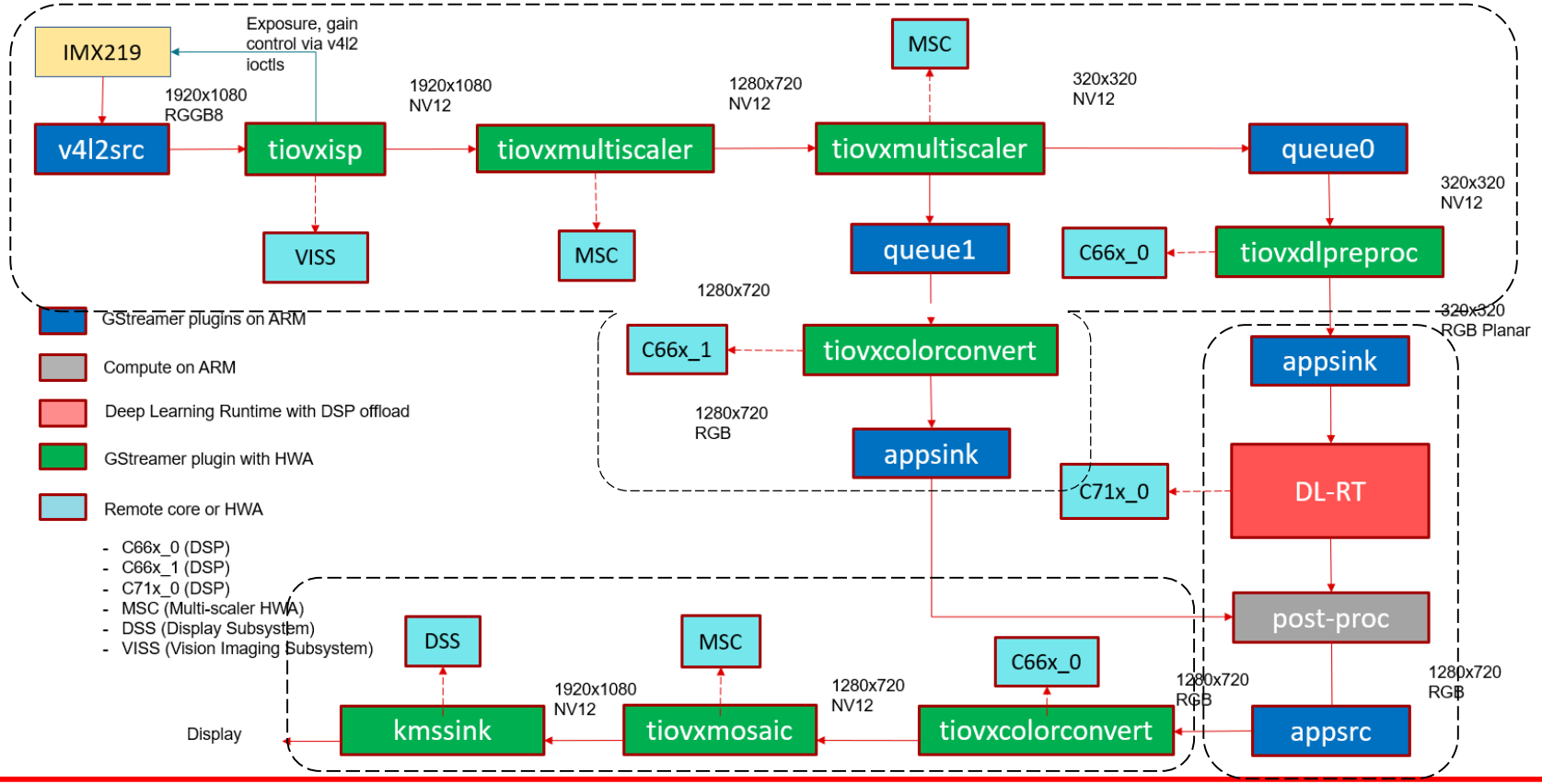
- TI OpenVx framework and libraries
- Vision Hardware accelerator drivers (VPAC/DMPAC)
- TI Deep Learning Runtime libraries
- Optimized DSP libraries (MMALIB, TIADALG, VXLIB)
- AUTOSAR MCAL and tools for MCU SW
- Safety Software Diagnostic Libraries (SDL)
- Baremetal and RTOS low-level drivers (PDK)

Edge AI SDK | Flexible use cases

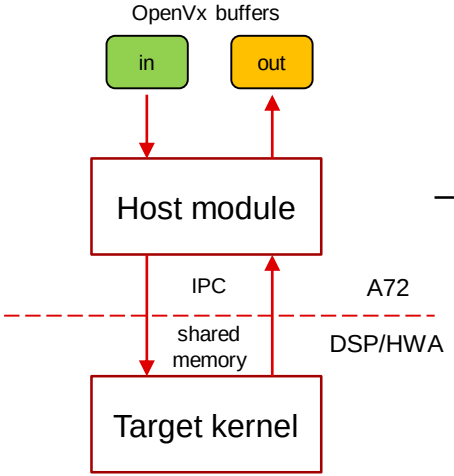


- Simple Gstreamer based Python and C++ application
- Flexible combinations of dataflows configurable using YAML
  - Single input – Single inference
  - Single input – Multi inference
  - Multi input – Multi inference
- Pre-validated with over 100+ models from edgeai-modelzoo
- End-to-End Zero data copy pipelines

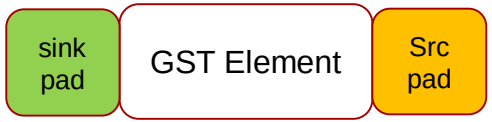
Edge AI SDK | Camera application end-to-end acceleration dataflow



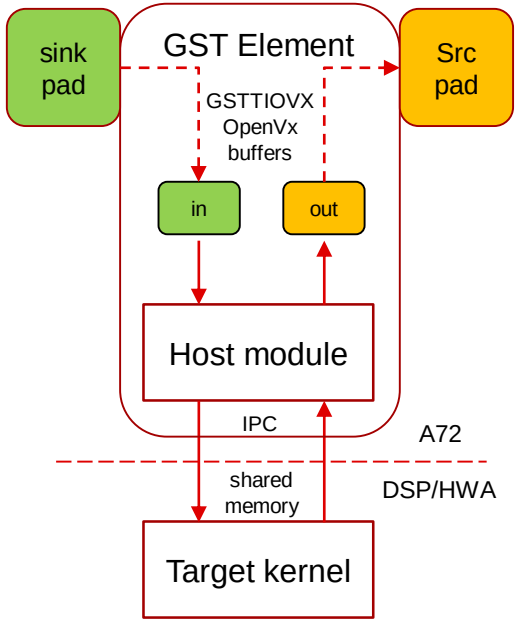
Edge AI SDK | Custom GStreamer elements using OpenVx



Typical OpenVx Node



Typical GStreamer Element



GStreamer element using OpenVx to access DSP/HWA

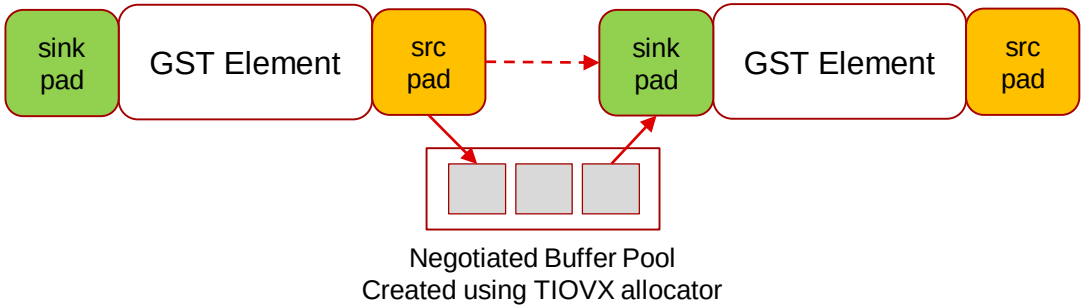
https://github.com/TexasInstruments/edgeai-gst-plugins/wiki

Edge AI SDK | Custom GStreamer foundational elements

Custom GStreamer Buffer Pool

GstBufferPool
- GstTIOVXBufferPool
  - GstImageBufferPool
  - GstTensorBufferPool
  - GstRawImageBufferPool

Enables zero buffer copy dataflows



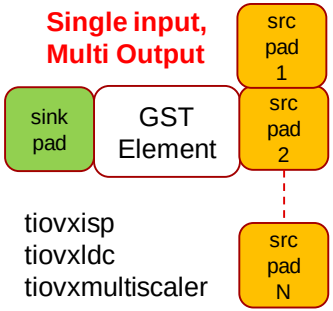
Custom GStreamer elements
- GstElement
  - GstBaseTransform
    - GstTIOVXSiso
    - GstTIOVXSiMo
    - GstTIOVXMiSo

Defines generic classes for mapping different OpenVx modules

Single input, Single Output

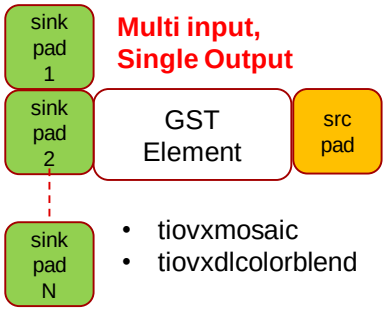
- tiovxcolorconvert
- tiovxdlpreproc

Single input, Multi Output



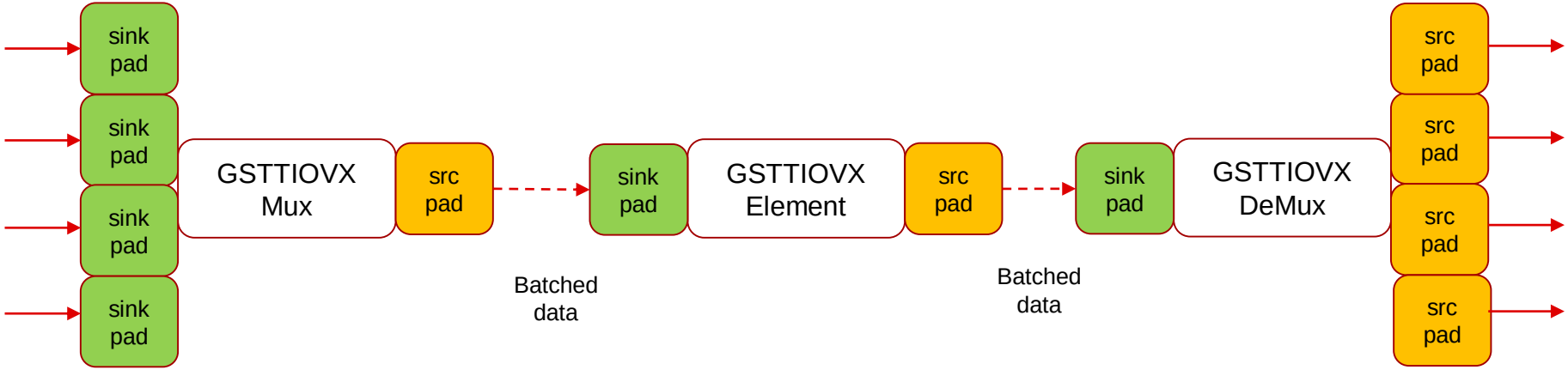
- tiovxisp
- tiovxldc
- tiovxmultiscaler

Multi input, Single Output



- tiovxmosaic
- tiovxdlcolorblend

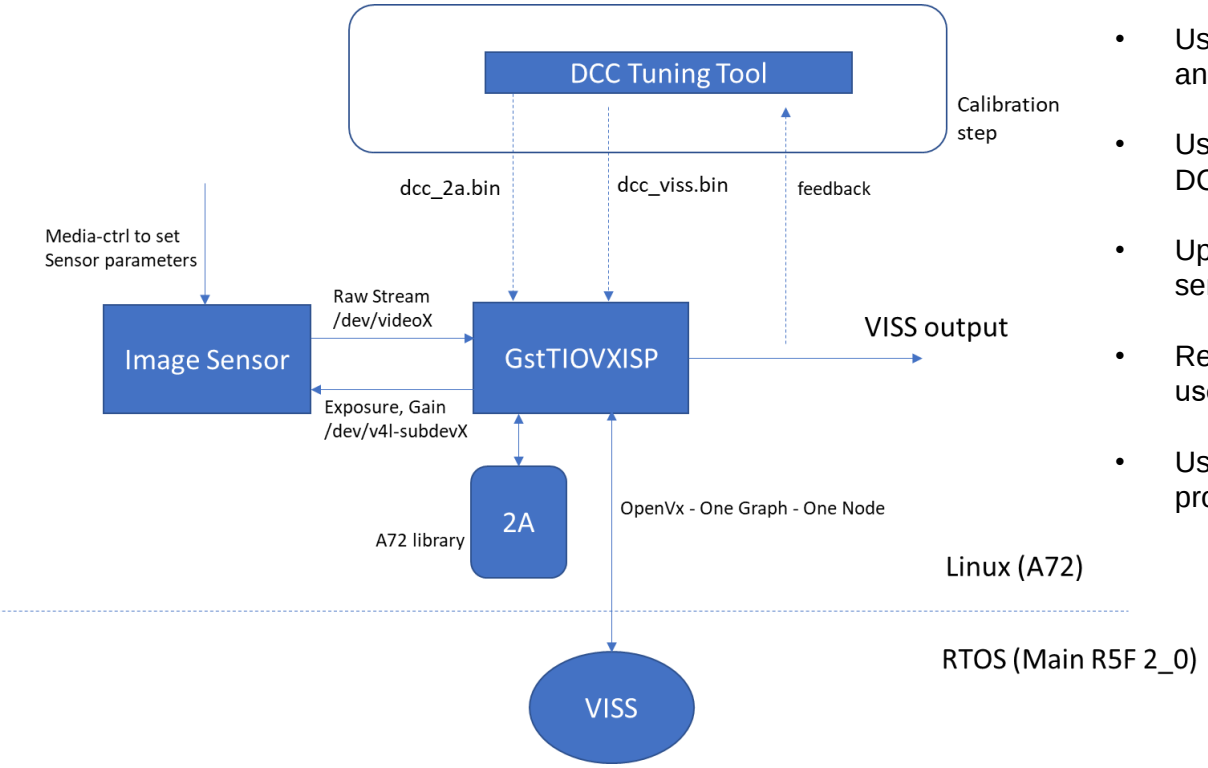
Edge AI SDK | Multichannel support



Features

- Custom Mux and Demux elements batches all inputs and outputs respectively
- Each GSTTIOVX element capable of handling batch of inputs using OpenVx replicate node feature
- A GstTIOVXMux element can be used while connecting a chain of GstTIOVX elements and process in batch
- A GstTIOVXDeMux element is required when the chain encounters a non TIVOX element
- This is mainly done to optimize performance and reduce the number of interrupts from each element to host core

Edge AI SDK | Using on-chip ISP



- Bring up the new sensor as v4l2 compliant driver
- Use standard media-ctrl APIs to set sensor parameters and dump raw output
- Use DCC tuning tool to tune the sensor and generate DCC binaries for 2A, VISS, LDC
- Update the GstTIOVXISP element to work with new sensor type, mostly adding the return path to sensor
- Replace 2A library on A72 with custom implementation, use wrapper API's to interface with ISP
- Use V4l2-subdev enumerated for each sensor to program back exposure and gain from 2A result

Current 8.1 SDK supports RpiV2 camera in 8-bit companded mode

Support will be extended to full 10-bit format in 8.2 SDK

FPD Link based IMX390 camera bringup in progress

Edge AI SDK | Performance

Model	FPS	Total time (ms)	Inference time (ms)	A72 Load (%)	DDR Read BW (MB/s)	DDR Write BW (MB/s)	DDR Total BW (MB/s)
ONR-CL-6150-mobileNetV2-1p4-qat	30.39	33.16	3.02	11.77	1778	842	2620
TFL-CL-0000-mobileNetV1-mlperf	30.30	33.15	2.00	9.94	1669	840	2509
TFL-OD-2020-ssdLite-mobDet-DSP-coco-320x320	30.37	33.16	5.01	14.14	1843	913	2756
TVM-CL-3410-gluoncv-mxnet-mobv2	30.31	33.16	2.00	12.34	1695	828	2523

Model	C71 Load (%)	C66_1 Load (%)	C66_2 Load (%)	MCU2_0 Load (%)	MCU2_1 Load (%)	MSC_0 (%)	MSC_1 (%)	VISS (%)	NF (%)	LDC (%)	SDE (%)	DOF (%)
ONR-CL-6150-mobileNetV2-1p4-qat	8	46	34	9	1	30.17	0	11.26	0	0	0	0
TFL-CL-0000-mobileNetV1-mlperf	5	46	34	9	1	35.99	0	11.11	0	0	0	0
TFL-OD-2020-ssdLite-mobDet-DSP-coco-320x320	16	49	34	8	1	29.92	0	11.14	0	0	0	0
TVM-CL-3410-gluoncv-mxnet-mobv2	7	47	33	8	1	30.47	0	11.12	0	0	0	0

Source : **CSI Camera with VISS (imx219)** Capture
Framerate : **30 fps** Resolution : **1080p** format : **SRGGB8**

Edge AI SDK | Roadmap

Sensor & Hardware \ Demo	Image Classification*	Semantic Segmentation*	Object Detection*	Single Input, Multiple Inference*	Multiple Input, Multiple Inference*	HW Decode + Inference*	ISP + Inference*	Inference + HW Encode*	Multi-channel ISP Process	Multi-channel ISP + GPU + Inference*
TDA4VM SK										
MMCSD/ USB/ PCIe (File Based)										
Ethernet (RTSP Streaming)										
USB Camera (C270/C920/C922)										
YUV Sensor (OV5640)										
Raw Sensor (Rpi v2 IMX219)										
Multi channel with FPDLink (IMX390 + UB953 + UB960)										
DSS (DP, HDMI)										
HW Acceleration										
VPAC: MSC										
VPAC: ISP + LDC										
Accelerated Deep Learning										
DSP Accelerated Pre Processing										
DSP Accelerated Post Processing										
HW Decode										
HW Encode										
GPU										

Legends: Available Now; Upcoming

* Integrated Demos with Capture + Pre-Process + Inference + Post Process + Display

Edge AI SDK | Demo

Call to action

- Get the SK-TDA4VM Starter Kit: https://www.ti.com/tool/SK-TDA4VM

- Get free support on www.e2e.ti.com

Build your Edge AI applicaton

- Use free Edge AI Cloud tool today while you wait for the Starter Kit www.ti.com/edgeaicloud