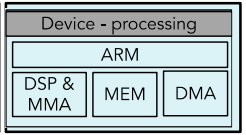
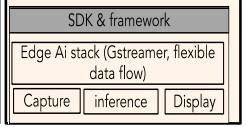
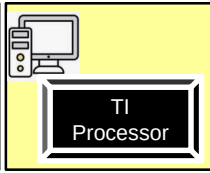
Analytics software Product

March-01-2024

Analytics software Product

Category		Tools/software	
Model Training & associated tools		**Model optimization tools (training env)** (Quantization aware training, sparsity, model surgery)	**Model zoo, Model training** (Object Detection, Human Pose estimation, 6D pose estimation, Depth, 3D OD,..)
Inference Tools		**Model compilation tools** Model import, Post training Quantization, Network Compiler	
Inference Runtime		**Model inference software** TFLite-RT, ONNX-RT, NEO AI DLR, TIDL-RT	
Edge AI, robotics SDK		**Edge AI SDK** End-to-end AI pipeline with camera, inference and display **Robotics SDK** ROS2/DDS, Robotics examples/use case	
Integrated Environment/ Tools		**Model Maker, Edge AI Studio : Model Composer, Model Analyzer**	

https://github.com/TexasInstruments/edgeai

Texas Instruments

Devices & Environment

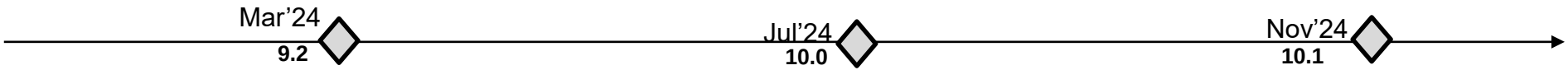
SOC name	MMA Version	ISA Family	Vector Width	TOPS	Availability from SDK
J721E	MMA1	C7100	512 (64B)	8	All
J721S2/AM68A	MMA2	C7120	512 (64B)	8	All
J784S4/AM69A	MMA2p1	C7120	512 (64B)	32	All
AM62A	MMA2-256	C7504	256 (32B)	2	All
AM67A/TDA4AEN (dual MMAs)	2x MMA2-256	C7524	256 (32B)	4	9.2

Module	Feature	Now	9.0 (Jul'23)	9.1 (Nov'23)	9.2 (Mar'24)	10.0 (Jul'24)	10.1 (Nov'24)	10.2 (Mar'25)	Comments
X86 (Tools)	Ubuntu version	18.04	22.04	-	-	>= 22.04	-	-	
	Python version	3.6	3.10	-	-	>= 3.10	-	-	
	Docker environment	Y	-	-	-	-	-	-	
Target board	OS (cortex-A)*	Linux	-	-	-	-	-	-	QNX - Limited features of edge AI offering (TIDL-RT)
	OS (C7x)	FreeRTOS	-	-	-	-	-	-	
	Distro	Yocto (Dunfell)	Yocto (Kirkstone)	-	-	>=Yocto (Kirkstone)	-	-	
	Docker environment (for Robotics)	18.04, 20.04	18.04, 20.04	22.04	-	>= 22.04	-	-	

Model Inference & associated Tools | Features (Available)

Devices	AM62A, TDA4VM, AM68A/TDA4AL/TDA4VL, AM69A/TDA4VH/TDA4VH
Accelerated Layers	Conv, deconv, batchnorm, Max Pool, average pool, Eltwise (Add, mul), Gemm, inner product, softmax, concat, split, flatten, argmax, reshape, transpose, upsample, depthtospace, Sigmoid, Hardsigmoid, Tanh, ELU, Relu, Prelu, Reducemean, scatter, gather, crop, flatten, pad, argmax, shuffle, data conversion
Models	**CNN based architectures** **Object detections**: SSD, Yolo, efficientdet, retinanet **Classification**: mobilenets, resnets, regnets, efficientnet, inceptionnet, resnext **Semantic Segmentation**: deeplabv3, aspp **6D object pose**: yolo based **Key point detection** (human pose estimation): yolo based, **Depth**: Midas, fast depth
Tools/ Frameworks	**Runtime**: ONNX-RT (1.14), TFLite-RT (2.8), TVM-Neo AI DLR (10.1), TIDL-RT, ONNX opset-18 **Application framework**: Gstreamer, ROS, **Middleware**: openVX **Development Environment** **Target OS**: Linux (Yocto - Kirkstone), Free RTOS **X86**: Ubuntu22.04, Python 3.10, Docker environment **Model optimization:** Post training quantization – **PTQ**
Others	Externally quantized models (QDQ format), Mixed precision, GPU accelerated host emulation inference Low Latency and high throughput (batch mode) Multi priority DNN handling

Model Inference & associated Tools | Features (Roadmap)



Devices	AM62A, TDA4VM, AM68A/TDA4AL/TDA4VL, AM69A/TDA4VH/TDA4VH, AM67A/TDA4AEN		
Accelerated Layers	**Transformer modules** Matmul, broadcasted (matmul, eltwise), 2D softmax, Layernorm, Patch embedding, Patch Merging, GeLU, SiLU scatterND	Additional optimization (latency and accuracy) of transformer modules Rotated NMS (CenterNet) Grid sample Multi C7x perf optimization	Deformable attention
Models	Deit, Swin, convnext DETR, Yolov8	Segformer Simple BEV, DETR-3D	BEVFormer/BEVFormer-v2, StreamPETR , Deformable DETR
Tools/FW	TFLite-RT ver upgrade (2.12)	Performance visualization	
Others		Partial batch mode Robustness improvement PTQ/QAT simplification	TIDL module safety Robustness improvement Memory Optimization

TI Confidential – NDA Restrictions

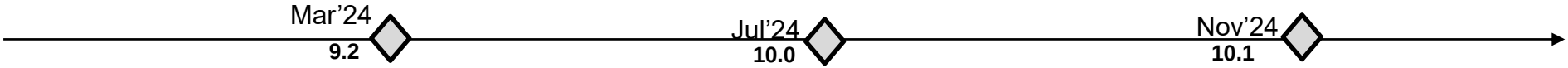
Texas Instruments

Model Training Tools | Features (Available)

Devices	AM62A, TDA4VM, AM68A/TDA4AL/TDA4VL, AM69A/TDA4VH/TDA4VH
Model Optimization Toolkit	Pytorch based Toolkits: Quantization Aware Training (QAT) Tool kit, Model Surgery Tool kit Model Pruning Toolkit
Model Zoo	**Collection of 60+ pre trained models/device across different category** Image Classification Models, Object Detection Models, Face Detection Models , Semantic Segmentation Models, Depth Estimation Models, 3D Object Detection Models, 6D Pose Estimation Models
Model Maker	Command line Integrated environment for training & compilation for different tasks such as Image Classification, Object Detection, Semantic Segmentation, Keypoint Detection Backend tool for Model Composer

Supported workflows documented here

Model Training Tools | Features (Roadmap)

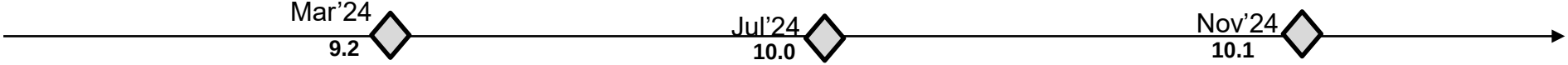


Devices	AM62A, TDA4VM, AM68A/TDA4AL/TDA4VL, AM69A/TDA4VH/TDA4VH, AM67A/TDA4AEN		
Model Optimization Toolkit	Model Surgery enhancement for Transformer networks	QAT enhancement for Transformer networks	
Model Zoo	~70 Models/SOC	~100 Models/SOC Transformer based sensor fusion, BEV Models	
Model Maker			Transformer/BEV Models

Edge AI, Robotics SDK | Features (Available)

Devices	AM62A, TDA4VM, AM68A, AM69A
Edge AI SDK	Linux based development environment for Edge AI applications Hardware accelerated open source inference: ONNX-Runtime, TFLite—Runtime, Neo AI DLR Gstreamer based edge AI application development capabilities with multi camera, multi channel Reference Application Type: Python, C++, OpTIFlow, OOB GUI Demo Smart Cameras: OV5640 Raw Cameras: IMX219, IMX390, OV2312 Fusion boards: Fusion1 V4L2 interface for Linux, accelerated multi media codecs Device Agent for Edge AI Studio
Edge AI application examples	Retail Checkout, Human pose, License plate recognition, 6D pose estimation, bar code reader, defect detection, pick& place
Robotics SDK	ROS2-Humble (Ubuntu 22.04) IMX390 CSI Camera and mmWave Radar ROS Driver 3D Perception with Object Detect + Stereo Disparity 6D Pose Estimation DNN Chain (for robot arm bin picking)

Edge AI, Robotics SDK | Features (Roadmap)



Devices	AM62A, TDA4VM, AM68A, AM69A, AM67A		
Edge AI SDK	openVX based edge AI application development capabilities with multi camera, multi channel Raw camera: Ox05B1S 8 camera use case on AM68A and AM69A with Fusion1	OpenMax interface for QNX 12 camera use case on AM69A with fusion mini and Fusion1	Performance optimizations
Edge AI application examples		Multi Camera-based BEV DNN	Transformer architecture based sensor fusion
Robotics SDK	Camera + Radar Fusion: 3D object detection with IMX219 + IWR6843ISK Human Pose Estimation DNN Chain	Multi Camera-based BEV DNN	Transformer architecture based sensor fusion demo

Edge AI integrated Tools

Model Composer

Release	Timeline (Model Maker)	Features/Tasks	Devices
9.1	Available	• Object Detection, Classification, Semantic Segmentation	AM62, AM62A, TDA4VM, AM68A, AM69A
10.1	30-Nov-24	• Object Detection, Classification, Semantic Segmentation • Docker image instead of source • New models (transformer based) – 1 for each task	AM62, AM62A, TDA4VM, AM68A, AM69A
11.1	30-Nov-25	• Object Detection, Classification, Semantic Segmentation, Key point detection • Docker image • Potentially new models for new task and additional task	AM62, AM62A, TDA4VM, AM68A, AM69A
TBD	Aug'24	• Tiny ML time series models (AFD and Motor diag) • Docker image instead of source	TMS320F28P55x

Model Analyzer

Release	Timeline (Model Maker)	Features/Tasks	Devices
9.2	31-Mar-24	• Object Detection, Classification, Semantic Segmentation • New models	AM62A, TDA4VM, AM68A, AM69A
10.2	31-Mar-25	• Object Detection, Classification, Semantic Segmentation • New models	AM62A, TDA4VM, AM68A, AM69A

Change Log

Date	Changes
05-Oct-23	**Descope** • Support of Fusion2 and Fusion2 mini board **Reduced scope** • [9.1] Reduced supported number of models in model zoo from 100 to 60 **Moved from 9.1 to 9.2** • Support of advanced model (SWIN, DETR) of Vision Transformer in TIDL • Robustness and optimization of Graph partition b/w CPU and C7x • Multi C7x power optimization • Support of key point based task in Edge AI Studio **Moved from 9.0 to 9.1** • H.264/H.265 codec integration on AM69A • Multi-channel optimization using ti-mux/ti-demux
23-Oct-23	**Moved from 9.1 to 9.2** • C75x(AM62A) HardSigmoid, TanH, ELU • C75x(AM62A) Sparsity • Rotated NMS for centerpoint style network • Transformer layers - mainly SWIN Layer Norm, Patch embedding, Patch Merging, window shift
01-Mar-24	Restructured the content

THANK YOU