赛灵思加速人工智能应用

陈志勇
高级技术市场经理
安富利电子科技
2019年8月18日， 厦门

AVNET
Reach Further™

About Avnet

- Founded in 1921
- Headquartered in Phoenix, Arizona
- AVT listed on the NYSE since 1960
- AVT listed on NASDAQ since 2018
- #128 on the FORTUNE 500 (U.S.) in 2018

Awards

- Top 10 for 2018 Gartner Supply Chain Top 25 - High-Tech industry category
- World's Most Ethical Company by Ethisphere Institute from 2014-2018

This is the new Avnet

We design, make, supply and deliver technology solutions.
We work with customers of every size, in every corner of the world.

15,000+
Employees worldwide

2,500+
Engineers around the world

1M+
Engineering community members

2.1M
Customers in 140+ countries

1,400+
Technology suppliers WW

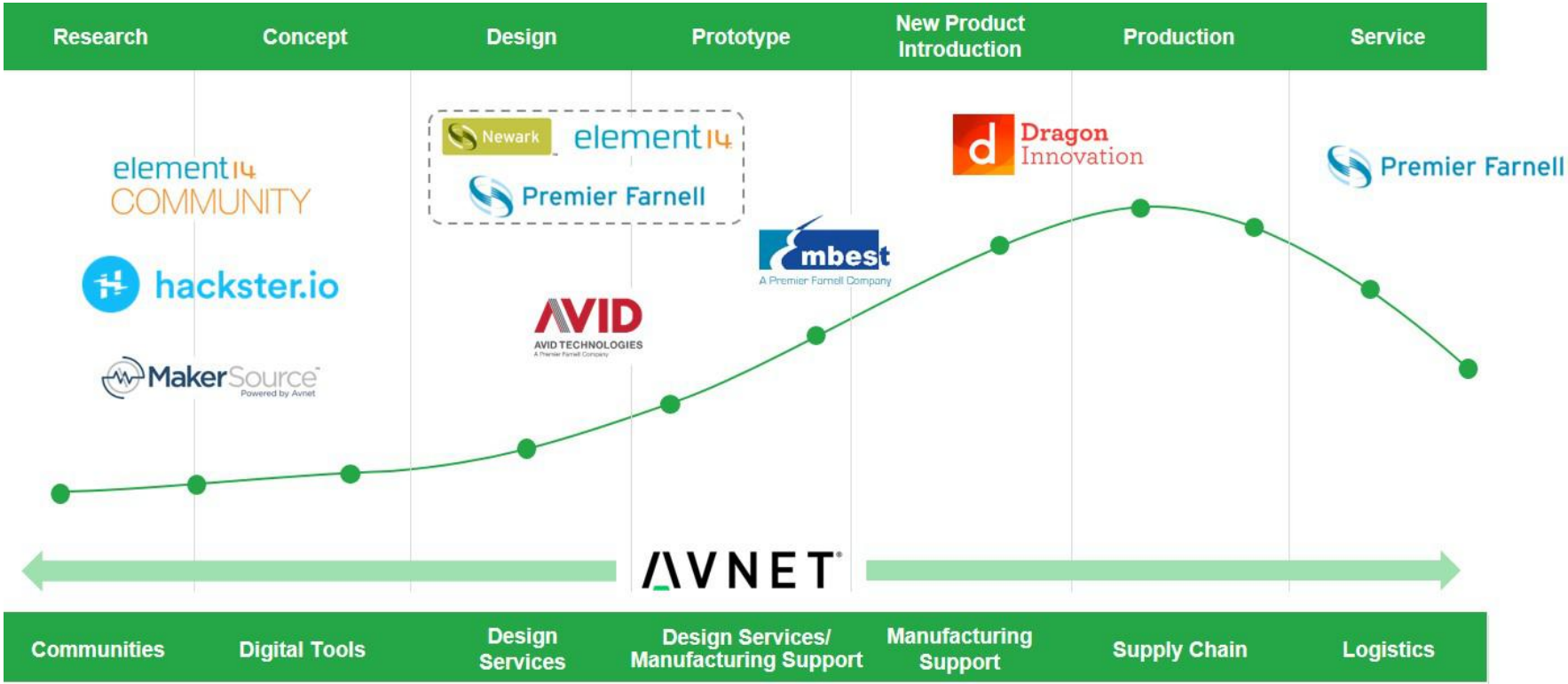
122B
Units shipped annually

125
Locations across the world

We guide today's ideas into tomorrow's technology.

AVNET

The Avnet ecosystem

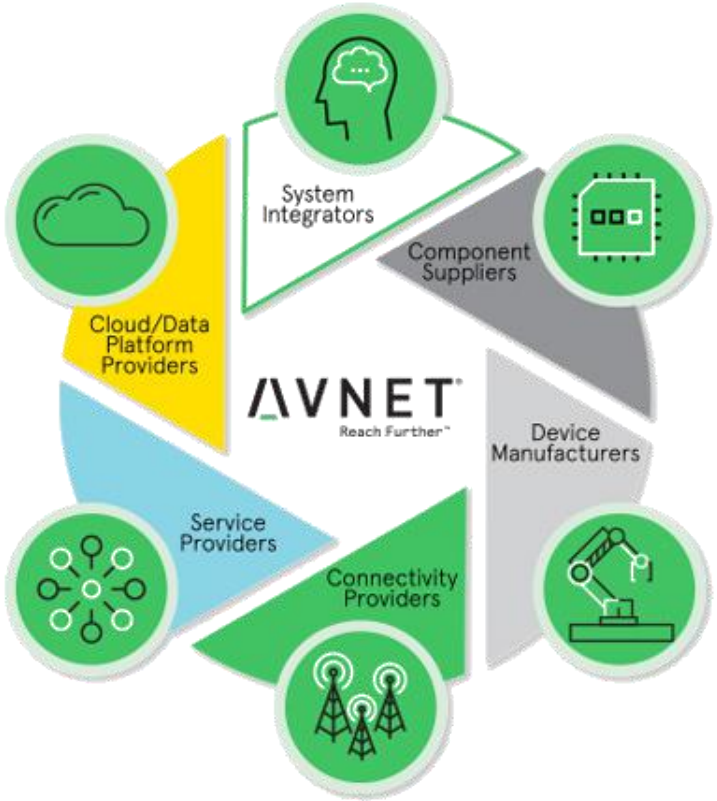


AVNET

Your one partner for IoT

A Key Enabler for IoT Development

We simplify complexities by connecting reliable partners to solve your challenge.

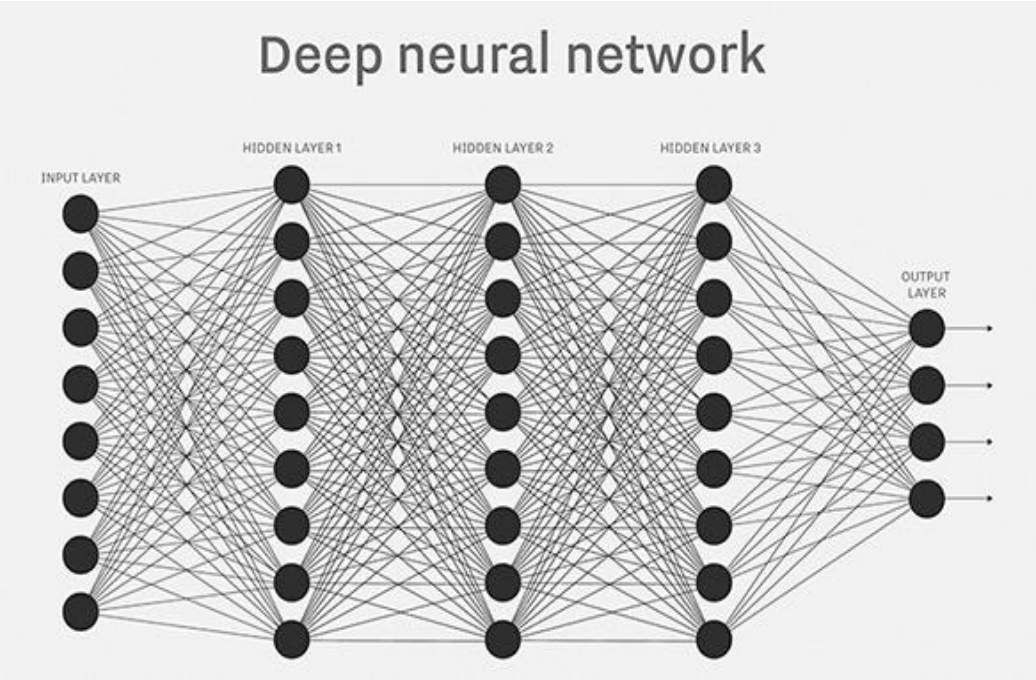


AVNET

Xilinx Machine Learning
Edge to Cloud

AVNET

Deep Learning



5-Layer Neural Network

Why now?

- New processors making DNN training feasible (Ops/$)
- Huge amounts of training data

Faster/Better Results

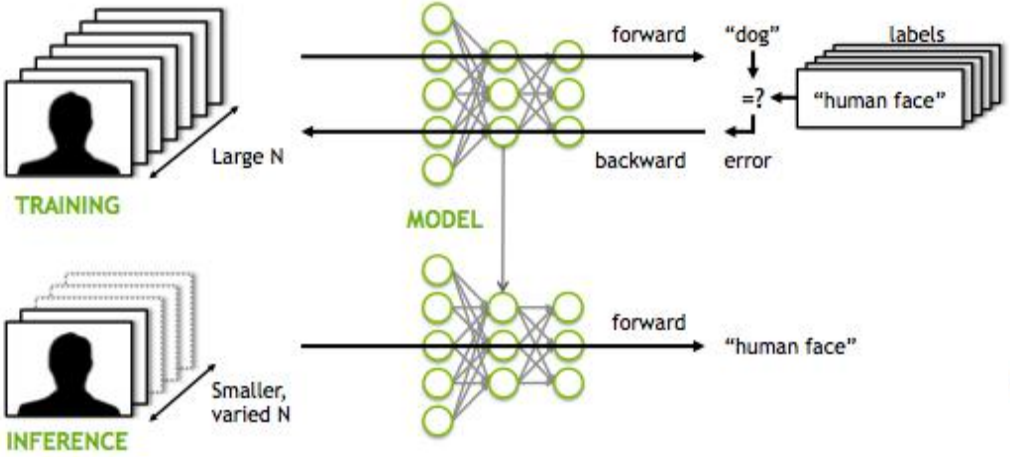
- Caused explosion in AI Research
- More Applications
- More Startups
- More Innovation
- More Acquisitions . . .

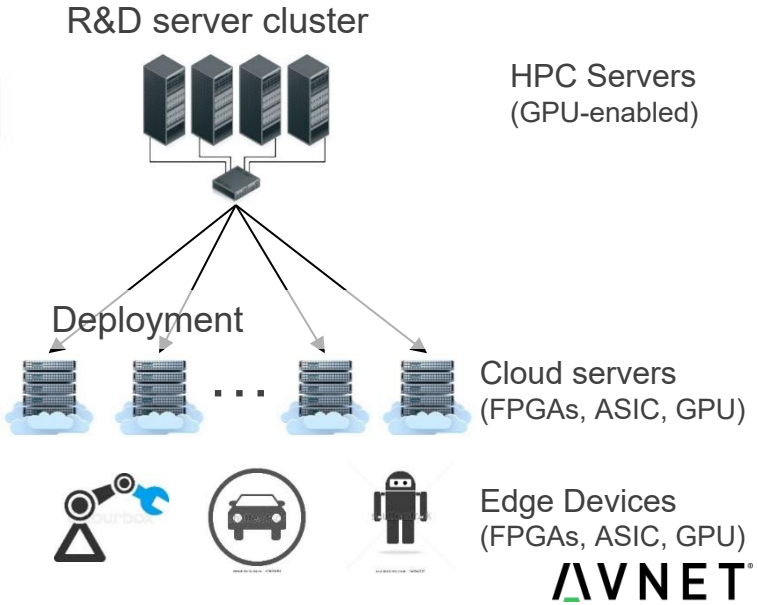
AVNET

Deep Learning: Training vs. Inference

Training: Process for machine to “learn” and optimize a model from data

Inference: Using trained model to predict/estimate outcomes from new observations





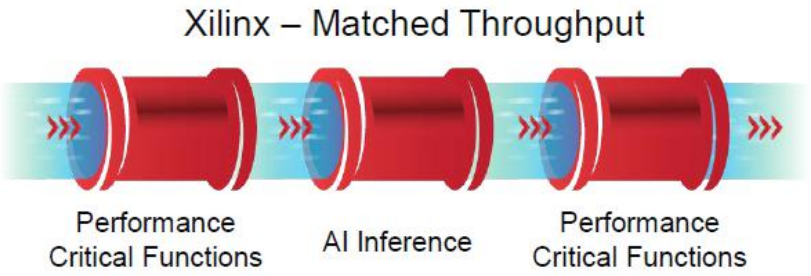
Delivering Adaptable ML Compute Acceleration

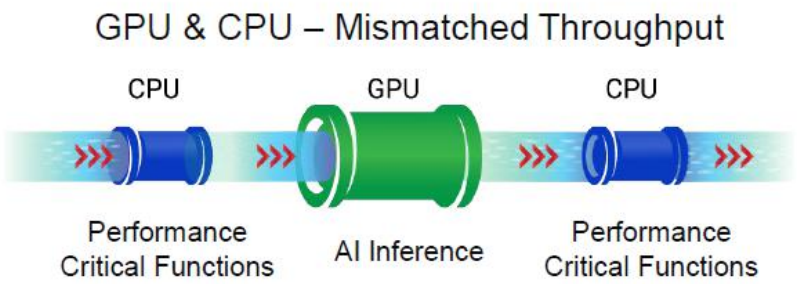
	CPU (Sequential)	GPU (Parallel)	FPGA / SoC / ACAP	Custom ASIC
				
SW Programmable	✓	✓	✓	✓
HW Adaptable	–	–	✓	–
Workload Flexibility	✓	✓	✓	–
Throughput vs. Latency	–	–	✓	✓
Device / Power Efficiency	–	–	✓	✓

AVNET

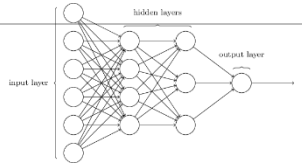
Accelerate the Whole Application Around AI Inference

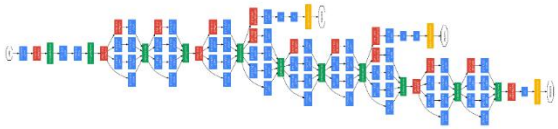
Application	Typical AI Networks	Non-AI Workloads
Smart Retail / Surveillance	CNN classification, detection, segmentation	multi-channel video decode
Speech Recognition	LSTM and BART	Speech clean-up, database lookup
Recommendation Engines	MLP	Keyword pre-post processing
Anomaly Detection	Random Forest	Smart NIC functions
Financial Tech	LSTM	Monte Carlo and other risk analysis models

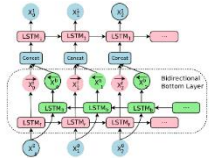




Deep Learning Models





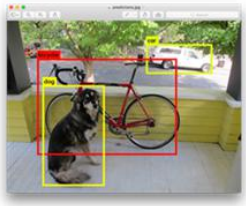


Multi-Layer Perceptron	Convolutional Neural Network	Recurrent Neural Network
• Classification • Universal Function Approximator • Autoencoder	• Feature Extraction • Object Detection • Image Segmentation	• Sequence and Temporal Data • Speech to Text • Language Translation

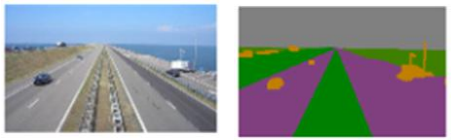
Classification

“Dog”

Object Detection



Segmentation

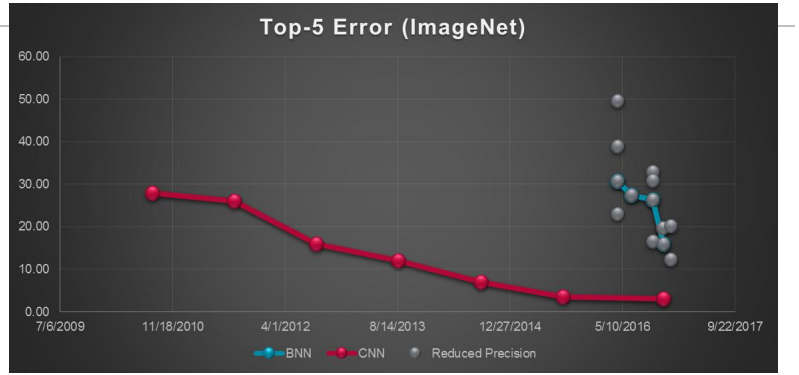


AVNET

Latest Inference Algorithmic Research

Optimize Compute with Reduced Precision CNNs & BNNs

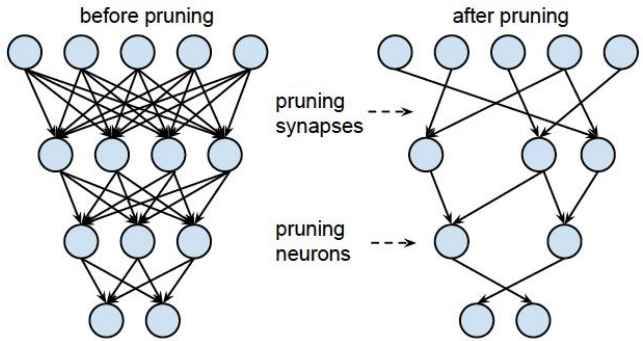
- 8 bit solution loses no significant accuracy
- BNNs are improving rapidly
- Custom Floating point provide significantly more compute density



Not only weights, but also sparsity pattern encodes information

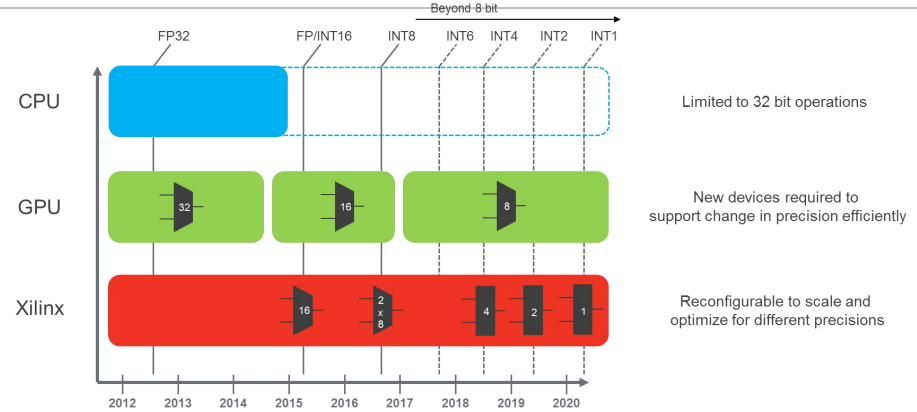
Model Pruning for higher effective performance: LSTM

- TIMIT 3-hour dataset – 20x
- LibreSpeech 100-hour dataset – 10x
- CustomerS1000-hour dataset – 20x
- CustomerS 3000-hour dataset – 5x

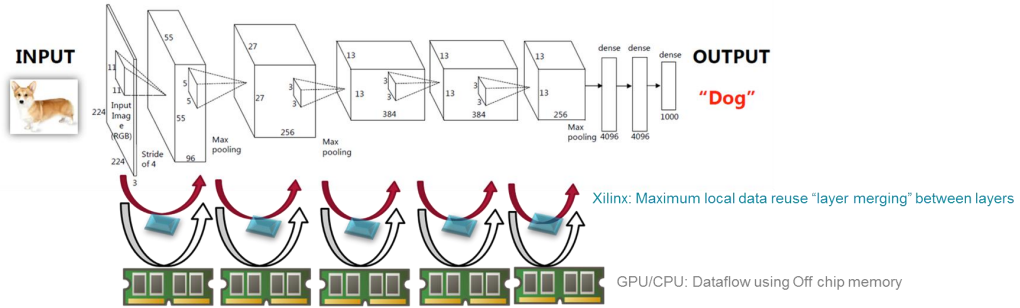


Xilinx Features for Implementing Efficient Inference Engines

Flexible Architecture for Any Precision

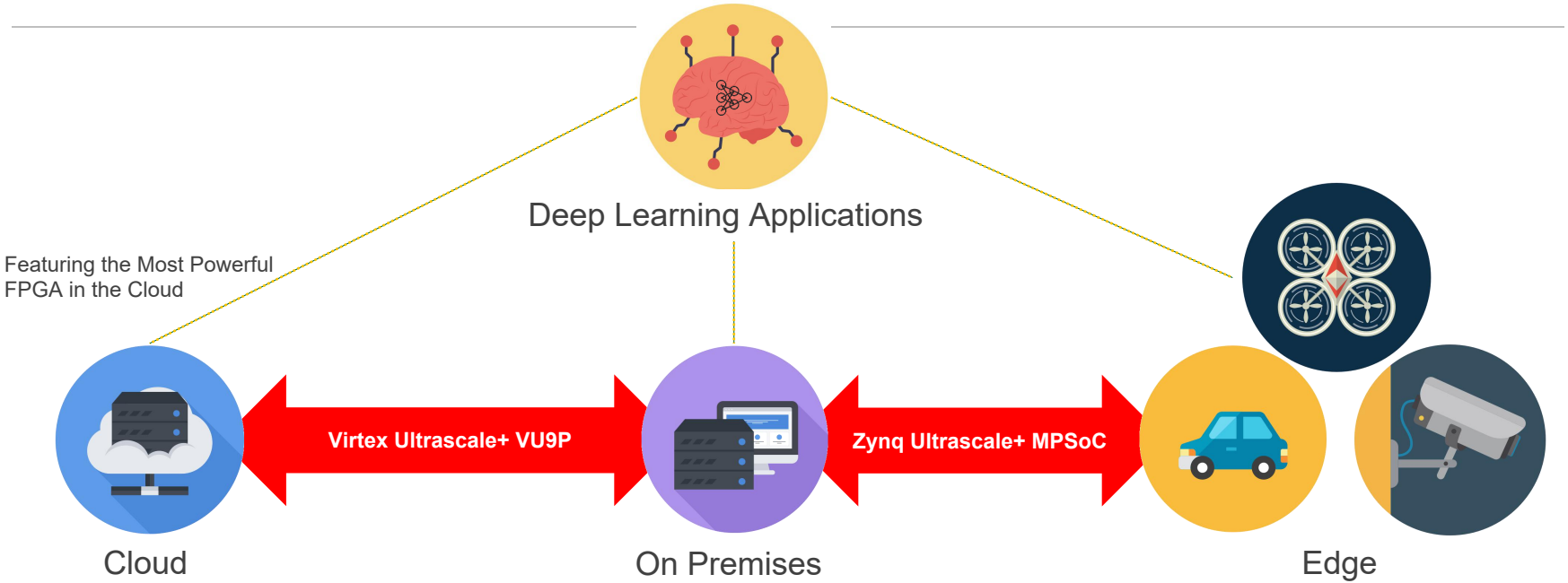


Flexible On-chip Memory for low latency



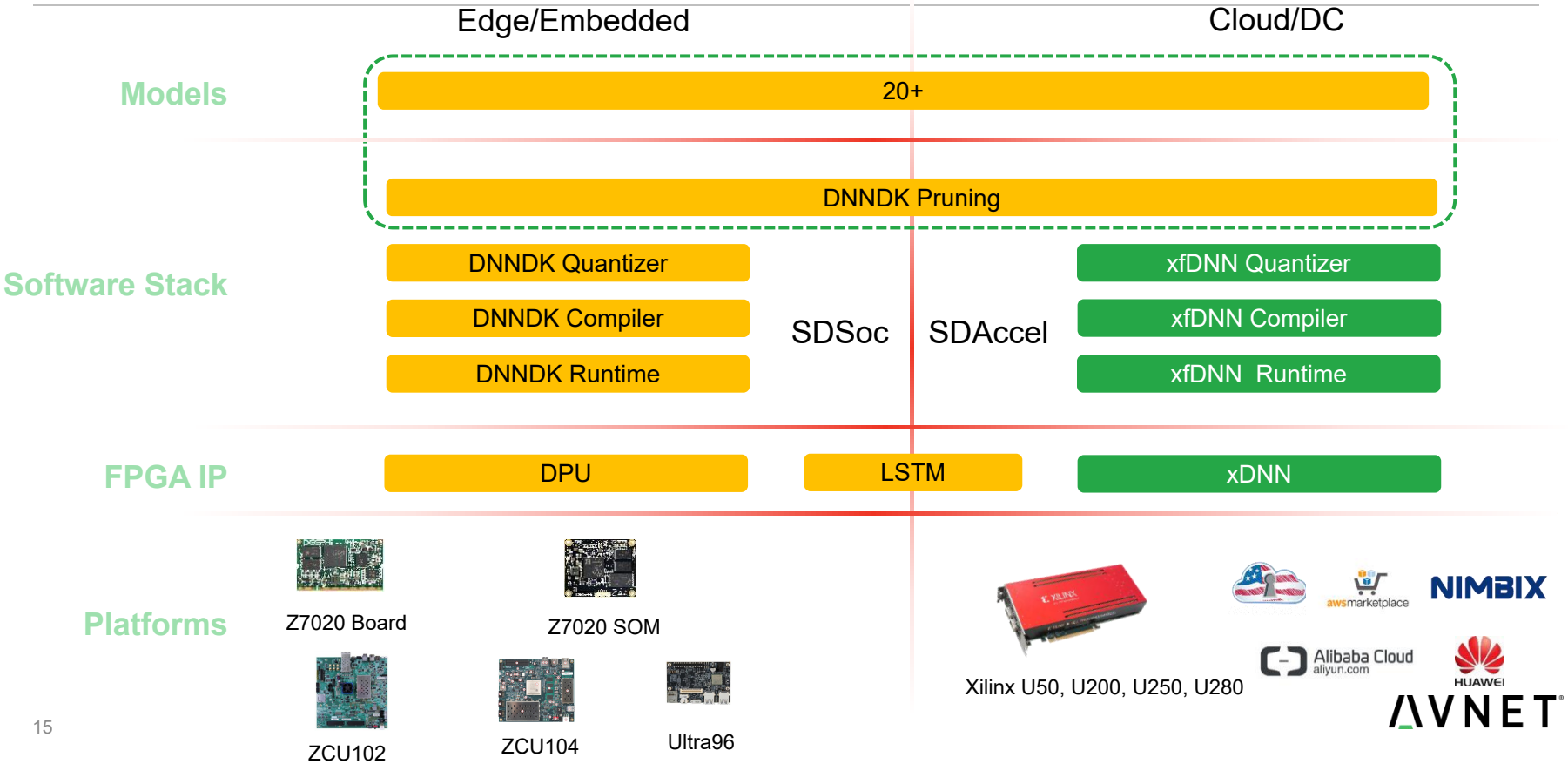
AVNET

Xilinx ML Solution from Edge/Embedded to Cloud/DC



AVNET

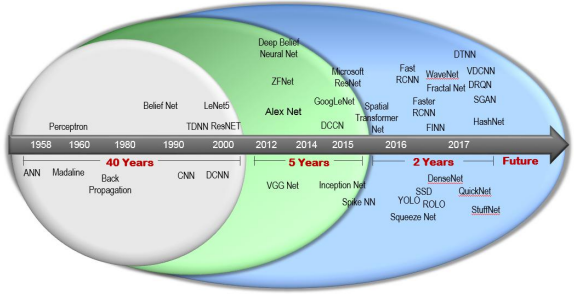
Xilinx ML Solution from Edge/Embedded to Cloud/DC

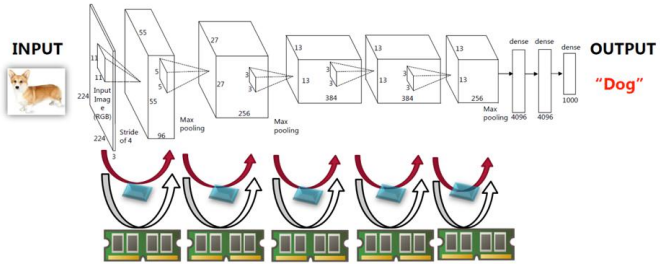


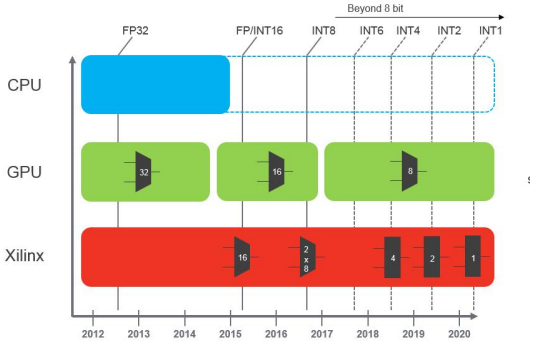
Xilinx ML for Edge/Embedded

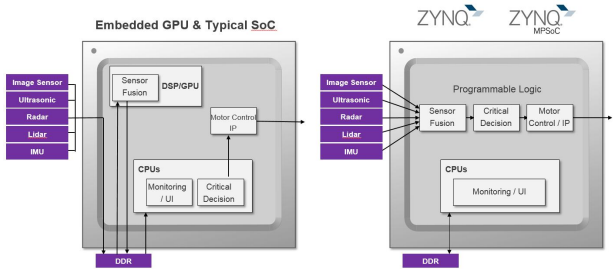
Xilinx Value Proposition for Edge/Embedded ML

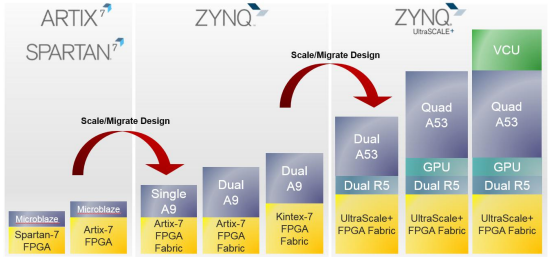
Xilinx offers the optimal tradeoff among latency, power, cost, flexibility, scalability & time-to-market for Edge/Embedded ML







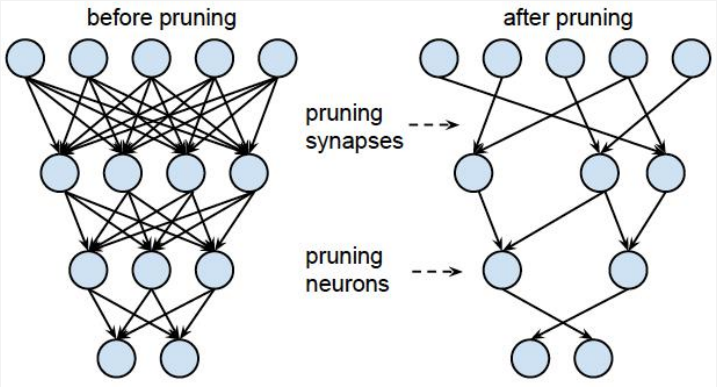




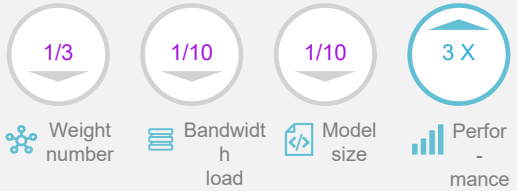
AVNET

Xilinx Pruning Overview

Deep compression Makes algorithm smaller and lighter



Highlight



Compression efficiency

Deep Compression Tool can achieve significant compression on **CNN** and **RNN**

Accuracy

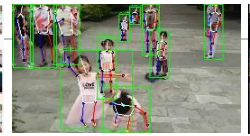
Algorithm can be **compressed 7 times without losing accuracy** under SSD object detection framework

AVNET

Xilinx Solution Stack for Edge/Embedded ML

Models


Face detection


Pose estimation


Video analytics


Lane detection


Object detection


Segmentation

Framework

Caffe

Darknet

TensorFlow

Tools

DEEPHi DNNDK

Compressio
Pruning
Quantizatio n

Compilati
Compiler
Assembler

Runtime
Core API
Loader
Driver
Profiler

ML HW Platforms


Z7020 Board


Z7020 SOM


ZCU102


ZCU104


Ultra96

AVNET

DPU Utilization

More DSP

Arch	LUTs	Registers	BRAM*	DSP
B512	17951	28280	69.5	97
B800	20617	35065	87	141
B1024	22327	39000	101.5	193
B1152	22796	40276	117.5	193
B1600	26270	50005	123	281
B2304	29592	57549	161.5	385
B3136	33266	69110	203.5	505
B4096	37495	84157	249.5	641

More LUT

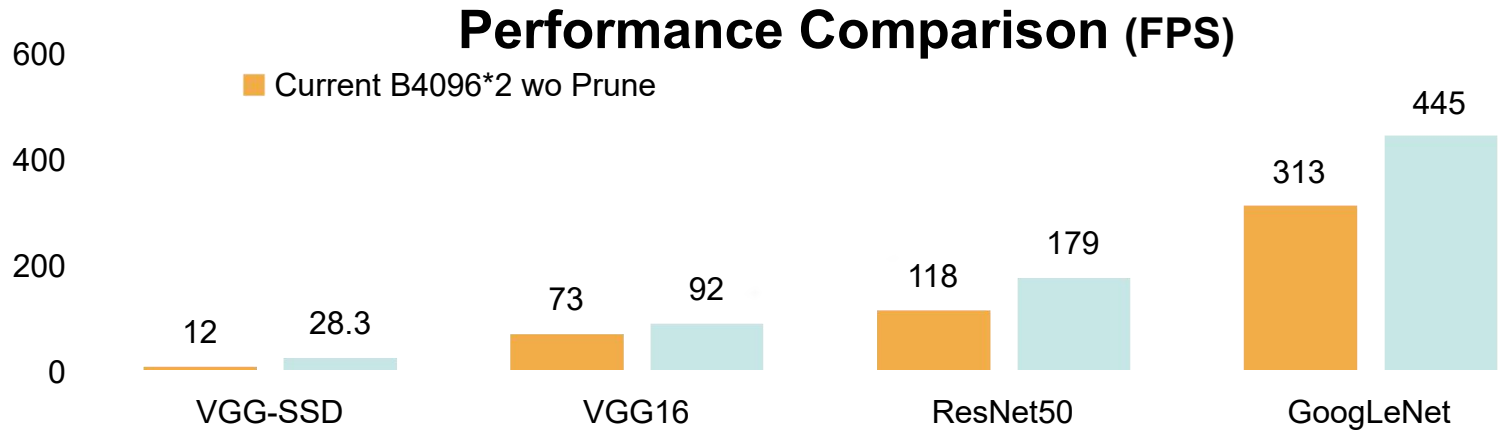
Arch	LUTs	Registers	BRAM*	DSP
B512	20759	33572	69.5	66
B1024	29155	49823	101.5	130
B1152	30043	49588	117.5	146
B1600	33130	60739	123	202
B2304	37055	72850	161.5	290
B3136	41714	86132	203.5	394
B4096	44583	99791	249.5	514

DPU provides flexible option depending on costumer's resources and continues to improve

** URAM also can be used by DPU if device supports, every URAM is roughly used as 3.7 BRAM*

AVNET

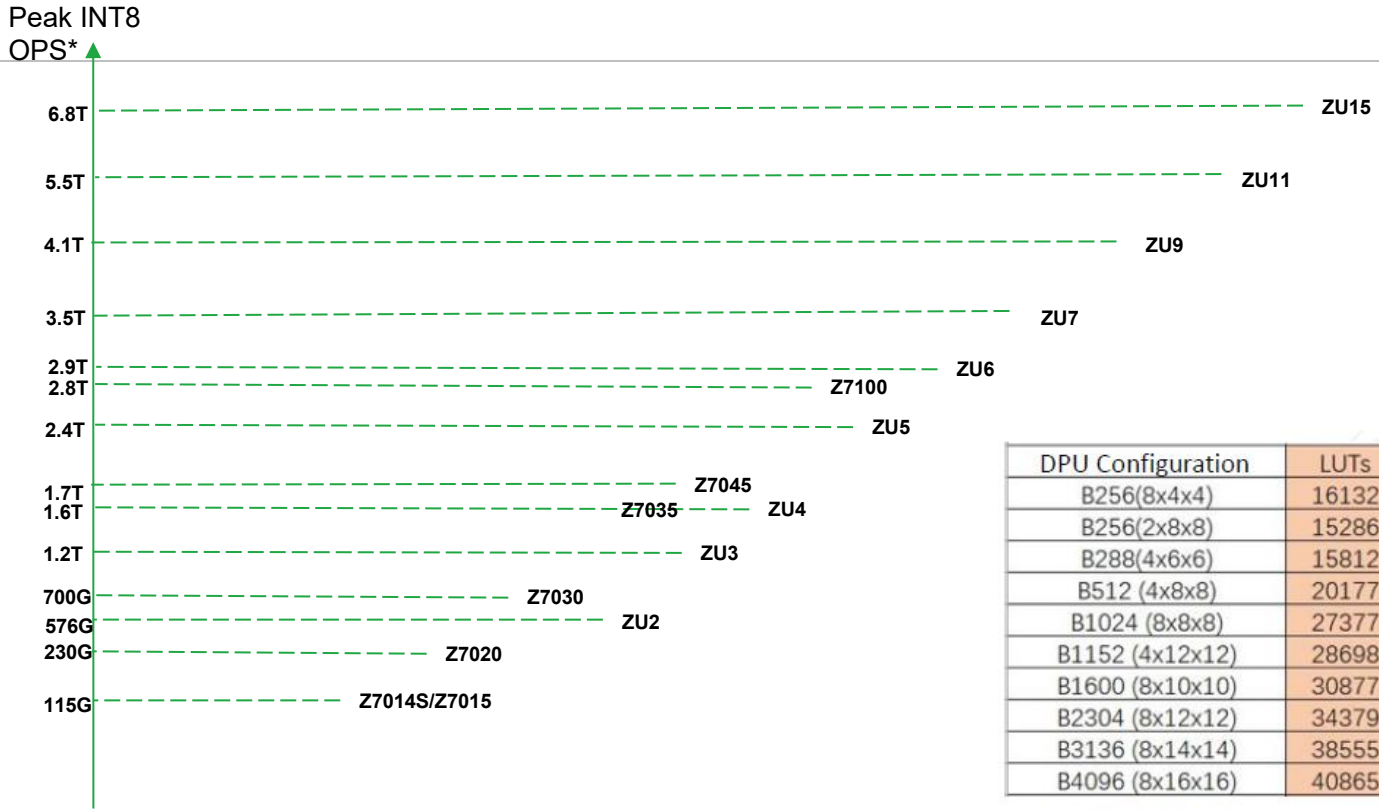
Perf Improvement with DPU



*The FPS of VGG-SSD of end to end performance
*The FPS of VGG16/ResNet50/GoogLeNet is of CONV part (w/o FC layer)

AVNET

DPU Scalability



DPU Configuration	LUTs	Registers	BRAM	DSP
B256(8x4x4)	16132	25064	43	66
B256(2x8x8)	15286	22624	53.5	50
B288(4x6x6)	15812	23689	46	62
B512 (4x8x8)	20177	31782	69.5	98
B1024 (8x8x8)	27377	46241	101.5	194
B1152 (4x12x12)	28698	46906	117.5	194
B1600 (8x10x10)	30877	56267	123	282
B2304 (8x12x12)	34379	67481	161.5	386
B3136 (8x14x14)	38555	79867	203.5	506
B4096 (8x16x16)	40865	92630	249.5	642

* B256/288/512/3136 work in progress

** With heterogenous DPUs*

AVNET

DNNDK Dev Flow

Five Steps with DNNDK

01 Model Compression

02 Model Compilation

03 Programming

04 Hybrid Compilation

05 Execution

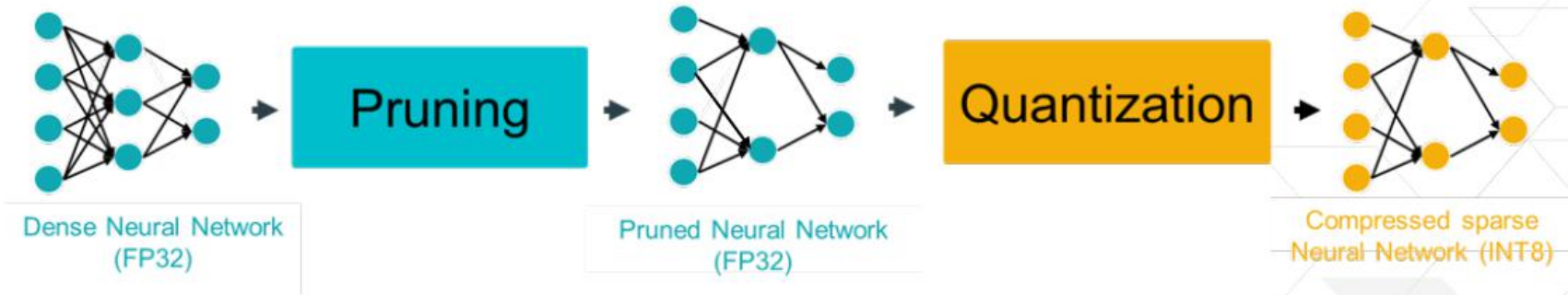
AVNET

第1步：使用Decent 进行模型压缩

decent – Deep compression Tool

decent_q – Quantization Tool

decent_p – Pruning Tool



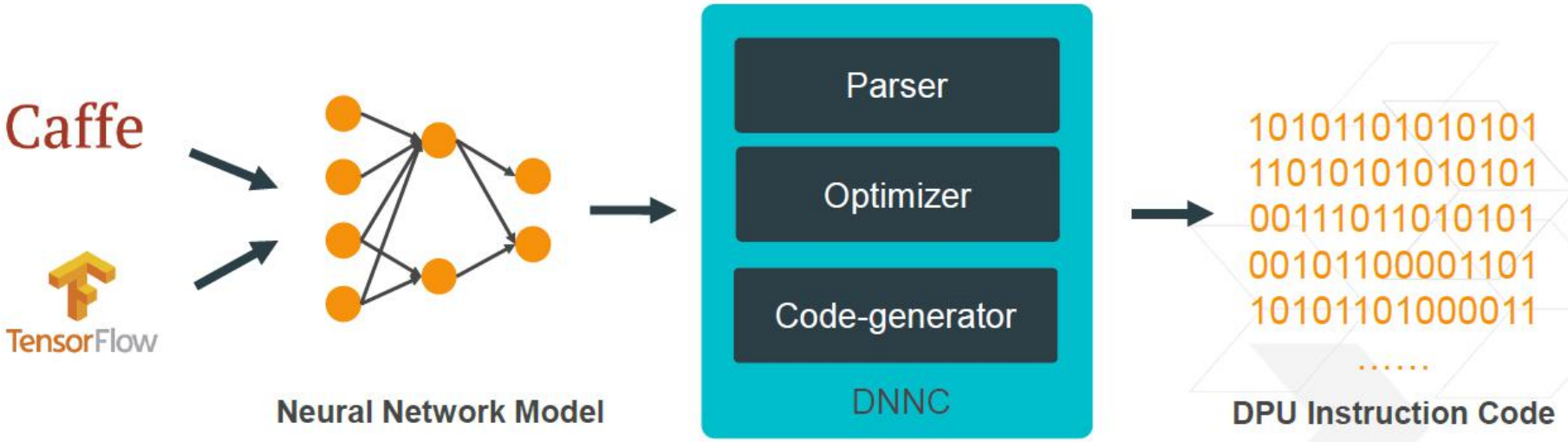
- Consists of two separate tools
  - Quantization Tool
  - Pruning Tool

- Effects
  - Compress model size 5x – 100x
  - Compress running time 1.5x – 10x

- Platform
  - Caffe, Darknet
  - TensorFlow
    - Quantization Tool Beta version
    - Pruning Tool Internal version

AVNET

第2步：使用 DNNC 进行模型编译



Programmable tensor-level DPU instruction set
- Compatible for Caffe/TensorFlow frameworks
- Flexible & scalable for various CNN layers

AVNET

Xilinx AI Developer Resource

- https://www.xilinx.com/products/design-tools/ai-inference/ai-developer-hub.html#edge

Edge AI Tools

Product	Documentation	Tool Download	File Size	MD5 Checksum
DNNDK	DNNDK User Guide (UG1327)	xlnx_dnndk_v3.0_190430.tar.gz	3.2 GB	56115c74eb43f0bdb142788cefa04a6d
	Xilinx AI SDK User Guide (UG1354)			
	Xilinx AI SDK Programming Guide (UG1355)			
DNNDK for SDSoC	DNNDK User Guide for SDSoC (UG1331)	xilinx_dnndk_v2.08_for_sdsoc_190214.tar.gz	667 MB	7f165aff5062497e4bb69b70773c49b1

Edge AI Evaluation Boards

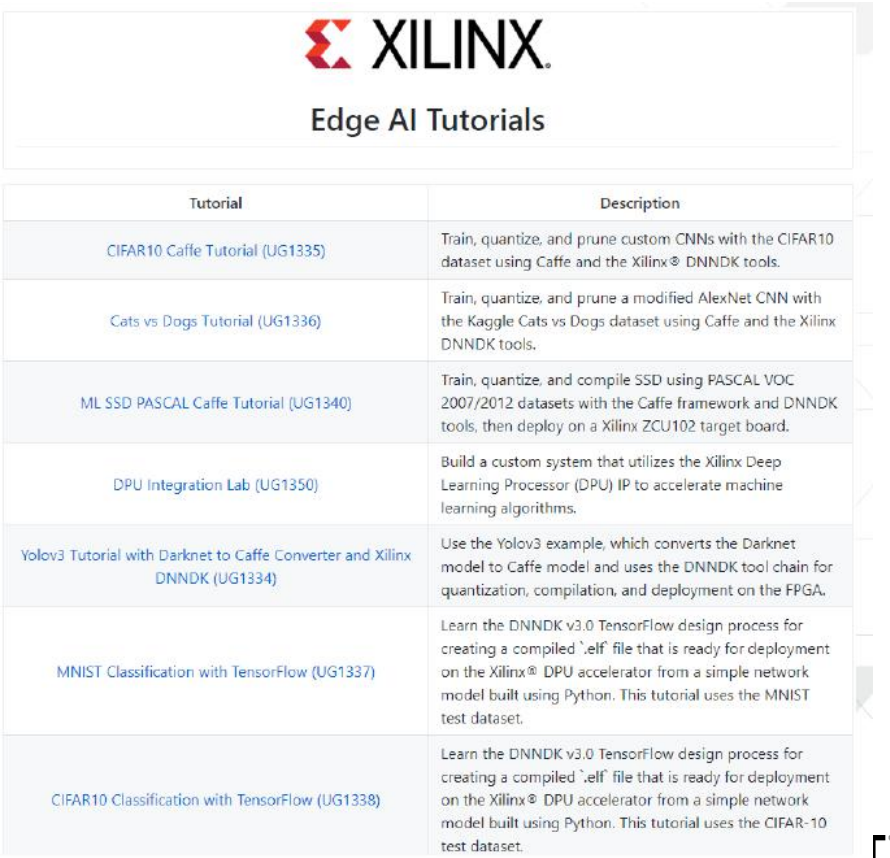
Product	Documentation	Image Download	DNNDK Version	File Size	MD5 Checksum
ZCU102 Kit	ZCU102 User Guide (UG1182)	xilinx-zcu102-prod-dpu1.4-2018.3-desktop-buster-2019-04-24.img.zip	v3.0	657 MB	d49eab4d293d8d1af40fcc369e1c4f53
		2018-12-04-zcu102-desktop-stretch.img.zip	v2.08	571 MB	d0d5faf8ece80b96f5591d09756d5a5d
ZCU104 Kit	ZCU104 User Guide (UG1267)	xilinx-zcu104-prod-dpu1.4-desktop-buster-2019-04-23.img.zip	v3.0	655 MB	503661dd1cc4549a562775034b95d0c8
		2018-12-04-zcu104-desktop-stretch.img.zip	v2.08	571 MB	eda2420c4afbc89efdeea741e0917e25
Avnet Ultra 96	Ultra 96 User Guide	xilinx-ultra96-prod-dpu1.4-desktop-buster-2019-04-23.img.zip	v3.0	652 MB	8beb24afe0af524e946c07e8551b61a4
		xilinx-ultra96-desktop-stretch-2018-12-10.img.zip	v2.08	565 MB	c5d2422063213b4bc4c18a3223c6adc8

Edge AI Targeted Reference Designs (TRD)

Product	Documentation	Image Download	File Size	MD5 Checksum
DPU TRD	DPU IP Product Guide (PG338)	zcu102-dpu-trd-2018-2-190322.zip	468 MB	3101cc91a5d121a8969613367b890b77

Xilinx Edge AI Tutorial

- https://github.com/Xilinx/Edge-AI-Platform-Tutorials



XILINX

Edge AI Tutorials

Tutorial	Description
CIFAR10 Caffe Tutorial (UG1335)	Train, quantize, and prune custom CNNs with the CIFAR10 dataset using Caffe and the Xilinx® DNNDK tools.
Cats vs Dogs Tutorial (UG1336)	Train, quantize, and prune a modified AlexNet CNN with the Kaggle Cats vs Dogs dataset using Caffe and the Xilinx DNNDK tools.
ML SSD PASCAL Caffe Tutorial (UG1340)	Train, quantize, and compile SSD using PASCAL VOC 2007/2012 datasets with the Caffe framework and DNNDK tools, then deploy on a Xilinx ZCU102 target board.
DPU Integration Lab (UG1350)	Build a custom system that utilizes the Xilinx Deep Learning Processor (DPU) IP to accelerate machine learning algorithms.
Yolov3 Tutorial with Darknet to Caffe Converter and Xilinx DNNDK (UG1334)	Use the Yolov3 example, which converts the Darknet model to Caffe model and uses the DNNDK tool chain for quantization, compilation, and deployment on the FPGA.
MNIST Classification with TensorFlow (UG1337)	Learn the DNNDK v3.0 TensorFlow design process for creating a compiled `.elf` file that is ready for deployment on the Xilinx® DPU accelerator from a simple network model built using Python. This tutorial uses the MNIST test dataset.
CIFAR10 Classification with TensorFlow (UG1338)	Learn the DNNDK v3.0 TensorFlow design process for creating a compiled `.elf` file that is ready for deployment on the Xilinx® DPU accelerator from a simple network model built using Python. This tutorial uses the CIFAR-10 test dataset.

Xilinx AI Model Zoo

- https://www.xilinx.com/bin/public/openDownload?filename=all_models_20190528.zip

Xilinx Model Zoo

ZCU102/ZCU104/Ultra96 performance number were generated with DNNDK v3.0&AI SDK v1.0

Application	Model	Model Download Link	File Size	MD5 Checksum	Backbone	Input Size	OPS per image	Parameters	Framework	Training Set
Image Classification	resnet50	https://www.xilinx.com/bin/public/op filename=resnet50_20190528.zip	203.97MB	37c12ca43e8fc105918d108741e66faa	resnet50	224*224	7.7G	25.6M	caffe_nv	ImageNet Train
Image Classification	inception_v1	https://www.xilinx.com/bin/public/op filename=inception_v1_20190528.zip	79.39 MB	d4861a76df1a8cd4c15e47789326befe	inception_v1	224*224	3.16G	7.02M	caffe_nv	ImageNet Train
Image Classification	inception_v2	https://www.xilinx.com/bin/public/op filename=inception_v2_20190528.zip	128.7 MB	16c103b72d62955cc7115e1b85c38c69	bn-inception	224*224	4G	10M	caffe_nv	ImageNet Train
Image Classification	inception_v3	https://www.xilinx.com/bin/public/op filename=inception_v3_20190528.zip	190.71 MB	eb5d4494d8c93bf2344b8c0104ff50c7	inception-v3	299*299	11.4G	23.8M	caffe_nv	ImageNet Train
Image Classification	mobileNet_v2	https://www.xilinx.com/bin/public/op filename=mobilenet_v2_20190528.zip	28.61 MB	1f11d595dc25c272af77174dbc445ce8	MobileNet_v2	224*224	608M	3.4M	caffe	ImageNet Train
Image Classification	tf_resnet50	https://www.xilinx.com/bin/public/op filename=tf_resnet50_20190528.zip	204.41 MB	97da4a71a02506bb100b5833288bd3ba	resnet50	224*224	6.97G	25.6M	tensorflow	ImageNet Train
Image Classification	tf_inception_v1	https://www.xilinx.com/bin/public/op filename=tf_inception_v1_20190528.zip	53.44 MB	0b3d6d03fc0143835d8880d1fd1ec322	inception_v1	224*224	3.0G	6.6M	tensorflow	ImageNet Train
Image Classification	tf_mobilenet_v2	https://www.xilinx.com/bin/public/op filename=tf_mobilenet_v2_20190528.zi	49.84 MB	a70f91359e52177305cffe3bab138d75	mobileNet_v2	224*224	1.17G	6.1M	tensorflow	ImageNet Train
ADAS Vehicle Detection	ssd_adas_pruned_0.95	https://www.xilinx.com/bin/public/op filename=ssd_adas_pruned_0.95_20190	8.23 MB	07968f6fbf1d0f2559acb4745919b4e6	VGG-16	480x360	6.3G	0.72M	caffe	bdd100k + private da
ADAS Pedstrain Detection	ssd_pedestrain_pruned_0.97	https://www.xilinx.com/bin/public/op filename=ssd_pedestrain_pruned_0.97	5.49 MB	01fc90acd723c7eba47879573b2b21fc	VGG-bn-16	640*360	5.9G	0.47M	caffe	coco2014_train_pers and crowndhuman

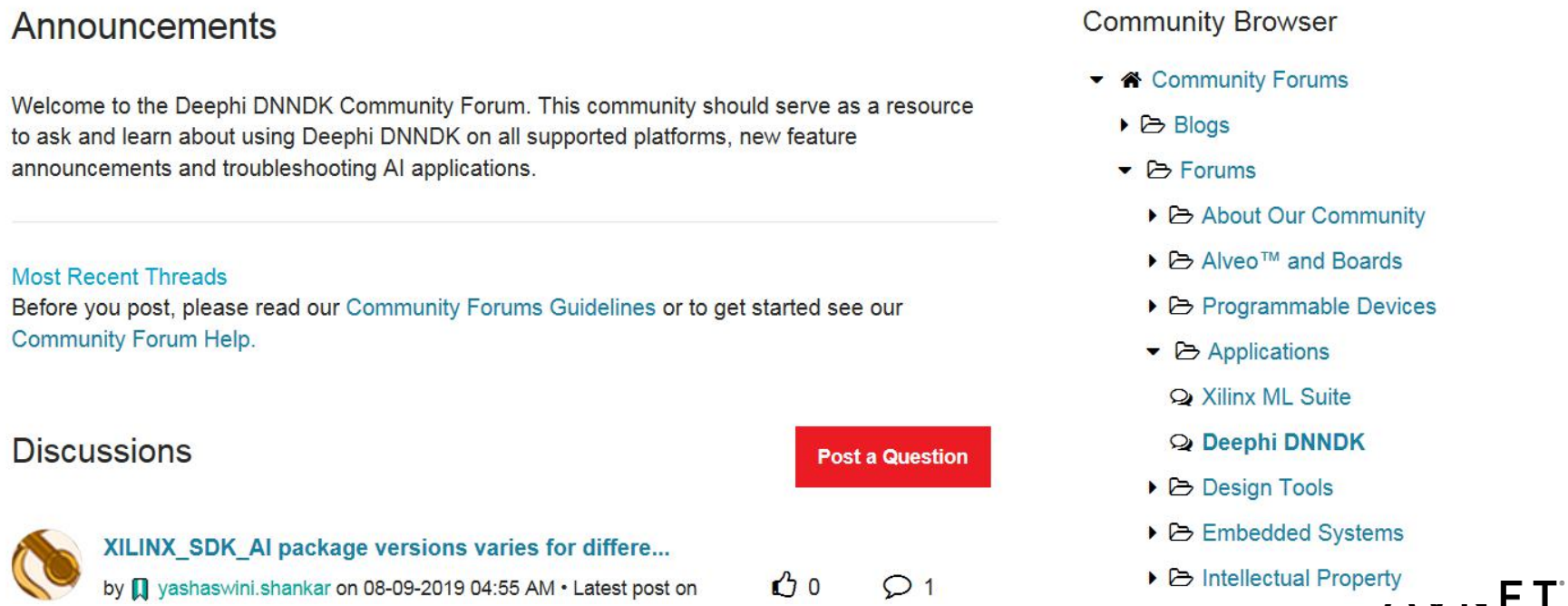
AVNET

Xilinx AI Demo Zoo

- AI SDK Demo
- 8-ch RTSP stream VCU+DPU demo
- Multi-model multi-camera Demo
- Demo Guide

Xilinx AI Forums

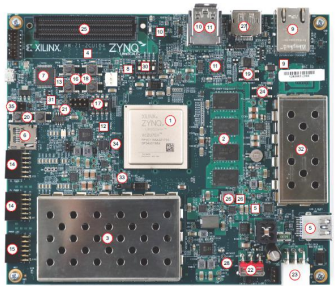
https://forums.xilinx.com/t5/Deephi-DNNDK/bd-p/Deephi



Out-of-box Supported Boards

- ZCU102
- ZCU104
- Avnet Ultra96

Ultra96

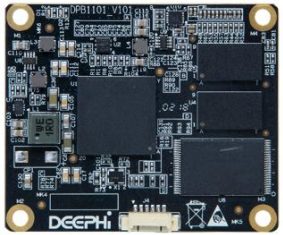


ZCU104

ZCU102

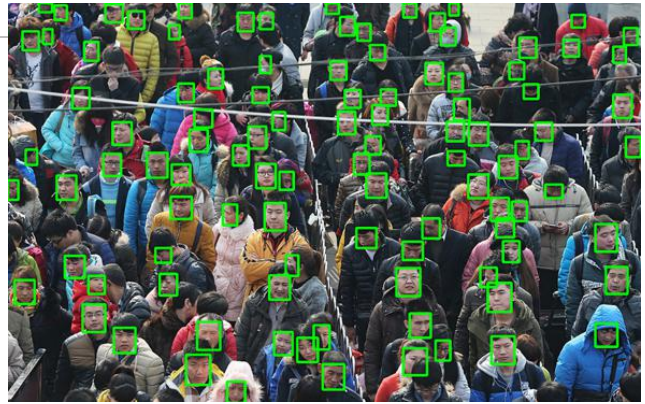
AVNET

Video Surveillance ML Solutions

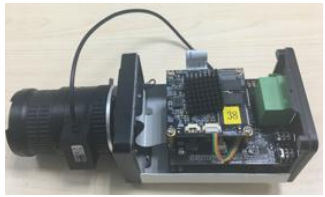


Intelligent
IP Camera Solution

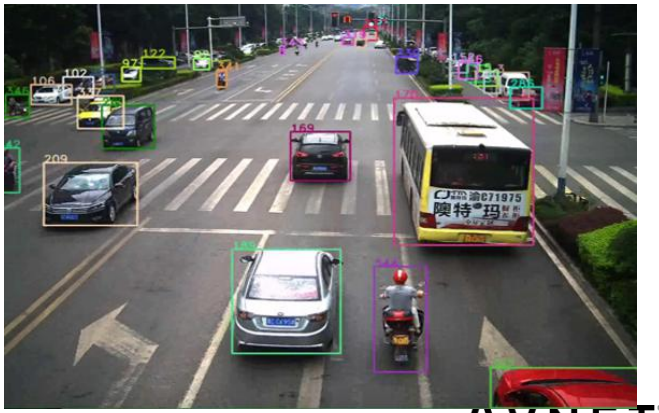
Video Analytics
Acceleration Solution



Face recognition camera
with Zynq7020

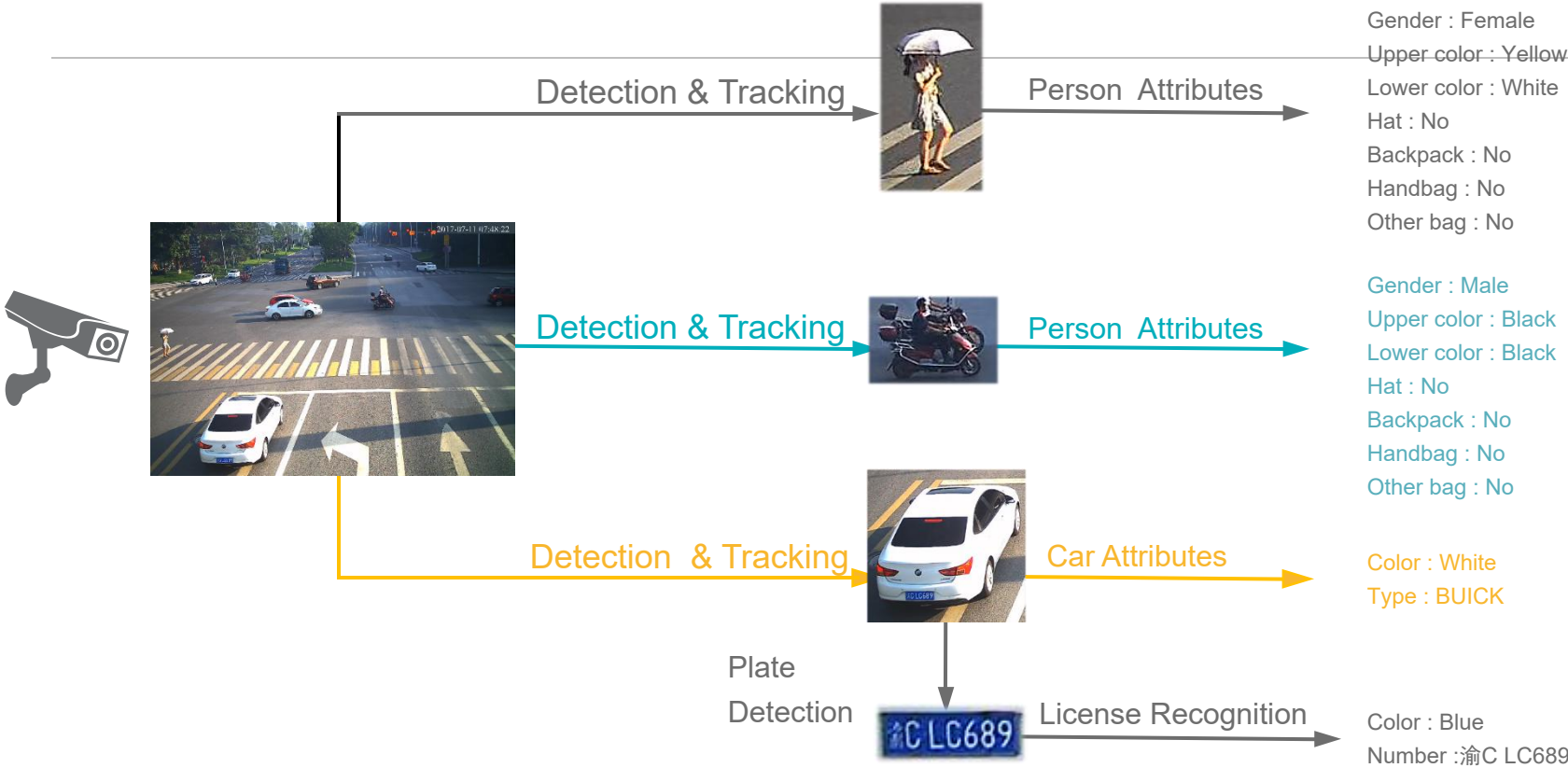


12-channel 1080P Video Analytics
with ZU9EG



AVNET

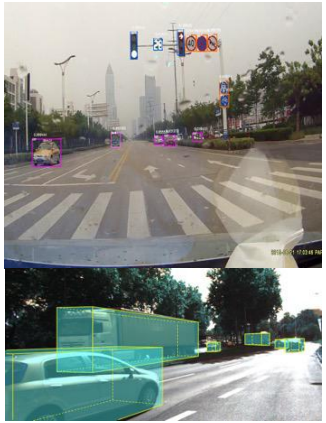
Video Surveillance ML Ref Design



AVNET

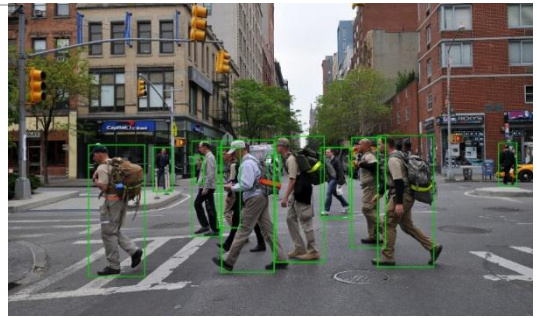
ADAS/AD ML Reference Design

2D/3D Object Detection



Lane Detection

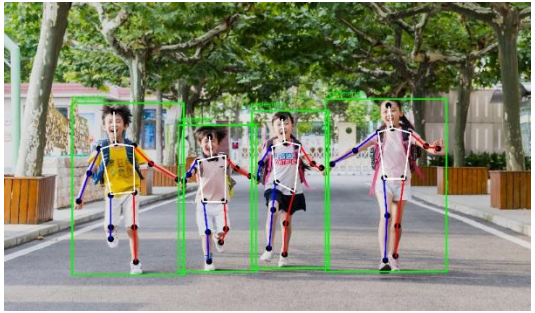
Pedestrian Detection



Segmentation + Detection

Segmentation

Pose Estimation



AVNET

8CH Detection Demo

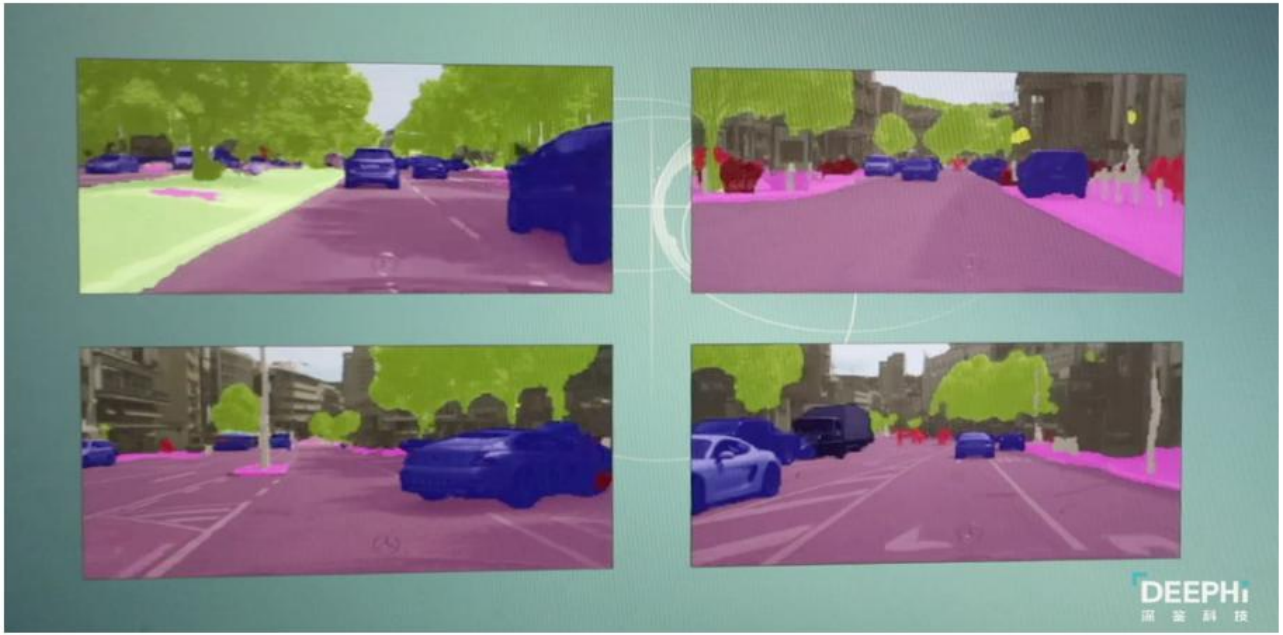
- Xilinx device
  - ZU9EG
- Network
  - SSD compact version
- Input image size to DPU
  - 480 * 360
- Operations per frame
  - 4.9G
- Performance
  - 30fps per channel

*Removed Video

AVNET

4-ch Segmentation + Detection Demo

- Xilinx device
  - ZU9EG
- Network
  - FPN compact version
  - SSD compact version
- Input image size to DPU
  - FPN – 512 * 256
  - SSD – 480 * 360
- Operations per frame
  - FPN – 9G
  - SSD – 4.9G
- Performance
  - 15fps per channel



*Removed Video

AVNET

Supported DNN (Deep Neural Network) by Applications

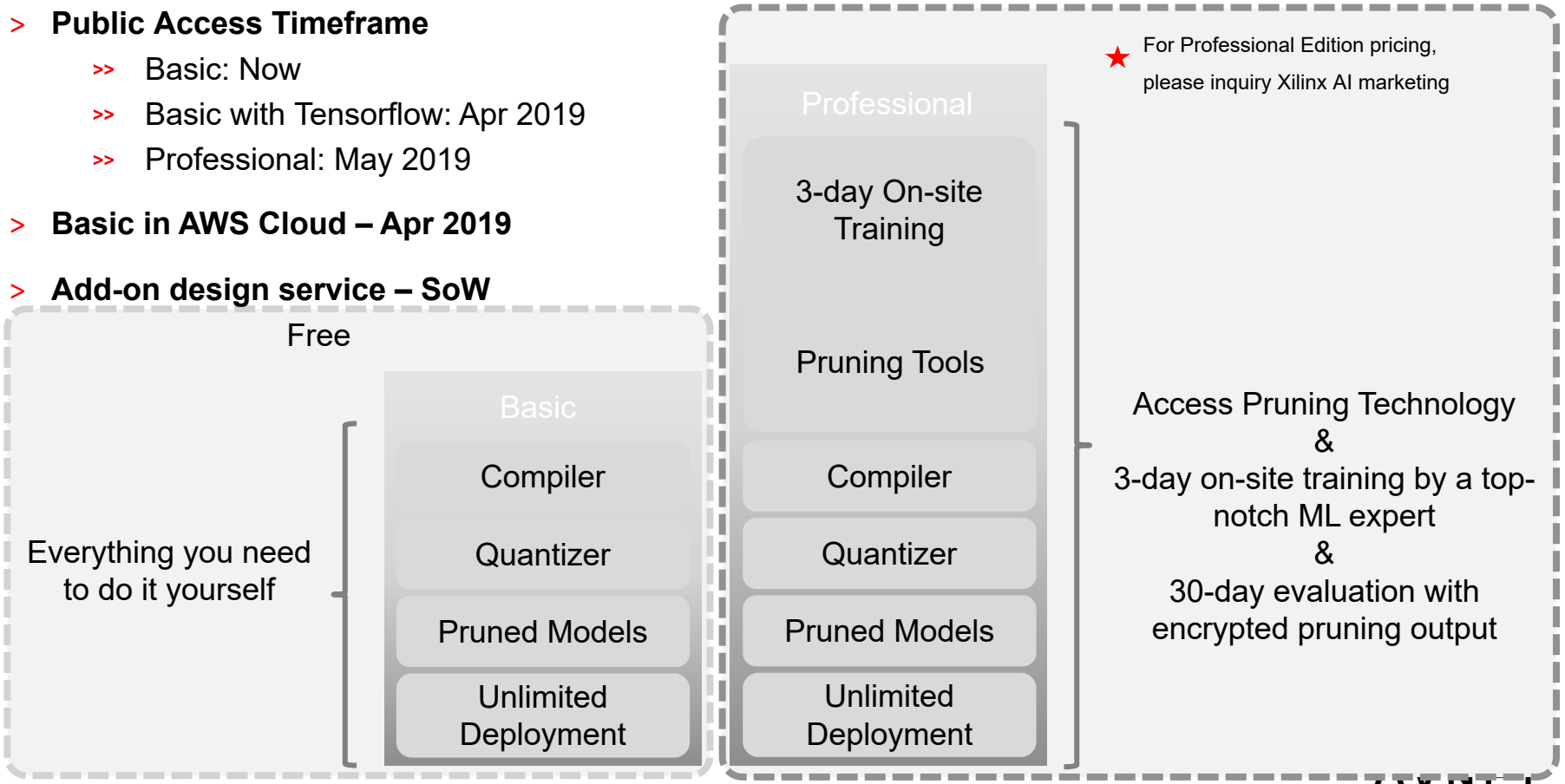
Application	Function	Algorithm	Developed	Pruned	Deployed
Face	Face detection	SSD, Densebox	✓	✓	✓
	Landmark Localization	Coordinates Regression	✓	N / A	✓
	Face recognition	ResNet + Triplet / A-softmax Loss	✓	✓	✓
	Face attributes recognition	Classification and regression	✓	N / A	✓
Pedestrian	Pedestrian Detection	SSD	✓	✓	✓
	Pose Estimation	Coordinates Regression	✓	✓	✓
	Person Re-identification	ResNet + Loss Fusion	✓		
Video Analytics	Object detection	SSD, RefineDet	✓	✓	✓
	Pedestrian Attributes Recognition	GoogleNet	✓	✓	✓
	Car Attributes Recognition	GoogleNet	✓	✓	✓
	Car Logo Detection	DenseBox	✓	✓	
	Car Logo Recognition	GoogleNet + Loss Fusion	✓	✓	
	License Plate Detection	Modified DenseBox	✓	✓	✓
	License Plate Recognition	GoogleNet + Multi-task Learning	✓	✓	✓
ADAS/AD	Object Detection	SSD, YOLOv2, YOLOv3	✓	✓	✓
	3D Car Detection	F-PointNet, AVOD-FPN	✓		
	Lane Detection	VPGNet	✓	✓	✓
	Traffic Sign Detection	Modified SSD	✓		
	Semantic Segmentation	FPN	✓	✓	✓
	Drivable Space Detection	MobilenetV2-FPN	✓		
	Multi-task (Detection+Segmentation)	Deephi	✓		

Supported Operators

- Conv
  - Dilation
- Pooling
  - Max
  - Average
- ReLU / Leaky Relu/ Relu6
- Full Connected (FC)
- Batch Normalization
- Concat
- Elementwise
- Deconv
- Depthwise conv
- Mean scale
- Upsampling
- Split
- Reorg
- Resize (Optional)
- Softmax (Optional)
- Sigmoid (Optional)

Basic and Professional Editions

- **Public Access Timeframe**
  - Basic: Now
  - Basic with Tensorflow: Apr 2019
  - Professional: May 2019
- **Basic in AWS Cloud – Apr 2019**
- **Add-on design service – SoW**

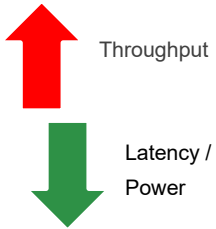


Availability

- **DNNDK & DPU**
  - DNNDK basic edition - Download from Xilinx.com
  - Pruning tool, separate upon request
  - DPU available for evaluation & system integration upon request
- **Demos & Ref Designs**
  - General: Resnet50, Googlenet, VGG16, SSD, Yolo v2/v3, Tiny Yolo v2/v3, Mobilenet v1/v2 etc..
  - Video surveillance: face detection & traffic structure
  - ADAS/AD: multi-channel detection & segmentation
  - DPU TRD (Work in progress)
- **Documentation**
  - DNNDK user guide – UG1327
  - DNNDK for SDSoC user guide – UG1331
  - Edge AI tutorials - https://github.com/Xilinx/Edge-AI-Platform-Tutorials
  - DPU product guide & tutorial (Work in progress)
- **Request or Inquiry**
  - Please contact Andy Luo, andy.luo@xilinx.com

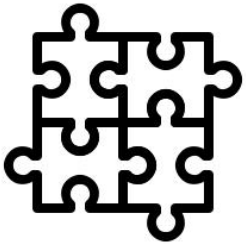
Xilinx ML for Cloud/DC

Xilinx Value Proposition for Cloud/DC ML



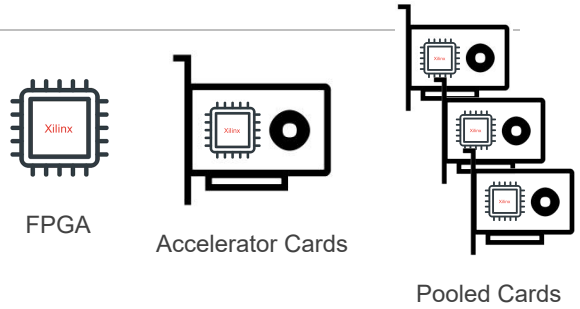
Customizable Performance

- Highest perf/watt
- Lower precision optimized
- Optimizations for Throughput and Latency



Flexibility

- Well suited for evolving Deep Learning Field
- Support for all types – CNN, DNN, LSTM
- Integrate with custom application in FPGA
- Full Software Stack for Applications



Scalable

- Configurable from 1 to many xDNN Engines
- Pool Xilinx cards for higher performance
- Deployed today on edge or cloud

AVNET

ML Suite Features

Based on xDNN v2

Supported Frameworks:

- Caffe
- MxNet
- Tensroflow

Examples

- DeepDetect REST Tutorial
- DeepDetect Webcam
- Image Classification
- x8 FPGA Pooling GoogLeNet v1 Demo on AWS F1 instance

xfDNN Tools

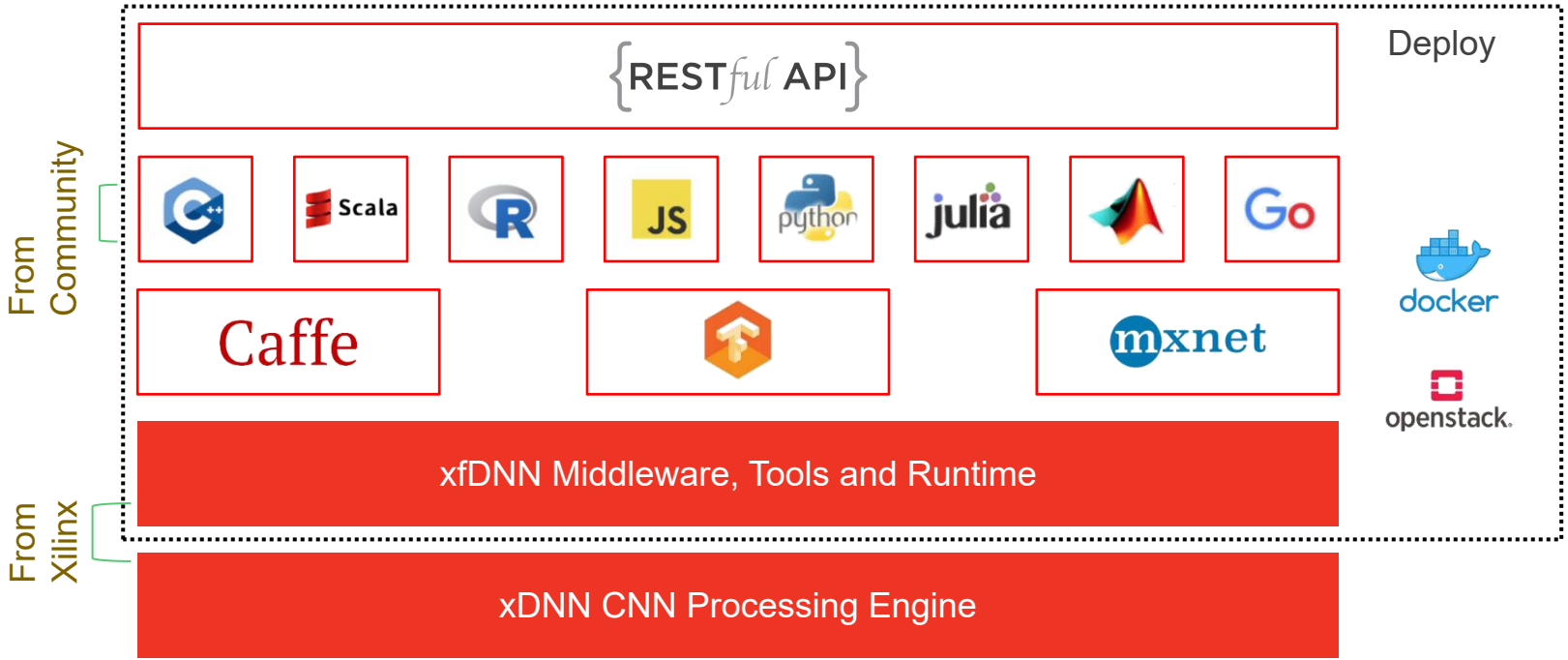
- Compiler
- Quantizer

Easy to use Python Interface

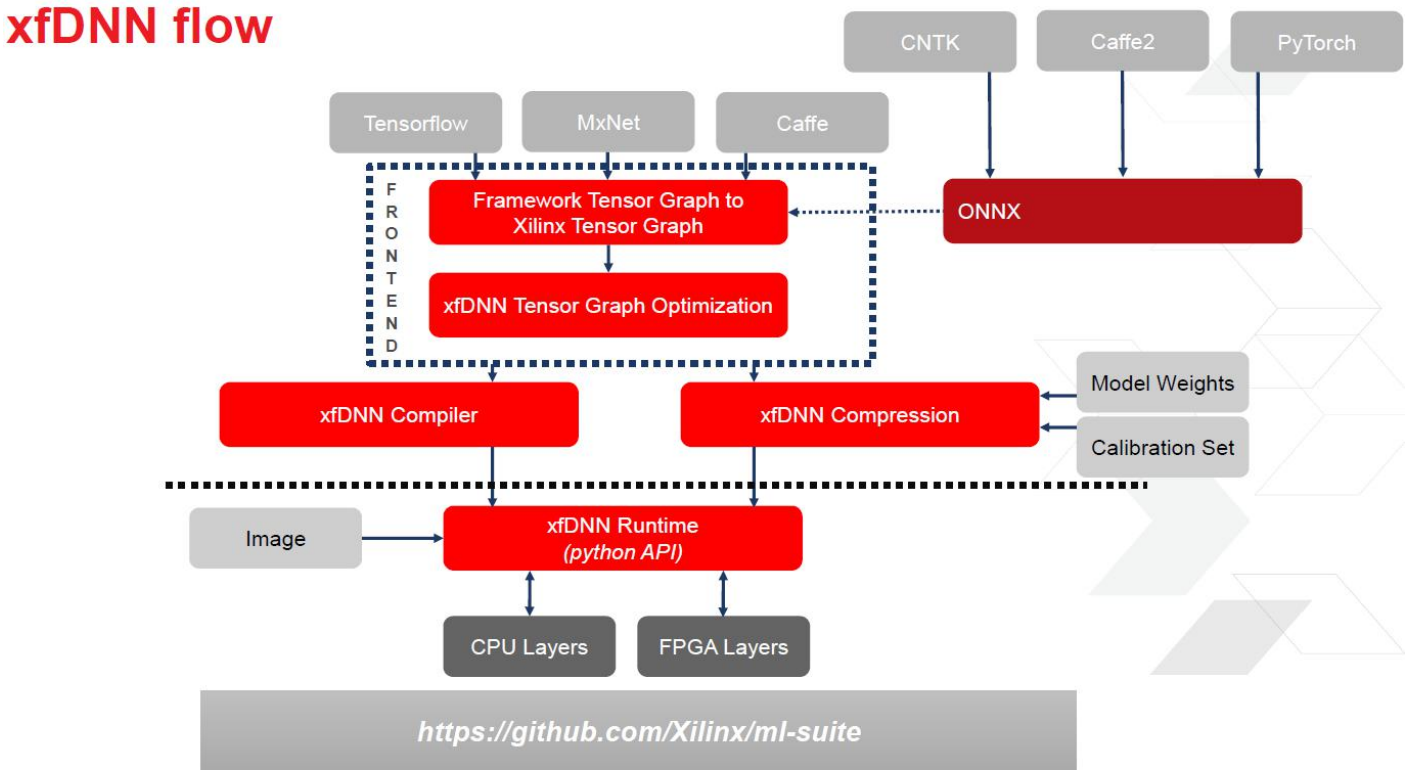
Precompiled Models

- 8/16-bit GoogLeNet v1
- 8/16-bit ResNet50

Seamless Deployment with Open Source Software



xfDNN Flow



AVNET

ML Suite Overlays with xDNN Processing Engines

Adaptable

> AI algorithms are changing rapidly

> Adjacent acceleration opportunities

Realtime

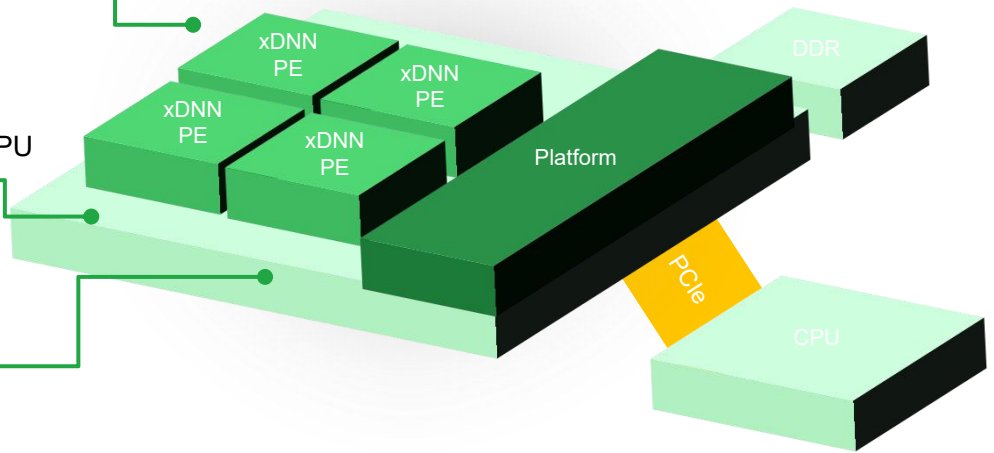
> 10x Low latency than CPU and GPU

> Data flow processing

Efficient

> Performance/watt

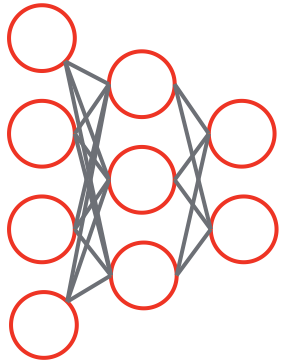
> Low Power



AVNET

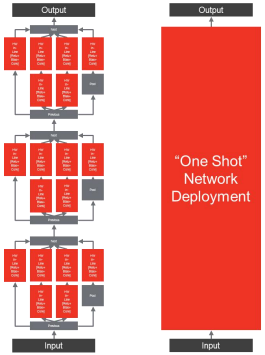
xfDNN Inference Toolbox

Graph Compiler



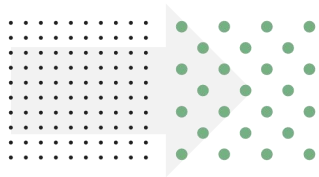
- Build network graphs from Frameworks
- Optimizes for Inference
- Generate code for xDNN IP
- HW/SW partition

Network Optimization



- Fused Layer Optimization
- On-Chip Memory enables Streaming
- "One Shot" Inference eliminates CPU calls

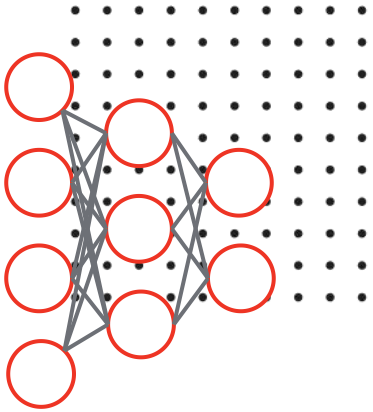
xfDNN Quantizer



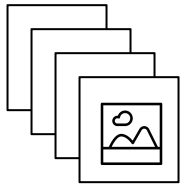
- Easily deploy pre-trained floating point models on 8 bit.
- Maintains accuracy without needing lengthy retraining
- Easy and Fast

AVNET

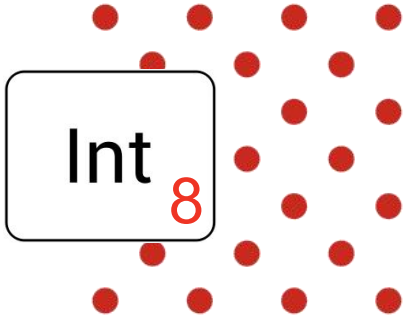
xfDNN Quantizer: Fast and Easy



1) Provide FP32 network and model
   - E.g., prototxt and caffemodel



2) Provide a small sample set, no labels required
   - 16 to 512 images



3) Specify desired precision
   - Quantizes to <8 bits to match Xilinx's DSP

Xilinx ML Processing Engine - xDNN

Features		Description	
Supported Operations	Convolution / Deconvolution / Convolution Transpose	Kernel Sizes	W: 1-15; H:1-15
		Strides	W: 1,2,4,8; H: 1,2,4,8
		Padding	Same, Valid
		Dilation	Factor: 1,2,4
		Activation	ReLU
		Bias	Value Per Channel
		Scaling	Scale & Shift Value Per Channel
	Max Pooling	Kernel Sizes	W: 1-15; H:1-15
		Strides	W: 1,2,4,8; H: 1,2,4,8
		Padding	Same, Valid
	Avg Pooling	Kernel Sizes	W: 1-15; H:1-15
		Strides	W: 1,2,4,8; H: 1,2,4,8
		Padding	Same, Valid
	Element-wise Add	Width & Height must match; Depth can mismatch.	
	Memory Support	On-Chip Buffering, DDR Caching	
	Expanded set of image sizes	Square, Rectangular	
	Upsampling	Strides	Factor: 2,4,8,16
Miscellaneous	Data width	16-bit or 8-bit	

- Programmable Feature-set
- Tensor level instructions
- 700M+ DSP Freq (VU9P)
- Customer network acceleration

Computer Vision On Xilinx FPGA

OpenCV for Xilinx 介绍

- Xilinx并没有自己的机器视觉算法，HLS中所有的算法来源都是OpenCV。
- 目前HLS提供的机器视觉算法函数，都只是opencv原版函数的一个重构，功能以及接口参数基本上同原opencv函数保持，适合于HLS综合成hdl代码硬件实现。
- 客户可以直接调用这些函数，也可以参考它们的实现，针对自己的算法做修改
- 毕竟opencv几千个函数，不可能所有的都提供HLS重构实现，Avnet可以协助客户做的：
  - 在客户已有自己的算法的前提下，Avnet 帮助客户评估在PS/PL实现的优劣（注重于从性能以及资源代价等方面去帮客户评估）。
  - 如果在PS实现，那么客户并不需要做太多工作，客户的代码可以比较容易的就移植到Zynq的ARM上。
  - 某些性能要求苛刻的场合，需要PL逻辑加速的，Avnet 会帮助客户用HLS把已有的软件算法转换成PL实现。
  - **Avnet 可以提供基于OpenCV 开源算法的Zynq平台实现方法培训**

xFopencv: HW Accelerated OpenCV Functions

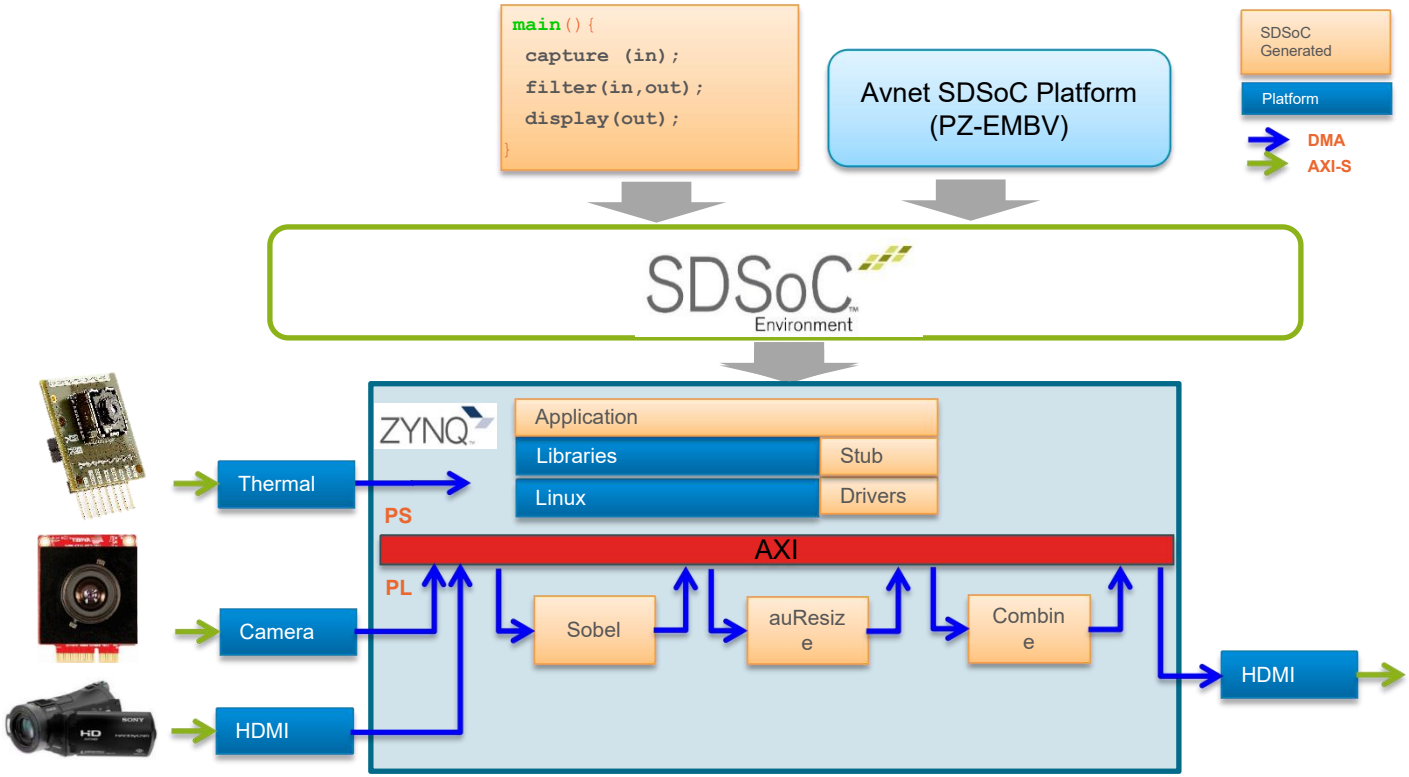
Level 1		Level 2		Level 3
Absolute difference	Channel combine	Box	Scale/Resize	Histogram of Oriented Gradients (HOG)
Accumulate	Channel extract	Gaussian	StereoRectify	ORB
Accumulate squared	Color convert	Median	Warp Affine	SVM (binary)
Accumulate weighted	Convert bit depth	Sobel	Warp Perspective	OTSU Thresholding
Arithmetic addition	Table lookup	Custom convolution	Fast corner	Mean Shift Tracking (MST)
Arithmetic subtraction	Histogram			LK Dense Optical Flow
Bitwise: AND, OR, XOR, NOT	Gradient Phase	Dilate	Harris corner	Canny edge detection
Pixel-wise multiplication	Min/Max Location	Erode	Remap	Image pyramid
Integral image	Mean & Standard Deviation	Bilateral	Equalize Histogram	Color Detection
Gradient Magnitude	Thresholding			StereoLBM

赛灵思高层次综合工具 （Vivado HLS）

- 是一种从 C -> RTL 语言的转换工具
- 全面覆盖 C、C++、OpenCL，能够进行浮点运算和任意精度浮点运算
- 可以输出 Verilog和 VHDL代码
- 可以通过制定约束（Directive）来提高运算性能和优化资源利用率
- 从算法验证到硬件实现的自动化工作流程
- 可以集成到嵌入式、System Generator 和 IP Integrator 中
- 适合C算法的工程师进行快速硬件验证
- 赛灵思提供基于OpenCV的库函数，适用于图像处理

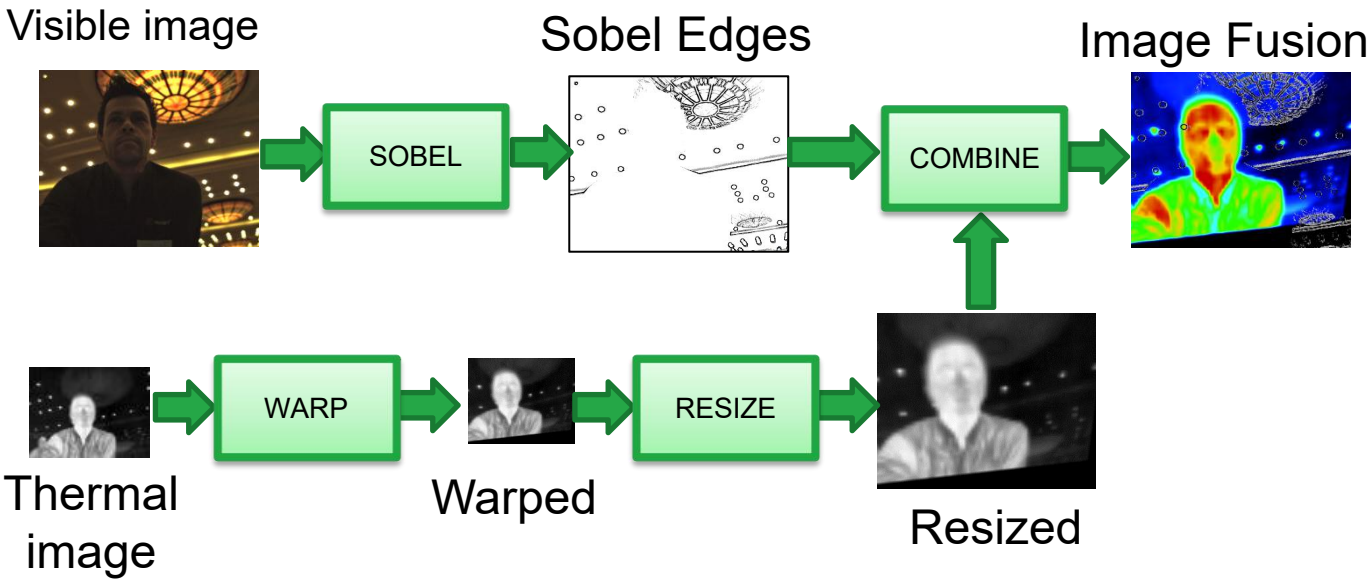
AVNET

Computer Vision / Sensor Fusion with the PicoZed Embedded Vision Kit

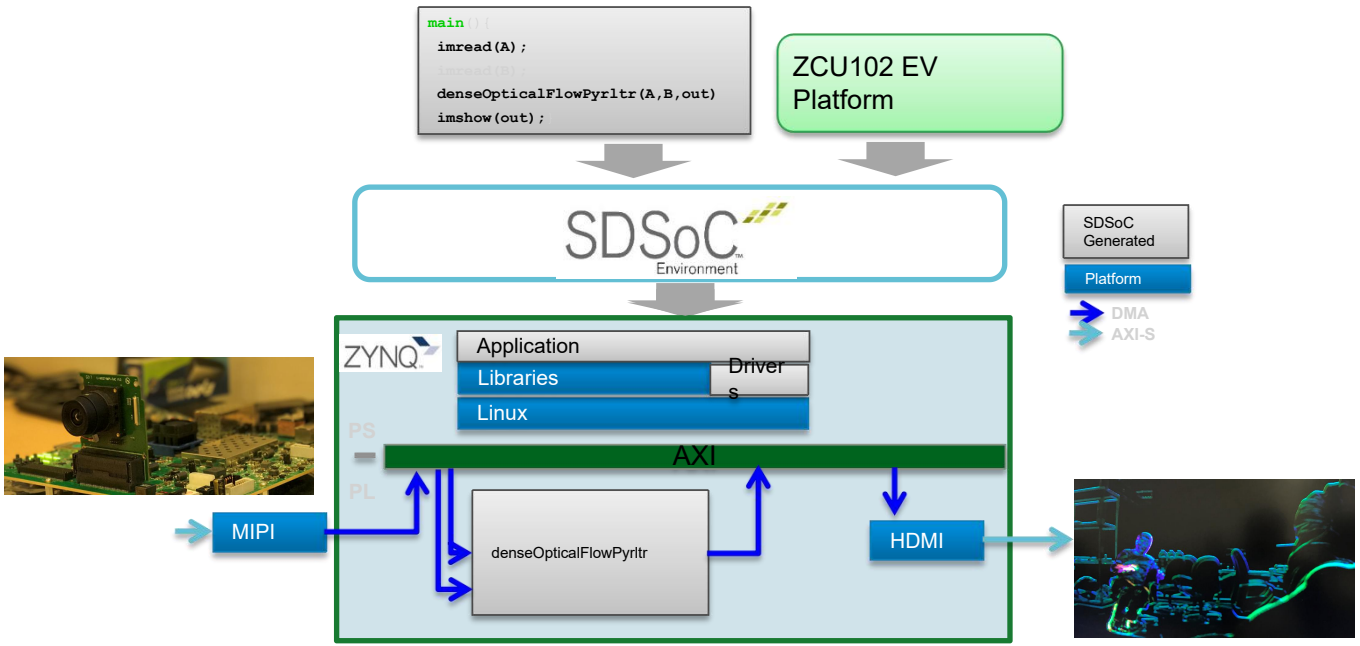


AVNET

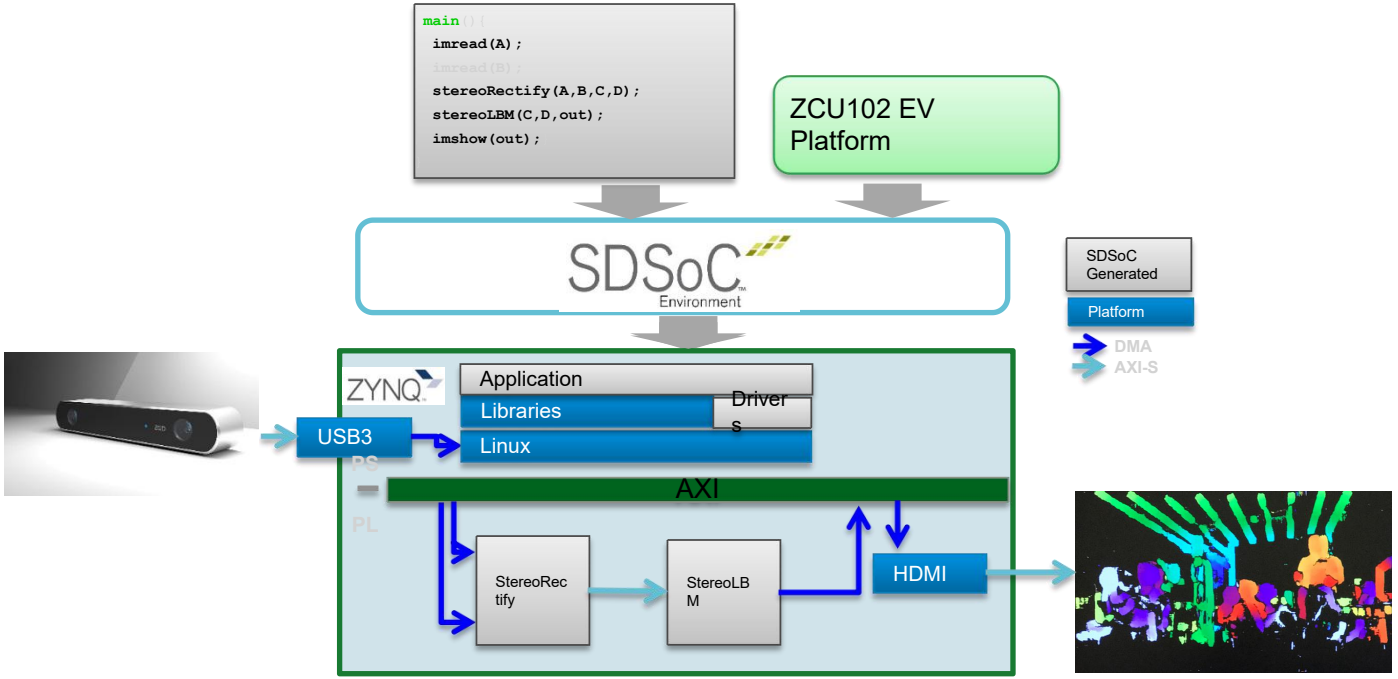
Computer Vision / Sensor Fusion with the PicoZed Embedded Vision Kit



Computer Vision Design Example: 4K60 Dense Optical Flow



Computer Vision Design Example: Stereo Disparity Map



AVNET

Avnet 开发板集锦 (http://ultrazed.org/)

- MicroZed: 基于Zynq 平台的袖珍开发板
- PicoZed: 基于Zynq 平台可用于生产的核心板
- UltraZed：基于 ZU3EG的学习板和核心板
- Ultra96：用于96 社区的 ZU3EG 开发板，也可用于生产。
- Zedboard：基于Zynq 平台的开发版，拥有最多客户群体的开发板。
- MiniZed：基于7Z007 单核平台的开发板
- Mini-ITX: 基于Zynq 平台的工业 Mini-ITX 开发板
- FPGA 系列：6SLX9, 7A35T/50T, KU040

Avnet 开发板 – Ultra96

- The Ultra96-V2 updates and refreshes the Ultra96 product that was released in 2018.
- Like Ultra96, the Ultra96-V2 is an Arm-based, Xilinx Zynq UltraScale+ ™ MPSoC development board based on the Linaro 96Boards Consumer Edition (CE) specification.
- Ultra96-V2 has been designed with a certified radio module from Microchip.
- Additionally, all components are updated to allow industrial temperature grade options. Additional power control and monitoring will be possible with the included Infineon Pmics.
- Support DNNDK, PYNQ.

Features

- Xilinx Zynq UltraScale+ MPSoC ZU3EG A484
- Micron 2 GB (512M x32) LPDDR4 Memory
- Delkin 16 GB microSD card + adapter
- PetaLinux environment available for download
- Microchip Wi-Fi / Bluetooth
- Mini DisplayPort (MiniDP or mDP
- 85mm x 54mm form factor
- Linaro 96Boards Consumer Edition compatible

Xilinx DPU IP Deployment in Ultra96 (XCZU3EG)

The tutorial is here:
https://github.com/Xilinx/Edge-AI-Platform-Tutorials/tree/master/docs/DPU-Integration

XILINX

Edge AI Tutorials

DPU Integration Tutorial

Introduction

This tutorial demonstrates how to build a custom system that utilizes the 1.3.0 version of Xilinx® Deep Learning Processor (DPU) IP to accelerate machine learning algorithms using the following development flow:

1. Build the hardware platform in the Vivado® Design Suite.
2. Generate the Linux platform in PetaLinux.
3. Use Xilinx SDK to build two machine learning applications that take advantage of the DPU.

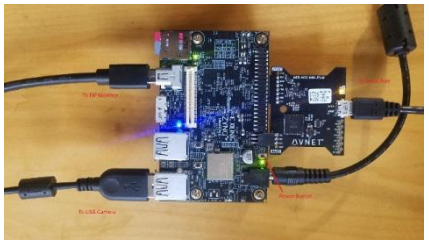
Note: The Ultra96 will be the targeted hardware platform.

Requirements for Using the Xilinx DPU

This section lists the software and hardware tools required to use the Xilinx® Deep Learning Processor (DPU) IP to accelerate machine learning algorithms.

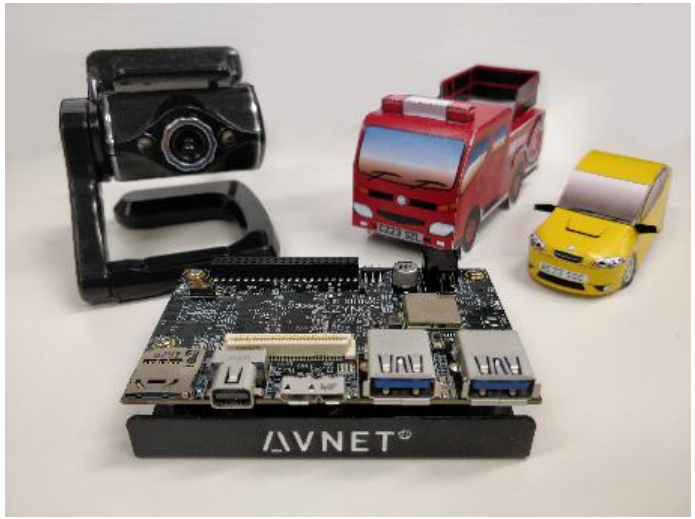
Software Requirements

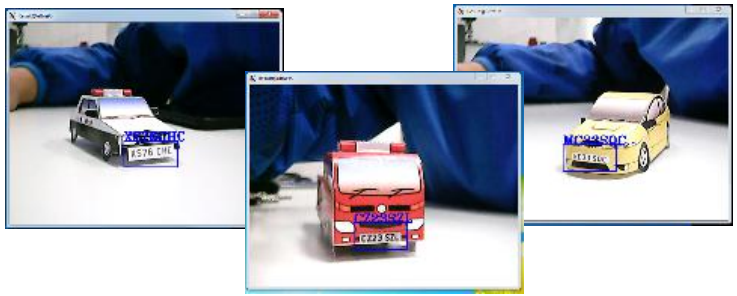
- Vivado® Design Suite 2018.2
- Board files for Ultra96 v1 should be installed
- Xilinx SDK 2018.2
- PetaLinux 2018.2



AVNET

自动车牌识别参考设计





特性

- 自定义深度学习网络
- Tensorflow 框架
- MMdnn 框架转换
- Xilinx DeePhi DNNDK 2.08
- Xilinx DeePhi DPU

- 精确度 92%
- 性能 37.6 plate/s

主要部件

- Avnet Ultra96 开发板
- Xilinx Zynq UltraScale+ MPSoC ZU3EG
- 2GB LPDDR4
- Wi-Fi
- USB 摄像机

目标应用

- 停车场自动化
- 公路计费

For further enquiries, please contact us at Asia-GDS-Contact@Avnet.com

AVNET

Avnet 产品 – 人脸识别相机

1080P 人脸识别一体摄像机

专用深度学习处理器，最大支持3万张人脸库

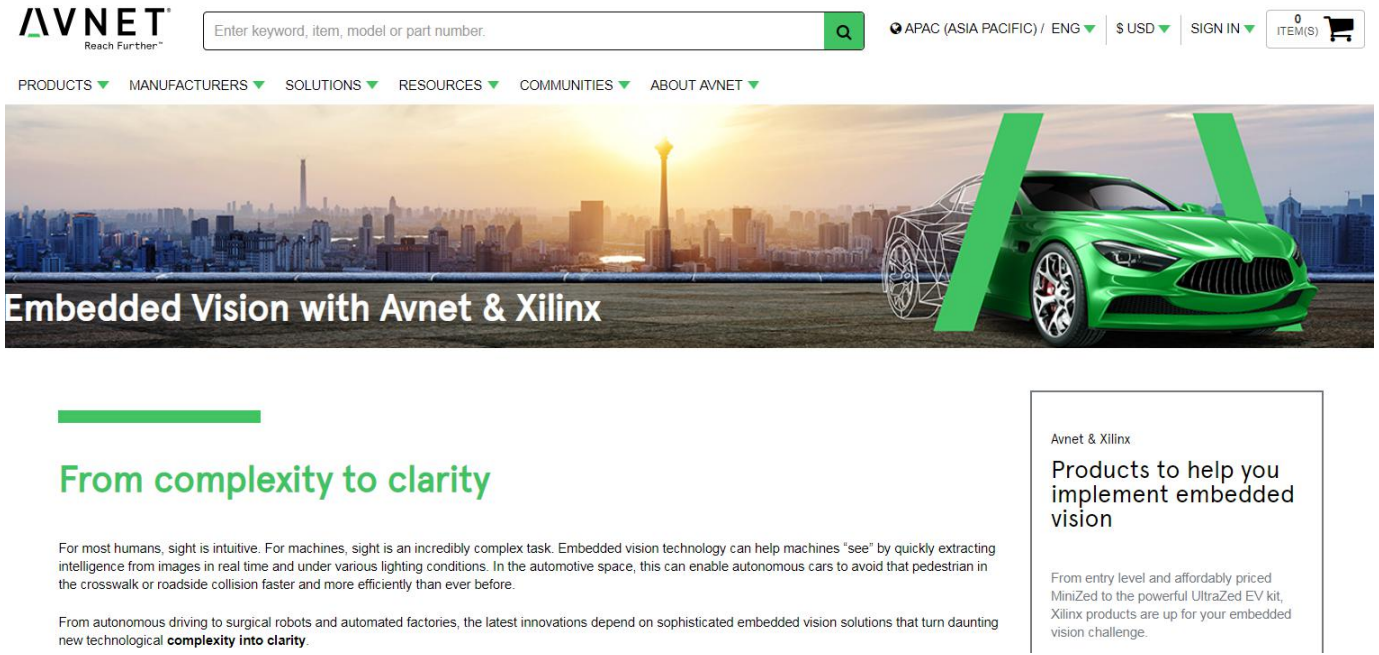
- 大规模神经网络算法，首创嵌入式端侧人脸识别比对一体；
- 专业IP Camera SOC, 4K uHD超高清H.264/H.265视频编码;
- 多路干接点输出可直接控制本地联动设备；
- 内置超级电容，断电可持续工作, 有效保护 TF卡内关键数据

主要应用领域：

- 智慧社区
- 超市及无人零售
- 门禁及道闸
- 交通站点
- 通行卡口

Avnet Embedded Vision Campaign

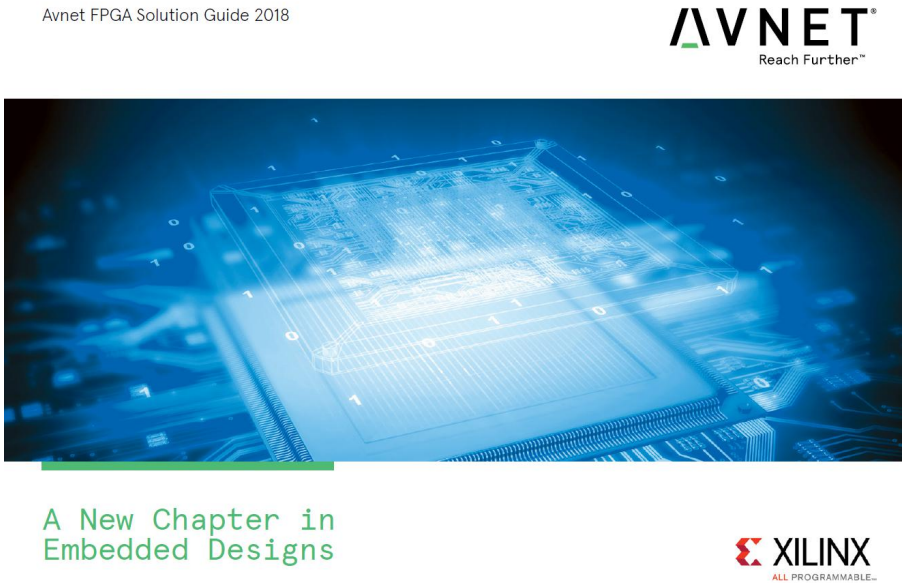
https://www.avnet.com/wps/portal/apac/products/c/embedded-vision/



AVNET

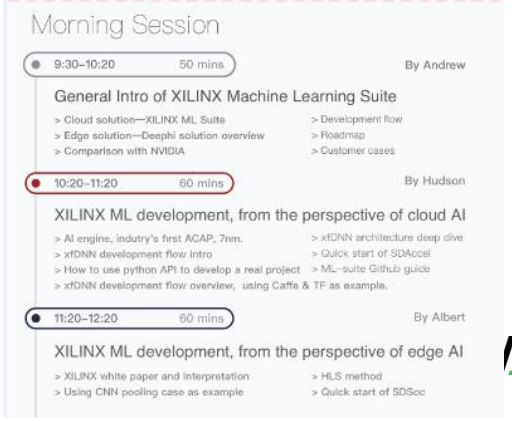
Avnet FPGA Solution Guide

- Reference Design
- IP Core
- Development Kits



Avnet AI Training and Solutions

- Training Event: AI workshop
  - Propose AI workshop to present Xilinx AI solution on Edge (DNNDK) and cloud (ML suite)
- Reference design:
  - Automatic Number Plate Recognition by ADS
  - DNNDK tool chain
  - Based on Ultra96
- Product:
  - Face recognition camera
  - Based on Deephi XC7Z020 module



AVNET

Thank you