



赛灵思工业物联网研讨会

XILINX IIoT SEMINAR

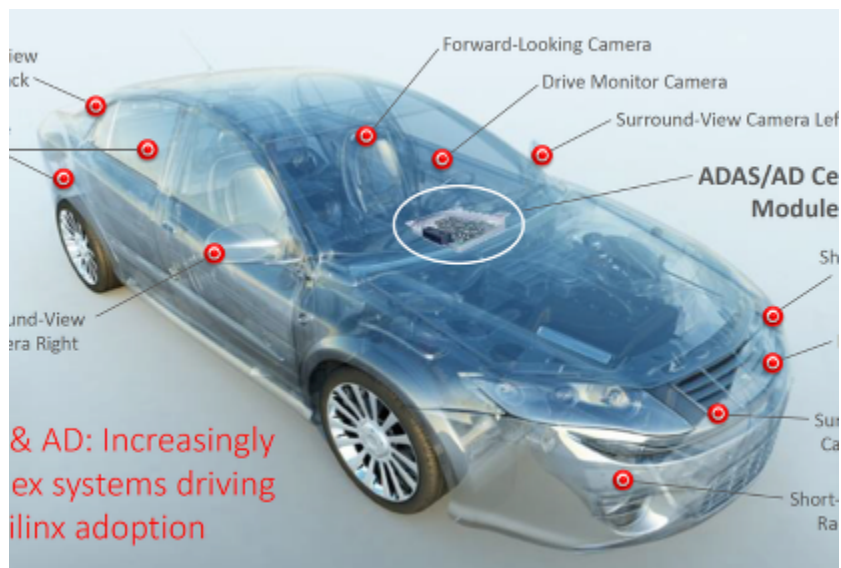
人工智能 — 给机器智慧之眼

刘竞秀

赛灵思 AI 市场开发总监

2019.05

商业模式发展趋势 – AI变现能力迅速增强



& AD: Increasingly
ex systems driving
ilinx adoption



技术发展趋势 - 万物互联+人工智能

1

数据大爆炸

- > 90% 非结构化数据
- > 视频及图像内容为主的数据
- > 需要更高的吞吐量与实时的计算能力

2

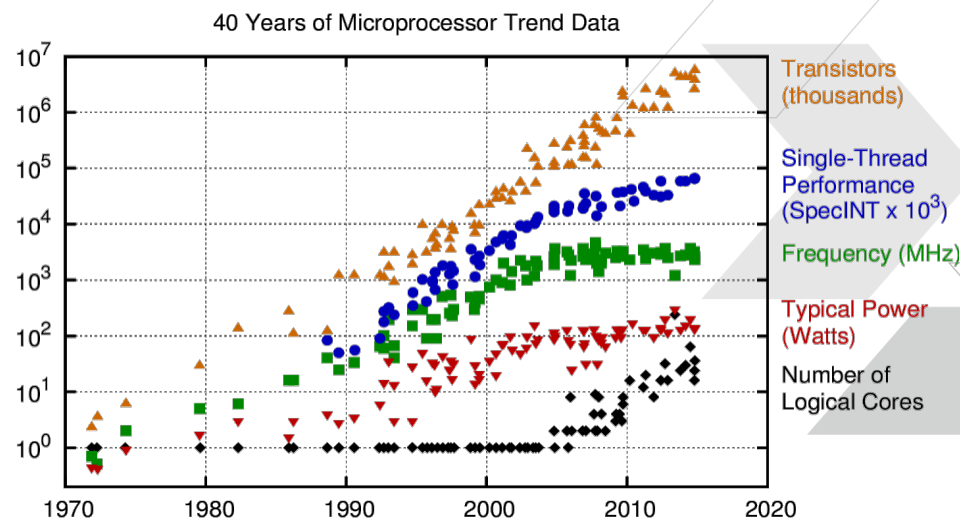
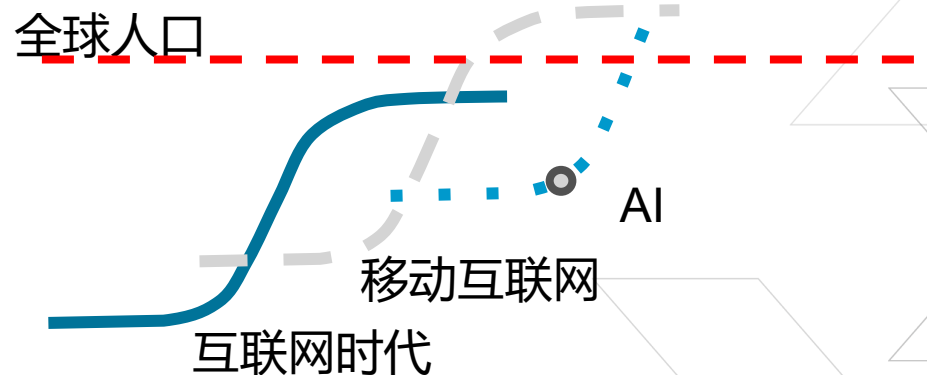
AI 黎明

- > 全行业的采用
- > 将新的智能注入应用
- > 从端到边缘到云

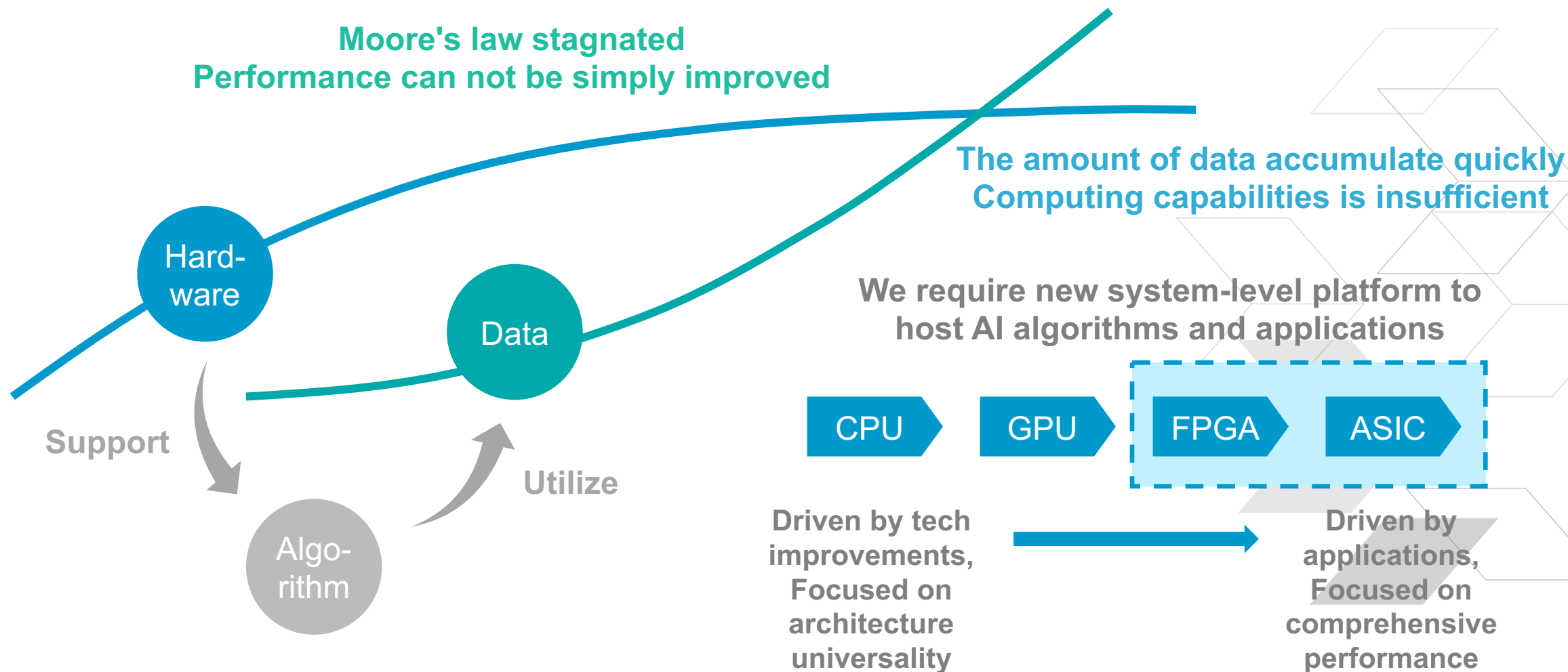
3

后摩尔定律计算时代

- > 带有加速器的异构计算
- > 广泛的应用需要不同的架构
- > 创新的速度超越了芯片的设计周期



高性能计算平台的发展趋势





赛灵思工业物联网研讨会
XILINX IIoT SEMINAR

我们的使命

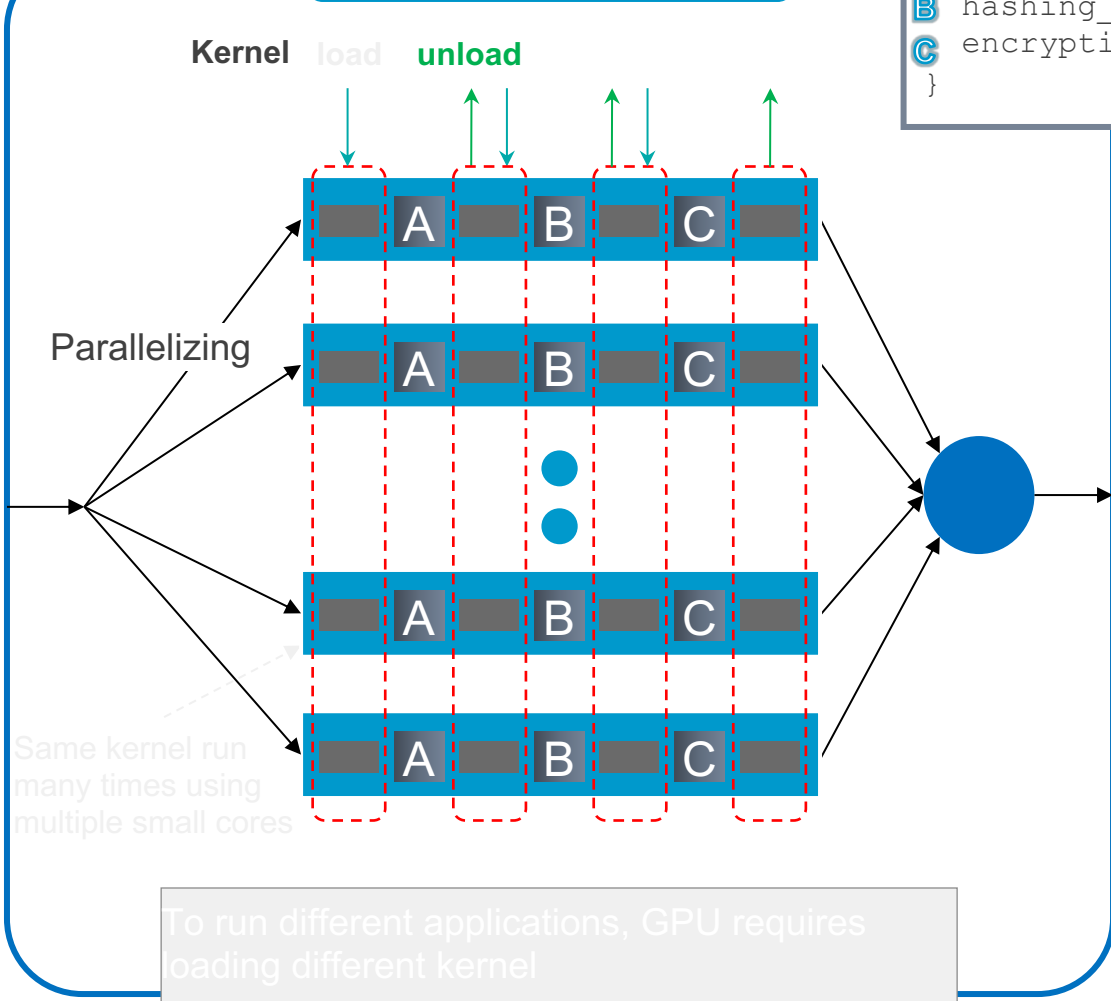
打造灵活应变、万物智能的世界



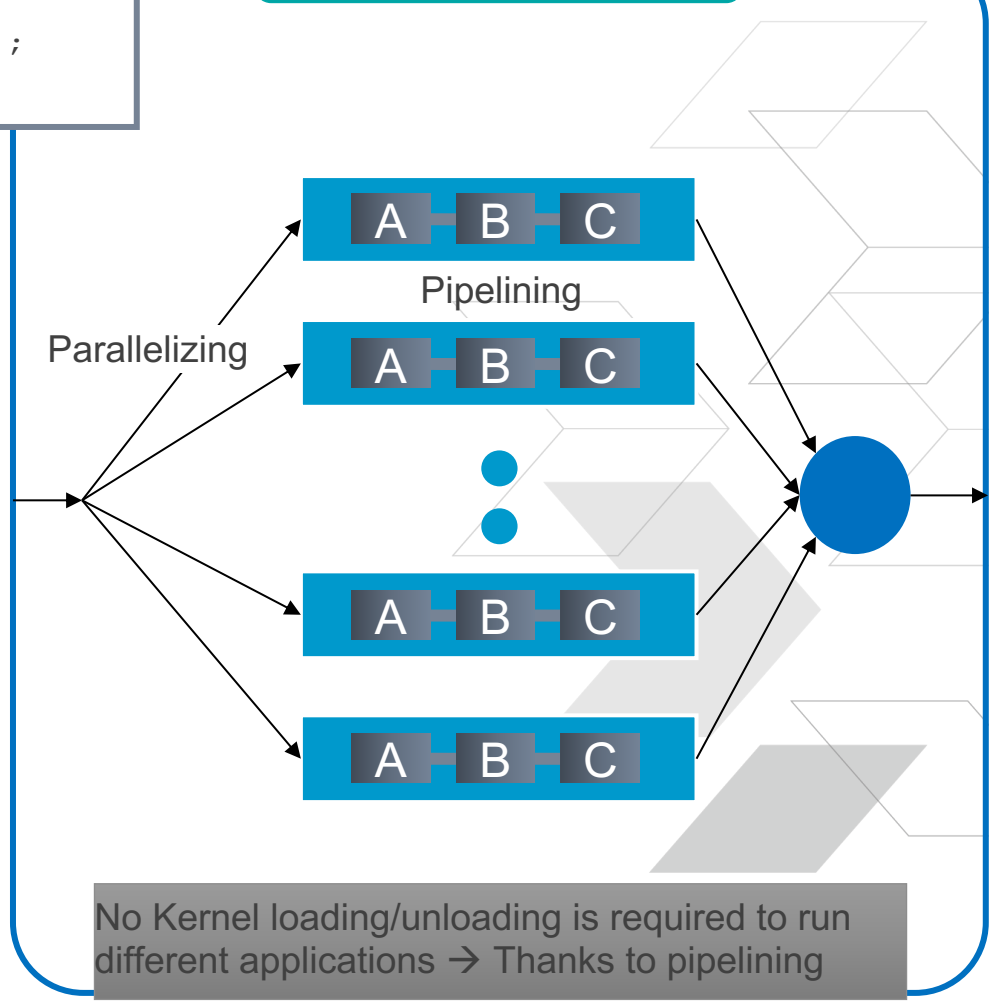
硬件可编程性成就性能更高的架构

```
For (i=0, i< num;++i)  
{  
  A classification_process();  
  B hashing_process();  
  C encryption_process();  
}
```

GPU Implementation

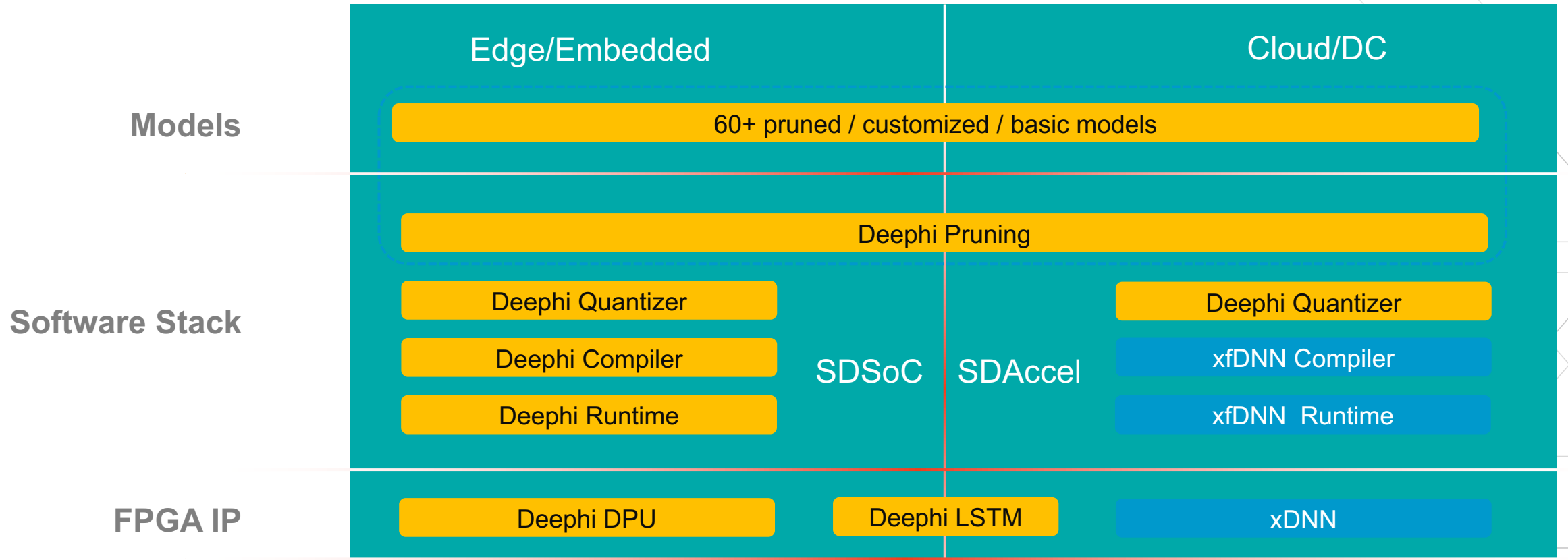


FPGA Implementation

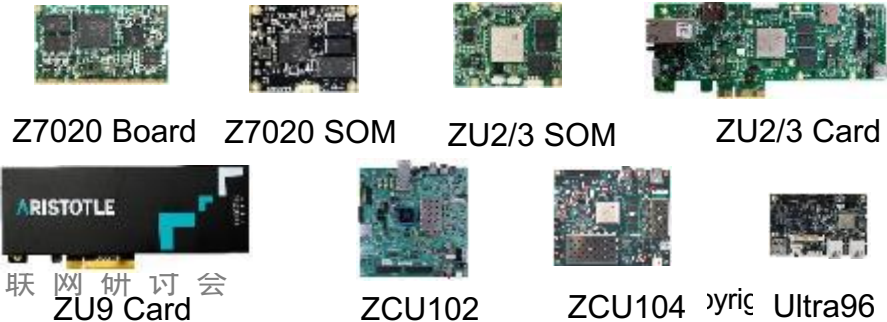


赛灵思人工智能整体解决方案

Xilinx AI Development



Platforms



算法模型集合 (Zoo)

69 Models in the zoo,
in which 34 models
are already deployed.

Network	Backbone	DPU Deployment	Input Size	OPs	Paras	Training Set	Val Set	Eva Norm Float	Eva Norm Fixed
Resnet50_v1	Resnet	Yes	224*224	7.7G	25.6M	ImageNet	ImageNet	0.7483	0.7338
Inception_v3	Inception	No	299*299	11.43G	23.8M	ImageNet	ImageNet	0.7401	0.7347
SSD	VGG16	No	300*300	62.77G	26.3M	Voc07+12	Voc07	77.19%	
RefineDet	VGG-16	Support	480*360	123.9G	29.6M	Coco2014	Coco2014	70.14%	
Densebox		Yes	320*320	492M		Private	Private	97.92%	97.50%
Yolo_v3	Yolo	Yes	512*288	53.7G	61.8M	Cityscape	Cityscape	53.7%	53.1%

- DPU deployment
 - **Yes:** the model is successfully deployed on DPU.
 - **Support:** the model is supported but not deployed. Similar model structure is deployed and test successfully.
 - **No:** the model is not supported by DPU right now mainly due to some special operations or layers.

DNNDK – 深度神经网络开发套件

> DECENT

- >> Flt32 to Int8 quantization with one line command

> DNNC

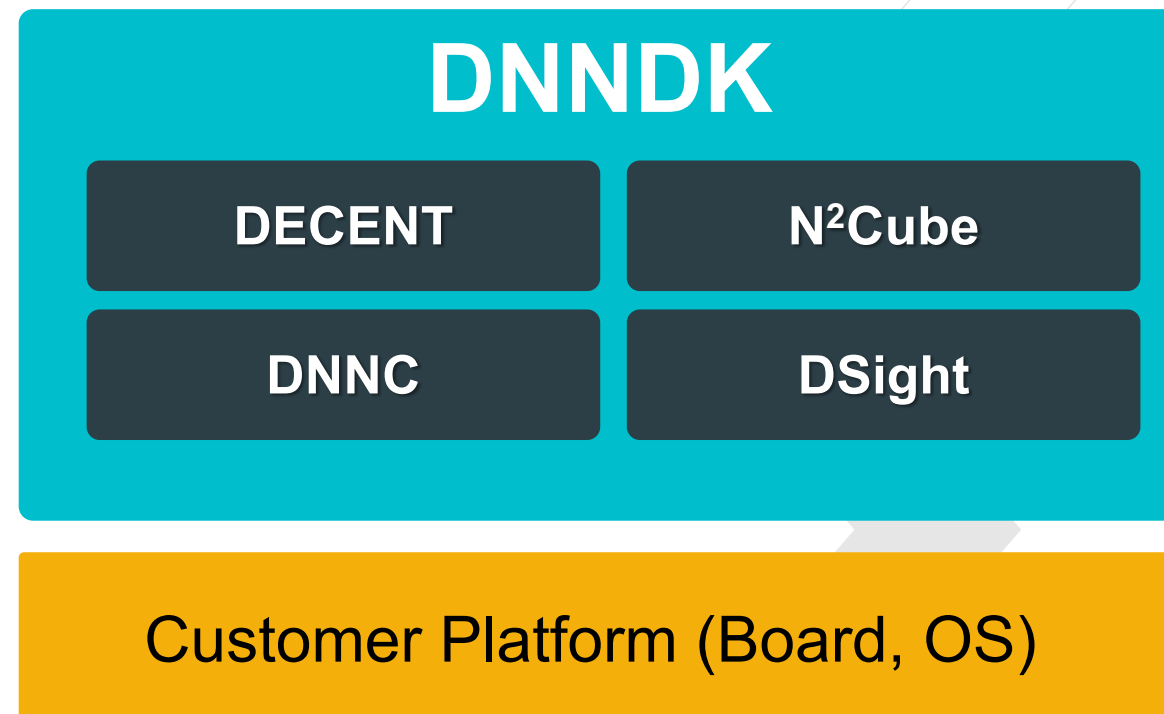
- >> Automatic layer fusion to avoid frequently data read and write

> Runtime N²Cube

- >> Various APIs to facilitate specific application

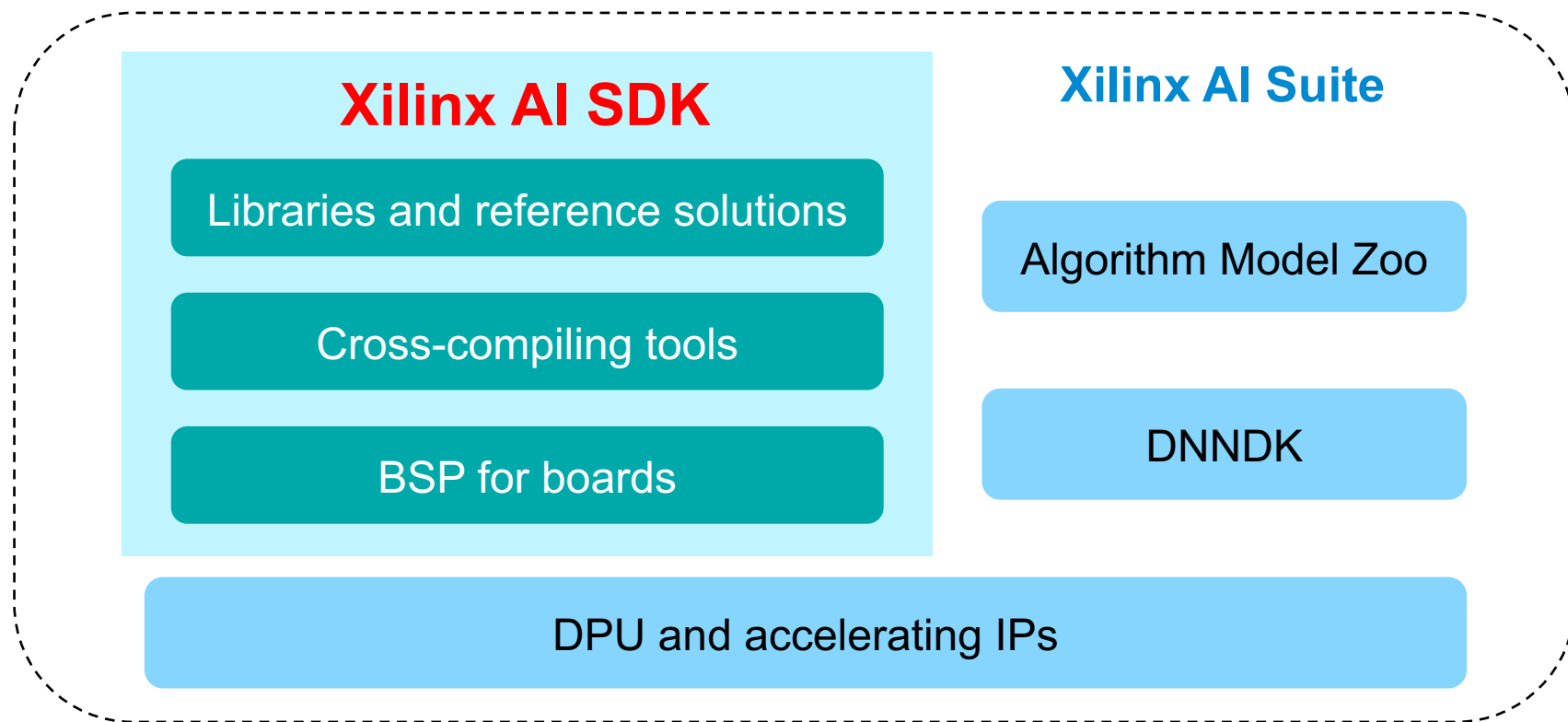
> Profiler Dsight

- >> Powerful tool as failure analysis and optimization

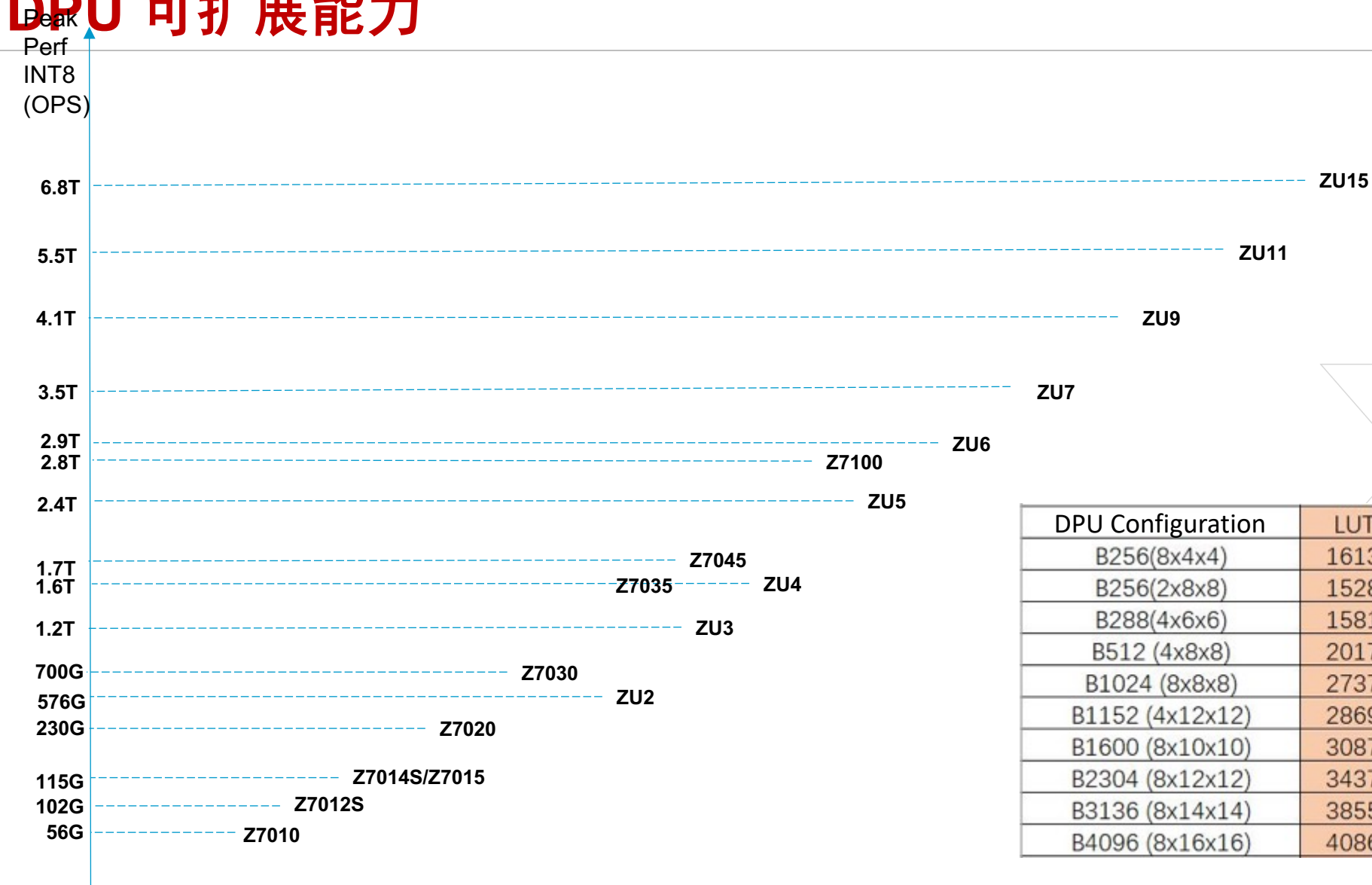


Xilinx AI SDK

- Xilinx AI SDK to enable low touch engagements for customers
 - Encourage top customers to use SDK to build applications and solutions
 - Only use generic SDK release to support low priority customers



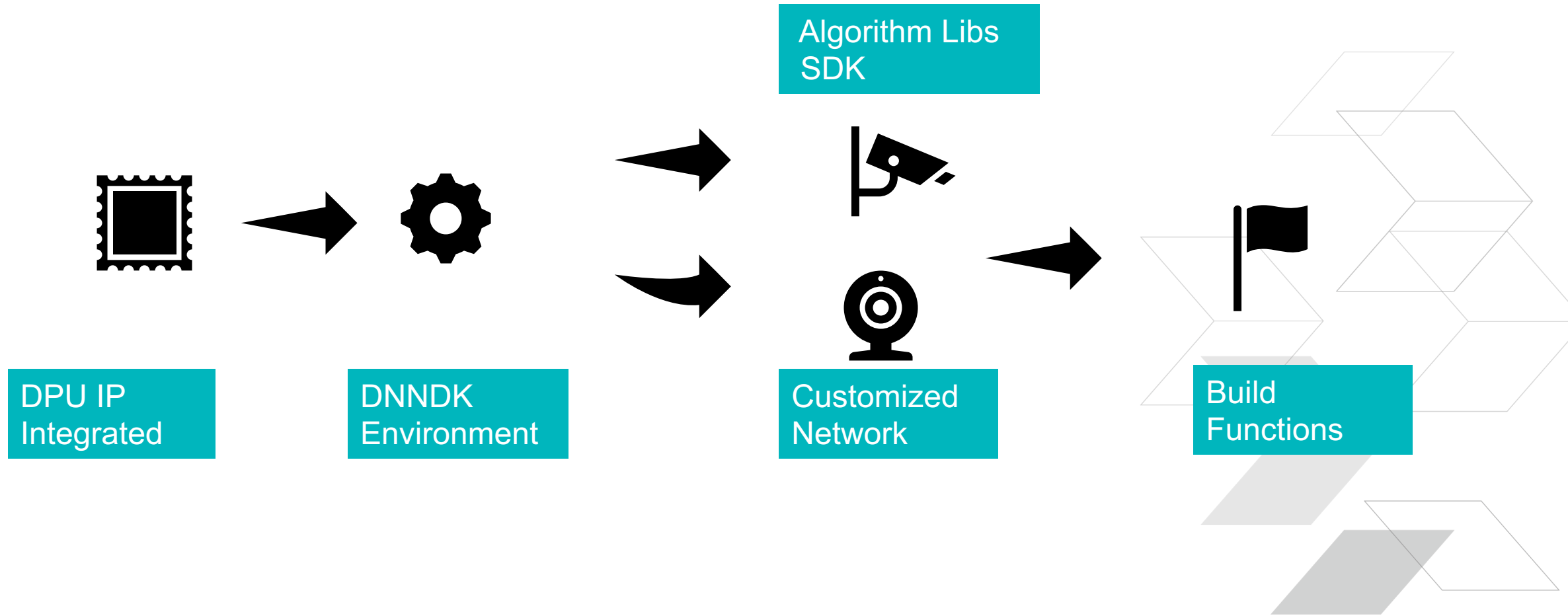
DPU 可扩展能力



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

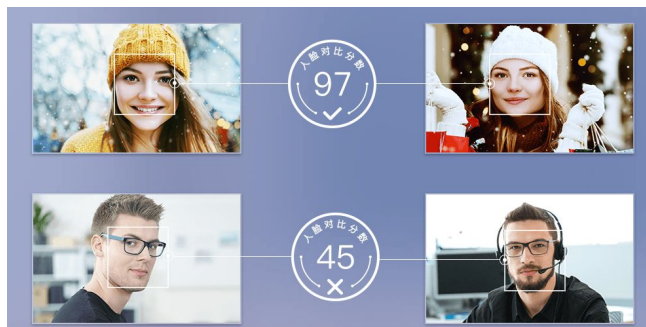
* B256/288/512/3136 work in progress

Develop Flow

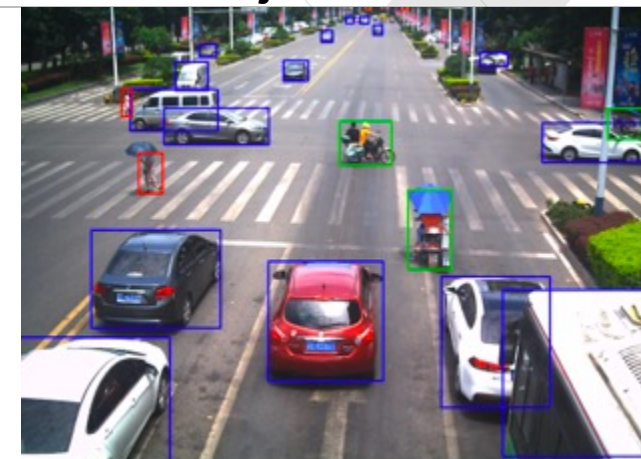


赛灵思人工智能模型库 – 视频分析

Face Det/Recognition



Object Detection



> Object Detection

>> Vehicle, Pedestrian, Cyclist

> Pedestrian Attributes Recognition

>> 5 binary attributes: gender, hat, backpack, hardback, other bag

>> 2 multi-class attributes: upper body color, lower body color

> Car Attributes Recognition

>> Type, Color

> Car Logo Detection

> Car Make Recognition

> License Plate Detection

> License Plate Recognition

> Tracking

>> Single Object Tracking

>> Multi Object Tracking



Pedestrian Attributes Recognition



Car Attributes Recognition

赛灵思人工智能模型库 – 视频分析



Car Logo Detection



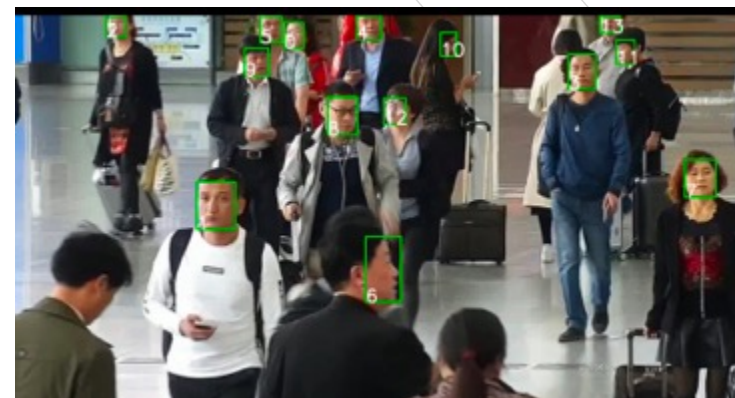
Car Make Recognition



License Plate Detection



License Plate Recognition



Tracking

赛灵思人工智能模型库 – 视频分析

- > Pedestrian Detection
- > Pose Estimation
 - >> Joints: head, neck, shoulder, elbow, wrist, hip, knee, ankle
- > Person Re-identification



Pedestrian Detection



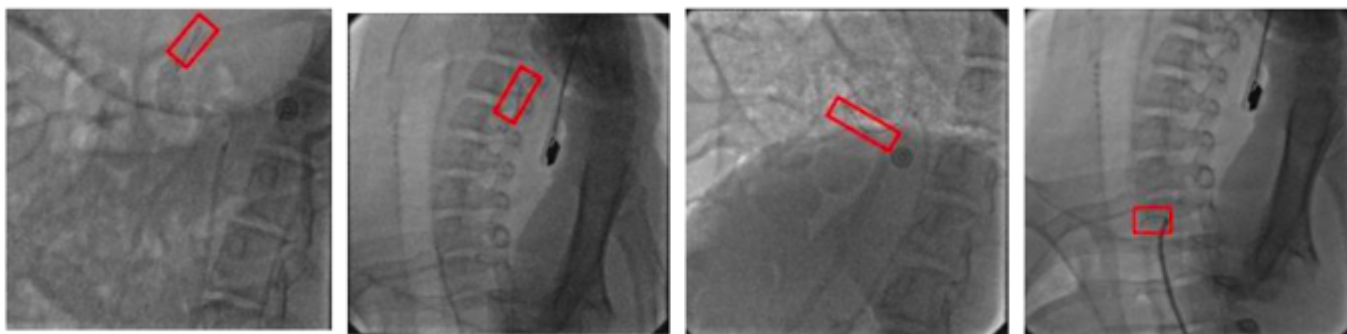
Person Re-identification



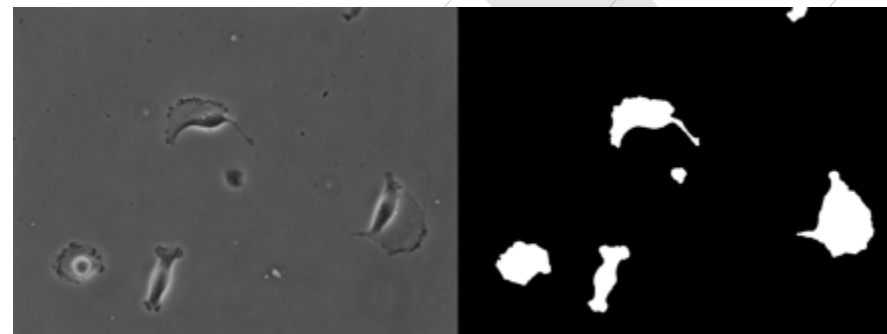
Pose Estimation

赛灵思人工智能模型库 – 医学影像分析

- > **Deep learning can facilitate clinicians and radiologists in diagnosis and treatment planning. Varieties of**
 - >> Modalities: x-ray, CT, PET, MRI, fMRI, ultrasound, microscopy ...
 - >> Target organs/tissues: brain, lung, liver, skin, breast, prostate, hear ...
 - >> Tasks:
 - Classification: Image/exam classification, object or lesion classification
 - Detection: organ, region and landmark localization, object or lesion detection
 - Segmentation: organ and substructure segmentation, lesion segmentation

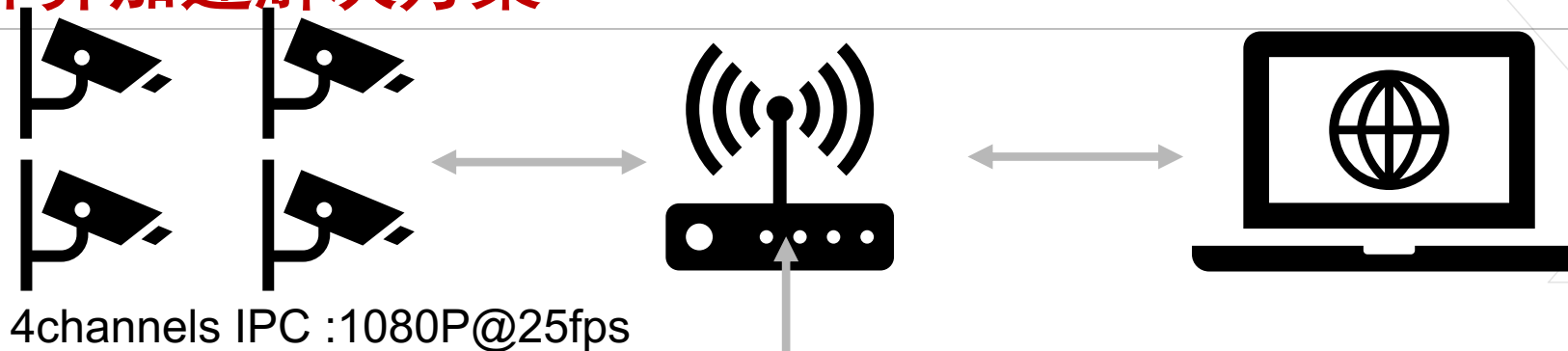


Results of Faster R-CNN in Guide-wire Detection

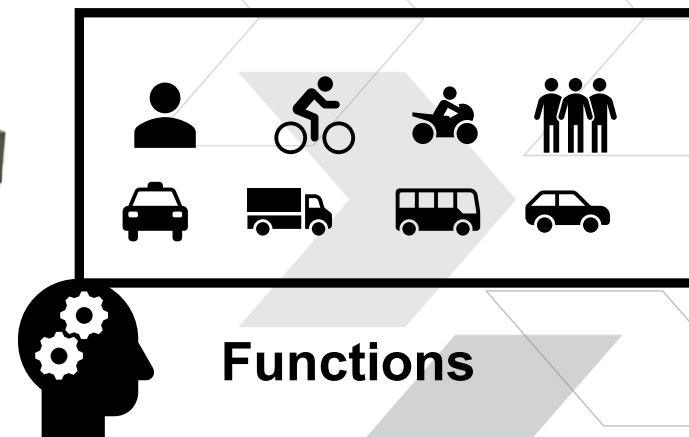
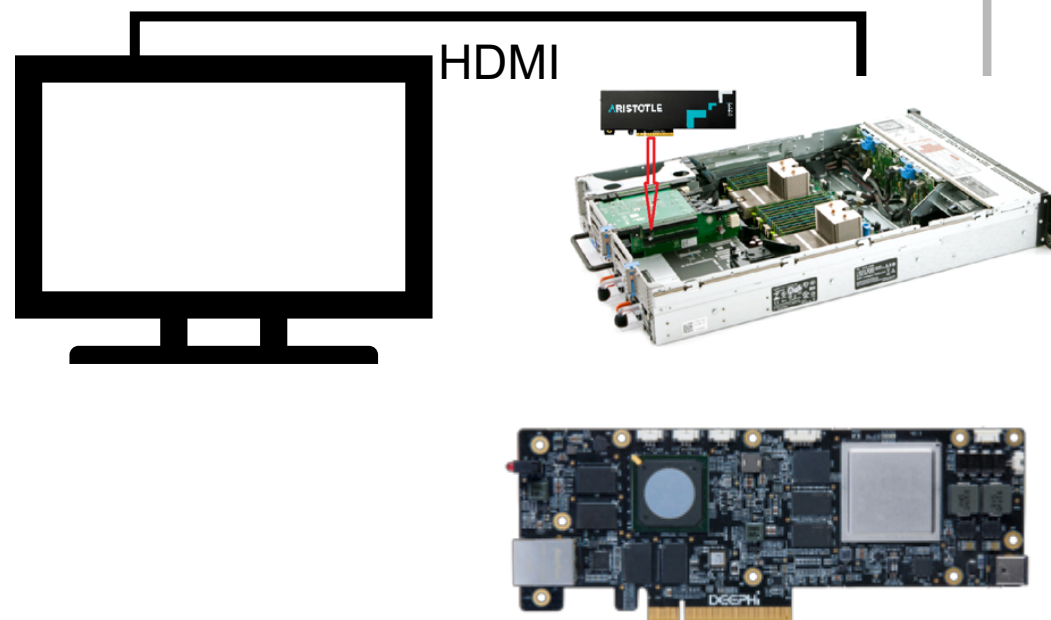


Results of U-Net in Cell Segmentation

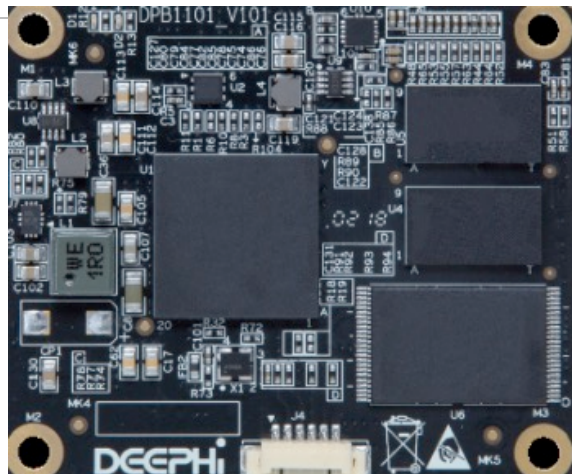
视频分析计算加速解决方案



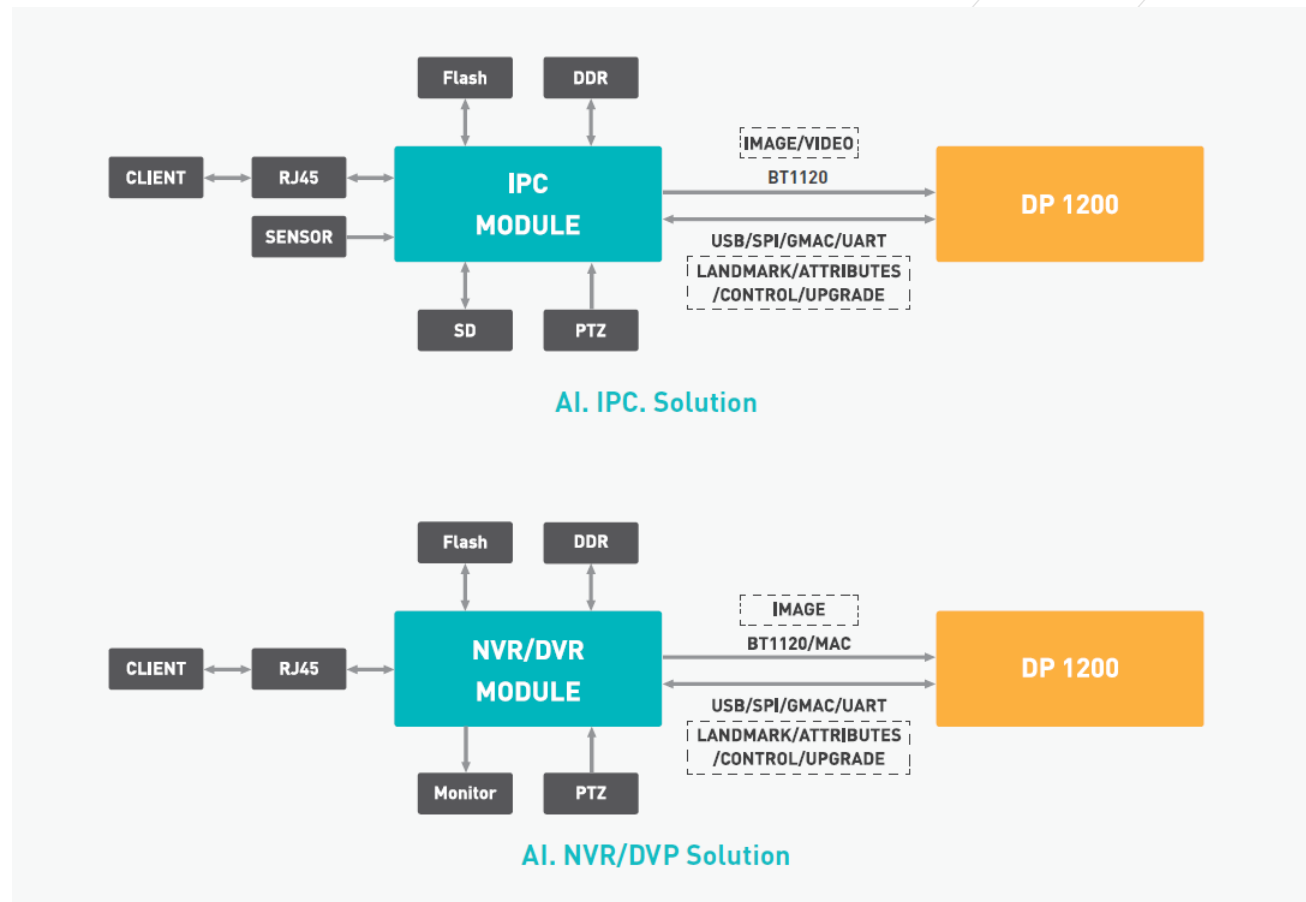
- **Platform:** MPSoC ZU9EG
- **Size:** 179*68.9*14.47 mm
- **DPU:** 2 * B4096
- **Peak perf.:** 4.1Tops (330Mhz)
- **Power:** 21W (whole board)



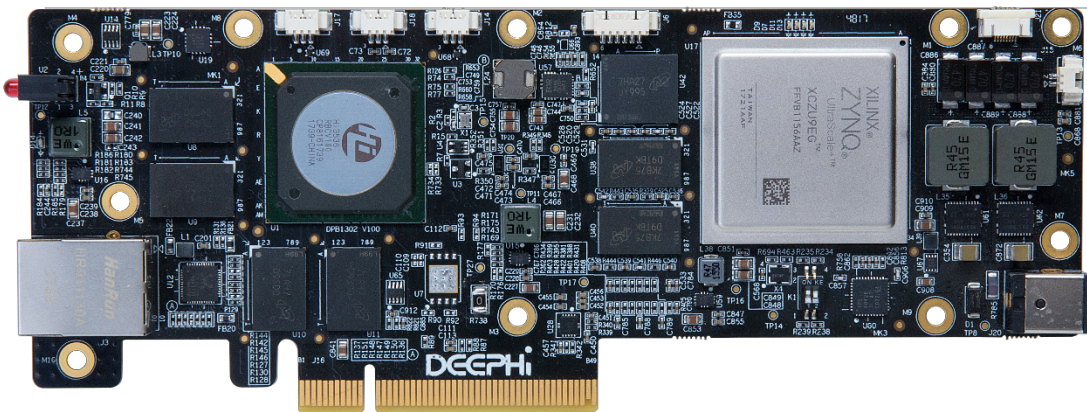
Solution DP-1200



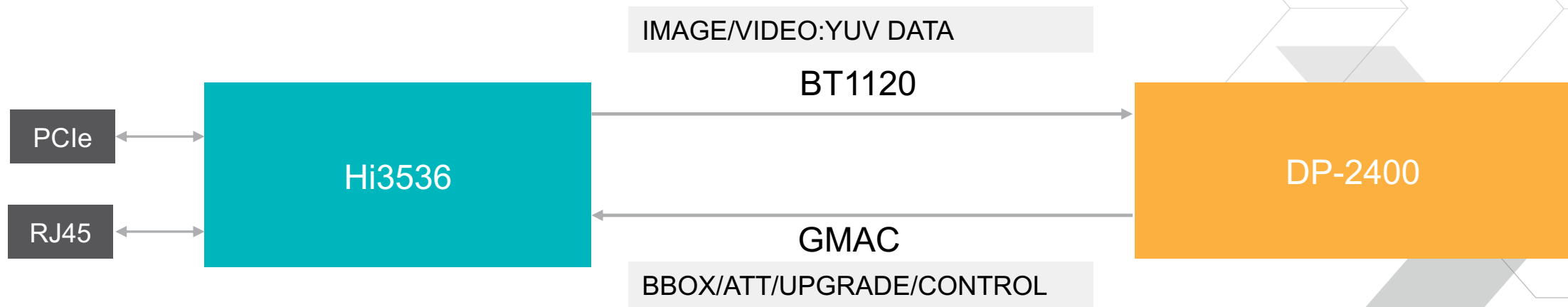
Attributes	References&Values
Face Processing	Up to 30 faces detected and captured per frame
Face Detection	Real-time face detection and bounding box regression
Image Format	JPG/BMP/PNG
Face recognition	1:1 1:N N:N
Interfaces	BT1120/UART/SPI/RJ45



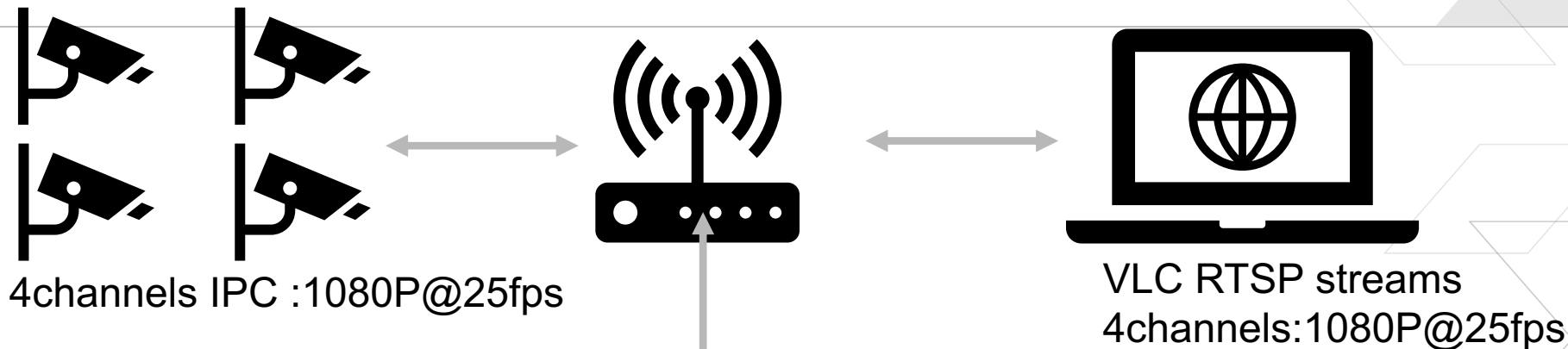
Solution DP-2400



Attributions	Performance&Values
FaceRecognition	9channels 1080P@30fps
ObjectDetection	12channels 1080P@30fps
PhatoAnalytic	126 images per second



8channels on ZCU104



Decoder:

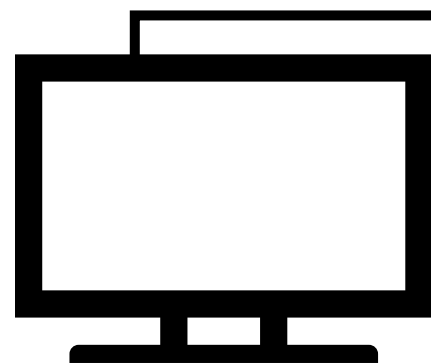
8channels realtime streams
Resolution 1080P@25fps
decoded;

DPU:

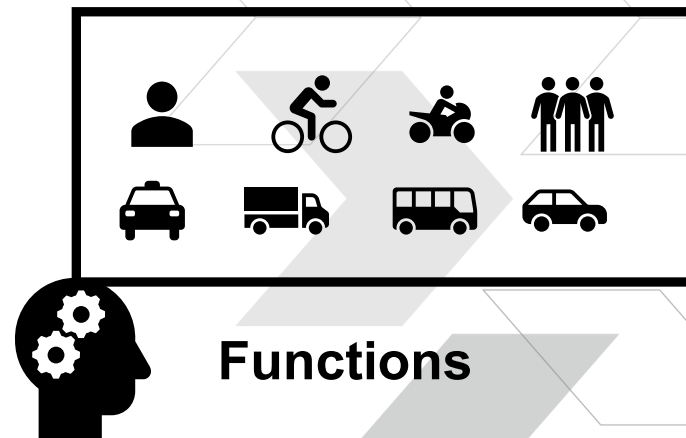
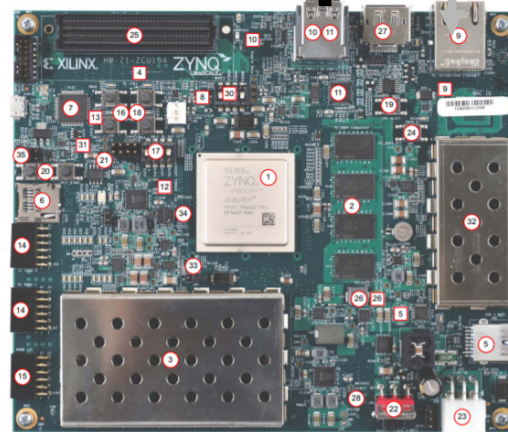
B4096 Single-core

Performance:

4channels for face
detection&&4channels for
traffic detetection;
12fps per channel;



HDMI





赛灵思工业物联网研讨会
XILINX IIoT SEMINAR

SDK Introduction

打造灵活应变、万物智能的世界

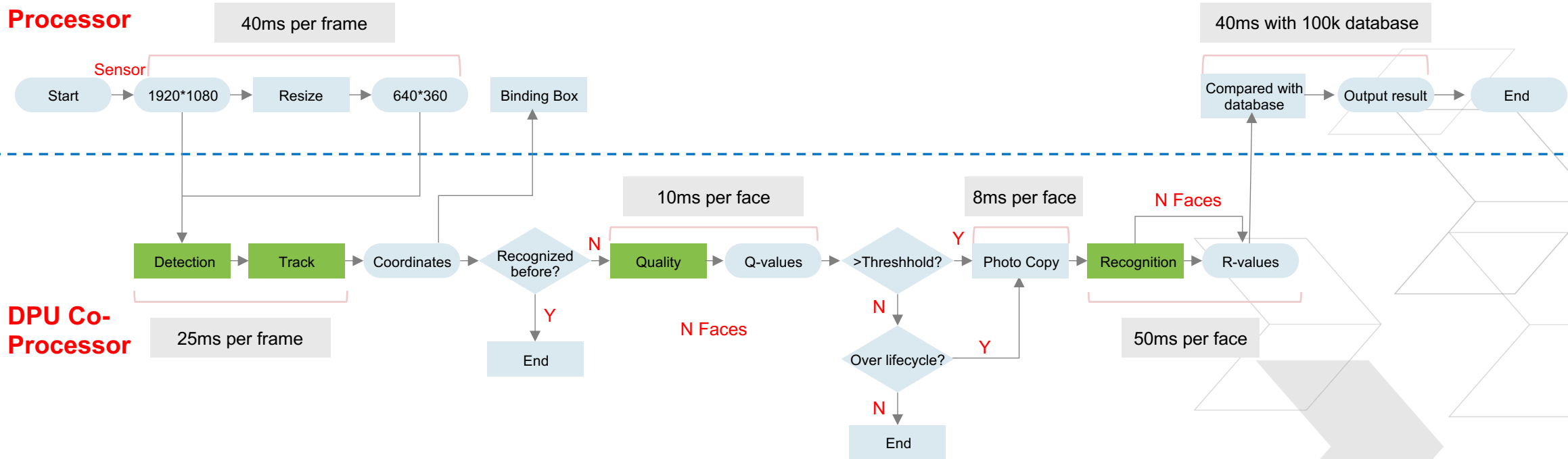


The SDK libs

LibsName	functions	LibsName	functions
lib_dpbase	基础库	lib_dpgraph	图表分析
lib_dpclassification	分类	lib_dpmap	匹配库
lib_dpcaratt	车辆属性	lib_dpmath	算法数学库
lib_dpdetseglane	检测+分割	lib_dpmultitask	多任务库
lib_dpfacetedetect	人脸检测	lib_dpobject	目标检测
lib_dpfacefeature	人脸属性提取	lib_dppedestrianatt	行人属性
lib_dpfacelandmark	人脸关键点	lib_dpplatedetect	车牌检测
lib_dpfacequality	人脸质量	lib_dpplatenum	车牌数字识别
lib_dpfacequality5pt	人脸质量+五点	lib_dpplaterecog	车牌识别
lib_dpfacerecog	人脸识别比对	lib_dprefinedet	Refinedet 检测
lib_dproadline	车道线	lib_dpsegmentation	语义分割
lib_dpsegdet	可行驶区域检测	lib_dpstructuredet	结构检测
lib_dpssd	SSD检测	lib_dpyolov3	Yolo_v3检测
lib_dptracker	跟踪

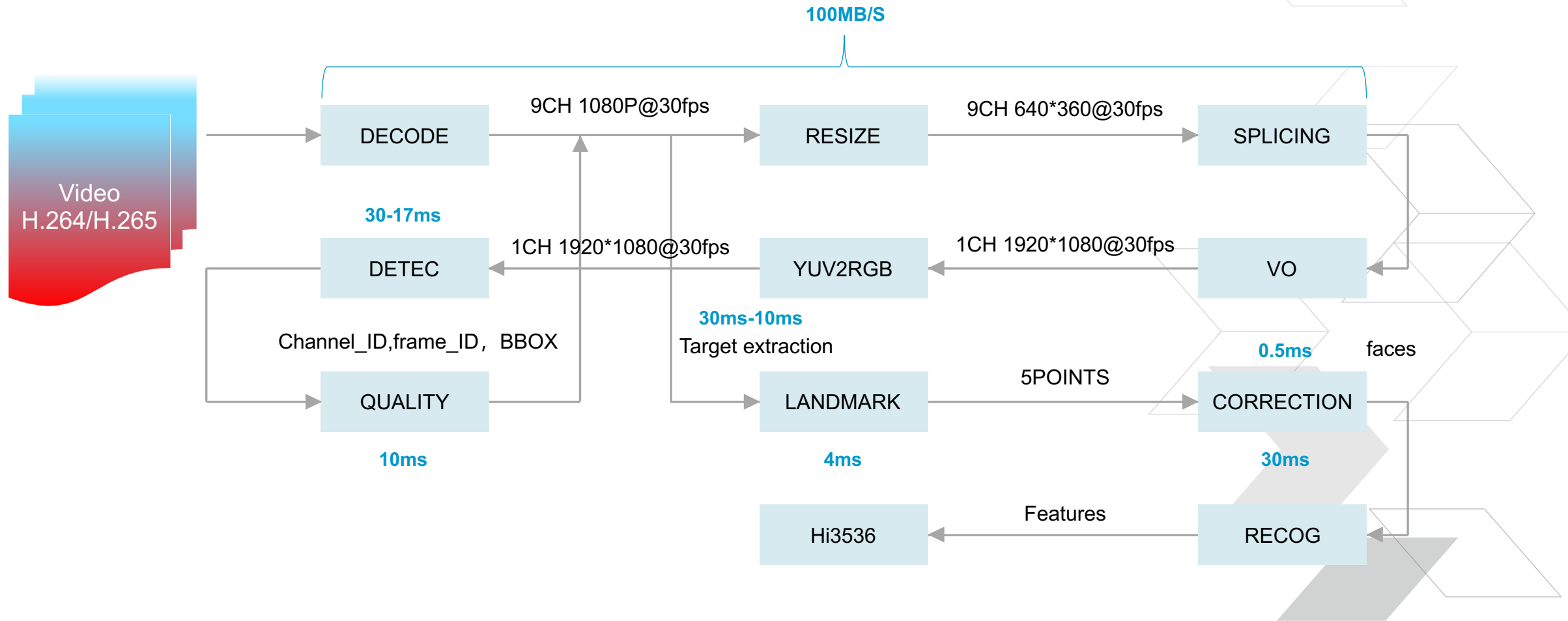
DP-1200 work flow on Face recognition

Main Processor



Using Zynq-7020, the average performance :
10 faces recognition per second

DP-2400 workflow on Face recognition



Adaptable.
Intelligent.



赛灵思工业物联网研讨会
XILINX IIoT SEMINAR

 XILINX®