

FPGA— 智能计算加速引擎

刘竞秀

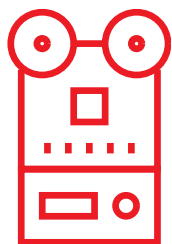
赛灵思 AI 市场开发总监

2019.6



计算的演变

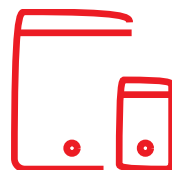
在所有市场中
通过异构架构加速新工作负载的趋势



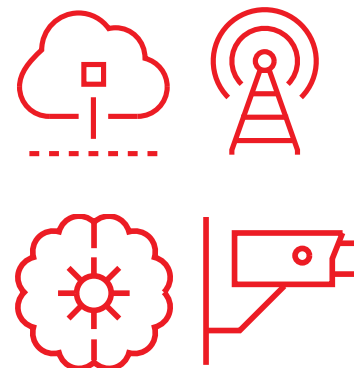
大型机时代
以百万为单位



PC 时代
以亿为单位

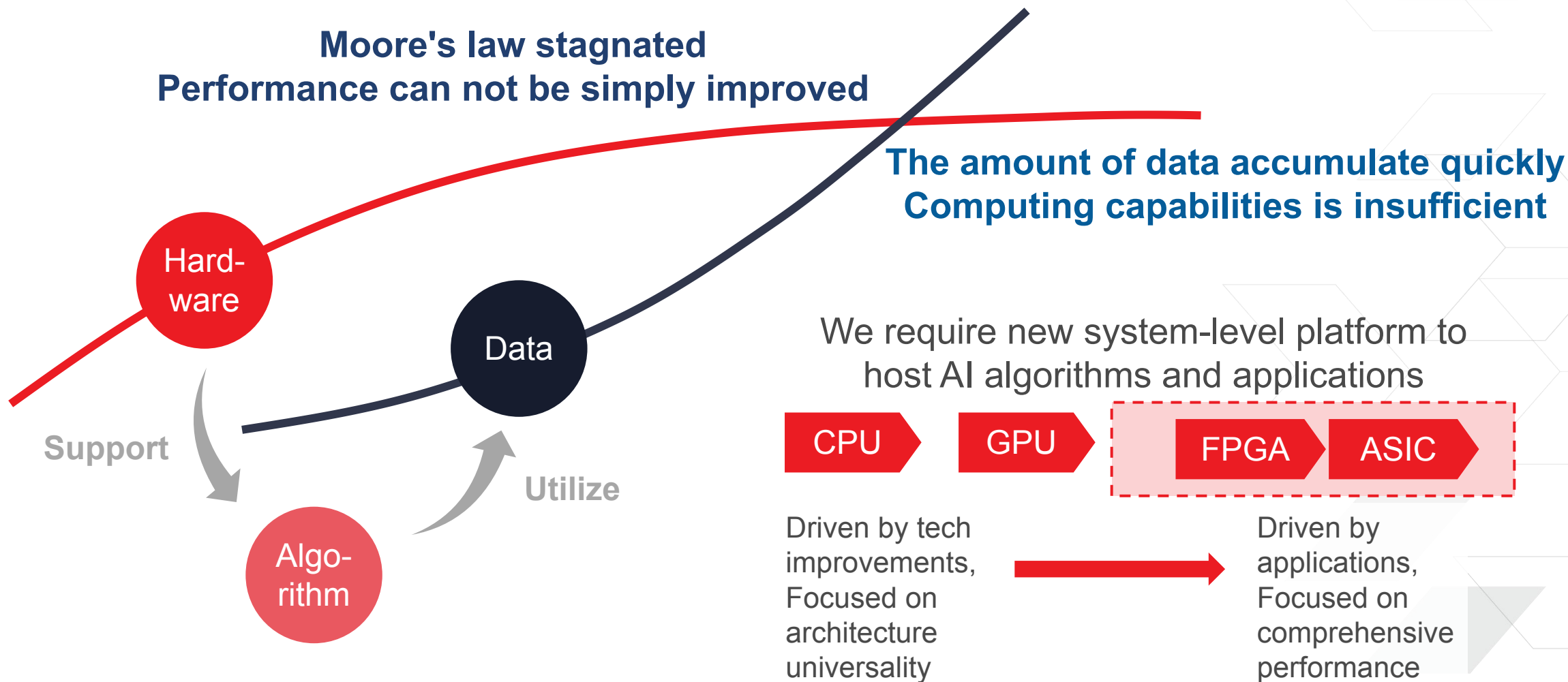


移动时代
以10亿为单位

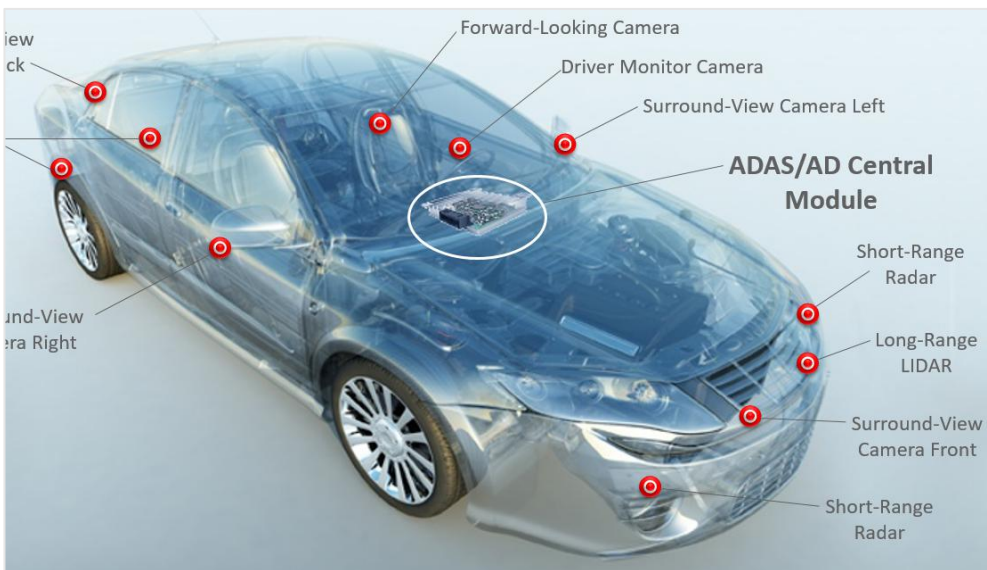


普适智能时代
以500亿为单位

高性能计算平台的发展趋势



商业模式发展趋势 – AI变现能力迅速增强



为什么我们的使命如此重要？

个性化医疗案例

- > 动态优化不同患者专用的基因组分
- > 分析时间从 1 天缩短至 20 分钟
- > 全基因组诊断诊治重症新生儿
- > 医疗数据和研究需要在全球范围内安全地访问

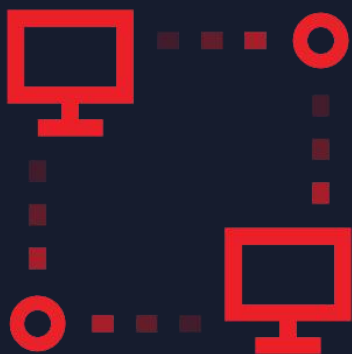


Rady
Children's
Hospital
San Diego



edico  genome

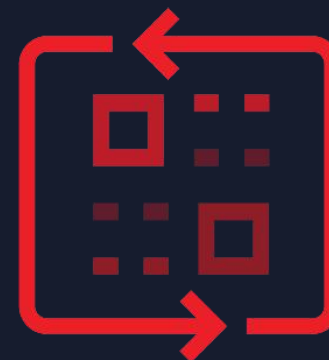
打造灵活应变万物智能的世界



万物智能与万物互联



超大规模领域的部署



动态变化的市场需要快速的创新
和部署

We Are Xilinx

1984 年成立



- 总 部
- 研发中心
- 销售和支持
- 制 造



25.4 亿美元 约 4,200 名

年收入



员工数量



2 万多家

客户数量



60 多项

行业第一



4,000 多项

专利

我们的传承：创新和行业第一的发展史

全球首个无晶圆（Fabless）
半导体企业



全球首款FPGA



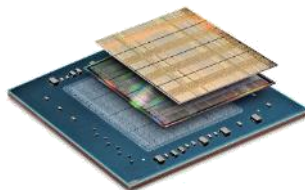
首款将处理器集成在FPGA 中



首个软件硬件都可以编程的 SoC



全球首款 2.5D IC FPGA



全球首个 ASIC 增强型设计套件



行业首款多处理器 SoC (MPSoC)



SDx 开发环境



首款 RFSoc



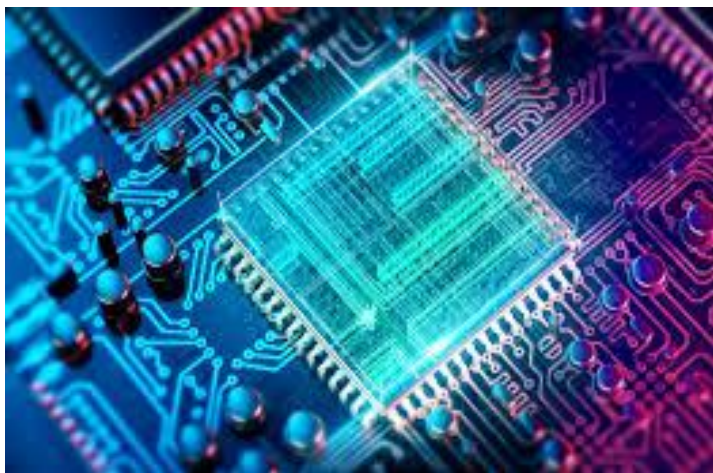
首款自适应计算加速平台 (ACAP)



赛灵思 战略



数据中心优先



计算加速



存储



SmartNIC 与网络加速

推出 Alveo 加速器卡



速度快

比 CPU 和 GPU 更快
超越 GPU 的时延优势



灵活应变

面向所有工作负载而优化
适应不断变化的算法



简单易用

云端及本地皆可部署
应用现已开放



不断扩大的数据中心计算生态系统

应用、工具 与社区

bitfusion

DEEPhi
深 鉴 科 技

edico genome

NGCODEC
NEXT GENERATION VIDEO COMPRESSION

RYFT

云端开发与部署 (FPGA 即服务 , FaaS)

aws

Alibaba Cloud

HUAWEI

Baidu 百度

NIMBIX

Tencent Cloud

技术与 系统

CCIX

AMD

arm

HUAWEI

IBM

Mellanox
TECHNOLOGIES

QUALCOMM

性能与服务器 CPU 对比

真实的 FaaS 用例

机器学习推断
语音识别



40倍

基因分析

全基因组分析：20分钟 vs. 33小时



100倍

视频直播
HEVC编码的帧速率



10倍

大数据分析

日志文件查询：40 分钟 vs. 60 小时



90倍

加速核心垂直市场的增长



Automotive
汽车



Wireless Infrastructure
无线基础设施



Wired Communications
有线通信



Audio, Video, & Broadcast
音频、视频与广播



Aerospace
航空航天



Industrial, Scientific & Medical
工业、科学与医疗



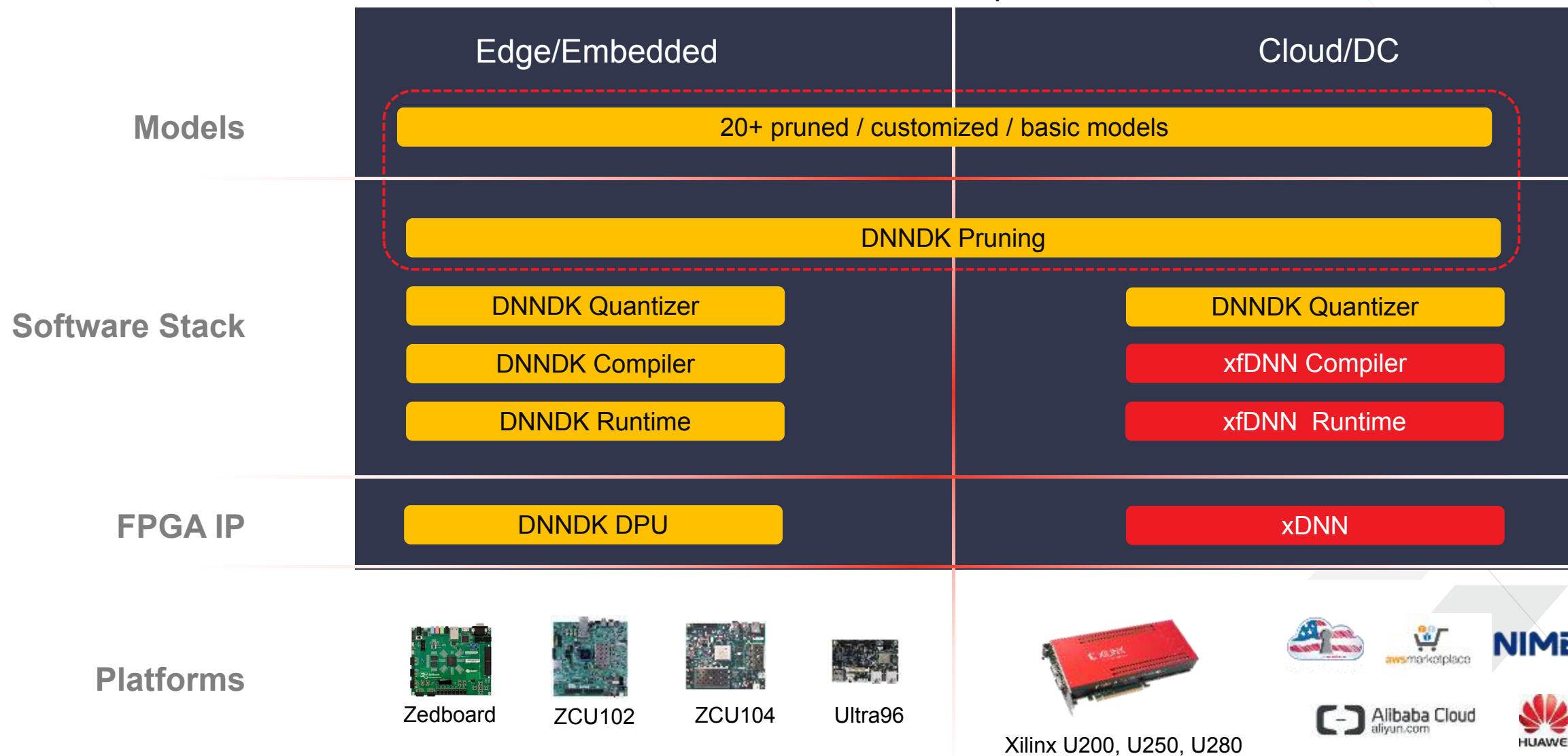
Test & Measurement, Emulation
测试、测量与仿真



Consumer
消费电子

赛灵思人工智能整体解决方案

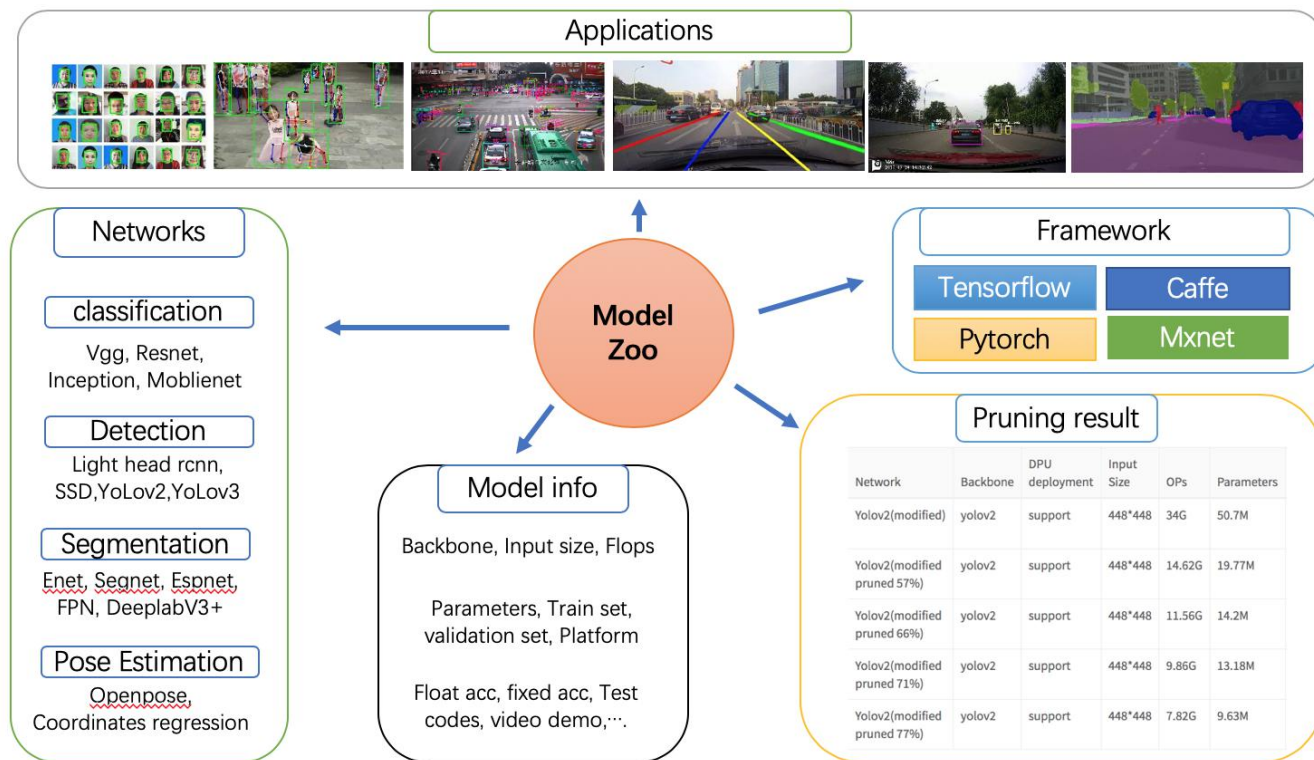
Xilinx AI Development



算法模型集合 (Zoo)

- > a repository to share algorithm models
- > aiming at serving as foundation for AI platform in Xilinx
- > incorporate the most popular and valuable **80+** neural networks
- > public version model zoo: **29** neural networks

Rich information

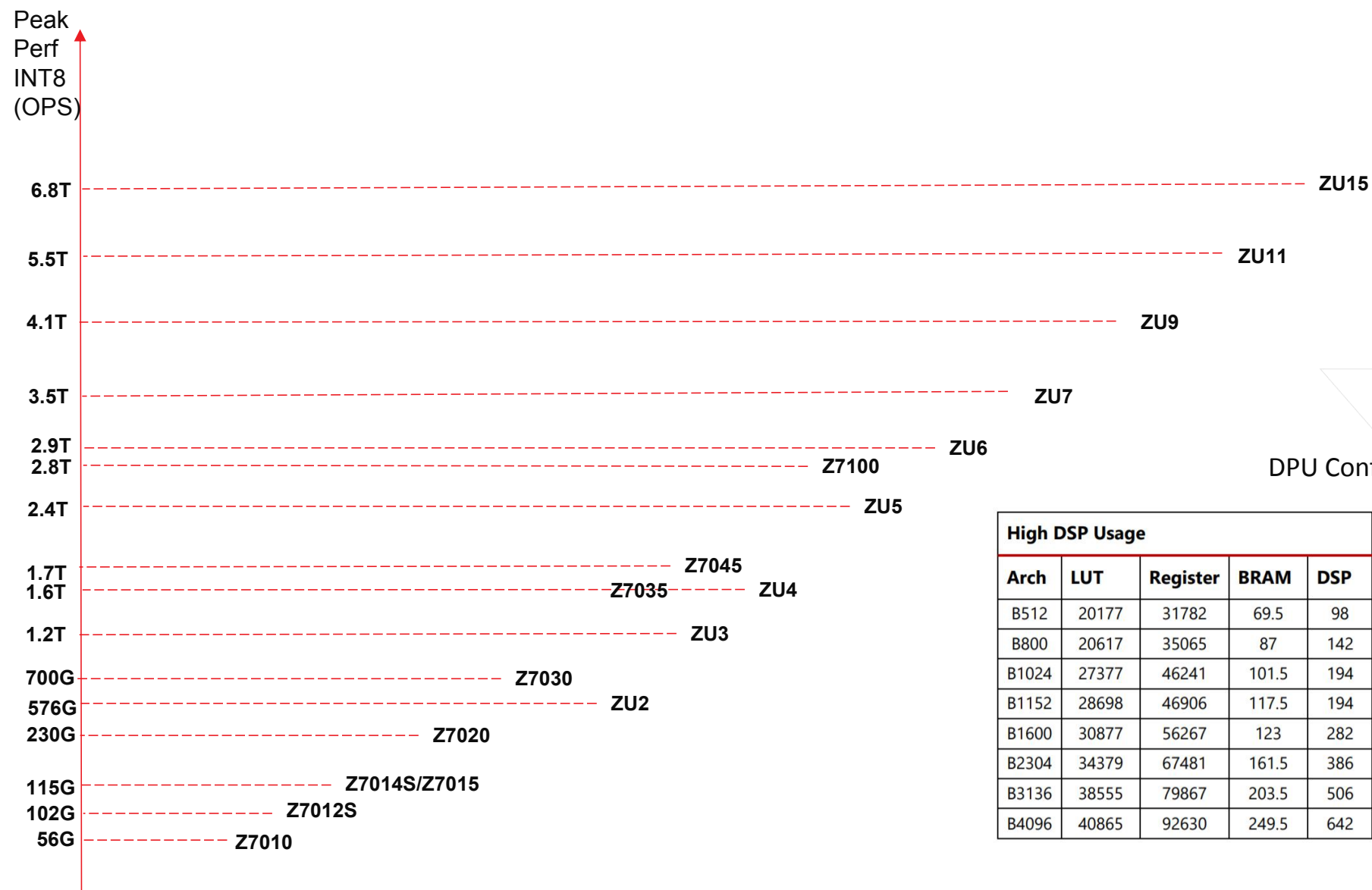


© Copyright 2019 Xilinx

Rich applications and scenarios

Application	Module
Face	Face detection
	Landmark Localization
	Face recognition
	Face attributes recognition
Pedestrian	Pedestrian Detection
	Pose Estimation
	Person Re-identification
Video Analytics	Object detection
	Pedestrian Attributes Recognition
	Car Attributes Recognition
	Car Logo Detection
	Car Logo Recognition
	License Plate Detection
	License Plate Recognition
	Object Detection
ADAS/AD	3D Car Detection
	Lane Detection
	Traffic Sign Detection
	Semantic Segmentation
	Drivable Space Detection

DPU 可扩展能力

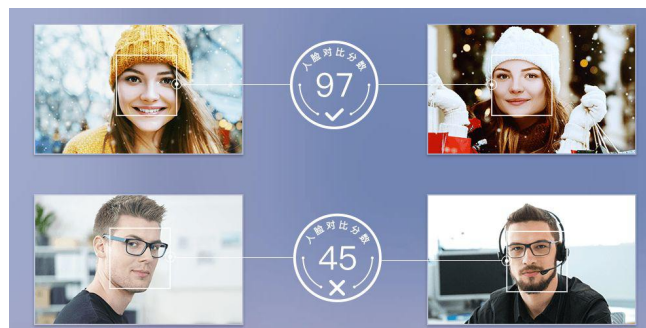


High DSP Usage					Low DSP Usage				
Arch	LUT	Register	BRAM	DSP	Arch	LUT	Register	BRAM	DSP
B512	20177	31782	69.5	98	B512	20759	33572	69.5	66
B800	20617	35065	87	142	B800	21050	33752	87	102
B1024	27377	46241	101.5	194	B1024	29155	49823	101.5	130
B1152	28698	46906	117.5	194	B1152	30043	49588	117.5	146
B1600	30877	56267	123	282	B1600	33130	60739	123	202
B2304	34379	67481	161.5	386	B2304	37055	72850	161.5	290
B3136	38555	79867	203.5	506	B3136	41714	86132	203.5	394
B4096	40865	92630	249.5	642	B4096	44583	99791	249.5	514

* B256/288/512/3136 work in progress

赛灵思人工智能模型库 – 视频分析

Face Det/Recognition



> 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

Object Detection



Pedestrian Attributes Recognition



Car Attributes Recognition

赛灵思人工智能模型库 – 视频分析



Car Logo Detection



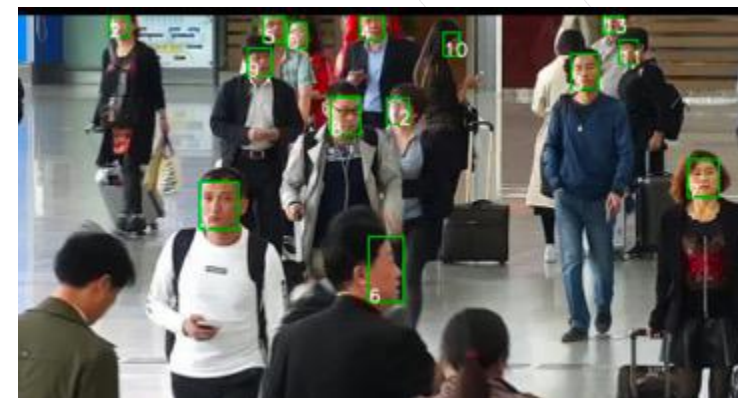
Car Make Recognition



License Plate Detection



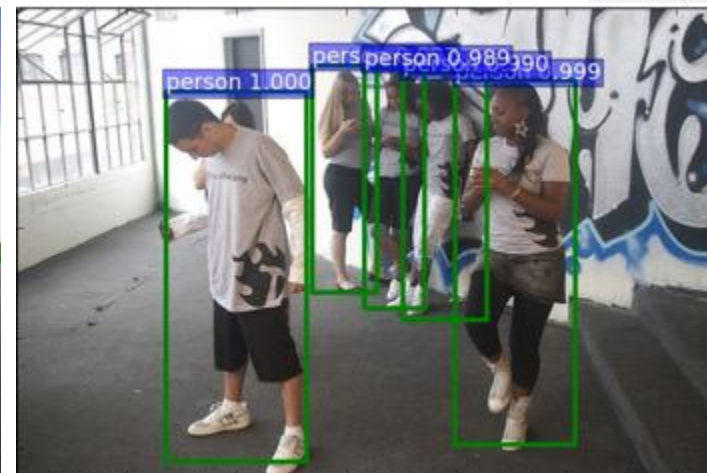
License Plate Recognition



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

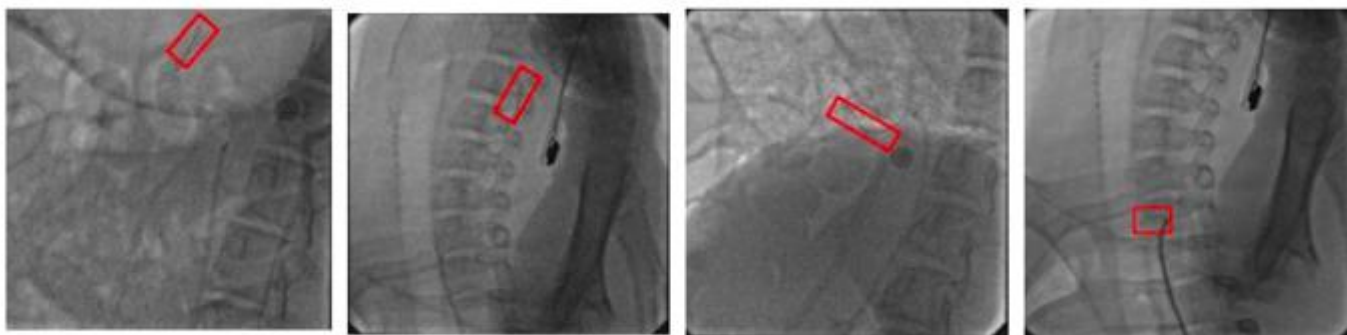


Pose Estimation

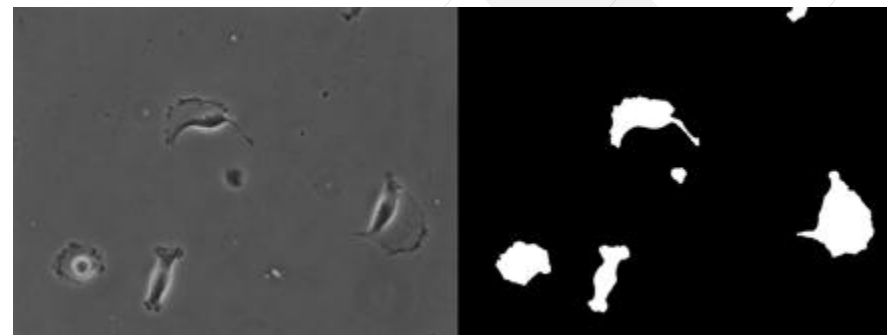
赛灵思人工智能模型库 – 医学影像分析

> 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



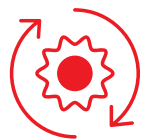
Results of Faster R-CNN in Guide-wire Detection



Results of U-Net in Cell Segmentation

用 Versal 驱动自适应的计算

全球首个自适应计算加速平台



标量处理引擎



灵活应变的硬件引擎



智能引擎
软件可编程,硬件灵活应变



突破性集成的高级协议引擎



赛灵思技术支持灵活的部署选项

器件



量产型开发板



FPGA as a Service (FaaS)



赋能软件开发

加速
开源框架



软件
应用开发者

加速库

Machine
learning



Database
analytics



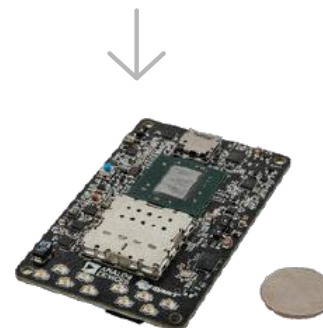
开发堆栈

开发环境



系统
开发者

开发板



A wide-angle photograph of a two-lane asphalt road stretching straight into the distance. The road is flanked by dark, flat land. In the background, a range of mountains with patches of snow is visible under a sky with soft, orange and yellow clouds, suggesting a sunset or sunrise. The overall mood is contemplative and forward-looking.

The Road Ahead

Building the
Adaptable Intelligent World