SSTG
安全产品技术部

滴滴出行

算丰AI芯片的产品设计与安全设计

芯之安全磐石 若广厦之地基

比特大陆 汤炜伟

共建安全新生态
2020滴滴网络安全峰会

2020.12.17

目录

01 比特大陆简介

01 比特大陆公司历程

经过七年成长的高性能计算芯片设计企业：
掌握7纳米和5纳米芯片设计能力的全球几家公司之一

区块链业务-蚂蚁矿机

全球市场占有率第一

人工智能业务——算丰

AI芯片和产品 / 深度学习专用TPU

2013

公司成立

推出55nm芯片

2014

推出28nm芯片

ANTPOOL矿池上线

2015

AI业务算丰启动

BTC.com矿池上线

2016

推出16nm芯片

2017

第一代AI芯片

年营收突破

25亿美元

2018

首款7nm芯片

第二代AI芯片

2019

第二款7nm芯片

第三代AI芯片

01 比特大陆算丰AI芯片

高算力，大带宽

- 8核A53，主控能力提高2~3倍
- INT8算力17.6T / 35.2T

强劲的视频编解码能力

- 最高支持38路1080P解码，提升4倍
- 视频前后处理能力提升4倍

BM1684

接口丰富，超低功耗

- 支持双千兆以太网口
- 功耗较上代产品降低50%以上

安全可靠，完美保障

- 新增安全引擎，保证数据安全、算法模型安全、设备安全

BM1682
第二代面向云端与边缘应用的AI推理芯片

BM1880
第一代面向终端应用的AI推理芯片

2016
组建AI芯片团队

2017
BM1680
工艺制程 28nm
峰值性能 2T FP32

2018
BM1682
工艺制程 28nm
峰值性能 3T FP32
视频解码 8路1080P硬解码

BM1880
工艺制程 28nm
峰值性能 1TOPS@INT8
视频解码 H.264解码

2019
BM1684
工艺制程 12nm
峰值性能 17.6T/16W
视频解码 32路1080P硬解码

2021
BM1686

02 AI芯片的设计初心

02 AI芯片的设计初心

Performance & Power	Application & Easy use	Lower cost & Safe
算力强、功耗适合 场景全覆盖 云端和边缘	机器视觉全算法支持 GPU迁移代价低 客户研发投入小	产品性价比优 客户数据安全 算法安全

02 全产品矩阵：云和边缘全覆盖

PeaceNet
算力云管理平台

算力资源池
算力调度 / 弹性伸缩

算法资源池
算法适配 / 算法仓库

应用网关
运营管理

BMNNSDK
一站式开发工具链

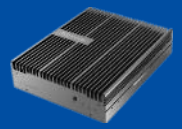
Compiler 编译器
BMNetC / BMNetT / BMNetP / BMNetM / Paddle-Lite

Runtime 运行时部署工具
BMRuntime Engine / BMCV / OpenCV /FFMpeg / BMLib

边缘

AI计算模组 SM5
算力 17.6TOPS@INT8

AI单芯加速卡 SC5H
算力 17.6TOPS@INT8



AI计算盒 SE5
算力 17.6TOPS@INT8

AI计算盒 SE3
算力 3TFLOPS@FP32

云

AI三芯加速卡 SC5+
算力 52.8TOPS@INT8

SC5+ AI服务器系列
X86平台，国产飞腾FT平台
一机多卡

AI超级算力中心
AI推理服务器集群
P级超强算力

BM1684
12nm 17.6TOPS@INT8

BM1682
28nm 3TFLOPS@FP32

02 AI芯片的设计：ASIC专用架构，算力、功耗、解码优于GPU

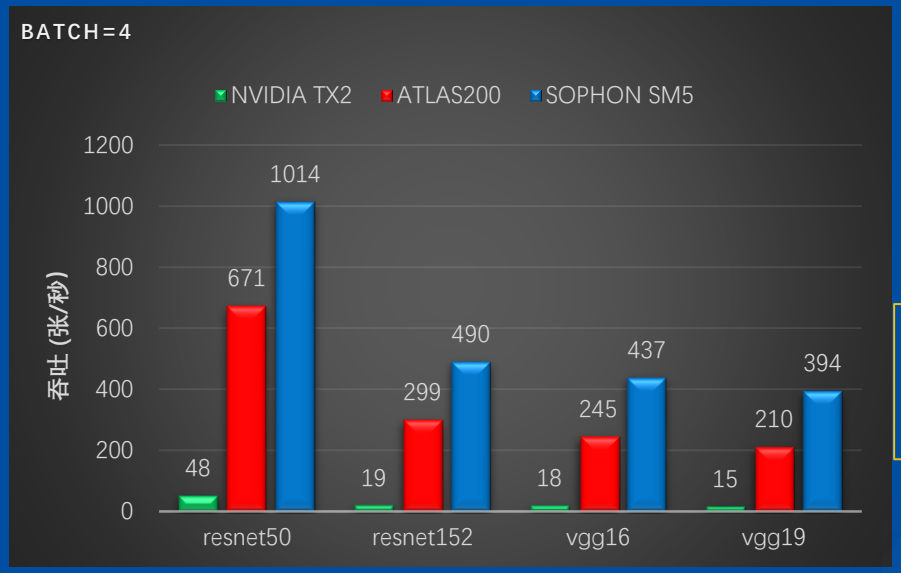
高算力，大带宽

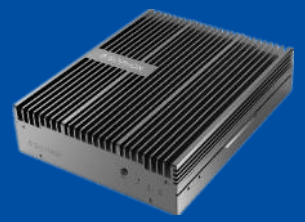
- 8核A53，主控能力提高2~3倍
- INT8算力17.6T / 35.2T

强劲的视频编解码能力

- 最高支持38路1080P解码，提升4倍
- 视频前后处理能力提升4倍

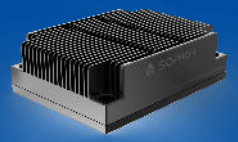
- ResNet超过GPU **20倍**
- VggNet超过GPU 25倍
- 能量效率超过旗舰 GPU **20倍**



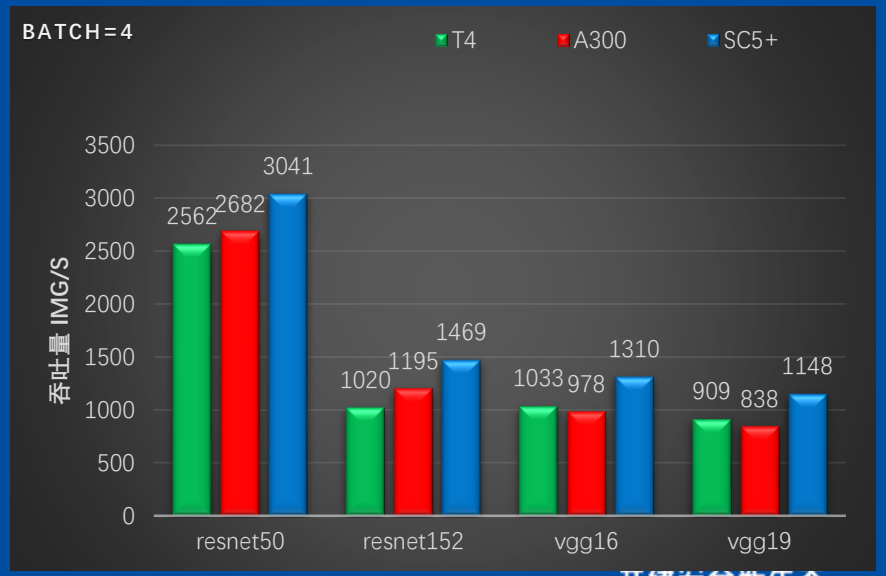


同等的被动散热边缘盒/模组（<20W）

同等的的半高半长板卡（<=75W）

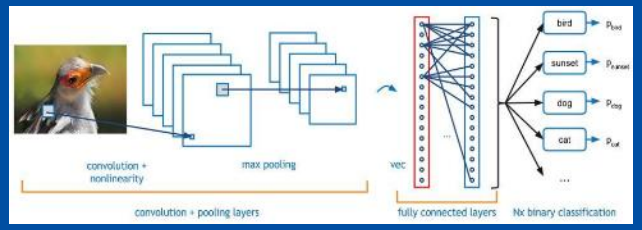


- ResNet超过旗舰 GPU **10%-32%**
- VggNet超过旗舰 GPU **8%-41%**
- 能量效率超过旗舰 GPU **70%**



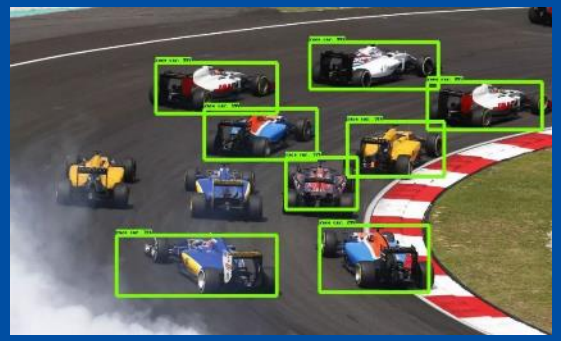
02 AI芯片的设计初心：机器视觉类AI算法全覆盖

图像分类



典型模型
- AlexNet、GoogLeNet、VGGNet、RESNET、DenseNet、mobilenet ……

目标检测



典型模型
- YOLO、SSD、Fast R-CNN ……

语义分割

语义分割：在语义上理解每个像素的角色

典型模型
- Deeplab，FCN、SegNet ……

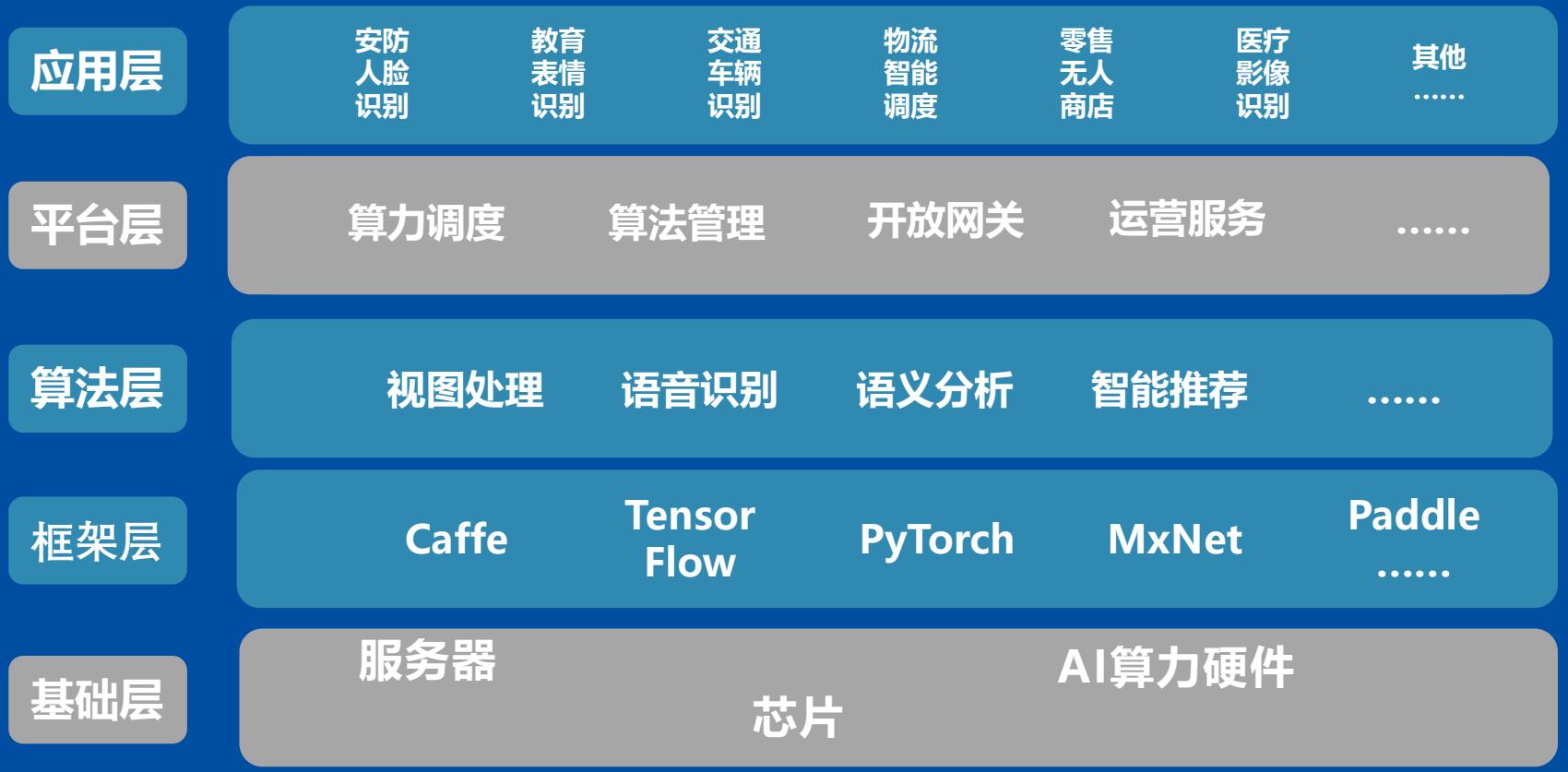
实例分割

语义分割的基础上，实例分割将不同类型的实

典型模型
- Mask R-CNN、Deeplab ……

行业/技术	图像分类	目标检测	语义分割	实例分割
安防	人体车辆属性分类	人脸检测、目标检测	行为姿态检测	行为姿态检测
互联网	电商图像识别、OCR	内容审查	视频广告	视频广告
消费电子	手机相册分类	手机相册分类	手机美颜	抠图、电影效果
汽车	人车分类、红绿灯	人车检测	可行驶区域感知	可行驶区域感知
医疗	病灶标注	病灶识别与标注	CT、X光辅助诊断	CT、X光辅助诊断

02 AI芯片的设计初心：框架完善支持、易开发易迁移



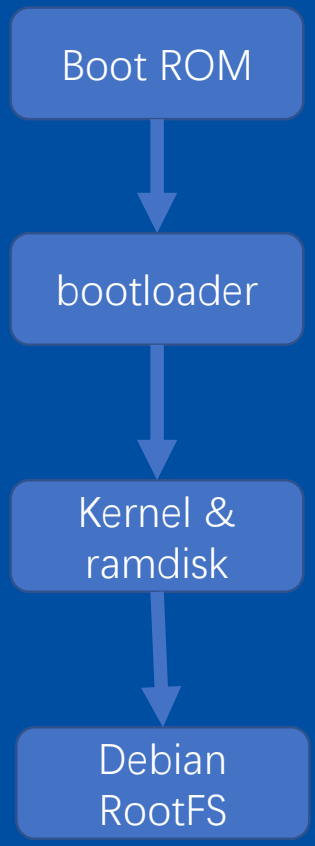
03

算丰AI芯片的安全六盾

03 最新一代AI芯片：BM1684的安全六盾

- Secure boot
- Crypto engines
- Secure key for customer
- Trust zone
- Secure firewall
- Chip unique ID

03 构建可信的系统：Secure boot 验证启动



片内固化
ARM trust firmware framework

- 从efuse读取root key digest，从SPI flash读取bootloader镜像和root key
- 用root key digest验证root key的合法性
- 用root key验证bootloader镜像的合法性

授信启动
U-boot verified boot framework

- bootloader镜像里包含了bootloader的code和public key
- 读取kernel + ramdisk的镜像
- 验证镜像的合法性

授信启动
Linux dm-verity

- ramdisk里包含了eMMC上根文件系统分区的root hash
- 采用了kernel的dm-verity机制，从root hash出发，对根文件系统做校验

验证启动

- 可采用非对称加密方法，对应用签名进行合法性验证

01 构建可信的系统：Trust zone 可信区

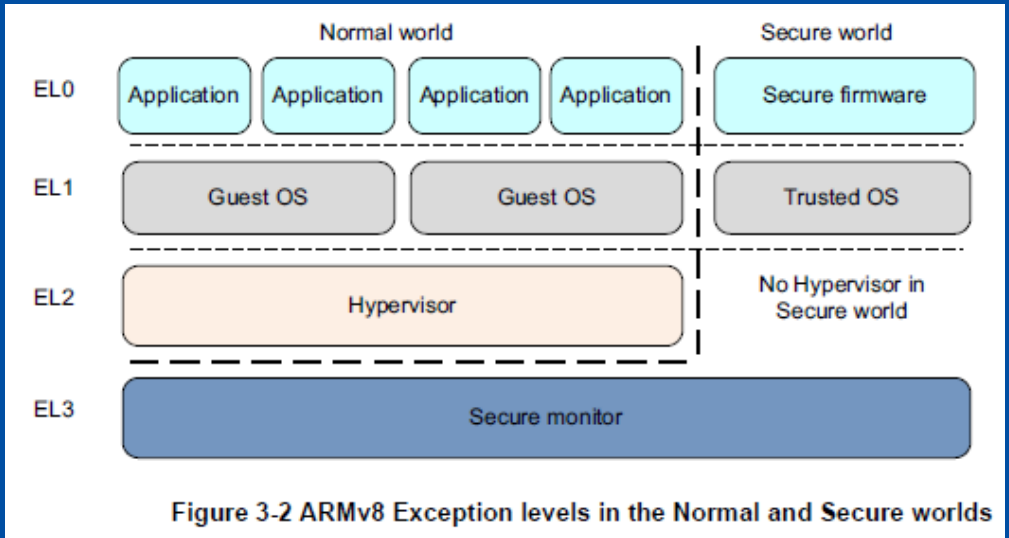
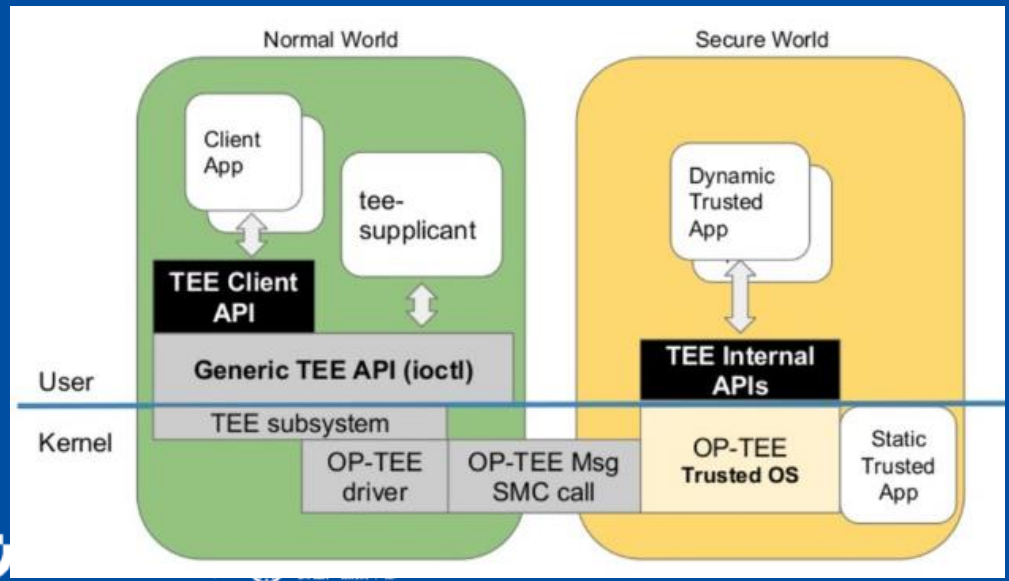
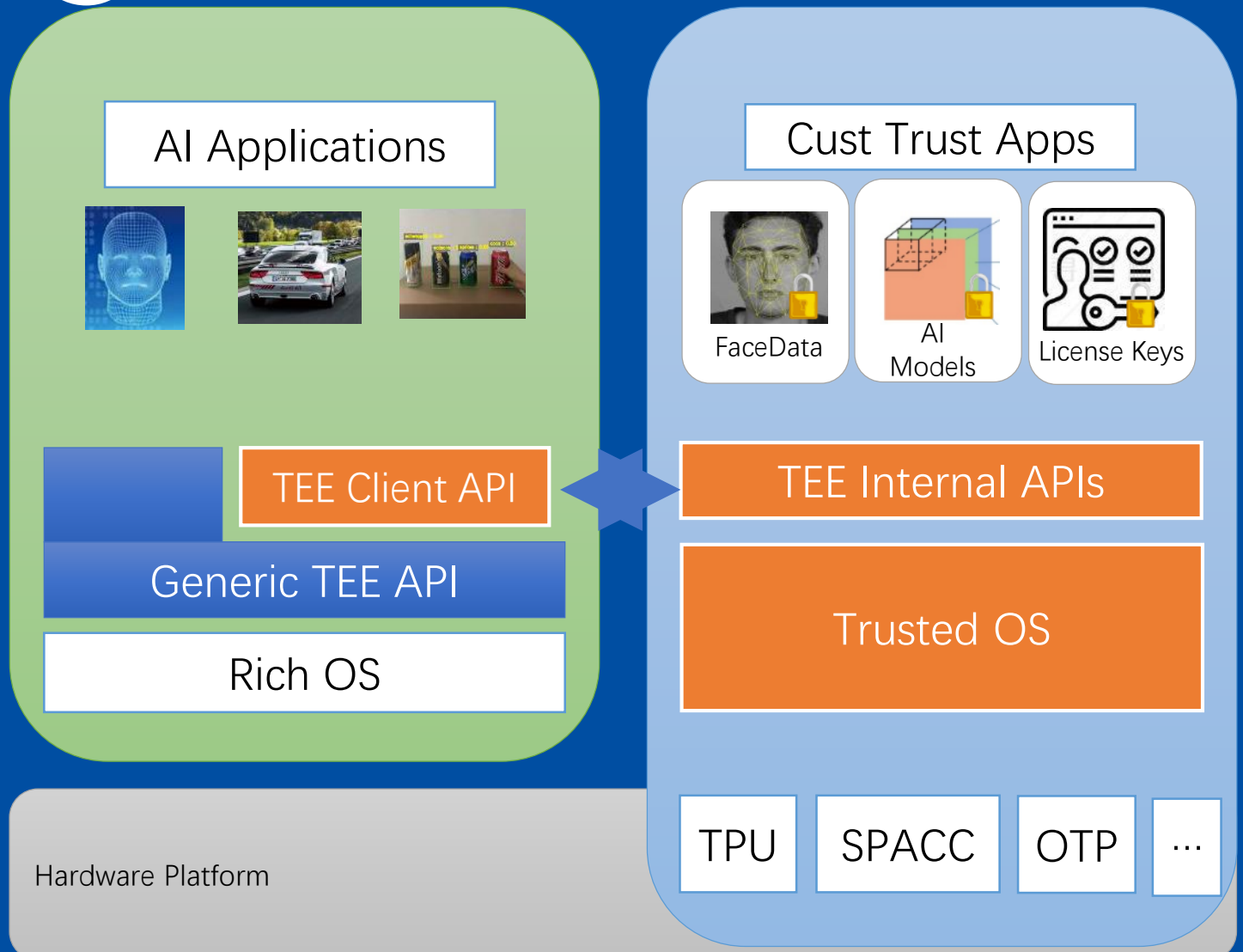


Figure 3-2 ARMv8 Exception levels in the Normal and Secure worlds



- TrustZone提供一套高效系统安全保护硬件架构，可以有效的抵御各种可能的攻击
- 它将SOC的资源划分为安全(secure world)和非安全(normal world)两个世界，关键的高敏感数据在安全世界执行，其它操作在正常世界执行
- 从芯片ROM开始，逐级建立信任链，每一步启动都需要最高特权级别执行和密钥验证，防止软件篡改替换
- BM1684遵循TrustZone设计规范进行了完整的实现

01 构建可信的系统：Secure Firewall & Secure key & Chip unique ID



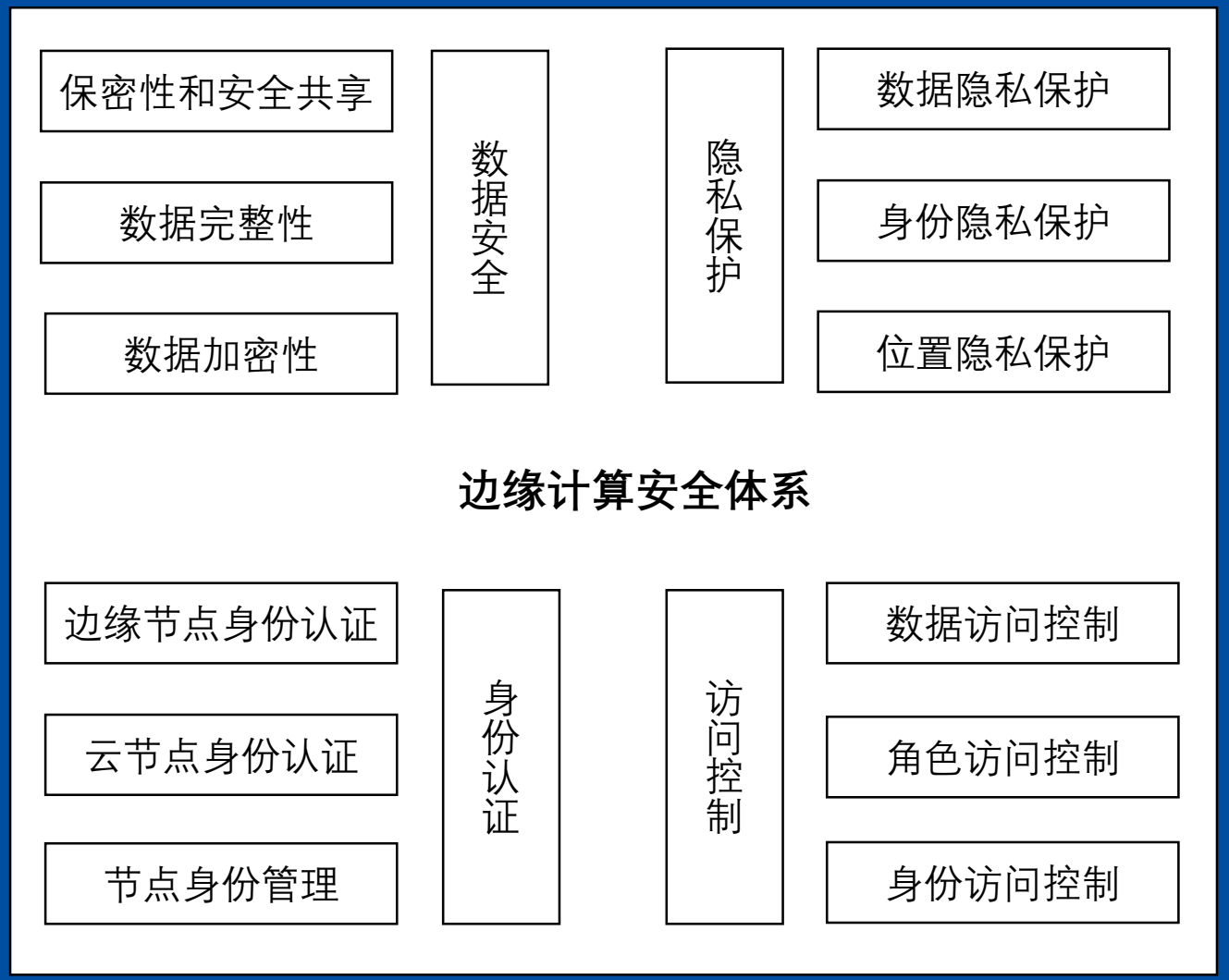
- 应用密钥
  - 客户根钥烧写至OTP区域，采用secure mode保护，仅供硬件加解密引擎读取
  - 加密后应用密钥由normal world传入secure world解密，用于相应数据的解密
- 数据：
  - 加密存储在非易失设备中(eMMC/SDD/HDD)
  - 使用时加载至secure world解密后使用
- 算法模型：
  - 加密存储在非易失设备中或由网络下发
  - 通过TA加载至secure world解密运行

01 构建可信的系统：Crypto Engines

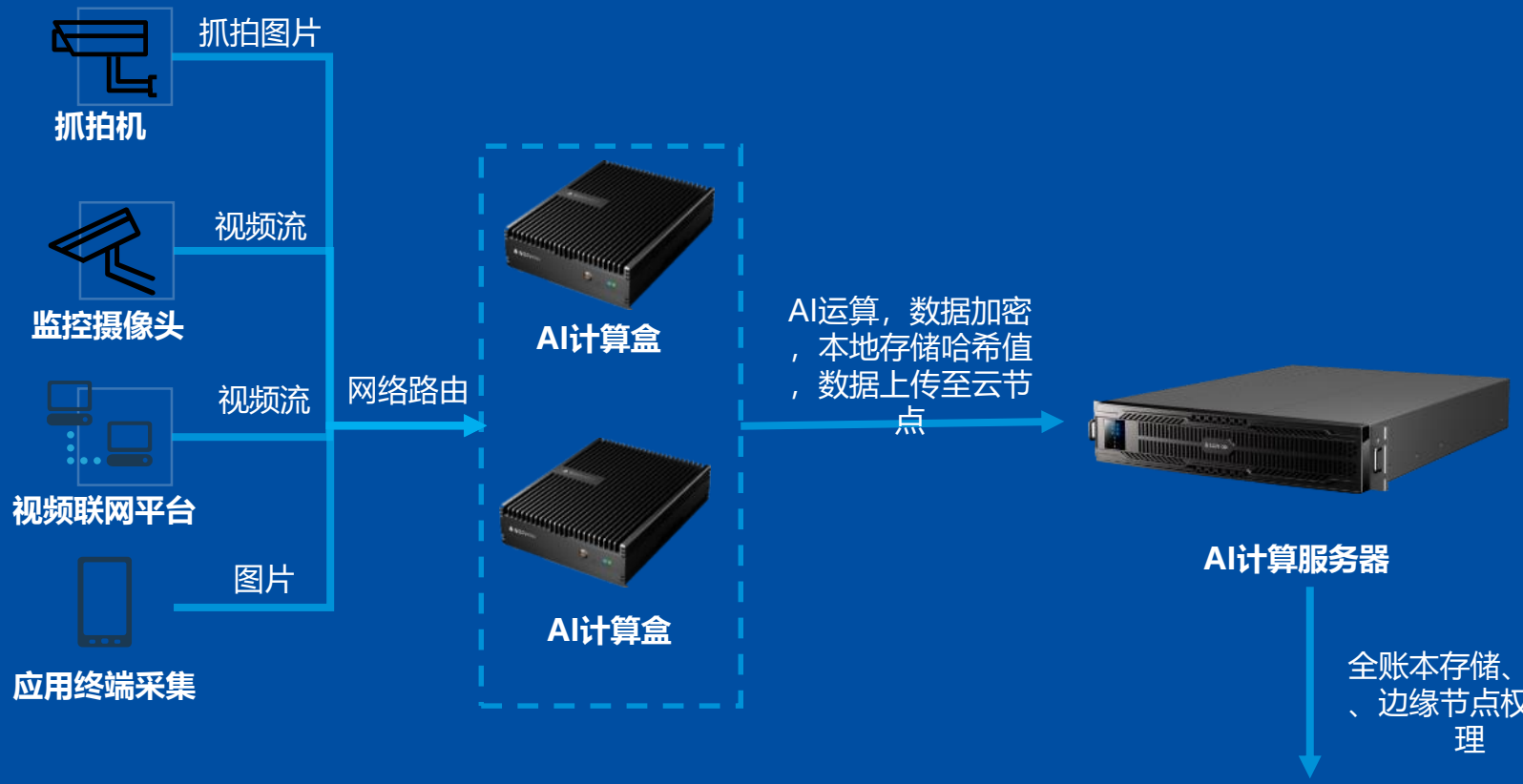
- Support crypto engines：
  - AES/DES;
  - SM4/SHA;
  - RSA/ECC;
- All above hardware accelerator
  - SPACC enabled one way read
  - Software (cpu) accelerator can read both ways not secure
- Secure key storage scheme support

04 区块链设想

01 边缘计算安全体系

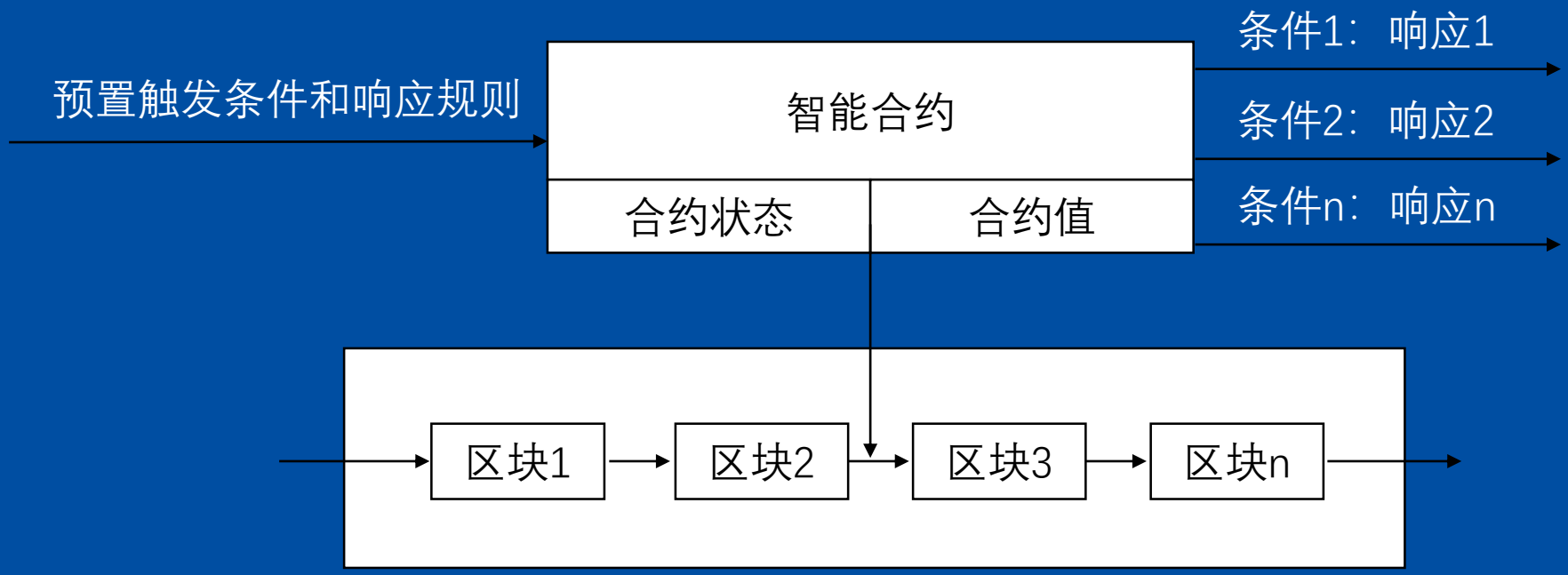

保密性和安全共享
数据完整性
数据加密性
数据安全
隐私保护
数据隐私保护
身份隐私保护
位置隐私保护
边缘计算安全体系
边缘节点身份认证
云节点身份认证
节点身份管理
身份认证
访问控制
数据访问控制
角色访问控制
身份访问控制

04 边缘计算典型网络节点和结构图



抓拍机、摄像头以及应用终端等终端节点负责采集数据，将数据通过网络路由上传到AI计算盒，边缘计算节点对终端采集的数据进行AI运算和处理，然后存储各个区块的哈希值。云节点负责账本存储、共识、路由和应用等任务， 云节点是云平台上划分的多个虚拟机， 虚拟机负责共识出块和全账本存储， 并可以通过开发区块链应用来管理和监控整个网络。

04 边缘计算和云：基于区块链的智能合约



智能合约封装了预定义的触发条件及响应规则、触发合约执行的场景（如达到特定时间或发生特定事件等）、特定情境下的应对动作等。在边缘计算场景中，可以采用智能合约来操作数据，实现区块链数据的生成、修改、删除等权限管理功能，智能合约为区块链部署到AI硬件产品上提供了灵活可编程的机制和算法。

结语

- 比特大陆作为最早的AI芯片厂家之一，芯片设计和软件设计健全；
- 从最底层的芯片和硬件层，提供多种安全框架的产品解决方案；而芯片级别的安全保护、隐私保护等是最基础的磐石；
- 以AI芯片构建的边缘计算方案，未来有机会通过区块链智能合约，构建安全认证和访问体系；
- 5G+AI时代来临，千亿级的物联网设备，安全问题会更加突出。我们愿与滴滴和各界朋友合作，打造一个具有算力之美的、安全的智能世界；

比特大陆AI生态圈


比特大陆
BITMAIN
福建省电子信息(集团)有限责任公司
海康威视 HIKVISION
千方科技 CHINA TRANSINFO
Dt Dream 数梦工场
NetPosa 东方网力
中海纪元
千视通
滴滴
Baidu 百度
SZMMO 深鲲科技
汉王 Hanvon
YISA 以萨技术
云从科技 CLOUDWALK

谢谢观看

THANK YOU