

Building on-Device AI using Google Coral

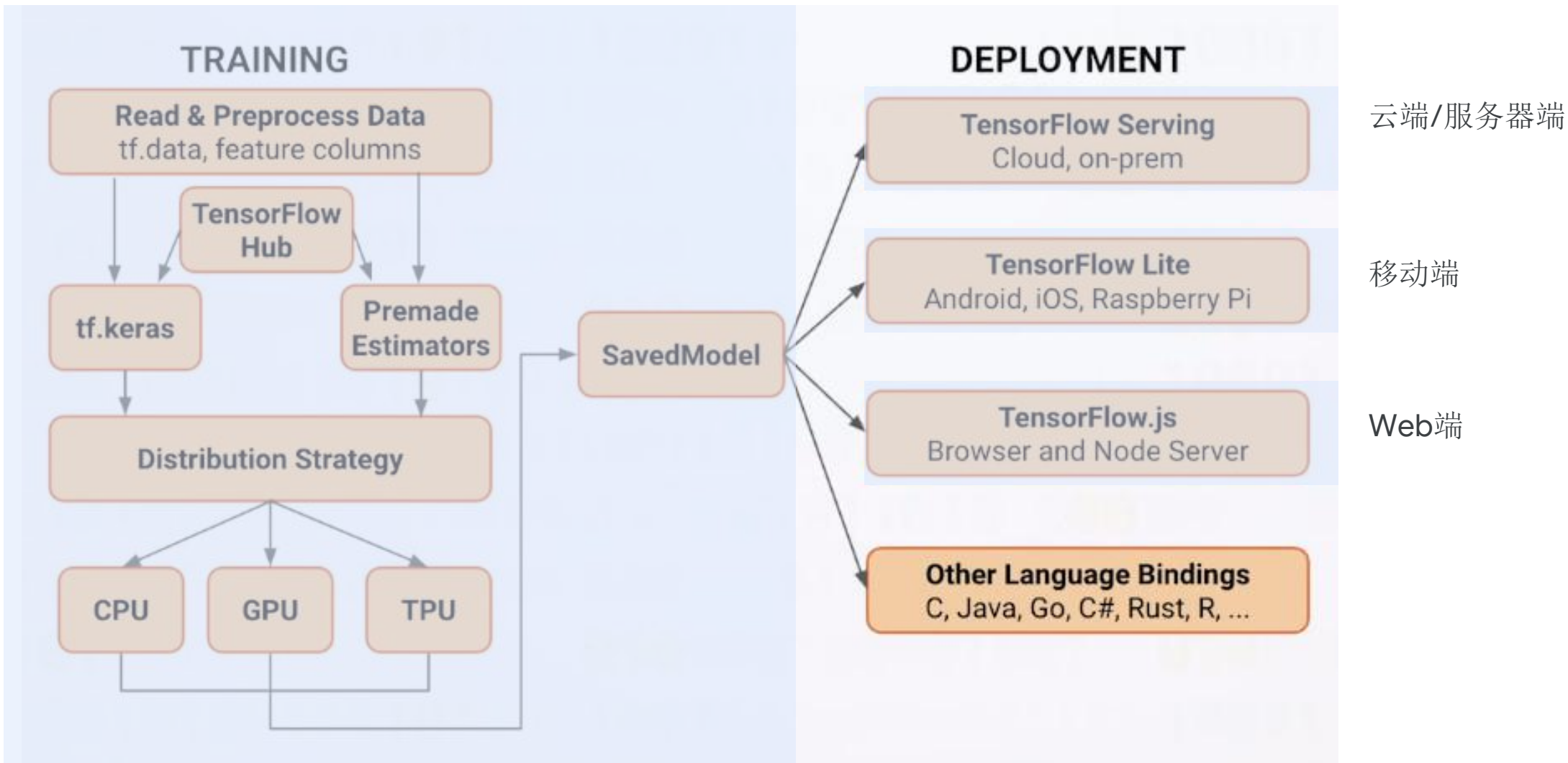


Eric Zhang



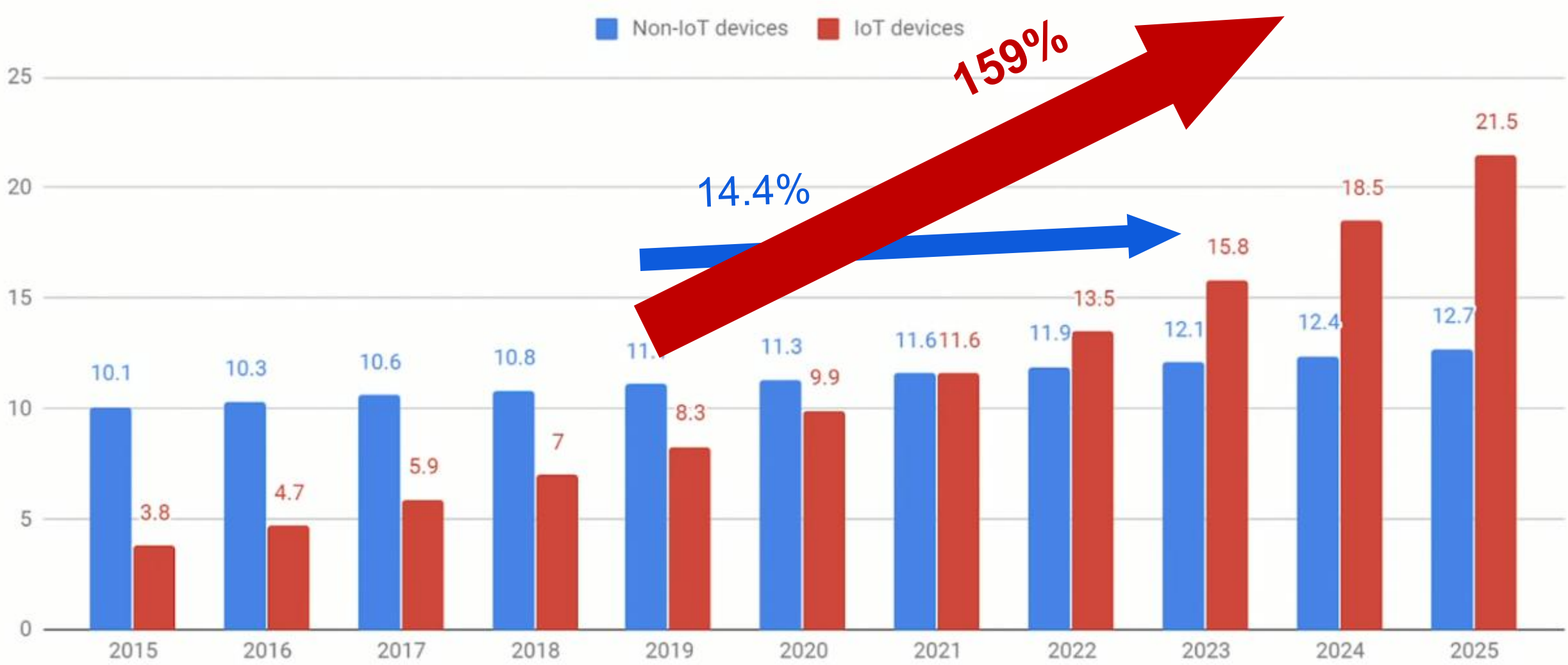
关于TensorFlow

模型算法



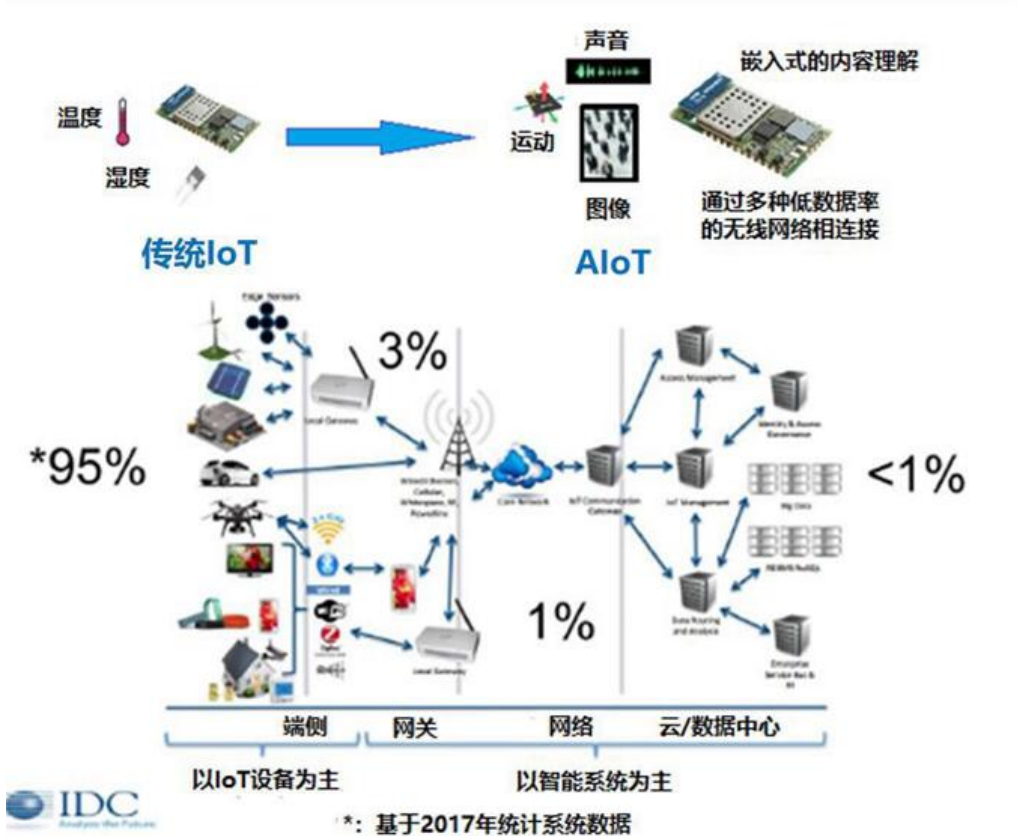
智能设备和物联网是未来最大潜力增长点

全球活跃的连接设备数量（十亿）



来源: IoT Analytics [State of 2018](#)

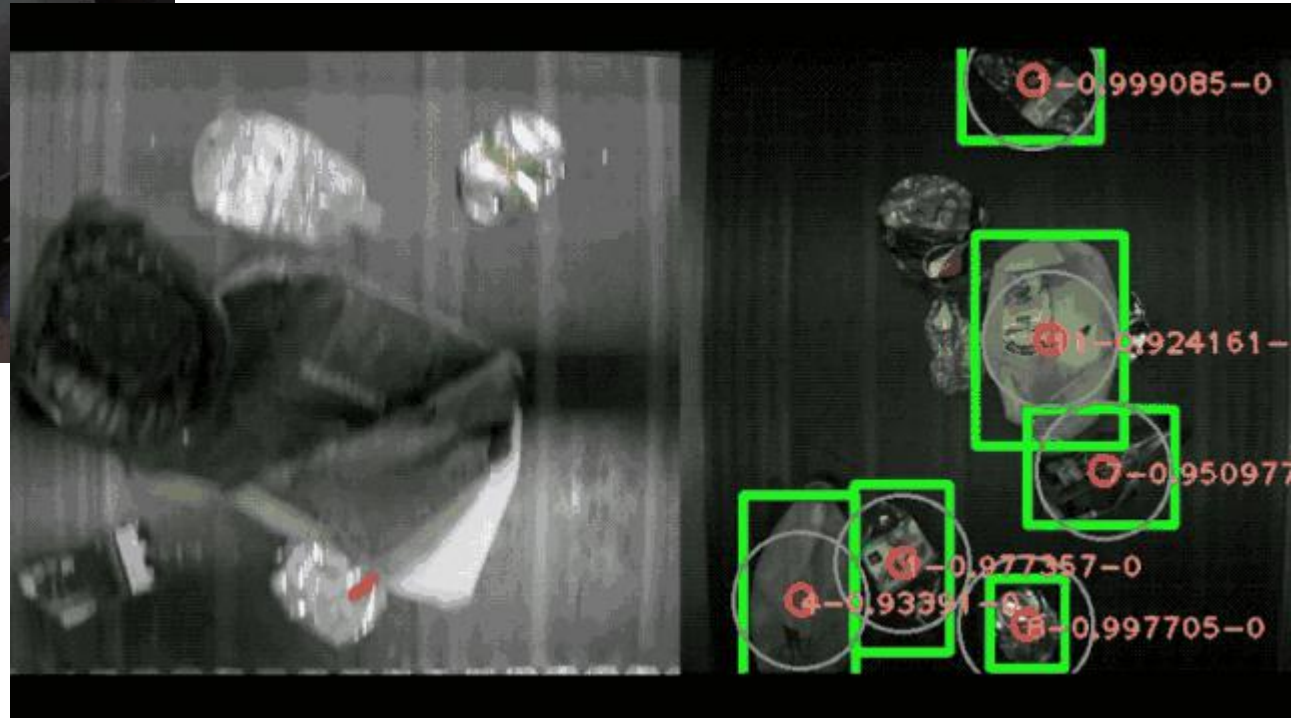
AIoT典型应用场景及特点



- AIoT: AI + IoT, 即在物联网设备中加入人工智能元素, 使物联网设备更智能, 从而解决实际问题。
- 传统IoT应用, 主要集中在传感器的数据采集, 仅做初步分析或不做分析将数据传送至云端, 设备ROI较低。
- AIoT增强端侧处理能力:
本地处理数据缩短系统反应时间。
节省海量数据传输及带宽耗用。
提高隐私性和安全性

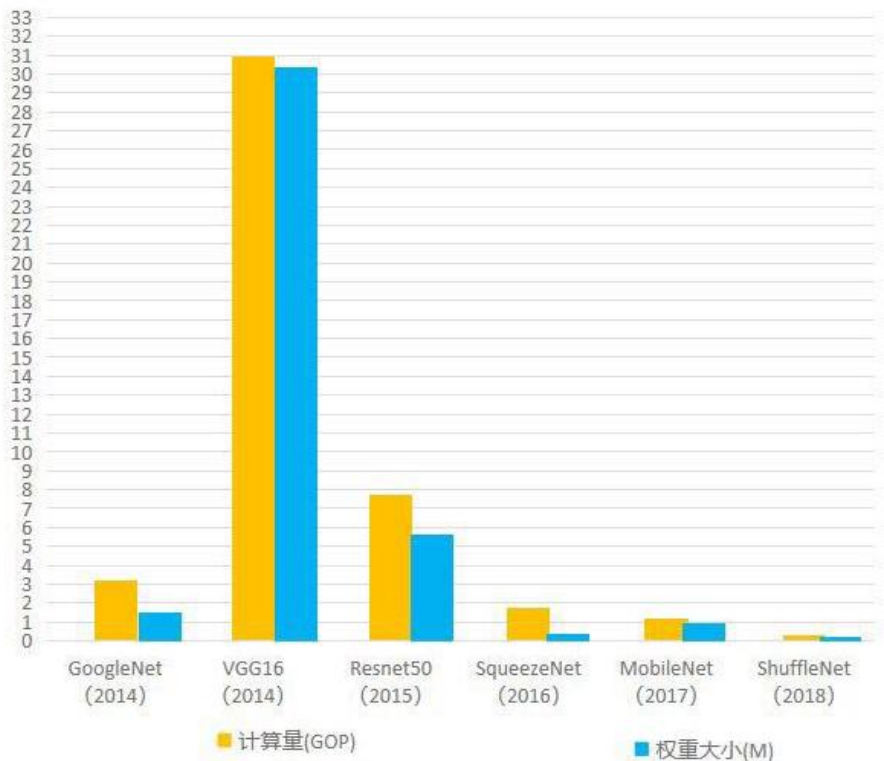
AIoT典型应用场景及特点

垃圾分拣机器人



人工智能将是推动高速增长的关键

神经网络模型计算量/权重大小对比(输入224*224*3)



- 伴随AI/ML研究的持续发展

- ✓ 模型设计方面：主流算法轻量化趋势明显
计算量及权重大小减少了几十倍
- ✓ 模型压缩方面：权重量化、剪枝、共享、哈夫曼编码
低秩分解、迁移/压缩卷积滤波器
知识蒸馏

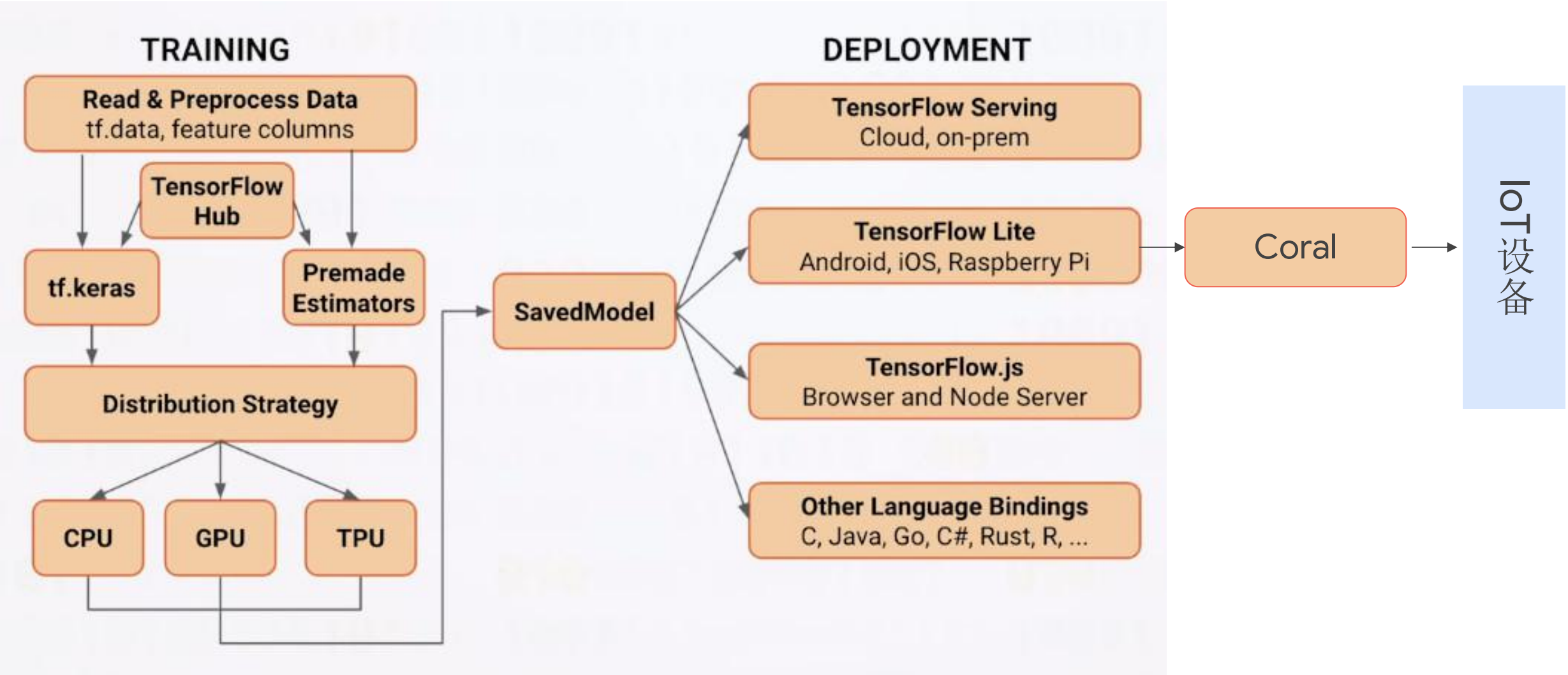
- 更多的AI能力在智能设备上应用带来了无限机会

- 随着边缘计算的发展，更多的AI能力从“云”上迁移到设备中

软：更适合终端设备的算法模型（更小，更快，更强）

硬：可在终端上做机器学习推理的硬件解决方案

关于TensorFlow



Introducing

Coral



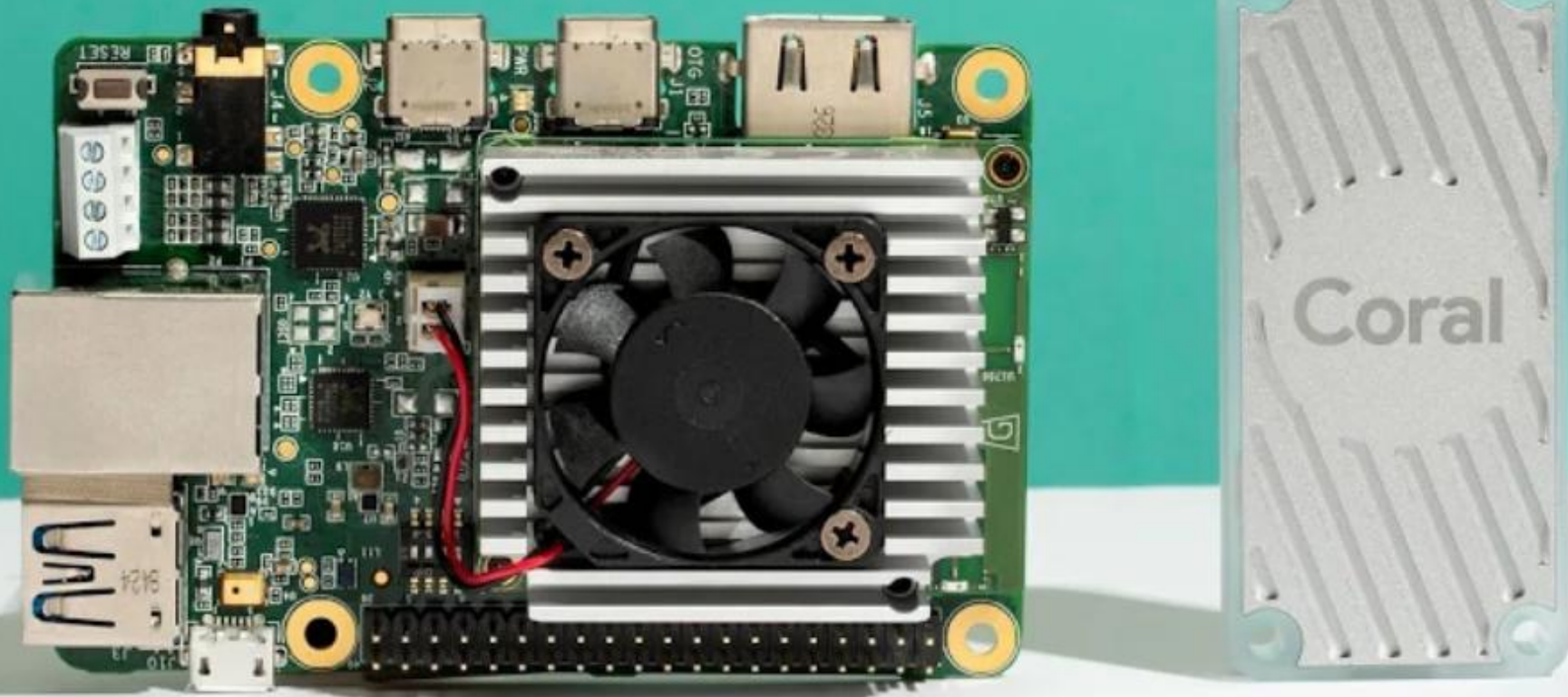
Google Coral 介绍

一个用于开发智能设备的机器学习应用平台

硬件组件为终端带来了高性能的机器学习加速。

软件工具通过TFLite帮助你开发基于机器学习模型和应用

利用成熟模型来快速设计实现原型并得到验证



Coral 命名由来



Coral 的灵感和使命

为人工智能应用的创新提供一个充满活力和高度协作的平台

设备端机器学习的优势

高性能：本地机器学习推理加速

更好的隐私保护：如有所需数据能够留存在设备端

设备上的应用可以访问本地数据：所有感知数据在设备端具有访问权限。

离线工作：可在无网环境工作或无需访问互联网的云服务。

节能：无需传输大规模数据到云上。不需要WIFI等耗能设备

Coral 产品套件

硬件&组件

板子

系统
周边配件

可以做机器学习推理能力的独立计算系统
或在现有系统上扩展机器学习处理功能

感知

音频
视频
环境感知
运动感知

丰富的周边感知组件能让开发者为模型抓取收集各种输入数据

软件&内容

工具

编译器
机器学习模型
SDK
OS

提供了平台开发所需的平台级工具包

文档

数据集
SDK参考
在线手册
样例

丰富的文档和示例让开发者了解如何使用产品和开发应用

Coral 硬件产品（当前可提供）

**Dev Board**

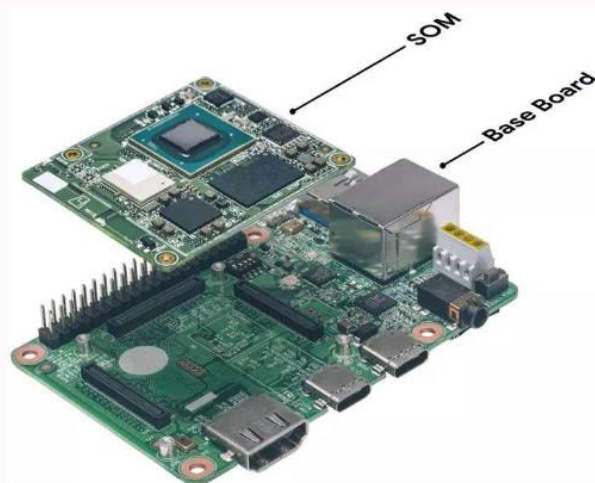
A single-board computer with a removable system-on-module (SOM) featuring the Edge TPU.

Price

\$149.99**USB Accelerator**

A USB accessory featuring the Edge TPU that brings ML inferencing to existing systems.

Price

\$74.99**Environmental Sensor Board**

An auto focus 25x25 mm 5-megapixel compatible camera module.

An accessory board that provides temperature, light, and humidity sensors for IoT applications.

Price

\$24.99

Price

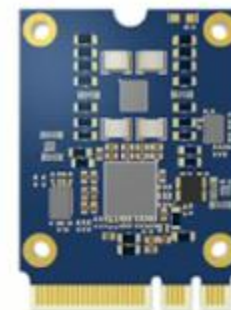
\$24.99**SOM**

A fully integrated System on Module in a 40mm x 48mm pluggable module with Edge TPU on board.

Price

\$114.99

Volume discount pricing available

**PCI-E Accelerator**

A PCI-E board for easy integration of Edge TPU into existing systems.

Available: Q3' 2019

Price

\$64.99

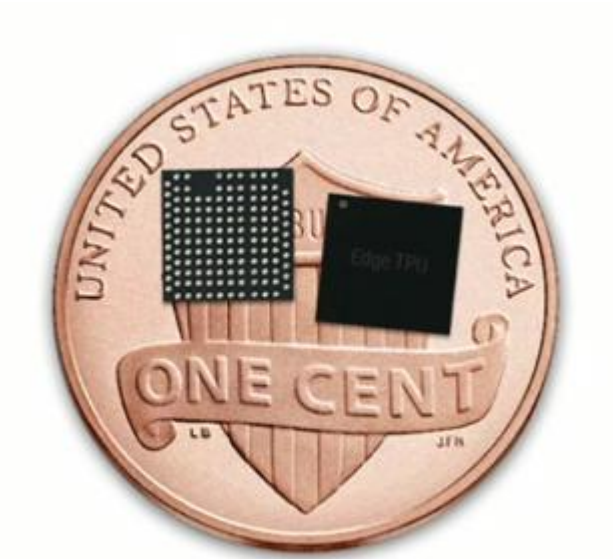
Volume discount pricing available

原型

生产

Google Edge TPU

Coral 板采用Google Edge TPU，一种专门用于在终端上进行机器学习推理的ASIC芯片



8/16位整形运算 | 2瓦特耗能 | 针对TFLite模型的优化

Edge TPU性能

Edge TPU处理器的处理性能可达 **4 万亿OPS**

性能对比（模型在终端上的推理）：

Model architecture	Desktop CPU*	Desktop CPU * + USB Accelerator (USB 3.0) <i>with Edge TPU</i>	Embedded CPU **	Dev Board † <i>with Edge TPU</i>
MobileNet v1	47 ms	2.2 ms	179 ms	2.2 ms
MobileNet v2	45 ms	2.3 ms	150 ms	2.5 ms
Inception v1	92 ms	3.6 ms	406 ms	3.9 ms
Inception v4	792 ms	100 ms	3,463 ms	100 ms

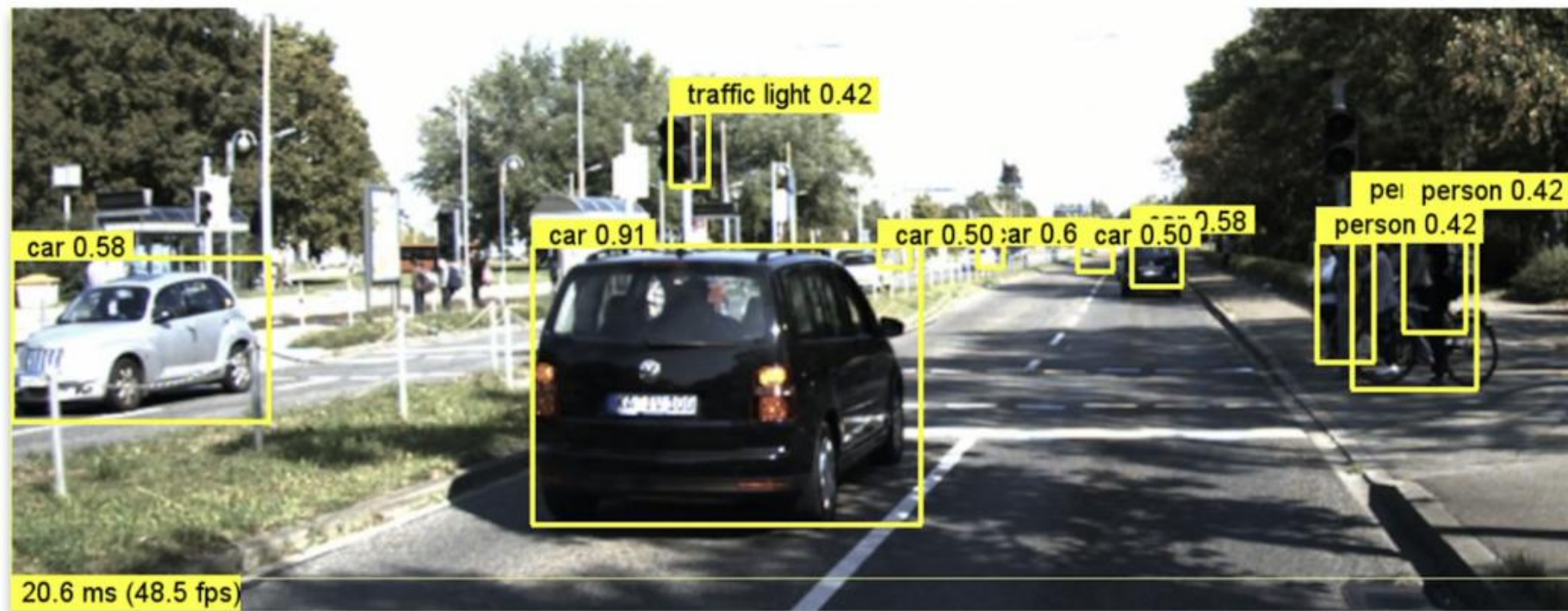
* Desktop CPU: 64-bit Intel(R) Xeon(R) E5-1650 v4 @ 3.60GHz ** Embedded CPU: Quad-core Cortex-A53 @ 1.5GHz

† Dev Board: Quad-core Cortex-A53 @ 1.5GHz + Edge TPU

Edge TPU实际应用

真实场景的示例，来自美国开发者的测试后评价

"I used canned Mobilenet V2 out of the box. I see 24 - 48 FPS with good accuracy - which is comparable with GTX 980 at 30 FPS. This is an impressive result!" - Ivo Stoyanov, SmartDrive Systems



Google Edge TPU与NVIDIA Jetson Nano



VS



MobileNet V2

配置	Time (s)	fps
i7-4870HQ	22,60	11,06
i7-7700K + GTX1080 2560CUDA	0,82	304,88
i7-7700K(tflite)	7,71	32,43
i7-7700K + Coral(tflite)	0,71	352,11
Jetson Nano 128CUDA	20,52	12,18
Jetson Nano (tflite)	23,02	10,86
RPi	120,81	2,07
RPi(tflite)	59,05	4,23
RPi + Coral(tflite)	3,31	75,53
Jetson Nano + Coral(tflite)	1,12	223,21

Coral Dev Board 开发板

Coral开发板

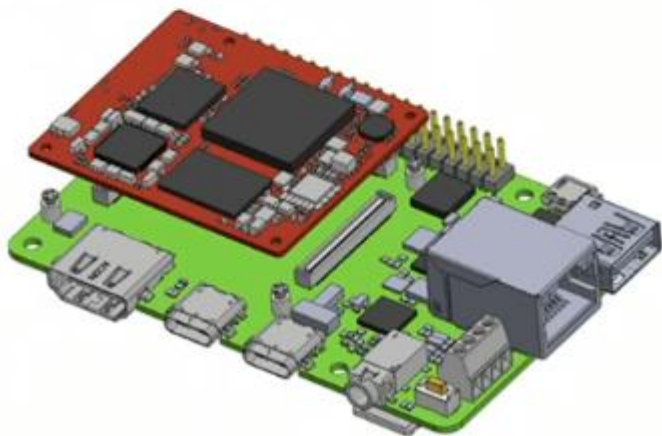


原型开发板，支持TFLite模型在上面直接做机器学习推理的功能。

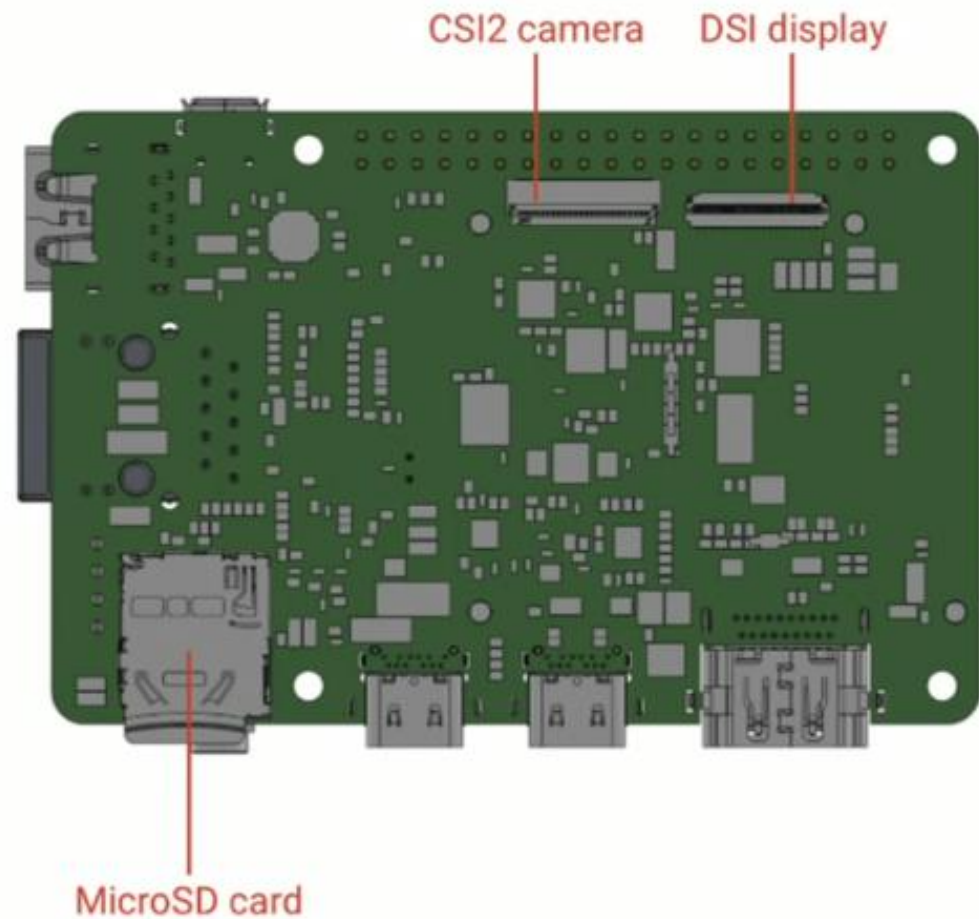
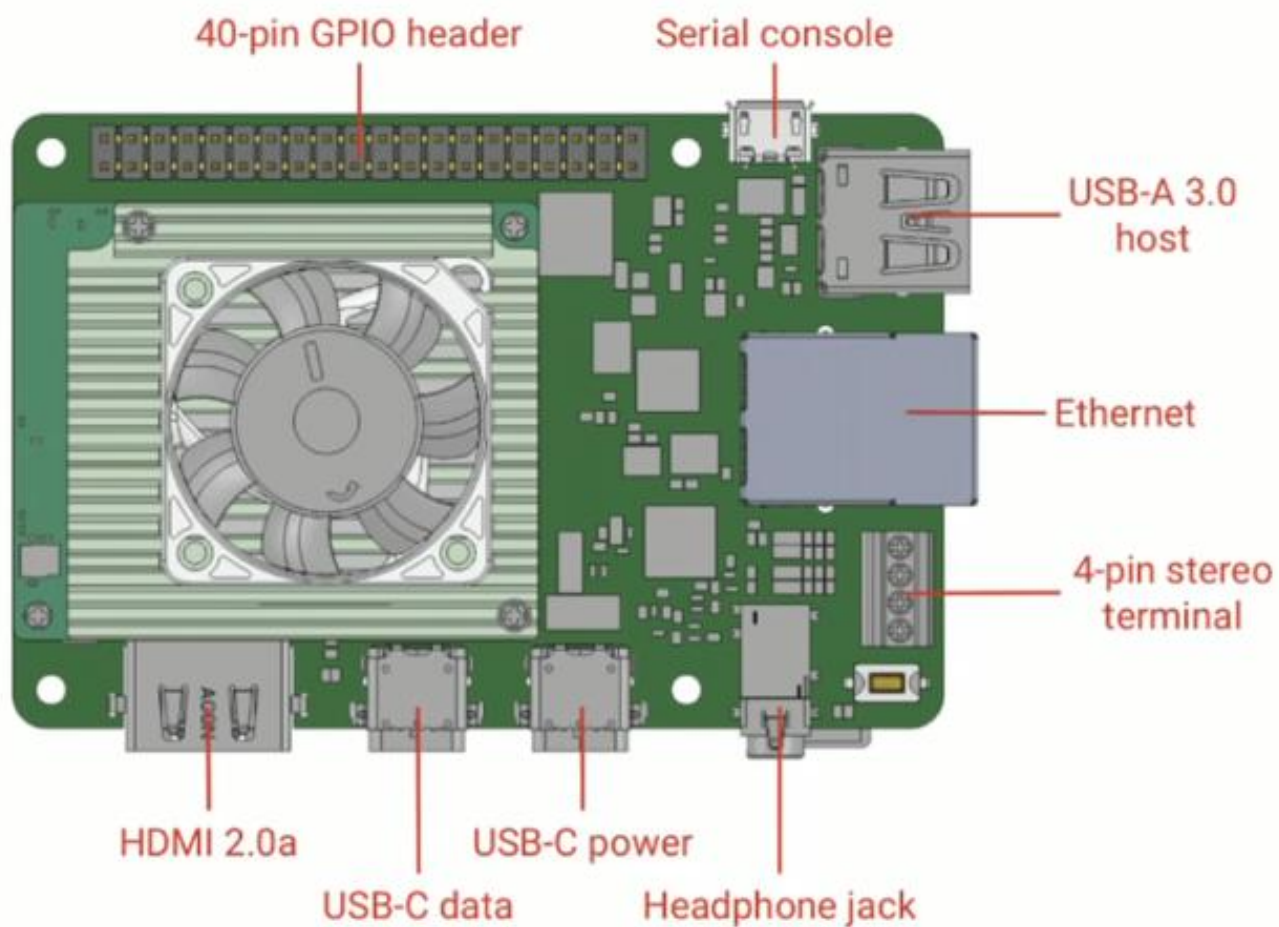
微型计算机：开发板上有CPU，GPU，内存及扩展SD存储，Linux系统且能够提供微型计算机的一切功能。

SoM（System on module）：采用的是SoM模块设计，将Coral Edge TPU用于兼容板上。

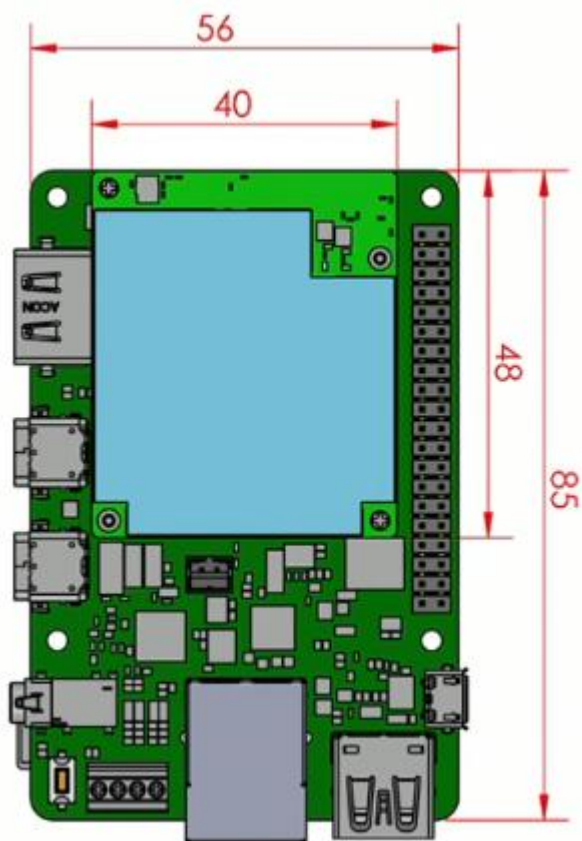
IO：底板为通用应用的需要提供多样IO接口。



Coral开发板设计布局

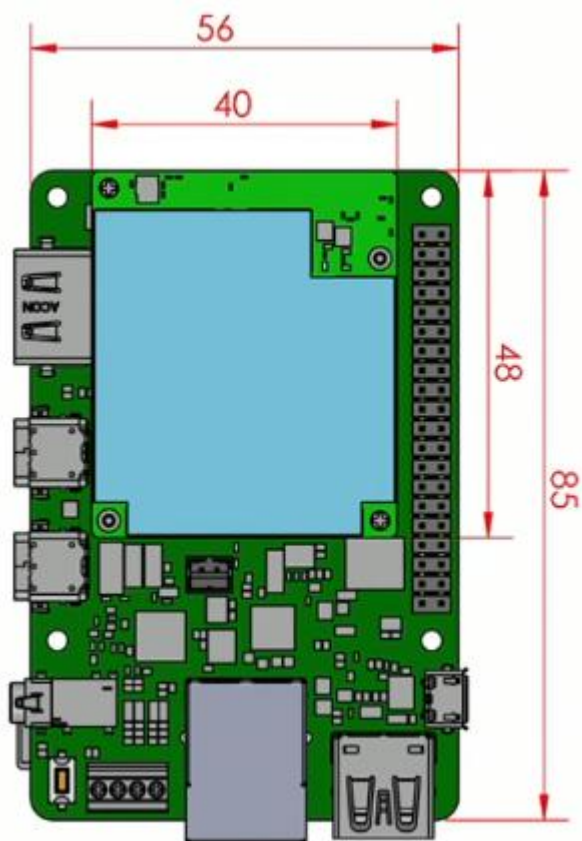


Coral开发板技术规格



- **Edge TPU模块 (SOM)**
 - CPU : Quad-core Cortex-A53, plus Cortex-M4F
 - GPU : C7000L GPU
 - TPU : Google Edge 机器学习加速处理器
 - 安全 : 加密处理器
 - RAM内存 : 1GB LPDDR4
 - 闪存 : 8GB eMMC
 - WiFi : Wi-Fi 2x2 MIMO(802.11b/g/ac 2.4/5GHz),Bluetooth
- **USB接口**
 - USB Type-C 电源 : 5V DC
 - USB 3.0 Type-C OTG口
 - USB 3.0 Type-A 口
 - USB 2.0 Micro-B serial console port

Coral开发板技术规格



- 音频接口
 - 3.5mm 音频插孔
 - 数字PDM麦克风
 - 2.54mm 4针端子 用于立体声扬声器
- 视频接口
 - HDMI 2.0a (全尺寸)
 - MIPI-DSI-39针 FFC连接器 (For Display)
 - MIPI-DSI2-24针 FFC连接器 (For Camera)
- Micro SD卡槽
- LAN: 千兆以太网网卡
- GPIO (40针扩展接口)
- 支持的OS: Debian Linux
- 支持的机器学习模型: Inception、MobileNet...

GPIO接口

40针 GPIO扩展口为外部设备接口而设计（兼容树莓派），可以扩展对接LED，开关，控制器，感知器等

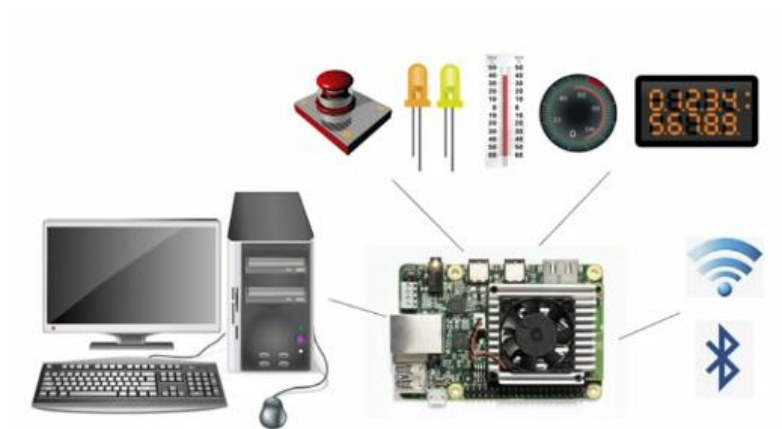
- ❑ Default pin functions, can be changed
- ❑ 5V and 3.3V
- ❑ GPIO
- ❑ PWM
- ❑ I2C x2
- ❑ SPI
- ❑ UART
- ❑ SAI (audio)

SOC SIGNAL	BASEBOARD SIGNAL	BASEBOARD SIGNAL	SOC SIGNAL
	3V3 power	1 2	5V power
I2C2_SDA	I2C2_SDA	3 4	5V power
I2C2_SCL	I2C2_SCL	5 6	Ground
UART3_TXD	UART3_TXD	7 8	UART1_TXD
	Ground	9 10	UART1_RXD
UART3_RXD	UART3_RXD	11 12	SAI1_TXC
GPIO6	GPIO_P13	13 14	Ground
PWM3	PWM3	15 16	GPIO_P16
	3V3 power	17 18	GPIO_P18
ECSPI1_MOSI	ECSPI1_MOSI	19 20	Ground
ECSPI1_MISO	ECSPI1_MISO	21 22	GPIO_P22
ECSPI1_SCLK	ECSPI1_SCLK	23 24	ECSPI1_SS0
	Ground	25 26	ECSPI1_SS1
I2C3_SDA	I2C3_SDA	27 28	I2C3_SCL
GPIO7	GPIO_P29	29 30	Ground
GPIO8	GPIO_P31	31 32	PWM1
PWM2	PWM2	33 34	Ground
SAI1_TXFS	SAI1_TXFS	35 36	GPIO_P36
NAND_DATA07	GPIO_P37	37 38	SAI1_RXD0
	Ground	39 40	SAI1_TXD0

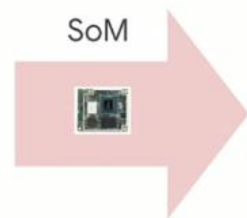
- Synchronous Audio Interface (SAI)
- Serial Peripheral Interface (SPI)
- General Purpose I/O
- 5V
- Inter-integrated Circuit (I2C)
- Universal Asynchronous Receiver-Transmitter (UART)
- Ground
- 3V3

Coral 开发板的使用方式

- 将开发板作为一个单板计算机来使用，直接接上IO外设来做原型开发。
- 部署AI应用和只用SoM板来扩展到生产中，绑定相应的软件和定制化的硬件。



原型及App开发



扩展到生产和将SoM嵌入到产品中

Coral USB 加速棒



MCU	Cortex-M0 MCU (power, thermal monitor)
TPU	Coral Edge TPU
RAM Memory	11 MB SRAM (on the Edge TPU ASIC)
Power	USB 3.0, Type-C connector
Data	USB 3.0, Type-C connector
Supported OS	Debian/Raspbian Linux
Supported ML	TF Lite
Supported Models	Inception, MobileNet, Daredevil



Coral 加速棒的优势

- 将设备机器学习能力赋能到很多机器上：

使用USB加速棒，让开发者很容易在Linux系统上获得Edge TPU特性功能

- 与很多主流硬件平台兼容：

Linux PC，笔记本，树莓派 和工业系统



Coral 产品套件

硬件&组件

板子

系统
周边配件

可以做机器学习推理能力的独立计算系统
或在现有系统上扩展机器学习处理功能

感知

音频
视频
环境感知
运动感知

丰富的周边感知组件能让开发者为模型抓取收集各种输入数据

软件&内容

工具

编译器
机器学习模型
SDK
OS

提供了平台开发所需的平台级工具和公用程序

文档

数据集
SDK参考
在线手册
样例

丰富的文档和示例让开发者了解如何使用产品和开发应用

Coral 软件组件

Coral为开发者开发机器学习应用提供了全栈的软件架构组件

Mendel OS

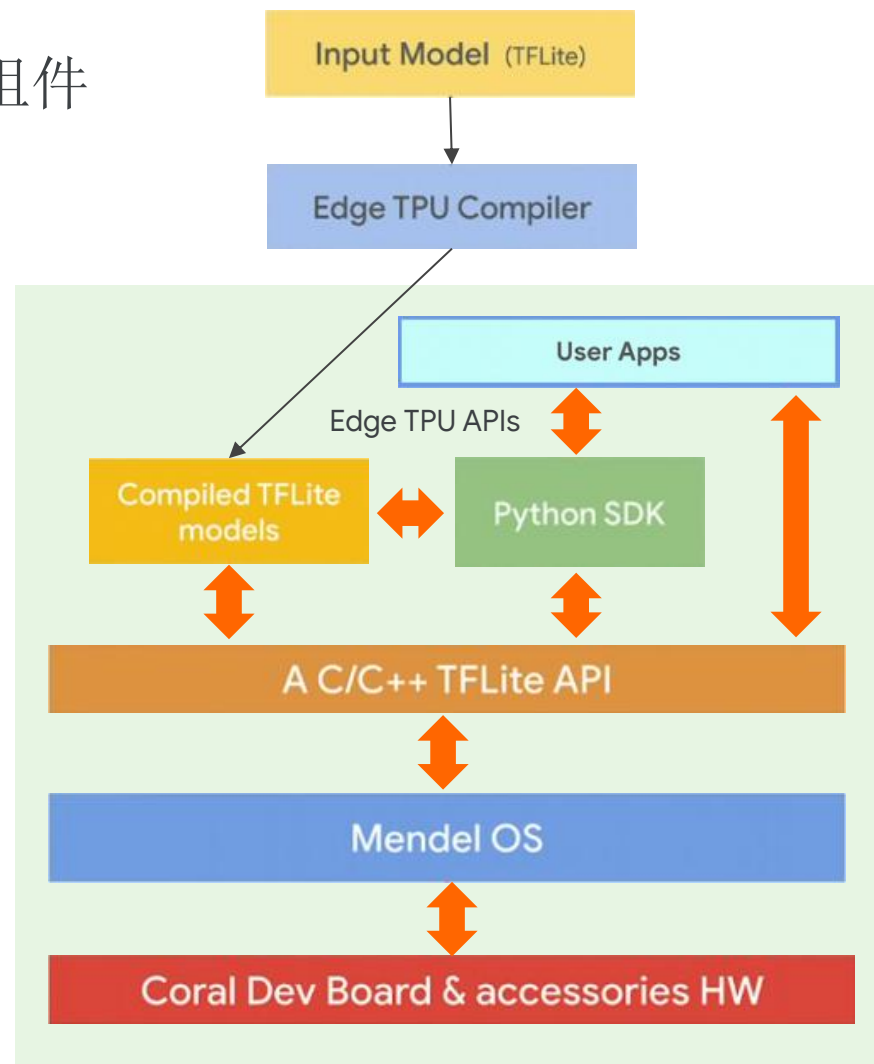
一个Debian Linux系统的分支

编译器

将TFLite 机器学习模型转成可在Coral产品上可运行的模型文件

Python SDK

为开发者提供了高级API函数实现对底层的来访问



Coral Python APIs

为开发者提供了易于上手的高级API模块，通过TFLite来实现模型推理的程序开发工作。

```
class edgetpu.basic.basic_engine.BasicEngine(model_path, device_path=None)
```

```
class edgetpu.classification.engine.ClassificationEngine(model_path, device_path=None)
```

```
class edgetpu.detection.engine.DetectionEngine(model_path, device_path=None)
```

```
class edgetpu.learn.imprinting.engine.ImprintingEngine(model_path)
```

Code Example in the Coral Object Detection Model Demo

```
# Initialize engine.
```

```
engine = DetectionEngine(args.model)
```

```
labels = ReadLabelFile(args.label) if args.label else None
```

Edge TPU已提供模型

Coral 上已有一些训练好的机器学习模型供大家使用

预编译好的TFLite 模型（.tflite文件）

已在Coral 设备上运行成功的模型，只需下载并部署到硬件上即可。

Python 模块已经预装在开发板设备上，但USB加速棒的话需要自己安装。

这些开放出来的模型用于非商业用途。

开放出来的模型： Image Classification、 Object Detection和Weight-imprinting model（迁移学习）

Coral 模型举例

Image Classification Models

MobileNet V1 (ImageNet)

Recognizes 1,000 types of objects

Dataset: ImageNet

Input size: 224x224

MobileNet V2 (ImageNet)

Recognizes 1,000 types of objects

Dataset: ImageNet

Input size: 224x224

MobileNet V2 (iNat insects)

Recognizes 1,000+ types of insect

Dataset: iNaturalist

Input size: 224x224

MobileNet V2 (iNat plants)

Recognizes 2,000+ types of plants

Dataset: iNaturalist

Input size: 224x224

MobileNet V2 (iNat birds)

Recognizes 900+ types of birds

Dataset: iNaturalist

Input size: 224x224

Inception V1 (ImageNet)

Recognizes 1,000 types of objects

Dataset: ImageNet

Input size: 224x224

Inception V2 (ImageNet)

Recognizes 1,000 types of objects

Dataset: ImageNet

Input size: 224x224

Inception V3 (ImageNet)

Recognizes 1,000 types of objects

Dataset: ImageNet

Input size: 299x299

Inception V4 (ImageNet)

Recognizes 1,000 types of objects

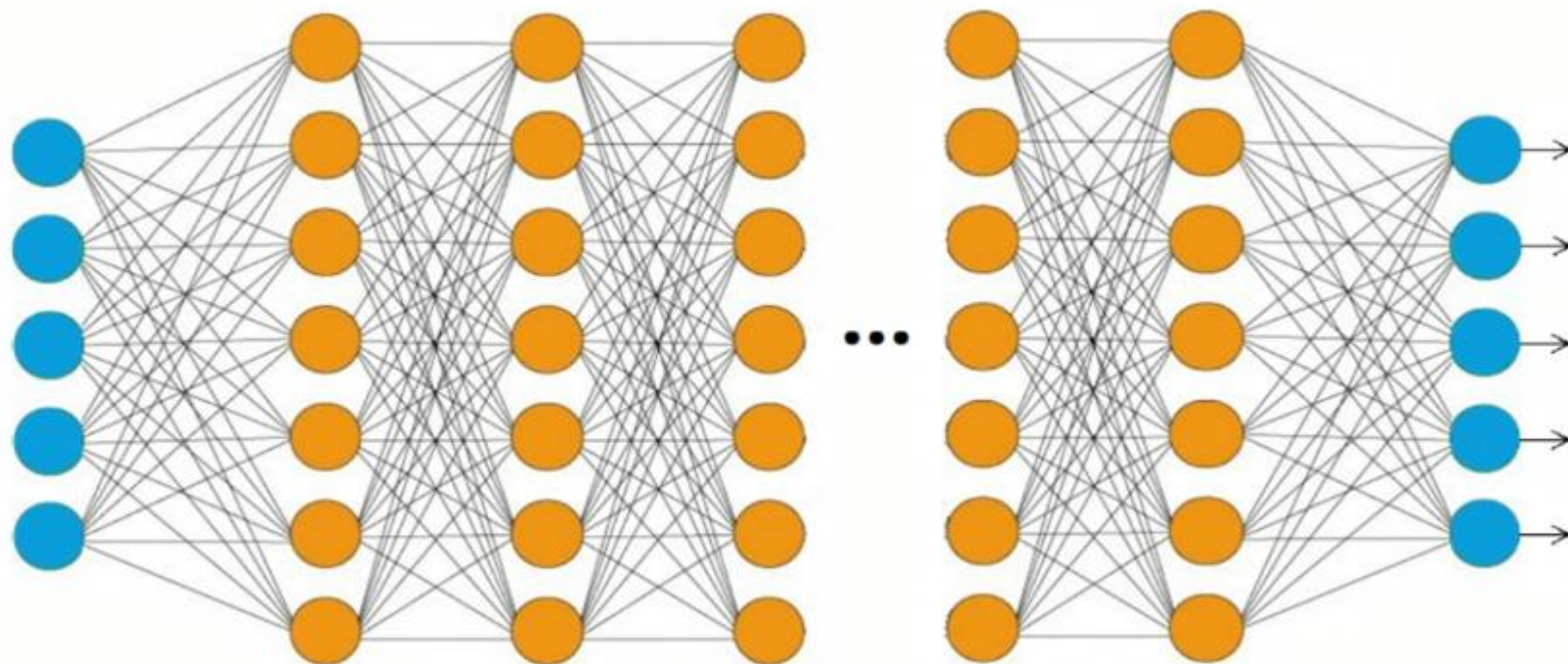
Dataset: ImageNet

Input size: 299x299

利用迁移学习来再训练模型

用迁移学习来训练模型可以节省时间和精力。

在预训练的模型上，通过调整神经网络的层结构和用自己的数据训练来创建自己的模型，并且大大减少训练成本。



Coral Code Example for transfer learning training

```
// Prepare your dataset: From the Docker /tensorflow/models/research/slim/  
directory
```

```
./prepare_checkpoint_and_dataset.sh --network_type mobilenet_v1
```

如何使用Coral机器学习模型

将所提供的Coral模型编译成属于自己的模型

方法1: 简单容易

使用预编译的TFLite模型

预训练模型

将模型上传至开发板

方法2: 用自己的数据提高准确度

用迁移学习定制模型

预训练模型+
自己的数据

方法3: 自己做模型进行编译

用Cloud TF编译器来
创建自己的模型

自己的模型

Edge TPU
编译器

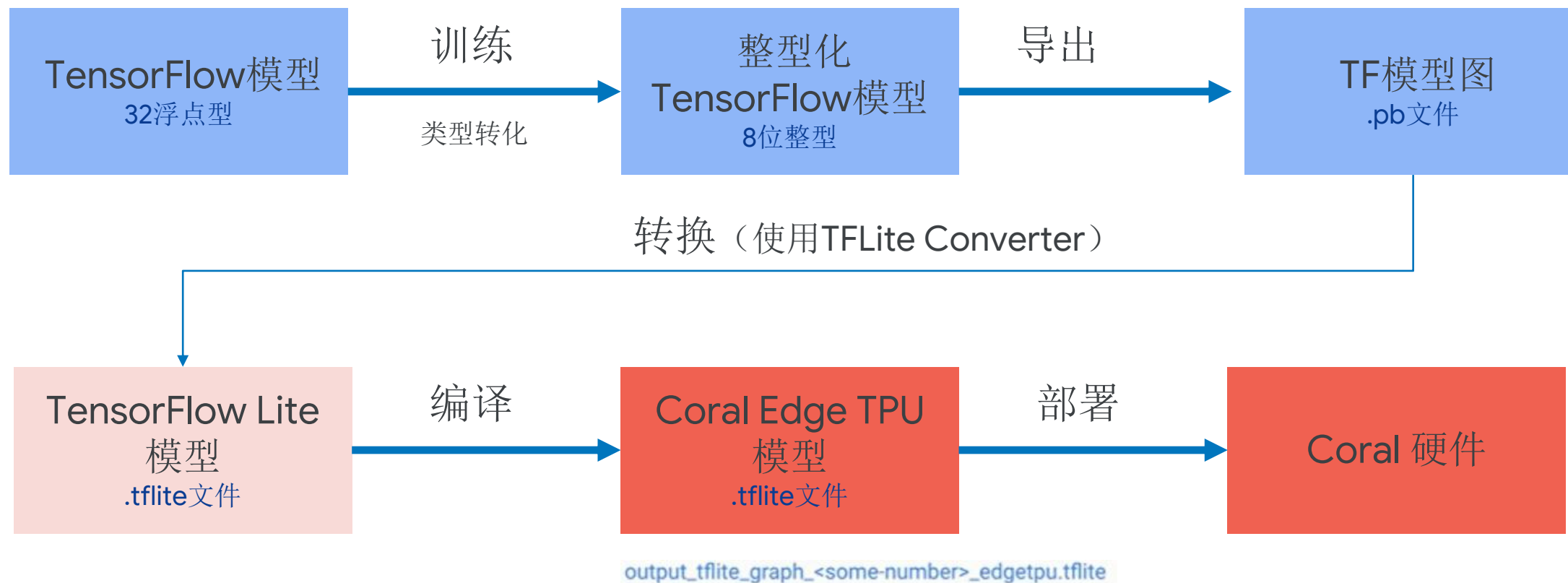
将模型上传
至开发板

Coral Dev
Board

1. GCP上Edge TPU编译器
2. 本地编译器 (2019.5.29发布)

创建定制化Coral模型的工作流

将所提供的Coral模型编译成属于自己的模型



利用Coral的潜在产品和领域

目前或未来领域使用的机器学习模型：

Object detection、 Object classification、 Audio processing



Consumer electronics



Warehousing management



Fleet management



Public services and kiosks



Home appliances



Consumer robots & Assistants



Automotives



Education Aids



Assembly line & packaging



Industrials & factory robots



Medical & health care industry



Scientific Research

更多的产品信息和文档

Coral 网址:

coral.withgoogle.com

Coral beta

Products Projects Docs Resources 

Build intelligent ideas with our platform for local AI

Our products are...



Private

Keep user data local and under your control



Fast

Accelerate AI locally, inference times down to 2 ms



Efficient

Low power demand to allow embedded applications



Offline

Deploy in the field where connectivity is limited

更多资源

Mendel Linux guide:

coral.google.com/docs/+/refs/heads/master/GettingStarted.md

Edge TPU model compiler: coral.withgoogle.com/web-compiler

Convert TF model to TensorFlow Lite:

www.tensorflow.org/lite/convert

Quantization-aware training:

[github.com/tensorflow/tensorflow/tree/r1.13/tensorflow/contrib/quantize](https://github.com/tensorflow/tensorflow/tree/r1.13/tensorflow/contrib/quantize#quantization-aware-training)
[#quantization-aware-training](https://github.com/tensorflow/tensorflow/tree/r1.13/tensorflow/contrib/quantize#quantization-aware-training)

Coral tag on StackOverflow: [\[google-coral\]](#)

Live Demo

Talk is cheap , Show me the Demo

总结

了解了Coral产品全家桶

学习了TFLite 模型

使用Coral 开发板和预编译模型来快速进行验证

通过迁移学习来重构模型

通过再训练模型来构建定制化模型

样例演示

THANKS



Eric Zhang

