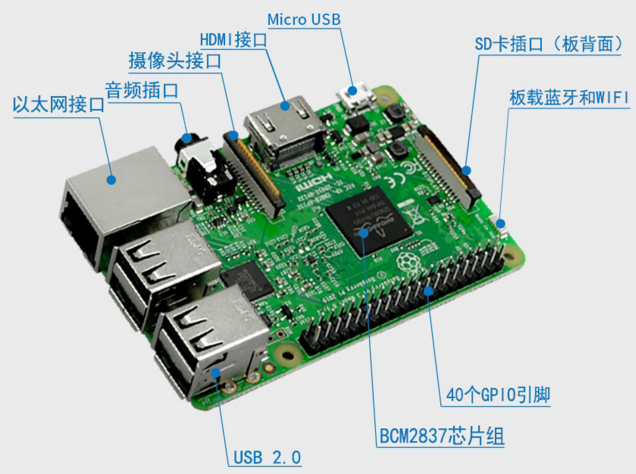
# 基于树莓派3b和Movidius神经计算棒的嵌入式人工智能平台

# 树莓派3b

Raspberry Pi 3 使用Broadcom的BCM2837 SoC，最高主频1.2GHz，4个Cortex A57内核，支持64位。提供一个100Mbps以太网接口，802.11n无线连接（板载天线），蓝牙4.1连接。

树莓派外观微型，可以适合很多场合。介个低廉，可以有更广阔的群体。接口丰富，可以用到很多开发中。



# 神经计算棒（NCS）

英特尔NCS是世界上首个基于USB模式的深度学习推理加速器，可以辅助人工智能应用开发者分析、调试、验证神经网络，并为深度学习推理做加速。

NCS内置的Myriad 2 VPU提供了强大且高效的性能，能够在计算棒上运行深度神经网络，并具有极低的功耗。在不联网，不使用云服务器的条件下，就可以在终端使用深度神经网络运行物体识别、视频分析等视觉相关的人工智能应用。低功耗以及低成本的特点让计算棒适用于非常多的场景，当下热门的人工智能设备，无人机、VR/AR头戴设备、智能摄像头等等，都有它的身影。



## 软件栈

Movidius为NCS提供了完整的SDK，包含Tool和API。

Tool部分有Profiler、Checker、Compiler三个工具。

1. Profiler工具可以分析的网络模型，并报告网络每一层的运行效率等，辅助开发者优化网络结构。
2. Checker工具可以在开发者开始部署网络之前进行测试，查看网络执行的结果和性能。
3. Compiler工具则是将深度网络模型（caffe模型或TensorFlow模型）转化为NCS能够识别的graph文件。

SDK中的API部分就是为深度学习开发者提供的NCS硬件调用接口，支持C和Python接口。

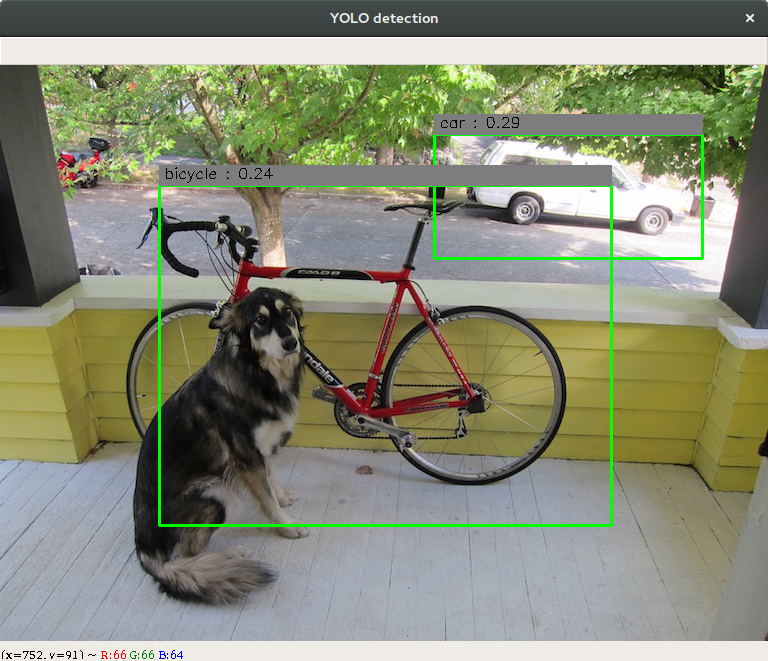
## 使用方式

通过训练得到的深度神经网络模型可以使用NCS SDK中的工具编译为能被计算棒使用的模型格式，通过调用SDK中的接口，可以方便的在主机（NCS所连接的计算机）与NCS之间通信。NCS利用训练好的网络模型计算出图像分析的结果，并传输到主机上，完成推理工作。

## Examples

Movidius为NCS提供了类似caffe model zoo和TensorFlow models的github主页。https://github.com/movidius/ncappzoo，提供了当前已经支持的caffe和TensorFlow网络模型。更多的支持还在持续开发中。

支持的caffe网络模型有AgeNet、AlexNet、GenderNet、GoogleNet、SSD\_MobileNet、SqueezeNet和TinyYolo。支持的TensorFlow网络模型有inception\_v1/2/3/4和mobilenet。



以caffe的GoogleNet为例介绍NCS使用方法，在ncsdk安装完成后进入到examples下的caffe目录。在GoogleNet目录下，可输入make help查看支持的操作。

### Compiler

Compiler工具负责将网络模型转换成NCS可以识别的graph文件。命令格式如下

1. Caffe

mvNCCompile network.prototxt [-w weights\_file] [-s Max Number of Shaves] [-in Input Node Name] [-on Output Node Name] [-is Input-Width Input-Height] [-o Output Graph Filename]

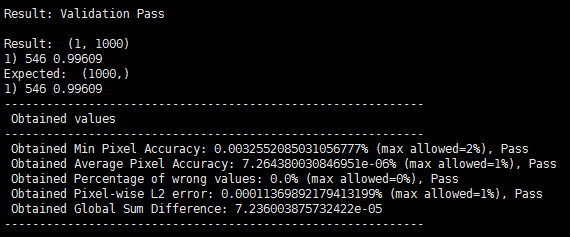
1. TensorFlow

mvNCCompile network.meta [-s Max Number of Shaves] [-in Input Node Name] [-on Output Node Name] [-is Input-Width Input-Height] [-o Output Graph Filename]

其中-s参数指定运行单元数量，Shave为NCS内加速运行单元。

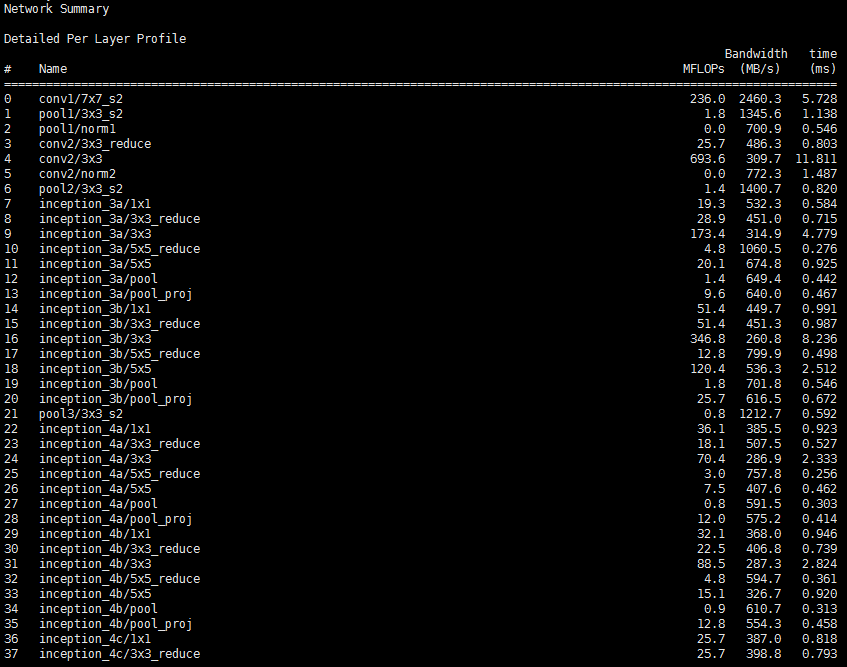
### Checker

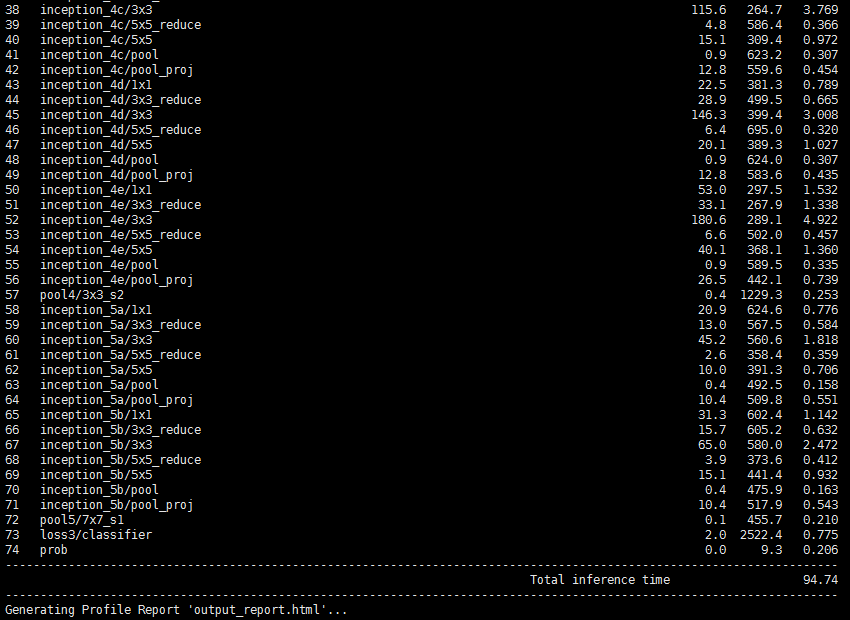
Checker工具可以对比在NCS和Caffe/TF运行网络的推理结果。



### Profiler

Profiler工具可以进行网络分析，给出每层网络的运行状况，包括Complexity(MFLOPs)，Bandwidth和time，如下图所示，也可以在生成的output\_report.html中进行查看。Html文件不仅包含上述数据，还包括数据流向。这些数据可以为进行网络优化提供依据和支持。





Html文件内容(局部)

