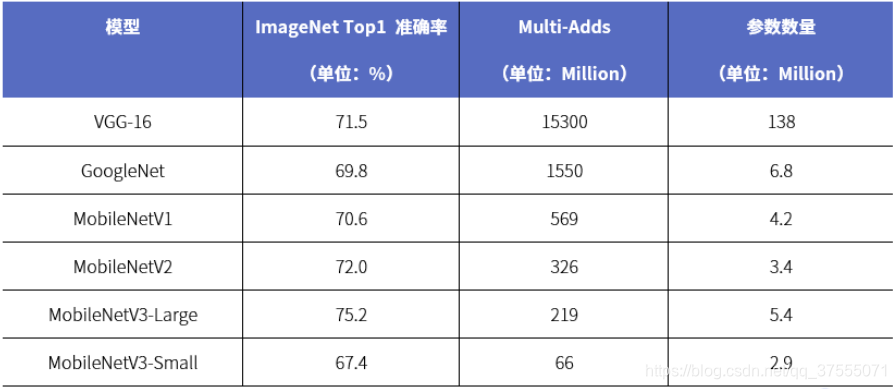
本文档记录于2025年6月11日，主要内容为首次接触到赵忠宇师兄的Mobile Net项目时想到的问题以及在网上搜索到的答案总结。阅读本文档需要基础的神经网络知识。

**第一部分：不涉及代码实现的概念性以及理论等问题。**

**1.什么是mobile net？**

Mobile net是轻量级神经网络的代表。相比起体积巨大和结构复杂的神经网络，mobile net拥有更少的参数量和运算量。Mobile Net可以在移动终端实现众多的应用，包括目标检测，目标分类，人脸属性识别和人脸识别等。

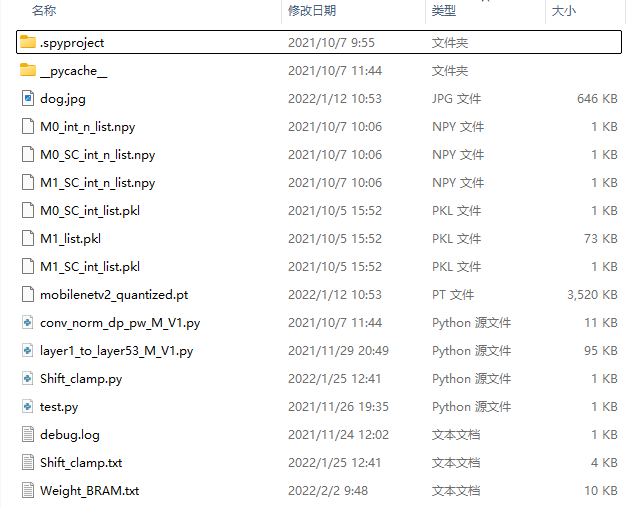


**2.Mobile net的结构是怎么样的？**

参考： [轻量级神经网络“巡礼”（二）—— MobileNet，从V1到V3 - 知乎](https://zhuanlan.zhihu.com/p/70703846)

**第二部分：工程文件**

1. **文件夹中的这些文件都是些什么？**



根据文件类型可以大致了解这些文件的作用：

前面两个文件夹属于配置文件

Dog.jpg是输入数据

NPY、PKL文件是存储数据的文件

PT文件是是PyTorch中用于保存张量数据的文件格式。与pth文件类似，pt文件也常用于模型的保存和加载，但更侧重于保存单个张量或一组张量数据。

Python源文件存储了python代码

Debug.log应该是日志信息，但是并没有存储什么有效信息

文本文档中：

Shift\_clamp里面有移位和截取操作的verilog代码

Weight\_BRAM里面有memory\_initialization\_radix=2;以及memory\_initialization\_vector，这个vector里面就是逐渐递增的从0000 0000 到1111 1110的八位二进制数，似乎指示了数据在内存中的地址

接下来逐个分析python源代码中的代码的含义。

Conv\_norm\_dp\_pw\_M\_V1根据文件名可以看出来是卷积-归一化-深度可分离卷积-逐点卷积-M-版本1。

Layer1\_to\_layer53\_M\_V1似乎代表了从第一层到第五十三层的神经网络。

Shift\_clamp是生成shift\_clamp.txt的代码

Test为测试文件

1. **能够理解python代码中的卷积的部分，但是不清楚量化具体过程**

**输入、权重减去零点值**

**卷积操作的结果乘以M0,加上M1，右移n位，最后加上输出的零点值。**