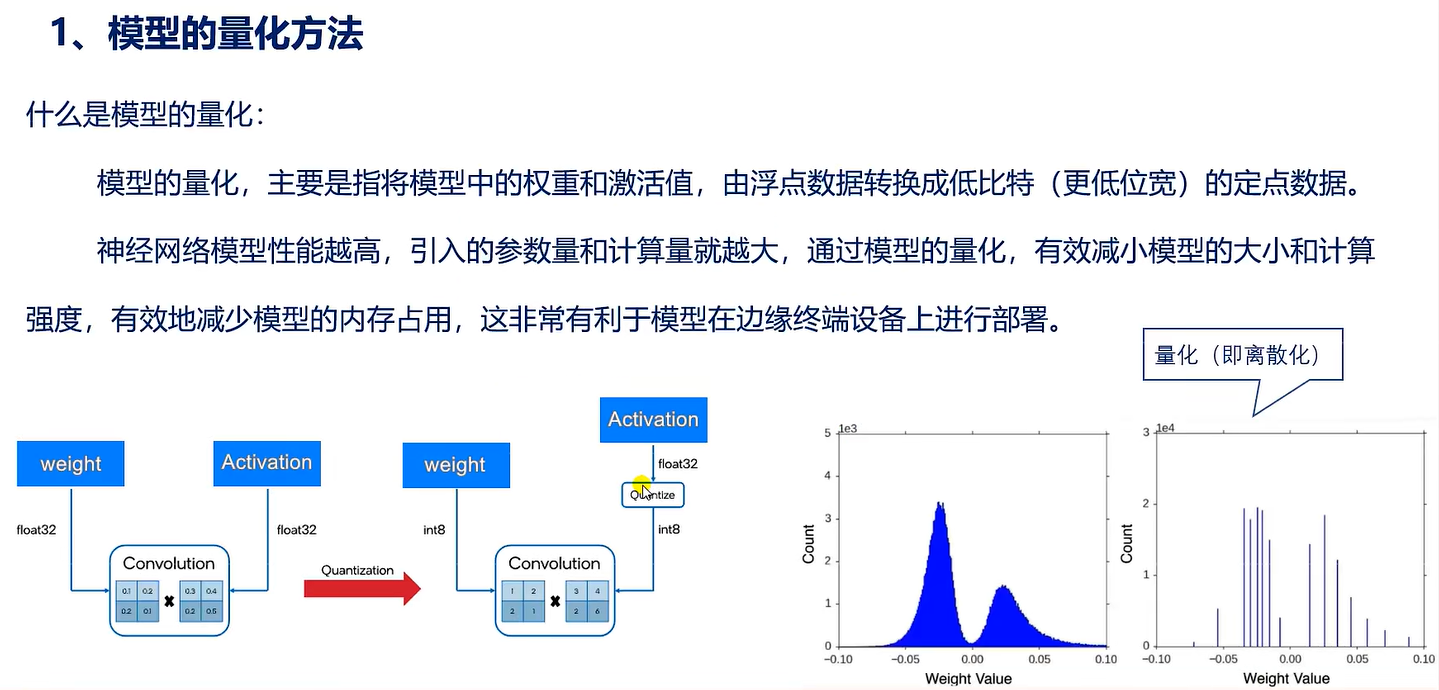
本讲课程讲解模型的量化功能。



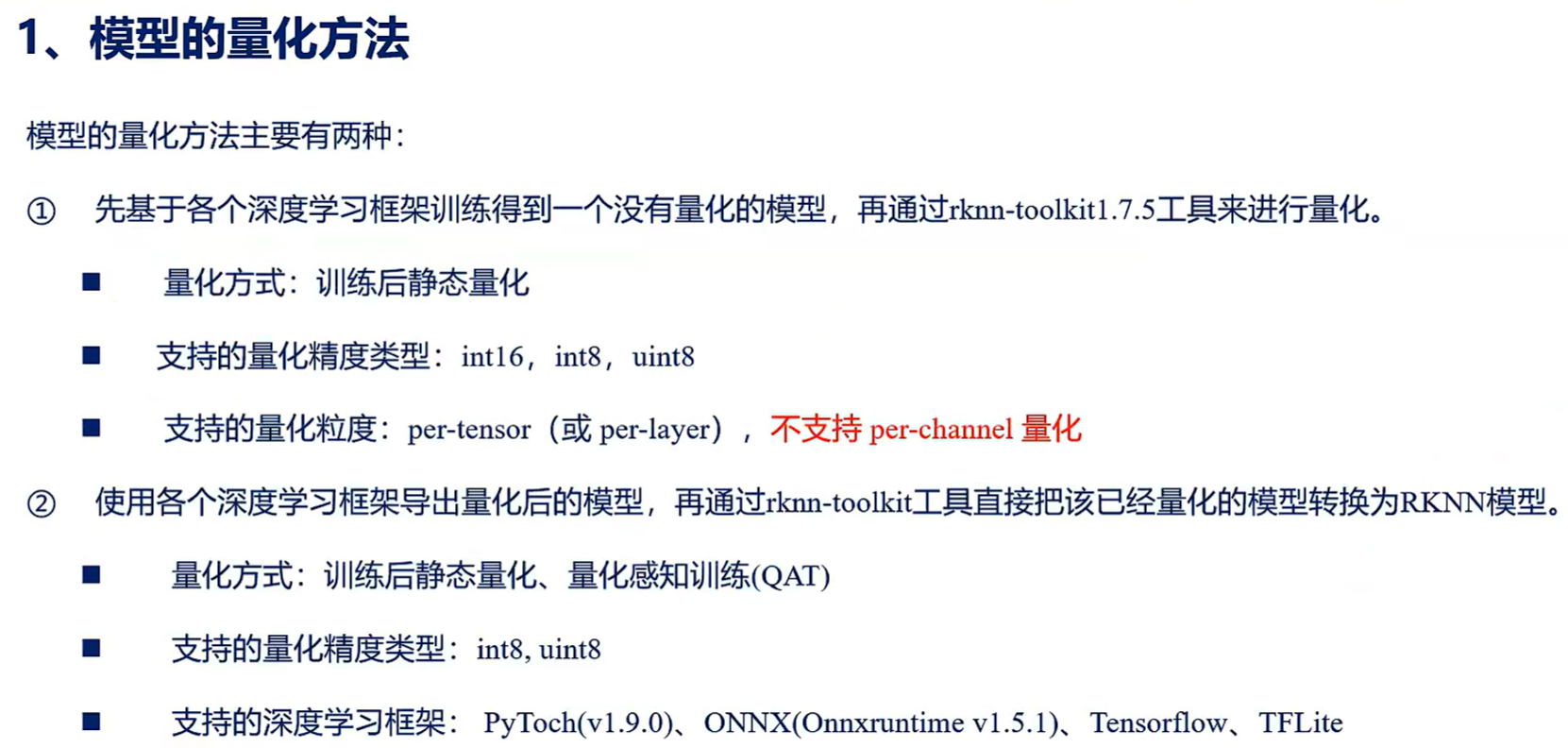
/\*

模型的量化方法：



模型的量化，主要是指将模型中的权重和激活值，由浮点数据转换成低比特（更低位宽）的定点数据。常见的就是转化为uint8。

其实模型的量化就是将连续数值的浮点模型的权重激活值，转化为有限多个离散值的过程。



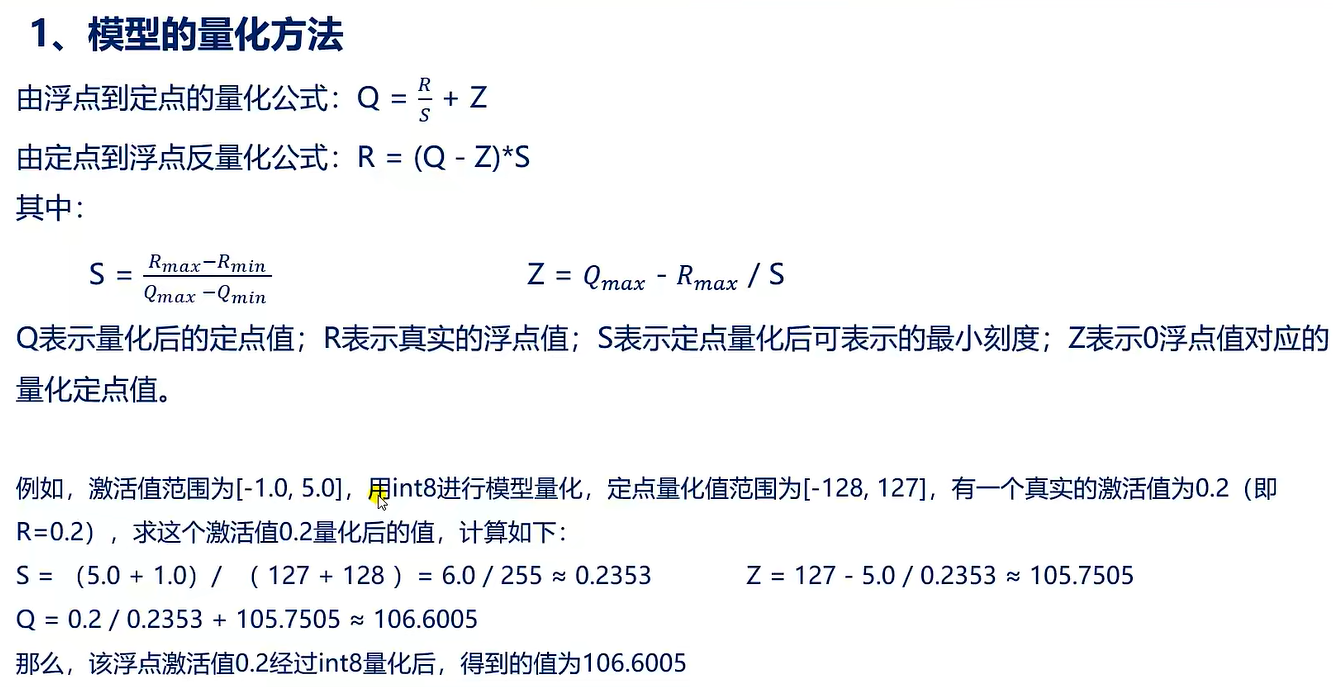
模型的量化方法主要就是2种。

1.先基于各个深度学习框架训练得到一个没有量化的模型（未量化的浮点模型），再通过rknn-toolkit1.7.5工具来进行量化。

2.使用各个深度学习框架导出量化后的模型，再通过rknn-toolkit工具直接把该已经量化的模型转换为RKNN模型。

我们主要使用方法一。





\*/

/\*

我们使用GPU进行模型训练时，是采用float32浮点数来训练的，训练得到的模型，即模型中的权重和激活值是float32类型，使用float32浮点数运算，能够保持精度。

一个原始的AlexNet网络模型，在imagenet数据集上进行训练时，该网络模型：

卷积层有5层：

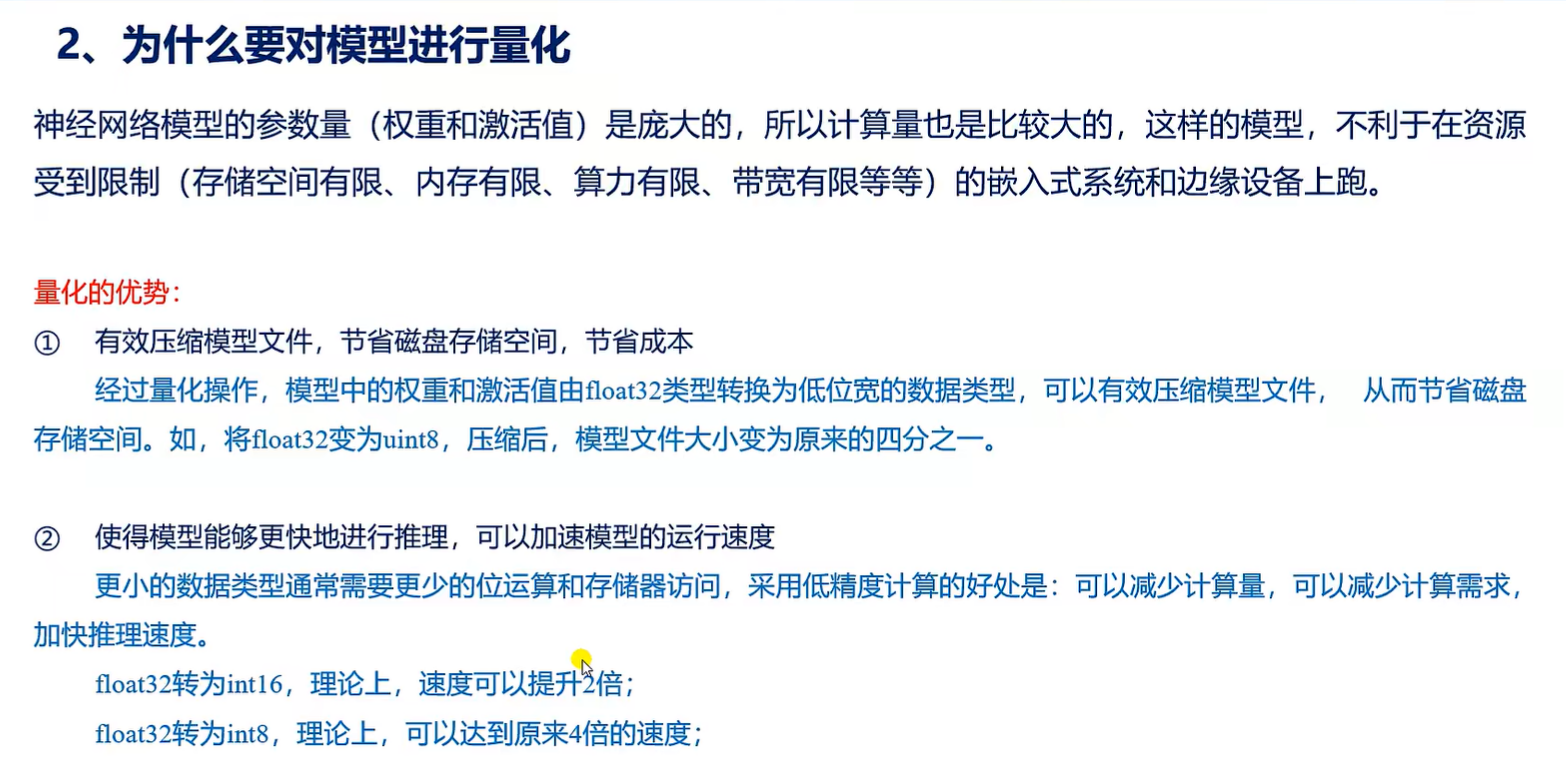
全连接层有3层；

池化层有3层；

神经元个数有65万个；

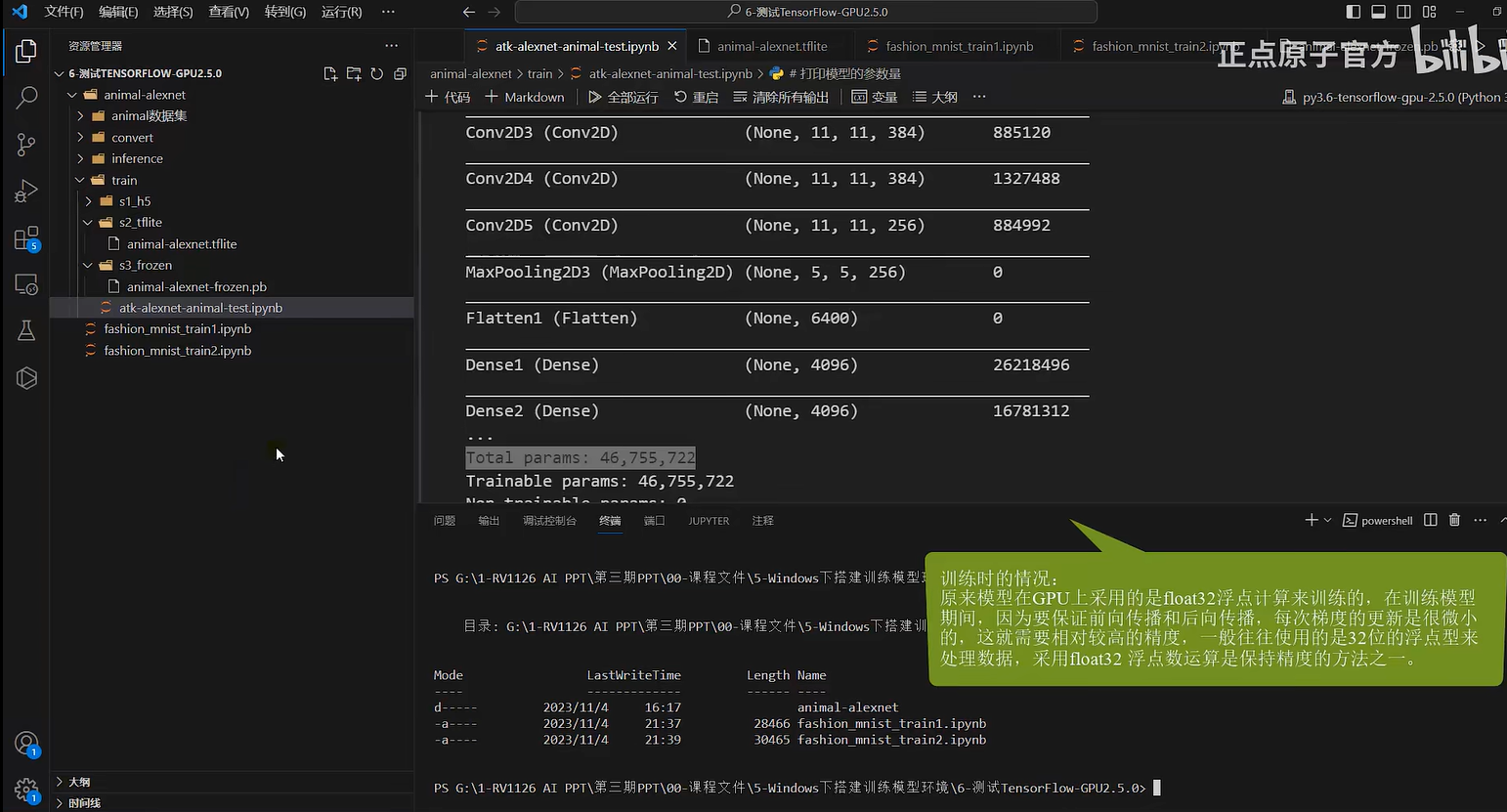
整个网络模型的参数大概有6000万个：

这样的模型，在没有量化之前，原始的模型大小接近200MB！

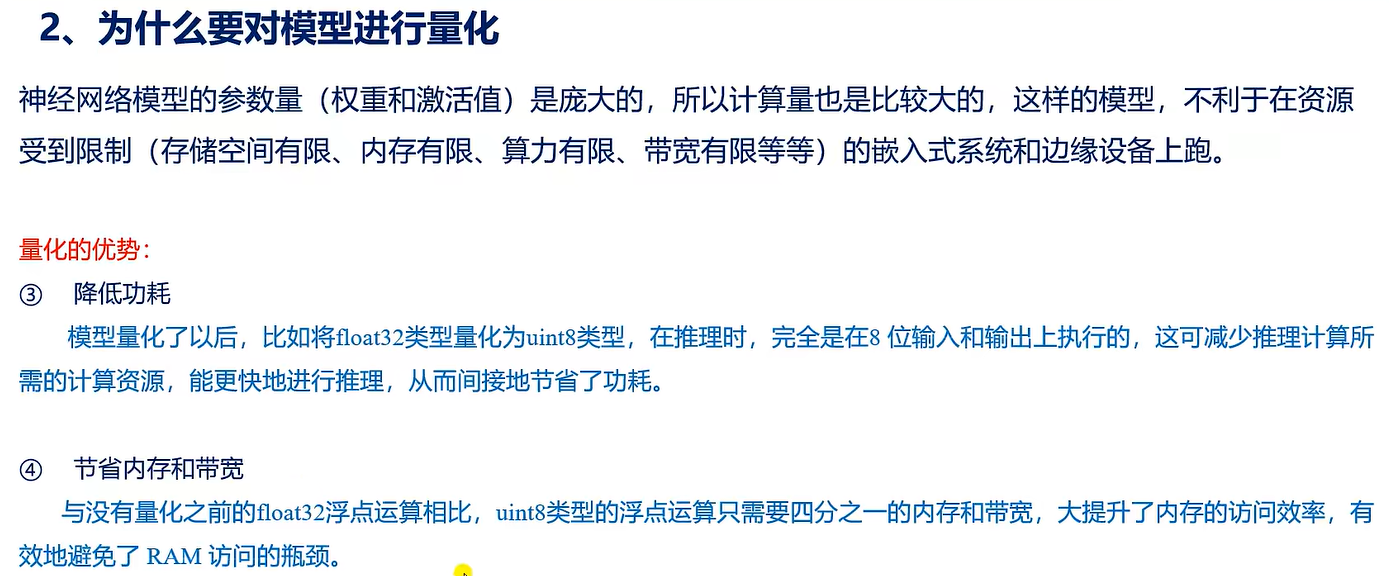


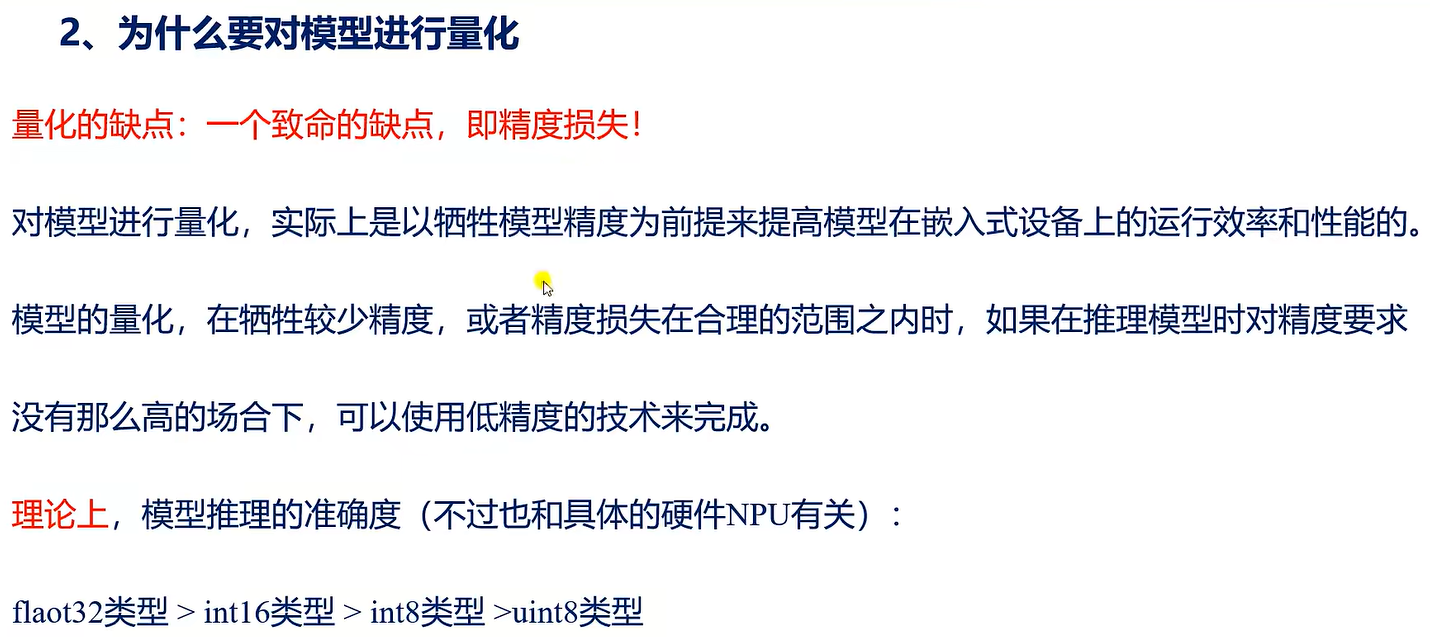
训练时的情况：

原来模型在GPU上采用的是float32浮点计算来训练的，在训练模型期间，因为要保证前向传播和后向传播，每次梯度的更新是很微小的，这就需要相对较高的精度，一般往往使用的是32位的浮点型来处理数据，采用float32浮点数运算是保持精度的方法之一。



量化后的速度提升，对于具有低延迟响应的应用来说非常重要，比如语音识别和实时视频分析应用。







\*/