****

**《人工智能应用基础》课程报告**

**题目名称：基于CIFAR-10数据集、简易人脸识别数据集的CNN多分类模型对比**

1. **摘要**
2. **所选数据集**

本次项目选用两个数据集，第一个数据集为CIFAR-10数据集，CIFAR-10是一个彩色图像数据集。一共包含10个类别的RGB 彩色图片，每个图片的尺寸为32×32 ，每个类别有6000个图像，数据集中一共有50000 张训练图片和10000 张测试图片。

第二个数据集为自己整理的简易人脸识别数据集，共61位同学的照片，每位同学10张照片，照片尺寸为150x150，每位同学的8张照片作为训练集，1张照片作为验证集，1张照片作为测试集。

1. **解决方案**

对于CIFAR-10数据集，采用MLP、CNN、CNN加图像增强、VGG模型+图像增强共计4种模型加以训练，对比四种模型的特点和准确率，对VGG模型进行简单的调参和改进期望取得更高的准确率。

对于简易人脸识别数据集，采用图像增强扩充数据集后，分别采用CNN模型、VGG-16模型，InceptionV3模型来进行训练，其中CNN模型主要测试不同的梯度下降函数、不同的激活函数、不同的卷积核数、不同的全连接神经元数对准确率的影响，对比得到较好的参数模型；VGG-16模型采用迁移学习、预训练等方法，提高识别的准确率；InceptionV3模型则更多的是对比的扩展，测试冻结不同的层数对训练模型的影响。

1. **实验结果概要**

在不断优化参数和神经网络结构的基础上，CIFAR-10数据集的准确率已经能够达到70%以上，简易人脸识别数据集已经能够达到93%左右的准确率。

同时我们分析了可能影响识别的特征，可能是由于背景等较大的特征影响了模型的整体推断。

VGG模型的参数较大，不能够很好的适配于CIFAR-10数据集的训练，所以我们对其filter和conv以及部分结构做了改变和调参，实现了一个类VGG模型的CIFAR-10多分类模型。

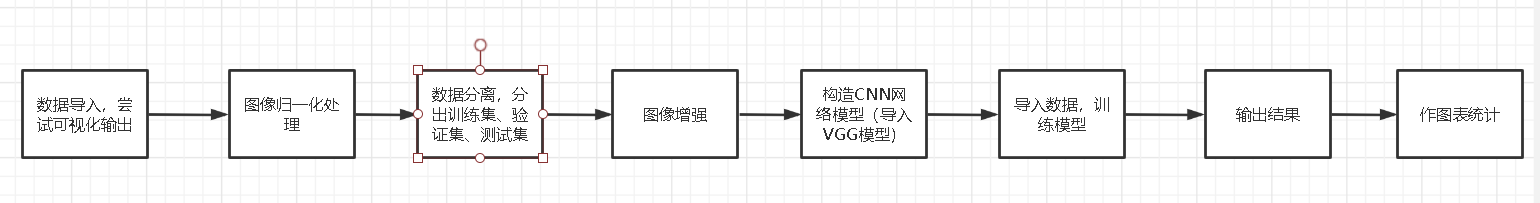
在其他各种参数的测试方面，我们发现Relu激活函数效果较好，同时tanh函数在一些的情况下也有很好的效果。同时RMSprop梯度下降函数也有较好的效果。卷积核的大小越小，对局部特征的抽取越好，正确率相应偏高等。

1. **任务分析**

我们的项目分为两个阶段，第一个阶段采用的数据集是CIFAR-10的标准数据集，采用四种神经网络模型（或者应用图像增强方法）试图对比不同的神经网络模型对识别和分类的影响，试图修改参数使问题得到较好的accuracy。

第二个阶段，由于CIFAR-10数据集规模较大，采用VGG模型训练时间较长，进行相关参数的调整测试效果消耗时间过长，故缩小数据集规模，采用收集到的61位同学的610张照片作为数据集，其中每位同学8张作为训练集，1张作为验证集，1张作为测试集。对61位同学的照片进行训练，达到人脸识别（照片分类）的效果，同时更细致的调节网络模型的参数和结构期望达到更好的效果。

**整体的代码流程图为：**



**第一阶段：**

基于CIFAR-10数据集的测试阶段，我们分别采用了：

1. **MLP（多层感知器）模型：**

MLP是最简单也是最常见的神经网络结构，是所有神经网络结构的基础，MLP神经网络是常见的ANN算法，由一个输入层，一个输出层和一个或多个隐藏层组成，MLP神经网络将一个特征向量作为输入，将该向量传递给隐藏层，通过权重和激励函数来计算结果，并将结果传递给下一层，最后将结果传递给输出层。

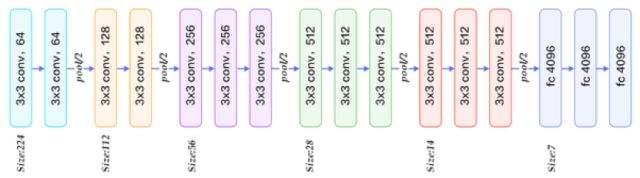
1. **简单CNN（卷积神经网络）模型：**

卷积神经网络主要是由输入层、卷积层、激活函数、池化层、全连接层、损失函数组成的，实质就是特征提取和决策推断的过程。要使特征提取更加准确，就需要将这些网络层结构进行组合。本次项目采用的是3层卷积层，每层卷积层链接1层池化层，最后进行Dropout和全连接层。

1. **图像增强+CNN模型：**

在简单CNN模型的基础上，我们添加了图像增强方法，改善图像的视觉效果，增强图像的有效信息，对图像增添翻转、对称等图片，相当于扩充数据集，期望在上一个方法的基础上提高识别的准确率。

1. **采用迁移学习，图像增强+VGG模型：**



VGG16的网络结构，VGG-16标准输入是224 \* 224，经过5次max pooling和13次卷积后变成7 \* 7 \* 512，cifar10数据集的图片大小只有32 \* 32，所以少进行一次max pooling，最后得到2 \* 2 \*512的结果传入全连接层。

**第二阶段：**

基于我们获取的61位同学的共610张照片数据集，分别采用下述方法：

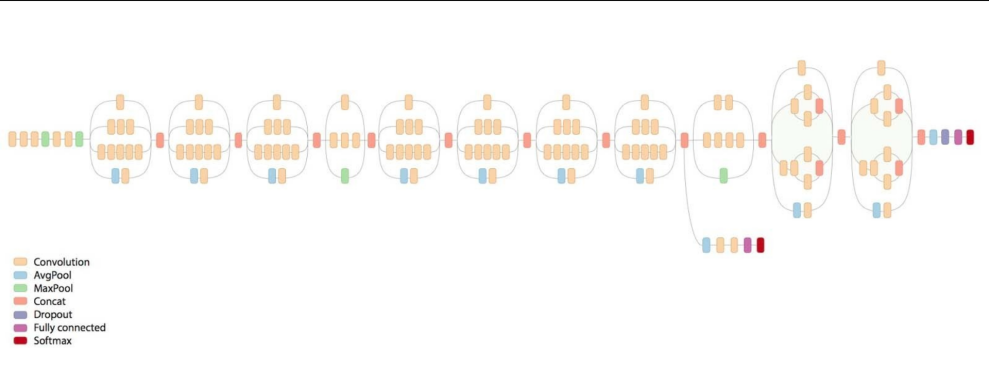
**1）CNN + 图像增强**

本数据集样本容量较小，故采用图像增强，对每个训练集的图片进行旋转等变换，对数据集进行扩充，并利用自己构建的简单CNN模型进行测试，同时测试Dropout的参数对准确度的影响、测试不同的梯度下降优化算法，包括RMSprop，adadelta等方法，分析对结果准确度的影响。同时，修改卷积核的大小对特征提取的区别、对比全连接神经元个数和激活函数的选择对准确率的影响。

**2）VGG模型**

在VGG16的基础上，对靠后层的卷积神经网络进行当前数据集训练和检验，同时根据需要改变全连接层，对数据进行分析，以及进一步调参。

**3）InceptionV3模型**

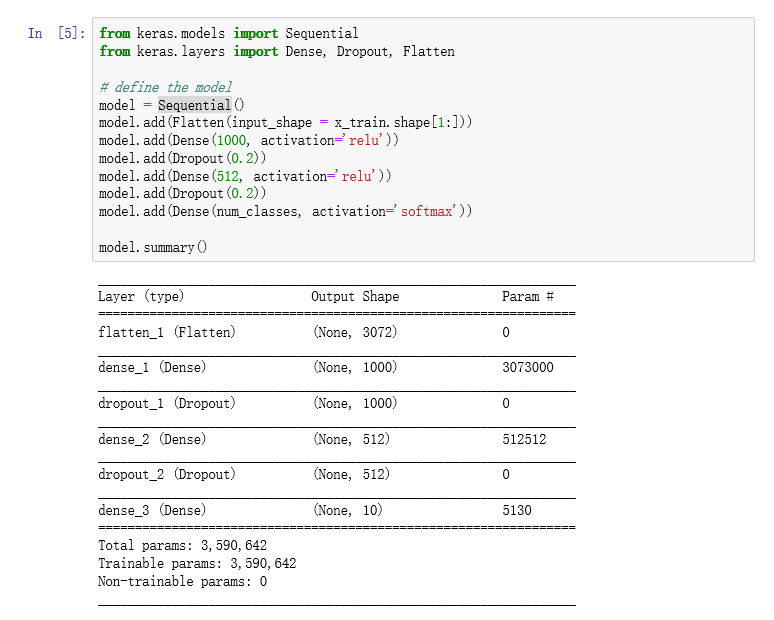


InceptionV3 网络是由 Google 开发的一个非常深的卷积网络，其主要思想是使用多个尺度的卷积核提取特征，并在同一模块中同时计算 1×1、3×3 和 5×5 卷积，然后将这些滤波器的输出沿通道维度堆叠并传递到网络中的下一层。我们设想在InceptionV3的基础上，对靠后层的卷积神经网络进行已有的数据集的训练和检验（前面的网络层数考虑冻结），以期达到比vgg更好的效果。

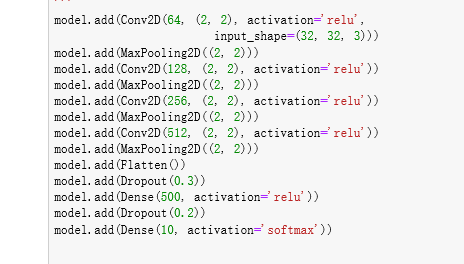
1. **方案设计**

提出自己的解决方案，以流程图、伪代码、框图等方式描述相应的算法设计思路。

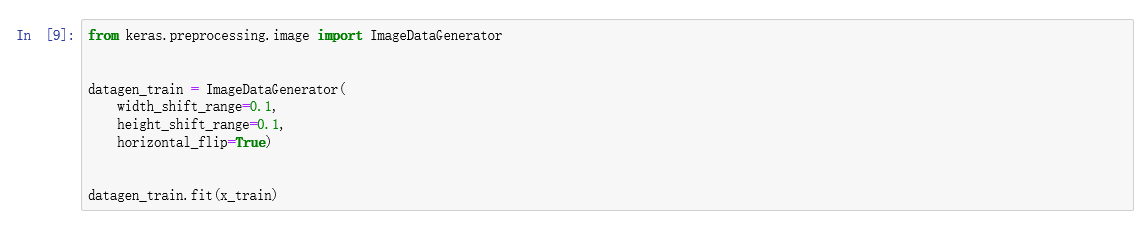
**CIFAR-10的MLP模型：**直接将数据展开成一维，用全连接层Dropout最后输出



**CIFAR-10的简单CNN模型：**采用relu作为激活函数，卷积层与池化层交替的网络结构。



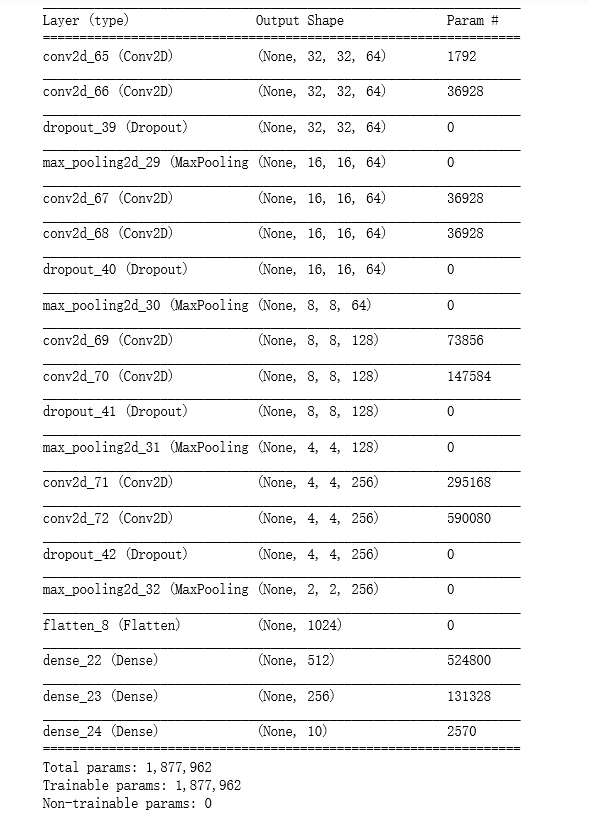
**CIFAR-10的数据增强+CNN的模型：**首先进行数据增强，这里简单的采用翻转的增强，因为CIFAR-10数据集中图像基本不存在颠倒的情况，但大多存在左右朝向混乱的情况，且数据集本身规模较大，不需要额外补充大量数据，故进行简单的增强，后续CNN神经网络结构与前者类似。



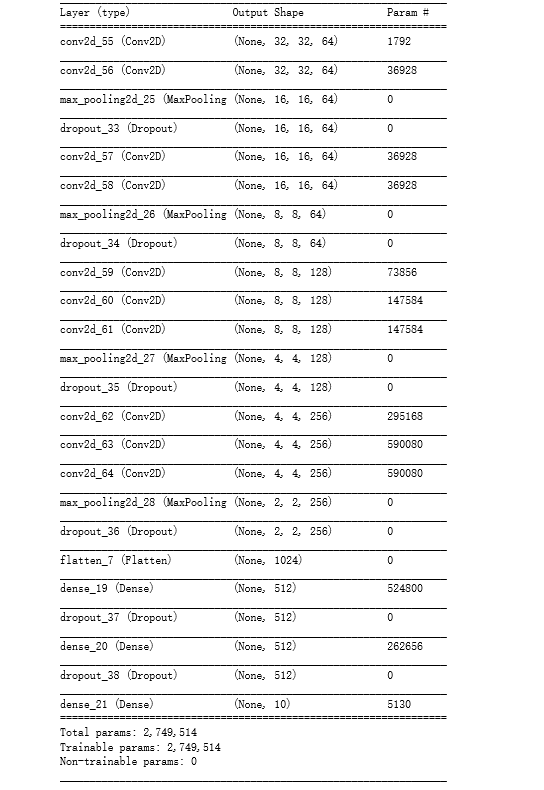
**CIFAR-10的数据增强+类VGG模型：**首先进行数据增强，后续神经网络结构采用迁移学习VGG模型，但是VGG-16的标准输入是224\*224，经过5次池化和13次卷积操作后变成7\*7\*512，而CIFAR-10数据集的尺寸是32\*32，所以我们可以少进行一次池化，最后得到2\*2\*512结果传入全连接层。

同时，由于标准的VGG-16模型含有大量的参数，直接使用会导致训练时间过长，epoch根本无法调大来增加准确率，所以本次是自己仿制的类VGG模型结构，整体结构接近VGG模型，但神经元数量及层数略有不同，总结来说共以下3种类VGG模型：



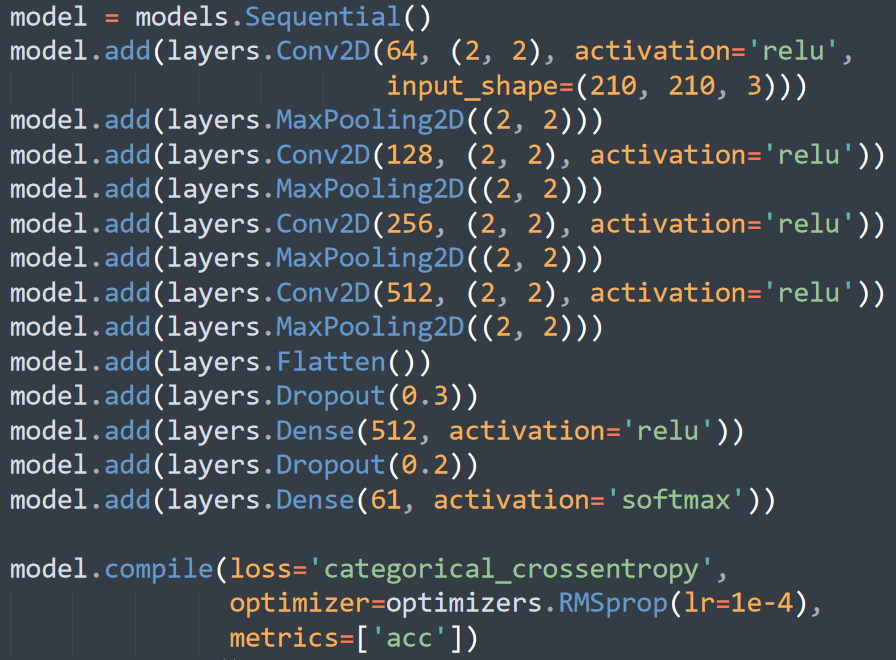




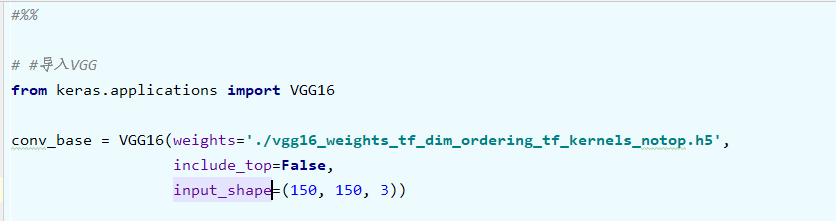


**人脸识别数据集CNN模型：**

采用简单的CNN模型，重点在于改变网络结构的激活函数、Dropout、梯度下降算法卷积核的大小、全连接层的神经元个数等参数去细致对比各参数和各函数对模型的影响，由于测量的次数较多，故此处只提供后续accuracy较高的一组模型的结构。

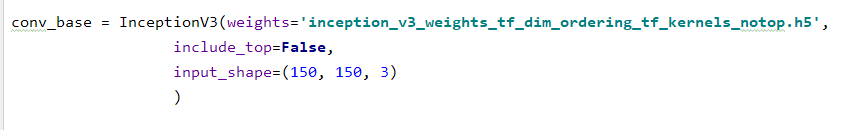


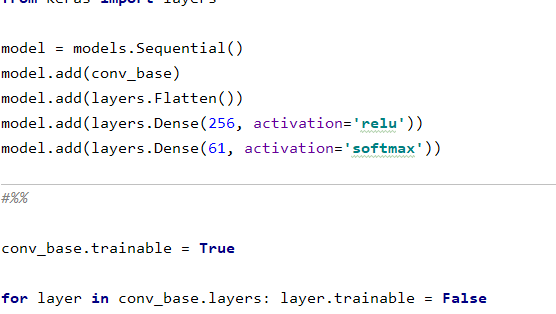
**人脸识别数据集VGG-16模型：**在VGG16的基础上，对靠后层的卷积神经网络进行已有的数据集的训练和检验，以达到既利用VGG对于边缘处理的优越性，又能对我们所面对的数据集测试集有较好的处理能力，获得较高的Test\_accuracy。

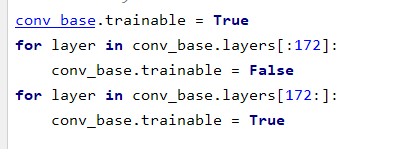




**人脸识别数据集InceptionV3模型：**对于引用的Inception\_V3，我们设计了不同的全连接层以及输出层，来寻找较优良的整体模型。我们所搜集到的资料往往描述将172层前面的部分完全冻结，并且这个172是一个比较好的参数，而我们在172周围进行了尝试。



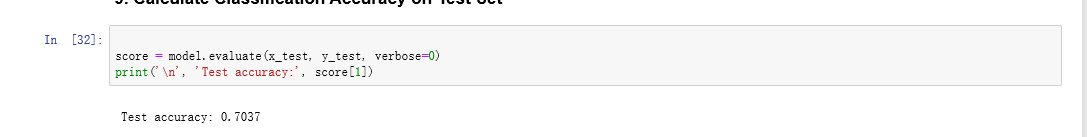




1. **实验结果及分析**

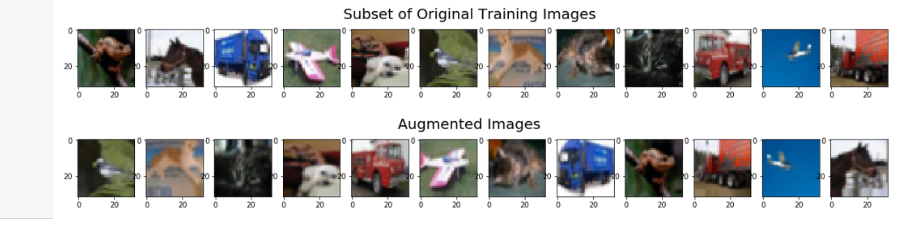
**（1）CIFAR-10训练准确率：**

CIFAR-10数据集训练中，DNN全神经元网络MLP模型的准确率能达到40%以上，简单CNN卷积神经网络的准确率能达到60%以上，图像增强后的简单CNN模型能达到65%~70%以上，图像增强后的类VGG网络的准确率在50%~75%以上，随参数的调整范围较大。



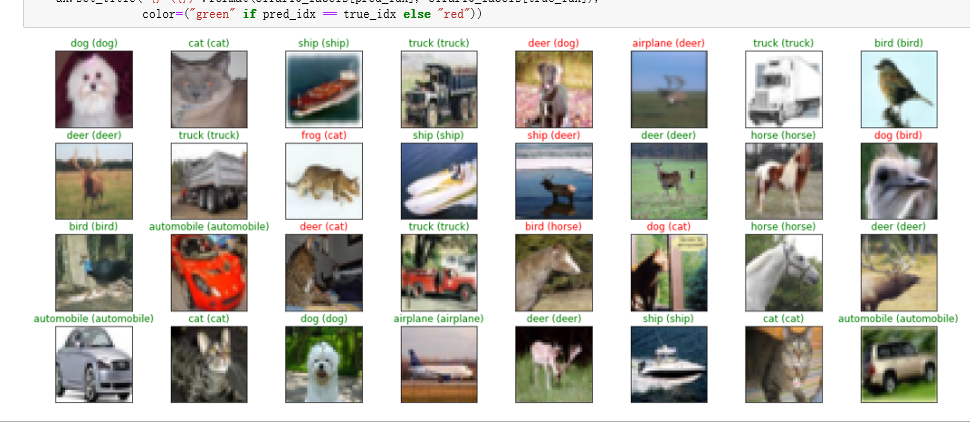
**（2）CIFAR-10数据增强：**

应用图像增强，可以发现，原训练集例如卡车头朝右，简单的图像增强后，通过翻转，我们的训练集中补充了卡车头朝左的数据，这是很关键的一点，因为我们训练的过程实际上是一个特征提取，识别推理的过程，即使我们的模型能够很完美的辨别 卡车和公交车 的细节特征，但如果初始卡车均为朝右，初始公交车均为朝左，测试集为一朝左的卡车，我们的模型依旧会将其识别为公交车。因为图片中类似朝向这样的特征非常明显，且不具备实际价值，所以应用图像增强能够有效避免类似情况，提高准确率。



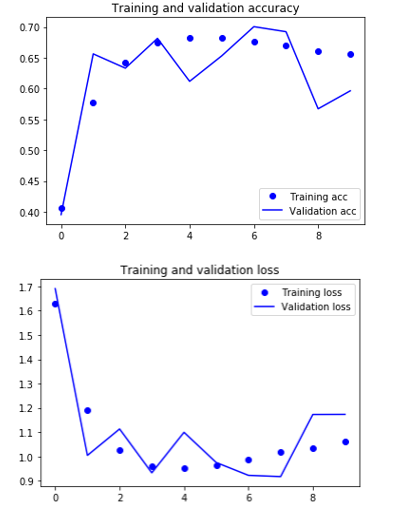
**（3）影响CIFAR-10数据集准确率的因素：**

分析可能影响识别准确率的原因，下图是部分验证集数据，分析识别错误的数据我们发现，大多是由于环境引起的。例如大部分deer的背景是绿色的森林、而图中的dog背景恰好也是偏绿的背景，则dog被误识别成为deer；同理第二行的deer由于背景偏蓝，被误识别成ship。卷积神经网络卷积层作用实际就是提取特征，对于像背景颜色这样的特征会对结果有很大的影响，这是很难避免的，只能依靠更深或者更多的训练次数来弥补。



**（4）CIFAR-10训练过程中acc和loss对于epoch轮数的变化：**

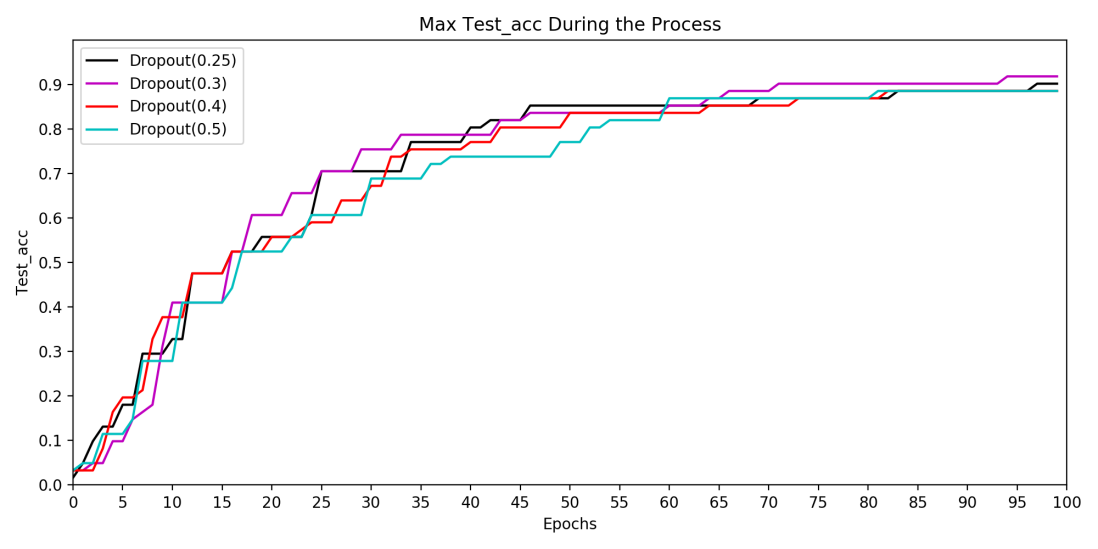
Acc和loss图像，观察acc图像可以看出，在前几次的训练中，准确率是有比较明显的上升，之后存在波动，在训练次数epoch达到7,8轮左右，开始出现准确度大幅下降的状态，即过拟合状态。相应的，误差loss也存在类似的波动情况，可以考虑修改学习率，修改激活函数、修改drop\_out函数使隐含层随机失活来使神经网络正则化等来优化。



由于使用VGG模型训练CIFAR-10数据带来的时间太久，后续我利用人脸识别数据集，进行后续的训练。

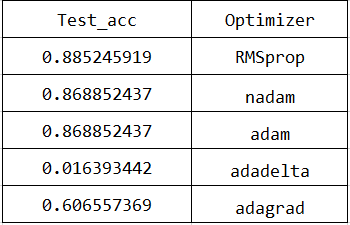
**（5）Dropout After Flatten的值对准确率的影响：**

利用VGG模型训练照片数据集，首先我测试了在展开层后进行Dropout的值对训练准确率的影响，发现在Dropout=0.3时，后续每轮训练的准确率较高，且训练效率也得到提升，当然由于本数据集规模有限，肯定存在一些误差。

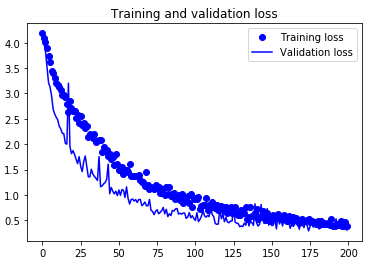


**（6）梯度下降算法的选择：**

经过测试，梯度下降优化算法中，adadelta梯度下降方向可能出现了问题，这直接导致程序并未找到合适的Loss最低点方向，进而导致准确率极低。RMSprop准确度最高。



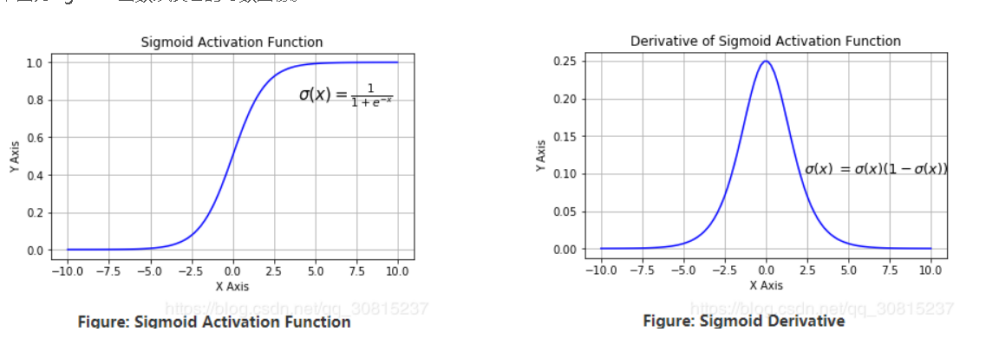
**（7）全连接层神经元个数和激活函数的选择对准确率的影响：**

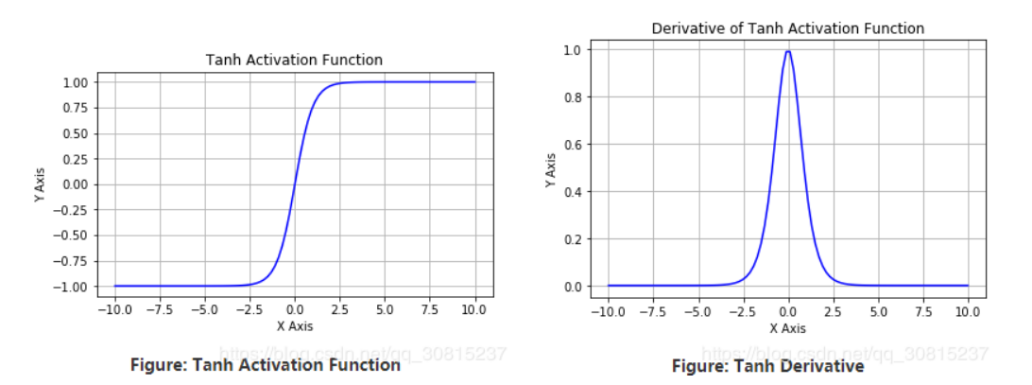
在训练和调参过程中，我们发现在我们编写的神经网络中，当全连接层神经元个数为1024、激活函数为tanh时，损失函数较小，即准确率较高、程序较优。

对比各个激活函数，主要运用到的有sigmoid、tanh、relu激活函数；

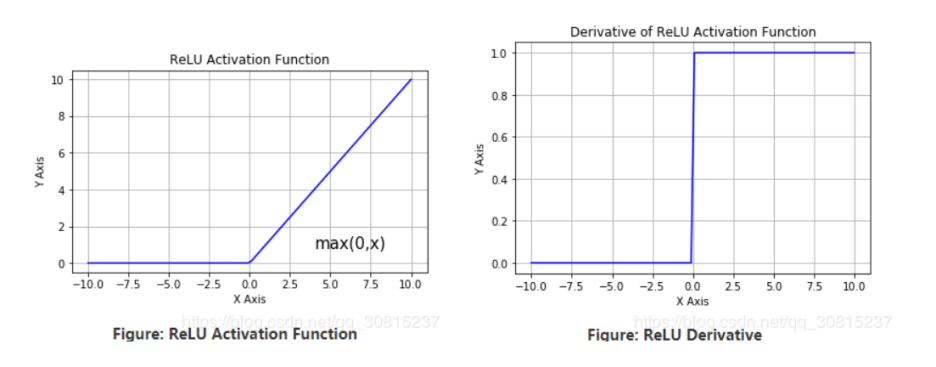
Sigmod函数：



Tanh函数：



Relu函数：

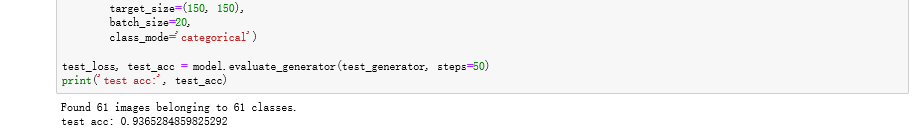


对比三种激活函数，relu函数是目前使用较频繁的激活函数，收敛速度更快、计算速度较快，在正数范围内可以有效减小梯度消失带来的问题，但是在负数范围内容易导致神经元失活，一般不会出现这样的情况，只有当学习率设置较大时会出现。

Tanh函数类似sigmod函数也是实数值压缩，与sigmod不同的是在-1到1之间输出范围是零均值的，而sigmod在-1到1之间函数值变化敏感，所以实际中tanh的效果是要优于sigmod的，但是二者同样存在梯度消失问题。

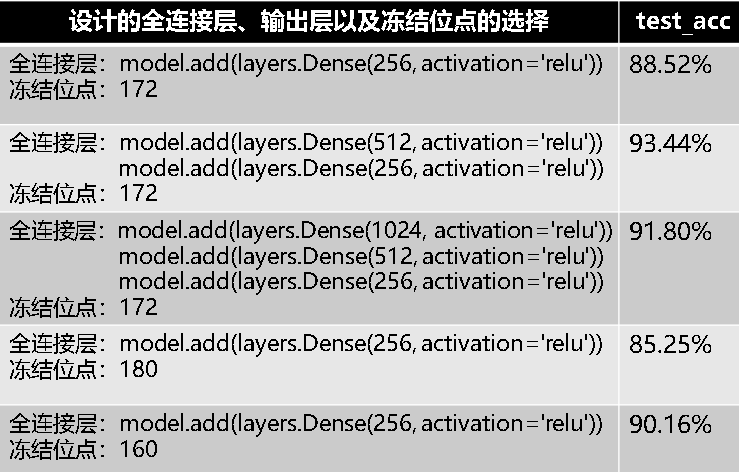
**（8）人脸识别VGG-16模型**

在本次VGG模型的训练之中，我们采用了预训练的方法，根据搭建的网络首先设一组初始化参数，开始训练后不断调整直到损失越来越小，训练过程中初始的参数会不断变化，并将当前阶段的训练模型的参数保存下来，如果后续还有类似任务可以直接使用保存下来的模型参数作为初始化参数进行训练。当然，实际训练中，也可以先冻结前面的一些层，训练几轮之后再解冻，进行联合训练，以此达到较好的效果，我们目前能够使用vgg16模型进行简单的预训练，达到最后的accuracy达到90%上下。



**（9）人脸识别InceptionV3模型：**

InceptionV3 网络是由 Google 开发的一个非常深的卷积网络，其主要思想是使用多个尺度的卷积核提取特征，并在同一模块中同时计算 1×1、3×3 和 5×5 卷积，然后将这些滤波器的输出沿通道维度堆叠并传递到网络中的下一层。我们在InceptionV3的基础上，对靠后层的卷积神经网络进行已有的数据集的训练和检验，以期达到比vgg更好的效果，我们所搜集到的资料往往描述将172层前面的部分完全冻结，并且这个172是一个比较好的参数，而我们在172周围进行了尝试，发现当值为180时，确实准确率更低，而当值取160时，准确率反而上升了 。



1. **总结**

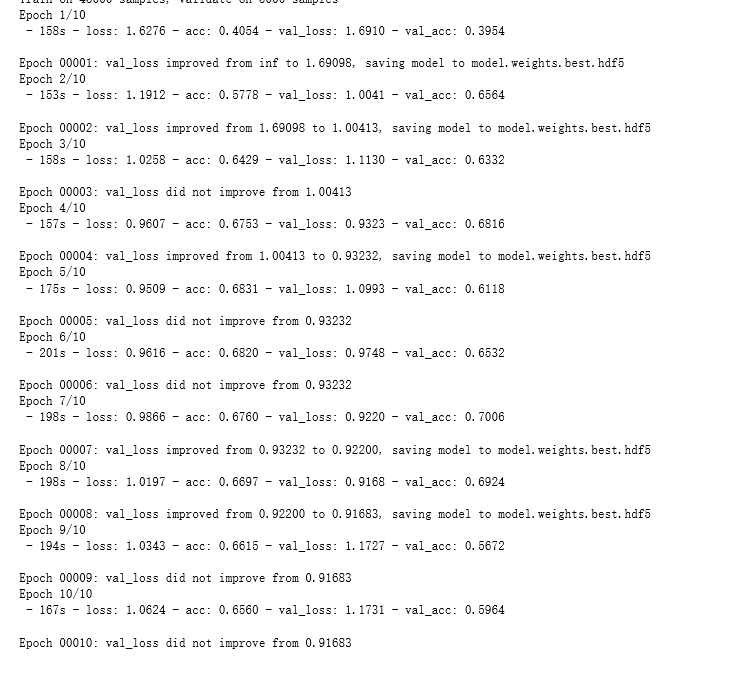
1.在实现方面，由于keras，tensorflow的集成度已经很高，几乎不需要自己去实现过多的代码，更多的是理解结构和各个参数的意义，难点不在于实现，而在于理论的学习，和参数的调节过程。

2.本次实验也涉及到了普通深度神经网络的测试，简单来说采用多层感知器全连接的方式，逐层提取特征，但是存在一些劣势，CNN则对相关问题有很好的优化，主要体现在局部感受野、权值贡献、卷积和池化过程。简单来理解，卷积层的作用就是提取局部特征，池化层的作用就是做一个采样，减少参数，全连接层的作用则相当于分类器，卷积层和激活函数的作用是提取特征并将特征映射到某一特征空间，而全连接层就是整合全局的信息，将分布式特征表示映射到标记空间中的作用。

3.产生一个问题， 卷积神经网络能够识别图像，其理论依据是卷机可以识别边缘、纹理等特征，然后通过不断的卷积，提取出抽象的特征，最终实现图像识别。但是一些诸如阴影之类的没有边缘和轮廓的东西可能会比较难以提取，需要对图像进行膨胀或者腐蚀，进行更强的增强，这也是说明了训练集的优劣对模型的准确率同样起到重要的影响，所以对训练集进行图像增强是比较有必要的。

4.我们发现不论是CIFAR-10数据集还是自己搜集的人脸识别数据集，都存在识别错误的情况，且比较明显。比较显而易见的影响是，CIFAR-10的分类中，背景起到影响识别的一大特征，在人脸识别中，衣服的颜色，是否佩戴眼镜等也成了影响识别的一大特征，即有一个比较大的特征阻碍了正确的分类造成识别不够准确。我的理解是，因为我们要找的目标特征太小了，因为卷积这个方法本来就是不断的提取特征，每一步卷积从某种程度上都会丢失一部分原图的信息，如果说当你要找的目标特征只有5\*4这么大的话，在卷积的过程中它就有可能被丢失掉，很难将它提取出来。像这种感觉用一些图像处理的方法做起来会好一些。

5.总结了一点调参的小事项1）epoch的取值不宜过大，过大将导致过拟合反而影响准确率，比较好的方法是先取一个较大的数值，在训练的过程中观察loss和acc的情况，如果acc有比较明显的降低，可以考虑终止并缩小epoch。（例如下图所示）2）卷积层和池化层的增添减少要注意神经元的递减，要注意卷积层的卷积核大小。3）要注意池化层window的大小 4）可以尝试不同的激活函数和loss函数，之后考虑optimizer等等



实验的建议，本学期上课更注重讲课，感觉平常的过程中有些忽视实验部分，在平常的过程中没有及时进行实验进行操作，导致比较生疏，理解也不够透彻，希望可以在平常的过程中安排一些小的，好操作的实验来配合讲课。