**第七次实验报告**

闫嘉依 2021214417

**作业标题：深度学习框架使用**

**作业说明**

CIFAR-10数据集由10个类的60000个32x32彩色图像组成，每个类有6000个图像。有50000个训练图像和10000个测试图像。数据集分为5个训练批次和1个测试批次，每个批次有10000个图像。与MNIST 数据集中目比， cifar-10有以下特点：CIFAR-10 是3通道的彩色RGB图像，而 MNIST 是1通道的灰度图像。CIFAR-10 的图片尺寸为32 × 32 ， 而MNIST 的图片尺寸为28 × 28 ，比MNIST 稍大。相比于手写字符， cifar-10 是现实世界中真实的物体，不仅噪声很大，而且物体的比例、特征都不尽相同。使用pytorch搭建一个基于cifar-10的分类模型，在实验报告给出模型解释以及训练集的loss曲线和accuarcy曲线。

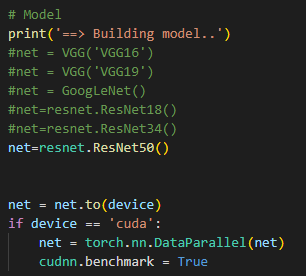
**数据集加载请参考** https://pytorch.org/docs/0.3.0/torchvision/datasets.html#cifar

**一、代码结构**

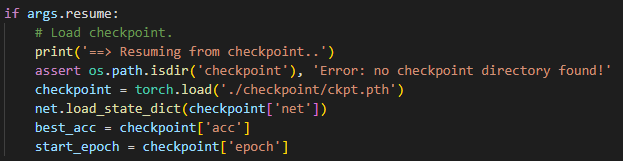
加载数据到data文件夹下，对图片每个通道进行归一化，训练集数据随机裁剪和翻转扩大样本，以减小目标比例大小的影响。



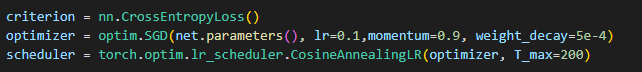
导入模型，放到cuda上



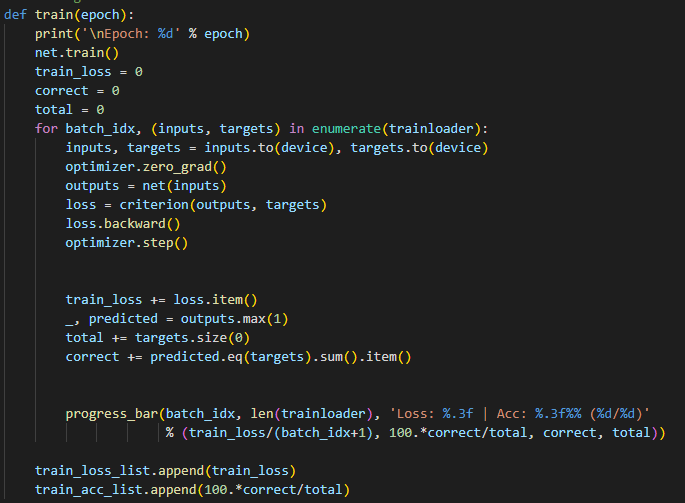
加载没训练完的模型



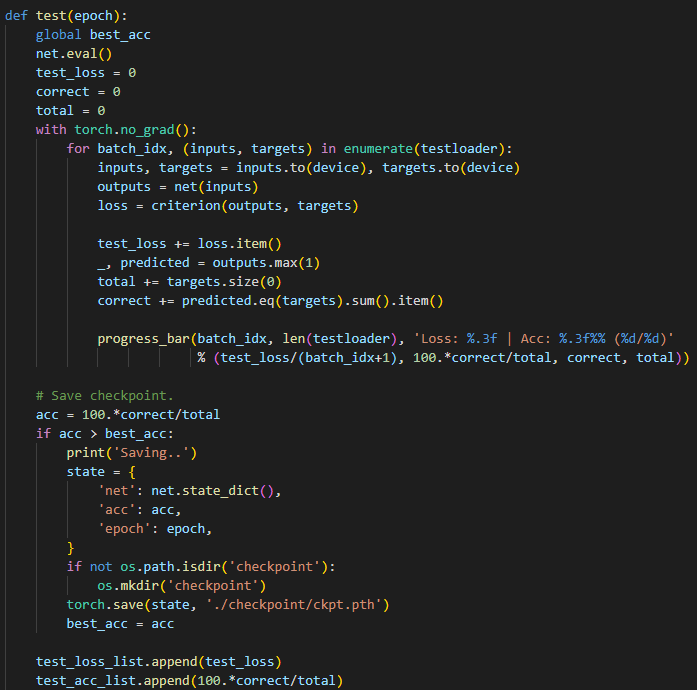
定义损失函数——交叉熵损失；优化器：SGD，其中学习率使用scheduler自动调节



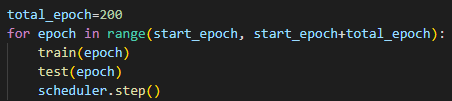
训练过程



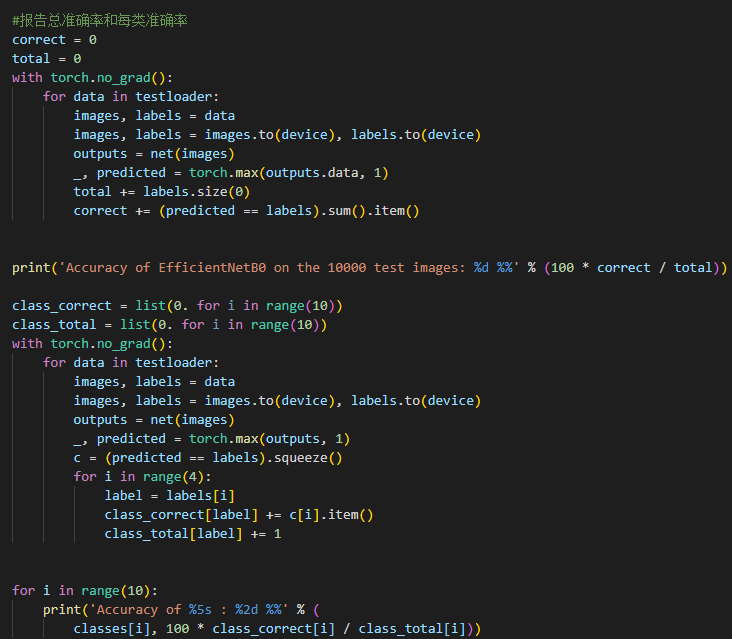
测试过程，测试的过程中保存该epoch结束的模型



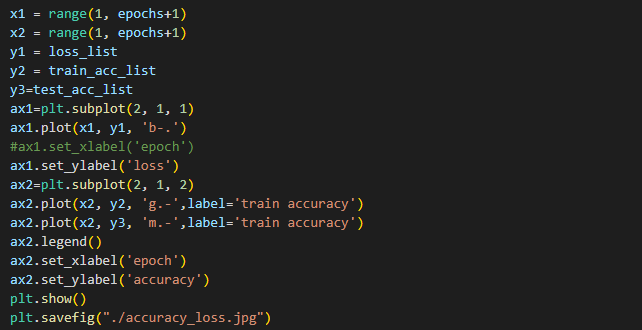
跑模型



报告总准确率和每类准确率



画acc和loss的图



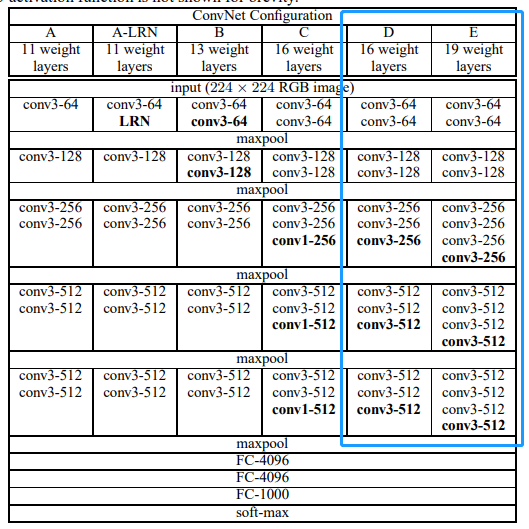
**二、网络结构**

本节对网络框架进行探索，由于数据集比较大，因此选取深层网络结构。

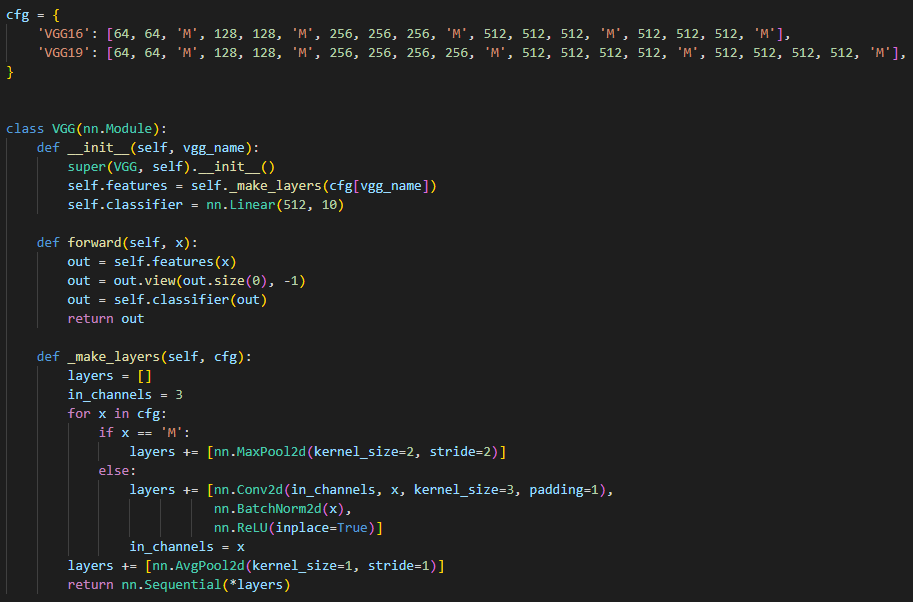
1. VGG

VGG模型说明增加网络的深度一定程度上能够影响网络最终性能。VGG的卷积核大小维3\*3，相比于大卷积核来说，一方面多层非线性层可以增加网络深度来保证学习更复杂的模式，另一方面参数量少，代价小。

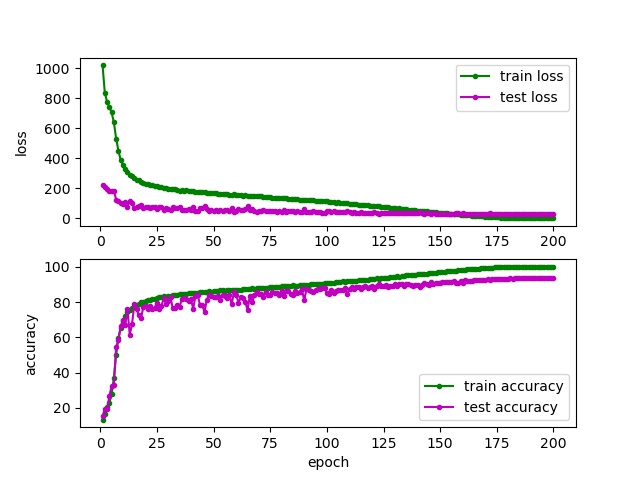
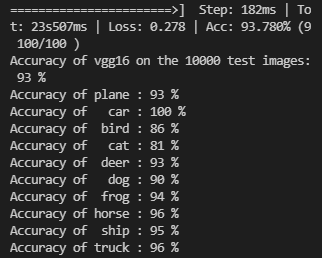
选取VGG16和VGG19分别进行实验，VGG16的网络有16层，VGG19有19层。模型结构如图所示，其中蓝框第一列为VGG16，第二列为VGG19。两个网络输入图像默认尺寸是224\*224，图像以stride=1，padding=1的过程通过一系列卷积层。网络的卷积核大小为3\*3，最大池化层以步长为2通过2\*2的窗口滑动，然后进入下一层网络。卷积层后是三层全连接层，最后通过softmax层输出预测结果。



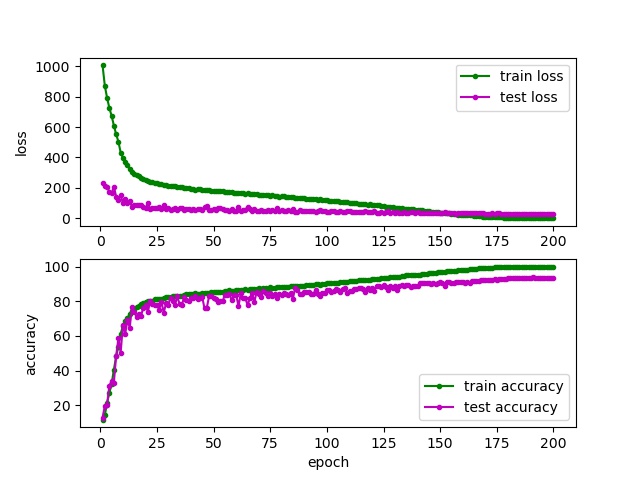
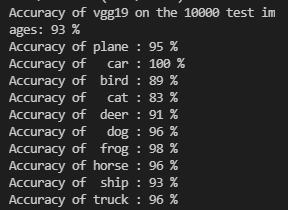
代码实现：每套conv-batchnorm-relu结构用循环表示



Vgg16结果：



Vgg19结果：

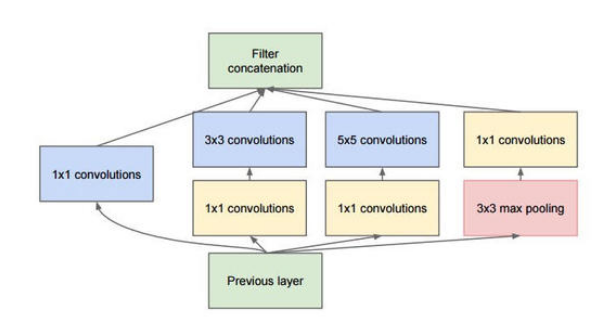


1. Googlenet

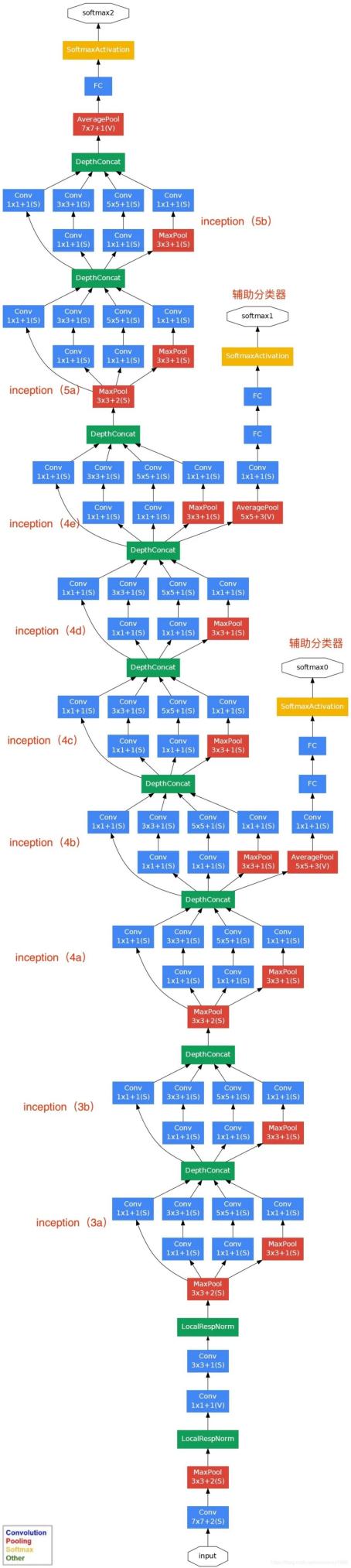
Ｇｏｏｇｌｅｎｅｔ有２２层，所以它是一个非常深的网络，而层数越多，参数量越多，网络容易出现过拟合，且计算量增大，因为滤波器的线性增加会导致计算量二次方递增。因此ｇｏｏｇｌｅｎｅｔ引入ｉｎｃｅｐｔｉｏｎ模块和Ｇａｐ，减少参数量。Ｇａｐ没有需要学习或优化的参数。

ｇｏｏｇｌｅｎｅｔ没有选用特定滤波器的大小，而是将１＊１、３＊３、５＊５的所有三个滤波器和３＊３的最大池化层应用于同一个ｉｎｃｅｐｔｉｏｎ中，并连接到单个输出向量中。１＊１的卷积会减少计算量，因为３＊３和５＊５的卷积核计算量更大。

在ｇｏｏｇｌｅｎｅｔ中，ｉｎｃｅｐｔｉｏｎ模块可以任意添加和删除，不影响后面的层。由于全连接层容易出现过拟合，因此用ｇａｐ层替代。



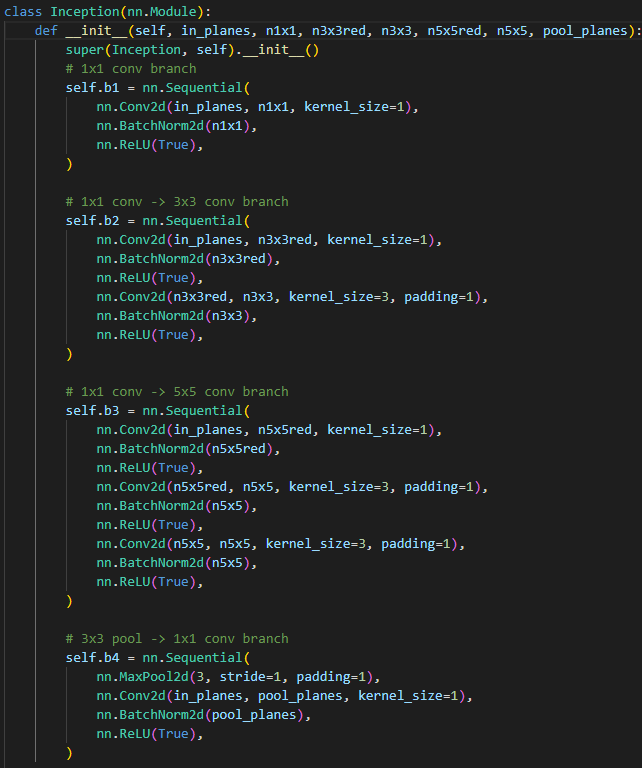
ｉｎｃｅｐｔｉｏｎ模块



ｇｏｏｇｌｅｎｅｔ结构

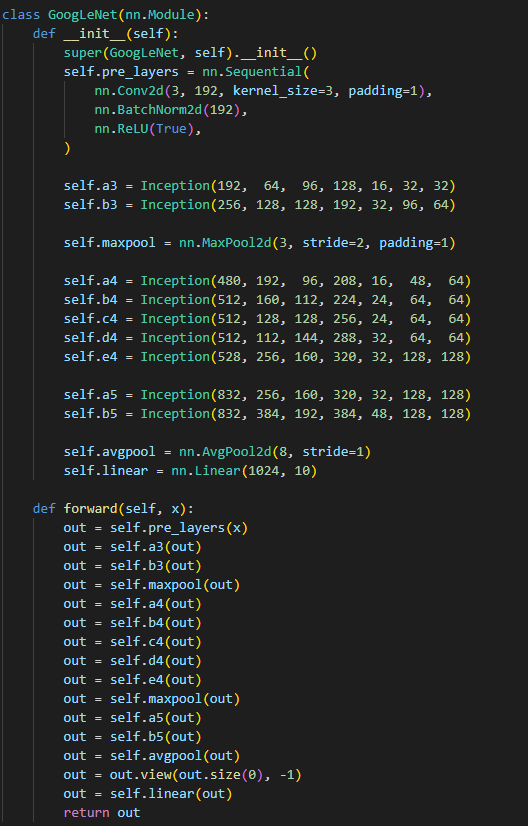
代码实现：

inception结构

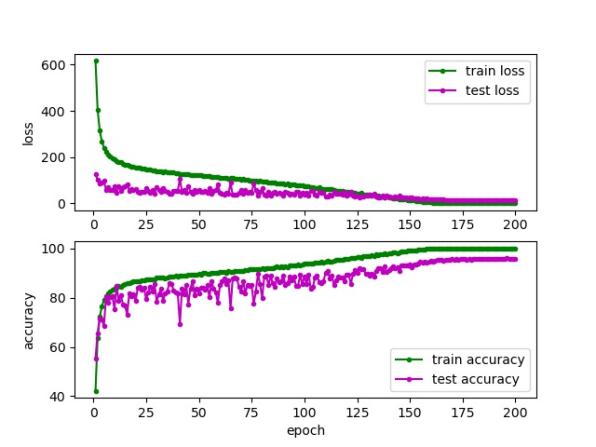
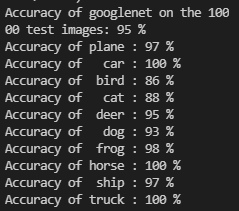




Googlenet



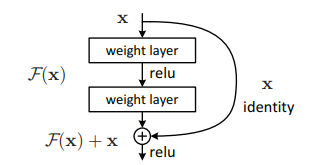
结果：



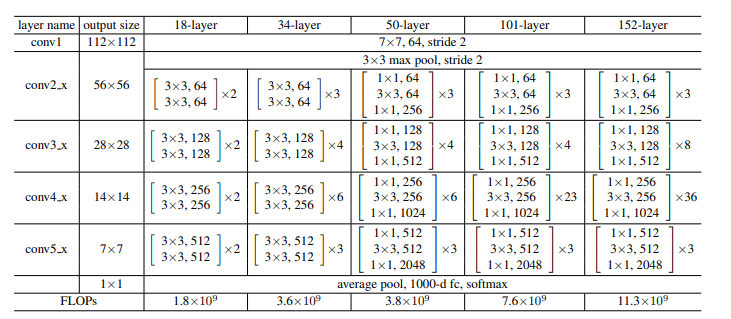
1. Resnet

一定深度过后，CNN会随着层数的加深导致更高的误差，因此性能会随着层数的增加而增加到一定深度，然后迅速下降。作者认为这未必是由梯度消失导致的，因为使用ｂａｔｃｈ　ｎｏｒｍａｌｉｚａｔｉｏｎ时也会出现这种情况。因此，模型在CNN中增加残差块，实现跳连接。

Ｒｅｓｎｅｔ是VGG反复重复后更深的网络，Ｒｅｓｎｅｔ有不同的深度，如１８层（Ｒｅｓｎｅｔ１８）、３４层（Ｒｅｓｎｅｔ３４）、５０层（Ｒｅｓｎｅｔ５０）、１０１层（Ｒｅｓｎｅｔ１０１）、１５２层（Ｒｅｓｎｅｔ１５２）。Ｒｅｓｎｅｔ的每个残差块都有３＊３的卷积层，最后一个卷积层后，添加一个ＧＡＰ层，最后增加全连接层实现分类。对深层次的网络，Ｒｅｓｎｅｔ使用ｂｏｔｔｌｅｎｅｃｋ特性提升效率。



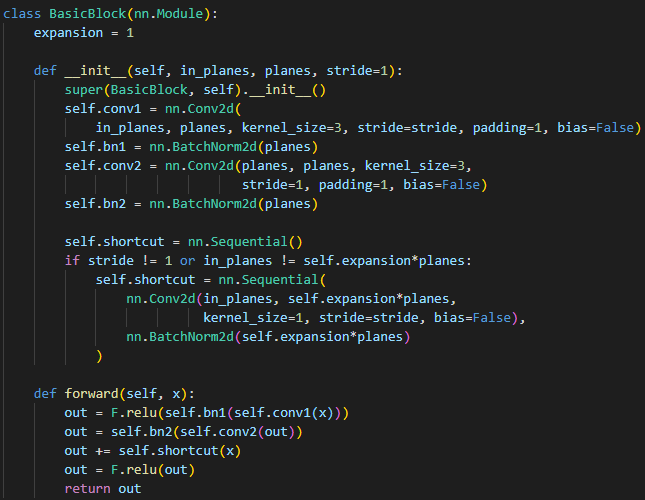
残差块



Ｒｅｓｎｅｔ网络结构

代码实现：

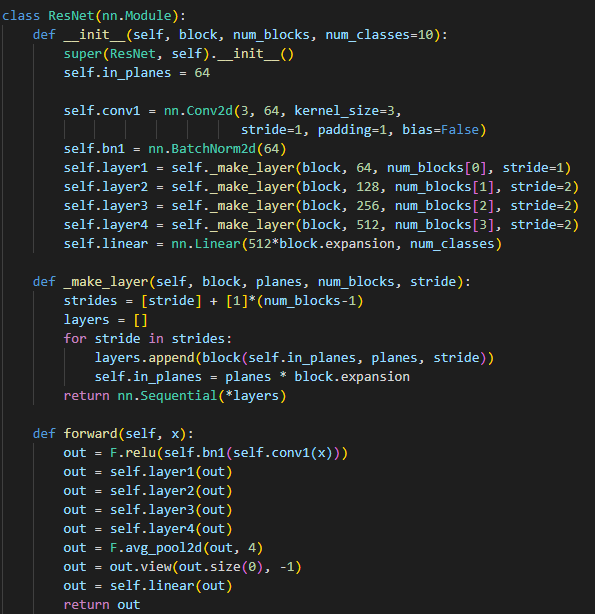
残差块：



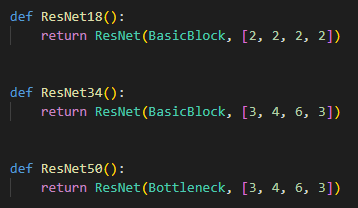
Bottleneck



Resnet网络

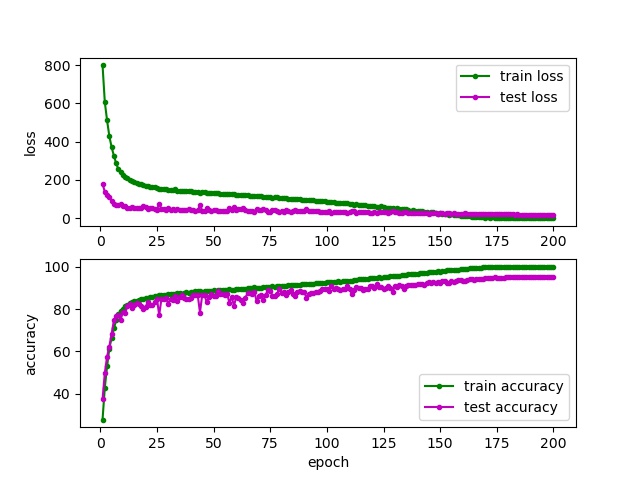
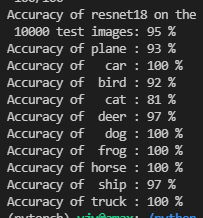


返回不同深度的模型

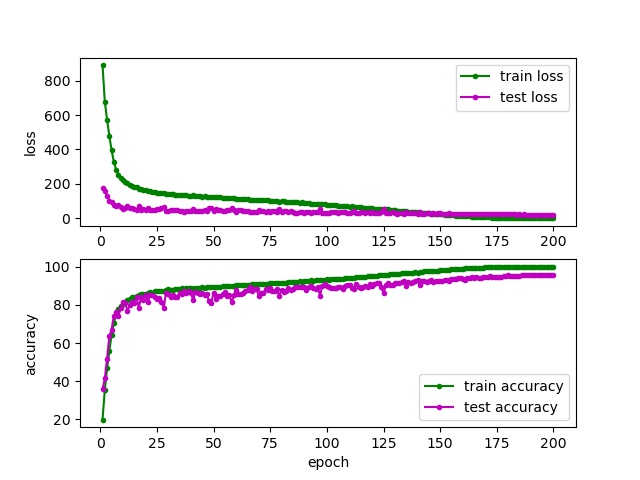
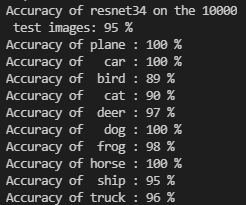


实验对resnet18、resnet34、resnet50模型展开探究。

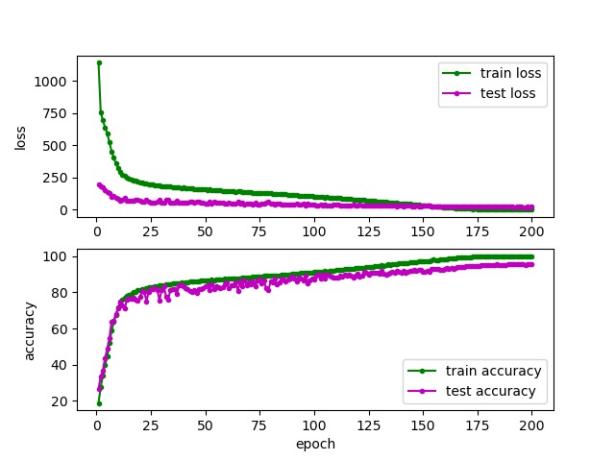
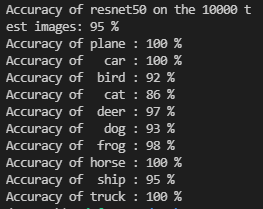
resnet18



resnet34



resnet50



1. **结果汇总与分析**

所有模型准确率如下：

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | plane | car | bird | cat | deer | dog | frog | horse | ship | truck | total |
| VGG16  VGG19  GoogleNet  Resnet18  Resnet34  Resnet50 | 0.93  0.95  0.97  0.93  **1**  **1** | 1  1  1  1  1  1 | 0.86  0.89  0.86  **0.92**  0.89  **0.92** | 0.81  0.83  0.88  0.81  **0.9**  0.86 | 0.93  0.91  0.95  **0.97**  **0.97**  **0.97** | 0.9  0.96  0.93  **1**  **1**  0.93 | 0.94  0.98  0.98  **1**  0.98  0.98 | 0.96  0.96  **1**  **1**  **1**  **1** | 0.95  0.93  **0.97**  **0.97**  0.95  0.95 | 0.96  0.96  **1**  **1**  0.96  **1** | 0.93  **0.95**  **0.95**  **0.95**  **0.95**  **0.95** |

从总准确率来看，模型准确率都在0.93以上，但差别不大，实验应该多保留几位小数能够看出差异。但是没时间重做了...T^T

从每个类别的准确率看，car类容易分类，所有模型准确率都能维持在100%，googlenet和resnet对horse和truck类分类性能更强，同时resnet对plane和dog的分类性能也更强。bird和cat更难分类，只有resnet对这些类别分类准确率能够达到0.9以上。可以看到，resnet结构优于googlenet结构，优于vgg结构。

对于resnet，尽管resnet18对更多的类别分类准确率很高，但是鲁棒性不强，对于cat类和plane类的分类准确率低于模型均值。更深层次的残差网络模型鲁棒性更强。