## 一、教师网络

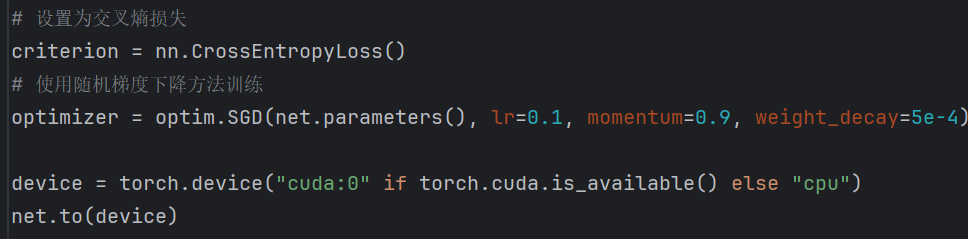
**1.架构分析**

（1）构建基础网络块类BasicBlock()

（2）构建ResNet网络类

（3）搭建教师网络net = ResNet(BasicBlock, [2,2,2,2])

（4）使用交叉熵损失，并使用随机梯度下降方法进行训练



**2.训练结果**

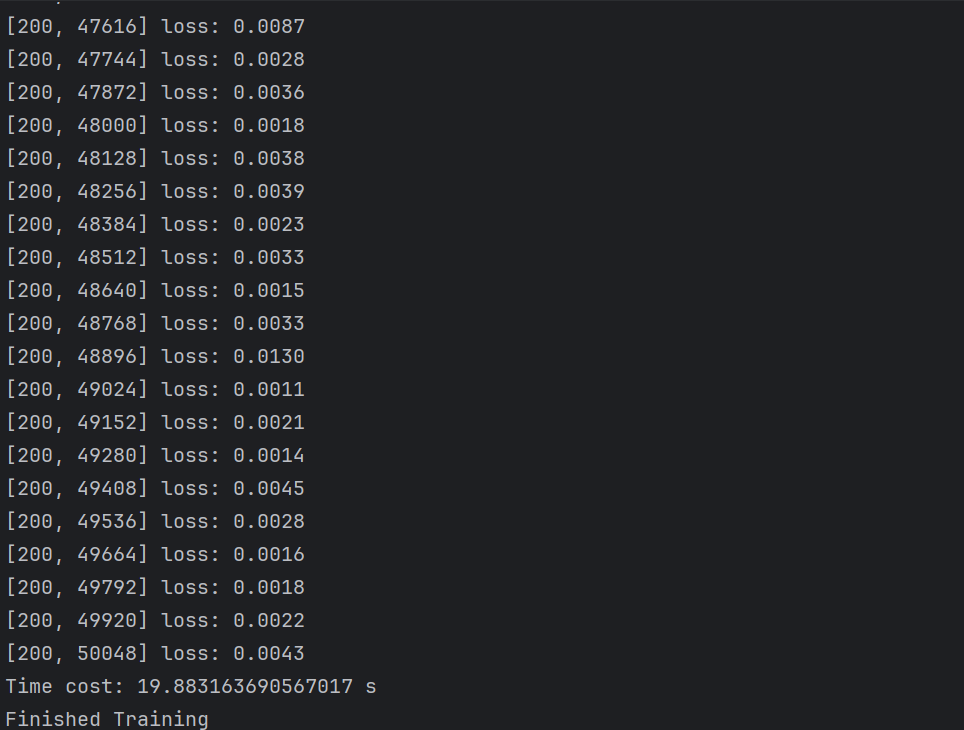
采用CIFAR10数据集进行训练，并对输入图像添加噪声，提高模型的泛化性：



设置批规模为128，迭代训练次数为200，变学习率训练，初始学习率为0.1当迭代次数大于100时，学习率变为0.01，当迭代次数大于150时，学习率变为0.001，保证训练精度更高，训练完成后保存训练参数到.pkl文件，以便学生模型学习使用。

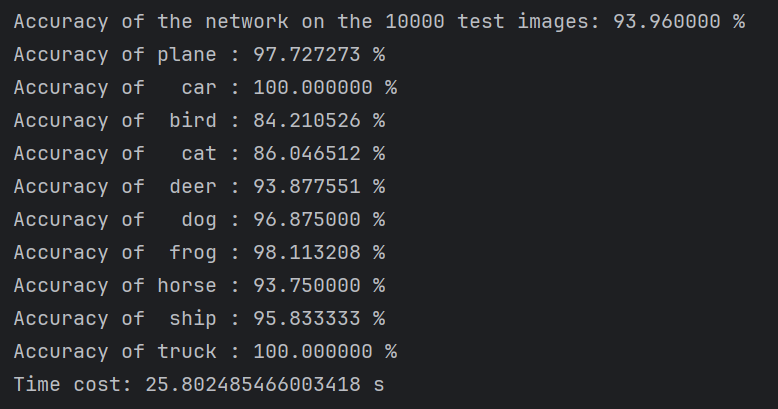


训练结果：



**3.测试结果**

教师网络测试精度达到97.73%



## 二、学生网络

**1.架构分析**

（1）构建基础网络块类BasicBlock()

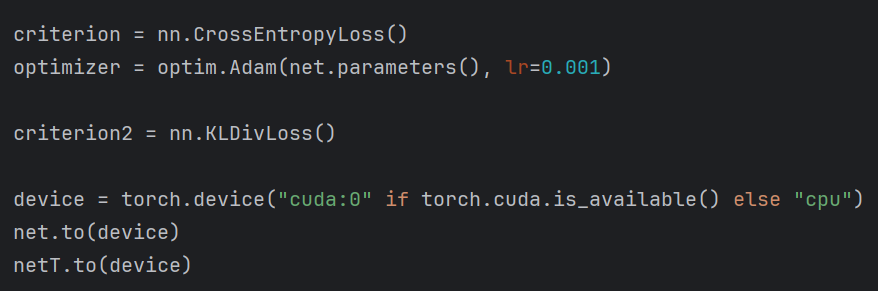
（2）构建ResNet网络类

（3）加载教师网络参数netT = torch.load('saved\_pkl/teacher.pkl')

（4）构建卷积层：\_ConvLayer();CNN()

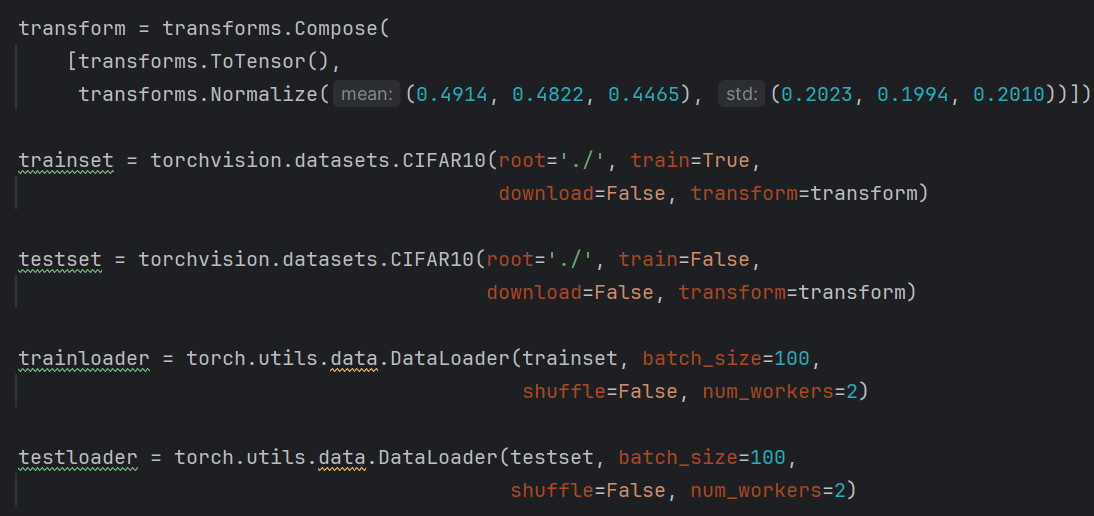
（5）搭建学生网络net = CNN()

（6）使用KD散度作为损失，并使用Adam方法进行训练



**2.训练结果**

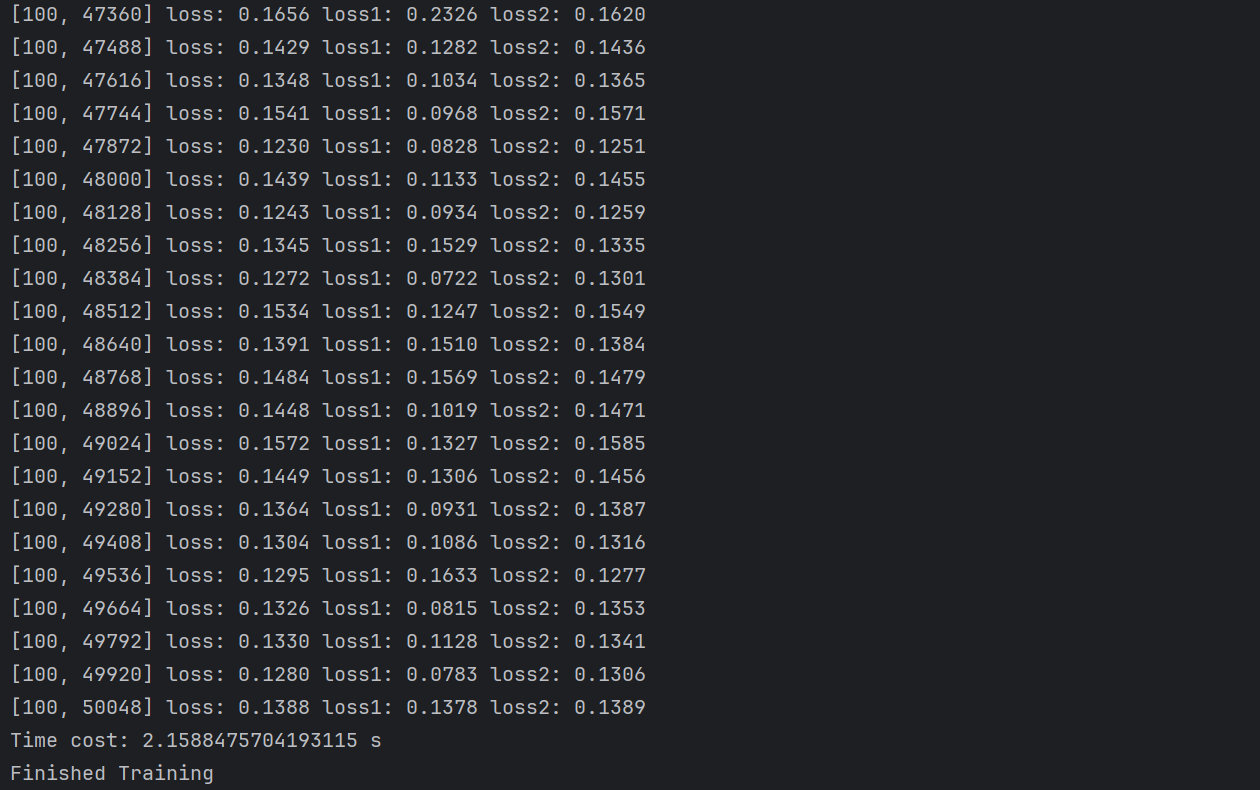
采用CIFAR10数据集进行训练，并不对输入图像添加噪声：



设置批规模为128，迭代训练次数为100，固定学习率训练，学习率为0.001，设置蒸馏温度为T=10，软硬损失权重为α=0.95；训练完成后保存训练参数到.pkl文件，以便后续使用。

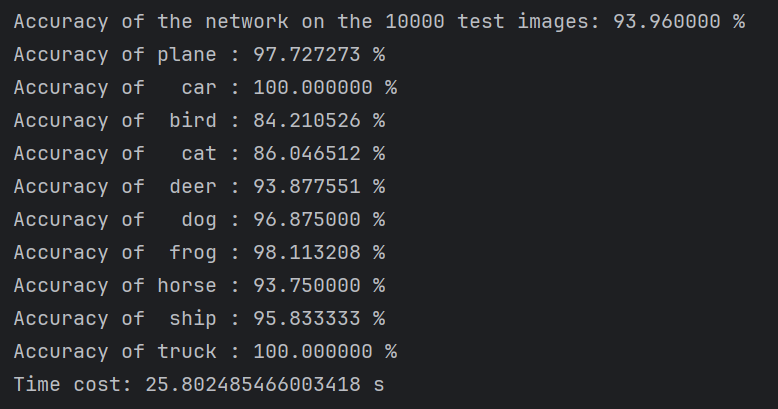


训练结果：



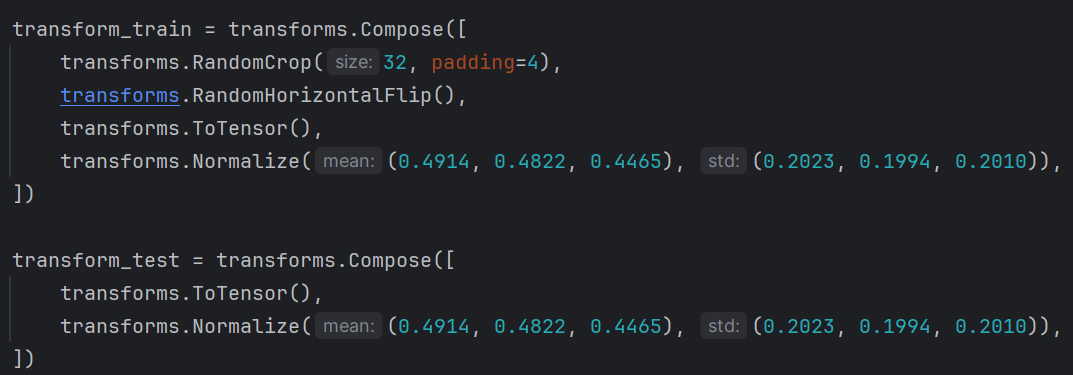
**3.测试结果**

不添加噪声的学生网络测试精度达到78.75%



**4.优化改进**

在学生网络训练时，也预先添加随机噪声，增强学生网络的泛化能力：



同时改变蒸馏温度T=2，进行训练，得到优化后的测试精度达到85.36%：

