**《专业方向前沿课》线上实验/课程大作业 报告**

**(2022-2023 学年第 2 学期)**

**第一单元 物体识别与卷积神经网络基础**

**学生姓名：**

**提交日期：2023 年 7 月19 日 学生签名：**

|  |  |  |  |
| --- | --- | --- | --- |
| **学 号** |  | **任课教师** |  |
| **学 院** | **软件学院** | **专业班级** |  |
| **课程名称** | **《计算机视觉》** |  |  |
| **教师评语：** | | | |
| **本论文成绩评定：** **分** | | | |

**说 明**

1、本次大作业报告需要配套完成线上实验的ipynb文件一同提交

2、ipynb需要包括实验手册2~5节所对应实验一到实验四的全部内容，包括各实验的中间结果，最后结果，以及开放题的实现与结果。

3、本课程大作业报告，重点摘录ipynb笔记中关于第二节、第五节开放性问题的代码细节与实验结果，并补充分析说明。

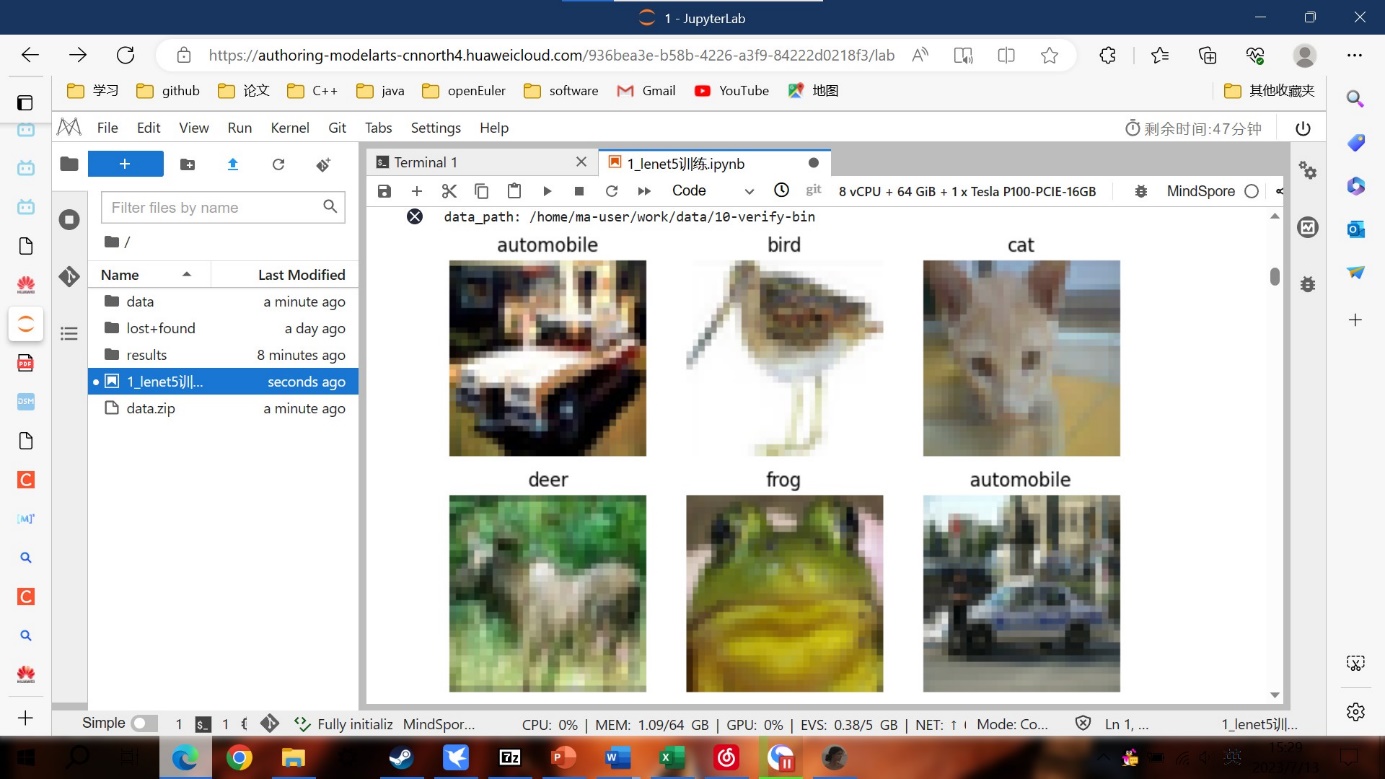
4、针对开放性问题的分析说明，本报告需要完整阐述所用模型或网络方法的修改方案，具体的参数设置，训练及测试的配置情况，数据使用情况，实验结果与分析评价。

1. **实验一到四完成情况与结果呈现**

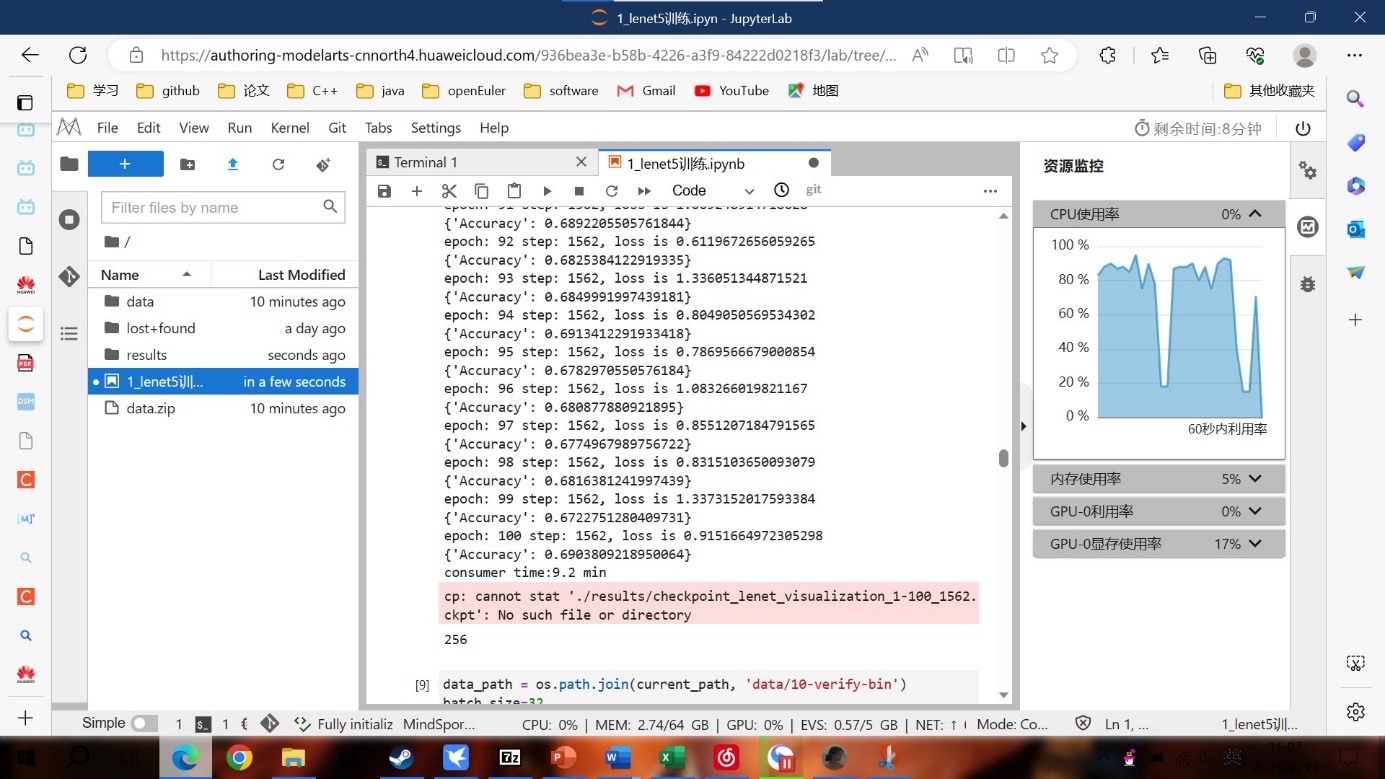
**1、实验一 LeNet-CIFAR10 分类**

1.1 实验过程

1.1.1 查看数据集

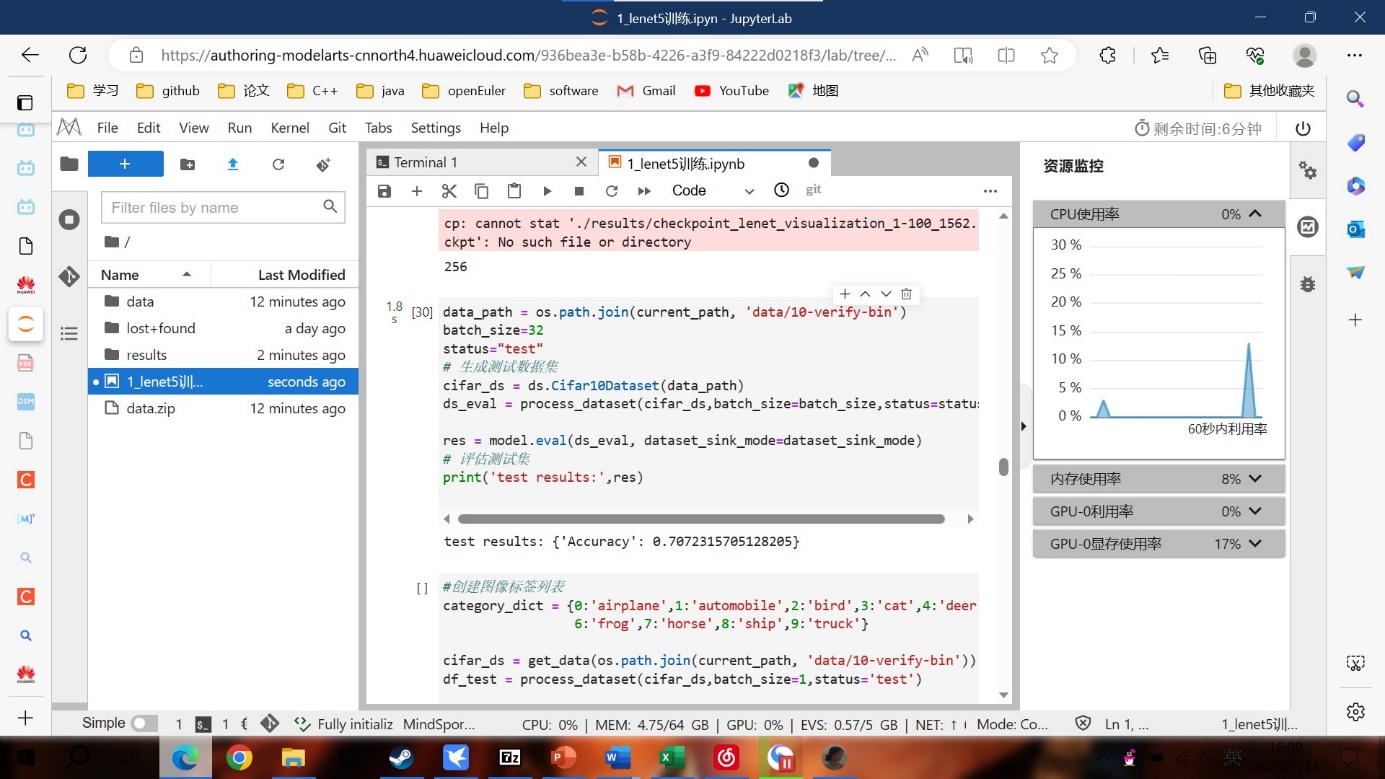


1.1.2 定义保存路径和训练



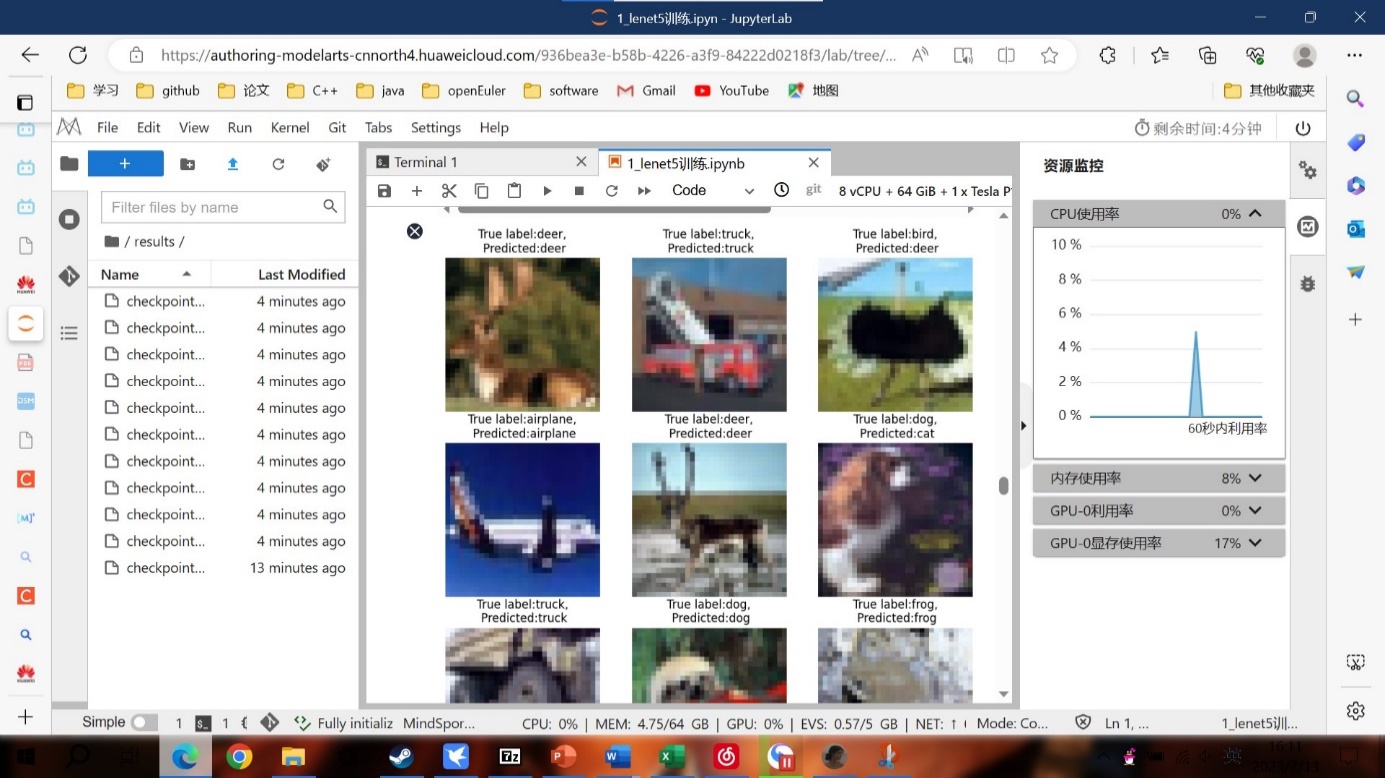
1.1.3 正确率计算

可以看到测试集与原数据集之间的标签预测正确率有0.7左右。



1.1.4 测试

随机选取九张图片，对比原标签与训练的标签，下图中仅有一个标签错误。



1.2 实验总结

本章提供了一个基于开源框架 Mindspore 的图像识别实验。该实验演示了如何利用开源框架Mindspore 完成 CIFAR-10 图像识别任务。本章对实验做了详尽的剖析，阐明了整个实验功能、结构与流程，详细解释了如何解析数据、如何构建深度学习模型、如何保存模型等内容。

1.3 思考题

1.3.1彩色图像有几个颜色通道？

答：3 个颜色通道，一般为 RGB。

1.3.2请列举一下常见的颜色空间。

答：RGB, HSV，Lab，CMYK 等。

1.3.3为什么我们在训练时会使用随机裁剪或者翻转的方式来处理图片？

答：为了增加网络的泛化能力，即使图片只有部分或者翻转了，也依然有能力识别出来。

1.3.4将图片归一化有什么好处？

答：加快模型的训练时候的收敛速度，也可以在一定程度上避免梯度消失或者爆炸。

1.3.5除了 Adam 优化器，请列举其他常见的优化器?

答：SGD, BGD, MBGD, Momentum, NAG(Nesterov Accelerated Gradient)，Adagrad，

Adadelta，RMSprop。

1.3.6什么叫一个 epoch?

答：一个 epoch 指代所有的数据送入网络中完成一次前向计算及反向传播的过程。

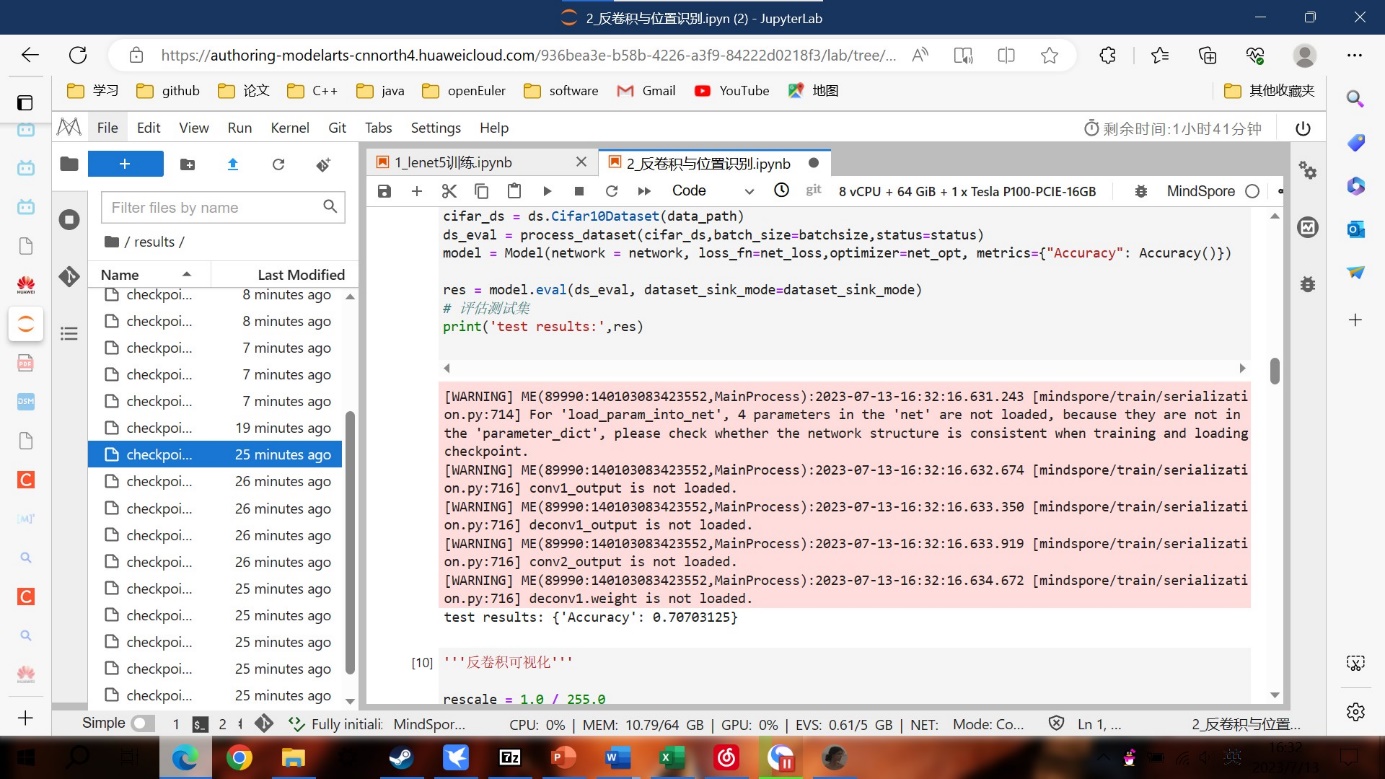
1.3.7如果模型过拟合，常见的处理方法有哪些?

答：L1，L2 正则化，Early stopping，数据增强，dropout 等。

**2、实验二 GardensPoint位置识别**

2.1 实验过程

2.1.1 导入预训练模型并计算正确率



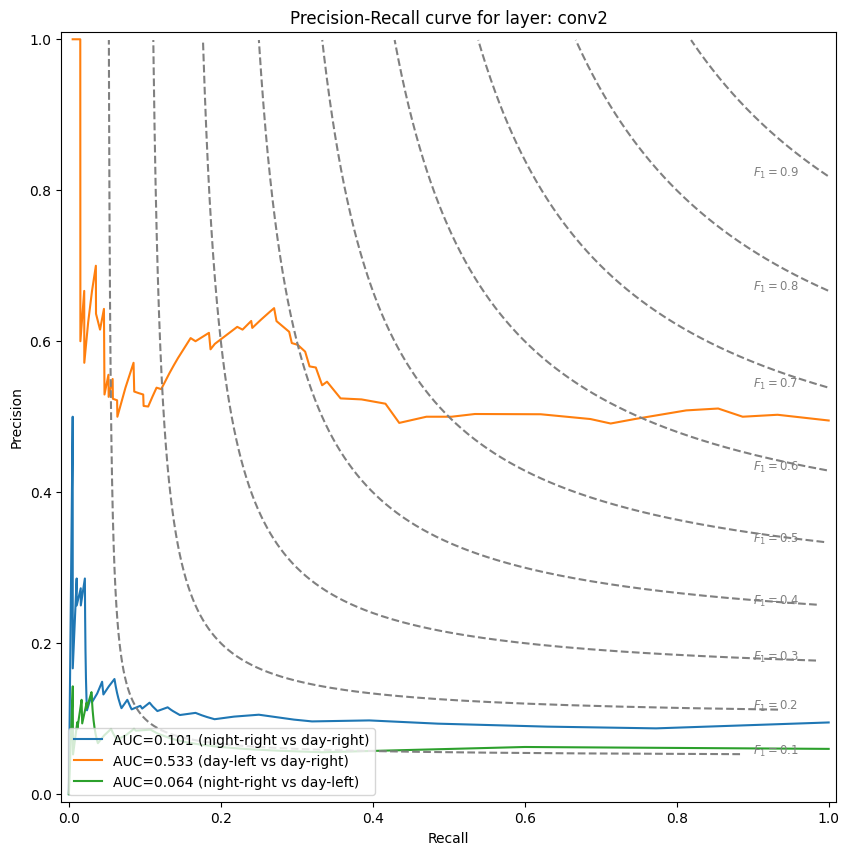
2.1.2 反卷积可视化

使用了一张飞机的图片作为测试，反卷积后仍能看见一些飞机的轮廓。

图表

描述已自动生成

2.1.3 PR曲线绘制



2.2 思考题

2.2.1 获取网络层特征的方法有哪些？

答：如果网络是通过 sequential 定义的，可以使用序号来提取不同层的特征；或者使用本实验中方法，定义网络层输出；或者定义 hook 函数。

2.2.2 传统特征提取方法有哪些？

答：SIFT、SURF、ORB 等。

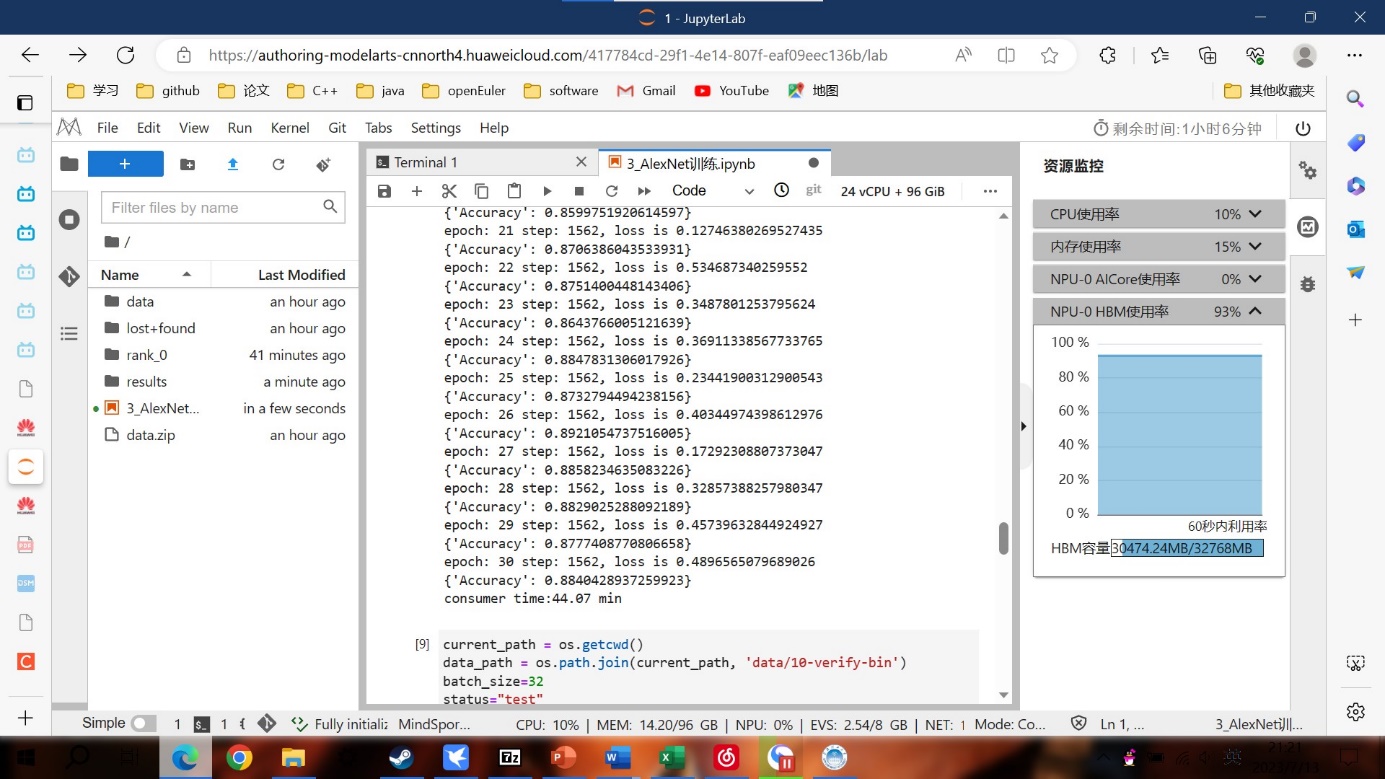
2.2.3 卷积网络不同层中提取的特征特点有哪些？

答：卷积网络中间层对外观变化更加鲁棒；最高层对视点变化更加鲁棒。

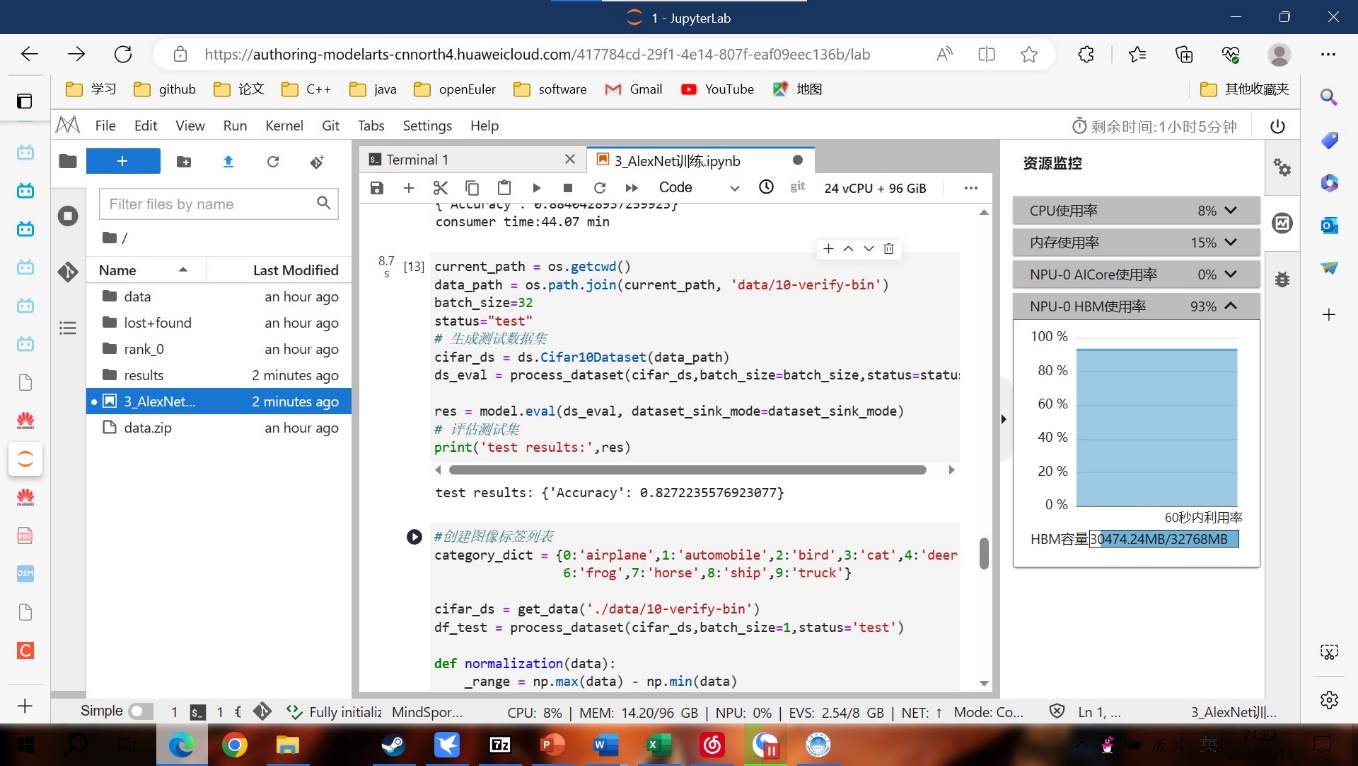
**3、实验三 AlexNet-CIFAR10分类**

3.1 实验过程

3.1.1 模型训练

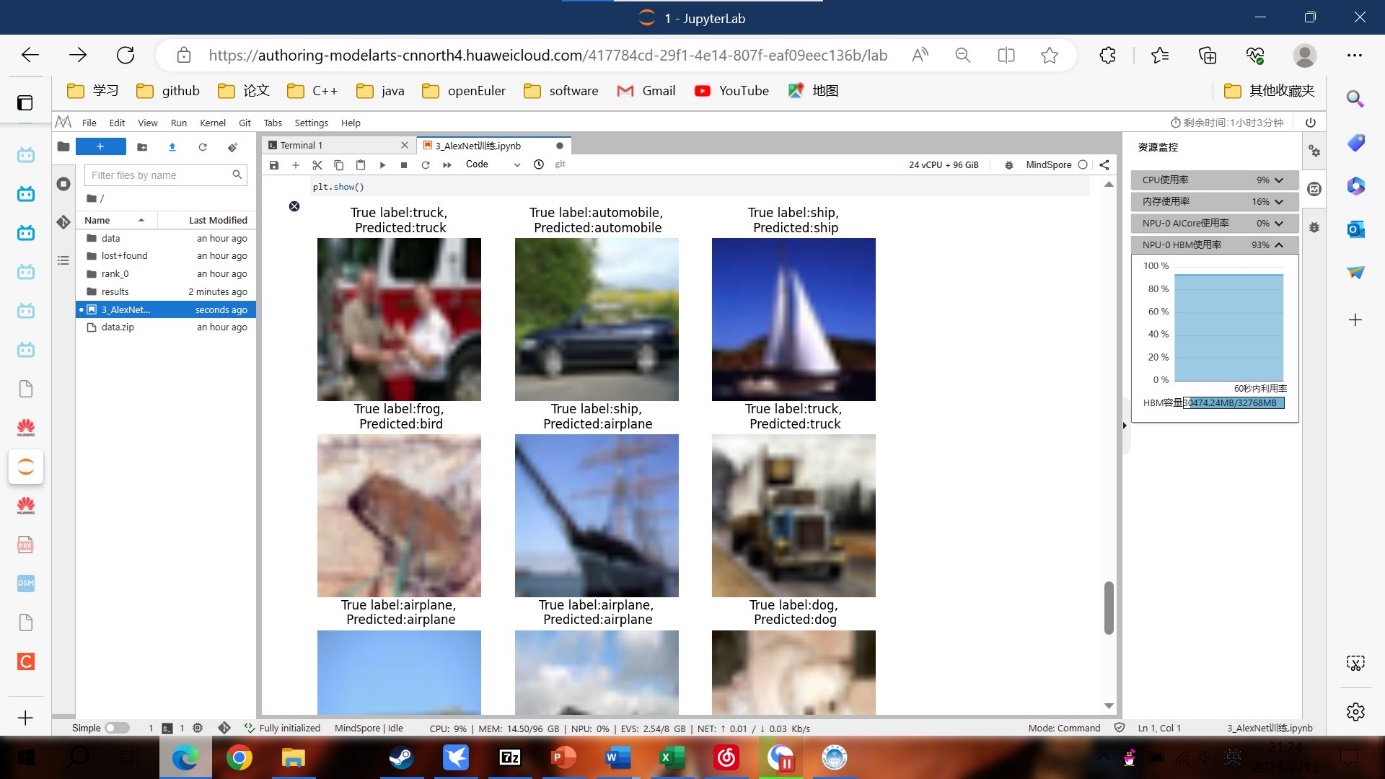


3.1.2 正确率计算



3.1.3 测试

随机选取九张图片对比真正标签与预测标签，下图中有两张图片预测错误。



3.2 实验总结

本章提供了一个基于开源框架 Mindspore 的图像分类实验。该实验演示了如何利用开源框架 Mindspore完成 CIFAR-10 图像分类任务。本章对实验做了详尽的剖析，阐明了整个实验功能、结构与流程，详细解释了如何解析数据、如何构建深度学习模型、如何保存模型等内容。

**4、实验四 反卷积可视化**

4.1 mindspore实现

4.1.1 训练

文本

描述已自动生成

4.1.2 实验结果

原图：



输出中间结果：

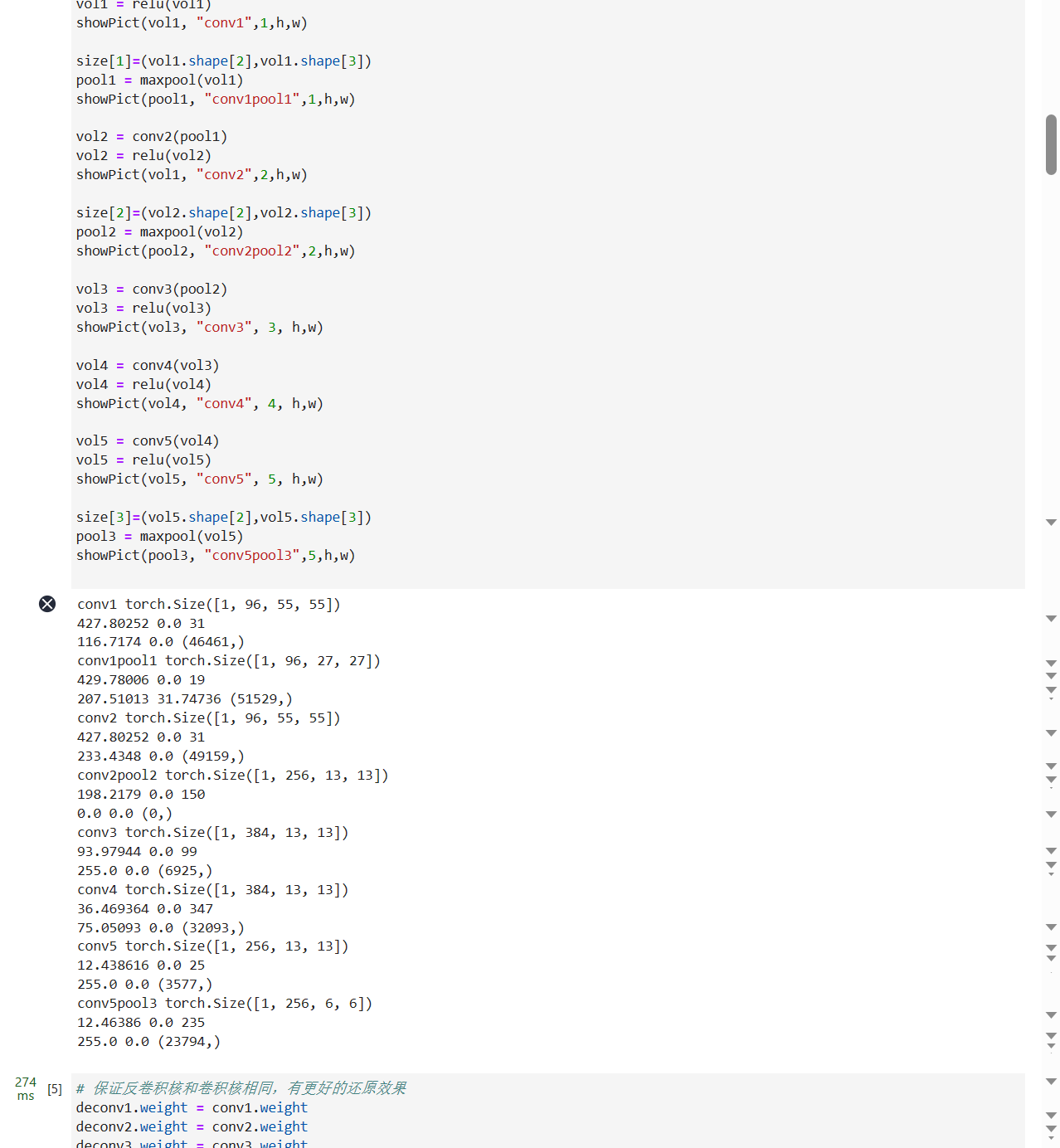
|  |  |  |
| --- | --- | --- |
| 一层反卷积输出： | 一层反卷积+一层反池化层输出： | 两层反卷积+一层反池化层输出： |
|  |  |  |
| 两层反卷积+两层反池化层输出： | 三层反卷积+两层反池化层输出： | 四层反卷积+两层反池化层输出： |
|  |  | 电脑萤幕  低可信度描述已自动生成 |
| 五层反卷积+两层反池化层输出： | 五层反卷积+三层反池化层输出： |  |
| 电子设备的屏幕  中度可信度描述已自动生成 |  |  |

4.2 pytorch实现

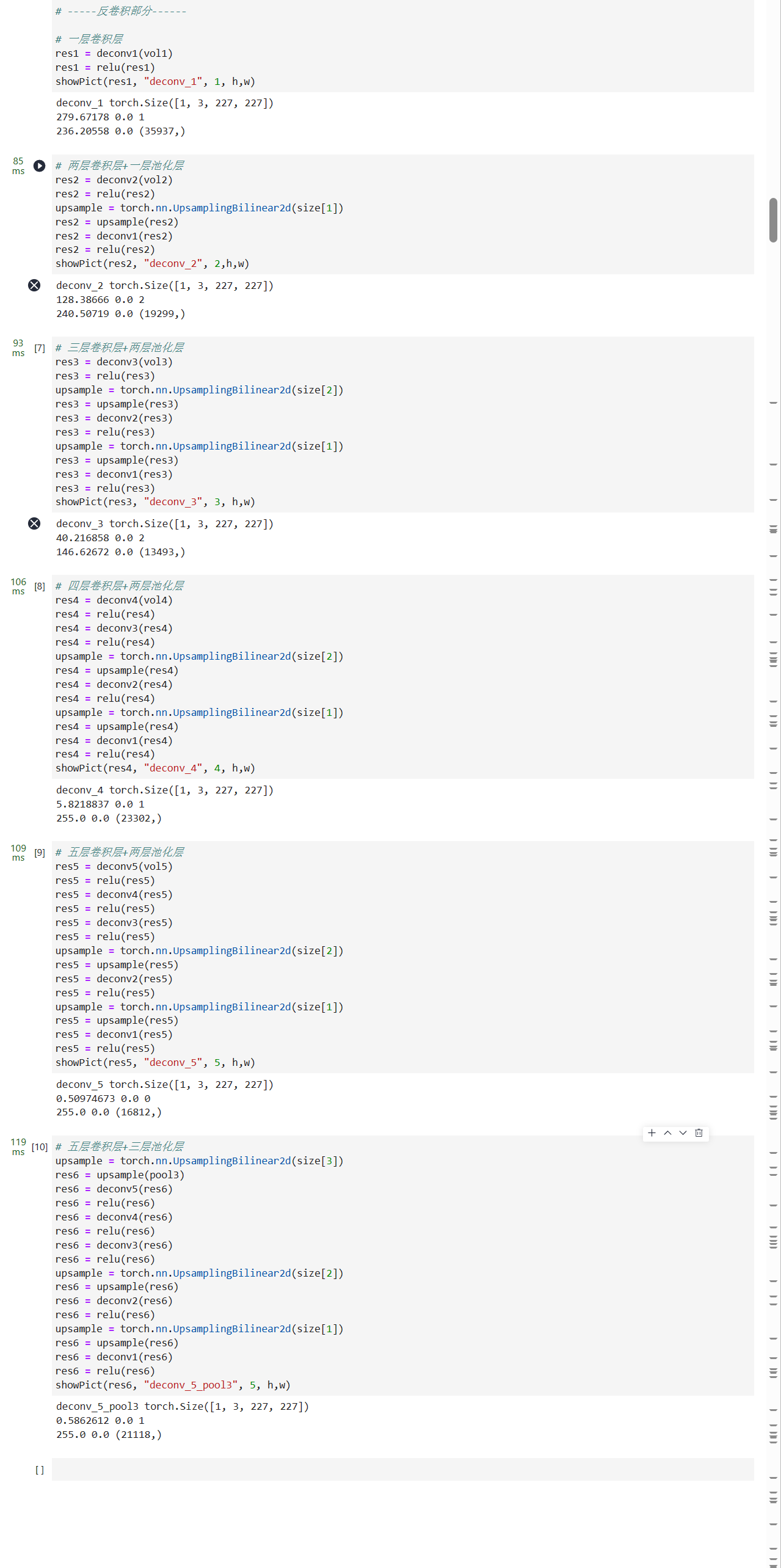
4.2.1 读取测试图片并输出相应数据



4.2.2 卷积



4.2.3 反卷积



4.2.4 输出中间结果

|  |  |  |
| --- | --- | --- |
| 一层反卷积输出： | 两层反卷积+一层池化层输出： | 三层反卷积+两层池化层输出： |
| 图片包含 建筑, 行李箱, 行李, 巴士  描述已自动生成 | 墙上有涂鸦  低可信度描述已自动生成 | 背景图案  描述已自动生成 |
| 四层反卷积+两层池化层输出： | 五层反卷积+两层池化层输出： | 五层反卷积+三层池化层输出： |
| 电子设备的屏幕  中度可信度描述已自动生成 | 电子设备的屏幕  中度可信度描述已自动生成 | 绿色的音响  中度可信度描述已自动生成 |

4.3 tensorflow实现

4.3.1 训练

图形用户界面, 应用程序

描述已自动生成

4.3.1 输出中间结果

|  |  |  |
| --- | --- | --- |
| 第一层：卷积--> 激活 --> 归一化 --> 最大池化 | 第二层：卷积--> 激活 --> 归一化 --> 最大池化 | 第三层：卷积--> 激活 |
|  |  |  |
| 第四层：卷积--> 激活 | 第五层：卷积--> 激活 --> 最大池化 |  |
| 电子设备的音响  低可信度描述已自动生成 | 夜晚的音响  低可信度描述已自动生成 |  |

4.4 实验总结

本章提供了一个反卷积对卷积中间结果进行可视化的实验。希望能从机器视角查看卷积结果。

1. **实验一开放题**

**【选用合适的网络，测试至少两种提升技巧，让测试集准确率提升至少两个百分点。】**

**选用lenet-5网络。**

* 1. **改进方法介绍**

1.1.1 增加迭代次数以提高准确率

1.1.2 增加了batch normal层数

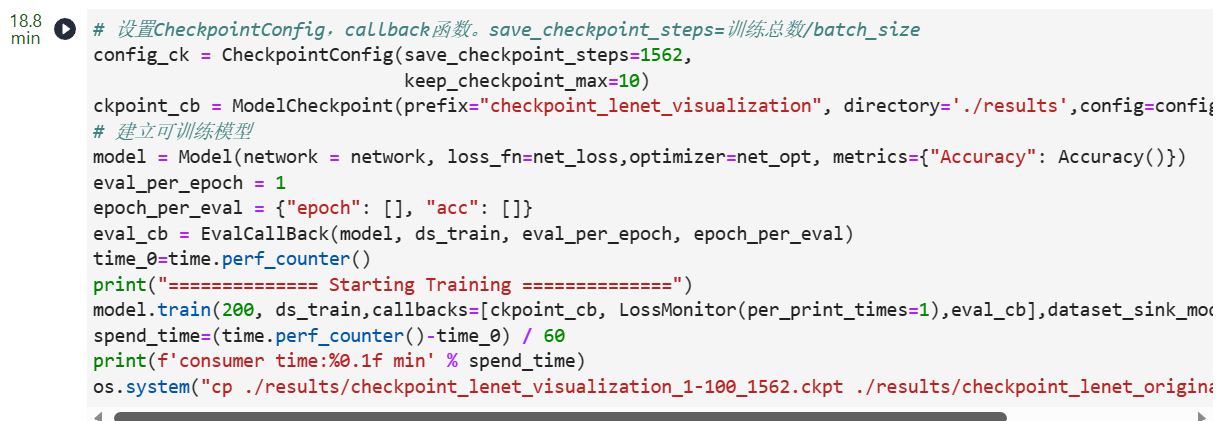
* 1. **实现**

1.2.1 在train函数中将epoch设置为200。

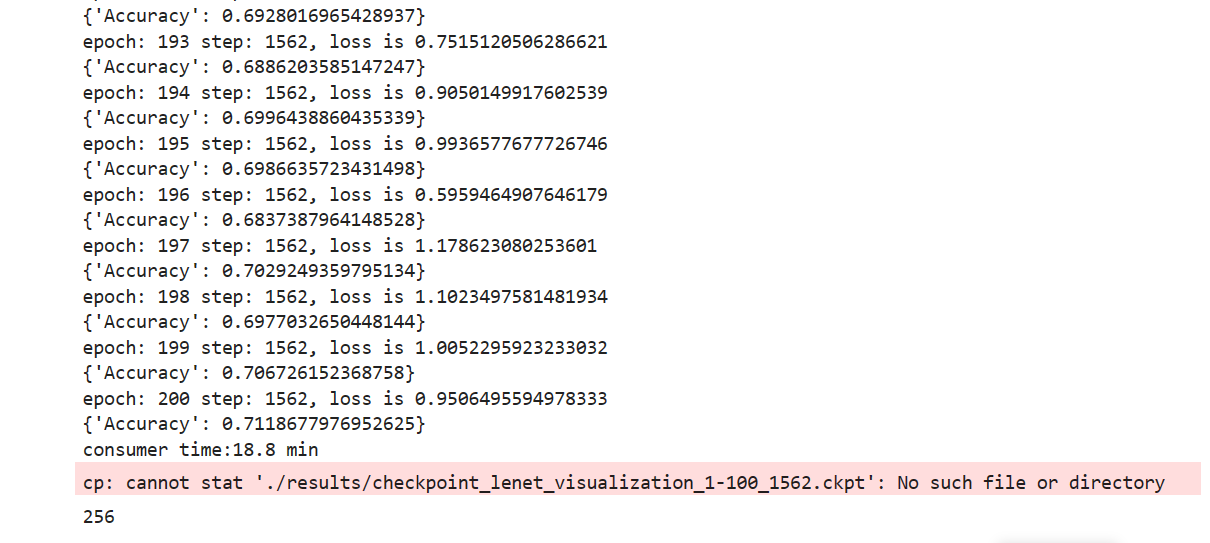
1.2.2 增加batch normal层数，其余参数保持不变。

* 1. **数据、参数与训练、设置说明**

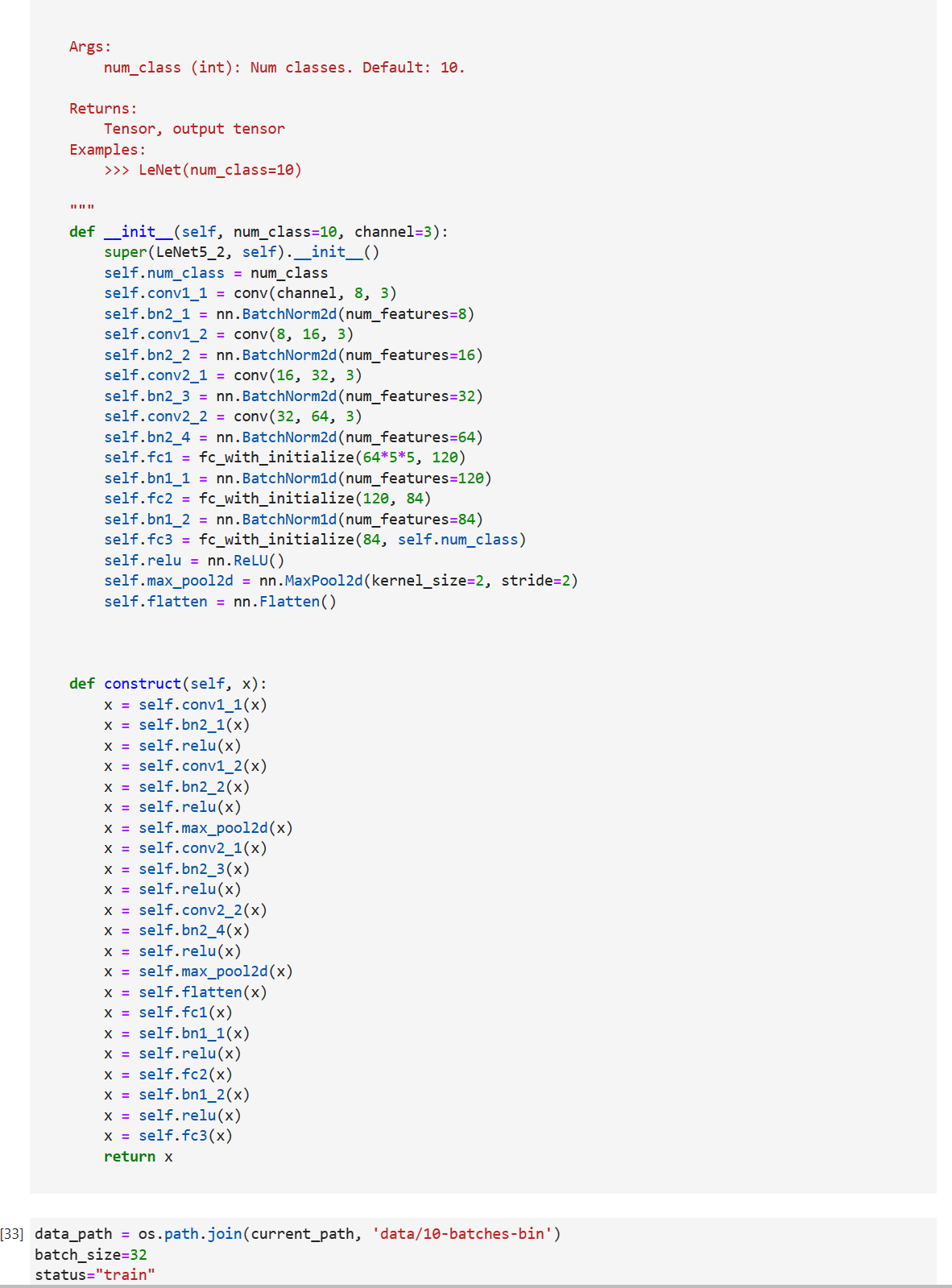
**1.3.1 修改epoch次数如下**

****

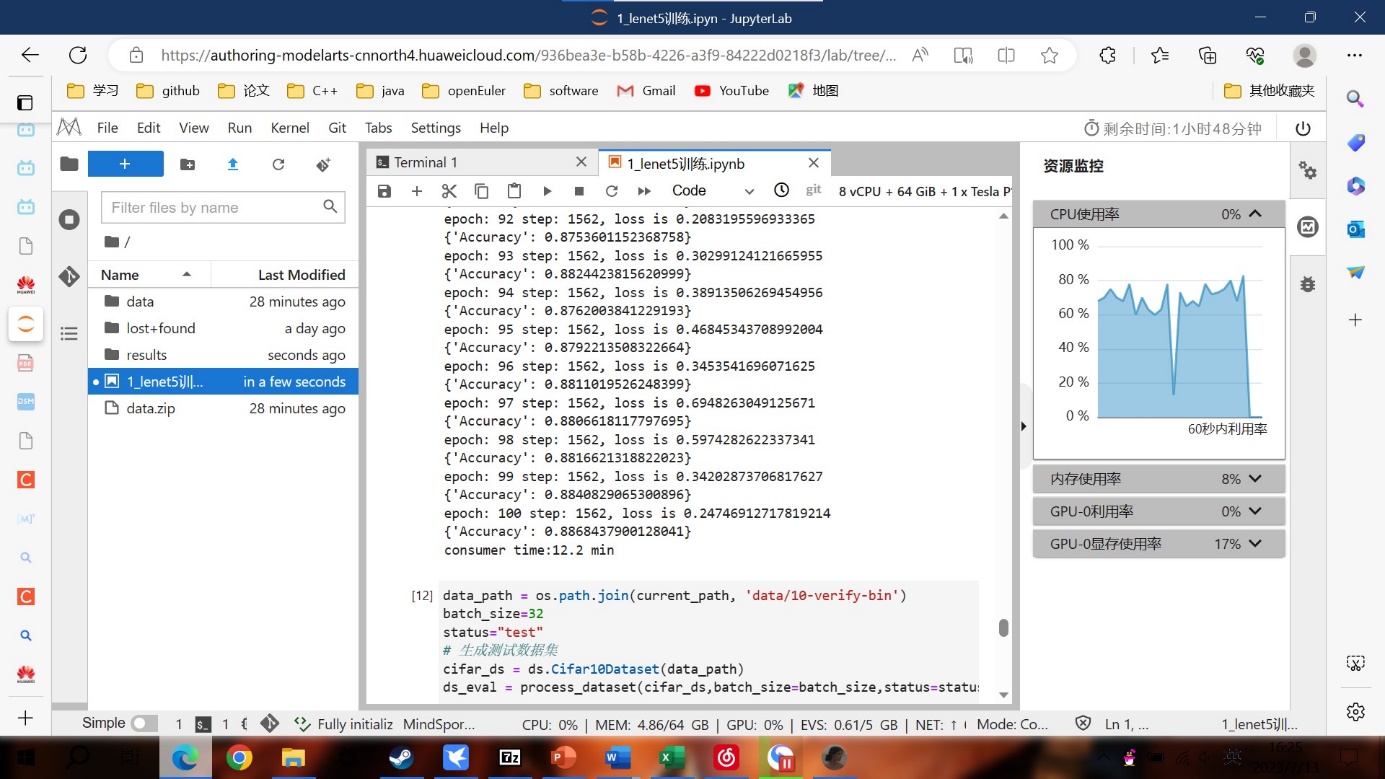
**训练：**

****

**1.3.2 增加batch normal代码如下**



**训练：**

****

* 1. **实验结果及分析**

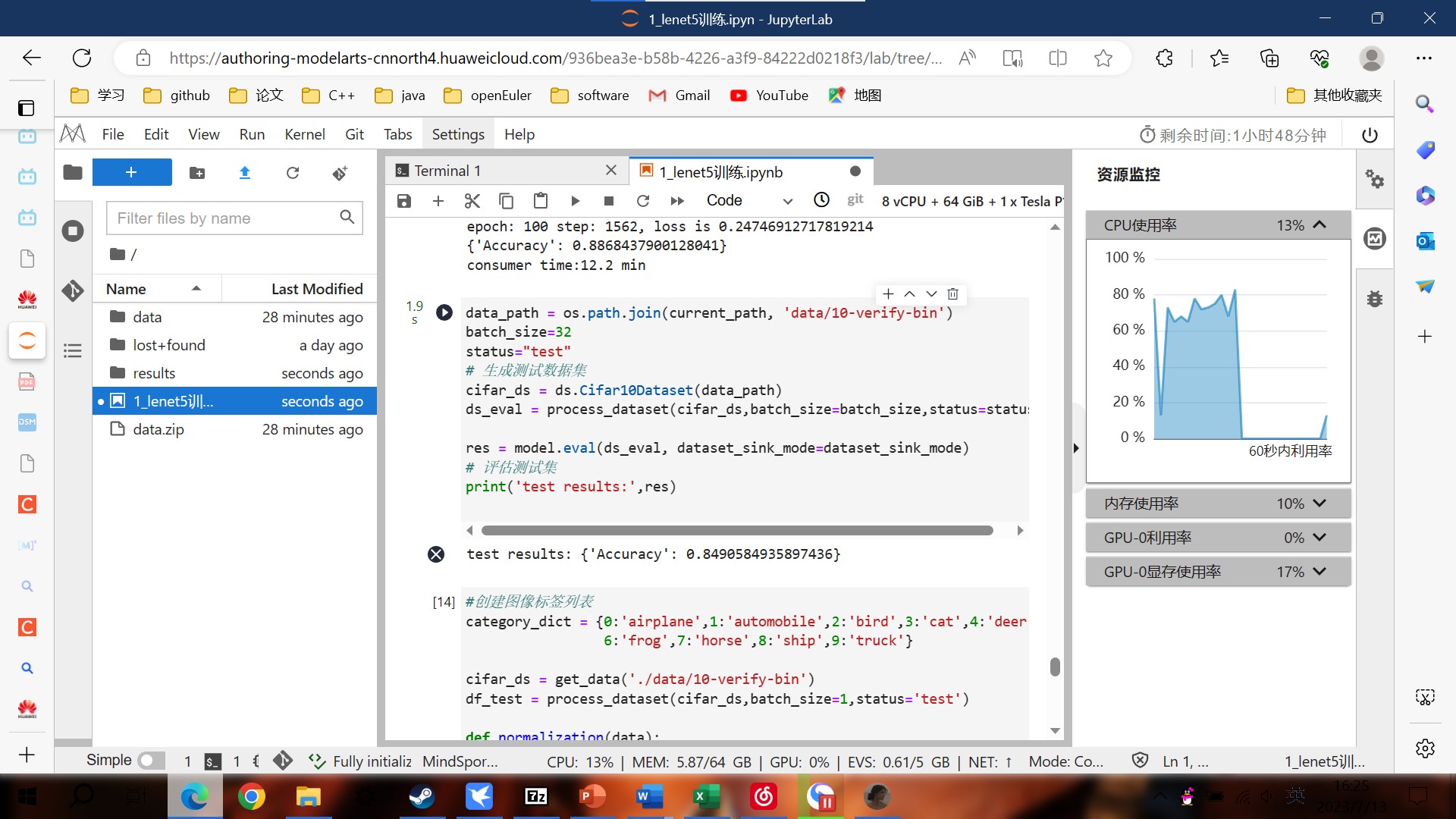
1.4.1 增加epoch次数

准确率为：可以发现准确率提升了两个百分点左右，不过根据训练时每个epoch的accuracy可以推测准确率已经趋于收敛，即使在增加epoch次数，准确率提升较小。



1.4.2 增加batch normal层

准确率如下图所示，较原先提高了将近十四个百分点，提升显著。



1. **实验四开放题**

**【选取典型测试案例，横向对比本单元验证的不同网络（LeNet, AlexNet）在不同层的特征输出情况。】**

* 1. **测试案例与各网络特征层介绍**

测试案例选用实验二与实验四中的测试图片进行比较。

Lenet：LeNet诞生于1998年，网络结构比较完整，包括卷积层、pooling层、全连接层。被认为是CNN的鼻祖。

输入32\*32\*1。

卷积层3个卷积的主要目的是使原始信号特征增强，并且降低噪音，卷积核 5\*5 。

下采样层2个（池化）：池化层的作用是降低网络训练参数及模型的过拟合程度。

全连接1个

输出层1个 10个类别 softmax。

Alexnet：AlexNet网络包含8层，其中包括5层卷积和2层全连接隐藏层，以及1个全连接输出层，其中

第一卷积层使用的 Kernel为11\*11, 后接3\*3 步幅为2的 池化层；

第二卷积层使用的 Kernel为5\*5，后接3\*3 步幅为2的池化层；

第三卷积层使用的 Kernel为3\*3，直接与第四层相连；

第四卷积层使用的 Kernel为3\*3，直接与第五层相连；

第五卷积层使用的 Kernel为3\*3，后接3\*3 步幅为2的池化层；

接下来，卷积部分结束后，其后紧跟两个输出均为4096的全连接隐藏层，最后为1个全连接输出层。

* 1. **特征输出实现及参数设置说明**

特征输出详见img文件夹，参数设置详见ipynb文件。

* 1. **实验结果对比及分析**

各层输出的图片详见img文件夹，随着卷积层数增加，图片丢失的信息越来越多，到了五层卷积层后，几乎无法看出原图片的影子。

1. **机时费用情况**

**请在此记录本次大作业完成时的机时券剩余金额：￥ 928.25 元**