**广州商学院**

**课 程 论 文**

**第 一 组**

**题目：文字识别**

课 程 名 称 **数字图像处理课程实验**

考 查 学 期  **2021学年 第二学期**

考 查 方 式 **课程论文**

姓 名 吴斯达

学 号 **201806140058**

专 业 **数据科学与大数据技术**

成 绩

指 导 教 师 王莎莎

摘要

利用深度学习中的卷积神经网络以及图像处理的技术，将从网上获取到字母图片数据集进行图像识别，包括对原图像的进行预处理，用tensorflow对进行图像识别时，还将采用卷积神经网络来构建模型，本次实验将分为未采用卷积神经网络进行图像识别，采用卷积神经网络以及采用卷积神经网络加数据增强这几个步骤，依次对比各种训练方法对图像识别的准确率。

**关键词：字母识别；卷积神经网络；图像处理；**

**1、绪论**

**1.1引言**

**深度学习在近年来越发火热，深度学习(DL, Deep Learning)是[机器学习](https://baike.baidu.com/item/%E6%9C%BA%E5%99%A8%E5%AD%A6%E4%B9%A0/217599" \t "https://baike.baidu.com/item/%E6%B7%B1%E5%BA%A6%E5%AD%A6%E4%B9%A0/_blank)(ML, Machine Learning)领域中一个新的研究方向，它被引入机器学习使其更接近于最初的目标——[人工智能](https://baike.baidu.com/item/%E4%BA%BA%E5%B7%A5%E6%99%BA%E8%83%BD/9180" \t "https://baike.baidu.com/item/%E6%B7%B1%E5%BA%A6%E5%AD%A6%E4%B9%A0/_blank)(AI, Artificial Intelligence)。 [1]**

**深度学习是学习[样本数据](https://baike.baidu.com/item/%E6%A0%B7%E6%9C%AC%E6%95%B0%E6%8D%AE/12726279" \t "https://baike.baidu.com/item/%E6%B7%B1%E5%BA%A6%E5%AD%A6%E4%B9%A0/_blank)的内在规律和表示层次，这些学习过程中获得的信息对诸如文字，[图像](https://baike.baidu.com/item/%E5%9B%BE%E5%83%8F/773234" \t "https://baike.baidu.com/item/%E6%B7%B1%E5%BA%A6%E5%AD%A6%E4%B9%A0/_blank)和声音等数据的解释有很大的帮助。它的最终目标是让机器能够像人一样具有分析学习能力，能够识别文字、图像和声音等数据。 深度学习是一个复杂的机器学习算法，在语音和图像识别方面取得的效果，远远超过先前相关技术。 [1]**

**深度学习在[搜索技术](https://baike.baidu.com/item/%E6%90%9C%E7%B4%A2%E6%8A%80%E6%9C%AF/1447197" \t "https://baike.baidu.com/item/%E6%B7%B1%E5%BA%A6%E5%AD%A6%E4%B9%A0/_blank)，[数据挖掘](https://baike.baidu.com/item/%E6%95%B0%E6%8D%AE%E6%8C%96%E6%8E%98/216477" \t "https://baike.baidu.com/item/%E6%B7%B1%E5%BA%A6%E5%AD%A6%E4%B9%A0/_blank)，机器学习，[机器翻译](https://baike.baidu.com/item/%E6%9C%BA%E5%99%A8%E7%BF%BB%E8%AF%91/411793" \t "https://baike.baidu.com/item/%E6%B7%B1%E5%BA%A6%E5%AD%A6%E4%B9%A0/_blank)，[自然语言处理](https://baike.baidu.com/item/%E8%87%AA%E7%84%B6%E8%AF%AD%E8%A8%80%E5%A4%84%E7%90%86/365730" \t "https://baike.baidu.com/item/%E6%B7%B1%E5%BA%A6%E5%AD%A6%E4%B9%A0/_blank)，[多媒体学习](https://baike.baidu.com/item/%E5%A4%9A%E5%AA%92%E4%BD%93%E5%AD%A6%E4%B9%A0/10528812" \t "https://baike.baidu.com/item/%E6%B7%B1%E5%BA%A6%E5%AD%A6%E4%B9%A0/_blank)，语音，推荐和个性化技术，以及其他相关领域都取得了很多成果。深度学习使机器模仿视听和思考等人类的活动，解决了很多复杂的模式识别难题，使得人工智能相关技术取得了很大进步。**

**1.2卷积神经网络**

**在无监督预训练出现之前，训练深度神经网络通常非常困难，而其中一个特例是卷积神经网络。卷积神经网**

**[IMG_256](https://baike.baidu.com/pic/æ·±åº¦å­¦ä¹ /3729729/0/bd3eb13533fa828b269a8a0ff31f4134960a5aa4?fr=lemma%26ct=single)**

**络受视觉系统的结构启发而产生。第一个卷积神经网络计算模型是在Fukushima(D的神经认知机中提出的，基于神经元之间的局部连接和分层组织图像转换，将有相同参数的神经元应用于前一层神经网络的不同位置，得到一种平移不变神经网络结构形式。后来，Le Cun等人在该思想的基础上，用误差梯度设计并训练卷积神经网络，在一些模式识别任务上得到优越的性能。至今，基于卷积神经网络的模式识别系统是最好的实现系统之一，尤其在手写体字符识别任务上表现出非凡的性能。 [5]**

**1.3字母识别**

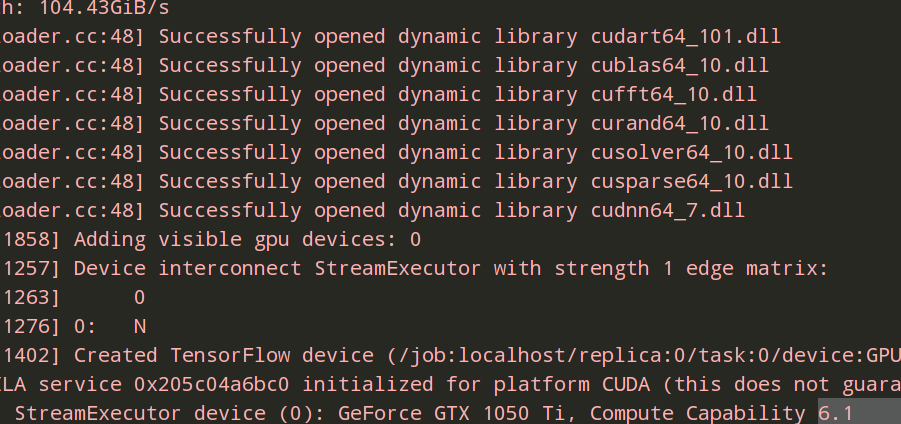
**在本次课程中我选择的题目为文字识别，所以我决定对选择文字识别中的字母识别来进行本次实验。利用计算机自动识别字符的技术，是模式识别应用的一个重要领域，文字识别技术可以为人们提供便利的将图片转为文字的方法，为了方便人们的使用，所以选择字母识别程序进行初步设计和研究学习。在本次实验，我使用的数据集是从网上下载的属于字母图片数据集的包含370000+张图片，利用该数据构建一个卷积神经网络来进行识别。**

**2、需求分析**

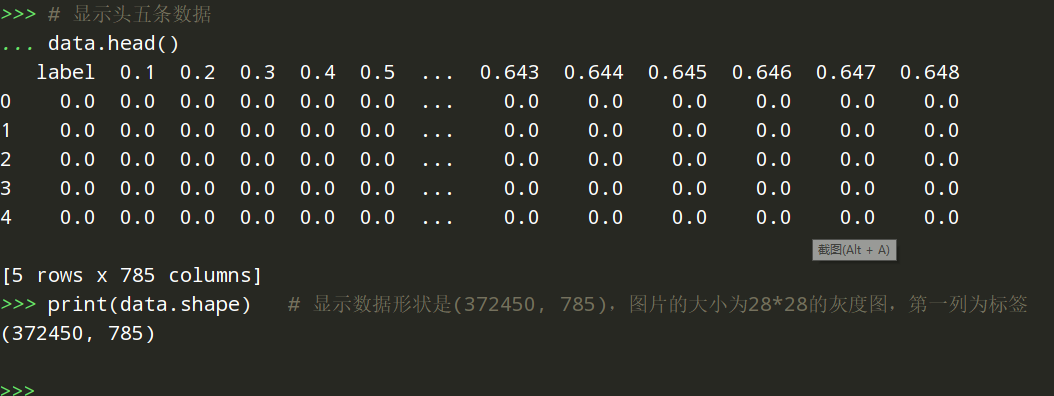
**在本次实验中，我需要将从网上下载下来的那包含370000+张图片的数据集投入到利用tensorflow构建的模型中去，并进行训练，使得模型训练准确达到百分之九十以上，并想办法提高图片的泛化能力，最终还要通过亲手写出字母的图像的方法来验证本次训练的模型的有效性。**

**3、总体设计**

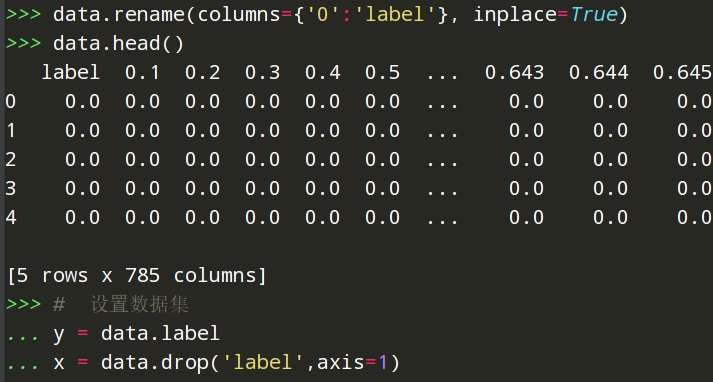
**首先为了提高本次实验的效率，决定安装tensorflow的GPU版本来进行本次实验，安装环境为CUDA10.1,CUDNN7.6,tensorflow-gpu2.3.0并运行相关程序验证安装是否成功，结果显示安装成功**

****

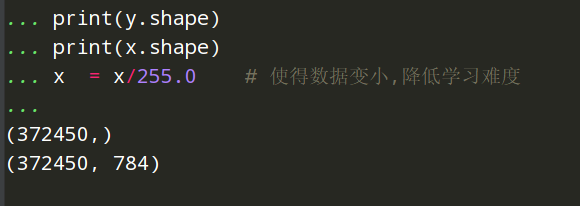
**安装成功后，我们开始对数据集进行预处理，本次实验采用的图片数据集是.csv格式，所以我们需要先查看该数据集的大小，形状**

****

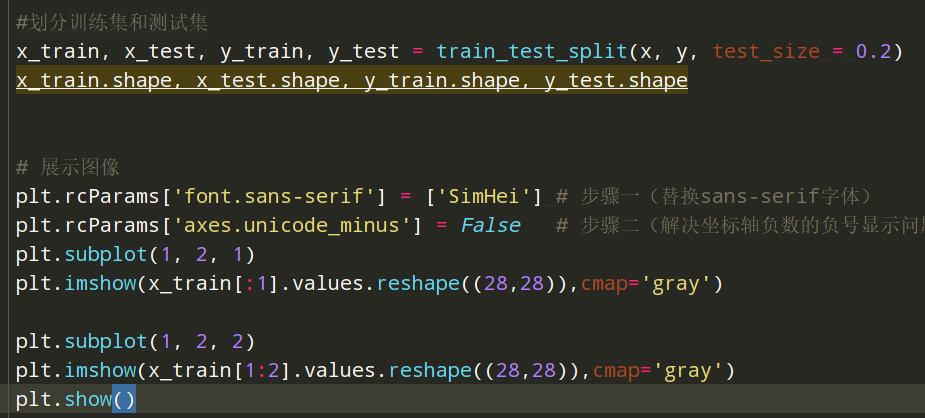
**在查看数据集的大小以及标签后我们开始对图片的标签进行处理，经过查看后发现第一列数据为标签，其余784列为图片像素值，图片为灰度图，大小为28\*28像素，将第一列重命名为label**

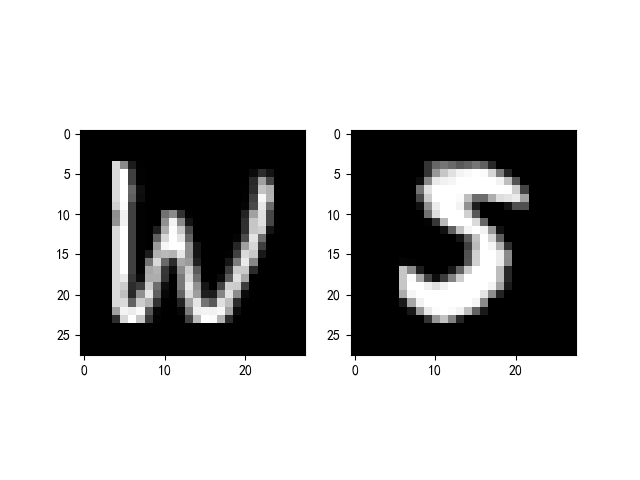


**之后我们将图片数据集的参数和标签分开，并将图片的像素值整体除以255，对数值进行归一化处理，使得原本0~255的灰度值变成0到1的数值，使得数据变小,降低学习难度，在将数据分开后，打印图片的训练参数以及标签**

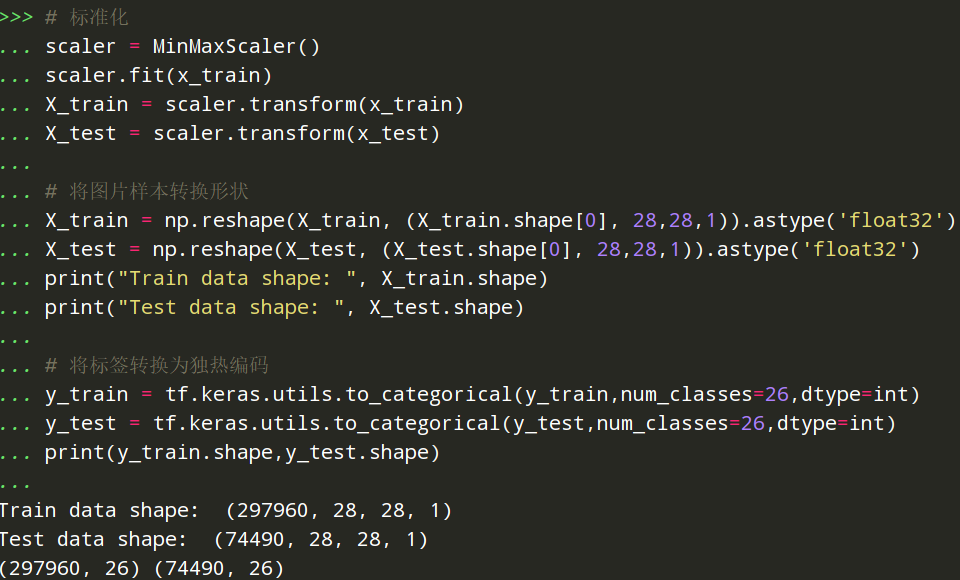
****

**接着我们用sklearn的train\_test\_split函数来划分训练集和测试集，并将图片打印出来，查看图片**

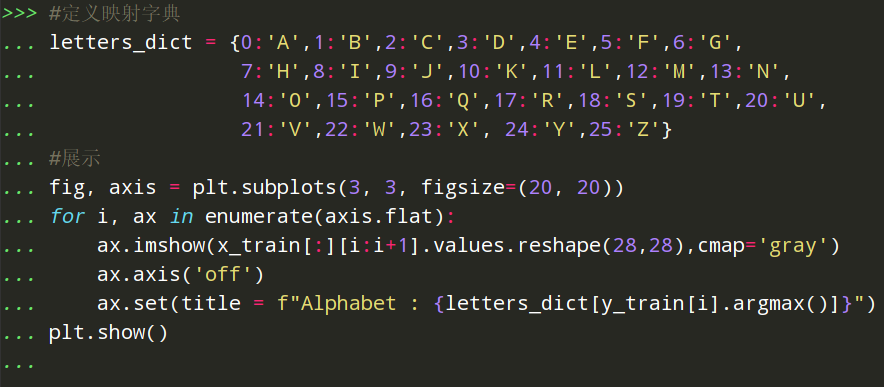
****

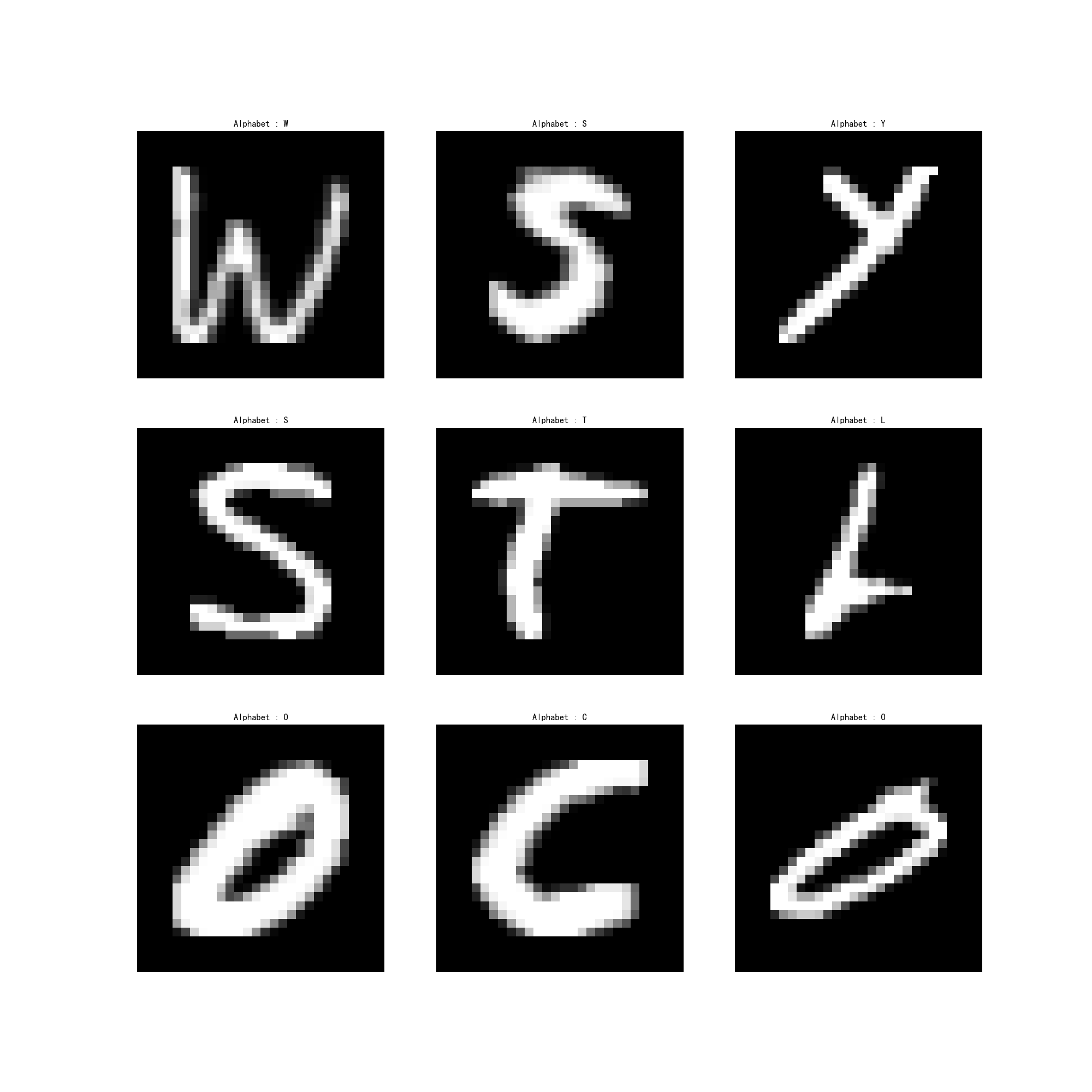


**可以看出图片可以正常显示，我们继续进行数据预处理，讲图片参数进行标准化，并转换形状，将标签转换为独热编码，重新打印数据的形状。**

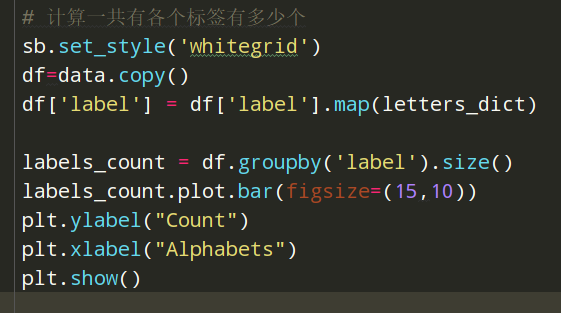
****

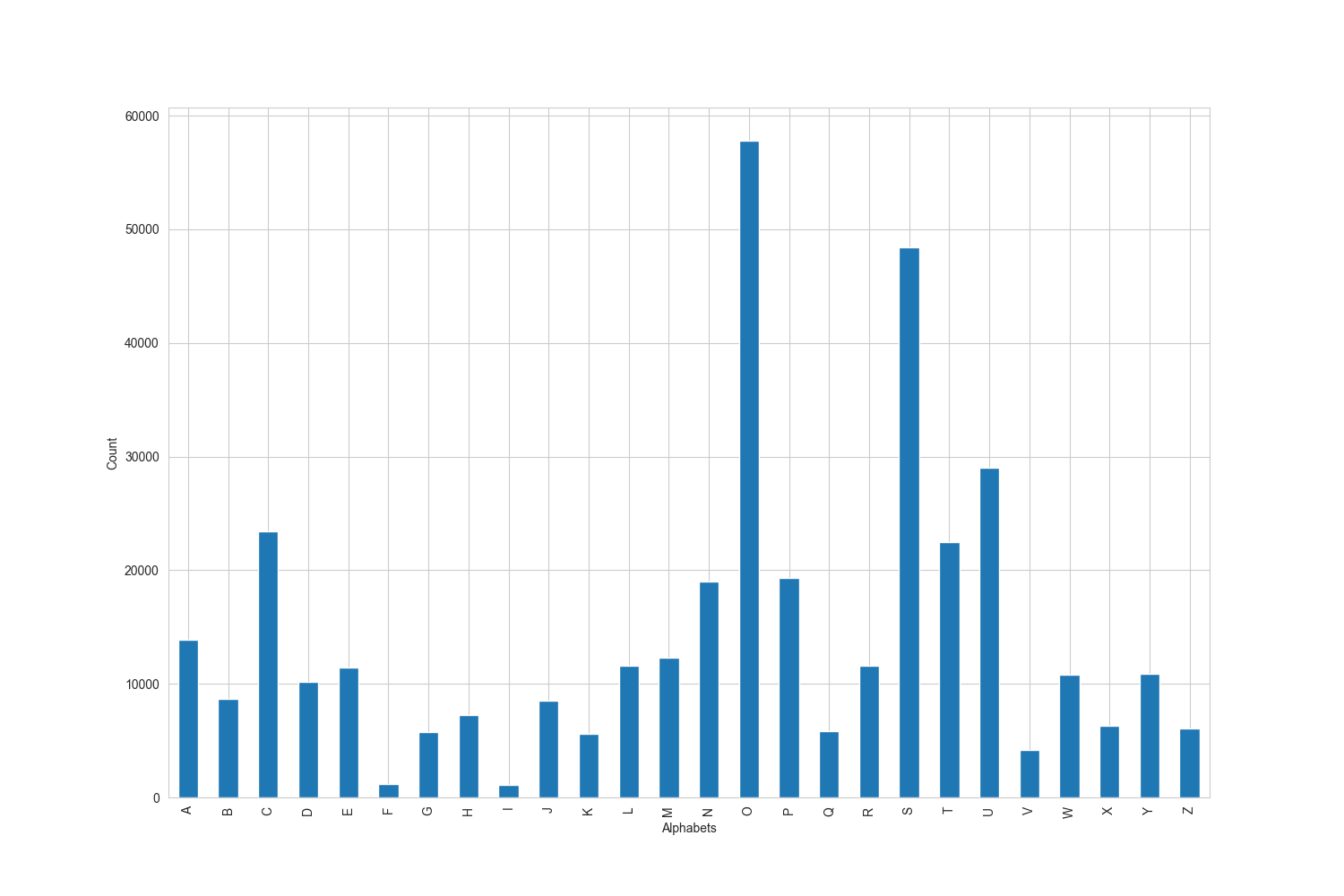
**接着我们定义字典映射，进一步展示图片以及对应的字母。**

****

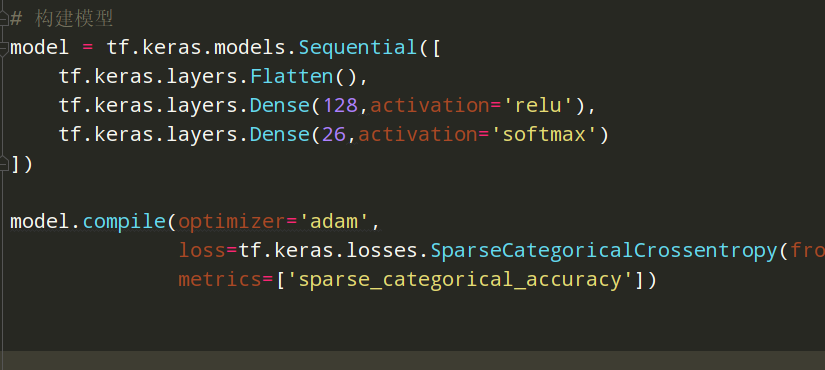


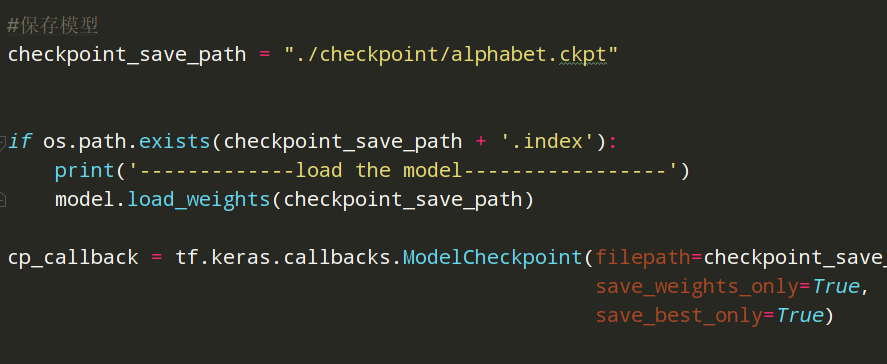
**我们接着展示每个字母对应的图片的个数**

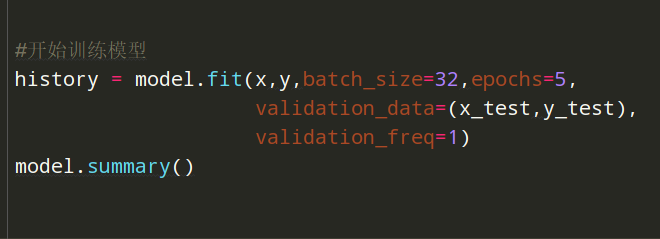




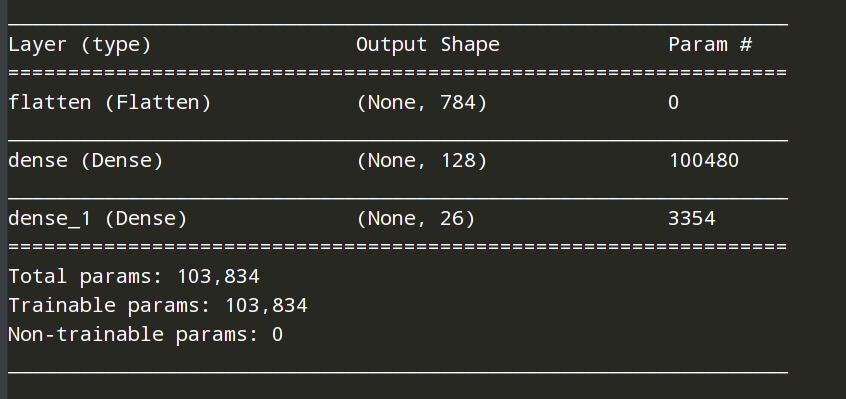
**在对数据集进行了预处理以及具体展示后我们开始构建神经网络模型，并将模型保存（此时先不保存,所以没有在模型训练那里加上callback参数），该神经网络模型一共有两层，第一层为拉直层，将数据从28\*28的二维数组拉直为一维的784大小的数组，激活函数为relu，第二层为全连接层，输入参数128，为了使输出符合概率分布，采用的激活函数为softmax。往下配置模型时，使用的优化器为adam，loss函数为sparsecategoricalcrossentory,fromlogit为False(输出为概率分布为false，否则为True)，评测指标为sparse\_categorical\_accuracy。训练模型时分为5个epoch，每次喂入32个batch数据。经过一个epoch训练一次。**



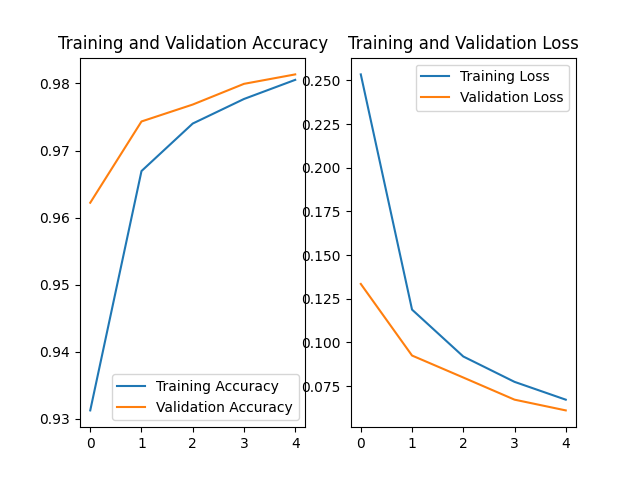




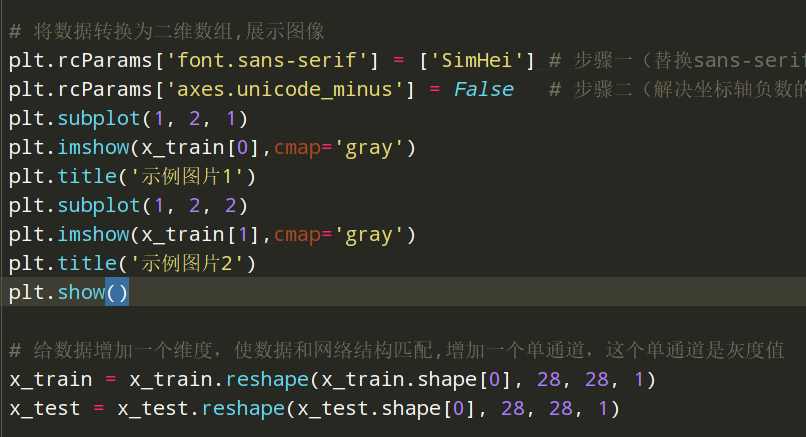
**在未构建卷积层下模型结构以及训练结果如下**

****

**模型的acc以及loss曲线**



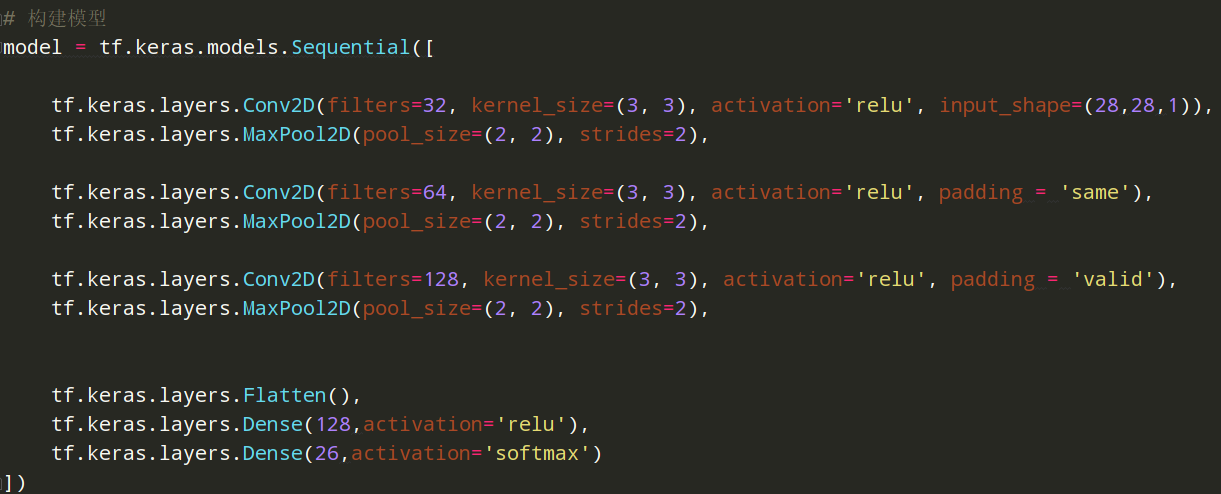
**模型以及达到要求，准确率超过百分之九十以上，接着我们开始构建卷积神经网络，卷积神经网络构建与前一步相似，不过要给数据增加一个维度，使数据和网络结构匹配,增加一个单通道，这个单通道是灰度值**

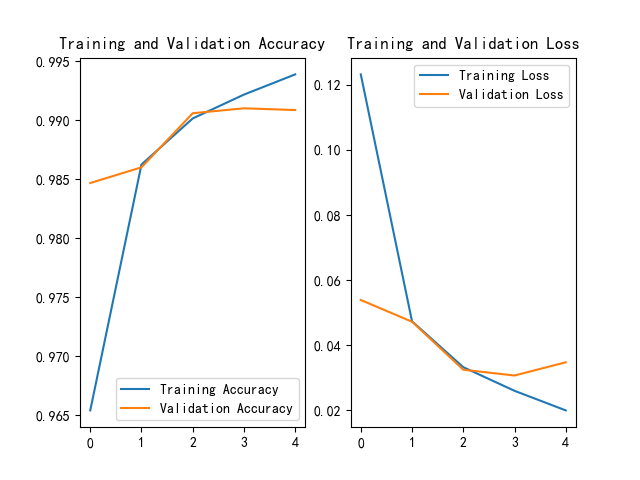
****

**（卷积计算可认为是一种有效提取图像特征的方法，一般会用一个正方形的卷积核，按照指定步长，在输入特征图上滑动，遍历输入特征图中的每个像素点。每一个步长，卷积核会与输入特征图出现重合区域，重合区域对应元素相乘、求和再加上偏置项得到输出特征的一个像素点，如果输入是单通道灰度图，使用的是3\*3的深度为1的卷积核，如果是三通的彩色图，使用的是3\*3\*3的卷积核，要输出特征图，要使得卷积核的通道数与输入特征图通道数一致，有几个卷积核，就有几个输出特征图，然后叠加在这张输出特征图后面）**

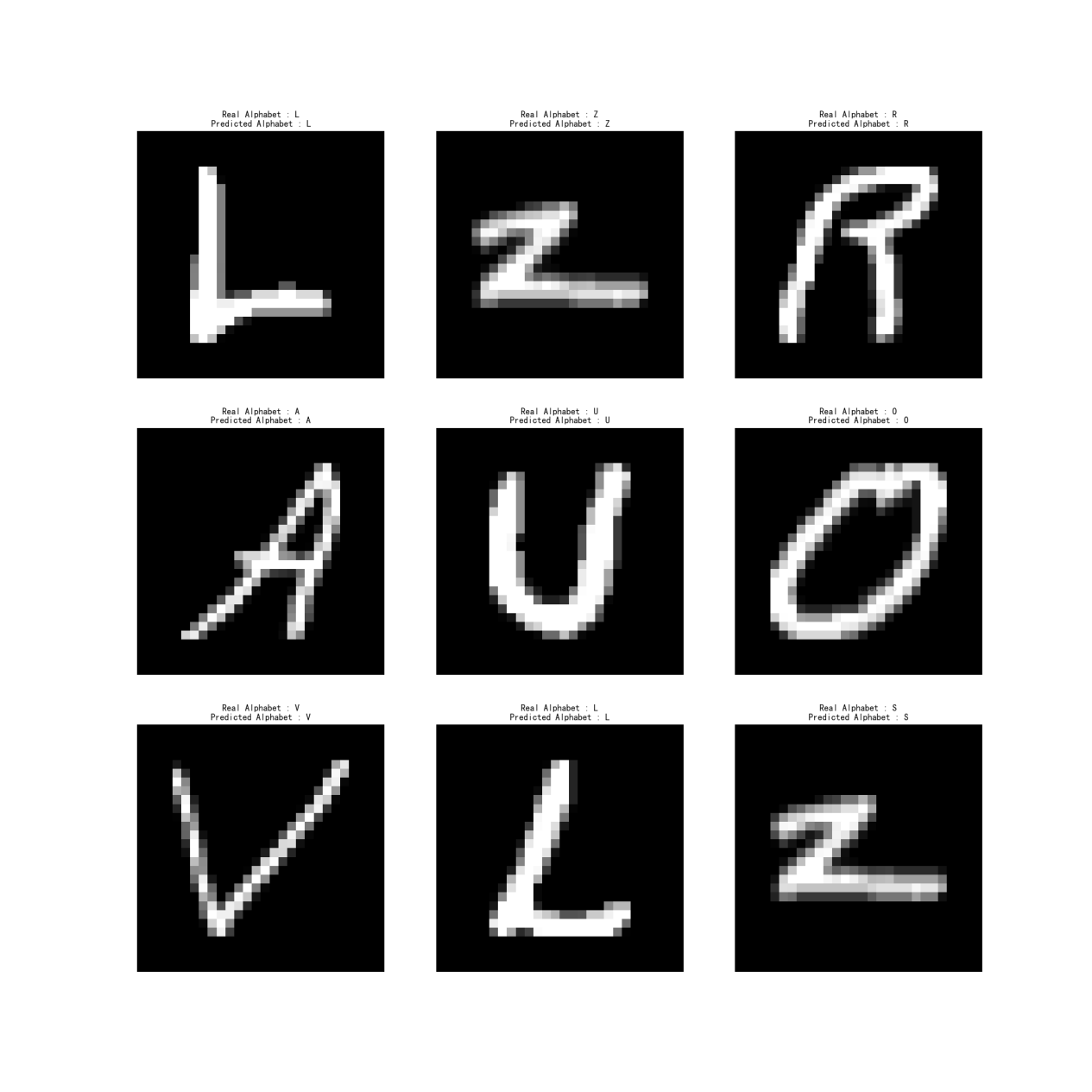
**卷积核个数为32，卷积核尺寸为3\*3，滑动步长为2，输入为28\*28\*1的单通道灰度图，池化层那里默认全零填充(输入特征图与输出特征图形状一致),选择最大池化（池化层用于减少特征数据量，最大池化可提取图片纹理，均值池化可保留背景特征。）**

**在构建模型时，加入了卷积层**

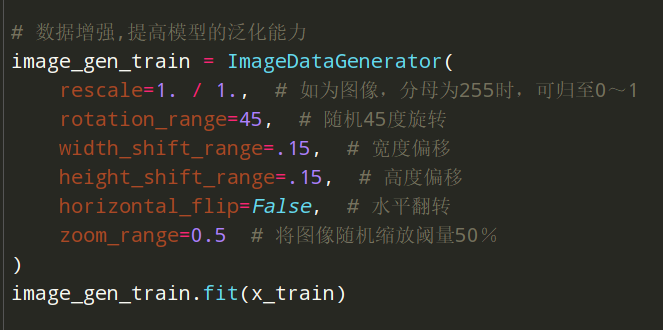
****

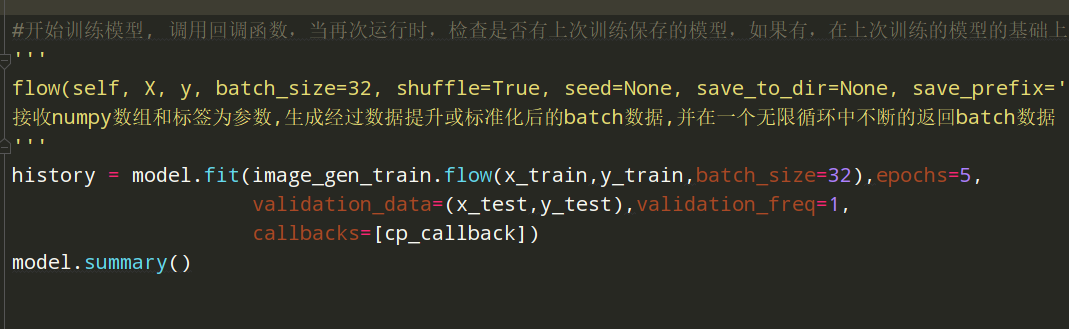


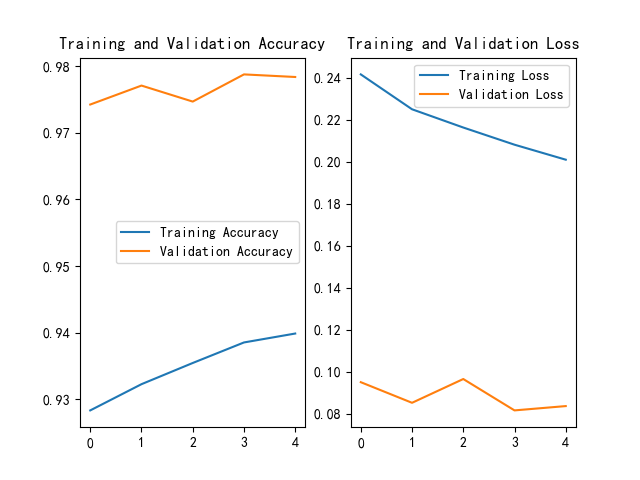
**本次模型构建并训练后，同时展示训练结果，查看部分图片的预测结果**

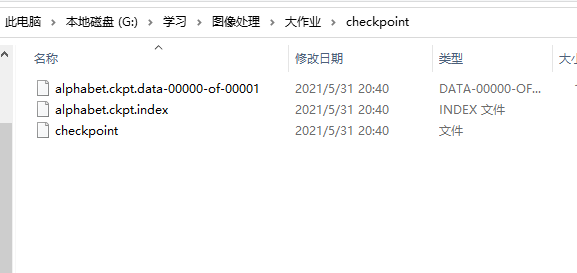


**本次测试预测结果的几张图片全部预测准确。为了提高图片的泛化能力，我决定进行数据增强，同时本次训练将会保留训练结果，将模型保存**



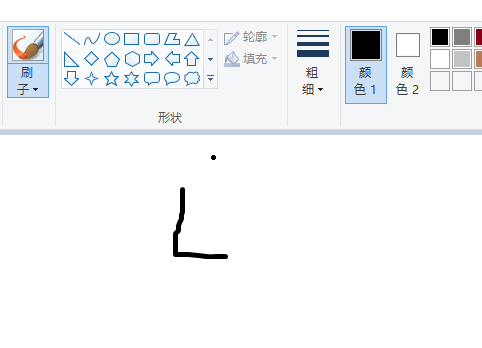


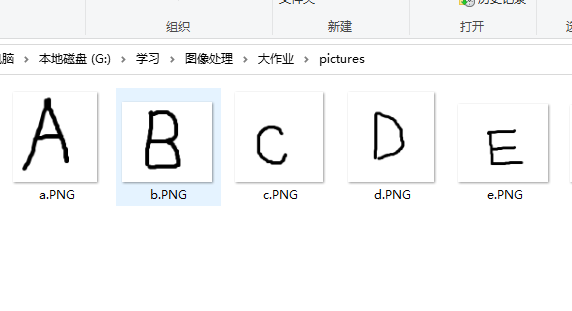


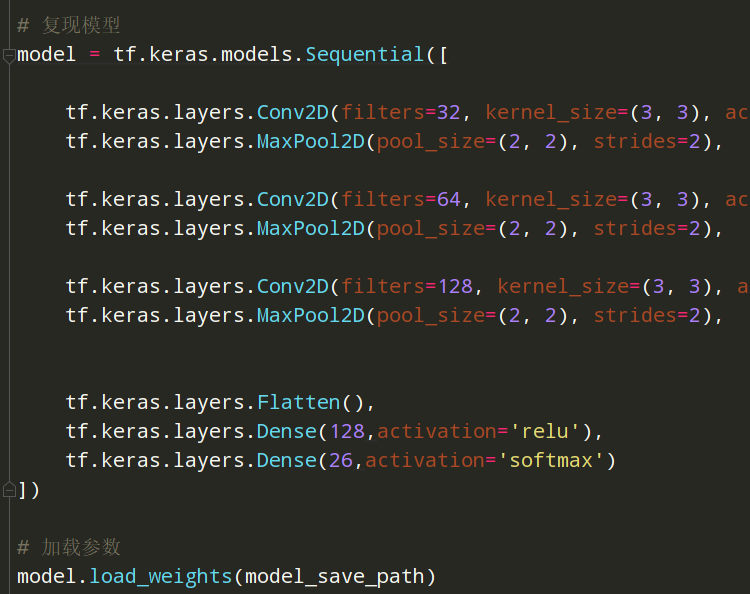


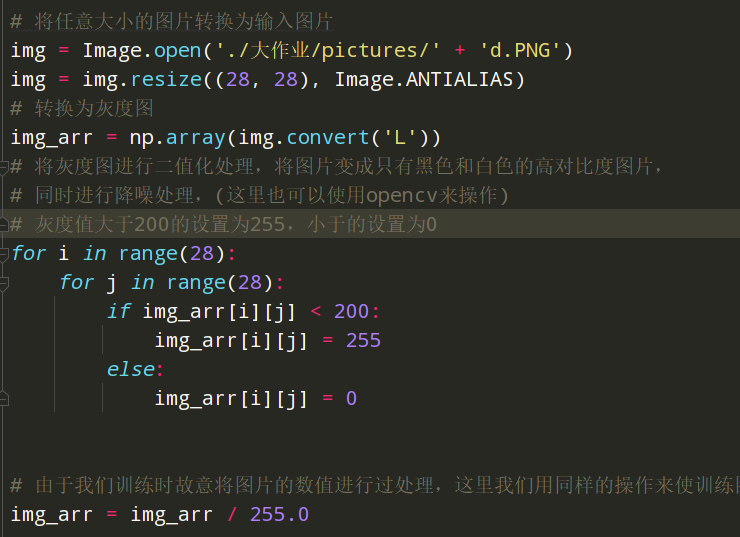
**最后的最后，为了证明本次训练模型的预测的准确性，通过手动绘制字母图片，**

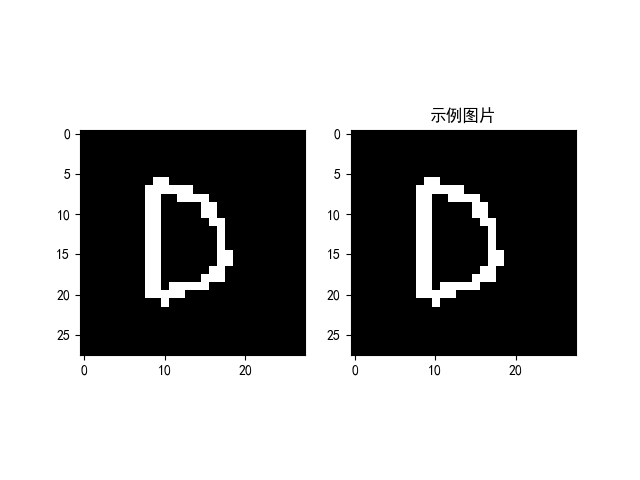
**并放入模型来进行预测**

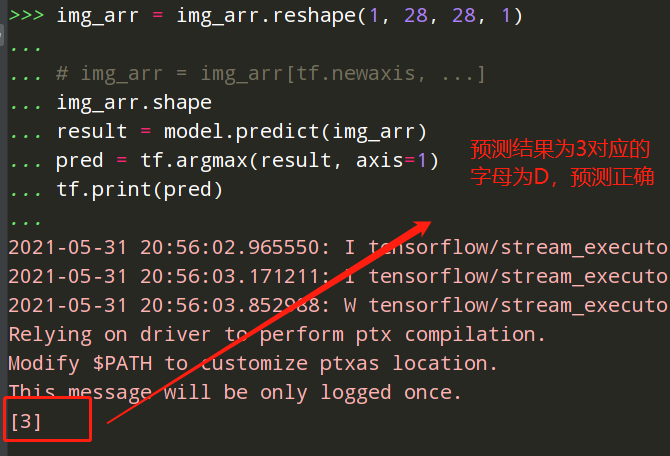






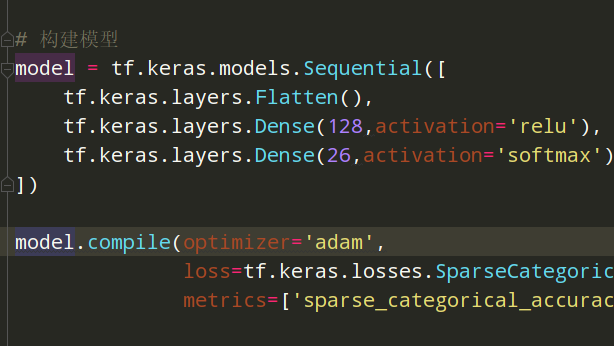




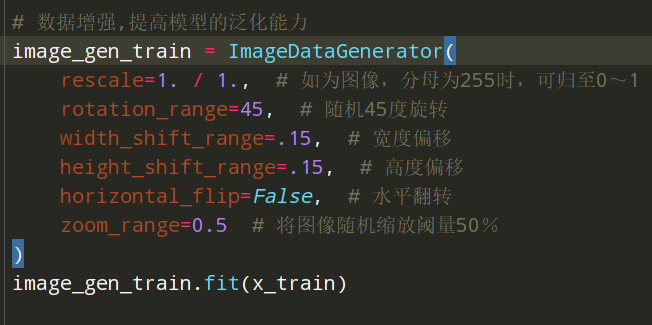


**4、源代码及注释说明**

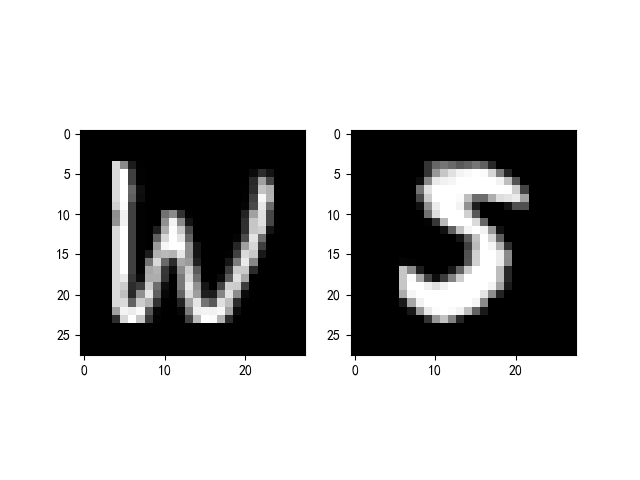
主要源代码展示

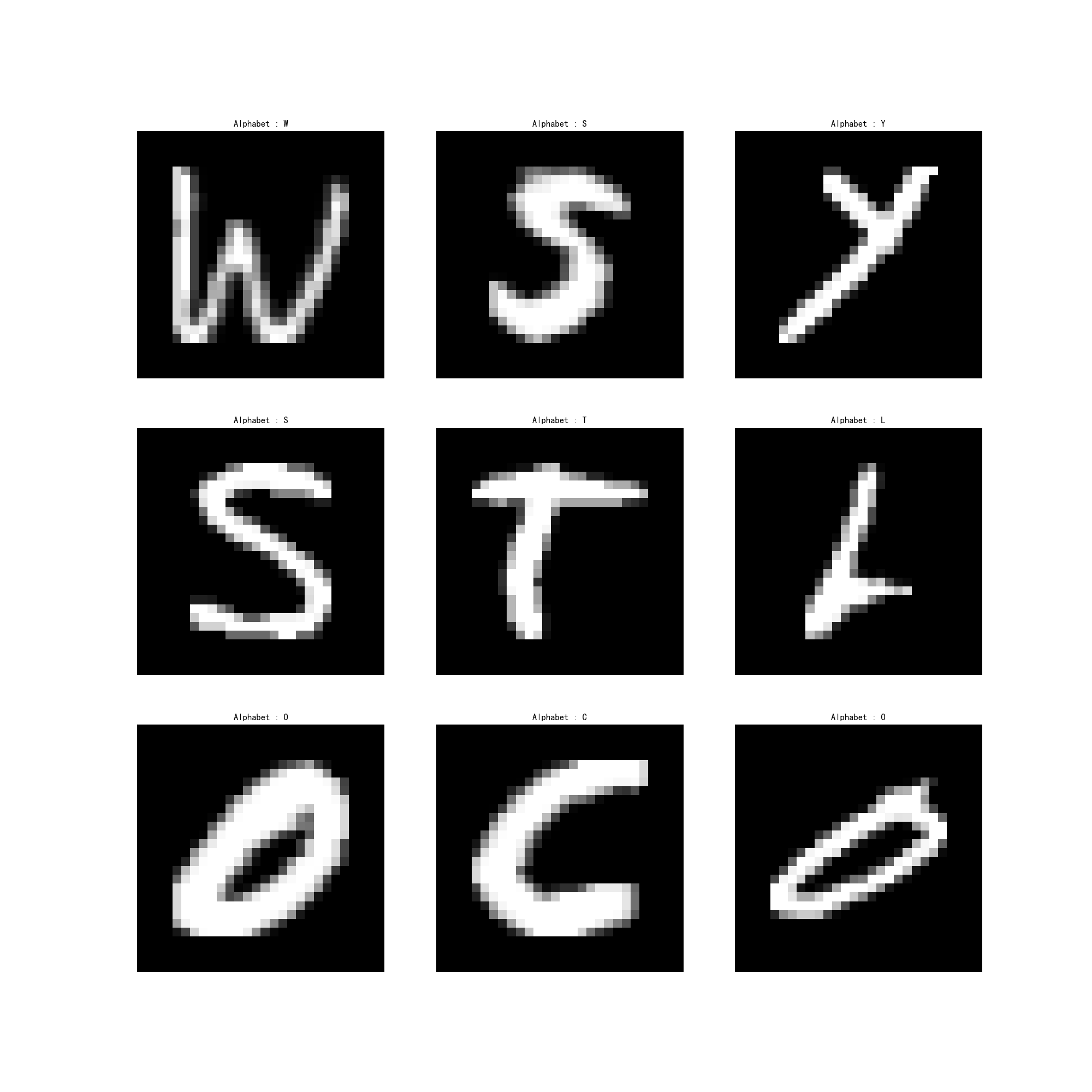


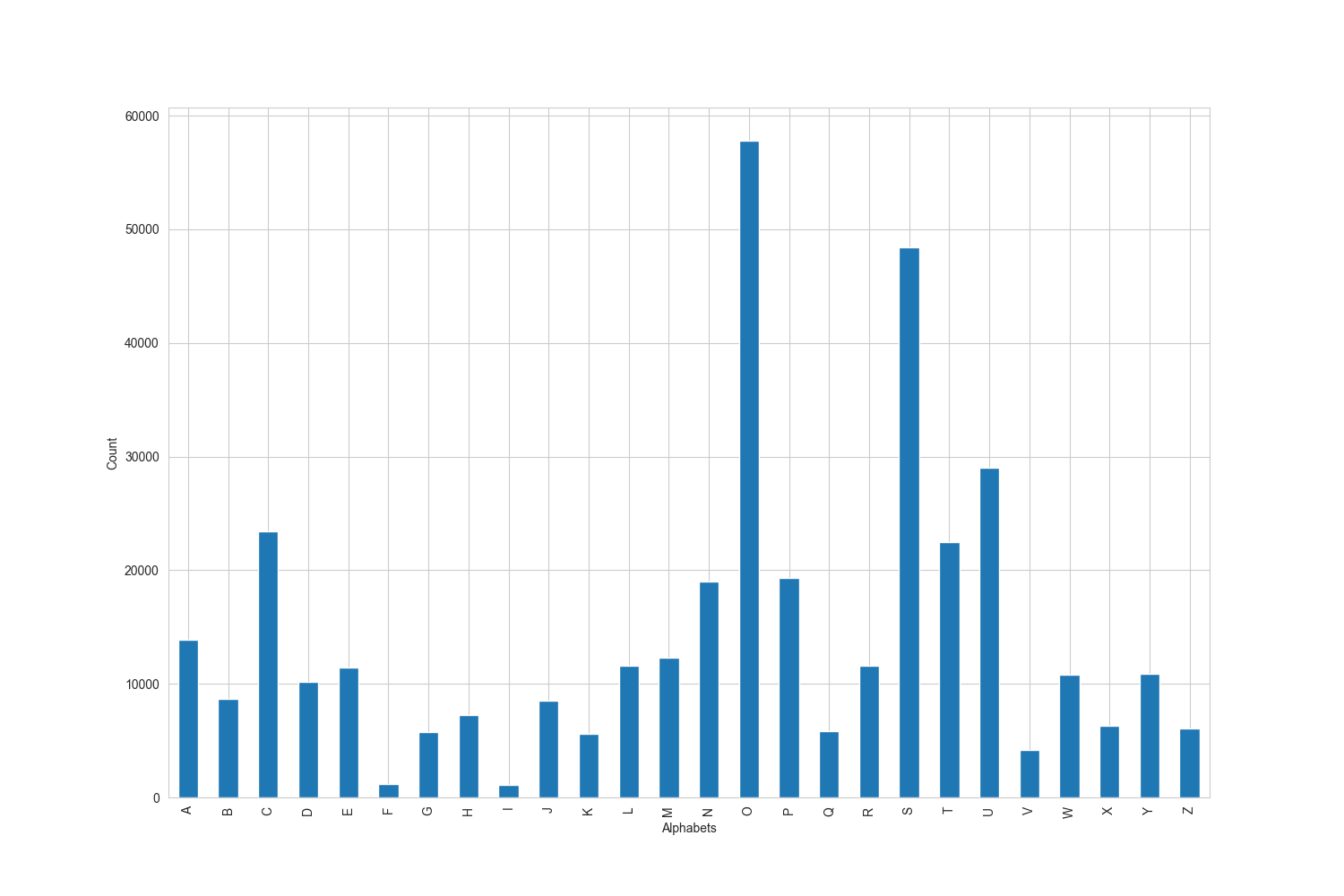


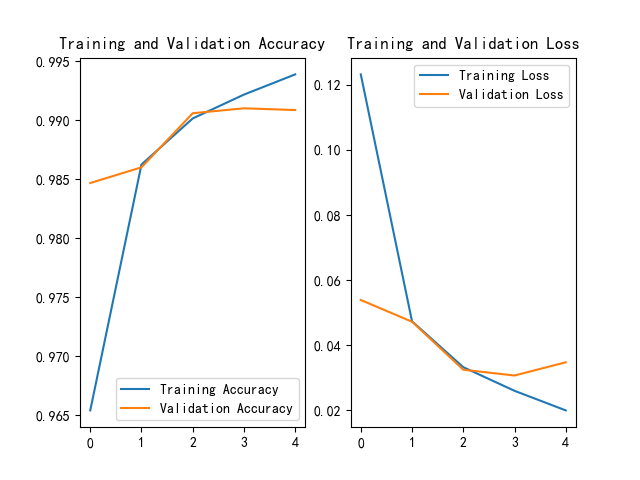


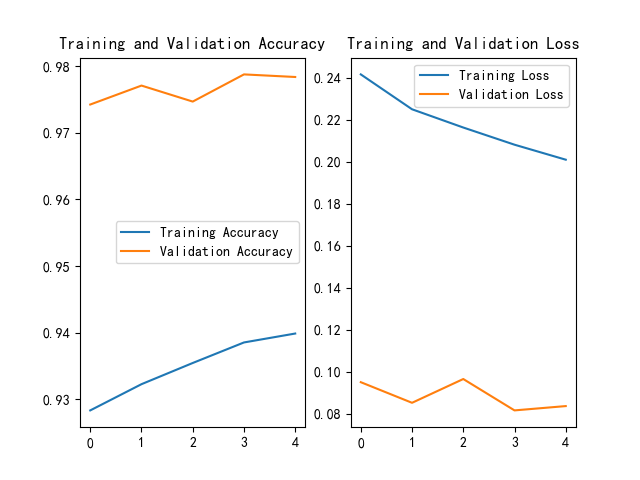
1. **测试数据及运行结果分析**

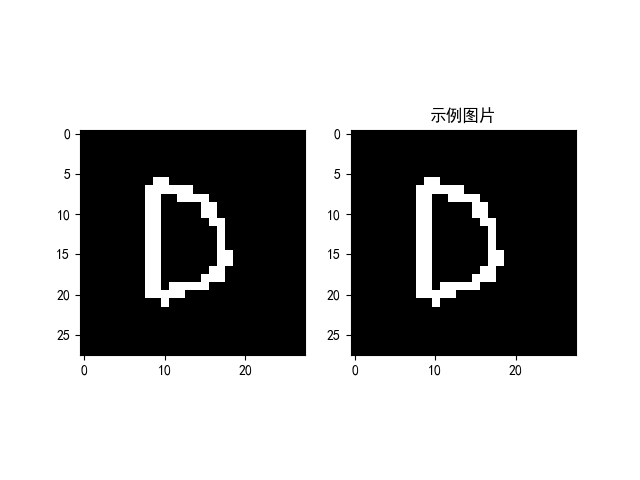












**6、总结与展望**

**在本次实验中，我尽可能详细的展示了字母识别的整个过程，在写本次实验的同时也重新学习了一遍卷积神经网络的过程，了解了整个构建卷积神经网络的构建步骤，并通过手写字母的办法来检验自己模型的结果，结果令人满意，模型的准确率达到了百分之九十以上，同时还提高了模型的泛化能力，但是在本次实验中，关于这部分的能力遗憾不能够清晰的展示，但是总体来说，本次实验结果达到了预期的目标，希望通过之后的学习，能够继续学习卷积神经网络的知识。**

**参考文献**

**课程论文（设计）评分表**

|  |  |
| --- | --- |
| 评分项目 | 得分 |
| **设计方案评价（20分）**  16-20：有完全实现设计方案，难度系数高，设计实现功能复杂  具有创新性、合理性，代码附有详细注释；  11-15：有完全实现设计方案，设计方案较先进，无明显错误；难度系数低，设计实现功能较复杂，设计具有新意、合理性，代码附有详细注释。  5-10：有基本实现设计方案，设计方案正确，但有少数失误；代码附有详细注释  4及以下：没有完全实现设计方案，难度系数很低，设计实现功能一般，设计具有合理性代码无注释 | 12 |
| **课程论文编写（40分）**  31-40：优秀表明报告清楚详细地说明了设计的过程，格式规范、内容完整，阐述清晰，层次分明。  21-30：良好表明报告说明了设计的过程，格式较规范、内容比较完整，阐述清晰，有层次。  10-20：中表明报告说明了设计的过程说明不太清晰，格式不规范、内容比较简单。  9分及以下：差表明报告说明了设计的过程说明不太清晰，格式不规范、内容比较简单。 | 32 |
| **答辩总结（40分）**  31-40：自己的设计阐述清晰，回答问题准确，基本概念非常清楚  21-30：自己的设计阐述较清晰，回答问题还算准确，基本概念比较清楚  10-20：自己的设计阐述清晰，基本能够回答问题，基本概念清楚  9分及以下：自己的设计阐述不清晰，回答问题不准确，基本概念不清楚 | 36 |
| **总得分**： | 80 |