# 计算机视觉实践-练习2实验报告

目录

[计算机视觉实践-练习2实验报告 1](#_Toc131886645)

[一、 实验目的 1](#_Toc131886646)

[二、 实验原理 1](#_Toc131886647)

[三、 实验步骤 3](#_Toc131886648)

[四、 数据集 4](#_Toc131886649)

[五、 程序代码 4](#_Toc131886650)

[六、 实验结果 8](#_Toc131886651)

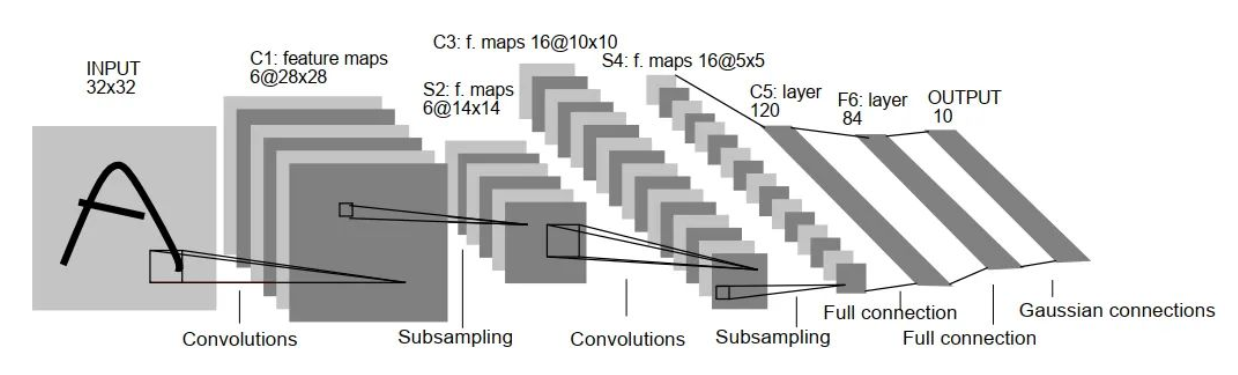
[七、 实验分析与总结 9](#_Toc131886652)

# 实验目的

* 熟悉卷积神经网络的基本结构，包括卷积层，池化层，激活函数及最后的全连接层等。
* 实现LeNet-5在MINIST数据集上的训练和测试

# 实验原理

LetNet-5是一个较简单的卷积[神经网络](https://so.csdn.net/so/search?q=%E7%A5%9E%E7%BB%8F%E7%BD%91%E7%BB%9C&spm=1001.2101.3001.7020)。其结构为输入的二维图像（单通道），先经过两次卷积层到池化层，再经过全连接层，最后为输出层。整个 LeNet-5 网络总共包括7层（不含输入层），分别是：C1、S2、C3、S4、C5、F6、OUTPUT。网络结构图如下图所示：



**输入层（Input Layer）**

输入层（INPUT）是 32x32 像素的图像，注意通道数为1。

**C1 层**

C1 层是卷积层，使用 6 个 5×5 大小的卷积核，padding=0，stride=1进行卷积，得到 6 个 28×28 大小的特征图。

**S2 层**

S2 层是降采样层，使用 6 个 2×2 大小的卷积核进行池化，padding=0，stride=2，得到 6 个 14×14 大小的特征图。S2 层其实相当于降采样层+激活层。先是降采样，然后激活函数 sigmoid 非线性输出。先对 C1 层 2x2 的视野求和，然后进入激活函数。

**C3 层**

C3 层是卷积层，使用 16 个 5×5xn 大小的卷积核，padding=0，stride=1 进行卷积，得到 16 个 10×10 大小的特征图。

**S4 层**

S4 层与S2一样也是降采样层，使用 16 个 2×2 大小的卷积核进行池化，padding=0，stride=2，得到 16 个 5×5 大小的特征图。

**C5 层**

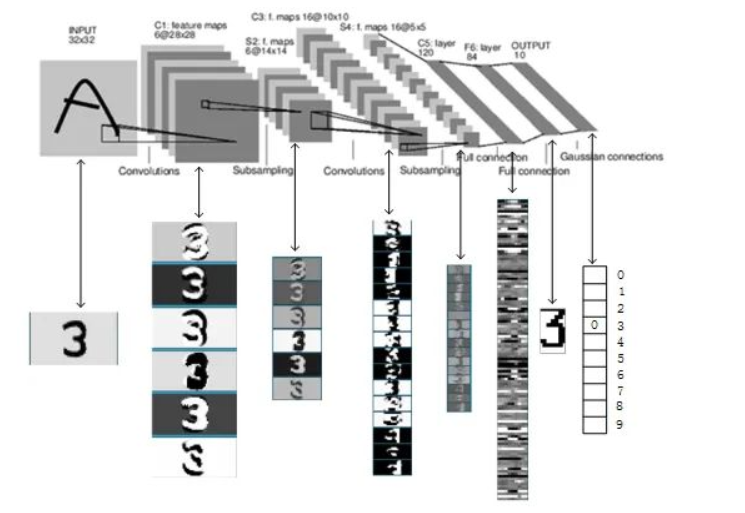
C5 层是卷积层，使用 120 个 5×5x16 大小的卷积核，padding=0，stride=1进行卷积，得到 120 个 1×1 大小的特征图。

**F6层**

F6 是全连接层，共有 84 个神经元，与 C5 层进行全连接，即每个神经元都与 C5 层的 120 个特征图相连。计算输入向量和权重向量之间的点积，再加上一个偏置，结果通过 sigmoid 函数输出。

**OUTPUT层**

最后的 Output 层也是全连接层，是 Gaussian Connections，采用了 RBF 函数（即径向欧式距离函数），计算输入向量和参数向量之间的欧式距离。整体过程如下图所示：



# 实验步骤

1. 下载并加载数据，并做出一定的预先处理
2. 搭建 LeNet-5 神经网络结构，并定义前向传播的过程
3. **将定义好的网络结构搭载到 GPU/CPU，并定义优化器**
4. **定义训练过程**
5. **定义测试过程**
6. 使用 MNIST 数据集进行训练
7. **保存模型**
8. **手写图片的测试**

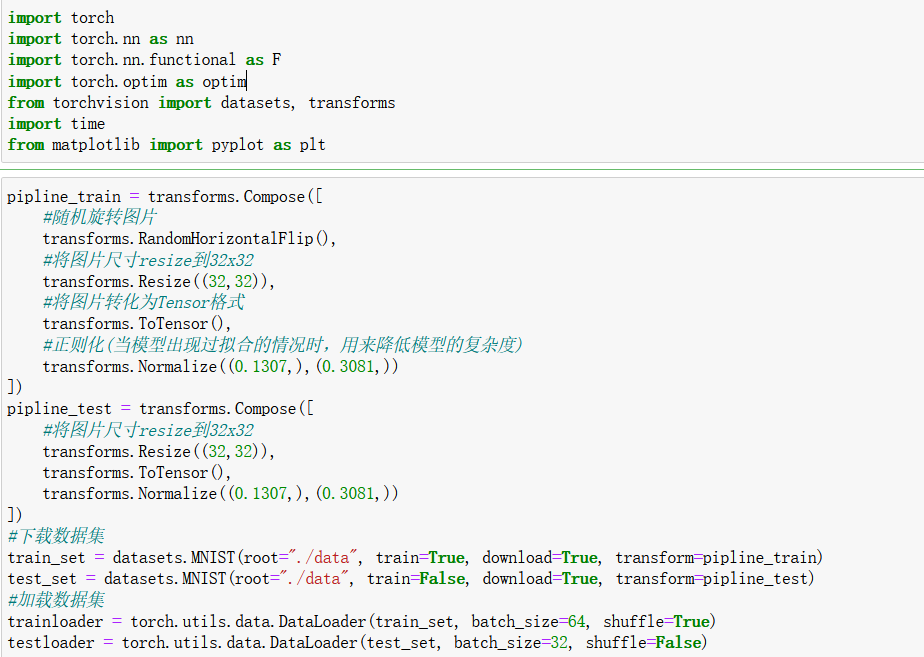
# 数据集

[MNIST](https://so.csdn.net/so/search?q=MNIST&spm=1001.2101.3001.7020)是一个非常有名的手写体数字识别数据集，训练样本：共60000个，其中55000个用于训练，另外5000个用于验证；测试样本：共10000个。MNIST数据集每张图片是单通道的，大小为28x28。部分展示如下：

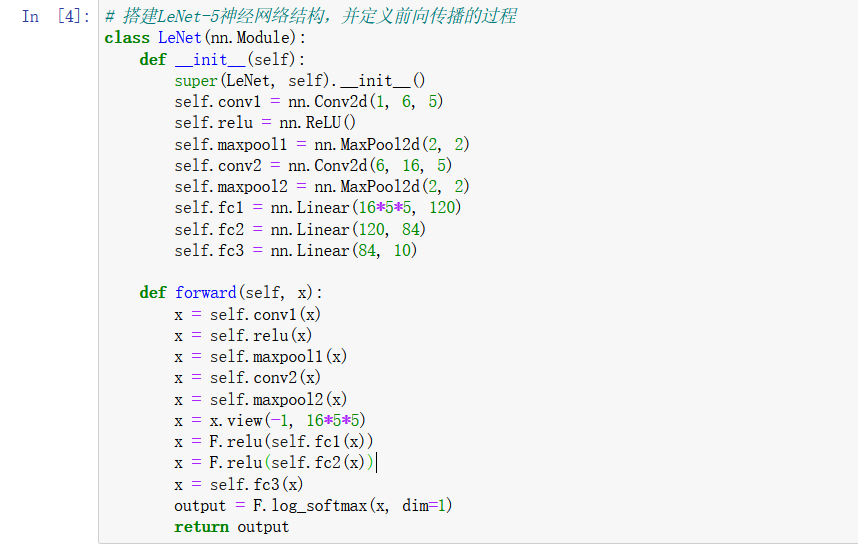


# 程序代码

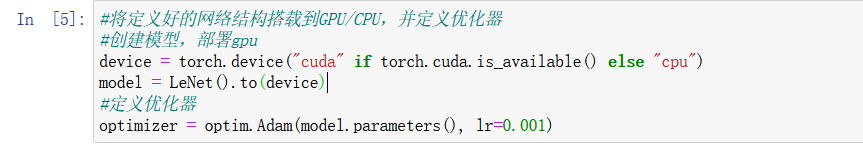
* 首先导入包并加载数据集，如下：



* 搭建LeNet-5神经网络结构，并定义前向传播的过程。



* 定义好的网络结构搭载到GPU/CPU，并定义优化器



* 定义训练过程



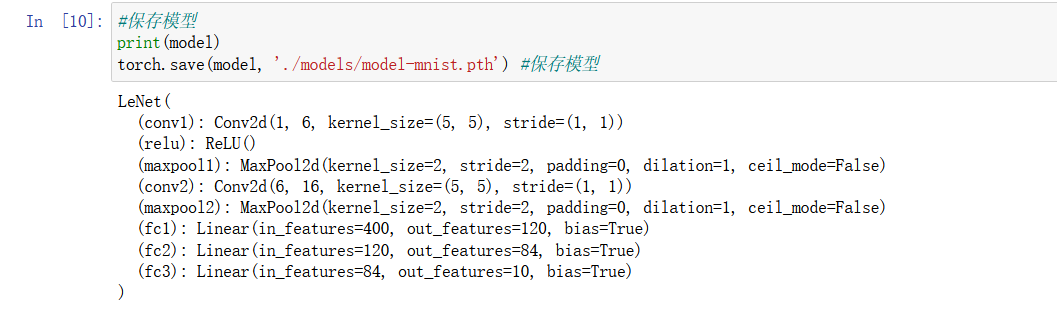
* 定义测试过程



* 运行调用



* 保存模型

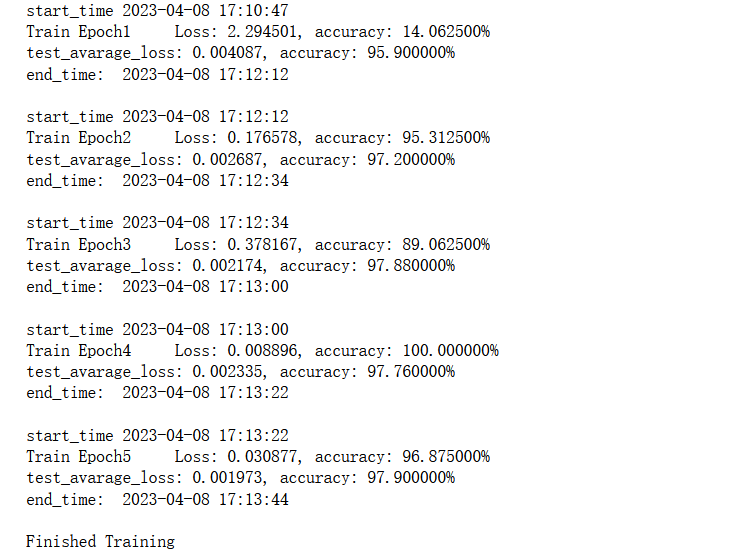


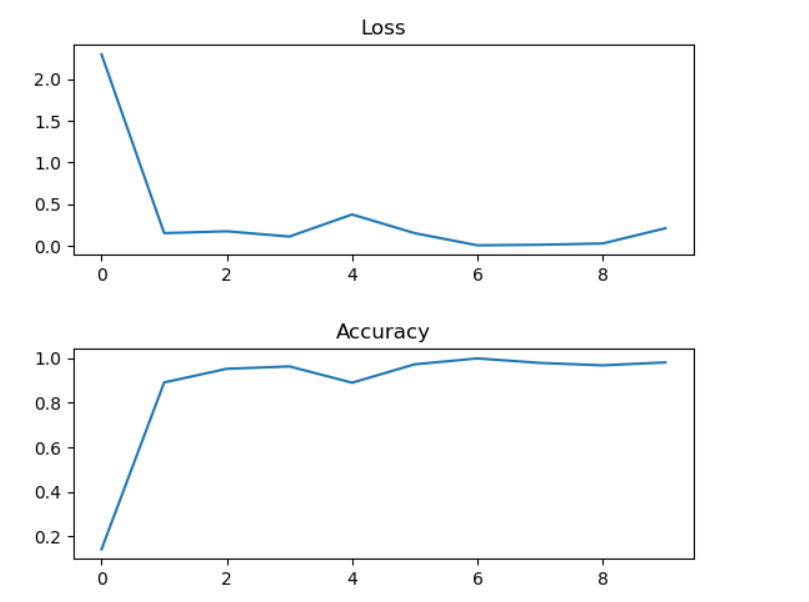
* 预测类别



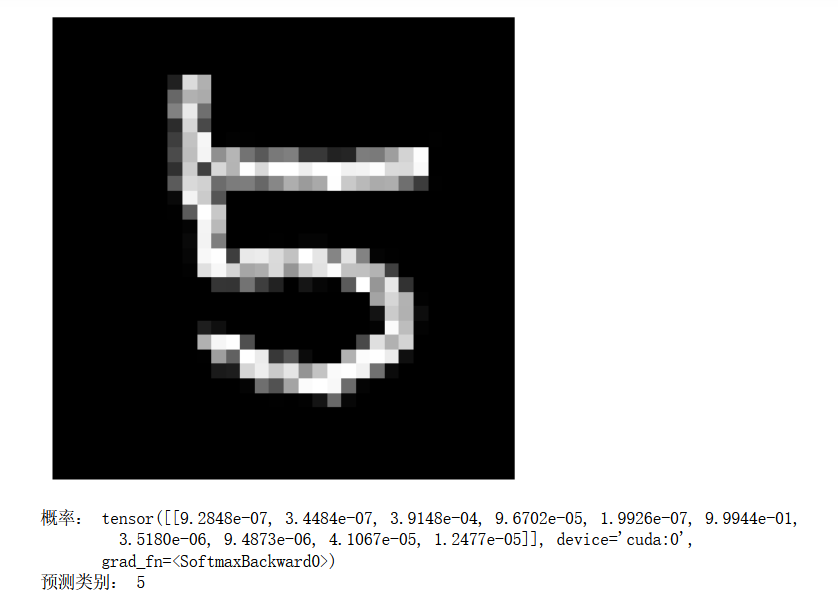
# 实验结果

* 训练结果如下：





* 预测结果如下：



# 实验分析与总结

LeNet-5 与现在通用的卷积神经网络在某些细节结构上还是有差异的，例如 LeNet-5 采用的激活函数是 sigmoid，而目前图像一般用 tanh，relu，leakly relu 较多；LeNet-5 池化层处理与现在也不同；多分类最后的输出层一般用 softmax，与 LeNet-5 不太相同。

LeNet-5 是一种用于手写体字符识别的非常高效的卷积神经网络。CNN 能够得出原始图像的有效表征，这使得 CNN 能够直接从原始像素中，经过极少的预处理，识别视觉上面的规律。然而，由于当时缺乏大规模训练数据，计算机的计算能力也跟不上，LeNet-5 对于复杂问题的处理结果并不理想。