基于卷积神经网络的手写数字识别研究

摘要：基于深度学习卷积神经网络在手写体数字识别中的应用，对卷积神经网络的结构进行介绍，同时使用MNIST数据集对设计的卷积神经网络模型进行训练与测试，训练后的模型在MNIST的测试集样本中的识别精度可以达到98.86%。实验结果表明，与传统的线性回归模型相比，卷积神经网络有更好的鲁棒性和抗干扰性，识别精度明显优于传统的线性回归模型。

关键词：卷积神经网络；MNIST；手写数字识；线性回归；深度学习

引言：

目前，由于手写数字的应用范围广泛，例如邮政编码手写体、智能手机的手写输入法等，使得手写数字识别研究成了深度学习的研究热点之一。手写数字识别研究的内容是让计算机准确、高效地识别人随机手写的数字，但由于不同人的书写方式千差万别，书写的大小与角度也不尽相同，进而增加了对手写数字识别的难度，因此传统线性回归模型不能正确精准地对手写体数字进行识别。在此背景下，由于卷积神经网络结构的不断优化以及计算机性能的不断提升，使得卷积神经网络在图像识别领域取得了重大的突破。卷积神经网络中可以将图像的每一个像素均看成一个特征，直接将手写数字图像作为卷积神经网络模型的输入并自动进行特征提取，再根据提取后的特征进行分类识别，不仅克服了传统方法中需要对图像实时预处理的弊端，还提高了对手写数字的识别精度。

一、卷积神经网络

卷积神经网络的核心思想是通过权值共享和局部感知域的方法减少网络的参数个数，缓解了网络模型过拟合的问题，获得了某种程度的尺度、位移、形变不变性，同时还降低了网络模型的复杂程度。卷积神经网络的基础架构主要包含了五个部分，分别是输入层、卷积层、池化层、全连接层与输出层。其网络结构如图1所示。

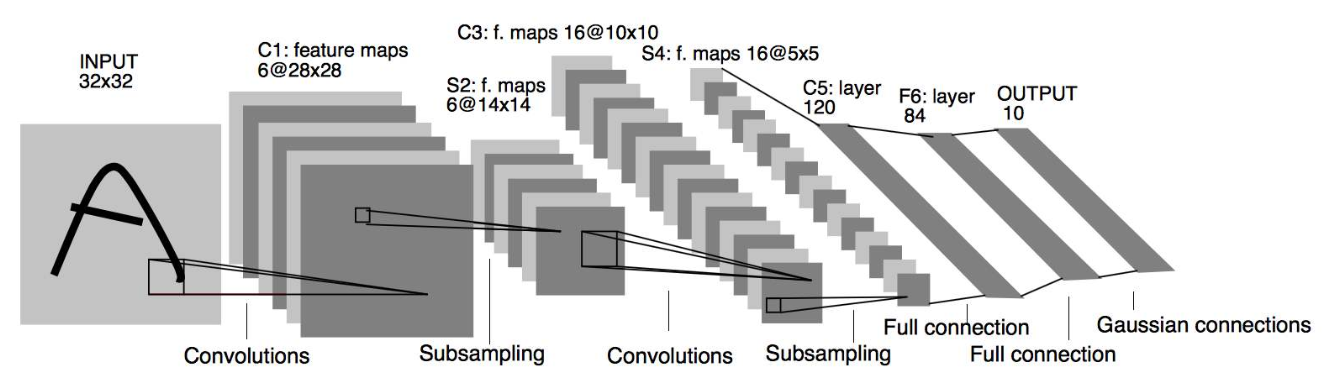


图1 卷积神经网络结构

1.1卷积层

卷积层又称为特征提取层，主要作用就是对输入的数据进行特征提取，其内部包含了多个卷积核，组成卷积核的每个元素都对应一个权重系数和一个偏差量，并且卷积核的个数决定了最终提取的数据特征数量。卷积运算的实质是用一个可训练的卷积核去卷积一个输入的图像，然后加一个偏置，得到卷积结果。卷积的详细过程可以理解为使用一个卷积核依次过滤图像的各个区域，从而得到这些区域的特征值。卷积的流程如图2所示。

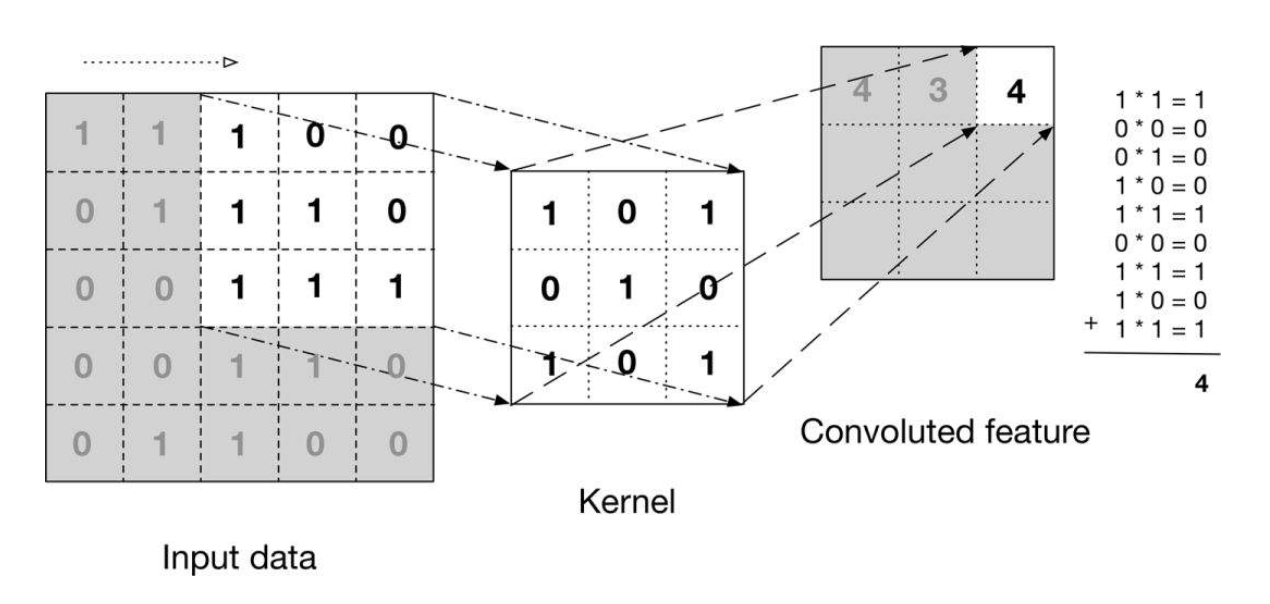


图2 卷积过程

1.2池化层

池化操作的主要目的是降维，以降低计算量，并在训练初期提供一些平移不变性。常用的两种池化操作是平均池化和最大值池化。

池化操作就是使用一个固定大小的滑窗在输入上滑动，每次将滑窗内的元素聚合为一个值作为输出。根据聚合方式的不同，可以分为平均池化和最大值池化。最大值池化如图所示。

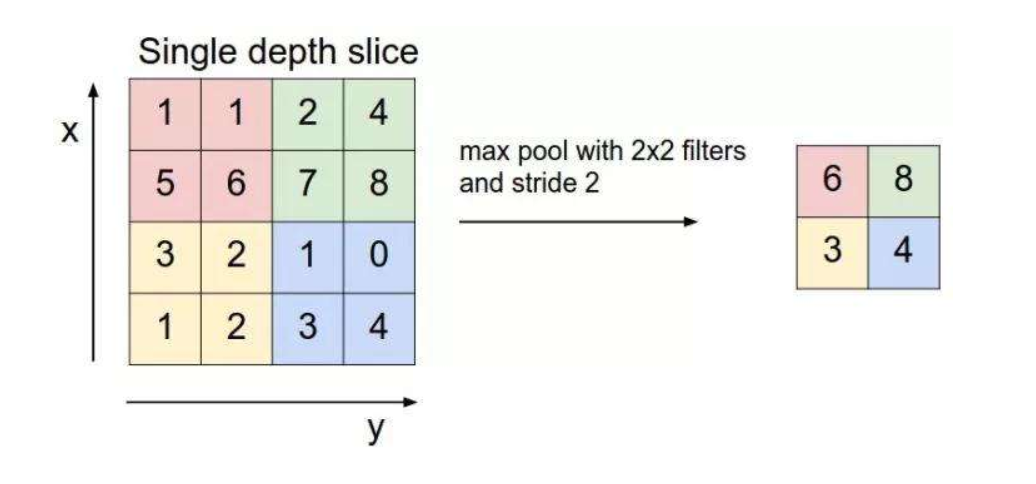


图3 池化

滑窗的大小k和滑动的步长s都会影响最终的输出。通常取,步长与滑窗大小相等，s=2。对于多通道的输入，池化是逐通道进行的，因此不会改变输入的通道数。对于输入，经过上述的池化操作后，输出为，它的长宽将减半。一般来说，对于任意大小的滑动窗口和步长可以使用下式计算。

激活函数是神经网络中的一个十分重要的概念，它的非线性使得神经网络几乎可以任意逼近任何非线性函数。如果不适用激活函数，无论神经网络有多少层，其每一层的输出都是上一层输入的线性组合，这样构成的神经网络仍然是一个线性模型，表达能力有限。

激活函数的选择可以多众多样，一个基本的要求是它们是连续可导的，可以允许在少数点上不可导。常用的激活函数包括S型激活函数和ReLU及其变种等。在隐藏层由于需要对图像特征的提取域降维，为了解决梯度消失的问题，使用了ReLU作为激活函数；在输出层为了避免出现重复分类的现象，使用了softmax作为激活函数。ReLU和softmax函数公式如式所示。

1.3全连接层

全连接层在整个卷积神经网络结构中位于池化层的后一层，显而易见其作用是连接经过卷积和降采样后提取出来的所有高级特征，并将最后的输出值送给分类器得出最终的分类结果。全连接层的最终目的是将学到的“分布式特征表示”映射到样本标记空间，其实现过程是通过对输入的特征做加权求和，再加上偏置量，然后通过softmax 激活函数来获得最终的结果，其表达式如式(4)所示。

上式中 表示全连接层中的权重系数;表示通过卷积层和池化层提取出来的特征；表示全连接层的偏置量。

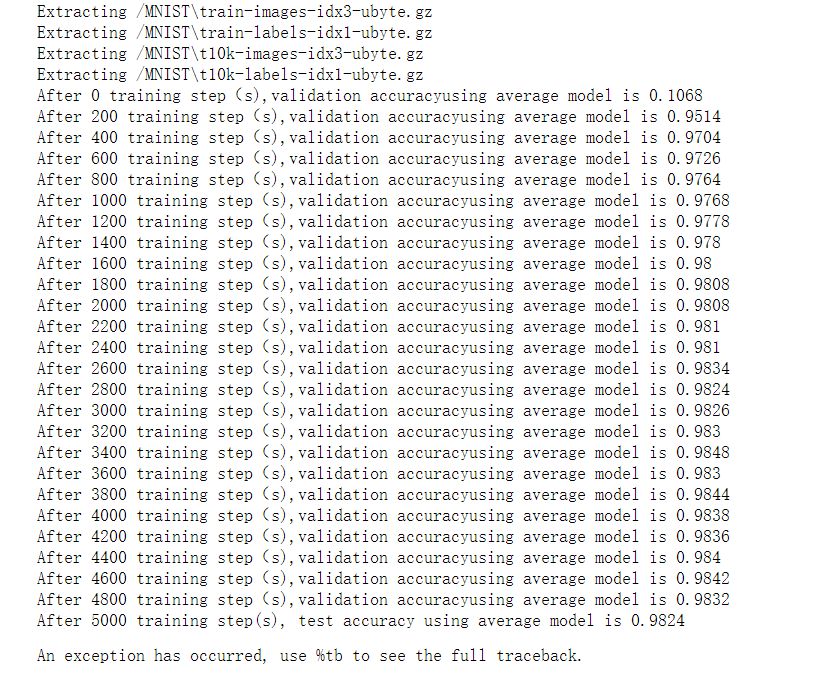
二、实验过程与分析

2.1 模型训练

本实验在Windows10，64位操作系统，Python3.7.0环境下使用MNIST数据集进行模型训练，其中每张图像大小为28×28像素。训练样本为MNIST中的60000张手写数字图像，为了提高模型识别的准确率总共训练5000次，总耗时1402秒；测试的样本为MNIST测试集中的10000张手写数字图像，在该测试集上的正确率为98.86%。卷积神经网络模型的训练过程如图所示。



随后我们使用了传统的线性回归模型进行实验如图所示



它的准确率是98.24%,没有卷积神经网络的结果好。

2.2 实验结果与分析

对10个类别的手写数字的测试表明，经过MNIST训练后的卷积神经网络模型对相对规范的手写数字的识别正确率达到98.8%以上，甚至出现识别率100%情况。对于在相同情况下训练的线性回归模型在手写数字相对规范的情况下对数字0、1、2、3、4、7的识别率也能达到94%以上，但对于数字5、6，9的识别率相对较低；对于数字8，由于其与数字3有较大的相似性，因此线性回归模型将数字8识别成了数字3，而卷积神经网络模型能够正确识别8和3。当手写数字的大小、形状和倾斜度发生变化时，卷积神经网络对数字0、1、2、3、4、7的平均识别率仍能大于在MNIST测试集中的99.25%的识别率，而对于数字5、6、8、9的识别率略有下降，但其平均识别率仍在97%以上，当手写数字发生严重变形时，卷积神经网络模型也会出现识别错误的情况；而在同样的条件下，线性回归模型的识别率大大降低且出现了较多识别错误的情况。因此足以证明，卷积神经网络相对于传统的线性回归模型有着更高的识别率，同时拥有更好的鲁棒性和抗干扰性。

三、总结

本文使用MNIST数据集对卷积神经网络模型进行训练，然后将训练好的模型用于在线手写数字的识别。实验结果表明，在MNIST的测试数据集上的识别准确率为98.86%，因此可以认为本实验的卷积神经网络模型达到了正确识别手写数字的效果。但训练出来的模型还存在一定的局限性，当手写体数字的位置、大小、倾斜度等发生较大变化时，会使得卷积神经网络模型对手写数字的识别精确率略低于MNIST测试数据集上的准确率，甚至会出现识别错误的情况。针对训练后的模型存在的局限性，可以通过对原始的数字图像进行缩放、平移、旋转等操作来扩充训练模型的数据集，进而使得训练后的模型对手写数字的识别有更高的精确率。

参考文献

[1] 马媛媛,史加荣.卷积神经网络及其在手写体数字识别中的应用[J].湖北工程学院学报,2017,37(6):66-72.

[2] 尹晓伟,王真真,孟庆林,等.基于改进的LeNet-5手写数字识别研究[J].信息通信,2019,32(3):17-18.

[3] 吕红.基于卷积神经网络的手写数字识别系统的设计[J].智能计算机与应用,2019,9(2):54-56,62.

[4] 邢萌.基于TensorFlow手写数字识别模型设计与实现[J].电子技术与软件工程,2019(2):56.

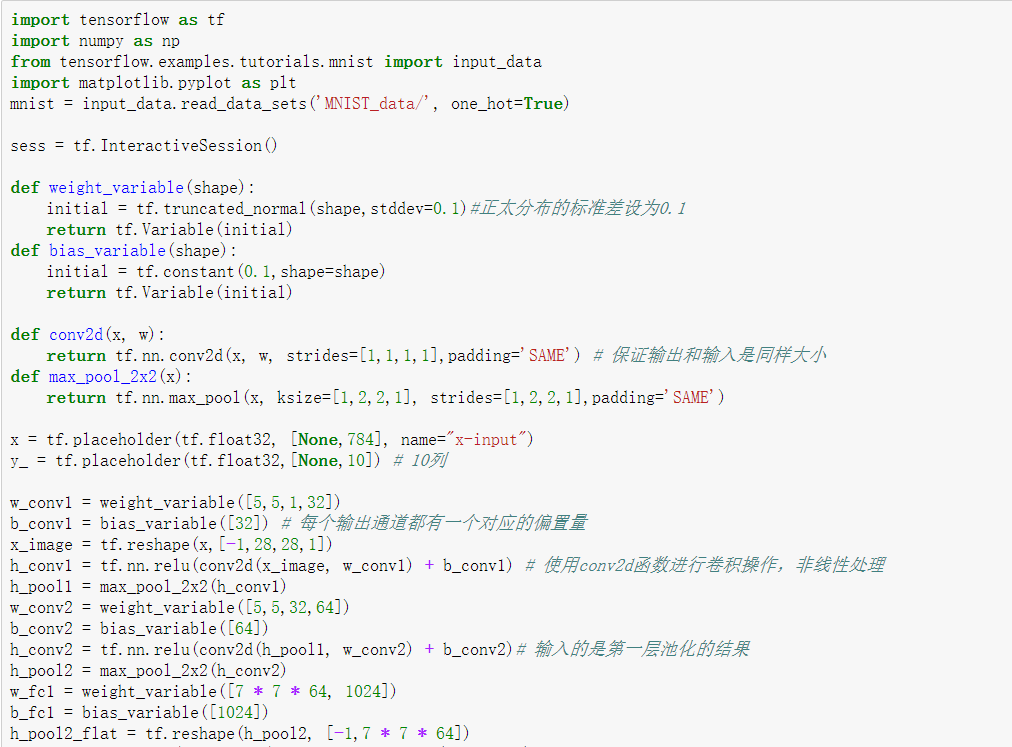
[5] 代贺,陈洪密,李志申.基于卷积神经网络的数字识别[J].贵州师范大学学报(自然科学版),2017,35(5):96-101.

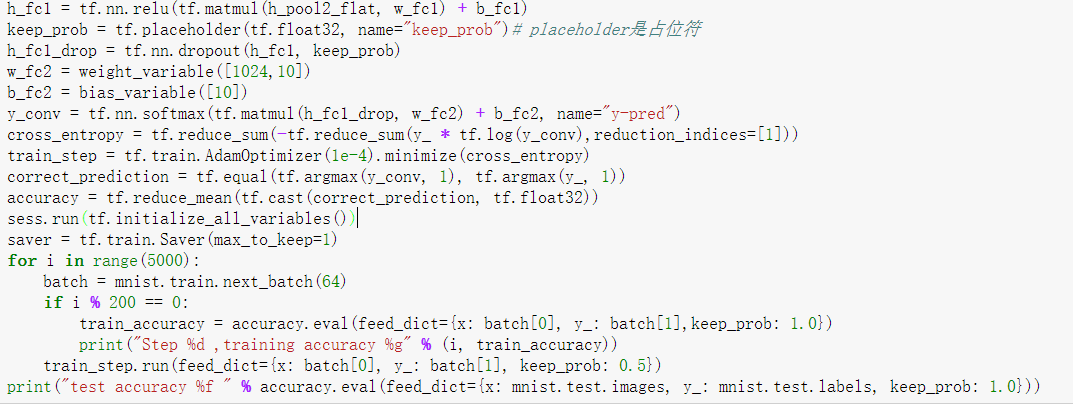
[6] 张涛,杨剑,宋文爱,等.改进卷积神经网络模型设计方法[J].计算机工程与设计,2019,40(7):1885-1890.

[7] 陈岩,李洋洋,余乐,等.基于卷积神经网络的手写体数字识别系统[J].微电子学与计算机,2018,35(2):71-74.

代码：

卷积神经网络





传统神经网络：





