

谭琬滢,左珊珊,邱佩琳,等. 基于深度卷积神经网络的手写数字识别研究[J]. 智能计算机与应用, 2024, 14(8): 138-142.  
DOI:10.20169/j.issn.2095-2163.240823

# 基于深度卷积神经网络的手写数字识别研究

谭琬滢, 左珊珊, 邱佩琳, 李启锐

(广东石油化工学院 计算机学院, 广东 茂名 525000)

**摘要:** 为了进一步提高手写数字的识别准确率,基于深度卷积神经网络对识别过程进行了研究。采用 PyTorch 框架构建了一个手写数字识别模型,然后,使用 MNIST 数据集对模型进行训练以及实验测试。实验结果表明,提出的模型对手写数字识别准确率达到 99.31%,与相关文献、LeNet-5 模型和 ResNet 模型相比分别提高了 11.01%、4.31%、3.31%、1.01%、4.31%、4.51%、10.71% 和 5.31%。

**关键词:** 卷积神经网络; 手写数字识别; MNIST; 深度学习; PyTorch

中图分类号: TP391

文献标志码: A

文章编号: 2095-2163(2024)08-0138-05

## A review of handwritten numeral recognition based on deep convolutional neural network

TAN Wanying, ZUO Shanshan, QIU Peilin, LI Qirui

(College of Computer Science, Guangdong University of Petrochemical Technology, Maoming 525000, Guangdong China)

**Abstract:** In order to further improve the recognition accuracy of handwritten digits, the recognition process was investigated based on deep convolutional neural networks. A handwritten digit recognition model was constructed using the PyTorch framework; then, the model was trained as well as experimentally tested using the MNIST dataset. The experimental results show that the proposed model achieves a handwritten digit recognition accuracy of 99.31%, which is an improvement of 11.01%, 4.31%, 3.31%, 1.01%, 4.31%, 4.51%, 10.71%, and 5.31%, respectively, when compared with the related literature, the LeNet-5 model, and the ResNet model.

**Key words:** convolutional neural network; handwritten numeral recognition; MNIST; deep learning; PyTorch

## 0 引言

随着科学技术的不断发展,以深度学习(Deep Learning, DL)为代表的的人工智能(Artificial Intelligence, AI)正被广泛应用于社会生产生活中。深度学习,即深度神经网络学习,是机器学习领域中一个新的研究方向,其被引入机器学习使其更接近于最初的人工智能目标<sup>[1]</sup>。作为一种人工智能技术,其原理类似于人类大脑的视觉神经网络系统,主要用来进行图像模式识别,如人脸识别、指纹识别等<sup>[2]</sup>。手写字识别也是模式识别的一种,在现实生活中有着广泛的应用,例如,邮件手写地址识别,计

算机、智能手机的手写输入法,照片扫描识别手写文字等等。手写数字识别是使计算机通过训练后,能够准确高效地识别各种手写数字。但是由于手写数字因人而异,风格千变万化,字体的大小和识别度差异明显,这些都是影响手写数字识别准确率的主要因素。因此,如何克服这些因素对手写数字识别准确率的影响是手写数字识别研究中需要解决的重要问题。

近年来,众多学者进行了手写数字识别方面的研究,也取得了一些积极的研究成果。例如:基于 MNIST 数据集,黄晨曦等<sup>[3]</sup>基于级联分类器研发了一个手写体数字识别系统,准确率可达 88.3%;万茹

基金项目: 广东石油化工学院大学生创新创业训练计划项目(73322156)。

作者简介: 谭琬滢(2001-),女,本科生,主要研究方向:图像识别;左珊珊(2002-),女,本科生,主要研究方向:机器学习;邱佩琳(2004-),女,本科生,主要研究方向:机器学习。

通讯作者: 李启锐(1982-),男,博士,副教授,主要研究方向:石化智能信息系统。Email: liqirui@gdupt.edu.cn

收稿日期: 2023-06-13

月等<sup>[4]</sup>通过优化BP神经网络,进一步提高手写数字的识别能力,准确率达到95%左右;杨济萍等<sup>[5]</sup>采用PCA“降维”,对脱机手写数字识别算法模型进行优化,准确率达96%;涂朴等<sup>[6]</sup>提出一个深度学习模型,并采用增加隐藏层和加入激活函数来优化模型,准确率达98.32%。虽然上述一些研究也是基于深度学习的,但是所使用的深度学习框架或神经网络有所不同。为获得更好的识别效果,本文通过使用卷积神经网络技术,结合图像处理技术,基于PyTorch深度学习框架构建手写数字识别模型;使用相同的MNIST数据集,对模型进行训练和测试,完成手写体数字的识别及应用,在大量的手写数字样本数据中得到需要较高的识别准确率。

## 1 相关技术

### 1.1 PyTorch

PyTorch是Facebook人工智能研究院于2017年初推出的一款新型深度学习框架,其前身为Torch。虽然PyTorch的底层和Torch框架一样,但其基于Torch提供了Python接口,即PyTorch可以看作是Torch的Python版本。相比Torch,PyTorch变得更加灵活,还具有强大的GPU加速功能,且能够支持动态神经网络<sup>[7]</sup>。此外,PyTorch的设计还由低到高遵循 $\text{tensor} \rightarrow \text{variable}(\text{autograd}) \rightarrow \text{nn.Module}$ 这3个抽象层次,这3个抽象层次分别为高维数组(张量)、自动求导(变量)和神经网络(层/模块),而且其相互联系紧密,可以同时进行修改和操作<sup>[8]</sup>,使得PyTorch较其他深度学习框架具有简洁、快速、易上手等特点。目前,PyTorch虽然年轻,但增长的势头非常猛烈,并通过与Caffe2的融合来进一步强化了自身的优势<sup>[9]</sup>,已经在智能聊天、文本翻译、文本和语音的转换以及图像识别等应用上取得了众多成就。

### 1.2 卷积神经网络

卷积神经网络(Convolutional Neural Network, CNN)是深度学习的典型模型,是由视觉系统的启发而产生的,其核心概念主要是稀疏连接、共享权值和池化<sup>[10]</sup>。卷积神经网络是受人类视觉神经系统启发设计出的网络结构,其非常擅长处理图像。卷积神经网络可以通过权值共享和稀疏连接等方法来减少参数,这和动物视觉神经细胞的工作原理比较类似,有利于实验人员用更小的训练集来训练,从而预防过度拟合并提高网络的识别准确率<sup>[11]</sup>。卷积神经网络的基础架构主要包含分别为输入层、卷积层、池化

层、全连接层与输出层5个部分。其中,卷积层、池化层、全连接层合称为卷积神经网络的隐含层。

#### 1.2.1 卷积层

卷积层又称为特征提取层,主要作用就是对导入的图片进行特征提取,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,通过对卷积核进行卷积运算可以在相应的图片中提取出特征,并输出特征图。卷积核的个数越多,卷积的层就越多,则输出的特征图中的特征就越复杂。

经过卷积操作后输出的图像大小会比输入时略小一点,可通过填充(Padding)操作使得输出与输入大小相同,通常以对称方式进行填充<sup>[12]</sup>。其运算后的效果类似于滤镜效果,可以帮助提取输入的不同特征,例如边缘、线条和角等<sup>[10]</sup>。

#### 1.2.2 池化层

池化层又称为下采样层,池化层一般出现在两个卷积层之间,主要作用是对上一层的特征图进行降采样<sup>[13]</sup>。若上一层输入单元尺寸不符合2的整数倍时,会采用全零填充法(Padding),即在外沿一圈补零的方法修正后再进行池化。通过池化层,可以有效减少数据量和模型参数量,使得模型的计算量得到减少,在一定程度上控制过拟合<sup>[14]</sup>。其中,最常用的池化是最大池化,其他常用的还有平均池化、重叠池化等。

卷积层和池化层可以在框架中多次叠加搭建,使得最后获得的特征图更精准复杂。

#### 1.2.3 全连接层

全连接层通常位于卷积神经网络的最后一层,其作用是将上一层的输出结果“拉平”成一组一维向量输出,来作为网络训练结果。可以理解是对输入特征图进行信息融合,将图像特征抽象出来进行整合<sup>[15]</sup>。全连接层会利用上一层处理完的特征对原输入图进行分类,其输出是通过对高级特征加权求和,并加上相应偏置量,通过归一化来增加网络的鲁棒性,最终完成分类概率的输出<sup>[16]</sup>。

#### 1.2.4 激活函数

卷积神经网络在进行数据分类时,对于数据的处理都是线性的,可以通过一条或多条直线进行分类,使用多个线性方程划出分类区域。但是在实际的应用中,大部分数据都并不是线性可分的,在卷积神经网络中引入激活函数,就能通过对空间进行折叠、翻转,实现数据从非线性可分到线性可分<sup>[17]</sup>。常见的激活函数有ReLU函数、Sigmoid函数、

Softmax 函数等。其中,ReLU 函数可以使部分神经元的输出变为 0,在一定程度上可以减少参数数量,从而降低计算量,使网络模型的训练更加容易<sup>[18]</sup>。

## 2 识别模型与流程

### 2.1 识别模型

本文构建的卷积神经网络模型如图 1 所示。各个层次的设计如下:

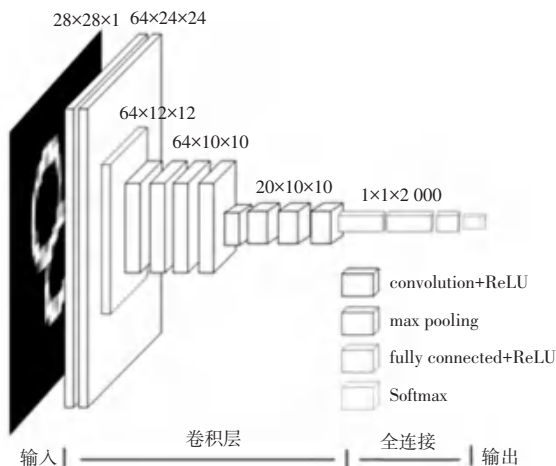


图 1 卷积神经网络模型结构

Fig. 1 Convolutional neural network model structure

**输入层:**本文使用的 MNIST 数据集每个样本都是一张  $28 \times 28$  像素的灰度手写数字图片,即  $28 \times 28$  的矩阵。

**卷积层 c1:**由于 MNIST 数据集的样本均为灰度图片,所以输入的通道数(in\_channels)为 1。第一层卷积层的卷积核过滤器尺寸(kernel\_size)为  $5 \times 5$ ,步长(stride)为 1,可得本层输出尺寸为  $28 - 5 + 1 = 24$ 。输出的通道数(out\_channels)为 64。

**池化层 s2:**输入为上一层卷积层 c1 的输出,即  $64 \times 24 \times 24$ 。本层池化层使用的是  $2 \times 2$  的过滤器,最后输出矩阵为  $64 \times 12 \times 12$ 。

**卷积层 c3:**本层和 c1 同为卷积层,但采用尺寸为  $3 \times 3$  的过滤器,步长依然为 1,可得本层输出尺寸为  $12 - 3 + 1 = 10$ ,输出矩阵为  $64 \times 10 \times 10$ 。

**卷积层 c4:**与 c1、c3 同为卷积层,与 c3 同样采用尺寸为  $3 \times 3$  的过滤器,步长为 1,但图像填充(padding)为 1,可得本层输出尺寸为  $10 - 1 + 1 = 10$ ,输出矩阵为  $20 \times 10 \times 10$ 。

**全连接层:**本层输入矩阵为  $20 \times 10 \times 10$ ,在本层将其拉直成为用一维空间的向量表示形式,长度为  $20 \times 10 \times 10 = 2\ 000$ ,以便在全连接层进行训练。

### 2.2 模型实现

PyTorch 框架下运用卷积神经网络模型实现手写数字识别应用主要包括加载数据、定义超参数、图像预处理等过程,实现步骤如下:

(1)加载必要的库(如 torch 等),可对接下来的操作提供一些必要的支持;

(2)定义超参数。设置每批处理的数据 BATCH\_SIZE 为 32,训练数据集的轮次 Epoch 为 20;

(3)数据预处理。对输入的手写数字图片进行预处理,使得数据都可以符合 PyTorch 中的数据格式。调用 Transforms 变换,将图片转换成 Tensor 格式,并对图片进行正则化,当模型出现过拟合时,可降低模型复杂度;

(4)载入数据。通过代码将 MNIST 数据集自动下载并加载到模型中。为提高模型训练精度,将加载的训练集和测试集使用 Shuffle 打乱;

(5)构建网络模型。搭建 7 层网络构成的网络模型,分别为输入层、卷积层 c1、池化层 s2、卷积层 c3、卷积层 c4、全连接层、和输出层;

(6)定义模型训练方法,进行模型训练。选择相应的优化算法,使用数据集中的数据对模型进行训练;

(7)定义测试方法,进行模型测试。选择相应的优化算法,再使用测试集的数据对模型能否准确识别进行测试验证,最终得出准确率和损失。

识别流程如图 2 所示。



图 2 手写数字识别流程图

Fig. 2 Handwritten digit recognition flowchart

### 3 实验验证与结果分析

#### 3.1 实验环境

本实验在 Windows11 64 位操作系统, GPU 为 NVIDIA CUDA 11.7.99 driver, Python3.9.0, PyTorch1.12, TensorBoard2.9.1 的环境下, 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。实验使用 MNIST 数据集对模型进行训练和测试。训练样本为 MNIST 数据集中的 60 000 张手写数字图像, 测试样本为 10 000 张。神经网络模型基于 PyTorch 框架实现。

#### 3.2 数据集

MNIST 数据集源于美国国家标准与技术研究所, 由 250 个不同人所手写的数字 0~9 构成<sup>[19]</sup>。作为机器学习领域中的一个经典数据集, 其由 60 000 个训练样本和 10 000 个测试样本组成, 每个样本都是一张 28×28 像素的灰度手写数字图片(黑底白字或白底黑字), 每个样本都有对应的标签, 用于描述样本。标签采用独热编码(one-hot Encoding)进行表示, 如标签数字 2 表示为 [0,0,1,0,0,0,0,0,0,0]。其中一部分数据集图像如图 3 所示。

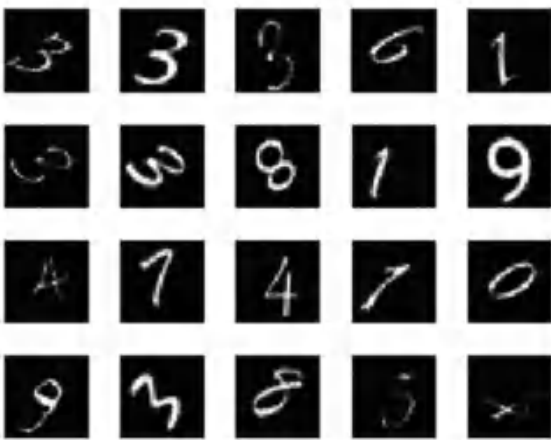


图 3 MNIST 数据集部分图像

Fig. 3 Partial image of the MNIST dataset

#### 3.3 实验结果分析

在模型的训练和学习过程中, 训练集的 Loss 值随 Epoch 的变化如图 4 所示。

从图 4 可以看出, 在经过大约 3 000 轮的迭代训练和学习后, Loss 值逐渐趋于平稳, 说明模型逐渐收敛且没有出现拟合的问题。

在训练过程中, 同时进行了模型的测试, 测试结果如图 5 所示。

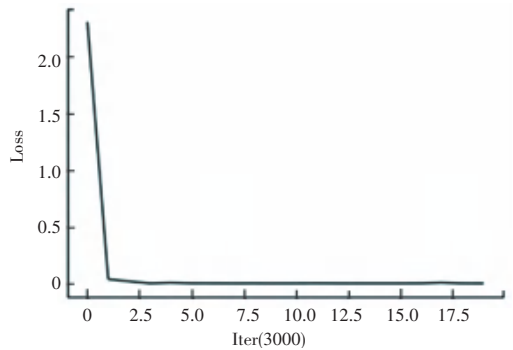


图 4 Loss 值

Fig. 4 Training Loss rate

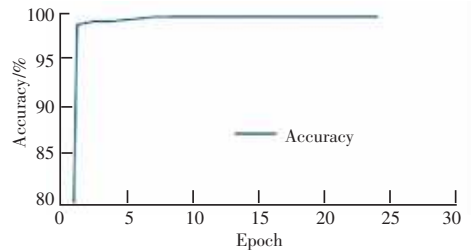


图 5 准确率

Fig. 5 Accuracy rate

进行了 50 次测试实验后取平均值的结果如图 5 所示。可以看出, 均过多轮 Epoch 训练后, 模型的准确率达到 99.31%。

在同样的数据集下, 本文所采用的手写数字识别方法与其他传统的手写数字识别方法的准确率对比情况如图 6 所示。可以看出, 相对于其他方法以及一些传统方法, 本文所提出的方法结果准确率最高。

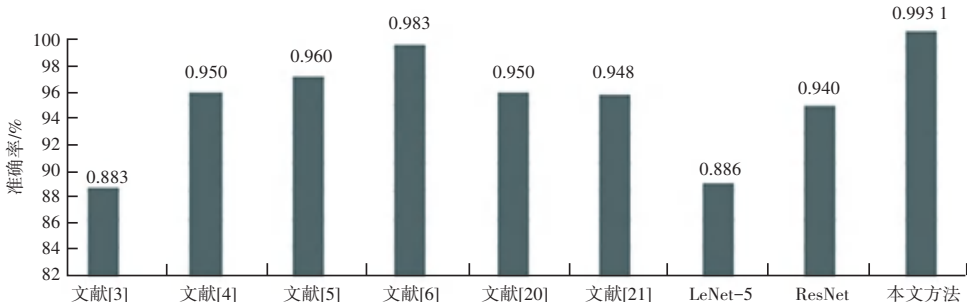


图 6 各模型在 MNIST 数据集上的测试准确率

Fig. 6 Test accuracy of each model on the MNIST dataset



## 4 结束语

本文介绍了深度学习的优势与应用,引入手写数字识别的应用,详细介绍了 PyTorch 的特点与优势,以及深度学习中的经典模型——卷积神经网络。并将两者相结合,再采用 MNIST 数据集作为实验数据,运用 Python 语言实现了手写数字识别的应用,取得了较好的识别效果。但由于实验所使用的计算机性能、搭建的模型能力有限,结果数据还尚待提升,下一步会继续研究使用更复杂的模型以及更好的算法进行优化,来减少损失。

## 参考文献

- [1] 陈先昌. 基于卷积神经网络的深度学习算法与应用研究[D]. 杭州: 浙江工商大学, 2014.
- [2] 吴炼. 基于卷积神经网络的边坡病害图像识别研究[D]. 南京: 南京邮电大学, 2023.
- [3] 黄晨曦, 郝泳涛. 基于级联分类器的手写体数字识别研究[J]. 电脑知识与技术, 2016, 12(20): 180-182.
- [4] 万茹月, 海玲, 谷铮, 等. 基于深度学习的手写文字识别[J]. 现代信息科技, 2021, 5(19): 4.
- [5] 杨济萍. 基于主成分降维模型的手写数字识别研究[J]. 网络安全技术与应用, 2021(3): 31-32.
- [6] 涂朴, 黄晨. 基于 Tensorflow 深度学习框架的手写体数字识别模型优化及应用[J]. 自动化技术与应用, 2020, 39(12): 110-114.
- [7] 黄一天, 陈芝彤. Pytorch 框架下基于卷积神经网络实现手写数字识别[J]. 电子技术与软件工程, 2018(19): 147.
- [8] 李刘刘. 基于生成式对抗网络的室内家居场景生成算法[D]. 杭州: 浙江工商大学, 2019.
- [9] 黄玉萍, 梁炜萱, 肖祖环. 基于 TensorFlow 和 PyTorch 的深度学习框架对比分析[J]. 现代信息科技, 2020, 4(4): 4.
- [10] 张新勇, 甘恒, 李昌夏, 等. 基于卷积神经网络的手写数字识别研究[J]. 电脑知识与技术: 学术版, 2021, 17(22): 3.
- [11] 韩蒙蒙. 基于卷积神经网络的手势识别算法研究[D]. 长春: 吉林大学, 2017.
- [12] 肖祥位. 基于 FPGA 的卷积神经网络实现与验证[D]. 成都: 中国科学院大学(中国科学院光电技术研究所), 2021.
- [13] 张远. 基于卷积神经网络的磁声耦合声信号处理算法研究[D]. 西安: 西安石油大学, 2023.
- [14] 刘赞. ReLU 激活函数下卷积神经网络的不同类型噪声增益研究[D]. 南京: 南京邮电大学, 2023.
- [15] 黎光艳. 基于轻量级神经网络的手写字符识别研究[D]. 杭州: 中国计量大学, 2022.
- [16] 邓国立, 杨春山. 基于 OpenCV-Python 的卷积神经网络手写数字识别研究[J]. 桂林航天工业学院学报, 2021, 26(4): 7.
- [17] 李慧. 手写体数字识别的卷积神经网络研究与 FPGA 实现[D]. 西安: 西安石油大学, 2021.
- [18] 宗春梅, 张月琴, 石丁. PyTorch 下基于 CNN 的手写数字识别及应用研究[J]. 计算机与数字工程, 2021, 49(6): 1107-1112.
- [19] 谢东阳, 李丽宏, 苗长胜. 基于改进 AlexNet 卷积神经网络的手写体数字识别[J]. 河北工程大学学报(自然科学版), 2021, 38(4): 102-106.
- [20] 王俊杰. 优化 BP 神经网络在手写体数字识别中的性能研究[J]. 电子设计工程, 2017, 25(6): 27-30.
- [21] 茹晓青, 华国光, 李丽宏, 等. 基于形变卷积神经网络的手写体数字识别研究[J]. 微电子学与计算机, 2019, 36(4): 5.