

文章编号:1673-9469(2021)04-0102-05

DOI:10.3969/j.issn.1673-9469.2021.04.015

# 基于改进 AlexNet 卷积神经网络的手写体数字识别

谢东阳, 李丽宏\*, 苗长胜

(河北工程大学 信息与电气工程学院, 河北 邯郸 056038)

**摘要:** 为了提高手写体数字的识别率,在 AlexNet 网络模型的基础上进行改进,引入 Inception-resnet 模块替换模型中的 Conv3 和 Conv4 来提升模型的特征提取能力;使用批归一化处理(BN)方法加快网络的收敛速度,防止过拟合;减少卷积核的数量,提升网络的训练速度。在 MNIST 数据集上进行训练与测试,实验结果表明改进的网络模型具有较高的检测精度,达到了 0.9966,证明了本算法的有效性。

**关键词:** 手写数字识别; AlexNet 卷积神经网络; Inception-resnet 模块; 批归一化处理

**中图分类号:** TP391.4

**文献标识码:** A

## Handwritten Number Recognition Based on Improved AlexNet Convolutional Neural Network

XIE Dongyang, LI Lihong\*, MIAO Changsheng

(School of Information and Electrical Engineering, Hebei University of Engineering, Handan, Hebei 056038, China)

**Abstract:** In order to improve the recognition rate of handwritten numbers, we have improved AlexNet network model in this paper. Conv3 and Conv4 were introduced to replace the model for Inception-resnet module, which improves the feature extraction capability of the model. The Batch Normalization (BN) method was used to accelerate network convergence and prevent overfitting, reducing the number of convolutional kernels and improving the training speed of the network. In this paper, training and testing are carried out on MNIST data sets. Experimental results show that the improved network model has a better detection accuracy of 0.9966, which proves the effectiveness of the algorithm.

**Key words:** handwritten number recognition; AlexNet convolutional neural network; Inception-resnet module; batch normalization

手写体数字识别是光学字符识别(Optical Character Recognition, OCR)的一个分支,是利用计算机将图片中的文字信息转换为计算机语言的过程。数字识别是一种图像分类问题,一直是机器深度学习的一个热点研究问题,在快递单号、财务报表、手写票据等领域广泛应用<sup>[1-2]</sup>。数字识别的重点在于算法对图像本身的特征提取,传统的分类算法都存在特征提取不充分的问题,如贝叶斯分类法、K最近邻算法、支持向量机(Support Vector Machines, SVM)、BP(Back Propagation)神经网络等<sup>[3]</sup>。

卷积神经网络(Convolutional Neural Network, CNN)的出现,极大地提高了手写数字识别的准确

率。CNN是由美国学者Cun提出的一种深度前馈人工神经网络,包括卷积层、池化层、全连接层和输出层<sup>[4]</sup>。CNN通过对提取到的图像特征自动学习,能够获得更高层次的特征表达,使学习到的特征信息具有更好的泛化能力<sup>[5]</sup>。邓长银等<sup>[6]</sup>通过改变LeNet-5模型的网络层数、激活函数以及输入图片的尺寸,相较传统方法识别率有所提高。茹晓青<sup>[7]</sup>等提出了一种基于形变卷积神经网络的识别方法,利用形变卷积代替VGG16模型中的普通卷积,提高了多变外观数字的识别精度。陈玄等<sup>[8]</sup>提出了一种融合卷积神经网络,通过融合两种网络的高级特征,增加网络层的高级尺寸,增强

收稿日期:2021-05-05

基金项目:河北省省级科技计划资助项目(20475702D);邯郸市科学技术局项目(19422031008-14)

作者简介:谢东阳(1995-),男,河南信阳人,硕士研究生,主要研究方向为计算机视觉。

\* 通讯作者:李丽宏(1973-),女,河北无极人,博士研究生,教授,主要研究方向为计算机视觉。

网络的表达能力,识别的精度高于两种模型单独使用。

尽管卷积神经网络在图像分类任务中表现优越,但是对于分辨率较小的图像,深度网络会浪费过多的计算资源。本文以 AlexNet 网络结构为基础,减小模型的卷积核尺寸与数量,引入 Inception-Resnet 模块提升算法的特征提取能力,并使用批标准化 (Batch Normalization, BN) 加快模型的训练速度。通过对实验结果的分析,验证了本文方法的有效性。

## 1 相关工作

### 1.1 AlexNet 网络

AlexNet 是由多伦多大学的 Krizhevsky 等<sup>[9]</sup>在 2012 年提出的,并在当年的 ImageNet 大赛上打破了图像分类的记录获得了比赛的冠军,使图像分类的准确率提高了一倍多。

AlexNet 的网络结构如图 1 所示, AlexNet 网络模型结构简洁,由 5 层卷积层、2 层隐藏全连接层和一层输出全连接层组成。整个网络结构使用 ReLu 代替 Sigmoid 作为激活函数,能有效地解决 Sigmoid 在层数较多的网络结构中出现的梯度弥散问题。同时网络在全连接层后添加 Dropout 操作,利用随机隐退神经元个数的方法,在网络的训练过程中减少模型的参数量,能够有效地防止过拟合现象的出现。在 Relu 激活函数后添加局部响应归一化 (Local Response Norm, LRN),通过对局部神经元创建响应竞争机制,提高了网络的泛化能力。使用重叠的最大池化层代替平均池化,很好地解决了平均池化的模糊问题,同时一定程度上提高了图像特征的丰富性。AlexNet 模型相比于 LeNet5 和传统的机器学习方法具有更高的识别精度,但是本身的参数量巨大,模型的训练时间长,识别速度慢。

### 1.2 Inception-Resnet 模块

Inception 模块将多个卷积层与一个池化层并联在网络中,模型在训练的过程中自主选择使用哪种方式获取特征信息,不需要人为地确定参数,是一种高效的稀疏结构,有利于提取到丰富的特征<sup>[10]</sup>。使用多种尺寸的卷积核,提取不同层次的特征,增加了特征的多样性。Inception 将不同尺寸的特征融合在一起有利于后面的分类任务。并利用小卷积代替大卷积,使用  $1 \times n, n \times 1$  的非对称卷积代替  $n \times n$  的卷积核,大大减少了网络的参数量,提升了网络的运算速度。Inception 结构增加了网络的宽度,在深层卷积神经网络中能提高网络的鲁棒性和泛化能力。

残差网络 (Residual Network, ResNet) 是 He<sup>[11]</sup>在 2016 年提出的一种深层卷积神经网络结构,用于解决在模型深度加大时出现的网络退化现象,提升网络的深度。ResNet 运用残差学习的思想,在数据传输过程中添加一个 shortcut 连接,使信息可以直接传输到任一层,保护了信息的完整性,能获得表达能力更强的特征,加速网络的训练。

Inception-Resnet-v2 模型是谷歌团队在 2016 年发布的卷积神经网络,是 Szegedy 在 Inception 结构的基础上与 ResNet 相结合提出的一种新的网络结构,使网络的深度进一步增加,并增强了网络的非线性,加速了网络的训练。

### 1.3 Batch Normalization

批归一化算法是由 Ioffe 等提出的一种训练优化方法,也是网络的一层,一般用在卷积层之后、激活函数之前,简称 BN 层。BN 的作用是在训练的过程中将每层神经元的输入数据归一化到均值为 0,方差为 1 的概率分布,同时添加两个超参数,然后才进行激活函数的运算,可以大大减少网络的训练时间。在训练过程中从总的训练数据集中

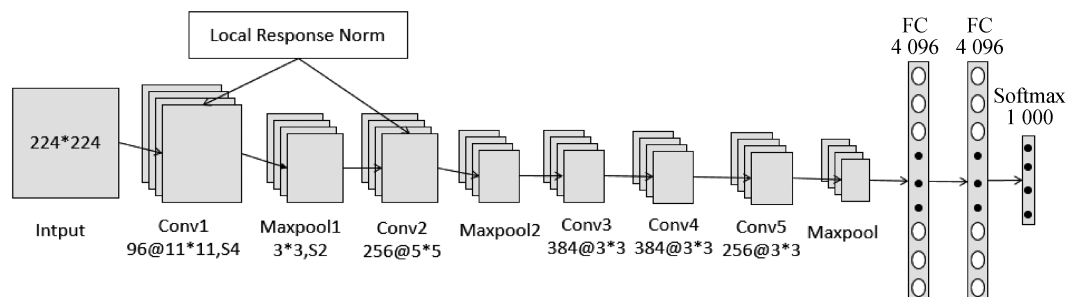


图 1 AlexNet 网络模型

Fig. 1 AlexNet network model

选取一个 mini-batch 数据集对网络进行训练,将这一个 batch 的输入做归一化处理。设  $x \in R_m^n$ , 其中  $m$  为批次大小,  $n$  为特征图的维度。计算每个 batch 的均值和方差:

$$\mu_k = \frac{1}{m} \sum_{i=1}^m x_i \quad (1)$$

$$\sigma_k^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu_k)^2 \quad (2)$$

对输入进行归一化处理

$$\hat{x}_k = \frac{x_k - \mu_k}{\sqrt{\sigma_k^2 - \varepsilon}} \quad (3)$$

式中:  $\varepsilon$  为一个极小值,防止分母为 0,  $k$  表示当前的维度。这样就将每个 batch 的输入数据归一化到  $N(0, 1)$  的正态分布。但是只对数据进行归一化可能会改变原本输入数据的特性或分布,会影响到当前层所学习到的特征,所以又引入了两个超参数对正态分布进行平移和缩放:

$$y = \gamma \cdot \hat{x} + \beta \quad (4)$$

式中:  $\gamma$  表示缩放,  $\beta$  表示平移,一般初始化  $\gamma = 1, \beta = 0$ 。

BN 层使模型可以使用大的学习率,加快网络的收敛速度;有类似于 Dropout 和正则化的效果,能防止网络的过拟合问题,提高模型的泛化能力。

## 2 算法改进

原始 AlexNet 模型的输入尺寸为  $224 \times 224$ , 为了使 AlexNet 模型能应用于 MNIST 数据集,将模型的第一个卷积层的卷积核从  $11 \times 11$  改为  $5 \times 5$ , 步长从 4 改为 1, 改变后的模型刚好能实现分辨率为  $28 \times 28$  的 MNIST 数据集的识别任务,得到了 99.33% 的识别精度。

更改后的 AlexNet 模型虽然能实现手写数字识别,但是 Conv1 和 Conv2 使用的是  $5 \times 5$  的大卷积核,相对于  $28 \times 28$  分辨率的图像来说尺寸较大,不能提取到足够的细节特征。为了提取更多的细节特征,使用两个  $3 \times 3$  的卷积核堆叠,以获取和  $5 \times 5$  卷积核相同的感受野,同时能减少模型的参数量。将模型中前两层最大池化的池化核从  $3 \times 3$  修改为  $2 \times 2$ , 缩小特征图尺寸的同时,进一步地减少了模型的参数量。

但是 AlexNet 的 Conv3、Conv4 和 Conv5 这三个卷积层的特征图尺寸都为  $3 \times 3$ , 特征图尺寸过小,通过卷积层的堆叠并不能提取到足够的特征信息,反而会浪费计算资源。为了获得更加丰富的特征信息,使用改进后的 Inception-resnet 结构替换

原模型中的 Conv3 和 Conv4,使模型能学习到不同尺寸的特征信息,提升模型的特征提取能力,有助于分类器的分类。改进后的 Inception-resnet 结构如图 2 所示,在模块中每层卷积层之后添加 BN 层,将数据规范化处理,增快网络的收敛速度。

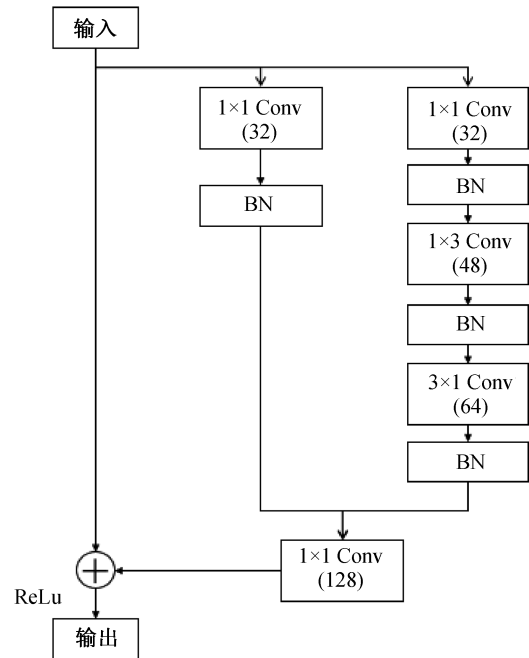


图 2 添加的 Inception-resnet 模块

Fig. 2 Inception-resnet module added

在 AlexNet 模型中使用了两个隐藏全连接层,由于全连接层的参数量巨大,因此删除一个全连接层,降低模型的参数量。为了使模型能够快速收敛,在每一个卷积层之后,激活函数之前使用 BN 对数据进行规范化处理,加速网络的训练。改进后模型的具体参数说明如表 1 所示。

表 1 改进模型具体参数

Tab. 1 Specific parameters of the improved model

结构名称	核大小	步长	填充方式	输出尺寸
Conv1-1	(3,3)	1	same	28×28×32
Conv1-2	(3,3)			
Maxpooling1	(2,2)	2	valid	14×14×32
Conv2-1	(3,3)	1	same	14×14×128
Conv2-2	(3,3)			
Maxpooling2	(2,2)	2	valid	7×7×128
Mixed	(1,1)/(1,3)/(3,1)	1	same	7×7×128
Maxpooling3	(3,3)	2	valid	3×3×128
Conv3	(3,3)	1	valid	3×3×256
Maxpooling3	(3,3)	2	valid	1×1×256
Fc	—	—	—	1 024
Softmax	—	—	—	10

### 3 实验与分析

#### 3.1 实验平台

本实验平台为戴尔笔记本,处理器为 Intel(R) Core(TM) i5,4 GB 内存,主频:1.7 GHz。运行环境为 Win10 64 操作系统,编译器为 Pycharm,编程语言为 Python。

#### 3.2 实验数据集

MNIST 数据集是一个手写体数字识别数据集,包含 50 000 条训练数据集和 10 000 条测试数据集。数据集包含 250 个不同人手写的阿拉伯数字 0~9,共 10 类手写体数字图像。每条数据集由图像和标签组成,每张图像的大小为  $28 \times 28 = 784$  的一维数组,标签类型为 One-Hot-Encoding 编码格式。部分数据集图像如图 3 所示。

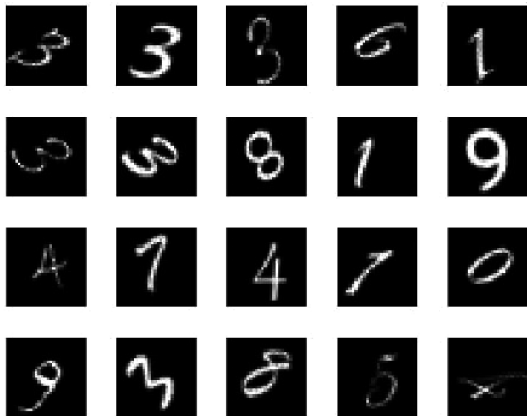
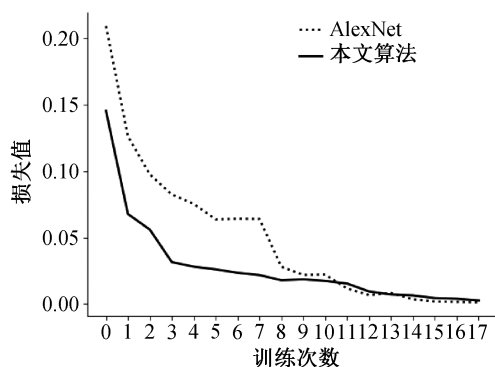


图 3 MNIST 数据集部分图像

Fig. 3 Partial images of the MNIST data set

#### 3.3 实验结果

本文以 AlexNet 模型为基础融合 Inception-ResNet-C 结构进行算法的改进,在 MNIST 数据集



(a) 训练损失函数曲线图

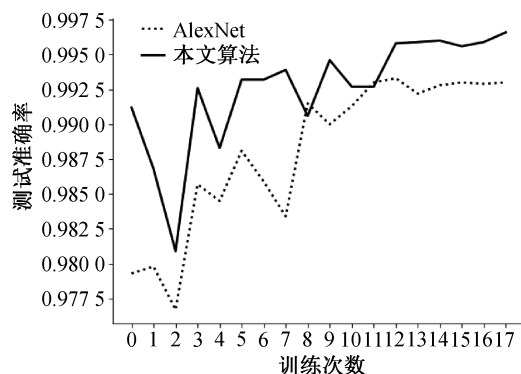
上进行验证,设置训练的 mini-batch 为 32,训练周期为 30 次,每次的训练样本数为 1 875。使用 Adam 优化器对梯度反向传播进行优化,初始学习率设置为 0.001,经过 2 个世代 val-Loss 不减少将学习率下调为原来的一半,加入 Early Stopping 机制,当 val-loss 经过 6 个世代不再减少时停止网络的训练。图 4(a)为 AlexNet 网络与本文算法在 MNIST 数据集上的训练损失函数曲线图,可以看出本文算法在训练过程中的收敛速度快于 AlexNet 网络,说明改进后的算法提高了模型的训练速度。图 4(b)为两种算法的测试准确率曲线图,可以看出改进后的算法在测试时的识别率高于 AlexNet 模型,说明改进后的方法有效地提升了手写数字识别的精度。

本文方法是在 AlexNet 基础上进行的改进,其参数量从  $2.1 \times 10^7$  减少到  $8 \times 10^5$  为原来的 1/26,模型训练一个世代的时间从 1 400 s 减少到 280 s,为原来的 1/5,大大加快了训练的速度,并且比 AlexNet 的准确率提高了 0.33%。为了验证本文算法的有效性,与其他算法进行比较,结果如表 2 所示。实验结果表明改进后的算法有效地提高了网络的检测精度。

表 2 不同算法的比较

Tab. 2 Comparison of different algorithms

方法	MNIST
文献[1]	0.993 6
文献[12]	0.989 0
文献[13]	0.992 2
文献[14]	0.993 4
文献[15]	0.993 7
LeNet5	0.992 2
AlexNet	0.993 3
本文方法	0.996 6



(b) 测试准确率曲线

图 4 模型训练与测试效果对比图

Fig. 4 Comparison of model training and test results

## 4 结论

本文提出基于改进 AlexNet 卷积神经网络的手写体数字识别方法,利用 Inception-resnet 模块提取多层特征并融合,提高了模型特征的学习能力;通过 BN 层对输入数据进行批归一化处理,有效地提升了模型的泛化能力;减少模型卷积核的数量,大大减少了网络的参数量,提升了模型的训练速度。利用 MNIST 手写体数字集进行实验,本文算法的检测精度达到了 99.66%,相比于 AlexNet 模型提升了 0.33%,与 LeNet5 模型相比提升了 0.44%,证明了本文算法的有效性。

### 参考文献:

- [1] 王梅,李东旭.基于改进 VGG-16 和朴素贝叶斯的手写数字识别[J].现代电子技术,2020,43(12):176-181+186.
- [2] 张成,戴俊峰,熊闻心.融合 LeNet-5 改进的扫描文档手写日期识别[J].计算机工程与应用,2021,57(9):207-211.
- [3] 杨旭,尚振宏.基于改进 AlexNet 的人脸表情识别[J].激光与光电子学进展,2020,57(14):243-250.
- [4] 黄健,张钢.深度卷积神经网络的目标检测算法综述[J].计算机工程与应用,2020,56(17):12-23.
- [5] 汪雅琴,夏春蕾,戴曙光.基于 LeNet-5 模型的手写数字识别优化方法[J].计算机与数字工程,2019,47(12):3177-3181.
- [6] 邓长银,张杰.基于改进 LeNet-5 模型的手写数字识别[J].信息通信,2018(1):109-112.
- [7] 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):47-51.
- [8] 陈玄,朱荣,王中元.基于融合卷积神经网络模型的手写数字识别[J].计算机工程,2017,43(11):187-192.
- [9] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet Classification with Deep Convolutional Neural Network [J]. Advances in neural information processing systems, 2012, 25(2):1097-1105.
- [10] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-V4, Inception-ResNet and the Impact of Residual Connections on Learning [C]//Thirty-first AAAI Conference on Artificial Intelligence. 2016:4278-4284.
- [11] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016:770-778.
- [12] 林仁耀,邓浩伟,兰红.卷积神经网络结合 SVM 的手写数字识别算法[J].通信技术,2019,52(10):2389-2394.
- [13] 胡貌男,邱康,谢本亮.基于改进卷积神经网络的图像分类方法[J].通信技术,2018,51(11):2594-2600.
- [14] 王建仁,马鑫,段刚龙,等.边缘智能背景下的手写数字识别[J].计算机应用,2019,39(12):3548-3555.
- [15] 侯杰,倪建成.基于 GoogLeNet 的手写体汉字识别[J].通信技术,2020,53(05):1127-1132.

(责任编辑 周雪梅)