

文章编号: 2096-1618(2019)01-0039-05

# 一种改进型卷积神经网络的图像分类方法

张 斌<sup>1</sup>, 王 强<sup>2</sup>

(1. 成都信息工程大学计算机学院, 四川 成都 610225; 2. 国网四川省电力公司德阳供电公司, 四川 德阳 618000)

**摘要:**基于 Keras 深度学习框架和卷积层取反操作, 提出一种改进型的卷积神经网络结构, 网络结构首层采用卷积层取反以增加有效特征信息的传递, 有效结合 Leaky ReLU 激活函数传递至下一层, 最后采用 Softmax 分类器实现图像分类。在两个公共数据集上, 同传统的卷积神经网络模型做对比实验, 实验结果表明, 改进的卷积网络模型是有效的。

**关 键 词:**深度学习; 卷积神经网络; 图像分类; 激活函数

**中图分类号:**TP301.6

**文献标志码:**A

**doi:**10.16836/j.cnki.jcuit.2019.01.009

## 0 引言

机器学习是一门交叉学科, 主要用来研究如何通过机器学习的算法来模仿或实现人类学习行为, 进而获取新的知识和技能的学科。机器学习系统需要通过大量的数据训练, 挖掘出大数据的价值。大数据的主要特征表现为: 数据量大、类型繁多、价值密度低、速度快并且时效高。随着互联网的发展, 目前已经全面进入大数据时代, 通过机器学习系统能够充分的挖掘大数据的价值, 并且在相关领域具有广泛的应用, 例如: 推荐系统<sup>[1]</sup>、欺诈检测、股票交易、机器视觉等。在机器视觉中图像识别<sup>[2]</sup>是一个主要的应用, 通过机器学习系统的感知过程来完成图像识别。其中, 特征提取<sup>[3]</sup>是最关键的部分, 良好的特征提取和表达, 对于算法的准确性起决定性的作用。

传统机器学习方法的特征提取部分通常都是靠人工选取。但人工选取特征存在许多缺点, 例如: 针对不同的数据, 需要使用不同的提取方法, 比较费时费力; 选取出较好的特征信息, 很大程度上是依靠经验和运气等。深度学习<sup>[4]</sup>作为机器学习的一个分支, 同传统的机器学习方法相比最大的优势就是不需要通过人工选取特征, 可以依靠多层网络结构对输入的数据进行层次化提取特征, 自动学习数据的特征。目前, 深度学习具有比较广泛的应用, 主要应用于语音识别<sup>[5]</sup>、图像分类<sup>[6]</sup>、目标检测<sup>[7]</sup>等领域。

在图像分类中, 图像特征的表达具有层次性, 同深度学习通过多层次的特征表达完美结合, 因此, 采用深度学习的方法进行图像分类具有较高的图像分类正确率。目前, 深度学习中广泛使用的网络模型有 3 种: 卷积神经网络(CNN)模型<sup>[8]</sup>、受限莫尔兹曼机器模型<sup>[9]</sup>

和自动编码器模型<sup>[10]</sup>。其中卷积神经网络不需要将特征提取和分类训练两个过程分开, 在训练的时候就自动提取的最有效的特征, 具有强大的特征提取能力。同时 CNN 训练的模型对缩放, 平移, 旋转等畸变具有不变性, 有着很强的泛化性。CNN 最大的特点在于卷积的权值共享结构, 可以大幅减少神经网络的参数量, 防止过拟合的同时又降低了神经网络模型的复杂度。因此文中主要研究深度学习模型中的卷积神经网络在自然图像分类中的应用。

## 1 卷积神经网络

卷积神经网络(CNN)是一种前馈的神经网络, 传统的卷积神经网络结构如图 1 所示, CNN 的适应性很强, 其网络结构与生物神经网络相似, 是通过一层一层的节点组织起来的, 其中每一个节点都是一个神经元。卷积神经网络可以利用空间结构关系减少需要学习的参数量, 从而提高反向传播算法的训练效率。

传统的卷积神经网络结构是由多个隐藏层交替组成的, 其中隐层主要由卷积层、激活函数和池采样层组成<sup>[11]</sup>。在卷积神经网络中, 第一个卷积层会直接接收图像像素级的输入, 每一个卷积操作只处理一小块图像, 进行卷积变化后再传到后面的网络, 每一个卷积操作只处理一小块图像, 进行卷积变化后再传到后面的网络, 每一层卷积都会提取数据中最有效的特征。对于激活函数的选取, 其中常见的激活函数有 Sigmoid, Tanh, ReLU 3 种。池采样层主要包含: 最大池化、均值池化、随机池化 3 种。隐含层中的卷积层和池采样层是实现卷积神经网络特征提取功能的关键, 通过一系列隐含层的特征提取, 最后通过全连接层拉成一维的特征向量, 作为分类器的输入, 最终通过分类器进行分类。

由图 1 可知, 首先卷积神经网络输入一个图像矩

阵  $A$ , 通过若干个的卷积核  $K_i^1 (i=1, 2, \dots, n)$  的卷积运算生成若干个特征图计算公式:

$$a_i^1 = f(u_i^1) = f(\text{conv2}(A, K_i^1, 'valid') + b_i^1) \quad (1)$$

其中,  $\text{conv2}(\cdot)$  函数表示输入图像矩阵  $A$  同卷积核  $K_i^1$  做相应的卷积运算;  $f(\cdot)$  函数表示激活函数,  $b_i^1$  表示偏置。

紧接着进行池采样层, 自定义一个池化窗口大小, 选择一种池化方式将所有的特征图进行池化:

$$a_i^2 = f(u_i^2) = f(\beta_i^2 \text{down}(a_i^1) + b_i^2) \quad (2)$$

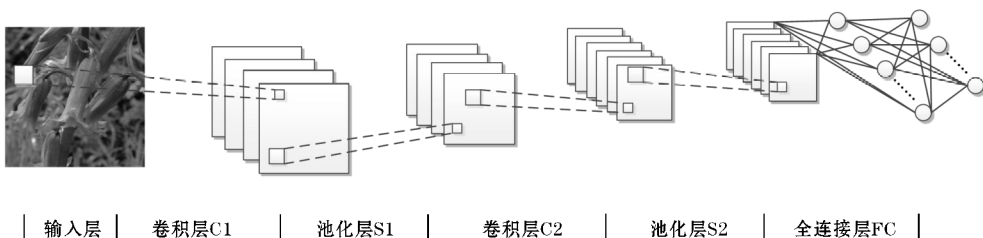


图1 传统的卷积神经网络结构

## 2 改进模型设计

通过以上对传统的卷积神经网络结构的介绍, 知道隐含层中的卷积层是提取图像特征信息的关键, 因此通过增加有效特征信息的提取将影响图像的分类正确率。文中基于 Keras 框架, 针对首层卷积层离输入层最近, 保留更多的有效特征信息将影响图像的分类效果。

如图2所示, 激活函数 ReLU 和 Leaky ReLU 的函数图像的比较, 通过函数图像可以发现当输入负值特征信息时, ReLU 激活函数将置为零即丢弃所有的负值特征信息, 而 Leaky ReLU 将以一个较小斜率的线性函数取值即保留了负值特征信息的向前传递。

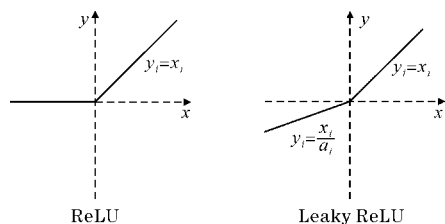


图2 ReLU 和 Leaky ReLU 激活函数

提出一种改进的模型设计, 主要是将首层卷积层进行改进, 并有效的结合 Leaky ReLU 激活函数, 增加向前传播的有效特征信息, 以有利于图像分类。传统的首层卷积层和改进后的首层卷积层对比图如图3所示。

由图3可知, 改进后的第一层卷积层将卷积特征信息进行取反操作, 并与原始特征信息一起作用于 Leaky ReLU 激活函数, 将其转移到下一层。这种改进的方法能够有效保留并增加负值特征信息的正向传递, 从而影响图像分类的效果。因为第一个卷积层会

其中,  $\beta_i^2$  表示连接的权重;  $\text{down}(\cdot)$  表示下采样函数。

接着再次进行卷积:

$$a_i^3 = f(u_i^3) = f(\sum_{j=1}^{F_1} \text{conv2}(a_j^2, k_{ij}^3, 'valid') + b_{ij}^3) \quad (3)$$

然后再次进行池采样层

$$a_i^4 = f(u_i^4) = f(\beta_i^4 \text{down}(a_i^3) + b_i^4) \quad (4)$$

最后, 将特征向量连接成一个长向量, 作为全连接层的输入。

直接接受图像像素级的输入, 图像通过多个不同的卷积核的滤波提取出局部特征, 随后的层结构都取决于第一层卷积核的滤波输出结果, 然后进行非线性的激活函数处理, 所以第一层卷积核的滤波输出结果将直接影响最终的图像分类效果。为保证更多的有利于图像分类的特征信息能够传递到下一层, 提出的改进网络模型结构使用了将第一层卷积层进行取反的操作方法。通过在两个公共数据集上实验, 验证了改进模型的有效性。

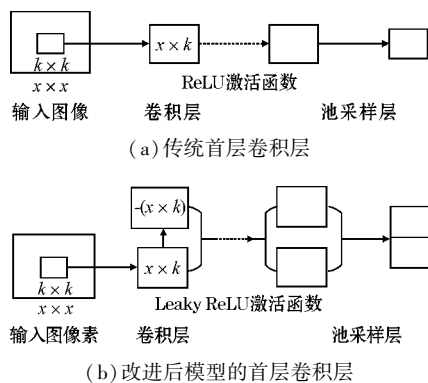


图3 传统首层卷积层和改进后模型的首层卷积层结构

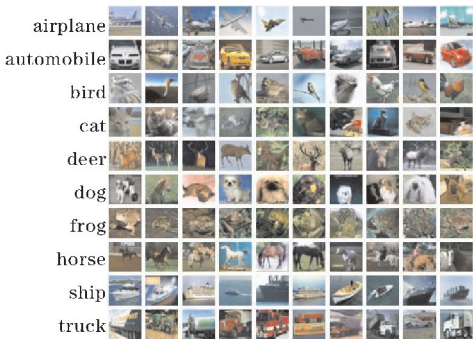
## 3 实验结果与分析

实验基于 Keras 和 Tensorflow<sup>[12]</sup> 来构建和训练所提出的 CNN 模型, 且卷积网络模型的训练使用了 GPU 并行处理。所有的实验都是在配备 Intel(R) Core i3-6100 CPU @ 3.70GHz 和 NVIDIA Tesla M40 GPU 的 PC 上运行的 Ubuntu 16.04 操作系统中进行的。

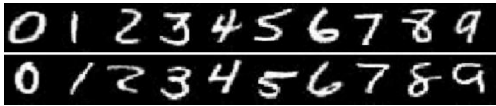
为验证 Improved CNN 模型的有效性, 将它同传统的 CNN 模型进行对比实验, 所有对比实验在同一个数

据集中都使用相同数量的卷积层和池采样层,并且每一层的参数设置也相同,实验结果验证了所提网络结构模型的有效性。

实验分别使用经典的 CIFAR-10 数据集和 MNIST 手写数字字符数据库进行对比。CIFAR-10 数据集由 10 个类的 60000 张彩色图像组成,每个类有 6000 个图像。其中 50000 张图像作为训练样本集,10000 张图像作为测试样本集。MNIST 手写数字字符数据库主要是由数字 0 ~ 9 的 60000 张训练集样本和 10000 张测试集样本组成。实验从字符数据库中随机抽取其中 6000 张图像作为训练样本集,另取 1000 张作为测试样本集,总共 6000 张图像,每张图片大小均为。两个数据集的部分样本示例如图 4 所示。



(a) CIFAR-10 部分图像



(b) MNIST 部分图像

图4  CIFAR-10 数据集与 MNIST 数据集

实验采用的 CNN 架构,包括输入层,3 个卷积层和 3 个池化层。除了在第一个卷积层后使用系数  $a = 1/3$  的 Leaky ReLu 激活函数外,在随后的每一个卷积层及全连接层后都使用一个 ReLU 激活函数,目的是将前面卷积核的滤波输出结果,进行非线性的激活函数处理。ReLU 激活函数<sup>[8]</sup>公式为

$$f(x)=\max(0,x)$$

(5)

当输入特征值输出为零时,训练后为零的神经元越多,它们越稀疏,提取的特征越具有代表性,泛化能力越强;当输入特征值输出等于输入时,无梯度耗散问题,收敛快。池化层作用是对激活函数的结果再进行池化操作,即降采样;使用最大池化方法以保留最显著的特征。针对 CIFAR-10 和 MNIST 两个不同的数据集,分别使用 1 个和 2 个全连接层,并且最后一个全连接层利用 Soft-max 回归模型进行分类,最终得到图像的分类结果。在训练过程中,选择 Adam optimizer<sup>[13]</sup>来优化网络损耗,2 个数据集的 Epoch 分别设置为 50 和 40,一次处理图片的最小输入的数量 (BATCH\_SIZE) 为 32。各层具体的参数设置如表 1、表 2 所示。

表 1  CIFAR-10 数据集 CNN 结构及各层参数设置

层号	类别	参数
Input	输入层	图像大小 32×32×3
Conv1	卷积层	滤波数 32
		卷积核 5×5
		步长 1
Pool2	池化层	激活函数 LeakyReLU
		pool_size 3×3
		步长 2
Conv3	卷积层	池化方法 max-pooling
		滤波数 32
		卷积核 5×5
Pool4	池化层	步长 1
		激活函数 ReLU
		pool_size 3×3
Conv5	卷积层	步长 2
		池化方法 max-pooling
		滤波数 64
Pool6	池化层	卷积核 5×5
		步长 1
		激活函数 ReLU
FC7	全连接层	pool_size 3×3
		步长 2
		池化方法 max-pooling
Output	输出层	神经元个数 64
		激活函数 ReLU
		神经元个数 10

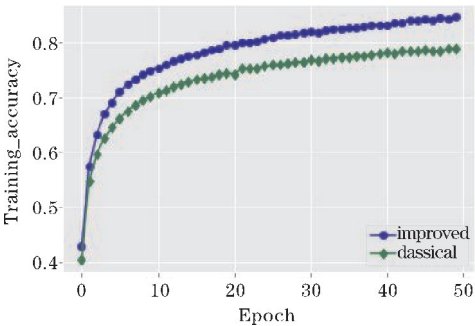
表 2  MNIST 数据集 CNN 结构及各层参数设置

层号	类别	参数
Input	输入层	图像大小 28×28×1
Conv1	卷积层	滤波数 32
		卷积核 3×3
		步长 1
Pool2	池化层	激活函数 LeakyReLU
		pool_size 2×2
		步长 2
Conv3	卷积层	池化方法 max-pooling
		滤波数 32
		卷积核 3×3
Pool4	池化层	步长 1
		激活函数 ReLU
		pool_size 2×2
Conv5	卷积层	步长 2
		池化方法 max-pooling
		滤波数 32
Pool6	池化层	卷积核 2×2
		步长 1
		激活函数 ReLU
FC7	全连接层	pool_size 3×3
		步长 2
		池化方法 max-pooling
FC8	全连接层	神经元个数 64
		激活函数 ReLU
		神经元个数 64
Output	输出层	激活函数 ReLU
		神经元个数 10
		神经元个数 10

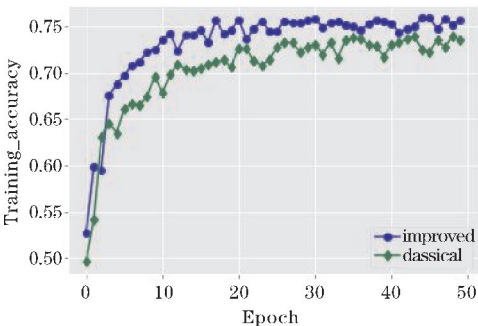


两个网络模型( Classical CNN, Improved CNN)在不同的数据集上分别经过 50 个和 40 个 Epoch 训练后,均已收敛,两个网络模型在不同数据集上的训练过程如图 5 和图 6 所示。如图 6 所示,在 CIFAR-10 数

据集中,CNN 无论在训练集的精度还是测试集上的精度,数值上都明显优于 Classical CNN,并且经过极少次数迭代后就收敛。

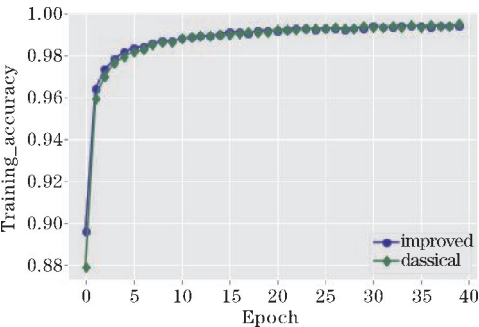


(a)代表训练精度曲线

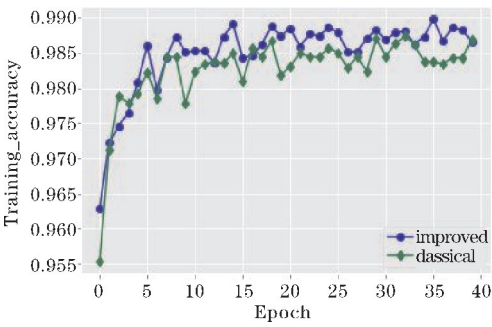


(b)代表测试准确度曲线

图 5 CIFAR-10 数据集准确率变化曲线



(b)代表训练精度曲线



(b)代表测试准确度曲线

图 6 MNIST 数据集准确率变化曲线

两个模型在不同数据集中的准确率的对比结果如表 3 所示。在数据集 CIFAR-10 中,Improved CNN 在训练集上的准确率明显高于传统 Classical CNN。Improved CNN 最高准确率达75.90%,传统方法的最高准确率仅为73.97%。另外,对模型的平均准确率也做了比较。Classical CNN 的平均准确率大概为70.37%,而文中提出的 Improved CNN 的在测试集上的平均准确率则达73.28%,相比 Classical CNN 提升

了2.91%。在数据集 MNIST 中,Improved CNN 在训练集上的准确率明显高于传统 Classical CNN。Improved CNN 最高准确率达98.98%,传统方法的最高准确率仅为98.74%。另外,对模型的平均准确率也做了比较。Classical CNN 的平均准确率大概为98.26%,而文中提出的 Improved CNN 的在测试集上的平均准确率则达98.50%,相比 Classical CNN 提升了0.24%。

表 3 2 个模型在不同数据集中的准确率的对比结果 %

数据集	模型结构	训练集准确率	测试集平均准确率	测试集最高准确率
CIFAR-10	Classical +ReLU	71.80	70.37	73.97
	Improved CNN	76.65	73.28	75.90
MNSIT	Classical +ReLU	98.63	98.26	98.74
	Improved CNN	98.71	98.50	98.98

4 结束语

提出一种基于 Keras 深度学习框架的改进型卷积神经网络模型结构 Improved CNN。同传统的 CNN 模型结构相比,提出的网络结构首层采用卷积层取反以

增加有效特征信息的传递,有效结合 Leaky ReLU 激活函数传递至下一层,最终有利于提高图像分类效果。在两个公共的数据集 CIFAR-10 和 MNIST 上验证了 Improved CNN 模型的图像分类效果,其实验结果均高于传统的 CNN 网络。实验结果证明改进的卷积网络模型是有效的。

## 参考文献:

- [1] Oh J, Lee S, Lee E. A User Modeling Using Implicit Feedback for Effective Recommender System [C]. International Conference on Convergence and Hybrid Information Technology. IEEE, 2008: 55–158.
- [2] Boureau Y L, Roux N L, Bach F, et al. Ask the locals: Multi-way local pooling for image recognition [J]. IEEE 2011 IEEE International Conference on Computer Vision (ICCV), 2011, 58(11): 2651–2658.
- [3] 郎波, 黄静, 危辉. 利用多层视觉网络模型进行图像局部特征表征的方法 [J]. 计算机辅助设计与图形学学报, 2015, 27(4): 703–712.
- [4] Le Cun, Y Bengio, G E Hinton, Deep learning [J]. Nature, 2015, 521(7553): 436–444.
- [5] Hinton G, Deng L, Dong Y, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups [J]. IEEE Signal Processing Magazine, 2012, 29(6): 82–97.
- [6] 吕刚, 郝平, 盛建荣. 一种改进的深度神经网络在小图像分类中的应用研究 [J]. 计算机应用与软件, 2014, 31(4): 182–184.
- [7] Ren S, He K, Grishick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137–1149.
- [8] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C]. International Conference on Neural Information Processing Systems, 2012: 1097–1105.
- [9] Hinton G E. A practical guide to training restricted boltzmann machines [M]. Springer Berlin Heidelberg, 2012: 599–619.
- [10] Zhang J, Shan S, Kan M, et al. Coarse-to-fine auto-encoder networks for real-time face alignment [C]. European Conference on Computer Vision, 2014: 1–16.
- [11] 王强, 李孝杰, 陈俊. Supplement 卷积神经网络的图像分类方法 [J]. 计算机辅助设计与图形学学报, 2018(3).
- [12] Abadi M, Agarwal A, Barham P, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems [J]. arXiv preprint arXiv, 2016.
- [13] Kingma D P, Ba J. Adam: A Method for Stochastic Optimization [J]. Computer Science, 2014.

## An Improved Convolution Neural Network Image Classification Method

ZHANG Bin<sup>1</sup>, WANG Qiang<sup>2</sup>

(1. College of Computer Science, Chengdu University of Information Technology, Chengdu 610225, China; 2. State Grid Sichuan Electric Power Company Deyang Power Supply Company, Deyang 618000, China)

**Abstract:** Based on Keras deep learning framework and convolution layer inverse operation, an improved Convolutional Neural Network structure is proposed in this paper. The first layer of the network structure uses convolutional layer inversion to increase the transmission of effective feature information. The LeakyReLU activation function is effectively combined to the next layer. Finally, the Softmax classifier is used to implement image classification. Compared with the traditional Convolution Neural Network model on two common datasets, the experimental results show that the improved convolution network model in this paper is effective.

**Keywords:** deep learning; Convolutional Neural Network; image classification; activation function