

基于卷积神经网络优化因素的图像分类研究

李尤丰

(金陵科技学院软件工程学院, 江苏 南京 211169)

摘要:为提高卷积神经网络的图像分类性能,通过图像数据集 CIFAR-10 的数据实验,验证了数据集、隐藏神经元数目和训练方法等优化因素对卷积神经网络的分类准确度、运行时间和网络参数数量的影响,从而为有效提高卷积神经网络图像分类性能的研究提供了参考。

关键词:卷积神经网络;网络优化;图像分类;CIFAR-10

中图分类号:O29;TP301 文献标识码:A 文章编号:1672-755X(2022)01-0026-06

Research on Image Classification Based on Optimization Factors in Convolutional Neural Network

LI You-feng

(Jinling Institute of Technology, Nanjing 211169, China)

Abstract: In order to improve the image classification performance of convolutional neural network, through the data experiment on image data set CIFAR-10, the effects of optimization factors such as data set, number of hidden neurons and training method on the classification accuracy, running time and network parameters of convolutional neural network are verified. Thus, it provides a reference for effectively improving the image classification performance of convolutional neural network.

Key words: convolutional neural network; network optimization; image classification; CIFAR-10

卷积神经网络(convolutional neural network, CNN)通过多个不同层次的感受野(receptive field)将视觉信息从视网膜传递到大脑,是一种通过模拟实验来实现视觉系统功能的生物策略算法。1998年卷积神经网络模型 LeNet-5^[1]在识别支票手写数字方面得到了成功应用,引起了研究者们对卷积神经网络图像分类性能的广泛关注。然而由于 LeNet-5 的参数量大、网络开销大、泛化能力差,使得卷积神经网络模型对图形分类的应用停滞不前。直到 2012 年,针对 LeNet-5 的缺陷 Krizhevsky 等^[2]提出了 AlexNet 模型,并在大型图像数据库 ImageNet^[3]的分类竞赛中夺得冠军,其分类准确度超过第二名 11%。这一成功使得卷积神经网络的图像分类应用成为关注的新焦点。同年,用于图像分类的新卷积神经网络模型如 VGG^[4]、GoogLeNet^[5]、NIN^[6]等被相继提出。随着 ResNet^[7]的残差学习能力被广泛认可,图像分类卷积神经网络模型如 RiR^[8]、Dense ResNet^[9]、Wide ResNet^[10]、Stochastic Depth^[11]、ResNeXt^[12]、Inception-ResNet^[13-14]等如雨后春笋般地应运而生。

收稿日期:2021-12-01

基金项目:金陵科技学院软件工程重点学科资助项目;金陵科技学院软件工程国家一流专业建设项目;金陵科技学院校级教育教改课题研究成果(JYJG202106)

作者简介:李尤丰(1977—),女,江苏宜兴人,副教授,硕士,主要从事图像分类等方面的研究。

随着卷积神经网络模型的迅猛发展,研究者们对已有模型的缺陷进行了相应改进,从而使得图像分类性能得到了提高。AlexNet 用两块 GPU 改进网络开销,用随机失活技术改进网络适应性;VGG 用小卷积核改进参数量大的问题,加深了网络深度;GoogLeNet 用 Inception 模块改进参数量和网络开销,大大降低了深度卷积神经网络的复杂度;ResNet 用恒等映射进行残差学习改善深度网络的退化。由此看出,要想针对网络的某种缺陷找到相应的改进方法,就需要研究卷积神经网络的哪些优化因素对网络的哪部分性能有影响。本文以最基本的卷积神经网络模型为例,用数据实验验证卷积神经网络模型的优化因素对模型图像分类性能的影响。

1 传统 CNN 的基本结构

传统 CNN 的基本结构分为输入层、卷积层、池化层、全连接层和输出层等,每一层的功能与作用不同,因此 CNN 模型是一种层网络模型。CNN 的工作过程大致为:把图像信息作为输入,与卷积层的多个卷积核进行卷积计算并加上偏置,得到相应的一系列特征面;然后将该结果传递给非线性激活函数,通过与激活函数的阈值比较得到卷积层的输出;最后传递给池化层进行池化操作,得到输入图像显著特征的简化模型。经过多个卷积层和池化层后,输入图像的特征逐步从低层的具体特征转化为高层的抽象特征,再经过全连接层的映射和输出层的分类运算得到输出向量。输出层的向量并不是一个确切的类别向量,而是一个特征分类概率向量,用于判断输入图像的分类。

2 实验设计

模型对数据集的拟合程度是评价 CNN 图像分类性能的一个重要指标,具体表现在训练集的拟合准确度 $fit(X)$ 与测试集的预测准确度 $Acc(X)$ 上。 $fit(X)$ 越高而 $Acc(X)$ 越低,说明网络存在过拟合,分类方差较大; $fit(X)$ 越低,说明网络欠拟合越严重。为找到 $fit(X)$ 、 $Acc(X)$ 与卷积神经网络模型的优化因素之间的关系,设计如下实验:

1) 改变数据集。在图像问题的分类中,有时可获取的有效数据量往往有限,比如医学图像数据。为测试图像数据量对卷积神经网络模型的分效率 ($fit(X)$ 、 $Acc(X)$) 影响,从比较同一数据集数据量改变前后的网络分类准确度来看,当数据量有限时,改变数据量的一种常用方式就是增强标准数据。它利用已有的有限图像数据,对数据进行拉伸、翻折、裁剪等预处理,提高数据集的规模和多样性,进而提高训练数据的效率。

2) 改变隐藏神经元数目。为测试卷积神经网络模型的规模对模型分类的准确度影响,在数据集不变的情况下,比较隐藏神经元数目在改变前后,模型分类准确度的变化。隐藏神经元数目的改变方式分为两种:①改变每层的隐藏神经元数目,即改变网络的宽度。表 1 给出了 4 种不同层规模的卷积神经网络模型,每一层的隐藏神经元数目逐步增加,但网络层数不变,以便于对比不同层规模对卷积神经网络模型分类准确度的影响。②改变隐藏神经元所属的层数,即改变网络的深度。在上述 4 种不同层规模的卷积神经网络模型上分别增加卷积层进行实验,具体网络结构和配置如表 2 所示。其中“层模型 1+”表示在表 1 的“层规模 1”的网络上增加了卷积层,其他依此类推。

表 1 卷积神经网络的 4 种不同层规模

层名称	输入规格	输出规格	层规模 1	层规模 2	层规模 3	层规模 4
卷积层	24×24	24×24	16	32	64	128
池化层	24×24	12×12	16	32	64	128
全连接层	12×12	1×1	512	1 024	2 048	2 048
输出层	1×1	10×1	—	—	—	—

3) 改变训练方法。当数据集和网络规模确定时,考虑不同的优化训练方法对卷积神经网络模型图像分类准确度的影响。考虑自适应矩估计法、随机梯度下降法、自适应梯度算法、柯西—牛顿法这 4 种常用训练算法对分类准确度的影响。

表 2 增加网络深度后的不同层规模

层名称	层模型 1+	层模型 2+	层模型 3+	层模型 4+
卷积层 1	3×3×16	3×3×32	3×3×64	3×3×128
	3×3×16	—	3×3×64	3×3×128
池化层 1	最大池化	最大池化	最大池化	最大池化
卷积层 2	—	3×3×64	3×3×128	3×3×256
	—	—	—	3×3×256
池化层 2	—	平均池化	平均池化	平均池化
全连接层	512	1 024	2 048	2 048
输出层	10	10	10	10

3 实验数据设置

3.1 数据集设置

由于图像数据集 CIFAR-10 相对比较简单,常被用于很多图像分类卷积神经网络模型实验,且比较契合于实验选用的传统 CNN 模型,因此被选作实验数据集。CIFAR-10 共有 60 000 张彩色图像,共包含 10 个类别,每个类别 6 000 张;用于训练集的图像 50 000 张,构成 5 个训练批,每个训练批有 10 000 张,剩下的 10 000 张图像用于测试批。训练批和测试批的组成方式为:先在每一类中随机抽取 1 000 张图像构成测试批,然后把剩下的 50 000 张图像随机分成 5 个训练批。由于训练批中的图像是随机划分的,因此其中的图像各自独立、不会重叠,使得每个训练批中的各类图像数量并不一定相同,每次实验时训练批和测试批的图像也很有可能不同。

3.2 实验环境设置

实验的学习框架为 TensorFlow^[15],并用 cuDNN^[16]对学习框架进行优化。单块 GPU 的型号为 NVIDIA Tesla K40C,使用的软件为 MATLAB R2014a,实验的训练算法为随机梯度下降法,动量系数为 0.9,采用数据批为 32 的小批量数据输入方式训练 10 000 次。模型的初始学习率为 0.01,当验证误差不再提高时,新学习率衰减 1/10,共衰减 3 次。权重初始化采用文献[17]的方法,权重衰减系数为 0.000 1,每层权重的学习率相同。为提高实验数据的稳定性,实验过程重复 10 次,并取最优值作为最终的分​​类准确度。

4 实验结果与分析

4.1 数据集对分类准确度的影响

数据增强前后的实验结果如图 1 所示。由图 1 可以看出:1)当数据经过标准数据增强后,整个测试过程的收敛速度变得很慢,但是测试准确度 $Acc(X)$ 有较大幅度的提高,这说明经过标准数据增强后,训练集的数据量和多样性都得到了显著提升,训练所得模型的泛化能力得到了提高。2)训练集和测试集分类准确度的差异 $fit(X) - Acc(X)$ 大大缩小,基本实现了两者同步提升,这表明有效的数据处理可以很好地解决模型的过拟合问题。3)经过标准数据增强后,图像数据更加多样化,模型看不到完全相同的图像,有效抑制了模型的过拟合,有效提升了卷积神经网络模型的分​​类准确度。因此,适当增加数据集的数据量和多样性,能有效提高模型的泛化能力。

4.2 隐藏神经元数目对分类准确度的影响

1)改变隐藏神经元的层规模时,实验结果如图 2 所示。在迭代次数 2 000~5 000 次范围内,模型在不同层规模间测试集的分​​类准确度差异较为明显,且随着层规模的增加,分​​类准确度也增加;但当迭代次数大于 5 000 次后,模型在不同层规模间测试集的分​​类准确度差异均越来越小。这是由于层规模增加时,需要训练的参数量大大增加,而参数量和计算量过于庞大会导致卷积神经网络模型在训练时给计算机硬件带来巨大负担,从而致使模型的学习训练效率降低,有时甚至无法实现整个卷积神经网络模型的学习训

练。因此在运行时间和计算量等允许的条件下,适当扩大层规模可以提高分类准确度;但当层规模达到一定程度后,增加层规模并未明显增加分类准确度的提高量,此时应考虑从其他方面改进模型的运行效率。

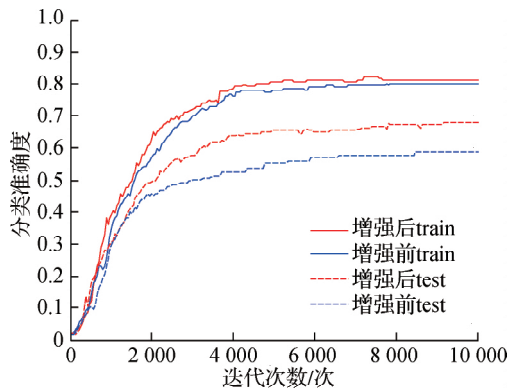


图 1 数据增强前后的实验结果

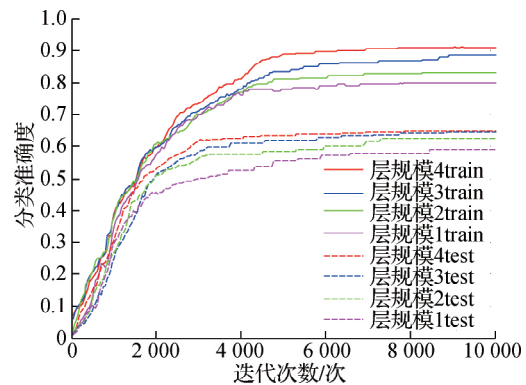


图 2 不同层规模的实验结果

2) 改变隐藏神经元所属的卷积层数时,实验结果如图 3 所示。卷积神经网络模型需迭代 6 000 次后才逐渐趋于稳定。究其原因,增加卷积层的同时增加了待训练的神经元数目和计算量,一定程度上增加了网络训练难度。在迭代次数 2 000~5 000 次内,层规模 4+测试集的分类准确度明显高于其他 3 种层规模,但当迭代次数大于 5 000 次后,层规模 4+测试集的分类准确度与其他 3 种层规模无明显差异。原因是增加卷积层数便增加了卷积神经网络模型的深度,较深的模型能够提取更抽象的图像特征,有助于模型理解更深层次的内容,在一定程度上能有效提高卷积神经网络模型的性能。但是当模型深度达到一定程度后,卷积神经网络模型的性能不再提高,甚至有所下降,这是由于模型训练的难度、分类效率提升的有限性、模型本身的局限性等因素影响了模型的性能。

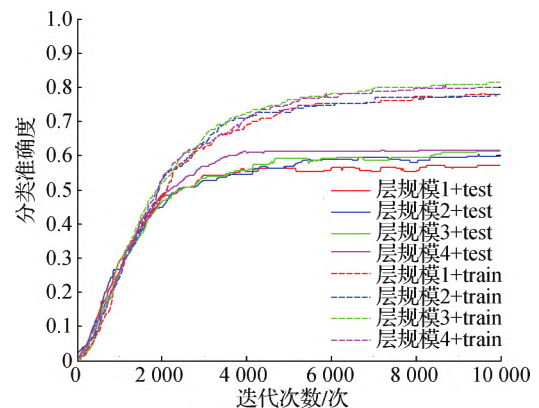


图 3 不同网络深度的实验结果

综合图 2、图 3 可以看出,在较小的迭代次数范围内,训练集和测试集的分类准确度曲线几乎重叠在一起,说明较小迭代次数范围内的卷积神经网络更易拟合。图 2 中训练曲线在不同层规模间的分类准确度差异较大,图 3 中训练曲线在不同层规模间的分类准确度差异较小,说明当层规模固定时增加一定的卷积层数有助于模型性能稳定。

3) 随着隐藏神经元数目的增多,模型的参数量、运行时间和分类准确度的变化趋势如图 4 所示。由图 4 可以看出,随着隐藏神经元数目的增加,模型的参数量、运行时间和分类准确度越来越大,当隐藏神经元数目达到一定规模时,参数量和运行时间陡增,而分类准确度越来越平缓,说明增加隐藏神经元数目获得的收益越来越少。这与 1) 和 2) 中层规模达到一定程度后分类准确度不再提高的结果一致。因此通过增加隐藏神经元数目获得更高的分类准确度是否合适,得看具体情况。

4.3 训练方法对分类准确度的影响

以均方误差为代价函数,通过常规正则化项构造目标损失函数,然后基于 4 种不同的梯度优化算法进行实验。为更好地比较训练方法,基于基本模型和层规模 4+基本模型进行实验,实验结果如图 5 所示。由图 5 可以看出,不同梯度优化算法的收敛速度存在明显差异。图 5(a)中,不同的训练算法很快收敛到平稳状态,图 5(b)中,不同训练算法收敛时均有明显振幅,说明模型复杂度提高时训练难度增加,这与前文隐藏神经元数目增加时模型训练难度增加的结论一致。综合图 5(a)和 5(b)可以看出,层规模 4+模型上自适应曲线仍有上升趋势,未明显平稳收敛,且分类准确度比基于基本模型的自适应曲线有所降低。原因是自适应算法涉及方差计算,在模型复杂时计算量明显增大,模型达到平稳状态的速度变慢,分类准确

度降低说明在简单的基本模型上自适应算法存在过拟合。4 种训练算法中,矩估计算法快速达到较高准确度,随机梯度算法收敛到较高准确度的速度较慢,柯西-牛顿算法的速度适中,自适应梯度算法表现良好,原因是自适应梯度算法具有一定的自我调整能力,对噪声、模型、数据和超参数表现出鲁棒性,适应能力较好。但是,自适应梯度算法的学习率依赖人工设置,特别在训练的后阶段,当约束项的分母越来越大时,梯度越来越小,且趋于零,很容易使训练提前结束。因此,要缩短训练时间,提高卷积神经网络模型训练的学习效率,需要根据具体情况选取合适的梯度下降算法。

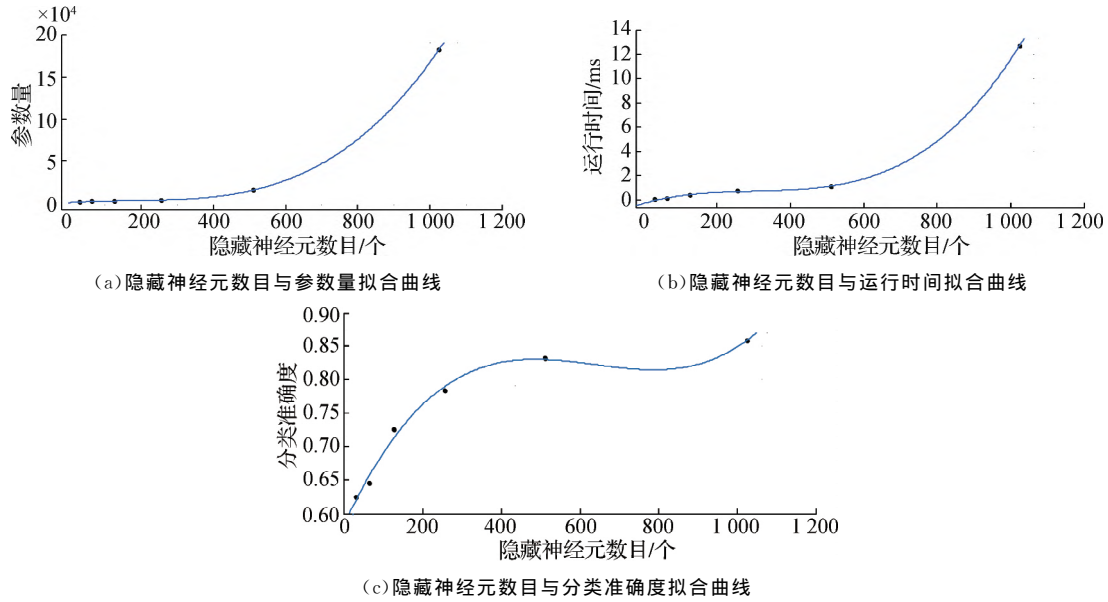


图 4 隐藏神经元数目对参数量、运行时间和分类准确度的影响

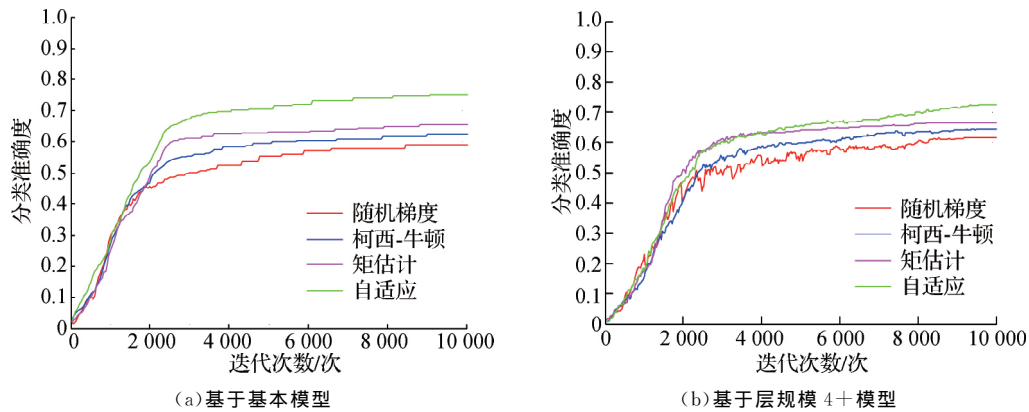


图 5 不同训练算法的实验结果

5 结 语

本文在数据集 CIFAR-10 上进行了一系列实验,实验结果表明,改变数据集、隐藏神经元数目和训练方法对卷积神经网络模型的图像分类准确度有一定的影响。增加数据量和数据多样性,不仅能提高数据利用效率,而且能提高卷积神经网络模型的图像分类准确度。当数据集固定时,通过增加隐藏神经元数目使网络宽度或者网络深度增加,也有助于提高卷积神经网络模型的图像分类准确度,当隐藏神经元数目增加到一定程度时,网络的参数量和计算量虽然迅速增加,但此时网络的增益相对较少。当数据集和网络规模固定时,采用适当的训练算法有助于提高网络模型的泛化能力,使训练准确度与测试准确度更接近,使模型运用于其他图像数据集的分类时具有较好的适应能力。

除了这几个因素外,改变模型的学习率、初始化权值等也对卷积神经网络模型分类准确度有一定的影响。应用时要多调试,尽量找到最适合的参数和方案。

参考文献:

- [1] LE C Y, BOTTOU I, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278–2324
- [2] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C]//Proceedings of Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2012: 1106–1114
- [3] DENG J, DONG W, SOCHER R, et al. ImageNet: a large-scale hierarchical image database[C]//Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2009: 248–255
- [4] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2016–04–05)[2021–08–15]. <https://blog.csdn.net/garfielder007/article/details/51029180>
- [5] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]//Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2015: 1–8
- [6] LIN M, CHEN Q, YAN S. Network in Network[J]. Computer Science, 2013, 12: 131–142
- [7] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// Proceedings of the 2016 IEEE International Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2016: 770–778
- [8] TARG S, ALMEIDA D, LYMAN K. Resnet in resnet: generalizing residual architectures[J]. Computer Science, 2016, 25: 16–29
- [9] HUANG G, LIU Z, LAURENS V, et al. Densely connected convolutional networks[EB/OL]. (2020–02–05)[2021–08–16]. https://blog.csdn.net/qq_38863413/article/details/104166211
- [10] ZAGORUYKO S, KOMODAKIS N. Wide residual networks[EB/OL]. (2017–11–08)[2021–08–16]. <https://blog.csdn.net/wangkun1340378/article/details/78474757>
- [11] HUANG G, SUN Y, LIU Z, et al. Deep networks with stochastic depth[EB/OL]. (2018–02–07)[2021–08–17]. <https://blog.csdn.net/u014451076/article/details/79283581>
- [12] XIE S, GIRSHICK R, DOLLÁR P, et al. Aggregated residual transformations for deep neural networks[EB/OL]. (2020–08–14)[2021–08–17]. https://blog.csdn.net/gongxifacai_believe/article/details/108002644
- [13] SZEGEDY C, IOFFE S, VANHOUCHE V, et al. Inception-v4, inception-ResNet and the impact of residual connections on learning[EB/OL]. (2019–10–31)[2021–08–18]. <https://blog.csdn.net/kxh123456/article/details/102828148>
- [14] SZEGEDY C, VANHOUCHE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2016: 2818–2826
- [15] ABADI M, BARHAM P, CHEN J, et al. TensorFlow: a system for large-scale machine learning[EB/OL]. (2019–10–09)[2021–08–18]. <https://blog.csdn.net/u011254180/article/details/82011756>
- [16] CHETLUR S, WOOLLEY C, VANDERMERSCH P, et al. CuDNN: efficient primitives for deep learning[EB/OL]. (2016–09–07)[2021–08–19]. <https://blog.csdn.net/u010620604/article/details/52464529>
- [17] HE K, ZHANG X, REN S, et al. Delving deep into rectifiers: surpassing human-level performance on imageNet classification[C]//Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2015: 1026–1034

(责任编辑:谭彩霞)