

引用格式: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. [RU X Q, HUA G G, LI L H. et al. Handwritten digital recognition based on deformable convolutional neural network [J]. Microelectronics & Computer,2019,36(4): 47-51.]

基于形变卷积神经网络的手写体数字识别研究

茹晓青,华国光,李丽宏,李 莉

(河北工程大学 信息与电气工程学院,河北 邯郸 056000)

摘 要: 本文引入形变卷积模块来增强网络对数字几何变换的建模能力,提出了一种基于改进的形变卷积神经网络手写体数字识别框架,在提高识别精度的同时,还有效的减少了训练的参数量,提高识别速度. 本文在手写体数据集及变换后的数据集中进行验证. 实验结果的分析以及与相应算法的比较,证明了本算法是有效的.

关键词: 手写体数字识别;卷积神经网络;形变卷积

中图分类号: TP391.4

文献标识码: A

文章编号: 1000-7180(2019)04-0047-05

Handwritten Digital Recognition Based on Deformable Convolutional Neural Network

RU Xiao-qing, HUA Guo-guang, LI Li-hong, LI Li

(School of information and electrical engineering, Hebei University of Engineering, Handan 056000, China)

Abstract: In this paper, the deformable convolution module is introduced to enhance the modeling ability of the network to digital geometric transformation, and an improved handwritten digital recognition framework based on deformable CNN is proposed. In addition to improving the recognition accuracy, the framework can effectively reduce the training parameters and improve the recognition speed. This paper demonstrate state-of-the-art performance competing methods on the handwritten dataset and the transformed dataset.

Key words: handwritten digital recognition; convolutional neural network; deformable convolution

1 引言

光学字符识别(Optical Character Recognition, OCR),是指对文本资料的图像文件进行分析识别处理,获取文字及版面信息的过程. 光学字符识别是图像处理、模式识别等领域的热点研究问题. 而手写体数字识别技术作为其中重要的分支,其目的在于快速识别图像中的数字并进行分析存储操作. 传统手写体数字方法无法完成多变外观数字识别任务,极大影响了手写体数字系统的识别精度.

近年来,卷积神经网络(Convolutional Neural

Networks, CNN)在视觉识别任务中取得了巨大的成功,能有效的解决手写体因个人因素造成的笔画粗细、断笔、粘连等导致特征提取不准确的问题^[1],使得识别系统对手写风格具有较强的鲁棒性. 本文以 VGG-16^[2]卷积神经网络为基础,提出了一种基于形变卷积结构的 CNN,通过在原有网络结构中附加卷积层学习图像特征中的偏移量,使卷积核一局部、密集和自适应的变形方式适应输入的图像目标,附加的卷积层仅仅增加了少量的参数,并不会对其速度造成影响. 这种形变卷积结构能很好地提升 CNN 适应几何变换的能力,从而提高手写体

收稿日期: 2018-07-28; 修回日期: 2018-09-12

基金项目: 河北省自然科学基金(sF2015402150);河北省教育厅资助项目(ZD2015087);邯郸市科学技术研究与发展计划项目(1721203049-1)

数字的识别精度. 为了进一步提升手写体数字识别的速度, 论文采用了较小的卷积核取代原有网络结构的卷积核. 通过实验结果的分析, 以及与相应算法的比较, 证明了本文方法的有效性.

2 神经网络

据文献[3,4] 可得, 现有的一些神经网络算法在手写体数字识别中取得了成功, 但这些神经网络算法采用的是全连接的邻接关系网络来完成识别工作, 将产生大量的参数, 极大的消耗了计算机的计算资源. 以一个 28×28 的灰度图像为例: 由神经网络的输出公式可得: 以参数 1 表示当前层, x 为输入参数, 则 Z 为激活函数的参数, g 表示为激活函数, b 表示偏置数.

$$Z^{[1]} = w^{[1]} x + b^{[1]} \quad (1)$$

$$A^{[1]} = g^{[1]}(Z^{[1]}) \quad (2)$$

当输入一个大小为 28×28 像素的图像时, 则神经网络有 784 个输入神经元, 则网络输入参数个数为 784. 此时参数数量并不大, 神经网络能较快的输出结果, 但是若输入图像像素为 1000×1000 时, 参数数量将会很大, 参数过多无疑会影响算法的性能, 此时若继续采用全连接的神经网络显然是不合适的. 并且使用全连接的网络对图像进行分析并不符合逻辑, 因为全连接的网络并不考虑图像的空间结构, 而是以完全相同的基础来对待相距很远和彼此接近的输入像素, 但是输入的目标字符是空间结构概念. 为了解决神经网络收敛速度慢以及空间结构的问题, 论文将引入卷积神经网络, 利用卷积神经网络的稀疏连接、权值共享、多特征图的特性来有效的处理输入的手写体数字.

2.1 卷积神经网络

卷积神经网络是一类特殊的前馈神经网络, 主要由卷积层和池化层交替连接的. 卷积神经网络对特征的提取和分类都是在卷积层和池化层中进行的, 所以对卷积层和层级的优化能提高特征提取的精度并优化分类效果. 卷积神经网络普遍结构如图 1 所示, 图中包含了两层卷积层, 分别是 C_1 和 C_2 以及两层池化层, 分别为 P_1 和 P_2 , 将目标图像输入网络中, 将输出图像的特征, 特征数目通常是由任务决定. 输出的特征通常会连接一层到两层的全连接层, 最后连接柔性最大化层进行分类.

2.1.1 卷积层

卷积层可看做是一个局部感受野, 也就是网络的部分连接. 在一幅图像中, 局部相邻近的部位相关性较强, 较远的位置相关性较弱. 因此, 神经网络中的神经元与其上一层神经元就不必进行全连接, 也

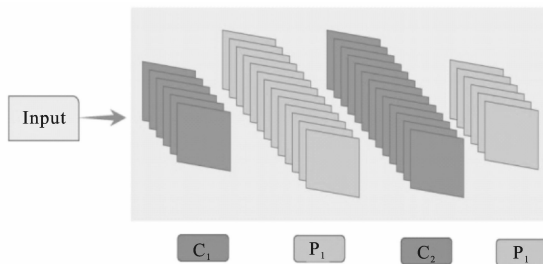


图 1 卷积神经网络结构

能较好的提取到局部的特征. 前层网络只需要局部感知, 在更高层将局部的信息综合起来就得到了全局的信息. 如上述提到的在全连接层的网络中, 输入被描述成一系列神经元, 而在卷积神经网络中, 是将其看做 28×28 的矩阵输入. 输入的目标图像的局部区域被视为神经元的局部感受野, 它是输入像素上的一个小窗口. 通过在输入的目标图像中交叉移动局部感受野就可以提取出更上一层的特征. 卷积核的操作过程具体如图 2 所示. 卷积操作过程如下: (1) 在输入特征图像上使用网格 R 进行局部区域采样. (2) 对采样值进行加权求和. 卷积操作通过网格 R 定义感受野的尺寸以及扩展率. 如: 设定卷积核的扩展率为 1, 则卷积核的大小为 3×3 , 则网络的采用偏移量为:

$$R = \{(-1, -1), (-1, 0), \dots, (0, 1), (1, 1)\} \quad (3)$$

对于输出的特征图 y 上的任意局部位置采样的值为:

$$y(p_0) = \sum_{p_n \in R} w(p_n) X(p_0 + p_n) \quad (4)$$

式中, p_n 表示的是网格 R 上进行迭代的位置, 并且每个网格位置 p_n 同时对应着权重参数 $w(p_n)$.

卷积层最大的特性是权值共享: 即为权值参数与阈值的共享. 如没有这个特性, 构造的神经网络会由于参数过多而损害算法的性能, 以由 10 个 $32 \times 32 \times 1$ 的特征图组成的特征图为例, 即每个特征图上有 1 024 个神经元. 每个神经元对应输入图像上一块 $5 \times 5 \times 3$ 的区域, 即一个神经元和输入图像的这块区域有 75 个连接, 即 75 个权值参数, 则共有 $75 \times 1\,024 \times 10 = 768\,000$ 个权值参数, 这是非常复杂的. 因此卷积神经网络引入“权值”共享原则, 即一个特征图上每个神经元对应的 75 个权值参数被每个神经元共享, 这样则只需 $75 \times 10 = 750$ 个权值参数, 而每个特征图的阈值也共享, 即需要 10 个阈值, 则总共需要 $750 + 10 = 760$ 个参数.

2.1.2 池化层

卷积神经网络除了使用卷积层, 也经常用到池化

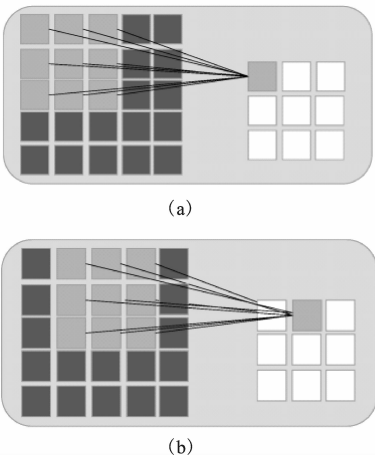


图 2 卷积核交叉移动过程

层来缩减模型的大小,提高网络模型的运算速度,同时提高所提取特征的鲁棒性.池化层的作用就是降低层的计算复杂度,简化从卷积层输出的信息.池化层类型分为平均池化与最大池化,常用到的是最大池化.最大池化过程中有一组超参数,即过滤器大小 f 及步长 s ,其中常用参数值是 $f=2,s=2$,其效果是相当于长度和宽度各减一半.另外,根据经验可得出,当最大池化输入大小为 $n \times n$ 时,设其填充 p (padding) 为 0,则输出大小为 $(\frac{n-f}{s}+1) \times (\frac{n-f}{s}+1)$.以 4×4 灰度图像为例,设其过滤器大小为 2×2 ,步长为 3,其最大池化过程如图 3 所示.

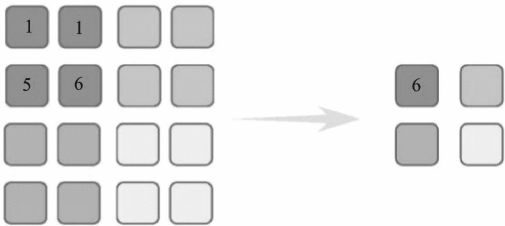


图 3 最大池化操作

3 形变卷积神经网络

在手写体数字识别任务中的一大挑战是如何处理数字的几何变换(包括尺度、位置、角度以及局部笔画的形变).CNN 因其固定的内在结构,导致了 CNN 无法很好的处理出现在手写体数字中的几何变换.一般的解决办法是进行数据增强,通过对数据集进行几何变化,使得数据集尽可能多的包含手写体数字可能的几何变化情况.这一方法的基础在于手写体数字的几何变换是已知的,对于未知的几何变换将无法很好的识别.为此,本文引入形变卷积^[5],使卷积层以自适应的变形方式适应输入的图

像目标,从而增强整体网络建模几何变换的能力.形变卷积核的本质是在原网格 R 的基础上加上偏移量,所加偏移量使通过附加的卷积层学习得到,具体如图 4 所示.

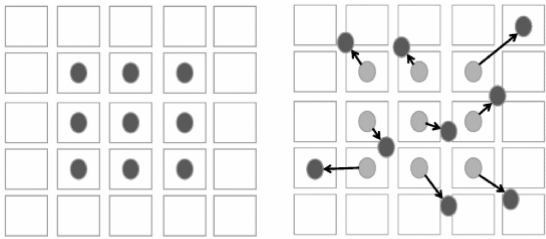


图 4 3×3 规则卷积抽样和形变卷积抽样

形变卷积的具体操作是在(4)式的基础上,网格 R 的每一个位置 p_n 增加了对应的偏移量 $\{\Delta p_n | n = 1, \dots, N\}$,其中 $N = |R|$.因此,形变卷积的输出结果为:

$$y(p_0) = \sum_{p_n \in R} w(p_n) X(p_0 + p_n + \Delta p_n) \quad (5)$$

形变卷积操作实现的关键是偏移量的学习.假设现有的卷积核,其大小为 3×3 ,为了学习到特征图的偏移量,我们构建了另一大小为 3×3 的卷积核,并用构建的卷积核处理输入的特征图,其输出的偏移量图大小与特征图大小一致,通道数为 $2 \times 3 \times 3$,分别表示了特征图上的网格 R 上对应采样点的水平方向和垂直方向的偏移.由于偏移量通常为非整数,因此需要通过双线性采样获取有效的采样网格 R ,配合卷积权重获取最终的结果,并且经过双线性采样后,在反向传播时能有效的求取梯度,从而学习到相关权重.具体的实现示例如图 5 所示.

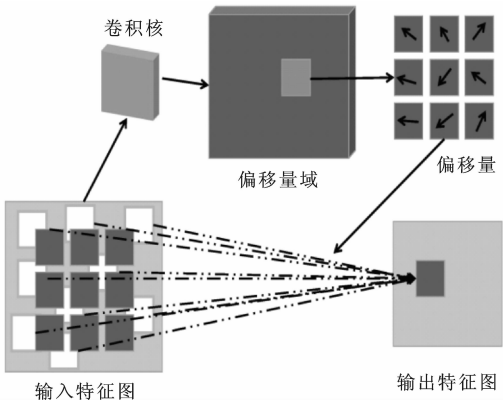


图 5 3×3 形变卷积

4 实验

实验平台为 python 3 上进行.本文验证算法使用的数据集为 MNIST 手写数字识别数据集.

4.1 实验数据集

MNIST 数据集来自美国国家标准与技术研究所(National Institute of Standards and Technology,NIST),是一个已经进行标记的手写体数字识别数据集.这个数据集由来自 250 个不同人手写的数字构成,其中 50%是高中学生,50%来自人口普查局的工作人员,包含了 0 到 9 总共 10 类的手写体数字图像,每个图像的大小为 28×28 像素,像素是由 0 到 1 的浮点数构成的,黑色越深表示数值越靠近 1.具体如图 6 所示.数据集共有 70 000 张手写体数字样本,其中 60 000 张为训练样本和 10 000 张为测试样本,部分示例如图 7 所示.

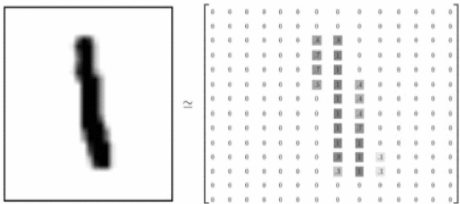


图 6 数据集图像表示结构



图 7 MNIST 数据集部分手写体数字

为了验证本文提出方法的有效性,通过仿射变换对 MNIST 数据句进行几何变换,获取经过变换后的 MNIST 手写体数据集进行手写体数字识别实验.通过与现有算法在正常 MNIST 数据集和变形 MNIST 数据集上进行对比实验.

4.2 实验训练细节

论文提出的基于形变卷积神经网络的总体结构如图 8 所示.本文以通过 ImageNet^[6]进行预训练的 VGG-16 网络为基础网络,将学习到的参数迁移到修改后的网络中,并进行修改.为了使网络获得能够适应手写体数字几何形变的能力,本文引入形变卷积核,取代 VGG-16 网络中的部分卷积核,同时使用 1×1 卷积核替换部分原有卷积核.实验采用的反向传播优化算法为 Adam 优化算法.卷积神经网络的各项参数为:学习率设置为 0.0001,迭代的次数为 10 000,批量归一化的大小为 32.具体而详细的网络结构被列在表 1 中(表中卷积核表示为 Conv,形变卷积核表示为 D-conv,池化表示为 Maxpool,

全连接表示为 Fcn).

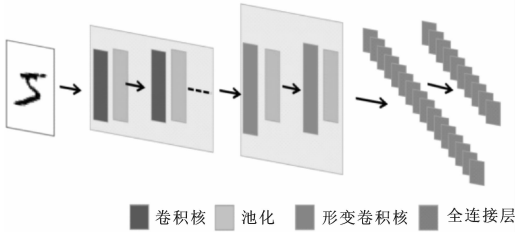


图 8 基于形变卷积神经网络结构

表 1 网络的具体设置

	1	2	3	4	5
结构名称	Conv1-1 Conv1-2	Maxpool	Conv3-1 Conv3-2	Maxpool	D-conv5-1 D-conv5-2
核大小	(3, 3) (3, 3)	(2, 2)	(3, 3) (1, 1)	(2, 2)	(3, 3) (3, 3)
	6	7	8	9	10
结构名称	Maxpool	D-conv7-1 D-conv7-2 D-conv7-3	Maxpool	Fcn-9	Fcn-10
核大小	(2, 2)	(3, 3) (3, 3) (1, 1)	(2, 2)	1024	10

4.3 实验结果分析

为了验证本文算法的有效性,我们在 MNIST 数据集上进行方法的评估.训练过程进行了 20 代,每次有 1 875 张训练样本,处理没一张样本所用的平均时间为 139ms 示.本文所提算法与其它算法的定量分析如表 2.本文提出算法在 MNIST 数据集中的准确率并不是最佳的,但是仍然取得了较优的精度.更值得注意的是本实验通过仿射变换获取了形变的 MNIST 数据集,并将基于 VGG-16 的常规卷积神经网络与基于形变卷积神经网络应用于形变的 MNIST 数据集,可以发现基于形变卷积神经网络能够很好的适应手写体数字的几何变换,在面对未知的变化时,能够更好的进行识别,并取得了 0.9476 的精度,比基于 VGG-16 的常规卷积神经网络高出了 0.3 的精度,这一结果表明了本文所提算法的有效性.

表 2 形变卷积神经网络与其它方法比较

方法	MNIST	形变 MNSIT
Lecun et al. [17]	0.9925	—
陈等人. [18]	0.9920	—
张等人. [19]	0.8900	—
王等人. [20]	0.9448	—
赵等人. [21]	0.9945	—
赵等人. [22]	0.9965	—
卷积神经网络	0.9972	0.6386
形变卷积神经网络	0.9948	0.9472

5 结束语

学习和适应手写体数字的几何变换能使数字分类达到了更好的效果,使得识别精度更高.因此,本文通过引入形变卷积操作,对 VGG-16 网络进行改进,并使用 MNIST 数据集进行训练,从而获取能有效提取和分类手写体数字的形变卷积神经网络.通过在 MNIST 数据集上进行实验,并与现有的方法以及卷积神经网络进行比较.实验结果表明,基于形变卷积神经网络的手写体数字方法具有较高的识别精度以及收敛速度,也侧面证明了对几何变换的处理能有效的提高特征提取精度.同时,本文的网络结构具有普适性,能够很好的泛化到多任务中.

参考文献:

[1] 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.

[2] KAREN SIONYAN, ANDREW ZISSERMAN. Very deep convolutional networks for large-scale image recognition[J]. Computer Science, 2014, arXiv preprint arXiv:1409.1556.

[3] 赵朋成,冯玉田,涂云轩.基于高倍特征深度残差网络的手写数字识别[J].电子测量技术,2018,6(41):

86-89.

[4] YANG F, JIN L, YANG W, et al. Handwritten/Printed Receipt Classification Using Attention-Based Convolutional Neural Network [C]// International Conference on Frontiers in Handwriting Recognition. IEEE, 2017:384-389.

[5] DAI JIFENG, et al. Deformable Convolutional Networks [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017:764-773.

[6] DENG J, DONG W, Socher R, et al. ImageNet: A large-scale hierarchical image database [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2009:248-255.

作者简介:

茹晓青 女,(1993-),硕士研究生.研究方向为图像处理、模式识别.

华国光 男,(1994-),硕士研究生.研究方向为计算机视觉、深度学习.

李丽宏(通讯作者) 女,(1974-),博士,副教授.研究方向为图像处理、计算机视觉、深度学习. E-mail:lilihgg@163.com.

李莉 女,(1984-),女,博士,讲师.研究方向为像处理、计算机视觉、模式识别.

(上接第 46 页)

参考文献:

[1] VERIKAS A, GELZINIS A, BACAUSKIENE M. Mining data with randomforests: a survey and results of new tests[J]. Pattern Recognition, 2011,44(2):330-349.

[2] AMARATUNGA D, CABRERA J, LEE Y. Enriched random forests[J]. Bioinformatics, 2008,24(18):2010-2014.

[3] MOORTHY K, MOHAMAD M S. Random forest for gene selection and microarray data classification[J]. Springer Berlin Heidelberg, 2012,7(3):142.

[4] 王燕,王兴芬,任俊玲.面向钓鱼网站敏感特征项选取的 IIGAIN 算法[J].计算机应用与软件,2016,33(4):297-301.

[5] HIPPI J, G NAKHAEIZADEH. Algorithms for association rule mining - a general survey and comparison [J]. Acm Sigkdd Explorations Newsletter, 2000,2(1):58-64.

[6] ZHANG W W, GONG J, LIU Q, et al. Lightweight

domain name detection algorithm based on morpheme features[J]. Ruan Jian XueBao/Journal of Software, 2016,27(9):2348-2364.

[7] 张茜,延志伟,李洪涛,等.网络钓鱼欺诈检测技术研究[J].网络与信息安全学报,2017,3(7):7-24.

[8] 王伟平,张兵.支持页面特征伪造识别的钓鱼网页检测方法[J].山东大学学报(理学版),2014,49(9):90-96.

[9] BERGHOLZ A, CHANG J H, PAASS G. Improved Phishing detection using model-based features [C]. Proceedings of the Nonference on Email and Anti-span(CEAS),2008.

[10] PhishTank. <http://www.phishtank.com>.

作者简介:

朱琪 男,(1994-),硕士研究生.研究方向为云计算与信息安全. E-mail:747116218@qq.com.

林果园 男,(1975-),博士,副教授.研究方向为网络空间安全、移动互联及其安全、云计算及其安全、信息系统及其安全.