基于卷积神经网络的商品图像精细分类

贾世杰,杨东坡,刘金环
(大连交通大学 电气信息学院,辽宁 大连 116028)

摘 要:针对某一类别商品图像的精细分类,研究并实现了深度学习中的卷积神经网络方法。所设计的卷积神经网络由 2 个卷积层、2 个亚采样层及 1 个完全连接层组成,特征平面的神经元只对其感受野的重叠区域做出反应,由反向传播算法调整网络参数最终完成学习任务。通过鞋类图像的精细分类实验表明,该方法平均分类正确率可达 91.5%。

关键词:卷积神经网络;商品图像;精细分类;亚采样层

中图分类号:TP391.4 **文献标志码**:A **文章编号**:1672-3767(2014)06-0091-06

Product Image Fine-grained Classification Based on Convolutional Neural Network

Jia Shijie, Yang Dongpo, Liu Jinhuan
(College of Electric Information, Dalian Jiaotong University, Dalian, Liaoning 116028, China)

Abstract: For the fine-grained classification of product image, convolutional neural network was explored and implemented as one deep learning method. The designed convolutional neural network consisted of two convolution layers, two subsampling layers and one fully connected layer, where the individual neurons were tiled to respond to overlapping regions in the visual field. The learning task was accomplished through adjusting the network parameters with back propagation algorithm. The proposed method has achieved an average accuracy of 91.5% in the fine-grained classification tests for the shoe images.

Key words: convolutional neural network; product image; fine-grained classification; subsampling layer

伴随着移动互联时代的到来,电子商务被广泛应用,许多购物网站也应运而生,如淘宝网、当当网、亚马逊、京东商城等,人们可以随时随地通过电脑或者手机等设备来选购自己满意的商品。所以,寻找商品的便捷性显得尤为重要。传统的商品查询方法是基于文本的,具有速度快、简单易行等优点,但这些标注仅仅说明商品的基本信息(元信息),如商品的名称、产地、尺寸、价格等,难以反映商品的完整特征。图像具有信息量大、表达简洁直观等优点,如果在网站中设置图片分类过滤器,无疑能方便用户进行浏览。

近年来,许多学者提出了基于内容的商品图像分类方法[1-3],这些方法无一例外是通过人为提取图像特征来实现图像分类,然而,很难有一种特征可以应用到所有商品图像的分类识别。并且,上述方法都是针对几十种或者几百种大范围商品图像来分类识别的,针对某一类别商品图像的精细分类却很少有人研究。

卷积神经网络是深度学习中一种高效的方法[4-7],已经成为众多科学领域的研究热点之一,特别是在模式识别领域,由于该网络可以直接输入原始图像,避免对图像进行复杂前期预处理,因而得到更为广泛的应用。本研究中主要识别以运动鞋为主的鞋类,这些鞋具有高度的相似性,因而可以称为精细分类。应用深度学习[8]中的卷积神经网络(convolutional neural networks, CNN)方法,克服传统的图像分类需要应用人工

收稿日期:2014-10-08
基金项目:辽宁省教育厅高等学校科学研究项目(L2014174)
作者简介:贾世杰(1969—),男,山东潍坊人,教授,博士,主要从事图像处理与模式识别方面的研究.
E-mail:jsj@djtu.edu.cn

提取特征的缺点，把原始图像像素作为输入，进而学习相应的特征。

1 卷积神经网络

人的大脑是深层的结构组织，通过对接收到的信息层层筛选抽样，最终得到对信息的认知。深度学习就是模拟人脑的一种网络结构，所应用的学习方式主要包括无监督学习和有监督学习，例如卷积神经网络是应用有监督学习方法，而深度信念网络(deep belief network，DBN)和栈式自编码神经网络(stack autoencode，SAE)是应用无监督学习方法。卷积神经网络是第一个真正成功使用多层次网络结构的深度学习算法。20 世纪 60 年代，Hubel 和 Wiesel 在研究猫科动物的大脑皮层时，发现通过一种独特的神经网络结构可以有效降低反馈神经网络的复杂性，进而提出了感受野的概念[9]；80 年代，Fukushim 在此理论基础上提出神经认知机的概念，也就是卷积神经网络的第一次网络实现[10]。

卷积神经网络是一个多层非全连接的神经网络，每层由多个二维平面组成，而每个平面由多个独立神经元组成。卷积神经网络包含两种特殊的网络结构：卷积层和亚采样层，结构如图 1 所示。其中，卷积层和亚采样层可以有多层，卷积神经网络的深度由此体现。卷积层中含有多个特征平面，完成特征抽取的任务，其中每个特征平面都代表上一层某方面的特征，特征平面由神经元构成，同一个特征平面的所有神经元共享一个连接权值。特征平面上的每个神经元只接受其相应感受野所传输的信号。每个卷积层后都会紧跟着一个亚采样层，由于输入的样本通过卷积层在特征空间进行重构，从而映射到高维空间，得到的高维特征映射是不能够直接作为特征使用的，因此要通过亚采样层进行降维，如果不对数据进行降维，则容易造成过拟合，还会导致维数灾难[11]。亚采样层特征平面上的每个神经元也是共享同一链接权重的。卷积神经网络通过权值共享大幅度减少需要训练的权值数目，从而大大降低对训练样本的需求。卷积层的特征平面与亚采样层的特征平面是一一对应的，亚采样层的神经元对卷积层上相应的感受野进行抽样(如取最值、均值等)，所以亚采样层上的神经元往往会大幅度减少。

卷积神经网络通过卷积层和亚采样层的相互配合来学习原始图像的特征，并且通过经典的 BP (back propagation)算法来调整参数，完成权值的更新，最终完成学习任务。BP 网络更新权值公式为[12]

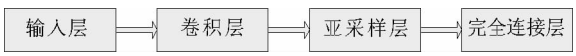

图 1 卷积神经网络结构示意图

Fig. 1 The diagram of CNN structure

$$w(t+1)=w(t)+\eta\delta(t)x(t)。\tag{1}$$

其中：$x(t)$为神经元的输出，$\delta(t)$表示该神经元的误差项，η 表示学习率。

卷积神经网络中卷积层的网络结构采用卷积的离散型[11]，表示为

$$x_\beta^\gamma = f\Big(\sum_{\alpha\in M_\beta} x_\alpha^{\gamma-1}k_{\alpha\beta}^\gamma + b_\beta^\gamma\Big)。\tag{2}$$

其中：M_β—输入特征的一个选择；k—卷积核；γ—网络的层数；b—每个输出特征映射添加的偏置，对于特定的输出映射，输入的特征映射可以应用不同的卷积核卷积得到[13]；f—卷积层神经元所用的激活函数，其中最常用的为 sigmoid 函数，类似的还有双曲正切函数，二者的不同在于 sigmoid 函数把$[-\infty,+\infty]$映射到$[0,1]$，而双曲正切函数则映射到$[-1,+1]$。

亚采样层的作用是对输入的特征映射进行采样，采样后，输入特征与输出的特征数目不会改变，但是输出特征的大小与输入特征相比会大幅减少。亚采样层表示为

$$x_\beta^\gamma = f(B_\beta^\gamma \mathrm{sub}(x_\beta^{\gamma-1})+b_\beta^\gamma)。\tag{3}$$

其中，sub(·)表示亚采样所用的函数，B 和 b 都是输出特征的偏置。f 的含义与卷积层的类似，表示亚采样层神经元的激活函数，可以取与卷积层一样的激活函数，也可以选择不用激活函数。

输入层感受野的特征经过卷积映射到新的特征空间，得到的特征作为亚采样层的输入。亚采样层对得到的特征进行抽样，最常用的方法称为池化(pooling)。池化就是把输入的特征图像分割为不重叠的矩形区域，而对相应的矩形区域做运算。对每个矩形取最大值的运算称为最大池化(max pooling)，取均值的运算

则称为均值池化(mean pooling)。池化作用不仅降低了上一层的复杂度,还可以防止过拟合。考虑到信息的损失情况,池化矩阵不宜取得过大。最后,把得到的多个特征映射转化为一个特征向量以完全连接的方式输出。卷积神经网络通过权值共享、局部感受野以及亚采样等过程来学习,具有较强的鲁棒性[14]。

2 基于卷积神经网络的商品图像精细分类

卷积神经网络在图像分类识别方面取得了较为理想的效果,例如手写字符识别、交通标志识别等。卷积神经网络的卷积层和亚采样层的设置非常灵活,不同的结构设置会得到不同的结果,因此针对不同类型的输入样本所应用的网络结构也不同。文献[7]给出的两个卷积网络应用实例中,第一个实例中输入图像大小为83×83,其采用的网络结构设置为:64c-10s-256c-6s,其中卷积核大小为9×9;另一个实例中输入图像大小为500×500,采用的网络结构设置为:20c-4s-20c-4s-200c,其中卷积核大小为7×7。文献[11]针对MNIST手写字符库,输入图像大小为28×28,其采用的网络结构设置为:4c-2s-12c-2s-26c,其中卷积核大小为5×5。文献[15]针对ImageNet图像库,输入图像大小为224×224,其采用的网络结构设置为:96c-5s-256c-3s-384c-384s-256c-3s,其中卷积核大小为5×5。文献[16]针对LFW(Labeled Faces in the Wild)图像库,输入图像大小为39×31,其采用的网络结构设置为:20c-2s-40c-3s-60c-2s-80c,其中卷积核大小为4×4和3×3。上述结构设置中c表示卷积层,s表示亚采样层,数字表示相应层所取特征映射个数。以上网络设置在其应用领域都取得了较理想的结果。经分析,卷积神经网络应用于图像分类时有以下规律:

1)实验数据集越大越复杂,所需网络层数越多;

2)后面的卷积层所取特征数目通常比前面的卷积层多;

3)输入图像的分辨率越大,所取的特征数目越多,亚采样层的抽样矩阵也越大;

4)卷积神经网络亚采样层数通常不超过3。

综合考虑以上规律,本研究设计的卷积神经网络包含2个卷积层、2个亚采样层及一个全连接层,取卷积核为5×5。图2所示为卷积神经网络用于分类的工作流程,包含训练过程和测试过程两部分,其中预处理完成灰度化和归一化等操作,以提高处理速度,规范输入样本;卷积层采用sigmoid函数作为激活函数,而亚采样层采用均值池化,无激活函数。

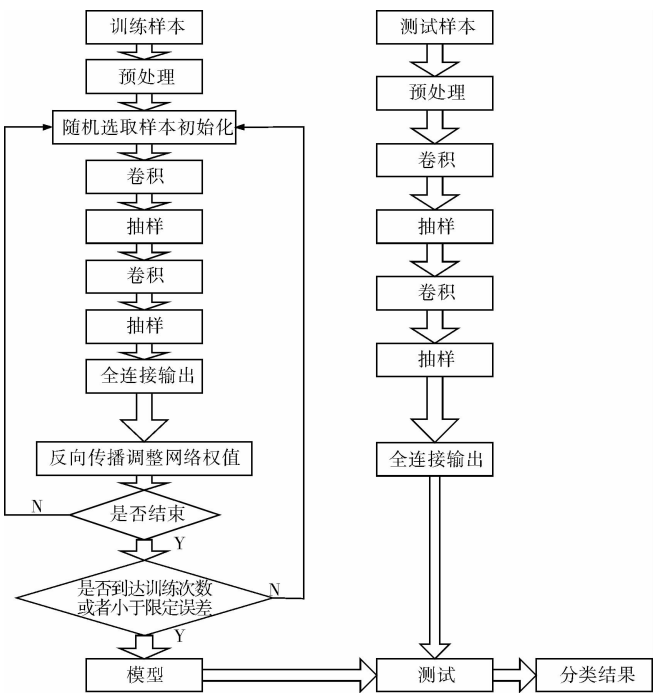

图2 卷积神经网络分类流程简图

Fig. 2 The diagram of classification with CNN

下面以图3为例,简要说明卷积神经网络提取特征的过程,输入的图像大小(像素)为100×100,通过第一个卷积层(C1)在特征空间重构后可获得6个特征映射,每个特征映射大小为96×96,因为卷积核的大小为5×5,特征映射是经过逐点获取的,所以经过卷积操作后每个特征映射的大小为(100−5+1)×(100−5+1)=96×96,而经过亚采样层(S1)的池化作用又可以得到6个大小为24×24的特征映射,因为池化矩阵的大小为4×4,而池化矩阵之间又是不重叠的,因此池化后的特征映射大小为(96/4)×(96/4)=24×24。S1的输出就是C2的输入,C2和S2也是通过这种计算得到新的特征映射数目。S1的特征映射通过卷积后得到含有12个特征映射的C2,每个特征映射大小为20×20,其中所用的12个卷积核大小为5×5。C2通过池化作用后得到含有12个特征映射的S2,每个特征映射大小为5×5,其中所用池化矩阵大小为4×4。

3 实验及结果分析

3.1 实验设置

实验全部在 Matlab2013a 环境下完成，Windows 7 操作系统，Intel Core CPU，主频 2.10 GHz，内存 3 GB。实验所用图像均取自淘宝网、当当网、亚马逊、京东商城等购物网站，以运动鞋为主，共 8 类，其中每类均在 600 幅以上，分辨率均在 200×200 像素以上，格式为 JEPG。部分图像样本如表 1 所示。

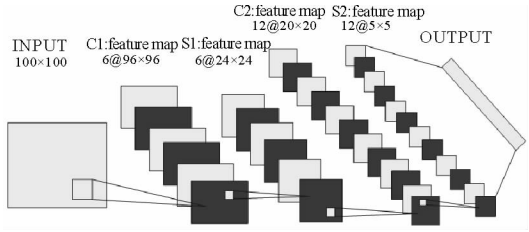

图 3 卷积神经网络特征提取过程

Fig. 3 The process of CNN feature extraction

所设计的卷积神经网络如图 3 所示，包含 2 个卷积层、2 个亚采样层和一个全连接层，选取卷积核大小为 5×5。为保证实验的客观性，实验图像均为随机选取，采用交叉验证法得到实验结果的平均正确率。

表 1 实验图库样本部分实例

Tab. 1 The samples of experimental images

登山鞋	跑步鞋	滑板鞋	篮球鞋	足球鞋	舞蹈鞋	训练鞋	帆布鞋
							

3.2 实验结果及分析

3.2.1 图像分辨率对分类结果的影响

采用不同分辨率的样本做训练，测试对分类结果的影响。输入训练样本为跑步鞋和训练鞋各自随机选取 500 张，测试图像随机选取 100 张。实验结果如表 2 所示，可以看到：

表 2 不同分辨率样本分类实验结果表

Tab. 2 The experimental results with different resolutions

分辨率/像素	100×100	84×84	68×68	52×52	36×36
平均准确率/%	84.0	75.5	76.0	64.0	50.0
训练时间/s	3 332.5	2 658.0	1 964.0	1 383.8	1 102.1
测试时间/s	9.9	7.4	4.8	3.7	2.8

1）随着图像分辨率的下降，由于信息的损失，导致分类识别正确率呈下降趋势；

2）图像分辨率越大，训练时间和测试时间越长。这是因为在相同的参数下，图像分辨率越大，通过卷积和采样得到的特征数目越多，处理时间就越长。

3.2.2 不同网络结构对分类结果的影响

网络结构具体设置如表 3 所示，实验结果如表 4 所示，通过分析可得：

1)亚采样层池化矩阵的选取对实验结果影响较大,若池化矩阵过小会导致亚采样层的池化作用不明显,因而分类效果不理想;

2)在一定范围内,特征映射数越多,分类的正确率就越高;

3)网络结构的改变,对训练时间影响较大,对测试时间影响较小。

3.2.3 迭代次数对分类结果的影响

卷积神经网络是通过迭代运算来求解权值的,经过多次迭代运算从而得到理想参数。应用不同迭代次数的实验结果如表5所示,具体分析如下:

1)在迭代次数很少的情况下,网络学习尚不充分,训练得到的模型也不理想,因此分类效果较差。随着迭代次数的增加,网络参数不断得到优化,分类识别准确率会随之上升,但是,随着迭代次数的增加,分类正确率上升幅度逐渐减小。当训练次数足够多时,网络参数就不会有大的变化,表示卷积网络已呈收敛状态,分类性能达到最优。

表3 不同网络结构的参数设置表

Tab.3 The parameters setup for different network structures

网络	C1	S1	C2	S2
Net1	12	4	24	4
Net2	10	4	20	4
Net3	8	4	16	4
Net4	6	4	12	4
Net5	6	2	12	2

表4 不同网络结构实验结果表

Tab.4 The experimental results with different network structures

网络	Net1	Net2	Net3	Net4	Net5
平均准确率/%	85.5	86.0	84.0	82.0	50.0
训练时间/s	10 484.1	7 796.6	5 632.2	4 956.23	6 079.9
测试时间/s	9.2	8.4	7.6	7.6	7.7

表5 不同迭代次数实验结果表

Tab.5 The experimental results with iterations

迭代次数	1	10	50	100	200	300	500
平均正确率/%	50.0	53.5	82.5	86.0	86.5	85.0	85.0
训练时间/s	111.10	749.10	4 047.44	7 796.59	15 428.98	25 721.50	42 139.80
测试时间/s	8.0	7.5	8.3	8.4	9.2	10.2	9.6

2)卷积神经网络训练所用时间和迭代次数是成正比的,但是测试时间却不受其影响。

3.2.4 本文方法与人工特征结合支持向量机方法的对比

人工特征采用塔式梯度直方图(pyramid histogram of oriented gradients,PHOG)和GIST(generalized search trees)特征,支持向量机(support vector machine, SVM)工具箱选择林智仁教授开发的libsvm3.17。实验中SVM选择高斯核函数,惩罚系数为1 000,核函数宽度为5,提取PHOG特征680维,提取GIST特征320维。实验中选取跑步鞋和舞蹈鞋为输入样本,随机选取训练样本400张,测试样本100张。

实验结果如表6所示,可得:

1)卷积神经网络的分类识别效果优于其他两种方法,应用不同的人工特征所得到的分类识别效果差别很大;

2)卷积神经网络在训练过程中所用的时间很长,但是其测试过程所用时间与其他方法对比却相近或者少于其他方法。

表6 不同机器学习方法实验结果表

Tab.6 The experimental results with different machine learning methods

机器学习方法	平均正确率/%	训练时间/s	测试时间/s
CNN	91.5	3 028.5	10.8
GIST+SVM	77.0	114.3	33.1
PHOG+SVM	85.5	36.2	8.8

4 结论

研究了卷积神经网络在商品图像精细分类中的应用,从输入图像分辨率、网络结构设置及迭代次数等方面测试了卷积神经网络的分类性能,实验所用图像取自当前流行的购物网站。本文方法不用预先人为提取

特征，所设计的由2个卷积层、2个亚采样层和1个全连接层组成的卷积神经网络，在商品图像精细分类实验中达到91.5%的平均分类正确率，优于使用人工特征＋SVM的分类结果。但是本文所设计的方法也存在一些不足：

1)只是研究了部分鞋类的分类效果，针对更多种类的商品精细图像分类还有待研究，因此在种类和数量上都需要补充，图像库还需完善。

2)只是应用了传统的卷积神经网络结构，采用了均值池化方法，塔式卷积神经网络及其他池化方法还有待研究。

3)因收敛速度慢导致训练时间较长，可以考虑将CNN与SVM相结合进行商品图像精细分类。

参考文献：

[1]贾世杰，邹娟，王茹香. 基于类词包技术的图像分类算法[J]. 化工自动化及仪表，2012(11)：204-209.
Jia Shijie，Zou Juan，Wang Ruxiang. An image classification algorithm base on class specific bag of words[J]. Control and Instruments in Chemical Industry，2012(11)：204-209.

[2]杨楠. 基于内容的商品图像分类技术研究[D]. 大连：大连理工大学，2011：6-50.

[3]贾世杰，孔祥维，付海燕，等. 基于互补特征和类描述的商品图像自动分类[J]. 电子与信息学报，2010，32(10)：2294-2300.
Jia Shijie，Kong Xiangwei，Fu Haiyan，et al. Auto classification of product image based on complementary features and class descriptor[J]. Journal of Electronics & Information Technology，2010，32(10)：2294-2300.

[4]Le Q V，Ndiam J，Coates A，et al. On optimization methods for deep learning[C]//The 28th International Conference on Machine Learning. Bellevue，Washington，June 28-July 2，2011：4-7.

[5]Mairal J，Koniusz P，Harchaoui Z，et al. Convolutional kernel networks[DB/OL]. [2014-06-12][2014-09-24]http://arxiv.org/abs/1406.3332.

[6]Szegedy C，Liu W，Jia Y，et al. Going deeper with convolutions[DB/OL]. [2014-09-17][2014-09-24]http://rxiv.org/abs/1409.4842.

[7]Lecun Y，Kavukcuoglu K，Farabet C. Convolutional networks and applications in vision[C]//IEEE International Symposium on Circuits and Systems. Pairs，May 30-June 2，2010：253-256.

[8]Bengio Y. Deep learning architectures for AI[J]. Foundations and Trends in Machine Learning，2009，2(1)：1-127.

[9]Hubel D H，Wiesel T N. Receptive fields，binocular interaction and functional architecture in the cat's visual cortex[J]. The Journal of Physiology，1962，160(1)：106-154.

[10]余凯，贾磊，陈雨强，等. 深度学习的昨天、今天和明天[J]. 计算机研究与发展，2013，50(9)：1799-1804.
Yu Kai，Jia Lei，Chen Yuqiang，et al. Deep learning：yesterday，today and tomorrow[J]. Journal of Computer Research and Development，2013，50(9)：1799-1804.

[11]李海峰，李纯果. 深度学习结构和算法比较分析[J]. 河北大学学报：自然科学版，2012，32(5)：538-544.
Li Haifeng，Li Chunguo. Note on deep architecture and deep learning algorithms[J]. Journal of Hebei University：Natural Science Edition，2012，32(5)：538-544.

[12]顾佳玲，彭宏京. 增长式卷积神经网络及其在人脸检测中的应用[J]. 系统仿真学报，2009(8)：2441-2445.
Gu Jialing，Peng Hongjing. Incremental convolution neural network and its application in face detection[J]. Journal of System Simulation，2009(8)：2441-2445.

[13]Bouvrie J. Notes on convolutional neural networks[DB/OL]. [2006-09-22][2014-09-24]. http://cogprints.org/5869/1/cnn_tutorial.pdf.

[14]Lecun Y，Bottou L，Bengio Y，et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE，1998，86(11)：2278-2324.

[15]Krizhevsky A，Sutskever I，Hinton G E. ImageNet classification with deep convolutional neural networks[C]//Advances in Neural Information Processing Systems，Annual Conference. Lake Tahoe，Nevada，Dec. 3-6，2012：1097-1105.

[16]Sun Y，Wang X，Tang X. Hybrid deep learning for face verification[C/OL]//2013 IEEE International Conference on Computer Vision. [2013-11-06][2014-09-24]. http://www.cs.utexas.edu/～grauman/papers/ suyog - jain-iccv2013.pdf.

（责任编辑：吕文红）