

基于改进的卷积神经网络在图片分类中的应用^①

赵新秋^② 贺海龙^③ 杨冬冬 段思雨

(燕山大学工业计算机控制工程河北省重点实验室 秦皇岛 066004)

摘 要 为解决传统的机器学习算法在图片分类中存在的准确率不高、耗时较长、难以适应复杂背景等问题,本文提出了基于改进的卷积神经网络(CNN) LeNet-5 的方法。该方法增加卷积核的个数能够有效地提取目标的特征,合理扩充数据集使得模型能够更快收敛,对比了不同激活函数对实验的影响,最后加入 dropout 层减少参数的个数。在对 101_ObjectCategories 图片库的 3 种目标(飞机、椅子、帆船)进行的分类实验中表明,本文提出的方法准确率高、分类效果良好、实用性强。

关键词 分类,卷积神经网络(CNN),LeNet-5,数据集,激活函数

0 引 言

随着社会的发展和生活水平的提高,人们对大规模图像的分类管理需求越来越大。图像分类是一种根据各自在图像信息中所反映的不同特征,把不同类别的目标区分开来的图像处理的方法。当前常用的分类方法有:基于尺度不变特征(scale invariant feature transform, SIFT)结合支持向量机分类器^[1],主要利用 SIFT 特征对目标的旋转、尺度和亮度变化保持不变的特征,以及支持向量机在样本数据线型不可分方面的优势,两者结合有利于增强分类算法的准确性;基于视觉词典模型的图像分类^[2]方法,通过对图像提取局部特征生成视觉码本,然后进行特征融合得到相应的图像表达向量,最后利用分类器进行分类。然而,这些传统的方法往往都依赖于图像特征的提取,当图像特征数量以及样本规模较大时,会消耗大量的时间。近年来,基于深度学习的方法应用在图像分类上取得了很大的成就。

深度学习^[3](deep learning)的概念起源于人工神经网络(artificial neural network, ANN),其基本思

想是通过构建深度学习网络将底层的输出作为下一层的输入,从大规模的数据中自主的学习到有效的高阶特征表示,最后将这些特征表示用于解决分类、回归和信息检索等特定的问题。得益于反向传播算法(back propagation, BP)^[4]的提出,引发了基于统计模型的机器学习^[5]的浪潮。20 世纪 90 年代,各种各样的浅层机器学习模型相继被提出,例如:支持向量机(support vector machine, SVM),boosting,最大熵方法(logistic regression, LR)等^[6],这些模型无论在理论分析还是应用中都获得了巨大的成功。2006 年 Hinton 等^[7]在《Science》上发表了一篇文章,开启了深度学习在学术界和工业界的浪潮,在 MNIST 数据库上取得了 98.8% 的识别率。自此基于卷积神经网络(convolutional neural network, CNN)的图像分类系统越来越受到研究人员的青睐。深度学习通过学习一种深层非线性网络结构,实现复杂函数的逼近,表征输入数据分布式表示,并展现了强大的从少数样本中学习数据集本质的能力。

LeNet-5 是一种典型的用来识别数字的卷积神经网络,当年美国大多数银行就是用它来识别支票上的手写数字。然而,传统的 LeNet-5 卷积神经网络

① 河北省自然科学基金(F2016203249)资助项目。
② 女,1969 年生,博士,副教授;研究方向:冶金综合自动化,智能控制;E-mail: zxq5460@ysu.edu.cn
③ 通讯作者,E-mail: 1250112436@qq.com
(收稿日期:2018-05-10)

络模型,输入图片像素大小必须为 32×32 ,网络结构简单,往往应用于比较简单的数据库,当样本特征较多或者背景较复杂时,该网络模型的识别率远远达不到人们的预期。针对以上问题,本文提出了基于改进的 LeNet-5 卷积神经网络。

1 卷积神经网络

1.1 卷积神经网络

卷积神经网络(CNN)^[7]是人工神经网络的一种,已成为当前语音识别和图像识别领域的研究热点。它的权值共享^[8]网络结构使之更类似于生物神经网络,降低了网络模型的复杂度,减少了权值的数量。卷积神经网络是为识别二维图像而特殊设计的一个多层感知器^[9],这种网络结构对平移、比例缩放、倾斜或者其他形式的变形具有高度不变性。根据学习规则^[10],某种特征训练次数越多,在以后识别过程中就越容易被检测。卷积神经网络的结构^[11]一般包括卷积层^[12]、下采样层^[13]和全连接层^[14]。

卷积层是卷积网络中的核心网络层,主要进行卷积操作,基于图像的空间局部相关性^[15]分别抽取图像的局部特征,通过将这些局部特征进行连接,可以形成整体特征。

下采样层主要是对卷积层的结果进行下采样操

作,主要作用是用来降低维度并保留有效信息,一定程度上避免过拟合^[16]。采样的策略有最大值区域采样和平均值区域采样,一般采用最大值区域采样。

全连接层连接所有特征,将输出值送给分类器。如果说卷积层、下采样层等操作是将原始数据映射到隐层特征空间的话,全连接层则起到将学到的“分布式特征表示”映射到样本标记空间的作用。

卷积网络的本质是一种输入到输出的映射,它能够学习大量的输入与输出之间的映射关系,而不需要任何输入和输出之间的精确数学表达式,只要用已知的模式对卷积网络加以训练,网络就具有输入输出之间的映射能力。

1.2 LeNet-5 卷积神经网络

LeNet-5 网络的输入为 32×32 的手写字体图片,包含 0 到 9 数字相当于 10 个类别的图片。3 个卷积层 C1、C3、C5 的卷积核个数分别为 6、16、120,卷积核个数的增加保证了对图像特征的充分提取;而且卷积核尺寸大小均为 5×5 ,移动步长均为 1;2 层池化层 S2、S4 均采用最大池化的方法,池化窗口大小均为 2×2 ,滑动窗口步长均为 2;F6 层是全连接层,包含 84 个神经元,该层中神经元的值是通过双曲正切函数计算输入向量和权重向量的点积得到的。最后一层为输出层,包含 10 个神经元,每个神经元代表一类。结构如图 1 所示。

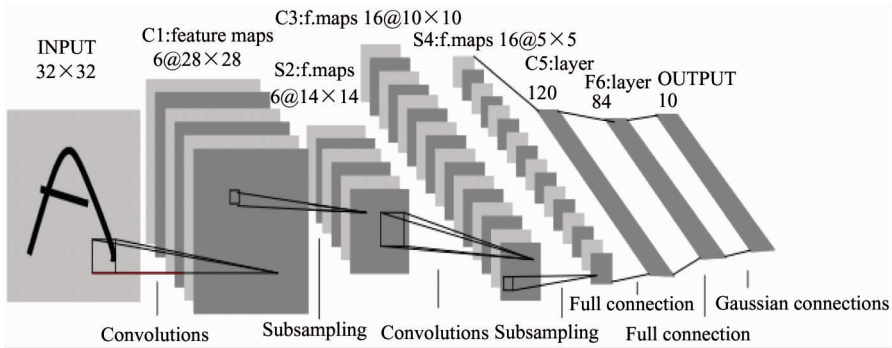


图 1 LeNet-5 卷积神经网络模型

但是由于该网络结构比较简单,只适用于背景简单、图像特征较少的图片,无法完成对背景复杂、特征较多的对象进行准确的分类,所以本文提出了基于改进的卷积神经网络。

1.3 改进的卷积神经网络

本文针对传统的卷积神经网络分类问题,提出了基于改进的卷积神经网络,结构如图 2 所示,该网络共有 8 层。

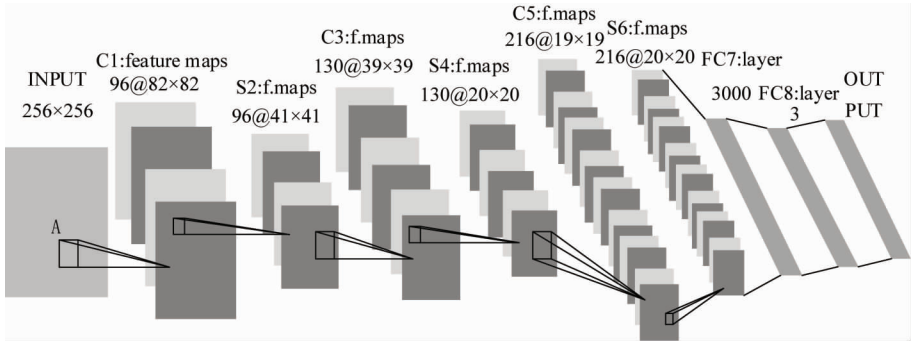


图 2 改进的卷积神经网络结构

与原有的 LeNet-5 网络不同的是本文将网络的输入改为 256×256 的图片,分辨率的增加能够保存样本的更多特征,有利于提高网络模型的分类效果。设输入图像的尺寸为 $W_i \times H_i$, 过滤器尺寸为 $w \times h$, 滑动步长为 s , 则输出图像的尺寸 W_0 和 H_0 的计算公式如下所示。

$$W_0 \times H_0 = [(W_i - w) / s + 1] \times [(H_i - h) / s + 1] \tag{1}$$

增加网络中卷积核的个数,3 个卷积层 C1、C3、C5 的卷积核个数分别为 96、130、216,在 CNN 中,对于卷积层的每一种输出特征图有:

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j\right) \tag{2}$$

式中, M_j 表示选择输入的特征图组合, k_{ij}^l 是输入的第 i 种特征图和输出第 j 种特征图之间连接所用的卷积核, b_j 是第 j 种特征图对应的偏置, f 是激活函数。每一个卷积核提取一种图像特征,考虑到图像背景的复杂度,卷积核个数的增加能够增强区分不同图片的能力。

增加 ReLu 层进行非线性激活,函数图像如图 3 所示。

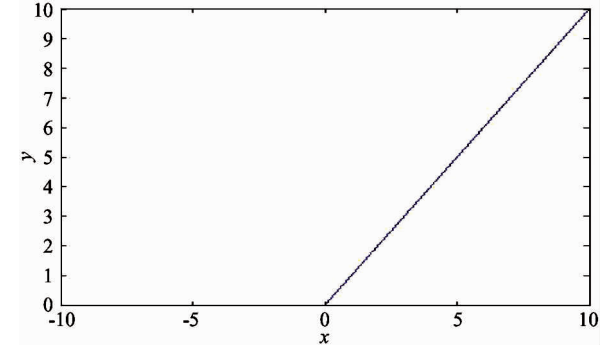


图 3 ReLu 函数曲线图

ReLu 函数得到的随机梯度下降法 (stochastic gradient descent, SGD)^[17] 收敛速度会比 Sigmoid/Tanh 快很多,而且相比于 Sigmoid/Tanh 需要计算指数等,ReLu 只需要一个阈值就可以得到激活值。

增加 dropout 层以提高网络的泛化能力,比例为 0.5,如图 4 所示。子图(a)是标准的一个全连接的神经网络,子图(b)是对子图(a)应用 dropout 的结果,它会以一定的概率随机地丢弃掉一些神经元。在训练的时候随机 dropout 一些神经元的激活,这样可以让模型鲁棒性^[18]更强,因为它不会太依赖某些局部的特征,局部特征有可能被丢弃。

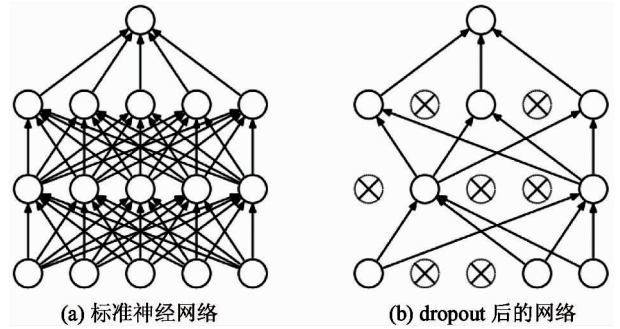


图 4 有无 dropout 层的网络结构

增加一个全连接层,用以减少图像特征信息^[19] 损失。FC7、FC8 均为全连接层,上一层的输出作为下一层的输入,最后输出 6 个神经元。全连接层结构如图 5 所示。

2 实验

2.1 实验环境

本实验中使用的处理器为 Intel(R) Core(TM) i5-7500 CPU @ 3.40 GHz,操作系统 win10 64 位,基

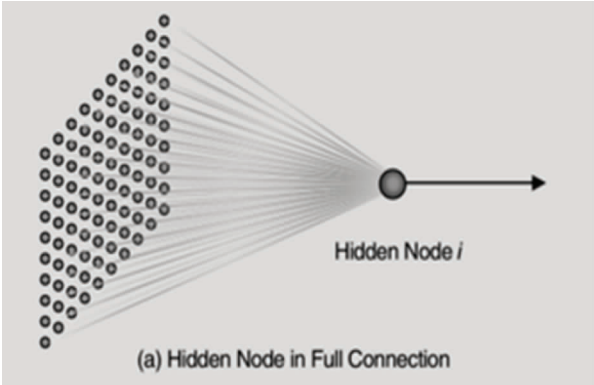


图 5 全连接结构图

于 caffe 框架,并使用 CUDA 进行加速。实验数据按照 4:1 的比例分为训练集和测试集,每次迭代次数为 5 000。

2.2 实验数据

实验数据来源于 101_ObjectCategories,是加利福尼亚理工学院 101 类数据库,共有 101 类图片,每个类别有 40 到 800 个图片,图片格式为 jpg 格式。本文从中选取了 3 类图片用做实验,如图 6 所示,分别是飞机、椅子和帆船,共计 976 张图片,每张图片均做归一化^[20]和灰度化处理^[21]。



图 6 样本图片

考虑到训练的样本数量对于最终实验结果的影响,本文采用图片翻转的方法对数据集进行合理地扩充,对每张图片进行水平翻转、垂直翻转、顺时针 90 度翻转,效果如图 7 所示。扩充之后的数据集图片为 3 754 张,有效地解决了因样本数量少造成实验误差大的问题,另外,在实验开始之前,统一将图片

的大小转换成 256 × 256。

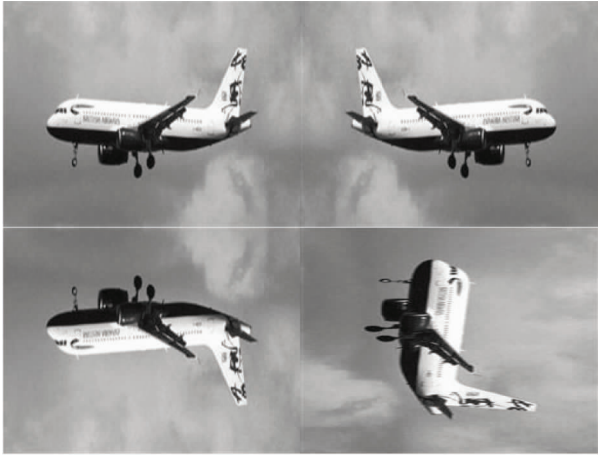


图 7 样本扩充图

2.3 实验结果对比

2.3.1 样本扩充前后对比

对于神经网络而言,训练集样本数量越多,测试准确率越高。本文将样本扩充前与样本扩充后进行了对比实验,扩充前样本数量为 976,在实验模型上的测试准确率为 89.1%;通过对样本进行水平翻转、垂直翻转、顺时针 90 度翻转后将样本数量增加至 3 754 张,扩充近 4 倍,实验模型上的测试准确率为 98.3%,如表 1 所示。

表 1 样本扩充前后对比实验

训练样本数	测试准确率(%)
976	89.1
3 754	98.3

由此可见,适当地扩充训练集有利于增强网络对样本特征的学习,提高网络的泛化能力。之后的实验均是建立在样本数量为 3 754 上进行的。

2.3.2 不同实验方法的比较

对比实验采用传统的特征 SIFT + 支持向量机(SVM)的方法与改进的卷积神经网络方法作比较,比较结果如表 2 所示。

表 2 两种方法对比结果

实验方法	测试准确率(%)
特征 SIFT + 支持向量机	83.20
改进的卷积神经网络	98.35

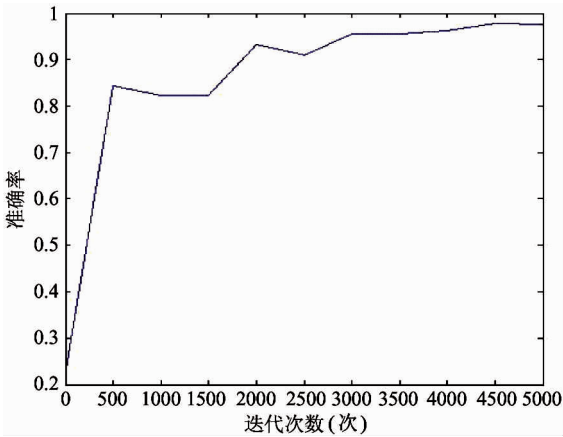


图 8 准确率图

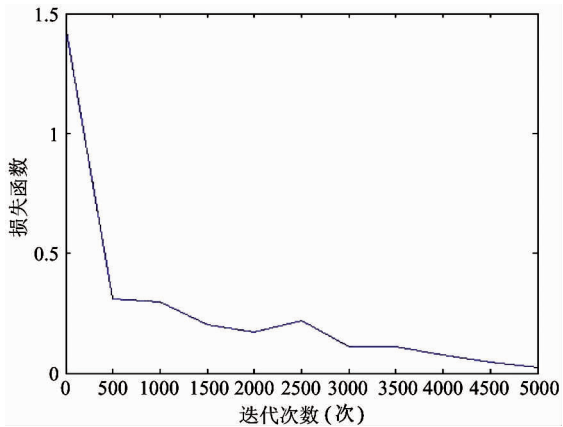


图 9 损失函数图

从表中可以看出传统的特征方法测试准确率仅为 83.20%，而改进后的卷积神经网络准确率已经达到了 95.35%，提高了 12 个百分点，另外传统的特征提取方法在背景复杂、特征点较多的情况下，分类效果不明显，没有卷积神经网络的准确率高。此外，图 8、9 分别为训练模型的准确率和损失函数图，从图中可以看出，随着迭代次数的增加，模型趋于收敛。

2.3.3 3 种激活函数的影响

神经网络中激活函数能够提供网络的非线性建模能力，网络中加入了激活函数之后，深度神经网络才具备了分层的非线性映射学习能力。所以本文探讨了 3 种不同的激活函数 Sigmoid、Tanh 和 ReLu 对实验的影响。其中 Sigmoid 是常用的 S 型非线性激活函数，其功能是把一个实数压缩到 0 至 1 之间，对中部区的信号增益大，对两侧区的信号增益小，数学

形式为 $f(x) = 1/(1 + e^{-x})$ 。但 Sigmoid 函数很容易出现梯度弥散的现象，无法完成对深层网络的训练。Tanh 也是常用的 S 型非线性激活函数，其功能是把实数压缩至 -1 到 +1 之间，对中部的信号增益较大，对两侧的信号增益较小，数学表达式为 $f(x) = (e^x - e^{-x})/(e^x + e^{-x})$ 。虽然克服了 Sigmoid 非 0 均值输出的缺点，但也存在梯度弥散的问题。ReLu 是一种非饱和线性修正函数，数学表达式为 $f(x) = \max(0, x)$ ，如果输入值 < 0 ，就令其等于 0，如果输入值 ≥ 0 则保持不变，它能够减少参数之间的依存关系，也缓解了过拟合问题。3 种激活函数对比实验结果如表 3 所示。

表 3 3 种激活函数对比实验

激活函数名称	测试准确率(%)
Sigmoid	74.38
Tanh	48.67
ReLu	98.35

从表格中能够清楚地发现，ReLu 相对于其他 2 种激活函数准确率更高。因为 Sigmoid、Tanh 都具有饱和性，造成梯度的消失，而 ReLu 能够在 $x > 0$ 时保持梯度不变，从而缓解梯消失的问题，所以本文选用 ReLu 作为激活函数。

2.3.4 有无 dropout 层的影响

由于 dropout 层能够在模型训练的时候随机让网络中的某些隐含层节点的权重不工作，但能使其权重保留下来，这样可以有效地防止模型的过拟合问题，提高网络的泛化能力。本文比较了有无 dropout 层对实验结果的影响，结果如表 4 所示。

表 4 有无 dropout 层对实验结果的影响

	测试准确率(%)
无 dropout 层	92.6
有 dropout 层	98.35

从实验结果能够发现，加入 dropout 层后，测试集准确率提升了将近 6 个百分点，提高了识别准确度。

3 结 论

本文在数据集不充足的情况下,合理地扩充了训练集,有效地减少了因样本数量少对实验造成误差的影响;增加了网络的深度以及卷积核的个数,使得模型能够更好地提取目标的特征;加入 dropout 层,有效地防止模型的过拟合,减少参数的数量。在传统的卷积神经网络上提出了改进,使得模型的准确率达到 98% 以上,与传统的基于尺度不变特征 (SIFT) + 支持向量机 (SVM) 的方法相比,提高了网络分类的准确率,具有更好的鲁棒性和泛化能力。另外,本文还探讨了 3 种不同的激活函数对实验的影响,证明了 ReLu 作为激活函数的优越性。由此可以得出,卷积神经网络在图像处理方面具有天然的优越性,但是本文的网络结构仅限于应用在指定的图片库上,而且网络的样本数量还有待提高,在过拟合问题上还有一些欠缺,如何针对特定的样本库设计特定的网络结构还需进行深入研究。随着理论与实践的发展,卷积神经网络在图像分类问题中将会得到更好地应用。

参考文献

- [1] Olgun M, Onarcan A O, Ozkan K, et al. Wheat grain classification by using dense SIFT features with SVM classifier[J]. *Computers & Electronics in Agriculture*, 2016, 122(C):185-190
- [2] 张永, 杨浩. 基于优化视觉词袋模型的图像分类方法[J]. *计算机应用*, 2017, 37(8): 2244-2247
- [3] Lecun Y, Bengio Y, Hinton G. Deep learning[J]. *Nature*, 2015, 521(7553): 436-436
- [4] 焦振. 误差反向传播神经网络 (BP 网络) 算法的启发式改进[J]. *安阳师范学院学报*, 2008(5): 47-49
- [5] 王珏, 石纯一. 机器学习研究[J]. *广西师范大学学报 (自然科学版)*, 2003, 21(2): 1-15
- [6] Farci P, Choo Q L, Kuo G, et al. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome[J]. *Journal of Hepatology*, 2002, 36(5): 582-585
- [7] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. *Science*, 2006, 313(5786): 504-507
- [8] Razavian A S, Azizpour H, Sullian J, et al. CNN features off-the-shelf: an astounding baseline for recognition[C]. In: *Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Columbus, USA, 2014. 512-519
- [9] 宋宜斌. 多层感知器的一种快速网络训练法及其应用[J]. *控制与决策*, 2000, 15(1):125-127
- [10] Wu C Y, Hsieh C Y, Chen S H, et al. Non-saturated binary image learning and recognition using the ratio-memory cellular neural network[C]. In: *Proceedings of the 7th IEEE International Workshop on Cellular Neural Networks and Their Applications*, Frankfurt, Germany, 2002. 624-629
- [11] 常亮, 邓小明, 周明全, 等. 图像理解中的卷积神经网络[J]. *自动化学报*, 2016, 42(9): 1300-1312
- [12] Teixeira F L, Chew W C, Oristaglio M L, et al. Perfectly matched layer and piecewise-linear recursive convolution for the FDTD solution of the 3D dispersive half-space problem[J]. *IEEE Transactions on Magnetics*, 2002, 34(5):2747-2750
- [13] Zhang N, Li L. Effects of common surfactants on protein digestion and MALDI MS analysis of the digested peptides using two-layer sample preparation[J]. *Rapid Communications in Mass Spectrometry*, 2004, 18(8):889-896
- [14] Ellwood B B, Benoist S L, El H A, et al. Impact ejecta layer from the mid-Devonian: possible connection to global mass extinctions. [J]. *Science*, 2003, 300(5626): 1734-7
- [15] 张凯, 李绪志. 基于区域相关性的高分辨率遥感图像融合算法[J]. *计算机辅助设计与图形学学报*, 2014, 26(6):889-895
- [16] 李彦冬, 郝宗波, 雷航. 卷积神经网络研究综述[J]. *计算机应用*, 2016, 36(9):2508-2515
- [17] Jin J, Fu K, Zhang C. Traffic sign recognition with hinge loss trained convolutional neural networks [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2014, 15(5):1991-2000
- [18] Sadegh N, Horowitz R. Stability and robustness analysis of a class of adaptive controllers for robotic manipulators [J]. *International Journal of Robotics Research*, 2016, 9(3):74-92
- [19] Gao X, Deng C, Li X, et al. Local feature based geomet-

ric-resistant image information hiding [J]. *Cognitive Computation*, 2010, 2(2):68-77

[20] Friston K J, Ashburner J, Frith C D, et al. Spatial registration and normalization of images [J]. *Human Brain Mapping*, 2010, 3(3):165-189

[21] Julian P, Dogaru R, Chua L O. A piecewise-linear simplicial coupling cell for CNN gray-level image processing [J]. *IEEE Transactions on circuits Systems*, 2002, 49(7):904-913

Application of improved convolution neural network in picture classification

Zhao Xinqiu, He Hailong, Yang Dongdong, Duan Siyu
(Key Laboratory of Industrial Computer Control Engineering of Hebei Province,
Yanshan University, Qinhuangdao 066004)

Abstract

In order to solve the problem that the traditional machine learning algorithm performs low accuracy, long time and difficult to adapt to the complex background in image classification, a novel method based on the improved convolutional neural network (CNN) LeNet-5 is proposed. The number of convolution kernels is increased to extract the characteristics of the target effectively, the data set is expanded reasonably to accelerate convergence of the model. Comparing with the effect of different activation functions, the dropout layer is increased to reduce the number of parameters. The classification experiment on the 3 targets (aircraft, chair, and sailboat) of the 101_Object-Categories image library shows that the proposed method has high accuracy, good classification effect and strong practicality.

Key words: classification, convolutional neural network (CNN), LeNet-5, data set, activation function