一种基于 GMP-LeNet 网络的车牌识别方法

林哲聪 张江鑫
(浙江工业大学信息工程学院 杭州 310023) (浙江省通信技术研究实验室 杭州 310023)

摘 要 车牌识别技术是智能交通管理系统的核心,对它的研究与开发具有重要的商业前景。传统的车牌字符识别方法存在特征提取复杂的问题,而卷积神经网络作为一种高效识别算法,对处理二维车牌图像具有独特的优越性。针对传统卷积神经网络 LeNet-5 识别车牌图像时,存在训练数据较少、全连接层参数冗余以及网络严重过拟合等一系列的问题,设计了一种全局中间值池化(GMP-LeNet)网络,其使用卷积层代替全连接层,利用 Network In Network 网络中的 1*1 卷积核进行通道降维,全局均值池化层直接将降维后的特征图馈送到输出层。实验证明,GMP-LeNet 网络能有效抑制过拟合现象,并具有较快的识别速度和较高的鲁棒性,车牌识别率达到了 98.5%。

关键词 车牌识别,卷积神经网络,LeNet-5,过拟合,池化
中图法分类号 TP181 **文献标识码** A

License Plate Recognition Method Based on GMP-LeNet Network

LIN Zhe-cong ZHANG Jiang-xin
(School of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China)
(Zhejiang Communication of Technology Research Laboratory, Hangzhou 310023, China)

Abstract As the core of intelligent traffic management system, the research of license plate recognition technology has important business prospects. The traditional license plate character recognition method has the problem of complex feature extraction. As an efficient recognition algorithm, convolution neural network has a unique superiority in dealing with two-dimensional license plate image. When the traditional convolution neural network LeNet-5 identifies the license plate image, there is a series of problems such as less training data, redundancy of the fully connection layer and over-fitting of the network. A global intermediate pool (GMP-LeNet) network was designed, which utilizes the convolution layer instead of the fully connection layer. The 1 * 1 convolution kernel learning from the NIN network is used to reduce channel dimension. Then the global mean pool layer feeds the feature graph to the output layer after the dimension reducing directly. Experiments show that GMP-LeNet network can suppress the over-fitting phenomenon effectively with a faster recognition speed and the higher robustness. The final license plate recognition rate is close to 98.5%.

Keywords License plate recognition, Convolution neural network, LeNet-5, Over-fitting, Pooling

1 引言

目前主流的车牌识别技术是通过神经网络[1]与模式识别[2]两种方式实现的。基于模式识别的车牌识别方法主要采用统计模式分类方法,它根据预处理的样本所提取的特征向量建立模式空间,利用样本模板进行模板匹配,选择最佳匹配作为结果来确定所归属的某个图像字符,有运算简单、识别速度快等优点,但对车牌图片的要求较高。基于神经网络的车牌字符识别采用了并行分布式系统,且具有自适应性好、识别率高、容错性较好及较强的自我调整能力等优点,但需要进行较长时间的网络训练,并且依赖于大量学习样本。

目前,国内外对车牌识别技术进行了广泛研究。文献[3]提出了改进型的基于欧拉数的模板匹配,去除冗余背景并填充有效背景,提高了最终的识别质量。Kim 等[4]使用提出的模糊神经网络算法,采用模糊聚类算法(fuzzy c-means algorithm)构建输入和中间层之间的网络结构,Max_Min 神经网络用于抑制层和输出层。MAHDI[5]提出了分类 SVM 算法来检测车牌,其性能接近于基于神经网络的车牌检测方法。文献[6]提出了多级二进制 SVM 分类方法,减少了传统分类器的使用个数,分类过程由粗到细,提高了分类的精度和效率。Abdullah 等[7]在 HSI 颜色空间中采用模糊识别技术来定位车牌,利用神经网络技术进行字符识别,取得了 93.7% 的识别率。文献[8]提出卷积神经网络(CNN)和长度特征(LF)的方法,通过改变激活函数提高网络的泛化能力。文献[9]融合多个卷积神经网络的识别结果与梯度下降训练算法来识别手写波斯/阿拉伯数字,实验结果达到 99.98% 的识别率。文献[10]对 LeNet-5 的 C5 层和输出层进行了简单的扩充,以匹配中文车牌。文献[11]在文献[10]的基础上,改进输入层、简化全连接层,提出了结构简单、识别速度较快的 SLeNet-5 网络。

本文针对中文车牌字符结构复杂、输出类别较多的问题,改进了传统 LeNet-5 卷积神经网络。通过采用全局中间值池

林哲聪(1993—),男,硕士,主要研究方向为卷积神经网络,E-mail:jesung@163.com;张江鑫(1964—),男,硕士,副教授,主要研究方向为光纤多媒体数字通信、有线电视技术等,E-mail:zjx@zjut.edu.cn。

化层替代全连接层，增加了 dropout 层，大大减少了可训练参数的数量，避免了过拟合，特别是在训练样本较少的情况下，仍能保持较高的识别性能。

2 基于 LeNet-5 网络的车牌识别方法

2.1 LeNet-5 卷积神经网络

卷积神经网络(CNN) 是一种深度学习(Deep Learning)网络结构，最初由 Hubel 和 Wiesel 提出，被广泛应用于模式分类领域。LeNet-5 作为一个经典的 CNN 模型，最早被应用于手写数字的识别。如图 1 所示，LeNet-5 由输入层、隐层和输出层组成。隐层分为 3 类：1)卷积层(C1，C3，C5)，用于特征提取；2)抽样层(S2，S4)，用于特征优化选取；3)全连接层(F6)，相当于传统多层感知机中的隐层。

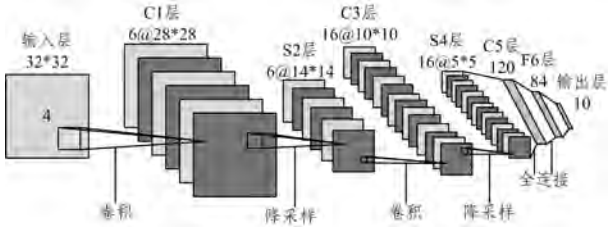

图 1 LeNet-5 结构图

2.1.1 卷积层

LeNet-5 中 C1，C3 和 C5 层是卷积层，通过一个局部感知野(即卷积核)提取该局部的特征，每个输出特征图通过权值共享来共享其所对应的卷积核，卷积层的计算公式为：

$$ConvOut_i = f(\sum_{j=1}^{N} input_j * KernelSet_i[j] + b_i^j) \quad (1)$$

其中，N 为输入特征的个数，i 表示输出特征图个数，每个卷积层中卷积核($KernelSet_i$)个数总与输出特征图 i 相等[12]。第 i 个特征图的输出($ConvOut_i$)是其所有输入特征图卷积对应的卷积核，加上直流偏置 b_i^j，经过激励函数 f 得到的。以 C1 层为例，输入 N 等于 1，包含 6 个特征图，因此 i 为 6。特征图中的每个神经元与输入中 5 * 5 的邻域相连，$KernelSet_i$ 为 5 * 5，因此得到的特征图大小为 28 * 28。

2.1.2 降采样层

S2，S4 是降采样层(池化层)，作用是对输入的特征图进行压缩，降低模型的耦合度，一方面使特征图尺寸变小，简化网络计算复杂度；另一方面压缩特征，提取最主要的特征。经典池化模型有两种最常见的模型：1)平均池化(Ave-Pooling)，它对池化域内所有值求和并取其平均数作为子采样特征图中的值；2)最大池化(Max-Pooling)只是取池化域中的最大值作为子采样的特征值。两者的算法表达式分别为：

$$AvePoolOut_i = f(\frac{1}{c^2}\sum_{m=1}^{c}\sum_{n=1}^{c} ConvOut_i^{mn} + b_2) \quad (2)$$

$$MaxPoolOut_i = f(\max_{m=1,n=1}^{c}\{ConvOut_i^{mn}\} + b_2) \quad (3)$$

其中，$ConvOut_i^{mn}$ 表示第 i 张输入特征图中(m，n)位置的元素，c 表示降采样池化域，b_2 为偏置。

2.2 LeNet-5 车牌识别方法

LeNet-5 车牌识别方法，顾名思义，就是将 LeNet-5 网络应用到车牌字符识别中，其流程如图 2 所示。首先需要准备车牌数据，将已分类的车牌字符图片打包成数据库文件，并构建 LeNet-5 网络，初始化学习率、迭代次数等参数；然后训练数据，将输入的图片数据库文件经过前向传输，得到输出层结果，计算误差并进行反向传输，更新每个层的权值。当迭代次数达到最大迭代次数时，结束训练，得到最终的模型文件。用户根据模型文件就可以识别出待识别的车牌字符。

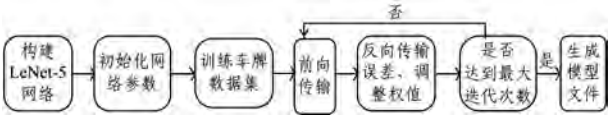

图 2 LeNet-5 车牌识别流程图

2.3 LeNet-5 网络识别中文车牌存在的问题

LeNet-5 适用于识别手写数字，识别率普遍达到 99%以上，而识别中文车牌字符却存在如下问题：

1)车牌字符比例为1:2，若将其水平拉伸，使其比例为1:1，则会导致字符变形，加大识别难度。

2)中文车牌存在汉字、数字和英文的组合，各地省份汉字一共有 32 个，由 0－9 这 10 个数字组成，英文字符由 A－Z(I、O 除外)24 个字母组成，分类总数一共为 66。由于输出层采用欧氏径向基函数，输出节点的增加更容易导致分类错误。

3)国内没有现成统一的车牌字符数据库，相比 CIFAR-100 和 mnist 数据库有 6 万个左右的训练样本，自身手动采集到的数据量远远达不到前者，因此网络很容易过拟合。

3 基于 GMP-LeNet 网络的车牌识别方法

3.1 LeNet-5 网络的改进

3.1.1 改进输入层

文献[13]比较了不同图像输入尺寸下的网络收敛性能，其将输入层的尺寸改为 32 * 16 以适应车牌比例。本文的一个设计是在 32 * 16 的基础上，在水平的左右各填充 8 个像素，如图 3 所示。这样既能保证长宽的对称，还不会使字符变形，有利于下一步进行特征提取。

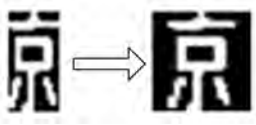

图 3 输入层改进

3.1.2 使用 dropout 层

我们增加了一个 dropout 层，以应对训练样本过少而导致的模型过拟合现象。dropout 层可以使模型在训练期间随机让网络某些隐含层节点的权重不工作，可以暂时认为不工作的节点不是网络结构的一部分，但是它的权重还会据需要保留下来。dropout 层会使训练样本的错误率提高，而使测试样本的错误率降低，提高网络的泛化能力。

3.1.3 用全局池化层代替全连接层

随着输出类别的增加，错误分类的概率(即误分率)会相应变大，一个可行的策略是增加 C5，F6 的个数，这种做法虽然能在一定程度上降低误分率，但会产生两个问题：一方面，传统 Lenet-5 的 C5 层有 120 * (16 * 5 * 5 + 1) = 48120 个可训练参数，F6 层有 84 * (120 * (1 * 1) + 1) = 10164 个可训练参数，这两个层的参数就已经占用了整个网络参数的 80%以上，增加 C5，F6 个数会加剧全连接参数的冗余，使网络更为庞大，增加训练时间，降低识别速度；另一方面，庞大的全连接层更容易出现过拟合的现象，换句话说，网络过度会拟合训练数据，错误地提取训练数据中的某些噪声作为特征，而忽略了

同一组训练样本的共有特征。过拟合主要的表现是，网络在训练阶段的识别率很高，而在实际测试阶段识别率会大幅下降。过拟合现象也是 Lenet-5 在进行手写字的识别时效果较好，而应用到其他数据集时却出现了较多问题的主要原因。

本文提出全局中间值池化(Global Mediate Pool)网络，简称为 GMP-LeNet。即增加了一个全局池化层代替全连接层，为每一个分类生成一个特征图，而不是在特征图顶部添加完全连接的层，取出每个特征图的池化值，并直接馈送到逻辑回归分类层。通过强制特征图和类别之间的对应，使特征图更容易地理解为类置信映射，此外在全局池化层中没有可优化的参数，因此在此层可避免过度拟合，对输入空间的变换更稳定。

3.2　GMP-LeNet 结构描述

GMP-LeNet 结构图如图 4 所示，从输入层到 S4 层相比 LeNet-5 基本保持不变，但由两个全连接层改为了两个卷积层和一个全局池化层。卷积层 C5 层由 100 个 3 * 3 特征图组成，C5 层中还包含了一个 dropout 层和激活函数层。激活函数采用 ReLu 函数，它的综合速率和效率都高于 Sigmoid 系函数[14]。卷积层 C6 层借鉴了 NIN(Network In Network)的结构[15]，使用了 1 * 1 的卷积核，使图片的深度从 100 降维为 66，主要作用是在丢失非常少信息的情况下降低数据的维度，实现两个通道的信息融合，强制将特征图个数与分类个数相对应。S7 层是全局池化层，将 66 个 3 * 3 特征图池化为1 * 1，组成特征向量，输入到 softmax 层。

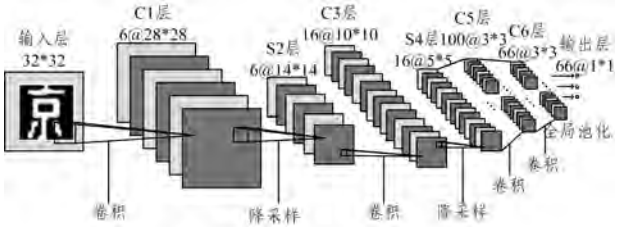

图 4　GMP-LeNet 结构图

表 1 列出了模型中各个层的网络参数，可以计算此时 C5 和 C6 层的可训练参数。C5 层有(100 * (16 * 3 * 3 + 1) = 14500 个可训练参数，C6 层有 66 * (100 * (1 * 1) + 1) = 13266 个可训练参数，相比 LeNet-5 的全连接层减少了一半以上的参数。

表 1　GMP-LeNet 网络参数

Layer 层	channels	Filter size	Filter stride	Padding	Input size
C1	6	5	1	—	32 * 32
S2	6	2	2	—	28 * 28
C3	16	5	1	—	14 * 14
S4	16	2	2	—	10 * 10
C5	100	5	1	1	5 * 5
C6	66	1	1	—	3 * 3
S7	66	3	1	—	3 * 3
Accuracy & Loss	66	—	—	—	1 * 1

3.3　GMP-LeNet 的中间值池化法

S7 层为全局池化层，作用是将 3 * 3 的灰度图像池化为 1 * 1的像素特征图，将其灰度值强制对应其类别，计算 loss 和 accuracy。池化方式的选择决定了特征图能否准确对应其输出类别。

Google-Net 的最后一层使用了 NIN 网络中的全局平均池化层[15]，而本文采用了一种相对折中的池化方式，即中间值池化法(Mediate Pool)[16]。这种算法可以兼顾平均池化法和最大池化法获取到的特征值，从而使模型产生的误差更小，稳定性更高，普遍适用于一般性图像。中间值池化算法的表达式为：

$$MediatePoolOut_i = f\left(\frac{1}{2}\left\{\frac{1}{c^2}\sum_{m=1}^{c}\sum_{n=1}^{c} ConvOut_i^{mn} + \max_{m=1,n=1}^{c} ConvOut_i^{mn}\right\} + b_2\right) \tag{4}$$

4　实验结果

本文的实验环境为 64 位 WIN7，Inter core i3 CPU 2.3 GHz，4 GB 内存，Caffe 框架，Visual Stdio 2013 平台。实验所需的车牌图像是在网络上大量搜集的经过定位、分割和阈值化，并最终进行归一化得到的。训练和测试的样例分别为 100 组和 10 组，一共 66 * (100 + 10) = 7260 组样本。

(1)准确率和学习率的关系曲线

图 5 显示了 GMP-LeNet 网络学习率 learning rate 和准确率 accuracy 的关系，初始学习率为 0.01，采用“inv”学习策略，学习率以负指数次方递减，最大迭代次数为 3000。从图 5 中可以看到，在前 600 次迭代中 accuracy 快速上升，随后开始震荡，随着学习步长降低，迭代到 1000 次时，accuracy 不再震荡，趋于稳定，并逐渐接近于 1。

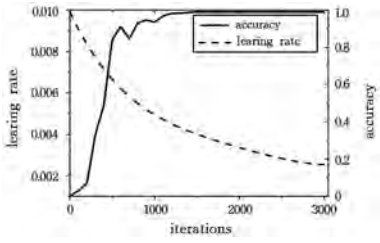

图 5　准确率和学习率曲线

(2)SLeNet、GMP-LeNet 的收敛曲线对比

文献[11]提出了 SLeNet-5 网络，它是为了适应中文车牌而专门改进了 LeNet-5。为了比较 SLeNet-5 和 GMP-LeNet 的收敛特性，将两种网络的可调节参数进行相同的设置，并采用相同的车牌数据集进行训练。最终得到的收敛曲线如图 6 所示，由图可知，GMP-LeNet 由于没有全连接层，参数更简化，loss 的收敛速度更快，收敛曲线更加稳定。

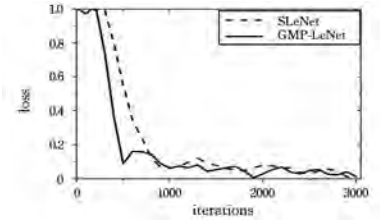

图 6　SLeNet 和 GMP-LeNet 的收敛曲线

(3)3 种 LeNet 网络的误分率性能折线图

图 7 对比了 SLeNet-5，GAP-LeNet(在 pool4 层采用全局均值池化方法)和 GMP-LeNet 网络的误分率性能。本文分别在训练阶段和测试阶段，针对数字英文(E)和汉字(CHN)，绘制了 loss 折线图，其中 train-loss(e)和 test-loss(e)分别为在训练阶段和测试阶段的数字英文的误分率，而 train-loss(chs)和 test-loss(chs)分别为在训练阶段和测试阶段的汉字字符的误分率。从图中可以看到，由于中文省份汉字形状复杂，无论在训练阶段还是测试阶段，3 种网络的数字英文字符

的误分率都低于汉字的误分率。测试集误差(test-loss)能反映该网络的实际识别能力,同一网络中从训练集到测试集上升的 loss 值在一定程度上反映了网络过拟合的程度。从图 7 可以看到,在 SLeNet 训练数量较少的车牌字符时,虽然对数字英文和汉字的训练误差分 SLeNet-5 别能达到 1%和 2%,但识别测试数据时,误差上升到了原来的两倍以上,可见 SLeNet-5 训练车牌时存在过度拟合,实际的车牌识别能力不高。GAP-LeNet 和 GMP-LeNet 网络由于都使用了全局池化层,并加入了 dropout 层惩罚训练时的节点权重,因此出现了训练集误差到测试集误差下降的情况,即(train-loss 比 test-loss 大),这种牺牲 train-loss 换取 test-loss 的方式是可取的,因为 train-loss 表示网络的拟合输入数据能力,而 test-loss 则代表网络的实际性能。针对两种全局池化 LeNet 网络,GAP-LeNet 和 GMP-LeNet 的过拟合现象都得到了有效抑制,而全局中间值池化算法每项的识别误差比全局均值池化低 0.5%~1%,这证明了改进全局池化方式是可行的。

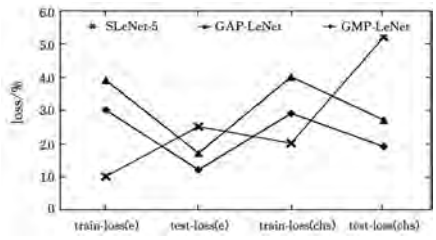

图 7 3 种 LeNet 网络的误分率性能

(4)3 种 LeNet 网络识别字符的时间对比

表 2 对比了 3 种网络的实际识别速度。SLeNet-5 由于参数众多,网络结构相对复杂,字符识别时间为 0.3 s。GAP-LeNet 和 GMP-LeNet 由于网络结构相似,利用特征图得到输出,识别时间需要约 0.26 s。

表 2 3 种网络的识别时间

LeNet 网络	识别每个字符的平均时间/s
SLeNet-5	0.305
GAP-LeNet	0.255
GMP-LNet	0.259

(5)常见识别算法的识别率对比

为了验证 GMP-LeNet 性能,搜集了一些使用不同车牌识别方法的文献,对比了这些算法在各自数据集下的识别率,如表 3 所列。经过改进的算法的字符测试集识别率都达到 93%以上,汉字的识别率普遍比数字字母低。其中,本文提出的 GMP-LeNet 性能相对最好,达到了 98%以上,能初步满足市面上对车牌识别率的要求。

表 3 几种车牌识别算法的识别率比较

(单位:%)

识别算法	数字和字母	汉字
模板匹配[3]	96.7	—
SVM 向量机[6]	97.5	95.3
BP 神经网络[17]	95.4	93.1
SLeNet-5	97.4	94.9
GAP-LeNet	98.4	97.3
GMP-LeNet	98.8	98.1

结束语 针对卷积神经网络全连接层参数繁多、训练样本过少、网络容易过拟合的现象,本文提出全局中间值池化方式。相比其他 CNN 模型,它的参数更为简洁,收敛速度和识别速度都有一定提升,测试集的误分率显著降低,过拟合现象也得以缓解,整个网络具有更好的泛化能力。此外,我们将 GMP-LeNet 和一些传统的车牌识别算法进行了比较,实验表明该车牌字符识别方法拥有较高的识别率,能够适用于实际的车牌识别系统之中。下一步的主要研究方向是改进卷积层优化特征提取,以更好地适应字符的扭曲和变形。

参 考 文 献

[1] 马爽,樊养余,雷涛,等.一种基于多特征提取的实用车牌识别方法[J].计算机应用研究,2013,30(11):3495-3499.

[2] 王晓雪,苏杏丽.数字图像处理在车牌识别中的应用[J].自动化仪表,2010,31(7):22-25.

[3] 陈玮,曹志广,李剑平.改进的模板匹配方法在车牌识别中的应用[J].计算机工程与设计,2013,13(5):1808-1811.

[4] KIM K B,CHO J H. Recognition System of Car License Plate using Fuzzy Neural Networks[J]. Journal of the Korea Society of Computer and Information,2007,12(5):313-319.

[5] MAHDI A,FATEMEH S,TABARI Z,et al. Automatic Iranian Vehicle License Plate Recognition System Based on Support Vector Machine (SVM) Algorithms[J]. Computer Engineering and Applications Journa,2013,2(1):161-174.

[6] 陈扬.数字图像模式识别在车牌自动识别中的应用研究[D].天津:天津大学,2017.

[7] ABDULLAH S N H S,KHALID M,YUSOF R,et al. License Plate Recognition using Multiclusterand Multilayer Neural Networks[J]. Information and Communication Tecimologies,2006,1:1818-1823.

[8] HE S,YANG C,PAN J S. The Research of Chinese License Plates Recognition Based on CNN and Length Feature[M]// Trends in Applied Knowledge-Based Systems and Data Science. Springer International Publishing,2016:389-397.

[9] AHRANJANY S S,RAZZAZI F. A Very High Accuracy Handwritten Character Recognition System forFarsi/Arabic Digits Using Convolutional Neural Networks[C]// Liverpool: International Conference on Bio-Inspired Computing. 2010:1585-1592.

[10] 赵志宏,杨绍普,马增强.基于卷积神经网络 LeNet-5 的车牌字符识别研究[J].系统仿真学报,2010,22(3):638-641.

[11] 张立,朱玉全,陈耿.基于卷积神经网络 SLeNet-5 的车牌识别方法[J].信息技术,2014,11:7-11.

[12] LIU Y J,HUANG H. Car Plate Character Recognition Using a Convolutional Neural Network with Shared HiddenLayers[C]// Chinese Automation Congress. Wuhan,2015:638-643.

[13] YUAN A Q,BAI G,JIAO L J. Offline Handwritten English Character Recognitionbased on Convolutional Neural Network [C]// International Workshop on Document Analysis Systems. Toronto,2012:125-129.

[14] KRIZHEVSKY A,SUTSKEVER I,HINTON G E. ImageNet classification with deep convolutional neural networks[C]// International Conference on Neural Information Processing Systems. Curran Associates Inc. 2012:1097-1105.

[15] LIN M,CHEN Q,YAN S C. Network In Network[D]. Banff: International Conference on Learning Representations,2014:1-10.

[16] 刘万军,梁雪剑,曲海成.不同池化模型的卷积神经网络学习性能研究[J].中国图象图形学报,2016,21(9):1178-1190.

[17] 郭荣艳,胡雪惠.BP 神经网络在车牌字符识别中的应用研究[J].计算机仿真,2010(9):299-301.