文章编号:1674-2869(2019)01-0089-04

基于卷积神经网络的验证码识别

张苏沛[1,2],刘 军*[1,2],肖澳文[1,2],杜 壮[1,2]
1. 智能机器人湖北省重点实验室(武汉工程大学),湖北 武汉 430205;
2. 武汉工程大学计算机科学与工程学院,湖北 武汉 430205

摘 要:针对传统验证码识别受字符分割限制的问题,将卷积神经网络应用到验证码的特征分析和识别中。使用验证码图像整体作为输入,对传统的LeNet-5的网络结构进行改进,构建一种端到端的卷积神经网络对图像由低级到高级逐层提取图像特征,选取ReLU作为激活函数,实现对验证码的识别。实验过程中设置对照组,研究不同因素对识别准确率的影响。测试结果显示,该模型能够进行端到端的识别,避免了字符分割方法流程过多导致的不足,在测试集上达到99%的识别率。结果表明训练次数的增加以及学习率的优化有助于提高卷积神经网络的准确率。

关键词:验证码;卷积神经网络;字符识别;学习率
中图分类号:TP317.4 **文献标识码**:A **doi**:10.3969/j.issn.1674-2869.2019.01.015

CAPTCHA Recognition Based on Convolutional Neural Network

ZHANG Supei[1,2],*LIU Jun*[*1,2],*XIAO Aowen*[1,2],*DU Zhuang*[1,2]
1. Hubei Key Laboratory of Intelligent Robot(Wuhan Institute of Technology), Wuhan 430205, China;
2. School of Computer Science & Engineering, Wuhan Institute of Technology, Wuhan 430205, China

Abstract: Aiming at the limitations of character segmentation in traditional completely automated public turing test to tell computers and humans apart (CAPTCHA) recognition we proposed an end-to-end convolutional neural network to characterize and identify CAPTCHAs. Firstly, a whole CAPTCHA image was used as an input, and then the convolutional neural network based on LeNet-5 was constructed to extract image features layer by layer from low-level to high-level. Finally, the ReLU function was selected as activation function to perform recognition task of CAPTCHA image. To study the effect of different factors on the recognition accuracy, a control group was provided in the experiments. The testing results show that the proposed methos realized the end-to-end recognition, thus avoiding the insufficiency caused by too many processes of character segmentation method and achieving 99% recognition rate on the test set. It is found that the increase of training times and the optimization of learning rate could improve the accuracy of convolutional neural network.

Keywords: captcha; convolutional neural network; character recognition; learning rate

验证码是一种区别用户是人或计算机的自动化测试方法,目前广泛应用于互联网,可以有效避免自动化程序滥用网站服务。验证码通常由字母和数字组成,为防止被机器自动识别,其分辨率通常较低,图片噪声较大。字符被一定程度地扭曲或倾斜,字符间往往存在粘连,“用户”需要识别并

收稿日期:2018-07-20
基金项目:智能机器人湖北省重点实验室开放基金(HBIR 201802);武汉工程大学第十届研究生教育创新基金
作者简介:张苏沛,硕士研究生。E-mail:zhangsupei@wit.edu.cn
***通信作者**:刘 军,博士,副教授。E-mail:liujun@wit.edu.cn
引文格式:张苏沛,刘军,肖澳文,等.基于卷积神经网络的验证码识别[J].武汉工程大学学报,2019,41(1):89-92.

键入正确的字符。人眼对验证码的识别率可以达到80%以上,但自动化程序识别准确度往往低于0.01%。这对于防止金融欺诈、电商刷单、恶意注册等批量化行为具有较好的效果[1-2]。

近年来,研究学者提出使用自动化方法识别验证码,如支持向量机(support vector machine,SVM)和旋转粗匹配算法[3-4]。上述方法一般有如下流程:图像预处理、二值化、去除离散噪声、字符分割、归一化、特征提取、训练和字符识别等[5]。这些方法模块之间相互独立,流程较为复杂,一旦某个模块出现故障,会直接影响最后的识别准确度。字符分割再识别的方法只针对复杂度较低、噪声较小的验证码图片,某些场景下字符相互堆叠,影响分割难度,并直接降低识别准确率。

受人类大脑处理信息方式的启发,有学者提出神经网络的方法。神经网络具有多层隐藏层结构,经过大量的训练,可以模拟人类的学习过程,实现机器智能化[6]。1998年,Le[7-8]首次提出卷积神经网络(convolutional neural network,CNN),该神经网络是一种前馈神经网络,与传统BP神经网络相比,层与层之间的神经元属与部分连接,而非全部连接。限于当时的计算机硬件条件,卷积神经网络未能广泛流行,而开销较少,效果较好的SVM[9]成为主要研究方法。21世纪初,计算机与互联网的快速发展使卷积神经网络的计算开销成为可能。基于LeCun提出的5层神经网络LeNet-5,陆续有研究者提出更深层的神经网络[10],并在各类图像识别比赛(如Kaggle)上取得优秀的成绩。近几年,类似ResNet等神经网络甚至实现了比人类识别更好的准确度[11-12]。

卷积神经网络最大的特点是将图片做为整体进行特征识别,省略了传统方法中图片预处理的过程。它包含了多层隐藏层结构,由底向上逐层学习更高层次的语义特征。目前,已有研究人员将卷积神经网络用于光学字符和大小写英文字符识别[13],并取得较高的识别效率。

基于上述分析,本文提出了基于LeNet-5的具有卷积结构的神经网络,针对验证码的特征进行学习和识别,能够有效避免人工设计梯度特征的缺陷。在缩短识别流程的基础上,一定程度上提高了验证码识别的准确度。整个系统可由如下步骤实现:1)利用标准第三方库生成验证码图片;2)利用大量验证码图片训练卷积神经网络,同时设置对照实验;3)利用训练完成的网络进行测试和识别。

1 卷积神经网络结构与设计

卷积神经网络是一种具有多层结构的前馈神经网络,通常应用于图像分析领域。对于一幅二维图像,不需要人工设计和手动提取特征,可以通过卷积层和池化层自动完成。与人类大脑的学习方法类似,神经网络可以通过提取出图像的具体特征来对图像进行判断。卷积层和池化层最大的便利性在于图像特征的大小及出现的位置,均不影响最终的提取。

1.1 卷积层

卷积层的工作方式可以理解为一个矩阵(也称为卷积核,大小通常为3×3,5×5)依次在图像上滑动并与对应位置的像素值做运算的过程,其最大特点是稀疏连接和权值共享[14]。稀疏连接是指每一层之间的神经元节点相互且非全连接,这是它与传统BP神经网络最大的区别,这一特点极大地降低了计算的复杂度并减少了权值的数量。权值共享是指卷积滤波器在对图像进行卷积操作得到特征图(feature map)之后,每一个滤波器共享同样的参数,包括相同的权重矩阵和偏置项。

1.2 池化层

卷积神经网络中,池化层往往紧随卷积层出现。输入图像经过卷积层得到特征图像之后,池化层可以进一步地提取图像特征。池化操作一般有3种:最大池化(max pooling)、平均池化(average pooling)以及随机池化(stochastic pooling)。最大池化指在邻域特征点内取最大值,这是最常用的池化操作也是本文选取的操作。池化层的存在可以有效降低特征图像的维度,降低复杂度,减少计算量,避免训练过程中的过拟合现象,保证平移不变性,提高模型的泛化性。

1.3 激活函数

激活函数在卷积神经网络中最大的作用是引入非线性因素[15],避免线性函数的局限性。常用的激活函数有3种:Sigmoid、Tanh以及ReLU。经研究指出[16],与Sigmoid和Tanh相比,ReLU不仅计算速度较快,避免了训练过程中过拟合的问题,而且可以更有效地实现深层的神经网络,防止梯度弥散。因此本文选用ReLU作为神经网络中的激活函数。

ReLU函数表达式为:

$$f(x)=\max(0,x) \tag{1}$$

其函数图像如图1所示。

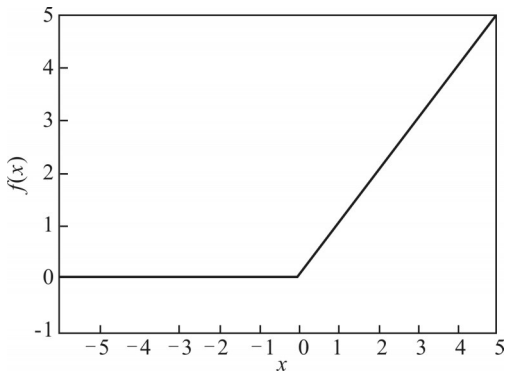

图1　ReLU函数图像

Fig. 1　Image of ReLU function

1.4　网络设计

在Le-Net 5的基础上，设计了7层的网络结构。除输入层以外，卷积层与池化层各有2层，前后交替排列，最后有2层全连接层。卷积层可以获取图像特征，池化层可以减少计算复杂度，避免过拟合；全连接层综合所有特征，输出不同类别的概率。另外激活函数ReLU引入的非线性因素，可以进一步增强网络的表达能力。本文卷积层中卷积核均为3×3，步长（strides）均为1；池化层核均为2×2。步长均为2。卷积层和池化层填充方式（padding）为“0”填充。网络中每一层的参数以及输出如下：

输入层，验证码图像经过灰度转换，大小为60像素×160像素。

C1层，32个卷积核对图像进行卷积操作，得到32张60×160的特征图，共有（3×3+1）×32=320个训练参数。

S2层，对C1层得到的特征图进行最大池化操作，得到32张30×80的特征图，共有（1+1）×32=64个训练参数。

C3层，64个卷积核对S2得到的特征图进行卷积操作，得到64张30×80的特征图，共有（3×3×32+1）×64=18 496个训练参数。

S4层，对C3层得到的特征图进行最大池化操作，得到64张15×40的特征图，共有（1+1）×64=128个训练参数。

F5层，全连接层，将15×40×64的特征图展开为38 400维的向量，输出1 024个节点。

F6层，全连接层，输入节点为1 024个，输出节点为62×4=248个（一张图片共4个字符，每个字符有62种可能）。

每一层参数及输出如表1所示。

表1　卷积神经网络各层参数

Tab.1　Parameters of each layer in CNN

名称	类别	窗口	步长	输出
C1	Convolution	3×3	1	60×160×32
S2	MaxPooling	2×2	2	30×80×32
C3	Convolution	3×3	1	30×80×64
S4	MaxPooling	2×2	2	15×40×64
F5	Full Connection	–	–	1 024
F6	Full Connection	–	–	248

2　实验部分

2.1　实验数据

卷积神经网络的训练需要大量的数据。由于缺少公共数据集，本文采用的图像由程序自动生成。验证码图像高度为60像素，宽度为160像素，字符集包括阿拉伯数字0~9，26个英文字母的大小写。每张图像包括4个随机字符并含有一定噪声，字符均有不同程度的扭曲、变形和粘连。数据集分为训练集和测试集，训练集包括40 000张图片，测试集包括1 000张图片。部分验证码图片如图2所示。

图2　部分验证码图片

Fig. 2　Some CAPTCHAs

2.2　实验结果

本文程序采用Python语言编写，基于Google开源机器学习框架Tensorflow。其中batch size为100，初始学习率（learning rate）为0.003，并在训练过程中不断衰减（每1 000次训练学习率下降为上一学习率的1/3），同时采用AdamOptimizer作为优化器。为验证学习率对训练结果的影响，训练过程中设置对照组（固定学习率）进行对比实验。

测试平台为Ubuntu 16.04，内存16 GB，GPU为NVIDIA GeForce GTX 1080Ti。两组实验迭代次数与准确率关系如表2所示。

实验结果表明，在学习率变化的情况下，迭代训练10 000次之后，模型已经达到了一个较好的准确度。随着迭代次数的增加，准确度继续提高；

尽管在训练过程中,准确率有一定程度的下降,但最终依然超过99%,并继续保持稳定。而当学习率保持固定不变时,模型性能较差。由此可以证明,变化的学习率可以提高卷积神经网络的训练效率。

表2 不同学习率与迭代次数下识别准确率的比较

Tab. 2 Recognition accuracy comparison under different learning rates and iterations

学习率	迭代数/次			
	10 000	20 000	30 000	40 000
变化学习率	0.941	0.984	0.953	0.992
固定学习率	0.86	0.915	0.923	0.95

测试集方面,除个别字符偶尔识别有误(如“0”,“o”与“O”),模型在大多数验证码上表现良好,测试集准确率在95%以上。

3 结 语

本文提出基于LeNet-5的网络模型对验证码进行识别,利用卷积神经网络对验证码的特征进行提取。实验表明:1)本文提出的卷积神经网络模型实现了端到端的识别,避免流程过多导致的设计缺陷,在验证码识别上具有较高的准确率;2)随着训练次数的增加,变化的学习率对于识别准确度的提高有明显帮助。

参考文献

[1] 王斌君, 王靖亚, 杜凯选,等. 验证码技术的攻防对策研究[J]. 计算机应用研究, 2013, 30(9): 2776-2779.

[2] 文晓阳, 高能, 夏鲁宁,等. 高效的验证码识别技术与验证码分类思想[J]. 计算机工程, 2009, 35(8): 186-188.

[3] 杨雄. 基于Python语言和支持向量机的字符验证码识别[J]. 数字技术与应用,2017(4):72-74.

[4] 高海昌,樊晔,王伟. 利用旋转归一化和粗匹配算法破解验证码[J]. 西安电子科技大学学报,2012,39(6): 78-83.

[5] 秦实宏,叶云丽. 复杂光照下的车牌定位方法[J]. 武汉工程大学学报,2015,37(11):69-73.

[6] 周志华, 陈世福. 神经网络集成[J]. 计算机学报, 2002, 25(1):1-8.

[7] LECUN Y, BOSERB, DENKER J S, et al. Handwritten digit recognition with a back-propagation network [C]// Advances in neural information processing systems. Denver: Morgan Kaufmann Publishers, 1990: 396-404.

[8] 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.

[9] CORTES C, VAPNIK V. Support-vector networks[J]. Machine Learning, 1995, 20(3):273-297.

[10] 汪家明,卢涛. 多尺度残差深度神经网络的卫星图像超分辨率算法[J]. 武汉工程大学学报,2018, 40(4):440-445.

[11] WU S, ZHONG S, LIU Y. Deep residual learning for image steganalysis [J]. Multimedia Tools and Applications,2018, 77(9):10437-10453.

[12] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [J]. IEEE Conference on Computer Vision and Pattern Recognition,2014(9):1-9.

[13] SHI B, BAI X, YAO C. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017,39(11): 2298-2304.

[14] 洪汉玉, 王澍, 朱浩,等. 低对比度嵌入型钢坯字符识别方法[J]. 武汉工程大学学报, 2012, 34(12): 38-43.

[15] 曲之琳,胡晓飞. 基于改进激活函数的卷积神经网络研究[J]. 计算机技术与发展,2017,27(12):77-80.

[16] 田娟,李英祥,李彤岩. 激活函数在卷积神经网络中的对比研究[J]. 计算机系统应用,2018,27(7): 43-49.

本文编辑:陈小平