

基于深度学习的车牌相似字符识别

潘翔 王恒

(浙江工业大学计算机科学与技术学院 杭州 310023)

摘要 针对车牌相似字符难以识别的问题,提出了基于深度学习的特征提取和识别方法。该方法首先对字符图像进行归一化处理;然后以归一化后的图像为输入,构建 5 层深度网络对相似字符由低层到高层的特征表达。在激励函数定义上采用对字符边缘特征敏感的卷积函数,从而能够对相似字符的局部差异进行分析。在实验部分与支持向量机(SVM)算法的分类效果进行比较,结果表明所提算法的识别率提高了 5%。

关键词 深度学习,深度卷积神经网络,相似字符识别

中图分类号 TP391.4 文献标识码 A

Similar Character Recognition of License Plates Based on Deep Learning

PAN Xiang WANG Heng

(College of Computer Science & Technology, Zhejiang University of Technology, Hangzhou 310023, China)

Abstract It is hard to recognize similar characters of a license plate, so a new method based on deep learning was proposed to extract features and recognize similar characters. Firstly, this method is to normalize the character images, and then the normalized images will be regarded as input. We built five-layers architecture of deep network and extracted similar characters featured in representing from low-level to high-level. The convolution function which is sensitive to character edge is adopted, so that it can analyze the local differences of similar characters. We compared this method with support vector machine (SVM) in the experiments. The results show that the accuracy rate of the proposed method increases 5%.

Keywords Deep learning, Deep convolutional neural network, Similar character recognition

1 引言

车牌识别是智能交通领域的一个基础研究问题,也是用于采集和分析车辆大数据信息的前提和基础。近年来,车牌识别研究取得了明显的进展。但是,对于复杂情况下的车牌识别问题,由于受光照、车速等影响,车牌识别的准确率还有待进一步提高。特别是其中存在如,“0,D,Q”,“8,B”等全局相似度高的相似字符。一些退化车牌的特征差异性非常不明显,这也是当前车牌识别的一个主要难题。

对于车牌识别,目前主要是采用判别式分析方法,这些方法包括基于模板匹配的字符识别^[1]、基于支持向量机的字符识别^[2]和神经网络的字符识别算法^[3]等。模板匹配法过程较为简单,但是受额外的因素干扰较大,比如获得的图像光照不均或形状改变,因此其识别率和鲁棒性不高。支持向量机和神经网络有较强的自学能力,能够在一定程度上克服字符形状的改变和位置偏移的缺陷。为了进一步提高车牌识别精度,研究人员在车牌预处理方面也做了很多研究。谢永祥等人在 HSV 颜色空间中通过结合纹理和模板特征来提高车牌定位和切分准确率,从而使得车牌更容易识别^[4]。类似地,洪涛等人针对车牌粘连字符自适应定位分割问题提出了高精度算法,同样也可以用于车牌字符切分^[5]。费继友等人则是提

出采用车牌边缘特征进行车牌智能校正和提取^[6]。赵伟等人则是采用 Retinex 理论进行车牌图像增强研究,提高字符质量。

另一方面,深度学习研究在模式识别领域产生了极大的推动作用。深度神经网络^[7]是一种含有多隐含层的机器学习模型并且具有特定的网络结构,与传统方法依赖人工设计特征相比,它避免了靠个人经验和知识设计特征算法,提升了分类识别的效率,目前比较常用的有深度卷积神经网络(Convolutional Neural Network, CNN)^[8]和深度信念网络(Deep Belief Network, DBN)^[9]。其中深度信念网络未考虑图像的二维结构,因此卷积神经网络更适合于二维结构的数据,可以直接将图像作为输入,自动学习特征,提取具有平移和形变不变性的特征表示^[10]。美国大学教授 Yan^[11]首次将深度学习应用于手写邮政编码数字识别的美国邮政服务中,提出了含有 3 层隐含层的网络结构,通过反向传播算法取得了良好的识别效果。Yan 经过不断研究于 1998 年提出了卷积神经网络(LeNet-5),它包含 6 层隐含层,与先前的网络结构相比,在很大程度上提高了手写数字识别的准确率^[12]。随着 LeNet-5 在手写数字识别上的巨大成功,不少学者开始将卷积神经网络用于手写字符识别^[13]、印刷字符识别^[14],并指出这种结构的模型在识别二维数据时比传统方法的表现更加优秀。

本文受浙江省自然科学基金(LY15F020024),浙江省文物局项目(2014014)资助。

潘翔(1977—),男,博士,教授,博士生导师,主要研究方向为模式识别、多媒体搜索、图形图像处理;王恒(1990—),男,硕士生,主要研究方向为图形图像处理。

Krizhevsky A 等^[15]于 2012 年首次将深度学习应用到 ILS-VRC 挑战赛中^[16],并在自己参与的组里取得第一名的成绩。在 2014 年的挑战赛中,几乎全部的参赛队伍都使用了深度卷积神经网络或者及其延生的方法^[17],而且取得了非常好的成绩。

受上述研究的启发,针对车牌相似字符提出深度识别网络。与已有判别式方法相比,深度识别网络对标注数据进行自适应学习和训练,从而抽取最能反映相似字符差异性的特征进行识别。这种方法大大提高了识别准确率。

2 深度卷积神经识别网络结构

本文采用深度卷积神经网络进行相似车牌字符识别。该网络采用反向传播结构,直接对相似字符图像进行特征提取。

本文使用的深度卷积神经网络利用局部感知和权值共享降低了网络的复杂度,也加快了网络的学习和训练的速度,并且使用反向传播算法使得网络更加收敛,得到了更好的训练模型。具有卷积结构的神经网络的最基本组成单元为单层卷积神经网络(见图 1)。首先在输入层之后连接多个单层卷积神经网络,完成了卷积神经网络最核心的组成部分,最后在它之后连接全连接层和分类器,即组成了一个深度卷积神经网络,并且每个前一层输出的数据作为后一层的输入。单层卷积神经网络一般由卷积、非线性变换和下采样 3 个最基本的过程组成。

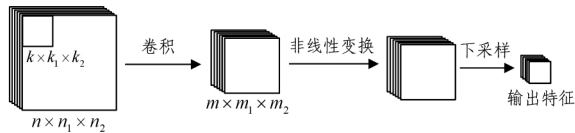


图 1 单层卷积神经网络结构

2.1 卷积层

卷积层的主要作用是特征的提取,它通过模拟生物的视觉皮层细胞,可对输入信号进行局部感知,这样和前一层的输入进行局部感受野^[18]相连,然后提取特征,这种局部感受野也称为卷积核。每个卷积核与前一层的特征图经过卷积操作后,能实现同一个映射面上输入数据的权值共享。同时这种权值共享的方式减少了参数的数量,也简化了模型的复杂度。这种局部感受野相当于特征的提取器。局部感受野的作用类似于一个 $n_1 \times n_2$ 大小的滑动窗口,并按一定的步长在字符图像上进行滑动,并且遍历了图像的所有位置。车牌的字符一般由粗的直线或曲线组成,形状特征明显,因此经过卷积之后,可以检测到字符图像的边缘、线段、角点等初级的视觉特征。因为卷积过程是对字符图像的所有位置上的特征进行提取,并映射到下一层,所以避免了字符局部特征的丢失,也使得不同的字符特征存在差异性,有利于字符的识别。

卷积过程以由 n 个大小为 $n_1 \times n_2$ 的二维数据组成的数据 x_i 作为输入,得到由 m 个 $m_1 \times m_2$ 大小的二维数据构成的输出结果 y_i 。该过程使用的卷积核 k 的大小为 $k_1 \times k_2$ (通常卷积核的大小为 5×5)。计算的形式如下:

$$y_j = f(\sum x_i * k_{ij} + b_j) \tag{1}$$

其中, $*$ 为卷积运算符, b_j 为偏置。

2.2 下采样层

下采样是对上一层输出的数据进行池化操作。池化操作

之后,图像的分辨率降低,网络能抵抗图像的缩放、平移等形变问题,同时还能减少特征的数量。由于拍摄角度的问题和图像分割的原因,获得车牌字符的位置和角度不一,通过本文采用的最大值池化后解决了上述问题,因此所提算法对字符图像的旋转、平移等有更好的鲁棒性。

经过下采样之后,得到与输入相同个数的输出特征图。下采样的计算形式如下:

$$x_i = f(\beta_i \text{down}(x_i) + b_i) \tag{2}$$

其中, $\text{down}()$ 表示下采样函数, β_i 表示下采样系数, b_i 为偏置。

2.3 非线性变换

非线性变换是将卷积过程获得的特征 x 作为输入,然后得到输入与输出的映射函数 $y = R(x)$ 。LeCun Yan 提出的 LeNet-5 网络中采用非线性函数 sigmoid,还有 sigmoid 改进版 tanh,softsign 等^[19]饱和非线性函数,但是目前深度学习多采用不饱和非线性函数 ReLU(Rectified Linear Units)^[20-21]。因为随着训练梯度的下降,ReLU 的收敛速度比使用线性饱和函数快很多,所以网络的训练速度同样得到了提升^[15]。4 种常见的非线性操作函数如下。

sigmoid:

$$f(x) = \frac{1}{1 + e^{-x}} \tag{3}$$

tanh:

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{4}$$

softsign:

$$f(x) = \frac{x}{1 + |y|} \tag{5}$$

ReLU:

$$f(x) = \max(0, x) \tag{6}$$

本文采用的是 5 层结构的网络(见图 2),第一层为输入层 I 是,数据大小为 28×28 字符的灰度图像。C1 层是个卷积层,对输入层经过大小为 5×5 的卷积核和可加偏置 b 的卷积操作之后获得 10 个 24×24 大小的特征图,其中特征图中的每个像素点与前一层的输入层相对应的一个 5×5 的区域局部相连。下采样层 S2 含有 10 个特征图,通过大小为 2×2 的最大池化值操作对卷积层 C1 的特征图进行池化后,S2 的特征图的大小降为 12×12 。同样地,对 S2 层的特征图经过 10 个大小为 5×5 的卷积核卷积操作后获得 10 个大小为 8×8 的特征图的 C3 层。S4 是第 2 个下采样层,经过 2×2 的最大池化采样后获得 10 个特征图并且每个特征图的大小为 4×4 。F6 包含 100 个神经元,它与 C5 是全连接的。最后一层 O 是输出层,由于是只针对两个字符的识别,因此输出层只含两个神经元,这两个神经元分别代表不同类别的字符。

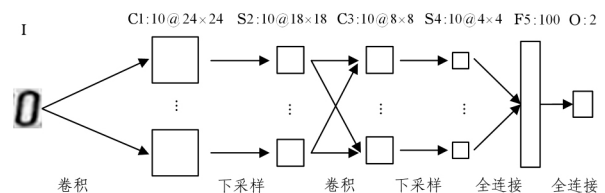


图 2 深度卷积神经网络

3 训练过程及特征的可视化

在输入层中已经定义了输入图片的大小为 28×28 ,但是获得的实验数据都是彩色图片,为了提高训练和测试的速度,首先要对图片进行灰度化处理。

深度卷积神经网络在训练模型时,通常由前向传播和反向传播两个过程构成。前向传播将二维结构的图像作为输入数据,数据在网络各层传播之后,最终获得数据的实际输出。

反向传播的主要目的是使网络更收敛。输入数据 X 在网络中传播之后,获得最后一层的特征图 Y 。将输出特征图 Y 与预设的标签 T 进行对比之后,获得两者的误差 E 。然后将误差 E 通过网络反向路径传播到网络的每个节点上,并更新对应的卷积核的权值 w 。随着网络训练的不断进行,权值的误差值随着迭代次数的增加而逐步减小,最终使该值收敛在一个稳定范围。反向传播更新权值,具体的推导过程可见文献[22-23]。

图 3 示出了输入字符分别在深度卷积神经网络的卷积层和下采样层中输出的特征图,并且特征由低层到高层变得越来越抽象。图 3 中的从左到右依次表示为输入层、卷积层 C1、下采样层 S2、卷积层 C3 和下采样层 S4。从各层提取的特征可以发现,深度卷积神经网络能有效地提取字符的特征,在底层能够获得字符的点、线段等轮廓特征,经过逐层提取之后,在高层获得抽象的语义特征。

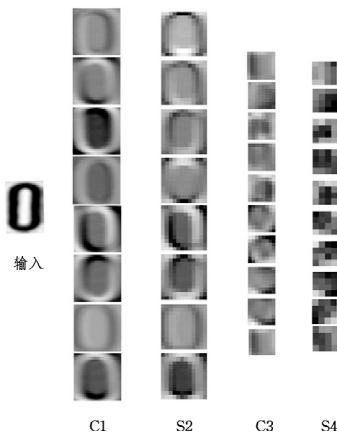


图 3 字符“0”在各层中输出的特征

4 实验分析和讨论

在收集了大量的车牌之后,将从中分割出的字符作为本文的实验数据。主要采集了在结构和形状上相似并且识别率不高的字符,如“0”与“D”,“8”与“B”。图 4 展示了部分的实验的字符图像,如某字符有噪声、某字符过度分割、某字符位置发生旋转。

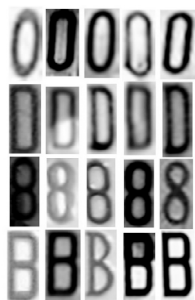


图 4 相似字符“0”与“D”,“8”和“B”

为了测试深度卷积神经网络在字符识别上的表现,将它与支持向量机(Support Vector Machine, SVM)做了对比实验。由于从车牌中分割出来的字符图像大小不一,因此需要对所有的字符进行预处理。相对于 SVM,首先将所有的样本转化为 32×32 的灰度图像并归一化。而对于卷积神经网络,统一设置所有字符的图像大小为 28×28 ,并且使字符的位置位于图像的中心,同时字符的大小相同。

使用林智仁教授的 LIBSVM^[24] 作为 SVM 的分类器,而卷积神经网络则使用图 2 所示的网络结构。为了验证不同大小的训练集对分类结果的影响,本文对每组相似字符各进行了 5 组实验,得到了如表 1 和表 2 所列的结果。

表 1 0 和 D 的正确率对比/%

训练集	测试集	SVM	卷积神经网络
1000	300	95.3	96.7
3000	600	96.0	97.8
5000	1000	95.2	98.9
7000	1500	94.7	97.5
10000	2000	94.2	98.5

表 2 8 和 B 的正确率对比/%

训练集	测试集	SVM	卷积神经网络
1000	300	97.0	97.1
3000	600	97.6	98.5
5000	1000	96.4	97.8
7000	1500	93.4	98.7
10000	2000	94.9	98.9

由表 1 和表 2 可知,当训练集的数量不断增加时,卷积神经网络的错误率下降明显,且比 SVM 的识别率更高,这表明“深度模型+大数据”的组合,对卷积神经网络识别性能的提升起了非常大的作用,因为海量的数据能使卷积神经网络学习到更丰富的内在信息。而 SVM 作为两层的浅层模型,在提取特征的过程中丢失了相似字符之间局部差异的特征。随着训练和测试样本数量逐渐上升,SVM 的准确率明显降低,这是由于 SVM 在大数据下获取图像的特征出现了过拟合的情况,在第四组“8”和“B”的字符识别中,发现许多“8”的字符被误识别成了“D”,但是对“D”字符的识别率较高,这是由于两层结构的 SVM 模型在特征表示上具有有限性。

为了验证深度卷积神经网络是否出现过拟合现象,选择“0”和“D”中训练集为 5000、测试集为 1000 的一组数据,训练误差随着实验迭代次数的上升而逐渐变小,最终该值稳定在可接受的范围内。

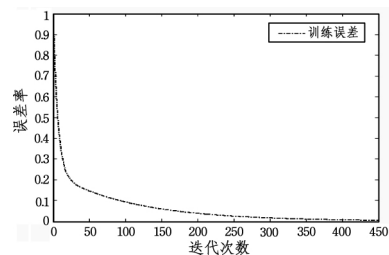


图 5 训练误差曲线

结束语 本文将深度学习应用到相似字符识别中,构建了自动学习特征提取和识别的深度卷积神经网络。通过实验对比,深度学习算法能自动学习相似字符的特征,克服了传统方法人为设计特征的缺点。在面对平移、旋转、有噪声的字符图像时,深度卷积神经网络依然能取得较好的识别效果,体现了高准确率和良好的稳定性。随着训练样本数量不断增加,

(下转第 247 页)

结束语 文中设计的改进SFS算法能够从单幅灰度影像中恢复出物体的三维形貌。从人造标准物的精度分析可以看出,与传统的线性迭代算法相比,改进的算法在保证一定的精度下执行效率高于传统算法。从实际环境基于平面拟合的精度分析可以看出,因为SFS方法都是基于理想的Lambertian漫反射模型,所以在实际环境下,重构精度较差。通过进一步改善光照模型,基于单幅影像的三维重构效果将更好,能够在真实环境下得到广泛的应用。

参考文献

- [1] HORN B K P, BROOKS M J. The variational approach to shape from shading[J]. *Computer Vision, Graphics, and Image Processing*, 1986, 33(2):174-208.
- [2] ZHANG R, TASI P S, CRYER J E, et al. Shape from Shading: a Survey [J]. *IEEE Transaction on Pattern Analysis and Machine*

Intelligence, 1999, 21(8):690-705.

- [3] 朴磊,李剑中,冯刚.从明暗恢复形状(SFS)的算法设计与精度分析[J]. *电子机械工程*, 2007, 23(6):53-56.
- [4] 郝平.基于图像灰度的自由曲面重构算法研究[D].大连:大连理工大学, 2005.
- [5] 王辉,李勇,洪霞,等.三维面行测量数据的计算全息可视化[J]. *光学学报*, 2003, 23(11):1384-1388.
- [6] 廖熠,赵荣椿.从明暗恢复形状(SFS)的几类典型算法分析与评价[J]. *中国图象图形学报*, 2001, 14(10):11-19.
- [7] 王国璋,宋玉贵.一种非朗伯表面SFS的快速粘性解算法[J]. *仪器仪表学报*, 2015, 36(7):1577-1583.
- [8] 付延安,刘海英,孟庆虎.基于明暗信息的胶囊内镜图像三维形状恢复[J]. *山东大学学报(工学版)*, 2012, 42(6):63-68.
- [9] 郭便,魏宏波.基于明暗恢复形状的加工表面形貌重构与粗糙度检测[J]. *工具技术*, 2011, 45(6):98-102.

(上接第231页)

深度卷积神经网络的准确率逐渐提高,表明通过大数据训练能学习到更丰富的内在特征信息。

但是与支持向量机相比,深度学习由于网络层次复杂,在识别上需要花费更多的时间。对于包含多个字符的车票,考虑到识别实时性,在后续研究中可以把支持向量机和深度识别相结合,先采用支持向量机进行第一层识别,然后对混淆字符进行二次识别,从而保证识别的实时性。另外也可以考虑采用深度学习进行车牌检测的应用研究。

参考文献

- [1] 马俊莉,莫玉龙,王明祥.一种基于改进模板匹配的车牌字符识别方法[J]. *小型微型计算机系统*, 2003, 24(9):1670-1672.
- [2] 张宏烈.支持向量机在字符识别中的应用研究[J]. *微计算机信息*, 2006, 22(4Z):245-247.
- [3] 张可,张高燕,吴苏,等.基于BP神经网络的字符识别系统[J]. *计算机与现代化*, 2009(1):63-65.
- [4] 谢永祥,董兰芳.复杂背景下基于HSV空间和模板匹配的车牌识别方法研究[J]. *图学学报*, 2014, 35(4):585-589.
- [5] 洪涛,梁伟建,卢玉凤.标牌粘连字符自适应定位分割重建与识别[J]. *中国图象图形学报*, 2014, 19(6):886-895.
- [6] 费继友,谢金路,李花,等.基于字符特征约束的自适应车牌校正提取[J]. *仪器仪表学报*, 2016, 37(3):632-639.
- [7] BENGIO Y. Learning Deep Architectures for AI[J]. *Foundations & Trends® in Machine Learning*, 2009, 2(1):1-127.
- [8] CUN Y L, BOSER B, DENKER J S, et al. Handwritten digit recognition with a back-propagation network[J]// *Advances in Neural Information Processing Systems*, 1989, 2:396-404.
- [9] HINTON G E, SALAKHUTDINOV R R. Reducing the Dimensionality of Data with Neural Networks[J]. *Science*, 2006, 313(5786):504-507.
- [10] DAHL J V, KOCH K C, KLEINHANS E, et al. Convolutional networks and applications in vision [C]// *IEEE International Symposium on Circuits & Systems*. IEEE, 2010:253-256.
- [11] LECUN Y, BOSER B, DENKER J S, et al. Backpropagation applied to handwritten zip code recognition[J]. *Neural Computation*, 1989, 1(4):541-551.
- [12] LÉCUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. *Proceedings of the*

IEEE, 1998, 86(11):2278-2324.

- [13] SERMANET P, CHINTALA S, LECUN Y. Convolutional neural networks applied to house numbers digit classification[C]// *International Conference on Pattern Recognition*. IEEE, 2012:3288-3291.
- [14] DENG H, STATHOPOULS G, SUEN C Y. Error-Correcting Output Coding for the Convolutional Neural Network for Optical Character Recognition [C] // *International Conference on Document Analysis and Recognition*. IEEE Computer Society, 2009:581-585.
- [15] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet Classification with Deep Convolutional Neural Networks[J]. *Advances in Neural Information Processing Systems*, 2012, 25(2):1097-1105.
- [16] DENG J, DONG W, SOCHER R, et al. ImageNet: A large-scale hierarchical image database[C]// *Computer Vision and Pattern Recognition*, 2009. IEEE, 2009:248-255.
- [17] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet Large Scale Visual Recognition Challenge[J]. *International Journal of Computer Vision*, 2015, 115(3):1-42.
- [18] HUBEL D H, WIESEL T N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex[J]. *Journal of Physiology*, 1962, 160(1):106-154.
- [19] GLOROT X, BENGIO Y. Understanding the difficulty of training deep feedforward neural networks[J]. *Journal of Machine Learning Research*, 2010, 9:249-256.
- [20] DAHL G E, SAINATH T N, HINTON G E. Improving deep neural networks for LVCSR using rectified linear units and dropout [C] // *IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013:8609-8613.
- [21] GLOROT X, BORDES A, BENGIO Y. Deep Sparse Rectifier Networks[C] // *Proceeding of the 14th International Conference on Artificial Intelligence and Statistics*. 2011:315-323.
- [22] HINTON G E. How neural networks learn from experience[J]. *Scientific American*, 1992, 267(3):144-151.
- [23] BOURIE J. Notes on Convolutional Neural Networks[R]. Massachusetts: Center for Biological and Computational Learning, 2006:38-44.
- [24] CHANG C, LIN C, et al. LIBSVM: A library for support vector machines[J]. *Acm Transactions on Intelligent Systems & Technology*, 2011, 2(3):389-396.