

doi: 10.3969/j.issn.0490-6756.2017.05.013

基于 Caffe 深度学习框架的车牌 数字字符识别算法研究

欧先锋^{1,2}, 向灿群^{1,2}, 郭龙源^{1,2}, 涂兵^{1,2}, 吴健辉^{1,2}, 张国云^{1,2}

(1. 湖南理工学院信息与通信工程学院复杂系统优化与控制湖南省普通高等学校重点实验室, 岳阳 414006;
2. 湖南理工学院异构计算及应用研究中心, 岳阳 414006)

摘要: 在车牌字符识别的某些场合中, 获得的字符通常存在切割不均匀、光照对比度强烈、遮挡严重等强噪声污染。针对被强噪声污染的数字字符, 提出一种基于 Caffe 深度学习框架的字符识别算法, 在 Caffe 框架下搭建卷积神经网络, 并对网络参数训练获得了一个鲁棒性强、识别精度高的网络结构。实验结果表明, 在低噪声、中度噪声、强噪声污染情况下, 文章中提出的方法相比当前典型的识别方法, 在数字字符识别上均具有较好的识别能力, 平均识别率高出将近 5%, 而在强噪声污染情况下, 识别效果具有更加明显的优势。

关键词: Caffe 框架; 车牌字符识别; 深度学习; 卷积神经网络

中图分类号: TP391.41 **文献标识码:** A **文章编号:** 0490-6756(2017)05-0971-07

Research of recognition of digital characters on vehicle license based on caffe deep learning framework

OU Xian-Feng^{1,2}, XIANG Can-Qun^{1,2}, GUO Long-Yuan^{1,2},
TU Bing^{1,2}, WU Jian-Hui^{1,2}, ZHANG Guo-Yun^{1,2}

(1. Key Laboratory of Optimization & Control for Complex Systems, College of Information & Communication Engineering, Yueyang 414006, China; 2. Research Center for Heterogeneous Computing & It's Application, Hunan Institute of Science & Technology, Yueyang 414006, China)

Abstract: In some license plate character recognition occasions, strong noises such as uneven cutting, strong illumination contrast and occlusion are inevitable. To solve this problem effectively, a character recognition algorithm based on Caffe deep learning framework was proposed in this paper. The convolutional neural network was built in the Caffe framework, and the network parameters were trained to obtain a network structure with high accuracy and robustness. Experiment results showed that the proposed algorithm has obvious advantages, the average recognition rate was about 5% higher when compared with the traditional digital character recognition method. And when there are strong noise pollution exists, the recognition results are better.

Keywords: Caffe framework; License plate character recognition; Deep learning; Convolutional neural network

收稿日期: 2016-12-21

基金项目: 湖南省自然科学基金项目(2017JJ3099, 2016JJ2064); 湖南省科技计划项目(2016TP1021); 湖南省研究生创新项目(CX2016B670); 湖南省教育厅科学研究项目(16C0723)

作者简介: 欧先锋(1983-), 男, 湖南郴州人, 博士生, 研究方向为图像处理、视频编码。E-mail: ouxf@hnist.edu.cn

通讯作者: 郭龙源。E-mail: guolongyuan@hnist.edu.cn

1 引言

车牌字符识别是公共交通、安全防护等领域的一项重要技术,具有重要的实际应用价值.近年来,国内外广大科研人员围绕车牌字符识别算法开展了广泛的研究,针对车牌上的字符提出了一些有针对性的算法,取得了一定的识别效果.在基于非机器学习方法上,文献[1]提出了基于评分模型的车牌字符识别方法.将待识别的字符进行分区,然后对分区的各个方格进行评分,采用类似模板匹配的方式,最终得分高的字符模板作为识别结果;针对当前大多数基于单一特征提取方法在车牌字符识别上的局限性,文献[2]提出了一种具有一定的鲁棒性的车牌字符多特征提取方法;文献[3]将尺度不变特征应用到车牌识别系统中来,有效的改善了不同环境和不同光照下的车牌识别效果;文献[4]提出基于相似聚类算法的识别方法.上述方法特点是在实际的应用过程中,不需要进行训练,然而也存在易受复杂环境中噪声等因素的干扰,且单个字符识别时间较长等问题.而在基于机器学习的识别方法上,文献[5]对支持向量机(SVM)分类器中的距离函数,拉格朗日函数以及二次凸优化函数进行了详细阐述,对 SVM 理论中典型的线性最优化二分类问题的求解进行了完整的推导,并将 SVM 运用于分类识别当中;文献[6]采用 BP 神经网络联合模板匹配进行车牌识别算法.这些方法针对特定的车牌字符,都能够取得比较好的识别效果,但是容易受到训练样本数量的影响.针对当前车牌字符识别中存在的这些典型问题,本文基于 Caffe 深度学习框架提出了一种车牌数字字符识别方法,在 Caffe 深度学习框架基础上搭建卷积神经网络来实现车牌字符识别.实验结果表明,文章中提出的方法具有很好的识别效果,尤其是在有噪声干扰的复杂环境下,该方法具有很强的鲁棒性,对车牌数字字符的识别率仍然能够达到 95% 以上.上述优点使得该方法在实际应用中具有较好的潜在应用价值.

2 Caffe 框架

近年来,几乎整个智能科学的研究者们都注意到一个技术名词——深度学习(Deep Learning).随着深度学习技术在学术界和工业界得到广泛认可,科研人员开发出了众多的深度学习开源框架,例如 Caffe^[7]、Theano、MXNet^[8]、Tensorflow^[9]、

Torch 等,Caffe 是其中较为成熟和完善的一类深度学习框架. Caffe 是由伯克利视觉和学习中心(BVLC)开发的基于 C++/CUDA/Python 实现的框架,它实现了前馈卷积神经网络架构,Caffe 框架以‘层’为单位对深度神经网络的机构进行了高度的抽象,通过一些精巧的设计显著优化了执行效率,并且在保持高效实现的基础上不失灵活性.因为利用了 MKL、OpenBLAS、cuBLAS 等计算库,支持 GPU 加速,它在处理海量数据时速度快;Caffe 非常适合做二维图像数据的特征提取,在图像处理过程中能够实现可视化操作;除此之外,Caffe 还提供一整套工具集,能够用于模型训练、预测、微调、发布、数据预处理以及良好的自动测试. Caffe 框架的发布极大的降低了深度学习研究和发展的难度^[10].

3 算法设计

3.1 车牌识别系统

车牌识别的过程中主要包含四个步骤:1) 车牌定位,对一幅图像中的车牌进行定位;2) 车牌提取,将定位成功后的车牌从原始图像中提取出来;3) 字符粗分割,将提取的车牌切割成单个字符;4) 字符识别,对切割出来的单个字符用基于 Caffe 的深度学习算法进行识别并得出最终识别结果.车牌识别的过程流程图如图 1 所示,为了方便描述,将预处理、牌照定位和字符粗分割统称为车牌图像预处理.



图 1 车牌识别过程流程图

Fig. 1 Flowchart of license plate recognition

车牌图像预处理只需得到分割后的单个数字字符,精度要求不需要很高,因此采用简单的垂直投影法进行快速粗分割便可得到符合识别要求的单个字符图.

3.2 卷积神经网络

卷积神经网络是基于动物视觉感受域设计而成,由卷积层、池化层和其它层构成.与传统的 BP 神经网络主要区别在于卷积神经网络采用了卷积核权值共享的方式代替了 BP 中的全连接方式,使得训练参数数量级的减少;池化层操作把输入层某一块的数据压缩成一个值,能进一步压缩数据;其

它层如归一化层能够模拟人脑的抑制作用,用在特殊场合有额外的效果. 卷积神经网络的训练同样采用 BP 算法.

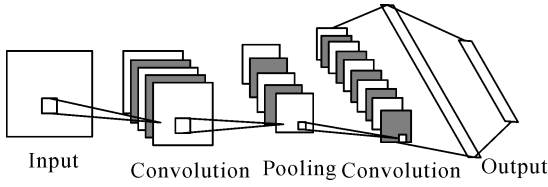


图 2 卷积神经网络
Fig. 2 Convolutional neural network

3.2.1 前向传播 如图 2 所示,典型的卷积神经网络中,包含了卷积层、采样层、全连接层. 在前向传播中,卷积层的上一层特征图被一个可学习的卷积核进行卷积,然后通过一个激活函数,就得到输出特征图,输出特征图的个数由卷积核的个数决定. 计算过程为

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

式(1)中, M_j 表示被卷积核 k_{ij}^l 作用的输入特征图集合; x_i^{l-1} 为 M_j 集合中的第 i 个输入特征图; b_j^l 为卷积核 k_{ij}^l 的偏置; $f(\cdot)$ 为激活函数,本文采用 ReLU 作为激活函数^[11]; x_j^l 为卷积输出特征图.

采样层可视为特殊卷积层,采样核根据算法需要而设定,不能学习,采样核对每个输入特征图一作用生成对应的输出特征图. 因此,输出特征图个数与输入特征图个数一致,但数据得到压缩,鲁棒性增强. 计算公式如下.

$$x_j^l = f(\beta_j^l \text{down}(x_j^{l-1}) + b_j^l) \quad (2)$$

式(2)中, $\text{down}(\cdot)$ 表示采样核函数,一般有均值采样与最大值采样两种采样函数,本文使用最大值采样; β_j^l 和 b_j^l 为每个输出特征图对应的乘性偏置和加性偏置,可在训练过程中学习, x_j^l 为采样输出特征图.

全连接层的前向传递与 BP 神经网络算法大同小异,在计算前需将输入特征图从二维图展开至一维向量,然后进行全连接计算. 计算公式为

$$y_j^l = f(W_{ij}^l \cdot x_i^{l-1} + b_j^l) \quad (3)$$

式(3)中, W_{ij}^l 为全连接权值; b_j^l 为全连接偏置,通过训练自动调节; y_j^l 为输出特征向量.

卷积神经网络在图像分类中具有突出表现,本文采用了 Softmax 作为分类损失函数^[12],简单而高效,更支持多分类. 在前向传播中,损失值计算公式如式(6).

$$y_j = y_i - \max(y_1, \dots, y_n) \quad (4)$$

$$p_i = \frac{e^{y_i}}{\sum_{j=1}^k e^{y_j}} \quad (5)$$

$$L(\theta) = -\frac{1}{m} [\sum_{i=1}^m \sum_{j=1}^k I\{y^{(i)} = j\} \log p_i] \quad (6)$$

式(4)中, y_i 为卷积神经网络最后一层节点的输出,该式保证了输出的最大值为 0; 式(5)对输出值进行归一化,并得出了每个输出节点的概率值; 式(6)为 Softmax 的代价函数, j 为 $y^{(i)}$ 对应的类别标签, $I\{\cdot\}$ 为示性函数,即当 $y^{(i)} = j$ 时输出为 1,其它输出为 0, m 为 batchsize 个数,最后求平均损失, θ 表示需要优化的参数,即 $k_{ij}^l, \beta_j^l, W_{ij}^l, b_j^l$ 等.

3.2.2 反向传播 网络各层参数利用反向传播获得的梯度进行调节,卷积层、采样层、损失层通过链式求导法则反向逐层计算梯度. 损失层梯度计算公式如式(7).

$$\nabla y_i^l = \frac{\partial L(\theta)}{\partial y_i} = \frac{\partial L}{\partial p_i} \cdot \frac{\partial p_i}{\partial y_i} = \sum_{l=1}^m (I\{y^{(i)} = l\} - p_i) \quad (7)$$

式(7)中, L 为最后一层输出层,全连接层梯度计算公式如下.

$$\nabla w_j^l = \frac{\partial L(\theta)}{\partial w_j^l} = \nabla y_j^l \cdot (y_j^l > 0) \cdot x_i^{l-1} \quad (8)$$

$$\nabla b_j^l = \nabla y_j^l \cdot (y_j^l > 0) \quad (9)$$

其中, ∇y_j^l 为第 l 层输出的梯度(下同),由于本文使用 ReLU 激活函数,梯度反向传导时需乘上 $(y_j^l > 0)$,即当第 l 层输出大于 0 时,梯度继续反馈,当输出小于 0 时梯度停止反馈. 采样层梯度计算如式(10).

$$\nabla \beta_j^l = \frac{\partial L}{\partial \beta_j^l} = \nabla y_j^l \cdot \text{up}(x_j^l) \quad (10)$$

式(10)中, $\text{up}(\cdot)$ 为下采样的反向操作; x_j^l 为 l 层的输出特征图,若层后全连接层,先将输出向量重构成特征图再进行梯度计算,偏置 b_j^l 的梯度计算与式(9)相同. 由此可以得到卷积层梯度为

$$\nabla k_{ij}^l = \frac{\partial L}{\partial k_{ij}^l} = \sum_{u,v} (\nabla y_j^l)_{uv} (y_i^{l-1})_{uv} \quad (11)$$

式(11)中, y_i^{l-1} 表示与卷积核 k_{ij}^l 作用了的输入特征图; ∇y_j^l 为与卷积核 k_{ij}^l 作用后生成的特征图的梯度. u, v 表示与卷积核 k_{ij}^l 有关的输入输出特征图集合.

获得各层网络参数的梯度后再逐层对所有参数进行更新,以完成网络的学习过程.

3.3 网络结构

文章中所使用的 Caffe 深度学习网络结构如图 3 所示,该结构在 LeNet 网络结构^[13]基础上做了调整. 网络的输入数据为 24×40 的彩色字符图片以及对应的标签,第一层为卷积层卷积核大小为

5×5,卷积步长为 1,经过第一层卷积层后,网络输出图片尺寸为 20×36;第二层为池化层,采用最大值池化,池化核大小为 2×2,步长为 2,经过池化后网络输出图片尺寸为 10×18;第三层为卷积层,卷积参数与第一层卷积参数相同,该层输出图片的尺寸为 6×14;第四层为池化层,相应参数与第二层池化层参数相同,不同的是该层采用均值池化,其输出图片尺寸为 3×7;第五层为全连接层,有 400 个神经元节点,与第四层输出神经元进行全连接,并通过 ReLU 激活函数输出;第六层为全连接层,采用 Softmax 作为损失函数输出,该层有 10 个输出神经元对应 10 类标签。

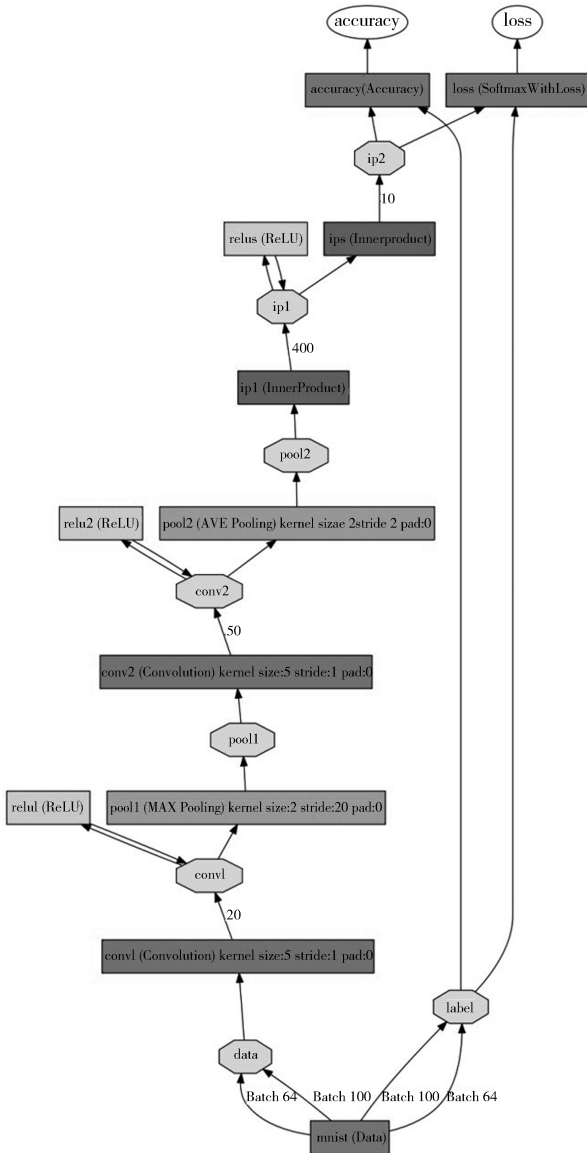


图 3 网络结构图
Fig. 3 Network structure

在该网络结构中,第一层卷积层的输出为 20 个特征图,第二层卷积层的输出为 50 个特征图.两层最

大池化层能够去除噪声,增加网络的鲁棒性,relu 层加速网络收敛,accuracy 层用于测试时计算测试精度,loss 层用于计算训练过程中的损失,并计算误差梯度。

4 仿真实验与分析

4.1 实验条件

本文的实验环境硬件平台为 Dell 5500 图形工作站,其基本配置为 Intel Core i7-4710MQ 四核处理器、NVIDIA GT940 2 G 独立显卡、8 G DDR3 内存、128 G 固态硬盘;所使用的软件仿真平台为 Matlab R2016a 以及 Caffe-windows 版本. 实验中通过将本文方法与相关文献中的车牌字符识别算法进行详细地对比与分析来说明本文提出的方法的有效性。

4.2 数据生成

本文所用数据集是从不同光照强度、倾斜角度、遮挡程度等条件下拍摄的 2000 副彩色车牌中对数字字符进行粗切割产生,生成了 1303 个‘0’, 1194 个‘1’,1165 个‘2’,944 个‘3’,728 个‘4’,924 个‘5’,985 个‘6’,743 个‘7’,1019 个‘8’,825 个‘9’. 切割出的图片未经任何处理,而且每幅图片中包含不同强度的噪声,如图 4 所示。

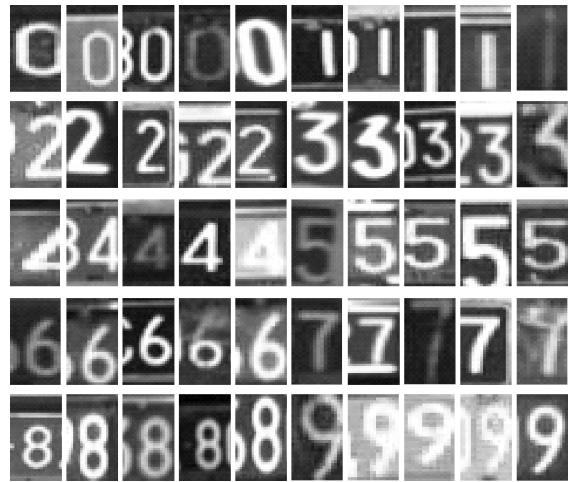


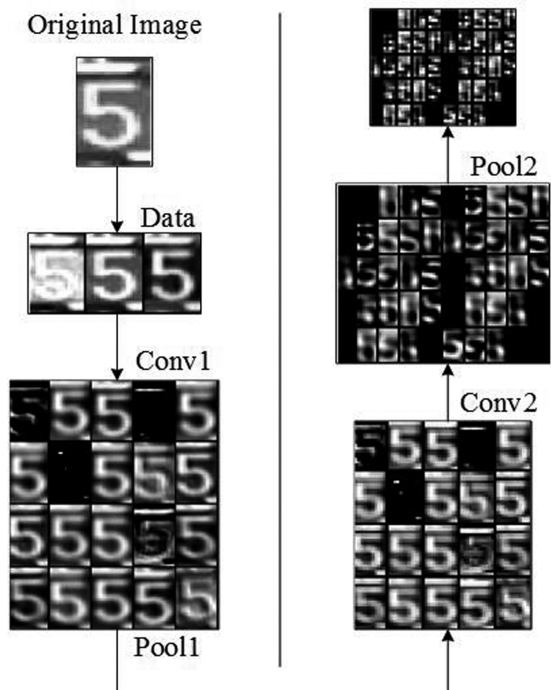
图 4 部分粗切割字符
Fig. 4 Partial character by rough segmentation

为了确保各个字符数量均衡,针对数量较少的字符额外的增加了它们的数量,使得每个字符的数量为 1300 个,并且对这 10 个字符分别贴上 0~9 的标签. 从每个字符中随机抽取 1000 张图片组成训练集,每个字符剩余的 300 张图片组成测试集. 并将训练集和测试集转换为 Caffe 所支持的 lmdb 或 leveldb 格式,转换过程中将图片归一化为 2440。

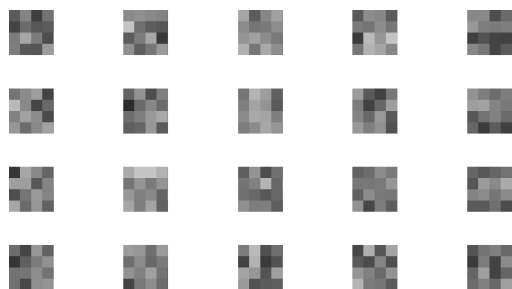
4.3 网络各层变化

测试图片通过网络的不同层生成对应的特征

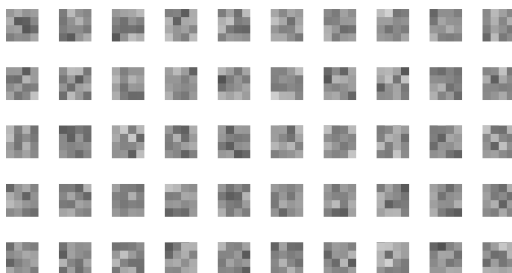
图, 本文使用 Matlab 将各层的权值及特征图提取出来, 如图 5 所示, 获得输入原始图片在整个网络中各层的变化过程.



(a) 测试样例在网络中逐层变化
(a) Test samples change layer by layer in the network



(b) conv1 卷积核
(b) convolution kernel of layer 1



(c) conv2 卷积核
(c) convolution kernel of layer 2

图 5 网络可视化

Fig. 5 Network visualization

图 5(b)和(c)分别是第一层卷积层的 20 个三维卷积核和第二层卷积层的 50 个三维卷积核, 以

彩色形式显示. 图 5(a)以被强噪声污染的字符‘5’为例, 描述了该字符在各层网络中的变化情况, 由图 5(a)可知, 层次越深, 噪声能够被滤除的也将越多, 使得整个方法具有很强的鲁棒性, 同时, 每个生成的特征图能够自动提取输入图片的不同特征. 字符‘5’的最终识别结果如图 6 所示. 图 6 中共有 10 个类别, 分别代表 0~9 字符, 第 6 个类别相似度最高即代表识别结果为字符‘5’.

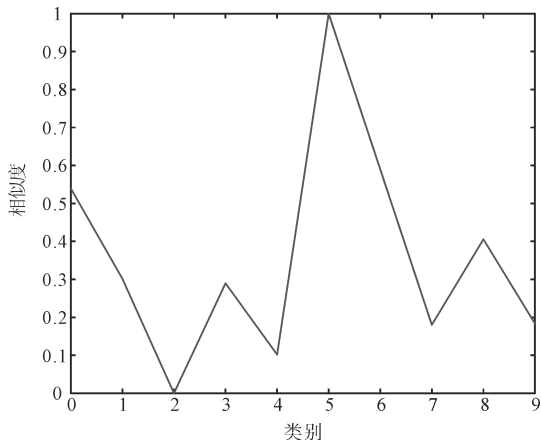


图 6 分类结果
Fig. 6 Classification result

4.4 训练曲线

对训练数据集进行训练, 并且训练集每迭代 500 次时就在测试集中随机抽取的 100 个测试样本进行测试, 并记录测试样本的损失值, 经过 2×10^4 次迭代后训练样本损失趋于 0, 测试样本损失持续减小, 如图 7 所示.

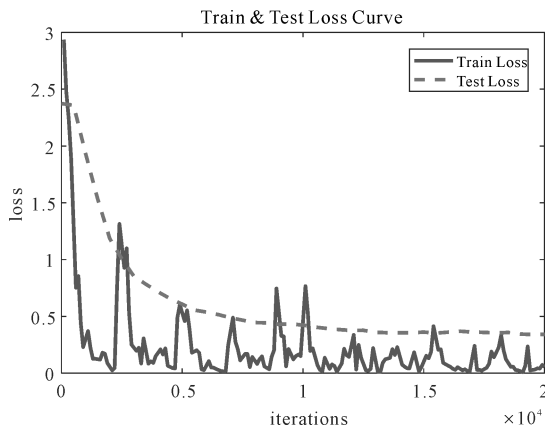


图 7 损失曲线
Fig. 7 Loss curve

Train loss 与 Test loss 不断下降, 网络处于学习状态, 能够训练出合理的网络; Train loss 下降 Test loss 不变, 说明网络过拟合; Train loss 与 Test loss 均不变, 说明网络学习遇到瓶颈, 需要调

整学习率或其它参数. Caffe 能够通过这些损失值的变化及时发现训练中的问题,降低训练成本.

4.5 识别结果与分析

我们对数字样本进行了筛选,将样本分为低噪声、中度噪声、强噪声三类,部分样本如图 8 所示.

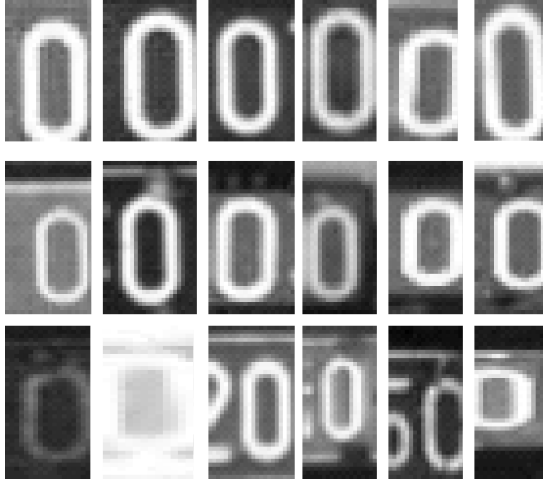


图 8 不同程度噪声干扰图片

Fig. 8 Different degrees of noise pictures

图 8 中以‘0’字符为例,第一行图片为低噪声污染图片,在光照,形状上有少许干扰,能够清楚的识别图片中的字符;第二行图片为中度污染图片,有明显的大小形状,杂物等干扰,但仍然能够识别图片中的字符;第三行图片为强噪声污染图片,光照范围变化剧烈、遮挡严重、干扰物质比重高,不容易识别图片中的字符.每类测试样本包含 500 张图片,其中字符 0~9 各有 50 张.分别用本文算法、文献算法^[14-17]对三类数字字符进行识别,识别结果如表 1 所示.

表 1 不同算法识别率比较

Tab. 1 The comparison of identification rate using different criteria

噪声强度	文献[14]	文献[15]	文献[16]	文献[17]	本文算法
低噪声	98.0±2.0	98.4±1.2	96.4±1.0	96.1±2.3	99.2±0.3
中等噪声	95.0±4.0	87.2±2.0	95.6±1.4	90.3±2.0	97.6±0.4
强噪声	88.0±4.0	56.5±3.4	91.2±2.6	67.7±2.6	93.8±0.6

由表 1 可知,本文提出的基于 Caffe 深度学习框架的车牌数字字符识别方法对不同噪声水平污染情况下的字符识别均具有较高的识别率,并且识别率波动范围小,尤其在被强噪声污染情况下的车牌字符识别中,该方法在识别率上明显高于其它算法,具有较高鲁棒性.

5 结 论

随着技术的发展,越来越多的深度学习开源框架展现在我们面前,这些工具不仅性能优异而且使用便捷.本文将深度学习开源框架 Caffe 应用于车牌数字字符识别中,不仅在识别精度上有所提升,而且具有很强的鲁棒性.限于样本库的缺乏,暂时无法实现对强噪声污染的字母及汉字字符的识别,这将是接下来的重点工作之一.

参考文献:

- [1] 莫林, 凌文彪, 张福元, 等. 基于评分模型的车牌字符识别方法[J]. 计算机应用与软件, 2015, 32: 203.
- [2] 何兆成, 余锡伟, 余文进, 等. 字符多特征提取方法及其在车牌识别中的应用[J]. 计算机工程与应用, 2011, 47: 228.
- [3] Zahedi M, Salehi S M. License plate recognition system based on SIFT features[J]. ProcComput Sci, 2011, 3: 998.
- [4] 韩春燕, 刘玉娇, 琚生根, 等. 中文微博命名体识别[J]. 四川大学学报: 自然科学版, 2015, 52: 511.
- [5] 周正松, 李瑶, 陶德元. 支持向量机的凸优化求解[J]. 四川大学学报: 自然科学版 2016, 53: 781.
- [6] Dong J, Sun M, Liang G, *et al.* The Improved neural network algorithm of license plate recognition[J]. Int J Signal Proc, 2015, 8: 49.
- [7] Jia Y Q, Shelhamer E, Donahue J, *et al.* Caffe: convolutional architecture for fast feature embedding[J]. Eprint Arxiv, 2014: 675.
- [8] Athey S, Imbens G. MXNet: a flexible and efficient machine learning library for heterogeneous distributed systems[J]. Statistics, 2015(1): 1.
- [9] Abadi M, Agarwal A, Barham P, *et al.* Tensorflow: large-scale machine learning on heterogeneous distributed systems[J]. 2016: 1.
- [10] 赵永科. 深度学习: 21 天实战 Caffe[M]. 北京: 电子工业出版社, 2016.
- [11] Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks[C]//Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS) 2011, Fort Lauderdale, USA JMLR: W&CP 15, 2011.
- [12] Kabani A W, El-Sakka M R. Object detection and localization using deep convolutional networks with softmax activation and multi-class log loss[M]/Ber-

- lin Heidelberg: Springer, 2016.
- [13] Lecun Y, Bottou L, Bengio Y, *et al.* Gradient-based learning applied to document recognition[J]. Proc IEEE, 1998, 86: 2278.
- [14] 李消霞, 张文静, 刘银萍, 等. 一种改进的 BP 神经网络车牌识别算法的研究[J]. 电脑知识与技术: 学术交流, 2016, 12: 223.
- [15] 王忠飞, 陈元正. 基于轮廓特征的车牌英文和数字识别方法[J]. 浙江工业大学学报, 2015, 43: 522.
- [16] Meng Q Y, Hu H, Bai Y. A method of character recognition based on general characteristic and connected regions[C]//Proceedings of the International Conference on Multimedia and Signal Processing. [s. l.]:IEEE Computer Society, 2011.
- [17] 宋加涛, 刘济林. 车辆牌照上英文和数字字符的结构特征分析及提取[J]. 中国图象图形学报, 2002, 7: 945.