

DOI:10.16644/j.cnki.cn33-1094/tp.2022.11.004

一种改进 IE-AlexNet 的少数民族文字图像识别方法*

杨秀璋¹, 周既松¹, 武 帅^{1,2}, 陈登建¹, 刘建义¹, 宋籍文³

(1. 贵州财经大学信息学院, 贵州 贵阳 550025; 2. 涟水县财政局; 3. 贵州高速公路集团有限公司)

摘 要: 提出一种融合自适应图像增强的 IE-AlexNet 少数民族文字图像识别方法。以阿拉伯文字为例, 计算各场景的图像噪声阈值, 利用六种方法实现图像增强, 然后构建一种改进 AlexNet 的神经网络模型, 实现阿拉伯文字识别。实验结果表明, 所提方法能有效检测复杂环境下的少数民族文字, 其 F1 值为 0.9559, 准确率为 0.9553, 均优于其他模型。同时, 对比实验突显图像增强的有效性, 具有一定应用价值。

关键词: 少数民族文字; 深度学习; 图像识别; 数字人文

中图分类号: TP391

文献标识码: A

文章编号: 1006-8228(2022)11-15-06

Research on image recognition of minority characters based on improved IE-AlexNet

Yang Xiuzhang¹, Zhou Jisong¹, Wu Shuai^{1,2}, Chen Dengjian¹, Liu Jianyi¹, Song Jiwen³

(1. School of Information of Guizhou University of Finance and Economics, Guiyang, Guizhou 550025, China;

2. Lianshui County Finance Bureau; 3. Guizhou Expressway Group Co., Ltd.)

Abstract: In this paper, an IE-AlexNet minority character image recognition method with adaptive image enhancement is proposed. Taking Arabic characters as an example, the image noise threshold of each scene is calculated, and six methods are used to realize image enhancement. An improved AlexNet neural network model is constructed to realize Arabic character recognition. Experimental results show that the proposed method can effectively detect minority characters in complex environments with an F1-score of 0.9559 and an accuracy rate of 0.9553, both of which are better than other models. The contrast experiment highlights the effectiveness of image enhancement, which has certain application value.

Key words: minority characters; deep learning; image recognition; digital humanities

0 引言

随着文本信息数字化程度不断提升, 模式识别技术被广泛运用于档案信息化应用中^[1]。部分档案由于以手写体为主, 需要对其进行手写体字符识别 (Optical Character Recognition, 简称 OCR), 以获取较为准确的文字信息^[2]。整个过程主要包括图像预处理、特征提取和分类器分类, 其中特征提取最为关键。传统手写体字符识别方法在提取目标书写体图像信息时, 能较好地提取手写体空间、字体轮廓信息, 但对

复杂环境下手写体图像的处理存在准确率低和识别效果不佳的现象^[3]。此外, 当前少数民族古文字研究集中于艺术考究和字形释义, 缺乏利用深度学习技术自动化识别文字, 且古文字主要通过古籍、雕刻、木刻或碑刻存在, 存在大量噪声, 数字化读取困难^[3]。针对上述情况, 本文提出一种改进 IE-AlexNet 神经网络的少数民族文字图像识别方法, 一定程度上提升了对复杂环境下手写体图像的关键特征提取效果, 提高模型整体分类效果。该方法有较好的鲁棒性和准确性。

收稿日期: 2022-04-11

*基金项目: 贵州省科技计划项目 (黔科合基础 [2019] 1041, 黔科合基础 [2020] 1Y279); 贵州省教育厅青年科技人才成长项目 (黔教合 KY 字 [2021] 135); 贵州财经大学 2021 年度校级项目 (No. 2021KYQN03)

作者简介: 杨秀璋 (1991-), 男, 苗族, 贵州凯里人, 博士生, 助教, 主要研究方向: Web 数据挖掘和人工智能。

通讯作者: 武帅 (1994-), 男, 汉族, 江苏淮安人, 硕士生, 助教, 主要研究方向: 数据分析、图书情报。

1 相关研究现状

现阶段,手写体字符识别的研究主要集中于对中文和英文手写体文字的识别,但针对少数民族古文字手写体识别研究相对较少。传统手写体字符识别方法在识别少数民族古文字手写体时,由于其手写体字型变化,数字化识别效果欠佳。因此,如何运用计算机视觉技术和档案数字化方法构建一个能够智能准确识别少数民族古文字的模型,一定程度上对提高少数民族档案数字化建设起到积极作用,具有较高的研究意义,是研究者们迫切需要解决的问题。

热依曼·吐尔逊等^[4]融合高斯模型和隐马尔可夫模型,设计了一种维吾尔语联机手写体识别系统,利用高斯模型模拟维吾尔语整词的静态特征和隐马尔可夫模型模拟书写笔迹的动态特征。王晓娟等^[5]针对数字手写体图像区域进行归一化处理,提出一种基于BP神经网络的图像识别方法。杨秀璋等^[6]针对水族古文字传承以刺绣、碑刻、木刻等为主,文字清晰度不高,数字化读取困难的现象,提出一种基于自适应图像增强和区域检测的水族文字提取与分割算法。姜文等^[7]针对手写体维吾尔文字字符识别过程中的特征提取环节,提出一种基于方向线素特征的手写体维吾尔文单字字符笔迹特征的KNN分类识别算法。杨秀璋等^[8]考虑到古文字的字形变化特点,提出一种改进卷积神经网络的阿拉伯文字图像识别方法。

此外,深度学习技术的不断成熟,为手写体字符识别提供了新的解决思路。本文在AlexNet神经网络的基础上融合图像增强技术,一定程度上提升了对阿

拉伯文字手写体关键特征的提取效果,提高了模型整体分类效果。将本文方法运用于手写体阿拉伯文字的识别,可以一定程度上拓宽手写体文字识别的研究范畴,同时也为本研究团队后期对水族文字识别提供理论基础,给少数民族数字化档案建设提供实际应用的可能性,这是我们研究的意义所在。

2 系统设计

本文提出一种融合自适应图像增强和深度学习的IE-AlexNet(Image Enhancement AlexNet)模型,并对复杂环境下的阿拉伯文字图像数据集进行实验,模型设计的研究内容如下。

2.1 总体框架

本文设计并实现了IE-AlexNet模型,其总体框架如图1所示。具体实现步骤如下。

- (1) 模拟现实场景采集并构造阿拉伯文字图像数据集,并将其随机划分为训练集和测试集。
- (2) 计算场景系数,构建自适应图像增强得算法实现去噪,增强图像质量,共包括形态学去噪(闭运算)、灰度伽马变换、灰度对数变换、直方图均衡化、ACE算法去雾和暗通道先验去雾六种算法。
- (3) 对所有图像进行预处理,包括图像向量表征、标准化处理、形状修改和编码转换等。
- (4) 构建IE-AlexNet模型该模型包括5层卷积层、3层池化层和3个全连接层,并优化模型超参数。
- (5) 通过构建Softmax分类器实现阿拉伯文字图像识别,对比少数民族文字识别的性能。

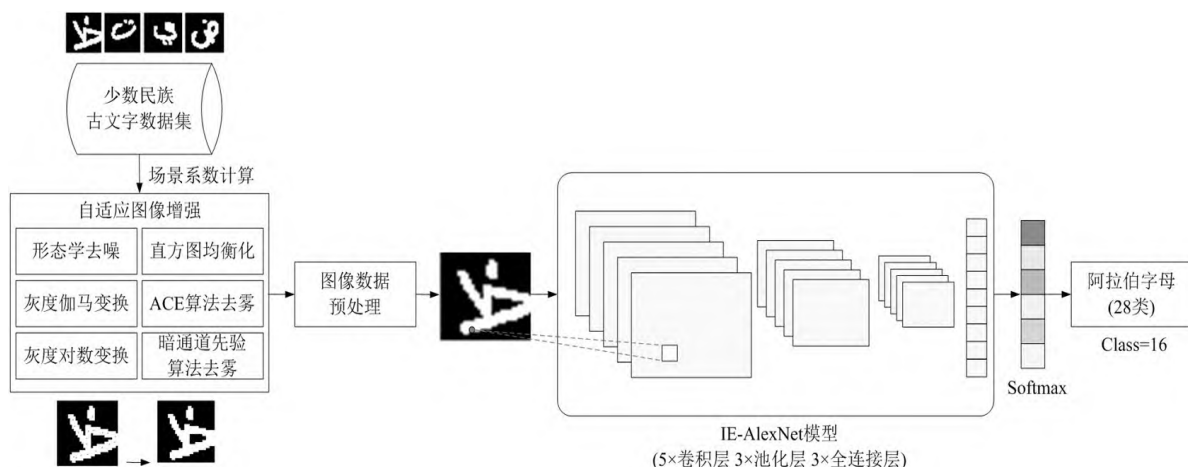


图1 复杂环境下少数民族文字图像识别的总体框架

2.2 AlexNet 模型

AlexNet 神经网络是由 Alex Krizhevsky 等人^[9]于 2012 年提出的首个应用于图像分类的深层卷积神经网络,并在当年举办的 ImageNet Large Scale Visual Recognition Competition (简称 ILSVC) 比赛中以 15.3% 的 top-5 测试错误率获得了分类任务的冠军,其

网络结构详如图 2 所示。该网络由五个卷积层和三个全连接层组成。卷积层用于提取特征,池化层用于实现特征降维,激活函数用于获取非线性特征,全连接层起到分类作用。输入图像经过卷积层特征提取和全连接层分类操作之后,输入到具有 1000 个节点的 Softmax 分类器中实现图像分类。

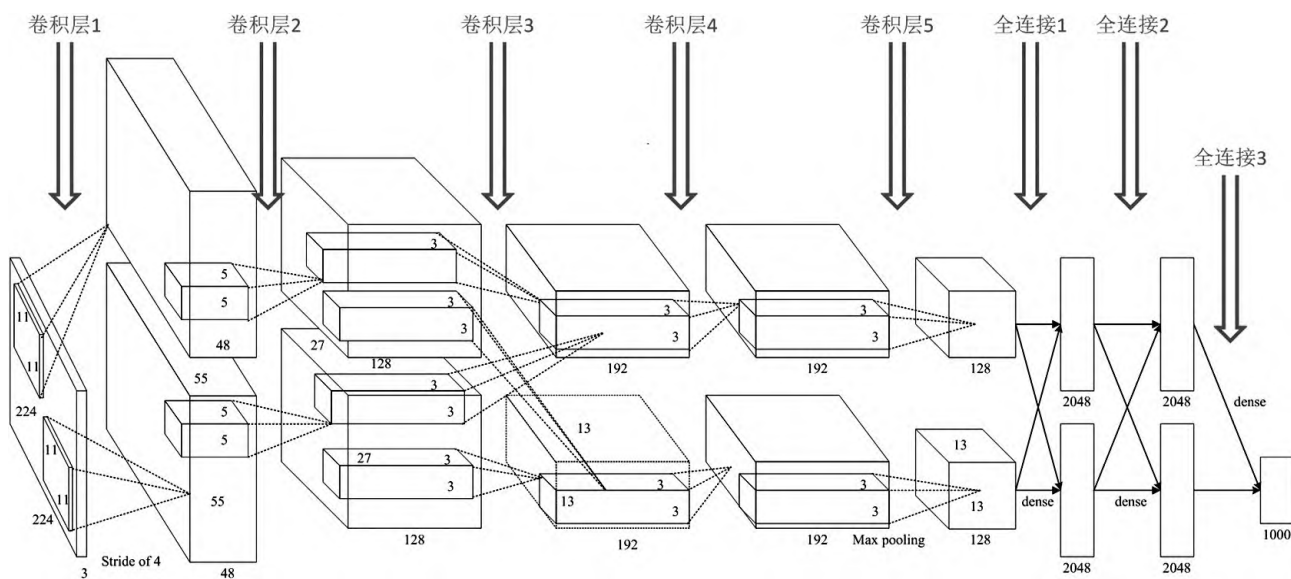


图 2 AlexNet 模型结构

2.3 自适应图像增强

在真实场景中,少数民族文字图像通常存在于古籍、木雕或碑刻中,因此存在大量的噪声,导致传统图像识别方法效果不理。本文结合真实场景噪声特点(含文字噪声、全局噪声、年代噪声、亮度噪声和光照影响),设计了一种自适应图像增强的算法。该算法通过计算场景系数,针对不同场景构建对应的阈值,再开展相应的图像增强处理,包括形态学、灰度伽马变换、灰度对数变换、直方图均衡化、自动色彩均衡(Automatic Color Equalization,简称 ACE)算法^[10]和暗通道先验(Dark Channel Prior,简称 DCP)去雾算法^[11]。

图 3 详细展示了四种场景下的阿拉伯文字图像处理效果。图 3(a)存在字内噪声,图 3(e)利用形态学闭运算有效去噪;图 3(b)存在全局噪声,通常存在于雕刻和木刻的扫描图像,图 3(f)利用 ACE 算法有效消除中心文字的内部噪声;图 3(c)属于年代噪声,年代久远的古籍会导致字体不清晰,图 3(g)是图像增强方法消除噪声的效果;图 3(d)是亮度或光线、过曝导致的噪声,图 3(h)利用暗通道先验去雾处理的效果图。

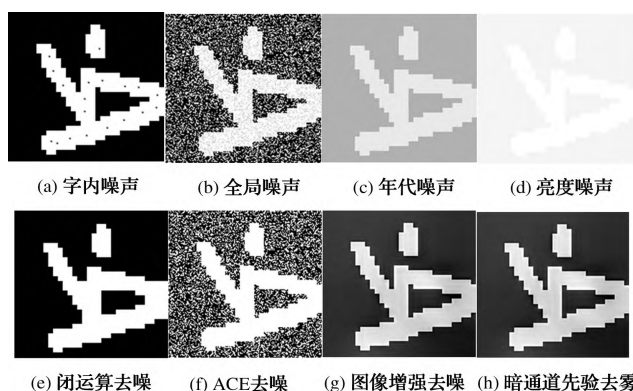


图 3 复杂环境下自适应图像增强算法处理的效果图

3 实验评估

本文进行了详细的对比分析,利用 Keras 和 Sklearn 构建模型。实验环境为 Windows 10 操作系统,处理器为 Inter(R) Core i7-8700K, GPU 为 GTX 1080Ti, 内存为 64GB。

3.1 数据集和模型参数

数据集是来自 Kaggle 的阿拉伯字母手写图像,并结合真实场景模拟噪声,最终生成如表 1 所示的数据集。其中,训练集共计 13440 幅字符图像,测试集共计

3360幅字符图像,涉及28类阿拉伯文字。同时,按照10%的比例进行噪声混淆。每幅图像大小为 32×32 ,并经过图像预处理修改为统一大小。

表1 阿拉伯手写文字数据集

类型	训练集	测试集
类别数量	28类	28类
样本数量	13440	3360
每类数量	480	120
混淆样本比例	10%	10%

为更好地评估IE-AlexNet模型,本文所有实验在相同的数据集及场景下实现。此外,为避免某些异常实验结果的影响,本文最终的实验结果为十次结果的平均值。

构建的AlexNet模型的超参数如表2所示,包括层类、核尺寸、通道数和函数。AlexNet共包括5个卷积层、3个池化层和3个全连接层,并且本文添加BN和Dropout函数,防止出现过拟合现象。

表2 模型超参数

层类	核尺寸	通道数	函数
Conv_1	11×11	96	ReLU、BN
Maxpool_1	3×3	96	
Conv_2	5×5	256	ReLU、BN
Maxpool_2	3×3	256	
Conv_3	3×3	384	ReLU
Conv_4	3×3	384	ReLU
Conv_5	3×3	384	ReLU
Maxpool_3	3×3	256	
FC_6		4096	
FC_7		4096	
FC_8		1000	

3.2 评价指标

实验评估指标为精确率(Precision)、召回率(Recall)、

F_1 值(F_1 -score)和准确率(Accuracy),它们是经典的分类评估指标,其计算过程如公式(1)~公式(4)所示。

$$precision = \frac{TP}{TP + FP} \quad (1)$$

$$recall = \frac{TP}{TP + FN} \quad (2)$$

$$f_1 - score = \frac{2 \times precision \times recall}{precision + recall} \quad (3)$$

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

3.3 实验对比分析

本文与经典机器学习、现有深度学习进行了详细的对比实验,实验结果如表3所示。其中,本文IE-AlexNet模型的精确率为0.9564,召回率为0.9554, F_1 值为0.9559,准确率为0.9553,均优于现有方法。

此外,本文方法的 F_1 值比机器学习中表现最好的SVM模型提升27.32%,比单层CNN和双层CNN模型提升3.49%和3.13%,比TextCNN模型提升0.87%,比文献[8]方法提升0.29%。该实验充分说明本文构建的IE-AlexNet能较好地识别阿拉伯文字图像,并应用于复杂场景下的少数民族古文字识别领域。

同时,本文对比了28种阿拉伯字母的识别效果,其对应字母、类别和 F_1 值如表4所示。表现最好的阿拉伯文字包括第0类、第1类、第4类、第11类和第22类,其 F_1 值分别为0.9917、0.9916、0.9876、0.9789、0.9746,这些文字相对于其他文字特点更明显,更容易被IE-AlexNet识别。而具有相似的文字识别效果相对较差,比如第10类和第9类,第7类和第8类,第2类和第3类等。

图4展示了28种阿拉伯字母的识别结果对应的混淆矩阵,蓝色对角线表示正确识别类别,红色区域是误报或漏报数量。

表3 各模型少数民族文字识别实验结果对比

类型	模型	precision	recall	f_1 -score	accuracy
机器学习	NB	0.5092	0.4938	0.5014	0.4938
	SVM	0.6871	0.6783	0.6827	0.6783
	KNN	0.6313	0.5702	0.5992	0.5702
	RF	0.6633	0.6539	0.6586	0.6539
	LR	0.5341	0.5199	0.5269	0.5199
深度学习	1-layer CNN	0.9222	0.9199	0.9210	0.9199
	2-layer CNN	0.9263	0.9229	0.9246	0.9229
	TextCNN	0.9478	0.9467	0.9472	0.9467
	4-layer CNN ^[16]	0.9536	0.9524	0.9530	0.9524
	本文方法(IE-AlexNet)	0.9564	0.9554	0.9559	0.9553

表4 IE-AlexNet 模型识别各类阿拉伯文字的实验结果

类别	F ₁ -score	类别	F ₁ -score	类别	F ₁ -score	类别	F ₁ -score
	0.9917 (0-class)		0.9465 (7-class)		0.9573 (14-class)		0.9593 (21-class)
	0.9916 (1-class)		0.9021 (8-class)		0.9754 (15-class)		0.9746 (22-class)
	0.9180 (2-class)		0.9170 (9-class)		0.9744 (16-class)		0.9874 (23-class)
	0.9322 (3-class)		0.8718 (10-class)		0.9524 (17-class)		0.9280 (24-class)
	0.9876 (4-class)		0.9789 (11-class)		0.9700 (18-class)		0.9746 (25-class)
	0.9675 (5-class)		0.9664 (12-class)		0.9383 (19-class)		0.9576 (26-class)
	0.9551 (6-class)		0.9717 (13-class)		0.9407 (20-class)		0.9630 (27-class)

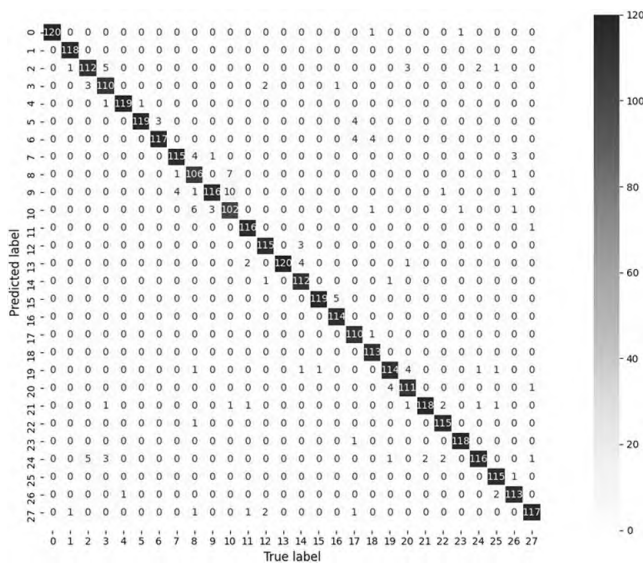


图4 IE-AlexNet 模型识别结果的混淆矩阵

3.4 图像增强及模型性能比较

为突出本文模型自适应图像增强的效果以及性能,本文分别进行了对比实验。表5展示了五种经典模型是否使用图像增强优化的前后效果。其中,使用自适应图像增强算法后,KNN 模型的 F₁ 值提升 14.46%,RF 模型的 F₁ 值提升 13.59%,单层 CNN 模型的 F₁ 值提升 11.97%,文献[8]模型的 F₁ 值提升 3.80%,本文 AlexNet 模型的 F₁ 值提升 3.49%。该实验充分说明本文方法能有效实现不同场景的图像增强,去除少数民族文字图像的噪声,具有更强的鲁棒性和准确率,并能有效识别阿拉伯文字或其他少数民族古文字,具有一定的应用前景和实用价值。

最后,本文对比了深度学习模型的训练误差随 Epoch 下降曲线,如图 5 所示。其中,IE-AlexNet 模型

表5 各模型迁移场景的情感分析实验结果对比

模型	是否图像增强优化	precision	recall	f1-score	F ₁ 值提升率
KNN	×	0.4862	0.4268	0.4546	14.46%
	√	0.6313	0.5702	0.5992	
RF	×	0.5266	0.5187	0.5226	13.59%
	√	0.6633	0.6539	0.6586	
1-layer CNN	×	0.8025	0.8002	0.8013	11.97%
	√	0.9222	0.9199	0.9210	
4-layer CNN (文献[8])	×	0.9164	0.9136	0.9150	3.80%
	√	0.9536	0.9524	0.9530	
AlexNet (本文方法)	×	0.9217	0.9204	0.9210	3.49%
	√	0.9564	0.9554	0.9559	

能以更快的速度下降并拟合,最终趋于0.1822为主。相较于其他模型,本文针对该数据集的图像识别效果更佳,性能更好。

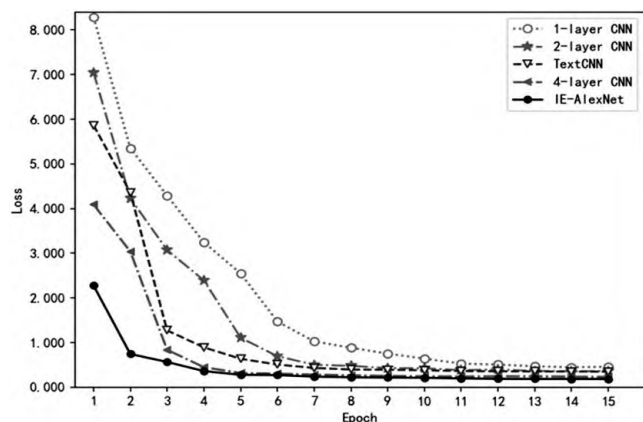


图5 实验误差变化曲线

4 结束语

传统图像识别方法较难识别少数民族文字,而且古文字主要以古籍、雕刻、木刻或碑刻而存在,利用深度学习技术自动化识别,存在大量噪声,数字化读取困难。

本文提出一种融合自适应图像增强的深度学习IE-AlexNet模型,旨在识别复杂场景下的阿拉伯文字图像。实验结果表明,本文方法能有效识别阿拉伯文字图像,并去除噪声,其 F_1 值为0.9559,准确率为0.9553,IE-AlexNet的 F_1 值比机器学习中表现最好的SVM模型提升27.32%,比单层CNN和双层CNN模型提升3.49%和3.13%,比TextCNN模型提升0.87%,比文献[8]方法提升0.29%。

该实验充分说明了本文构建的IE-AlexNet能较好地识别阿拉伯文字图像,其应用于复杂场景下的少数民族古文字识别领域,有较好的鲁棒性和准确率,有一定的应用前景和实用价值。

参考文献(References):

- [1] 李明华. 关于建立国家电子档案战略备份中心的提案[J]. 中国档案, 2022(3):20
- [2] 李远沐, 王展青. 一种改进的小批量手写体字符识别算法[J]. 小型微型计算机系统, 2020, 41(7):1541-1546
- [3] 杨秀璋, 夏换, 于小民. 一种基于水族濒危文字的图像增强及识别方法[J]. 计算机科学, 2019, 46(S2):324-328
- [4] 热依曼·吐尔逊, 吾守尔·斯拉木. 一种维吾尔语联机手写识别系统[J]. 中文信息学报, 2014, 28(3):112-115, 122
- [5] 王晓娟, 白艳萍. 基于BP神经网络的手写体数字的识别方法研究[J]. 数学的实践与认识, 2014, 44(7):112-116
- [6] 杨秀璋, 武帅, 夏换, 于小民. 基于自适应图像增强技术的水族文字提取与识别研究[J]. 计算机科学, 2021, 48(S1):74-79
- [7] 姜文, 卢朝阳, 李静. 基于方向线素特征的手写体维文字符识别[J]. 微电子学与计算机, 2013, 30(10):97-100
- [8] 杨秀璋, 施奕, 李娜, 等. 一种改进卷积神经网络的阿拉伯文字图像识别方法[J]. 信息技术与信息化, 2021(9):6-11
- [9] Krizhevsky A, Sutskever I, Hinton G E. 2012 AlexNet[J]. Adv. Neural Inf. Process. Syst., 2012:1-9
- [10] RIZZI A, GATTA C, MARINI D. A new algorithm for unsupervised global and local color correction[C]. Pattern Recognition Letters, 2003(124):1663-1677
- [11] He K M, Sun J, Tang X O. Single image haze removal using dark channel prior[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(12):2341-2353

(上接第14页)

参考文献(References):

- [1] 薛玉洁, 周杰, 刘雪阳. 单载波至多载波MIMO-OFDM实现高频谱效率方案研究[J]. 南京邮电大学学报, 2019, 39(2): 55-62
- [2] 张家骥. 基于MATLAB的OFDM系统设计与仿真[J]. 信息通信, 2019(12):84-85

- [3] 教青青. MIMO-OFDM系统模型分析研究[J]. 现代工业经济和信化, 2017, 7(5):96-97
- [4] 方志成. MIMO-OFDM通信系统模型设计[J]. 信息通信, 2017(2):205-206
- [5] 兰洪光, 马芳, 韦笑. 基于5G的数据链关键技术研究[J]. 战术导弹技术, 2022(4):1-16