

基于PSO改进的BP神经网络字符识别算法研究

郭成¹, 胡小平², 彭向前¹, 陈立锋¹

¹湖南科技大学, 机电工程学院, 湖南 湘潭

²湖南科技大学, 机械设备健康维护湖南省重点实验室, 湖南 湘潭

收稿日期: 2022年7月4日; 录用日期: 2022年7月28日; 发布日期: 2022年8月5日

摘要

针对字符识别算法中常规的BP神经网络算法在训练识别多种类字符时易造成过拟合而导致训练失败, 字符识别率不高, 传统的模板匹配法对字符识别的效果也不好的问题, 提出一种基于PSO改进的BP神经网络字符识别算法。采用粒子群优化算法(PSO)计算BP神经网络的初始权值和偏置, 提高了网络的收敛速度和训练过程的稳定性; 在常规BP神经网络的基础上加入正则化算法, 防止网络训练的过拟合。在选择网络层节点参数的细节上, 简化输出层节点数, 减少网络的训练参数。最后通过对印刷体字符样本进行识别实验, 识别准确率达到98.97%, 证明该方法可有效地提高字符的识别率。

关键词

字符识别, PSO算法, BP神经网络, 正则化

Research on Improved BP Neural Network Character Recognition Algorithm Based on PSO

Cheng Guo¹, Xiaoping Hu², Xiangqian Peng¹, Lifeng Chen¹

¹School of Mechanical Engineering, Hunan University of Science and Technology, Xiangtan Hunan

²Hunan Provincial Key Laboratory of Mechanical Equipment Health Maintenance, Hunan University of Science and Technology, Xiangtan Hunan

Received: Jul. 4th, 2022; accepted: Jul. 28th, 2022; published: Aug. 5th, 2022

Abstract

In order to solve the problem that the conventional BP neural network algorithm in character rec-

文章引用: 郭成, 胡小平, 彭向前, 陈立锋. 基于 PSO 改进的 BP 神经网络字符识别算法研究[J]. 人工智能与机器人研究, 2022, 11(3): 248-257. DOI: 10.12677/airr.2022.113026

ognition algorithm is easy to cause over-fitting and leads to training failure, the character recognition rate is not high, and the effect of traditional template matching method on character recognition is not good, an improved BP neural network character recognition algorithm based on PSO is proposed. Particle swarm optimization algorithm (PSO) is used to calculate the initial weight and offset of BP neural network, which improves the convergence speed of network and the stability of training process. Regularization algorithm is added based on conventional BP neural network to prevent over-fitting of network training. In the details of selecting the node parameters of the network layer, the number of nodes in the output layer is simplified and the training parameters of the network are reduced. Finally, through the recognition experiment on the printed character samples, the recognition accuracy reaches 98.97%, which proves that this method can effectively improve the character recognition rate.

Keywords

Character Recognition, PSO Algorithm, BP Neural Network, Regularization

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

光学字符识别(Optical Character Recognition, 简称 OCR), 是一项将光学设备和软件系统相结合的自动检测技术, 其中主要包含字符定位、字符分割和字符识别等模块, 通过信息采集设备对产品进行拍摄或扫描, 将产品的字符信息转换成机器间传输的通讯信号[1]。随着智能化工业革命的到来, OCR 技术逐渐运用到更多的行业领域, 如车牌识别、工件喷码字符识别、票据信息识别等[2]。应用领域的宽泛性决定了 OCR 技术的研究目标就是要求更强的鲁棒性, 更快的识别速度以及更高的识别精度。

近年来, 光学字符识别最常用的方法是模板匹配法和神经网络法。模板匹配法对相似字符的区分能力较差, 且匹配的速度慢, 识别率不高, 因此针对工业字符识别一般采用神经网络法, 其中 BP 神经网络的应用最为广泛。王逸铭等[3]提出一种基于 BP 神经网络的扫描电镜图像字符识别方法, 用以识别电镜下的字符信息; 柴伟佳等[4]提出了一种基于联想记忆算法与 BP 算法相结合的混合神经网络方法进行车牌字符识别。基于 BP 神经网络的字符识别算法在工业中的应用普遍存在字符样本获取繁琐; 并且常规的 BP 神经网络在字符类别较多时, 网络训练过程中容易陷入局部最小值而导致训练不足, 也可能出现训练过度的过拟合现象, 网络的泛化能力不高, 从而影响字符识别的速度和准确率。因此设计一种字符提取便捷, 克服 BP 网络训练不足, 过拟合等问题的准确高效的字符识别算法对工业制造自动化领域具有重大意义。

结合现有的研究基础, 针对常规 BP 神经网络存在的问题, 提出一种基于 PSO 算法改进的 BP 神经网络, 并在网络内部加入正则项, 优化目标函数, 减少输出层节点数, 从而减少网络的训练参数量, 避免类别过多对网络复杂度的影响。改进后的神经网络模型不但提高了网络训练速度, 减少了训练过程中的过拟合现象, 还提高了网络的泛化能力, 提高了字符识别的速度和准确率。最后结合 C++和 MFC 设计了一个字符识别系统, 使在进行字符检测时可以在界面中灵活调节分割阈值和字符极性以适应不同背景下的字符识别。实验测试了不同背景、不同极性的字符图片, 结果表明, 该字符识别系统能够有效地适应不同背景下的字符识别。

2. 系统方案设计

2.1. 算法流程设计

算法主要由字符特征提取和网络模型构建两部分构成。由于使用的输入图像是不同背景下的字符串图像，故需要先对图像进行预处理，主要包括图像增强和滤波[5] [6] [7]，可以消除背景噪声影响，为字符分割提供更高质量的图片。将预处理后的图像进行阈值分割，并通过轮廓面积筛选分割出单个字符，并提取相应的字符特征。最后将归一化后的字符特征向量输入字符训练模型，得到权值和偏置数据后进行字符识别。整体算法流程如图 1 所示。

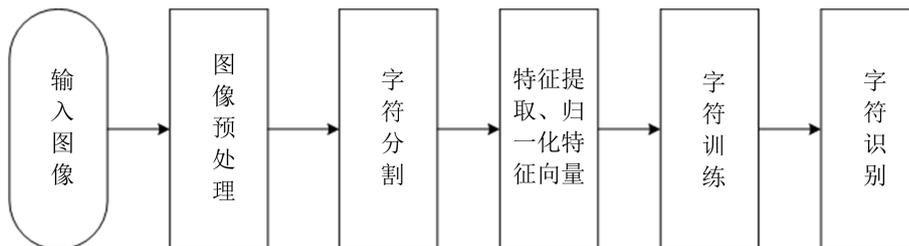


Figure 1. The flow of character recognition algorithm
图 1. 字符识别算法流程

2.2. 系统框架设计

为实现不同背景下不同工业产品的字符识别，字符识别系统主要包括四个功能：图像显示功能，预处理参数设置功能，字符提取及创建字符库功能，字符训练和字符识别功能。图像显示功能包括显示输入图像及每一步骤的处理结果图像；预处理参数设置功能主要调节分割阈值和字符极性，保证不同背景下字符分割的准确性，字符提取及创建字符库功能是将单个字符设置分类标签并保存到字符库；字符训练和字符识别功能主要是用改进的 BP 神经网络算法训练和识别字符。系统界面如图 2 所示。



Figure 2. System interface
图 2. 系统界面

3. 图像预处理及字符分割

3.1. 图像预处理

在工业检测过程中采集的图像都会存在一定的噪声或模糊，因此需要对图像进行一定的预处理，得到便于字符分割的高质量图片。Gamma 变换图像增强能够有效地提高图像对比度，但传统的伽马校正通常会改变输入图像的亮度且难以在过暗和过亮区域同时取得良好的增强效果。为解决此类问题，采用了一种基于双边伽马校正的保亮度图像增强方法[8]，该方法在避免过增强、保亮度和突出细节三方面均取得良好效果，能够保留图像的重要信息。增强前后的效果如图 3 所示。



Figure 3. Contrast picture of character image enhancement on glass surface: (a) Original drawing; (b) Enhanced effect picture

图 3. 玻璃表面字符图像增强对比图：(a) 原图；(b) 增强效果图

3.2. 字符分割

经过增强后的图像可以通过选取合适的二值化阈值将字符与图像背景分割，利用形态学方法消除分割后产生的孤立点和杂色，利用 Opencv 中的轮廓查找算法提取图像中的轮廓信息，根据轮廓面积筛选出字符轮廓，筛选后的字符轮廓分别计算出最小外接矩形，根据最小外接矩形的角度和位置信息精准定位字符，并将字符经过仿射变换矫正，然后单独分割保存到相应的字符库中。图 4 所示为字符分割结果。



Figure 4. Character extraction result

图 4. 字符提取结果

4. PSO-BP 神经网络模型

4.1. 特征提取

字符特征提取就是对经字符分割得到的字符图像数据进行变换，将原始图像数据模式变为变换空间中的数据模式，并且变换后的数据模式的本质特征对模式类别具有良好的区分度[9]。将分割后的字符先进行归一化处理，使字符图像大小统一为 12*16，然后利用 Sobel 算子分别对图像 X 方向和 Y 方向进行边缘信息提取，以及融合 X 和 Y 方向的边缘信息图像，得到 X 方向，Y 方向，及融合边缘图像，提取三幅图像的区域特征信息得到 36 维特征向量，同时对图像进行下采样得到大小为 10*14 的图像，获取其像素值并进行归一化处理得到 140 维特征向量，最后再对原图像进行二值化处理后获取字符图像的左右区域像素占比、上下区域像素占比，中心区域像素占比，中间四列，中间五行的像素比例、所有黑色像素占比，以及上下左右四部分的黑色像素占比共 10 维特征向量，共得到 186 维的特征向量作为 BP 神经网络

络的输入向量。

4.2. PSO-BP 神经网络模型

BP 神经网络是根据误差反向传播算法训练的多层前馈神经网络，通常由一个输入层、几个隐藏层和一个输出层组成[10]。其基本结构如图 5 所示。

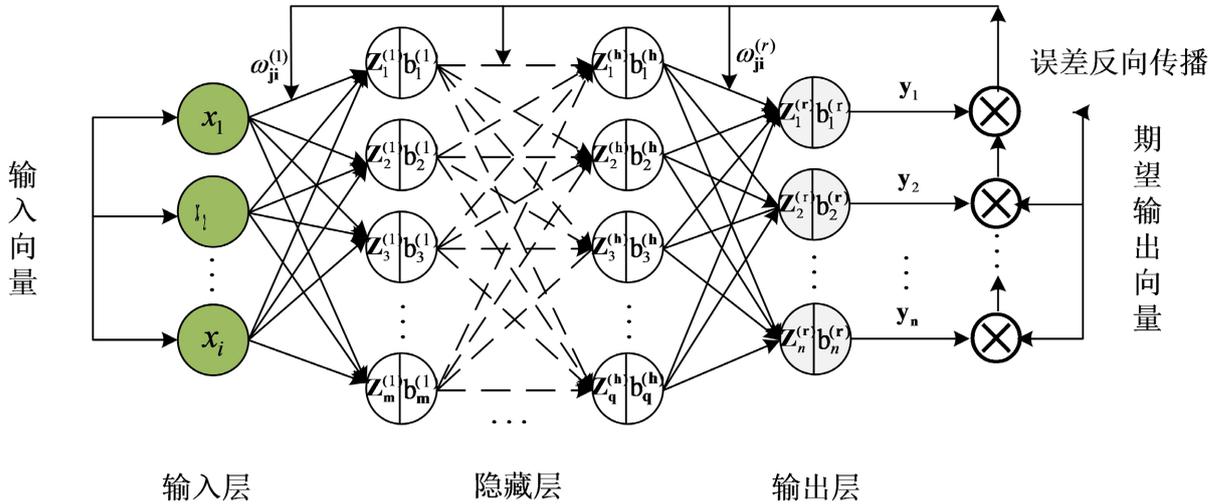


Figure 5. Basic structure of BP neural network
图 5. BP 神经网络基本结构

本文提出的 BP 神经网络模型的训练过程包括五个步骤，相比于传统的 BP 神经网络模型，本文模型优化了网络初始化方式，以及神经网络的目标函数，具体的训练识别过程如下：

(1) 网络初始化。根据系统的实际输入和输出，确定网络输入层，隐藏层和输出层的节点数。初始化输入层，隐藏层和输出层神经元之间的连接权值和偏差阈值，确定学习速率和神经元激活函数。

根据提取到的字符特征个数，确定模型的输入层个数为 186；现有的基于 BP 神经网络的字符识别方法在选择输出神经元个数时都是选择识别的字符类别数作为输出神经元的个数，此处为了减少模型的计算参数，从而降低模型的复杂度，用字符的二进制 ASCLL 值作为字符的标签，每个字符的二进制 ASCLL 值有 7 位，因此设定网络的输出神经元个数为 7，结合经验公式：

$$n_h = \sqrt{n_o + n_i} + a \tag{1}$$

并通过试验得出网络的隐藏层神经元个数为 48 时网络的收敛效果最佳。（ n_h ， n_o 和 n_i 分别为隐藏层神经元个数、输出层神经元个数和输入神经元个数， a 为 1~10 的常数）。

国内外现有的基于 BP 神经网络的字符识别算法选择的输出节点数为字符类别数，对于 26 个大小写字母及 10 个数字，在训练过程中需要计算更新的参数量为：

$$N = n_i * n_h + n_h * n_o + n_h + n_o \tag{2}$$

由式(2)可知，在输入层和隐藏层节点数不变的情况下，减少输出层节点数可以一定程度上提升了网络的训练速度。

确定各层神经元个数后，需要初始化各层神经元之间的权值和偏置。传统的 BP 神经网络一般采用随机初始化的方式生成(-1, 1)的高斯分布作为网络各层的初始权值和偏置。这种初始化方法容易造成每次的训练效果有较大的差异，且可能造成梯度的消失或梯度爆炸，使最终字符识别的效果不稳定。为提

高字符识别效果的稳定性, 本文运用 PSO 算法得出最适于网络收敛的初始权值和偏置。

PSO 算法的优势在于容易实现并且没有许多参数的调节。目前已被广泛应用于函数优化、神经网络训练、模糊系统控制以及其他遗传算法的应用领域[11]。

尽管 BP 神经网络在误差反向传播及更新权重和偏差方面具有显著的机制, 但初始权重和偏差仍将决定网络的收敛方向[12] [13]。因此, PSO 算法将 BP 网络的均方误差作为适应度函数来连续地训练网络的权值和偏差, 最后将最优权重和偏差集作为 BP 网络的初始权值和偏置输入到网络中, 以生成最优的字符识别模型。具体实施步骤如下:

1、首先建立粒子群与网络的权值和偏差的对应关系, 即一个粒子代表一个权值和偏差, 计算每个粒子的维数。

2、初始化网络的其他参数: 粒子群大小、最大迭代步数、学习因子、粒子初始位置和速度、网络学习率和目标精度。

3、通过 PSO 算法获得最优的权值和偏差集, 并将优化结果作为网络的初始权值和偏置输入 BP 神经网络进行训练。

(2) 计算网络各层的输入输出。其中第 P 层的第 j 个神经元的输入:

$$Z_j^{(p)} = \sum_{i=1}^k w_{ji} * O_i + b_j^{(p)} \quad (3)$$

其中 w_{ji} 表示连接前一层的第 i 个节点与 P 层第 j 个节点之间的权值, k 为前一层网络的节点数, O_i 表示前一层网络第 i 个节点的输出, $b_j^{(p)}$ 表示 P 层第 j 个节点的偏差阈值。

假设激活函数为 $f(\dots)$, 则神经元的输出为:

$$y_j^{(p)} = f(Z_j^{(p)}) \quad (4)$$

本模型选用的激活函数为:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (5)$$

(3) 确定目标函数, 计算输出误差和均方误差。传统的 BP 神经网络用均方误差来度量神经网络在数据上的损失函数, 也就是待优化的目标函数。根据上述公式, 计算输出层节点 j 的输出值 y_j , 结合期望输出值 Y_j , 计算输出误差:

$$e_j = Y_j - y_j \quad (6)$$

和均方误差:

$$E_j = \frac{1}{2} \sum_{j=1}^l e_j^2 \quad (7)$$

标准的 BP 神经网络的训练准则是要求所有样本的期望值与输出值的误差平方和小于给定的足够小的允许误差 ϵ 。但实际应用中随着拟合误差的减小, 开始预测误差也随着减小, 但随着拟合误差减小到某个值之后, 预测误差反而增大, 说明泛化能力降低。这就是 BP 神经网络建模过程中遇到的“过拟合”现象[14]。

本文中通过在目标函数基础上增加一个 L2 正则项来减小方差, 从而减轻“过拟合”现象。改进后的目标函数如式(8)所示:

$$E_j = \frac{1}{2} \sum_{j=1}^l e_j^2 + \frac{\lambda}{2} \sum_i^N \omega_{ji}^2 \quad (8)$$

其中 λ 为正则化系数。

(4) 更新权值和偏置。根据改进的网络预测误差更新网络连接权重:

$$w_{ji} = w_{ji} + \eta y_j^{(p)} (1 - y_j^{(p)}) O_i \sum_{k=1}^m (w_{jk} e_{jk} + \lambda w_{jk}) \quad (9)$$

更新网络节点阈值:

$$b_j^{(p)} = b_j^{(p)} + \eta y_j^{(p)} (1 - y_j^{(p)}) \sum_{k=1}^m (w_{jk} e_{jk} + \lambda w_{jk}) \quad (10)$$

其中 η 为学习率。

(5) 比较误差精度或迭代次数, 看是否满足设定目标, 若不满足, 返回到(2)。

PSO-BP 神经网络模型的训练流程如图 6 所示。

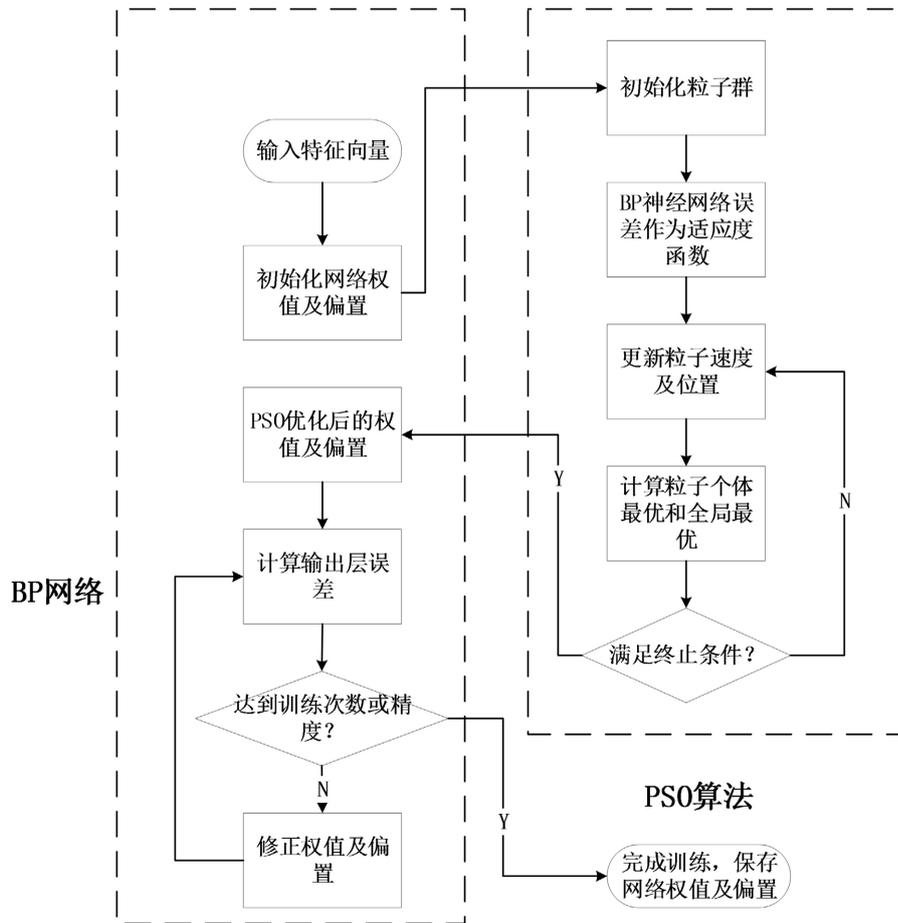


Figure 6. BP neural network training flow chart
图 6. BP 神经网络训练流程图

5. 实验及结果分析

为验证所提出的改进的有效性, 采用工业相机采集了大量字符图片, 通过字符分割得到 2480 张单个字符的图像, 每种字符仅有 40 张图片, 充分体现了小样本训练, 并对 20 张含有多个字符的图片进行识别测试。对几种 BP 网络模型与所提出的网络模型进行实验结果的对比分析。从模型训练时间, 训练误差和字符识别准确率(所有字符的平均识别率)三个方面对模型的性能进行了评估。

5.1. PSO-BP 神经网络参数确定

为得到最优的网络模型，需要通过实验确定模型的主要参数：网络输出节点数、PSO 算法迭代次数、BP 网络学习率和正则化系数。

(1) 网络输出节点

改进的网络模型创新性提出以字符的二进制 ASCLL 值作为字符的类别标签，以其二进制位数作为输出节点数，对比现有算法中以字符类别数作为输出节点数，对两种模型进行实验对比，实验结果如表 1 所示。

Table 1. The influence of the number of network output nodes on the model

表 1. 网络输出节点数对模型的影响

输出节点数	训练时间(s)	训练误差	准确率
7	11.2	0.0002	98.58%
62	15.45	0.00015	98.64%

实验结果表明。输出节点的简化能够加快网络的训练速度，并且对网络的精度影响不大。

(2) PSO 算法迭代次数

PSO 算法进行一次完整的迭代后就会使粒子更靠近最优解，迭代次数过少则达不到优化效果，迭代次数过多则会影响训练速度。为选择合适的迭代次数，采用几种不同的迭代次数对模型进行训练，实验结果如表 2 所示。

Table 2. The influence of iterations of PSO algorithm on the model

表 2. PSO 算法迭代次数对模型的影响

迭代次数	训练时间(s)	训练误差	准确率
50	11.43	0.0025	91.58%
100	13.45	0.00015	98.20%
150	15.65	0.00012	96.77%
200	20.78	0.0001	95.90%

通过实验，在保证检测精度的前提下选用 100 次迭代的方案能使模型更快的收敛。

(3) 学习率

学习速率(learning rate)的调整是神经网络参数非常重要的一部分。学习速率代表了神经网络中随时间推移，信息累积的速度。为寻求最优的学习率，设置了几种不同的学习率对模型进行训练，实验结果如表 3 所示。通过对实验结果进行比较，最终选择网络的学习率为 0.4。

Table 3. The influence of learning rate on the model

表 3. 学习率对模型的影响

学习率	训练时间(s)	训练误差	准确率
0.7	14.36	0.024	92.97%
0.6	15.06	0.018	95.16%
0.5	15.26	0.0018	96.55%
0.4	15.34	0.00012	98.39%
0.3	17.26	0.0001	97.18%

(4) 正则化系数

为选择合适的正则化系数，设置了几种正则化系数进行模型的训练，结果如表 4 所示。当正则化系数为 0.0001 时，网络模型的识别率最高，达到了 98.97%。

Table 4. The influence of regularization coefficient on the model

表 4. 正则化系数对模型的影响

正则化系数	训练时间(s)	训练误差	准确率
0.01	11.4	0.05	88.71%
0.005	11.56	0.01	90.32%
0.001	11.85	0.002	93.55%
0.0005	12.14	0.0008	97.06%
0.0001	12.47	0.0001	98.97%
5×10^{-5}	13.69	0.00001	96.13%

5.2. 改进效果对比分析

通过字符训练和识别的实验，对改进前后和现有的几种字符识别模型对比，其结果如表 5 所示。

Table 5. Comparison of improved experimental results

表 5. 改进实验结果对比

网络模型	训练时间(s)	训练误差	准确率
BP	16.56	0.001	83.87%
PSO-BP	13.45	0.0005	94.82%
BP + 正则化	12.15	0.0001	95.11%
本文模型	10.26	0.0001	98.97%
LSSM + BP [4]	16.49	0.0001	97.40%
PCA-SVM [15]	19.68	0.0001	96.55%

从实验的结果可知，在常规 BP 神经网络字符识别模型的基础上，减少网络输出节点数，加入 PSO 算法优化初始权值，在 BP 网络的目标函数中加入正则化项，都能使原有模型在时效的提高和字符识别的准确率上有明显的改善。

6. 总结

为实现工业字符识别的精度，速度和稳定性，着重探讨了对 BP 神经网络字符识别算法模型进行优化。结合 PSO 算法，在传统 BP 神经网络的基础上加入正则化项，在设置网络输出节点数时改变常规的分类标签，减少网络的输出节点数，在简化模型结构的同时，仅运用少量字符样本进行模型训练，保证了字符识别的准确率，并运用 MFC 搭建了一个简单的字符识别系统，在工业字符识别的运用上具有一定的可行性。

基金项目

国家自然科学基金资助项目(61572185); 汽车玻璃外观缺陷检测关键算法研究(E57110); 湖南省教育厅科研项目重点项目资助(19A170)。

参考文献

- [1] 黄方. 人工神经网络的字符识别[J]. 电子测试, 2020(16): 76-77+95.
- [2] 张成, 戴俊峰, 熊闻心. 融合 LeNet-5 改进的扫描文档手写日期识别[J]. 计算机工程与应用, 2021, 57(9): 207-211.
- [3] 王逸铭, 于永洲, 孙光峤. 基于神经网络模型的扫描电镜图像字符识别方法[J]. 制造业自动化, 2020, 42(7): 18-20.
- [4] 柴伟佳, 王连明. 一种基于混合神经网络的车牌字符识别方法[J]. 东北师大学报(自然科学版), 2018, 50(1): 63-67.
- [5] Liu, M., Jiang, Q. and Hu, J. (2021) Detection of Highway Lane Lines and Drivable Regions Based on Dynamic Image Enhancement Algorithm under Unfavorable Vision. *Computers and Electrical Engineering*, **89**, Article ID: 106911. <https://doi.org/10.1016/j.compeleceng.2020.106911>
- [6] Huang, Z., Wang, Z., Zhang, J., Li, Q. and Shi, Y. (2021) Image Enhancement with the Preservation of Brightness and Structures by Employing Contrast Limited Dynamic Quadri-Histogram Equalization. *Optik*, **226**, Article ID: 165877. <https://doi.org/10.1016/j.ijleo.2020.165877>
- [7] Qi, Y., Yang, Z., Lian, J., Guo, Y., Sun, W., Liu, J., Wang, R. and Ma, Y. (2021) A New Heterogeneous Neural Network Model and Its Application in Image Enhancement. *Neurocomputing*, **440**, 336-350. <https://doi.org/10.1016/j.neucom.2021.01.133>
- [8] 胡钰, 李甜甜, 黄梁松, 李玉霞. 基于双边伽马校正的保亮度图像增强方法[J]. 计算机应用与软件, 2019, 36(5): 204-210+241.
- [9] Singh, S., Sarangi, P.K., Singla, C. and Sahoo, A.K. (2021) Odia Character Recognition System: A Study on Feature Extraction and Classification Techniques. *Materials Today: Proceedings*, **34**, 742-747. <https://doi.org/10.1016/j.matpr.2020.04.680>
- [10] Chen, H., Wang, Y., Zuo, M., Zhang, C., Jia, N., Liu, X. and Yang, S. (2022) A New Prediction Model of CO₂ Diffusion Coefficient in Crude Oil under Reservoir Conditions Based on BP Neural Net-Work. *Energy*, **239**, Article ID: 122286. <https://doi.org/10.1016/j.energy.2021.122286>
- [11] Qin, H., Yang, R., Guo, C. and Wang, W. (2021) Fault Diagnosis of Electric Rudder System Using PSOFOA-BP Neural Net-Work. *Measurement*, **186**, Article ID: 110058. <https://doi.org/10.1016/j.measurement.2021.110058>
- [12] Wang, J., Wen, Y.Q., Gou, Y.D., Ye, Z.Y. and Chen, H. (2017) Fractional-Order Gradient Descent Learning of BP Neural Networks with Caputo Derivative. *Neural Networks*, **89**, 19-30. <https://doi.org/10.1016/j.neunet.2017.02.007>
- [13] Li, Z., Qin, Y., Hou, S.D., Zhang, R. and Sun, H.X. (2020) Renewable Energy System Based on IFOA-BP Neural Network Load Forecast. *Energy Reports*, **6**, 1585-1590. <https://doi.org/10.1016/j.egyr.2020.10.048>
- [14] 张永宏, 吴鑫. BP 神经网络在图像字符识别中的改进和应用[J]. 南京信息工程大学学报(自然科学版), 2012, 4(6): 526-529.
- [15] 杨济萍. 基于主成分降维模型的手写数字识别研究[J]. 网络安全技术与应用, 2021(3): 31-32.