

• 计算机科学与技术 •

DOI:10.15961/j.jsuese.202300579



基于多分支结构的手写字图像特征提取自适应算法

郭晓静¹, 赵小源^{2*}, 邹松林²

(1. 中国民航大学 工程技术训练中心, 天津 300300; 2. 中国民航大学 电子信息与自动化学院, 天津 300300)

摘要:飞机地面维护工卡是维修操作和归档的重要依据,分步完成其手工填写和数字化存储具有重要价值。为减少飞机运行安全隐患,受行业规范限制,工卡通常设计成可离线部署工作的识别模型。工卡书写不但字符类别数目多,还存在大量汉字、英文混用情形,导致字符特征提取困难且识别精度不高。为了针对性地提升平均识别准确率和速度,减少结构相似字、结构复杂字等的错误识别,本文提出一种多分支卷积与特征融合提取结构。利用深层卷积的多尺度特征提取优势,引入改进的重参数化多分支结构来改善图像全局、局部特征提取效果;采用全卷积实现区域空间特征与图像深层特征融合,在分类过程中,提出融合全卷积分类器结构,依据字符特征复杂程度不同自适应分类,改善相似字与复杂字类间、类内的分类识别效果。与主流的手写字识别方法相比,改进后网络结构的存储大小为 69.1 MB;在汉字数据集上的实验表明,识别精度与速度均大幅提升,模型首次预测准确率和前 5 次预测准确率分别达到 97.50% 和 99.79%。模型对相似字符、中英文字符的识别模型优势明显,在包含了中英文和数字的数据集上,改进后结构存储大小为 69.2 MB,实验结果中首次预测准确率达到 97.23%,推理速度达到 1 400 张/s,对飞机地面维护工卡识别等特定领域有一定价值。

关键词:脱机手写汉字识别;全卷积;重参数化结构;空间特征融合;重参数化多分支卷积算法

中图分类号: TP394.41

文献标志码: A

文章编号: 2096-3246(2025)03-0247-09

随着民航业发展,航班密度和数量大幅增加,地面维护保障压力日益增大。维修工程师在日常的维护、维修、放行等工作结束后,需要填写飞行维修记录本、放行工卡等作为工作日志。虽然互联网和移动终端智能设备的不断发展使无纸化逐渐成为主流,但出于行业安全性考虑且受航空公司规模、经济效益等限制,业内仍需分步完成人工手写、数字化存档工作^[1]。目前,中国国航、东航、南航等数十家公司可实现机务人员维修记录的在线填写、上传及数字化存档^[2]。但在多数中小型航空公司和机场,由于经济效益和信息化程度有限,或是需要综合考虑在线传输技术的应用领域和范围以减少工作频段干扰造成的飞行安全隐患,一般需要脱机离线分步完成日志的书写、采集及存储^[3]。例如:2023 年,美国国家航空航天局(NASA)航空安全报告系统(ASRS)曾提及,在某机场 5G 网络推广运行期间,有机组发现在起飞和进近阶段,机长一侧的无线电高度表和部分机载设备系统出现指示错误造成导航

系统工作受扰。由此可见,研究特定领域内手写字符的脱机识别及模型轻量化具有应用价值^[4]。

飞机地面维护工卡是维修操作和归档的重要依据,工卡书写因存在大量汉字、英文混用情形,导致字符特征提取困难且识别精度不高。区别于标准印刷体,手写字易受字体结构和人为因素影响,导致字符类内呈特征差异化,类间差异减小,影响字符分类识别结果^[5]。传统的字符分类识别有标准模板匹配方法和基于深度学习的图像分类识别方法^[6-8]。模板匹配方法从标准字体图像中提取特征,通过特征比对实现字符识别。判别函数和分类器的改进^[9-11]改善了识别效果,但由于识别效果受模板特征匹配阈值影响,准确率仅能达到 90%^[12]。基于深度学习的方法采用卷积运算提取深层特征,识别对象从 90 年代的纯数字发展到当前的汉字文本^[13],所采用模型在平均识别精度方面提升较大。黄婉蓉等^[14]利用各卷积层所提取特征计算其注意力权重,提出注意力机制卷积网络(AT-CNN)模型,在汉字

收稿日期:2023-07-29 修回日期:2023-11-02 网络出版日期:2024-04-19

作者简介:郭晓静(1980—),女,副教授。研究方向:智能检测;图像处理。E-mail:13820869553@139.com

*通信作者:赵小源,E-mail:2021021095@cauc.edu.cn

数据集(HWDB1.1)上的识别准确率达到95.05%,但对相似字符的区分度低。Xie等^[15]利用上下文时序推理,提出预处理增强管道模型,识别准确率达到96.72%,但识别速度较慢。周於川等^[16]通过改进SqueezeNet实现脱机手写汉字的识别,为提高速度减小了模型体积,导致模型精度下降至96.32%。进一步地,Ding等^[17]提出重参数化网络RepVGG,该模型利用多分支训练、单分支推理实现目标分类与检测,比ResNet推理速度提高近1倍,在ImageNet准确率达到80.21%以上,但训练复杂性提高。Trang等^[18]利用改进的RepVGG分类功能实现新冠肺炎分型,准确率达到95.40%,与原始RepVGG相比提高了3.60%,但忽略了图像特征的空间信息,缺乏空间一致性。为了提取典型图像特征及空间特征,解决字符分类、语义分割、小目标细粒度分类等识别精度不高问题,Shiri等^[19]保留特征层,引入全卷积^[20]提取特征位置,在MNIST数据集上准确率达到92.86%,但对于大规模复杂数据,仅进行全卷积操作会导致部分特征信息损失。

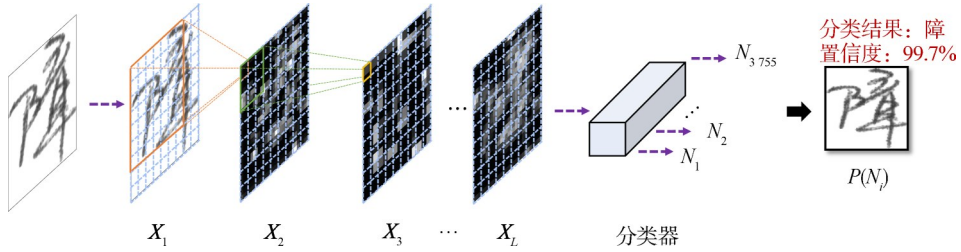


图1 多层卷积实现图像特征提取与分类识别

Fig. 1 Multi-layer convolution for image feature extraction and classification recognition

图1中,卷积层数 L 为14,假设有汉字图像 i ,其标签实际分类为 N_i ,模型预测分类为 \hat{N}_i ,特征提取函数记为 $F(\cdot)$,分类运算记为 $H(\cdot)$,卷积运算记为 $\text{Conv}(\cdot)$,则有式(1)~(4)成立:

$$\hat{N}_i = H(\hat{X}_i) \quad (1)$$

$$\hat{X}_i = F_i(X_{i,1}, X_{i,2}, \dots, X_{i,5}) \quad (2)$$

$$X_{i,j} = F_j(X_{i,j,1}, X_{i,j,2}, X_{i,j,3}) \quad (3)$$

$$X_{i,j,d} = \text{Conv}_d(X_{i,j-1}) \quad (4)$$

式(1)~(4)中, X_i 为第 i 类字符图像实际特征, \hat{X}_i 为模型提取图像特征, $X_{i,j}$ 为第 j 层特征输出, $X_{i,j,d}$ 为第 j 层、第 d 个通道特征输出。

利用卷积运算实现各层各通道的特征提取,特征输出如图2所示。图2使用更大的卷积核有利于增大感受野,增加单个卷积核可获得的信息;更多的卷积层有利于提取聚合图像深层抽象特征,各层的通道数增加有利于提高该层细粒度特征。特征提取结果经由分类器通过概率值得出分类结果。因此,图像输出分类与实际分类的匹配程度易受总卷积层数和各层融合特

综上,在符合民航行业规范标准前提下,研究飞机地面维护工卡的手写字特征提取与识别,综合考虑常用字、相似字及复杂字的特征提取效果和识别速度,以RepVGG结构为基础,利用改进的多分支重参数化结构建立多尺度自适应模型,具有理论研究基础。

1 问题提出

手写字识别属于基于图像特征的分类问题。单通道逐层卷积对于不同复杂程度的字符特征提取存在差别,卷积层越深,越有利于网络学习不同抽象层次的特征,而浅层卷积更能获得图像细节及边缘特征。因此,构建手写字脱机识别模型,实质是通过增加卷积通道数、卷积层数,设计多层多通道卷积模型,优化特征提取和分类算法,以改善类间、类内特征提取效果,提高分类识别精度和推理速度。

本文采用的汉字手写字数据集(HWDB1.0-1.1^[21])共包含3755类,分类数代表不同汉字种数。利用多层卷积实现手写字分类模型,过程如图1所示。

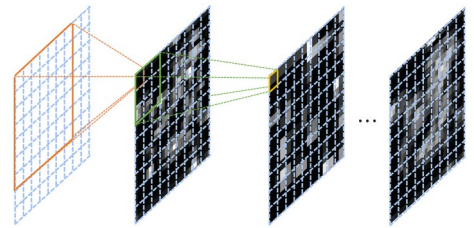


图2 各层各通道特征提取及卷积运算

Fig. 2 Layer and channel feature extraction and convolution operation

征影响。为了在提高模型分类识别精度的同时避免运算成本过大,需要合理确定卷积层数、通道数以及卷积核大小。运算成本越高,模型分类识别速度越慢。运算成本包含模型参数量(P_a)与内存访问成本(C_{MA}),计算如式(5)~(6)所示:

$$P_a = C_{L-1}(m^2 C_L + 1) \quad (5)$$

$$C_{MA} = hw(C_{L-1} + C_L) + m C_{L-1} C_L \quad (6)$$

式(5)~(6)中, C_{L-1} 为 $L-1$ 层卷积核数量, C_L 为 L 层卷积核数量, m 为卷积核大小, h 、 w 分别为特征图的高和宽。可以看出,减少卷积核数量和大小,能够大幅减少

参数量,实现模型压缩;在 C_L, C_{L-1} 相等时,内存访问成本出现最小值。由于内存访问成本在模型训练测试时不便直接计算,后续实验中采用每秒处理的实际图片张数作为其表征量。

对于本文研究的手写字图像识别,由于存在较多的相似字、复杂结构字,也有部分简单字,其特征尺度差异较大,因此,构建分类识别模型的关键在于解决3个方面问题。首先,需要合理优化卷积层数、各层通道数,避免深度过大、通道数过多导致的过拟合和运算量过大;其次,需要统一各卷积层全通道特征矩阵维数,避免局部、全局特征融合导致的识别精度降低;此外,需要优化多卷积层网络结构,避免简单字特征冗余导致的运算效率下降,以及复杂字特征不全导致的识别精度下降。

基于以上问题,本文提出一种基于重参数化的多分支卷积字符识别算法(re-parameterized multi-branch convolutional algorithm, RMCA)。

2 分类识别模型设计

2.1 重参数化卷积结构设计

图3为特征提取器的重参数化卷积结构,该模块采用多分支架构,每个卷积层包含3个分支。

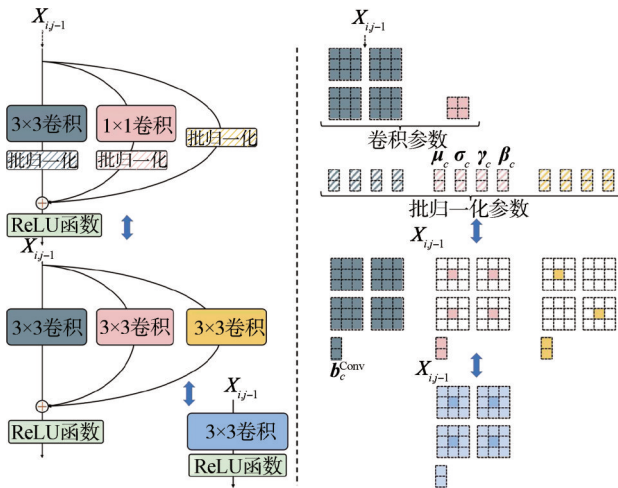


图3 重参数化卷积结构

Fig. 3 Re-parameterized convolutional structure

图3中,为了满足不同复杂程度手写字图像特征提取需要,采用重参数化结构进行优化。如图3所示,第*i*类图像第*j*层输入特征矩阵 $X_{i,j-1}$ 通过3分支(即各自独立的核大小 3×3 、 1×1 的卷积和短接分支)后分别归一化后进行特征融合,则该层图像特征与核大小 3×3 的卷积输出特征维数相同。卷积与归一化按式(7)~(8)计算。

$$\text{Conv}(X_{i,j}) = W_c X_{i,j-1} + b_c^{\text{Conv}} \quad (7)$$

$$\begin{aligned} X_{i,j} &= \sum_{c=1}^3 \text{BN}(\text{Conv}_c(X_{i,j-1})) = \\ &= \sum_{c=1}^3 (\gamma_c \cdot \frac{\text{Conv}_c(X_{i,j-1}) - \mu_c}{\sigma_c} + \beta_c) = \\ &= \sum_{c=1}^3 (\gamma_c \cdot \frac{W_c \cdot X_{i,j-1} + b_c^{\text{Conv}} - \mu_c}{\sigma_c} + \beta_c) = \\ &= \sum_{c=1}^3 (\gamma_c^* W_c X_{i,j-1} + b_c^{\text{Conv}*}) = \\ &= \sum_{c=1}^3 (W_c^* X_{i,j-1} + b_c^{\text{Conv}*}) \end{aligned} \quad (8)$$

式(7)~(8)中, W_c 为*c*分支中的可学习权重矩阵, μ_c 为样本均值, σ_c 为标准差, γ_c 为缩放系数,BN表示批归一化操作, b_c^{Conv} 和 β_c 分别表示*c*分支的原始卷积偏置和第*c*分支批归一化层^[22]的偏置,带*参数表示运用重参数化转换后的等效矩阵。

由图3可以看出,归一化融合输出特征 $X_{i,j}$ 作为下一层卷积块的输入矩阵,可等效为将各分支原卷积核 W_c 填充为同等大小为 3×3 的卷积核 W_c^* 后再完成同维数卷积特征的融合。在采用重参数化结构后,将原多分支不同卷积核卷积运算后的特征融合,等效为统一尺度的卷积核矩阵运算,可以大幅减少参数量和计算量,提高模型拟合能力。由此可知,增加重参数化的设计有利于改善图像融合特征并减少模型运算量。

为了兼顾复杂字、简单字图像特征提取效果,避免卷积层数增加造成全局特征损失,本文设计改进的重参数化结构将使用更大的卷积核以改善融合图像矩阵尺度。

将第1个卷积块重参数化结构改为4分支,结构如图4所示。

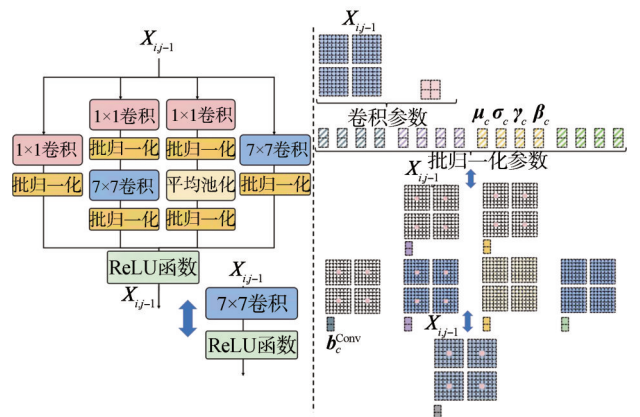


图4 改进的重参数化卷积结构

Fig. 4 Improved Re-parameterized convolutional structure

各分支卷积核尺寸分别设置为 1×1 和 7×7 组合,因此可将该层卷积核尺寸一致等效为 7×7 。卷积层数设置为两层,其他各层按原重参数化结构,卷积核等效为 3×3 ,以避免增加模型计算量。各层卷积与归一化

运算,除首个卷积块分支数 c 变化外,其他各层均满足式(8)。

2.2 分类器融合结构设计

通过多层卷积后的全连接层实现特征融合,易造成全局图像特征信息损失或是部分层的特征信息输出损失,导致分类器判别错误。因此,引入指定层空间特征,以优化分类器的融合结构。为了避免特征冗余导致的运算效率下降,也防止特征提取不充分导致的识别精度下降,设计图 5 所示分类器融合结构,采用全卷积替代全连接层,引入特征提取器的第 4 阶段空间特征至分类器。

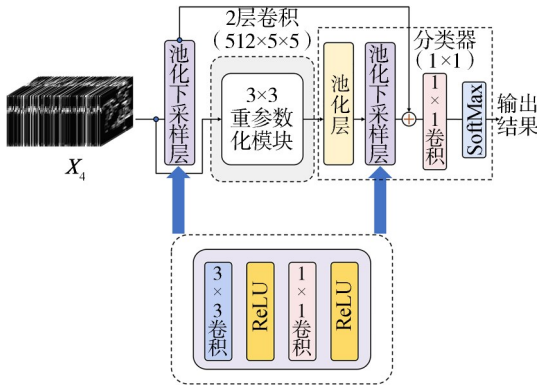


图 5 分类器融合结构

Fig. 5 Classifier fusion structure

由式(3)可知,改进前,第 4 阶段特征输出为 $X_{i,4}$,第 5 阶段特征输出为 $X_{i,5}$ 。改进后,将第 4 阶段输出特

征沿两个路径传播:1) $X_{i,4,k_1}$ 作为第 5 阶段输入矩阵,得到输出特征矩阵 ($X_{i,5,k_1}$);2) $X_{i,4,k_2}$ 经卷积下采样、池化操作 (Pool()) 后得到空间特征矩阵 (Pool($X_{i,4,k_2}$))。两分支特征矩阵拼接 (Concat()), 得到模型输出的融合特征 \hat{X}_i , 从而输出分类值 \hat{N}_i 。具体运算如式(9)~(10)所示。

$$X_{i,4,k_2} = S_{k_2}(C_L(X_{i,4})) \quad (9)$$

$$\hat{X}_i = \text{Concat}[C_L(\text{Pool}(X_{i,5,k_1})), \text{Pool}(X_{i,4,k_2})] \quad (10)$$

式(9)~(10)中, $C_L(\cdot)$ 为下采样卷积运算, $S(\cdot)$ 为特征分支。下采样卷积核大小均为 3×3 , 池化卷积核大小为 1×1 。上述分类器融合设计中,所采用的特定卷积层特征按不同比例融合的算法,可由分类器输出结果获得的两分支特征贡献程度实现分支比例的自适应调节,该方法有利于改善不同复杂程度中文字符的分类识别精度,使模型能够准确提取空间特征,避免多层卷积导致的全局特征输出损失。在本文实验中,比例分支 $k_1:k_2$ 取 4:1。

本文设计构建的手写汉字分类识别模型如图 6 所示。图 6 中, n 为模块层数,模型主要由特征提取器和融合分类器两部分构成。输入图像经过预处理操作(统一尺寸大小和归一化)后输入模型。特征提取器采用 5 阶段的多分支重参数化结构,其中,第 1 阶段采用 4 分支,其余 4 阶段采用 3 分支。之后,融合分类器接收特征提取器的输出,生成最终的分类结果。

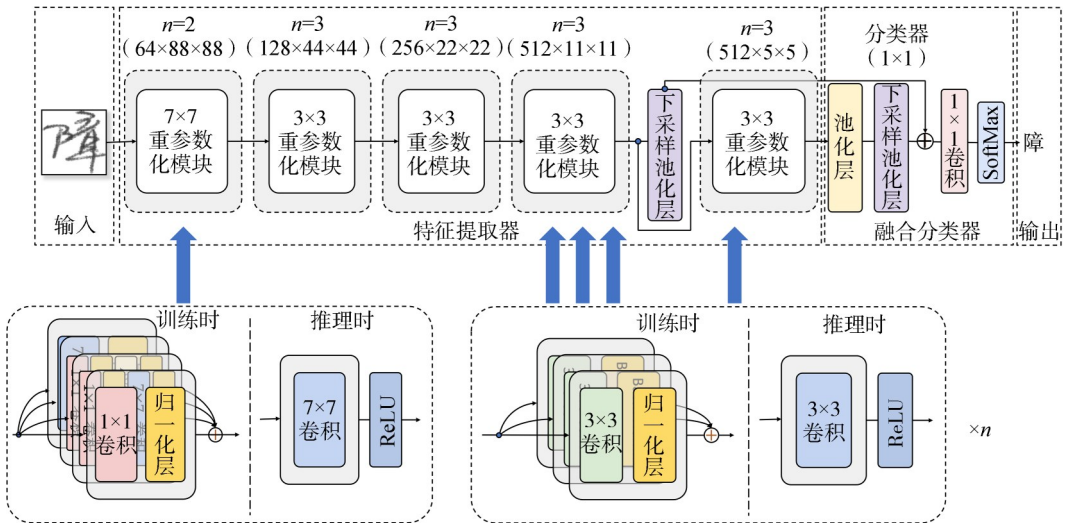


图 6 重参数化手写字符识别模型

Fig. 6 Re-parameterized handwritten character recognition model

3 模型训练与验证

实验采用 Ubuntu 18.04 操作系统,运行内存为 64 GB, CPU 型号为 Intel(R) Xeon(R) CPU @ 2.20GHz, GPU 型号为 Tesla A100-SXM4,显存大小 40 GB。学习率初

值设置为 0.000 2,按余弦规律衰减,批量大小设为 512。实验训练数据集分为两组,第 1 组采用中国科学院自动化研究所公开的手写汉字数据集 (HWDB1.0-1.1),共 3 755 类,含中文手写字符样本数 267 万余张。为便于不同方法下的对比实验,测试采用 ICDAR-2013 数

据集^[23],含中文手写字符样本数22.4万张,记为1组。此外,针对民航领域中英文缩写、数字签名交替出现的特点,在3755类手写汉字的基础上添加英文字符大小写(52类)和数字(10类)各300张图像,新增图片按4:1比例划分训练和测试集,构建共3817类的混合字符集,样本数268万余张,记为2组,覆盖了维修工单中出现的高频字符集。

3.1 验证实验

将输入图像尺寸统一为96像素×96像素,并对像素灰度值进行标准化处理。经计算,样本像素灰度值均值和标准差分别为0.883和0.201。实验得到1组模型准确率和训练损失曲线,如图7所示。由图7(a)可以看出:训练集的识别准确率在5轮后迅速上升,达到99.00%,最终达到99.87%;测试集的识别准确率第1轮已达到94.90%,最终识别准确率达到97.50%。由图7(b)可知,训练损失曲线在前两轮中迅速下降,在第3轮后缓慢下降并趋于平稳,最终收敛到1.35。

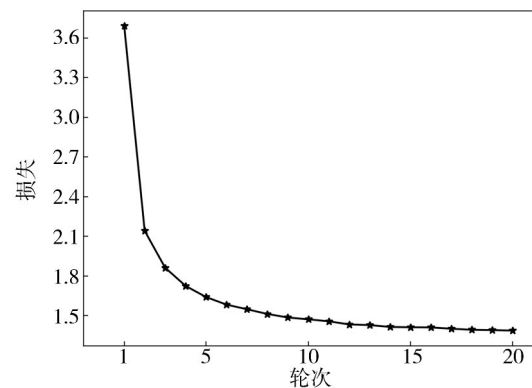
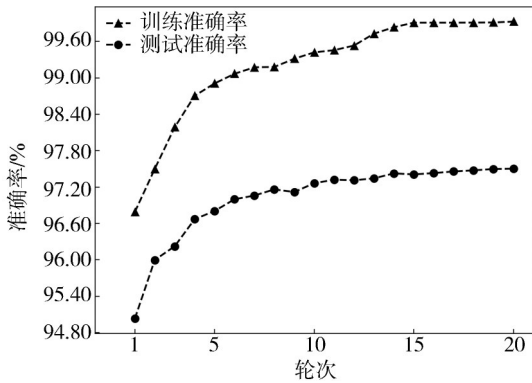


图7 训练和测试准确率与损失曲线(1组)

Fig.7 Accurately and loss curves in training and test(Group 1)

另外,针对手写汉字中相似字符识别效果,按照偏旁部首、字型结构、书写特点等相似程度,在ICDAR2013测试集中选取相似字符1000类,每类字符60张测试图片,构建相似字符测试集。实验结果表明,本文模型在该相似字符测试集的首次预测准确率达到93.14%。与文献[24]的结果相比,准确率提升4.54%。

3.2 消融实验

针对1组,设计了4组消融实验,分别比较本文模型及其在依次移除了分类器融合、全卷积分类器、7×7卷积等模块后的字符分类识别精度,结果见表1。本文模型的首次预测准确率和前5次预测准确率分别达到97.50%和99.79%,与改进前相比准确率明显提高。在特征提取器第1阶段卷积核尺寸变为7×7后,模型首次预测准确率提高0.29%;引入特征融合的分类器设计后,模型首次预测准确率提高0.68%。

表1 消融实验结果(1组)

Tab.1 Results of ablation experiments (Group 1)

模型	首次预测准确率/%	前5次预测准确率/%
重参数化模型	96.53	99.21
重参数化+7×7卷积	96.82	99.32
重参数化+7×7卷积+全卷积分类器	96.95	99.64
重参数化+7×7卷积+全卷积分类器+分类器融合(本文模型)	97.50	99.79

对比各模型第4、5层卷积输出特征效果如图8所示,由图8可知,改进后模型图像特征较丰富。对比笔画较少且特征稀疏的汉字、结构复杂笔画较多的汉字的特征提取效果,采用类激活图分别输出第4、5层特征,结果如图9所示。

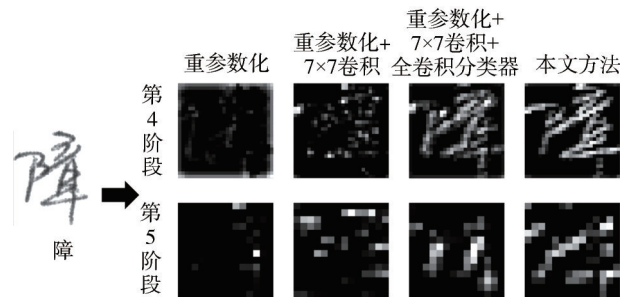


图8 改进前后第4、5层输出特征图

Fig.8 Layer 4 and 5 output feature maps before and after improvement

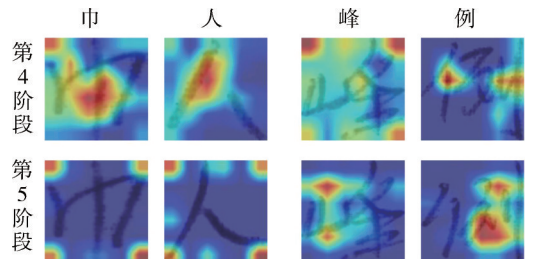


图9 不同结构字第4、5层特征图差异

Fig.9 Differences in feature maps for layers 4 and 5 of different structural characters

由图9可知,特征稀疏的汉字第4层特征较第5层丰富,而结构复杂的汉字第5层仍具有丰富的细节特征,改进前后特征提取效果增强。对飞行记录卡图像

扫描件所截取的手写内容进行测试,结果如图 10 所示。图 10 中,结果稳定置信度均在 90% 以上。综上,

本文模型分类器结构的改进符合预期设计,能够实现自适应分类识别,提高了手写字符分类识别精度。



图 10 飞机地面维护工卡手写识别测试

Fig. 10 Aircraft maintenance job card recognition test

3.3 对比实验

为了对比本文模型在领域内的识别精度和推理

速度,选择 10 种手写字识别模型作对比实验,结果见表 2。

表 2 与其他字符识别模型的性能对比结果

Tab. 2 Performance comparison results with other character recognition models

模型名称	训练集	模型体积/MB	推理速度/(张·s ⁻¹)	首次预测准确率/%
CNN+Similarity Ranking ^[9]	HWDB1.1	—	—	95.58
AT-CNN ^[14]	HWDB1.1	—	—	95.05
Skew Correction Based on ResNet ^[25]	HWDB1.1	—	—	95.50
Lightweight-MobileNetV2 ^[8]	HWDB1.0+1.1	48.34	—	96.12
Augmentation Pipeline CNN ^[15]	HWDB1.0+1.1	—	—	96.72
RepVGG ^[17]	HWDB1.0+1.1	—	1 295	90.60
VGG16 ^[26]	HWDB1.0+1.1	—	913	94.23
EfficientNet ^[27]	HWDB1.0+1.1	—	1 062	92.10
ResNet50 ^[28]	HWDB1.0+1.1	—	1 203	89.07
Ensemble-CNN-voting ^[29]	HWDB1.0+1.1	950.00	—	96.79
RMCA(本文方法)	2组	69.20	1 408	97.23
	1组	69.10	1 410	97.50

由表 2 可知:本文模型大小低至 69.10 MB,略大于轻量化模型 Lightweight-MobileNetV2(模型大小为 48.34 MB),但本文模型首次预测准确率达到 97.50%,推理速度达到 1 410 张/s,相较其他模型,推理速度和

准确率优势明显。与同为重参数化模型的 RepVGG 相比,准确率提升 6.90%,推理速度提升 8.80%;与 ResNet50 相比,准确率提升 8.43%,推理速度提升 17.20%。

可见,本文方法运行效率较高,适用于算力较低

的设备,识别离线手写汉字的场景。对比实验结果验证了模型的准确性和分类可靠性。与文献[30]仅纯数字识别方法相比,本文模型不仅在纯数字识别方面精度相当,而且中英文混合识别准确率为97.23%。

4 结 论

本文提出一种改进的重参数化卷积结构(RMCA),设计了手写字特征提取与分类识别模型。在手写字、英文及数字数据集上的实验表明,模型平均分类识别精度、识别效率较同类方法提升明显,能够满足在实际场景需要,便于实现脱机部署与应用。所作改进主要体现在:

1)设计多层、多分支重参数化结构及其卷积核参数,通过等效结构简化模型,既保持高精度,也提高了推理速度。

2)设计自适应全卷积分类结构,融合多层特征,改善空间特征提取效果,增加了类间的特征区分度,提高了相似字符、复杂字识别准确率,总体识别准确率显著提高。

3)通过实验验证了多分支重参数化模型在提升字符识别问题准确率方面的有效性,获得了较高的识别精度和推理速度,对比其他方法具有一定优势。

后续会进一步研究模型轻量化与硬件感知量化方法,优化移动端部署效率,在保持精度的前提下压缩模型规模。

参考文献:

[1] Thulasy T N,Nohuddin P N,Abd Rahim N,et al.Skill set issues in aircraft maintenance from industrial revolution 4.0 context:A document analytics survey[J].Human Systems Management,2022,41(4):503–516.

[2] Sun Di,Gao Sai,Yan Chao.Study on civil aviation maintenance intelligent job card[J].Aviation Maintenance & Engineering,2023(5):86–88.[孙焱,高赛,阎超.航空维修领域智慧工卡的研究与应用[J].航空维修与工程,2023(5):86–88.]

[3] Chen Jiashu.Research on coexistence of IMT and high-altitude communication platform system in 5G candidate frequency band[D].Beijing:Beijing University of Posts and Telecommunications,2019.[陈家书.5G 候选频段内 IMT 与高空通信平台系统的共存研究[D].北京:北京邮电大学,2019.]

[4] Liu Chenglin,Jin Lianwen,Bai Xiang,et al.Frontiers of intelligent document analysis and recognition:Review and prospects[J].Journal of Image and Graphics,2023(8):2223–2252.[刘成林,金连文,白翔,等.文档智能分析与识别前沿:回顾与展望[J].中国图象图形学报,2023(8):2223–2252.]

[5] Li Yunqing,Du Jun,Hu Pengfei,et al.A method of radical form and hierarchical structure based handwritten Chinese character error correction[J].Journal of Image and Graphics,2023,28(8):2382–2395.[李云青,杜俊,胡鹏飞,等.结合部首字形和层级结构的手写汉字纠错方法[J].中国图象图形学报,2023,28(8):2382–2395.]

[6] Jin Lianwen,Zhong Zhuoyao,Yang Zhao,et al.Applications of deep learning for handwritten Chinese character recognition:A review[J].Acta Automatica Sinica,2016,42(8):1125–1141.[金连文,钟卓耀,杨钊,等.深度学习在手写汉字识别中的应用综述[J].自动化学报,2016,42(8):1125–1141.]

[7] Zhuo Tiantian,Sang Qingbing.Application of attention mechanism and composite convolution in handwriting recognition[J].Journal of Frontiers of Computer Science & Technology,2022,16(4):888–897.[卓天天,桑庆兵.注意力机制与复合卷积在手写识别中的应用[J].计算机科学与探索,2022,16(4):888–897.]

[8] Shi Pingping,Huang Hongqiong.Lightweight MobileNetV2 offline handwritten Chinese character recognition based on attention mechanism[C]//Proceedings of the International Symposium on Robotics,Artificial Intelligence,and Information Engineering(RAIIIE2022).Hohhot:SPIE,2022:528–535.

[9] Zou Junyi,Zhang Jinliang,Wang Ludi.Handwritten Chinese character recognition by convolutional neural network and similarity ranking[EB/OL].(2019–08–30)[2023–07–01].<http://arxiv.org/abs/1908.11550.pdf>.

[10] Kimura F,Takashina K,Tsuruoka S,et al.Modified quadratic discriminant functions and the application to Chinese character recognition[J].IEEE Transactions on Pattern Analysis and Machine Intelligence,1987,9(1):149–153.

[11] Wei Xiaohua,Lu Shujing,Lu Yue.Compact MQDF classifiers using sparse coding for handwritten Chinese character recognition[J].Pattern Recognition,2018,76:679–690.

[12] Luan Shangzhen,Chen Chen,Zhang Baochang,et al.Gabor convolutional networks[J].IEEE Transactions on Image Processing,2018,27(9):4357–4366.

[13] Su Tonghua,You Hongming,Liu Shuchen,et al.FPRNet:End-to-end full-page recognition model for handwritten Chinese essay[C]//International Conference on Frontiers in Handwriting Recognition.Cham:Springer,2022:231–244.

[14] Huang Wanrong,He Kai,Liu Kun,et al.Handwritten Chinese character recognition based on attention mechanism[J].Laser & Optoelectronics Progress,2020,57(8):081002.[黄婉蓉,何凯,刘坤,等.基于注意力机制的手写体中文字符识别[J].激光与光电子学进展,2020,57(8):081002.]

[15] Xie Canyu,Lai Songxuan,Liao Qianying,et al.High perfor-

- mance offline handwritten Chinese text recognition with a new data preprocessing and augmentation pipeline[C]//International Workshop on Document Analysis Systems. Cham:Springer,2020:45–59
- [16] Zhou Yuchuan, Tan Qinhong, Xi Chuanlong. Offline handwritten Chinese character recognition of SqueezeNet and dynamic network surgery[J]. Journal of Chinese Computer Systems, 2021, 42(3): 556–560. [周於川, 谭钦红, 奚川龙. SqueezeNet 和动态网络手术的脱机手写汉字识别[J]. 小型微型计算机系统, 2021, 42(3): 556–560.]
- [17] Ding Xiaohan, Zhang Xiangyu, Ma Ningning, et al. RepVGG: Making VGG-style ConvNets great again[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Nashville: IEEE, 2021: 13728–13737.
- [18] Trang K, Nguyen A H, Ton That L, et al. Improving RepVGG model with variational data imputation in COVID-19 classification[J]. IAES International Journal of Artificial Intelligence(IJ-AI), 2022, 11(4): 1278.
- [19] Shiri P, Baniyasi A. Convolutional fully-connected capsule network(CFC-CapsNet): A novel and fast capsule network[J]. Journal of Signal Processing Systems, 2022, 94(7): 645–658.
- [20] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence. Boston: IEEE, 2016: 640–651.
- [21] Liu Chenglin, Yin Fei, Wang Dahan, et al. Online and offline handwritten Chinese character recognition: Benchmarking on new databases[J]. Pattern Recognition, 2013, 46(1): 155–162.
- [22] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[EB/OL]. (2015–02–11)[2023–07–01]. <http://arxiv.org/abs/1502.03167.pdf>.
- [23] Yin Fei, Wang Qiufeng, Zhang Xuyao, et al. ICDAR 2013 Chinese handwriting recognition competition[C]//Proceedings of the 2013 12th International Conference on Document Analysis and Recognition. Washington: IEEE, 2013: 1464–1470.
- [24] Shao Yunxue, Gao Guanglai, Wang Chunheng. A connection reduced network for similar handwritten Chinese character discrimination[C]//Proceedings of the 2016 15th International Conference on Frontiers in Handwriting Recognition(ICFHR). Shenzhen: IEEE, 2017: 54–59.
- [25] Huang Zetao, Zhang Qian. Skew correction of handwritten Chinese character based on ResNet[C]//Proceedings of the 2019 International Conference on High Performance Big Data and Intelligent Systems(HPBD&IS). Shenzhen: IEEE, 2019: 223–227.
- [26] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2014–09–04)[2023–07–01]. <http://arxiv.org/abs/1409.1556.pdf>.
- [27] Tan Mingxing, Le Q V. EfficientNet: Rethinking model scaling for convolutional neural networks[EB/OL]. (2019–05–28)[2023–07–01]. <http://arxiv.org/abs/1905.11946.pdf>.
- [28] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Deep residual learning for image recognition[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Las Vegas: IEEE, 2016: 770–778.
- [29] Chen Li, Wang Song, Fan Wei, et al. Beyond human recognition: A CNN-based framework for handwritten character recognition[C]//Proceedings of the 2015 3rd IAPR Asian Conference on Pattern Recognition(ACPR). Kuala Lumpur: IEEE, 2016: 695–699.
- [30] Zhang Yaping, Liang Shan, Nie Shuai, et al. Robust offline handwritten character recognition through exploring writer-independent features under the guidance of printed data[J]. Pattern Recognition Letters, 2018, 106: 20–26.

Handwritten Words Image Character Extraction Adaptive Algorithm Based on the Multi-branch Structure

GUO Xiaojing¹, ZHAO Xiaoyuan^{2*}, ZOU Songlin²

(1. Engineering Techniques Training Center, Civil Aviation University of China, Tianjin 300300, China;

2. College of Electronic Information and Automation, Civil Aviation University of China, Tianjin 300300, China)

Abstract:

Objective The aircraft ground maintenance job card is the essential reference for maintenance operations and records. It requires handwritten image identification and digital storage. Due to the limitations of maintenance rules or manuals in civil aviation, a mixture of Chinese and English words often forms complex sentences on the same job card, which creates difficulties in word character extraction and reduces recognition precision. This study applies a new method of multi-branch convolution, the Re-parameterized and Multi-branch Convolution Algorithm (RMCA), to enhance the recognition of complex structures and similar words, improving mean average precision (MAP) and identification efficiency. This study addresses several problems in handwritten word identification. First, the number of layers in the deep convolution network affects the results of characteristic extraction. Second, features extracted from different layers represent varying dimensions in the feature matrix. Third, Chi-

nese words demonstrate varying levels of complexity.

Methods The identification precision index is defined as the mean average precision of Top1 and Top5, and the identification efficiency index is expressed as memory access cost (MAC) to evaluate the proposed model. However, calculating MAC during the model training process presents challenges. Therefore, MAC is replaced with the number of processed image pieces per second. The improved RMCA algorithm utilizes the strengths of deep convolution to extract image characteristics related to boundaries and fine details. Deep convolution layers are known for learning features at different abstraction levels, while lower layers capture more localized details. In addition, the dimension of the convolutional kernel influences the receptive field and local features within a certain layer. The core of the handwritten word identification model lies in the added convolutional channels and layers, along with an adaptive identification algorithm designed for identical and similar words in handwritten images. Higher identification precision and efficiency serve as reference indices for evaluating the model. The improved RMCA algorithm applies four branches in the initial layers, which differs from the original re-parameterized structure. The kernel sizes in the four branches are set to 11 and 77, equivalent to a variable dimension kernel of 77. The following of a fully connected layer can cause the loss of boundary or specific layer features, making it challenging to meet identification requirements for words with complex or simple structures. Hence, the improved RMCA algorithm utilizes spatial features. The fully connected layer is replaced with fully convolutional layers, and the spatial features from the fourth layer are passed to the classifier. This design enables the improved model to adapt automatically to various word structure complexities. The improved model comprises four functional components. The innovation of this study lies in several aspects. First, the enhanced re-parameterized structure across multiple stages and branches achieves an effect equivalent to variable convolution. Second, the refined classifier with fully convolutional layers combines features from specific intermediate layers with the output layer, resulting in improved precision for complex and similar words. The feature extraction performance is enhanced. Compared to feature outputs from the fourth and fifth layers of traditional models, the conclusion is confirmed. The simple Chinese word characteristics in the fourth layer are more abundant than in the fifth, whereas complex handwritten Chinese words contain similarly detailed features across both layers.

Results and Discussions The training image datasets for the experiment consist of two groups. Group 1 includes the HWDB1.0–1.1, comprising 3 755 classes of Chinese words, totaling 2.68 million images. The test dataset is the ICDAR–2013, containing 224 thousand Chinese word images. Group 2 extends Group 1 by incorporating English uppercase and lowercase letters (52 classes) and digits from 0 to 9 (10 classes). The test dataset expands the ICDAR–2013 with additional images of English letters and digits. The experimental results presented in this study demonstrate improvements in evaluation indices compared to other models, including the pre-improvement model. The experiments are divided into two categories: ablation and comparative. The results of the ablation experiment indicate that the Top1 and Top5 precision indices improved to 97.50% and 99.79%, respectively. Specifically, altering the 77 kernel in the first layer increases precision by 0.3%, while modifying the classifier results in a 0.6% gain. Group 2 achieves a Top1 recognition accuracy of 97.23%. The results of the comparative experiment, based on ten traditional models, showed that the proposed model occupies 69.1 MB of storage, slightly more than the 48.34 MB of the lightweight MobileNetV2. However, the model achieves superior precision (97.50%) and identification speed (1 410 cards per second). Compared to the original Rep–VGG and ResNet50, the improved model increases precision by 6.90% and 8.43%, respectively. Identification speed improves by 8.8% and 17.2%, respectively.

Conclusions These results confirm that the proposed method enhances word recognition precision and efficiency. Experiments involving similar word identification yield consistent findings. In the long term, the improved model proves applicable in the field of aircraft maintenance job cards and other specialized areas requiring handwritten word identification.

Key words: offline handwritten Chinese character recognition; fully convolutional network; re-parameterized structure; spatial feature fusion; re-parameterized multi-branch convolutional algorithm

(编辑 吴芝明)

引用格式: Guo Xiaojing, Zhao Xiaoyuan, Zou Songlin. Handwritten words image character extraction adaptive algorithm based on the multi-branch structure[J]. *Advanced Engineering Sciences*, 2025, 57(3): 247–255. [郭晓静, 赵小源, 邹松林. 基于多分支结构的手写字图像特征提取自适应算法[J]. *工程科学与技术*, 2025, 57(3): 247–255.]