

基于 OCR 技术的复杂背景下工件标识字符识别方法

刘 基, 赵志诚, 王晓东

(太原科技大学电子与信息工程学院, 山西太原 030024)

摘要: 针对铸造工艺过程中白模挂涂、铸件补漆后工件标识字符模糊不清、背景复杂, 从而导致识别难度增加的问题, 提出了一种基于卷积循环神经网络与连接时域分类的端对端光学字符识别模型。该模型在卷积层基于VGGNet16进行了简化和改进, 又利用旋转、加噪、调整亮度和对比度的数据增广方法解决了样本数量少的问题。根据试验对比选择合适的模型参数, 实现了复杂背景下铸造工件标识字符的识别。结果表明, 改进后的字符识别模型稳定、识别率高, 对相似字符具有较好的鲁棒性。

关键词: 铸造工件; 工件标识符; 光学字符识别; 卷积循环神经网络; 连接时域分类

近年来, 随着计算机技术的发展与普及, 现代化技术水平也迅速提高, 以计算机技术为依托的图像处理技术随之快速崛起, 并广泛应用到了各个领域。现今, 图像处理技术几乎与人们的生活息息相关, 而图像识别技术更是大幅度地给我们的生活带来便捷, 在多个工业生产过程中减轻了工作人员的劳动强度, 减少了工业生产过程的错误率, 同时大幅度提高生产效率^[1]。作为计算机视觉的重要领域, 利用光学字符识别技术可以从图像中识别出字符信息, 但传统的OCR技术, 如模板匹配算法^[2-5]存在抗干扰能力差, 识别准确率不高等问题。目前, 深度学习技术因具有抗干扰能力强的优势, 得到了广泛的应用, 且在字符识别^[6]、人脸识别^[7]、语音识别^[8]等方面已经取得了很好效果。

在消失模铸造过程主要包括: 聚乙烯珠粒预发泡和熟化、成形后烘干并组合形成白模、挂涂(喷耐火涂料)后烘干、造型、浇注、抛丸、机加工等。在整个铸造加工过程中, 往往通过在白模上印刻标识字符以实现铸造工件的身份识别, 并为智能制造和产品全生命周期管理提供索引依据。但由于白模和工件上的标识字符是物体表面一种凹凸字符, 且字符与背景同色, 成熟的识别“同底色”字符的方法大多只针对单一产品。另外, 工艺过程中白模需经过挂涂处理, 使字符变得更加模糊, 甚至人眼也无法分辨, 这无疑大幅度增加了字符识别的难度。如图1和图2分别为白模挂涂和工件补漆后的标识字符。虽然成熟的字符识别方法很多, 如车牌识别系统^[9-11]和中文手写识别算法^[12-14], 但都具有一定的局限性, 无法适用于多个场景。因此, 设计一套针对同底色字符识别系统成为了铸造工艺过程亟待解决的问题。

本文以某铸造加工企业白模生产、白模挂涂、工件补漆过程中对工件号识别问题为背景, 提出了一种基于卷积神经网络和循环神经网络, 并与连接时域分类(Connectionist temporal classification, CTC)模型相结合的光学字符识别(optical character recognition, OCR)方法, 并运用Python语言编程实现对工件标识字符的识别。

1 训练数据增广

训练数据对深度网络性能有重要影响, 一般地, 如果训练数据越丰富、对问题

作者简介:

刘 基(1994-), 男, 硕士生, 主要研究方向为机器视觉。E-mail: 9815751330@qq.com

中图分类号: TP273

文献标识码: A

文章编号: 1001-4977(2021)

07-0855-06

基金项目:

山西省重点研发计划项目(201803D121025); 山西省留学人员科技活动择优资助项目(2020005)。

收稿日期:

2021-01-04 收到初稿,

2021-02-01 收到修订稿。

空间的覆盖度越高,那么训练所得的深度学习泛化性越好、性能越优。虽然实际采集的数据真实可靠,但它通常存在采集成本高、数据分布难以控制等问题,因此通常采用数据增广的方法丰富优化训练数据^[15-17]。本文在收集真实数据的基础上,采用旋转、加噪、调整亮度和对比度3种策略增广训练数据。

1.1 旋转

因为在生产过程铸件的位置变化以及挂涂、补漆后的铸件都会堆叠到一块,容易造成采集所得字符的角度不尽相同,所以本文通过旋转已有字符图像来模拟不同的情况,来保证训练集能尽量全面覆盖问题空间。旋转即以图片中心为顶点按一定的角度做圆周运动。本文分别按 15° 、 30° 、 45° 、 60° 、 75° 的角度进行不切边旋转,其中旋转 30° 的增广结果如图3所示。

1.2 加噪

因为采集的工件字符是在不同环境拍摄的,所以总会存在诸如光源扰动等情况,使得拍摄出的字符图像会出现噪声,因此本文对原始图片进行加噪处理,尽量模拟可能发生的噪声扰动^[15]。常见的添加噪声方式是在图像中添加高斯噪声、椒盐噪声,经过对工件上字符图像的统计观察同时也为了添加噪声的字符图片更符合真实情况将两种添加噪声的方式分别加入图像扩充数据,高斯加噪的增广结果如图4所示。

1.3 调整亮度和对比度

因为被拍摄的不同,工件所在的位置以及不同时间点下光照条件不相同,所以现实中拍摄的工件字符图像总会存在字符明暗程度不同的情况,因此本文通过调整样本数据亮度和对比度的操作尽量模拟各种光照情况。调整图片的亮度和对比度即在改变图像的宽度、高度、通道数所对应的值,调整后的图像如图5所示。

2 基于CRNN-CTC框架的工件标识字符识别

在解决字符识别问题上,本文将图像通过CNN (Convolutional Neural Network) 抽取的特征序列输入RNN (Recurrent Neural Network) 网络,结合CNN表达深层次特征能力、RNN善于处理文本序列、CTC解决时序类数据的分类问题的优点,实现端到端的训练与识别。

本文在模型设计上基于CRNN-CTC的网络结构层次,根据工件号字符的特点,针对于网络中的CNN进行了改进,以便于解决复杂背景下工件号的识别问

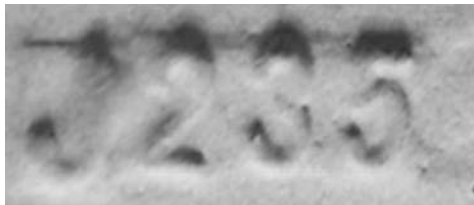


图1 白模挂涂后的工件号

Fig. 1 The workpiece number after the white pattern is painted

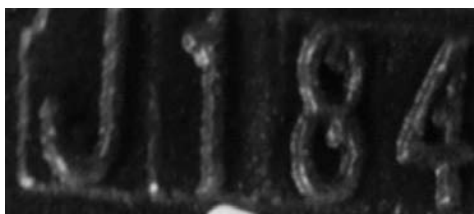


图2 工件补漆后的识别号

Fig. 2 The workpiece identifier after the finished casting is painted



图3 旋转30度

Fig. 3 30 degrees rotation



图4 加入高斯噪声

Fig. 4 Adding gaussian noise

题。改进后的CRNN的网络结构由三部分组成，自底向上为深度卷积层、循环层和转录层^[13]。CRNN结构如图6所示。

2.1 深度卷积层

构建的CNN网络中，如何保证模型运算速度以及正确率的前提下，选择合适的CNN结构是至关重要的。因此需要进行试验对比，调试影响模型训练速度以及识别准确率的网络结构和各个参数，以达到最优的效果。

在神经网络结构中，合适的卷积层数及卷积核数的选择能够更好地学习原始信号样本所隐含的内部特征，进而提高模型的准确率^[19]。文献[20]通过研究网络层数与卷积核尺寸间的关系问题发现，网络层数越多，网络的性能越好，对训练样本的特征提取和学习效果也就越好。在试验过程中发现，由于样本维度较低，持续增加网络层数和卷积核的数量极易出现过拟合现象，不仅未能提高网络训练正确率，反而使之大幅度降低。因而，在深度卷积层，本文选择的是基于VGGNet16的简化深度卷积神经网络模型，使用VGGNet16中 3×3 的卷积核大小。简化后的网络层数包括8个Conv卷积层（Convolutional Layer），8个ReLU激活层（Activation Layer），3个最大池化层（Max Pooling Layer），简化后的卷积核数分别为16，32，64，128，然后对处理过的图像特征提取，得到包含时序信息的序列特征图 $x = \{x^1, x^2, \dots, x^T\}$ ，其中 x 每一列 x^T 为 $x^t = \{x^t_1, x^t_2, \dots, x^t_m\}$ ，时间序列 t 都从 $t=1$ 开始，即 $1 \leq t \leq T$ ， m 和 T 分别指的是特征图的高和宽。因为仅需提取图像的序列特征，所以本文将所搭网络结构的全连接层（Fully Connected Layer）舍去，利用模型中的卷积层和最大池化层等来从输入图像中提取特征序列。深度卷积层的网络结构如图7所示。

2.2 循环层

在循环层，本文选择了深度双向长短时记忆网络（Deep Bidirectional Long-Short Term Memory）。对于传统的RNN模型，能够处理一定的短期依赖，但无法处理长期依赖问题。因为当序列较长时，序列后部的梯度很难反向传播到前面的序列，这就产生了梯度消失问题。为解决此问题，提出了长短时记忆（Long-Short Term Memory，LSTM）网络，双向LSTM的基本

思想是每一个训练序列向前和向后分别是两个循环神经网络，而且这两个都连接着一个输出层。这个结构提供给输出层输入序列中每一个点的完整的过去和未来的上下文信息。因此，对于来自深度卷积层的输入序列特征，双向LSTM能更准确的预测，将结果输出。

本文采用3层双向LSTM网络堆叠的结构，然后将序列特征图 $x = \{x^1, x^2 \dots x^T\}$ 作为输入，且LSTM的每一个时间片后接softmax分类器，得到输出结果 $y = \{y^1, y^2 \dots y^T\}$ ， y 是一个后验概率矩阵，其中 y 的每一列 y^t 为 $\{y^t_1, y^t_2, \dots, y^t_n\}$ ， n 代表需要识别字符的长度。



图5 调整亮度和对比度
Fig. 5 Adjusting brightness and contrast

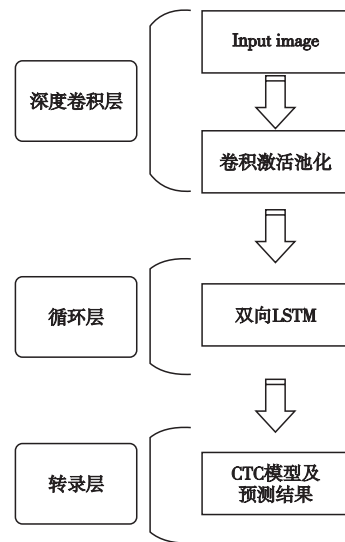


图6 CRNN网络结构
Fig. 6 CRNN network architecture



图7 深度卷积层网络结构
Fig. 7 Deep convolutional layer network architecture

2.3 转录层

转录层本文采用的是CTC模型，是将LSTM网络预测的特征序列的所有可能的结果进行整合，使得输入序列和输出序列对齐，转换为最终结果的过程。本文是在双向LSTM网络的最后连接上一个CTC模型，做到端对端的识别，即按照每帧预测 $y=y_1, \dots, y_T$ 对标签序列 l 定义概率，且输出 l 的概率，如公式（1）所示。

$$p(I|y) = \sum_{\pi: B(\pi)=1} p(\pi|y) \quad (1)$$

其中，序列到序列的映射函数 $B(\pi)=1$ 表示将序列映射到序列 I ，其中 $\pi \in L^T$ ，序列 π 的概率定义为 $p(\pi|y) = \prod_{t=1}^T y_t^{\pi_t}$ ， $y_t^{\pi_t}$ 为 t 时刻有标签 π_t 的概率。

本文在计算出标签序列的条件概率之后，采用了基于词典的转录方法。在数据处理阶段，基于工件标识字符建立字典 D 。在基于字典的模式中，每个测试采样与词典 D 相关联。基本上，通过选择词典中具有公式（1）中定义的最高条件概率的序列来识别标签序列，如公式（2）如下：

$$I^* = \arg \max_{I \in D} p(I|y) \quad (2)$$

其中， $\max_{I \in D} p(I|y)$ 表示在词典 D 约束下具有的最大条件概率 $p(I|y)$ 。再将目标函数 O 设置为最小化条件概率的负对数似然，如公式（3）所示：

$$O = \min_x (-\sum_{L_i, L_i \in x} \log p(L_i|y_i)) \quad (3)$$

其中， $x = \{L_i, L_i\}$ 表示训练集；其中 L_i 表示训练图像； L_i 为真实的标签序列； y_i 是输入训练图像经过深度卷积层和循环层生成的序列。因此，可以在成对的图像和序列上进行端对端训练，去除了在训练图像中手动标记所有单独组件的过程。

因为学习率的取值难以确定，往往通过人的经验来设置，所以选择可以自适应调节学习率的优化器。自适应优化器Adam结合了Adagrad善于处理稀疏梯度和RMSprop善于处理非平稳目标的优点，在试验过程中，在全局学习率为0.005时，训练集和测试集准确率更高，相较于设置其他学习率的值，效果更好。因此在选择Adam优化器来最小化CTC损失函数，并将全局学习率设置为0.005。

3 试验分析

本文通过收集真实数据增广处理后，因为全连接层的存在，需将训练集字符图像设置成固定大小，本文设置为 $3 \times 224 \times 224$ ，然后对图像进行归一化处理，使像素值转换为0~1之间，加快了收敛速度，同时防止梯度消失，使模型更加稳定，最终处理好的数据输入至所搭建的网络中，并将网络的训练次数设置为150次，进行批量化训练，批次大小为50，以训练次数为横坐标，每批次的损失值和准确率为纵坐标，训练集和测试集的结果分别如图8和图9所示。

可以看出，网络模型在测试集中准确率与训练集中相差无几，并且损失函数值略低于训练集，证明网络模型训练效果较好，未出现过拟合或欠拟合的情况，而且在训练次数越来越多时，识别的准确率越来越高，经过150次训练后，本文所搭建的网络模型平均准确率达到98.8%，具有较高的工件标识字符识别效率。为了更好地体现本文所搭建的CRNN-CTC模型框架在识别公司复杂背景下工件号的具体效果，在相同数据集的条件下，本文使用改进后的CNN与相关文献的算法进行了对比试验。不同算法准确率如表1所示，进一步说明了本文所搭建的基于深度学习中CRNN-CTC的端对端OCR识别模型对于某公司复杂工件标识字符的识别具有良好的效果。

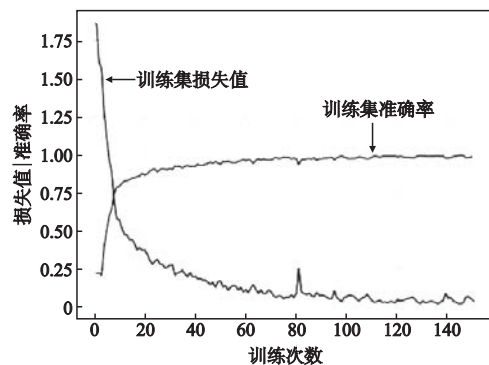


图8 训练集损失值和准确率

Fig. 8 Training set loss value and accuracy

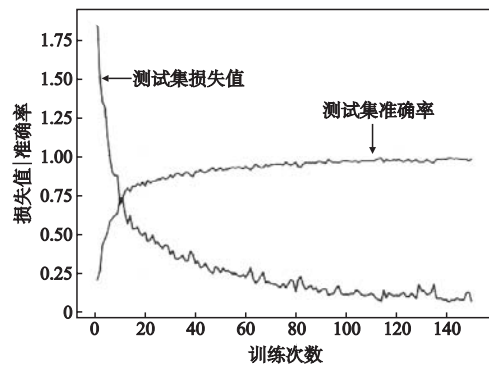


图9 测试集的损失值和准确率

Fig. 9 Loss values and accuracy of test set

表1 各算法准确率对比

Table 1 Comparison of accuracy of each algorithm

算法	准确率/%
AlexNet+LSTM	93.2
VGGNet16+LSTM	89.7
模板匹配法	82.5
本文算法	98.8

4 结束语

以铸造加工过程白模生产、白模挂涂、铸件补漆过程中对工件号识别问题为背景,并结合铸件上工件标识字符以及生产过程场景变化的特点,设计了基于深度学习中CRNN与CTC相结合的识别工件号的OCR模型,结合CNN的优点,对VGGnet16的网络层进行简化

和改进,最终识别准确率达到98.8%,具有较高的工件号识别效率。但是网络模型各项参数需通过多次试验对比才能取得相对最优的结果。因此,需要进一步研究该网络模型应用于识别工件号时自适应参数选择的能力,进一步提高运算效率。

参考文献:

- [1] GAO Zhiyu, LIU Bailin, GU Hongxian, et al. Research on the application of convolutional neural networks in the image recognition [J]. International Journal of Advanced Network, Monitoring and Controls, 2020, 5 (2): 31-38.
- [2] 曾理, 高艳, 毕碧. 铁路货车铸件工件号倾斜DR图像的识别方法 [J]. 中国铁道科学, 2012, 33 (2): 95-99.
- [3] 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.
- [4] 谷秋颇, 白艳萍. 基于模板匹配的车牌数字与字母识别 [J]. 数学的实践与认识, 2011, 41 (3): 66-71.
- [5] ZHONG Fuqiang, HE Songping, LI Bin. Blob analyzation-based template matching algorithm for LED chip localization [J]. The International Journal of Advanced Manufacturing Technology, 2017, 93 (1-4) 55-63.
- [6] 黄方. 人工神经网络的字符识别 [J]. 电子测试, 2020 (16): 76-77.
- [7] 任志玲, 薛新根. 基于深度学习的人脸识别 [J/OL]. 控制工程: 1-5 [2020-12-29].
- [8] 刘娟宏, 胡彧, 黄鹤宇. 端到端的深度卷积神经网络语音识别 [J]. 计算机应用与软件, 2020, 37 (4): 192-196.
- [9] 张荣梅, 张琦, 陈彬. 基于改进LeNet5的车牌识别算法 [J]. 科学技术与工程, 2020, 20 (12): 4775-4779.
- [10] ZHAO Youting, YU Zhi, LI Xiyang. Evaluation methodology for license plate recognition systems and experimental results [J]. IET Intelligent Transport Systems, 2018, 12 (5): 375-385.
- [11] 陈强, 吴锴, 柏云涛. 基于深度学习的车牌识别系统开发研究 [J]. 现代计算机, 2019 (35): 63-66.
- [12] 张新峰, 闫昆鹏, 赵珣. 基于双向LSTM的手写文字识别技术研究 [J]. 南京师大学报 (自然科学版), 2019 (3): 58-64.
- [13] 张冀, 郑传哲. 基于多尺度CNN-RNN的单图三维重建网络 [J]. 计算机应用研究, 2020 (11): 3487-3491.
- [14] LI Zhiyuan, XIAO Yi, WU Qi, et al. Deep template matching for offline handwritten Chinese character recognition [J]. The Journal of Engineering, 2020 (4): 120-124.
- [15] 罗月童, 黄文, 饶永明, 等. 基于深度学习的芯片表面字符识别方法 [J]. 合肥工业大学学报 (自然科学版), 2020 (4): 464-468.
- [16] 王钰清, 陆文凯, 刘金林, 张猛, 苗永康. 基于数据增广和CNN的地震随机噪声压制 [J]. 地球物理学报, 2019 (1): 421-433.
- [17] ZHANG Xiao. Parameter expanded data augmentation for analyzing correlated binary data using multivariate probit models [J]. Statistics in Medicine, 2020, 39 (25): 3637-3652.
- [18] 石鑫, 董宝良, 王俊丰. 基于CRNN的中文手写识别方法研究 [J]. 信息技术, 2019 (11): 141-145.
- [19] 许爱华, 杜洋, 袁涛. 基于深度学习的电机轴承故障诊断研究 [J]. 组合机床与自动化加工技术, 2020 (3): 46-50.
- [20] SHI Baoguang, BAI Xiang, YAO Cong. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39 (11): 2298-2304.

Recognition of Workpiece Identifier on Complex Background Using OCR Technology

LIU Ji, ZHAO Zhi-cheng, WANG Xiao-dong

(School of Electronic Information Engineering, Taiyuan University of Science and Technology, Taiyuan 030024, Shanxi, China)

Abstract:

The workpiece characters is blurred and has a complex background after the white pattern and the casting are painted, which leads to the difficulty of identifying the workpiece characters is gradually increasing. In order to solve the problem, the author proposes an end-to-end optical character recognition (OCR) model based on convolutional recurrent neural network (CRNN) and connectionist temporal classification (CTC), and the model is simplified and improved based on VGGNet16 in small amount of data by augmented method of rotate, adding noise, adjusting brightness and contrast. According to the experimental comparison, the appropriate model parameters are selected to recognize the identification characters of cast workpiece in complex background. The results show that this improved model is stable and has high recognition rate, and has great robustness for similar characters.

Key words:

cast workpiece; workpiece identifier; optical character recognition; convolutional recurrent neural network; connectionist temporal classification
