

DOI: 10.13382/j.jemi.B2306810

基于改进 DBNet 和 SVTR 算法的连铸板坯号检测与识别*

刘乐 张晓松 黄锋 方一鸣

(燕山大学智能康复及神经调控河北省重点实验室 秦皇岛 066004)

摘要:针对钢铁连铸产线板坯号识别字符区域小、光照变化复杂、板坯号图像质量差等问题,提出了一种基于深度学习的连铸板坯号检测与识别两阶段算法。首先,基于采集的连铸产线板坯图像,制备用于板坯号检测与识别的数据集;其次,在板坯号检测阶段,基于 DBNet 算法设计一种 AD-PAN 特征融合结构,以增强检测算法的多尺度特征融合能力和扩大感受野,提高板坯号定位精度;再次,在板坯号识别阶段,引入 SPIN 矫正网络和 SVTR 板坯号识别网络进行端到端训练,使其能够主动转换输入亮度,并改善字符间以及字符与背景间色彩失真的问题。最后,在自制的板坯号检测与识别数据集上进行了对比实验。实验结果表明,本研究提出的算法能够有效定位辊道上不同位置的板坯,并且在复杂背景下对板坯号进行鲁棒识别。其中,板坯号检测 Hmean 数值为 97.92%,板坯号识别的准确率为 97.33%,验证了本文所提算法具有较高的板坯号检测与识别精度。

关键词:板坯号识别;DBNet;特征金字塔融合;端到端网络;SPIN 矫正;SVTR

中图分类号:TP391.4 **文献标识码:**A **国家标准学科分类代码:**510

Detection and recognition of continuous casting slab numbers based on improved DBNet and SVTR algorithms

Liu Le Zhang Xiaosong Huang Feng Fang Yiming

(Key Laboratory of Intelligent Rehabilitation and Neromodulation of Hebei Province,
Yanshan University, Qinhuangdao 066004, China)

Abstract: In order to tackle the challenges associated with small character regions, complex lighting variations, and poor image quality in the identification of slab numbers in steel continuous casting production lines, a two-stage algorithm is proposed for slab number detection and recognition, utilizing deep learning techniques. Firstly, datasets for slab number detection and recognition are prepared based on the collected slab images of a continuous casting production line. Secondly, in the slab number detection stage, an AD-PAN feature fusion structure based on the DBNet algorithm is designed to enhance the multi-scale feature fusion capability and expand the receptive field of the detection algorithm, thereby improving the localization accuracies of the slab numbers. Thirdly, in the slab number recognition stage, the proposed algorithm incorporates a SPIN correction network and a SVTR slab number recognition network for end-to-end training, enabling them to actively transform input brightness and improve color distortion among characters and between characters and backgrounds. Finally, comparative experiments are conducted on self-made datasets for slab number detection and recognition, and experimental results demonstrate that the algorithm proposed in this study is capable of effectively locating slab at different positions on the roller table and robustly recognizing slab numbers in complex backgrounds. The Hmean value for slab number detection is 97.92%, and the accuracy for slab number recognition is 97.33%, confirming the high accuracy of slab number detection and recognition achieved by the proposed algorithm.

Keywords: slab number recognition; DBNet; feature pyramid fusion; end-to-end network; SPIN correction; SVTR

收稿日期: 2023-08-08 Received Date: 2023-08-08

* 基金项目: 秦皇岛市科学技术研究与发展计划项目(202302B048)、河北省创新能力提升计划项目(22567619H)、河北省教育厅科学研究项目(ZD2021102)资助

0 引言

连铸板坯是钢铁生产过程中重要的中间产品。板坯在到达热轧区入口前需要对喷涂在板坯头部的板坯号进行识别,以确保即将进入热轧的板坯与生产计划中的板坯信息一致^[1]。目前钢铁企业主要通过人工完成板坯号的比对,由于连铸板坯生产是连续不间断的,故该方法存在效率低,失误率和遗漏率较高等问题;另外,板坯码垛及吊运过程中易造成板坯号字迹脱落,而工业现场场景和光线复杂多变,字符喷涂质量不稳定,以及热钢表面与板坯号的弱对比度等问题,均给板坯号的检测与识别带来了一定的困难和挑战。

针对工业场景下字符检测与识别问题,常用方法可分为单字符识别和序列字符识别两种。其中,单字符识别方法首先将图像中的字符串分割成单个字符,然后逐一识别,最后再进行字符组合。传统的单字符识别方法主要是利用手工设计特征分割字符,如文献[2]提出一种利用字符特征智能校正提取车牌的方法,实现了对字符轮廓的自适应检测;然而该方法对于倾斜角度较大的车牌需要人为调整约束条件。文献[3]针对食品包装点阵字符漏印问题,提出一种基于卷积神经网络识别反馈的检测算法,克服了传统滑动窗口多尺度和目标位置不准确的影响。然而,字符的形状和大小在实际场景中变化较大,文献[2]和[3]所使用的传统滑动窗口方法只能处理水平字符,而对于非水平字符则无法准确检测和定位。进一步,基于深度学习的单字符识别方法可以通过神经网络自动学习特征,在大数据集和复杂场景下表现更佳。如:文献[4]针对钢坯表面点阵字符运动失真问题,提出一种多方向线扫描方法来确定相邻点阵字符间的边界,并采用点云配准方法来识别易变形字符;然而该方法对光照环境要求较高,存在配准精度不稳定的问题。文献[5]考虑光照变化对钢卷标签字符识别效果的影响,级联了四种不同的滤波器对感兴趣区域进行滤波处理,有效提高了字符的清晰度和辨识度;然而该方法中滤波器主要针对光照变化进行处理,对于其他类型噪声的去除效果有限。

需要说明的是,上述单字符识别方法^[2-5]在实际使用过程中往往存在一定的局限性,如:传统单字符识别方法存在依赖手动设计特征、分割准确性不高、对光照和噪声敏感等问题;基于深度学习的单字符识别方法需要对每个字符进行标注,对数据需求量较大,且在训练和调优时复杂性较高。而基于深度学习的序列字符识别方法可以直接将输入字符图像映射为目标字符串序列,在字符识别方面表现出了较好的性能。如:文献[6]使用 CRNN 算法对芯片表面标识码序列进行识别,由于其不涉及字

符级的标注和分割,有效减少了标注工作量;然而该算法在处理长序列字符时有一定局限性。文献[7]针对工业场景中弯曲金属零件字符定位问题,给出一种优化的特征注意网络,有效提高了字符候选区域的质量;然而该方法生成的候选区域可能无法覆盖小目标的完整区域,进而导致漏检。文献[8]用 ResNet^[9] 取代了 CTPN 中的 VGG16 网络,并使用注意力机制来获得上下文信息,有效提高了检测算法在多语言场景下的鲁棒性;然而该算法对模糊文本或大字体字符的检测效果较差。文献[10]针对工业字符识别任务,提出一种基于残差结构和“挤压-激活”的卷积循环神经网络识别算法,有效提高了对字符特征的学习能力及识别的准确性;然而该算法在处理复杂场景或具有极端光照、模糊等问题的字符时效果较差。文献[11]针对污水处理设备铭牌参数信息识别准确率低的问题,使用 DenseNet+BLSTM+CTC 算法对文本序列进行识别;然而当文本行中包含较多字符时,该算法识别效果较差。文献[12]提出一种用于曲面金属工件字符识别的图像金字塔融合方法,并从多次曝光的图像序列中合成高质量的细节增强图像;然而该方法是基于相机与目标场景相对静止的前提,并不适用于相机与拍摄目标有相对移动的图像序列。

基于上述分析,为了提高字符检测的特征表达能力和更好地应对复杂工业场景对字符识别的影响,本文提出一种基于改进 DBNet^[13] (differentiable binarization net) 和 SVTR^[14] (scene visual text recognition) 算法的连铸板坯号检测与识别算法。具体:在 DBNet 算法基础上采用残差注意力卷积和空洞卷积设计一种 AD-PAN (attention and dilation convolutions path aggregation network) 特征金字塔结构,以通过多尺度特征融合和扩大感受野来解决连铸板坯号区域小、多尺度的问题,进一步提升板坯号检测的鲁棒性;整合 SPIN (structure-preserving inner offset network) 矫正网络和 SVTR 序列识别网络,形成一个端到端网络来进行模型训练,以克服字符区域光照不均匀、模糊等问题,提高板坯号识别的精度。此外,在某钢厂的连铸产线上采集板坯图像,并构建用于板坯号检测与识别的数据集,而后在该数据集上进行对比试验,验证所提算法具有较高的板坯号检测与识别精度。

1 板坯号检测与识别数据集制备

由于目前还没有公开的连铸板坯号检测与识别数据集供本文使用,故本文在某钢厂热轧区入口处安装 1 个摄像头(如图 1)采集辊道上的板坯图像(带有板坯号)。在采集板坯图像时,常存在如下问题:1)高温板坯表面有很高的反射率,导致图像中会出现过曝(高光)和欠曝(阴影)的情况,其中,高光区域缺乏像素细节,而阴影区

域则缺乏清晰度。2) 钢厂生产车间内存在光源不均匀的情况, 导致图像中有些区域过亮, 有些区域过暗。3) 环境光, 如: 室内照明、天窗光线等, 也会对板坯图像采集产生干扰, 导致图像中出现噪点或色彩失真。为确保数据集的多样性, 采集的图像包含了季节变化、光照强度变化和板坯成型时间的差异。具体: 在 6 个月时间内选取了 10 个工作日共 2 500 个场景, 每个场景包含 1~3 块板坯, 且板坯可出现在图像中辊道上的任意位置; 采集的板坯图像的宽度为 1 920 pixels, 高度为 1 080 pixels。另外, 喷印在板坯头部的板坯号由 11 个字符组成, 其中, 板坯号第三位和倒数第三位是 A~D 间的字母, 其余字符是 0~9 间的数字。

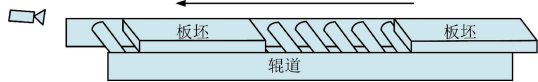


图 1 板坯图像采集示意图

Fig. 1 Schematic diagram of slab image acquisition

对采集的板坯图像板坯号包围边框和板坯号字符标注, 得到板坯号包围边框四点坐标和板坯号字符标签文件。由于板坯号在检测阶段被认为是有效标签, 但不代表其在识别阶段也有效, 因此在识别阶段, 需在检测标注的基础上进一步剔除缺失字符的标签。经过筛选, 对处理好的 2 443 个标签数据分别构建板坯号检测和板坯号识别两个数据集, 其中, 板坯号检测数据集的标签是由板坯图像以及板坯号包围边框的四点坐标构成, 而板坯号识别数据集的标签是由板坯号包围边框裁剪之后的板坯号字符图像和板坯号字符构成。两个数据集均按 7 :

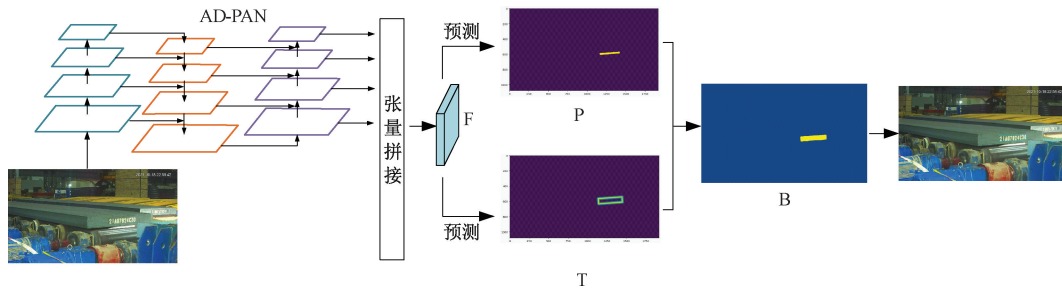


图 3 改进的 DBNet 板坯号检测算法结构图

Fig. 3 Structure diagram of improved DBNet slab number detection algorithm

具体: 采集的板坯图像经过 AD-PAN 结构得到 4 张大小为输入图像高度和宽度 1/4、1/8、1/16、1/32 的特征图; 将这些特征图分别上采样至输入图像高度和宽度的 1/4 大小, 并对它们进行特征融合得到特征图 F, 进而由特征图 F 得到概率图 P 和阈值图 T; 将概率图 P 和阈值图 T 经过可微分二值化操作, 得到近似二值图 B; 最后, 根据二值图 B 计算损失函数, 并通过反向传播算法进行

1 : 2 比例划分训练集、验证集和测试集。

2 板坯号检测与识别算法整体结构

本文采用改进的 DBNet 算法来输出每张板坯图像中板坯号所在的包围框; 再经过裁剪之后, 将板坯号字符图像输入到 SPIN^[15]-SVTR 网络中, 以修正包围框出现的颜色或亮度问题, 并识别其包含的字符; 最后将原始图片、包围框位置以及字符内容同时输出。板坯号检测与识别算法整体结构如图 2 所示。

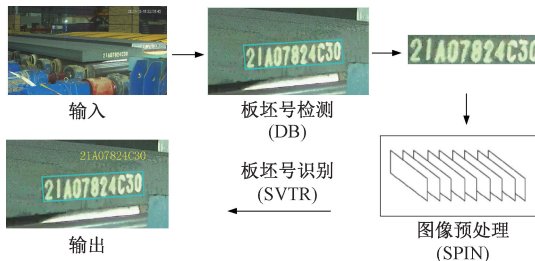


图 2 板坯号检测与识别算法整体结构

Fig. 2 The overall structure of the algorithm for slab number detection and recognition

3 基于改进 DBNet 的板坯号检测算法

考虑到板坯号字符区域相对于板坯图像占比约为 0.3%, 属于相对较小物体检测。因此本节采用残差注意力卷积和空洞卷积设计一种 AD-PAN 结构, 以增强多尺度上的定位能力和扩大感受野。改进的 DBNet 板坯号检测算法如图 3 所示。

模型训练。

图 3 中的 AD-PAN 结构具体如图 4 所示, 其主要作用是对图像特征进行路径增强和聚合, 进而使底层的定位信息更容易传播。具体: 输入的图像通过骨干网络生成 4 个尺度特征映射 $\{C_1, C_2, C_3, C_4, C_5\}$; 为解决骨干网络输出特征表达能力不足的问题, 引入了残差注意力 (residual squeeze excitation, RSE) 机制, 并将横向链接中

的卷积层替换为残差注意力卷积； C_5 通过残差注意力卷积生成特征映射 P_5 ，并通过自上而下的传播途径与较低层次的其他特征融合得到 $\{P_2, P_3, P_4, P_5\}$ ；使用 $\{N_2, N_3, N_4, N_5\}$ 来表示对应于 $\{P_2, P_3, P_4, P_5\}$ 新生成的特征图，这里的 N_2 和 P_2 完全相同。此外，PANet^[16] (path aggregation network) 缩短了底层与顶层特征之间的信息路径。

图 4 中的残差注意力卷积结构如图 5 所示。由于在 AD-PAN 结构中进行上采样和下采样时卷积核的大小不同，故对 C_i 进行 1×1 卷积处理，以获得相同的通道维数 (256)；对 P_i 则是进行 3×3 卷积处理，此时 P_i 的通道数也均为 256。然后特征图依次经过 1×1 卷积层、ReLU 激活层、 1×1 卷积层和 Sigmoid 激活层，为每个特征图生成相应的空间权值。生成的权值映射和合并通道后的特征映射经过哈达玛乘积操作，将得到的特征添加到输入特征图中，得到最终特征图。由于该特征图具有丰富的多尺度上下文信息，在一定程度上缓解了由于通道数量减少而造成信息丢失的问题。

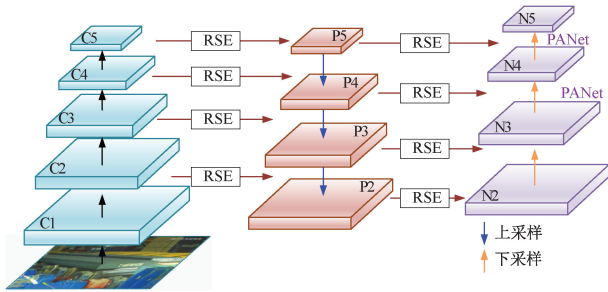


图 4 AD-PAN 结构图

Fig. 4 AD-PAN structure diagram

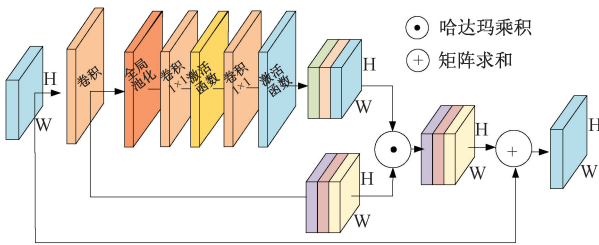


图 5 残差注意力卷积结构图

Fig. 5 RSE convolution structure diagram

进一步，本节在 AD-PAN 结构中设计如图 6 所示的树形分层结构来进行多尺度与上下文信息整合。具体：树形分层结构通过空洞卷积为输入特征图提供不同大小的感受野，利用 add 操作融合来自 3 个支路的特征信息，以提高多尺度图像预测能力。其中，3 个平行分支中的空洞卷积具有相同的卷积核大小，但扩张速率不同，具体来说，每个空洞卷积的核为 3×3 ，不同分支的膨胀率 d

别为 1、3、5。多分支空洞卷积特征融合可以帮助模型有效提取目标的边缘特征，增强边界的清晰度和精确度，对于小目标的定位非常关键。

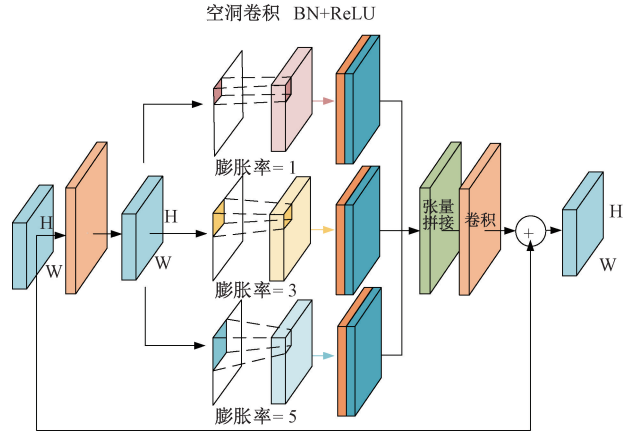


图 6 树形分层结构图

Fig. 6 Tree hierarchical structure diagram

4 SPIN-SVTR 端到端板坯号识别算法

在连铸板坯生产过程中，受光照条件、背景噪声等因素影响导致板坯号字符对比度低，喷涂机工作不稳定导致字符模糊或出现的滴液线 (如图 7 所示) 等问题，均给板坯号的准确识别带来了很大挑战。对此，本节引入基于卷积神经网络的 SPIN 矫正网络对板坯号字符图像在色彩空间上进行变换，并与板坯号识别网络整合成 SPIN-SVTR 端到端网络，使其在模型训练时可以自适应地提取板坯号识别网络所需输入图像的有效特征，从而提高整个模型的板坯号识别性能。

改进的 DBNet 算法可以从含复杂工业场景的板坯图像中分割出板坯号包围框，并通过前置如图 8 所示的 SPIN 矫正网络学习输入图像的映射函数，以对板坯号字符图像颜色进行处理。SPIN 矫正网络将板坯号字符图像中具有相同亮度的所有像素表示为结构模式，该网络由 SPN (structure preserving network) 和 AIN (auxiliary inner-offset network) 两部分组成。其中，SPN 可处理板坯号字符图像 (如图 7(a)、(b)) 中由于光照或高温使得文本模式亮度分散而引起的色彩失真问题。AIN 是一个辅助网络，可用于区分板坯号字符图像 (如图 7(c)、(d)) 中由于文本模式受到滴液线干扰、字符模糊等噪声模式引起的色彩偏移问题。

SPN 本质上是通过过滤输入图像的亮度级别来实现图像结构保留。原始图像中所有具有相同亮度级别的像素在变换后的图像中仍具有相同的亮度级别，本节将结构模式定义为图像中亮度级别为 c 的集合

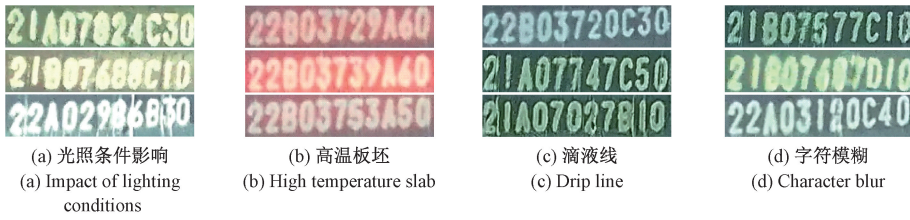


图 7 连铸板坯号不同类型字符

Fig. 7 Different types of characters for continuous casting slab numbers

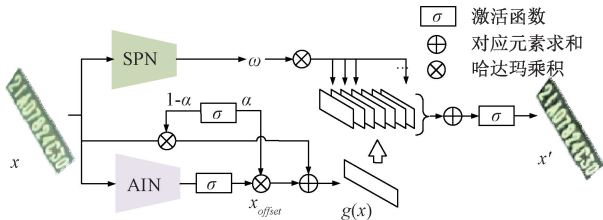


图 8 SPIN 矫正网络

Fig. 8 SPIN correction network

$\{(i, j) | x(i, j) = c\}$ 。SPN 利用逐像素方式将图像像素亮度变换为不同的强度级别,进而把文本模式和非文本模式分离开,同时将文本模式映射到相近亮度级别,以此来聚合不同的文本模式。在 SPN 中,给定输入图像 $x \in I$,令 $x' \in I'$ 表示变换后的图像,则 I 至 I' 的变换 T 可定义为如式(1)所示。

$$x'(i, j) = T[x(i, j)], \quad (1)$$

式中: $x(i, j)$ 和 $x'(i, j)$ 分别为输入和输出图像在坐标点 (i, j) 上的亮度。SPN 的图像转换具体如式(2)所示。

$$x' = T(x) = \text{sigmoid}\left(\sum_i W_{(i)}^s H_{\text{SPN}}(x) x^{\beta_i}\right), \quad (2)$$

式中: W^s 和 $H_{\text{SPN}}(\cdot)$ 分别为 SPN 中最后一个全连接层的部分权重和特征提取器, $W^s H_{\text{SPN}}(\cdot)$ 为最后一个全连接层输出的 $(2K + 1)$ 维特征向量。

由于 SPN 没有考虑到噪声模式的亮度可能与文本模式的亮度相似,从而导致出现模式混淆的问题。而 AIN 可以在每个坐标点 (i, j) 上生成色彩偏移 $x_{\text{offsets}}(i, j)$ 来减轻模式混淆,具体: AIN 首先将图像分成小块,然后预测每个块的偏移值。需要说明的是,所有偏移值都由 sigmoid 函数激活,并通用上采样操作映射到输入图像的大小。在 AIN 中, SPN 将在每个坐标点上生成如式(3)所示的色彩偏移:

$$\begin{cases} g(x) = (1 - \alpha) \otimes x + \alpha \otimes x_{\text{offsets}} \\ \alpha = \text{sigmoid}(W^z H_{\text{SPN}}(x)) \\ x_{\text{offsets}} = W^a H_{\text{AIN}}(x) \end{cases} \quad (3)$$

式中, W^z 为 SPN 中最后一个全连接层的部分权重, W^a 为 AIN 的权重参数, $H_{\text{AIN}}(\cdot)$ 为特征提取器, \otimes 为哈达玛乘积。 α 为一个可学习的更新门,用于接收来自 SPN 的

信号,并感知不同任务的难度,同时兼顾输入图像 x 与预测的色彩偏移之间的平衡。 $g(x)$ (或 x) 为更新(或输入)的图像。在 AIN 的协助下,本节对更新后的图像进行如式(4)所示的增强转换:

$$x' = T(x) = \text{sigmoid}\left(\sum_i \omega_i (g(x)) \beta^i\right) \quad (4)$$

经过 SPIN 处理的板坯号字符图像如表 1 所示。

表 1 SPIN 预处理图像对比效果

Table 1 Comparative effects of SPIN preprocessing on images

图号	输入字符图像	SPIN 预处理后
a		
b		
c		
d		
e		
f		
g		

从表 1 可以看出,经过 SPIN 预处理后,有效提高了图像字符的像素亮度,并抑制了背景亮度,使得目标区域的灰度级范围更窄,背景区域的灰度级范围更广,且背景和目标的亮度差更加明显。

板坯号识别网络中采用的 SVTR 算法是在 PaddleOCRv3^[17] 中引入一种基于 transformer 编码和 CTC (connectionist temporal classification) 解码的字符识别网络,它通过多级逐步下采样的方式获取多尺度特征,并利用自注意力机制和多层感知机对特征进行处理和融合。SVTR 算法结构如图 9 所示。

在图 9 的编码过程中,经过 SPIN 矫正网络的板坯号字符图像首先通过 Patch Embedding 层进行特征提取并通过多层 Mix Block 堆叠使板坯号字符图像的特征信息逐步传递和聚合。需要说明的是,在 Mix Block 堆叠中会进行 Merging 下采样操作,用于减小特征图的高度,降低计算复杂度。 Mix Block 分为 Local Mix Block 和 Global Mix Block 两部分,其中, Local Mix Block 通过滑动窗口的方式进行自注意力计算,用于编码字符形态特征以及字

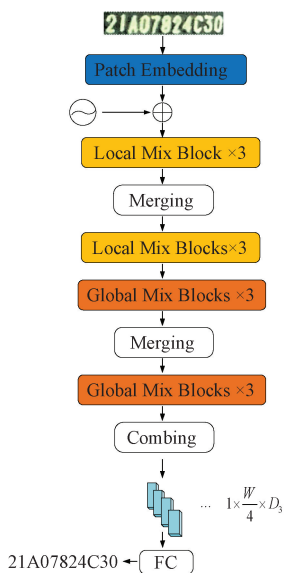


图 9 SVTR 算法结构图

Fig. 9 SVTR algorithm structure diagram

符不同部分之间相关特征;而 Global Mix Block 则是一个全局自注意力机制,用于评估所有字符特征之间的依赖性,并弱化非字符特征的影响。最后,全连接层将经过 Mix Block 处理的字符特征输出为一个特征序列。进一步,在 CTC 解码过程中,将特征序列中连续重复的标签合并成单个标签,并去除空白标签,最终选择得分最高的字符序列作为输出。

5 实验结果及对比分析

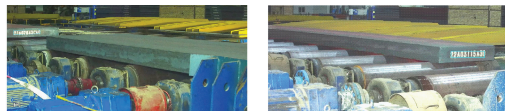
本文所提算法基于 Paddlepaddle-GPU 2.3.1 框架,编程语言版本:Python 3.8.13,运行环境:Linux Ubuntu 18.04 操作系统,CPU 型号: Intel (R) Xeon (R) Gold 6248R,CPU 主频:3.00 GH,服务器内存:128 GB DDR4,显卡型号:RTX 3090,显存:24 GB。

5.1 改进的 DBNet 板坯号检测算法测试与评价

在板坯号检测阶段采用改进的 DBNet 算法对板坯图像进行网络训练,将输入图像的大小统一调整为 640×640 ,骨干网络为 ResNet18,设置学习率为 0.000 01,输入批量大小为 8,采用 Adam 优化器,最大迭代次数为 1 200。对于板坯号检测任务,本节使用 Precision, Recall 和 Hmean 3 个指标来衡量网络的性能。

为了验证本文所改进 DBNet 板坯号检测算法的有效性,将其与原 DBNet 算法进行了对比实验研究,实验结果如图 10 所示。其中,图 10(a)为原 DBNet 算法的板坯号检测结果,图 10(b)为本文所改进 DBNet 算法的板坯号检测结果。可以看出:在图 10(a)中,包围框并没有很好地选中全部字符,因此存在漏检的问题。漏检的本质原

因在于,一部分字符的特征没有响应,字符检测算法在训练时没有学习到漏检部分字符的特征。而在图 10(b)中,由于本文所改进的 DBNet 板坯号检测算法有效提高了对字符区域的特征表达能力,故包围框很好地选中了全部字符,较好地解决了原 DBNet 算法存在的漏检问题。



(a) 原DBNet算法检测结果

(a) Original DBNet algorithm detection results



(b) 改进DBNet算法检测结果

(b) Improved DBNet algorithm detection results

图 10 板坯号检测结果对比

Fig. 10 Comparison of slab number detection results

本文所改进 DBNet 板坯号检测算法与原 DBNet 算法、文献 [8]、主流开源文本检测算法 EAST^[18]、PSENet^[19] 的评价指标结果对比如表 2 所示。

表 2 板坯号检测算法性能比较

Table 2 Performance comparison of slab number detection algorithms

方法	Precision/%	Recall/%	Hmean/%	FPS	参数量/M
文献 [8]	95.79	93.44	94.60	25.61	67.6
EAST	93.47	93.85	93.66	39.97	49.3
PSENet	97.21	92.20	97.30	35.54	47.5
DBNet	98.66	90.98	94.67	32.89	49.5
本文算法	99.15	96.72	97.92	19.45	74.1

从表 2 可以看出:本文改进的 DBNet 板坯号检测算法在损失一部分 FPS 和增加一部分参数量的情况下, Precision、Recall 和 Hmean 3 个指标均优于表 2 中的其他检测算法,表明了本文算法具有更好的鲁棒性与精确性。具体:本文改进的 DBNet 算法在召回率上较原 DBNet 算法提升了 5.74%,在精度上提升了 0.49%,在 Hmean 上提升了 3.25%。

5.2 SPIN-SVTR 端到端板坯号识别算法测试与评价

在板坯号识别阶段,自制的板坯号识别数据集大部分图片的宽高比为 5:1,故本节将图片的输入大小设置为 $[32, 160]$,并将板坯号出现的 '0'、'1'、'2'、'3'、'4'、'5'、'6'、'7'、'8'、'9'、'A'、'B'、'C'、'D',共 14 个字符存为网络训练的自定义字典。此外,设置学习率为 0.000 01,输入批量大小为 128,采用 Adam 优化器,最大迭代次数为 500。

字符识别任务常用的评价指标有:词准确率 (Word

Accuracy, 用 W_a 表示) 和字符准确率 (Char Accuracy, 用 C_a 表示)。其中, 词准确率表示正确识别单词与单词总数的比值, 如式(5)所示:

$$W_a = W_t / W_{all} \quad (5)$$

字符准确率表示准确识别的字符与字符总数的比值, 如式(6)所示:

$$C_a = C_t / C_{all} \quad (6)$$

式中: W_t 为正确匹配的单词总数, W_{all} 为待识别单词总数, C_t 为正确匹配的字符总数, C_{all} 为待识别字符总数。

将本文所改进的 SPIN-SVTR 端到端板坯号识别算法与原 SVTR 算法、文献[6]、[11]和主流开源文本识别算法 ABINet^[20]的评价指标结果对比如表 3 所示。

表 3 板坯号识别算法性能评价

Table 3 Performance evaluation of slab number recognition algorithm

方法	Wa/%	Ca/%	FPS	参数量/M
文献[6]	97.72	79.71	93.0	110.6
文献[11]	99.59	95.90	95.4	144.8
ABINet	99.03	90.98	130.3	147.6
SVTR	99.48	94.47	141.1	90.7
本文算法	99.76	97.33	128.8	99.9

从表 3 可以看出: 本文改进的 SPIN-SVTR 端到端板坯号识别算法与原 SVTR 算法相比, 在损失少量 FPS 和增加少量参数量的情况下, 图像的识别率更高。具体: 在字符准确率方面提升了 2.86%, 在词准确率方面提升了 0.28%。本文改进的 SPIN-SVTR 端到端板坯号识别算法在参数量和 FPS 上都优于其他识别算法, 具体: 相较于文献[6]、[11]和 ABINet 算法, 在字符识别准确率方面分别高出了 1.31%、0.17%、0.73%; 在板坯号识别准确率方面分别高出了 6.93%、1.43%、5.66%, 验证了本文所改进板坯号识别算法的有效性。板坯号示例识别结果如表 4 所示。

表 5 消融实验结果





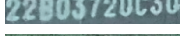



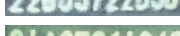
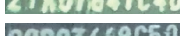

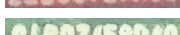



Table 5 Ablation experiment results

PAN	RSECONV	树形分层结构	Precision/%	Recall/%	Hmean/%	FPS	参数量/M
√			98.24	91.80	94.91	31.41	50.5
√	√		99.56	93.03	96.18	27.29	51.1
√	√	√	99.15	96.72	97.92	19.45	74.1

从表 5 可以看出, 原 DBNet 算法只加入 PAN 时, 检测精度会稍微降低, 但提高了召回率和 Hmean 指标, 同时模型参数量增加了 1 M, FPS 略有降低; 再继续加入 RSECONV 注意力机制后, 在少量提高模型参数量和减少 FPS 的情况下, 精度、召回率和 Hmean 指标均有所提升, 即: 分别提升了 1.32%、1.17%、1.27%, 从而有效抑制了

表 4 板坯号示例识别结果

Table 4 Example identification results of slab numbers

图号	板坯号	结果
1		21A07725A30
2		22B03714C10
3		22B03602D60
4		21B07688C30
5		22B03720C30
6		21B07630A40
7		22A03016C50
8		21B07628A30
9		22B03722D30
10		21A07841C40
11		22B03668C50
12		22B03729A60
13		21B07659D40
14		22B03602D60
15		21B07672C10

从表 4 可以看出: 本文所改进的 SPIN-SVTR 端到端识别算法对光照差异较大图(1)、(13), 字符模糊图(2)、(4)、(8), 高温板坯图(12), 滴液线图(3)等都有很高的识别精度。需要说明的是, 将图(7)中第七个字符“9”误识别为“1”, 可能是由于滴液线干扰严重; 后期可通过扩充数据集中此类数据, 来进一步提高模型对此类板坯图像的识别能力。

5.3 消融实验

为验证本文所提算法中改进内容在检测阶段的有效性而进行了消融实验, 即: 在自制的板坯号测试集上对比了 PAN、RSECONV 和树形分层结构的有效性。消融实验结果如表 5 所示。

背景信息, 提高了目标特征的表达能力; 再进一步加入树形分层结构后, 召回率和 Hmean 分别提升了 3.69% 和 1.74%。此外, 使用不同空洞率的空洞卷积有效提高了模型对正样本的覆盖率, 尽管模型参数量有所增加, FPS 有所减少, 但也满足连铸板坯生产实际中 FPS 大于 15 的要求。综合上述实验结果可知, 使用本文改进后的检测

算法可以有效改善板坯号特征提取能力,提升板坯号的检测能力。

6 结 论

本文基于改进的 DBNet 和 SPIN-SVTR 算法研究了连铸板坯号检测与识别问题。其中,在板坯号检测算法基础上设计 AD-PAN 结构,使用残差注意力卷积和空洞卷积,极大地提高了板坯号检测精度;板坯号识别算法整合 SPIN 矫正网络进行端到端训练,有效缓解了由于采集的板坯号字符图像中光照强度变化和背景噪声导致的颜色失真问题。另外,基于自制的板坯号检测与识别两个数据集进行了对比实验研究,实验结果表明本文所提算法具有如下 3 个显著优点:1)表征力:改进的板坯号检测模型对图像中轨道上任意位置的板坯均有效,且板坯号检测的 Hmean 达到 97.92%。2)鲁棒性:SPIN 预处理方法有效提升了光照条件变化、字符中滴液线等低质量字符的图像质量。3)准确率:本文所提板坯号识别算法的识别率达到 97.33%。在下一步的研究工作中,可利用知识蒸馏策略对算法开展轻量化研究,从而减少模型参数数量,以便其在移动设备上部署,并实现更高效的板坯号检测与识别。

参考文献

- [1] ZHAO Q J, CAO P, TU D W. Toward intelligent manufacturing: label characters marking and recognition method for steel products with machine vision [J]. *Advances in Manufacturing*, 2014, 2: 3-12.
- [2] 费继友, 谢金路, 李花, 等. 基于字符特征约束的自适应车牌校正提取 [J]. *仪器仪表学报*, 2016, 37(3): 632-639.
FEI J Y, XIE J L, LI H, et al. Adaptive license plate correction and extraction based on character feature constraint [J]. *Chinese Journal of Scientific Instrument*, 2016, 37(3): 632-639.
- [3] 曹泽卫, 欧阳, 林冬婷, 等. 基于 CNNs 识别反馈的点阵字符检测方法 [J]. *电子测量与仪器学报*, 2020, 34(8): 159-166.
CAO Z W, OU Y, LIN D T, et al. Dot matrix character detection method based on CNNs recognition feedback [J]. *Journal of Electronic Measurement and Instrumentation*, 2020, 34(8): 159-166.
- [4] GE J H, LIU L CH, SUN J X, et al. Automatic recognition of hot spray marking dot-matrix characters for steel-slab industry [J]. *Journal of Intelligent Manufacturing*, 2023, 34(2): 869-884.
- [5] CALDEIRA T, CIARELLI P M, NETO G A. Industrial optical character recognition system in printing quality control of hot-rolled coils identification [J]. *Journal of Control, Automation and Electrical Systems*, 2020, 31(1): 108-118.
- [6] 赵杰, 高林, 李玮. 基于 OCR 的芯片表面标识码自动识别系统设计 [J]. *电子测量技术*, 2022, 45(5): 111-117.
ZHAO J, GAO W, LI W. Design of automatic identification system for chip surface identifiers based on OCR [J]. *Electronic Measurement Technology*, 2022, 45(5): 111-117.
- [7] GUAN T K, GU CH CH, LU CH SH, et al. Industrial scene text detection with refined feature-attentive network [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(9): 6073-6085.
- [8] NIU S Z, LI X X, WANG M S, et al. A modified method for scene text detection by ResNet [J]. *Computers, Materials and Continua*, 2020, 65(3): 2233-2245.
- [9] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C]. *IEEE conference on Computer Vision and Pattern Recognition*, 2016: 770-778.
- [10] ZHANG X SH, WANG Y. Industrial character recognition based on improved CRNN in complex environments [J]. *Computers in Industry*, 2022, 142: 103732.
- [11] 郭毛琴, 谢红薇, 张效良. 污水处理设备铭牌中字符检测与识别 [J]. *计算机工程与设计*, 2022, 43(10): 2094-2910.
GUO M Q, XIE H W, ZHANG X L. Character detection and recognition of sewage treatment equipment nameplat [J]. *Computer Engineering and Design*, 2022, 43(10): 2094-2910.
- [12] XIANG ZH, WU H X, ZHOU D. Recognition of characters on curved metal workpiece surfaces based on multi-exposure image fusion and deep neural networks [J]. *Neurocomputing*, 2022, 512: 117-129.
- [13] LIAO M H, WAN ZH Y, YAO C, et al. Real-time scene text detection with differentiable binarization [C]. *Proceedings of the AAAI conference on artificial intelligence*, 2020, 34(7): 11474-11481.
- [14] DU Y K, CHEN ZH N, JIA C Y, et al. SVTR: Scene text recognition with a single visual model [C]. *European Conference on Artificial Intelligence*, 2022: 884-890.
- [15] ZHANG CH W, XU Y L, CHENG ZH ZH, et al. SPIN: Structure-preserving inner offset network for scene text recognition [C]. *Proceedings of the AAAI conference on artificial intelligence*, 2021, 35(4): 3305-3314.

- [16] LIU SH, QI L, QIN H F, et al. Path aggregation network for instance segmentation [C]. IEEE CVF Conference on Computer Vision and Pattern Recognition, 2018: 8759-68.
- [17] LI CH X, LIU W W, GUO R Y, et al. PP-OCRv3: More attempts for the improvement of ultra lightweight ocr system[J]. ArXiv Preprint,2022, ArXiv: 2206. 03001.
- [18] ZHOU X Y, Yao C, WEN H, et al. East: An efficient and accurate scene text detector[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 5551-5560.
- [19] NGUYEN H, TRAN D, NGUYEN K, et al. PSENet: Progressive self-enhancement network for unsupervised extreme-light image enhancement[C]. Proceedings of the IEEE CVF Winter Conference on Applications of Computer Vision, 2023: 1756-1765.
- [20] FANG SH CH, XIE H T, WANG Y X, et al. Read like humans: Autonomous, bidirectional and iterative language modeling for scene text recognition [C]. Proceedings of the IEEE CVF Conference on Computer Vision and Pattern Recognition, 2021: 7098-7107.

作者简介



刘乐(通信作者),2015年于燕山大学获得博士学位,现为燕山大学自动化系副教授,主要研究方向为复杂非线性系统智能优化控制、基于深度学习的图像识别算法及在流程工业中应用研究。

E-mail: leliu@ysu.edu.cn

Liu Le (Corresponding author) received his Ph. D. degree from Yanshan University in 2015. Now he is an associate professor at the Department of Automation, Yanshan University. His main research interests include intelligent optimization control

of complex nonlinear systems, image recognition algorithm based on deep learning and its application in process industry.



张晓松,2019年于河北大学工商学院获得学士学位,现为燕山大学硕士研究生,主要研究方向为场景文本检测与识别。

E-mail: Delay0123456789@163.com

Zhang Xiaosong received his B. Sc. degree from Industrial and Commercial College, Hebei University in 2019. Now he is a M. Sc. candidate at Yanshan University. His main research interests include scene text detection and recognition.



黄锋,2021年于华东交通大学获得学士学位,现为燕山大学硕士研究生,主要研究方向为目标检测与跟踪。

E-mail: hfajcd@163.com

Huang Feng received his B. Sc. degree from East China Jiaotong University in 2021. Now he is a M. Sc. candidate at Yanshan University. His main research interests include object detection and tracking.



方一鸣,2003年于燕山大学获得机械与电子工程博士学位。现为燕山大学自动化系教授,主要研究方向为连铸轧钢自动化智能化技术及应用,图像识别、智能优化算法及在流程工业中应用。

E-mail: fymingysu@163.com

Fang Yiming received his Ph. D. degree from Yanshan University in 2003. Now he is a professor at the Department of Automation, Yanshan University. His main research interests include intelligent technology and application of continuous casting and rolling steel automation, image recognition, intelligent optimization algorithm and application in process industry.