

## 基于改进 PSENet-CRNN 的工件编码字符识别算法

熊新炎<sup>1</sup> 马宏伟<sup>1</sup>

XIONG Xinyan MA Hongwei

## 摘要

由于传统识别方法在面对工业环境的复杂性时表现出较低的准确率和鲁棒性,提出了一种基于改进 PSENet-CRNN 的工件编码字符识别算法。结合了 PSENet 在字符分割方面的优势和 CRNN 在序列建模方面的特长,旨在提高在复杂工业环境下的字符识别性能。通过改进网络结构、优化训练策略等手段,新算法有效解决了字符粘连、光照不均、背景干扰等问题,显著提升了识别准确率和鲁棒性。实验结果表明,新算法在多种评价指标上均优于传统方法和其他先进算法,为智能制造领域提供了一种高效、稳定的字符识别解决方案。

## 关键词

工件编码; 字符识别; PSENet; CRNN; 评价指标

doi: 10.3969/j.issn.1672-9528.2024.05.016

## 0 引言

随着工业 4.0 时代的到来,工件编码字符识别作为工业生产智能制造中的关键一环,对于工业生产制造具有重要意义。传统的字符识别方法在面对复杂多变的工业环境时,往往表现出较低的识别准确率和鲁棒性。鉴于此,研发一种高效且稳定的工件编码字符识别算法显得尤为迫切和重要,这对于提升工业生产效率和减少人为错误具有重要意义。

近年来,深度学习在图像处理与模式识别领域展现出了令人瞩目的进步。特别是 CRNN 模型,巧妙地将 CNN 与 RNN 相结合,在自然场景下的文字识别任务中取得了卓越的性能。此外, PSENet 作为一种创新的字符识别算法,通过逐步扩展字符的分割范围,有效地解决了字符间粘连的难题,为字符识别提供了全新的思路。然而,在现实应用场景中,工件编码字符的识别面临着诸多挑战。光照不均、背景干扰以及字符磨损等问题常常导致识别准确率的下降。尽管 PSENet 算法在字符识别方面具有一定的优势,但在应对这些复杂情况时,其表现仍显不足<sup>[1]</sup>。

因此,本文提出一种基于改进 PSENet-CRNN 的工件编码字符识别算法。该算法旨在结合 PSENet 在字符分割方面的优势和 CRNN 在序列建模方面的特长,通过改进网络结构、优化训练策略等方式,提高算法在复杂工业环境下的字符识别性能。本研究不仅有望为智能制造领域提供一种新的字符识别解决方案,也能够推动深度学习在字符识别技术中的进

一步发展。

## 1 基于改进 PSENet 的工件编码字符检测定位算法

## 1.1 ShuffleNetv2

在工业环境中,检测精度和速度至关重要,因此轻量级模型更受欢迎。VGG-16、ResNet、DenseNet 等网络虽特征提取能力强,但参数量和计算量大<sup>[2]</sup>。ShuffleNetv2 通过组卷积和通道混合改进 ResNet,显著减少计算量,实现高效字符识别<sup>[3]</sup>。组卷积通过分组操作降低计算复杂度。传统卷积和组卷积分别如图 1(a)和图 1(b)所示。

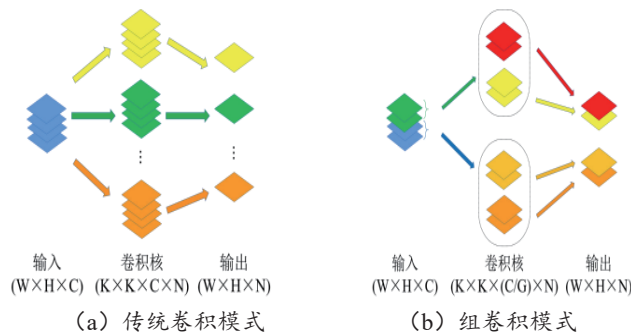


图 1 传统卷积与组卷积对比

从图 1 中可知,传统卷积参数数量为  $K \times K \times C \times N$ ,而组卷积减少至  $K \times K \times (C/G) \times N$ ,降低了  $G$  倍的参数量和计算量,加速了网络训练与推理<sup>[4]</sup>。然而,卷积操作在特征提取的过程中,往往会造成通道间的信息隔离,这在一定程度上限制了特征的全面表达。为了打破这种局限,提升通道间的信息交流效率,提出了通道混合机制,如图 2 所示,ShuffleNet 网络单元结构如图 3 所示。

1. 哈尔滨商业大学 黑龙江哈尔滨 150028

[基金项目] 博士科研项目“基于多维微型传感器人体动作识别技术研究”(2017BS023)

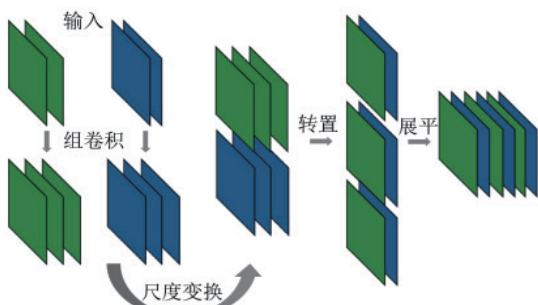
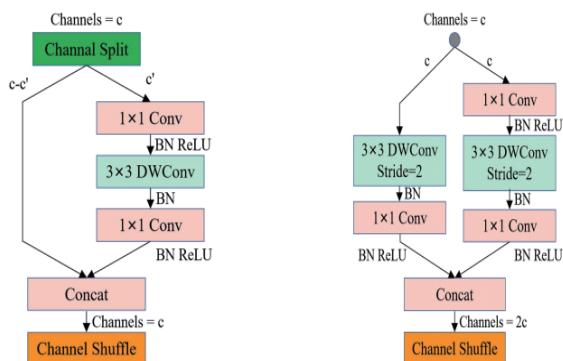


图2 通道混合图



(a) 普通 ShuffleNet 单元 (b) 下采样 ShuffleNet 单元

图3 通道混合图

图3(a)的 ShuffleNet 单元分离特征通道，其中一路不卷积，另一路用三通道卷积，汇合用 concat 替代 ResNet 的 add，并混合通道。图3(b)为下采样和通道维数扩展，不分离通道以保证 concat 后通道翻倍，用步长2的 DWConv 缩小特征图，实现下采样。特征复用增强表达能力。

## 1.2 改进 PSENet

基于语义分割的渐进式尺度扩展网络 PSENet 通过像素级别的分割来构建完整的字符行，从而有效区分背景和字符实例<sup>[5]</sup>。因此，通过将标注矩形按照  $d_i$  个像素数进行缩放，见式(1)，可以得到缩放后的真实分割图。

$$d_i = \frac{A(p_n) \times (1-r^2)}{P(p_n)} \quad (1)$$

$$r_i = 1 - \frac{(1-m)(n-i)}{n-1} \quad (2)$$

式中： $d_i$  为需要缩放的像素数值， $A(\cdot)$  为标注框的面积， $P(\cdot)$  为标注框周长， $r_i$  为缩放比例， $m$  为最小缩放比例， $n$  为分割结果总个数。

## 2 基于改进 CRNN 的工件编码字符检测识别算法

为了解决单字符检测中可能出现的漏检问题，本文在识别部分引入了 CRNN 算法，它通过充分利用序列的前后向信息来增强识别能力。然而，CRNN 在识别过程中对于形状相似的字符容易发生误识别，因此为了有效解决上述情况，本

文在多头注意力机制的基础上对 CRNN 进行优化和改进，从而提升其识别性能。

### 2.1 多头自注意力机制

Transformer 模型的独特之处在于其运用了多头自注意力机制，因此其在序列建模任务中得到了广泛的应用<sup>[6]</sup>。多头注意力机制所携带的丰富语义信息对于模型的性能至关重要。为了更深入地提升注意力机制对位置信息的敏感度，从而增强字符识别的准确性，本文在研究中特别引入了一种创新的相对距离感知位置编码技术。这种位置编码技术能够捕捉到字符间的相对距离信息，并将其融入注意力机制中，使模型在处理字符序列时能够更准确地关注到字符的位置特征<sup>[7]</sup>。选取4个多头自注意力的头模型训练和算法性能对比实验，如图4。

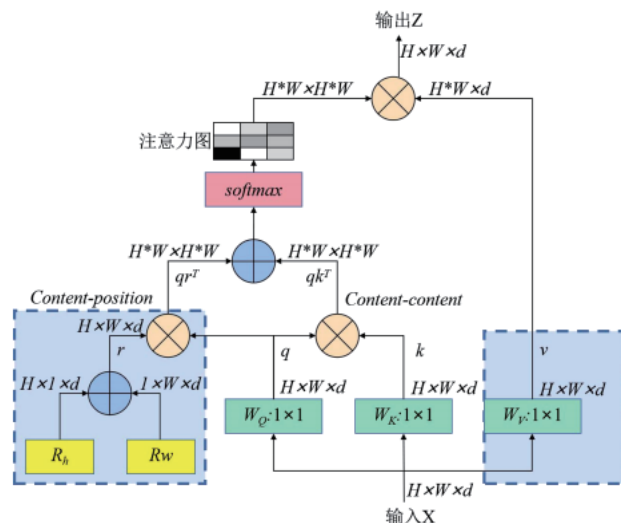


图4 多头自注意力机制

通过引入位置信息  $h$  和  $w$ ，增强了模型对序列中元素位置关系的感知能力。在多头自注意力机制中，内容信息注意力矩阵与位置信息注意力矩阵被巧妙地结合起来。首先，内容信息通过多头注意力机制计算得到一系列注意力权重，这些权重反映了不同元素之间的内容相关性。同时，位置信息  $h$  和  $w$  也被转换为相应的注意力矩阵，以捕捉序列中元素的位置依赖关系。随后，内容信息注意力矩阵与位置信息注意力矩阵进行相加操作，将内容相关性和位置关系信息融合在一起。这样，模型在关注内容的同时，也能够充分考虑到元素在序列中的位置信息。经过相加操作后获得的注意力矩阵，为了进一步确保其中的值在合理范围内并凸显关键信息，采用 softmax 函数对其进行归一化处理。这一步骤能够使得注意力矩阵中的元素值更加符合概率分布，即所有元素值均位于0到1之间，并且和为1，从而更有效地体现出各个位置的重要性程度，为后续的特征提取和识别任务提供有力的支持。将其转化为注意力激活图  $H \times H \times H \times N$ 。这张激活图直观地展示了模型中不同元素之间的注意力分布，为后续的序列建模任务提供了重要的指导信息。最后将激活图与  $v$  做矩阵

乘法得到输出  $Z:H \times W \times d^{[8]}$ 。多头自注意力计算公式为：

$$Attention(q, k, v) = \text{soft max} \left( \frac{q(k^T + r^T)}{\sqrt{d_k}} \right) \quad (3)$$

式中： $d_k$  表示  $q$  矩阵的列数。 $d_k$  使得梯度训练更加稳定。

多头注意力机制通过采用多个独立的注意力单元进行并行计算，这一设计有助于模型更全面地捕捉图像中的信息，提高识别的准确性。同时，为了避免过拟合问题，采用了集成学习的策略，将多个注意力单元的输出进行融合，以获得更加稳定可靠的结果。在线性转换后得到的  $q$ 、 $k$ 、 $v$ ，每个注意力机制只负责处理最终输出中的一部分信息，它们之间相互独立，协同工作，共同构成了这一强大的多头注意力机制<sup>[9]</sup>。为达到模型的轻量化和减少输入宽度降采样的需求，本文选择 ResNet18 作为基础架构，并巧妙地融合了 MHSA 技术，以便更高效地提取特征。

## 2.2 损失函数改进

CRNN 对于 CTC Loss 在训练样本的标签制作上无需严格对齐，它通过序列到标签的转换方式实现识别<sup>[10]</sup>。然而，CTC Loss 对于形近字的识别能力存在局限，难以准确区分。因此，在处理形近字识别任务时，需要采取额外的策略或方法以提升识别的准确性。为了解决上述问题，借鉴 Wen 等人<sup>[11]</sup>Center Loss 式(4)的思想。

$$L_{center} = \frac{1}{2} \sum_{i=1}^m \|x_i - c_{yi}\|_2^2 \quad (4)$$

式中： $x_i$  为样本字符特征， $c_{yi}$  为样本中心值。

构造总损失函数为：

$$L_{total} = L_{CTC} + \lambda L_{center} \quad (5)$$

式中： $\lambda$  表示 Center Loss 的权重。

## 3 实验结果与分析

### 3.1 实验流程

基于改进 PSENet-CRNN 的字符检测算法实验具体流程如下。

- (1) 初始化参数并训练 PSENet 定位网络。
- (2) 加载标注数据，提取 ShuffleNetv2 特征图  $F$ 。
- (3) 经卷积、上采样和 Sigmoid 处理，得到不同尺寸的分割掩膜，通过损失函数训练网络，得到 PSENet-SF 模型。
- (4) 初始化并训练多头自注意力 CRNN 识别网络。
- (5) 加载数据，缩放图片高度后提取 ResNet18-MHSA 特征图，输入双向 LSTM 分类，利用 CTC 损失函数训练，得到 CRNN-RMC 模型。

(6) 加载测试集，利用 PSENet-SF 定位，裁剪字符区域，再输入 CRNN-RMC 识别，验证模型性能。

在实验过程中，使用框准率 (precise boundingbox rate, PBR) 和图片识别率 TIIR (total image identification rate, TIIR) 评估算法性能。PBR 计算公式为：

$$PBR = \frac{N_{test}}{N_I} \quad (6)$$

式中： $N_{test}$  代表测试集中定位成功的图片总数， $N_I$  表示图片总数。

TIIR 计算公式为：

$$TIIR = \frac{I_P}{N_I} \quad (7)$$

式中： $I_P$  表示完全正确检测的图片， $N_I$  表示图片总数。

在本文的预先训练过程中，设定了迭代步数为 30 000 次，以确保模型能够充分学习并优化其参数。为了确保模型在训练过程中能够稳定收敛，设置学习率为 0.001，这一策略有助于模型逐步优化其权重，实现更加精确的识别效果。同时，为了进一步提升模型的泛化能力，采用滑动平均策略来更新模型的参数。这一策略通过平滑参数的变化，有效减少了过拟合的风险，使得模型在面对不同场景和输入时能够展现出更加稳定和可靠的性能。此外，为了更有效地调整学习率以适应不同的训练阶段，采用了指数衰减的方式来更新学习率。具体的操作为每迭代 10 000 步，学习率会按照 0.95 的衰减率进行降低。这种策略有助于在训练初期快速收敛，而在后期则能够更精细地调整模型参数。为了验证本文提出的改进网络的有效性和适用性，使用钢管序列号数据集作为主要实验对象进行验证。

### 3.2 不同定位算法的实验对比

将 PSENet-SFHufflev2-FPN 命名为 PSENet-SF，为了检测 PSENet-SF 的定位性能，与 AdadvancedEAST 等当前 11 种主流定位算法进行对比实验，数据集采用钢管序列号数据集，实验数据如表 1。

表 1 不同定位算法实验对比

| 定位算法网络          | PBR/% | Speed/ms |
|-----------------|-------|----------|
| Adadvanced EAST | 95    | 117      |
| TextBoxes++     | 63    | 89       |
| Faster R-CNN    | 85    | 258      |
| YOLOv3(608*608) | 58    | 78       |
| SSD(300*300)    | 58    | 118      |
| DETR            | 93    | 139      |
| PixelLink       | 58    | 128      |
| SegLink         | 73    | 160      |
| YOLOv4(608*608) | 69    | 60       |
| DBNet           | 95    | 145      |
| CTPN            | 99    | 430      |
| Faster R-CNN-FF | 89    | 274      |
| PSENet          | 99    | 86       |
| PSENet-SF       | 100   | 38       |

### 3.3 迁移学习 CRNN 实验对比

为了深入探究 CRNN 识别算法在采用迁移学习后的性能提升，本文将设计并实施一系列实验，用来对比不同方法下模型的识别效果，从而找出最佳的优化策略。在本次实验中，采用 ICPR 数据集作为迁移学习的源数据集，首先在该数据

集上对 CRNN 模型进行预训练，然后将其应用于钢管序列号数据集进行进一步的训练。其结果将如表 2 所示。

表 2 CRNN 迁移学习实验对比

| 基础算法网络          | Epoch | TIIR/% | Speed/ms |
|-----------------|-------|--------|----------|
| CRNN-VGG16-nf   | 402   | 58     | 45       |
| CRNN-VGG16-f    | 43    | 87     | 45       |
| CRNN-VGG16-s-nf | 402   | 65     | 45       |
| CRNN-VGG16-s-f  | 42    | 85     | 45       |
| CRNN-VGG16-e-f  | 42    | 93     | 45       |

从实验结果表中可以得到，采用迁移学习策略的模型相较于未采用迁移学习的模型，在整张识别率上均呈现出更高的性能。不仅如此，迁移学习的应用还使得模型在训练过程中所需的 epoch 数量显著减少，从而提高了训练效率。另一方面，原图数据集增强的方法能更好地展现字符场景的多样性，展现出了对字符场景的鲁棒性，因此在效果上优于采用模拟生成数据的方法。

3.4 定位结合识别实验对比

将改进的识别算法重新命名为 CRNN-RMC，选取 3.2 节中实验结果排名前列的几种定位算法与改进的识别算法进行结合，实现工件编码字符识别功能。定位结合识别实验对比结果如表 3。

表 3 定位结合识别实验对比

| 基础算法网络                | TIIR/% | Speed/ms |
|-----------------------|--------|----------|
| DBNet+CRNN-RMC        | 90     | 169      |
| AdvancedEAST+CRNN-RMC | 87     | 158      |
| CTPN+CRNN-RMC         | 91     | 479      |
| DETR+CRNN-RMC         | 87     | 172      |
| Faster R-CNN-FF       | 79     | 318      |
| PSENet-SF+CRNN-RMC    | 97     | 73       |

经过 3.2 节、3.3 节和 3.4 节的实验验证发现，PSENet-SF 与 CRNN-RMC 的结合在定位与识别任务中展现出了卓越的性能。相较于其他算法组合，这一组合在整张识别率和检测速度上均展现出了显著的优势。特别是在处理钢管序列号数据集时，其整张识别率实现了大幅提升，效果尤为显著。这一优异表现主要归功于 PSENet-SF 在定位精度上的显著提升，以及 CRNN 在捕捉长距离字符间相关性方面的卓越能力。这些优势使得该组合算法在实际应用中具有更高的可靠性和实用性，为工业环境下的字符识别问题提供了有效的解决方案。

4 结语

针对智能制造中工件编码字符识别的挑战，本文提出了一种结合 PSENet 定位和 CRNN 识别的算法。为解决传统候选框定位的不精确问题，本文采用基于分割的 PSENet 算法精准定位字符区域，并引入 ShuffleNetv2-FPN 以提升定位效率。针对单字符检测易导致的漏检，本文利用 CRNN 算法，

结合字符序列的前后信息提升识别率。为克服 CRNN 可能存在的超识别问题，本文创新性地融入位置信息嵌入的 MHSA 模块至 ResNet18 中。另外，针对字符污染和低分辨率引发的形近字误识别问题，本文利用 Center Loss 与 CTC Loss 作为辅助损失函数，旨在增强网络对形近字符的区分能力，从而优化字符识别的准确性。实验表明，该算法在满足工业需求的同时，显著提高了字符识别的准确性和鲁棒性。这一研究为智能制造领域的字符识别技术提供了新的解决方案，并推动了深度学习在该领域的应用发展。

参考文献：

[1] 王宇. 复杂背景下工业字符检测和识别方法研究 [D]. 西安：西安建筑科技大学,2023.

[2] 李帆. 基于深度学习的嵌入式喷码检测系统设计与实现 [D]. 桂林：广西师范大学,2022.

[3] LI W, WANG S, ULLAH I, et al. Multiple attention-based encoder-decoder networks for gas meter character recognition[J]. Scientific reports, 2022, 12(1): 10371.

[4] 巴珊珊. 基于深度学习的工业字符检测与识别方法研究 [D]. 广州：华南理工大学,2022.

[5] WU X, CHEN Q, XIAO Y, et al. LCSEgNet: an efficient semantic segmentation network for large-scale complex chinese character recognition[J]. IEEE transactions on multimedia, 2021, 23: 3427-3440.

[6] 李建明. 基于卷积神经网络的工业字符检测方法研究 [D]. 无锡：江南大学,2021.

[7] 王光军. 基于神经网络的自然场景中的字符识别算法的研究与实现 [D]. 成都：电子科技大学,2020.

[8] 潘杰. 基于深度学习的票据识别系统的设计与实现 [D]. 武汉：华中科技大学,2020.

[9] 王宇. 复杂背景下工业字符检测和识别方法研究 [D]. 西安：西安建筑科技大学,2023.

[10] 治骞. 基于机器视觉的工业字符识别方法研究 [D]. 西安：西安理工大学,2022.

[11] WEN Y, ZHANG K, LI Z, et al. A discriminative feature learning approach for deep face recognition[C]//European Conference on Computer Vision. Cham: Springer, 2016: 499-515.

【作者简介】

熊新炎（1980—），男，湖北监利人，博士，讲师，研究方向：机器视觉。

马宏伟（1993—），通信作者（email: 568254137@qq.com），男，山西大同人，硕士，研究方向：机器视觉。

（收稿日期：2024-02-29）