基于残差优化和内容自适应的文本识别算法

王敏[1,2],黎永顺[1],欧翔[2],曹冉[2],吴佳[1]

(1.南京信息工程大学 电子与信息工程学院,南京 210044;2.安徽建筑大学 电子与信息工程学院,合肥 230601)

摘 要:西林瓶标签信息在保障患者用药安全和高效的药物管理方面发挥着关键作用.针对传统的文本识别网络对药瓶标签图像中长文本和模糊文本的识别性能差的问题,提出了一种基于残差优化和内容自适应的文本识别算法.在传统文本识别网络的基础上,采用多尺度残差特征提取模块来代替原有的特征提取卷积网络,通过优化 ResNet 网络的下采样过程并引入多尺度特征融合模块,增强了特征提取能力.同时,加入卷积注意力模块提升了网络对文本的关注,增强了网络对低分辨率文本的识别能力.其次,在序列建模阶段,融合多层双向内容自适应递归单元和自注意力机制,提升了长文本序列的建模能力.实验结果表明,与卷积递归神经网络文本识别网络相比,本算法识别准确率提高了 3.92%,相较于其他文本识别网络相比均有一定的提升.

关键词:文本识别;CRNN;特征提取;内容自适应循环单元;自注意力机制

中图分类号:TP391.4 **文献标志码**:A **文章编号**:1000-2367(2026)02-0030-08

西林瓶作为一种常见的药物包装形式,广泛用于液体药物的储存和分配中,其标签信息的准确识别对于医疗行业而言至关重要[1].西林瓶标签通常包含了药物的重要信息,如药物名称、生产批号、有效期、质量浓度等,这些信息在保障患者用药安全和高效的药物管理方面发挥着关键作用[2-4].

近些年来,深度学习以其出色的特征学习和表达能力在自然场景文本处理领域取得了显著的成绩,效果和效率都已远超人类水平,涌现出一批优秀的文本识别算法[5],主要分为基于序列预测和基于注意力机制.其中基于序列预测的文本识别算法是将单词视为不同长度的序列,先对单词序列进行预测,然后对预测结果进行连接得到完整的文本信息.SHI 等[6]提出了一种基于图像序列识别的端到端的卷积递归神经网络,主要由卷积层、循环层和转录层三部分组成.卷积层使用循环神经网络[7]进行预测,并采用连接主义时间分类层[8]对序列进行解码,最终得到字符序列的预测结果,该方法能高效处理时序数据,具有良好鲁棒性和无需字符位置标签等优势,但存在计算开销大、长序列依赖问题、对大规模数据的依赖及难以应付复杂背景噪声等问题.WANG 等[9]提出了一种门控递归卷积神经网络.该模型在递归卷积层上添加门来控制上下文调制,平衡前馈信息和循环信息,并建立高效双向长短时记忆网络[10],提高了文本识别准确率,但是计算复杂度高,在对长序列处理时可能遇到梯度消失问题,且对数据量和硬件资源要求较大.为消除噪声对解码器效果的影响,XUE 等[11]于 2023 年提出将场景文本识别分解为两个相关联的任务,先从图像中检测候选字符再通过对检测到的候选字符中的单词进行解码来识别场景文本,从字符语义而不是图像噪声特征直接学习,可以有效纠正错误检测到的候选字符.基于注意力的文本识别算法在编-解码结构中使用注意力机制实现序列的循环解码,使模型在学习过程中能实现参数权重的自动调整,让注意力集中于感兴趣的特征,排除无用信息

收稿日期:2024-11-02;**修回日期**:2025-03-02.

基金项目:国家自然科学基金(41775165;41775039);安徽省高校杰出青年科研项目(2023AH020022).

作者简介(通信作者):王敏(1983—),女,陕西咸阳人,南京信息工程大学教授,博士,研究方向为信号与信息处理,E-mail:yu0801@163.com.

引用本文:王敏,黎永顺,欧翔,等.基于残差优化和内容自适应的文本识别算法[J].河南师范大学学报(自然科学版),2026,54(2):30-37.(Wang Min,Li Yongshun,Ou Xiang,et al.Residual optimization and content-adaptive text recognition algorithm[J].Journal of Henan Normal University(Natural Science Edition),2026,54(2):30-37.DOI:10.16366/j.cnki.1000-2367.2024.11.02.0002.)

的干扰.LI 等[12]提出了一种不规则文本识别方法,在基于 LSTM 的编解码器框架中引入二维注意力模块,增强了模型对不规则文本的识别能力,但是在处理长文本或大规模数据时,训练效率较低,二维注意力机制的复杂性也可能影响模型优化,且在复杂背景或噪声下,识别精度可能下降.LEE 等[13]提出了自注意文本识别网络,在 Transformer 结构的基础上,使用二维自注意力机制来描述场景文本图像中字符的二维空间依赖性,实现对任意排列和大字符间距的文本的识别.ZHENG 等[14]提出了多域字符距离感知(multi-domain character distance perception,MDCDP),遵循交叉注意机制,利用位置嵌入对视觉特征和语义特征进行查询,使用位置嵌入来遵循交叉注意机制以及查询视觉和语义信息,随后生成包含多领域的字符距离信息的内容感知嵌入.堆叠多个 MDCDP 结构形成 CDistNet 文本检测网络,解决在困难文本处理时特征和字符错位的问题.

这些算法在自然场景文本处理领域均表现出较好的处理性能,但在医学药品信息提取方面的应用较少,而且大多数西林瓶标签缺少可识别的条形码,使得采用常规的方法进行标签信息提取十分困难.本文针对传统的文本识别网络对药瓶标签图像中长文本和模糊文本的识别性能差的问题,提出了一种基于残差优化和内容自适应的文本识别算法.

1　本文方法

卷积递归神经网络(CRNN)结合了卷积神经网络(CNN)和递归神经网络(RNN),用于时序依赖的图像序列或文本识别任务.CRNN 通过卷积层提取图像特征,再通过长短时记忆(LSTM)或门控循环单元(GRU)进行时序建模,捕捉字符的顺序信息.最后,利用 CTC 损失函数解码预测结果,无需字符位置标签,适用于不规则文本和手写体识别,但是计算复杂度较高,尤其是在处理长序列时,训练和推理速度较慢.本文在 CRNN 文本识别模型的基础上,提出了一种基于残差优化和内容自适应的西林瓶标签文本识别算法(residual optimization and content-adaptive text recognition network,ROACNet),其网络结构如附录图 S1 所示.首先提出了多尺度残差特征提取网络(multi-scale residual feature extraction network,MRFENet)来代替原有的特征提取网络,MRFENet 在 ResNet34 的基础上对下采样过程进行优化,并添加特征融合网络,充分融合各级特征信息,获得同时包含丰富语义信息和细节信息的特征图像,提高了网络对模糊文本的识别能力;然后在循环层中提出了基于自注意力的内容自适应循环模块(self-attention-based content adaptive loop module,SCALM)对特征序列进行预测,在 SCALM 模块中使用双向内容自适应循环单元代替原始的 BiLSTM 模块,并在序列预测前添加自注意力模块,提高网络对整体文本重要性的关注,提高了网络对长文本序列的识别能力.最后通过转录层,将循环层输出的概率矩阵中的每一个向量转换为标签序列,最终输出识别到的文字结果.

1.1　基于多尺度特征的深度残差网络的特征提取

针对传统文本识别网络对模糊文本识别效果差和文本特征的提取能力差的问题,提出了基于多尺度的深度残差网络的特征提取网络(MRFENet).MRFENet 模块在 ResNet34 的基础上通过堆叠卷积层进行下采样操作并采用添加 padding(在输入图像的边缘添加额外的像素值)的 3×3 卷积操作代替 7×7 的卷积下采样操作,提高对小尺寸文本的识别能力.同时在残差块之间添加卷积注意力模块(CBAM[15],convolutional block attention module),提高网络的特征提取能力,并增加多尺度融合模块,融合多尺度特征信息,获得同时包含高级语义信息和底层细节特征的特征图像,充分利用各级特征信息,提高网络对不同尺寸文本的识别能力,具体网络结构如附录图 S2 所示.

1.1.1　特征提取网络优化

由于在处理文本的识别任务时,原始图像的分辨率较低,且其包含的特征信息较少.因此,本文 MRFENet 的特征提取阶段采用在原有的 ResNet34 网络的基础上进行优化得到 IDResNet34 网络,具体网络结构见表 1.在 conv1 残差块中,将 7×7 的卷积换成了 3×3 步距为 1,填充为 1 的卷积,并且使用堆叠的卷积核为 3×3,通道数为 32,步距为 2×1 的残差结构代替原始的 3×3 的池化操作进行下采样.conv3_x、conv4_x 和 conv5_x 残差块,都先通过 1×1 的卷积层后再采用 3×3 的卷积进行下采样,提高了网络的训练效果和计算

速率.

1.1.2 多尺度特征融合

西林瓶标签文本识别需要对较多的模糊文本进行识别.为更好地利用细节特征,提出了多尺度融合模块,网络结构如图 S2 所示.conv2_x,conv3_x,conv4_x,conv5_x 输出 4 个大小不同的特征图 $F1,F2,F3,F4$.将不同尺寸的特征图自下而上进行特征融合,生成包含高级语义信息和细节特征的特征图.设融合后特征为 $Q_i(i=1,2,3,4)$,其计算过程为:

表 1 网络结构

Tab. 1 Network structure

Layer name	ResNet34		IDResNet34	
conv1	7×7,64	stride 2	3×3,64,pad1	stride1×1
	3×3max pool	stride 2	$\begin{bmatrix}3\times3, & 32\\ 3\times3, & 32\end{bmatrix}\times3$	stride2×1
conv2_x	$\begin{bmatrix}3\times3, & 64\\ 3\times3, & 64\end{bmatrix}\times3$	stride 2	$\begin{bmatrix}3\times3, & 64\\ 3\times3, & 64\end{bmatrix}\times3$	stride2×2
conv3_x	$\begin{bmatrix}3\times3, & 128\\ 3\times3, & 128\end{bmatrix}\times4$	stride 2	$\begin{bmatrix}1\times1, & 128\\ 3\times3, & 128\end{bmatrix}\times4$	stride2×2
conv4_x	$\begin{bmatrix}3\times3, & 256\\ 3\times3, & 256\end{bmatrix}\times6$	stride 2	$\begin{bmatrix}1\times1, & 256\\ 3\times3, & 256\end{bmatrix}\times6$	stride2×1
conv5_x	$\begin{bmatrix}3\times3, & 512\\ 3\times3, & 512\end{bmatrix}\times3$	stride 2	$\begin{bmatrix}1\times1, & 512\\ 3\times3, & 512\end{bmatrix}\times3$	stride2×1

$$Q_i = D(F_i) \oplus C(F_i), \tag{1}$$

其中,D 是下采样过程(通过卷积操作时设置较大的步幅来实现空间分辨率的降低),C 是 1×1 的卷积操作(1×1 卷积是一种特殊的卷积操作,使用大小为 1×1 的卷积核(滤波器)对输入特征图进行卷积),$\oplus$ 是元素相加运算.

1.2 基于自注意力内容自适应循环的序列建模

为了解决传统卷积循环神经网络(convolutional recurrent neural network,CRNN)[6]因编码和解码使用固定长度矢量引起的处理长序列模型性能下降的问题,提出了一种基于自注意力内容自适应循环的序列模型(sequence of content adaptive loop modell,SCALM),结构如图 1 所示,采用 BiCARU 模块代替原始 BiLSTM 网络进行特征序列的标签预测,并引入自注意力机制,使模型能够选择性地关注输入序列中信息量最大的部分,提高模型对长序列的处理能力.

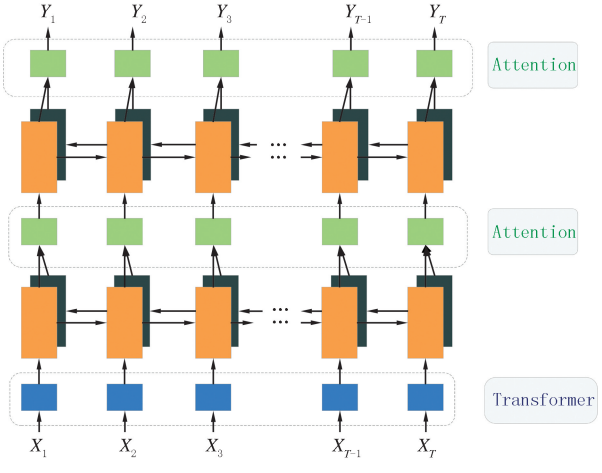


图1 SCALM模型的网络结构

Fig. 1 Network structure of SCALM

1.2.1 BiCARU

内容自适应循环单元(content adaptive recurrent unit,CARU)在传统 RNN 的基础上引入了内容自适应门,利用权值的概念来分析隐藏状态的转换,同时考虑单词文本本身和文本整体内容的重要性,可以提高模型文本识别的能力,结构如图 2 所示.其中 $(W;B)$ 是类似式(2)的线性函数,tanh 和 σ 分别为范围和范围(−1,1)和(0,1)内的 S 型激活函数.tanh 是为了防止训练中的分歧,σ 的结果作为后续操作的参考量.但 CARU 只使用过去的上下文,为充分利用两个不同方向的上下文信息,本文选择采用两个方向的 CARU 组成双向的 BiCARU 模块,并将两个 BiCARU 模块进行堆叠,产生深度 BiCARU 模块,实现对序列预测,输出对应所有字符的概率分布向量.

首先,输入由当前单词映射得到的特征 x,用于生成下一个隐藏状态,同时传递到内容自适应门:

$$x^{\cdots} = Wv^{\cdots} + B. \tag{2}$$

CARU 结合 $h^{\cdots}$ 和 $x^{\cdots}$ 相关参数来产生一个新的隐藏状态 n:

$$n''' = \tanh((Wh''' + B) + x'''). \tag{3}$$

z''' 是一个更新门,用于隐藏状态的转化:

$$z''' = \sigma(Wh''' + B + Wv''' + B), \tag{4}$$

$$l''' = \sigma(x''') \odot z'''. \tag{5}$$

l''' 使用 Hadamard 算子更新门 z''' 与输入特征 x''',称为内容自适应门,防止输入特征对隐藏状态的稀释:

$$h''' = (1 - l''') \odot h''' + l''' \odot n''', \tag{6}$$

其中,h''' 为输入序列,n''' 为当前时刻的隐藏状态,两者结合生成下一个隐藏状态 h'''.

CARU 选择根据当前文本对隐藏状态进行加权,并引入内容自适应门,来减轻对长期内容的依赖.CARU 中上述流程主要分为 3 个部分:内容状态(content-state):通过线性层生成新的隐藏状态 n,等价简单的递归神经网络.文本权重(word-weight):生成当前文本的权重 σ,用以连接权重和词性之间的关系.内容权重(word-weight):生成当前内容的权重 z,其目的是克服对长期内容的依赖性.CARU 直接将文本权值发送给提议的门,并将其乘以内容权值,通过这种方式,内容自适应门同时考虑到了单词和内容的重要性.

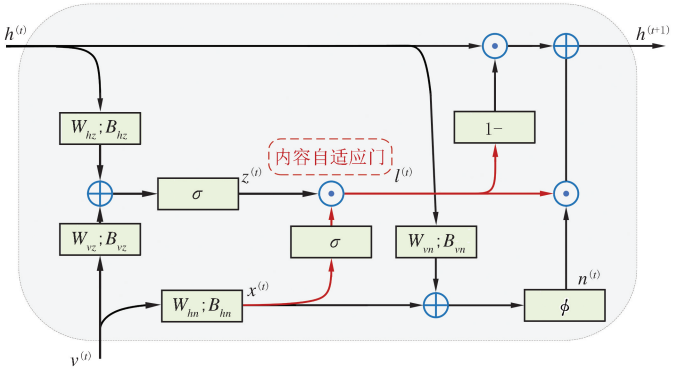


图2　CARU结构图

Fig.2　CARU structure

1.2.2　自注意力机制

由于中文文本类型多,结构相似且复杂,为进一步提高文本识别的准确率,在上述基础上引入自注意力机制(self-attention mechanism[16-17]).自注意力机制通过获取全局序列文本的注意力,提高对长序列文本的识别能力.

改进后的 BiCARU 网络结构如图 3 所示,将特征提取网络获取的特征序列 X_{T} 输入到自注意力机制中.首先,对每个输入元素,通过线性变换生成用于计算该位置对其他位置的注意力分布的查询向量($\boldsymbol{Q}$)、用于提供其他位置信息注意力分布的键向量($\boldsymbol{K}$)和与该位置相关的信息注意力分布的值向量($\boldsymbol{V}$)3 种不同的向量,然后通过式(7)计算得到注意力的权重,并将每个位置的 $\boldsymbol{V}$ 向量按照对应的注意力权重进行加权求和得到注意力表示.接着将原始输入序列与得到的注意力表示相加,并对得到结果进行归一化,从而得到编码结果 X.再将经过自注意力机制编码的结果送入到 BiCARU 中捕捉序列的局部语义信息和长期依赖关系,从而得到输出结果 y,用于后续的标签序列预测.上述计算过程如下式:

$$\mathrm{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \mathrm{softmax}(\frac{\boldsymbol{QK}}{\sqrt{d}})\boldsymbol{V}, \tag{7}$$

$$X_{\cdots} = \mathrm{self\_attention}(X), \tag{8}$$

$$y = \mathrm{BiGARU}(X). \tag{9}$$

2　实验结果与分析

2.1　实验设置及评价指标

2.1.1　实验设置

本文实验编程语言为 python3.7,深度学习框架为 pytorch1.8.1,实验平台使用 Linux 系统,CPU 为 Intel(R)Core(TM)i7-11200H CPU@2.40 GHz,GPU 为 NVIDIA RTX2080Ti,11 G 运行内存.

2.1.2 评价指标

文本识别的评价指标是行识别准确率(a),即把一个文本行作为基本单位,计算识别正确的单行文本数量 N 占总文本行数量 M 的比值

$$a=\frac{N}{M}\times 100\%. \tag{10}$$

2.2 数据集介绍

目前公开的文本检测和识别数据集主要针对街景文本且语言大多为英文,与西林瓶标签有较大的差别.本文采用 MV-CL022-40GC 线扫描相机进行西林瓶图像采集,依次进行裁剪、尺度归一化、空白填充后,再利用百度飞桨的半自动化图形标注工具 PPOCRLabel 进行标注和人工校对,最终得到 7 384 张单行文本图片数据集,创建了西林瓶标签文本检测数据集(vial label text recognition dataset,VLTRD).对获得的数据集进行划分得到训练集和测试集,其中 6 784 张为训练集图片,600 张为测试集图像.附录图 S3 为西林瓶标签文本检测数据集中的文本识别标注示意图.

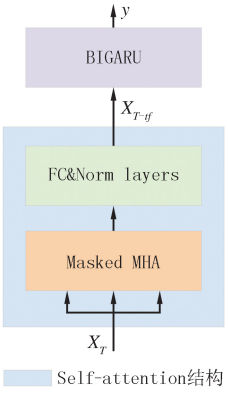
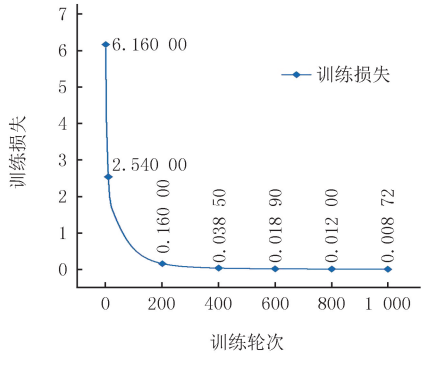


图3 改进后的BiCARU网络结构

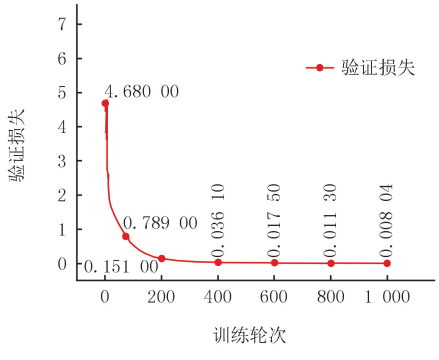
Fig.3 Improved bicaru network structure

2.3 模型训练

使用 VLTRD 数据集进行训练时,6 784 张图片作为实验的训练集,另外 600 张图片作为实验的测试集,采用 Adam 作为训练的优化器,训练周期设置为 1 000 轮,训练的 batchsize 设置为 64,初始学习率设置为0.001.训练得到网络的损失函数和网络模型精度曲线如图 4 和图 5 所示.



(a) 训练集损失函数曲线



(b) 测试集损失函数曲线

图4 网络模型的损失函数图

Fig.4 Loss function diagram of network model

从图 4 和图 5 可以看出,在训练开始时,网络损失函数值快速下降,当训练到 400 轮之后,网络损失函数值波动呈现震荡,模型逐渐收敛.同时,随着训练开始,精度一直在提升,当训练到 400 轮左右时,精度曲线开始震荡,此时模型开始收敛.

2.4 消融实验对比分析

为验证本文提出网络模型的有效性,通过对比验证不同改进措施在创建的 VLTRD 数据集以及现有标准数据集 ICDAR2013,SVT,CTW1500 和Ⅲ T5K 上的效果,设置了若干消融实验,其中 ICDAR2013 和Ⅲ T5k数据集包含不同序列长度的文本,可用于检测本文方法对不同序列长度文本检测能力的提升.SVT 和 CTW1500 数据集主要是通过采集街景图片进行截取得到,图片质量不高,存在较多模糊文本,可用于检测本文方法对模糊文本检测精度的提升.

2.4.1 参数优化后的残差网络对识别结果的影响

使用优化后的残差网络代替原本的特征提取网络 VGG16,其他模块保持不变,来验证使用优化的深度残差网络能否提高识别的准确率,识别结果见表 2.

从表 2 可知,在 4 种数据集上,将特征提取网络从 VGG16 修改为本文优化的深度残差网络后,识别效

果分别提升了 2.92%、3.04%、2.66%、2.82%和 2.70%.

2.4.2　增加多尺度特征融合对识别结果的影响

在传统卷积循环神经网络结构中加入了多尺度特征融合模块，其他模块保持不变，来验证使用多尺度融合模块是否能提高识别的准确率，识别结果见表 3.

从表 3 可知，在 4 种数据集上，增加了多尺度特征融合模块后，本方法的识别效果分别提升了 1.64%、2.26%、1.22%、1.55%和 1.26%.

从表 2、表 3 共同看出，本方法在 ICDAR2013、Ⅲ T5K 数据集上，通过引入基于多尺度特征深度残差网络的特征提取模块，有效地提高了网络对长文本的检测能力.

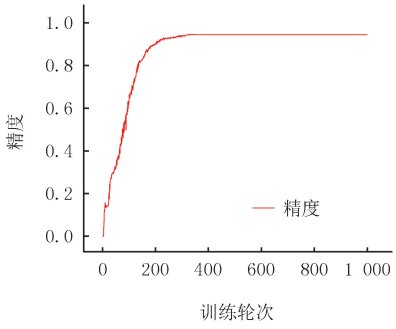


图5　网络模型精度曲线

Fig. 5　Accuracy curve of network model

2.4.3　使用 BiCARU 对识别结果的影响

选择使用 BiCARU 来代替 CRNN 网络中的 BiLSTM，其他模块保持不变.即在不改变网络特征提取方式和转录编码的基础上，验证使用 BiCARU 进行序列建模能否提高模型识别的准确率，实验结果见表 4.

表 2　深度残差网络对识别结果的影响

Tab. 2　Effect of depth residual network on recognition results %

网络	ICDAR2013	Ⅲ T5K	SVT	CTW1500	VLTRD
VGG16	86.70	78.20	80.80	61.82	90.70
本文	**89.62**	**81.24**	**83.46**	**64.60**	**93.40**

表 3　多尺度融合模块对识别结果的影响

Tab. 3　Effect of multi-scale fusion module on recognition results %

特征融合	ICDAR2013	Ⅲ T5K	SVT	CTW1500	VLTRD
无	86.70	78.20	80.80	61.82	90.70
有	**88.34**	**80.46**	**82.02**	**63.37**	**91.96**

从表 4 中可知，在 4 种数据集上，使用 BiCARU 进行序列建模后，模型识别准确率分别提升了 1.26%、1.76%、0.62%、1.66%和 1.14%.

2.4.4　加入自注意力机制对识别结果的影响

选择在序列建模的过程中加入自注意力机制来获取全局的特征信息，其他模块保持不变，实验结果见表 5.

表 4　BiCARU 对识别结果的影响

Tab. 4　Effect of BiCARU on recognition results %

模块	ICDAR2013	Ⅲ T5K	SVT	CTW1500	VLTRD
BiLSTM	86.70	78.20	80.80	61.82	90.70
BiCARU	**87.96**	**79.96**	**81.42**	**63.48**	**91.84**

表 5　自注意力机制对识别结果的影响

Tab. 5　Effect of self attention mechanism on recognition results %

自注意力	ICDAR2013	Ⅲ T5K	SVT	CTW1500	VLTRD
无	86.70	78.20	80.80	61.82	90.70
有	**87.52**	**80.14**	**82.46**	**63.16**	**91.56**

从表 5 可以得知，在 4 种数据集上，特征提取后插入自注意力机制模块，其他模块保持不变，模型识别效果的准确率分别提升了 0.82%、1.94%、1.66%、1.34%和 0.86%.表 4、表 5 共同表明，本文方法在 SVT 和 CTW1500 数据集上，引入基于自注意力内容自适应循环的序列建模模块，可以有效提高网络对模糊文本的识别能力.

2.5　不同算法识别效果性能对比

为进一步验证本文提出的基于残差优化和内容自适应文本识别算法的模型性能，分别将本文模型与 CRNN[6]，GRCNN[9]，SAR[12]，AON[18]，ACE[19]，CRNN-PR[20]，MSF-CRNN[21]等文本识别模型进行实验，同样在上述 5 种数据集进行识别效果对比，实验结果见表 6.

表 6　不同算法的识别效果性能对比

Tab. 6　Comparison of recognition effects of different algorithms %

模型	CRNN[6]	GRCNN[9]	SAR[12]	AON[18]	ACE[19]	CRNN-PR[20]	MSF-CRNN[21]	本文
ICDAR2013	86.7	—	—	—	89.70	90.90	**94.90**	93.24
Ⅲ 5K	78.2	80.8	91.5	87	82.30	89.80	**91.30**	88.72
SVT	80.8	81.5	84.5	82.80	82.60	84.30	**90.10**	88.62
CTW1500	61.82	64.56	66.94	67.62	67.40	68.34	70.28	**70.52**
VLTRD	90.7	91.2	91.6	91.86	91.60	92.82	93.42	**94.62**

从表 6 可以看出,本文模型与基础 CRNN 网络的识别效果相比,在 ICDAR2013、SVT、Ⅲ5K 和 VLTRD 等数据集上,识别的准确率分别提升了 6.54%、10.52%、7.82%、8.7%和 3.92%.本文模型的识别结果与 GRCNN、SAR、AON 等文本识别模型相比,在 ICDAR2013、SVT、Ⅲ5K 和 VLTRD 等数据集上均有不同程度的提升,从而证明了本文提出的方法可以有效提高文本识别算法的准确率.

还可以看出,虽然本文提出的算法与最新的算法相比,在 ICDAR2013、SVT、Ⅲ5K 等数据集上体现的优势不大,但是在 VLTRD 标签文本的识别上展示出了较大的优势;与 CRNN、GRCNN、SAR、AON、ACE、CRNN-PR 和 MSF-CRNN 等算法相比,识别准确率分别提升了 3.92%、3.42%、2.76%、3.02%、1.80%、1.94%和 1.20%.说明本文提出的算法在西林瓶标签信息识别任务中具有显著的优越性.

附录图 S4 展示了本文提出的算法在模糊文本和长文本识别上的优越性.可以看出,与原始的 CRNN 文本识别模型相比,本文改进后的文本识别网络对长文本和模糊文本的识别能力有了明显的提高,改善了原始网络对相近文本识别能力弱的问题,有效提高了模型的文本识别能力.

3 结 论

本文主要实现了一种基于残差优化和内容自适应的文本识别算法模型.该方法在原始 CRNN 网络的基础上,提出了两个改进模块:一是通过在 ResNet 网络的基础上进行网络结构和参数的优化,并添加多尺度的特征提取网络,提出了 MRFENet 特征提取网络.提高了网络的特征提取能力和对模糊文本的识别能力.二是融合 BiGRAU 和自注意力机制,提出 SCALM 模块进行序列预测,使得网络能够同时考虑单词本身和文本整体的重要性,通过自注意力机制捕获不同元素之间的关系,提高了文本识别的准确性.通过消融实验及识别效果,证实了本文提出算法的有效性.该方法在特征提取和序列建模上有所进展,但计算复杂度仍较高,尤其在长文本和大规模数据处理时效率较低,主要源于 CBAM 注意力机制和多尺度融合模块增加了参数量和计算开销.未来研究应关注优化计算效率、减小模型规模,并提升在复杂背景和噪声下的鲁棒性.

附录见电子版(DOI:10.16366/j.cnki.1000-2367.2024.11.02.0002).

参 考 文 献

[1] 朱俏俏,周向荣,句秋月,等.立卧一体口服液瓶装盒系统的设计与研究[J].包装与食品机械,2023,41(2):74-79.
ZHU Q Q,ZHOU X R,JU Q Y,et al.Design and research on vertical and horizontal integrated oral liquid bottle packing system[J].Packaging and Food Machinery,2023,41(2):74-79.

[2] 唐丛,刘宗明.食品包装视觉信息传达[J].食品与机械,2023,39(12):92-99.
TANG C,LIU Z M.Research progress of visual information communication in food packaging:a systematic review[J].Food & Machinery,2023,39(12):92-99.

[3] 陈冰峰,蔡美玲.药物图标标签在手术室高警示药品管理中的应用效果分析[J].海峡药学,2023,35(10):87-90.
CHEN B F,CAI M L.Application effect analysis of drug icon label in high warning drug management in operating room[J].Strait Pharmaceutical Journal,2023,35(10):87-90.

[4] 合理用药国际网络中国中心组临床安全用药组,中国药理学会药源性疾病学专业委员会,药物不良反应杂志社.医疗机构药品条码技术应用相关用药错误防范指导原则[J].药物不良反应杂志,2020,22(1):6-11.

[5] 陈晓龙,陈显龙,袁建平,等.基于深度学习的电力设备铭牌识别[J].广西大学学报(自然科学版),2018,43(6):2216-2226.
CHEN X L,CHEN X L,YUAN J P,et al.Electricity equipment nameplate recognition based on deep learning[J].Journal of Guangxi University(Natural Science Edition),2018,43(6):2216-2226.

[6] SHI B G,BAI X,BELONGIE S.Detecting oriented text in natural images by linking segments[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR).Honolulu:IEEE,2017:3482-3490.

[7] SCHUSTER M,PALIWAL K K.Bidirectional recurrent neural networks[J].IEEE Transactions on Signal Processing,1997,45(11):2673-2681.

[8] GRAVES A,FERNÁNDEZ S,GOMEZ F,et al.Connectionist temporal classification:labelling unsegmented sequence data with recurrent neural networks[C]//Proceedings of the 23rd International Conference on Machine Learning-ICML'06.Pennsylvania:ACM,2006:369-376.

[9] WANG J F,HU X L.Gated recurrent convolution neural network for OCR[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems.California:ACM,2017:334-343.

[10] ZHANG S，ZHENG D Q，HU X C，et al.Bidirectional long short-term memory networks for relation classification[C]//Pacific Asia Conference on Language，Information and Computation.[S.l.：s.n.]，2015.

[11] XUE C H，HUANG J X，ZHANG W Q，et al.Image-to-character-to-word transformers for accurate scene text recognition[J].IEEE Transactions on Pattern Analysis and Machine Intelligence，2023，45(11)：12908-12921.

[12] LI H，WANG P，SHEN C H，et al.Show，attend and read：a simple and strong baseline for irregular text recognition[J].Proceedings of the AAAI Conference on Artificial Intelligence，2019，33(1)：8610-8617.

[13] LEE J，PARK S，BAEK J，et al.On recognizing texts of arbitrary shapes with 2D self-attention[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops(CVPRW).Seattle：IEEE，2020：2326-2335.

[14] ZHENG T L，CHEN Z N，FANG S C，et al.CDistNet：perceiving multi-domain character distance for robust text recognition[J].International Journal of Computer Vision，2024，132(2)：300-318.

[15] WOO S，PARK J，LEE J Y，et al.CBAM：convolutional block attention module[C]//Computer Vision-ECCV 2018.Cham：Springer，2018：3-19.

[16] SHAW P，USZKOREIT J，VASWANI A.Self-attention with relative position representations[EB/OL].[2024-10-13].https：//arxiv.org/abs/1803.02155

[17] 杨蓓，梁鑫，尹航，等.基于自注意力机制的大规模 MIMO 信道状态信息特征向量反馈方法[J].电信科学，2023，39(11)：128-136.
YANG B，LIANG X，YIN H，et al.Self-attention mechanism-based CSI eigenvector feedback for massive MIMO[J].Telecommunications Science，2023，39(11)：128-136.

[18] CHENG Z Z，XU Y L，BAI F，et al.AON：towards arbitrarily-oriented text recognition[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.Salt Lake City：IEEE，2018：5571-5579.

[19] XIE Z C，HUANG Y X，ZHU Y Z，et al.Aggregation cross-entropy for sequence recognition[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR).[S.l.]：IEEE，2020：6531-6540.

[20] BAEK J，MATSUI Y，AIZAWA K.What if we only use real datasets for scene text recognition toward scene text recognition with fewer labels[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR).[S.l.]：IEEE，2021：3112-3121.

[21] ZOU L，HE Z H，WANG K，et al.Text recognition model based on multi-scale fusion CRNN[J].Sensors，2023，23(16)：7034.

Residual optimization and content-adaptive text recognition algorithm

Wang Min[1,2]，Li Yongshun[1]，Ou Xiang[2]，Cao Ran[2]，Wu Jia[1]

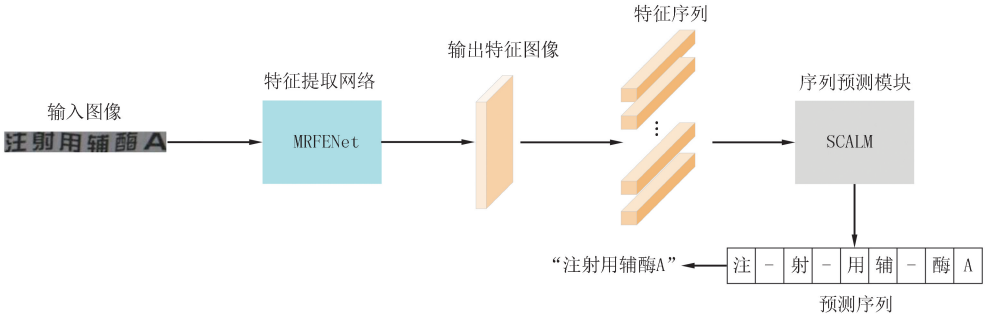
(1. School of Electronic and Information Engineering，Nanjing University of Information Science & Technology，Nanjing 210044，China；2. School of Electronic and Information Engineering，Anhui Jianzhu University，Hefei 230601，China)

Abstract：The labeling information of penicillin bottles plays a crucial role in ensuring patient medication safety and efficient drug management. A text recognition algorithm based on residual optimization and content adaptation is proposed to address the problem of poor recognition performance of traditional text recognition networks for long and fuzzy text in medicine bottle label images. On the basis of traditional text recognition networks，by optimizing the down sampling process of ResNet network and introducing the multi-scale feature fusion module，the ability of feature extraction is enhanced. At the same time，the addition of CBAM attention mechanism has improved the network's attention to the text，a multi-scale residual feature extraction module is adopted to replace the original feature extraction convolutional network，enhancing the network's ability to recognize low resolution text. Secondly，in the sequence modeling stage，the fusion of multi-layer bidirectional content adaptive recursive units and self attention mechanisms enhances the modeling capability of long text sequences. The experimental results show that compared with the CRNN text recognition network，this algorithm has improved the recognition accuracy by 3.92%，which is a certain improvement compared to other text recognition networks.

Keywords：text recognition；CRNN；feature extraction；CARU；self-attention mechanim

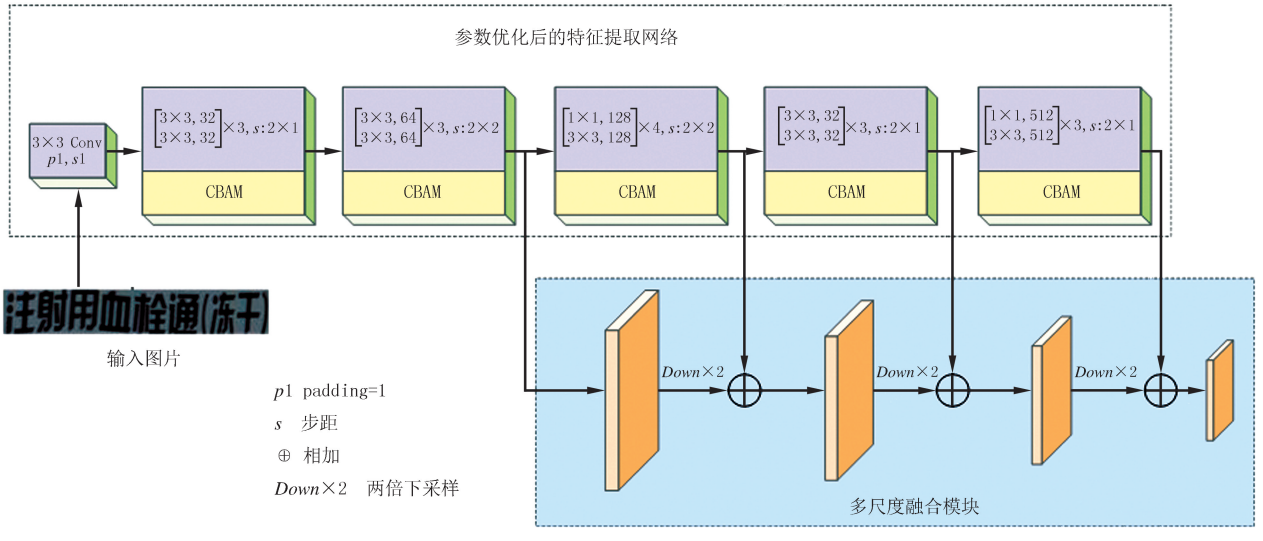
[责任编校　陈留院　杨浦]

附　录



图S1　基于残差优化和内容自适应的文本识别网络结构

Fig. S1　Structure of text recognition network based on residual optimization and content adaptation

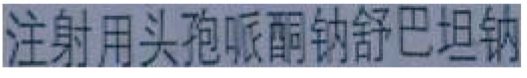


图S2　MRFENet网络结构

Fig. S2　MRFENet network structure

图S3　文本识别数据集标注示意图

Fig.S3　Labelimg of text recognition dataset



CRNN：注射用头孢派同舒巴日坦钠
本文：注射用头孢哌酮钠舒巴坦钠

CRNN：杨予江草业集团有限公司
本文：扬子江药业集团有限公司

CRNN：注射用头包西工钠
本文：注射用头孢西丁钠

图S4　文本识别效果对比

Fig.S4　Comparison of text recognition effect