

基于双态非对称网络的卷烟激光码识别^①



梁尚荣¹, 王慧琴¹, 马琦², 王可¹, 文钰栋²

¹(西安建筑科技大学 信息与控制工程学院, 西安 710055)

²(咸阳市烟草公司 礼泉分公司, 咸阳 713200)

通信作者: 王慧琴, E-mail: hqwang@xauat.edu.cn

摘要: 卷烟激光喷码识别是烟草稽查工作的重要手段. 本文提出一种基于双态非对称网络的烟码识别方法, 针对畸变烟码训练样本不足导致模型泛化能力弱的问题, 设计非线性局部增强方法 (nonlinear local augmentation, NLA), 通过在烟码图像边缘设置可控基准点进行空间变换, 生成有效畸变训练样本以增强模型泛化能力; 针对烟码与背景图案特征相似导致识别精度低的问题, 提出双态非对称网络 (dual-state asymmetric network, DSANet), 将 CRNN 的卷积层划分为训练模式和部署模式, 训练模式通过引入非对称卷积优化特征权重分布, 增强模型关键特征提取能力; 为保证实时性, 部署模式设计 BN 融合和分支融合方法, 通过计算融合权重并初始化卷积核, 将卷积层等效转换回原始网络结构, 降低用户端推理时间; 最后, 在循环层中引入自注意力机制, 通过动态调整序列特征权重, 进一步加强模型对烟码特征的提取能力. 通过对比实验, 该方法具有更高的识别精度和速度, 其识别精度达到 87.34%.

关键词: 卷烟激光码; 数据增强; 文本识别; 非对称卷积; 注意力机制

引用格式: 梁尚荣, 王慧琴, 马琦, 王可, 文钰栋. 基于双态非对称网络的卷烟激光码识别. 计算机系统应用. <http://www.c-s-a.org.cn/1003-3254/9729.html>

Cigarette Laser Code Recognition Based on Dual-state Asymmetric Network

LIANG Shang-Rong¹, WANG Hui-Qin¹, MA Qi², WANG Ke¹, WEN Yu-Dong²

¹(College of Information and Control Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China)

²(Liquan Branch, Xianyang Tobacco Co. Ltd., Xianyang 713200, China)

Abstract: Cigarette laser code recognition is an important tool for tobacco inspection. This study proposes a method for recognizing cigarette codes based on a dual-state asymmetric network. Insufficient training on samples of distorted cigarette codes leads to the weak generalization ability of the model. To address this issue, a nonlinear local augmentation (NLA) method is designed, which generates effective training samples with distortion to enhance the generalization ability of the model through spatial transformation using controllable datums at the edges of cigarette codes. To address the problem of low recognition accuracy due to the similarity between cigarette codes and their background patterns, a dual-state asymmetric network (DSANet) is proposed, which divides the convolutional layers of the CRNN into training and deployment modes. The training mode enhances the key feature extraction capability of the model by introducing asymmetric convolution for optimizing feature weight distribution. For real-time performance, the deployment mode designs BN fusion and branch fusion methods. By calculating fusion weights and initializing convolutional kernels, convolutional layers are equivalently converted back to their original structures, which reduces user-side inference time. Finally, a self-attention mechanism is introduced into the loop layer to enhance the extraction capability of the model for cigarette code features by dynamically adjusting the weights of sequence features. Comparative experiments show that this method has higher recognition accuracy and speed, with the recognition accuracy reaching 87.34%.

^① 基金项目: 陕西省烟草公司咸阳公司科技项目 (2022610425240008)

收稿时间: 2024-06-03; 修改时间: 2024-06-28; 采用时间: 2024-07-18; csa 在线出版时间: 2024-11-15

Key words: cigarette laser code; data augmentation; text recognition; asymmetric convolution; attention mechanism

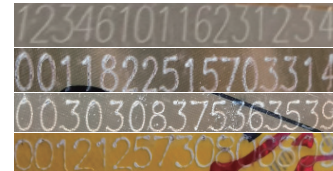
中国烟草行业在国民经济中扮演着重要的角色,凭借着巨大的市场需求,中国国家烟草公司也成为全球最大的香烟制造商.根据国家统计局的数据,中国烟草行业的年销售额超过千亿元人民币,占国内烟草产品销售总额的70%以上,此外,烟草行业还直接或间接带动了大量就业岗位.总体来看,中国烟草行业在国民经济中具有不可替代的地位和作用,对国家具有重要意义.

面对烟草行业的庞大经济效益与复杂产业链,国家层面必须对各个环节实施严密管控,杜绝走私、仿冒和非法跨境交易等违法行为,烟码因此产生.烟码特指每条卷烟上激光喷印的两组各含16位数字字母组合的编码,在条烟出厂之前,由烟草企业采用专业设备喷印在烟盒上,用作识别其身份的独特印记,例如,烟码“2102516306073605 XYXC610425101770”表示的信息为:该条烟于2022年10月25日出库,由咸阳烟草(XYXC)专卖局配发给零售许可证代码为“610425101770”的零售网点销售,是尾号“163060736”的烟箱中第5条烟.该工作目前以人工的方式进行,工作人员需要肉眼识别并手动抄录32位激光码,再转录系统,整个过程费时费力且容易出错.因此,烟码识别流程的智能化成为迫切的需求,利用计算机文本识别模型,能够提高烟码识别效率,为稽查人员有效地做好烟码记录和管理工作提供便利,是一个具有重要意义的研究方向.

如图1所示,采集的烟码图像主要有以下特点:(1)烟码畸变.受激光喷印工艺局限或拍摄角度等因素影响,烟码数据中存在字符形状扭曲、倾斜、透视等现象,如图1(b)所示.由于数据集中缺少畸变烟码样本,模型得不到充分学习从而对这类数据泛化能力弱;(2)背景复杂.如图1(c)所示,不同品牌的卷烟产品,其包装常以鲜艳多元的色彩搭配繁复的图案装饰,同时打码的位置非标准化分布,背景图案与烟码字符特征相似导致烟码特征信息表达不突出,使得模型在特征提取中很难将背景与烟码字符进行有效区分.

目前自然场景文本识别方法在深度学习方面的研究主要分为两种:基于分割和无需分割的方法.基于分割的方法需要先将图像按字符分割再进行单字符识别,

该类方法依赖于分割精度,计算量大、具有较大的精度损失,并且会忽略上下文信息.



(a) 规则清晰烟码图像



(b) 畸变烟码图像

(c) 复杂背景烟码图像

图1 不同类型烟码数据图

无需分割的方法可以实现端到端的识别,烟码识别对实时性要求高,而现有的识别模型无法同时兼顾速度和精度,对复杂场景鲁棒的网络必定伴随着复杂的结构设计和大量的参数,这使得现有模型在烟码识别中面临两方面的核心挑战:一方面,尽管高精度算法在识别准确性上表现出色,但其速度往往难以满足实时应用所要求的高效响应,导致实用性受限;另一方面,对于复杂场景中出现的文本,虽然轻量化算法在资源效率上具备优势,但在识别准确率上却有损失.考虑到实时性的需求,本文选择基于序列的方法作为烟码识别的基准模型,但该类方法结构简单,在识别烟码特定数据时存在缺乏畸变训练样本使模型泛化能力弱、背景图案与烟码字符特征相似导致识别准确率低等问题,需要针对以上问题对网络进行进一步改进和优化.针对以上问题,本文主要贡献如下.

(1) 提出了一种非线性局部增强方法 NLA, 该方法专门针对序列文本图像实现字符级多样性增强,创造出对模型训练更有效的难度样本,增强模型对畸变烟码的识别能力.

(2) 提出双态非对称网络 DSANet. 该网络改进 CRNN 卷积层,通过引入非对称卷积增强特征提取能力,并通过设计融合策略使用户端在不增加额外推理时间的情况下提升识别精度.

(3) 改进 CRNN 循环层, 引入自注意力机制, 通过动态调整序列特征权重, 进一步加强模型对烟码特征的提取能力。

1 相关工作

1.1 传统文本识别算法

传统的文本识别方法主要采用单字符识别的方式。在烟码领域, 王杰^[1]和杜梦圆^[2]研究了基于支持向量机 (support vector machine, SVM) 的烟码分类模型, 实现了烟码的单字符识别。倪晶^[3]通过模版匹配和 HOG 特征实现了单个烟码字符的识别。吕妃^[4]通过模版匹配和 LBP 实现了烟码字符识别。胡承东等^[5]将网格与 Zernike 矩特征 BP 神经网络相结合的方法应用于烟码识别, 解决了笔画粗细变化及字符旋转后无法识别的问题。

1.2 基于深度学习的文本识别算法

基于深度学习的文本识别方法可以划分为基于分割和无需分割两大类。基于分割的方法在烟码识别领域有诸多的应用, 彭召等^[6]通过设计分类器网络实现了单个烟码字符的识别。谢志峰等^[7]通过构建多任务学习框架, 将一对一识别转换为一对多识别, 降低了字符切分造成的精度损失。无需分割的方法主要分为 3 种: 基于序列特征的方法、基于注意力机制的方法和基于多模态特征的方法^[8]。

基于序列特征的经典方法有 Shi 等^[9]提出的一种结合 CNN 和 RNN, 并用 CTC^[10]解码器的 CRNN 网络, 该网络采用全卷积运算并支持端到端训练, 适用于不同长度的文本。蔡景轩^[11]将 CRNN 模型应用于烟码识别领域, 实现了无分割烟码识别。受启发于 Vision Transformer, Du 等^[12]提出一种基于单个视觉模型的自然场景文本识别网络 (single visual text recognition, SVTR), 不需要额外的序列处理步骤。

基于注意力机制的方法通常是编码器-解码器^[13]的结构。最初由 Shi 等^[14]引入空间变换网络 (spatial Transformer network, STN)^[15]至文本识别领域, 旨在矫正图像。后续研究中, 同一作者团队构建了 ASTER 模型^[16], 其革新之处在于以双向解码器结构取代了单向设计, 使得基于注意力机制的预测单元能够同步捕捉来自双端的信息流。Yang 等^[17]则设计了对称约束矫正网络 (stacked cross refinement network, ScRN), 该网络利用文本的中轴线特征精确刻画文本曲度, 提升了模型对高度弯曲文本的适应能力。Luo 等^[18]提出的

MORAN 网络无需依赖特定几何约束条件, 即可精准估算图像中各局部区域的位移参数, 有效应对不规则文本形态。陈佐瓚等^[19]设计了一种集成了注意力机制的双监督模型, 通过结合自适应卷积技术与文本焦点模块, 引导模型专注于文本的关键部分。Zhong 等^[20]提出了一种新的 SGBANet 网络, 该网络利用 GAN 技术, 通过模仿清晰图像特征来提升复杂图像中文本清晰度。

基于多模态特征的方法在提取视觉特征的基础上同时引入了语义信息。Fang 等^[21]提出 ABINet 网络, 通过联合优化视觉模块和语言模块, 实现视觉特征与语言特征的深度融合, 以提升识别结果精确度。Chu 等^[22]在 ABINet 的基础上, 通过叠加多个视觉组件形成迭代视觉建模单元, 通过迭代反馈来优化视觉特征的增强。Wang 等^[23]在 Transformer 模型的基础之上, 添加了一个自适应寻址与聚合模块, 通过多尺度特征整合来优化文本预测。Atienza^[24]基于 ViT^[25], 提出了 ViTSTR 网络。该网络只使用编码器架, 通过简化架构来提升效率和效能。Yan 等^[26]在 ViTSTR 基础上引入辅助的金字塔 Transformer 模块, 通过细化特征捕捉和降低计算负荷来提升模型效率。

2 非线性局部增强

由于数据集中缺少畸变烟码训练样本, 导致识别模型对此类变化特征学习不足, 因此需要更多畸变样本用于识别模型训练, 大规模收集数据成本太高且不能覆盖所有的可能性, 形态矫正会带来巨大计算量, 与之相比, 数据增强是一种成本较低的方式且独立于识别模型。目前对文本图像的增强还是以增加噪声干扰或进行几何变换为主, 该类方法将图像作为整体进行增强, 对文本多样性没有改变。针对以上问题, 本文提出了一种非线性局部增强方法 NLA, 通过局部非线性变换改变每个字符形态, 生成具有多样性的有效畸变样本。

该方法首先将烟码图像平均划分为 N 个部分, 在图像的顶部及底部边界初始化 $2(N+1)$ 个基准点 p , 然后, 设定每个基准点的移动区域为以 p 为圆心, 半径为 R 的圆形区域, 通过在移动区域内对基准点进行基于移动最小二乘法的相似性变换, 实现将基准点移动到 q 来增强图像, 如图 2 所示。

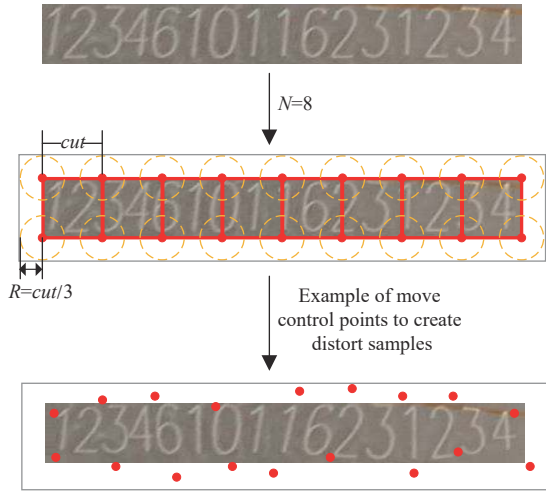


图2 NLA 增强方法

NLA 方法中每个基准点 p 的运动操作 $(\Delta x, \Delta y)$ 与两个因素有关: (1) 运动方向, 即 $(\Delta x, \Delta y)$ 的符号; (2) 运动的距离, 即 $(|\Delta x|, |\Delta y|)$. 通过实验证明, 移动距离太大时产生的样本扭曲过度, 反而降低了识别模型的鲁棒性. 因此, 将移动的距离限制在半径范围 R 内随机产生. 避免了过度增强导致的不确定性.

基于以上思想, 本文通过对输入图像进行基于移动最小二乘法 (moving least squares, MLS)^[27] 的相似性变换实现字符形变的效果. 对于给定图像中的一个点 u , 它的相似性变换是:

$$f(u) = (u - p_*)M + q_* \quad (1)$$

其中, $M \in R^{2 \times 2}$ 是线性变换矩阵, 满足 $M^T M = \lambda^2 I$, λ 为缩放系数, 并且 p_* 和 q_* 是初始基准点 p 和移动基准点 q 的加权质心:

$$p_* = \frac{\sum_{i=1}^{2(N+1)} w_i p_i}{\sum_{i=1}^{2(N+1)} w_i}, \quad q_* = \frac{\sum_{i=1}^{2(N+1)} w_i q_i}{\sum_{i=1}^{2(N+1)} w_i} \quad (2)$$

点 u 的权重 w_i 为:

$$w_i = \frac{1}{|p_i - u|^{2\alpha}}, \quad u \neq p_i \quad (3)$$

注意, 当 u 接近 p_i 时, 权重 w_i 会增加. 这表示 u 取决于相距其最近的基准点的运动. w_i 是有界的. 如果 $u = p_i$, 则 $f(u) = q_i$. 本文设置 $\alpha = 1$.

上文的最优相似性变换函数 $f(u)$ 可通过最小化 MLS 优化模型 E 得到:

$$E = \sum_{i=1}^{2(N+1)} w_i |f_u(p_i) - q_i|^2 \quad (4)$$

薄板样条变换 (TPS)^[28] 在形状矫正方面已经取得了成功, 但 TPS 的变换呈现不均匀的缩放和剪切, 不保留图像真实性, 这对烟码的增强是不可取的, 因为会破坏烟码的可读性. 此外, 通过比较刚性变换和相似性变换发现, 相似性变换是对烟码最理想的变换. 如图 3 所示, 刚性变换很真实地保留了相对形状, 但相似性变换更适合于烟码图像的增强, 因为它为每个字符提供了更灵活但真实的变换, 所以本文选择相似性变换策略.



图3 相似性变换和刚性变换结果对比

3 双态非对称网络

3.1 DSANet 概述

烟码图像中存在背景图案与烟码特征相似导致特征混淆的情况, CRNN 模型中的 3×3 卷积均匀提取空间特征不利于突出关键特征信息, 而非对称卷积和自注意力机制能够提取更丰富、多样化的特征表示, 有助于提升模型对物体形状和细节的感知能力, 从而提高模型对存在背景干扰图像的特征提取能力. 因此, 本文提出了一种双态非对称网络 DSANet, 通过引入非对称卷积和自注意力机制增强特征提取能力, 使用户端在不增加额外推理时间的情况下提升识别精度. 网络结构如图 4 所示, 由卷积层、循环层和转录层 3 部分组成, 其中卷积层划分为训练模型和部署模型.

训练模式下设计并行非对称卷积 (parallel asymmetric convolution block, PACB) 模块替换所有 3×3 卷积层, PACB 中的水平和垂直卷积核能够加强卷积核骨架位置权重, 以此降低背景信息干扰, 提高烟码特征提取能力, 同时, 非对称卷积形状灵活, 能更好地处理旋转图像, 提高模型泛化能力^[29]; 同时, 为了稳定和加速训练过程, 防止过拟合, 在 PACB 的每个分支中都加入批归一化层.

但 PACB 使速度大幅降低, 相较于基准模型, 训练速度和推理速度降低了 40.85% 和 16.38%, 因此, 部署模式设计 BN 融合和分支融合方法, 通过计算融合权重并初始化原始网络卷积核, 将 PACB 模块等效转换回 3×3 卷积结构, 使用户端在不增加额外推理时间的情况下提升识别精度; 其次, 在循环层引入自注意力机制, 进一步增强关键特征并抑制背景噪声.

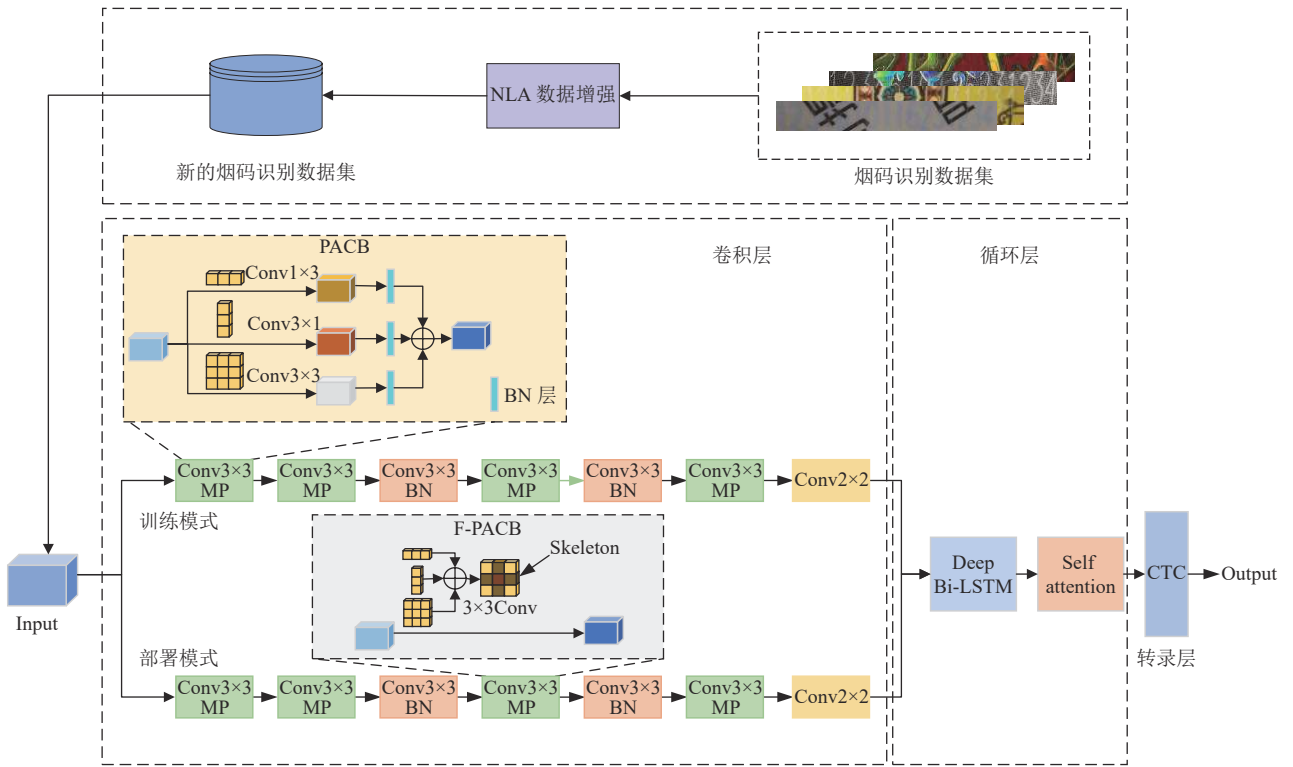


图4 DSANet 模型结构图

3.2 卷积层

3.2.1 设计并行非对称卷积模块

PACB 由 3 个非对称卷积分支构成, 如图 4 所示, 为减少过拟合, 加速训练过程, 在每一层最后都进行批归一化处理, 提高泛化能力. 假设输入图片 $M \in R^{W \times H \times C}$, 输出特征图为 $O \in R^{R \times T \times D}$, 则通过 3 个分支得到的 3 个中间特征 $O^{(i)}, i = \{1, 2, 3\}$ 的具体计算公式如下:

$$O^{(1)} = BN(Conv_{1 \times 3}(M)) \quad (5)$$

$$O^{(2)} = BN(Conv_{3 \times 1}(M)) \quad (6)$$

$$O^{(3)} = BN(Conv_{3 \times 3}(M)) \quad (7)$$

$O^{(1)}, O^{(2)}, O^{(3)}$ 相加得到融合特征图 O^{fusion} :

$$O^{fusion} = O^{(1)} \oplus O^{(2)} \oplus O^{(3)} \quad (8)$$

其中, \oplus 表示特征图在相应位置上的元素级相加.

3.2.2 分析卷积核可加

PACB 模块能够很好地提升模型性能, 但是也增加了近 1.4 倍的参数量, 影响模型的实时性, 本节试图将 PACB 中的分支结构合并为单分支标准卷积结构, 这样就可以等效转换回基准模型, 不会增加额外的推理时间了. 下面分析可行性.

研究得到卷积操作的一个特性: 对同一输入应用尺寸各异但相互兼容的 2D 卷积核, 若保持恒定步长并确保输出特征图尺寸一致, 各卷积结果累加所得的特征图等同于先将这些卷积核在相应位置相加形成一个等效卷积核, 再用它进行一次卷积操作所得的输出^[30]. 如式 (10) 所示. 也就是说, 对于 2D 卷积核而言, 即使核的尺寸不同, 可加性理论上依然适用:

$$I * K^{(1)} + I * K^{(2)} = I * (K^{(1)} \oplus K^{(2)}) \quad (9)$$

其中, I 为输入, $K^{(1)}$ 和 $K^{(2)}$ 是两个具有兼容性的 2D 卷积核, \oplus 表示核参数在相应位置上的元素级相加.

兼容性是指较小的核可以内嵌到较大的内核中, 例如, 1×3 和 3×1 卷积核与 3×3 的卷积核是兼容的, 可以将 1×3 和 3×1 卷积核添加到 3×3 的核上. 层 p 和层 q 的卷积满足以下条件时其核具有兼容性:

$$M^{(p)} = M^{(q)}, H_p \leq H_q, W_p \leq W_q, D_p = D_q \quad (10)$$

如图 5 所示, 该属性可以通过 2D 卷积的滑动窗口进行验证, 对于卷积核 $F^{(j)}$, 输出的 j 通道特征图上的某一点 y 的计算公式如式 (11) 所示:

$$y = \sum_{c=1}^C \sum_{h=1}^H \sum_{w=1}^W F_{h,w,c}^{(j)} X_{h,w,c} \quad (11)$$

其中, X 是输入图片 M 上相应的滑动窗口, 结合式 (9), 显然, 如果两个不同的卷积核产生的同尺寸特征图上对应点共享相同的滑动窗口 X , 则可加性成立.

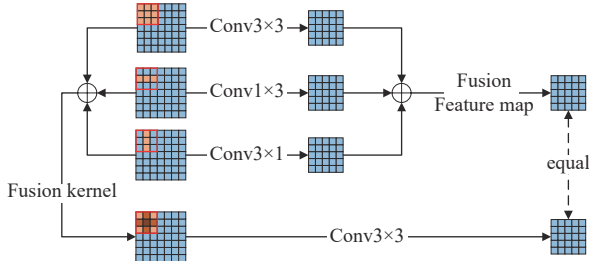


图 5 滑动窗口卷积

3.2.3 设计 BN 融合和分支融合方法

第 3.2.2 节分析得出了卷积具有可加性, 因此, 当模型训练完成后, 部署模式下通过融合方法将 PACB 模块等效转换回标准 3×3 卷积, 如图 4 中的 F-PACB 模块, 转换后产生的输出结果相同. 这种转换通过 BN 融合和分支融合两步实现, 如图 6 所示, 下面从某个通道 j 上进行的 2D 卷积分析融合过程.

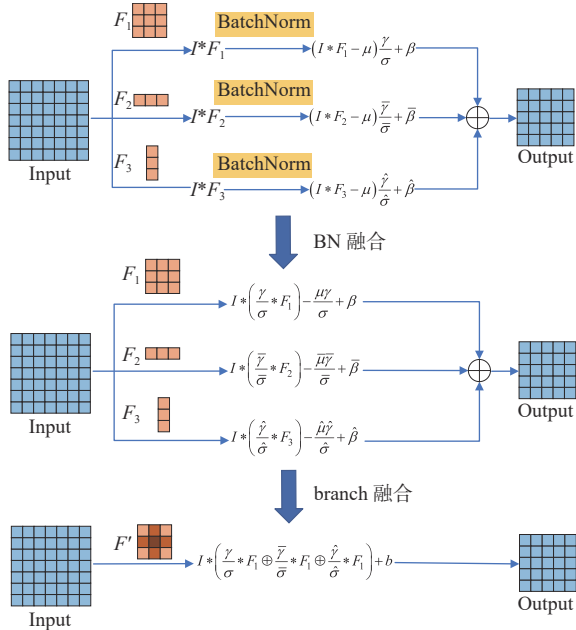


图 6 PACB 融合过程

对于特征图 O 的 j 通道输出计算方式为:

$$O_{:::j} = \sum_{k=1}^C M_{:::k} * F^{(j)}_{:::k} \quad (12)$$

其中, $*$ 表示卷积算子, $M_{:::k}$ 表示尺寸 $W \times H$ 图像的第 k 个通道, $F \in R^{W \times H \times C}$ 表示三维卷积核.

每层最后进行的批归一化处理表示为:

$$O_{:::j} = \left(\sum_{k=1}^C M_{:::k} * F^{(j)}_{:::k} - \mu_j \right) \frac{\gamma_j}{\sigma_j} + \beta_j \quad (13)$$

其中, μ_j 和 σ_j 为批归一化中的通道均值和标准差值, γ_j 和 β_j 分别为学习到的比例因子和偏差项.

BN 融合: 由于卷积层的一致性特性, 批归一化能够等效地整合进卷积层, 如式 (13) 所示, 对于每个分支, 可以构造一个新的卷积核 $\frac{\gamma_j}{\sigma_j} F^{(j)}$, 并添加一个偏置项 $-\frac{\mu_j \gamma_j}{\sigma_j} + \beta_j$, 从而实现相同的输出效果.

分支融合: 通过将非对称核添加到标准卷积核的相应位置, 成功地将 3 个经过 BN 融合的分支合并为一个标准卷积层, 这一过程是通过融合后的权值进行初始化实现的. $F'^{(j)}$ 表示融合的卷积核, b_j 表示得到的偏差, $F'^{(j)}$ 的计算公式如式 (14) 和式 (15) 所示:

$$F'^{(j)} = \frac{\gamma_j}{\sigma_j} F^{(j)} \oplus \frac{\tilde{\gamma}_j}{\tilde{\sigma}_j} \tilde{F}^{(j)} \oplus \frac{\hat{\gamma}_j}{\hat{\sigma}_j} \hat{F}^{(j)} \quad (14)$$

$$b_j = -\frac{\mu_j \gamma_j}{\sigma_j} - \frac{\tilde{\mu}_j \tilde{\gamma}_j}{\tilde{\sigma}_j} - \frac{\hat{\mu}_j \hat{\gamma}_j}{\hat{\sigma}_j} + \beta_j + \tilde{\beta}_j + \hat{\beta}_j \quad (15)$$

其中, $\tilde{F}^{(j)}$ 和 $\hat{F}^{(j)}$ 分别为对应的 1×3 和 3×1 非对称卷积核, \oplus 是核参数在相应位置上的元素相加. 因此, 可以得到 j 通道输出如式 (16) 所示:

$$O_{:::j} + \tilde{O}_{:::j} + \hat{O}_{:::j} = \sum_{k=1}^C M_{:::k} * F'^{(j)}_{:::k} + b_j \quad (16)$$

其中, $O_{:::j}$, $\tilde{O}_{:::j}$ 和 $\hat{O}_{:::j}$ 分别表示原始 3×3 , 1×3 和 3×1 分支的输出.

在 3D 卷积核中, 对每个通道执行 2D 卷积核融合的操作, 可以将 PACB 等效的转换为一个标准层. 然而, 这种转换仅对部署模型下有效, 因为训练过程中的动态是不同的.

3.3 循环层

3.3.1 深层双向长短时记忆网络

卷积层从输入烟码图像中提取空间特征, 而循环层可以捕捉序列中随时间变化的时序特征, 可以筛选并保留长序列中关键信息, 同时摒弃无关紧要的部分. 该网络由 Bi-LSTM 构成, 具备同时参考序列当前时刻之前与之后信息的能力, 通过融合双向隐藏层以生成最终输出, 该模型内部所有前后向神经元权重均实现共享, 避免增加额外模型复杂度. 如图 7 所示, 以下介绍该网络结构.

图 7 中 $X = \{x_1, x_2, x_3, \dots, x_t\}$ 为输入特征序列, h_t 表示 LSTM 单元, 对于单向 LSTM 而言, 其遗忘门、输

入门、输出门、循环状态以及单元输出的计算公式如下:

$$f_t = \sigma(W_f [\bar{y}_{t-1}, x_t] + b_f) \quad (17)$$

$$i_t = \sigma(W_i [\bar{y}_{t-1}, x_t] + b_i) \quad (18)$$

$$o_t = \sigma(W_o [\bar{y}_{t-1}, x_t] + b_o) \quad (19)$$

$$\tilde{C}_t = \tanh(W_C [\bar{y}_{t-1}, x_t] + b_C) \quad (20)$$

$$\bar{y}_t = o_t * \tanh(f_t * C_{t-1} + i_t * \tilde{C}_t) \quad (21)$$

其中, x_t 和 \bar{y}_t 分别表示 t 时刻 LSTM 单元的输入和输出特征, W_f, W_i, W_o, W_C 和 b_f, b_i, b_o, b_C 分别表示遗忘门、输入门、输出门以及循环状态中权重和偏置. 两个相同结构不同方向的 LSTM 构成 Bi-LSTM, 所以 Bi-LSTM 的预测结果计算公式如式 (22) 所示:

$$\tilde{y}_t = \sigma(\bar{w}\bar{y}_t + \hat{w}\hat{y}_t) \quad (22)$$

其中, \bar{y}_t 和 \hat{y}_t 表示 t 时刻正向和反向 LSTM 单元的输出结果, \bar{w} 和 \hat{w} 表示可训练权重, \tilde{y}_t 表示 t 时刻 Bi-LSTM 单元的最终输出结果, $\sigma(\cdot)$ 表示激活函数.

Deep Bi-LSTM 由两层 Bi-LSTM 叠加而成, 如图 7 所示, 下层 Bi-LSTM 的输出 $\tilde{Y} = \{\tilde{y}_1, \tilde{y}_2, \tilde{y}_3, \dots, \tilde{y}_t\}$ 作为上层 Bi-LSTM 的输入, 两层计算过程相同, 最终输出特征序列 $Y = \{y_1, y_2, y_3, \dots, y_t\}$.

3.3.2 引入自注意力机制

如图 7 所示, 自注意力机制 (self attention mechanism)^[31]接收前一层网络生成的特征图 $Y \in R^{T \times C}$, $T = H \times W$ 作为输入, C 为通道数, 通过分析查询矩阵 Q 、键矩阵 K 和值矩阵 V 之间的相互作用, 重点关注序列特征的整体相关性. Q, K 和 V 分别由输入张量 Y 和权重矩阵的点积运算得出, 如式 (23) 所示:

$$Q = YW^q, K = YW^k, V = YW^v \quad (23)$$

其中, $W^q \in R^{C \times D}$, $W^k \in R^{C \times D}$, $W^v \in R^{C \times D}$ 分别表示 Q, K 和 V 的权重矩阵.

查询矩阵 Q 和键矩阵 K 转置的点积运算得到注意力权重, 经 $Softmax(\cdot)$ 函数得到含有注意力信息的特征矩阵 A . A 与值矩阵 V 的点积通过权重矩阵 W_o 恢复为 C 通道, 得到注意力输出 Z , 如式 (24) 和式 (25) 所示, 设定缩放因子 $\sqrt{d_k}$ 稳定模型梯度:

$$A = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) \quad (24)$$

$$Z = (AV)W_o, W_o \in R^{D \times C} \quad (25)$$

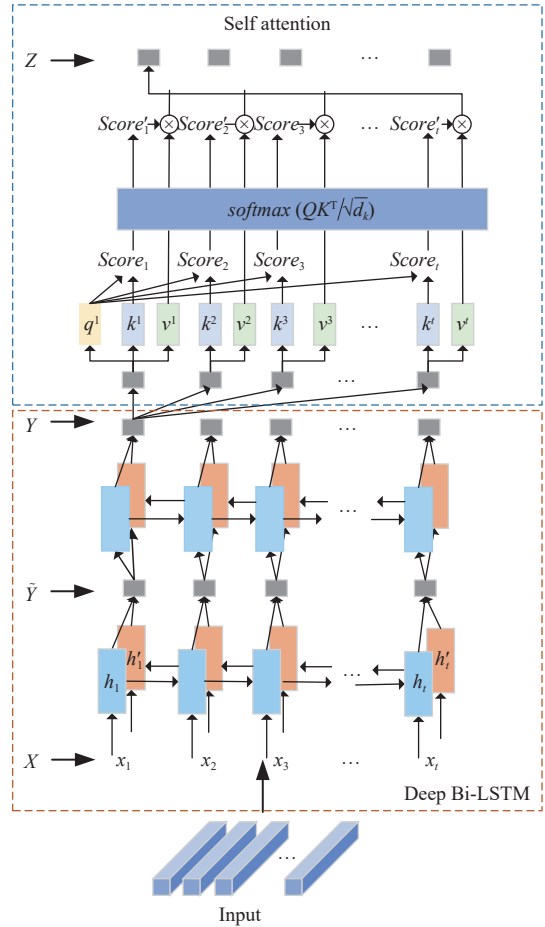


图 7 序列处理模型结构图

3.4 转录层

转录层采用 CTC 解码器, 这种技术通过引入空符号, 有效地解决了高层网络获得的预测标签序列 $Y = \{y_1, y_2, \dots, y_t\}$ 与实际图像标签不对齐的问题. CTC 解码器在进行无词典转录时, 通常依赖于最优路径策略, 这种策略的核心是寻找一条最大概率的字符串路径, 将这个路径的对应序列作为最优序列输出, 其计算公式如下:

$$p(a|Y) = \prod_{t=1}^T p(a_t|Y_t) = \prod_{t=1}^T Y_{a_t}^t \quad (26)$$

$$L^* = \beta(\arg \max_a p(a|Y)) \quad (27)$$

其中, $p(a|Y)$ 为任意一个可能输出路径 a 的概率, $Y_{a_t}^t$ 是 t 时刻输出标签为 a_t 的概率. 当找到最大概率的标签序

列 a 时,通过映射 β 移除重复字符和空白字符,最终得到烟码识别标签 L .

4 实验

4.1 数据集

由于缺乏公开的烟码数据集,本文采用的数据集是在烟草专卖局拍摄得到,卷烟品牌包括兰州、中华、苏烟、云烟、黄鹤楼等.利用 Labelme 对烟码位置区域及烟码字符进行人工标注,均以整行标注的形式给出.经筛选过滤,获得如下数据集:

CCD (cigarette code detection): 检测数据集,包含 6 913 张原始烟码图像,其中训练集 5 531 张,验证集 691 张,测试集 691 张.烟码字符区域标注形式采用 4 个顶点表示的矩形框,进行单行标注.

CCR (cigarette code recognition): 识别数据集,包含 13 826 张单行烟码字符区域图像,其中训练集 11 062 张,验证集 1 382 张,测试集 1 382 张.由 CCD 标注区域裁剪获得,烟码字符采用整行标注的形式.

4.2 实验设备

实验环境如下: Windows 10 操作系统, AMD Ryzen 9 5900X 12-Core 处理器, 3.70 GHz 主频, 32 GB 内存, NVIDIA GeForce RTX 3090 显卡.实验采用 Python 作为编辑语言,所用深度学习框架为 PyTorch 1.7.1.

DSANet 训练模型输入尺寸为 384×32 ,采用 Adam 优化器进行端到端的训练,设置初始学习率为 0.001, batchsize 为 64,训练迭代次数为 100 次.

4.3 评价指标

本文实验采用单字符识别率 CRA (character recognition accuracy)、单词识别率 WRA (word recognition accuracy) 和推理速度 FPS 作为评价标准.其中单字符识别率是指模型对单个烟码字符的识别准确率,如式 (28) 所示:

$$CRA = \frac{C_r}{C} \quad (28)$$

其中, C 为测试烟码中的单字符总数, C_r 表示正确识别的烟码字符数.与之相似的 WRA 在本文中表整行识别正确的样本数量占总测试样本数量的比例.

4.4 消融实验

(1) NLA 变换方式

为了验证 NLA 方法中选择相似性变换的有效性,分别以仿射变换、刚性变换改变 NLA 的变换方法应

用于 CCR 数据集,设置 $N = 8$, $R = cut/3$,进行了 4 组消融实验,以各方法增强后的数据集训练的基准模型最终的性能衡量各方法的有效性.

表 1 的消融实验结果表明,使用仿射变换、刚性变换和相似性变换增强的识别模型 WRA 值比 DSANet 分别高出 4.5、4.6 和 9.94 个百分点,仿射变换与相似性变换类似,但不具有保角性,会产生过度变形的样本.刚性变换仅限于几何变换,虽然对于一般的对象能更好地保留真实性,但并未增加烟码字符的多样性,对于模型性能的提升程度有限.因此,相似性变换更适合用于增强烟码字符的多样性.

表 1 NLA 变换方式消融实验 (%)

Method	CRA	WRA
DSANet	95.76	77.40
DSANet+Affine	96.60	81.90
DSANet+Rigid	96.63	82.00
DSANet+NLA (Ours)	98.50	87.34

(2) 主要参数 N 和 R 的取值

NLA 方法中涉及两个重要的参数:分割部分数 N 和移动半径 R ,它们的取值对 NLA 的增强效果起着至关重要的作用,文中取值为 $N = 8$, $R = cut/3$.本节实验分别对两个关键参数的取值进行研究,假设烟码图像的长度为 img_W ,宽度为 img_H ,将其划分为 N 份则每一份的长度为 $cut = img_W/N$, $R \in (0, cut]$.

表 2 实验设定 $N = 8$ 研究 R 的取值,实验结果表明,对于 384×32 的图像最优移动半径取值为 $R = cut/3$, R 过小变换幅度小,不足以生成有效样本,若 $R > img_H/2$,则说明上下边框对应的一对控制点的变换区域在 y 轴方向有重叠部分,这样就会导致过度形变,产生无效训练样本.

表 3 实验设定 $R = cut/3$ 研究 N 的取值,由实验结果可以发现 N 最佳取值为 8,因为烟码识别样本有极端宽高比的特点,有 16 个字符,划分为 8 个部分时,每两个字符共享一组控制点,两端字符独自占有一组控制点,这样产生的文本适当的增加了每个字符的形变,但不会造成过度形变.

NLA 方法生成的畸变烟码图像如图 8 所示,包括扭曲、透视、倾斜和拉伸 4 种.同时,为了测试 NLA 对识别模型性能的提升,选取畸变烟码图像进行了测试,部分结果如图 9 所示,其中, (a) 为 ground truth; (b) 为 DSANet; (c) 为 DSANet+NLA.可以看出,使用 NLA 增

强的 DSANet 在处理畸变烟码时相比于未增强模型有更好的效果.

表 2 分割部分数 N 消融实验 (%)

N	CRA	WRA
1	94.6	78.40
2	95.55	82.21
3	95.91	83.64
4	96.02	84.10
5	96.24	84.95
6	96.35	85.40
7	97.53	86.83
8	98.50	87.34
9	96.87	83.28
10	97.00	84.00

(3) 非对称核对旋转图片的有效性

实验中的 NLA-*表示在 NLA 增强后的数据集上训练所得的模型. 为验证非对称卷积核对旋转图片的鲁棒性, 以 WRA 作为评价指标对水平核、垂直核进行消融实验. 测试数据包含原始图、逆时针 90° 旋转、 180° 旋转和上下翻转 4 种. 表 4 显示, 移除任何一种非

对称卷积核都会降低模型性能, 但相比于仅使用 3×3 卷积的模型, 任意加入其一均可提升模型性能. 仅加入水平核的模型在 3 个旋转测试集上相较于卷积模型识别精度分别提高 1.06、1.27 和 1.19 个百分点, 仅加入垂直核的模型精度分别提高 0.61、0.52 和 0.8 个百分点, 可以看出, 水平核对 180° 和上下翻转图像有更高的精度.

表 3 移动半径 R 消融实验 (%)

R	CRA	WRA
$\frac{cut}{2}$	95.63	82.50
$\frac{cut}{3}$	98.50	87.34
$\frac{cut}{4}$	96.14	84.54
$\frac{cut}{5}$	95.31	81.24

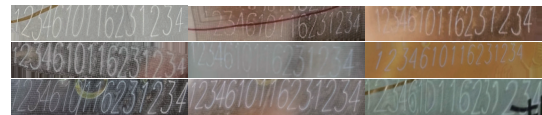


图 8 NLA 生成的增强样本

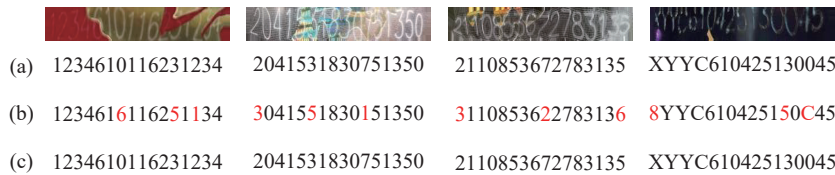


图 9 畸变烟码识别结果对比图

表 4 NLA-DSANet 方法下的旋转畸变消融实验

Horizontal kernel	Vertical kernel	Original input	Rotate 90°	Rotate 180°	Up-down flip
—	—	85.75	58.00	58.95	58.86
—	√	86.98	58.61	59.47	59.66
√	—	86.90	59.06	60.22	60.05
√	√	87.34	59.20	59.89	60.10

水平核和垂直核分别提取水平和垂直方向的特征, 图像旋转 180° 后, 不会改变这些特征的基本方向, 对它们而言, 仍然可以提取特定方向上不会因旋转显著改变的特征. 对于翻转图像, 如图 10 所示, 与 3×3 方形卷积核相比, 水平卷积核在翻转图像对称位置上产生相同的结果, 也就是说, PACB 的一部分仍然可以提取出正确的特征. 以上实验证明了 PACB, 特别是其中的水平核, 更为稳健, 可以提高模型泛化能力.

(4) 标准核骨架重要性

以下实验通过改变非对称卷积核的融合位置, 将

PACB 中的非对称核添加到 3×3 卷积核的骨架及边缘 (右、下边), 对 3×3 核骨架位置的重要性进行验证.

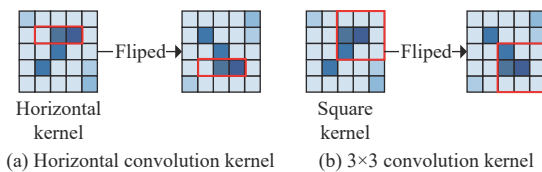


图 10 不同形状卷积核图像上提取特征的区域

表 5 的实验结果显示, 非对称卷积核添加到骨架位置时 CRA 和 WRA 值比添加至边缘处高 1.05 和 0.42 个百分点, 说明卷积核的骨架位置更重要. 图 11 为同一批次出库的复杂背景烟码图像识别结果对比图, 可以看出非对称卷积加强骨架位置对复杂背景干扰烟码图像的有效性.

图 11 中“*”表示未识别出, 可以看出非对称卷积的加入相比于未加入的模型效果更好, 而将非对称卷

积核添加至骨架位置相比于添加至边缘位置对背景干扰严重的图像识别效果更好。综合实验结果可得: (1) 在 3×3 卷积核中, 骨架相较于边缘位置更重要; (2) PACB 模块能强化骨架特征, 同时有助于抑制背景噪声影响; (3) 相较于增强骨架的 PACB 模块, 增强边缘位置会导致模型性能下降。

表 5 非对称卷积核添加位置消融实验 (%)

Model	CRA	WRA
NLA-DSANet (skeleton)	98.50	87.34
NLA-DSANet (border)	97.55	86.92

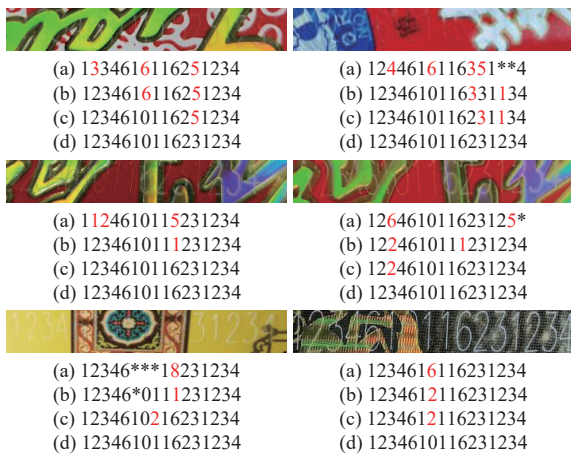


图 11 复杂背景烟码图像识别结果对比图

(5) DSANet 算法整体消融实验

为了评估 DSANet 中各改进模块的有效性, 进行消融实验, 训练数据如图 12 所示, 本文方法的整体损失稳定下降并低于其他模型, 精度稳定上升且始终优于其他模型, 证明该方法具有更好的收敛性能。

由表 6 可知, 训练模式较基准模型在 CRA 和 WRA 上分别提升 1.26 和 2.04 个百分点, 训练速度和 FPS 分别降低了 45.24% 和 19.58%, 仅 PACB 的引入就使训练速度和 FPS 降低 40.85% 和 16.38%, 说明 PACB 对速度影响很大; 部署模式的准确率与训练模型持平, 但 FPS 仅比基准模型下降 4.9%, 推理速度仅次于基准模型, 因为基准模型中还加入了注意力机制, 所以推理速度与 NLA-CRNN+SA 相同。以上结果说明, PACB 卷积核的融合可以消除其带来的额外计算量, 使模型等效恢复到原始的速度。因此本文提出的识别模型能够维持推理速度的同时提升识别准确率。

4.5 对比实验

为了进一步验证本文方法的有效性, 在同一烟码

识别数据集上使用 CRNN (baseline)、替换骨架网络的 CRNN-ResNet50^[32]、SVTR、基于注意力机制并带有矫正功能的 ASTER、MORAN、基于多模态的方法 ABINet、基于烟码识别研究的两个经典方法, 以及 NLA-DSANet 方法, 评价以上模型的识别性能并进行比较, 结果如表 7 所示, NLA-DSANet 模型的整行识别率 WRA 达到了 87.34%, 与上述模型相比, 分别提高了 13.94、12.04、6.69、8.81、11.24、4.34、9.86、18.74 个百分点; 推理速度仅次于 CRNN。实验结果表明, 该方法在烟码识别任务中获得了较高的识别精度, 且不损失过多推理时间, 能够达到推理时间和精度的平衡。

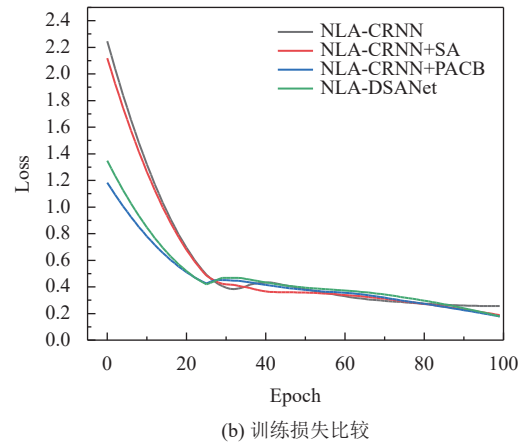
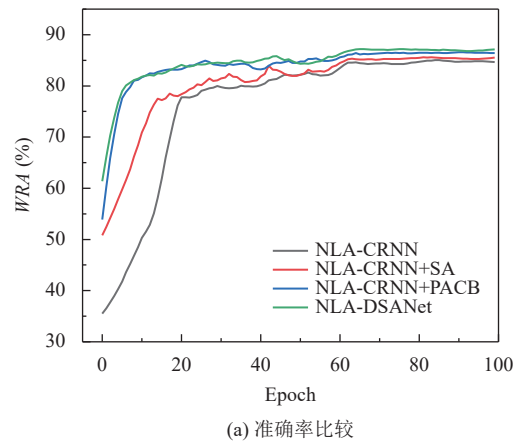


图 12 训练数据对比图

表 6 DSANet 整体消融实验

Method	CRA (%)	WRA (%)	Training speed (samples/s)	FPS (f/s)
NLA-CRNN	97.24	85.30	864.80	78.95
NLA-CRNN+SA	97.33	85.75	809.40	75.01
NLA-CRNN+PACB	97.51	86.76	511.50	66.02
NLA-CRNN+SA+PACB	98.50	87.34	473.60	63.49
NLA-CRNN+SA+F-PACB	98.50	87.34	—	<u>75.01</u>

表7 对比实验

Method	CRA (%)	WRA (%)	FPS (f/s)
CRNN	93.35	73.40	78.95
CRNN-ResNet50	93.82	75.30	74.00
SVTR	96.37	80.65	42.75
ASTER	95.97	78.53	66.01
MORAN	94.03	76.10	63.49
ABINet	96.50	83.00	51.36
姜思明等 ^[33]	94.37	77.48	69.6
谢志峰等 ^[7]	92.15	68.60	42.10
NLA-DSANet (Ours)	98.50	87.34	<u>75.01</u>

5 结论

针对畸变烟码样本不足使模型泛化能力弱、背景图案与烟码字符特征相似导致识别准确率低的问题,本章提出了一种基于双态非对称网络的烟码识别方法。首先,设计非线性局部增强模块,通过在烟码图像边缘设置可控基准点进行空间变换,生成有效的畸变烟码训练样本,增强模型泛化能力;其次,设计双态非对称网络,该网络在CRNN模型的基础上将卷积层划分为训练模式和部署模式,训练模式设计并行非对称卷积模块替换所有 3×3 卷积结构,通过非对称卷积优化权重分布,增强模型烟码特征提取能力,为保证实时性,部署模式设计BN融合和分支融合方法,通过计算融合权重并初始化原始网络卷积核,将PACB等效转换回 3×3 卷积结构,使用户端在不增加额外推理时间的情况下提升识别精度;最后,在循环层中引入自注意力机制,通过动态调整序列权重,增强模型对烟码区域的关注度。实验表明,与其他识别算法相比,本章算法识别精度达到87.34%,与CRNN相比提升了13.94个百分点,在烟码数据集上具有更高的识别精度。

参考文献

- 1 王杰. 复杂背景下激光防伪码图像识别研究 [硕士学位论文]. 武汉: 武汉理工大学, 2016.
- 2 杜梦圆. 复杂背景下的32位激光防伪码识别算法研究 [硕士学位论文]. 武汉: 武汉理工大学, 2014.
- 3 倪晶. 低质量激光防伪码识别及其系统实现 [硕士学位论文]. 武汉: 武汉理工大学, 2017.
- 4 吕妃. 复杂背景下的激光防伪码切分与识别算法设计 [硕士学位论文]. 武汉: 武汉理工大学, 2018.
- 5 胡承东, 速永仓, 曹玲芝. 烟包喷码字符识别系统研究. 机械工程与自动化, 2010(3): 117–119.
- 6 彭召, 丁晓, 刘哲, 等. 基于先验知识的香烟防伪数字串识

别. 计算机应用与软件, 2019, 36(12): 189–194, 244.

- 7 谢志峰, 吴佳萍, 章曙涵, 等. 基于深度神经网络的烟码智能识别方法. 计算机辅助设计与图形学学报, 2019, 31(1): 111–117.
- 8 曾凡智, 冯文婕, 周燕. 深度学习的自然场景文本识别方法综述. 计算机科学与探索, 2024, 18(5): 1160–1181.
- 9 Shi BG, Bai X, Yao C. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(11): 2298–2304. [doi: 10.1109/TPAMI.2016.2646371]
- 10 Graves A, Fernández S, Gomez F, *et al.* Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks. Proceedings of the 23rd International Conference on Machine Learning. Pittsburgh: ACM, 2006. 369–376.
- 11 蔡景轩. 基于深度学习的烟草码定位识别算法研究与实现 [硕士学位论文]. 杭州: 浙江大学, 2021.
- 12 Du YK, Chen ZN, Jia CY, *et al.* SVTR: Scene text recognition with a single visual model. Proceedings of the 31st International Joint Conference on Artificial Intelligence. Vienna: ACM, 2022. 884–890.
- 13 Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal: MIT Press, 2014. 3104–3112.
- 14 Shi BG, Wang XG, Lyu PY, *et al.* Robust scene text recognition with automatic rectification. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 4168–4176.
- 15 Jaderberg M, Simonyan K, Zisserman A, *et al.* Spatial Transformer networks. Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal: MIT Press, 2016. 2017–2025.
- 16 Shi BG, Yang MK, Wang XG, *et al.* ASTER: An attentional scene text recognizer with flexible rectification. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(9): 2035–2048. [doi: 10.1109/TPAMI.2018.2848939]
- 17 Yang MK, Guan YS, Liao MH, *et al.* Symmetry-constrained rectification network for scene text recognition. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019. 9147–9156.
- 18 Luo CJ, Jin LW, Sun ZH. MORAN: A multi-object rectified attention network for scene text recognition. Pattern Recognition, 2019, 90: 109–118. [doi: 10.1016/j.patcog.2019.01.020]

- 19 陈佐瓚, 徐兵, 丁小军, 等. 基于 Encoder-Decoder 框架的双监督机制自然场景文本识别. 计算机工程与应用, 2022, 58(6): 128–133.
- 20 Zhong DJ, Lyu SJ, Shivakumara P, *et al.* SGBANet: Semantic GAN and balanced attention network for arbitrarily oriented scene text recognition. Proceedings of the 17th European Conference on Computer Vision. Tel Aviv: Springer, 2022. 464–480.
- 21 Fang SC, Xie HT, Wang YX, *et al.* Read like humans: Autonomous, bidirectional and iterative language modeling for scene text recognition. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 7098–7107.
- 22 Chu XJ, Wang YT. IterVM: Iterative vision modeling module for scene text recognition. Proceedings of the 26th International Conference on Pattern Recognition (ICPR). Montreal: IEEE, 2022. 1393–1399.
- 23 Wang P, Da C, Yao C. Multi-granularity prediction for scene text recognition. Proceedings of the 17th European Conference on Computer Vision. Tel Aviv: Springer, 2022. 339–355.
- 24 Atienza R. Vision Transformer for fast and efficient scene text recognition. Proceedings of the 16th International Conference on Document Analysis and Recognition. Lausanne: Springer, 2021. 319–334.
- 25 Dosovitskiy A, Beyer L, Kolesnikov A, *et al.* An image is worth 16x16 words: Transformers for image recognition at scale. Proceedings of International Conference on Learning Representations. Washington: ICLR, 2021.
- 26 Yan XM, Fang ZH, Jin YC. Augmented Transformers with adaptive n-grams embedding for multilingual scene text recognition. arXiv:2302.14261, 2023.
- 27 Schaefer S, Mcphail T, Warren J. Image deformation using moving least squares. ACM Transactions on Graphics, 2006, 25(3): 533–540. [doi: 10.1145/1141911.1141920]
- 28 Bookstein FL. Principal warps: Thin-plate splines and the decomposition of deformations. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1989, 11(6): 567–585. [doi: 10.1109/34.24792]
- 29 连哲, 殷雁君, 米增, 等. 用于场景文本检测的非对称迭代细化预测网络. 计算机工程与应用. <http://kns.cnki.net/kcms/detail/11.2127.TP.20231229.1051.002.html>. (2024-01-02).
- 30 Ding XH, Guo YC, Ding GG, *et al.* ACNet: Strengthening the kernel skeletons for powerful CNN via asymmetric convolution blocks. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019. 1911–1920.
- 31 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010.
- 32 罗灿. 融合残差网络的改进 CRNN 模型的银行回执单识别算法研究 [硕士学位论文]. 重庆: 重庆大学, 2022.
- 33 姜思明, 谭升达, 陈冠达, 等. 基于 BiSeNet 和 YOLOv5s 的条烟 32 位激光码识别方法研究. 电子技术与软件工程, 2023(6): 163–168.

(校对责编: 孙君艳)