

文章编号: 2096-1618(2023)04-0422-07

# 基于 CRNN 改进的中文手写体文本行识别

舒珊珊, 郑晓旭, 文成玉

(成都信息工程大学通信工程学院, 四川 成都 610225)

**摘要:** 中文手写体文本行识别可以将纸质书写内容转换为可编辑的电子内容。对于手写体书写随意性大、中文字符种类多,且基于字符分割的方法识别准确率不高这些问题,提出基于卷积循环神经网络改进的端到端的中文手写体识别方法。首先将图片传入基于改进的 Inception 结构的特征提取网络,该网络首先改进 GoogLeNet 模型,然后在此基础上又改进添加卷积模块的注意力机制模块和 Inception 组合结构,改进后的模型能更好地提取图片的有效特征;之后将提取到的图片特征传入循环层,即两层双向长短时记忆网络进行预测;最后将预测序列传入转录层,经过连接时序分类进行转录输出。在 CASIA-HWDB2 数据集的实验结果表明,该方法能获得 95.12% 的识别准确率,证明方法的可行性。

**关键词:** 手写体识别; 卷积循环神经网络; 卷积模块的注意力机制模块; 双向长短时记忆网络; 连接时序分类

中图分类号: TP39

文献标志码: A

doi: 10.16836/j.cnki.jcuit.2023.04.008

## 0 引言

光学字符识别与手写体识别都是模式识别的研究重点,其主要应用在身份证识别、车牌识别、街景文字识别和自动化阅卷系统等方面,即将各种实体文本转换为更自由、方便操作和存储的电子文档。相较于光学字符识别,不同书写者的书写风格各不相同且书写时易存在字符粘连等现象,这些增加了手写体识别的难度。

在 手 写 体 识 别 中,单 字 符 手 写 体 识 别 技 术 已 经 比 较 成 熟,文 本 行 手 写 体 识 别 还 在 发 展。基 于 分 割 的 手 写 体 识 别 先 将 文 本 行 文 字 分 割 成 单 字 符 或 者 利 用 滑 动 窗 口,框 出 每 个 单 字 符,然 后 利 用 单 字 符 分 类 器 对 分 离 得 到 的 单 字 符 进 行 识 别,最 后 利 用 语 言 模 型 等 整 合 文 本 以 输 出<sup>[1]</sup>。为 避 免 文 本 分 割 造 成 的 分 割 不 当 和 分 割 错 误 累 积 等 问 题,无 需 分 割 的 端 到 端 手 写 体 文 本 行 识 别 模 型 被 提 出。文 献[2]提 出 将 MDLSTM-RNN+CTC 模 型 应 用 于 中 文 文 本 行 识 别,综 合 运 用 了 4 个 方 向 的 特 征。但 该 方 法 效 率 不 是 很 好,在 ICDAR2013 数 据 集 上 能 达 到 83.5% 的 识 别 准 确 率,如 结 合 语 言 模 型,则 该 网 络 能 达 到 89.4% 的 识 别 准 确 率。文 献[3]提 出 卷 积 循 环 神 经 网 络 (convolutional recurrent neural network, CRNN) 模 型,在 未 使 用 语 言 模 型 的 情 况 下,其 在 ICDAR2013 数 据 集 上 能 达 到 89.6% 的 识 别 准 确 率。为 进 一 步 提 升 中 文 文 本 行 识 别 准 确 率,本 文 提 出 基 于 CRNN 改 进 的 中 文 文 本 行 识 别 算 法,主 要 贡 献 有 以 下 2 点:

(1) 采 用 端 到 端 无 分 割 的 CRNN 模 型 框 架,改 进

CRNN 模 型 的 卷 积 层 (convolutional neural network, CNN),将 基 于 VGG 模 型 的 CNN 改 为 基 于 GoogLeNet 模 型 的 CNN。通 过 消 融 实 验 对 比,改 进 后 的 模 型 在 HWDB2 小 数 据 集 上 能 取 得 94.2% 的 识 别 准 确 率。

(2) 在 将 CRNN 的 CNN 替 换 为 GoogLeNet 模 型 的 基 础 上,再 改 进 添 加 CBAM 模 块 和 Inception 组 合 结 构。通 过 消 融 实 验 对 比,改 进 后 的 模 型 在 HWDB2 小 数 据 集 上 能 取 得 94.75% 的 识 别 准 确 率 且 在 HWDB2 全 部 数 据 集 上 能 取 得 95.12% 的 识 别 准 确 率。

## 1 相关工作

### 1.1 CRNN 模型简介

CRNN 模 型 在 文 本 识 别 中 应 用 广 泛,模 型 结 构 为: 卷 积 神 经 网 络 层 + 循 环 层 + 转 录 层。假 设 传 入 图 像  $I \in R^{H \times W \times 1}$ ,因 为 数 据 集 为 白 纸 黑 字 形 式,灰 度 图 足 以 表 征 图 片 信 息,这 里 的 输 入 为 灰 度 图。图 片 经 过 CNN 网 络 提 取 特 征,得 到 特 征  $f \in R^{1 \times 1 \times C}$ ,即 经 过 CNN 后 得 到 高 为 1,宽 为 1,维 度 为  $C$  的  $T$  个 特 征;然 后,将 这  $T$  个 特 征 传 入 循 环 层 网 络,这 里 是 两 层 双 向 长 短 时 记 忆 网 络 (bidirectional long short-term memory network, BiLSTM),特 征 图 经 过 BiLSTM 处 理 得 到 序 列 的 预 测 输 出  $Y = [y_1, y_2, \dots, y_T]$ ,  $T \in R^L$ ,其 中  $L$  为 文 本 字 典 中 字 符 的 总 个 数 (包 含 空 白 占 位 符);最 后,将  $Y$  传 入 转 录 层 进 行 转 录 输 出,得 到 除 去 冗 余 信 息 的 最 终 文 本 识 别 结 果。CRNN 模 型 结 构 见 图 1。

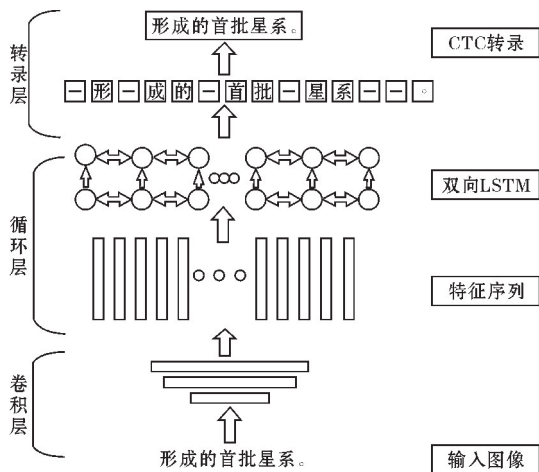


图1 CRNN 模型结构

## 1.2 卷积神经网络层

卷积层、池化层和全连接层等构成传统的卷积神经网络。卷积层利用卷积核对特征图进行局部信息特征提取,同一层的卷积核为一个固定的权值矩阵,通过该卷积核作用于特征图再经过激活函数即可得到提取到的特征输出。随着卷积层深度的增加,提取到的特征也由低级特征转变为更抽象的高级特征<sup>[4]</sup>。卷积公式如下:

$$Z=f(W^i \cdot X+b^i) \quad (1)$$

其中  $X$  为输入特征向量,  $W^i$  为第  $i$  层卷积层的权值矩阵,  $b^i$  为第  $i$  层卷积层的偏置量,  $f(\cdot)$  为激活函数。常用的激活函数有 ReLU, Tanh, Sigmoid 3 种。

池化层起到特征筛选的作用,其对提取到的特征进行降维处理,即在池化层池化核大小的特征矩阵中选取一个代表值。全连接层可以整合卷积层或者池化层中具有类别区分性的局部信息<sup>[5]</sup>。

CRNN 模型中的卷积神经网络采用的是去除全连接层的 VGG 网络,其利用卷积层和池化层提取筛选特征。然后将得到的特征传入循环层。本文对 CRNN 模型中的 CNN 部分进行了改进,其他部分保持不变。

## 1.3 循环层

循环神经网络 (recurrent neural network, RNN) 具有记忆功能,可用于处理序列信息。但对于长序列信息,其容易出现梯度消失和梯度爆炸现象,作为改进,提出了长短期记忆网络 (long short-term memory, LSTM) 和门控循环单元 (gated recurrent unit, GRU)。对长短期记忆网络而言,其特殊的“门”结构将短期记忆与长期记忆结合起来<sup>[6]</sup>,可以弥补 RNN 的不足。

$$f_{\text{update}} = \sigma(W_u \cdot (a^{t-1}, x^t) + b_u) \quad (2)$$

$$f_{\text{forget}} = \sigma(W_f \cdot (a^{t-1}, x^t) + b_f) \quad (3)$$

$$f_{\text{output}} = \sigma(W_o \cdot (a^{t-1}, x^t) + b_o) \quad (4)$$

式(2)、(3)、(4)分别为更新门、遗忘门和输出门表达式,式中  $\sigma$  为 sigmoid 激活函数,  $W_f$  为遗忘门权重矩阵,  $b_f$  为遗忘门偏置,  $a^{t-1}$  为上一个单元的隐藏层输

出,  $x^t$  为本单元的输入,其他类似。上一个单元的隐藏层输出和本单元的输入决定三个“门”的具体值。

$$C^t = \tanh(W_c \cdot (a^{t-1}, x^t) + b_c) \quad (5)$$

$$C^t = f_{\text{update}} \circ C^t + f_{\text{forget}} \circ C^{t-1} \quad (6)$$

$$a^t = f_{\text{output}} \circ C^t \quad (7)$$

式中“ $\circ$ ”为按元素相乘操作,上一个单元的隐藏层输出与本单元的输入在  $\tanh$  激活函数的作用下决定单元状态  $C$  的候选值。通过更新门与单元状态候选值作用,遗忘门与上一个单元状态值作用决定经过本单元后的单元状态值。输出门与单元状态值作用决定隐藏层的输出值。可以看出,更新门决定候选值是否被网络存储记忆,遗忘门决定上一时刻的单元状态被遗忘或被继续记忆。

在 CRNN 网络中,其循环层使用的是双向 LSTM 结构,它既从前到后进行预测又从后往前进行预测,照顾到了序列前后的信息,对文本信息预测具有很好的作用。

## 1.4 转录层

在 CRNN 模型中,其转录层使用的是连接时序分类 (connectionist temporal classification, CTC) 算法。CTC<sup>[7]</sup> 是一种用于序列建模的工具,可以解决不定长输出问题和输入序列与输出序列的对齐问题。CTC 包含两部分内容,CTC 损失计算和 CTC 解码输出。在 CRNN 模型中,计算网络损失用的是 CTCloss 这个指标,通过 CTCloss 来进行模型的反向传播,不断更新模型参数以获得更好的结果。CTCloss 计算公式如下:

$$p(\pi|x) = \prod_{t=1}^T y_{\pi_t}^t, \forall \pi \in (L')^T \quad (8)$$

其中  $L'$  为包含 CTC 特有的空白占位符和文本字典中字符的总的集合,  $\pi$  可以视为一条解码路径<sup>[8]</sup>,即一个序列有  $T$  个输出,每个输出的维度为  $L'$ 。式(8)是将时间步  $t$  时路径  $\pi$  的概率进行累加。

$$p(L|x) = \sum_{\pi \in \beta^{-1}(L)} p(\pi|x) \quad (9)$$

CTC 引入一个序列到序列的映射函数  $\beta$ ,式(9)表示按照解码规则,能得到标签文本输出的,所有路径的概率累加和。

CTC 解码分为贪心搜索解码和束搜索解码。文献[9]发现两种解码方式得到的结果相差不多,但是束搜索代码当束大小设置为 10 时解码时间是贪心搜索算法的 3 倍左右,本文采用的是贪心搜索算法解码。

## 2 中文手写体文本行识别算法

### 2.1 基于改进的 Inception 结构的特征提取网络

CRNN 模型的原卷积神经网络结构很简便,对于提取图像特征来说还有改进的空间,近年来不少学者尝试对其进行改进。2017 年,文献[10]提出由 CNN、RNN 和注意力机制组成的模型并通过实验推论出在文本识别中并不是深度越深越好,它提出:图像分类需

要非常复杂、深层次的图像特征,但对于文本识别而言这些特征反而不是很有益。该文献使用的3种CNN模型是 Inception-V2、Inception-V3 和 Inception-Resnet-V2。2018年,文献[11]提出一种越南身份证识别方法,该方法采用 Inception-V3 和基于注意力机制的 BiLSTM 模型进行文本识别。该方法在运用 Inception 结构的同时,在循环层运用了注意力机制。2021年,文献[12]提出基于 CRNN 的车牌识别方法,在文本识别模块中其引入 STN 和基于残差学习与注意力机制的 CRNN 模型。

通过学习与实验验证,本文选取 GoogLeNet 网络作为卷积神经网络的主干网络,并在其基础上进行改进,形成改进的基于 Inception 结构的卷积神经网络。

## 2.2 基于 GoogLeNet 模型的 CNN

Inception-V2、Inception-V3 等网络是在 GoogLeNet 网络的基础上发展出来的,本文选取 GoogLeNet 网络作为 CRNN 模型的特征提取网络。在原网络的基础上,去除了原模型中的辅助分类器和网络后面的全连接层等操作,同时去除了原模型中后面3个 Inception 结构。最后,经过特征提取网络后的通道数为528。

在应用 GoogLeNet 网络时,本文进行了3种尝试。第一种为保持原模型结构不变。第二种为仿照 CRNN 模型,CRNN 模型在后面两层池化层操作时,将池化层步距改为(2,1),这样可以保持图片高不变的情况下将图片的宽缩减为原来的1/4。本文将 Inception 结构后的两层池化层步距改成(1,2)和(2,1),这样经过两个池化层后才将图片的高和宽各缩减为原来的一半。第三种为将前两个卷积操作后的池化层步距改为(1,2)和(2,1),其他保持不变。

本文对这个过程进行了实验验证,最后证明第三种方法效果最好。最后确定的适用的 GoogLeNet 模型结构见图2,其在网络前面部分采用以上策略有效保留了图片的特征信息。

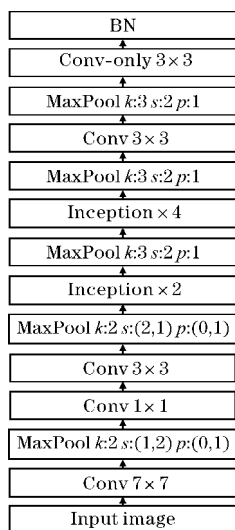


图2 基于 GoogLeNet 模型的 CNN

## 2.3 CBAM 模块

在网络进行特征提取时,为使网络更关注目标区域,注意力机制被提出并得到广泛使用。CBAM 结构<sup>[13]</sup>既包含通道注意力又包含空间注意力,同时其符合即插即用的条件,使用起来非常方便。本文引入 CBAM 模块来提升特征提取网络的表达能力,以使得网络更关注文本行中文字所在的区域。CBAM 结构如图3所示。

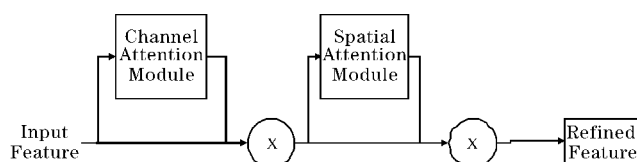


图3 CBAM 结构

CAM 模块公式:

$$M_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) \quad (10)$$

在 CAM 模块:输入特征分别经过一个平均池化和一个最大池化,都得到特征为:1×1×C 的特征图。两个特征图分别经过多层感知机 (multi-layer perceptron, MLP) 结构,最后将经过 MLP 的两部分特征相加,用 sigmoid 激活函数处理。其中 MLP 结构为两个卷积层,第一个卷积层进行降维处理,第二个卷积层再将通道维度恢复到初始维度。在该模块压缩了空间维度,通道维度保持不变,该模块关注图片中有意义的信息。

SAM 模块公式:

$$M_s(F) = \sigma(f^{7 \times 7}([\text{AvgPool}(F) + \text{MaxPool}(F)])) \quad (11)$$

SAM 模块:在通道维度求其最大池化和平均池化结果,分别得到 channel 为1的特征图然后将两个特征图进行拼接后,经过一个卷积核大小为7×7,步距为1,padding 为3的卷积层,得到的特征图 channel 维度为1后经过 sigmoid 激活函数处理。在该模块压缩了通道维度,空间维度保持不变,该模块关注目标的位置信息。

CBAM 模块:先将输入特征  $F_1$  与  $F_1$  经过 CAM 模块得到的结果  $\text{CAM}_{F_1}$  相乘得到  $F_2$ ,然后将  $F_2$  与  $F_2$  经过 SAM 模块得到的结果  $\text{SAM}_{F_2}$  相乘。即先经过通道注意力处理再通过空间注意力处理。即:

$$F_{\text{out}} = (F_2 \otimes (\text{SAM}_{F_2})), F_2 = (F_1 \times \text{CAM}_{F_1}) \quad (12)$$

式中 $\otimes$ 表示对应元素逐个相乘。

## 2.4 改进的 Inception 组合结构

通过研究 Inception-V2、Inception-V3 和 Inception-



Resnet-V2<sup>[14]</sup> 网络结构,发现其在图片大小不同时采用不同的 Inception 模块进行处理。参考其网络结构,本文对 GoogLeNet 网络模型进行如下改进,改进的 Inception 结构都包含卷积、批归一化和 Relu 函数激活 3 个操作。

图 4 的 Inception 结构融合了不同尺度的特征信息,同时 1×1 的卷积结构起到了降维、减少计算量的作用。

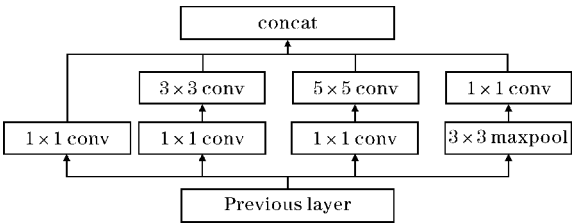


图 4 Inception 结构

图 5 的 Inception\_B 结构首先将 5×5 的卷积变为两个 3×3 卷积的叠加,这样操作保持感受野不变的条件起到减少计算量的作用。然后将 3×3 卷积替换为 1×3 和 3×1 两个非对称卷积的叠加,该操作同样起到减少运算量的作用。

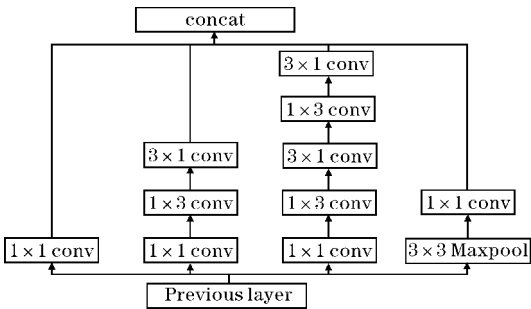


图 5 Inception\_B

图 6 的 Inception\_B\_c 结构与图 5 处理方法类似,不同之处在于它将全串联结构改为串并联结合的形式,进一步增加了网络的宽度。

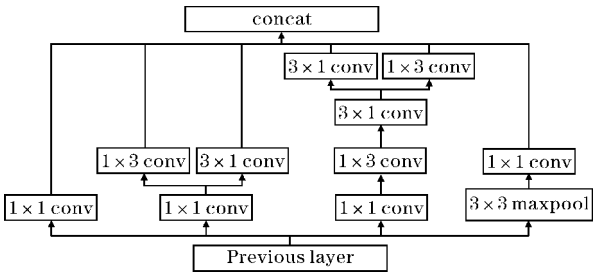


图 6 Inception\_B\_c

图 7 的 Inception\_Resnet\_B 结构在采用非对称卷积的同时,添加了残差结构。残差单元在深度学习是为了防止随着网络深度加深产生的梯度消失和梯度爆炸问题,同时其起到特征复用的作用。

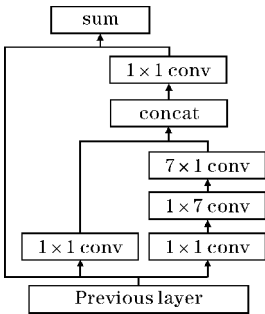


图 7 Inception\_Resnet\_B

因为非对称卷积最好用于图片大小在 12×12 ~ 20×20, 在网络后半段时图片的高处于这个区间,可以使用。将这些结构组合起来,形成了本文提出的新的 Inception 特征提取网络结构。网络结构如表 1 所示。

表 1 基于 Inception 结构的特征提取网络

Type	Configurations
BatchNormalization	—
Convolution_only	#maps:528, k:3×3, s:1, p:0
MaxPool	Window:3×3, s:2, p:1
Convolution	#maps:528, k:3×3, s:1, p:1
Inception_B	IN:512, OUT:528
Inception_B	IN:512, OUT:512
Inception_B_c	IN:512, OUT:512
Inception_B_c	IN:480, OUT:512
Inception_Resnet-B	IN:480, OUT:480
CBAM	—
MaxPool	—
Inception	IN:256, OUT:480
Inception	IN:192, OUT:256
MaxPool2	Window:2×2, s:2×1, p:0×1
Conv3	#maps:192, k:3×3, s:1, p:1
Conv2	#maps:128, k:1×1, s:1, p:1
MaxPool1	Window:2×2, s:1×2, p:1×0
Conv1	#maps:64, k:7×7, s:1, p:3
Input	W:Width, H:48, Gray Images

这里 Convolution 操作包含卷积层、批归一化层和激活函数。而 Convolution-only 只包含卷积操作。IN 表示网络输入特征维度,OUT 表示网络输出特征维度。 $k$  表示卷积核大小, $s$  表示步距, $p$  表示 0 填充大小。maps 表示卷积核个数>window 表示池化操作时的滑窗大小。

### 3 实验结果与分析

#### 3.1 数据集

实验采用的数据集为 CASIA-HWDB2.0-2.2 手写

体数据集。该数据集来自中科院自动化研究所。它包含 GB2312-80 标准一级中文字符 2703 类,共 52230 行文本图片。数据集解析后可以得到文本行图片及其对应的标签,标签和文本为相同的命名。

因为 CASIA-HWDB2 数据集图片数目较多,进行一些消融实验对比时,选取小数据集。小数据集的组成方式为:从 HWDB2.0-2.2Train 数据集里各取 10000 张图片作为小数据集的训练集,从 HWDB2.0-2.2Test 数据集里各取 1000 张图片作为小数据集的测试集,即共 3 万张训练集图片,3 千张测试集图片。

### 3.2 评价指标

字符串编辑距离 (Levenshtein Distance) 常被用来当字符级别手写体识别的评估方式,可计算识别结果与标签的相似程度。字符串编辑距离:对于两个字符串,让两个字符串变得一样所进行的插入、替换和删除操作的次数,每进行一次改动,距离+1。最终确定的识别准确率计算公式:

$$AR = \frac{N - \text{Ins} - \text{Sub} - \text{Del}}{N} \quad (13)$$

式中  $N$  为文本行标签所对应字符的数目,Ins、Sub、Del 分别为预测字符串转换为标签需要插入、替换和删除的单词个数。

同时采用字符错误率来进行辅助评估。计算公式如下:

$$\text{CER} = \frac{\text{Sub} + \text{Ins} + \text{Del}}{N} \quad (14)$$

### 3.3 实验环境设置

本文实验在 Windows 操作系统下进行,python 版本为 python3.8, GPU 型号为 NVIDIA RTX3060,显存为 6GB, CUDA 版本为 11.3,采用 Pytorch1.11.0 深度学习框架。算法均在 GPU 加速中运行。

选择无需设置学习率的 Adadelta 梯度下降算法,实验的 batchsize 大小均为 4。

### 3.4 数据处理方式

因为解析后得到的图片大小不一,而 CRNN 模型要求输入图片高度为 16 的整数倍,决定将所有图片统一为一样的大小。图片统一前,先将数据集中白边过多的图片进行适度裁剪。然后进行数据处理,首先将图片高度设置为一固定值对所有图片进行等比例缩放,然后计算出此状态下图片宽的最大值,将所有图片的宽加白边设置成宽的最大值大小,最终确定图片大小为 1420×48。本文实验的图片大小都为 1420×80。

### 3.5 模型训练与测试

在小数据集上对 CRNN 原模型进行实验验证,CRNN 原模型中经过特征提取网络后特征维度为 512,这里改为 528,以与经过 GoogLeNet 网络后的特征维度相同。图 8 即为结果展示,迭代 35 个 epoch 后,识别准确率达到 90% 左右。

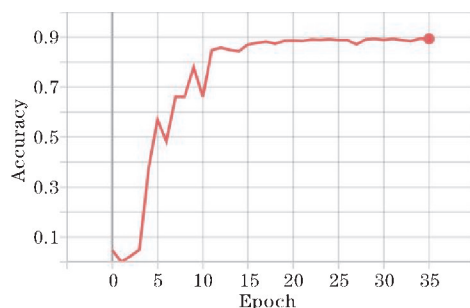


图8 CRNN 模型运行结果

在小数据集上,进一步验证提出的 3 种 GoogLeNet 模型改进方法,结果如图 9 所示。图中蓝色曲线、绿色曲线、粉色曲线分别为第一种改进、第二种改进和第三种改进的识别准确率结果展示。可以看出 3 种方法的识别准确率均高于基于 CRNN 模型的识别准确率。3 种方法中,第三种方法比前两种方法略好,识别准确率在 94.2% 左右。通过增加网络宽度能更多地获取更丰富的网络特征,从而提高识别准确率。

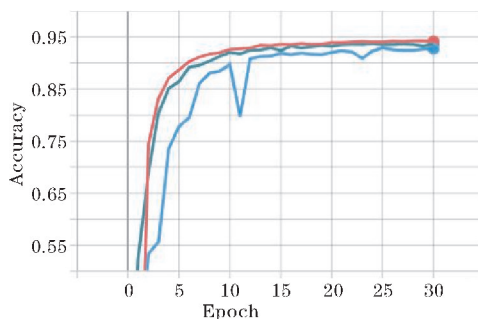


图9 改进的 GoogLeNet 模型运行结果

最后,在小数据集上验证基于改进的 Inception 结构的特征提取网络,识别准确率结果如图 10 所示。图中淡蓝色曲线准确率为改进的 GoogLeNet 模型的识别准确率,为 94.25% 左右,深蓝色曲线为在 GoogLeNet 模型中添加 CBAM 模块的模型,准确率为 94.5% 左右,橙色曲线为在 GoogLeNet 模型中添加 CBAM 模块和改进的 Inception 结构的模型,准确率为 94.75% 左右。可以看出,改进后的模型比原模型略好。通过添加注意力模块有助于卷积时更好地提取应该关注的文字区域特征。采用改进的 Inception 组合结构除了减少计算量之外还增加了网络的判别能力。

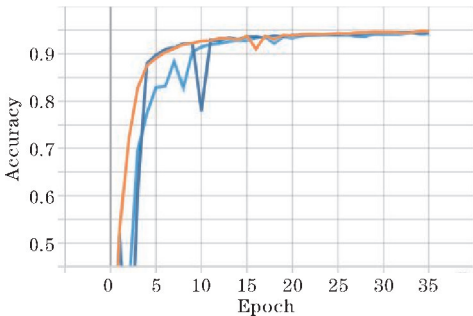


图 10 基于改进的 Inception 结构的网络模型运行结果

同时,在 HWDB2 全部数据集上运行本文的模型,识别准确率结果如图 11 所示,最终准确率稳定在 95.12% 左右。字符错误率结果如图 12 所示,最终稳定在 4.85% 左右。

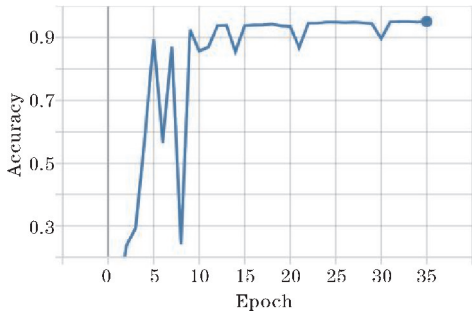


图 11 识别准确率

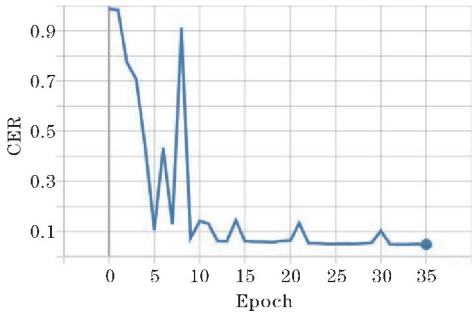


图 12 字符错误率

将本文的实验结果与其他模型进行对比,如表 2 所示。文献[15]提出的基础模型达到的准确率为 91.76%,经过训练样本预处理后准确率达到 94.21%,增大数据集后达到的准确率为 94.9%,本文并未对数据集进行预处理以及添加数据集等操作。本文提出的方法在识别准确率上较其他模型均有所提升。在原模型的基础上改进的卷积神经网络有较好的提取图像特征能力以致对预测解析出的文本序列准确率有提升帮助。

表 2 实验结果对比

模型	数据集	识别准确率/%
CRNN	HWDB2. X	90.00
文献[15]	HWDB2. X	94.90
文献[16]	HWDB2. X	91.32
文献[17]	HWDB2. X	90.9
本模型	HWDB2. X	95.12

4 结束语

针对中文文本行识别过程中遇到的问题,在 CRNN 模型的基础上,替换其 CNN 模块,将其改为基于 GoogLeNet 网络的 CNN。在此基础上,又进一步进行改进,添加了 CBAM 模块和改进的 Inception 结构。实验结果表明,本文的改进对网络性能确实有提升。在接下来的工作中,将继续探索,尝试改进循环层和转录层结构来达到更好的识别效果,或尝试使用语义分析的方法进行后处理来进一步提升识别准确率,或进一步增加对数据集的处理、增加数据等。

参考文献:

[1] 金连文,钟卓耀,杨钊,等.深度学习在手写汉字识别中的应用综述[J].自动化学报,2016,42(8):1125-1141.

[2] Messina R, Louradour J. Segmentation-free handwritten Chinese text recognition with LSTM-RNN[C]. Proceedings of the 13th IAPR International Conference on Document Analysis and Recognition, IEEE. 2015:171-175.

[3] Shi B, Xiang B, Cong Y. An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition[J]. Proceedings of IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016, 39(11):2298-2304.

[4] 周飞燕,金林鹏,董军.卷积神经网络研究综述[J].计算机学报,2017,40(6):1229-1251.

[5] T N Sainath, A. Mohamed, B. Kingsbury and B. Ramabhadran, Deep convolutional neural networks for LVCSR[C]. Proceedings of 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. 2013:8614-8618.

[6] 夏瑜潞.循环神经网络的发展综述[J].电脑知识与技术:经验技巧,2019,15(21):182-184.

[7] Graves Alex, Santiago Fernández, Faustino J, et al. Connectionist Temporal Classification: Labeling Unsegmented Sequence Data with Recurrent Neural Networks[C]. Proceedings of the 23rd International Conference on Machine Learning, ICML 2006, 2006:369-376.

[8] 蔡斯琪.不定长中文文本图像的识别算法研究

- [D]. 北京:北京交通大学,2021.
- [9] 张显杰,张之明. 基于卷积神经网络和 Transformer 的手写体英文文本识别[J/OL]. 计算机应用. <https://kns.cnki.net/kcms/detail/51.1307.tp.20220304.1230.006.html>.
- [10] Wojna Z, Gorban A N, Lee D S, et al. Attention-based Extraction of Structured Information from Street View Imagery[C]. Proceedings of IEEE Computer Society, IEEE Computer Society, IEEE Computer Society. 2017:844-850.
- [11] Liem H D, Minh N D, Trung N B, et al. FVI: An End-to-end Vietnamese Identification Card Detection and Recognition in Images[C]. Proceedings of 2018 5th NAFOSTED Conference on Information and Computer Science (NICS). IEEE, 2018:338-340.
- [12] 刘高洪,孙博洋,刘宗伟,等. 基于 CRNN 的车牌识别方法[J]. 计算机科学与应用, 2021, 11(11):2804-2816.
- [13] Woo S, Park J, Lee J Y, et al. CBAM: Convolutional Block Attention Module[C]. Proceedings of 15th European Conference on Computer Vision, ECCV 2018, 2018: 3-19.
- [14] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning[C]. Proceedings of 31st AAAI Conference on Artificial Intelligence, AAAI 2017, 2017:4278-4284.
- [15] 杨亚锋. 基于 CNN-RNN 框架的脱机手写中文文本行识别模型及其加速和压缩方法的研究[D]. 广州:华南理工大学, 2019.
- [16] 石鑫,董宝良,王俊丰. 基于 CRNN 的中文手写识别方法研究[J]. 信息化建设. 2019, 43(11): 141-144.
- [17] 马洋洋. 基于深度学习的端到端脱机手写体识别技术研究[D]. 西安:陕西师范大学, 2021.

## Improved Chinese Handwritten Text Line Recognition based on CRNN

SHU Shanshan, ZHENG Xiaoxu, WEN Chengyu

(College of Communicating Engineering, Chengdu University of Information Technology, Chengdu 610225, China)

**Abstract:** Chinese handwritten text line recognition converts paper writing into editable electronic content. For the problems of random handwriting, the variety of Chinese characters, and low recognition accuracy of the method based on character segmentation. This paper proposes an improved end-to-end Chinese handwriting recognition method based on Convolutional Recurrent Neural Network (CRNN). First, the picture is passed to the feature extraction network based on the improved Inception structure, the network first improved the GoogLeNet model, and then added the attention mechanism module (CBAM) and the Inception combined structure, after the improvement the model can do better in extracting the effective features of the picture. Then the extracted picture features were passed to the recurrent layer, a two-layer bidirectional long-short-term memory network (BiLSTM), for prediction. Finally, the predicted sequence was passed to the transcription layer, the Connectionist Temporal Classification (CTC), for transcriptional output. Experiments use the CASIA-HWDB2 dataset. The results show that the method can obtain a recognition accuracy of 95.12%, which proves the feasibility of the method.

**Keywords:** handwritten chinese text recognition; CRNN; CBAM; BiLSTM; CTC