李昭仪，于淼，于晓鹏. 基于满文不定长字元的识别方法研究[J]. 智能计算机与应用,2025,15(12):123-130. DOI:10.20169/j.issn.2095-2163.24032705

基于满文不定长字元的识别方法研究

李昭仪，于　淼，于晓鹏
(吉林师范大学 数学与计算机学院，吉林 四平 136000)

摘　要：本文提出一种基于卷积循环神经网络(Convolutional Recurrent Neural Network，CRNN)的满文识别方法,在识别任务上表现稳定。为了使模型更快收敛,并且进一步提高识别准确率,将权重初始化方法更换为针对 ReLU 激活函数的 Kaiming normal,并且对网络中卷积层结构进行修改,改进后的网络对本文制作的满文字元数据集的识别率从 99.7%提高到 99.76%,对满文单词数据集的识别率从 99.15%提高到 99.61%,对满文字元和单词复合数据集的识别率从 99.39%提高到 99.68%。此外,实验结果表明在提高识别率的同时训练速率也有明显提升。

关键词：满文识别；CRNN；深度学习
中图分类号：TP391.4　　**文献标志码**：A　　**文章编号**：2095-2163(2025)12-0123-08

Research on recognition method based on Manchu indefinite length characters

LI Zhaoyi，YU Miao，YU Xiaopeng
(College of Mathematics and Computer，Jilin Normal University，Siping 136000，Jilin，China)

Abstract：This paper proposes a Manchu recognition method based on Convolutional Recurrent Neural Network(CRNN)，which performs stably on the recognition task. In order to make the model converge faster and further improve the recognition accuracy，the weight initialization method is changed to Kaiming normal for ReLU activation function. The recognition rate of the improved network for the Manchu character dataset made in this paper is increased from 99.7% to 99.76%，from 99.15% to 99.61% for the Manchu word data set，and from 99.39% to 99.68% for the full text and word composite data set. In addition，the experimental results show that the training rate is also significantly improved while the recognition rate is improved.

Key words：Manchu recognition；CRNN；deep learning

0　引　言

在历史的长河中,形成了大量的满文古籍文献,具有重要的历史文化价值。由于年久日深,满文材料已经遭到了不同程度上的损坏,因此需要将这些资料转化为方便大家学习查阅的电子版进行保存。然而,目前认识和了解满文的人越来越少,急需对满文的识别进行研究。

1　相关研究

1.1　满文基础知识

满文是一种拼音文字,由元音和辅音组成。满文共有 12 字头,每组字头约有 100 余字,满文单词由 12 个字头拼接变形而成。满文的书写特征是从上到下沿着一条直线竖排书写,成文规则是从左到右竖写。满文字元在满文单词中的不同位置书写规则不同,分别是独立体、词首型、词中型、词尾型,以字母 a 为例,如图 1 所示。

图 1　字母 a 在满文单词中不同位置的写法

Fig. 1　Different positions of the letter a in Manchu words

1.2　CRNN

CRNN 主要用于端对端识别长度不固定的文本序列,应用该网络模型时,无需提前对文字图像进行逐字切割[1],CRNN 由特征提取、序列分析和序列解码 3 个部分。该卷积循环神经网络使用的是双向长

作者简介：李昭仪(1999—)，女,硕士研究生,主要研究方向:深度学习,图像检测与识别研究；于晓鹏(1964—),男,学士，教授,主要研究方向:深度学习,数值模拟研究。
通信作者：于　淼(1990—),男,博士,讲师,主要研究方向:深度学习,数值模拟研究。Email:ym@jlnu.edu.cn。

收稿日期：2024-03-27

短期记忆网络(Bidirectional Long - Short Term Memory, BiLSTM[2])和卷积神经网络(Convolutional Neural Network, CNN)对文字图像上的文字进行特征提取。

CRNN 的结构如图 2 所示,该模型可实现较长的文本序列识别。

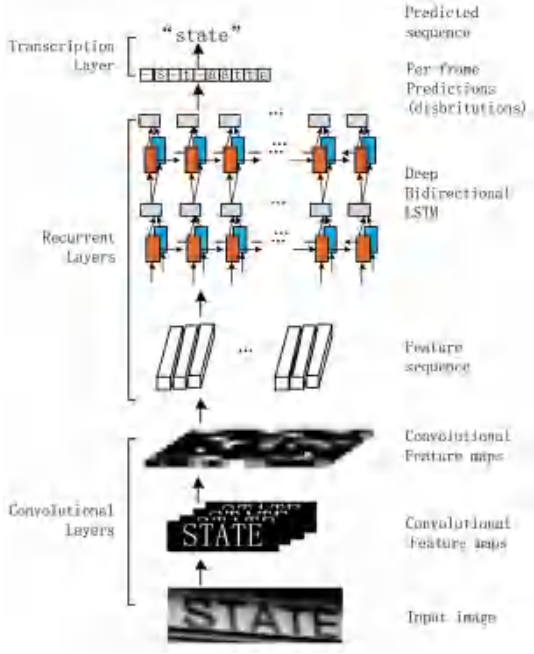

图 2 CRNN 结构图

Fig. 2 CRNN structure diagram

1.2.1 CNN

在 CRNN 中使用深度 CNN,对输入的图像提取较像特征并转化为特征图,该特征图作为下一步骤循环层的输入,根据循环神经网络(Recurrent Neural Netwok, RNN)接入数据信息的需要,经过调整得到特征向量序列。该调整过程称为 Map - to - Sequence,如图 3 所示。

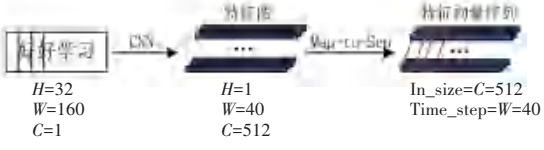

图 3 Map-to-Sequence 过程图

Fig. 3 Map-to-Sequence process diagram

为了使提取的特征图尺寸相同,因此先要在输入图像时将图像尺寸调整到固定大小,在满文字元或单词图像进入 CRNN 网络模型训练之前需要做归一化处理[3]。由于卷积神经网络中卷积层和最大池化层的存在,使其具有平移不变性的特点,如图 4 所示。

图 4 特征向量对应感受野图示

Fig. 4 Feature vectors correspond to receptive fields

其中每个特征向量的维度就是该特征图的通道数。在提取特征信息的同时对于已经成功提取的特征信息进行整合,最终以特征向量的形式输出,CRNN 网络会将输出的特征向量转换成特征序列输送到循环层以便于后续网络模型进行标签预测。

1.2.2 RNN

循环神经网络(RNN)虽然可以克服训练数据是连续且长短不一的序列时的训练问题,但到真正满文识别实验中,RNN 经常在训练方面存在诸多问题。由于 RNN 在多层网络中训练长序列时存在梯度消失和梯度爆炸问题,所以 CRNN 中使用的是长短期记忆网络,LSTM 的基本单元结构如图 5 所示,其是由一个单元模块、输入门、输出门和遗忘门组成。

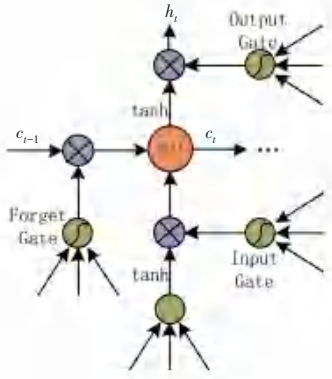

图 5 LSTM 基本单元结构

Fig. 5 Basic unit structure of LSTM

循环层由一个双向 LSTM 循环神经网络构成,预测特征序列中每一个特征向量的标签分布,其将卷积层得到的特征序列 $x = [x_1, x_2, \cdots, x_T]$ 中的每一个分量 x_i 映射生成一个标签分布 y_i。采用循环层有以下几点好处:第一,循环层可以提取上下文信息。第二,经过循环层得到的误差在反向传播时可以传回卷积层。第三,也是最重要的一点,RNN 可以处理长短不一的序列,而这是卷积层所不具备的。所以这里使用 LSTM 网络模型来代替传统的 RNN,LSTM 是 RNN 的一种变体,也称长短时记忆结构,通过门机制(包括遗忘门、输入门和输出门)将短时

记忆和长时记忆结合到一起[4-5]，在一定程度上解决了这问题。深度双向 LSTM 结构如图 6 所示。

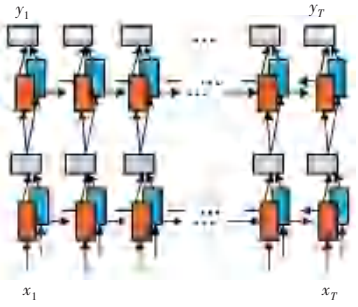

图 6　深度双向 LSTM 结构

Fig. 6　Depth bidirectional LSTM structure

1.2.3　CTC

循环层完成了对序列标签的预测工作，并向转录层传递，得到最终的识别结果。该过程用到的方法是连接时序分类 CTC 算法[6]。假设输入的满文单词序列长度为 T，每个样本点 $t(T > t)$ 在循环层的最后一层都会输出一个 Softmax 向量，表示该样本点的预测概率，所有样本点的这些概率传输给 CTC 模型后，经过对数据处理，就可以得到最终的满文单词序列标签。

2　方法设计

2.1　实验方法

本文实验使用的 CRNN 网络包括特征提取、序列分析和序列解码 3 个模块，使用的是双向长短期记忆网络（BiLSTM）和卷积神网络（CNN）对文字图像上的文字进行特征提取，之所以使用双向长短期记忆网络是因为 LSTM 是单向结构，无法兼顾上下文信息，其内部结构如图 7 所示。

在对文字图片进行识别时将两个 LSTM 分别沿正向、反向拼接形成BiLSTM，可以根据后面的信息预测出更准确的序列。最后引入语音识别领域的 CTC（Connectionist Temporal Classification）[7] 到图像处理序列长度不固定对齐问题，在 RNN 循环层获得到的预测标签通过 CTC 转录层形成最终的识别结果。

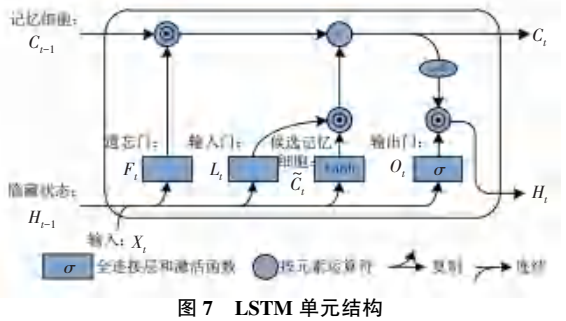

图 7　LSTM 单元结构

Fig. 7　LSTM cell structure

2.2　实验改进

2.2.1　权重初始化方法优化

CRNN 对输入图片的推理过程如图 8 所示。为了提高 CRNN 网络的性能，得到更高的准确率，本文采用更适配的权重初始化方法对网络进行改进，形成本文的 KM_CRNN。

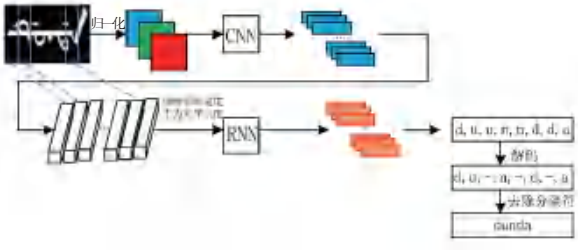

图 8　CRNN 识别图片信息过程

Fig. 8　Process of CRNN identifying picture information

为加快模型收敛速度，并进一步提升训练后的模型的性能，本文引入一种新的权重初始化方法 Kaiming normal[8]，其作用顺序如图 9 所示。

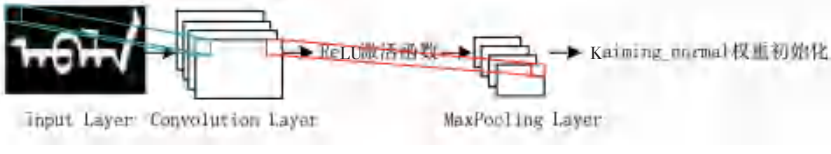

图 9　Kaiming normal 作用顺序图

Fig. 9　Sequence diagram of Kaiming normal action

Kaiming normal 初始化方法是由 He Kaiming 团队提出，该方法通过均值为 0 和标准偏差为 stddev = sqrt(1/fna_in) 的截断正态分布中抽取权重值来初始化神经网络权重。

Kaiming normal 的初始化公式为：

$$\mathcal{N}\left(0, \sqrt{\frac{2}{(1+a^2)n_l}}\right) \quad (1)$$

其中，n_l 表示第 l 层的神经元数量，a 表示 ReLU 激活函数的斜率，通常取值 0。

从第一层一直往前进行前向传播，可以得到某层的方差：

$$\mathrm{Var}(y_i) = \mathrm{Var}(y_1)\left(\prod_{i=0}^{L}\frac{1}{2}n_i\mathrm{Var}(w_i)\right) \quad (2)$$

这里的 x_1 就是输入的样本，将其归一化处理，

所以 $\mathrm{Var}(x_1)=1$，现在让每层输出方差等于 1，即：$\frac{1}{2}n_i\mathrm{Var}(w_i)=1$ 且 $\mathrm{Var}(w_i)=\frac{2}{n_i}$，基于此，正向传播时，Kaiming 初始化的权重分布为：

均匀分布：$W \sim U[-\sqrt{\frac{6}{n_i}},\sqrt{\frac{6}{n_i}}]$

高斯分布：$W \sim N[0,\frac{2}{n_i}]$

因为反向传播时 $\Delta x_i=\hat{w}_i\Delta y_i$，其中 Δ 表示损失函数并对其求导 $\hat{w}_i$ 为参数，可以得到公式：

$$\begin{aligned}\mathrm{Var}(\Delta x_i) &= \hat{n}\mathrm{Var}(\hat{w}_l\Delta y_i)=\hat{n}[\mathrm{Var}(\hat{w})\mathrm{Var}(\Delta y_i)+\\ &\mathrm{Var}(\hat{w}_l)(E\Delta y_i)^2+\mathrm{Var}(\Delta y_i)(E\hat{w}_i)^2]=\\ &\hat{n}\mathrm{Var}(\hat{w}_l)\mathrm{Var}(\Delta y_i)=\\ &\frac{1}{2}\hat{n}\mathrm{Var}(\hat{w}_l)\mathrm{Var}(\Delta x_{i+1}) \qquad (3)\end{aligned}$$

其中 $\hat{n}$ 表示反向传播时输出通道数，最后得到 $\frac{1}{2}\hat{n}_i\mathrm{Var}(w_l)=1$ 即 $\mathrm{Var}(w_l)=\frac{2}{\hat{n}_l}$。因此反向传播时，Kaiming 初始化的权重分布为：

均匀分布：$W \sim U[-\sqrt{\frac{6}{n_i}},\sqrt{\frac{6}{n_i}}]$

高斯分布：$W \sim N[0,\frac{2}{n_i}]$

该方法考虑了 ReLU 对输出数据分布的影响，使得输入和输出数据方差一致，通过该方法随机初始化神经网络的权重，来提高神经网络的训练效果。在神经网络中，初始化权重是必不可少的一步，初始化权重的作用是为了给每个神经元不同的权重，使不同神经元彼此之间不会过度依赖，可以保证神经网络的可靠性和鲁棒性，本文所用方法显著提升了网络模型的收敛速度，在缩短网络运行时间的同时提升了识别的准确率。

2.2.2 网络结构优化

CRNN 的网络结构如图 10 所示。从图中可以看出，CRNN 网络卷积层中使用了 4 次最大值池化操作，前两次的池化层的窗口尺寸为 2×2，后两个池化层的窗口尺寸为 1×2，也就是图片的高度进行了 4 次减半调整，而宽度只进行了前两次池化发生的两次减半调整。

由于卷积层输入的满文字元图片尺寸较小，包含的像素特征信息较少，经过上述的网络卷积池化操作之后，极有可能造成图片的特征信息在深层网络中丢失。因此本文实验中对 CRNN 网络卷积层结构进行优化，减少池化层和卷积层，并且使用一个 5 * 5 卷积代替两个 3 * 3 卷积，形成了本文的 SOKM_CRNN。结构优化后的网络结构如图 11 所示。

图 10 CRNN 网络结构图

Fig. 10 CRNN network structure diagram

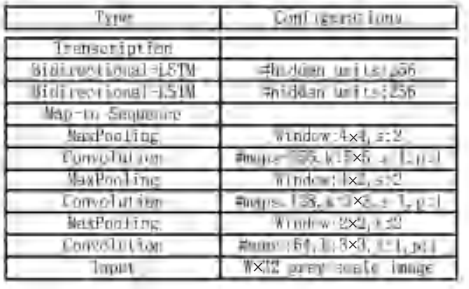

图 11 优化后的网络结构

Fig. 11 Network structure after optimization

通过减少参数数量，可以降低计算复杂度，并且可以有效防止网络模型训练过程中出现过拟合的情况，有效地提升了识别效果。

3 实验结果分析

3.1 评估指标

本文将从相同字元情况下，不同识别方法对于满文字元和满文单词的识别准确率，以及训练耗费时长作为评估指标。

3.2 数据集及实验参数设置

本文的满文字元训练参数见表 1。根据数据集的划分，本文的满文识别部分共设置 5 000 个 Epoch，Batch Size 为 64，初始学习率为 0.01。

表 1 满文识别训练参数

Table 1 Manchu recognition training parameters

参数名称	参数
学习率	0.01
Epoch	5 000
Batch Size	64

3.3　对比实验及实验结果分析

本文分别采用 VGG16、CRNN、将权重初始化方法更换为 Kaiming normal，在使用 Kaiming normal 的基础上优化网络结构 4 种方法进行训练，模型识别结果见表 2。改进识别网络之后，满文字元的识别准确率达到 99.27%，满文单词的识别率达到 99.58%，整个数据集的识别率达到 99.44%，在提高识别率的同时也提高了训练速率，在训练满文单词数据集时每训练一轮较改进之前减少 0.01 s，在训练满文字元数据集时每训练一轮较改进之前减少 3.65 s，在训练满文复合数据集时每训练一轮较改进之前减少 0.18 s。

表 2　改进算法的实验结果

Table 2　Experimental results of the improved algorithm

识别方法	是否有满文字元	是否有满文单词	准确率/%	运行耗时/(s · Epoch^{-1})
VGG16	√	×	67.30	83.75
	×	√	1.10	5.60
	√	√	60	89
CRNN	√	×	99.70	8.52
	×	√	99.15	1.67
	√	√	99.39	7.29
KM_CRNN	√	×	99.27	4.87
	×	√	99.58	1.66
	√	√	99.44	7.11
SOKM_CRNN	√	×	99.76	3.60
	×	√	99.61	1.46
	√	√	99.68	4.87

1) VGG16

2014 年牛津大学 Karen Simonyan 和 Andrew Zisserman 提出 VGG 卷积神经网络[9]，VGG16 由 13 个卷积层和 3 个全连接层构成，结构示意图如图 12 所示。

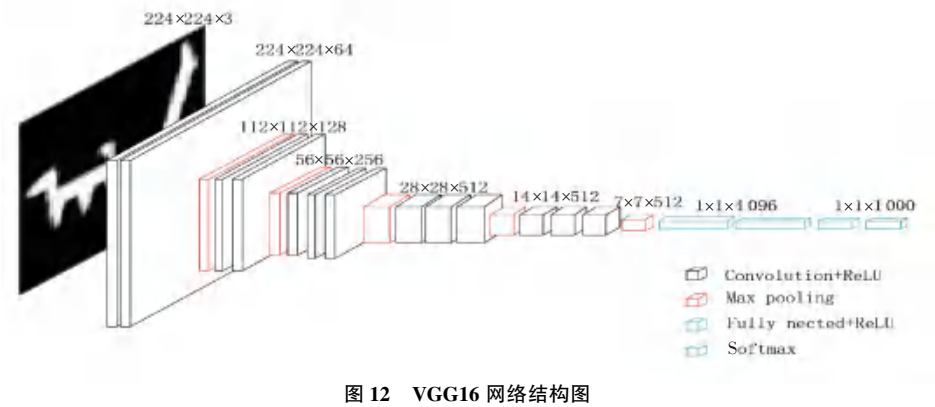

图 12　VGG16 网络结构图

Fig. 12　VGG16 network structure diagram

通过训练可知，准确率在 13 个 Epoch 后达到饱和。图 13 是训练到 Epoch = 12 时训练数据集的准确率。从图中可以得到在字元训练集的准确率达到 67.3%，既包含字元又包含单词的复合数据集准确率达到 60.0%。由于 VGG16 是一个通用的图像分类模型，其在处理各种不同类型的图像上表现良好，但对于字符串识别这种特定任务，需要更适配的模型结构和训练策略，因此单词训练集的准确率仅达到 1.1%左右。

2) CRNN

本文测试了 CRNN 对满文字元、单词、单词+字元数据集的识别能力，根据实验结果得到关于损失值的结果图，如图 14 所示。从该图中可以看出在 Epoch = 4 800 之后，数据集的损失值降到 0 左右。

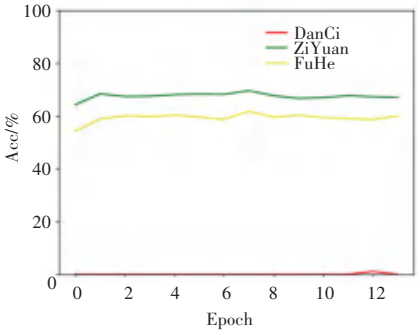

图 13　VGG 在不同满文数据集的识别准确率

Fig. 13　Recognition accuracy of VGG in different Manchu data sets

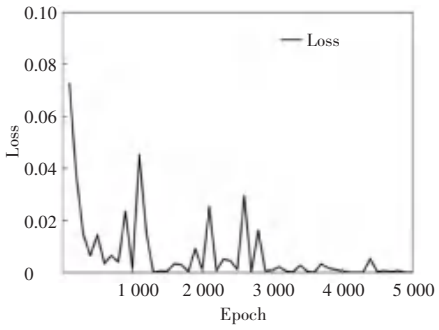

图 14 CRNN 在复合数据集上训练时的损失值

Fig. 14 Loss value of CRNN when training on a composite data set

观察对数据集中各类数据的识别情况,可以大概分为以下几类:

如图 15 所示,该图是利用 CRNN 训练好的模型进行字元识别的结果,经统计,最终对整个字元数据集识别的准确率达到 99.7%。

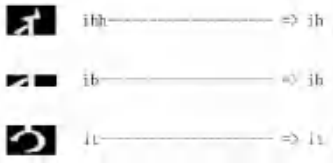

图 15 CRNN 识别满文字元结果

Fig. 15 CRNN recognition result of full text

如图 16、图 17 所示,分别是利用 CRNN 训练好的模型进行满文单词识别的正确结果和错误结果。左侧是图像,右图是检测结果。从图中可以看出,当满文单词的笔画之间分隔清晰时,可以得到很好的识别效果,但是当笔画复杂、笔画之间粘连严重或者字元之间极其相似时,就会有识别出错的情况。

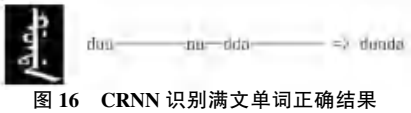

图 16 CRNN 识别满文单词正确结果

Fig. 16 CRNN recognition of Manchu words correct results

图 17 CRNN 识别满文单词错误结果

Fig. 17 Error results of CRNN recognition of Manchu words

经过对识别结果的统计,主要错误集中在把字元 d 做词首错认为字元 m 做词首的情况以及把字元 c 做词首错认为字元 z 做词首的情况。最终对整个满文单词数据集识别的准确率达到 99.15%。而对将字元和单词图像综合在一起并且新添加部分待识别图片之后得到的数据集进行测试时,得到识别的准确率达到 99.39%。

3) KM_CRNN

在上文测试了 CRNN 对满文字元、单词、单词+字元数据集的识别能力的基础上,测试了改进的 CRNN——KM_CRNN 的识别能力,综合二者在各个数据集上的识别结果得到识别准确率的结果对比图,如图 18 所示。改进之后的网络在训练到 Epoch = 5 000 时,识别字元的准确率为 99.27%,识别单词的准确率达到 99.58%,识别单词加字元复合数据集的准确率达到 99.44%。

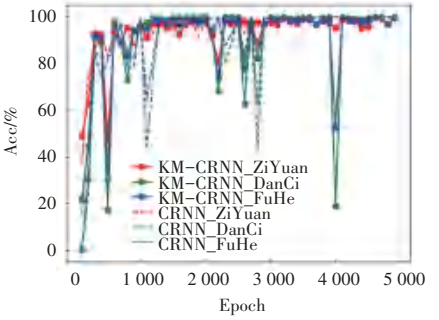

图 18 CRNN 和 KM_CRNN 在不同数据集上识别的准确率结果对比图

Fig. 18 Comparison of recognition accuracy results of CRNN and KM_CRNN on different data sets

如图 19 所示,分别是对满文字元 a 做词首、词中、词尾时识别情况,经统计,最终对整个字元数据集识别的准确率达到 99.27%。

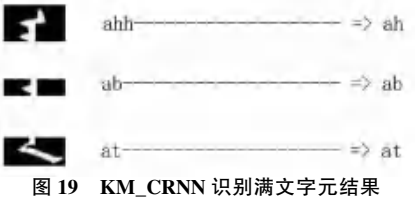

图 19 KM_CRNN 识别满文字元结果

Fig. 19 KM CRNN recognition result of full text

如图 20、21 所示,是利用 KM_CRNN 训练好的模型进行满文单词识别的结果。上文利用 CRNN 识别单词 DZUSI 错误的结果,在 KM_CRNN 方法的识别下,会得到正确的结果,进一步提高了识别的准确率。对整个单词数据集的识别准确率从 99.15% 提高到 99.58%。

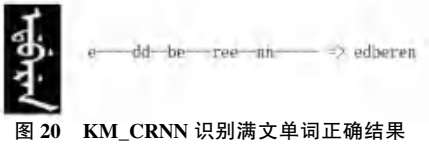

图 20 KM_CRNN 识别满文单词正确结果

Fig. 20 Correct result of KM CRNN recognition of Manchu words

图 21 KM_CRNN 识别满文单词改正结果

Fig. 21 Correction results of KM CRNN recognition of Manchu words

4）SOKM_CRNN

从图 18 中可以看出 KM_CRNN 在识别准确率方面表现优于 CRNN，在以上网络结果基础上，综合分析 CRNN、KM_CRNN 和 SOKM_CRNN 三者网络识别结果，得到了以下识别准确率对比图。如图 22～图 24 所示，从图中可以看出，SOKM_CRNN 网络的识别准确率表现最为亮眼。

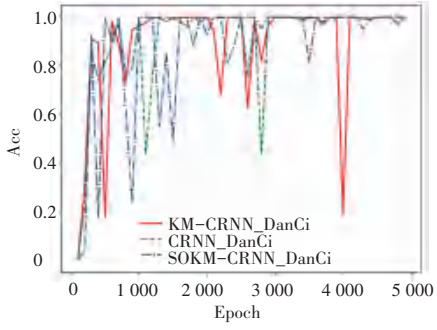

图 22　CRNN、KM_CRNN 和 SOKM_CRNN 识别满文单词准确率对比

Fig. 22　Comparison of the accuracy of CRNN, KM_CRNN and SOKM_CRNN in recognizing Manchu words

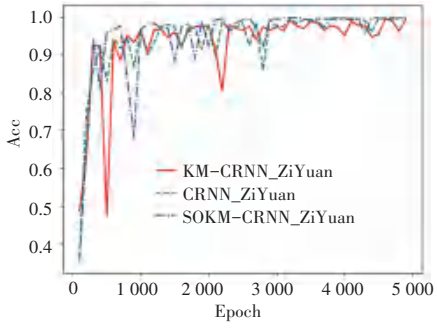

图 23　CRNN、KM_CRNN 和 SOKM_CRNN 识别满文字元准确率对比

Fig. 23　Comparison of the accuracy of CRNN, KM_CRNN and SOKM_CRNN for identifying Manchu character elements

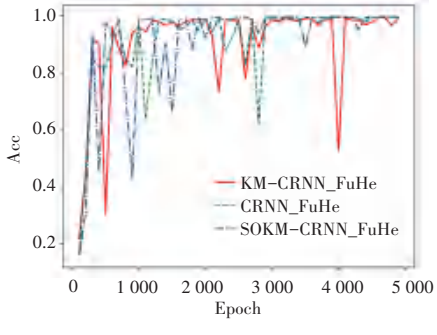

图 24　CRNN、KM_CRNN 和 SOKM_CRNN 识别满文单词+字元复合数据集准确率对比

Fig. 24　Comparison of accuracy of CRNN, KM_CRNN and SOKM_CRNN in recognition of Manchu word + character complex dataset

经过对识别结果的统计，发现对于满文字元 g 在满文单词 BOIHOLONGGOLO 中时，CRNN 和 KM_CRNN 均会对该字元识别错误，识别成 k，而 SOKM_CRNN 可以对该错误进行更改。如图 25 所示是 SOKM_CRNN 的识别结果。

图 25　SOKM_CRNN 识别满文字元的改进结果

Fig. 25　Improved results of SOKM_CRNN for recognizing Manchu character elements

在对满文单词识别结果进行统计发现对于满文单词 TSU 的识别结果在使用 CRNN 和 KM_CRNN 时均会发生识别错误，利用 CRNN 识别的结果是 MZU，而利用 KM_CRNN 识别的结果为 MU，而 SOKM_CRNN 可以对该错误进行更改，识别为对的结果。如图 26 所示是 SOKM_CRNN 的识别结果。

图 26　SOKM_CRNN 识别满文单词改进结果

Fig. 26　Improved results of SOKM_CRNN recognition of Manchu words

如图 27 所示，该图展示了 CRNN、KM_CRNN 和 SOKM_CRNN 分别在复合数据集进行训练耗时情况，从图中可以看出，本文最后改进的 SOKM_CRNN 模型 训练时长 显著降低，通过对实验结果统计，平均每训练 100 轮耗费时长减少 242 s。

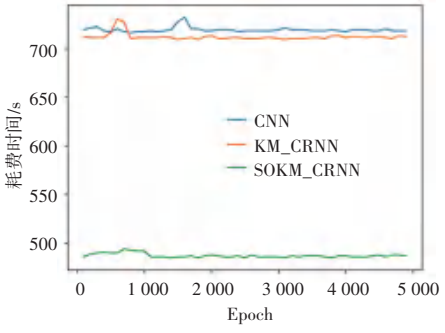

图 27　CRNN 和 KM_CRNN 训练时耗费时间

Fig. 27　Time spent on training CRNN and KM CRNN

4　结束语

本文的识别方法是先将满文单词切分成要识别的字元序列，再将该序列放到 CRNN 中的卷积层来提取特征，再到循环层更好地来处理序列信息，在循环层中使用双向 LSTM，提高了准确率，最后将信息输入到转录层将前面通过 CNN 层和 RNN 层得到的预测序列转换成标记序列，得到最终的识别结果。本文提出的相较于传统的 CRNN 识别网络 KM_

CRNN 识别网络在满文单词识别任务中节省了大量的网络模型训练时间,同时模型对于满文单词识别的准确率有了明显的提升。

在下一步的研究工作中,将考虑引入 Attention 注意力机制模块,进一步提高识别的准确率,以获取更优的实验结果。

参考文献

[1] SHI B, BAI X, YAO C. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017,39(11):2298-2304.

[2] HOCHREITER S, SCHMIDHUBER J. Long Short - Term Memory[J]. Neural Computation, 1997, 9(8):1735-1780.

[3] 周乐. 基于 YOLOv4 以及 CRNN+CTC 算法的中文车牌识别系统[D]. 南京:南京邮电大学, 2023. DOI: 10. 27251/d. cnki. gnjdc. 2022. 000671

[4] HUANG Z, XU W, YU K. Bidirectional LSTM-CRF models for sequence tagging[J]. arXiv preprint arXiv,1508. 01991,2015.

[5] MA X, HOVY E. End-to-end sequence labeling via Bidirectional LSTM- CNNs - CRF[J]. Transactions of the Association for Computational Linguistics,2016,4:267-280. DOI:10. 13140/RG. 2. 1. 2182. 5685

[6] 谢旭康. 基于端到端的语音识别模型研究及系统构建[D]. 无锡:江南大学,2023. DOI:10. 27169/d. cnki. gwqgu. 2022. 000190

[7] 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. IMLS, 2006: 369-376.

[8] HE K, ZHANG X, REN S, et al. Delving deep into rectifiers: Surpassing human - level performance on imagenet classification [C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2015: 1026 - 1034. DOI: 10. 1109/ICCV. 2015. 123

[9] SIMONYAN K, ZISSERMAN A . Very deep convolutional networks for large - scale image recognition[J]. arXiv preprint arXiv, 1409. 1556, 2014.