第 58 卷 第 12 期/2021 年 6 月/激光与光电子学进展 研究论文

激光与光电子学进展 先进成像

基于 CTC-Attention 脱机手写体文本识别

马洋洋，肖冰*

陕西师范大学计算机科学学院，陕西 西安 710062

摘要 针对脱机手写体书写随意、字符分割困难和识别精度依赖字典等问题，提出了一种基于 CTC-Atention 脱机手写体文本识别算法。利用卷积神经网络(CNN)与双向长短时期记忆网络(BLSTM)实现对图像的特征编码，然后使用基于链接时序分类(CTC)模型和基于注意力机制(Attention-based)模型的多任务学习(MTL)框架实现对特征序列的解码。在训练过程中利用 CTC 模型和注意力机制模型同时训练，有效地解决了 CTC 预测局部信息时忽略了整体信息，以及注意力机制解码不受约束的问题。在经典的手写英文单词数据集 IAM 上进行实验，结果表明，该方法的字符准确率达到了 93.4%，单词准确率达到了 81.8%，证明了提出方法的可行性。

关键词 图像处理；脱机手写体文本识别；链接时序分类；注意力机制；多任务学习

中图分类号 TP391 **文献标志码** A **doi**：10.3788/LOP202158.1210007

Offline Handwritten Text Recognition Based on CTC-Attention

Ma Yangyang, Xiao Bing*

College of Computer Science, Shaanxi Normal University, Shaanxi, Xi'an, 710062 China

Abstract Aiming at the problems of casual writing of the offline handwritten text, difficulty in character segmentation, and the dependence of recognition accuracy on a dictionary, an offline handwritten text recognition algorithm based on connectionist temporal classification (CTC)-attention is proposed. The convolutional neural network and bidirectional long short-term memory are used to encode the image features. Multitask learning framework based on CTC and Attention-based models is used to decode feature sequences. In the training process, the CTC model and the attention mechanism model are used to train at the same time, which effectively solves the problem of ignoring the overall information when CTC predicts local information, and the problem of unconstrained decoding of the attention mechanism. Experiments on IAM dataset, i.e., the classical handwritten English word dataset, showed that the character accuracy rate of the proposed method is 93.4%, and the word accuracy rate is 81.8%, proving the proposed method's feasibility.

Key words imaging processing; offline handwritten text recognition; connectionist temporal classification; attention; multitask learning

OCIS codes 100.3008; 100.4996; 100.2000

1 引 言

文字识别是计算机视觉研究领域的分支之一，俗称光学字符识别(OCR)，是利用光学技术和计算机技术将纸质文档中的文字转换成为黑白点阵的图像文件，并通过识别技术将图像中的文字转换成文本格式，供进一步编辑加工的技术，属于模式识别和人工智能，是计算机科学的重要组成部分。OCR 技术的兴起是从印刷体字体识别开始的，印刷体字体的成功识别为后来手写体的发展奠定了坚实的基础。手写体识别较印刷体识别难度高，而在手写体识别中脱机手写体识别难度又比联机手写体识别难

收稿日期：2020-08-24；**修回日期**：2020-09-28；**录用日期**：2020-10-14

基金项目：中央高校基本业务(GK201803058)

* **E-mail**：16392603@qq.com

度高，尽管最近有新的进展，但是个人书写的显著差异和手写字符的不精确性质使识别任务变得困难，脱机手写体识别仍是一个开放的研究问题。

目前，脱机手写体单词识别问题的研究主要针对以下两个方面[1]的区别：1）从手写图像中提取特征的策略不同；2）对分类器的输出进行解码以预测构成给定单词的字符序列的方法不同。

脱机手写体识别中主要通过两种策略来提取图像特征，分别是利用不同的计算机视觉技术进行检测和将图像的像素序列直接用作原始特征。Doetsch 等[2]使用第一种策略。Kozielski 等[3]在 8 pixel×32 pixel 大小的图像中采用主成分分析提取信息。Bideault 等[4]利用定向梯度直方图（HOG）进行特征提取。Graves 等[5]考虑了每个像素提取的特征，例如平均矩和其他矩，以及重心、转换和其他聚合。Bluche 等[6]直接使用像素特征作为模型的输入。Sueiras 等[1]提出基于注意力机制（Attention-based）的模型并使用水平滑动窗口来识别手写体文本，模型中直接使用像素特征作为模型的输入。

脱机手写体识别对手写预测输出进行解码，以将其转换为识别手写单词的字符序列的方法有两种。第一种方法是使用统计学方法隐马尔可夫模型（HMM）。Doetsch 等[2]提出将 HMM 以强制对齐的方式应用于训练集，以生成字符长度统计。Bluche 等[7]利用卷积神经网络（CNN）直接处理图像，使用 HMM 在迭代过程中给出对齐方式。Bianne-Bernard 等[8]提出基于 HMM 考虑动态信息和上下文信息的识别器。第二种是使用 Graves 等[5]提出的链接时序分类（CTC）目标函数。Graves 等[5]提出使用递归神经网络和链接时序分类的离线手写识别器。Voigtlaender 等[9]提出堆叠交替卷积和多向 MDLSTM 的多层，使用具有连接时序分类损失的 Softmax 层来处理输入和输出序列之间的对齐方式。Bluche 等[10]提出将 CNN 与 HMM 联合使用的手写单词识别方法。之后，他又提出双向长短时记忆（BLSTM）和 CTC 结合的手写体文本识别方法[11]。

注意力机制开始是在机器翻译中使用，并在序列学习的任务中发挥重要作用[12-13]，近年来在语音识别和手写体文本识别中得到进一步推广。2017 年，Bluche 等[6]提出基于注意力机制的端到端（end-to-end）手写体文本识别模型。2018 年，Sueiras 等[1]提出基于注意力机制的序列到序列（Seq2Seq）模型并使用水平滑动窗口进行了手写体文本识别。2018 年，Kang 等[14]提出基于混合注意力机制的序列到序列的手写体文本识别模型。

近年来，深度学习在手写识别中取得了较好的效果[15]。目前流行的手写体文本识别框架有 CRNN＋CTC 和 CNN＋Seq2Seq＋Attention。CTC 模型借鉴了马尔可夫假设，有效地解决了序列的动态对齐问题，注意力机制模型的编码-解码体系中[16]通过一个注意力机制解决了图像特征序列和标签之间的对齐问题。基于注意力机制的解码器网络使用注意力机制找出输出序列中每个字符与编码器网络针对图像输入生成的隐藏状态之间的对齐方式。在每个输出位置，解码器网络对每个输入时间计算隐藏状态与编码网络状态之间的匹配分数，以形成时间对分布，然后将其用于提取相应编码器网络状态的平均值。这种基本的时间注意力机制太灵活，因为它允许极不连续的对齐。但在脱机手写体文本识别中，特征输入和相应的字符输出通常以相同的顺序进行。

本文提出一种基于 CTC-Attention 多任务学习（MTL）文本识别算法，CTC 允许使用动态规划，能高效地计算严格单调的对齐方式。在训练过程中，将 CTC 目标作为正则化附加到基于注意力机制的编码器网络中，减少了不规则对齐字符的数量。在解码过程中，将基于注意力机制的分数和基于 CTC 分数结合在一个单词波束搜索算法中，以消除不规则对齐。

2 相关研究

2.1 CTC 模型

基于链接时序分类（CTC）模型是一个潜在变量模型，它单调地将输入序列映射到较短长度的输出序列，在输出序列和最终标签之间增加了多对一的空间映射。假设模型输出长度为 L 的字符序列 $C=\{c_l\in U\mid l=1,2,\cdots,L\}$，$U$ 是所有字符集合。CTC 使用额外符号“U_{blank}”表示字母边界，用于处理字母符号的重复。与 HMM/DNN 类似，带有附加空白符号的逐帧字符序列表示为 $Z=\{z_t\in U\cup U_{\text{blank}}\mid t=1,2,\cdots,T\}$。根据条件独立性假设，后验概率 $p(C\mid X)$ 的因式分解可表示为

$$p_{\text{ctc}}(C\mid X)\overset{\text{def}}{=}\sum_{Z}\prod_{t}p(z_t\mid z_{t-1},C)p(z_t\mid X)\frac{p(C)}{p(Z)},\tag{1}$$

式中：$p(z_t\mid X)$ 表示逐帧后验分布；$p(z_t\mid z_{t-1},C)$

表示转移概率；$p(C)$表示基于字符的语言模型；$p(Z)$表示隐藏状态序列的先验概率。对于所有输入 X 的帧后验分布，本研究定义了 CTC 的目标函数 $p_{\mathrm{ctc}}(C|X)$。$p(z_t|X)$通过 BLSTM 网络进行建模：$p(z_t|X)=\mathrm{Softmax}(\mathrm{Lin}(\boldsymbol{h}_t))$，$\boldsymbol{h}_t=\mathrm{BLSTM}(X)$，$\mathrm{Softmax}(\cdot)$是激活活函数，$\mathrm{Lin}(\cdot)$是线性层，将隐藏的矢量 $\boldsymbol{h}_t$ 转换为$(|U|+1)$二维矢量的线性层（+1 表示 CTC 中引入的空白符号）。

2.2　基于注意力机制的模型

和 CTC 方法比较，基于注意力机制的方法不需要条件独立性假设，而是根据概率链式法直接估计后验概率 $p(C|X)$，计算式为

$$p_{\mathrm{att}}(C|X)\overset{\mathrm{def}}{=}\prod_{l=1}^{L}p(c_l|c_1,c_2,\cdots,c_{l-1},X),\tag{2}$$

式中，$p_{\mathrm{att}}(C|X)$是基于注意力方法的目标函数。$p(c_l|c_1,c_2,\cdots,c_{l-1},X)$计算式为

$$p(c_l|c_1,\cdots,c_{l-1},X)=\mathrm{Decoder}(\boldsymbol{r}_l,\boldsymbol{q}_{l-1},c_{l-1}),\tag{3}$$

$$\boldsymbol{r}_l=\sum_{t=1}^{T}a_{lt}\boldsymbol{h}_t,\tag{4}$$

$$a_{lt}=\mathrm{Attention}(\{a_{l-1}\}_{t=1}^{T},\boldsymbol{q}_{l-1},\boldsymbol{h}_t),\tag{5}$$

$$\boldsymbol{h}_t=\mathrm{Encoder}(X),\tag{6}$$

式中：$\mathrm{Attention}(\cdot)$是基于内容的具有卷积特征的注意力机制，$a_{lt}$ 是注意力权重，它表示每个输出 c_l 的隐藏向量 $\boldsymbol{h}_t$ 的软对齐，c_{l-1} 和 $\boldsymbol{q}_{l-1}$ 分别是先前递归网络的输出和隐藏向量；$\boldsymbol{r}_l$ 是基于隐藏向量的加权总和在(4)式中形成按字母顺序的隐藏向量；$\mathrm{Decoder}(X)$通过 BLSTM 网络进行建模：$\mathrm{Decoder}(\cdot)=\mathrm{Softmax}(\mathrm{Lin}(\mathrm{LSTM}(\cdot)))$；$\mathrm{Encoder}(X)$是将输入 X 转换为逐帧的隐藏向量 $\boldsymbol{h}_t$，即 $\mathrm{Encoder}(X)=\mathrm{BLSTM}(X)$。

3　基于 CTC-Attention 的脱机手写体文本识别算法

CTC 和注意力机制模型在识别任务中能够取得良好的识别性能，但在脱机手写体文本识别中，还存在问题。CTC 模型的不足之处在于需要假设标签内部之间的条件具有独立性，且每次输出都是独立的单个字符概率，忽略了整体信息的不足。基于注意力机制模型在计算注意力权重时由所有帧的加权和表示，未引入任何对齐约束条件，致使解码时易产生错位，且存在纯粹数据驱动对长序列输入难以训练的问题[17]。而脱机手写体文本识别需要考虑数据的整体信息，解码时需要考虑输入和输出的对齐原则。针对这两种模型的不足和脱机手写体文本识别的需求，本文提出了基于 CTC-Attention 的多任务框架来解决这些问题。本文提出的基于 CTC-Attention 的脱机手写体文本识别算法模型包括两部分，即利用 CNN 与 BLSTM 实现对图像特征编码的编码器和利用 CTC-Attention 的多任务学习框架的解码器。网络整体设计如图 1 所示。

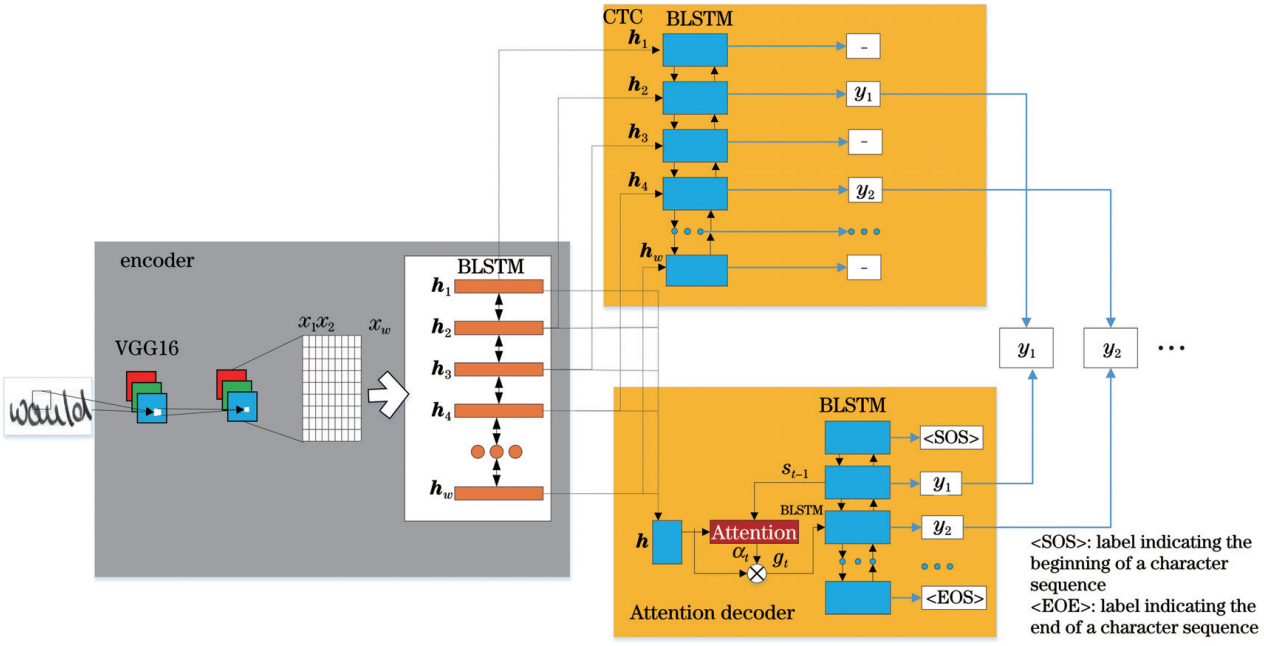

图 1　基于 CTC-Attention 的 MTL 框架

Fig. 1　MTL framework based on CTC-Attention

3.1 编码器

本文方法中，编码器采用 VGG＋BLSTM 结构，如图 2 所示。手写体文本图像视觉上不像真实世界图像那么复杂，背景相对简单。视觉几何群(VGG)网络的主要贡献在于使用非常小的 3×3 的卷积核进行网络设计，并将网络深度增加到 16～19 层，该网络在 2014 年 ImageNet 比赛中获得了定位第 1、分类第 2 的好成绩，具有很好的泛化能力。本研究使用基于 VGG-16-BN 深度卷积神经网络作为特征提取模块。在序列识别问题中，大多数情况不仅需要考虑当前获得的信息，同时要考虑之前存储的先验信息的附加信息作为当前的输入，故需要采用递归神经网络(RNN)，它是一种专门处理序列问题和具有时间依赖关系问题的网络。LSTM 是一种特殊的 RNN，可以解决梯度消失的问题，能够轻松学到长期以来的信息，且还可以自己学习哪些信息需要保留，哪些信息需要忘记。而 BLSTM 网络可以同时考虑序列的下文信息，完成序列的编码任务。

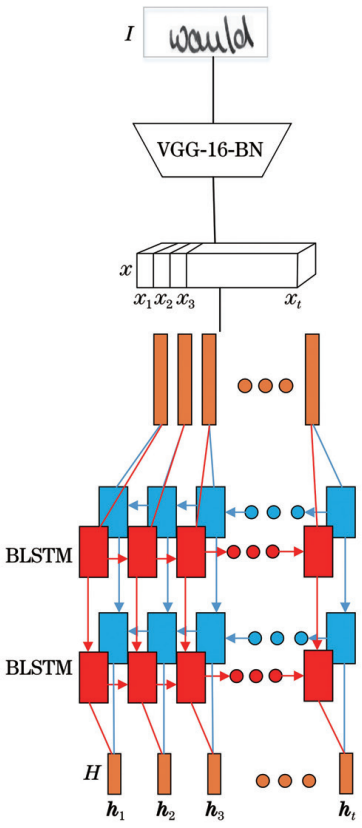

图 2 CNN＋BLSTM 结构

Fig. 2 CNN＋LSTM structure

在 VGG＋BLSTM 编码器中，利用 VGG-16-BN 网络学习空间信息，将图像 I 转换为中间层特征 x，然后将其重塑为二维特征图。利用两层的 BLSTM 网络学习序列特征，获得 H，H 是共享 x 相同宽度的编码器的输出，每个元素 $\boldsymbol{h}_i \in H$ 是每个时间步长的 BLSTM 的输出，将进一步用于解码。

越深的网络，提取的图像特征越抽象，越具有语义信息，即意味着拥有更强的表达能力。在编码器的特征提取模块，VGG-16-BN 网络包含 13 个卷积层、批归一化(BN)层和 4 个池化层。卷积层对输入图像信息通过加权来提取空间特征，池化层在降低特征图维数的同时，保留了特征图中最强的特征，BN 层作为单独的一层能够使网络远离饱和区，增强非线性。本质上解决了反向传播过程中的梯度问题，优化了网络模型。BLSTM 输入层包含了 2 个方向相反且独立的 LSTM，输出层结果包含了 2 个 LSTM 的结果。脱机手写体文本识别不仅要考虑当前字符的上文信息，也要考虑下文信息。BLSTM 能保证网络兼顾上下文信息并提取序列特征，从而在编码过程获得更好的特征信息[18]。

多任务学习(MTL)是一种机器学习模式，通过借助其他的辅助任务来帮助提升目标任务的学习效果，通过同时训练多个任务使得多个任务能够相互影响。

编码器利用 CTC 和注意力机制模型同时训练，使用 CTC 目标函数作为辅助任务来训练注意力模型的编码器，使用 CTC 和注意力模型共享编码器网络。编码器训练的目标函数为

$$L_{\mathrm{MTL}}=\lambda \log\, p_{\mathrm{ctc}}(C \mid X)+(1-\lambda)\log\, p_{\mathrm{att}}(C \mid X), \tag{7}$$

其中，λ 是可调节的参数，满足 $0 \leqslant \lambda \leqslant 1$，用于对两个目标函数进行线性插值。

3.2 解码器

基于 CTC-Attention 的脱机手写体文本识别，不仅在编码器训练中结合了 CTC 和注意力机制，在解码过程中也通过波束搜索算法将 CTC 和注意力机制的得分结合在一起，进行联合解码，以进一步消除不规则对齐。假设(1)式中的 p_{ctc} 和(2)式中的 p_{att} 分别是 CTC 和注意力模型给出的序列概率，则联合解码目标可以定义为

$$\hat{C}=\arg\max_{C \in U^{*}}\{\lambda \log\, p_{\mathrm{ctc}}(C \mid X)+(1-\lambda)\log\, p_{\mathrm{att}}(C \mid X)\}, \tag{8}$$

其中，$\hat{C}$ 是解码器根据给定的输入 X，找到的最有可能的字符序列。

在波束搜索过程中，解码器需要计算每个部分

的得分。设 Ω_l 是一个长度为 l 的一组局部推断；在波束搜索开始时，Ω_0 仅包含一个带有起始符号〈SOS〉的假设。从 $l=1$ 到 $L_{\max}$，Ω_{l-1} 中每个部分的推断得分通过附加可能的单个字符来扩展，并且新推断的得分存储在 Ω_l 中，其中 $L_{\max}$ 是假设要搜索的最大长度。每个新推断的得分在对数域的计算公式为

$$\alpha(h',X)=\alpha(g,X)+\log p(c'|g_{l-1},X), \tag{9}$$

其中，g_{l-1} 是 Ω_{l-1} 中的部分推断得分，c' 是附加到 g 的字母，h' 是由 $h'=g\cdot c'$ 得到的新得分。如果 c' 是表示序列末尾的特殊符号〈EOS〉，则 h' 被添加到 $\hat{\Omega}$ 中但不添加到 Ω_l 中，其中 $\hat{\Omega}$ 表示一组完整的推断。最后，$\hat{\Omega}$ 的计算表达式为

$$\hat{C}=\arg\max_{h'\in\hat{\Omega}}\log\alpha(h',X)。 \tag{10}$$

在波束搜索过程中，允许 Ω_l 仅保持有限数量且具有较高分数的推断得分，以提高搜索效率。

联合解码需要在波束搜索中将 CTC 和基于注意力机制的分数组合起来，其中：第一步使用波束搜索得到一组完整的推断，在此阶段仅考虑基于注意力机制的序列概率；第二步使用 CTC 和注意力概率对此完整的推断重新打分，其中 CTC 概率通过 CTC 前向算法所得。重打分步骤计算公式为

$$\hat{C}=\arg\max_{h'\in\Omega}\left[\lambda\alpha_{\text{ctc}}(h',X)+(1-\lambda)\alpha_{\text{att}}(h',X)\right], \tag{11}$$

其中，

$$\begin{cases}\alpha_{\text{ctc}}(h',X)\overset{\text{def}}{=}\log p_{\text{ctc}}(C|X)\\ \alpha_{\text{att}}(h',X)\overset{\text{def}}{=}\log p_{\text{att}}(C|X)\end{cases}。 \tag{12}$$

4　实验结果

4.1　数据集

本研究所用的数据集是 IAM 手写体英文单词数据集，它是手写体文本识别任务最常用的数据集。IAM 数据集包含 115320 个由 657 个作者提供的带标签的手写体英文单词图像，并带有大写字母、小写字母、数字和一些标点符号，但是它们不包括完整的 ASCII 可打印字符表。本课题组对原始数据进行筛选，最后获得了训练集 47981 张图像和验证集 7554 张图像，用于实验。表 1 是原始数据及标签示例。

4.2　评价标准

手写体文本采用的是标准性能指标：字符错误率（CER，R_{CE}）和字错误率（WER，R_{WE}）。CER 用于计算 Levenshtein 距离，它是将一个字符串替换为另一个字符串所需的字符，计算公式为

表 1　原始数据集示例

Fig. 1　Example of original dataset

Data	Label
	HR
	charged
	negotiate

$$R_{\text{CE}}=\frac{S+I+D}{N}, \tag{13}$$

式中，S 表示字符替换个数，I 表示字符插入个数，D 表示字符删除个数，N 表示标签字符串的字符个数。WER 用来计算识别结果与真实标签的相似度。识别出的字符串通过替换、插入和删除操作转换为标签字符串，计算公式为

$$R_{\text{WE}}=\frac{S_{\text{W}}+I_{\text{W}}+D_{\text{W}}}{N_{\text{W}}}, \tag{14}$$

式中，S_{W} 表示单词替换个数，I_{W} 表示单词插入个数，D_{W} 表示单词删除个数，N_{W} 表示标签字符串中单词个数。

4.3　实验细节及分析结果对比

实验环境：操作系统为 Linux 16.04，GPU 型号为 TITAN Xp，CUDA 版本为 10.2，采用 Pytorch 深度学习框架。其中，批量大小（Batchsize）设为 50。所有图像以原始长度和高度的比例，调整高度为 64 pixel，最长单词长度为 1011 pixel，对图像右侧填充零，图像的大小设置为 64 pixel×1011 pixel。

本文提出的 MTL 框架中，引入一个超参数 λ，用来平衡 CTC 模型和 Attention 模型的权重。$\lambda=1$ 时，只使用 CTC 模型解码；$\lambda=0$ 时，只使用 Attention 模型解码；λ 的取值范围是 0～1。实验中设计了 0，0.2，0.5，0.8，1.0 共 5 个 λ 值来进行实验对比分析。不同 λ 的单词准确率曲线如图 3 所示。

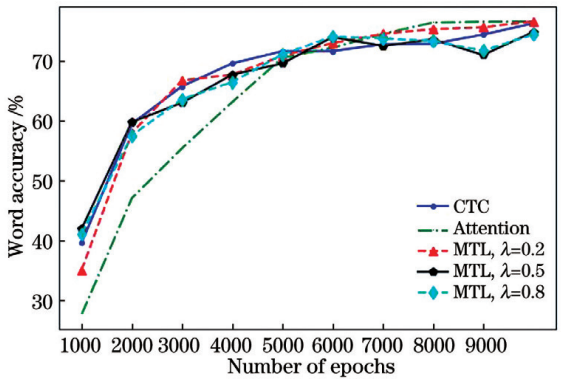

图 3　不同 λ 的单词准确率曲线

Fig. 3　Word accuracy curves of different λ

图 3 所示为不同 λ 值随着迭代次数的增加在验证集上单词准确性的学习曲线。实验结果表明，当 λ＝0.2 时，单词准确率达到最好。注意到，当 λ＝0.2 时，网络迅速学习并趋于稳定，较大的 λ 即赋予 CTC 损失更多的权重时，网络收敛效果不是很好，单词识别准确率会受到一定的影响。

4.4　实验结果及分析

表 2 是 MTL 模型在 IAM 脱机手写体单词数据集上识别的结果对比。实验结果表明，当参数 λ＝0.2 时，字符错误率和单词错误率均达到最小。

表 3 给出近年来几种流行方法在 IAM 脱机手写英文单词数据集上的识别结果对比，它们大多数在原始数据中使用了预处理操作。Mor 等[19]在实验中过滤掉字符和短单词，模型训练使用训练集数据和验证集数据。Krishnan 等[20]不仅使用了语言模型，而且在自己合成的数据上进行预训练后再训练模型。Stuner 等[21]的训练过程使用了 240 万的单词词典(Lexicon)和语言模型。Wiginton 等[22]对原始数据进行预处理，清理了标点符号和大写字母，使用轮廓规范化并对测试集进行了扩充。Bluche 等[6]额外使用 CTC 损失进行了预训练，获得了更有意义的特征表示。Sueiras 等[1]校正了图像中的线条偏斜和倾斜，根据基线和语料库线条对字符的高度进行了归一化处理。我们的模型没有使用任何的预处理操作，但识别结果还是比较令人满意的。

表 2　MTL 在验证数据集上的字符错误率和单词错误率

Table 2　CER and WER of MTL on valid dataset

unit：%

Model	CER(valid)	WER(valid)
CTC	9.7	27.6
Attention	7.1	19.1
MTL(λ＝0.2)	6.6	18.2
MTL(λ＝0.5)	7.4	19.8
MTL(λ＝0.8)	10.4	29.2

表 3　几种流行方法在 IAM 数据集上识别率比较

Table 3　Comparison of the recognition rate of several popular methods on the IAM dataset

Method	Author	Pre-processing	Lexicon	Language model	Pre-train	CER	WER
RNN+CTC	Mor *et al*[19]						20.90
	Krishnan *et al*[20]				Synthetic	6.34	16.19
	Stunner *et al*[21]		2.4 million	√		4.77	13.30
	Wiginton *et al*[22]	√		√		6.07	19.07
Attention	Bluche *et al*[6]				CTC	12.60	
	Sueiras *et al*[1]	√				8.80	23.80
CTC_Attention	Ours					**6.60**	**18.20**

5　结　　论

传统的脱机手写体文本识别方法需要进行复杂的预处理和精心设计的特征提取方法，导致识别精度低，泛化能力差。本文提出基于 CTC 模型和 Attention 模型的多任务学习(MTL)框架，不需要复杂的预处理操作，弥补了 CTC 解码忽略整体信息的不足，解决了 Attention 模型解码不受约束的问题。本实验结果表明，提出的框架在 IAM 脱机手写体英文数据集上字符准确率达到了 93.4%，单词准确率达到了 81.8%。接下来的研究工作，将致力于脱机手写体行级文本识别，使用合适的网络结构进行训练以达到更高的识别率。脱机手写体文本识别是模式识别、人工智能领域中一个重要研究方向，该课题的研究具有重要的理论价值和实用意义，值得继续研究。

参　考　文　献

[1]　Sueiras J, Ruiz V, Sanchez A, et al. Offline

continuous handwriting recognition using sequence to sequence neural networks[J]. Neurocomputing, 2018, 289: 119-128.

[2] Doetsch P, Kozielski M, Ney H, et al. Fast and robust training of recurrent neural networks for offline handwriting recognition[C]//2014 14th International Conference on Frontiers in Handwriting Recognition, September 1-4, 2014, Hersonissos, Greece. New York: IEEE Press, 2014: 279-284.

[3] Kozielski M, Doetsch P, Ney H et al. Improvements in RWTH's system for off-line handwriting recognition[C]//2013 12nd International Conference on Document Analysis and Recognition, August 25-28, 2013, Washington, DC, USA. New York: IEEE Press, 2013: 935-939.

[4] Bideault G, Mioulet L, Chatelain C, et al. Spotting handwritten words and REGEX using a two stage BLSTM-HMM architecture[J]. Proceedings of the IEEE, 2015, 9402: 94020G.

[5] Graves A, Schmidhuber J. Offline arabic handwriting recognition with multidimensional recurrent neural networks [J]. Advances in Neural Information Processing Systems, 2008: 545-552.

[6] Bluche T, Louradour J, Messina R, et al. Scan, attend and read: end-to-end handwritten paragraph recognition with MDLSTM attention[C]//2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), November 9-15, 2017, Kyoto, Japan. New York: IEEE Press, 2017: 1050-1055.

[7] Bluche T, Ney H, Kermorvant C, et al. Feature extraction with convolutional neural networks for handwritten word recognition[C]//2013 12nd International Conference on Document Analysis and Recognition, August 25-28, 2013, Washington, DC, USA. New York: IEEE Press, 2013: 285-289.

[8] Bianne-Bernard A L, Menasri F, Mohamad R A H, et al. Dynamic and contextual information in HMM modeling for handwritten word recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(10): 2066-2080.

[9] Voigtlaender P, Doetsch P, Ney H, et al. Handwriting recognition with large multidime nsional long short-term memory recurrent neural networks [C]//2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR), October 23-26, 2016, Shenzhen, China. New York: IEEE Press, 2016: 228-233.

[10] Bluche T, Ney H, Kermorvant C, et al. Tandem HMM with convolutional neural network for handwritten word recognition[C]//2013 IEEE International Conference on Acoustics, Speech and Signal Processing, May 26-31, 2013, Vancouver, BC, Canada. New York: IEEE Press, 2013: 2390-2394.

[11] Graves A, Liwicki M, Fernández S, et al. Anovel connectionist system for unconstra ined handwriting recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(5): 855-868.

[12] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//NIPS'17: Procee dings of the 31st International Conference on Neural Information Processing Systems, December 4-9, 2017, Long Beach, CA, USA. Canada: NIPS, 2017: 6000-6010.

[13] Zhu M K, Lu X L. Human action recognition algorithm based on Bi-LSTM-Attention model[J]. Laser & Optoelectronics Progress, 2019, 56(15): 151503.
朱铭康, 卢先领. 基于 Bi-LSTM-Attention 模型的人体行为识别算法[J]. 激光与光电子学进展, 2019, 56(15): 151503.

[14] Kang L, Toledo J I, Riba P, et al. Convolve, attend and spell: an attention-based sequence-to-sequence model for handwritten word recognition[M]//Brox T, Bruhn A, Fritz M, et al. Lecture notes in computer science. Cham: Springer, 2019, 11269: 459-472.

[15] Fang D B, Feng G, Cao H Y, et al. Handwritten formula symbol recognition based on multi-feature convolutional neural network[J]. Laser & Optoelectronics Progress, 2019, 56(7): 072001.
方定邦, 冯桂, 曹海燕, 等. 基于多特征卷积神经网络的手写公式符号识别[J]. 激光与光电子学进展, 2019, 56(7): 072001.

[16] Huang W R, He K, Liu K, et al. Handwritten Chinese character recognition based on attention mechanism[J]. Laser & Optoelectronics Progress, 2020, 57(8): 081002.
黄婉蓉, 何凯, 刘坤, 等. 基于注意力机制的手写体中文字符识别[J]. 激光与光电子学进展, 2020, 57(8): 081002.

[17] He W J, Liu J B, Pan M, et al. Natural scene text recognition algorithm based on Attention-CTC[J]. Electronic Science and Technology, 2019, 32(12): 32-36.
和文杰, 刘敬彪, 潘勉, 等. 基于 Attention-CTC 的自然场景文本识别算法[J]. 电子科技, 2019, 32(12): 32-36.

[18] Yang H J, Yan Z, Wu Z L, et al. Extraction method of interest text in image based on recurrent neural network [J]. Laser & Optoelectronics Progress,

2019, 56(24): 241501.
杨恒杰, 闫铮, 邬宗玲, 等. 基于循环神经网络的图像特定文本抽取方法[J]. 激光与光电子学进展, 2019, 56(24): 241501.

[19] Mor N, Wolf L. Confidence prediction for lexicon-free OCR[C]//2018 IEEE Winter Conference on Applications of Computer Vision (WACV), March 12-15, 2018, Lake Tahoe, NV, USA. New York: IEEE Press, 2018: 218-225.

[20] Krishnan P, Dutta K, Jawahar C V, et al. Word spotting and recognition using deep embedding[C]// 2018 13rd IAPR International Workshop on Document Analysis Systems (DAS), April 24-27, 2018, Vienna, Austria. New York: IEEE Press, 2018: 1-6.

[21] Stuner B, Chatelain C, Paquet T, et al. Handwriting recognition using cohort of LSTM and lexicon verification with extremely large lexicon[J]. Multimedia Tools and Applications, 2020, 79(45/46): 34407-34427.

[22] Wigington C, Stewart S, Davis B, et al. Data augmentation for recognition of hand written words and lines using a CNN-LSTM network[C]//2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), November 9-15, 2017, Kyoto, Japan. New York: IEEE Press, 2017: 639-645.