Computer Science and Application 计算机科学与应用, 2024, 14(5), 48-60
Published Online May 2024 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2024.145113

Hans汉斯

基于门控卷积和堆叠自注意力的离线手写汉字识别算法研究

罗序良，吴毅良*，刘翠媚，郭凤婵

广东电网有限责任公司江门供电局，广东 江门

收稿日期：2024年4月16日；录用日期：2024年5月14日；发布日期：2024年5月22日

摘　要

针对离线手写文本识别(HTR)在自然语言处理领域中的重要性以及其广泛应用于帮助视障用户、人机交互和自动录入等方面的实际需求，本研究提出了一个全新的模型。该模型在门控卷积网络的基础上引入了堆叠自注意力编码器－解码器，用于离线识别手写的汉字文本。由于书写风格的多样性、不同字符之间的视觉相似性、字符重叠以及原始文档中的噪音等挑战，设计准确且灵活的HTR系统具有相当大的难度，特别是当处理较为复杂、包含大量字符的文本时，算法的学习能力显得不足。为了解决这一问题，我们提出的模型包括特征提取层、编码器层和解码器层。其中，特征提取层从输入的手写图像中提取高纬度的不变特征图，而编码器和解码器层则相应地转录出文本。实验结果显示，该模型在HCTD数据集上的字符错误率(CER)为6.72，单词错误率(WER)为11.11；在HCWD数据集上的实验结果CER为6.22和WER为7.17。相对于其他研究者的模型，本文设计的模型在手写汉字识别率上提升了11%。

关键词

汉字识别，自注意力编码器－解码器，门控卷积，离线手写文本识别

Research on Offline Handwritten Chinese Character Recognition Algorithm Based on Gated Convolution and Stacked Self-Attention

Xuliang Luo, Yiliang Wu*, Cuimei Liu, Fengchan Guo

Jiangmen Power Supply Bureau, Guangdong Power Grid Co., Ltd., Jiangmen Guangdong

Received: Apr. 16th, 2024; accepted: May 14th, 2024; published: May 22nd, 2024

*通讯作者。

文章引用: 罗序良, 吴毅良, 刘翠媚, 郭凤婵. 基于门控卷积和堆叠自注意力的离线手写汉字识别算法研究[J]. 计算机科学与应用, 2024, 14(5): 48-60. DOI: 10.12677/csa.2024.145113

Abstract

In light of the significance of offline handwritten text recognition (HTR) in the field of natural language processing and its wide-ranging applications in meeting the practical needs of assisting visually impaired users, enabling human-computer interaction, and facilitating automated data entry, this study proposes a novel model. The model integrates the stacked self-attention encoder-decoder on the basis of gated convolution networks for recognizing offline handwritten Chinese characters. Given the challenges posed by diverse writing styles, visual similarities among different characters, character overlap, and noise in original documents, designing an accurate and flexible HTR system is notably difficult, especially when dealing with complex text containing a large number of characters, where algorithms often demonstrate limited learning capabilities. To address this issue, our proposed model comprises feature extraction, encoder, and decoder layers. The feature extraction layer extracts high-dimensional invariant feature maps from the input handwritten images, while the encoder and decoder layers transcribe the text accordingly. Experimental results demonstrate that the model achieves a character error rate (CER) of 6.72 and a word error rate (WER) of 11.11 on the HCTD dataset; and on the HCWD dataset, the CER is 6.22 and the WER is 7.17. Compared to models developed by other researchers, our designed model shows an 11% improvement in handwritten Chinese character recognition accuracy.

Keywords

Chinese Character Recognition, Self-Attention Encoder-Decoder, Gated Convolution, Offline Handwriting Text Recognition

1. 引言

手写文字识别(HTR)一直以来都是图像处理和自然语言处理领域中最具吸引力和挑战性的研究领域之一。它被广泛应用在各种应用程序中，用于将手写图像转换成可编辑文本的用户界面，提高 HTR 系统的识别性能可以改进文字处理领域的自动化流程。

根据参考文献[1]，手写文字识别主要分为离线识别和在线识别两种类型。在离线识别中，手写特征从扫描图像中提取，而在在线识别中，特征则同时从笔迹轨迹和生成的图像中提取。由于不同书写者书写风格的差异、不同字符的视觉相似性、字符之间的重叠以及手写字符的复杂特征，在从手写文档中提取特征方面存在一定困难。此外，原始文档的背景复杂也是离线手写识别中的另一个挑战。这表明，从在线手写识别的笔迹轨迹输入中提取特征要比从扫描图像进行离线识别要好得多。因此，离线手写识别需要更复杂的方法来准确提取特征并提高识别性能。

在过去几十年中，国内外学者提出了多种 HTR 系统，并取得了显著的成果。例如，已经采用了隐马尔可夫模型(HMM) [2]和 HMM-神经网络混合模型[3]来识别手写文档。然而，由于 HMM 的独立假设，匹配提取的特征与标签存在一定的局限性，即使在 HMM-NN 混合系统的情况下稍微放宽，仍然存在大范围输入问题。因此，研究者最近提出了深度神经网络(DNN)方法，用于改进浅层机器学习技术中的分割、特征提取、分类和识别问题[4]。深度学习方法已经被应用于离线 HTR 的研究，并在拉丁文、阿拉伯

文和梵文书写中展现出显著的结果，甚至能够实现多语言识别[5]。在文献[6]中，提出了使用带有双向门控循环单元(GRU)的门控卷积神经网络(Gated CNN)进行离线手写文本识别的方法，并取得了良好的效果。

在手写汉字识别技术中，与其他文字(如英文字母)相比，汉字的识别具有多个挑战。汉字的复杂笔画结构、字形变化与连写、字形相似性、多音字和多义字以及书写风格的多样性，增加了识别的复杂度。为了提高对汉字的准确识别能力，手写字体识别技术需要使用笔迹分析、字形特征提取和上下文语境分析等多种技术手段。

本研究提出了一种新型的门控卷积和堆叠自注意力编码器–解码器网络(GCSEN)，旨在识别离线手写的汉字文本。在本模型中，我们采用门控卷积神经网络(Gated CNN)从手写图像中提取特征，并利用称为 Transformer 的堆叠自注意力编码器–解码器来转录文本。我们将这些模型整合在一起，因为门控卷积神经网络在从复杂数据集中提取特征方面表现出较高性能，并且堆叠自注意力编码器–解码器通过避免递归在语言建模方面表现优异[7]。本文的模型包括特征提取层、编码器和解码器层。此外，我们对其他三种最近提出的基于循环结构的网络模型进行了性能评估。这些模型分别为具有 LSTM 的卷积神经网络(CNN) [8]、具有长短期记忆(LSTM)的门控 CNN [9]和具有门控循环单元(GRU)的门控 CNN [6]。为了模拟所选择和提出的模型，我们从网络上收集了一组离线手写文本数据集。通过预处理所收集的数据集，我们准备了手写汉字文本行数据集(HCTD)和手写汉字单字数据集(HCWD)，并对我们的系统进行了评估。

2. 相关工作

2.1. 汉字特征分析

汉字作为中国传统的书写系统具有悠久的历史，可以追溯到数千年前的甲骨文和金文。汉字经历了演变和发展，形成了多种书法风格和字体变体，每个汉字都有独特的组成结构和书写风格，深受中国文化、历史和艺术的影响。对汉字的特征分析涉及笔画的组成、书写顺序、相互关系以及字形的变体。汉字的书写具有多样性和复杂性，而不同的书写风格如楷书、行书、草书等则展现了汉字的丰富多彩。图 1 为手写汉字的示例。

Figure 1. Example of handwritten Chinese characters
图 1. 手写汉字示例

此外，汉字的形体结构复杂，通常由多个笔画组成，包括多种连笔和隶变的书写形式。部分汉字因结构复杂、形态相似或多音多义而识别难度有所增加。因此，手写识别汉字需要考虑字符结构、笔画路径、笔画顺序以及连笔特征等因素的综合分析。有效的汉字识别算法需要能够准确捕捉这些特征，并结合上下文语境，提高识别的准确率和鲁棒性。

2.2. 手写文本识别技术概述

构建手写文本识别(HTR)系统的主要技术研究包括分割(字符、单词或文本行级别)、特征提取和分类任务。采用分割技术对输入图像的字符、单词或文本行进行检测并分割。由于在手写汉字中会存在一些连字，因此在字符层级的分割难度有所增加。因此，我们进行了单词层级和文本行层级的分割。在增强图像质量方面，采用的预处理技术如中值滤波[10]和高斯平滑[11]有效提升分割算法的性能。

此外，特征提取和分类子任务涉及多种机器学习技术，如 HMM、支持向量机(SVM)和神经网络[12]。最近，基于深度神经网络的模型已被广泛应用，并在高维自动特征图提取和手写文本识别方面展现出了有前景的结果[13]。采用端到端方式来执行特征提取和识别任务。与传统的机器学习方法不同，基于深度神经网络的方法使用了最少的预处理[14]。已有一些关于离线手写文字识别系统 HTR 的研究工作。大多数采用传统机器学习方法，但识别结果仍需改进以用于实际应用。以下是一些先前的研究成果：Assabie [15]提出了一个不限定作者的手写识别系统，利用原始笔画的特征和特殊关系，通过原始笔画的特殊关系来衡量提出模型的准确性。利用三个不同来源的数据集，分别达到了 87%、76%和 81%的识别结果。在文献[16]中，使用 HMM 构建了一个独立于作者的阿姆哈拉语离线手写单词识别系统，利用方向场张量来检测文本行并从文本行中提取特征。对于每个字符，原始的结构特征被存储为模型的训练和测试的特征列表。文献[17]提出了基于 inceptions 结构神经网络的离线手写汉字识别方法，具有结构简单，易于进行网络深度扩展，训练参数少的优点。并且该方法采用随机梯度下降优化算法来提高识别的准确率。在文献[18]中，采用卷积神经网络中的 LeNet-5 网络模型对手写汉字进行识别。该方法能高效、稳定地从有噪声图像中识别出文字，同时经高斯滤波与 PCA 滤波处理后的图像识别精确度更高。

与英语相比，为汉字文字系统设计健壮的离线手写文本识别(HTR)系统面临一些挑战，如字符数量和视觉上相似字符的问题。汉字系统包括 3700 多个常用汉字，并且其笔画结构相较于英语更为复杂。因此，字符数量的增加会对 HTR 系统产生影响，需要更多的内存和计算资源。此外，汉字中存在一些视觉上相似的字符，对计算机而言非常难以识别。

本文提出了使用门控卷积神经网络从手写文本行/单词图像中提取特征，并利用 Transformer 网络来转录相应的文本，从特征提取层提取的特征图。为了训练和测试所提出的模型，我们准备了一个专门的离线汉字手写文本行数据集和手写汉字单词数据集。

3. 模型设计

本研究引入了一种新的门控卷积神经网络架构，结合了堆叠的自注意力编码器–解码器模型，用于识别离线手写的汉字文本。此外，我们对当前各种文字系统的最新模型进行了广泛调研，包括 Puigcerver 提出的 CNN-1D-LSTM [8]，Bluche 和 Messina 提出的门控 CNN-1D-LSTM 模型用于英语[9]，以及适用于阿拉伯语等文字系统的 HTR Flor++ (门控 CNN-1D-GRU) [6]。

本文提出的模型由三个主要组件组成：特征提取层、编码器层和解码器层，其架构如图 2 所示。编码器–解码器层加入了堆叠的多头自注意力机制，然后对接上基于位置编码的全连接网络，即 Transformer。这一架构近年来在自然语言处理任务中备受关注，并不断取得新的成果。

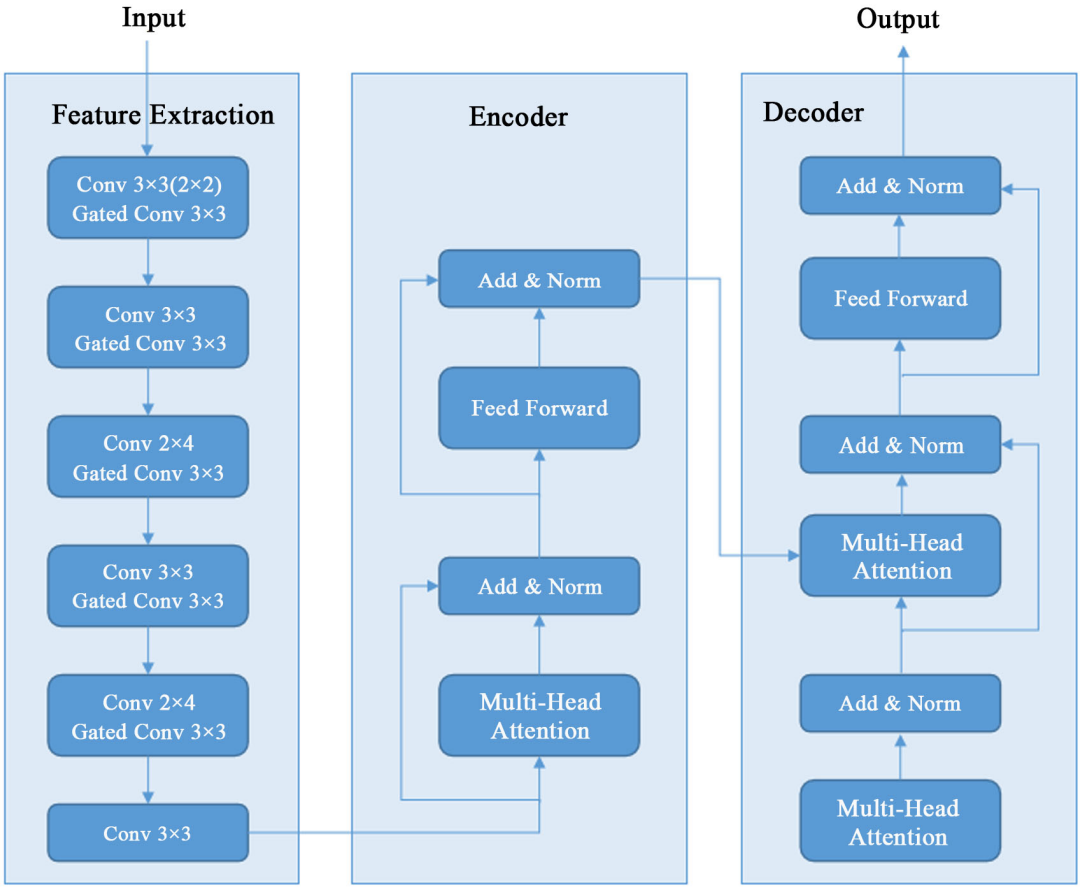


Figure 2. The improved model architecture proposed in this paper
图 2. 本文改进的模型架构

在手写文本识别中，与自然语言处理中使用 Transformer 的关键区别在于，在将图像输入 Transformer 网络之前，需要将图像转换为序列表示。为了解决这一问题，我们引入了一个特征提取单元，位于编码器层之前，负责从输入图像中提取特征图的序列。这一预处理步骤确保了图像中的视觉信息能够被 Transformer 网络适当地结构化，以便进行后续处理。虽然我们提出的模型包括三个基本层，即特征提取、编码器和解码器层，但需要强调的是，该模型是以端到端的方式进行训练的。这些层中的每一层在识别过程中都发挥着独特的作用，下面我们将简要概述每一层。

3.1. 特征提取层

如图 2 所示，特征提取层是我们模型的基础组件。值得注意的是，自然语言处理和手写文本识别的输入数据性质有着显著的区别。对于手写文本识别而言，输入为手写文本图像，需要经过预处理转换成空间特征图，然后才能传递到模型的后续层。

为了将输入的手写图像转换成空间特征图，我们使用了门控卷积神经网络架构，具体细节请参考文献[9]。网络架构中含有 11 个卷积神经网络层和 5 个门控卷积层。在每个卷积层中，我们使用 PReLU 激活函数和批次重归一化的处理方法。此外，在最后的三个门控卷积层中，我们采用了概率为 0.2 的 dropout 技术，防止模型过拟合，增强了模型的泛化能力。

通过门控和卷积层，我们提取了不同维度的特征图，包括 16、32、40、48、56 和 64。此外，我们

还选择了不同的核大小：第三和第五层采用了 2 x 2 的卷积核，在其余的卷积层上使用了 3 x 3 的卷积核。这些特征图尺寸和核大小的选择是经过精心设计的，以捕捉和表示手写文本的不同空间特征，使得模型的后续层能够有效地处理这些信息。

3.2. 编码器层

在特征提取完成之后，提取的特征图传递到编码器层，由编码器层进行下一步的处理。编码器层由六个类似的堆叠层组成，每个堆叠层旨在将输入的特征图转换为更高级的表示，以捕捉与后续处理相关的信息。在每个编码器层内，存在两个主要单元：自注意力单元和前馈神经网络单元，它们共同逐步完善特征表示。

自注意力单元是编码器的关键组成部分，它接受先前编码器层生成的一组输入编码，并评估每个编码相对于其他编码的重要性。这一过程根据它们相互之间的相关性分配权重给输入编码，从而生成了一组输出编码，携带了对特征图中空间关系和上下文依赖的精细理解。随后，每个编码的输出被分别传递到前馈神经网络单元，该单元对特征编码应用非线性变换，增强了它们的表征能力，并进一步完善了其中包含的信息。

这些单元的输出随后被传递给下一个编码器单元，或者在最后一个编码器单元的情况下，传递到解码器层。信息在单元之间流动的这种结构，使得模型能够逐步构建手写文本的抽象和上下文化的表示，这对于准确的识别和理解至关重要。编码器分析空间和上下文信息的能力使之成为模型识别手写文本的基本组成部分。

3.3. 解码器层

解码器层作为模型的最后一个部分，其结构与编码器层类似，共由六个堆叠的模块组成。这些模块与编码器层的模块有共同的特征，但额外增加了一个子层，用于在编码器堆栈的输出上运行多头注意力。这一新增部分对解码器的功能至关重要，因为它使得解码器在生成最终输出序列时能考虑编码器输出中编码的上下文和关系。

除了本文提出的模型外，另外我们还评估了其他三个最近开发的模型的性能。第一个模型，如文献[8]所述，模型包含了卷积神经网络(CNN)、双向长短时记忆(BLSTM)和连接主义时间分类(CTC)。该模型被称为 CNN-1D-LSTM，参数约为 960 万，包括五个卷积层和五个 BLSTM 层。每个卷积层包含 3x3 的卷积核、批次归一化处理、ReLU 激活函数和 2 x 2 的最大池化。此外，为了减少过拟合，还应用了概率为 0.5 的 dropout，并采用 RMSProp 算法[19]进行参数更新，根据每批次的 16 幅图像的 CTC 损失梯度进行逐步调整。

第二个模型采用了不同的方法，利用卷积编码器处理输入图像，并使用双向 LSTM 解码器预测字符序列。由 Bluche 和 Messina [9]提出的门控卷积循环神经网络(Gated CNN-BLSTM)，设计更为紧凑，参数约为 73 万。Gated CNN-BLSTM 包括八个卷积层，其中包括三个门控卷积层和两个 BLSTM 层。该模型使用了 tanh 激活函数和 RMSProp 优化器，然后通过 CTC 损失的梯度逐步调整参数。

第三个模型，名为 FLOR++ [6]，包含了由 11 个卷积层组成的架构，若是包括 6 个门控卷积层和 2 个双向门控循环单元(BGRU)层。类似前面讨论的模型，FLOR++在门控和传统卷积块中提取不同维度(16、32、40、48、56 和 64)的特征图。在第三和第五个卷积层中，使用 2 x 4 的卷积核，而在其他卷积层中则使用 3x3 的卷积核。该模型在所有卷积层中应用了 He 均匀初始化器、PReLU 激活函数和批次重归一化，并对最后三个门控卷积层采用了 0.2 的 dropout 以解决模型过拟合问题。FLOR++的视觉模型采用两个双向 GRU 层、0.5 的 dropout，并用密集层交替，并且使用 CTC 来计算损失和转录预期的字符。这些多样

化的模型为我们在各种语言的手写文本识别方法的全面探索做出了重要贡献。

4. 实验与结果分析

本节主要展示并分析本文提出的模型以及第 3.3 节中介绍的另外三个最新技术模型的实验成果。这种严格的评估对于衡量模型在手写文本识别任务中的有效性至关重要。并将详细讨论实验使用的数据集制作和具体的实验设置，以及从每个模型中得出的结果。这些结果不仅突出了模型的性能，还为了解其优势和潜在改进领域提供了有价值的见解，最终促进了手写文本识别技术的进步。

4.1. 数据准备

虽然像英语、阿拉伯语和瑞典语这样的语言已经有大量公开获取的数据集，大大促进了各自领域的研究，但对于汉语手写文本识别来说，缺乏这样的资源给我们的研究带来了独特的挑战。因此，我们精心进行了数据集的创建和准备工作，以促进我们的实验。本节概述了我们研究基础的数据集收集方法所涉及的过程。

首先，我们收集了日常工作生活中使用的手写数据，包括信件、公告、通知等。收集到所需要的资料后，我们通过扫描设备将其转换为图像数据。接着，基于互联网上存在大量数据的事实，我们利用自动化的爬虫技术从网络上抓取手写字体的图片数据，其中包括书法爱好者提交到网上的数据以及一些学生发布的作业内容。当然，我们也意识到部分收集到的图片数据不符合实际需求，例如背景过于复杂、字体模糊等，因此需要经过人工清理。总体来看，本文共收到了超过 38,000 份手写汉字图片数据，其中包含超过 50 万个汉字，手写汉字数据集示例如图 3 所示。

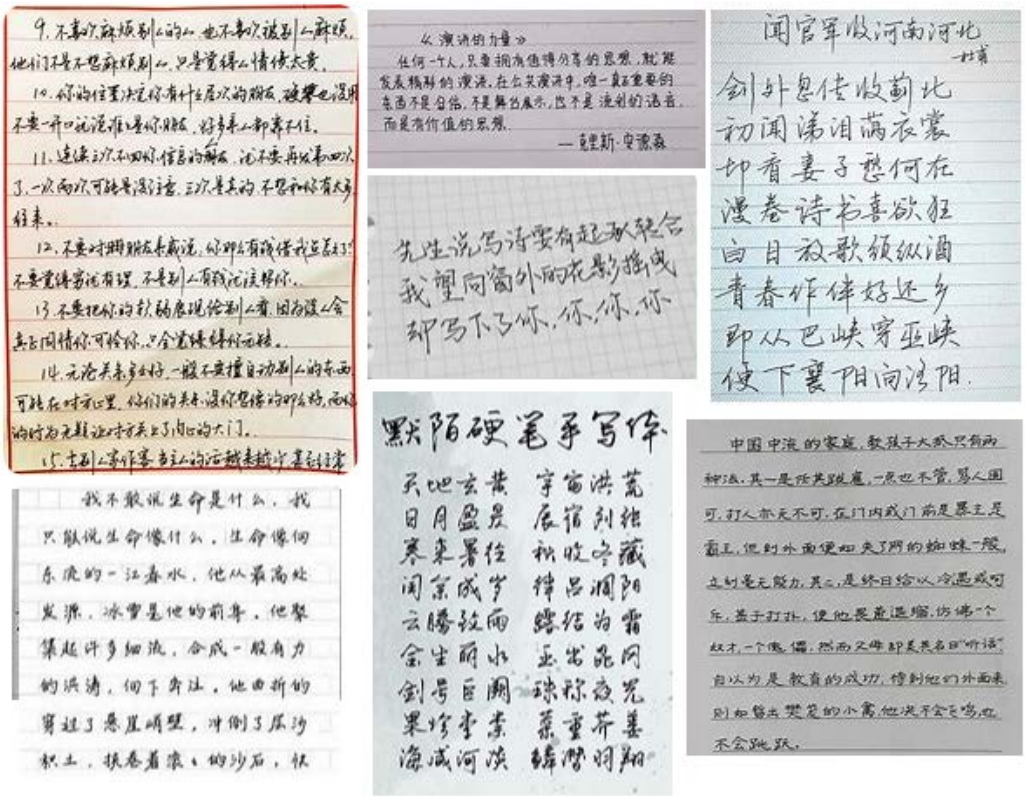
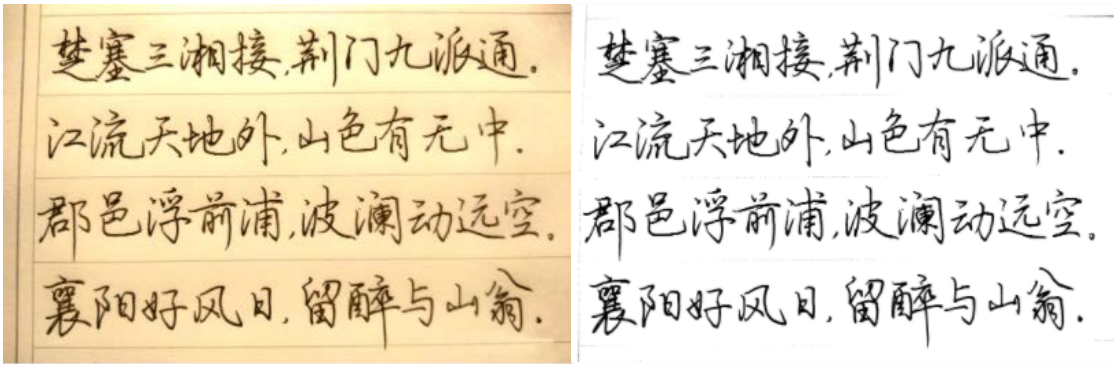


Figure 3. Collection of handwritten Chinese character data
图 3. 手写汉字数据收集

在完成数据整理后，我们使用 VIA 标注工具[20]对图像进行了详细标注，主要以段落级别进行标记。随后，我们从图像中裁剪出手写部分，并隔离核心内容，以便进行进一步的分析和处理。为进一步提高质量，裁剪后的图像经历了一系列的预处理步骤，以优化其用于后续识别任务。这些预处理包括将输入的裁剪图像转换为二进制表示，并可选择将其规范化为灰度图像。此外，为了纠正任何图像倾斜，我们使用了纠偏算法。在这一预处理阶段，OCRopus 工具箱[21]发挥了关键作用，它提供了一整套工具，用来增强输入的手写裁剪图像。图 4 展示了裁剪后的图像以及其对进行二值化处理后输出图像。

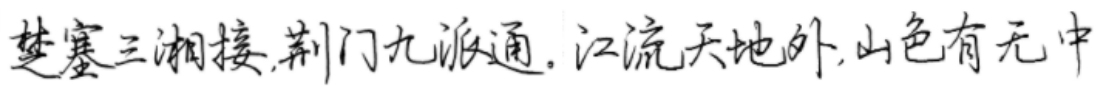


(a) 原始图像　　(b) 二值化图像

Figure 4. Cropped images and binarized images
图 4. 裁剪后的图像和二值化图像

在数据裁剪完成后，我们将注意力转向了单词和文本行分割的重要任务。这些任务旨在从预处理图像中提取有意义的文本行。为了完成此任务，我们使用了 OCRopus 工具箱中的分割模块进行处理。该工具箱提供了一系列基于 Python 的文档分析和识别工具，简化了分割过程。

在数据预处理完成后，创建用于训练的基础真实数据是一个重大挑战。为了解决这一问题，我们利用 OCRopus 工具箱的核心组件 OCRopus-gtedit HTML 和 extract 命令为每个分割的文本行图像准备了基础真实文本数据。图 5 展示了一个样本分割的文本行图像以及相应的基础真实数据。



(a) 原始文本行图像

楚塞三湘接 荆门九派通 江流天地外 山色有无中

(b) 真实数据

Figure 5. The original text line and actual data
图 5. 原始的文本行与真实数据

在这个分割文本行的情况下，如果出现了多行被连接成单个单元的情况，会导致分割错误，这样的情况会被严格丢弃，确保了数据的完整性。因此，在 3000 个分割的文本行中，HCTD 数据集包括了 2900 个文本行图像，而由于没有基础真实数据，有 100 个文本行图像被排除在外。

为了进一步丰富 HCWD 数据集，我们对选定的分割文本行图像进行了二次分割，采用了基于轮廓的算法。该算法不仅完成了分割任务，还标记了文本行，利用了之前 HCTD 中提取的分割文本行图像和它们的基础真实数据。图 6 展示了一个分割文本行中检测到的单词图像的示例，此外，算法的伪代码见算

法 1。这一严格的数据准备过程确保了我们研究汉字手写文本识别所需的高质量数据集的可用性。

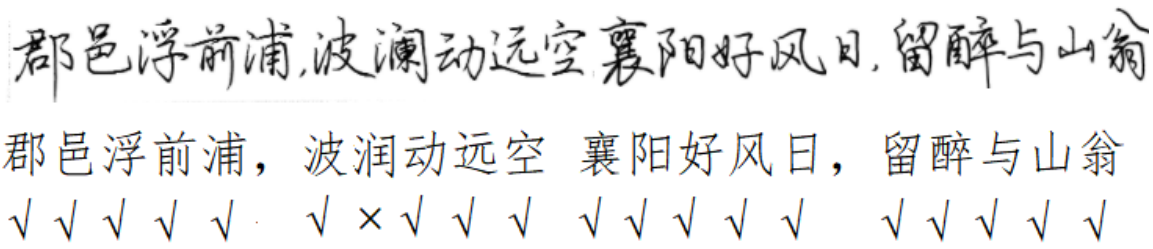


Figure 6. Detected Chinese characters from the input text line image
图 6. 从文本行输入图像中检测到的汉字

算法 1：字检测和半自动标签.

1. I←读取文本行图像
2. gt←读取真实数据(GT)
3. I←重置 I 的高度//本文中高度设置为 50
4. Kernel←创建各向异性过滤器核//本文中设置 kernels size = 25，sigma = 11，theta = 7
5. If←在 I 上进行 2D 过滤//使用已创建的过滤器核和 OpenCV 中的 filter2D 方法
6. Ithr←在 If 上进行阈值处理
7. components←检测 Ithr 中的连接组件//使用 OpenCV 中的 findContours 方法
8. words←gt.split(‘‘) //使用空格对数据进行分割
9. For (i, c) in enumerate (components):
   a. If contour_area(c) > minArea//本文中 min area = 100
      i. box ← boundingRect (c)
      ii. (x, y, w, h) ← box
      iii. Iw ← I [y:y + h, x:x + w]
      iv. 保存图像 Iw 与它对应的标签 words [i]

半自动标记方法的有效性高度依赖于检测程序的准确性。正如图 6 所示，最初的 6 个连续单词被精确检测和标记，显示了该方法简化标记过程的潜力。然而，如果两个连续的单词被错误地检测为一个单词，就会导致后续的标记错误。为了解决这个问题，还需要对数据重新进行手动标记。

为了确保数据集的整体质量和可靠性，我们进行了细致的手动重新标记工作，涉及纠正最初由于检测错误而错误标记的单词标签。由于这一勤勉的策划过程，我们制作了一个名为 HCWD 的全面数据集，共包括 10,550 个单词。另外，为了增强我们模型的鲁棒性和通用性，我们通过合成的方法制作了合成文本行数据共 290,000 个，合成的单词数据共 500,000 个。这些合成数据库对于微调模型、调整权重和优化超参数以增强识别性能起到了重要作用。表 1 中列出了数据集的具体划分结果。

Table 1. Dataset statistics
表 1. 数据集统计

数据集	总数	训练	测试	验证
HCTD	2900	2340	560	-
HCWD	10,550	8532	2108	-
合成的文本行	290,000	174,000	58,000	58,000
合成的单词	500,000	300,000	100,000	100,000

4.2. 实验设置

本文的训练设备采用一台高性能的图形工作站，其硬件配置包括英特尔 Core i9-13900K (3.60 GHz) CPU、128 GB RAM 和 GeForce RTX 4090 24G GPU，操作系统为 Ubuntu。集成开发环境采用了 Python 3.6、Keras 库与 TensorFlow 框架进行搭建。为了确保模型的鲁棒性，我们对所有提出的网络进行了预训练，使用了合成生成的汉字书写文本行图像。为了保持一致性和优化，我们采用了 RmsProp 优化器，并将批处理大小设置为 8，以防止过拟合并提高效率。我们还配置了早期停止机制，在验证损失值在 20 个轮次内没有显著改善时触发。

本文提出的模型识别性能的评估基于两个关键指标：字符错误率(CER)和单词错误率(WER)。这些指标作为基本的衡量标准，用于评估手写识别系统的准确性和可靠性。CER 通过计算 Levenshtein 距离来量化字符识别的准确性，该距离测量对齐识别文本与基础真实文本所需的字符级操作(替换、插入和删除)的累积数量。较低的 CER 值表示字符级别识别的准确性更高，表明识别字符与真实字符之间的匹配更接近。类似地，WER 通过量化识别文本中词语转录的准确性来衡量识别文本和实际内容之间的差异。它计算了对齐识别单词与基础真实文本所需的单词级操作(替换、插入和删除)的累积数量。较低的 WER 值反映了手写词语转录的准确性更高，表明识别单词与真实内容之间的对应更为紧密。

在手写识别的背景下，实现较低的 CER 和 WER 值表明系统在解码手写文本方面更准确和可靠。这些指标在评估系统性能时至关重要，它提供了有关其准确转录手写字符和单词的能力以及识别模型的整体有效性的依据。

4.3. 实验结果

本文的第一个实验旨在展示所提出的模型在使用文本行数据集时的识别性能。实验结果见表 2，其中列出了每个模型在手写文本识别环境中的表现数据。

Table 2. Experimental results on the HCTD dataset
表 2. 在 HCTD 数据集上的实验结果

网络模型	WER (%)	CER (%)
CNN-LSTM	13.5	8.15
GNN-LSTM	12.8	7.12
GNN-GRU	12.51	6.92
GCSEN	11.11	6.72

根据表 2 中的结果，明显可见本研究提出的模型 GCSEN 在单词错误率(WER) 11.11%和字符错误率(CER) 6.72%的情况下表现出色，突显了该模型在手写汉字识别方面的优势。相比之下，虽然 GNN-GRU 模型的表现不俗，但其 WER 为 12.80%，CER 为 7.12%。值得注意的是，GNN-GRU 模型优于 CNN-LSTM 模型，凸显了前者在汉字识别方面的优势。

为了进一步研究合成生成数据集对我们的识别模型性能的影响，我们进行了一项从头开始使用手写数据集进行训练的实验。实验结果显示，所有模型的性能都显著提升，因为所提出的模型的 WER 和 CER 值均减少了约 11%。这些结果突显了使用合成数据集对模型进行预训练后对整体识别性能的提升起到了积极的作用。合成数据的利用不仅增强了模型的适应性，还有助于减少识别错误，进一步强调了这种方法在手写文本识别领域的有效性。

本研究的第二个实验着重于评估所提出的模型在单词识别方面的性能，采用了 HCWD 数据集。实验结果详见表 3，展示了模型在识别该特定数据集中的手写文本方面的能力。

Table 3. Experimental results on the HCWD dataset
表 3. 在 HCWD 数据集上的实验结果

网络模型	WER(%)	CER(%)
CNN-LSTM	11.75	7.55
GNN-LSTM	9.24	6.46
GNN-GRU	9.08	6.41
GCSEN	7.17	6.22

通过表 3 中的结果，明显可见本文提出的 GCSEN 模型在汉语手写文本识别领域持续展现出优越性。该模型超越了先前提出的其他模型，在识别准确性和效率方面提供了更高水平。这一结果强调了所提出的模型在应用于 HCWD 时的稳健性和多功能性，重新确立了其处理各种风格和形式的手写文本的能力，进一步巩固了其作为汉语手写文本识别领先解决方案的地位。

第二个实验的积极结果进一步强化了所提出的模型在增强手写文本识别领域的承诺，特别是在复杂手写汉字的背景下，并强调了其在从手写文件中准确识别文本的各种应用中的潜力。我们实验的结果突显了基于 Transformer 网络相对于先前最先进的模型的显著优势。除了改进的识别性能外，基于 Transformer 的模型还具有更好的参数效率。这意味着它在需要更少计算资源的情况下取得了显著的结果，使其成为更高效和可扩展的解决方案。

然而，需要承认，汉字与许多其他文字一样，存在独特的挑战。这些文字包含具有相似语音但不同视觉形状的字符，真实值与输入图像之间的差异构成了识别性能的挑战。模型识别视觉上相似字符的能力是一个不断改进的过程。在文本行和基于词的实验中，观察到所提出的网络在处理共享相似形状的字符或处理样本训练数量有限的字符时存在局限。此外，连字的存在，即具有特定形状和含义的字符组合，也显著影响了模型的识别性能。

为了更好地解决这些限制并增强模型的能力，需要扩展更多样化的训练数据集。该数据集应涵盖更广泛范围的每个字符样本，特别关注因视觉相似性或训练数据稀缺性而带来挑战的字符。持续完善和扩展数据集有助于提高字符区分和识别准确性，进而推动模型在识别手写汉字方面的性能。

5. 结论

近年来，随着机器学习技术的进步和大规模数据集的可用性，离线手写文本识别(HTR)取得了显著成就，促进了高效识别模型的设计。然而，对于以汉字为基础的语言，如汉语，HTR 领域仍相对未被充分探索，需要显著改进。本文通过精心准备了两个关键数据集：一个包含 10550 个词的 HCWD 和一个包含 2900 个文本行的 HCTD，这些数据集经过精心收集。利用这些数据集，我们进行了一系列实验，以识别单个词和完整的文本行。我们的方法利用了门控卷积作为特征提取层，随后采用强大的 Transformer 网络将提取的特征转录成文本。此外，我们对最近提出的模型进行了彻底分析，包括 CNN-LSTM、GNN-LSTM 和 GNN-GRU 网络。实验结果表明，我们的模型在手写汉字和文本行识别方面具有优势。凭借经过精心准备的手写测试数据集，所提出的 GCSEN 模型在 HCWD 上取得了显著成效，其字符错误率(CER)为 6.22%，单词错误率(WER)为 7.17%；在 HCTD 数据集上的结果 CER 为 6.72%，WER 为 11.11%。相对于 GNN-GRU 模型，本文的模型的识别性能提升约 11%。展望未来，我们计划通过整合语言建模技术和扩展我们的数据集，进一步提升手写词和文本行的识别能力。

基金项目

本文由“南网高层次人才特殊支持计划”项目资助。

参考文献

[1] Liu, C.-L., Yin, F., Wang, D.-H. and Wang, Q.-F. (2013) Online and Offline Handwritten Chinese Character Recognition: Benchmarking on New Databases. *Pattern Recognition*, **46**, 155-162. https://doi.org/10.1016/j.patcog.2012.06.021

[2] Natarajan, P., Saleem, S., Prasad, R., MacRostie, E. and Subramanian, K. (2008) Multi-Lingual Offline Handwriting Recognition Using Hidden Markov Models: A Script-Independent Approach. In: Doermann, D. and Jaeger, S., Eds., *Arabic and Chinese Handwriting Recognition*, Springer, Berlin, 231-250. https://doi.org/10.1007/978-3-540-78199-8_14

[3] España-Boquera, S., Castro-Bleda, M.J., Gorbe-Moya, J. and Zamora-Martinez, F. (2011) Improving Offline Handwritten Text Recognition with Hybrid HMM/ANN Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **33**, 767-779. https://doi.org/10.1109/TPAMI.2010.141

[4] Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012) ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, **25**, 1-9.

[5] Zhao, Y., Zhang, X., Fu, B., Zhan, Z., Sun, H., Li, L. and Zhang, G. (2022) Evaluation and Recognition of Handwritten Chinese Characters Based on Similarities. *Applied Sciences*, **12**, Article No. 8521. https://doi.org/10.3390/app12178521

[6] Flor, A., Neto, D.S., Leite, B., Bezerra, D. and Toselli, A.H. (2020) HTR-Flor++: A Handwritten Text Recognition System Based on a Pipeline of Optical and Language Models. Association for Computing Machinery, New York.

[7] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I. (2017) Transformer: Attention Is All You Need. *Proceedings of the Advances in Neural Information Processing Systems* 30, Long Beach, 4-9 December 2017, 5998-6008.

[8] Puigcerver, J. (2017) Are Multidimensional Recurrent Layers Really Necessary for Handwritten Text Recognition? *Proceedings of the International Conference on Document Analysis and Recognition*, *ICDAR*, Kyoto, 2 July 2017, Volume 1, 67-72. https://doi.org/10.1109/ICDAR.2017.20

[9] Bluche, T. and Messina, R. (2017) Gated Convolutional Recurrent Neural Networks for Multilingual Handwriting Recognition. *Proceedings of the International Conference on Document Analysis and Recognition*, *ICDAR*, Kyoto, 2 July 2017, Volume 1, 646-651. https://doi.org/10.1109/ICDAR.2017.111

[10] Huang, T.S., Yang, G.J. and Tang, G.Y. (1979) A Fast Two-Dimensional Median Filtering Algorithm. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **27**, 13-18. https://doi.org/10.1109/TASSP.1979.1163188

[11] Praveen, K.S., Babu, K.P. and Sreenivasulu, M. (2016) Implementation of Image Sharpening and Smoothing Using Filters. *International Journal of Scientific Engineering and Applied Science*, **2**, 7-14.

[12] Xu, S., Wu, Q. and Zhang, S. (2020) Application of Neural Network in Handwriting Recognition. *IEEE Transactions on International Conference of Stanford University*, Stanford, 20-22 December 2020, 1-3.

[13] Bluche, T., Louradour, J. and Messina, R. (2017) Scan, Attend and Read: End-to-End Handwritten Paragraph Recognition with MDLSTM Attention. *Proceedings of the International Conference on Document Analysis and Recognition*, *ICDAR*, Kyoto, 9-15 November 2017, Volume 1, 1050-1055. https://doi.org/10.1109/ICDAR.2017.174

[14] Soomro, M., Farooq, M.A. and Raza, R.H. (2017) Performance Evaluation of Advanced Deep Learning Architectures for Offline Handwritten Character Recognition. *Proceedings of the* 2017 *International Conference on Frontiers of Information Technology*, Islamabad, 18-20 December 2017, 362-367. https://doi.org/10.1109/FIT.2017.00071

[15] Assabie, Y. and Bigun, J. (2008) Writer-Independent Offline Recognition of Handwritten Ethiopic Characters. *Proceedings of the* 11*th International Conference on Frontiers in Handwriting Recognition* (*ICFHR*), Montréal, 19-21 August 2008, 652-657.

[16] Assabie, Y. and Bigun, J. (2009) HMM-Based Handwritten Amharic Word Recognition with Feature Concatenation. *Proceedings of the* 2009 10*th International Conference on Document Analysis and Recognition*, Barcelona, 26-29 July 2009, 961-965. https://doi.org/10.1109/ICDAR.2009.50

[17] 陈站, 邱卫根, 张立臣. 基于改进 inception 的脱机手写汉字识别[J]. 计算机应用研究, 2020, 37(4): 1244-1246. https://doi.org/10.19734/j.issn.1001-3695.2018.09.0784

[18] 张静娴, 冷青轩, 陈航, 等. 基于图像滤波预处理的卷积神经网络汉字识别[J]. 电工技术, 2023(24): 69-73. https://doi.org/10.19768/j.cnki.dgjs.2023.24.021

[19] Tieleman, T. and Hinton, G. (2012) Lecture 6.5-rmsprop: Divide the Gradient by a Running Average of Its Recent Magnitude. *COURSERA*: *Neural Networks for Machine Learning*, **4**, 26-31.

[20] Dutta, A. and Zisserman, A. (2021) The VIA Annotation Software for Images, Audio and Video. *Proceedings of the* 27*th ACM International Conference on Multimedia*, Nice, 1 January 2021, 2276-2279.

https://doi.org/10.1145/3343031.3350535

[21] Breuel, T.M. (2008) The OCRopus Open Source OCR System. *Proceedings of the Document Recognition and Retrieval XV, SPIE,* San Jose, 27 January 2008, 120-134.