

基于汉字拆分嵌入和二部图的残损碑文识别

蔺广逢^{*①} 吴娜^① 贺梦兰^① 张二虎^① 孙强^②

^①(西安理工大学印刷包装与数字媒体学院 西安 710048)

^②(西安理工大学自动化与信息工程学院 西安 710048)

摘要: 古籍碑刻承载着丰富的历史文化信息,但是由于自然风化侵蚀和人为破坏使得碑石上的文字信息残缺不全。古碑文语义信息多样化且样例不足,使得学习行文语义补全识别残损文字变得十分困难。该文试图从字形空间语义建模解决补全残损汉字进行识别理解这一挑战性任务。该文在层级拆分嵌入(HDE)编码方法的基础上使用动态图修补嵌入(DynamicGrape),对待识别汉字的图像进行特征映射并判别是否残损。如未残损直接转化为层级拆分编码,输入二部图推理字节到部件节点的边权重,比对字库编码识别理解;如残损需要在字库里检索可能字和部件,对汉字编码的特征维度进行选择,输入二部图推理预测可能的汉字结果。在自建的数据集以及中文自然文本(CTW)数据集中进行验证,结果表明二部图网络可以有效迁移和推理出残损文字字形信息,该文方法可以有效对残损汉字进行识别理解,为残损结构信息处理开拓出了新的思路和途径。

关键词: 残损碑文; 碑文预测; 碑文识别; 残损文字识别; 二部图神经网络

中图分类号: TN911.73; TP18

文献标识码: A

文章编号: 1009-5896(2024)02-0564-10

DOI: 10.11999/JEIT230893

Damaged Inscription Recognition Based on Hierarchical Decomposition Embedding and Bipartite Graph

LIN Guangfeng^① WU Na^① HE Menglan^① ZHANG Erhu^① SUN Qiang^②

^①(Faculty of Printing, Packaging and Digital Media, Xi'an University of Technology, Xi'an 710048, China)

^②(Faculty of Automation and Information Engineering, Xi'an University of Technology, Xi'an 710048, China)

Abstract: Ancient inscriptions carry rich historical and cultural information. However, due to natural weathering and man-made destruction, the text information on the inscriptions is incomplete. The semantic information of ancient inscriptions is diverse and the text examples of ancient inscription are insufficient, which make it very difficult to learn the semantic information between Chinese characters for recognizing damaged characters. The challenging task of damaged characters recognition and understanding by Chinese character spatial semantic modeling is attempted to be solved in this paper. Based on Hierarchical Decomposition Embedding(HDE), the proposed DynamicGrape performs feature mapping on damaged character image and determines whether it is damaged. If character is not damaged, its image is directly converted into hierarchical decomposition embedding to reason the edge weight of the bipartite graph for recognizing Chinese character. If character is damaged, it is necessary to search for possible Chinese characters and components in the encoding set, select the feature dimension of HDE from image mapping, and input the bipartite graph to infer the possible Chinese character. In the self-built dataset and Chinese Text in the Wild(CTW) dataset, the experimental results show that the bipartite graph network can not only transfer and infer Chinese character pattern of damaged characters effectively, but also precisely recognize and understand damaged Chinese characters. It opens up new ideas for the damaged structure information processing.

Key words: Damaged inscription; Inscription prediction; Inscription recognition; Damaged text recognition; Bipartite graph neural networks

收稿日期: 2023-08-14; 改回日期: 2023-12-08; 网络出版: 2023-12-18

*通信作者: 蔺广逢 lgf78103@xaut.edu.cn

基金项目: 国家自然科学基金(61771386), 陕西省重点研发计划(2020SF-359), 陕西省自然科学基金(2021JM-340)

Foundation Items: The National Natural Science Foundation of China (61771386), Key Research and Development Program of Shaanxi (2020SF-359), Natural Science Basic Research Plan in Shaanxi Province of China (2021JM-340)

1 引言

为了推进古籍碑刻数字化建设, 本文针对因时间长久, 保存不当等原因造成碑刻内容不同程度缺损的问题进行研究, 提出残损汉字的字形预测与识别理解问题。

从字形角度观测, 汉字是一个可拆分结构, 能够建立从汉字到部件, 再从部件到笔画的层次关系。残损汉字是指字形不完整的汉字。残损汉字因为缺失的位置不同, 即使是同一种风格的同一个汉字, 它的字形表示差异也会很大。

根据汉字结构研究的不同层次, 把现有深度学习下的方法分为字符级方法、部件级方法和笔画级方法。字符级方法^[1-14]在不涉及对字符内在空间结构认知, 利用循环神经网络和连接时间分类(Connectionist Temporal Classification, CTC)的方法^[1-4]、利用编解码器和注意力机制的方法^[5-9]以及利用图像分割思想的方法^[10-12]。解决汉字行识别的重复问题, 但它们都需要大量的训练样本。另外最新的研究在标记原型学习^[13]的基础上提出了一种字符级上下文解耦网络(Character-Context Decoupling Framework, CCDF)^[14], 减少词级样本中上下文信息在新字符的视觉表征带来的影响, 提升词级的开放集文本识别能力。部件级方法^[15-24]利用单个汉字图像识别, 构建单个字符的部件和部件之间的空间关系来识别字符, 这些方法相比字符级识别方法更加灵活, 但也对字符的内在结构有一定的先验知识要求。笔画级方法^[25-29]在字符笔画序列的基础上对汉字进行五笔编码, 使用编码向量代表汉字, 而向量中的维度代表笔画, 最终利用距离度量相似度, 对单个汉字图像进行识别。但笔画级别的拆分会损失字符内部的一部分信息, 导致笔画序列和字符类别无法一一对应。

残损汉字的识别的主流方法是利用生成对抗网络(Generative Adversarial Network, GAN)对残损汉字进行修复, 然后进行识别。但上述思路对残损汉字数据集的依赖、对汉字图像背景要求高等问题。

在字符层面研究残损碑文识别, 因其视觉表征差异大, 相同风格碑文样本少, 不足以建立好的判别模型, 不能达到好的识别结果。此外, 从文本语义对残损汉字进行识别, 因古碑文语义丰富且样例不足, 很难建立有效的语言语义模型。在碑文识别任务中, 特别是考古中常会遇见需要定义的新字符等, 如何处理这些新字符和新文字是模式识别与文字识别技术与应用面临的一个重要挑战^[30]。

本文的创新点在于针对少样本或零样本的残损碑文图像的单个字符图像进行识别, 将其转换为统

一风格(楷体)的条件下, 尝试利用汉字内部结构信息对残损碑文进行预测识别。本文有3点贡献:

(1)利用基于Unicode 9.0标准的表意文字描述序列规则对汉字结构进行分解, 通过编码表示汉字的内部结构, 首次通过建模二部图表示部件级汉字内部空间结构语义信息, 学习并传播汉字部件结构空间结构语义信息。

(2)从缺损值插补的角度, 把对残损汉字编码的补全转化为二部图的边预测任务, 对残损汉字字形进行预测和识别。

(3)实验表明提出的方法成功地解决了在少样本或零样本情况下, 对残损碑文进行预测识别的难题, 为碑文图像识别领域带来了新的研究思路和方法。

在接下来的内容描述几个关键点: 第2节绍基于汉字编码的二部图建模和残损汉字识别网络框架; 第3节介绍实验数据集、实验环境和实验结果及分析; 第4节总结本文所做工作, 并且对残损文字识别的未来研究方向进行了展望, 最后讨论了在其他领域的可能潜在应用价值。

2 汉字编码与二部图建模

本文在汉字编码的基础上利用二部图网络推理字形结构传播与迁移, 建立最终残损汉字字形的预测识别模型。分别从汉字编码、二部图建模汉字编码、和残损汉字识别网络详细介绍本文方法。

2.1 汉字编码

汉字作为一个可拆分的结构, 拆分等级从高到低, 依次为笔画→部件→汉字。当汉字拆分为笔画序列时, 笔画作为汉字拆分的最小单元, 难以提供高级语义, 而当拆分汉字至部件等级, 可以保证一定的部件分布结构。但仅以部件集合或部件序列来表示汉字, 可能存在重复, 例如, “旱”字与“盱”字就会被视作同一个字。针对此问题, 需要利用部件不同的结合方式, 即结构来表示并区分不同的字。因此, 只有同时利用部件和结构才能实现汉字的唯一表示。本文应用了层级拆分嵌入(Hierarchical Decomposition Embedding, HDE)是Cao等人^[20]提出的一种汉字编码方法。它能够同时处理汉字内部的结构和部件信息, 保证汉字编码的唯一性。

HDE编码方法可以在识别过程中把少样本字符转换成汉字部件, 在汉字部件空间中提供了识别任意字符的可能。

2.2 二部图建模字符到部件映射关系

在完成汉字编码后, 为了建立从残损汉字到部件的预测, 需要学习汉字和部件的映射关系。因此在二部图上建模从字符到部件的异质结点结构, 将缺失部件的修复问题转化成在二部图上的边预测问题。

在借助汉字编码表示字符和部件与结构关系的二部图中,节点由两类不同的节点构成,单个字符作为一种类型节点,部件集和结构集作为另一种类型节点。字符节点的初始特征向量只在拥有对应部件的维度为1,其中,1代表该维度对应的部件存在,0代表不存在,如图1中橙色向量所示。部件和结构节点的初始特征向量也是one-hot向量,其在代表对应部件或结构的维度上为1,如图1中蓝色向量所示。二部图的边仅存在于不同类型节点之间,相同类型节点间不存在边。具体来说,HDE编码向量中非零维度的值作为对应字符节点和结构或部件节点的加权边。在图2的矩阵中,行表示不同的汉字的HDE编码(灰色的区域表示不存在对应的结构和部件),列表示不同汉字包含的结构和部件。

2.3 残损汉字识别网络

得到汉字编码并借助二部图建模字符到部件与结构的映射关系后,需要利用图网络挖掘汉字内部信息,用以复原残损汉字的编码,因此本文提出残损汉字识别网络DynamicGrape,网络结构如图3所示。

DynamicGrape输入的数据包括汉字图像和图像上的部件两部分内容。一方面对汉字图像进行处理,经过特征提取模块得到的特征同时送入残损判断模块和双向嵌入迁移模块。残损判断模块利用图像特征判断汉字图像中字符是否残损,双向嵌入迁移模块将特征空间的特征映射到编码空间内,统一图数据的向量空间。另一方面,图像上的部件称为部件类别集,部件类别集在得到残损判断模块的结果后开始处理。首先利用部件类别集在汉字编码集检索得到字符候选集,当残损判断模块判断待识别汉字为残损字符时,需要利用字符候选集在汉字编码集上通过部件检索策略得到待识别汉字可能存在

的部件候选集;当残损判断模块判断待识别汉字为完整字符时,直接使用输入模型的部件类别集。最后,将已知的汉字编码和映射到编码空间的待识别汉字编码同时送入图网络进行优化。图网络对二部图优化后,利用优化后的节点表征预测部件类别集(完整情况)/部件候选集(残损情况)对应的二部图边关系,用以重新构建待识别汉字编码。使用距离和字符候选集对应的编码进行度量,距离最小的汉字编码所对应的字符类就是待识别汉字的最终识别结果。

2.3.1 特征提取

当前模型输入的图像是单个汉字的图像,图像尺寸受到限制,为了兼顾特征的有效提取,特征提取模块选择的是更换了全连接层的Resnet18模型 $f(\cdot)$ 。

输出的特征向量 F_c 会送入残损判断模块和双向嵌入迁移模块。

2.3.2 残损判断

为判别汉字图像残损与否,在得到特征向量 F_c 后送入残损判断模块。残损判断模块仅由1层全连接层构成,通过Softmax得到最终输出概率 P_{dama} ,确定汉字残损情况。其损失函数 L_{dama} 采用交叉熵函数。

2.3.3 双向嵌入迁移

为了图数据上的向量空间统一,需要将特征空间下的特征向量映射到编码空间上。零样本中文文本识别网络(Zero-shot Chinese Text Recognition Network, ZCTR N)^[21]所提出的双向嵌入迁移模块不仅能够实现特征空间向编码空间的映射,还能实现编码空间到特征空间的映射,即双向嵌入来约束信息的迁移与变换。此模块的输入是特征空间下的特征向量,输出是编码空间下的待识别汉字编码。

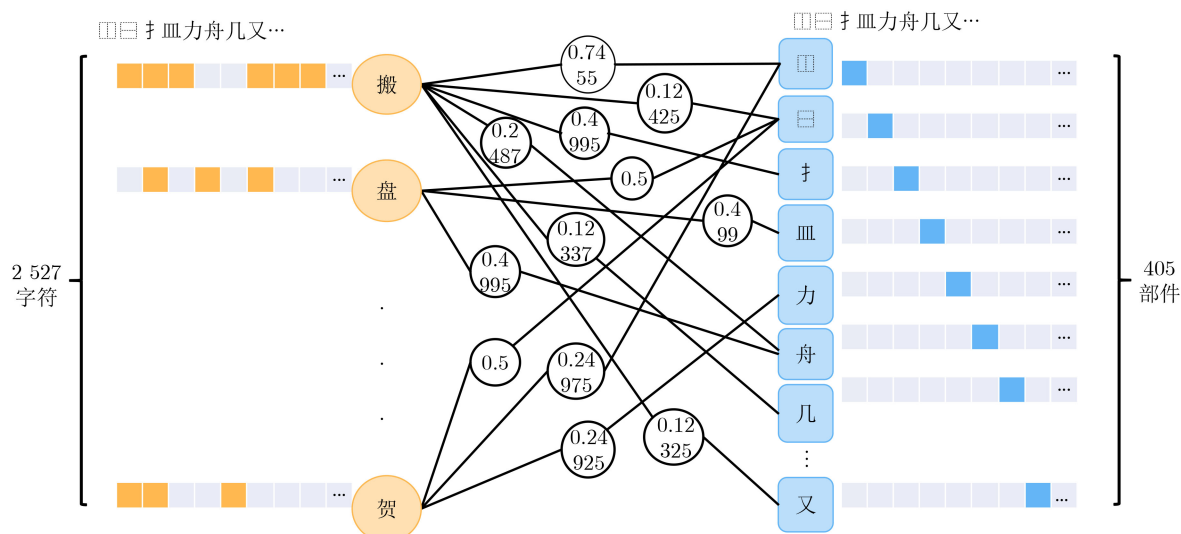


图1 二部图建模汉字编码集

双向嵌入迁移模块采用1层全连接层构成，为了保证不会有负值出现，采用修正线性单元 (Rectified Linear Unit, ReLU)作为激活函数。重构损失如式(1)

$$L_{re} = \frac{1}{N} \|\mathbf{F}'_c - \mathbf{F}_c\|_2^2 \quad (1)$$

这里 \mathbf{F}'_c 是双向嵌入迁移模块恢复的特征向量， N 是特征空间下特征向量的维数。

双向嵌入迁移模块用于特征空间向编码空间的映射，在待识别汉字编码并入二部图之前，让待识别汉字的编码和二部图内的汉字编码置于同一编码空间中，有利于保证图数据节点表征和拓扑结构的统一。

2.3.4 部件检索策略

在处理汉字图像的同时，需要使用部件检索策

	𠂇	𠂈	扌	皿	力	舟	几	...	又
搬	0.749	0.124	0.499			0.248	0.123	...	0.123
盘		0.5		0.499		0.499		...	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
贺	0.249	0.5			0.249			...	

图2 HDE汉字编码集

略对部件类别集进行处理，得到字符候选集和部件候选集。

字符候选集的获取需要利用输入的部件类别集在汉字编码集上进行检索，如图4所示。在以表格可视化的汉字编码集中，部件对应表格列，汉字对应表格行。给定部件类别集 R_C ，针对部件类别集内的一个单独部件 r_c ，利用 r_c 所对应的部件列，筛选出该部件列中非零单元格对应的所有汉字行，构成该部件的字符候选子集 sub_c 。对所有部件类别集中的部件筛选完毕后，得到字符候选子集集合 $sub_{cs} = \{sub_c^1, \dots, sub_c^i, \dots, sub_c^K\}$ 。最终对字符候选子集集合作交集操作，得到字符候选集 sub_C

$$sub_C = sub_c^1 \cap \dots \cap sub_c^i \cap \dots \cap sub_c^K \quad (2)$$

其中， sub_c^i 是利用第 i 个部件在汉字编码集检索后得到的字符候选子集， K 是部件类别集 R_C 内部件的数量。

部件候选集的检索过程只有在残损判断模块判断文字图像上的文字为残损情况下才会进行，需要利用字符候选集在汉字编码集上进行检索，如图5所示，在以表格可视化的汉字编码集中，部件对应表格列，汉字对应表格行。针对字符候选集 sub_C 中的一个单独字符，利用该字符所对应的汉字行，筛选出该汉字行中非零单元格对应的部件列，构成该

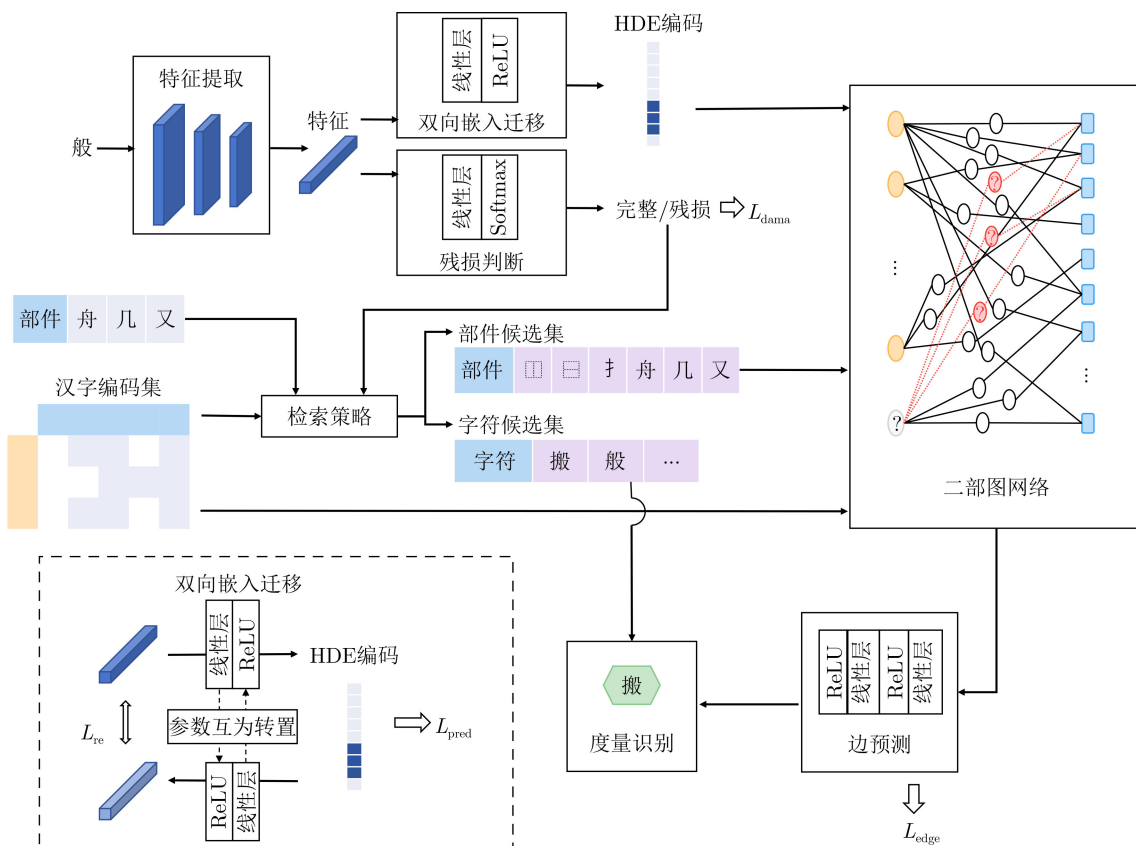


图3 DynamicGrape网络结构图

字符的部件候选子集 sub_r 。对所有字符候选集中的汉字筛选完毕后,得到部件候选子集集合 $sub_{rs} = \{sub_r^1, \dots, sub_r^i, \dots, sub_r^S\}$ 。最终对部件候选子集集合合作并集操作,得到部件候选集 sub_R 。

$$sub_R = sub_r^1 \cup \dots \cup sub_r^i \cup \dots \cup sub_r^S \quad (3)$$

其中, sub_r^i 是利用第*i*个字符在汉字编码集检索后得到的部件候选子集, *S*是字符候选集 sub_C 内字符的数量。

部件检索策略在使用过程中,一方面要求DynamicGrape在检索策略前有部件集合的输入或预测来提供检索依据,另一方面它能够得到残缺汉字可能存在的部件,并且缩小度量识别时识别候选字符集的数量,提高识别准确率。

2.3.5 添加边嵌入的GraphSAGE

在将二部图数据送入图网络之前,需要将待识别汉字的信息加入到二部图数据中。具体来说,首先构建待识别汉字的节点特征,该节点特征为部件类别集对应维度为1的1维向量。该字符节点和部件类别集内的部件节点相连,每个边的权值是该汉字编码对应部件的维度值。值得注意的是,针对每一个待识别汉字,图数据都会单独处理一次。因此,在识别不同的汉字时,图数据并不一致。

在完成二部图数据构建后,送入图网络中优化二部图的节点表征和拓扑结构。

图网络中,图采样和聚集方法(Graph Sample and AGgrEgate, GraphSAGE)^[31]拥有一定的对未知数据的泛化能力,能够适应图数据的变化。Grape^[32]认识到边中蕴含的信息,在GraphSAGE的基础上增加了对边信息的利用。为兼顾适应图数据变化以及边信息的利用,使用以Grape作为图网络实现的基础对二部图的优化。

具体来说,给定二部图 $G=(V, E)$, *V*表示节点集合, *E*表示边集合。对于第*l*层,输入的是第*l*-1层输出的节点表征 $h_v^{(l-1)}$ 和边嵌入 $e_{uv}^{(l-1)}$,聚合函数如式(4),输出经过聚合的节点表征 $n_v^{(l)}$

$$n_v^{(l)} = AGG_l \left(\sigma \left(P^{(l)} \cdot \text{CONCAT} \left(h_v^{(l-1)}, e_{uv}^{(l-1)} \right) \right) \right) \quad (4)$$

$|\forall u = N(v)|$

其中, AGG_l 是聚合函数, σ 是激活函数, $P^{(l)}$ 是可训练参数, CONCAT 是拼接函数, *N*是节点邻域。

经过节点更新得到新的节点嵌入,节点更新函数如式(5)

$$h_v^{(l)} = \sigma \left(Q^{(l)} \cdot \text{CONCAT} \left(h_v^{(l-1)}, n_v^{(l)} \right) \right) \quad (5)$$

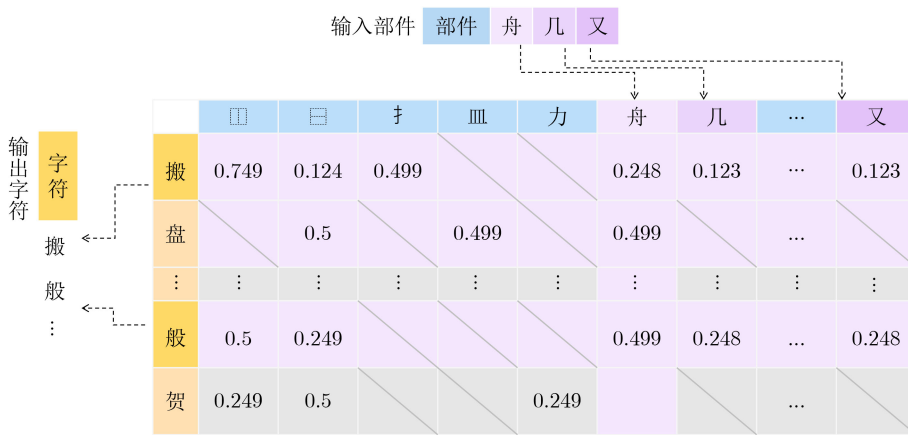


图4 从输入部件到输出字符的检索过程示例

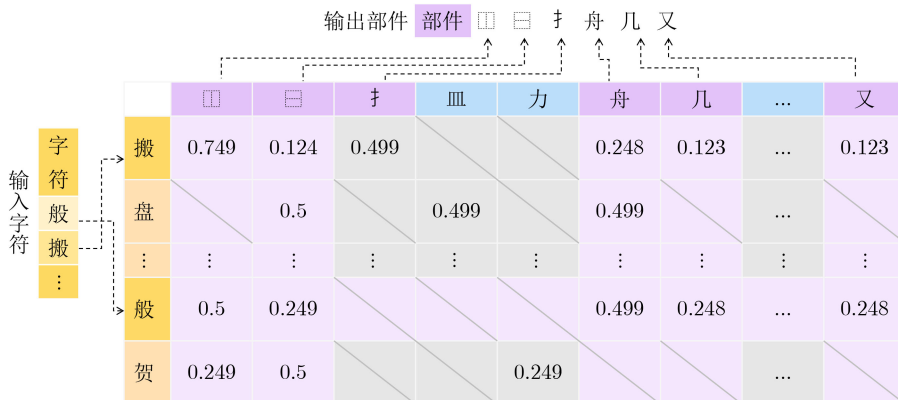


图5 从输入字符到输出部件的检索过程示例

其中 $Q^{(l)}$ 是可训练参数。

最后, 对边嵌入进行更新, 得到 $e_{uv}^{(l)}$, 边嵌入更新函数如式(6)

$$e_{uv}^{(l)} = \sigma \left(W^{(l)} \cdot \text{CONCAT} \left(e_{uv}^{(l-1)}, h_u^{(l)}, h_v^{(l)} \right) \right) \quad (6)$$

其中, $W^{(l)}$ 是可训练参数, $h_u^{(l)}$ 是边 $e_{uv}^{(l)}$ 的字符节点表征, $h_v^{(l)}$ 是边 $e_{uv}^{(l)}$ 的部件或结构节点表征。

现得到第 l 层输出节点表征 $h_v^{(l)}$ 与边嵌入 $e_{uv}^{(l)}$ 。图网络实现对字符节点和部件或结构节点的表征优化, 学习到了汉字和部件的映射关系, 为下游边预测建立从残损汉字到部件的预测打好基础。

2.3.6 边预测与度量识别

经过图网络对二部图的优化后, 利用由多层感知器构成的边预测模块在优化后的节点表征 $h_v^{(L)}$ 上得到待识别汉字节点和部件与结构节点新的边关系 \hat{E}_{uv} , 从而构建待识别汉字的对应编码 hde_{char} 。

在完成待识别汉字节点和部件与结构节点边关系的重构后, 一方面可以用于构建损失函数, 另一方面, 用于字符识别。损失函数的构建如式(7)所示

$$L_{\text{pred}} = \frac{1}{M} \|\hat{E}_{uv} - E_{uv}\|_2^2 \quad (7)$$

其中, E_{uv} 是以加权边形式使用的汉字编码, M 是汉字编码需要的部件集和结构集的总计数量。

在得到待识别汉字的编码后, 利用余弦距离和字符候选集内字符对应的汉字编码进行度量, 余弦距离如式(8)。最终, 和待识别汉字编码距离最小的编码所对应的字符是DynamicGrape的识别结果

$$D(\text{hde}_{\text{char}}, \text{hde}_{\text{subc}}) = \frac{\text{hde}_{\text{char}}^T \text{hde}_{\text{subc}}}{\|\text{hde}_{\text{char}}\| \times \|\text{hde}_{\text{subc}}\|} \quad (8)$$

其中, hde_{subc} 是字符候选集中字符对应的汉字编码构成的集合。

3 实验

3.1 数据集

采用中文自然文本(Chinese Text in the Wild, CTW)数据集^[33]和自建数据集部件-编码数据集验证本文网络(数据和源码地址为<https://github.com/eightyninth/DynamicGrape>)。

CTW数据集是自然场景下的中文字符数据集, 包含32 285张高分辨率图像。该数据集共包含1 018 402个文字实例, 分为3 850类, 用6种属性来描述文字在自然场景下的情况。数据集的划分情况如下: 训练集占75%, 验证集占5%, 分类测试集和检测测试集各占10%。本文删去了CTW训练集和验证集中不常见字符类别的文字实例, 来适应汉字分解规则下得到的汉字编码。因此, CTW的文

字类别从3 850类降为2 091类, 文字实例从832 933个降为255 838个。

部件-编码数据集(Component-Code Dataset, CCD)共包含2 527个汉字编码以及对应的汉字部件分割图像。在此基础上把汉字分为残损和完整两种情况, 其中, 残损的情况仅对应缺失一个部件的情况, 完整的情况不作处理。最终, 数据集每个字符类仅存在一个完整字符图像。整个数据集分为训练集、验证集和测试集。部件-编码数据集进行数据两种不同划分。一种是零样本(zero shot)情况, 一个字符类的所有图像会同时分属一个集合, 保证训练集, 验证集, 测试集的字符类互不相交; 另一种是多样本(many shot)的情况, 会打乱所有汉字图像, 随机分配图像, 此时, 同一字符类可能会同时出现于训练集、验证集和测试集之中。训练集、验证集和测试集的字符样本数量分别为6 542, 2 022以及1 021。实验初始数据集图像大小为96×96, 学习率为0.001, batch size设置为128, 模型的最大迭代次数为5 000。

3.2 先进方法的对比实验

为验证DynamicGrape的性能, 和现有方法卷积递归神经网络(Convolutional Recurrent Neural Network, CRNN)^[1]、稠密卷积网络(Dense convolutional Network, DenseNet)^[34]、残差网络(Residual Nets, ResNet)^[35]、部件计数网络(Radical Counter Network, RCN)^[15]、零样本中文识别网络(Zero-shot Chinese Text Recognition Network, ZCTRNet)^[21]和字符上下文解耦框架(Character Context Decoupling Framework, CCDF)^[14]在部件-编码数据集两种情况以及在CTW数据集上进行对比, 实验结果如表1所示。

对实验结果进行分析, 可以得到以下结论: 第一, 相较于其他识别方法, DynamicGrape在大多数指标上能够得到最好的识别准确率。第二, 因为CRNN, ResNet和DenseNet并不是零样本方法, 因此它们在部件-编码数据集的零样本情况下失效, 而在部件-编码数据集的多样本情况以及CTW数据集上能得到很好的识别准确率。第三, RCN, ZCTRNet和CCDF可以处理零样本字符识别, 但是没有考虑残损恢复, 而且利用距离进行识别, 面临包括模型预测编码过于稠密、距离计算字符类候选数量过多等问题。这些问题会导致RCN和CCDF失效、ZCTRNet识别准确率过低。

3.3 消融实验

DynamicGrape主要由以下部分组成, 特征提取模块、残损判断模块、双向嵌入迁移模块、图网

络以及边预测模块。为分析每个因素对模型效果的影响,设计消融实验,对模型中引入的不同模块进行分析验证,在保持输入数据和其他训练参数不变,模型在对模块进行取舍的情况下,分析其残损判断的准确率,加权边预测的偏差以及最终模型识别的准确率。

消融实验分为以下几种,第1种,模型DynamicGrape w/o damaged & BiTrans,表示模型DynamicGrape删去了残损判断模块和双向嵌入迁移模块;第2种,模型DynamicGrape w/o damaged,表示模型DynamicGrape删去了残损判断模块;第3种,DynamicGrape w/o BiTrans,表示模型DynamicGrape删去了双向嵌入迁移模块;第4种,模型DynamicGrape是提出的完整方法,实现了完整的识别过程。

消融实验的评估在部件-编码数据集的零样本情况和多样本情况以及CTW数据集上进行,实验结果如表2、表3和表4所示。其中,平均平方误差

(Mean Squared Error, MSE)和平均绝对误差(Mean Absolute Error, MAE)描述了边权值预测的偏差,它们反应了汉字编码复原的偏差,而识别准确率描述了模型最终的识别结果。

对实验结果进行观测,可以得到以下结论:

第一,在部件-编码数据集的零样本情况以及CTW数据集中,DynamicGrape缺少双向嵌入迁移模块能够得到编码复原最优值。具体来说,在部件-编码数据集的零样本情况下,MAE能够达到0.093 4,在CTW数据集上,MAE能够达到0.010 3;在部件-编码数据集多样本情况下,DynamicGrape能够得到编码复原最优值,MAE能够达到0.064 9。

第二,在所有数据集中,DynamicGrape w/o damaged得到最高的完整文字识别率。这是因为DynamicGrape中残损判断模块判断准确率的低下,导致了部件检索策略对完整字符编码添加了不存在的部件,影响了度量识别过程。

第三,在部件-编码数据集的零样本情况下,DynamicGrape能够得到最优的残损文字识别准确以及所有文字识别准确率;在部件-编码数据集的多样本情况下,DynamicGrape w/o damaged能够得到最优的残损文字识别准确以及所有文字识别准确率。这里的原因和结论二相同,是完整模型的残损判断模块判断准确率的低下造成的。

3.4 字形的残损比例分析

在残损汉字识别方法中,修复过程对文字图像中字形残损的比例(Damaged Rate, DR)有一定要求,本文在单个部件图像的基础上构建部件-编码数据的图像,因此能够得到单个部件字形的像素统计,根据字形的残损比例对部件-编码数据集中的字符图像进行分类统计。

利用部件-编码数据集的字符图像残损比例统计,验证DynamicGrape对不同残损比例字符图像的识别准确率。由实验结果可知,随着字形残损比例的不断增加,DynamicGrape的识别准确率在不断下降。为了更直观地展示该趋势,以图6展示部件-编码数据集不同数据划分下不同残损比例的识别准确率。图中个别与整体趋势相反的结果是由该字形残损比例下样本量过少导致的。

4 结论

本文针对残损碑文汉字小样本辨识问题,提出基于图网络的残损汉字识别方法DynamicGrape。不同于仅利用部件序列,造成汉字表示重复的情况,本文所提方法利用结构和部件对汉字进行编码表示,在保证充分利用汉字字符内部信息的同时对汉字唯一表示。

表1 先进方法的对比实验(%)

模型	识别准确率		
	完整汉字	残损汉字	所有汉字
部件-编码数据集 零样本			
CRNN ^[1]	0	0	0
DenseNet ^[34]	0	0	0
ResNet ^[35]	0	0	0
RCN ^[15]	0	0	0
ZCTR ^[21]	0.401 6	0.201 9	0.252 0
CCDF ^[14]	0	0	0
DynamicGrape	99.407 1	55.606 9	66.567 8
部件-编码数据集 多样本			
CRNN ^[1]	0.269 4	0	0.197 8
DenseNet ^[34]	76.350 1	54.814 8	60.534 1
ResNet ^[35]	81.750 5	64.915 8	69.386 7
RCN ^[15]	0	0	0
ZCTR ^[21]	0.189 8	0.137 3	0.151 2
CCDF ^[14]	0	0	0
DynamicGrape	97.579 1	50.909 1	63.303 7
CTW数据集			
CRNN ^[1]	66.057 3	/	66.057 3
DenseNet ^[34]	79.375 4	/	79.375 4
ResNet ^[35]	80.429 4	/	80.429 4
RCN ^[15]	0	/	0
ZCTR ^[21]	0.142 7	/	0.142 7
CCDF ^[14]	82.485 4	/	82.485 4
DynamicGrape	97.254 4	/	97.254 4

表 2 部件-编码 零样本数据集下的消融实验

模型	MSE	MAE	残损判断 准确率(%)	识别准确率(%)		
				完整汉字	残损汉字	所有文字
验证集						
DynamicGrape w/o damaged & BiTrans	0.017 6	0.105 5	/	99.407 1	56.068 6	66.913 9
DynamicGrape w/o damaged	0.017 8	0.115 6	/	99.407 1	55.606 9	66.567 8
DynamicGrape w/o BiTrans	0.019 7	0.085 2	95.796 2	98.419 0	59.102 9	68.941 6
DynamicGrape	0.020 9	0.104 2	74.975 3	82.608 7	66.160 9	70.277 0
测试集						
DynamicGrape w/o damaged & BiTrans	0.016 0	0.096 2	/	98.809 5	63.589 1	72.282 1
DynamicGrape w/o damaged	0.016 8	0.109 0	/	98.809 5	63.199 0	72.184 1
DynamicGrape w/o BiTrans	0.021 8	0.093 4	93.535 7	98.015 9	63.979 2	72.380 0
DynamicGrape	0.021 0	0.105 3	75.318 3	80.158 7	71.001 3	73.261 5

表 3 部件-编码 多样本数据集下的消融实验

模型	MSE	MAE	残损判断 准确率(%)	识别准确率(%)		
				完整汉字	残损汉字	所有文字
验证集						
DynamicGrape w/o damaged & BiTrans	0.015 5	0.103 9	/	98.137 8	51.380 5	63.798 2
DynamicGrape w/o damaged	0.018 4	0.115 9	/	97.765 4	51.447 8	63.748 8
DynamicGrape w/o BiTrans	0.013 8	0.068 0	94.906 0	97.579 1	51.447 8	63.699 3
DynamicGrape	0.011 8	0.061 5	26.508 4	97.579 1	50.707 1	63.155 3
测试集						
DynamicGrape w/o damaged & BiTrans	0.016 8	0.107 8	/	98.550 7	51.275 2	64.054 8
DynamicGrape w/o damaged	0.019 6	0.119 0	/	98.550 7	51.677 9	64.348 7
DynamicGrape w/o BiTrans	0.014 7	0.070 6	92.752 2	97.826 1	50.335 6	63.173 4
DynamicGrape	0.013 4	0.064 9	26.934 4	98.188 4	51.275 2	63.956 9

表 4 CTW数据集下的消融实验

模型	MSE	MAE	残损判断准确率(%)	识别准确率(%)		
				完整汉字	残损汉字	所有文字
DynamicGrape w/o damaged & BiTrans	0.002 2	0.024 9	/	97.247 9	/	97.247 9
DynamicGrape w/o damaged	0.001 1	0.014 1	/	97.352 0	/	97.352 0
DynamicGrape w/o BiTrans	0.001 1	0.010 3	/	97.033 2	/	97.033 2
DynamicGrape	0.001 6	0.013 3	/	97.254 4	/	97.254 4

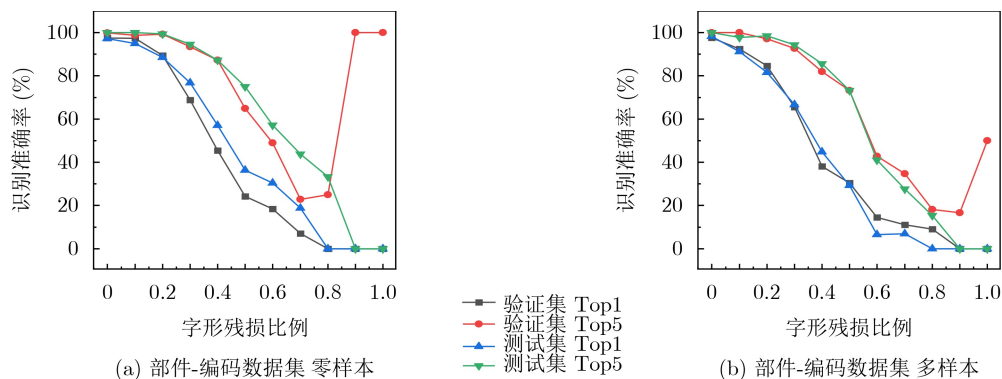


图 6 部件-编码数据集不同数据划分下不同残损比例的识别准确率趋势

DynamicGrape首先对汉字进行残损判别,帮助后续残损字符的编码重建。然后将汉字图像空间转换到特征空间进一步映射到编码空间。最后转化汉字编码到二部图上进行表示,在加入待识别汉字节点和对应加权边后利用图网络进行优化和信息迁移,完成对残损待识别汉字对应的编码进行构建,与字符候选集中汉字编码对比得到最终识别结果。通过对比实验以及噪声实验显示出部件的结构和组成对于残损汉字识别的重要影响。

虽然本文方法构建了残损碑文识别的较完整方法,在数据集中取得了显著的实验效果,但是,仍存在3个局限性。一个是,要求检测出碑文的残损文字,并对其进行高效的去噪处理;二是,假设残损汉字缺失了一个整部件(实际情况可能缺失整部件的部分);三是,残损程度限制在字形的40%以下。因此,本文方法是对实际应用的一种较理想假设下做出的研究,其重点在于通过汉字部件信息迁移构建残损文字。在未来的工作中,我们将进一步研究模型对自然残损碑文和部件集合的识别,增加模型对实际数据处理的普适性。

本文的研究成果不仅在碑文图像识别领域具有重要意义,还在其他领域有潜在的广泛应用价值。针对少样本或零样本情况下的问题,本文的方法可以被拓展应用于具有空间或时序关联关系的数据缺失的恢复与推理中。如特定空间结构复杂的图像补全,气候时序结构数据的预测和特定高速目标轨迹结构预测与跟踪等。

参 考 文 献

- [1] SHI Baoguang, BAI Xiang, and YAO Cong. 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. doi: [10.1109/TPAMI.2016.2646371](https://doi.org/10.1109/TPAMI.2016.2646371).
- [2] BUŠTA M, NEUMANN L, and MATAS J. Deep TextSpotter: An end-to-end trainable scene text localization and recognition framework[C]. 2017 IEEE International Conference on Computer Vision, Venice, Italy, 2017: 2204–2212. doi: [10.1109/ICCV.2017.242](https://doi.org/10.1109/ICCV.2017.242).
- [3] LIU Xuebo, LIANG Ding, YAN Shi, *et al.* FOTS: Fast oriented text spotting with a unified network[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 5676–5685. doi: [10.1109/CVPR.2018.00595](https://doi.org/10.1109/CVPR.2018.00595).
- [4] LIU Yuliang, CHEN Hao, SHEN Chunhua, *et al.* ABCNet: Real-time scene text spotting with adaptive Bezier-curve network[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020: 9809–9818. doi: [10.1109/CVPR42600.2020.00983](https://doi.org/10.1109/CVPR42600.2020.00983).
- [5] SHI Baoguang, WANG Xinggang, LYU Pengyuan, *et al.* Robust scene text recognition with automatic rectification[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 4168–4176. doi: [10.1109/CVPR.2016.452](https://doi.org/10.1109/CVPR.2016.452).
- [6] LI Hui, WANG Peng, SHEN Chunhua, *et al.* Show, attend and read: A simple and strong baseline for irregular text recognition[C]. The 33rd AAAI conference on artificial intelligence, Honolulu, USA, 2019: 8610–8617. doi: [10.1609/aaai.v33i01.33018610](https://doi.org/10.1609/aaai.v33i01.33018610).
- [7] QIAO Zhi, ZHOU Yu, YANG Dongbao, *et al.* SEED: Semantics enhanced encoder-decoder framework for scene text recognition[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020: 13528–13537. doi: [10.1109/CVPR42600.2020.01354](https://doi.org/10.1109/CVPR42600.2020.01354).
- [8] HE Tong, TIAN Zhi, HUANG Weilin, *et al.* An end-to-end TextSpotter with explicit alignment and attention[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 5020–5029. doi: [10.1109/CVPR.2018.00527](https://doi.org/10.1109/CVPR.2018.00527).
- [9] WANG Wenhai, XIE Enze, LI Xiang, *et al.* PAN++: Towards efficient and accurate end-to-end spotting of arbitrarily-shaped text[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(9): 5349–5367. doi: [10.1109/TPAMI.2021.3077555](https://doi.org/10.1109/TPAMI.2021.3077555).
- [10] LIAO Minghui, ZHANG Jian, WAN Zhaoyi, *et al.* Scene text recognition from two-dimensional perspective[C]. The 33rd AAAI Conference on Artificial Intelligence, Honolulu, USA, 2019: 8714–8721. doi: [10.1609/aaai.v33i01.33018714](https://doi.org/10.1609/aaai.v33i01.33018714).
- [11] YU Deli, LI Xuan, ZHANG Chengquan, *et al.* Towards accurate scene text recognition with semantic reasoning networks[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020: 12113–12122. doi: [10.1109/CVPR42600.2020.01213](https://doi.org/10.1109/CVPR42600.2020.01213).
- [12] LYU Pengyuan, LIAO Minghui, YAO Cong, *et al.* Mask TextSpotter: An end-to-end trainable neural network for spotting text with arbitrary shapes[C]. The 15th European Conference on Computer Vision, Munich, Germany, 2018: 67–83. doi: [10.1007/978-3-030-01264-9_5](https://doi.org/10.1007/978-3-030-01264-9_5).
- [13] LIU Chang, YANG Chun, QIN Haibo, *et al.* Towards open-set text recognition via label-to-prototype learning[J]. *Pattern Recognition*, 2023, 134: 109109. doi: [10.1016/j.patcog.2022.109109](https://doi.org/10.1016/j.patcog.2022.109109).
- [14] LIU Chang, YANG Chun, and YIN Xucheng. Open-set text recognition via character-context decoupling[C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, USA, 2022: 4523–4532. doi: [10.1109/CVPR52688.2022.00448](https://doi.org/10.1109/CVPR52688.2022.00448).
- [15] LI Yunqing, ZHU Yixing, DU Jun, *et al.* Radical counter network for robust Chinese character recognition[C]. The 25th International Conference on Pattern Recognition, Milan, Italy, 2021: 4191–4197. doi: [10.1109/ICPR48806](https://doi.org/10.1109/ICPR48806).

- 2021.941291.
- [16] WANG Wenchao, ZHANG Jianshu, DU Jun, *et al.* DenseRAN for offline handwritten Chinese character recognition[C]. 2018 16th International Conference on Frontiers in Handwriting Recognition, Niagara Falls, USA, 2018: 104–109. doi: [10.1109/ICFHR-2018.2018.00027](https://doi.org/10.1109/ICFHR-2018.2018.00027).
- [17] ZHANG Jianshu, ZHU Yixing, DU Jun, *et al.* Radical analysis network for zero-shot learning in printed Chinese character recognition[C]. 2018 IEEE International Conference on Multimedia and Expo, San Diego, USA, 2018: 1–6. doi: [10.1109/ICME.2018.8486456](https://doi.org/10.1109/ICME.2018.8486456).
- [18] WU Changjie, WANG Zirui, DU Jun, *et al.* Joint spatial and radical analysis network for distorted Chinese character recognition[C]. 2019 International Conference on Document Analysis and Recognition Workshops, Sydney, Australia, 2019: 122–127. doi: [10.1109/ICDARW.2019.40092](https://doi.org/10.1109/ICDARW.2019.40092).
- [19] WANG Tianwei, XIE Zecheng, LI Zhe, *et al.* Radical aggregation network for few-shot offline handwritten Chinese character recognition[J]. *Pattern Recognition Letters*, 2019, 125: 821–827. doi: [10.1016/j.patrec.2019.08.005](https://doi.org/10.1016/j.patrec.2019.08.005).
- [20] CAO Zhong, LU Jiang, CUI Sen, *et al.* Zero-shot handwritten Chinese character recognition with hierarchical decomposition embedding[J]. *Pattern Recognition*, 2020, 107: 107488. doi: [10.1016/j.patcog.2020.107488](https://doi.org/10.1016/j.patcog.2020.107488).
- [21] HUANG Yuhao, JIN Lianwen, and PENG Dezhi. Zero-shot Chinese text recognition via matching class embedding[C]. The 16th International Conference on Document Analysis and Recognition, Lausanne, Switzerland, 2021: 127–141. doi: [10.1007/978-3-030-86334-0_9](https://doi.org/10.1007/978-3-030-86334-0_9).
- [22] YANG Chen, WANG Qing, DU Jun, *et al.* A transformer-based radical analysis network for Chinese character recognition[C]. The 25th International Conference on Pattern Recognition, Milan, Italy, 2021: 3714–3719. doi: [10.1109/ICPR48806.2021.941243](https://doi.org/10.1109/ICPR48806.2021.941243).
- [23] DIAO Xiaolei, SHI Daqian, TANG Hao, *et al.* RZCR: Zero-shot Character Recognition via Radical-based Reasoning[C]. The Thirty-Second International Joint Conference on Artificial Intelligence, Macao, China, 2023.
- [24] ZENG Jinshan, XU Ruiying, WU Yu, *et al.* STAR: Zero-shot Chinese character recognition with stroke-and radical-level decompositions[EB/OL]. <https://arxiv.org/abs/2210.08490>, 2022.
- [25] GAN Ji, WANG Weiqiang, and LU Ke. Characters as graphs: Recognizing online handwritten Chinese characters via spatial graph convolutional network[EB/OL]. <https://arxiv.org/abs/2004.09412>, 2020.
- [26] GAN Ji, WANG Weiqiang, and LU Ke. A new perspective: Recognizing online handwritten Chinese characters via 1-dimensional CNN[J]. *Information Sciences*, 2019, 478: 375–390. doi: [10.1016/j.ins.2018.11.035](https://doi.org/10.1016/j.ins.2018.11.035).
- [27] CHEN Jingye, LI Bin, and XUE Xiangyang. Zero-shot Chinese character recognition with stroke-level decomposition[EB/OL]. <https://arxiv.org/abs/2106.11613>, 2021.
- [28] YU Haiyang, CHEN Jingye, LI Bin, *et al.* Chinese character recognition with radical-structured stroke trees[EB/OL]. <https://arxiv.org/abs/2211.13518>, 2022.
- [29] CHEN Zongze, YANG Wenxia, and LI Xin. Stroke-based autoencoders: Self-supervised learners for efficient zero-shot Chinese character recognition[J]. *Applied Sciences*, 2023, 13(3): 1750. doi: [10.3390/app13031750](https://doi.org/10.3390/app13031750).
- [30] 杨春, 刘畅, 方治屿, 等. 开放集文字识别技术[J]. 中国图象图形学报, 2023, 28(6): 1767–1791. doi: [10.11834/jig.230018](https://doi.org/10.11834/jig.230018).
YANG Chun, LIU Chang, FANG Zhiyu, *et al.* Open set text recognition technology[J]. *Journal of Image and Graphics*, 2023, 28(6): 1767–1791. doi: [10.11834/jig.230018](https://doi.org/10.11834/jig.230018).
- [31] HAMILTON W L, YING Z, and LESKOVEC J. Inductive representation learning on large graphs[C]. The 31st International Conference on Neural Information Processing Systems, Long Beach, USA, 2017: 1025–1035.
- [32] YOU Jiakuan, MA Xiaobai, DING D Y, *et al.* Handling missing data with graph representation learning[C]. The 34th International Conference on Neural Information Processing Systems, Vancouver, Canada, 2020: 1601.
- [33] YUAN Tailing, ZHU Zhe, XU Kun, *et al.* A large Chinese text dataset in the wild[J]. *Journal of Computer Science and Technology*, 2019, 34(3): 509–521. doi: [10.1007/s11390-019-1923-y](https://doi.org/10.1007/s11390-019-1923-y).
- [34] HUANG Gao, LIU Zhuang, VAN DER MAATEN L, *et al.* Densely connected convolutional networks[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 4700–4708. doi: [10.1109/CVPR.2017.243](https://doi.org/10.1109/CVPR.2017.243).
- [35] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, *et al.* Deep residual learning for image recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 770–778. doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).

蒯广逢：男，副教授，研究方向为图像处理与模式识别。
吴娜：女，硕士生，研究方向为图像处理与模式识别。
贺梦兰：女，硕士生，研究方向为图像处理与模式识别。
张二虎：男，教授，研究方向为图像处理与模式识别。
孙强：男，副教授，研究方向为情感计算、智慧气象与计算机视觉。

责任编辑：余蓉