

基于多标签语义分割的硬笔字笔画提取^①

余嘉云¹, 李丁宇², 徐占洋², 王晶弘², 林巍³

¹(南京师范大学, 南京 210023)

²(南京信息工程大学, 南京 210044)

³(江苏少儿春互联教育科技有限公司 南京技术研发中心, 南京 210032)

通信作者: 李丁宇, E-mail: 1063765113@qq.com



摘要: 汉字作为中华文化的载体, 因其复杂的结构区别于其他文字. 笔画作为汉字的基本单元, 在硬笔字评价中起到至关重要的作用. 正确提取笔画, 是硬笔字评价的首要步骤. 现有的笔画提取方法多数是基于规则的, 由于汉字的复杂性, 这些规则通常无法顾及所有特征, 且在评价时无法根据笔顺等信息与模板字笔画匹配. 为了解决这些问题, 该文将笔画提取转化为多标签语义分割问题, 提出了多标签语义分割模型 (M-TransUNet), 利用深度卷积模型以汉字为单位任务进行训练, 保留了笔画原有结构, 避免了笔画段组合的二义性, 同时得到了硬笔字的笔顺, 有利于笔画评价等下游任务. 由于硬笔字图像只分为前景和背景, 没有额外颜色信息, 所以更容易产生 *FP* (false positive) 分割噪声. 为解决此问题, 本文还提出了一种针对笔画分割结果的局部平滑策略 (local smooth strategy on stroke, LSSS), 淡化噪声的影响. 最后, 本文对 M-TransUNet 的分割性能以及效率进行了实验, 证明了本文算法在很小性能损失的情况下, 极大地提升了效率. 同时对 LSSS 算法进行了实验, 证明其在 *FP* 噪声消除的有效性.

关键词: 硬笔字; 笔画提取; 多标签语义分割; 局部平滑策略

引用格式: 余嘉云, 李丁宇, 徐占洋, 王晶弘, 林巍. 基于多标签语义分割的硬笔字笔画提取. 计算机系统应用, 2024, 33(9): 174-182. <http://www.c-s-a.org.cn/1003-3254/9620.html>

Stroke Extraction for Chinese Handwriting Character Based on Multi-label Semantic Segmentation

YU Jia-Yun¹, LI Ding-Yu², XU Zhan-Yang², WANG Jing-Hong², LIN Wei³

¹(Nanjing Normal University, Nanjing 210023, China)

²(Nanjing University of Information Science & Technology, Nanjing 210044, China)

³(Nanjing Technology R & D Center, Jiangsu Children's Spring Interconnection Education Technology Co. Ltd., Nanjing 210032, China)

Abstract: As the carrier of Chinese culture, Chinese characters are distinguished from other scripts by their complex structure. As the basic unit of Chinese characters, strokes play a vital role in the evaluation of Chinese handwriting characters. The correct extraction of strokes is the primary step in evaluating Chinese handwriting characters. Most existing stroke extraction methods are based on specific rules, and due to the complexity of Chinese characters, these rules usually cannot take into account all the features, and cannot match the strokes of template characters based on stroke order and other information during evaluation. To address these issues, this study transforms stroke extraction into a multi-label semantic segmentation problem and proposes a multi-label semantic segmentation model (M-TransUNet), which utilizes a deep convolutional model to train with Chinese characters as a unit task, retaining the original structure of the strokes and avoiding ambiguity in stroke segment combinations. At the same time, the stroke order of the Chinese handwriting characters is obtained, which is conducive to downstream tasks, such as stroke evaluations. Since the handwriting images

① 收稿时间: 2024-01-11; 修改时间: 2024-02-29, 2024-04-23; 采用时间: 2024-04-29; csa 在线出版时间: 2024-07-30

CNKI 网络首发时间: 2024-07-31

are only divided into foreground and background without additional color information, they are more prone to generating *FP* segmentation noise. To solve this problem, this study also proposes a local smooth strategy on strokes (LSSS) for the stroke segmentation results to dilute the impact of noise. Finally, this study conducted experiments on the segmentation performance and efficiency of M-TransUNet, demonstrating that the algorithm significantly enhances efficiency with minimal performance loss. Additionally, experiments were carried out on the LSSS algorithm to demonstrate its effectiveness in eliminating *FP* noise.

Key words: Chinese handwriting character; stroke extraction; multi-label semantic segmentation; local smooth strategy

汉字是中华文化的载体,学习书法是学习中华文化的重要途径.汉字作为世界上最复杂的文字之一,初学者总是需要在老师指导下学习书法.国内中小学对书法教育越来越重视,对于书法教育资源的需求也随之增长,人们急切地需要计算机来进行书法辅助教学.而作为汉字的基本单位,笔画在硬笔字评价中起到至关重要的作用.

硬笔字笔画提取向来是一项艰巨的任务,汉字的结构复杂,笔画类型繁多,且笔画交叉的情况很多.传统方法多是通过规则判定交叉区域附近笔画段归属,进行笔画的拆分,大致分为两类.

其中一类是细化得到骨架,通过像素的排列规则进行骨架笔画的提取^[1-3].通过解决交叉点的分裂问题^[4]、拆解笔画段并以一定规则重组笔画^[5],对硬笔字的骨架进行笔画提取. Lin 等人^[6]将骨架拆为笔画段,通过计算所有交叉点的笔画段的入射方向,基于双向图确定需要连接的笔画段对.李建华等人^[7]将基本笔画被简化为4种类型:横、竖、撇、捺.根据骨架与水平线夹角确定笔画类型.在处理特定类型的笔画时,其余类型的笔画被视为噪声.刘相聪等人^[8]利用 CPD 算法配准目标字与模板字的交叉点与拐点来确定笔画段归属,并引入笔段权重以处理复杂汉字结构.这类方法非常依赖骨架的质量,交叉点的分裂以及毛刺等问题,通过规则是无法完全避免的,且丢失了笔锋、顿笔等信息,对于关键点配准的方法,甚至无法保证两个点集数量相等.

另一类是通过硬笔字图像原有的特征进行提取. Cao 等人^[9]使用了 PBOD 算法获取笔画延续性质,改善了重叠笔画的分离并保持了连通性.史伟等人^[10]根据笔段外围轮廓的一阶微分特性,获得其边界特征,根据特征动态调整搜索方向,更精确地匹配笔段的实际轮廓.程立等人^[11]使用水平垂直差分法检测汉字轮廓,形成封闭的多边形轮廓.提取轮廓上的凸点和凹点.根据凹点的位置、角度、距离等条件,将相交笔画的凹点进行配

对并连接.章夏芬等人^[12]则是在骨架基础上,将笔画轮廓同时引入到汉字的笔画分割中.通过精确定位每个笔画的轮廓范围,根据交叉区域的特征,将它们分配给相应的轮廓笔画.这类方法相较基于骨架的方法,利用了更多的信息,提高了容错率,但仍然摆脱不了人为制定规则的束缚,无法涵盖手写硬笔字可能出现的所有情况.

近年来也出现了基于神经网络的笔画提取算法. Liu 等人^[13]将笔画提取转化为实例分割问题,采用两阶段策略^[14],并提供了大量数据集,将数据集内的所有汉字共同训练,提高了效率和鲁棒性,但由于实例分割的特性,可以正确划分笔画的类别,但无法得知属于硬笔字的哪个笔画,从而影响与模板字笔画对比评价.张巍等人^[15]则是利用条件生成对抗网络^[16],将每个字的每个笔画作为一项训练任务,可以选择生成具体某个笔画,但由于汉字笔画繁多,此方法训练任务繁重,且在提取硬笔字笔画时需要加载多次模型权重,提取笔画时间过长.这类方法通过学习样本笔画的特征来更好地适应形态复杂多变的手写硬笔字.相较于传统方法,此类方法有着很好的鲁棒性.

因此本文将使用神经网络对笔画进行提取.语义分割作为经典的神经网络分支任务,获得了大量研究者的关注,其中的代表网络 U-Net^[17]更是被广泛应用与研究,其全卷积的编码解码、跳跃连接结构,使其可以在结构轻量化的同时,保证高级语义与低级语义的正确表达.在此基础上,研究者们对其进行了多种改进, TransUNet^[18]则是向 U-Net 加入了 self-attention 机制^[18],在保证高分辨率特征的基础上,捕获了远距离的特征依赖.

然而 TransUNet 只能解决单标签的分割问题,无法解决笔画相交的情况.多标签语义分割是最适合硬笔字笔画提取的方法.且硬笔字图像通常为二值图,没有额外的颜色信息,对于神经网络分割算法而言,只能依赖于空间结构信息,可供学习的特征有限,相较于其他 RGB 图像的分割任务更容易出现 *FP* 分割噪声,影

响下游细化与笔画评价任务。

为解决上述问题, 本文使用多标签语义分割模型 M-TransUNet, 将汉字作为单位任务, 笔画作为类别进行训练. 对于分割噪声, 本文还提出了一种针对笔画分割结果的局部平滑策略 LSSS, 消除分割噪声的同时, 避免整体图像被模糊, 保证了笔画边缘的清晰.

1 算法分析

由于硬笔字的特殊性, 两个甚至多个不同笔画之间可能存在相交, 如图 1 的中段相交 (图 1(a)) 和顶点相交 (图 1(b)), 其中顶点相交导致的笔画相连, 可以规定相交区域属于其中一个笔画, 不会影响其他笔画的连贯性; 对于中段相交的笔画, 如果规定相交区域属于其中一个笔画, 则其他笔画会出现断裂, 破坏了笔画应有的连贯性. 将每个汉字的每一个笔画作为单位任务, 每个任务只做前景背景的二分类任务可以避免这个问题, 但随之而来的是成倍的训练成本, 以及提取笔画时, 每个笔画都需要加载模型权重. 因此, 本文使用多标签语义分割进行笔画提取. 尽管多标签语义分割相较于二分类的单标签语义分割, 增加了很多需要学习的特征, 而且硬笔字图像本身缺乏颜色以及灰度信息, 只能依赖于空间结构信息, 导致分割噪声有所增加. 本文提出了一种针对笔画的局部平滑策略 LSSS, 消除了部分分割噪声.

1.1 多标签语义分割网络 M-TransUNet

本文参考了 TransUNet 网络^[18], 结构见图 2, 其在

U-Net^[17]结构上加入了 Transformer 模块. 输入图像尺寸为 224×224 , 首先利用卷积层提取特征图, 将特征图分为 14×14 块并展平成一维序列; 其次将其输入 Transformer layer^[19], 经过 12 层多头自注意力模块和全连接层, 捕获全局上下文信息; 然后将 Transformer layer 的输出调整为二维特征图, 并进行 3 次上采样, 在上采样的过程中分别与下采样期间卷积层 28×28 、 56×56 与 112×112 的特征图进行拼接, 保留高分辨率语义信息; 最后上采样恢复到原图 224×224 大小, 经过输出层的 3×3 卷积核将通道缩减至类别数, 使用 Softmax 与 argmax 得到分割掩膜.

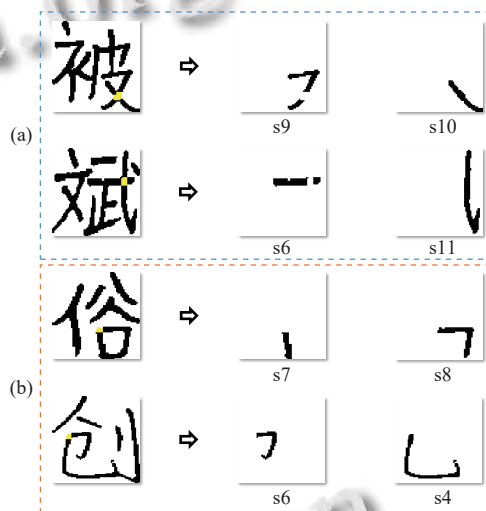


图 1 笔画相交示意图

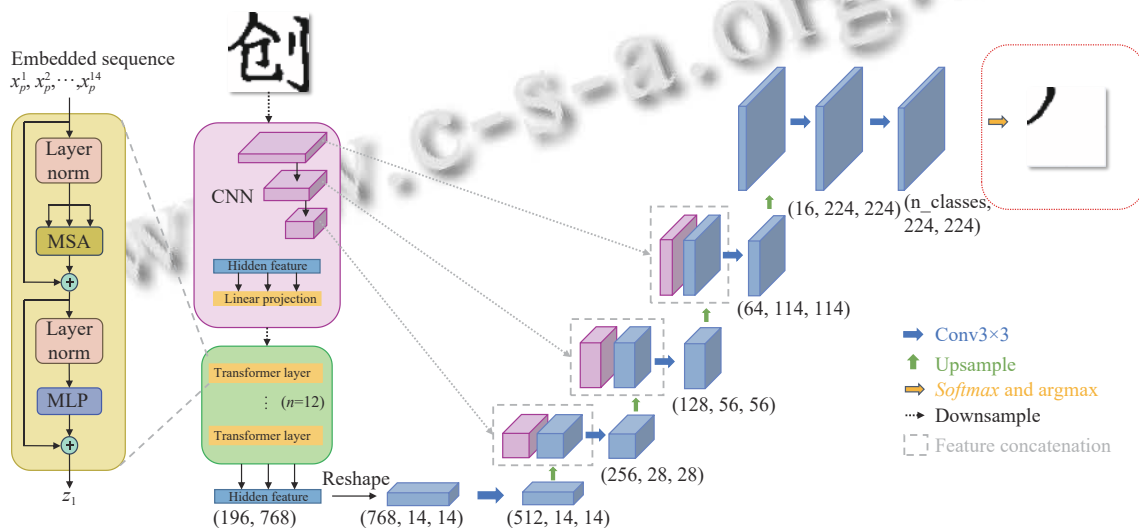


图 2 TransUNet 网络框架

通常, Transformer 将输入特征图当作一维序列, 并且只注重全局信息的捕获, 缺少对局部高分辨率特征

的关注. 而 U-Net 则受限于网络的深度, 以及卷积网络本身的性质, 无法关注到全局信息. TransUNet 则是结

合了 Transformer 与 U-Net 的优点,通过自注意力机制获得全局上下文信息的同时,得益于 U-Net 的跳跃连接设计,保留了足够的高分辨率特征.在硬笔字笔画提取的任务中,输入图像为二值图,只分为硬笔字前景和背景,相交笔画之间不存在颜色或灰度的差异,提取笔画只能依靠笔画自身的形状结构,以及硬笔字的全局结构.而 TransUNet 的特性刚好满足硬笔字笔画提取的需求.在单标签分类或语义分割任务中,为了得到一个得分最大的类别,通常希望有较大的得分差距. *Softmax* 激活函数见式 (1):

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_{c=1}^C e^{z_c}} \quad (1)$$

由于指数函数 e^{z_i} 在 $z_i > 0$ 时,其导数 e^{z_i} 非常大, z_i 之间的微小差异,会导致极大的激活值差异,较大的值会抑制较小的值,这使得 *Softmax* 激活函数广泛应用在单标签分类以及语义分割任务中.而在多标签语义分割的任务中,每个像素可能同时属于多个类别,为了得到一个或多个得分大于阈值的类别,我们不希望得分较高的类别抑制得分稍低的类别,类别之间的得分应当是独立的. *Sigmoid* 激活函数经常应用于二分类任务,

见式 (2):

$$\sigma(z_i) = \frac{1}{1 + e^{-z_i}} \quad (2)$$

可以将神经网络输出的实数映射至 $[0, 1]$, $\sigma(z_i) > 0.5$ 则认为属于 1 所代表的类别,否则认为属于 0 所代表的类别.多标签多分类任务则可以看作是 C 个二分类任务的合集,因此相较于 *Softmax*, *Sigmoid* 是更适合于多标签语义分割的激活函数.

本文提出了多标签 TransUNet (M-TransUNet),将 TransUNet 的输出层激活函数由 *Softmax* 改为 *Sigmoid*,分别计算不同类别的得分,以 0.5 为阈值,判断像素是否属于当前类别,其过程见图 3.由于硬笔字图像背景为白色像素,前景为黑色像素,得分小于 0.5 认为属于当前笔画类别,得分大于 0.5 则认为不属于当前笔画.多标签任务等价于多个二分类任务,应当使用二分类交叉熵函数平均值 (式 (3)) 作为损失函数,其中 N 为像素个数, y_j 为单个像素真实值, \hat{y}_j 为单个像素预测值.

$$L_{\text{BCE}} = \frac{\sum_{j=1}^N -y_j \times \log(\hat{y}_j) - (1 - y_j) \times \log(1 - \hat{y}_j)}{N} \quad (3)$$

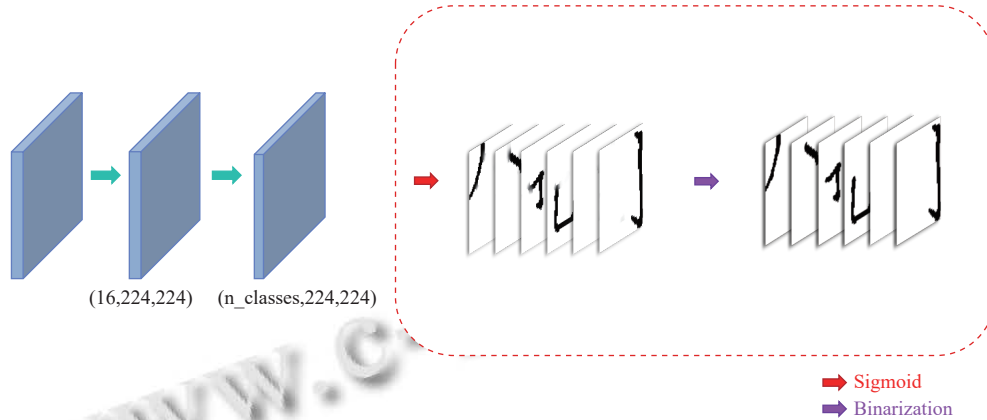


图 3 M-TransUNet 输出层

然而二分类交叉熵函数可以很好地代表单个像素的预测值与其真实值的差异,却无法准确表达语义分割任务中预测掩膜与真实掩膜的相似度. *DiceLoss* (式 (4)) 可以表达预测掩膜与真实掩膜之间的相似度,但对单个像素的差异则不敏感.

$$L_{\text{DiceLoss}} = 1 - \frac{2 \sum_{j=1}^N y_j \hat{y}_j}{\sum_{j=1}^N y_j + \sum_{j=1}^N \hat{y}_j} \quad (4)$$

L_{DiceLoss} 更加关注于 y_j 和 \hat{y}_j 接近 1 的掩膜相似程

度,对于接近 0 的 y_j 和 \hat{y}_j 则非常不敏感.而我们需要关注的是黑色笔画像素,因此,本文对 *DiceLoss* 做了反转 (reverse) 变式:

$$L_{\text{DiceLossR}} = 1 - \frac{2 \sum_{j=1}^N (1 - y_j)(1 - \hat{y}_j)}{\sum_{j=1}^N (1 - y_j) + \sum_{j=1}^N (1 - \hat{y}_j)} \quad (5)$$

因此,本文采用混合损失函数:

$$L_{\text{Mix}} = \frac{L_{\text{BCE}} + L_{\text{DiceLossR}}}{2} \quad (6)$$

1.2 局部平滑策略 LSSS

虽然硬笔字的笔画有先后顺序,但不存在笔画被遮挡的情况,因此笔画的分割结果应该是一整块连续的黑色像素.然而本文所使用的硬笔字图像为二值图,没有额外的颜色信息,只能依赖于空间结构信息,相较于其他的分割任务更容易出现 *FP* 分割噪声,如图 4,这样的噪声会对下游的笔画细化及评价任务产生影响.局部回归已很好地应用在图像处理领域^[20],图像的像素排列是规则有序的,每一个像素与周围像素都有一定的联系,局部回归可以利用周围邻域内的所有像素值,拟合中心像素最有可能的值.这样会使图像更加平滑,可以一定程度上消除小面积的 *FP* 分割噪声.

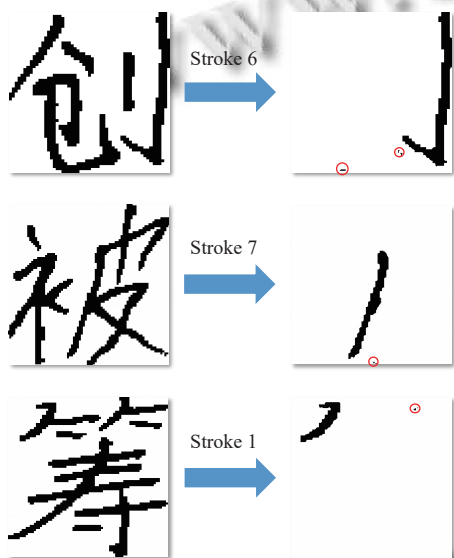


图 4 分割噪声示意图

硬笔字图像不需要额外的灰度信息,硬笔字前景像素之间的灰度梯度为 0,如图 5 “撇”和“捺”相交部分;硬笔字前景和背景之间的灰度梯度为 1,如图 5 “捺”的边界.当进行笔画提取时,我们仍然希望笔画与背景之间保持接近 1 的灰度梯度,而局部回归这类利用邻域像素拟合中心像素值的平滑方法,会使得背景与前景之间的边缘变得模糊.

因此,本文提出了一种只针对硬笔字前景区域内的局部平滑策略 LSSS,利用硬笔字图像原有的二值特征,不需要额外的算法就可以分辨出前景区域,而分割后的笔画仍属于硬笔字前景的一部分.设定一个 $H \times H$,

H 为单数的窗口,将窗口内的像素分为 $\frac{H+1}{2}$ 级,我们将每一级视为一个整体.为了避免笔画边缘模糊,我们只计算第 n 级中属于硬笔字前景区域内的有效像素平均值.

$$A_n = \frac{S_n}{K_n} \quad (7)$$

其中, K_n 为有效像素个数, S_n 为有效像素灰度值之和.不同级权重由内级向外级逐渐衰减, α 为衰减系数,每一级的权重如式 (8):

$$W_n = \frac{\alpha^{n-1}}{\sum_{a=1}^{\frac{H+1}{2}} \alpha^{a-1}} \quad (8)$$

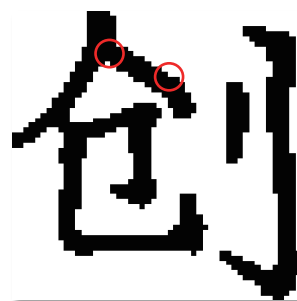


图 5 灰度梯度示意图

这样就得到了中心像素拟合值.

$$p = \sum_{n=1}^{\frac{H+1}{2}} A_n \times W_n \quad (9)$$

当某一级不存在硬笔字前景区域内的有效像素时,其权重 W^* 需要平均分配给其余所有级.

$$p^* = p \times (1 + W^*) \quad (10)$$

算法流程见图 6.由于排除了图像背景区域的影响,消除了小面积 *FP* 误分割,也保证了笔画边缘的清晰.

2 实验

2.1 实验数据

实验所使用的数据来自于中小学学生,均为楷体硬笔字.为验证笔画的数量以及笔画的交叉次数对多标签语义分割性能的影响,以“创”“被”和“筹”这 3 个字作为实验对象.其中“创”字笔画数量较少,笔画相交较少且为端点相交;“被”字笔画数量适中,笔画相交适中,同时包含端点相交和中段相交;“筹”字笔画数量较多,同一笔画相交次数较多,且多为中段相交.每个字包含

240 张图片, 其中“创”字因其包含 6 个笔画, 因此包括 1 440 张独立的笔画图片 (240 张全字图片乘以每个字的笔画数)。同理, “被”字和“筹”字分别包含 10 个和 13 个笔画, 分别有 2 400 张和 3 120 张笔画图片。所有实验数据集共包含 6 960 张笔画图片。所有的数据集都

是人工标注, 并确保数据的准确性和一致性。标注的方法是在每张图片中, 除当前关注的笔画外, 其余的笔画都被擦除。训练集、验证集比例为 8:2, 均为二值化后的硬笔字图片, 并且通过旋转和翻转对训练集进行扩充。

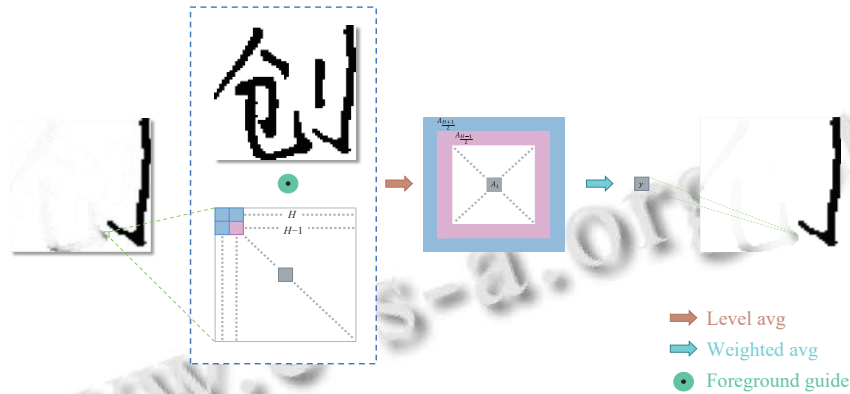


图6 LSSS 算法流程图

2.2 实验过程

本实验均使用随机梯度下降算法, 初始学习率设置为 0.004, 衰减率为 1×10^{-6} , 动量为 0.9, 迭代次数为 200。首先对 TransUNet 进行训练, 以“创”字为代表, 其中每个笔画作为一个单独的训练任务, 共训练 6 次, 并按照笔画排序。然后对 M-TransUNet 进行训练, 标签由单个硬笔字所有笔画的人工标签拼接而成, 形成一个 $n \times w \times h$ 的三维矩阵, 其中 n 代表笔画数量, 并且符合笔画顺序。仍然以“创”字为代表, 我们只需建立一个训练任务, 得到一个三维输出, 从第 1 个维度将输出矩阵分为 6 个二维矩阵。

得到神经网络输出结果后, 本文分别对单标签输出与多标签输出使用 LSSS 算法, 窗口长度 H 设置为 8, 衰减系数设置为 0.9。最后对平滑后的图像做二值化处理。

2.3 实验结果

2.3.1 分割性能

本文通过 $F1$ score ($F1$) 和 Accuracy (Acc) 对实验结果进行评估, $F1$ 在正负样本不平衡的情况下仍可以很好地评价分割结果。

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \quad (11)$$

$$F1 = \frac{2TP}{2TP+FP+FN} \quad (12)$$

表 1–表 3 分别记录了“创”“被”以及“筹”字使用 TransUNet 进行分割得到的单笔画、所有笔画平均以及使用 M-TransUNet 分割的整体的 $F1$ 和 Acc 。其中 TransUNet 对 3 个字的 $\overline{F1}$ 分割性能已经达到了 0.977 0、0.965 5、0.965 1, 对于书写规范的硬笔字已达到人工提取笔画的效果。M-TransUNet 需要学习更多的笔画特征, 在分割性能上有所下降, 但 $\overline{F1}$ 仍可以达到 0.967 7、0.951 2、0.951 3。

表 1 “创”笔画提取性能

笔画	Acc	F1
s1	0.9992	0.9814
s2	0.9986	0.9593
s3	0.9985	0.9726
s4	0.9982	0.9717
s5	0.9991	0.9823
s6	0.9993	0.9945
Avg	0.9988	0.977
Multi avg	0.9977	0.9677

2.3.2 分割效率

此外, 本实验还记录了单标签与多标签语义分割的效率对比。在单标签语义分割任务中, 每个笔画需要单独训练一个模型, 这不仅消耗大量的训练时间, 而且在预测时也需要加载多个模型, 增加了时间成本。而多标签语义分割只需一次训练便可同时处理多个笔画, 极大地节省了训练时间并减少了预测时加载模型的时间。本实验对比了“创”“被”和“筹”字使用 TransUNet,

M-TransUNet 以及张巍等人^[15]方法 (CGAN) 训练及预测所消耗时间, 见图 7. 其中训练时间单位为 min, 预测时间单位为 s. 实验平台为 RTX2060.

表 2 “被”笔画提取性能

笔画	Acc	F1
s1	0.9997	0.9859
s2	0.9991	0.9813
s3	0.9975	0.9532
s4	0.9984	0.9349
s5	0.9986	0.9535
s6	0.998	0.967
s7	0.9984	0.9755
s8	0.9982	0.9626
s9	0.9973	0.9674
s10	0.9985	0.9741
Avg	0.9984	0.9655
Multi avg	0.9967	0.9512

其中表 4 记录了“被”字通过 M-TransUNet 分别使用二分类交叉熵 (BCE)、DiceLoss 以及混合损失函数 (Mix) 达到的分割性能.

表 3 “筹”笔画提取性能

笔画	Acc	F1
s1	0.9991	0.9713
s2	0.9987	0.9804
s3	0.9982	0.9727
s4	0.9992	0.9734
s5	0.9983	0.9776
s6	0.9979	0.9628
s7	0.9986	0.9653
s8	0.9981	0.9423
s9	0.9975	0.9527
s10	0.9973	0.9479
s11	0.9987	0.9637
s12	0.9993	0.9603
s13	0.9997	0.9756
Avg	0.9985	0.9651
Multi avg	0.9975	0.9513

2.3.3 损失函数消融实验

我们还测试了不同损失函数对于训练结果的影响,

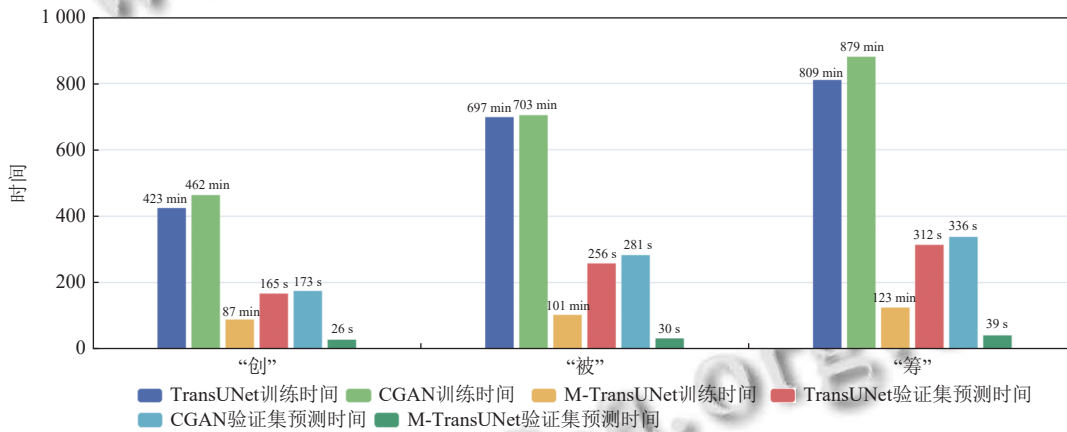


图 7 训练及预测时间对比图

表 4 损失函数性能影响

损失函数	Acc	F1
BCE	0.9832	0.7214
DiceLoss	0.9372	0.9286
Mix	0.9967	0.9512

当单独使用 BCE 时, 由于正负样本的不均衡, 导致输出结果偏向白色背景, F1 性能大幅下降. 单独使用 DiceLoss 则对于每个像素约束不够充分, Acc 性能与 F1 性能都有所下降. 而 Mix 在 Acc 以及 F1 性能都优于单独使用一种损失函数.

2.3.4 LSSS 性能影响

图 8 展示了测试集中典型的 FP 误分割, 笔画为“创”第 4 画. 对于小面积的分割噪声 (图 8(a)), 像素本

身灰度值更加接近 0, 其周围像素灰度值多接近 1, 在转换为硬标签后如图 8(b). 经过 LSSS 处理后, 硬笔字前景区域内的像素灰度值变化更为平滑, 前景与背景交界处灰度梯度并没有受到影响, 被误分割的灰度值接近 0 的像素受周围灰度值接近 1 的像素的影响, 其灰度值也更加接近于 1 (图 8(c)), 转换为硬标签后如图 8(d).

表 5 记录了 TransUNet 与 M-TransUNet 输出结果经过 LSSS 处理后的表现, 可以发现 LSSS 在没有牺牲分割性能的情况下, 有效地消除了部分小面积误分割, 对下游细化以及评价任务更加友好. 图 9 展示了本文算法对笔画提取的最终结果.

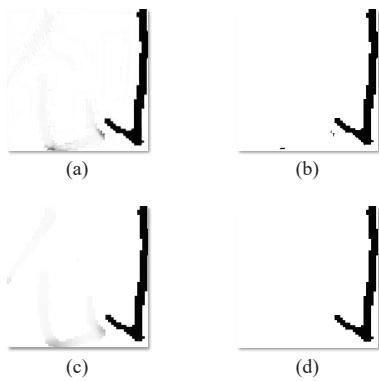


图8 平滑效果示意图

2.3.5 笔画提取性能对比

由于传统算法缺少客观的性能评估标准, 本文方法无法与传统算法进行直接的比较. 通常传统算法会人为判断一个笔画是否提取正确, 并计算提取正确的

笔画在所有笔画中所占比例. 阳平等^[21]工判断分割正确率为 86.7%, Xu 等人^[1]工判断 KT 字体笔画提取正确率为 92.74%, 且传统方法一个笔画出现分割错误, 会影响其他笔画. 而本文方法不存在多检漏检的情况, 提取的笔画与实际笔画是一一对应的, 笔画对应的正确率为 100%.

表5 经 LSSS 处理后的笔画提取性能

测试用例	Acc	F1
“创” Avg _{LSSS}	0.9991	0.9785
“创” Multi avg _{LSSS}	0.9980	0.9699
“被” Avg _{LSSS}	0.9987	0.9682
“被” Multi avg _{LSSS}	0.9976	0.9593
“筹” Avg _{LSSS}	0.9989	0.9659
“筹” Multi avg _{LSSS}	0.9978	0.9524



图9 测试集笔画提取效果展示

因此, 本文统计了“创”“被”“筹”这 3 个字使用 M-TransUNet 和张巍等人^[15]方法 (CGAN) 所有笔画验证集中 F1 超过 0.88 的笔画所占比例以作为正确提取率, 以及 Xu 等人^[1]所用方法 (Traditional) 基于本文数据集的提取正确率见表 6.

表6 笔画提取正确率 (%)

笔画提取算法	“创”	“被”	“筹”
Traditional	83.19	71.33	61.09
CGAN	98.92	97.31	96.31
M-TransUNet	97.84	96.53	94.95

本文方法相较于传统方法, 正确率有大幅提升. 相较于张巍等人的方法^[15], 本文方法正确率略有下降, 但得益于多标签的语义分割, 在训练时间及预测时间上有着大幅的优势. Liu 等人^[13]通过实例分割对笔画进行分割, 无法获取硬笔字的笔顺信息, 且无法避免多检漏

检的情况, 后续无法与模板字相应笔画匹配与比较, 本文方法在训练时通过标签的类别维度信息, 确定了硬笔字的笔顺, 输出的笔画可以直接与相应的模板字笔画进行比较与评分.

3 结束语

在本文中, 我们将硬笔字笔画提取转化为了语义分割问题, 通过形态不一的数据集让神经网络学习笔画的深层特征, 提高鲁棒性与准确性, 并且使用多标签语义分割网络 M-TransUNet 代替单标签, 大大减少了训练时间, 同时可以根据标签顺序确定笔画的顺序, 满足了下游笔画评价任务的必要条件. 对于分割噪声, 本文还提出了针对笔画的局部平滑策略 LSSS, 在消除分割噪声的同时, 避免了笔画边缘模糊, 下游笔画细化以

及评价任务无需对分割结果进行进一步处理. 实验结果表明了本文方法的可行性, 对于硬笔字智能评价具有重要的意义.

参考文献

- 1 Xu ZY, Liang Y, Zhang QN, *et al.* Decomposition and matching: Towards efficient automatic Chinese character stroke extraction. Proceedings of the 2016 Visual Communications and Image Processing. Chengdu: IEEE, 2016. 1–4.
- 2 范勇峰, 李成城, 林民, 等. 基于局部信息的手写汉字笔画提取. 内蒙古师范大学学报(自然科学汉文版), 2023, 52(2): 181–188.
- 3 荀恩东, 吕晓晨, 安维华, 等. 面向书写教学的手写汉字图像笔画还原. 北京大学学报(自然科学版), 2015, 51(2): 241–248.
- 4 Liao CW, Huang JS. Stroke segmentation by Bernstein-Bezier curve fitting. Pattern Recognition, 1990, 23(5): 475–484. [doi: [10.1016/0031-3203\(90\)90068-V](https://doi.org/10.1016/0031-3203(90)90068-V)]
- 5 Fan KC, Wu WH. A run-length-coding-based approach to stroke extraction of Chinese characters. Pattern Recognition, 2000, 33(11): 1881–1895. [doi: [10.1016/S0031-3203\(99\)00173-9](https://doi.org/10.1016/S0031-3203(99)00173-9)]
- 6 Lin F, Tang XO. Off-line handwritten Chinese character stroke extraction. Proceedings of the 2002 International Conference on Pattern Recognition. Quebec City: IEEE, 2002. 249–252.
- 7 李建华, 王宏, 闫文芝, 等. 一种新的汉字细化和笔画提取方法. 第六届全国信息获取与处理学术会议论文集(1). 北京: 《仪器仪表学报》杂志社, 2008. 230–233.
- 8 刘相聪, 李壮峰, 姜杰, 等. 基于 CPD 算法与笔段权重的楷体字笔画提取研究. 计算机应用与软件, 2022, 39(8): 204–212, 219.
- 9 Cao RN, Tan CL. A model of stroke extraction from Chinese character images. Proceedings of the 15th International Conference on Pattern Recognition. Barcelona: IEEE, 2000. 368–371.
- 10 史伟, 傅彦, 陈安龙, 等. 一种动态的汉字笔段提取方法. 计算机应用研究, 2008, 25(7): 1998–2000.
- 11 程立, 王江晴, 李波, 等. 基于轮廓的汉字笔画分离算法. 计算机科学, 2013, 40(7): 307–311.
- 12 章夏芬, 刘佳岩. 用爬虫法提取书法笔画. 计算机辅助设计与图形学学报, 2016, 28(2): 301–309. [doi: [10.3969/j.issn.1003-9775.2016.02.013](https://doi.org/10.3969/j.issn.1003-9775.2016.02.013)]
- 13 Liu LZ, Lin KY, Huang SX, *et al.* Instance segmentation for Chinese character stroke extraction, datasets and benchmarks. arXiv:2210.13826, 2022.
- 14 He KM, Gkioxari G, Dollár P, *et al.* Mask R-CNN. Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 2961–2969.
- 15 张巍, 张筱, 万永菁. 基于条件生成对抗网络的书法字笔画分割. 自动化学报, 2022, 48(7): 1861–1868.
- 16 Isola P, Zhu JY, Zhou TH, *et al.* Image-to-image translation with conditional adversarial networks. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 1125–1134.
- 17 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. Proceedings of the 18th Medical Image Computing and Computer-assisted Intervention. Munich: Springer, 2015. 234–241.
- 18 Chen JN, Lu YY, Yu QH, *et al.* TransUNet: Transformers make strong encoders for medical image segmentation. arXiv:2102.04306, 2021.
- 19 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010.
- 20 Zou BJ, Dai YL, He Q, *et al.* Multi-label classification scheme based on local regression for retinal vessel segmentation. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2021, 18(6): 2586–2597. [doi: [10.1109/TCBB.2020.2980233](https://doi.org/10.1109/TCBB.2020.2980233)]
- 21 阳平, 娄海涛, 胡正坤. 一种基于骨架的篆字笔划分割方法. 计算机科学, 2013, 40(2): 297–300. [doi: [10.3969/j.issn.1002-137X.2013.02.066](https://doi.org/10.3969/j.issn.1002-137X.2013.02.066)]

(校对责编: 孙君艳)