

自然场景下乌金体藏文的检测与识别^①



洪松, 高定国, 三排才让, 取次

(西藏大学 信息科学与技术学院, 拉萨 850000)

通讯作者: 高定国, E-mail: gdg@utibet.edu.cn

摘要: 自然场景乌金体藏文文本信息作为高度浓缩的高层语义信息, 不仅具有较大的研究和实用价值, 而且可以用于协助藏文场景文本理解领域的研究. 目前针对自然场景下乌金体藏文的检测与识别的相关研究甚少, 本文在人工收集的自然场景乌金体藏文图像数据集的基础上, 对比了目前常见的文字检测算法在自然场景乌金体藏文上的检测性能以及在不同特征提取网络下基于序列的文字识别算法 CRNN 在自然场景乌金体藏文图像数据集上的识别准确率并分析了在 314 张真实自然场景下乌金体藏文识别出错的特殊例子. 实验表明本文在文字检测阶段采用的可微分的二值化网络 DBNet 在测试集上具有更好的检测性能, 该方法在测试集上的准确率、召回率、F1 值分别达到了 0.89、0.59、0.71; 在文字识别阶段采用 MobileNetV3 Large 作为特征提取网络时, CRNN 算法在测试集上的识别准确率最高, 达到了 0.4365.

关键词: 自然场景; 乌金体藏文; 检测; 识别

引用格式: 洪松, 高定国, 三排才让, 取次. 自然场景下乌金体藏文的检测与识别. 计算机系统应用, 2021, 30(12): 332-338. <http://www.c-s-a.org.cn/1003-3254/8262.html>

Detection and Recognition of Wujin Style Tibetan Scripts in Natural Scenes

HONG Song, GAO Ding-Guo, SAMPEL Tsering, QU Ci

(Information Science and Technology Academy, Tibet University, Lasa 850000, China)

Abstract: As a highly condensed high-level semantic information, the text information of Wujin style Tibetan scripts in natural scenes not only has great research and practical value, but also can be used to assist researchers with text understanding in Tibetan scenes. At present, there are few related studies on the detection and recognition of Wujin style Tibetan scripts in natural scenes. Based on the manually collected image data set of Wujin style Tibetan scripts in natural scenes, this study compares the detection performance of common text detection algorithms on such scripts. The recognition accuracy of the sequence-based text recognition algorithm, CRNN, under different feature extraction networks is also compared on the image data set collected. Examples of recognition failure during the recognition of Wujin style Tibetan scripts in 314 real natural scenes are analyzed as well. Experiments show that the differentiable binary network, DBNet, used in the text detection stage has better detection performance on the test set. The accuracy, recall, and F1 value of this method on the test set reach 0.89, 0.59, and 0.71, respectively; when MobileNetV3 Large is used as the feature extraction network in the text recognition stage, the CRNN algorithm has the highest recognition accuracy of 0.4365 on the test set.

Key words: natural scene; Wujin Tibetan script; detection; recognition

① 基金项目: 西藏大学研究生高水平人才培养计划 (2018-GSP-020); 青海省藏文信息处理与机器翻译重点实验室/藏文信息处理教育部重点实验室开放课题 (2020Z001)

Foundation item: High-level Talent Training Plan of Graduates of Tibet University (2018-GSP-020); Open Fund of Qinghai Provincial Key Laboratory of Tibetan Information Processing and Machine Translation/Key Laboratory of Tibetan Information Processing, Ministry of Education (2020Z001)

收稿时间: 2021-02-09; 修改时间: 2021-03-18; 采用时间: 2021-04-30

1 引言

目前深度学习广泛应用于计算机视觉的各个领域,基于深度学习的自然场景文字检测与识别一直都是一项具有挑战性的研究.但是目前针对自然场景文字检测与识别的相关研究主要集中在中英文中,且取得了不错的成绩.在自然场景文字检测的研究中,对于水平文本的检测,Tian等人提出的CTPN^[1]网络,该网络结合了CNN与LSTM深度网络,该网络可以学习字符的上下文关系从而提高文本检测的准确率.对于倾斜文本的检测,Shi等人提出了SegLink网络^[2],该网络首先将文本分解成片段和链接这两个局部可检测的元素,然后通过端到端训练的全卷积神经网络,可以在多个尺度上密集地检测这两个元素,最终通过邻近连接方法将各小文字块连接成单词.Yang等人提出了IncepText网络^[3],该网络提出了一种基于变形的PSROI池化模块来处理多方向的文本检测,能更好的提高大尺度以及扭曲变形文字图像的文本检测正确率,使模型具有更好的鲁棒性.随后也有一些研究人员提出了DMPNet网络^[4]、EAST网络^[5]、PixelLink网络^[6],支持倾斜文本的检测.Liao等人提出了一个可微二值化模块^[7],它可以在分割网络中执行二值化过程,通过与DB模块的优化,分割网络可以自适应地设置二值化阈值,既简化了后处理,又提高了文本检测的性能.在自然场景文字识别的研究中,对于规则文本的识别,应用最广泛的主要是Shi等人提出的一种基于序列的文字识别网络CRNN^[8],该网络能够处理任意长度的序列,且不涉及字符分割和水平尺度归一化,能进行端到端的训练.对于非规则文本的识别,主要是Shi等人提出的RARE网络^[9],该网络由空间变换网络STN和序列识别网络SRN组成,网络模型能够识别多种类型的不规则文本,包括透视文本和弯曲文本.

目前自然场景下的中英文检测与识别相关的研究已经取得了不错的成果,但是自然场景下乌金体藏文的检测与识别的相关研究仍处于起步阶段.为了给藏文识别的相关研究奠定基础,早期一些研究人员主要研究了印刷体和手写体藏文的识别.在印刷体和手写体藏文识别的研究中,一些研究人员针对藏文字符的特点在藏文字符轮廓特征提取^[10,11]、印刷体藏文字符特征提取^[12,13]、手写体藏文字符特征提取^[14]等方面取得了不错的成果,随后相关研究人员在此基础上进一

步研究了印刷体和手写体藏文的识别^[15-17].但是在印刷体和手写体藏文识别的研究中主要是采用传统的图像处理方法对藏文文本图像进行二值化、去噪等预处理操作,再对图像中的藏文字符进行切分,然后将切分后的藏文字符图片进行归一化后送入特征提取模块提取单个藏文字符的特征,最后通过传统的模式分类方法得到识别结果.从中可以看出传统的藏文识别方法存在多个中间环节,同时要求藏文图像质量较好且能切分出藏文字符.而自然场景下的乌金体藏文图像质量较差且文字存在透视变换、噪声干扰大,传统的藏文识别方法显然不适合自然场景下乌金体藏文的识别.

目前,得益于深度学习技术的发展,利用深度学习技术对自然场景下藏文的检测与识别进行研究是当下的发展趋势.赵栋材^[18]在传统的文字识别方法基础上,增加了基于BP网络的训练方法,通过对大量的木刻藏文经书字符的训练,实现了普通干扰情况下木刻经文的识别.梁弼等人^[19]提出了一种基于HMM分类器的联机手写藏文识别的方法,并设计了3种不同的HMM分类器对藏文字丁进行识别.王维兰等人^[20]提出了一种新的多层联系子层递归神经网络模型并融合藏文字丁的空间结构特征来进行联机手写藏文识别,实验结果说明该方法可以更好地表征藏文字的各笔划特征以及笔划间的空间结构关系.虽然这些学者在早期利用深度学习技术对木刻藏文经书、藏文字丁、手写体藏文进行了研究,但是公开发表的自然场景下藏文的识别文章只有仁青东主^[21]等人研究的基于深度学习的自然场景下藏文的识别,在该研究中虽然采用了基于深度学习的序列识别方法CRNN^[8]与二维串识别技术的结合,但是在该研究中对自然场景藏文文本行进行识别时仍然要对藏文文本行图像进行块切分,并没有完全体现出CRNN算法的优势,同时在该研究中对自然场景下藏文的检测只是一笔带过,没有做藏文检测实验.

因此本文在没有开源的自然场景乌金体藏文图像的情况下,人工收集和标注了自然场景乌金体藏文图像数据集,并在此基础上探究适合自然场景乌金体藏文检测与识别方法.本文在自然场景乌金体藏文检测阶段对比了不同的场景文字检测算法,最后采用可微分的二值化网络DBNet^[7],在乌金体藏文识别阶段采用基于序列的端到端文字识别算法CRNN^[8],在该阶段采用算法合成的全藏字作为字典库,并采用修改后的

MobileNetV3 Large^[22] 作为特征提取网络. 最后在该方法上做了基础性实验, 实验表明在自然场景下乌金体藏文的识别中 CRNN 算法可以进行端到端的识别, 并不需要对藏文文本行进行块切分, 简化了识别流程.

2 算法思想

本文中自然场景下乌金体藏文的识别系统主要分为两个阶段, 一是乌金体藏文文本检测, 在该阶段中采用可微分的二值化网络 DBNet^[7]; 二是乌金体藏文文本识别, 在该阶段中采用基于序列的文字识别算法 CRNN^[8].

2.1 乌金体藏文文本检测理论

如图 1 所示, 在乌金体藏文检测阶段将图片输入网络后, 经过特征提取网络 ResNet-18 提取特征和上采样融合并通过 con-cat 操作后得到图 1 中蓝色的特征图 F , 然后使用 F 预测出概率图 P 和阈值图 T , 最后通过可微分的二值化算法计算出近似二值图 \hat{B} , 从而得出乌金体藏文检测结果. 在乌金体藏文检测阶段, 采用了 Liao 等人^[7] 在中英文检测实验中在传统的 ResNet-18 网络^[23] 的 conv_3_x, conv_4_x, conv_5_x 后加入可变形卷积增大感受野的范围, 以此来适应纵横比较大的乌金体藏文检测.

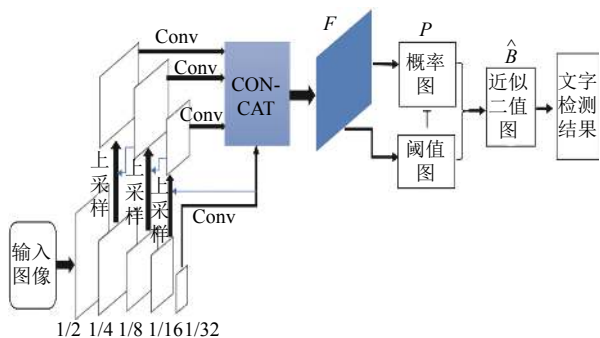


图 1 DBNet 网络结构图

乌金体藏文检测阶段中可微分的二值化过程如式 (1) 所示, 其中 \hat{B} 表示近似的二值图, (i, j) 表示概率图中的坐标, P 、 T 是网络学习的概率图、阈值图, k 是一个梯度缩放因子. 式 (1) 之所以会改善网络性能, 可以从梯度的反向传播来解释, 定义一个 $f(x)$ 如式 (2) 所示, 其中 $x = P_{i,j} - T_{i,j}$. 在使用交叉熵损失的情况下, 正样本的损失记为 L_+ 如式 (3), 负样本的损失记为 L_- 如式 (4). 正、负样本的损失对于输入 x 的偏导数分别为式

(5)、式 (6) 所示. 从微分式子可以看出 k 是梯度增益因子, 梯度对于错误预测的增益幅度很大.

$$\hat{B}_{i,j} = \frac{1}{1 + e^{-k(P_{i,j} - T_{i,j})}} \quad (1)$$

$$f(x) = \frac{1}{1 + e^{-kx}} \quad (2)$$

$$L_+ = -\ln \frac{1}{1 + e^{-kx}} \quad (3)$$

$$L_- = -\ln \left(1 - \frac{1}{1 + e^{-kx}} \right) \quad (4)$$

$$\frac{\partial L_+}{\partial x} = -kf(x)e^{-kx} \quad (5)$$

$$\frac{\partial L_-}{\partial x} = kf(x) \quad (6)$$

网络训练的损失函数如式 (7) 所示, L_s 为概率图的损失、 L_b 为二值图的损失、 L_t 为阈值图的损失. 其中 α 和 β 的取值分别为 1.0 和 10. 其中 L_s 和 L_b 采用二值交叉熵损失如式 (8) 所示, 其中 S_l 是经过采样的数据集. L_t 采用的是 L_1 损失如式 (9) 所示, 其中 R_d 为标注框经过扩充后得到的框里的一组像素的索引.

$$L = L_s + \alpha \times L_b + \beta \times L_t \quad (7)$$

$$L_s = L_b = \sum_{i \in S_l} y_i \log_2 x_i + (1 - y_i) \log_2 (1 - x_i) \quad (8)$$

$$L_t = \sum_{i \in R_d} |y_i^* - x_i^*| \quad (9)$$

2.2 乌金体藏文文本识别理论

在乌金体藏文识别阶段中首先根据检测结果裁剪出乌金体藏文文本块, 如图 2 所示将裁剪出的乌金体藏文文本块送入 MobileNetV3 Large 网络提取文字特征, 然后使用 LSTM 对特征序列中的每帧进行预测, 最后采用 CTC 进行转录并输出最后的识别结果.

因为藏文不仅可以横向叠加, 还可以纵向叠加, 所以在乌金体藏文识别阶段借鉴 Du 等人^[24] 在中英文识别实验中 MobileNetV3 Large 网络的配置, 该配置修改了原始 MobileNetV3 Large 中除第一个下采样特征图的步幅, 将步幅为 (2, 2) 修改为 (2, 1), 将第二个下采样特征图的步幅从 (2, 1) 修改为 (1, 1), 以此来保留更多的水平和垂直方向上的信息. 该网络配置如表 1 所示, 除了修改下采样特征图的步幅, 该网络配置也调整了最大池化层的窗口大小和步幅, 将原来 7×7 的窗口改

为 2×2 ,步幅由1改为2.表1中exp size表示膨胀参数;SE表示Squeeze-and-Excite结构;NL表示非线性,HS表示h-swish激活函数,RE表示ReLU激活函数;s表示步幅;Bneck模块中包括 1×1 卷积、 3×3 卷积或者 5×5 卷积、SE结构.

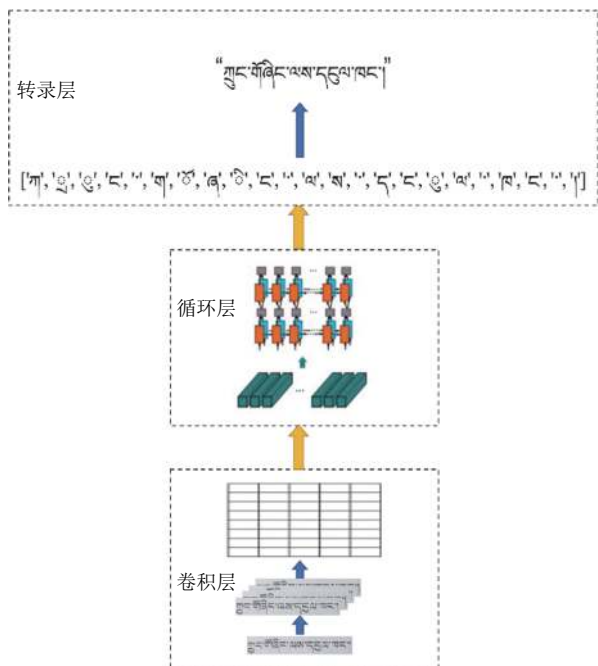


图2 CRNN 结构图

表1 MobileNetV3 Large 网络配置表

Operator	exp size	out	SE	NL	s
Conv2d, 3×3	—	16	—	HS	2
Bneck, 3×3	16	16	—	RE	1
Bneck, 3×3	64	24	—	RE	(2, 1)
Bneck, 3×3	72	24	—	RE	1
Bneck, 5×5	72	40	√	RE	(2, 1)
Bneck, 5×5	120	40	√	RE	1
Bneck, 5×5	120	40	√	RE	1
Bneck, 3×3	240	80	—	HS	1
Bneck, 3×3	200	80	—	HS	1
Bneck, 3×3	184	80	—	HS	1
Bneck, 3×3	184	80	—	HS	1
Bneck, 3×3	480	112	√	HS	1
Bneck, 3×3	672	112	√	HS	1
Bneck, 5×5	672	160	√	HS	(2, 1)
Bneck, 5×5	960	160	√	HS	1
Bneck, 5×5	960	160	√	HS	1
Conv2d, 1×1	—	960	—	HS	1
Max Pool, 2×2	—	—	—	—	2

用式(10)来表述训练数据集,其中 I_i 表示训练图片, l_i 表示图像中藏文音节序列的ground truth.目标函

数的构建是通过最小化ground truth的条件概率的负对数似然函数,如式(11)式所示.其中 y_i 是循环层和卷积层从 I_i 生成的序列,该目标函数直接计算输入图像和它对应的ground truth之间的损失函数,因此该网络可以进行输入图像与真实的标签序列的端到端的训练.

$$x = \{I_i, l_i\} \tag{10}$$

$$O = - \sum_{l_i, l_i \in x} \log_2 p(l_i | y_i) \tag{11}$$

3 实验与分析

由于缺乏开源的自然场景乌金体藏文文本图像数据集,本实验中网络训练所用的数据集全是在拉萨街头人工拍摄的乌金体藏文文本图像.这些乌金体藏文文本图像主要包括了交通标志牌、店铺广告牌、公交站牌等自然场景乌金体藏文文本图像.最终收集了449张乌金体藏文文本图像,4321张自然场景乌金体藏文文本块图像.

3.1 乌金体藏文检测

在自然场景乌金体藏文检测实验中,将前期收集的449张自然场景乌金体藏文图像尺寸统一为 608×608 ,然后将这449张图片进行数据增强最终扩充为1796张,并将这1796张图像随机分为训练集1600张和测试集196张,使用Labelme软件对实验数据集进行了标注.在此基础上对比了CTPN^[1]、EAST^[5]场景文字检测算法与本文所采用的可微分的二值化网络DBNet在乌金体藏文检测上的效果,其中成功检测到乌金体藏文的部分检测效果如图3所示,图3(a)为CTPN检测效果,图3(b)为EAST检测效果,图3(c)为DBNet检测效果.在测试集上的性能测试结果如表2所示.



图3 成功检测到乌金体藏文的部分图片展示

因为在本文收集的自然场景乌金体藏文检测实验数据集图像中很多乌金体藏文文本行都是倾斜文本,完全水平的乌金体藏文文本较少.从图3中可以看出,

本文所采用的可微分二值化网络 DBNet 在真实自然场景下乌金体藏文的检测效果最好,且能检测出倾斜文本. CTPN 网络只对水平文本检测效果较好,不适合倾斜文本的检测. 而 EAST 算法能检测倾斜文本,但是在实际的测试中出现很多误检. 所以在表 2 的性能测试中,CTPN 网络的性能较差,而相比于 EAST 网络本文所采用的检测网络的准确率最高.

表 2 几种检测算法在乌金体藏文文本检测实验上的性能对比

实验方法	Precision	Recall	F1
CTPN	0.45	0.34	0.39
EAST	0.83	0.64	0.72
DBNet	0.89	0.59	0.71

3.2 乌金体藏文识别

在自然场景乌金体藏文识别实验中,将前期收集的 4321 张自然场景下乌金体藏文文本图像块尺寸统一为 32×100,采用算法合成的全藏字作为转录的字典,其中包含 18785 个藏文音节. 在此基础上依次进行了 3 组实验,但是该实验无法与现有相关工作做对比,主要有以下两个方面的原因,一方面因为与自然场景下乌金体藏文识别相关的论文较少,无法与其他研究成果进行比较;另一方面因为目前没有开源的自然场景藏文数据集,都是基于自制的数据集进行实验,没有可比性. 这 3 组实验验证了本文采用的方法在自然场景乌金体藏文识别上的可行性. 网络训练结束后在测试集上的性能测试如表 3 所示.

表 3 不同特征提取网络在乌金体藏文文本识别实验上的性能对比

特征提取网络	数据集组合形式	识别准确率
MobileNetV3 Large	训练集2000张	0.0397
	测试集126张	
ResNet-34	训练集4195张	0.3333
	测试集126张	
MobileNetV3 Large	训练集4195张	0.4365
	测试集126张	

采用 3.1 节中 DBNet 网络训练出的乌金体藏文检测模型和实验 3 训练出的乌金体藏文识别模型在 314 张自然场景乌金体藏文图像上的部分正确检测到乌金体藏文文本并完全识别正确的结果可视化如图 4 所示,图 4(a) 的识别结果为“ལུ་ཁྱེད་མཉེ་མཉེ་ཚོ།”;图 4(b) 的识别结果为“གཉེས་ཀྱི་འོད་སྤྲོད་ལྷན་ཁང་།”;图 4(c) 的识别结果

为“ལང།”和“ལུ་ཁྱེད་མཉེ་མཉེ་ཚོ།”. 但是其中也存在一些错误识别的结果,部分特殊示例如图 5 和图 6 所示. 图 5(a) 的识别结果为“ཉེས་ཀྱི་འོད་སྤྲོད་ལྷན་ཁང་།”,第 3 个音节的基字识别错误,正确的应为“ཉ”;第 5 个音节中“ལྷ”不符合藏字构字规则;图 5(b) 的识别结果为“ཉེས་ཀྱི་འོད་སྤྲོད་ལྷན་ཁང་ཚོ་ལང།”,第 2 个音节错误,应为“ཉེས”,其中“ལྷ”是特殊藏字,第 5 个音节错误,“ལྷ”不符合藏字构字规则;图 5(c) 的识别结果为“ལྷན་ཁང་ལྷན་ཁང་ལྷན་ཁང་།”,第 3 个音节后加字下多了一个构件,第 4 个音节“ལྷན་ཁང་”不符合藏字构字规则. 图 6(a) 的识别结果为“ལུ་ཁྱེད་མཉེ་མཉེ་ཚོ་ལང།”,第 1 音节少了下加字“ལ”,第 2 音节少了下加字“ལ”;图 6(b) 的识别结果为“ལུ་ཁྱེད་མཉེ་མཉེ་ཚོ་ལང།”,第二个音节错误,缺下加字,应为“ལྷ”;图 6(c) 的识别结果为“ལུ་ཁྱེད་མཉེ་མཉེ་ཚོ་ལང།”和“ལུ་ཁྱེད་མཉེ་མཉེ་ཚོ་ལང།”,后者中第 4 个音节缺下加字.



(a) 示例 1 (b) 示例 2 (c) 示例 3

图 4 正确的检测结果和识别结果示例



(a) 示例 a (b) 示例 b (c) 示例 c

图 5 识别结果错误示例 1



(a) 示例 a (b) 示例 b (c) 示例 c

图 6 识别结果错误示例 2

从表 3 中可以看出,随着数据集规模的增大,最终在测试集上的识别准确率有了很大的提升. 在实验 2 和实验 3 中采用了不同特征提取网络进行实验,实验结果表明实验 3 采用的特征提取网络在乌金体藏文识

别上的效果更好. 通过分析在 314 张自然场景下乌金体藏文图像上的实际识别效果, 从中可以发现本文训练出的模型对结构简单的藏文字符能够识别正确, 如图 4 所示, 证明了该方法在自然场景乌金体藏文识别上的可行性. 但是由于实验数据集规模太小, 实际的测试结果中也存在一些识别错误的情况, 部分示例如图 5 和图 6 所示. 从中发现了一些特殊的现象如下: (1) 识别错误的结果中有些音节如“ལྷོ་”、“ལྷོ་”、“བཙམས་”等不符合藏字构字规则, 如图 5 所示; (2) 下加字漏识别现象比较常见, 其中音译词“ལྷོ་”也会识别错误, 如图 6 所示; (3) 出现特殊藏字“ལྷོ་”时会识别错误; (4) 一些结构极其相似的藏文字符如“བ”、“བ”、“བ”以及“ཨ”、“ཨ”等在识别时容易出错. 其中现象 (2) 和现象 (3) 是因为特殊藏字“ལྷོ་”和音译词“ལྷོ་”不符合全藏字中藏字的构字规则, 但也属于藏字. 现象 (1) 和现象 (4) 是因为训练集规模太小, 网络没有学习到复杂结构藏字和结构相似藏文字符的特征.

4 总结与展望

本文采用基于残差网络 ResNet-18 的可微分的二值化网络 DBNet 用于自然场景乌金体藏文的检测, 采用基于序列的识别算法 CRNN 用于自然场景乌金体藏文的识别, 在人工收集的乌金体藏文图像数据集的基础上进行了初步的实验, 实验表明了该方法在乌金体藏文检测与识别上具有可行性. 通过分析实验结果在现有方法的基础上可以改善的方面主要有以下几点: (1) 根据藏文构字特点, 在 CRNN 的转录层中如何将藏文构字规则引入其中, 以避免识别结果中出现不符合藏字构字规则的情况; (2) 完善词典, 加入特殊藏字以及音译词; (3) 如何合成多样化的自然场景乌金体藏文本块以扩充训练集提高网络的识别准确率.

参考文献

- 1 Tian Z, Huang WL, He T, *et al.* Detecting text in natural image with connectionist text proposal network. Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 56–72.
- 2 Shi BG, Bai X, Belongie S. Detecting oriented text in natural images by linking segments. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017. 3482–3490.
- 3 Yang QP, Cheng ML, Zhou WM, *et al.* Inceptext: A new inception-text module with deformable psroi pooling for multi-oriented scene text detection. arXiv: 1805.01167, 2018.
- 4 Liu YL, Jin LW. Deep matching prior network: Toward tighter multi-oriented text detection. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017. 3454–3461.
- 5 Zhou XY, Yao C, Wen H, *et al.* East: An efficient and accurate scene text detector. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017. 2642–2651.
- 6 Deng D, Liu HF, Li XL, *et al.* Pixellink: Detecting scene text via instance segmentation. Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence. New Orleans: AAAI, 2018.
- 7 Liao MH, Wan ZY, Yao C, *et al.* Real-time scene text detection with differentiable binarization. Proceedings of the AAAI Conference on Artificial Intelligence. New York: AAAI, 2020. 11474–11481.
- 8 Shi BG, Bai X, Yao C. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(11): 2298–2304. [doi: 10.1109/TPAMI.2016.2646371]
- 9 Shi BG, Wang XG, Lv PY, *et al.* Robust scene text recognition with automatic rectification. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016. 4168–4176.
- 10 党红恩. 藏文字形轮廓特征优化识别提取仿真研究. 计算机仿真, 2016, 33(11): 341–344. [doi: 10.3969/j.issn.1006-9348.2016.11.074]
- 11 孙淑娟, 房培玉. 基于蚁群算法的现代藏文字符轮廓提取技术研究. 微计算机应用, 2008, 29(5): 38–41.
- 12 普次仁. 多种印刷字体藏文字符的特征提取方法研究. 西藏大学学报(自然科学版), 2008, 23(1): 25–28.
- 13 刘真真, 王茂基, 李永忠, 等. 基于分形矩的印刷体藏文特征提取方法. 模式识别与人工智能, 2008, 21(5): 654–657. [doi: 10.3969/j.issn.1003-6059.2008.05.013]
- 14 朱利娟, 云中华, 边巴旺堆. 基于极坐标变换的脱机手写藏文字符特征提取方法. 计算机应用与软件, 2018, 35(3): 162–166. [doi: 10.3969/j.issn.1000-386x.2018.03.031]
- 15 王维兰, 丁晓青, 陈力, 等. 印刷体现代藏文识别研究. 计算机工程, 2003, 29(3): 37–38, 94. [doi: 10.3969/j.issn.1000-3428.2003.03.014]
- 16 李旺平. 藏文 OCR 识别方法研究 [硕士学位论文]. 武汉:

- 武汉理工大学, 2017.
- 17 李刚. 印刷体藏文识别系统的研究与实现 [硕士学位论文]. 上海: 上海交通大学, 2007.
- 18 赵栋材. 基于 BP 网络的木刻藏文经书文字识别研究. 微处理机, 2012, (5): 35–38, 43. [doi: [10.3969/j.issn.1002-2279.2012.05.011](https://doi.org/10.3969/j.issn.1002-2279.2012.05.011)]
- 19 梁弼, 王维兰, 钱建军. 基于 HMM 的分类器在联机手写藏文识别中的应用. 微电子学与计算机, 2009, 26(4): 98–101, 104.
- 20 王维兰, 陈万军. 基于笔划特征和 MCLRNN 模型的联机手写藏文识别. 计算机工程与应用, 2008, 44(14): 91–93, 194. [doi: [10.3778/j.issn.1002-8331.2008.14.025](https://doi.org/10.3778/j.issn.1002-8331.2008.14.025)]
- 21 仁青东主, 尼玛扎西. 基于深度学习的自然场景藏文识别研究. 高原科学研究, 2019, 3(4): 96–103.
- 22 Howard A, Sandler M, Chen B, *et al.* Searching for MobileNetV3. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE, 2019. 1314–1324.
- 23 He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016. 770–778.
- 24 Du YN, Li CX, Guo RY, *et al.* PP-OCR: A practical ultra lightweight OCR system. arXiv: 2009.09941, 2020.