

# 基于深度学习的藏文现代印刷物版面检测技术研究

吴燕如<sup>1,2</sup>, 珠杰<sup>1,2</sup>, 管美静<sup>1,2</sup>

(1. 西藏大学信息科学技术学院, 西藏 拉萨 850000)

(2. 藏文信息技术国家地方联合中心, 西藏 拉萨 850000)

**[摘要]** 针对藏文现代图书版面中的文本行分布不均匀、现代藏文字体差异较大的问题, 提出了一种基于 Faster R-CNN 的版面文本行检测算法. 通过在整理标注的数据集上训练, 用 ResNet-50 网络提取出藏文现代图书版面特征信息. 为了有效提高模型的泛化能力, 在 COCO 数据集下的网络模型中进行迁移学习. 实验结果表明, 该方法可对藏文现代印刷物的版面实现文本行的定位, 检测准确率为 83%, 召回率为 95%, 明显提高了版面检测的精确度.

**[关键词]** 深度学习, 藏文现代印刷物, Faster R-CNN, 版面检测

**[中图分类号]** TP391 **[文献标志码]** A **[文章编号]** 1672-1292(2021)01-0044-05

## Research on Layout Inspection Technology of Modern Tibetan Prints Based on Deep Learning

Wu Yanru<sup>1,2</sup>, Zhu Jie<sup>1,2</sup>, Guan Meijing<sup>1,2</sup>

(1. School of Information Science and Technology, Tibet University, Lhasa 850000, China)

(2. National and Local Joint Center for Tibetan Information Technology, Lhasa 850000, China)

**Abstract:** Aimed at the uneven distribution of text lines in the layout of modern Tibetan books and the large differences in modern Tibetan fonts, a layout text line detection algorithm based on Faster R-CNN is proposed. By training on collated and labeled data set, we use the ResNet-50 network to extract the feature information of the Tibetan modern book layout. In order to effectively improve the generalization ability of the model, transfer learning is performed in the network model under the COCO dataset. The experimental results show that this method can realize text line positioning on the layout of modern Tibetan printed materials, with a detection accuracy rate of 83% and the recall rate of 95%, which significantly improves the accuracy of layout detection.

**Key words:** deep learning, modern Tibetan prints, Faster R-CNN, layout detection

近年来, 国家高度重视藏文化资源的保护和珍藏<sup>[1]</sup>. 优秀的藏文化资源中藏文现代印刷物是重要的保存对象. 从藏文印刷物中检测版面信息对于藏文化实现数字化存储具有重要意义<sup>[2]</sup>. 目前藏文印刷物版面分辨率较低, 版面中文本行也比较密集, 增加了版面检测的难度.

当前国内外对中文和英文中文本区域检测已经有了一定的研究, Epshtein 等<sup>[3]</sup>提出了笔画宽度变换的文本检测算法, Pan 等<sup>[4]</sup>提出让 MSER 和卷积神经网络相结合的检测方法, 但这些方法均不能有效解决文本分辨率较低的问题. Zhu 等<sup>[5]</sup>提出了使用训练出的级联强分类器对图像中的滑动窗口进行分类, 实现文本区域的检测, 该方法虽然提高了检测精度, 但增加了训练难度. 在现有的研究中, 对藏文现代印刷物版面检测还相对较少, 但对于中英文自然场景下的文本检测和物体检测的研究已经比较成熟, 取得了不错的成效. 因此, 本文利用 Faster R-CNN 检测算法研究藏文现代印刷物的版面检测问题.

深度学习方法本身具有较强的非线性拟合能力, 在计算机视觉领域得到了广泛应用<sup>[6]</sup>. 基于深度学习的目标检测方法对网络结构不断改进, 主要形成了 R-CNN 检测系列<sup>[7]</sup>和单阶段检测系列<sup>[8]</sup>, 前者主要

收稿日期: 2020-08-08.

**基金项目:** 西藏大学研究生“高水平人才培养计划”项目(2017-GSP-131)、西藏自治区高等教育教学改革研究重点项目、多学科融合的新工科创新创业教育体系研究项目、藏语文传承与发展之藏汉双向机器翻译平台建设项目、计算机及藏文信息技术国家团队及重点实验室建设项目(藏大财指[2018]81号)、国家重点研发计划重点专项(2017YFB140220).

**通讯作者:** 珠杰, 博士, 教授, 博士生导师, 研究方向: 藏文信息处理、数据挖掘. E-mail: 790139756@qq.com

是基于候选区域的方法,后者借鉴了回归的思想. 2013 年, GIRSHICK 等<sup>[9]</sup>提出 R-CNN 检测算法,实现了将神经网络的方法应用到目标检测上. 2015 年, GIRSHICK<sup>[10]</sup>又提出了 Fast R-CNN 算法,主要是在 R-CNN 和 SPP-Net 检测算法的基础上加以改进. Faster R-CNN 网络实现了用神经网络的方法提取建议区域<sup>[11]</sup>,有效减少了需要计算的特征,加快了检测速度和精确度. 单阶段检测方法主要有 YOLO<sup>[12]</sup>和 SSD 方法<sup>[13]</sup>,直接通过特征图得到类别得分和位置.

实际应用中,R-CNN 系列检测速度虽然没有单阶段方法快,但检测准确率较高<sup>[14]</sup>. 本文选取 Faster R-CNN 模型<sup>[15]</sup>作为藏文现代印刷物中版面的定位方法,在手工整理的藏文现代图书版面数据集上划分训练集和测试集,通过增加候选框的数量,作为文本区域的定位方法.

1 藏文现代印刷物数据集的构建

本文选取一部分藏文现代图书做为原图像,样本具有文字区域多而其他类别区域相对较少的特点,只对现代图书版面中的文本行区域进行检测. 生成的样本库有 1 320 张图片,图片像素较低的为 374 \* 541,像素较高的为 876 \* 1 300,图片中包含的文本行个数在 5-40 之间. 具体藏文现代图书版面示例如图 1 所示.

藏文图书版面搜集整理之后,通过人工对数据集进行标注. 使用 labellmg 数据标注工具,对整理的数据集完成标注,制作的数据集格式均为 Pascal Voc 格式. 标注出每一部分的文本行所在的最小外接矩形,并标注出类别标签,作为网络训练中评估的参考标准.

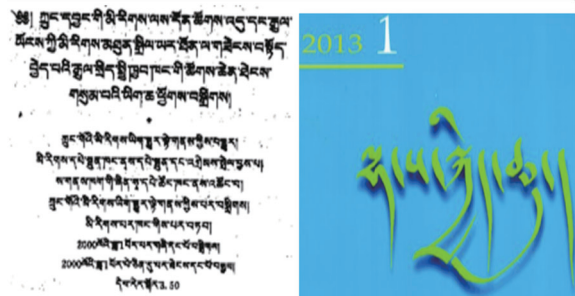


图 1 采集到的藏文现代图书示例  
Fig. 1 Examples of collected Tibetan modern books

2 Faster R-CNN 的藏文现代图书文本区域定位算法

Faster R-CNN 检测方法在结构上主要由 3 个部分组成:特征提取、RPN 网络、ROI Pooling. 具体流程如图 2 所示.

2.1 特征提取

在实现过程中采用经典的 ResNet-50 网络,通过 5 部分卷积操作、2 次池化操作、3 层全连接层,最后由 softmax 完成整个输出,得到整张图片的特征. 这样避免了特征的重复计算,加快了训练速度. 卷积层提取到的特征图用于后续网络的输入.

2.2 RPN 网络

RPN 网络和 SelectSearch 一样都是用来生成候选框,但传统方法生成的候选框数量较多,需要时间较长. RPN 网络中只包含卷积层,该网络的位置在 Conv5-3 之后,用神经网络的方法大大提高了候选框的生成速度. 针对藏文现代图书的定位问题,在 Conv5-3 特征图上采用大小为 3 \* 3 的 filter,设置为步长 1 的滑动卷积,这样每个窗口就映射成一个 256 维的向量. 256 维向量并行进入全连接层,分别对滑动窗口生成的建议区域进行分类和回归.

对卷积特征图上的每个像素点设置 20 种不同的候选窗口,根据藏文现代图书中文本行大小长短的不同,经改进使用 64 \* 64、128 \* 128、256 \* 256、512 \* 512 的窗口面积,每个面积下设置 5 种不同的缩放,比例分别为 1:2、1:5、1:1、2:1、5:1,这样就生成了 20 个尺度的候选框,这样分类层对于一个像素点生成的候选框可以生成 40 个得分,用来判断候选框包含目标或者不含有目标的概率. 回归层对于每个像素点生成的候选框共产生 80 个位置坐标,再用非极大值抑制的方式对生成候选框进行筛选,用回归方法对候选框位置进行调整,得到更精确的建议区域. RPN 网络产生的损失如式(1)所示:

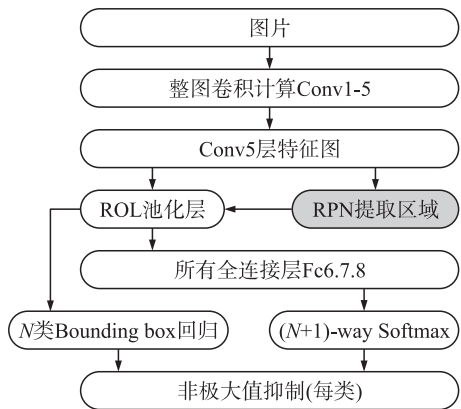


图 2 Faster R-CNN 检测流程图  
Fig. 2 The detection flow Chart of Faster R-CNN

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{\text{cls}}} \sum L_{\text{cls}}(p_i, p_i^*) + \frac{1}{N_{\text{reg}}} \sum_i p_i^* L_{\text{reg}}(t_i, t_i^*), \quad (1)$$

式中,  $i$  为第  $i$  个锚点框;  $p_i$  为锚点框为目标的概率值;  $p_i^*$  为锚点框对应的真实框的类别;  $t_i$  为预测框的坐标位置;  $t_i^*$  为对应的真实框的坐标位置;  $L_{\text{cls}}$  为类别损失, 即判断预测的建议区域是前景或背景的概率, 实质上是一个二分类问题, 其定义如式(2)所示:

$$L_{\text{cls}}(p_i, p_i^*) = -\log[p_i^* p + (1 - p_i^*)(1 - p_i^*)]. \quad (2)$$

$L_{\text{reg}}$  为回归部分的损失, 具体定义如式(3)<sup>[16]</sup>所示:

$$L_{\text{reg}}(t_j, t_j^*) = \sum_{x,y,w,h} \text{smooth}_{L1}(t_i - t_i^*). \quad (3)$$

### 2.3 ROI Pooling

RPN 网络生成的候选区域对应映射在特征图上, 形成的映射区域均被划分为  $7 * 7$  大小的子图, 这样不同大小的建议区域被转化为相同大小的感兴趣池化图<sup>[17]</sup>, 并进入全连接层, 用 softmax 对其类别进行预测, 并对边框位置进行回归, 获得更精确的边框位置. 该过程的损失仍是分类损失和回归损失, 整体损失定义如式(4)<sup>[18]</sup>所示:

$$L(p, u, t, v) = L_{\text{cls}}(p, u) + \lambda \mu L_{\text{loc}}(t, v), \quad (4)$$

式中,  $u$  为感兴趣区域所属的类别;  $p$  为属于类别的概率值;  $t$  为建议框的位置坐标;  $v$  为对应的真实框的位置坐标.

## 3 实验结果与分析

### 3.1 实验环境及数据

本文实验硬件环境为 intel i7 处理器, 运行内存 32G, 显卡为 NVIDIA GeForce RTX2080, 操作系统为 Windows10 平台, 实验采用 TensorFlow 框架, Python 语言, 采用 Labellmg 软件对藏文现代图书进行手动标注. 实验采用了藏文图书 1 200 张作为训练集, 120 张作为测试集.

### 3.2 实验评估指标

本文采用准确率  $P$  (precision)、召回率  $R$  (recall) 和  $F$ -值对实验结果进行评估<sup>[19]</sup>. 准确率是识别正确的框数量占有所有识别到的框数量的比例, 召回率是识别正确的框数量占有所有真实框数量的比例, 准确率  $P$ 、召回率  $R$ 、 $F$ -值的具体定义分别如下所示:

$$P = \frac{TP}{TP + FP}, \quad (5)$$

$$R = \frac{TP}{TP + FN}, \quad (6)$$

$$F\text{-值} = \frac{2 * P * R}{P + R}, \quad (7)$$

式中,  $TP$  为正确识别的框的个数;  $FP$  为检测错误的框的个数;  $FN$  为正样本漏检的个数.

### 3.3 改进的 Faster R-CNN 网络训练

改进的 Faster R-CNN 网络在训练过程中使用的初始化参数来自 COCO 数据集的预训练模型<sup>[20]</sup>. 训练中学习率初始化为 0.001, 衰减系数为 0.94, 动量值为 0.89, 总迭代次数为 50 000. 在相同的实验条件下与 SSD 检测模型训练过程的损失进行对比, 查看训练过程的日志文件可以看出实验过程中的损失变化, 具体的损失曲线如图 3 所示.

由图 3 可知, 随着训练次数的增加, 网络训练的损失不断降低. 藏文现代图书版面在 40 000 次迭代后开始收敛; 当完成 50 000 次迭代时, 藏文现代图书的训练损失率降至最低值 0.82, 损失基本趋于稳定. 该数据集在 SSD 模型训练过程中损失不断降低, 当迭代至 30 000 次时, SSD 模型也处于收敛状态, 此时训练损失为 0.4. 可以看出, SSD 模型训练的收敛速度比改进的 Faster R-CNN 快很多.

### 3.4 藏文图书版面检测效果

采用改进后的 Faster R-CNN 对测试集进行测试, 典型的藏文现代图书版面中文本行的检测效果如



图 4 所示.

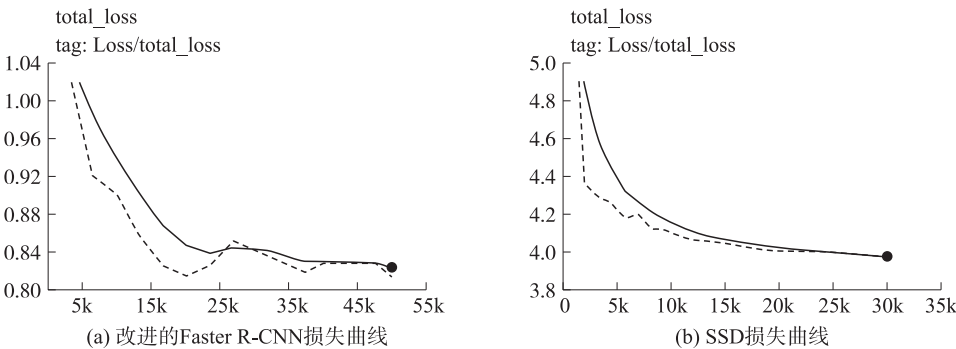


图 3 损失曲线图  
Fig. 3 Loss curve

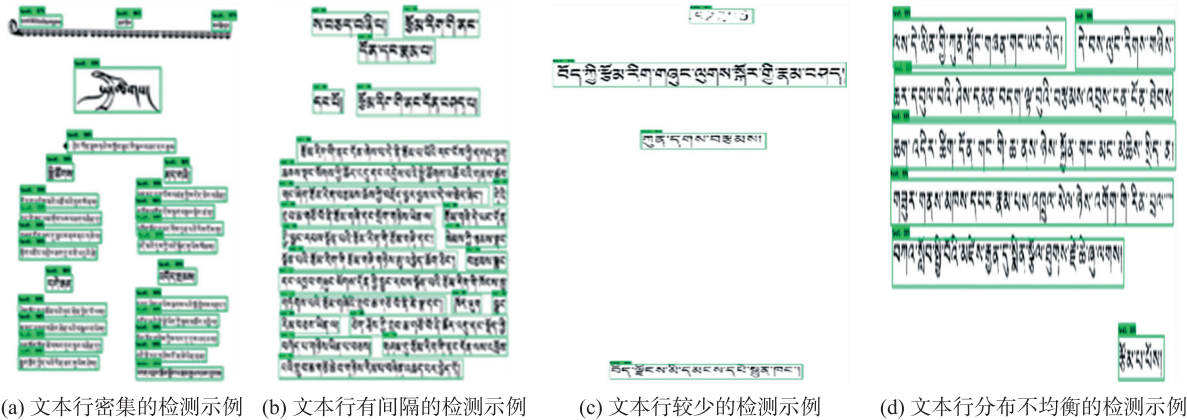


图 4 藏文现代图书版面检测效果示例  
Fig. 4 Example of detection effect of modern Tibetan book layout

由图 4 可知,矩形框所在的位置为预测框,每个矩形框对应一个预测的准确度.改进的 Faster R-CNN 不仅可有效检测出藏文现代图书中的文本行,还可检测出排版不同版面的文本行,检测效果并未受到文本行的长度、数量和整体文本行分布的影响.在字体样式差异较大的情况下,改进的 Faster R-CNN 也能有效识别文本行.

在改进的 Faster R-CNN 和 SSD 实验基础上,本文进行了原始的 Faster R-CNN 实验.3 种检测模型在该数据集上的检测性能对比如表 1 所示.

由对比可知,SSD 模型的准确率和召回率要比 Faster R-CNN 低很多,SSD 对较长的文本行和字体样式差异较大的文本行召回效果较差;原始的 Faster R-CNN 模型的准确

率和召回率都没有改进后的 Faster R-CNN 检测方法高.改进后的 Faster R-CNN 模型在本文的数据集上具有一定的准确率和召回率性能优势,相比原始的 Faster R-CNN、SSD 模型具有良好的应用效果.

为了验证改进后的方法在藏文现代图书数据集上的有效性,本文对改进的 Faster R-CNN 与 Faster R-CNN 模型应用在图像检测领域的性能进行了对比.文献[19]中 Faster R-CNN 对精密零部件检测,该实验最终准确率为 87.8%,召回率为 80.3%;文献[21]中 Faster R-CNN 对目标人物出现的位置进行检测,该实验最终在基础网络为 ResNet-101 的训练中准确率达到 94.2%,平均精度为 66.8%;文献[22]在基础网络为 ResNet-50 的训练中对蓝莓成熟果检测的准确率为 94%,而召回率只有 77%.由此可知,本文改进的 Faster R-CNN 模型在藏文现代图书数据集训练时的召回效果较好,整体性能较高.

4 结论

本文以藏文现代图书作为研究对象,建立了藏文现代图书标注的数据集,在深度学习的 TensorFlow 框

表 1 数据集在两种模型上的性能对比

Table 1 Performance comparison of the data set on two models

检测模型	准确率	召回率	F-值
原始的 Faster R-CNN	0.802 8	0.912 0	0.869 7
改进的 Faster R-CNN	0.830 0	0.953 6	0.887 6
SSD	0.753 6	0.829 0	0.791 1

架上训练 Faster R-CNN 检测网络,并用训练好的 COCO 数据集下的模型进行迁移学习.为了解决藏文现代图书版面中文本行分布不均匀的问题,本文采用了多个版面差异较大的数据集进行训练,并改变了原始的 Faster R-CNN 中 anchor 的面积和长宽比例,有效解决了数据集中文本行分布不均匀的检测问题.由实验结果可以看出:

(1)改进的 Faster R-CNN 在藏文现代图书版面的检测上,当图片中的文本行比较密集或文本行较为稀疏的情况下,相比 SSD 网络模型具有较好的检测效果;

(2)当版面中文本行信息较少的情况下,SSD 对长文本行的检测出现错误,改进的 Faster R-CNN 检测方法仍具有良好的检测效果;

(3)在训练中迭代次数相同时,SSD 模型的收敛速度远比改进的 Faster R-CNN 快,但检测准确率和召回率都没有改进的 Faster R-CNN 检测方法高.由此可知,改进后的 Faster R-CNN 对该数据集具有良好的适应性.

本文在实验过程中,只采用了藏文现代图书建立数据集,由于藏文数据集现有资源收集难度较大,实验并没有与其他类型的藏文现代印刷物的版面进行对比,在整个藏文印刷物版面数据集上没有很好的通用性,这是今后在实验过程中仍需进一步探索的问题.

### [参考文献] (References)

- [1] 索南草. 浅谈藏文典籍文化的传承与保护[J]. 时代教育, 2014(13): 116-117.
- [2] 张西群. 面向藏文历史文献的版面分割方法研究[D]. 北京: 北京工业大学, 2018.
- [3] EPSSTEIN B, OFEK E, WEXLER Y. Detecting text in natural scenes with stroke width transform[C]//IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE, 2010.
- [4] PAN Y F, HOU X W, LIU C L. A hybrid approach to detect and localize texts in natural scene images[J]. IEEE Trans Image Process, 2011, 20(3): 800-813.
- [5] ZHU J, CHEN X J, YUILLE A L, et al. DeePM: a deep part-based model for object detection and semantic part localization[C]//ICLR 2016. San Juan, Puerto Rico, 2016.
- [6] 张学鹏. 基于深度学习的图像语义分割方法研究与实现[D]. 成都: 电子科技大学, 2018.
- [7] CHEN C Y, LIU M Y, ONCEL T, et al. R-CNN for small object detection[C]//ACCV2016. Taipei, China, 2016.
- [8] 赵丹凤. 基于通用对象估计的目标检测与模糊车牌识别算法研究[D]. 南京: 南京邮电大学, 2016.
- [9] GIRSHICK R, LANDOLA F, DARRELL T, et al. Deformable part models are convolutional neural networks[C]//IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE Computer Society, 2015.
- [10] GIRSHICK R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE, 2015.
- [11] 张勋, 陈亮, 朱雪婷, 等. 基于区域卷积神经网络 Faster R-CNN 的手势识别方法[J]. 东华大学学报(自然科学版), 2019, 45(4): 559-563.
- [12] 李响, 苏娟, 杨龙. 基于改进 YOLOv3 的合成孔径雷达图像中建筑物检测算法[J]. 兵工学报, 2020, 41(7): 1347-1359.
- [13] 蒋强卫. 基于卷积神经网络的双目视觉物体识别与定位研究[D]. 哈尔滨: 哈尔滨工程大学, 2018.
- [14] 曾健. 基于深度学习的汽车门板焊点识别算法研究及应用[D]. 广州: 华南理工大学, 2019.
- [15] 张新, 郭福亮, 梁英杰, 等. 基于 R-CNN 算法的海上船只的检测与识别[J]. 计算机应用研究, 2020, 37(增刊 1): 314-315, 319.
- [16] 孙朝云, 裴莉莉, 李伟, 等. 基于改进 Faster R-CNN 的路面灌装封裂缝检测方法[J]. 华南理工大学学报(自然科学版), 2020, 48(2): 84-93.
- [17] 贺颖. 变换逼近理论指导下的卷积神经网络及其应用[D]. 唐山: 华北理工大学, 2019.
- [18] SHI J H, CHANG Y J, XU C H, et al. Real-time leak detection using an infrared camera and Faster R-CNN technique[J]. Computer & Chemical Engineering, 2020, 135: 106780.
- [19] 孙海铭, 时兆峰, 李晗, 等. 基于 Faster R-CNN 的精密零部件的识别方法[J]. 飞控与探测, 2020, 3(2): 26-36.
- [20] 王莹, 丁鹏. 基于深度学习的交通信号灯检测及分类方法[J]. 汽车实用技术, 2018(17): 89-91.
- [21] 周华平, 殷凯, 桂海霞, 等. 基于改进的 Faster R-CNN 目标人物检测[J]. 无线电通信技术, 2020, 46(6): 712-716.
- [22] 朱旭, 马洪, 姬江涛, 等. 基于 Faster R-CNN 的蓝莓冠层果实检测识别分析[J]. 南方农业学报, 2020, 51(6): 1493-1501.

[责任编辑: 严海琳]