

基于 Faster-RCNN 和改进 ResNet 实现的甲骨文识别算法

任启华, 郭珂言, 李浩然

2024 年 4 月 15 日

摘要

本文讨论了在学科融合发展盛行的当下, 如何将计算机与甲骨文研究相结合——通过人工智能高效实现甲骨文信息处理。在甲骨文的信息化处理研究中, 甲骨文分割与识别一直是一大难点, 在过去的数年间, 人们相继提出了许多种算法试图实现这一过程, 包括 VGG16, Inceptionv4, RCNN 等等, 但都存在一定的问題。甲骨文拓片的分割与识别受到多重干扰因素的影响, 常见的有点状噪声, 人工纹理等等。在这篇文章中, 我们主要采取了 Faster-RCNN 和改进的 ResNet 以及优异的图像预处理技术, 来有效的实现对甲骨文的分割与识别, 并且, 我们认为, 这种结合方法有很大的推广价值

代码网址: <https://github.com/vincent1234123/->, 里面有我们的项目代码, 省略了部分数据预处理的步骤

目录

1	问题的重述	1
2	问题的分析	1
2.1	问题一的分析	1
2.2	问题二和问题三的分析	1
2.3	问题四的分析	1
3	模型的建立与求解	2
3.1	预处理	2
3.2	拓片的分割	2
3.3	甲骨文的识别	3
4	结果分析与检验	3
5	模型的评价与推广	4
6	实验结果	5
7	其他工作	6
8	参考文献	8

1 问题的重述

如果将问题中的要求整合，则一共有四个阶段——预处理，建模，分割，识别。首先我们需要对甲骨文图像干扰因素进行初步判别和处理，从而建立甲骨文图像预处理模型，对附件 1 进行图像预处理，然后对原始拓片图像进行分析，实现对甲骨文字单的分割，并进行模型评估（其中附件 2 为已标注分割的数据集），并对附件 3 中的原始拓片实施这一过程，然后单独上传，最后对拓片进行文字识别（可利用附件 4 中的例子加上自己查询资料）

2 问题的分析

2.1 问题一的分析

经过背景知识和相关资料的查找，我们可以知道，甲骨文干扰因素主要分为三类：点状噪声，人工纹理和固有纹理。人工纹理是指通过特殊的施工工艺在物体表面形成的纹理，而固有纹理，顾名思义，即为拓片本身材质所形成的纹理。在这里我们采取了现代图像处理的方法，来实现高效，准确的图像预处理，可以有效屏蔽掉多种纹理噪声。

2.2 问题二和问题三的分析

甲骨文拓片分割是甲骨文数字化工程的基础问题，主要是利用数字图像处理和计算机视觉技术，在拓片的复杂背景中提取所需要的文字区域，在如今人们提出了多种方法来达到高效的分割，例如，何颖提出利用小波变换与 FCM 相结合，还有 Mask R-CNN 等技术，而在本文中，主要利用了 FASTER rcnn 和 NMS 来实现这一目的

2.3 问题四的分析

甲骨文是中国最古老的文字之一，历史的悠久也表明了它的字形与现代汉字存在很大差异，它更倾向于象形字，这也大大增加了识别难度。在过去的工作中，Guo 等人提出了基于卷积神经网络的甲骨文字分类方法，还有较为常见的就是基于深度度量学习的跨模态甲骨文字识别方法，而在这篇文章中，我们主要利用已知的样例与在网上搜集到的一些资料，利用 ResNet 算法实现了高效的甲骨文识别。

3 模型的建立与求解

3.1 预处理

预处理部分我们分为了 3 个部分：线性灰度变换，otsu 二值化，8 连通算法来进行纹理的清洗

线性灰度变换我们采取了三种不同的方式

第一种方式是将 RGB 图像转为灰度图，即将 3 通道图像改为单通道图像，因为甲骨文图像仅需黑白即可，所以我们认为将三通道转为单通道不但可以降低数据的复杂度同时我们还可以保留绝大多数的有用信息

第二种方式是图像灰度的对比加强，这里我们使用该种方法，加强了前景背景之间的对比，实现了图像加强，进一步完善了图像信息

第三种方式是图像灰度反转，我们观察到了数据集中大多数图片是黑底白字，但也有些图片是白底黑字，这里我们应用该方法，让数据的类型达到统一，方便了模型的识别

之后我们用到了 otsu 二值化，这里我们将灰度图中每个像素点的值域从 0 255 转换到了非 0 即 1，也是在保留了大部分信息的情况下简化了数据，有利于模型的准确识别

最后我们用到了八连通算法来消除纹理等噪声，具体实现为先用 dfs 遍历连通块，再通过对连通块大小和长宽比的分析进行筛选，有效去除了部分噪声，方便了模型的识别

经过上述预处理步骤，我们从原始图像得到了复杂度低的，保留了绝大多数有效信息的图像，这样不但提升了模型的效果，同时提升了模型的训练速度

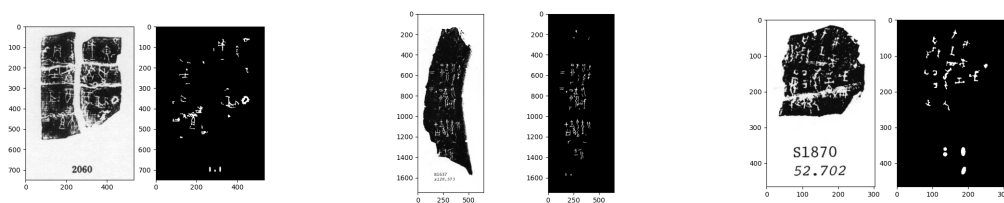


图 1: Experimental results of the authors

3.2 拓片的分割

图像分割部分我们选用了既有较高的准确率，同时也有较快的速度的 Faster-RCNN 网络，在 Faster-RCNN 网络中，该模型主要包括三个部分：负责提取图像特征的 backbone（主干），提出 proposal 的 RPN 以及最后进行识别分类的 ROI Pooling

层整体上我们在预训练的 resnet50-faster-rcnn 上进行微调，将识别的目标转为了甲骨文文字，取得了不错的效果

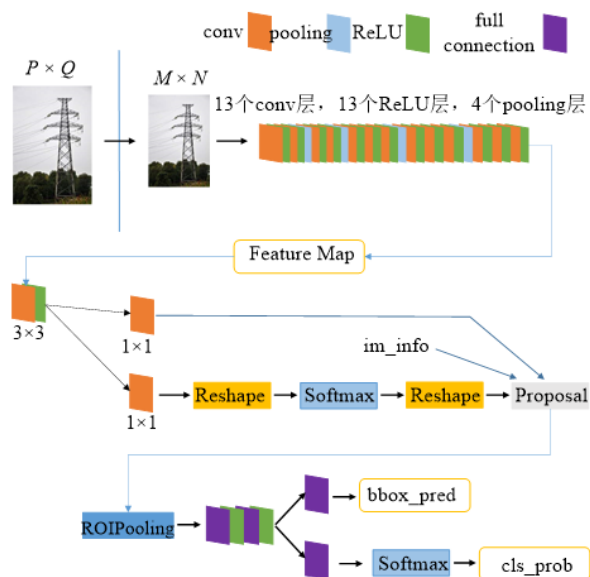


图 2: Faster-RCNN

3.3 甲骨文的识别

识别部分我们选用了稍加改进的 resnet50 进行识别，参数初始化我们选用了何恺明提出的 kaiming-normal 初始化方法，这样让模型参数能够更快的收敛，提升了训练速度，训练时我们采用了 Adam 优化方法，有效实现了模型修正。通过较为完备的图像预处理方法以及 resnet50 架构，我们取得了不错的效果

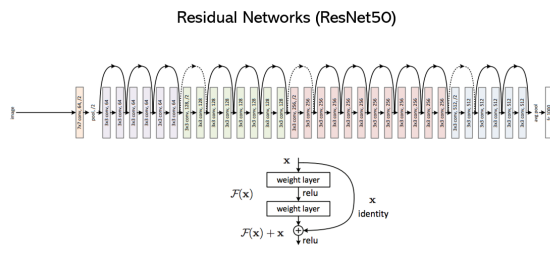


图 3: resnet50

4 结果分析与检验

结果评价与检验：1. 图像预处理成功减小了数据复杂度，并且保留并加强了原有图像的信息，既提升了训练速度又提升了模型效果，取得了不错的成果 2. 甲骨文图

像分割上我们通过 Faster-RCNN 实现了不错的识别效果，能够有效将甲骨文框选出来，甚至超过了人眼的识别能力 3. 甲骨文文字识别上，我们通过应用先进的图像预处理技术和 resnet50 以及训练技术，我们模型的 top 1 accuracy 达到了 74.68%，top 5 accuracy 达到了 90.35% 分别比高旭硕士在其论文中提到的准确率提高了 6.11% 和 2.92%，这是个不错的进展！综上所述，我们提出的方法实现了从甲骨文拓片到甲骨文文字识别整套一体化流程，并取得了相当不错的成果

表 1: 实验结果

Model	accuracy top 1	accuracy top 5
VGG16	26.91%	44.02%
Inception-ResNet-v2	33.43%	52.00%
ResNet-101	36.25%	58.03%
高旭 ResNet	68.57%	87.43%
our model	74.68%	90.35%

5 模型的评价与推广

预处理方面我们参考了多篇论文，选用了多种方法，既减小了数据复杂度又保留了绝大多数信息，实现了较为优异的数据预处理效果。图像分割上，我们采用了较为先进的 Faster-RCNN 算法，实现了高效且有效的分割，在最后的识别方面，利用优异的数据预处理技术，我们的模型无需长时间的调参和训练就可以达到十分优异的水准，综上所述，我们所建立的一从拓片预处理，到分割，再到识别的一体化模型，可以有效解决甲骨文识别问题。

此外由于 Faster-RCNN, resnet50 在 torchvision 中已有现成的实现，这让我们的模型在其他领域也可以快速部署，迅速推广，并且实现较为不错的效果

代码网址：<https://github.com/vincent1234123/->

6 实验结果

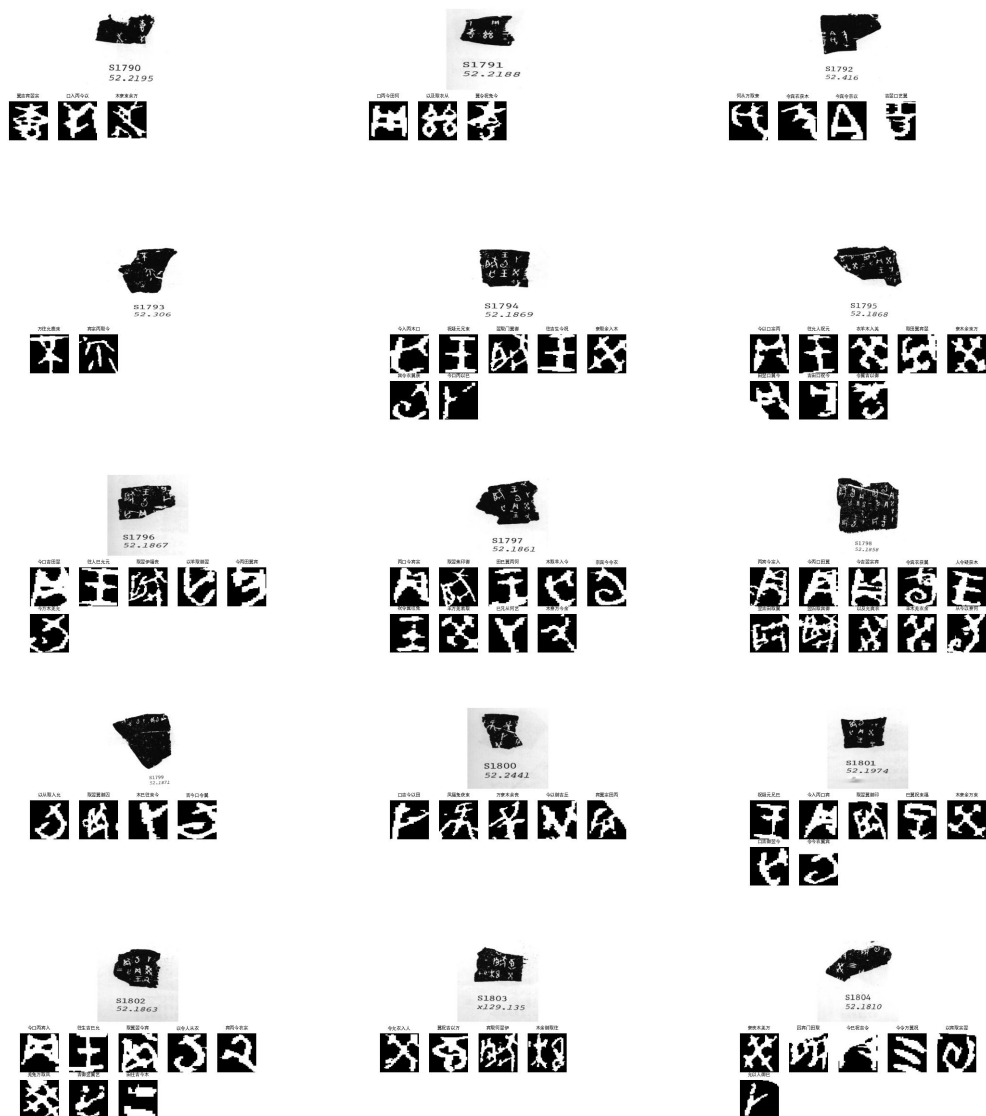


图 4: Experimental results of the authors



图 5: Experimental results of the authors

7 其他工作

此外，我们还研究了用于识别的小样本方法，通过 Siamese Network 来进行小样本分类，具体分类部分采用了 DN4，在 5-way-1-shot 测试中取得了 50.63% 的正确率，小样本学习方法具有更好的推广能力和更大的应用范围，这让我们模型的迁移能力大大提升

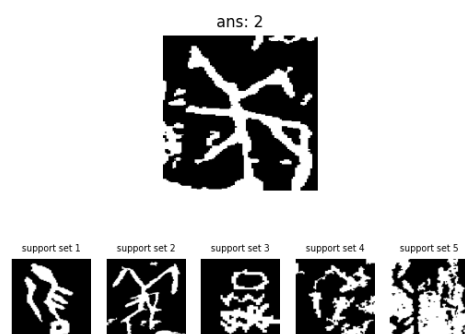


图 6: few shot pic 1

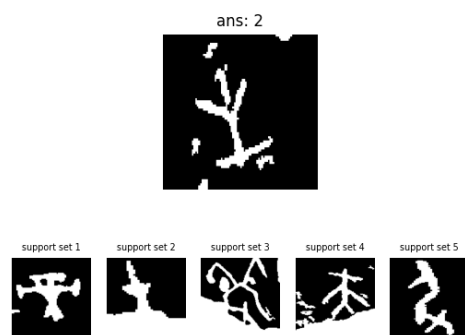


图 7: few shot pic 2

8 参考文献

- 【1】宋传鸣, 乔明泽, 边缘梯度协方差引导的甲骨文字修复算法【J】。辽宁师范大学学报 (自然科学报), 2023,46 (02): 194-207
- 【2】高旭, 基于卷积神经网络的甲骨文识别研究与应用【D】。吉林大学, 2021
- 【3】何凯明 et.al Deep Residual Learning for Image Recognition【C】。Proceedings of the IEEE conference on computer vision and pattern recognition.2016:770-778
- 【4】Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks
- 【5】Wenbin Li:Revisiting Local Descriptor based Image-to-Class Measure for Few-shot Learning.