

| | |
|------|-----------|
| 队伍编号 | MC2409918 |
| 题号 | B |

基于深度学习的一种可能的高效甲骨文图像识别

摘要

甲骨文是中国古代商代晚期至西周初期的一种古文字系统，最早出现在龟甲和兽骨上，后来也出现在青铜器上。作为中国最早的已知文字之一，甲骨文以其象形而著称，其文字结构中融合了丰富的象征性图画，并逐渐演变成具有抽象含义的符号。这些刻有卜辞的甲骨文记录了古代社会的政治、军事、宗教等方面，为我们提供了窥见古代中国社会风貌和思想的珍贵资料。然而甲骨文的识别难度大，给考古工作带来了很大的挑战。但随着计算视觉和神经网络技术的发展，使用计算机处理甲骨文图片来识别文字成为可能，并以高准确性和高速度备受关注。由于甲骨文遗迹经历长时间风吹雨晒，部分字显示不全，噪点过多，这对识别产生了巨大的挑战。本文中我们创新地采用了中值滤波、大津阈值法以及形态学的膨胀操作预处理了图片，不仅过滤了消除白色噪点和大背景噪点还可以连接边缘，使得字体连续完整。我们使用了 Faster R-CNN 对预处理的图片进行图像识别，先使用网络的主干部分（backbone）重新设置大小（resize）、归一化，这样就出现了足够多的特征，再使用生成网络（Region Proposal Network）处理图像得到由候选框（bounding boxes）组成的一组提议（proposal），它比锚框精确，这是第二次调整，我们的目的是框选可能的甲骨文区域。我们采用数字编号替代甲骨文汉字目录名称，以解决中文路径兼容性问题，并简化文件路径。随后，我们选择了 FPN 作为模型，因其多尺度特征提取、特征层级结合和减少信息丢失等优势，特别适合处理甲骨文碎片的小物体识别任务。最后，我们对结果进行了评价，我们的模型取得了不错的分割识别效果。

关键词：中值滤波 大津阈值法 形态学膨胀操作 Faster R-CNN 特征金字塔网络 机器学习 深度学习 图像识别

Contents

| | | |
|----------|--------------|-----------|
| 1 | 背景综述 | 1 |
| 1.1 | 综述 | 1 |
| 1.2 | 问题重述 | 2 |
| 2 | 基本假设 | 2 |
| 2.1 | 中值滤波 | 2 |
| 2.2 | 大津阈值法 | 3 |
| 2.3 | 膨胀操作 | 3 |
| 2.4 | 特征金字塔 (FPN) | 3 |
| 2.5 | Faster R-CNN | 3 |
| 3 | 方法 | 3 |
| 3.1 | 概述 | 3 |
| 3.2 | 第一问 | 5 |
| 3.2.1 | 图像预处理——中值滤波 | 5 |
| 3.2.2 | 图像预处理——大津阈值法 | 5 |
| 3.2.3 | 图像预处理——膨胀操作 | 5 |
| 3.2.4 | Faster R-CNN | 5 |
| 3.3 | 第二问 | 7 |
| 3.3.1 | 图像分割模型的建立 | 7 |
| 3.3.2 | 模型评估的关键指标 | 7 |
| 3.3.3 | 目标检测 | 7 |
| 3.4 | 第三、四问 | 8 |
| 4 | 结果 | 8 |
| 4.1 | 第一问 | 8 |
| 4.2 | 第二问 | 9 |
| 4.3 | 第三问 | 10 |
| 4.4 | 第四问 | 10 |
| 5 | 讨论 | 10 |
| 5.1 | 优缺点概要 | 10 |
| 5.1.1 | 优点 | 10 |
| 5.1.2 | 缺点 | 10 |
| 6 | 附录 | 12 |
| 6.1 | 查重 | 12 |
| 6.2 | 源代码 | 12 |

1 背景综述

1.1 综述

甲骨文 (Oracle Bone Inscriptions, OBIs) 是中国最早的一种文字系统, 是在龟甲和兽骨上刻写的一种古老文字, 出现于商朝时期 (约公元前 16-11 世纪)。其中, 甲骨是指用牛或羊等动物的肩胛骨或腹甲制成的, 是当时的占卜载体。甲骨文是表意文字, 记录了当时王室的政治、军事、宗教、历史等各种信息。其中最著名的是所谓的“甲骨文王卜”(或“甲骨文卜辞”), 即当时的王室占卜记录。甲骨文作为中国最早的记录历史的文字载体, 对研究商代社会、文化、历史具有重要价值。20 世纪初商周时期的考古发现为研究甲骨文提供了大量实物材料 [9, 10]。随着考古学研究的深入, 甲骨文的识别和解读变得越来越重要。甲骨文作为中国古代文化的珍贵遗产, 记录了丰富的历史信息, 然而, 传统的甲骨文识别方法面临着诸多挑战。

最常见的甲骨文检索方法是拓片法, 其中通过将一张纸放在被检查物体上, 并用滚墨擦拭纸张, 以再现甲骨文的表面。图 1 显示了几个甲骨和甲骨文的示例。图1中的三张照片分别显示了一个甲骨的正面、侧面和背面。在正面视图图像中可见的几个字符, 用红色方框标出。图1b 和 c 显示了正面-侧面甲骨的拓片图像和专家绘制的图像, 图中的字符更加清晰。图1d-g 展示了一组甲骨文的示例, 这些示例中红色方框标示了“今日之雨”一词, 而图中的铭文则以现代汉语的详细转录形式呈现, 同时附有英文含义。

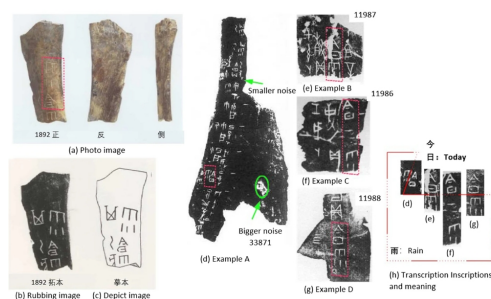


Figure 1: 甲骨文示例

甲骨文字符与现代汉字略有相似, 但这并不意味着我们可以理解甲骨文。就这些字符的质量而言, 它们的大小不均匀, 存在大小噪声, 一些字符有缺损, 字符的倾斜程度也不一致。噪声指的是拓片图像中不属于字符的白色区域。图1d 显示了接近字符附近的较小噪声, 右侧则有一个较大的噪声。这些噪声与字符非常相似, 并严重干扰了字符的检测和识别。

为了解决这些问题, Mengting Liu 等人曾提出一种新的深度卷积神经网络, 实现了对大规模甲骨文数据集的快速准确图像检索。其模型能够自动提取图像特征并从训练数据中学习。实验结果表明, 所提出的方法在识别率方面表现出色。特别是, 前 5 个预测类别的准确率达到 94.2%, 这显著降低了考古研究者的搜索推测空间, 提高了甲骨文识别的效率和准确性。

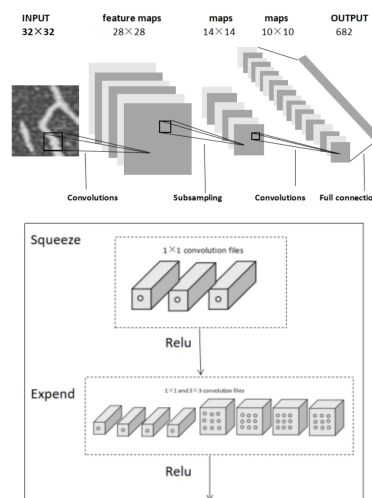


Figure 2: 基于深度卷积神经网络的甲骨文识别 (上图) 和 Fire 模块结构 (下图)

Liu Meng 等人基于单镜头多框检测器构建了甲骨文识别模块, 他们发现单镜头多框检测器在小目标检测方面表现较弱。本研究将单镜头多框检测器应用于甲骨文检测, 并对误检测进行了分析, 以实现更高的准确性。

2.2 大津阈值法

大津阈值法的基本假设包括图像的灰度直方图呈双峰分布，即存在两个主要的灰度集合；通过寻找最佳阈值，使得图像在该阈值处分割后的两个类别之间的类间方差最大化，以实现有效的图像分割；同时，大津法假设图像中的像素之间相互独立，简化了其计算过程 [1]。

2.3 膨胀操作

膨胀操作的基本假设包括结构元素的选择适应于所处理的图像特征，每个图像像素周围存在与结构元素相交的像素，以及在膨胀操作中选择输出像素值为与结构元素相交位置像素值的最大值 [8]。

2.4 特征金字塔 (FPN)

特征金字塔 (Feature Pyramid Network, FPN) 的基本假设是，在目标检测和分割任务中，多尺度特征是至关重要的。由于传统深度卷积神经网络 (CNN) 在高层级会丢失细节信息，在低层级会丢失语义信息，FPN 通过构建特征金字塔来解决这一问题。这一金字塔通过融合不同层级的特征，使得网络能够在不同尺度下捕获目标的特征，从而提高了目标检测和分割的性能 [4]。

2.5 Faster R-CNN

Faster R-CNN 假设了在图像中存在着大量的候选目标区域，这些区域可以通过区域建议网络 (Region Proposal Network, RPN) 来提出。通过共享的卷积特征提取网络，Faster R-CNN 能够从输入图像中提取出丰富的特征表示。然后，这些特征被用于两个任务：一是用于生成候选框，即 RPN 的任务；二是用于目标检测网络，这个网络对生成的候选框进行分类和位置回归。[7]。

3 方法

3.1 概述

* 由于算力有限，我们只采用了一半的训练集，但经过随机抽样发现效果极佳，本论文的所有结果基于一半的训练集。问题二和问题三用的 'oracle_bone_detection_model.pth' 模

型,问题四用的 'oracle_bone_detection_model.pth' 模型。

数据集预处理:

首先，我们对数据集进行预处理。为了解决操作系统可能对中文路径支持的问题，我们将甲骨文汉字及其异体字目录名称由中文转换为数字编号 (0 至 75)。这种方法简化了文件路径，避免了因编码问题导致的错误，并且通过编号可以快速索引到对应的汉字类别。以下是一个示例重命名操作：

- 万 → 0
- 丘 → 1
- 丙 → 2
- 喪 → 3
- 乘 → 4
- 亦 → 5
- 人 → 6

这样的简化操作有助于文件管理系统更高效地处理文件路径和访问。

其次，我们对图像进行预处理和特征提取。这一阶段的目的是准备图像数据，使其适合进行更高级的处理和分析。关键步骤包括：

- 数字编码去除：清除图像中的非甲骨文元素，如数字编码，这对于减少模型训练和推理阶段的噪声非常重要。
- 二值化处理：简化图像，使甲骨文的形状和边界更加明显，有利于后续的特征提取和分析。
- 边缘检测：Canny 算法的应用帮助描绘甲骨文的精确轮廓，对于理解其形态学特征至关重要。
- 降噪与细化处理：通过中值滤波和大津算法降低图像噪声并优化文字边缘，提高最终图像质量。

FPN 模型:

FPN 是一种高效的深度学习架构，特别适合处理图像中的小物体，如甲骨文碎片。它的优势包括：

- 多尺度特征提取: FPN 通过在不同尺度上提取特征, 增强了对小到大不等大小物体的识别能力。甲骨文字符因其多变的大小和形状, 能从 FPN 的多尺度特征提取中受益, 有助于捕捉各种大小字符的特征信息。
- 特征层级结合: FPN 结合了低层 (细节丰富) 和高层 (语义丰富) 的特征, 能够同时处理全局形状特征和局部细节特征, 这对于甲骨文的分类尤为重要, 因为甲骨文的辨识往往依赖于形状和细节的结合。
- 减少信息丢失: 通过特征金字塔结构, FPN 避免了在不同层级间频繁的上采样和下采样操作, 从而减少信息的丢失。这对于保持甲骨文图像中的细节信息至关重要。
- 处理不同尺度的能力: 由于甲骨文的大小和比例可能会有很大的变化, FPN 的能力处理不同尺度的物体使得模型在识别不同大小的甲骨文时表现更佳。

Faster R-CNN 模型和评估:

接着我们建立 Faster R-CNN 模型和评估, 我们将标注数据转换成模型能理解的格式, 确保训练数据的质量和准确性。整个训练过程涵盖了特征提取、区域建议网络和 ROI 池化层的应用。通过一系列指标 (IoU, 准确率, 精确率, 召回率, F1 分数) 来衡量模型的表现, 确保它能准确地定位和分类甲骨文。

测试流程:

步骤 1: 准备测试数据

- 选择测试图像: 例如, 使用图像 `'w01793.jpg'`, 该图像含有多个甲骨文目标。
- 确保标注数据可用: 测试前确保有关于该图像的正确标注信息, 标注数据应详细记录每个甲骨文的位置和类别。

步骤 2: 应用模型进行目标检测

- 加载模型: 使用之前训练好的模型 (如 `'oracle_bone_detection_model.pth'`) 进行测试。

- 执行目标检测: 对选定的测试图像进行目标检测, 模型会输出每个甲骨文的位置和可能的类别。

步骤 3: 目标分类

- 分类识别: 根据检测结果中的位置信息, 将图像中相应的区域送入分类器进行分类识别。这一步骤验证模型是否能准确地识别出各个甲骨文的具体类别。
- 记录预测结果: 将每个目标的预测类别及其置信度保存, 用于后续分析。

步骤 4: 结果评估

- 比对标注数据: 将模型的预测结果与实际的标注数据进行比对, 计算准确率、召回率、F1 分数等评估指标。
- 分析误差: 识别错误分类的案例, 分析可能的原因, 如模型训练不足、特征提取不充分等。

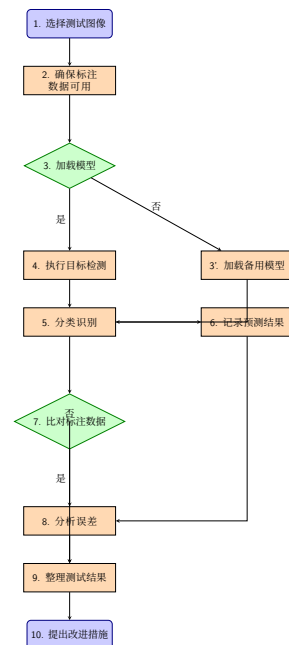


Figure 5: 基本思路

3.2 第一问

3.2.1 图像预处理——中值滤波

中值滤波是一种非线性图像平滑滤波方法。该算法针对每个像素，在其邻域内选择一个模板，该模板由其周围几个像素组成。与卷积类似，中值滤波是一种邻域运算，但不同之处在于它不涉及加权求和。相反，它对模板中的像素按照灰度级进行排序，然后选择其中间值作为输出像素的值。算法表达式为：

$$g(x, y) = \text{med}(f(x - i, y - j)) \quad (i, j) \in W$$

上式中， $f(x - i, y - i)$ 为输入像素灰度值， $g(x, y)$ 为输出像素灰度值， W 为滑动模板窗口，可以是方形、圆形等 [3]。

3.2.2 图像预处理——大津阈值法

大津阈值法是一种自动确定图像二值化阈值的技术，在图像分割领域被广泛应用。该方法由日本学者大津展之于 1979 年提出。其基本原理是通过最大化类间方差或最小化类内方差来找到最佳阈值 τ [1]。具体来说，我们先统计图像灰度像素的个数，设 n_i 是图像中灰度为 i 的像素数量，8 位图像的每个像素通常使用一个字节（8 位）来表示其灰度值，因此有 $n_0, n_1, n_2 \dots n_{255}$ ，且 n_i 有两类，一类 $n_i > \tau$ ，另一类 $n_i < \tau$ ，图像中像素灰度为 i 的概率为： $p_i = \frac{n_i}{n_0 + n_1 + \dots + n_{255}}$ ，且有： $\sum_{i=0}^{255} p_i = 1$ 。设阈值为 t 时，像素被分到 A 的概率为 $p_A(t)$ ，分配到 A 的像素的平均灰度为 $m_A(t)$ ，同样有像素被分到 B 的概率为 $p_B(t)$ ，分配到 B 的像素的平均灰度为 $m_B(t)$ ，灰度级 t 的累加均值 m ，整个图像的灰度值 m_G ，则有：

$$p_A(k) * m_A(k) + p_B(k) * m_B(k) = m_G$$

$$p_A(k) + p_B(k) = 1$$

根据方差的概念，方差的表达式写成：

$$\sigma^2 = p_A(k)(m_A(k) - m_G)^2 + p_B(k)(m_B(k) - m_G)^2$$

于是有：

$$\sigma^2 = p_A(k)p_B(k)(m_A(k) - m_B(k))^2$$

其中：

$$p_A(t) = \sum_{i=0}^t p_i$$

$$p_B(t) = \sum_{i=t+1}^{255} p_i$$

$$m = \sum_{i=0}^t i p_i$$

$$m_G = \sum_{i=0}^{255} i p_i$$

$$m_A(t) = \frac{\sum_{i=0}^t i p_i}{p_A(t)}$$

$$m_B(t) = \frac{\sum_{i=t+1}^{255} i p_i}{p_B(t)}$$

于是有：

$$\sigma^2 = \frac{(m_G * p_A(t) - m)^2}{p_A(t)(1 - p_A(t))}$$

求出上式的最大灰度级 m ，即为大津阈值法的阈值 τ 。

3.2.3 图像预处理——膨胀操作

膨胀操作是一种形态学操作，通过滑动结构元素在图像上的不同位置并检查它与相应图像区域的重叠来扩张图像。在膨胀过程中，如果结构元素的中心与图像的某一像素位置重叠，那么该位置就会被视为输出图像的一个元素，并且该元素的灰度值将取决于结构元素和图像区域之间的最大值。因此，膨胀操作会扩大图像中的白色区域，通常用于连接物体、填充空洞等任务中 [8]。

$$(A \oplus B)(x, y) = \bigcup_{(s, t) \in B} A(x - s, y - t)$$

3.2.4 Faster R-CNN

我们使用 Faster R-CNN (Region-based Convolutional Neural Network, 基于区域的卷积神经网络) 来对图像进行预处理，框选可能的甲骨文区域。

Faster R-CNN 是一种用于目标检测的深度学习模型，它由 Ross B. Girshick 在 2016 年提出，是 R-CNN 系列模型的一种改进版本。它由

两个主要组件组成：区域建议网络（RPN）和区域分类网络（Region-based CNN, R-CNN）。Faster R-CNN 的主要特点是引入了 RPN，它能够在单一的网络中同时生成目标的边界框和类别概率。

在 Faster R-CNN 中，首先通过卷积神经网络（Convolutional Neural Network, CNN）提取图像特征。然后，RPN 在这些特征图上滑动窗口，为每个窗口提供候选目标框的位置建议。接着，这些候选框被输入到分类网络中进行目标分类和边界框的精确定位。整个模型通过端到端的方式进行训练，使得 RPN 和分类网络可以共享卷积层，从而提高了模型的速度和准确性 [7]。

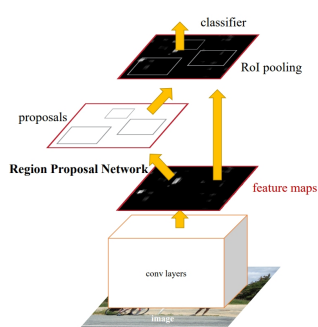


Figure 6: Faster R-CNN 基本结构 [7]

Faster R-CNN 中，区域建议网络（Region Proposal Network, RPN）负责生成候选目标框的位置建议，RPN 的核心思想是利用卷积神经网络直接生成区域建议（region proposal），其方法本质上是一种滑动窗口的思想，它只需在最后的卷积层上进行一次滑动，利用锚框机制和边界框回归可以得到多尺度、多长宽比的区域建议。具体而言，对于每个 3x3 的滑动窗口，RPN 同时预测输入图像中的多种尺度（例如 128、256、512 像素）和多种长宽比（例如 1:1、1:2、2:1）的区域建议，这种映射机制被称为锚框（anchor）。之后它再利用基础网络对图像进行一系列的卷积和池化操作，从而提取特征并生成原始的特征图（灰色区域）。接着，在原始的特征图后面添加一个卷积层和整流线性单元（Rectified Linear Unit, ReLU）激活函数 $f(x) = \max(0, x)$ ，得到将用于生成区域建议的特征图。对于特征图中的

每个点，都会将其映射回原始图像，从而得到一个基准点。然后，根据预先设置的尺度大小和长宽比，在围绕每个基准点的位置生成 k 个锚框。特征图中的每个点都会输出与这 k 个锚框相关的信息，包括是否包含目标以及回归出 k 个区域建议的坐标值。

在训练 RPN 时，首先对于得到的 $W \times H \times k$ 个锚框进行正负样本的划分。具体操作如下：

- 将与任何一个 ground-truth 边界框（gt box）具有最大 IOU（Intersection over Union，交并比）的锚框，以及与任何一个 gt box 的 IOU 大于 0.7 的锚框，作为正样本。
- 将剩余的与 gt box 的 IOU 小于 0.3 的锚框作为负样本。
- 其余的锚框被忽略，不参与训练。

对于每个锚框，后续会接上一个用于二分类的 softmax 层（ $\text{softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^N e^{z_j}}$ ， z_i 是原始的未归一化的输出）和一个用于边界框回归的网络。其中，二分类 softmax 层用于判断锚框是否包含目标的概率，而边界框回归器用于调整锚框的 4 个坐标值，以更准确地拟合目标边界框。因此 RPN 的损失函数可以定义为：

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

其中：

- p_i 是预测的锚框是否包含对象的概率。
- t_i 是预测的锚框的回归参数。
- p_i^* 是表示锚框是否为正样本的标签。
- t_i^* 是表示正样本锚框的回归目标的标签。
- N_{cls} 是正负样本的数量。
- N_{reg} 是正样本的数量。
- L_{cls} 是二元交叉熵分类损失函数。
- L_{reg} 是平滑的 L1 回归损失函数。
- λ 是用于平衡分类损失和回归损失的超参数。

Ross B. Girshick 使用 robust loss (smooth L1) 作为回归损失函数 L_{reg} ，其表达式如下：

$$smooth_{L1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases}$$

由于图片输入的大小通常不能保证一致，ROI (Region of Interest, 直译为“感兴趣区域”) 池化通常应用于目标检测任务中，用于提取每个候选目标区域的固定大小的特征表示。ROI 池化的操作如下：

- 对于每个候选目标区域，首先将其划分为一个固定大小的网格（通常是 $N \times N$ ）。
- 然后，在每个网格单元格中进行池化操作（如最大池化），得到一个固定大小的特征。
- 最后，将所有网格内的池化结果拼接成一个固定大小的特征向量作为该候选区域的特征表示。

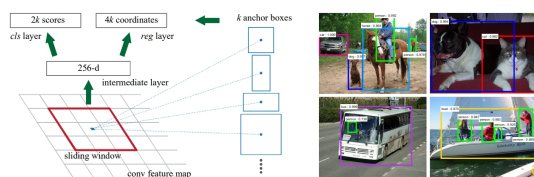


Figure 7: 左图：区域提议网络（RPN）。右图：使用 RPN 提议在 PASCAL VOC 2007 测试集上进行的示例检测。该方法可以检测各种尺度和长宽比的对象。

整个 Faster R-CNN 的训练过程可以分为 4 个步骤：

1. **初始化 RPN 网络**：使用在 ImageNet 数据集上预训练好的模型来初始化模型参数，并训练 RPN 网络。这一步的目标是训练一个用于生成候选区域的区域提议网络 (Region Proposal Network, RPN)。
2. **训练 Fast R-CNN**：使用在 ImageNet 数据集上预训练好的模型初始化模型参数，并

将第一步中训练好的 RPN 网络生成的候选区域作为输入，训练一个 Fast R-CNN。这一步的目标是训练一个用于目标检测的区域分类网络。

3. **微调 RPN 网络**：使用第二步训练好的 Fast R-CNN 的网络参数初始化 RPN 网络，但是将 RPN 与 Fast R-CNN 共享的网络层的学习率设置为 0，仅微调 RPN 独有的网络层。这一步的目标是进一步优化 RPN 网络，使其更适合生成准确的候选区域。
4. **微调 Fast R-CNN**：固定共享的网络层，仅微调 Fast R-CNN 所独有的全连接层。这一步的目标是进一步优化 Fast R-CNN 的目标检测性能。

3.3 第二问

3.3.1 图像分割模型的建立

我们使用训练好的 Faster R-CNN 处理数据图片，附件中我们上传了训练好的模型 `'oracle_bone_detection_model.pth'`。

3.3.2 模型评估的关键指标

我们采用的评估指标包括准确率、精确度、召回率、F1 分数、交并比、平均精度均值、Dice 系数和结构相似性指标 (Structural Similarity Index, SSIM)。准确率衡量了模型在整体上的分割性能，精确度衡量了模型在预测为正例的样本中的准确性，而召回率衡量了模型对正例样本的识别能力。F1 分数综合考虑了精确度和召回率，而交并比则衡量了模型预测结果的覆盖程度。平均精度均值通过绘制精度-召回率曲线并计算曲线下面积来评估对象检测模型的性能。Dice 系数用于衡量分割结果的覆盖程度，而 SSIM 通常用于评估图像重建质量。

3.3.3 目标检测

Feature Pyramid Network (FPN) 的结构主要包括三个部分：bottom-up, top-down 和 lateral

connection。

Bottom-up 过程:

Bottom-up 过程是将图片输入到 backbone ConvNet 中提取特征的过程。对于不同层级的特征，有的保持尺寸不变，有的尺寸减小一半。以 ResNet 为例，将卷积块 conv2, conv3, conv4, conv5 的输出定义为 $\{C_2, C_3, C_4, C_5\}$ 。这些特征图的尺寸分别是原图的 $\frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32}$ 倍。

Top-down 过程:

Top-down 过程是将高层得到的特征图进行上采样后传播到低层的过程。通过上采样，高层的语义信息可以传播到低层特征，使得低层也包含丰富的语义信息。本文使用最近邻上采样，即在原有像素的基础上插入新像素，方法简单但可能会导致不连续性。

Lateral Connection:

Lateral connection 主要包括以下步骤:

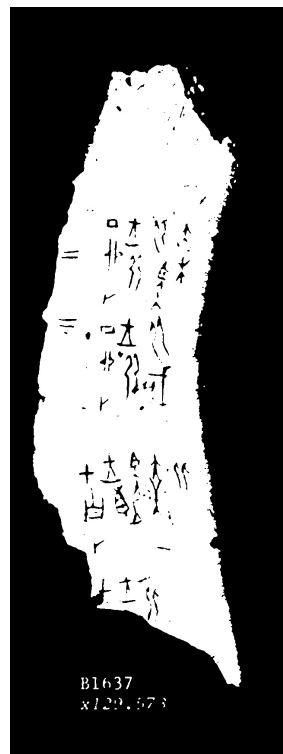
1. 对于每个 stage 输出的特征图 C_n ，进行 1×1 卷积以降低维度。
2. 将降维后的特征与上一层采样得到的特征图 P_{n+1} 相加，采用逐元素相加的方式。
3. 对相加结果进行 3×3 卷积，以消除上采样产生的混叠效应，同时输出维度统一为 256。

3.4 第三、四问

第三、四问基于第一、二问。

4 结果

4.1 第一问



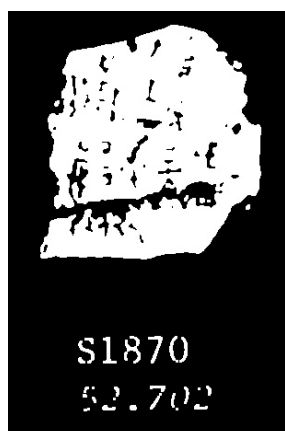


Figure 9: 第一问结果, 由上到下的文件名分别是 'h02060.jpeg', 'w01637.jpeg', 'w01870.jpeg'



4.2 第二问

我们使用 'oracle_bone_detection_model.pth' 对 '2_Train' 文件夹的文件进行处理, 以下是我们随机选取的三张照片, 剩下的结果见附件。考虑到时间和算力的原因, 我们只提供了 365 张照片的结果。

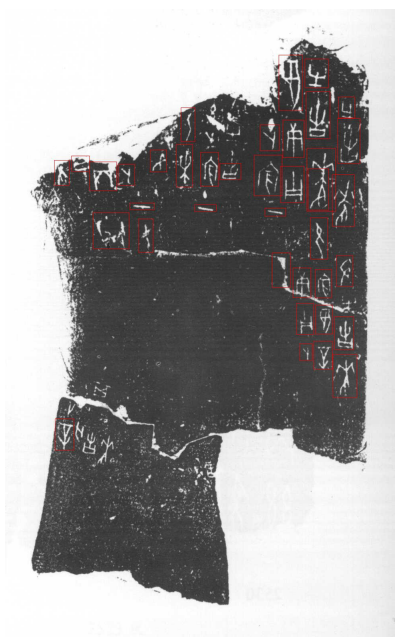


Figure 10: 使用训练的模型分割的部分结果

4.3 第三问

见附件。

4.4 第四问

见附件。

5 讨论

5.1 优缺点概要

5.1.1 优点

- **高准确性：**Faster R-CNN 在甲骨文检测框选上表现出色，具有较高的检测准确性和较低的误检率，经过测试，表现良好。
- **快速检测速度：**尽管是基于深度学习的复杂模型，但 Faster R-CNN 在推理阶段的检测速度相对较快，可以在实时或高效率的应用中使用。
- **良好的通用性：**Faster R-CNN 适用于各种不同类型的目标检测任务，对不同的甲骨文照片其性能表现良好。

5.1.2 缺点

- **计算资源需求高：**训练 Faster R-CNN 模型通常需要大量的计算资源，包括高性能的 GPU 和大量的训练时间。
- **困难样本处理：**甲骨文目标过小且部分和噪点高度重叠，部分不连通，Faster R-CNN 可能存在一定的检测困难，需要更多的样本和优化算法比如梯度下降算法来处理。
- **定位精度不足：**对于甲骨文来说，由于噪点和其独特的文字需要高精度目标定位，Faster R-CNN 难免存在一定程度的定位误差，需要进一步改进和调优。

References

- [1] Sunil L Bangare et al. “Reviewing Otsu’s method for image thresholding”. In: *International Journal of Applied Engineering Research* 10.9 (2015), pp. 21777–21783.
- [2] Guang Deng and LW Cahill. “An adaptive Gaussian filter for noise reduction and edge detection”. In: *1993 IEEE conference record nuclear science symposium and medical imaging conference*. IEEE. 1993, pp. 1615–1619.
- [3] Thomas Huang, GJTGY Yang, and Gregory Tang. “A fast two-dimensional median filtering algorithm”. In: *IEEE transactions on acoustics, speech, and signal processing* 27.1 (1979), pp. 13–18.
- [4] Tsung-Yi Lin et al. “Feature pyramid networks for object detection”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 2117–2125.
- [5] Lin Meng. “Recognition of Oracle Bone Inscriptions by Extracting Line Features on Image Processing.” In: *ICPRAM*. 2017, pp. 606–611.
- [6] Lin Meng et al. “Oracle bone inscription detector based on ssd”. In: *New Trends in Image Analysis and Processing-ICIAP 2019: ICIAP International Workshops, BioFor, PatReCH, e-BADLE, DeepRetail, and Industrial Session, Trento, Italy, September 9–10, 2019, Revised Selected Papers 20*. Springer. 2019, pp. 126–136.
- [7] Shaoqing Ren et al. “Faster r-cnn: Towards real-time object detection with region proposal networks”. In: *Advances in neural information processing systems* 28 (2015).
- [8] Pierre Soille et al. *Morphological image analysis: principles and applications*. Vol. 2. 3. Springer, 1999.
- [9] 张国硕. “甲骨文在商代非档案说”. In: *档案学研究* 2 (1999), pp. 64–68.
- [10] 曹定云. “殷墟甲骨文研究百年回顾与展望”. In: *社会科学战线* 5 (1997), pp. 160–168.

6 附录

6.1 查重

PaperYY® 检测报告单-打印版

检测文献: BMC2409918 (免费版)

文献作者: 郑嘉诚, 季平衡, 毛雅楠

报告时间: 2024-04-16 07:09:31

段落个数: 3

报告编号: YY202404160709277575

检测范围: 中国期刊库 中国图书库 硕士论文库 博士论文库 会议论文库 报纸库
网友专利库 网友标准库 网友共享库 个人对比库 网页库 百科库



总文字复制比: 18.6%

去除引用文献复制比: 18.6%

去除本人已发表文献复制比: 18.6%

单篇最大文字复制比: 2.8%

重复字数: 1,903

总字数: 10,205 (不含参考文献)

总段落数: 3 (不含参考文献)

前部重合字数: 0

疑似段落数: 2

后部重合字数: 1,903

单篇最大重复字数: 290

疑似段落最小重合字数: 268

6.2 源代码

在提交给组委会的附件中我们提供了数据和原始代码。