

队伍编号	BMC2409052
题号	B

## 基于 YOLO 模型的甲骨文字符自动检测识别研究

### 摘要

甲骨文是中国迄今为止发现的最早的成熟文字系统，在中国古代历史和文化中占据极其重要的地位，是中华优秀传统文化的根脉，对汉字的从古至今的发展和中华文明的历史传承起到了关键作用。甲骨文字符检测识别一直是甲骨文研究中重要的研究内容，目前，这些工作都需要甲骨学专家们人工完成，效率低下，且耗费昂贵的专家资源。因此研究甲骨文字符的自动检测识别很有价值。<sup>1</sup>

本文以 **YOLOv8 模型** 为基本方法，对甲骨文字符的自动检测识别进行研究。

对于问题一，本文分三步构建了甲骨文预处理方法。由于对拓片图像中甲骨文的识别极易受点状噪声、人工纹理、固有纹理等元素的干扰，我们首先对图像进行了**高斯滤波处理**，去除点状噪声。然后我们对图像进行了**均值滤波平滑处理**，并通过**直方图修正**方法增强图像，突出了有价值的甲骨文部分，最后我们用 **BRISK(Binary Robust Invariant Scalable Keypoints)**模型提取图像特征。

对于问题二，本文建立一个快速准确的甲骨文图像分割模型，实现对甲骨文原始拓片图像的单字分割，并从不同维度进行模型评估。本文以题目提供的数据为主，经过图像预处理、筛选后形成 YOLO 模型的**训练数据集**。经过多次调整训练以及针对特别缺陷的改进，得到一个能从拓片图像中识别甲骨文单字的 YOLO 模型 A。本模型的**精确率为 80.9%**，**召回率为 89.0%**，基于精确率和召回率的调和平均数的 **F1 分数为 84.7%**，三者均属于较高层次，PCC，RCC，PR-Curve 曲线同样表明了模型的优秀性能。可以认为问题二模型建立较为成功。

对于问题三，本文利用已有模型对附件 3 中的 200 张甲骨文原始拓片图片进行单字分割，并将结果**保存到指定文件**和附件的 results 的文件夹中。

对于问题四，由于题目给定的数据集中不同类别数据规模差距大，本文对数据量大的类别进行**欠采样**，规定最大规模为 200；同时，**扩大训练效果不理想的类别**的数据规模，补充数据的来源为 **OBCD (Oracle Bone Character Detection)**。得到足够数据后，以甲骨文单字图像为数据，其对应的现代汉字为标签，设置各类别**训练集和验证集**的比例为 **8: 2** 构建数据集。经过训练，得到一个能够识别甲骨文单字图像对应的现代汉字的 YOLO 模型 B。本文首先使用模型 A 从拓片中识别甲骨文单字并分割，将得到的甲骨文单字图像输入模型 B 识别，得到其对应的现代汉字。从而实现了对甲骨文原始拓片的文字识别。本文将测试集中的图像，经历以上流程输出的结果，保存在附件的 results 文件夹中。

**关键词：**甲骨文，YOLO 模型，卷积神经网络，图像识别

## 目录

一、问题重述 .....	1
1.1 问题背景 .....	1
1.2 问题提出 .....	1
一、 问题分析 .....	2
2.1 问题一的分析 .....	2
2.2 问题二的分析 .....	2
2.3 问题三的分析 .....	2
2.4 问题四的分析 .....	2
三、模型假设与符号说明 .....	2
3.1 模型假设 .....	2
3.2 符号假设 .....	3
四、模型的建立与求解 .....	4
4.1 问题一模型的建立与求解 .....	4
4.2 问题二模型的建立与求解 .....	9
4.3 问题三模型的建立与求解 .....	12
4.4 问题四模型的建立与求解 .....	13
五、模型的分析与检验 .....	14
5.1 问题一模型的分析与检验 .....	14
5.2 问题二模型的分析与检验 .....	14
5.3 问题三模型的分析检验 .....	20
5.4 问题四模型的分析检验 .....	20
六、 模型的评价与改进 .....	21
6.1 模型的优点 .....	21
6.2 模型的不足 .....	21
6.3 模型的推广和改进 .....	21
附件 .....	22
参考文献 .....	23

## 一、问题重述

### 1.1 问题背景

甲骨文是中国迄今为止发现的最早的成熟文字系统，印刻在兽骨上而得名。对甲骨文的研究具有重要价值，不仅丰富了中国文明起源的历史根基，也为世界文化的研究提供了重要媒介。当今中国，在党的领导下与我国政府大力支持下，甲骨文的研究开启了新篇章，应用人工智能和大数据技术于甲骨文数字化建设和全息性研究具有重要意义，成为处理甲骨文信息领域的重要热点。[2]

### 1.2 问题提出

甲骨文自动检测识别首先要对原始拓片图像的甲骨文位置进行识别，过程中往往受到点状噪声、人工纹理和固有纹理等干扰元素的严重影响，且甲骨文图像来源广泛，包括拓片、拍照、扫描、临摹等，而不同的图片来源对甲骨文识别的干扰元素不同，为降低干扰因素对图片识别与切割的影响我们对图片进行预处理。为实现对甲骨文图像的单个文字的自动分割和识别，建模解决以下问题：

问题一：已有未知来源的甲骨文图片，要求建立模型对甲骨文图片进行预处理，实现对甲骨文图像识别中的干扰元素的初步判别和处理，提取图像特征。

问题二：凭借附件 2 (Train 文件夹) 中已标注分割的数据集为依据，建立一个准确、高速的甲骨文图像分割模型，使其能自动对不同的甲骨文原始拓片图像进行单字分割，并从多个维度来评价建立的模型。

问题三：利用问题二建立的甲骨文图像分割模型，对附件 200 张甲骨文原始拓片图像进行模型处理，记录其中数据，将结果保存到“Test\_results.xlsx”中，上传至竞赛平台。

问题四：基于前三问对甲骨文原始拓片图像的单字分割研究，采用合适的方法进行甲骨文原始拓片的文字识别，对测试集中的 50 张甲骨文原始拓片图像进行文字自动识别，并以适当结果呈现。

## 一、问题分析

### 2.1 问题一的分析

本问提出对甲骨文的数字化识别中的基础问题是使甲骨文拓片实现图像分割,而甲骨文拓片分割往往受到点状噪声、人工纹理和固有纹理三类元素的干扰,因此在甲骨文拓片识别之前需对其进行预处理以降低这三种因素的影响,过程一、高斯滤波去噪,二、用中值滤波方法作平滑处理,三、应用直方图修正方法进行图像增强,四、将图像二值化处理突出甲骨文部分。预处理之后应用 BRISK 模型对图像特征提取,方便图像处理。

### 2.2 问题二的分析

本问主要目的是对甲骨文原始拓片图像进行分析,需要建立一个甲骨文图像分割模型,要求其快速准确,能够自动对甲骨文原始拓片图像进行单字分割。基于卷积神经网络(CNN)开发的 YOLO 模型因其出色的特征提取性能、良好的泛化能力和易于部署的特点成为了本文的训练建立模型的选择。本文同时对分割后的数据结果进行检测和分析,从混淆矩阵、F1 曲线、PCC 图像、RCC 图像、PR 曲线、损失函数图像六个维度进行模型评估,以测试模型的性能。

### 2.3 问题三的分析

本问主要是对问题二模型的应用检验,通过将 Test 文件夹中 200 张甲骨文原始拓片图像进行自动分割,测试问题二模型的训练效果。

### 2.4 问题四的分析

测试集中的原始拓片图像可以通过问题二中的模型分割为甲骨文单字图像,故本问主要部分是对甲骨文单字含义的识别。训练集中给出了数十个现代汉字所对应的甲骨文单字图像集,可以训练新的 YOLO 模型,学习不同现代汉字对应的甲骨文单字的特征,从而实现对甲骨文单字含义的识别。训练集中不同的汉字对应的甲骨文单字图像的数量差距较大,可以分别进行过采样或欠采样来调节权重,以保证模型对于不同文字的识别效果。

## 三、模型假设与符号说明

### 3.1 模型假设

为了提高模型的计算速度和准确度,本文做出如下基本假设:

假设 1: 题目中给出的甲骨文标注数据真实可靠,不存在人工误差。

假设 2: Recognize-测试集中所有甲骨文都有训练集中的字与它对应。

假设 3: 甲骨文样本都是单字陈列,无多字重叠的情况。

### 3.2 符号假设

符号	说明	单位
$I'(x,y)$	滤波后的新像素值	
$I(i,j)$	原始图像中位置为 (i,j)的像素值	
$N(x,y)$	以 (x,y) 为中心的邻域内的像素点集合	
$M(x,y)$	中值滤波目标中值	
$\Omega(x,y)$	所有像素灰度值的统计中值	
$P_i(r_i)$	各灰度级的概率密度	
$S_k(r_k)$	累计分布函数	
$g_k$	直方图均衡化最后的输出灰度级	
$Precision$	精确率	
$Recall$	召回率	
$F_1$	对 $Precision$ 和 $Recall$ 进行整体评价指标	

## 四、模型的建立与求解

### 4.1 问题一模型的建立与求解

#### 4.1.1 问题一模型的去噪处理

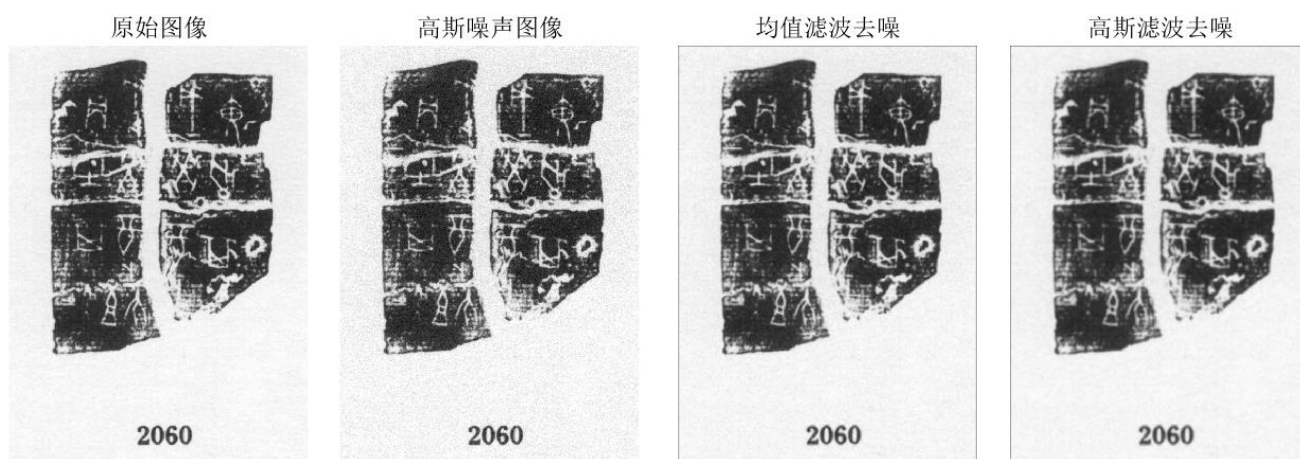


图 1 去噪处理

图像去噪是图像处理中非常重要的环节。目的是减少或移除图片中的噪音提高图像质量，是针对点状噪声干扰因素的处理。

#### 1. 均值滤波

均值滤波是一种最简单的线型滤波，其基本思想是用像素点及其邻域内像素点的平均值来替换该像素点的值。因此，均值滤波的操作可表述为：

$$I'(x,y) = \frac{1}{M} \sum_{(i,j) \in N(x,y)} I(i,j) \quad (4.1.1)$$

其中， $I'(x,y)$ 是经过滤波后的新像素值， $I(i,j)$  是原始图像中位置为  $(i,j)$ 的像素值， $N(x,y)$  是以  $(x,y)$  为中心的邻域内的像素点集合， $M$  是邻域内的像素数量。

均值滤波模型具有简单且计算速度快的优点，但它总是会模糊图像的边缘和细节。而在甲骨文拓片图片的识别中因诸多因素的影响对细节精确度要求较高，因此我们考虑应用其他类型的滤波器。

#### 2. 高斯滤波

高斯滤波是一种常用的线性滤波技术，其基本思想是，图像中的每个像素点的灰度值可以通过对其邻域内像素点的灰度值加权平均来估计，权重由高斯函数给出，距离中心像素点越远的像素点权重越小<sup>[3]</sup>。可表述为：

$$I(x, y) = \frac{1}{2\pi\sigma^2} \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} I(i, j) \exp\left(-\frac{(i^2+j^2)}{2\sigma^2}\right) \quad (4.1.2.1)$$

其中：

$I(x,y)$  是经过滤波后的新像素值。

$I(i,j)$  是在位置 $(i,j)$ 的原始图像的像素值。

$\sigma$ 是高斯分布的标准差，决定了滤波器的宽度，即权重的衰减速率。

$\exp$  表示指数函数。

求和是对所有整数  $i$  和  $j$  进行的，在实际应用中通常会限制在一个足够大的邻域内，例如  $[-r,r]$  的范围，其中  $r$  是一个足够大的整数，以确保高斯函数在邻域外接近于零。高斯滤波器的大小(即  $r$  的值)和标准差  $\sigma$  是影响滤波效果的两个重要参数。较大的  $\sigma$  值会导致更平滑的图像，同时也会更多地模糊图像的细节；较小的  $\sigma$  值则会产生较少的模糊，但同时去噪效果可能不明显。模型经过多次的参数调整与比较，最后确定了最适宜甲骨文精确识别的高斯滤波参数，并进行了多次试验，最终决定应用高斯滤波对甲骨文拓片进行去噪处理。

#### 4.1.2 问题一模型的平滑处理



图 2 平滑处理

中值滤波是一种非线性的数字滤波技术，我们利用中值滤波对去噪后图像进行平滑处理，其基本思想是在图像的每个像素点及其邻域内，取所有像素点的灰度值的中值作为该像素点的新灰度值<sup>[4]</sup>，不同于均值滤波，中值滤波不会平滑图像的边缘和细节能更好的保留图像的结构和对比度。

若 $(x,y)$ 是一个已经给定的像素位置，原始图像所在位置是 $(x,y)$ 的像素灰度值是 $I(x,y)$ ，滤波器窗口的大小是 $N$ (例如，对于 $3 \times 3$ 的窗口， $N=9$ )。滤波器窗口内的像素集合可以表示为：

$$\Omega(x,y) = \{I(x+i,y+j) \mid (i,j) \in S\} \quad (4.1.2.1)$$

，其中 $S$ 是滤波器窗口内所有可能的像素位置。在操作中值滤波的过程中目标是找到一个中值 $M(x,y)$ ，作为 $\Omega(x,y)$ 中所有像素灰度值的统计中值。在数学上，中值 $M(x,y)$ 可以表示为：

$$M(x,y) = \text{median}(\Omega(x,y)) \quad (4.1.2.2)$$

这里，应用了`median`函数用于表示集合 $\Omega(x,y)$ 中的像素灰度值在进行排序后位于中间位置的值。如果 $N$ 是奇数，中值就是排序后正中间的值；如果 $N$ 是偶数，则中值通常是中间两个值的平均值。

中值滤波的表达方式简单直观，但在计算过程中需要排序每个窗口内的每个像素值，导致计算的复杂程度较高，在处理复杂图像和大量图像时尤为明显。

#### 4.1.3 问题一模型的增强

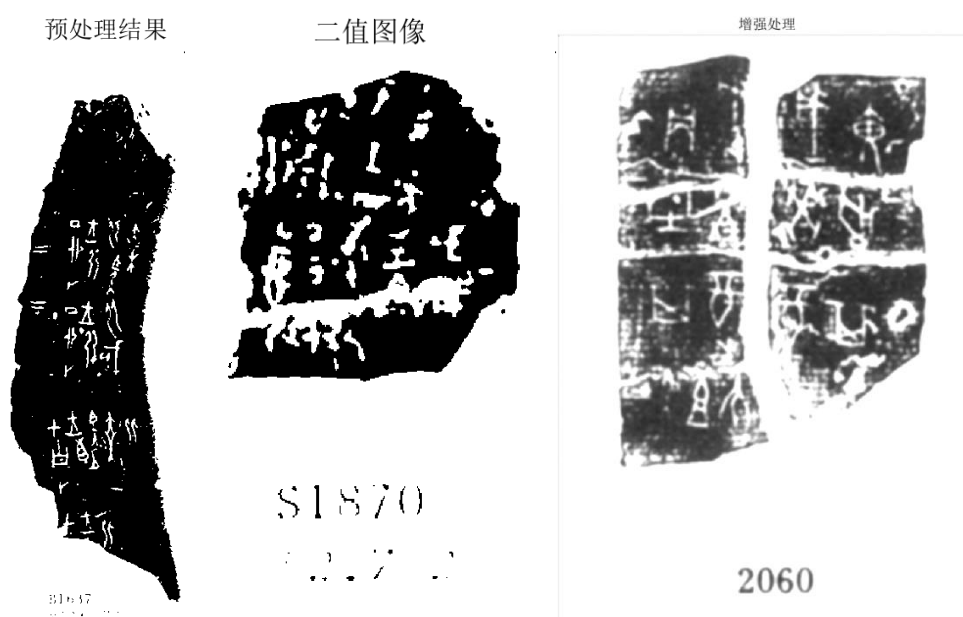


图 3 增强处理

图像增强的主要目的是改善图像的视觉效果，突出感兴趣的特征，以及丰富信息量。具体来说，图像增强技术可以用于提高图像的对比度以及视觉体验，同时还可以抑制不感兴趣的或者无关的特征，扩大图像中不同物体之间的差别，以便于更好地进行图像分析和识别，满足特定要求，在本题中，我们要对甲骨文来加强处理，对一些噪点等不重要的信息做抑制处理，所以我们采用图像增强方法中的直方图均衡化。<sup>[5]</sup>

直方图均衡化是一种常用的图像增强技术，其主要目的是改善图像的对比度。这种方法基于图像的直方图分布，通过拉伸或压缩直方图来优化图像的对比度。直方图是一个统计图表，用于表示图像中每个像素强度(或颜色)的频率分布。在灰度图像中，



直方图显示了从黑色(强度值 0)到白色(强度值 255)的每个强度级别的像素数量。累积直方图(CDF)是直方图的累积和,表示小于或等于某个强度值的像素数量占总像素数量的百分比。直方图均衡化的目标是使累积直方图尽可能接近线性分布,从而使得图像的对比度在整个强度范围内均匀分布。这个过程包括以下步骤:

计算累积直方图:对于给定的图像,首先计算每个强度级别的累积直方图。

归一化累积直方图:将累积直方图的值归一化到 0 到 1 的范围内,这样可以得到每个强度级别的概率。

创建均衡化映射:通过计算累积直方图的 CDF,创建一个映射函数,该函数将原始图像中的每个强度值映射到新的强度值。这个映射函数通常是线性的,确保累积直方图是均匀分布的。

应用映射:使用映射函数将原始图像中的每个像素强度值转换为新的强度值,生成增强后的图像。

其一般实现过程为

(1)统计原始输入图像各灰度级的像素数目 $n_i$ ,  $i=0,1,\dots,L-1$ ,其中  $L$  为灰度总级数;

(2)计算原始图像直方图,即各灰度级的概率密度,  $P_i(r_i) = \frac{n_i}{n}$ ,  $n$  为原始图像的总像素数目;

(3)计算累计分布函数

$$S_k(r_k) \approx \sum_{i=0}^k P_i(r_i) \quad k=0,1,\dots,L-1 \quad (4.1.3.1)$$

(4)计算最后的输出灰度级

$$g_k = \frac{\text{INT}[(g_{\max}-g_{\min})S_k(r_k)+g_{\min}+0.5]}{L-1} \quad k=0,1,\dots,L-1 \quad (4.1.3.2)$$

式中  $\text{INT}[]$ 是取整运算符。令  $g_{\min} = 0, g_{\max} = L - 1$ ,则计算式简化为

$$g_k = \frac{\text{INT}[(L-1)S_k(r_k)+0.5]}{L-1} \quad k=0,1,\dots,L-1 \quad (4.1.3.3)$$

(5)用  $f_k$ (原图像的灰度级函数)和  $g_k$  的映射关系,修改原图像的灰级度,获得输出图像,其直方图为近似均匀分布。

直方图均衡化实质上是减少图像的灰度级来加大对比度,图像经均衡化处理之后,图像变得清晰,直方图中每个像素点的灰度级减少,但分布更加均匀,对比度更高。但同时将原始函数的累积分布函数作为变换函数,只能产生近似均匀的直方图。并且,在某些情况下,并不一定需要具有均匀直方图的图像。

#### 4.1.4 问题一模型的二值化处理

图像二值化是将灰度图像转换为只包含两种像素值的图像的过程,通常是黑和白。在二值化图像中,所有像素值高于某个阈值的像素被设置为一个值(例如 255,代表白色),而所有低于阈值的像素被设置为另一个值(例如 0,代表黑色)。图像的二值化处理过程简化了图像的复杂性,使得对图像的进一步分析和处理变得更加容易,尤其是

在需要突出图像中的特定结构或特征时。对与精确识别甲骨文拓片具有重要意义。最基本的二值化方法是阈值二值化法，设置一个阈值将像素分为两类。所有高于阈值的像素被设置为一个值(通常是最大值)，所有低于阈值的像素被设置为另一个值(通常是最小值)。其操作可以表示为：

$$B(x,y) = \begin{cases} 0 & \text{if } I(x,y) < T \\ 255 & \text{if } I(x,y) > T \text{ 或 } I(x,y) = T \end{cases} \quad (4.1.4.1)$$

其中， $I(p)$ 是原始图像中像素  $p$  的灰度值， $J(p)$ 是二值化后图像中像素  $p$  的值， $B$  代表背景像素值(通常是 255)， $W$  代表前景像素值(通常是 0)。应用阈值后，图像中的每个像素只可能有两个值，即  $B$  或  $W$ 。这样，图像就被简化为只有两种颜色的图像，可以更清晰地看到图像中的重要特征。本模型中为了更加突出模型的特点我们主要进行了补体变换、二值化和最终的按位取反，大大提高了模型的可视化效果，和图像处理算法的效率和准确性。

#### 4.1.5 对问题一模型的特征提取

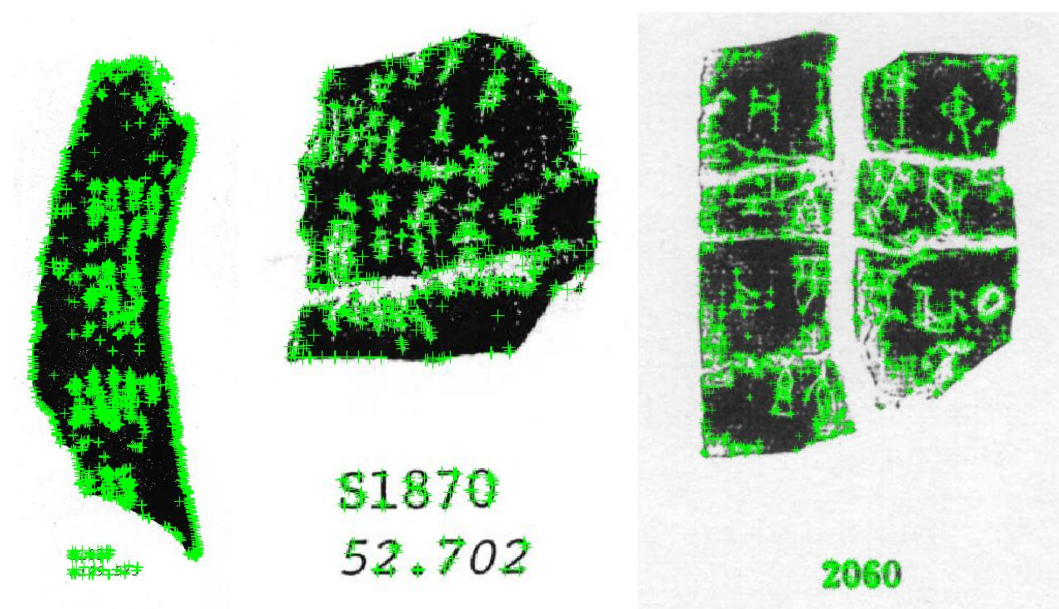


图 4 模型一的特征提取

特征提取是计算机视觉和图像处理领域的核心环节，在简化数据、突出重要信息、提高鲁棒性、促进图像匹配和识别、支持机器学习和模式识别、增强图像分析的可解释性、提高算法效率方面有重要作用。在模型中我们主要应用 BRISK(Binary Robust Invariant Scalable Keypoints)算法<sup>[6]</sup>提取甲骨文拓片预处理后的照片中的关键点和描述子，BRISK 算法可以表述为：

##### (1)尺度空间特征点检测

通过构建图像金字塔实现多尺度空间并检测关键点，其中包括多个 octaves 和 intra-octaves.每个 octave 由原始图像通过因子 2 进行下采样得到，而每个 intra-octave

的采样因子为 1.5。在每个 octave 和 intra-octave 层中，使用 FAST9-16 算法来检测潜在的兴趣区域，通过设定一个阈值 T 来区分显著的特征点。

## (2)非极大值抑制

在检测到潜在的特征点后，算法进行非极大值抑制以筛选出最具代表性的点。对于每个检测到的点，算法检查其同层以及上下层邻域内的 FAST 响应值，确保选定的点在邻域内是最大的。用以确保了特征点的显著性和尺度不变性。

## (3)特征点位置和尺度的精确估计

对于通过非极大值抑制筛选出的点，BRISK 算法通过二次函数拟合来精确估计其位置和尺度。这一步骤涉及最小二乘法，通过最小化误差来确定最佳拟合参数。

## (4)特征点方向的确定

为了实现旋转不变性，算法需要确定每个特征点的主方向。这是通过对特征点周围的采样区域进行旋转，使得最大梯度方向与水平方向对齐来实现的。特征点的方向可以表示为：

$$\theta = \arctan\left(\frac{g_y}{g_x}\right) \quad (4.1.5.1)$$

其中， $g_x$ 和 $g_y$ 是沿着特征点主方向的梯度分量。

## (5)特征点描述子的生成

BRISK 算法生成的描述子是一个二进制向量，长度为 512 位。描述子的生成基于关键点周围的固定样本模式点，这些点均匀分布在以关键点为中心的同心圆上。对于每一对样本点，算法比较它们的亮度值，并根据比较结果生成二进制位。具体来说，如果点 $p_j$ 的亮度大于点 $p_i$ ，则对应的二进制位设为 1，否则设为 0。

## (6)特征点匹配

在特征点匹配阶段，BRISK 算法使用汉明距离(Hamming distance)来衡量两个描述子之间的相似性。汉明距离是两个等长二进制串之间不同位的数量，它提供了一种有效的方式来快速匹配特征点。

BRISK 算法较为成功有效的检测和描述了图像中的关键点，成功实现了对甲骨古文拓片的有效信息提取，为数字化分析识别图像奠定了基础。

## 4.2 问题二模型的建立与求解

问题二中需要实现对不同甲骨文原始拓片图像进行自动单字分割。本文利用题目附件中的甲骨文原始拓片图像及甲骨文的坐标建立了基本数据集，删除了缺失甲骨文坐标的部分图像。

## (1)YOLOv8 检测算法

YOLO(You Only Look Once)是一种 one-stage 目标检测算法，即仅需要“看”一次就可以识别出图片中物体的边界框和 class 类别。可以用于物体检测、实例分割和图像分类等任务。

YOLOv8 是一个 SOTA 模型，它建立在 Yolo 系列历史版本的基础上，并引入了新的功能和改进点，成功实现了模型的性能和灵活度的提高，使其成为实现图像分割、姿态估计、目标检测等任务的最佳选择。其具体创新点包括一个新的骨干网络、一个新的 Anchor-Free 检测头和一个新的损失函数，其可在 CPU 到 GPU 的多种硬件平台上运行。这个算法的核心思想是将目标检测问题转化为单个的回归问题。YOLO 算法将图像划分为一个固定数量的网格，然后在每个网格中预测对象的边界框和类别。这些边界框通过回归预测出来，每个边界框包含了一个对象的位置和大小信息，以及对该对象所属类别的置信度。

YOLOv8 目标检测算法主要核心组件分为 backbone、head 和 neck。

backbone 从输入图像中提取高级语义特征，以捕捉物体的形状、纹理和上下文信息；neck 的主要目标是将不同尺度的特征图进行融合，以便更好地检测不同大小和尺度的目标。融合后的特征图被传递给 head 进行目标检测。最新的单目标检测算法摒弃 Anchor Boxes，改用 Anchor-Free，解决了正负样本不平衡问题和定位不精准问题。同时 YOLOv8 借鉴 YOLOv7 中 ELAN 的并行梯度流思想，将 YOLOv5 中的 C3 模块改为 C2f，在增强梯度流的同时保持 YOLOv8 模型的轻量化。在解耦头改进上，YOLOv8 借鉴了 YOLOX 的解耦思想实现类别特征和位置特征的解耦合，提高检测效率。在标签分配上 YOLOv8 采用了和 YOLOv6 相同的策略，提升目标检测模型在复杂场景下的准确性和鲁棒性，并增加对目标位置和大小不确定性的建模能力。YOLOv8 作为更强大的单目标检测算法，更好地集成了单目标检测算法的优势，使得算法能够感知到不同尺度的目标，可以更好地适用在文字这类不同目标检测上。<sup>[7]</sup>

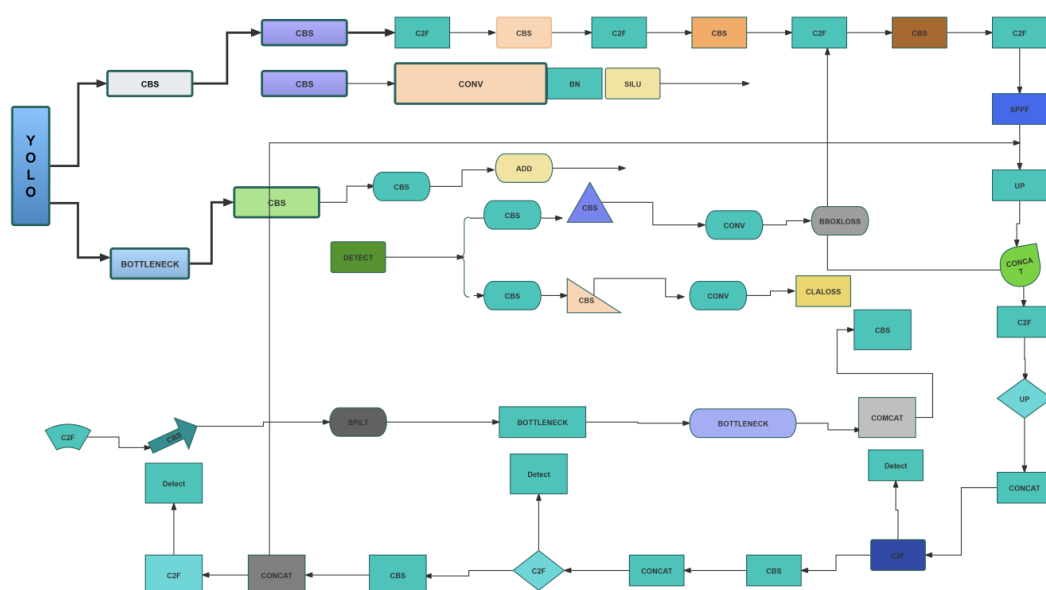


图 5 YOLO

YOLO 算法把输入图像分成  $S \times S$  个方格,每个方格输出一个  $B \times 5 + C$  维的张量。这里,  $B$  是每个方格预测方框(Bounding box)的数目,  $C$  表示需要检测的对象类别数目。如果检测目标的中点落在某个方格内,那么目标物体就由该方格进行检测与输出。

YOLO 算法采用了相对坐标的方法,并使用 Logistic 激活函数,使每个坐标的取值落在区间 0 到 1。目标检测网络在每个边界框处预测 5 个值,分别是  $t_x$ 、 $t_y$ 、 $t_w$ 、 $t_h$  和  $t_o$ 。如果单元格相对图像左上角偏移记为  $(C_x, C_y)$ , 并且先验边界框宽度、高度分别记为  $p_w$ 、 $p_h$ , 对应的预测结果为:

$$b_x = \sigma(t_x) + C_x \quad (4.2.1)$$

$$b_y = \sigma(t_y) + C_y \quad (4.2.2)$$

$$b_w = p_w e^{t_w} \quad (4.2.3)$$

$$b_h = p_h e^{t_h} \quad (4.2.4)$$

$$Pr(object) * IOU(b, object) = \sigma(t_o) \quad (4.2.5)$$

### (1)卷积操作

YOLOv8 使用卷积操作来提取图像特征。卷积操作通过将一个小尺寸的滤波器与图像进行叠加,以提取图像中的局部特征。在 YOLOv8 中,卷积操作使用 ReLU 激活函数来增加非线性,并使用批量归一化来减少数据分布的偏差。

公式:

$$y = x * w + b + \frac{k*(s-h)}{M} \quad (4.2.6)$$

其中  $y$  是卷积后的输出,  $x$  是输入图像,  $w$  是滤波器的权重,  $b$  是偏置项,  $k$  是空间偏差项,  $s$  是输入图像的大小,  $h$  是滤波器的大小,  $M$  是批量归一化的参数。

### (2)池化操作

YOLOv8 使用池化操作来减小特征图的尺寸,以提高模型的效率和准确性。池化操作可以将特征图中的像素值进行聚合和压缩,以减少计算量和参数数量。YOLOv8 中常用的池化方式包括最大池化和平均池化。

数学公式表示为:

$$P = \left\{ \frac{f(i,j)}{\sum(i,j)} \right\} \quad (4.2.7)$$

其中  $P$  是池化后的特征图,  $f(i,j)$  是输入特征图中的像素值,  $\sum(i,j)$  是特征图中的所有像素值的和。

### (3)损失函数

损失函数是衡量模型预测误差的指标,用于优化模型的参数以提高预测精度。在 YOLOv8 中,常用的损失函数包括分类损失和定位损失。分类损失用于衡量模型的分类精度,定位损失用于衡量模型的定位误差。YOLOv8 还使用交叉损失和 Huber 损失来处理不同类型的数据分布和数据倾斜问题。

数学公式表示为:

$$L = -\sum y_i * \log(p_i) + \sum q_{ij} * (x_j - b_j)^2 \quad (4.2.8)$$

其中 L 是总损失函数,  $y_i$  是真实标签,  $p_i$  是模型预测的概率分布,  $q_{ij}$  是真实标签和模型预测的交集部分的数量,  $x_j$  和  $b_j$  分别是预测和真实的目标位置坐标。

### (4)反卷积操作

YOLOv8 中还使用反卷积操作来对特征图进行上采样,以恢复目标的原始尺寸和分辨率。反卷积操作与卷积操作相反,它通过将权重反向应用于特征图上的像素值,以重建原始特征图。

数学公式表示为:

$$x = W^T * P * S^{-1} + b \quad (4.2.9)$$

其中 x 是反卷积后的输出特征图, W 是反卷积权重矩阵, P 是池化后的特征图, S 是上采样的尺寸。

作为一种先进的目标检测算法, YOLOv8 拥有高性能、Anchor-Free 检测、多样化的任务支持、可扩展性、多尺度模型等优势,同时也存在对小目标检测的挑战、对遮挡和旋转的鲁棒性有待提高、计算的资源要求受限、新算法的挑战和改进等问题。我们经过综合考虑,认为对其的应用可以在对甲骨文拓片图片的识别建模中发挥最优效果,具体请看本文 5.2 对问题二建模的分析与检验。

## 4.3 问题三模型的建立与求解

问题三建立在问题二的基础上,对问题二中建立的甲骨文原始拓片图像进行自动单字分割模型进行应用,实现了对附件 3(Test 文件夹)中的 200 张甲骨文原始拓片图像进行自动单字分割,并将处理结果成功存放在“Test\_results.xlsx”中,此文件已经成功单独上传至竞赛平台。

4. 4 问题四模型的建立与求解

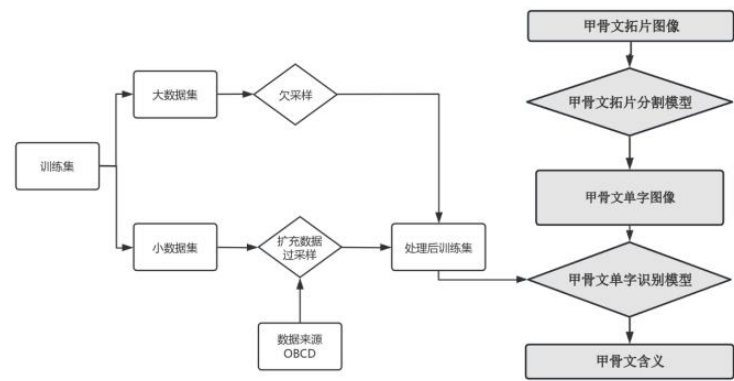


图 6 问题四流程图

在问题四的研究中，再次采用了 YOLO 模型作为目标检测算法，用于处理图像中的物体检测和识别任务。本问需要处理的部分为对甲骨文单字含义的识别。题目附件提供的练习集以现代汉字命名文件夹，文件夹下为该现代汉字对应的甲骨文单字图像集。本文以现代汉字为其对应的甲骨文单字图像的 label，标注范围为整张图像（也即认为整张甲骨文单字图像即代表该现代汉字，事实上也确实如此）进行处理，建立起问题四模型训练的基本数据集。同时，本文在初步训练后发现，由于各个汉字数据集规模差距巨大（几千张到几十张），导致不同汉字的识别效果差距也较大。为了矫正这种情况，数据集针对甲骨文图像较多的汉字进行欠采样，使其进入数据集的数量不大于 200；同时对于部分识别效果非常不理想的汉字，本文从 OBCD（Oracle Bone Character Detection）中获取了足够的扩充数据集的图像（图像保存在上传的附录的 datasets\_pack 文件夹中）。

在训练 YOLOv8 模型时，高质量的数据标注起着至关重要的作用。这些标注不仅为模型提供了学习的基础，而且确保了模型能够从图像中正确地提取甲骨文字符的特征。通过卷积神经网络(CNN)<sup>[8]</sup>模型能够学习到图像中的局部特征，并结合多尺度检测来识别不同大小的甲骨文字符。这一过程涉及到模型对图像特征的深入理解和学习，以便能够准确地预测字符的位置和类别。

在训练过程中，还对模型的超参数进行了调整。这种调整有助于提高模型的准确率和召回率，从而确保模型能够有效地识别和定位甲骨文字符。超参数的优化是一个关键步骤，因为它直接影响到模型的性能和最终的识别效果。

使用训练好的 YOLOv8 模型对甲骨文原始拓片图像进行识别时，首先需要对图像进行分割，这一步骤将图像中的单个字符分离出来，为后续的识别分析打下基础。然后，模型对分割后的图像进行识别<sup>[9]</sup>，最终输出每个字符所属于的类别。

处理后的图像存储在“附件\results\results\_04”中，其中每个文件夹（以测试集中的图像文件命名）包括甲骨文拓片图片原图像，和切割后的甲骨文单个文字图像，甲骨文单字图像的文件名即为模型识别得到的现代汉字，若显示“none”则说明该甲骨文的含义不存在于训练集给出的现代汉字中、亦或是模型预测的失误。



在本次识别过程中，发现有部分甲骨文字符并没有被正确识别。经过观察，认为这可能是因为这些字符在数据集中没有对应的图像，或者字符被大面积的刮痕覆盖，导致置信度过低而无法准确识别。这一发现指出了数据集的完整性和质量对于模型性能的重要性，同时也提示了在未来工作中需要对数据集进行进一步的扩充和清洗，以提高模型的识别能力。

## 五、模型的分析与检验

### 5.1 问题一模型的分析与检验

现对 4.1 中建立的甲骨文拓片图片预处理模型进行不同模型的比较、分析和检验。  
在问题一模型的预处理中：

(1) 对模型进行了去噪处理，通过应用均值滤波和高斯滤波两种方法进行了对比，并使用信噪比(SNR)10、峰值信噪比(PSNR)或结构相似性(SSIM)等指标来评估去噪效果，最后认为应用高斯滤波对甲骨文拓片图片进行预处理更能较为完整的保留图片的边缘和细节特征，对甲骨文拓片图片的进一步处理如分割、识别等产生的不利影响较小。

(2) 为了对图片进一步去噪处理，并且提高图片的计算机视觉效果较少误检、漏检的情况，我们应用中值滤波对图片进一步进行平滑处理。

(3) 前两步预处理后，甲骨文拓片图片存在一定的失真情况，因此我们应用直方图增强作为一种复原技术，尝试恢复图像的原始分布，并在一定程度上改善图像的质量。

(4) 为简化图像，方便对图像的进一步处理和分析，方便对图像的特征提取、目标分割，模型对图像进行了二值化处理。同时为了图像有更明显的视觉效果，方便判断模型的建立是否成功，我们进一步对二值化结果进行了取反的操作，以提高对比度。

(5) 为更好的分析、理解、分类和识别图像，应用 BRISK 模型对图片进行特征提取，成功实现数据降维、增强了鲁棒性、提高了识别准确性、简化了决策过程、压缩率图像、实现目标跟踪与检测。

最终问题一模型所得对甲骨文拓片图片的预处理结果较为成功，大大方便了计算机视觉对图片识别处理，建模较为成功。

### 5.2 问题二模型的分析与检验

现对 4.2 中建立的 YOLOv8 模型对图像分割的结果进行不同维度的模型评估。

YOLOv8 在训练完成之后，会在 run 目录下把训练的过程一些参数与结果示意图保存下来，其目标检测性能指标如下：



## (1) 混淆矩阵

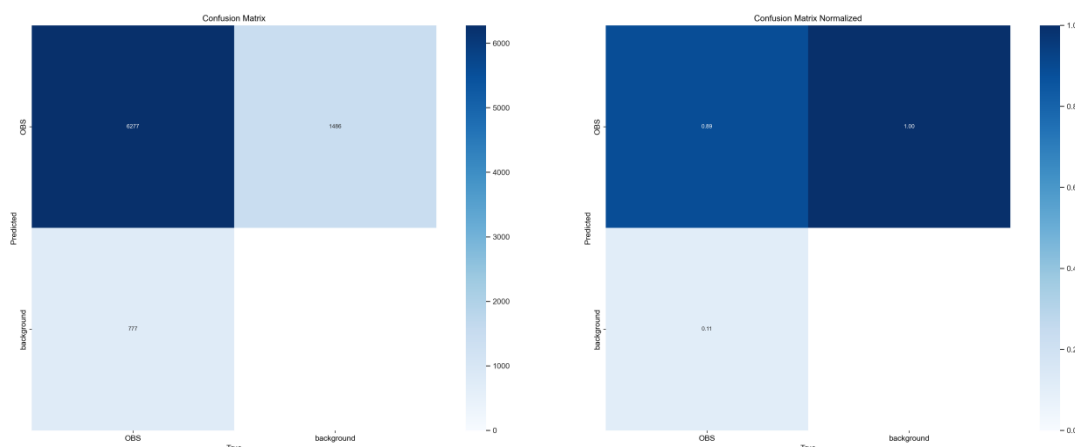


图 7 混淆矩阵

为了更加直观了解分类模型的错误类型，添加混淆矩阵进行检验，重点在了解模型是否混淆了两个不同的类别，是否错误的将一个类别预测为另一个类别。这种详细的分析有助于克服仅使用分类准确率带来的局限性。

图 6 中其中各个值从上到下，从左到右分别为

TP(True Positive)表示将正类预测为正类的数量，即正确预测的正类样本数。

FP(False Positive)表示将负类预测为正类的数量，即错误预测的负类样本数。

FN(False Negative)表示将正类预测为负类的数量，即错误预测的正类样本数。

TN(True Negative)表示将负类预测为负类的数量，即正确预测的负类样本数。

精确率(Precision)和召回率(Recall)是常用于评估分类模型性能的指标，其计算方法如下：

精确率(Precision):

$$Precision = \frac{TP}{TP+FP} \quad (5.2.1)$$

计算得模型的 Precision=80.9%

精确率是指在所有被模型预测为正例(Positive)的样本中，实际为正例的比例。它衡量了模型在正例预测中的准确性。

召回率(Recall):

$$Recall = \frac{TP}{TP+FN} \quad (5.2.2)$$

计算得模型的 Recall=89.0%

召回率是指实际为正例在所有被模型预测为正例(Positive)的样本的比例。它衡量了模型在正例预测中的准确性。

## (2) F1 曲线

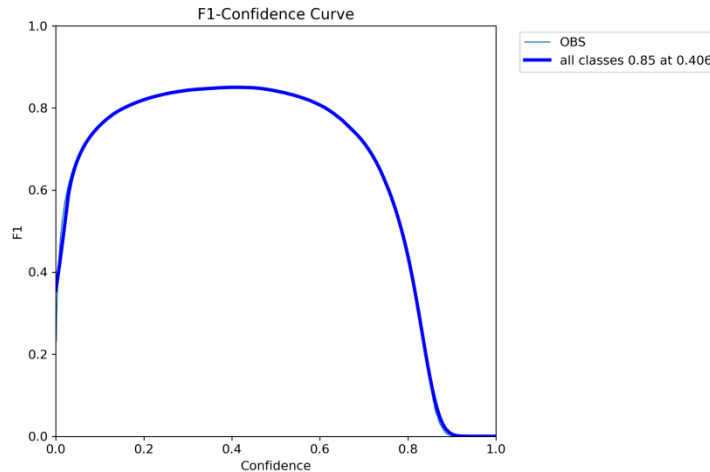


图 8 F1 曲线

为了能够评价不同算法的优劣，在 Precision 和 Recall 的基础上提出了 F1 值的概念，来对 Precision 和 Recall 进行整体评价。F1 的定义为：

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (5.2.3)$$

$$F_1 = \frac{2TP}{2TP + FN + FP} \quad (5.2.4)$$

F1 曲线是多分类问题中常用的一种性能评估工具。它基于精确率和召回率的调和平均数的 F1 分数，取值范围介于 0 和 1 之间。1 代表最佳性能，而 0 代表最差性能。通常情况下，通过调整置信度阈值(判定为某一类的概率阈值)，可以观察到 F1 曲线在不同阈值下的变化。在阈值较低时，模型可能将许多置信度较低的样本判定为真，从而提高召回率，同时降低精确率。而在阈值较高时，只有置信度很高的样本才被判定为真，使得模型的类别判定更为准确，从而提高精确率。

理想状态下，F1 曲线显示在置信度为 0.4-0.6 的区间内取得了较好的 F1 分数。表明在这个范围内，模型在平衡精确率和召回率方面表现较为理想。

计算得模型的  $F_1 = 84.7\%$

### (3) Precision-Confidence 曲线

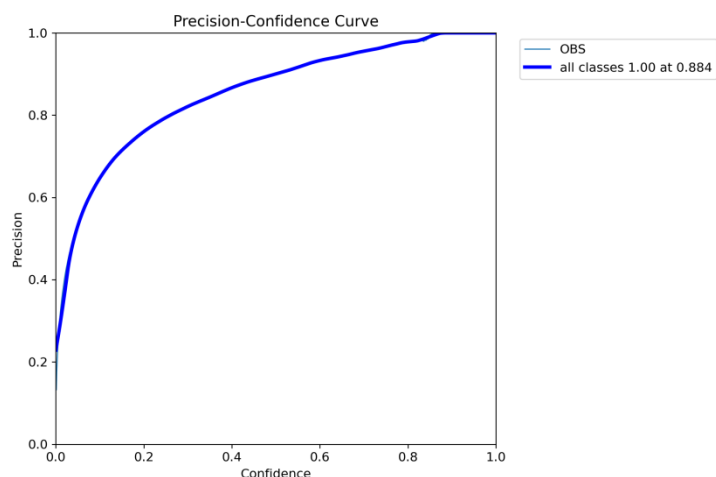


图 9 PCC

PCC 图的横坐标表示检测器的置信度，纵坐标表示精度(或召回率)。曲线的形状和位置反映了检测器在不同信心水平下的性能。

在 PCC 图中，当曲线向上并向左弯曲时，表示在较低置信度下仍能保持较高的精度，说明检测器在高召回率的同时能够保持低误报率，即对目标的识别准确性较高。

相反，当曲线向下并向右弯曲时，说明在较高置信度下才能获得较高的精度，这可能导致漏检率的增加，表示检测器的性能较差。

因此，PCC 图对于评估检测器在不同信心水平下的表现提供了有用的信息。在图中，曲线向上并向左弯曲是期望的效果，而曲线向下并向右弯曲则表示改进的空间。

如图的曲线向上并向左弯曲，能发现检测器在 0.4~0.6 时的置信度下已经能保持较高的精度，表明了对目标的识别准确性较高。

### (4) Recall-Confidence 曲线

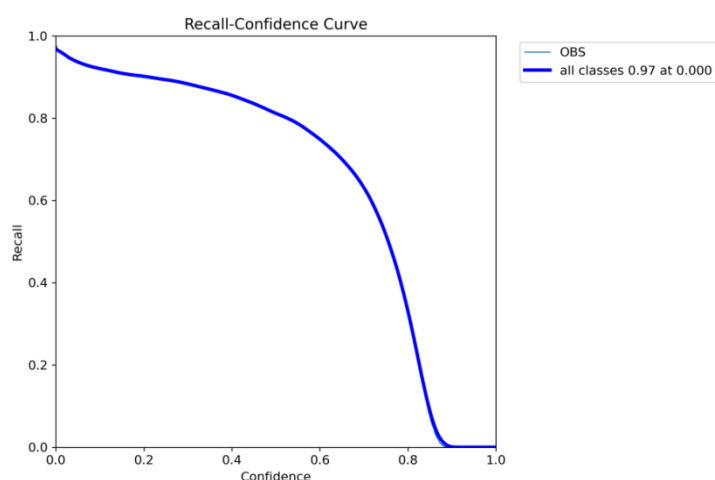


图 10 RCC

在理想情况下，希望算法在保持高召回率的同时能够保持较高的精度。

在 RCC 图中，当曲线在较高置信度水平下呈现较高召回率时，说明算法在目标检测时能够准确地预测目标的存在，并在过滤掉低置信度的预测框后依然能够维持高召回率。这反映了算法在目标检测任务中的良好性能。

同时应注意到，RCC 图中曲线的斜率越陡峭，表示在过滤掉低置信度的预测框后，获得的召回率提升越大，模型的检测性能得到提高。

在图表中，曲线越接近右上角，表示模型性能越好。当曲线靠近图表的右上角时，说明模型在保持高召回率的同时能够维持较高的精度。因此，RCC 图可用于全面评估模型性能，帮助找到平衡模型召回率和精度的合适阈值。

## (5)P-R 曲线

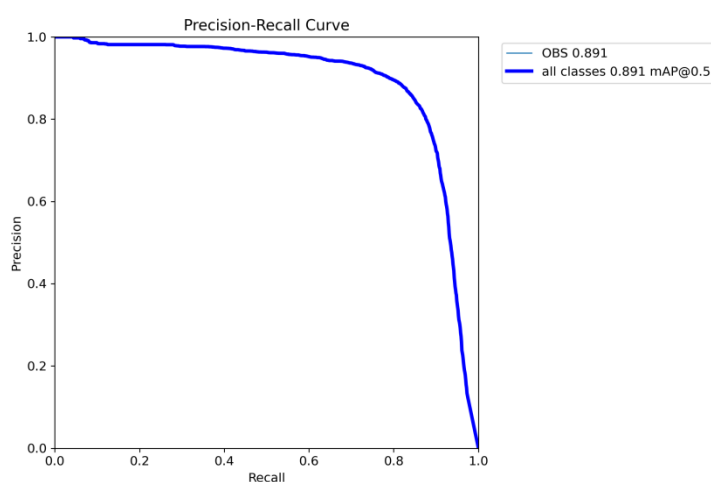


图 11 PR-curve

PR-curve 是精确率(Precision)和召回率(Recall)之间的关系。精确率表示预测为正例的样本中真正为正例的比例，召回率表示真正为正例的样本中被正确预测为正例的比例。

在 PR Curve 中，横坐标表示召回率，纵坐标表示精确率。通常情况下，当召回率降低时，精确率会升高，反之亦然。PR Curve 反映了这种取舍关系。曲线越靠近右上角，表示模型在预测时能够同时保证高的精确率和高的召回率，即预测结果较为准确。相反，曲线越靠近左下角，表示模型在预测时难以同时保证高的精确率和高的召回率，即预测结果较为不准确。

通常，PR Curve 与 ROC Curve 一同使用，以更全面地评估分类模型的性能。PR Curve 提供了对模型在不同任务下性能表现的更详细的洞察。

如图的曲线接近图标的右上角，表示模型在高召回率的同时维持了较高的精度，检测性能较好，预测结果较为准确。

## (6) 损失函数

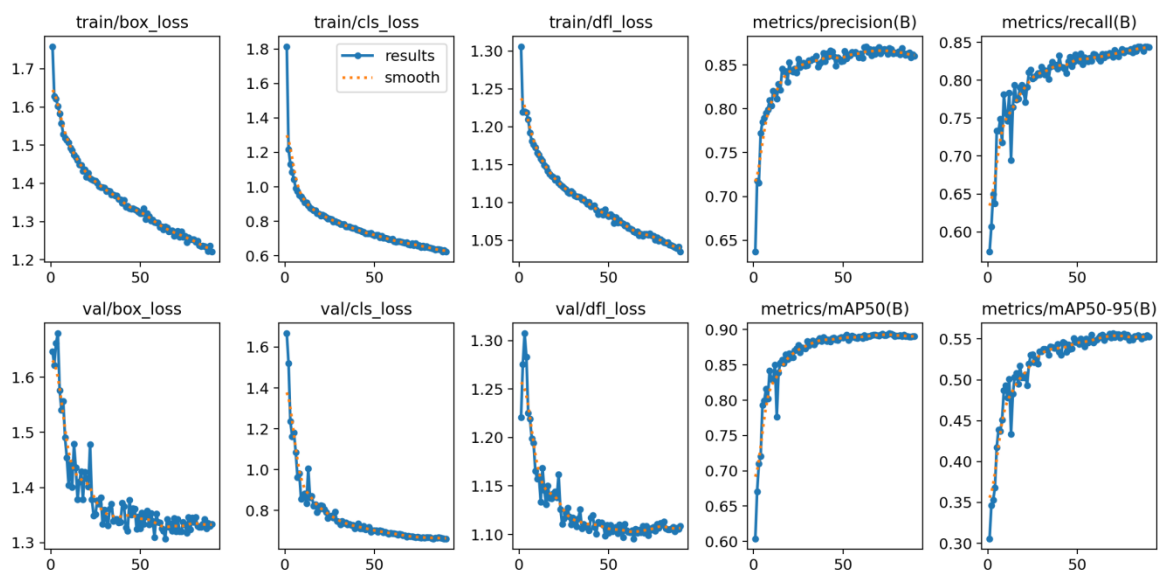


图 12 损失函数

损失函数在目标检测任务中扮演关键角色，它用于衡量模型的预测值与真实值之间的差异，能够直接影响模型性能。以下是一些与目标检测相关的损失函数和性能评价指标的解释：

定位损失(box\_loss)是衡量预测框与标注框之间的误差，通常使用 **GIoU(Generalized Intersection over Union)**来度量，其值越小表示定位越准确。其目的是通过最小化定位损失，使模型能够准确地定位目标。

置信度损失(obj\_loss)是计算网络对目标的置信度，通常使用二元交叉熵损失函数，其值越小表示模型判断目标的能力越准确。其目的是通过最小化置信度损失，使模型能够准确判断目标是否存在。

分类损失(cls\_loss)是计算锚框对应的分类是否正确，通常使用交叉熵损失函数，其值越小表示分类越准确。其目的是通过最小化分类损失，使模型能够准确分类目标。

**Precision(精度)**是正确预测为正类别的样本数量占有所有预测为正类别的样本数量的比例。其目的是衡量模型在所有预测为正例的样本中有多少是正确的。

**Recall(召回率)**是正确预测为正类别的样本数量占有所有真实正类别的样本数量的比例。目的是衡量模型能够找出真实正例的能力。

**mAP(平均精度)**: 是使用 **Precision-Recall** 曲线计算的面积，**mAP@[.5:.95]** 表示在不同 **IoU** 阈值下的平均 **mAP**。其综合考虑了模型在不同精度和召回率条件下的性能，是目标检测任务中常用的评价指标。

在训练过程中，通常需要关注精度和召回率的波动情况，以及 **mAP@0.5** 和 **mAP@[.5:.95]** 评估训练结果。这些指标可以提供关于模型性能和泛化能力的有用信息。

模型评估汇总

性能指标	精确率 (Precision)	召回率 (Recall)	$F_1$	PCC	RCC	PR-curve	损失函数
指标数据	80.9%	89.0%	84.7%				
性能评价	优	优	优	优	中	优	优

最终分析得问题二模型建立较为成功，可以对甲骨文原始拓片图像进行较为快速的自动单字分割，并有较高的精确度与准确性，检测器有较高的性能，建模较为成功。

### 5.3 问题三模型的分析检验

利用问题二模型建立的甲骨文拓片图片切割模型，对附件 3(Test 文件夹)中的 200 张甲骨文原始拓片图像进行自动单字分割，完成之后与原图进行比对，发现仍有部分人工纹理被误判为甲骨文。由于人工纹理的特征为在图像中灰度值较高，显灰白色，基于这个特点，本文对原始图像进行预处理，主要是将灰度值较高的像素用白色像素覆盖。经过对比，预处理之后图像中人工纹理明显减弱，模型预测失误率减少。

在对结果图像分析之后，发现模型预测正确度较高，切割字体的完整度较强，模型训练结果较为成功。

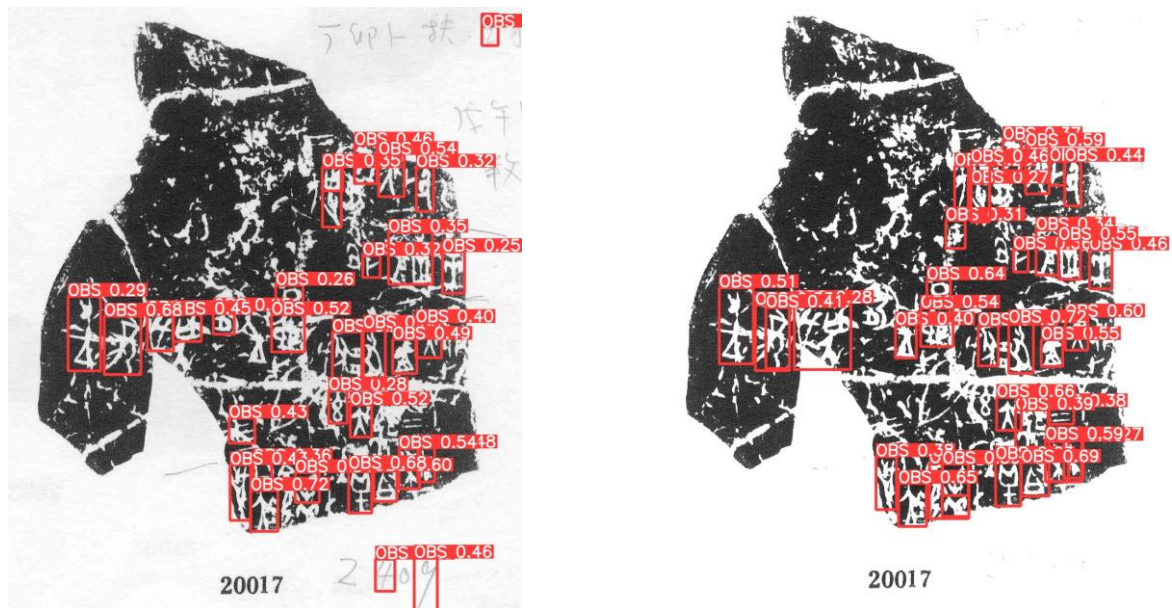


图 13 预处理结果对甲骨文拓片图像识别的影响

### 5.4 问题四模型的分析检验

问题四模型在综合应用了问题一、二、三模型的基础上增加了对分割区域的识别部分，其综合性能一定程度依赖于图片预处理结果和甲骨文拓片图片分割模型的精确

度和速度(具体见 5.1 问题一模型的分析检验和 5.2 问题二模型的分析检验)。

就问题四而言，对甲骨文分割结果的精确识别一方面依赖于数据集的数量和准确度，应用越多的甲骨文拓片的精确文字照片进行训练，模型对于甲骨文拓片图片的识别能力在一定程度上就越强，另一方面甲骨文的精确识别还依赖于应用模型的识别能力和速度，依赖于更优化的模型 YOLOv8 虽然就目前而言在图片的分割处理和精确识别中拥有较大优势，但若在处理较小图片和更为复杂图片中，如文字重叠，就会表现乏力，还需要对其不断改进提高增强性能。

## 六、模型的评价与改进

### 6.1 模型的优点

(1)在问题一甲骨文拓片图片预处理模型中，较为综合的考虑了各种可能影响机器识别甲骨文图片的因素，可以应用于各种情况下的甲骨文拓片图片，预处理具有普遍性且效果较好。

(2)在问题二甲骨文的准确分割模型中，应用处理该问题性能相对较强的图像处理模型 YOLOv8 模型，应用卷积神经网络进行训练提高了模型性能，并对各种预处理图片进行训练最终确定最适宜 Train 文件夹中甲骨文拓片原始图象分割的模型，使模型有该情况下较高的综合性能。

### 6.2 模型的不足

(1)甲骨文原始拓片文字识别模型中训练集较少且部分训练集不够准确，导致模型在识别方面仍有一定的性能提升的空间。

(2)甲骨文拓片照片受拍摄时声音、光线或甲骨文拓片本身收到损坏等因素导致诸多文字存在不完整、不清晰等情况无法对其精确识别。

(3)由于计算资源有限，不能在更大规模的神经网络中训练模型，且训练参数中的 epoch、batch 数值都较小，模型的精度难以在有限的时间内提高。

### 6.3 模型的推广和改进

(1)可以增加模型数据集的数量，提高数据集的准确度，提高模型性能。

(2)可以考虑应用更多神经网络算法（如 MobileNet、R-FCN、FPN 等）与 YOLO 模型结合，完善模型不足的部分，以期获得更好的性能。

(3)通过循环式生成对抗神经网络(CycleGAN)或者 AIGC，建立一个可以自行学习甲骨拓片风格的、可以生成具有丰富多样性的字符图片的算法，来合成数据提供给识别模型强化训练，以解决甲骨文字符数据获取的困难。

(4)超参数优化：继续优化模型的超参数，使用如网格搜索、随机搜索或贝叶斯优化等方法来寻找最佳的超参数组合。



## 附件

(1) 文件夹 `code`:

- 1) 问题一代码——matlab 实现
- 2) 问题二代码——python 实现
- 3) 问题三代码——python 实现
- 4) 问题四代码——python 实现

(2) 文件夹 `datasets_pack`: 第四题补充数据集

(3) 文件夹 `models_data`

- 1) 第二问模型数据
- 2) 第四问模型数据

(4) 文件夹 `results`:

- 1) `results1`——问题一结果
- 2) `results2`——问题二结果
- 3) `results3`——问题三结果
- 4) `results4`——问题四结果



## 参考文献

- 
- [1] 王浩彬.基于深度学习的甲骨文检测与识别研究[D].华南理工大学,2019.001469.
- [2] 莫伯峰,张重生.人工智能在古文字研究中的应用及展望[J].中国文化研究,2023,(02):47-56.2023.02.012.
- [3] 周作梅,李俊杰.基于高斯滤波的低照度图像信息增强方法[J].信息与电脑(理论版),2022,34(17):202-204.
- [4] 肖莉.一种基于中值滤波的图像预处理改进算法[J].高师理科学刊,2023,43(10):41-44.
- [5] AU Zheng, Hao.TI The research of image' color enhancement alogrithms base on histogram equalization.DT Dissertation/Thesis.PY 2009
- [6] AU Cheng, ML.AU Cheng, ML.AF Cheng, Min-Lung.Matsuoka, Masashi.TI An Enhanced Image Matching Strategy Using Binary-Stream Feature Descriptors.SO IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.PY 2020.
- [7] 鲍禹辰,徐增波,田丙强.基于 YOLOv8 改进的服装疵点检测算法[J/OL].东华大学学报(自然科学版).2023.0296.
- [8] 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,2021.003746.
- [9] 张泽骏,李健钊,吴武林,钟洁凤,吴亚榕,林钦永.基于分量分割与 YOLOv8s 的红火蚂蚁巢识别[J].现代信息科技,2023,7(20):66-74