

# 基于改进 YOLOv5s 的白酒瓶盖瑕疵检测

王军<sup>1,2</sup>, 万书东<sup>1</sup>, 程勇<sup>1,2</sup>

(1.南京信息工程大学 软件学院, 南京 210044; 2.南京信息工程大学 科技产业处, 南京 210044)

**摘要:** **目的** 瓶装白酒生产过程中, 瓶盖表面瑕疵会影响产品外观质量。针对白酒瓶盖表面瑕疵检测效率低和目标检测效果差的问题, 提出一种基于 YOLOv5s 的改进算法 DTS-YOLO。**方法** 首先, 在主干网络中引入可变形卷积, 以提高模型对极端长宽比瑕疵的检测精度。其次, 引入 Transformer 编码块, 使网络聚焦于提取图像的全局信息。最后, 在颈部网络构建 C3SE-Lite 模块, 将 C3 模块嵌入 SE 注意力模块的同时引入 Ghost 卷积, 减少参数数量的同时, 增强对瓶盖瑕疵的检测能力。**结果** 实验结果表明, 本文所提方法相较于基础网络, 参数量减少了 10%, 平均精度均值达 95%, 平均检测速度达 30 帧/s。**结论** 本文方法有效实现了白酒瓶盖表面瑕疵快速、准确地检测, 可广泛应用于瓶装白酒生产过程中瓶盖表面检测。

**关键词:** YOLOv5s; 瑕疵检测; 可变形卷积; Transformer 编码块; 注意力机制

**中图分类号:** TB487 **文献标志码:** A **文章编号:** 1001-3563(2024)07-0180-09

**DOI:** 10.19554/j.cnki.1001-3563.2024.07.023

## Liquor Bottle Cap Defect Detection Based on Improved YOLOv5s

WANG Jun<sup>1,2</sup>, WAN Shudong<sup>1</sup>, CHENG Yong<sup>1,2</sup>

(1. School of Software, Nanjing University of Information Science and Technology, Nanjing 210044, China;

2. Science and Technology Industry Division, Nanjing University of Information Science and Technology, Nanjing 210044, China)

**ABSTRACT:** In production of bottled liquor, there are usually surface defects on the bottle cap that affect the quality of the product. The work aims to propose an improved algorithm model DTS-YOLO based on YOLOv5s to solve problems of low detection efficiency of blemishes on the surface of liquor bottle caps and poor detection of small targets. First, deformable convolution was introduced into the backbone network to improve the detection accuracy of the model for extreme aspect ratio defects; Secondly, the Transformer coding block was incorporated into the backbone network to make the backbone network focus on the extraction of global information of the image; Finally, influenced by Inspired by the C3 module in YOLOv5s, the C3SE-Lite module was designed. The C3 module was embedded in the SE attention module and the GhostConv convolution was introduced at the same time, so that the model could reduce the number of parameters while enhancing the ability to detect defects. The experimental results showed that under the premise of reducing the number of parameters by 10%, the average precision of the method in this paper reached 95%, and the average detection speed was 30 f/s. The method presented in this paper can effectively detect the surface defects of bottle caps quickly and accurately, and can be widely applied to the surface detection of bottle caps during the production of bottled liquor.

**KEY WORDS:** YOLOv5s; flaw detection; deformable convolution; Transformer coding block; attention mechanism

当前,白酒产业竞争激烈,我国白酒企业数量由 2016 年的 1 593 家减至 2022 年 963 家<sup>[1]</sup>。厂家为提升产品竞争力以避免被淘汰,在注重酒本身口感的同时,越来越重视白酒包装的外观设计。但是,在瓶装白酒生产过程中,受生产环境、设备和制造工艺等因素影响,白酒瓶盖出现多种瑕疵,如破损、变形、坏边、打旋、断点和喷码异常等,极大影响瓶装白酒的外观质量,进而影响消费者的购买欲望。从生产角度来看,瑕疵种类繁多,目前主要依赖人工检测,无法实现实时高精度检测,还消耗大量人力、物力。因此,在瓶装白酒生产过程中,实现白酒瓶盖表面瑕疵快速、准确地检测能够起到非常重要的作用,既可以加快生产速度,又可以提高厂家竞争力。传统的机器学习方法,在一定程度上解决了瑕疵检测问题,但是检测精度和速度都有待进一步提高,不适用于复杂或多样化的瑕疵检测<sup>[2]</sup>。随着计算机视觉技术的发展,基于深度学习的目标检测方法逐渐成为主流。

目前基于深度学习的目标检测算法主要分为 2 类,一类是基于区域建议的两阶段目标检测算法,另一类是基于回归思想的单阶段目标检测算法。两阶段目标检测算法主要包括 Fast R-CNN 系列<sup>[3-4]</sup>和 Mask R-CNN<sup>[5]</sup>,使用感兴趣区域从输入图像中执行分类和边界框回归。两阶段检测器的检测性能好,但因模型参数多、计算量大,检测速度较低,难以满足工业实际场景需求。相比之下,具有高检测速度的单阶段算法更受欢迎,如 YOLO (You Only Look Once) 系列和 RetinaNet<sup>[6]</sup>目标检测算法。白酒瓶盖表面瑕疵尺寸跨度大,小目标瑕疵多,难以检测,从而导致漏检和误检的问题。为此,段禄成等<sup>[7]</sup>和李玉洁等<sup>[8]</sup>采用改进 YOLOv3 算法检测瓶盖瑕疵,然而其忽略瑕疵尺度跨度大的问题,对全局特征的提取效果不佳。此外,

车璇等<sup>[9]</sup>采用 RetinaNet 算法检测瓶盖瑕疵,但该方法对瑕疵的注意力较弱,无法完成对复杂背景瑕疵的准确识别任务。研究者们<sup>[7-9]</sup>提出的模型参数量较大,未充分考虑到模型后期部署问题,导致模型实用性差。本文综合考虑白酒瓶盖表面瑕疵检测过程中存在瑕疵尺寸变化大、小目标检测效果差等问题,提出基于 YOLOv5s 改进的算法模型 DTS-YOLO,兼顾检测实时性和实用性。本文的主要贡献如下:

- 1) 为应对白酒瓶盖中极端尺寸比例瑕疵,本文在主干网络 (backbone) 中采用 DCNv2 (Deformable Convolutional Networks v2) 卷积层<sup>[10]</sup>,通过学习每个采样点的权重和控制卷积核采样点的偏移量,减少无关因素干扰,进而提高模型精度。
- 2) 为提高对白酒瓶盖瑕疵中众多小目标瑕疵的检测能力,本文在主干网络 (backbone) 末端 C3 中引入 Transformer 编码器<sup>[11]</sup>,通过融合全局信息来给予更高级别的检测特征,进而加强模型对瑕疵的检测性能。
- 3) 为增强模型对瑕疵的检测能力,并将大量的训练资源集中在关键特征上。本文构建 C3SE-Lite 模块,在提高模型精度的同时降低模型参数量,更易于设备部署。

## 1 YOLOv5 原理

YOLOv5 是一种经典的单阶段目标检测算法。根据网络深度和特征图宽度的不同,YOLOv5 可细分为 YOLOv5n、YOLOv5s、YOLOv5m、YOLOv5l 和 YOLOv5x 5 种版本,以适配不同的检测任务。本文综合考虑模型的准确性和参数量大小,采用 YOLOv5s 作为基准模型并对其进行改进。图 1 展示了 YOLOv5 网络模型的整体结构框架。

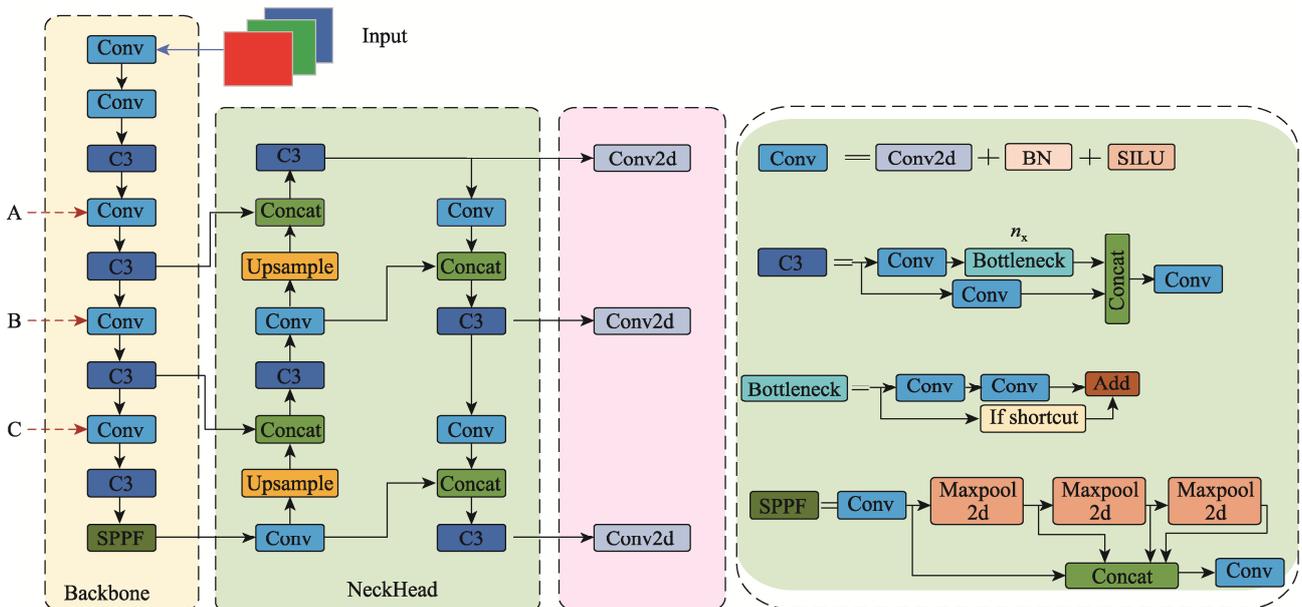


图 1 YOLOv5 网络结构  
Fig.1 YOLOv5 network structure

YOLOv5 网络结构分为 4 个主要部分：输入端 (Input)、主干网络 (Backbone)、颈部网络 (Neck)、输出端 (Head)。输入端 (Input) 采用 Mosaic 数据增强技术、自适应图像缩放和自适应锚框技术<sup>[12]</sup>。主干部分 (Backbone) 通常采用性能优越的分类器网络。从 YOLOv5v6.0 版本开始，为便于模型部署，主干网络 (Backbone) 不再包含 Focus 模块，主要由 C3、Conv 以及 SPPF 模块组成。颈部层网络 (Neck) 由特征金字塔网络 (Feature Pyramid Networks, FPN) 和路径聚合网络 (Path Aggregation Network, PAN) 组成，该模型通过融合不同尺度的特征信息，从而解决多尺度目标检测的问题。输出端 (Head) 主要由损失函数和非极大值抑制 (Non-Maximum Supression, NMS) 2 个部分组成<sup>[13]</sup>。

## 2 算法优化

### 2.1 可变形卷积模块

在瓶装白酒生产过程中，受到生产环境、设备和制造工艺等多种因素影响，产生许多细小瑕疵，包括许多极端细长瑕疵。由于瑕疵的尺度复杂多变，而传统卷积中的卷积核大小和比例是固定的，只能对特征图的特定位置进行特征学习和下采样，在检测瑕疵任务中存在一定局限性。因此，为缓解在卷积过程中图像几何变换带来的影响，本研究将 DCNv2 引入主干网络 (Backbone)。

可变形卷积 (Deformable Convolutional Networks, DCN)<sup>[14]</sup>在图像分类、目标检测和图像分割等任务中表现出色。它通过改变卷积核的形状，学习输入特征映射中采样点的偏移量，从而提高模型的性能。然而，第一代可变形卷积存在从目标感兴趣区域扩展到外部区域的问题，导致性能下降。为解决上述问题，DCNv2 中提出调制变形卷积，该模块能够学习每个采样点的权重，同时控制卷积核采样点的偏移量，减少无关因素影响，提高模型的精度<sup>[15]</sup>。以下是 DCNv2 模块的操作流程：

当采用  $3 \times 3$  的卷积核，将其定义为  $R$ ， $R$  的大小为二维。式 (1) 表示卷积操作的计算方式：

$$R = \{(-1, 1), (0, 1), \dots, (0, 1), (1, 1)\} \quad (1)$$

DCNv2 在提取输入图像的特征图时共有 2 个步骤。首先，对输入图片进行传统卷积操作，得到原始的特征图。然后，通过对原始特征图进行卷积运算，产生可变形卷积偏移，进而生成的特征图可以适应各种形状的目标物体。式 (2) 表示计算正常卷积操作的输出特征图的过程：

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n) \quad (2)$$

式中： $y$  为输出特征图； $p_0$  为传统卷积核的中心点； $p_n$  为传统卷积核的采样点； $x$  为输入特征图。使用可变

形卷积核计算输出特征图的公式如式 (3) 所示。

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n + \Delta p_n) \Delta m_n \quad (3)$$

式中： $\Delta p_n$  为调整后的偏移量； $\Delta m_n$  为权重系数，其余变量与传统卷积操作相同。可变形卷积引入采样点的位置偏移，使输出特征图能够更好地表示不规则目标的特征。偏移量  $\Delta p_n$  根据目标特征的分布来移动区域  $R$  中的点，并且由于偏移是通过将输入特征图与另一个卷积层进行卷积生成的，因此通常表示为小数。通过对偏移执行双线性插值，将可变形卷积的公式转换为式 (4)。

$$x(p) = \sum_q G(q, p) \cdot x(q) \quad (4)$$

式中： $q$  为偏移后的采样点位置； $p$  为整数网格点； $G(q, p)$  为从双线性插值操作中得到的采样点位置的整数形式。可变形卷积的结构如图 2 所示。

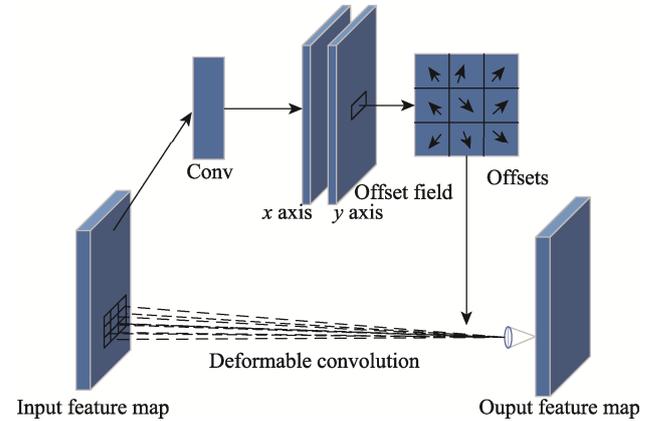


图 2 可变形卷积结构  
Fig.2 Deformable convolution structure diagram

### 2.2 Transformer 模块

根据观察实际生产情况，本研究发现相同类型的瑕疵通常是由相同的操作步骤产生。因此，相同类型的瑕疵图像会在一定范围上表现出相似特征，而不同类型的瑕疵图像则具有不同的范围特征。因此，本研究在主干网络 (Backbone) C3 模块中引入 Transformer 模块，以合并全局信息，通过赋予模型更高级别的检测特征，提高模型的精度<sup>[16]</sup>。

Transformer 模块由 Encoder 和 Decoder 2 个部分组成<sup>[17]</sup>。本研究仅使用 Encoder 模块，其具体步骤如下：

1) 将图片特征转换为编码器所需要的向量序列数据。

2) 将第 1 步的序列数据与位置编码相加，构成输入部分。

3) 将输入的向量中 3 个  $Q$ 、 $K$ 、 $V$  矩阵进行归一化处理送入 Encoder 中的多头注意力机制中。多头注意力机制中的各个头相互独立，经过式 (5)，可得到与头数相同数量的权重矩阵。最后，将权重矩阵与特

征图相乘, 使网络能够更专注于重要特征。最终, 将多头注意力机制的输出结果与第 2 步的数据进行残差连接。

$$Attention(Q, K, V) = softmax(\frac{Q \times K^T}{\sqrt{d_k}}) \times V \quad (5)$$

式中:  $Q$  为查询向量;  $K$  为键向量;  $V$  为值向量;  $d_k$  为 3 个输入向量的维度;  $softmax()$  为归一化函数, 将矩阵值映射到 0 和 1 之间。

4) 将第 3 步的输出送入前馈神经网络, 并将其与第 3 步的输出进行残差连接, 得到最终的输出结果。

在白酒瓶盖上, 小目标瑕疵比例较高。通常情况下, 卷积神经网络 (CNN) 的深层特征对于小目标的区分能力更强。此外, 主干网络 (Backbone) 末端的特征图分辨率较低, 将 Transformer 模块应用于低分辨率特征图, 可以减少模型的存储和计算成本, 更适合在工业场景中部署和应用。所以, 本研究将 Transformer 模块融入主干网络 (Backbone) 的末端 C3 模块中, 并将其命名为 C3TR, 其结构如图 3 所示。该方法能够有效利用 Transformer 模块合并全局信息, 赋予模型更高级别的检测特征, 提高模型对小目标瑕疵检测的准确性。

### 2.3 C3SE-Lite 模块

在神经网络训练过程中, 不同的卷积通道用于提取物体的不同特征。如果网络将大量训练资源投入到无关信息中, 会致使模型的训练效率变低、精度下降。在白酒瓶盖瑕疵检测任务中, 瓶盖瑕疵基本局限于小目标上, 很少出现尺寸较大的目标。然而, 原始模型对瑕疵注意力不足, 常常导致瑕疵漏检或误检等问题。

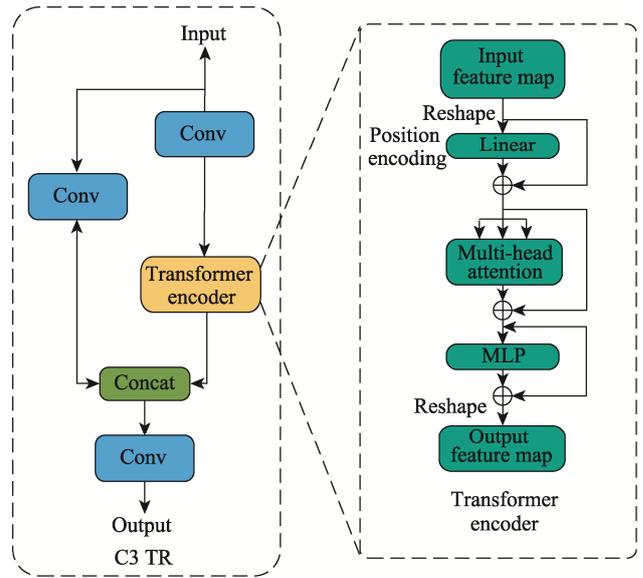


图 3 C3TR 结构  
Fig.3 C3TR structure diagram

为增强模型对白酒瓶盖瑕疵的关注和位置感知能力, 本研究引入 SE 注意力模块<sup>[18]</sup>, 将其嵌入到颈部网络 (Neck) 的所有 C3 模块中。SE 模块通过为不同的卷积通道分配不同的权重, 以提升网络特征提取的效率和性能, 其结构如图 4 所示。

$F_{sq}$  表示压缩 (Squeeze) 模块,  $U_c$  表示输入特征层, 将  $U_c$  通过全局平均池化压缩为一个维度为  $1 \times 1 \times C$  的特征向量, 这个特征向量表示每个通道的全局信息, 如式 (6) 所示。

$$z_c = F_{sq}(U_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (6)$$

接下来是激励模块 (Excitation), 如式 (7) 所示。  
 $s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 \cdot z)) \quad (7)$

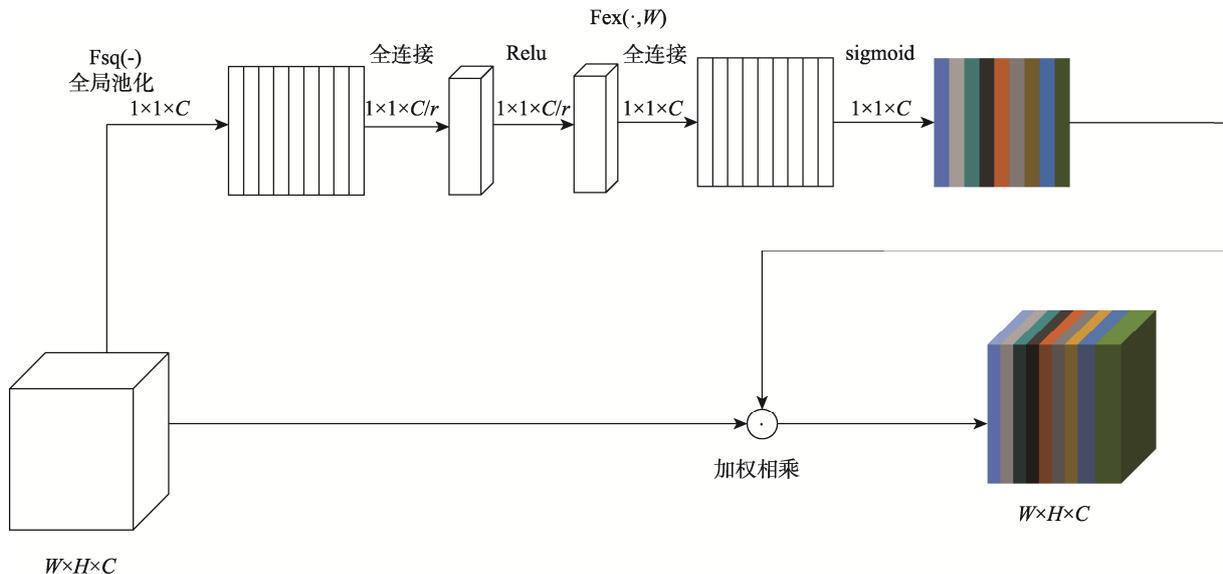


图 4 SE 结构  
Fig.4 SE structure diagram

首先是降维操作,  $W_1$  的维度为  $C/r$ ,  $r$  是一个缩放系数; 缩放后经过 ReLU 激活函数, 接着是升维操作, 激活后进行全连接, 恢复通道数为  $C$ ; 之后经过 sigmoid 函数得到最终权重向量; 最后将原输入特征层乘上得到的权值向量, 为每个通道赋予不同的权重<sup>[19]</sup>。

本研究在颈部网络 (Neck) 的 C3 模块中嵌入 SE 注意力模块, 模型的层数加深, 增大了计算量和参数量。在深层网络中, 大量堆叠卷积层致使模型存在冗余特征图, 进而增加参数量和计算成本。为解决这个问题, Han 等<sup>[20]</sup>所提出的 GhostNet 采用一种名为 Ghost Module 的模块化设计, 用较少的参数和计算量提取大量特征, 并有效地减少特征提取中的冗余计算<sup>[21]</sup>。因此, 将 C3 模块中嵌入注意力机制的 Bottleneck 结构中的

卷积操作全部替换为 GhostConv, 并将整个 C3 模块命名为 C3SE-Lite, 如图 5 所示。通过这样的改进, 在降低模型参数数量的同时仍能提升模型的精度。

GhostConv 将传统卷积操作分解为 2 个阶段: 传统卷积和轻量级线性变换。传统卷积仍使用传统的单步卷积, 但卷积核数量减少一半, 因此输出通道减少一半, 从而计算量减少一半。轻量级线性变换是对第 1 阶段提取的特征图逐个进行卷积核为 3 或 5 的卷积。最后, 将 2 个阶段得到的特征图进行拼接, 从而获得与普通卷积相当的特征表达能力。

综上所述, 为提升 YOLOv5s 在白酒瓶盖瑕疵检测任务中的性能, 本文对 YOLOv5s 进行改进, 提出了一种名为 DTS-YOLO 的算法, 其结构如图 6 所示。

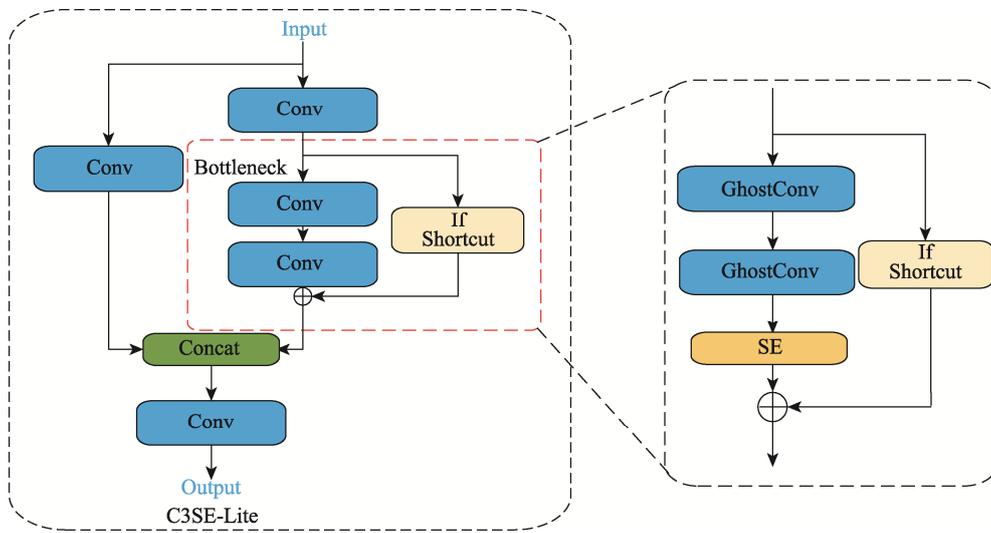


图 5 C3SE-Lite 结构  
Fig.5 C3SE-Lite structure diagram

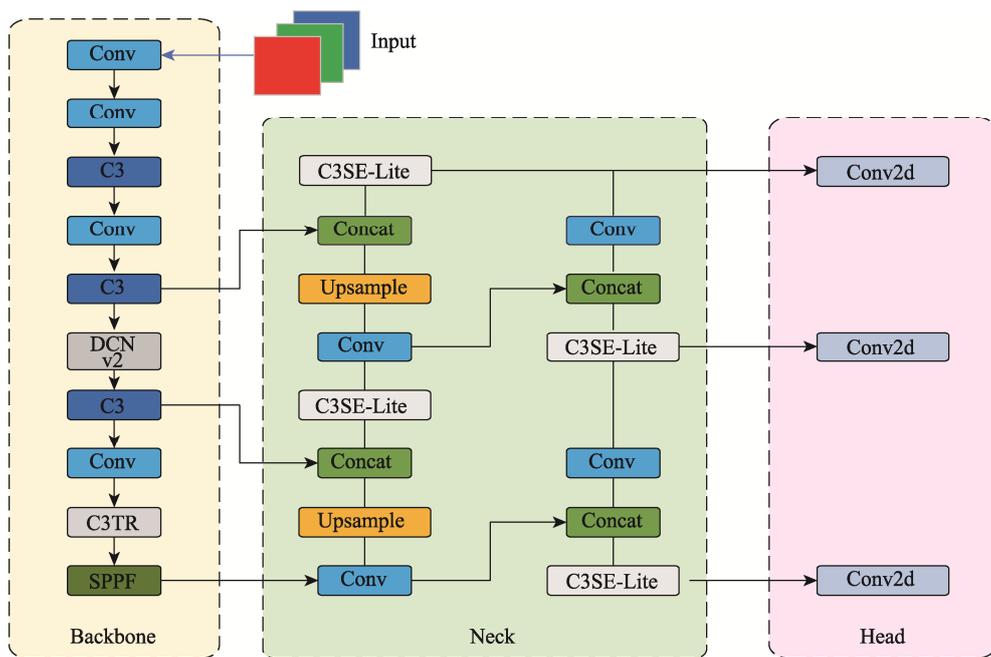


图 6 DTS-YOLO 结构  
Fig.6 DTS-YOLO structure diagram

### 3 实验设置与结果分析

#### 3.1 实验环境

本实验所采用的操作系统为 Windows 10; CPU 型号为 intel(R) Xeon(R) Gold 5218R CPU@2.10 GHz, 内存为 128 GB; 显卡为 NVIDIA A10 24 GB; 深度学习框架为 PyTorch 3.9, CUDA 版本为 11.7。模型的训练过程中使用 SGD 优化器, 学习率设置为 0.01, 训练次数为 300 次, 动量为 0.937, 权重衰减为 0.000 5, Batch-size 值设置为 8。

#### 3.2 数据集

本实验采用天池公开的企业生产数据集作为实验数据集。该数据集包含共计 2 969 张酒瓶瓶盖瑕疵图像, 每张图片中包含一种或多种瑕疵, 其中瑕疵类型分为 7 种, 包括瓶盖破损、瓶盖变形、瓶盖坏边、瓶盖打旋、瓶盖断点、喷码正常以及喷码异常, 其目标瑕疵数量分别为 1 619、705、656、480、614、489、199。数据集部分图片如图 7 所示。

该数据集中存在大量小目标以及极端长宽比瑕疵, 很大程度上增加了检测任务的难度。此外数据集中图片数量较少, 但检测类别众多, 因此面临着较大挑战。为此, 本实验采用一系列的数据增强方法对原始数据集进行处理, 包括随机旋转、水平翻转和平移。随机旋转通过在图像中心进行随机角度的旋转, 提升模型对不同角度物体的辨识能力。水平翻转将图像进行左右翻转, 引入更多的数据多样性。平移通过在图像水平或垂直方向移动图片, 增强模型对物体位置的鲁棒性。通过这些方式, 扩充数据集的规模, 总共得到 5 938 张图像, 其中包括 9 524 个目标瑕疵。扩充后的数据集按照 9:1 的比例进行划分, 分别用作训练集和测试集。

#### 3.3 评价指标

为评价模型综合的检测性能, 本文选取召回率 (recall)、平均精度均值 (mAP)、参数量 (parameters)、浮点计算次数 (FLOPs) 作为评价指标。

$$R = \frac{T_p}{T_p + F_n} \quad (8)$$

式中:  $T_p$  为模型成功的预测正例;  $F_n$  为被模型错误预测为负例的正例;  $R$  为召回率。

$$m_{AP} = \frac{\sum_{j=0}^n A_{P(j)}}{n} \quad (9)$$

式中:  $A_{P(j)}$  为第  $j$  类瑕疵的平均精度;  $n$  为瑕疵类别数目;  $m_{AP}$  为平均精度均值。

#### 3.4 实验结果与分析

##### 3.4.1 DCNv2 最佳插入点实验

为达到最佳检测性能, 本研究在主干网络 (backbone) 中选择 3 个位置 (分别标记为 A、B、C) 来尝试替换为 DCNv2。图 1 的 Backbone 中展示了插入点。通过实验验证, 发现 B 处展现出最佳效果 (表 1), 因此, 本研究选择在 B 处引入 DCNv2, 并进行后续实验。

表 2 展示了引入 DCNv2 后模型对不同瑕疵的影响。从召回率看, 引入可变形卷积有助于减少模型在小目标和极端长宽比瑕疵上的忽略, 平均召回率提升。从检测精度看, 瓶盖变形、瓶盖坏边和瓶盖断点的检测精度均有提升。这些瑕疵涵盖了大量极端长宽比的情况, 改进后的模型增强了对这些瑕疵的识别能力。然而, 在瓶盖破损、瓶盖打旋以及喷码正常这 3 类瑕疵检测中, 检测精度略有下降。瓶盖破损中存在大量小目标, 其检测精度下降主要是因为模型在提升特征提取能力的同时, 也学习到了数据集中未标注的细小

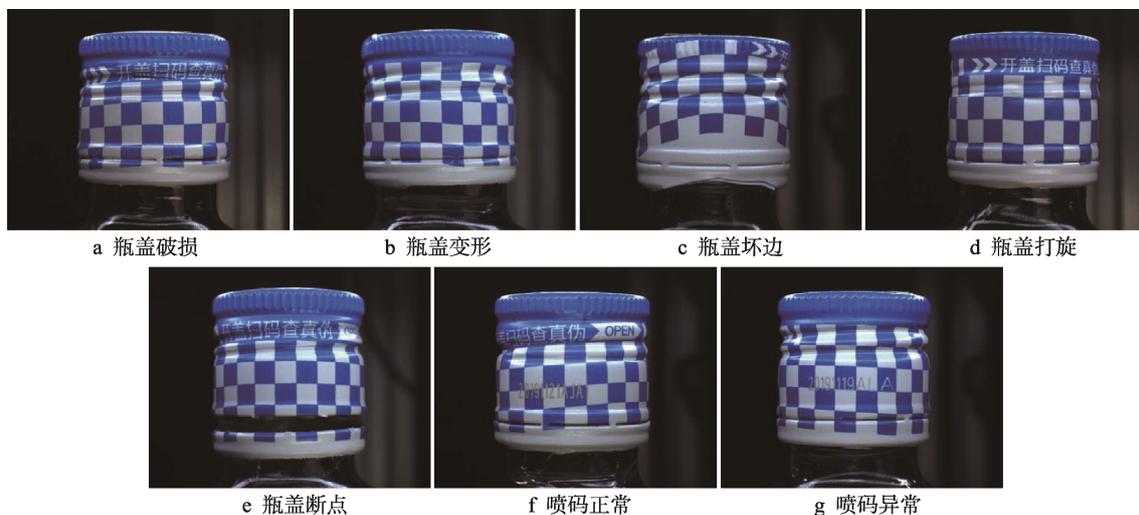


图 7 酒瓶瓶盖瑕疵分类

Fig.7 Classification chart of wine bottle cap defects

表1 DCNv2不同位置对比  
Tab.1 Comparison of different positions of DCNv2

DCNv2 位置	召回率/%	平均精度均值/%
无	90.8	93.7
A	91.0	93.2
B	91.2	94.3
C	90.3	94.1

表2 YOLOv5s与YOLOv5-DCNv2检测结果对比  
Tab.2 Comparison of YOLOv5s and YOLOv5-DCNv2 detection results

瑕疵种类	YOLOv5s		YOLOv5-DCNv2	
	召回率/%	平均精度/%	召回率/%	平均精度/%
瓶盖破损	80.5	90.1	81.1	88.0
瓶盖变形	82.5	89.2	82.6	91.0
瓶盖坏边	78.9	82.8	79.1	85.7
瓶盖打旋	100.0	99.5	100.0	99.4
瓶盖断点	93.9	96.1	95.7	97.3
喷码正常	100.0	98.9	99.6	98.8
喷码异常	100.0	99.5	100.0	99.5
平均值	90.8	93.7	91.2	94.3

杂质和其他小目标瑕疵,从而导致误检。瓶盖打旋和无此类瑕疵时图像中特征相似,喷码正常和喷码异常瑕疵相似,因此造成模型少量误检,精度略微下降。总体而言,引入DCNv2后,模型减少对瑕疵的忽略,提高对极端长宽比瑕疵的检测能力,模型的平均精度提高了0.6个百分点,平均召回率提高了0.4个百分点。考虑到瑕疵之间的高度相似性,模型需要进一步合并全局信息,增加对瑕疵的关注度,利用引入DCNv2后模型检测到被忽略的瑕疵,减少误检,进一步提高模型精度。

### 3.4.2 消融实验

为验证各种改进策略的有效性,本文进行了消融实验,对比分析不同模块的加入对整个模型性能的影响,选取平均精度均值(mAP)、召回率(recall)、参数量(parameters)3个指标作为度量标准。观察可得,通过将主干网络(Backbone)部分卷积替换为DCNv2后的YOLOv5-A模型,增强模型对极端尺寸瑕疵的检测能力(表3)。尽管参数量略有增加,但召回率和平均精度均值都有所提升,分别提高了0.4个百分点和0.6个百分点。在主干网络(Backbone)末端引入Transformer编码器的YOLOv5-B模型,合并全局信息。在几乎不增加参数量的情况下,平均精度均值提高了0.3个百分点,召回率下降了0.1个百分点。将颈部网络(Neck)的C3模块替换为C3SE-Lite的YOLOv5-C模型。因为嵌入的SE模块使得模型更加关注待检测的瑕疵部分,并且通过引入GhostConv来减少模型参数。YOLOv5-C在参数量减少 $7.4 \times 10^5$ 的同时,召回率提高了1.2个百分点,平均精度均值提高了0.3个百分点。在同时更改主干网络(Backbone)部分卷积为DCNv2和在主干网络(Backbone)末端引入Transformer编码器的YOLOv5-D模型。虽然参数量有轻微的上升,但是较原模型的召回率上升0.4个百分点,平均精度均值上升了0.8个百分点,较仅加入DCNv2的YOLOv5-A模型平均精度均值上升0.2个百分点。在此基础上将颈部网络(Neck)的C3全部替换为C3SE-Lite模块的DTS-YOLO,参数量下降至 $63.1 \times 10^5$ 。整个模型的平均精度均值上升至95.0%,召回率上升至92.5%。

综上所述,改进模型有效提高了检测精度,相较于基础网络,DTS-YOLO具有更小的参数量、更高的精度和更低的漏检率,可应用于工业设备部署。

图8为YOLOv5s和DTS-YOLO在天池白酒瓶盖瑕疵数据集上的检测结果。对比两者可以观察到,YOLOv5s在检测时发生漏检、误检以及冗余框的问题。而DTS-YOLO在精准定位瑕疵的同时,具有更高的置信度。DTS-YOLO在NVIDIA A10上对图片的检测速度可以达到30.0帧/s,满足了实时性检测(25.0帧/s)的需求。

表3 消融实验结果  
Tab.3 Ablation experiment results

Model	DCNv2	C3TR	C3SE-Lite	召回率/%	平均精度均值/%	参数量/ $10^5$
YOLOv5s				90.8	93.7	70.2
YOLOv5-A	√			91.2	94.3	70.6
YOLOv5-B		√		90.7	94.0	70.2
YOLOv5-C			√	92.0	94.0	62.8
YOLOv5-D	√	√		91.2	94.5	70.6
DTS-YOLO	√	√	√	92.5	95.0	63.1

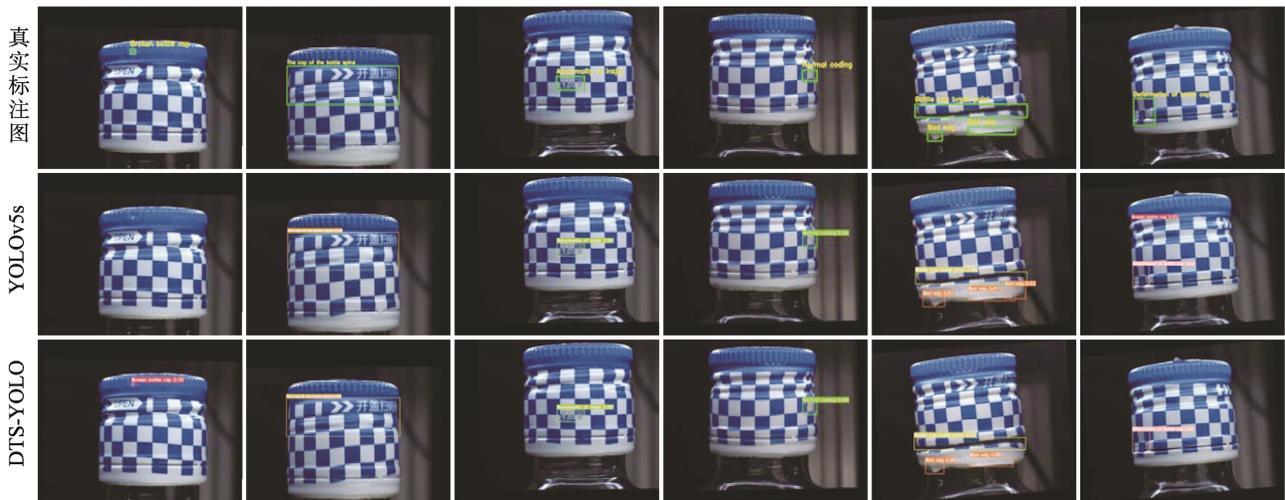


图 8 检测结果对比  
Fig.8 Comparison of test results

### 3.4.3 对比实验

为验证 DTS-YOLO 模型的先进性和有效性, 将 DTS-YOLO 算法与 Faster-RCNN、SSD、YOLOv3、YOLOv3tiny、YOLOv7<sup>[22]</sup>、YOLOv7tiny、YOLOXs 和 YOLOv5s 算法进行比较。考虑到实际应用场景的需求, 本文选取平均精度均值 (mAP)、召回率 (recall)、参数量 (parameters) 和浮点计算次数 (FLOPs) 4 个指标作为对比实验的度量标准。

观察可得, 相较于传统网络, DTS-YOLO 在平均精度均值方面表现最优 (表 4)。虽然从模型参数量来看, YOLOv7tiny 优于本文模型, 但其在召回率和准确率方面分别低于本文模型 4.7 个百分点和 3.8 百分点。从模型浮点计算次数来看, YOLOv3tiny 优于本文模型, 但其在召回率和准确率方面分别低于本文模型 3.6 个百分点和 5.0 百分点。综上所述, 与经典的目标检测算法相比, 本文算法在召回率、平均精度均值、模型参数量、浮点计算次数 4 个方面取得了更好的平衡, 更适合部署和实际应用。

表 4 各种网络模型性能对比

Tab.4 Performance comparison of various network models

Model	召回率/%	平均精度均值/%	参数量/10 <sup>5</sup>	浮点计算次数/10 <sup>9</sup>
Faster-RCNN	—	89.0	—	—
SSD	—	80.0	—	—
YOLOv3	90.0	92.1	615.5	155.4
YOLOv3tiny	88.9	90.0	86.8	13.0
YOLOv5s	90.8	93.7	70.2	15.8
YOLOv7	91.3	93.0	372.2	105.2
YOLOv7tiny	87.8	91.2	60.3	13.2
YOLOXs	—	90.9	89.4	26.7
DTS-YOLO	92.5	95.0	63.1	13.2

## 4 结语

在瓶装白酒生产过程中, 提高瓶盖瑕疵的检测精度和检测速度对瓶装白酒生产厂家具有重要意义。针对目前目标检测算法在白酒瓶盖瑕疵检测任务中检测效果差, 本文基于 YOLOv5s 提出一种改进算法 DTS-YOLO。DTS-YOLO 通过融合 DCNv2 来提高模型对极端长宽比瑕疵的检测能力; 引入 Transformer 编解码块使网络能够提取图像的全局信息, 增强模型对瑕疵的检测能力; 最后, 在颈部网络构建 C3SE-Lite 模块, 增加模型对瑕疵的关注度。通过在天池公开的瓶盖瑕疵数据集上进行实验, 本文提出的 DTS-YOLO 算法平均精度达 95.0%, 召回率达 92.5%, 检测速度为 30.0 帧/s。然而, 在更小目标瑕疵检测方面有待进一步加强。因此, 下一步研究将提高模型对极端长宽比瑕疵和小目标瑕疵的检测精度, 调整锚框大小, 并持续加强特征融合, 以实现模型对瑕疵目标准确、快速地检测。

### 参考文献:

[1] 侯隽. 行业进入“寡头时代”白酒中小企业机会在哪儿 [J]. 中国经济周刊, 2023(8): 86-87.  
HOU J. Industry Entering the Era of Oligopoly: Where Are the Opportunities for Small and Medium-sized Enterprises in the Liquor Industry?[J]. China Economic Weekly, 2023(8): 86-87.

[2] SABERIRONAGHI A, REN J, EL-GINDY M. Defect Detection Methods for Industrial Products Using Deep Learning Techniques: A Review[J]. Algorithms, 2023, 16(2): 95.

[3] GIRSHICK R. Fast R-Cnn[C]// Proceedings of the IEEE International Conference on Computer Vision. New

- York: IEEE Computer Society, 2015: 1440-1448.
- [4] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. *IEEE Trans Pattern Anal Mach Intell*, 2017, 39(6): 1137-1149.
- [5] HE K, GKIOXARI G, DOLLÁR P, et al. Mask r-cnn[C]// *Proceedings of the IEEE International Conference on Computer Vision*. New York: IEEE Computer Society, 2017: 2961-2969.
- [6] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal Loss for Dense Object Detection[J]. *IEEE Trans Pattern Anal Mach Intell*, 2020, 42(2): 318-327.
- [7] 段禄成, 谭保华, 余星雨. 基于改进 YOLOv3 的酒瓶盖瑕疵检测算法[J]. *电子测量技术*, 2022, 45(15): 130-137.
- DUAN L C, TAN B H, YU X Y. Defect Detection for Wine Bottle Caps Based on Improved YOLOv3[J]. *Electronic Measurement Technology*, 2022, 45(15): 130-137.
- [8] 李玉洁, 韩进, 刘恩爽. 一种金属类酒瓶盖瑕疵质检算法[J]. *中国科技论文*, 2022, 17(11): 1236-1244.
- LI Y J, HAN J, LIU E S. A Quality Detection Algorithm for Metal Wine Bottle Cap Defects[J]. *China Sciencepaper*, 2022, 17(11): 1236-1244.
- [9] 车璇, 朱文忠, 李韬, 等. 基于改进 RetinaNet 的白酒瓶盖缺陷检测方法[J]. *国外电子测量技术*, 2023, 42(4): 173-180.
- CHE X, ZHU W Z, LI T, et al. Defect Detection Method of Liquor Bottle Caps Based on Improved RetinaNet[J]. *Foreign Electronic Measurement Technology*, 2023, 42(4): 173-180.
- [10] ZHU X, HU H, LIN S, et al. Deformable Convnets v2: More Deformable, Better Results[C]// *Proceedings of the IEEE/ CVF Conference on Computer Vision and Pattern Recognition*. New York: IEEE Computer Society, 2019: 9308-9316.
- [11] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is All You Need[J]. *Advances in Neural Information Processing Systems*, 2017, 30: 5998-6008.
- [12] WANG L J, CAO Y Y, WANG S, et al. Investigation into Recognition Algorithm of Helmet Violation Based on YOLOv5-CBAM-DCN[J]. *IEEE Access*, 2022, 10: 60622-60632.
- [13] 叶宇星, 孙志锋, 马风力, 等. 基于改进 YOLOv5s 的腌制蔬菜真空包装缺陷检测[J]. *包装工程*, 2023, 44(9): 45-53.
- YE Y X, SUN Z F, MA F L, et al. Vacuum Packaging Defect Detection of Pickled Vegetables Based on Improved YOLOv5s[J]. *Packaging Engineering*, 2023, 44(9): 45-53.
- [14] DAI J, QI H, XIONG Y, et al. Deformable Convolutional Networks[C]// *Proceedings of the IEEE International Conference on Computer Vision*. New York: IEEE Computer Society, 2017: 764-773.
- [15] PARK H, PAIK J. Pyramid Attention Upsampling Module for Object Detection[J]. *IEEE Access*, 2022, 10: 38742-38749.
- [16] GUO Z X, WANG C S, YANG G, et al. MSFT-YOLO: Improved YOLOv5 Based on Transformer for Detecting Defects of Steel Surface[J]. *Sensors*, 2022, 22(9): 3467.
- [17] 李想, 特日根, 仪锋, 等. 针对全球储油罐检测的 TCS-YOLO 模型[J]. *光学精密工程*, 2023, 31(2): 246-262.
- LI X, TE R G, YI F, et al. TCS-YOLO Model for Global Oil Storage Tank Inspection[J]. *Optics and Precision Engineering*, 2023, 31(2): 246-262.
- [18] HU J, SHEN L, SUN G. Squeeze-and-excitation Networks[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. New York: IEEE Computer Society, 2018: 7132-7141.
- [19] 许思昂, 李艺杰, 梁桥康, 等. 基于改进 YOLOv5 算法的 PCB 裸板缺陷检测[J]. *包装工程*, 2022, 43(15): 33-41.
- XU S A, LI Y J, LIANG Q K, et al. Bare PCB Defect Detection Based on Improved YOLOv5 Algorithm[J]. *Packaging Engineering*, 2022, 43(15): 33-41.
- [20] HAN K, WANG Y, TIAN Q, et al. Ghostnet: More Features from Cheap Operations[C]// *Proceedings of the IEEE/ CVF Conference on Computer Vision and Pattern Recognition*. New York: IEEE Computer Society, 2020: 1580-1589.
- [21] SUN Z, YANG H, ZHANG Z, et al. An Improved YOLOv5-Based Tapping Trajectory Detection Method for Natural Rubber Trees[J]. *Agriculture*, 2022, 12(9): 1309.
- [22] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable Bag-of-freebies Sets New State-of-the-art for Real-time Object Detectors[C]// *Proceedings of the IEEE/ CVF Conference on Computer Vision and Pattern Recognition*. New York: IEEE Computer Society, 2023: 7464-7475.