

# 基于锚点辅助与细粒度多尺度特征的违禁品检测<sup>\*</sup>

黎作鹏<sup>1,2,3</sup> 刘佳祥<sup>1</sup> 张少文<sup>1</sup>

(1. 河北工程大学信息与电气工程学院 邯郸 056038; 2. 邯郸学院河北省光纤生物传感与通信器件重点实验室 邯郸 056005;  
3. 邯郸学院信息工程学院 邯郸 056005)

**摘要:** 针对目前 X 光违禁物品之间遮挡、噪声干扰和检测精度低等问题,以 YOLOv8s 网络为基础模型,提出一种融合了锚点辅助训练策略和细粒度多尺度特征的违禁品检测模型。在网络中采用 C3\_Res2Net 模块替换 C2f 模块,通过集成不同层次的特征来增强多尺度,以增加网络层的感受野范围,获取细粒度层面的特征,解决违禁品之间存在遮挡带来检测精度低问题;采用滑动平均的 Slide Loss 目标类别损失函数和改进的边框损失函数尝试为困难样本分配更高的权重,降低高质量锚框的竞争力的同时,减小了低质量示例产生的有害梯度,同时聚焦于普通质量的锚框,提高检测器的整体性能,使得具有更好的抗噪声干扰能力;在训练前期使用 ATSS 和 Task-Aligned Assigner 协同训练机制的标签分配策略,利用 Anchor-based 的预设信息,达到稳定模型训练的目的;在训练的后期采用锚点辅助训练策略充分发挥了结合不同 Anchor 网络的各自优势,从而进一步提升了模型检测精度。本文所改进模型在公开数据集 SIXray、HiXray 上进行了训练和测试,mAP<sub>50</sub> 分别达到 94.9%、83.7%,mAP<sub>50:95</sub> 为 73.1%、52.2%。结果表明,本文所改进模型具有较高的检测准确性和稳定性。

**关键词:** 违禁品检测;YOLOv8;细粒度;滑动平均;锚点辅助;协同训练;标签分配

**中图分类号:** TP391.4;TN911.73 **文献标识码:** A **国家标准学科分类代码:** 520.6040

## X-ray prohibited detection method based on anchor-aided and granular level multi-scale features

Li Zuopeng<sup>1,2,3</sup> Liu Jiaxiang<sup>1</sup> Zhang Shaowen<sup>1</sup>

(1. School of Information and Electrical Engineering, Hebei University of Engineering, Handan 056038, China;

2. Hebei Key Laboratory of Optical Fiber Biosensing and Communication Devices, Handan University, Handan 056005, China;

3. School of Information Engineering, Handan University, Handan 056005, China)

**Abstract:** Aiming at the current problems of occlusion, noise interference and low detection accuracy between prohibited X-ray objects, a contraband detection model integrating anchor-aided training strategy and fine-grained multi-scale features was proposed based on the YOLOv8s network. In the network, the C3\_Res2Net module is used to replace the C2f module. By integrating features at different levels to enhance multi-scale, the receptive field range of the network layer is increased, and features at the fine-grained level are obtained to solve the problem of low detection accuracy caused by occlusion between contraband items; the sliding average Slide Loss target category loss function and the improved border loss function are used to try to assign higher weights to difficult samples, which reduces the competitiveness of high-quality anchor frames while reducing the harmful gradients generated by low-quality examples. At the same time, the focus is on anchor frames of ordinary quality to improve the overall performance of the detector and make it have better anti-noise interference ability; in the early stages of training, the ATSS (Adaptive Training Sample Selection) and Task-Aligned Assigner collaborative label assignment strategies are used, leveraging anchor-based preset information to stabilize model training; in the later training stages, an anchor-aided training strategy further enhances detection accuracy by exploiting the respective advantages of various anchor networks; the improved model was trained and tested on the public SIXray and HiXray datasets, achieving mAP<sub>50</sub> scores of 94.9% and 83.7%, and mAP<sub>50:95</sub> scores of 73.1% and 52.2%, respectively. The results demonstrate that the proposed model offers high accuracy and stability in contraband detection.

**Keywords:** contraband detection; YOLOv8; granular level; sliding average; anchor-aided; collaborative training; label assignment

## 0 引言

随着社会的不断进步和科技的飞速发展,公共安全成

为了一个社会日益关注的问题。在公共场所,特别是交通枢纽、机场、车站、地铁等人流密集的地方,违禁物品的携带和使用可能对人们的生命和财产安全造成严重威胁。

收稿日期:2024-11-16

<sup>\*</sup> 基金项目:国家自然科学基金(62101174)、河北省高等学校科学技术研究项目(BJK2022025)、河北省光纤生物传感与通信器件重点实验室项目(SZX2022010)资助

但大多数X光违禁物品的检查主要依靠人工筛选的方式,然而,依靠人工进行筛选可能因为视觉疲劳而出现漏检、误检,其次人眼在处理复杂场景存在一定的局限性。因此,高效可靠的违禁物品检测算法对于确保公共安全至关重要。

在X光违禁品检测早期主要运用机器学习算法对违禁物品进行分类和检测。常用的机器学习算法包括支持向量机<sup>[1]</sup>(support vector machines, SVM)、随机森林<sup>[2]</sup>(random forest)和视觉词袋算法<sup>[3]</sup>等。这些方法需要手动设计特征信息,并经过特征选择和分类器训练。然而,这些方法在处理姿态与角度变化、小目标检测和复杂背景干扰等问题时存在一定的局限性。随着深度学习技术的发展,近年来,相关人员对违禁物检测开展了研究,例如:1)在姿态与角度变化等问题方面,游玺等<sup>[4]</sup>在Cascade R-CNN<sup>[5]</sup>模型的基础上加入空间自适应注意力机制,设计多尺度自适应候选区域生成网络,改进级联检测头;董乙杉等<sup>[6]</sup>在YOLOv5<sup>[7]</sup>模型的基础上,融合反向瓶颈并加入注意力机制设计出一种深度可分离主干网络,有效缓解目标因角度与姿态变化带来的特征提取困难问题;2)在针对小目标检测问题方面,王昱晓等<sup>[8]</sup>在SSD<sup>[9]</sup>网络模型上,提出一种新的特征融合框架并用于X光违禁品检测;李松等<sup>[10]</sup>在对YOLOv7<sup>[11]</sup>模型进行改进中,增加微小物体的检测头,提高对小目标的敏感度;陈冠豪等<sup>[12]</sup>在YOLOv8<sup>[13]</sup>的Neck部分加入Swin-CS模块,Swin-CS模块通过融合Swin Transformer的滑动多头自注意力和CBAM(channel and spatial attention module)注意力,不仅扩大了感受野,还补充了物体间的多维交互信息,增强了小目标的特征提取能力;李广睿等<sup>[14]</sup>提出了一种改进的HRNet多尺度特征融合网络模型,增强对于小目标的检测与识别;3)在针对复杂背景干扰问题方面,贾兴旺等<sup>[15]</sup>通过将结合卷积与注意力模块引入到YOLOv5网络模型中,增强模型对检测目标的感知能力,有效抑制复杂背景的干扰;孙嘉傲等<sup>[16]</sup>在YOLOv5网络模型中加入坐标注意力通过将位置信息嵌入到通道注意力来提高网络对违禁品的位置感知能力;左景等<sup>[17]</sup>在YOLOv8网络模型的Backbone中添加设计空间和通道重构机制(spatial and channel reconstruction attention module, SCAM)增强关键语义信息提取,通过对特征图空间和通道的重构抑制背景干扰;马昌嵩等<sup>[18]</sup>提出一种反向加权融合多尺度特征的违禁品检测算法,通过反向自适应地引导融合多尺度上下文特征以此来解决复杂背景干扰问题。

上述相关研究已取得了有效的进展,仍然存在以下问题:1)由于目标之间存在遮挡,导致模型出现误检现象;2)没有更多关注困难样本造成噪声干扰,从而导致模型出现漏检现象;3)在实际应用中检测精度还需要进一步提高。

由于YOLO是一种高效的目标检测网络,适用于实时检测任务,同时它在速度和精度之间取得了较好的平衡,因

此更适合在X光违禁物检测这一实时性与精确度要求较高的任务中使用,由于YOLOv8模型是当前较新的YOLO系列模型之一,因此,本文将以YOLOv8为基础模型针对以上3个问题开展研究,主要工作如下:

1)构造C3\_Res2Net模块,在卷积块中构造分层残差,通过引入了一种新的特征图分割操作,沿着通道维度将输入特征图分成多个组。使得可以进行跨组特征聚合,从而网络模型能够捕获更丰富的空间信息来处理重叠遮挡问题。

2)引入滑动平均的Slide Loss函数作为目标类别损失函数以及使用非单调聚焦系数和InnerIoU来改进CIoU使得模型具有更好的抗噪声干扰能力,减少在违禁品检测中误检的几率。

3)使用协同训练机制标签分配策略达到稳定模型训练以及全新的锚点辅助训练策略利用不同目标检测学习范式增加模型检测精度。

## 1 网络结构

本文在YOLOv8s网络模型的基础上进行改进工作包括:1)将C2f模块替换为C3\_Res2Net模块,可以有效缓解目标之间遮挡带来的特征提取困难问题,增强模型对违禁品特征的提取能力;2)在分类分支使用改进的Slide Loss损失函数<sup>[19]</sup>提高困难样本的表征学习能力,而在回归分支对CIoU Loss<sup>[20]</sup>进行改进,减小低质量样本对模型训练的负面影响,并提高模型收敛速度和泛化能力;3)在网络的分类头和回归头上分别添加Anchor-based辅助分支,充分利用不同目标检测方法,进一步提高了模型的检测精度;4)在网络训练的初始阶段使用自适应训练样本选择(adaptive training sample selection, ATSS)<sup>[21]</sup>和Task-Aligned Assigner<sup>[22]</sup>协同训练机制标签分配策略,有利于增加模型的稳定性,改进后的网络结构如图1所示。

### 1.1 细粒度多尺度处理

现有的基本构建块主要关注如何提高检测效率或引入新的训练技巧,尤其是跨阶段部分(cross stage partial, CSP)<sup>[23]</sup>块和高效层聚合网络(efficient layer aggregation networks, ELAN)<sup>[24]</sup>块,忽视了多尺度特征表示的重要性。

图2为C2f模块,其参考了CSP以及ELAN的思想进行的设计,可以使其在保证轻量化的同时获得更加丰富的梯度流信息。

然而,C2f模块因为通过减少卷积层的数量和使用更简单的结构来降低计算复杂性,但对于特征提取能力和更好的处理复杂任务的能力相对较弱,而运用CSP思想设计的C3模块能在保持较高计算复杂性的同时,提供较好的特征提取能力,其次,C2f模块忽视了一个重要的问题,即编码多尺度特征。其中一种有效的设计原则是Res2Net<sup>[25]</sup>,它通过单个残差块内构造类似于残差的分层连接,从而在更精细的水平上提高多尺度表示能力。

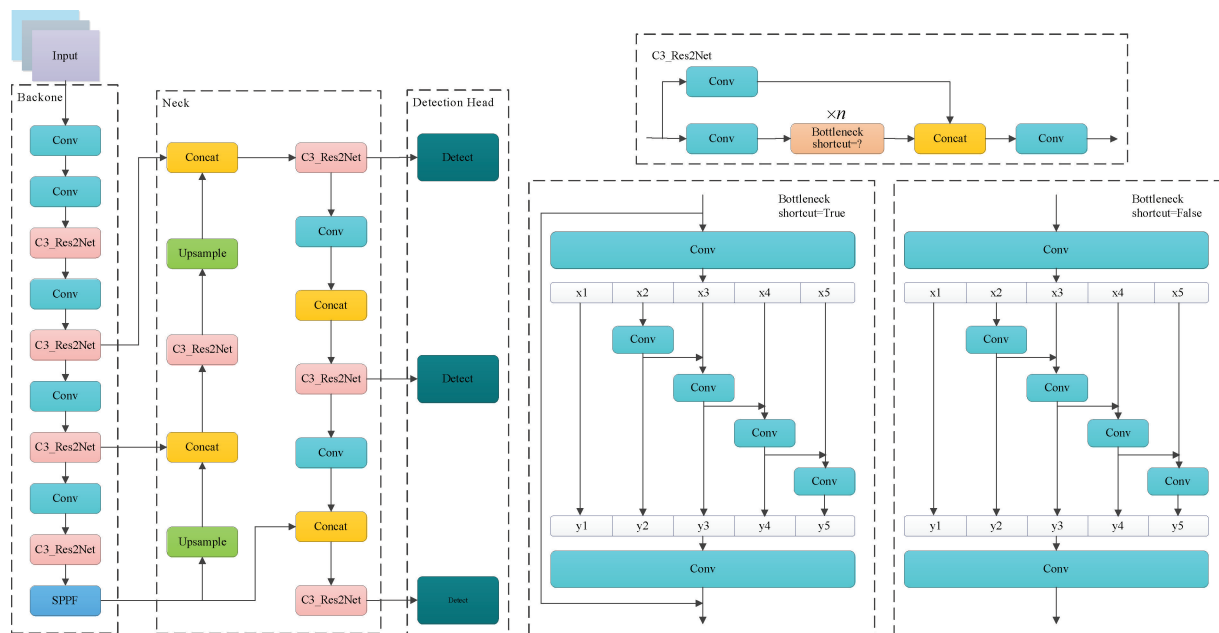


图 1 改进后的网络结构

Fig. 1 Improved network structure

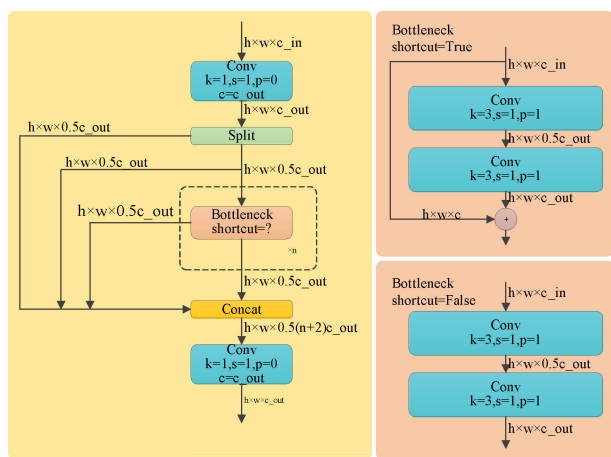


图 2 C2f 模块示意图

Fig. 2 C2f module diagram

综上所述,为了能够处理复杂任务提供较好的特征提取能力,选用 YOLOv5 中的 C3 模块替换 YOLOv8 中的 C2f 模块,并对 Res2Net 进行改进,将改进的 Res2Net 融入到 C3 模块中 Bottleneck 部分形成 C3\_Res2Net 模块, C3\_Res2Net 模块如图 3 所示。

在改进的 Res2Net 中将大卷积核融入到 Res2Net 中并且增加 Res2Net 的通道维度,具体为经过一个  $1 \times 1$  卷积操作,将特征图均匀地分割为 5 个特征子集,用  $X_i$  表示,其中  $i \in \{1, 2, 3, 4, 5\}$ 。与输入特征图相比,每个特征子集  $X_i$  具有相同的空间大小,但通道数为  $\frac{1}{5}$ 。除  $X_1$  外,每个  $X_i$  都有对应的  $5 \times 5$  卷积,记为  $K_i()$ 。用  $Y_i$  表示

$K_i()$  的输出。特征子集  $X_i$  与  $K_{i-1}()$  的输出相加,然后输入  $K_i()$ 。因此,  $Y_i$  可以写为:

$$Y_i = \begin{cases} X_i, & i = 1 \\ K_i(X_i), & i = 2 \\ K_i(X_i + Y_{i-1}), & 2 < i \leq 5 \end{cases} \quad (1)$$

每个  $5 \times 5$  卷积算子  $K_i()$  都可能从所有特征分割  $\{X_j, j \leq i\}$  接收特征信息。每次特征分割  $X_j$  经过  $5 \times 5$  卷积算子时,输出结果可以具有比  $X_j$  更大的感受野。由于组合爆炸效应, Res2Net 模块的输出包含不同数量和不同组合的感受野大小,使得在改进的 Res2Net 模块中,分割以多尺度方式处理,这有利于全局和局部信息的提取,解决违禁品中物体与物体之间遮挡的问题。为了更好地融合不同尺度的信息,连接所有分割并传递给  $1 \times 1$  卷积。通过分割和串联策略,卷积操作能够更高效地处理特征。为了减少参数数量省略初次分割的卷积,这也可以视为一种特征重用的方法。

## 1.2 损失函数改进

损失函数的计算包括 2 个分支:分类分支和回归分支,分类分支采用目标类别损失  $L_{cls}$ ,回归分支采用边框损失  $L_{box}$  和(Distribution Focal Loss)<sup>[26]</sup> 损失  $L_{dfl}$ , 3 个 Loss 采用一定权重比例加权,其损失计算公式为:

$$L = \lambda_1 L_{cls} + \lambda_2 L_{box} + \lambda_3 L_{dfl} \quad (2)$$

式中:  $\lambda_1, \lambda_2, \lambda_3$  为平衡系数。

### 1) 目标类别损失函数

目标类别损失  $L_{cls}$  采用二值交叉熵损失函数进行计算,其计算公式为:

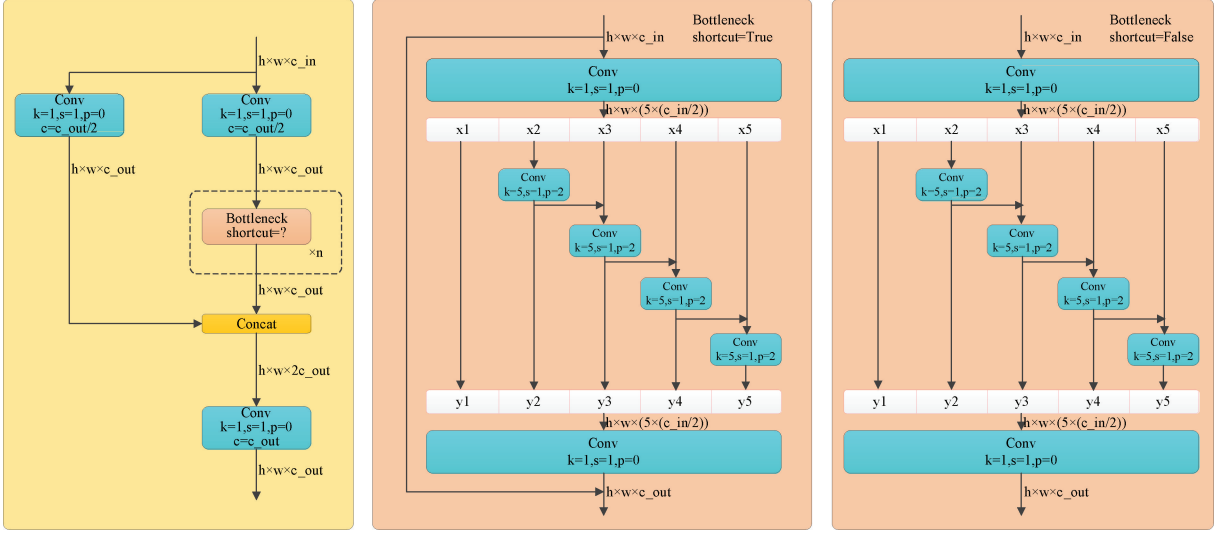


图 3 改进的 C3\_Res2Net 模块示意图

Fig. 3 Schematic diagram of the improved C3\_Res2Net module

$$L_{cls}(O, C) = - \frac{\sum_{i \in pos} \sum_{j \in cls} (O_{ij} \ln(\hat{C}_{ij}) + (1 - O_{ij} \ln(1 - \hat{C}_{ij})))}{N_{pos}} \quad (3)$$

$$\hat{C}_{ij} = \text{Sigmoid}(C_{ij}) \quad (4)$$

式中:  $O_{ij} \in \{0, 1\}$ , 表示预测目标边界框  $i$  中是否真实存在第  $j$  类目标, 0 表示不存在, 1 表示存在。  $C_{ij}$  为预测值,  $\hat{C}_{ij}$  为  $C_{ij}$  通过 *Sigmoid* 函数得到的目标概率,  $N_{pos}$  为正样本数。

但是二值交叉熵损失函数忽视了样本不平衡问题, 即大多数情况下简单样本数量较多, 而困难样本相对稀疏。使用 Slide Loss 有利于改善样本不平衡问题, Slide Loss 可以表示为:

$$f(x) = \begin{cases} 1, & x \leq \mu - 0.1 \\ e^{1-\mu}, & \mu - 0.1 < x < \mu \\ e^{1-x}, & x \geq \mu \end{cases} \quad (5)$$

然而在 Slide Loss 中将边界框的 IoU 值的平均值作为阈值  $\mu$ , 但将平均值作为阈值会受到噪声和异常值的影响, 因此引入指数滑动平均值作为阈值用于平滑 IoU 的变化, 使得阈值  $\mu$  能够更稳定地反映模型的性能。改进后的计算方法如下:

$$\mu = \lambda\mu + (1 - \lambda)\text{auto\_iou} \quad (6)$$

式中:  $\lambda = \alpha(1 - e^{-\frac{i}{\beta}})$ 。  $\mu$  初始化为 1 是为了保证在训练开始时设置一个最优的基线,  $\lambda$  为衰减因子, *auto\_iou* 是每个 batch-size 中预测框和 gt 框的平均 IOU 的值,  $i$  为迭代的次数,  $\alpha, \beta$  为两个超参数。

将二值交叉熵损失函数替换为改进的 Slide Loss 函数有利于改善样本不平衡问题。

## 2) 边框损失函数

在原模型中使用的边框损失 CIoU Loss 将目标与预测之间的距离、重叠率、尺度以及长宽比都考虑进去。记

锚框为  $\vec{B} = [x \ y \ w \ h]$ , 目标框为  $\vec{B}_{gt} = [x_{gt} \ y_{gt} \ w_{gt} \ h_{gt}]$ , 结合图 4, 其计算过程如下:

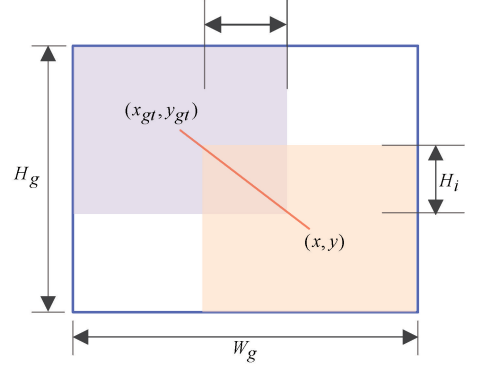


图 4 最小的包围盒(蓝色)和中心点的连接(红色)  
Fig. 4 The smallest bounding box(blue) and the connection of the center point(red)

并集的面积:

$$S_u = w_h + w_{gt}h_{gt} - W_iH_i \quad (7)$$

IoU 用于度量目标检测任务中预测框与真实框的重叠程度, IoU Loss 定义为:

$$\mathcal{L}_{IoU} = 1 - IoU = 1 - \frac{W_iH_i}{S_u} \quad (8)$$

CIoU Loss 在 IoU Loss 的基础上构造惩罚项, 并将惩罚项定义为中心点连接的归一化长度, 并且增加了纵横比一致性的考虑, 如式(9)所示。

$$\mathcal{L}_{CIoU} = \mathcal{L}_{IoU} + \frac{(x - x_{gt})^2 + (y - y_{gt})^2}{W_g^2 + H_g^2} + \alpha v \quad (9)$$



式中:

$$\nu = \frac{4}{\pi^2} \left( \arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2,$$

$$\alpha = \frac{\nu}{\mathcal{L}_{IoU} + \nu}.$$

然而 CIoU Loss 只是致力于强化边界框损失的拟合能力。但是,在目标检测训练集中含有低质量示例,如果一味地强化边界框对低质量示例的回归,显然会危害模型检测性能的提升。因此引入动态非单调聚焦机制<sup>[27]</sup>,使用离群度替代 IoU 对锚框进行质量评估,并提供了明智的梯度增益分配策略。该策略在降低高质量锚框的竞争力的同时,也减小了低质量示例产生的有害梯度。

定义离群度以描述锚框的质量,其定义为:

$$\beta = \frac{\mathcal{L}_{IoU}^*}{\mathcal{L}_{IoU}} \in [0, +\infty) \quad (10)$$

式中:  $\mathcal{L}_{IoU}^*$  为动量  $m$  的滑动平均值,  $\mathcal{L}_{IoU}^*$  中的 \* 表示从计算图中分离。

利用  $\beta$  构造非单调聚焦系数  $r$  并将应用于 CIoU 中:

$$\mathcal{L}_{nmf-CIoU} = r \mathcal{L}_{CIoU}, r = \frac{\beta}{\delta \alpha^{\beta-\delta}} \quad (11)$$

式中:  $\alpha, \delta$  为两个超参数。

在 CIoU Loss 引入非单调聚焦系数之后为了弥补边框回归损失泛化性不足以及收敛速度慢的问题,引入 Inner-IoU<sup>[28]</sup>,使用尺度因子 ratio 控制生成不同尺度的辅助边框用于计算损失,将其应用至现有 CIoU Loss 中能够获得更快更为有效的回归结果。

如图 5 所示,gt 框与锚框分别表示为  $B^{gt}$  和  $B$ 。gt 框和 gt 框内部的中心点用  $(x_c^{gt}, y_c^{gt})$  表示,而  $(x_c, y_c)$  则表示锚框和内部锚框的中心点。gt 框的宽度和高度分别表示为  $w^{gt}$  和  $h^{gt}$ ,而锚框的宽度和高度分别表示为  $w$  和  $h$ 。变量“ratio”对应的是尺度因子。Inner-IoU 的定义如下:

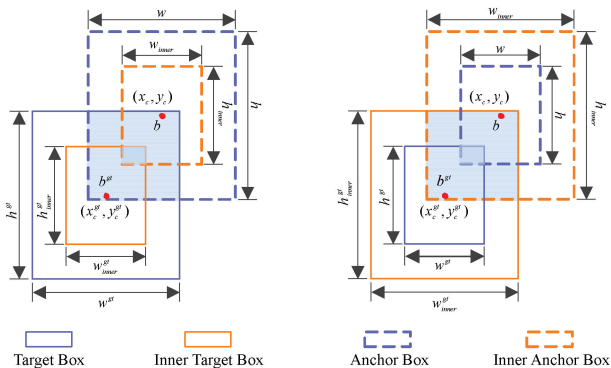


图 5 Inner-IoU 示意图

Fig. 5 Inner-IoU diagram

$$b_l = x_c - \frac{w \times ratio}{2}, b_r = x_c + \frac{w \times ratio}{2} \quad (14)$$

$$b_t = y_c - \frac{h \times ratio}{2}, b_b = y_c + \frac{h \times ratio}{2} \quad (15)$$

$$inter = (\min(b_r^{gt}, b_r) - \max(b_l^{gt}, b_l)) \times (\min(b_b^{gt}, b_b) - \max(b_t^{gt}, b_t)) \quad (16)$$

$$union = (w^{gt} \times h^{gt}) \times (ratio)^2 + (w \times h) \times (ratio)^2 - inter \quad (17)$$

$$IoU^{inner} = \frac{inter}{union} \quad (18)$$

将 Inner-IoU 应用至带有非单调聚焦系数的 CIoU 的边框回归损失函数中,  $\mathcal{L}_{Inner-nmf-CIoU}$  的定义如下:

$$\mathcal{L}_{Inner-nmf-CIoU} = \mathcal{L}_{nmf-CIoU} + IoU - IoU^{inner} \quad (19)$$

### 1.3 锚点辅助训练策略

基于深度学习的目标检测技术从学习范式上主要可分为 Anchor-free 和 Anchor-based 两大类,这两类方法针对不同尺度的目标检测上分别存在不同的优势,在 YOLOv8 中采用基于 Anchor-free 的检测方式。由于 Anchor-free 不需要预设 Anchor,只需要对不同尺度的特征图的目标中心点和宽高进行回归,可以更灵活地适应不用形状和尺寸的目标,从而可以避免一些由于 Anchor 设置不合理导致的漏检问题。然而,由于 Anchor-free 的检测方式每个位置只预测一个框,会导致一些重叠或遮挡区域无法被检测从而使得精度并不如 Anchor-based 的检测方式而 Anchor-based 的检测方式使用 Anchor 机制产生密集集 Anchor box,使得网络可直接在此基础上进行目标分类及边界框坐标回归,密集集 Anchor box 可有效提高网络目标召回能力。

因此,在 YOLOv8 采用的 Anchor-free 训练范式的基础上,进一步融合 Anchor-based 训练范式,构建锚点辅助训练策略,充分发挥了结合不同 Anchor 网络的各自优势,并对全阶段网络进行映射及优化,最终实现了 Anchor 的统一,从而进一步提升了模型检测精度。基于锚点辅助训练策略的示意图如图 6 所示,具体来说,一方面,在网络的分类头和回归头上分别添加 Anchor-based 辅助分支,在训练阶段,该分支与 Anchor-free 分支分别进行独立的 Loss 计算,之后会对 Loss 进行相加,各自反向传播进行网络的优化。通过 Anchor-based 辅助分支,为网络训练引入额外的内嵌指导信息,并与 Anchor-free 分支的信息进行整合,从而达到对结合不同 Anchor 网络的全方位融合的目的,进一步挖掘网络自身的潜力,充分发挥其效能。另一方面,在网络标签匹配的过程中引入同特征点密集采样的机制。通过扩大每次样本匹配过程中所选取候选框的范围,增加候选框中正样本的数量,并且对同一特征点重复投放采样点,进一步提升在训练过程中候选框的质量。与此同时,在网络的每一层中还会搭配原始的 Anchor-free 分支,进一步提升候选框的多样性。除此之外,只在训练过程中

$$b_l^{gt} = x_c^{gt} - \frac{w^{gt} \times ratio}{2}, b_r^{gt} = x_c^{gt} + \frac{w^{gt} \times ratio}{2} \quad (12)$$

$$b_t^{gt} = y_c^{gt} - \frac{h^{gt} \times ratio}{2}, b_b^{gt} = y_c^{gt} + \frac{h^{gt} \times ratio}{2} \quad (13)$$

引入额外的辅助分支,在测试过程中不予使用。最终在不增加推理时间的情况下,提升网络精度。

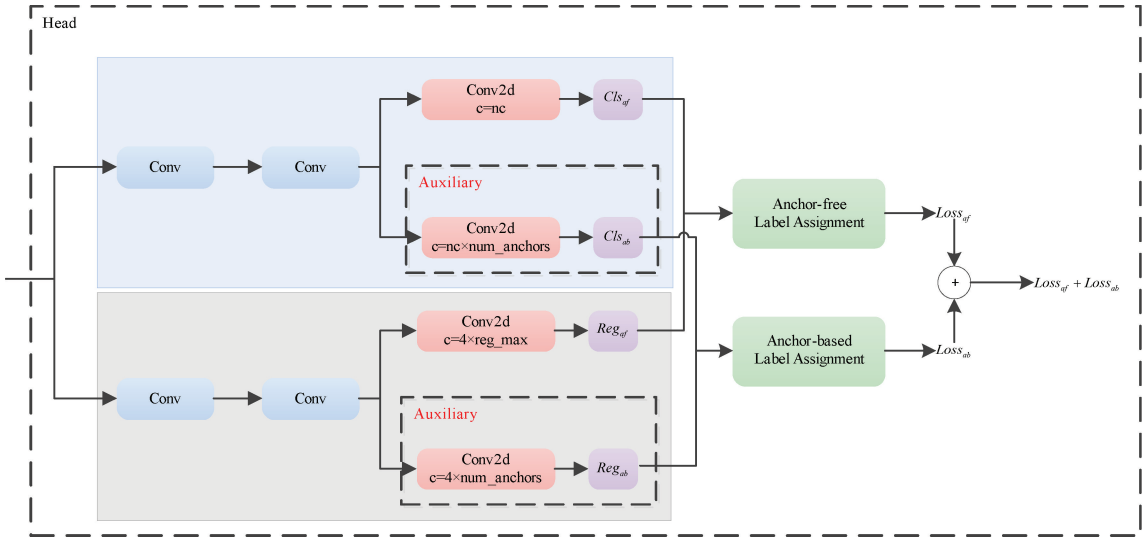


图6 锚点辅助训练策略示意图

Fig. 6 Schematic diagram of anchor-aided training strategy

#### 1.4 协同训练机制标签分配策略

YOLOv8 原模型使用的是基于 Anchor-free 检测方法,标签分配策略使用 Task-Aligned Assigner<sup>[22]</sup>实现正负样本的匹配,此标签匹配算法根据训练实例的锚点分配应该满足以下规则:对齐的锚点应该能够预测高分类分数,并进行精确的定位;未对齐的锚点应该具有较低的分类分数并随后被抑制,从而设计出使用分类得分与预测框和 gt 的 IOU 的高阶组合来衡量任务对齐程度,其公式如下:

$$t = s^{\alpha} + u^{\beta} \quad (20)$$

式中:  $s$  和  $u$  分别表示分类分数和 IOU 值。 $\alpha$  和  $\beta$  用于控制锚点对齐度量中这两个任务的影响。然而,由于在模型的训练前期分类分数以及预测框和 gt 的 IOU 的值比较低,并且基于 Anchor-free 检测方法因 bbox 的回归相对困难,使得模型的预测结果不太稳定,因此,在模型的训练前期使用 Anchor-based 检测方法,使用 ATSS 和 Task-Aligned Assigner 协同训练机制标签分配策略,利用 Anchor-based 的预设信息,进而达到稳定模型训练的目的。在模型的训练后期使用锚点辅助训练策略,标签分配策略使用 Task-Aligned Assigner 标签分配。

## 2 实验与结果分析

### 2.1 实验配置与训练策略

本文实验使用 PyTorch2.0.0 深度学习框架,并采用 Python3.8 编程语言,操作系统为 Ubuntu20.04,显存为 12 GB,显卡为 NVIDIA GeForce RTX3080Ti,输入图像大小为  $640 \times 640$ ,Batchsize 设置为 16,不采用预训练权重训练。

### 2.2 评价指标

实验采用平均精度均值(mean average precision,

mAP)、平均精度(average precision, AP)、参数量(parameters)、计算量作为实验结果的评价指标。精确度(precision)表示模型预测的所有目标中,预测正确的比例,其表达式为:

$$P_{\text{Precision}} = \frac{T_p}{T_p + F_p} \quad (21)$$

召回率(Recall)表示所有真实目标中模型预测正确的目标比例:

$$R_{\text{Recall}} = \frac{T_p}{T_p + F_N} \quad (22)$$

式中:  $T_p$  是 IoU 大于阈值的检测框数量,同一个真实框只计算一次。 $F_p$  是 IoU 小于等于阈值的检测框数量或者是检测到同一个真实框的多余检测框的数量。 $F_N$  是没有检测出的真实框的数量。

通过对每种类别的准确度和召回率绘制的 P-R 曲线积分可得到该类别的平均精度(average precision, AP)值:

$$AP_i = \int_0^1 P_{\text{Precision}} dR_{\text{Recall}} \quad (23)$$

式中:  $R_{\text{Recall}}$  是 P-R 曲线的横坐标,  $P_{\text{Precision}}$  是 P-R 曲线的纵坐标。

对  $N$  个类别求平均值可得到平均精度均值(mean average precision, mAP),其公式为:

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (24)$$

### 2.3 数据集介绍

SIXray 数据集<sup>[29]</sup>包含具有不同场景和角度的 X 射线图像,这些图像是从真实的安检场景中获取。此数据集包含 1 059 231 张 X 射线图像,其中 8 929 张图像针对 6 种不同类别进行手动标注:Gun(枪)、Knife(刀)、Wrench(扳

手)、Pilers(钳子)、Scissors(剪刀)、Hammer(锤子)。使用标注的 8 929 张图像按 7:3分为训练集、验证集实验。

HiXray 数据集<sup>[30]</sup>是由北京航空航天大学软件开发环境国家重点实验室、科大讯飞于 2021 年公开的高质量 X 光行李违禁品数据集,其数据集包含 8 类常见的机场禁止携带的物品:PO1(锂离子方形电池)、PO2(锂离子圆柱形电池)、WA(水)、LA(笔记本电脑)、MP(手机)、TA(平板电脑)、CO(化妆品)、NL(废金属打火机)共 45 364 张真实场景下的 X 射线图像,实验中用于训练的图像共 36 295 张,用于验证的图像共 9 069 张。

2.4 对比实验

本文所提模型与当前部分目标检测模型进行了比较,包括 Faster R-CNN、SSD、RetinaNet、FCOS、YOLO 系列等模型,以及较新的在 YOLO 基础上改进的关于违禁品的模型,从而进一步验证本文所提模型的准确性与高效性,并且在 SIXray、HiXray 数据集上评估其性能并对不同类别下的 AP 进行了分析,使用 mAP<sub>50</sub> 作为评估指标,实验结果如表 1、表 2 所示。

表 1 和表 2 的实验结果显示,在和经典的目标检测模型相比可以观察到,在 SIXray 上的实验,本文所提模型优于其他模型,并在一些类别(Knife、Pilers 和 Scissors)上取得了显著的改进;在 HIXray 上的实验,本文所提模型在大

表 1 SiXray 数据集对比实验							
Table 1 Comparative experiment on SiXray dataset							
模型	AP					mAP <sub>50</sub>	%
	Gun	Knife	Wrench	Pilers	Scissors		
Faster R-CNN <sup>[31]</sup>	90.1	80.0	79.3	85.3	88.3	84.6	
ResNet50 <sup>[32]</sup>	90.6	87.8	63.6	84.8	57.4	76.8	
SSD	—	—	—	—	—	82.9	
RetinaNet <sup>[33]</sup>	88.7	76.6	79.4	86.3	88.0	83.8	
FCOS <sup>[34]</sup>	—	—	—	—	—	82.0	
YOLOv4-tiny <sup>[35]</sup>	88.3	80.1	60.3	66.9	48.5	57.3	
YOLOv5s	98.5	88.6	89.1	93.3	92.3	92.4	
YOLOv5m	98.7	89.7	91.6	94.3	92.0	93.3	
YOLOX-S <sup>[36]</sup>	90.4	84.7	77.9	85.4	80.0	69.7	
YOLOv7-tiny	98.8	86.9	85.3	90.0	88.6	89.9	
YOLOv8s	98.8	91.6	90.6	93.7	92.3	93.4	
SUN	99.1	86.8	88.7	92.5	89.7	91.1	
WANG <sup>[37]</sup>	94.3	90.8	91.6	94.2	95.4	93.3	
YOLOv9s <sup>[38]</sup>	99.2	92.1	92.1	94.8	94.2	94.5	
YOLOv10m <sup>[39]</sup>	98.7	91.2	90.6	93.9	91.9	93.3	
YOLOv11m	98.9	93.0	92.3	95.3	93.4	94.6	
Ours	98.9	93.3	92.3	95.1	94.8	94.9	

表 2 HiXray 数据集对比实验									
Table 2 Comparative experiment on HiXray dataset									
模型	AP								mAP <sub>50</sub>
	PO1	PO2	WA	LA	MP	TA	CO	NL	
SSD	87.3	81.0	83.0	97.6	93.5	92.2	36.1	0.01	71.4
FCOS	88.6	86.4	86.8	89.9	88.9	88.9	63.0	13.3	75.7
RetinaNet	73.5	76.7	76.2	82.3	79.8	81.5	50.6	12.7	66.7
YOLOv5s	95.5	94.5	93.8	98.1	98.2	96.5	63.6	11.6	81.5
YOLOv5m	96.0	95.1	93.4	98.2	98.2	96.5	65.5	18.7	82.7
YOLOv7tiny	95.8	94.9	92.9	97.0	98.2	96.5	64.2	4.31	80.5
YOLOv8s	95.8	94.1	94.5	98.0	98.2	96.0	67.1	11.2	81.9
SUN	96.2	95.0	95.4	97.8	98.1	94.2	73.3	17.0	83.3
ZUO	96.4	94.7	94.2	97.6	98.6	96.3	69.2	15.7	82.9
YOLOv9s	98.1	97.3	95.1	95.8	96.3	93.4	67.8	10.1	81.7
YOLOv10m	98.3	97.9	93.9	95.9	96.9	94.4	68.6	12.5	82.3
YOLOv11m	98.3	97.9	95.6	95.8	96.7	94.4	69.9	19.4	83.5
Ours	96.6	95.7	94.5	97.9	98.4	97.0	67.5	21.8	83.7

部分类别上优于其他模型。

另外,实验对比的 YOLO 系列改进模型中,文献[16]在 YOLOv5s 网络模型中加入坐标注意力来提高网络对违禁品的位置感知能力;文献[17]在 YOLOv8s 网络模型的 Backbone 中添加设计空间和通道重构机制抑制背景干扰文献;文献[37]中在 YOLOv8m 模型中引入 Soft-NMS(non-

maximum suppression 非最大抑制)来抵抗包含遮挡的情况,经实验对比分析证实本文模型优于上述模型。

需要指出的是,表 2 的 mAP<sub>50</sub> 的值明显偏低,主要是由于 HiXray 数据集关于 CO(化妆品)与 NL(废金属打火机)的图像数量偏少、物品特点等因素导致 mAP<sub>50</sub> 被拉低。

2.5 损失函数与 C3\_ResNet(网络模型)参数测试

本小节通过实验测试不同的边框损失函数来说明改进的损失函数的准确性;通过比较不同通道数和卷积核的大小来验证 C3\_Res2Net 的性能。

边框损失函数为衡量模型在预测目标边框位置的准确性,使用不同的边框损失函数会对网络模型的性能产生影响。实验分别采用 CIoU Loss、有非单调聚焦系数的 CIoU Loss 无 Inner-IoU、有非单调聚焦系数的 CIoU Loss 有 Inner-IoU、4 种情况来研究改进不同的边框损失函数对网络模型性能和推理时间的影响,结果如表 3 所示,其中  $\alpha$  代表是否使用非单调聚焦虚数,  $\beta$  代表 IoU+inner IoU,Time 为模型在 2 787 张验证集图像上的平均单张推理时间,FPS 为模型每秒能处理的图像数量以检测网络的运算实时性。

表 3 改进边框损失函数结果

Table 3 Results of improved border loss function				
Component	mAP <sub>50</sub> /%	mAP <sub>50,95</sub> /%	Time/ms	FPS
CIoULoss	93.4	71.5	4.3	233.1
$\alpha$ CIoULoss	93.7	71.6	4.2	237.1
CIoULoss+ $\beta$	93.6	71.2	4.3	233.2
$\alpha$ CIoULoss+ $\beta$	93.9	72.1	4.3	234.3

表 3 表明,当结合 2 种方法对 CIoU Loss 进行改进后,网络模型获得最优的检测性能。这说明使用有非单调聚焦系数与 Inner-IoU 的 CIoU Loss 损失函数预测目标边框位置的准确性时,能够减小模型预测边框与真实边框之间的差异,因此本文中使用了改进的 CIoU Loss,以达到更优的检测效果。

Res2Net 的通道数与卷积算子的卷积核大小会对模型的性能产生影响,虽然增加这些参数可以提高模型特征表达能力,但也会导致模型参数量与计算复杂度增加,从而影响模型的运算实时性。为探究 Res2Net 中通道划分个数和卷积算子的卷积核大小对网络性能的影响,将 Res2Net 中通道个数设置为 4、5、6,共 3 种情况,并且卷积算子分别使用  $3\times3$ 、 $5\times5$  和  $7\times7$  卷积核进行实验,结果如表 4 所示。表 4 中  $N$  为划分的通道个数,Kernel 为卷积算子的卷积核大小,GFLOPs 为模型所需浮点运算次数。

从结果可见,模型的参数量、计算量、推理时间等随着通道个数增加和卷积核大小的增大而提高,当划分的通道个数为 6 且卷积算子的卷积核为  $7\times7$  时,模型拥有最优的检测性能,mAP<sub>50</sub> 值达到了 94.3%,这说明通过增加通道个数与增大卷积算子的卷积核,能够提升网络对特征的提取能力,使其学习到更复杂丰富的特征表达,进而提高检测性能。此外,卷积算子采用越大的卷积核能够实现更高的检测性能,这说明较大的卷积核可以增加有效感受野,提升模型的特征感知能力。虽然增加通道数与增大卷积

表 4 不同通道数和卷积核大小实验结果

Table 4 Experimental results with different channel numbers and convolution kernel sizes						
$N$	Kernel	mAP <sub>50</sub> /%	mAP <sub>50,95</sub> /%	Parameter	GFLOPs	Time
4	$3\times3$	93.0	71.0	8 549 462	22.0	1.81
4	$5\times5$	93.6	72.1	10 212 422	26.2	8.40
4	$7\times7$	93.8	72.5	12 706 862	32.4	7.80
5	$3\times3$	93.5	72.4	9 872 754	25.3	6.70
5	$5\times5$	93.8	72.5	13 231 474	33.7	9.20
5	$7\times7$	94.0	73.3	18 269 554	46.3	9.60
6	$3\times3$	93.7	72.6	11 795 930	30.0	7.90

算子的卷积核会增加模型的参数量与计算量,从而降低网络模型的运算实时性,但其仍能满足实时检测的要求,为了在性能和参数量之间保持平衡,本文选择通道数为 5,卷积算子的卷积核为  $5\times5$ 。

感受野在目标检测领域备受关注,正如人的视野一样,感受野体现的是卷积核对当前输入特征图的感知范围,如果范围很小,那么接收的信息是片面的、局部的。如果增大感受野,那么就可以获得更多的全局信息,从而更有利于对当前局面的判断。本文为了说明在 Res2Net 模块中卷积算子的卷积核大小对于有效感受野的影响,在通道个数 5 的情况下对卷积算子分别使用  $3\times3$ 、 $5\times5$ 、 $7\times7$  卷积核共 3 种情况进行实验,结果如表 5 所示,  $t$  表示测试模型输出最后特征中心点和输入图像相关区域点的比例,在 2 787 张测试集图像上随机选取 50 张图像,并且缩放放到  $1\,024\times1\,024$  的大小。

表 5 不同卷积核在不同输出特征中心点和输入图像相关区域点的比例实验结果

Table 5 Experimental results of the ratio of different convolution kernels at different output feature center points and input image related area points %				
卷积核	$t=20\%$	$t=30\%$	$t=50\%$	$t=99\%$
$5-3\times3$	4.1	6.5	13.9	90.5
$5-5\times5$	4.6	7.3	15.9	97.2
$5-7\times7$	5.4	8.7	18.7	97.8

可以发现随着卷积算子的卷积核的增大,模型输出最后特征中心点和输入图像相关区域点的比例值也在增加,说明大的卷积核可以带来更大的感受野,从而体现了大卷积核在基于 CNN 的视觉识别任务中的有效性,图 7 为感受野的可视化热力图。

2.6 消融实验

为了进一步探讨各改进策略对模型的贡献,在原始的 YOLOv8s 模型的基础上加入各个改进策略进行消融实验



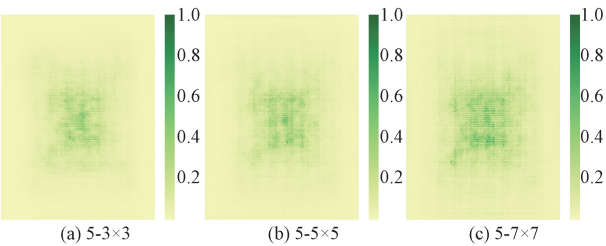


图 7 卷积核感受野的热力图

Fig. 7 Heat map of convolution kernel receptive field

并在 SIXary 数据集上进行验证,实验结果如表 6 所示。使用锚点辅助结合 ATSS 和 Task-Aligned Assigner 协同训练机制标签分配的方法对网络进行优化,模型的  $mAP_{50}$  提升了 0.7%, $mAP_{50,95}$  提升了 0.7%,使用锚点辅助策略充

分发挥了结合不同 Anchor 网络的各自优势,进一步提高模型的检测精度,ATSS 和 Task-Aligned Assigner 协同训练机制标签分配利用了 Anchor-based 的预设信息,进而达到稳定模型训练的目的;改进的边框损失函数与分类损失函数使得网络的  $mAP_{50}$  提升了 0.5%, $mAP_{50,95}$  提升了 0.6%,说明改进边框损失函数使得模型更加关注难例,改善样本不平衡问题,改进的分类损失函数使模型更加关注锚定普通质量的框并且弥补了泛化能力不足的问题,提高了模型定位对象的能力,还能够根据不同检测器与检测任务进行自我调整;将 C2f 模块替换为 C3\_Res2Net 模块后使得网络模型的  $mAP_{50}$  提升了 0.4%, $mAP_{50,95}$  提升了 1.1%,说明 C3\_Res2Net 模块在特征层内部进行多尺度的卷积,形成不同感受野,更好地处理细粒度的图像细节。

表 6 SIXray 数据集上的消融实验研究

Table 6 Ablation Experimental Study on SIXray Dataset

模型	$mAP_{50}/\%$	$mAP_{50,95}/\%$	Parameter	GFLOPs
Baseline	93.4	71.5	11 127 906	28.4
Baseline+C3_Res2Net	93.8	72.5	13 231 474	33.7
Baseline+Loss	93.9	72.1	11 127 906	28.4
Baseline+锚框辅助+ATSS	94.1	72.2	11 105 460	56.4
Baseline+锚框辅助+ATSS+Loss+C3_Res2Net	94.9	72.9	13 294 372	68.2

通过使用 C3\_Res2Net 模块(A)、改进损失函数(B)、引入锚点辅助训练策略(C)、ATSS 和 Task-Aligned Assigner 协同训练机制标签分配策略(D)、逐层增加,实验结果如表 7 所示,模型的性能不断提升, $mAP_{50}$  值提升至 94.9%, $mAP_{50,95}$  提升至 72.9%,较原始模型提高了 1.5%和 1.4%,这说明本文提出的改进策略可以有效提高网络性能。

步降低,表明改进策略的有效性。



图 8 模型改进前后可视化对比图

Fig. 8 Visual comparison of the model before and after improvement

表 7 在 SIXray 数据集上逐层消融实验

Table7 Layer-by-layer ablation experiments on the SIXray dataset

A	B	C	D	$mAP_{50}/\%$	$mAP_{50,95}/\%$	Parameter	GFLOPs
×	×	×	×	93.4	71.5	11 127 906	28.4
✓	×	×	×	93.7	71.2	13 294 372	68.2
✓	✓	×	×	94.0	71.6	13 294 372	68.2
✓	✓	✓	×	94.4	72.2	13 294 372	68.2
✓	✓	✓	✓	94.9	72.9	13 294 372	68.2

为直观展示所提模型的有效性,图 8 展示了原始模型与所改进模型在 SiXray 数据集上的结果进行了可视化对比。第 1 行图为原料模型检测结果,第 2 行图为原料改进后的检测结果,绿色、蓝色和红色框分别表示正确、误检和漏检预测。纵向对比可知,本文所改进的模型有效地提高了原始模型对违禁品的检测能力,模型的误检率与漏判率进一

3 结 论

以 YOLOv8s 网络作为基线模型,将 C2f 模块替换为 C3\_Res2Net 模块在粒度级别上表示多尺度特征,并增加了每个网络层的感受野范围,应对目标之间存在的遮挡问题而导致模型出现误检现象;在目标类别损失函数中使用基于滑动平均的 Slide Loss 损失函数使得模型更多关注困难样本,在边框回归损失函数引入非单调聚焦系数和辅助边框,来防止低质量示例产生较大的有害梯度并且弥补了

模型泛化能力不足增加其收敛速度,提高模型的抗噪声干扰能力;采用锚点辅助策略充分发挥了不同 Anchor 检测方法的优点,提升了模型的精度,使用协同训练机制标签分配提高了模型的稳定性;改进后的网络目标检测模型在识别违禁品方面具有良好的准确性和稳健性。该模型在 SIXray、HiXray 数据集进行了训练与测试,并与 Faster R-CNN、SSD、FCOS、RetinaNet、YOLO 系列等目标检测模型进行了对比,提高了原始模型的性能,实验结果显示其精度优于当前大部分目标检测模型。

## 参考文献

- [1] BASTAN M, YOUSEFI M R, BREUEL T M. Visual words on baggage X-ray images [C]. Computer Analysis of Images and Patterns: 14th International Conference, 2011: 360-368.
- [2] 王宇,邹文辉,杨晓敏,等.基于计算机视觉的X射线图像异物分类研究[J].液晶与显示,2017,32(4):287-293.  
WANG Y, ZHOU W H, YANG X M, et al. X-ray image illegal object classification based on computer vision[J]. Chinese Journal of Liquid Crystals and Displays, 2017, 32(4): 287-293.
- [3] TURCSANY D, MOUTON A, BRECKON T P. Improving feature-based object recognition for X-ray baggage security screening using primed visual words[C]. 2013 IEEE International conference on industrial technology (ICIT), IEEE, 2013: 1140-1145.
- [4] 游玺,侯进,任东升,等.融合空间注意力的自适应安检违禁品检测方法[J].计算机工程与应用,2023,59(21):176-186.  
YOU X, HOU J, REN D SH, et al. Adaptive security check prohibited items detection method with fused spatial attention [J]. Computer Engineering and Applications, 2023, 59(21): 176-186.
- [5] CAI Z, VASCONCELOS N. Cascade R-CNN: Delving into high quality object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 6154-6162.
- [6] 董乙杉,郭靖圆,李明泽,等.基于反向瓶颈和 LCBAM 设计的 X 光违禁品检测[J].计算机科学与探索,2024,18(5):1259-1270.  
DONG Y SH, GUO J Y, LI M Z, et al. An X-ray prohibited items detection model based on inverted bottleneck and attention mechanism [J]. Journal of Frontiers of Computer Science and Technology, 2024, 18(5): 1259-1270.
- [7] GLENN J, ALEX S, JIRKA B, et al. Ultralytics/yolov5: v6.1-YOLOv5. P6 [EB/OL]. [2022-02-22]. <https://github.com/ultralytics/yolov5>.
- [8] 王昱晓,张良.多尺度特征融合的安检图像危险品检测[J].激光与光电子学进展,2021,58(8):152-159.  
WANG Y X, ZHANG L. Dangerous goods detection based on multi-scale feature fusion in security images[J]. Laser & Optoelectronics Progress, 2021, 58(8): 152-159.
- [9] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]. European Conference on Computer Vision. Cham, Amsterdam: Springer, 2016: 21-37.
- [10] 李松,亚森江·木沙.改进 YOLOv7 的 X 射线图像违禁品实时检测[J].计算机工程与应用,2023,59(12):193-200.  
LI S, YASENJIANG M SH. Improved YOLOv7 X-ray image real-time detection of prohibited items[J]. Computer Engineering and Applications, 2023, 59(12): 193-200.
- [11] 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]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464-7475.
- [12] 陈冠豪,潘广贞.基于 YOLOv8s 的 X 光违禁品识别[J].计算机系统应用,2024,33(12):240-247.  
CHEN G H, PAN G ZH. X-ray contraband identification based on YOLOv8s [J]. Computer Systems & Applications, 2024, 33(12): 240-247.
- [13] JOCHER G. Ultralytics YOLOv8 [EB/OL]. [2023-01-10]. <https://github.com/ultralytics/ultralytics>.
- [14] 李广睿,刘琼,张熠卿,等.面向安检 X 光图像的违禁物品语义分割与识别研究[J].电子测量与仪器学报,2024,38(2):1-9.  
LI G R, LIU Q, ZHANG Y Q, et al. Semantic segmentation and recognition of contraband for security X-ray images [J]. Electronic Measurement Technology, 2024, 38(2): 1-9.
- [15] 粟兴旺,王小明,黄金玻,等.基于可变形卷积与注意力机制的 X 光安检违禁品检测[J].电子测量技术,2023,46(10):98-108.  
SU X W, WANG X M, HUANG J B, et al. Prohibited items detection based on deformable convolution and attention mechanism in X-ray security inspection[J]. Electronic Measurement Technology, 2023, 46(10): 98-108.
- [16] 孙嘉傲,董乙杉,郭靖圆,等.自适应与多尺度特征融合的 X 光违禁品检测[J].计算机工程与应用,2024,60(2):96-102.  
SUN J AO, DONG Y SH, GUO J Y, et al. Detection of X-Ray contraband by adaptive and multi-scale feature fusion [J]. Computer Engineering and Applications, 2024, 60(2): 96-102.
- [17] 左景,石洋宇,卢树华.基于轻量化卷积和 SCAM 改进的 X 光违禁品检测[J/OL].计算机科学与探索,1-17 [2024-09-01]. <http://kns.cnki.net/kcms/detail/11.5602.TP.20240711.1543.007.html>.  
ZUO J, SHI Y Y, LU SH H. Improved X-ray prohibited items detection method based on

- lightweight convolution blocks and SCAM attention mechanism[J/OL]. Journal of Frontiers of Computer Science and Technology, 1-17 [2024-09-01]. <http://kns.cnki.net/kcms/detail/11.5602.TP.20240711.1543.007.html>.
- [18] 马昌嵩,裴晓芳,周磊,等.反向加权融合多尺度特征的X射线图像违禁品检测[J].国外电子测量技术,2024,43(4):170-180.  
MA CH S, PEI X F, ZHOU L, et al. Reverse weight fusion of multi-scale features for prohibited object detection in X-ray images [J]. Foreign Electronic Measurement Technology, 2024, 43(4): 170-180.
- [19] YU Z, HUANG H, CHEN W, et al. Yolo-facev2: A scale and occlusion aware face detector[J]. Pattern Recognition, 2024, 155: 110714.
- [20] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression [C]. AAAI Conference on Artificial Intelligence, 2020.
- [21] ZHANG S F, CHI C, YAO Y, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 9759-9768.
- [22] FENG C J, ZHONG Y J, GAO Y, et al. Tood: Task-aligned one-stage object detection [C]. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE Computer Society, 2021: 3490-3499.
- [23] WANG C Y, LIAO H Y M, WU Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020: 390-391.
- [24] WANG C Y, LIAO H Y M, YEH I H. Designing network design strategies through gradient path analysis[J]. ArXiv preprint arXiv:2211.04800, 2022.
- [25] GAO S H, CHENG M M, ZHAO K, et al. Res2net: A new multi-scale backbone architecture [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 43(2): 652-662.
- [26] LI X, WANG W H, WU L J, et al. Generalized focal loss: Towards efficient representation learning for dense object detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(3): 3139-3153.
- [27] TONG Z, CHEN Y, XU Z, et al. Wise-IoU: bounding box regression loss with dynamic focusing mechanism[J]. ArXiv preprint arXiv: 2301.10051, 2023.
- [28] ZHANG H, XU C, ZHANG S J. Inner-IoU: More effective intersection over union loss with auxiliary bounding box [J]. ArXiv preprint arXiv: 2311.02877, 2023.
- [29] MIAO C, XIE L, WAN F, et al. Sixray: A large-scale security inspection x-ray benchmark for prohibited item discovery in overlapping images[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 2119-2128.
- [30] TAO R, WEI Y, JIANG X, et al. Towards real-world X-ray security inspection: A high-quality benchmark and lateral inhibition module for prohibited items detection [C]. IEEE/CVF International Conference on Computer Vision, 2021: 10923-10932.
- [31] REN SH Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(6): 1137-1149.
- [32] HE K M, ZHANG X Y, REN SH Q, et al. Deep residual learning for image recognition [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [33] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]. IEEE International Conference on Computer Vision, 2017: 2980-2988.
- [34] TIAN Z, SHEN C, CHEN H, et al. FCOS: Fully convolutional one-stage object detection [C]. IEEE/CVF International Conference on Computer Vision, 2019: 9627-9636.
- [35] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolo v4: Optimal speed and accuracy of object detection[J]. ArXiv preprint arXiv:2004.10934, 2020.
- [36] ZHANG G, LIU S, WANG F, et al. Yolo x: Exceeding yolo series in 2021 [J]. ArXiv preprint arXiv:2107.08430, 2021.
- [37] WANG Z S, WANG X H, SHI Y T, et al. Lightweight detection method for X-ray security inspection with occlusion [J]. Sensors, 2024, 24(3): 1002.
- [38] WANG C Y, YEH I H, LIAO H Y M. YOLOv9: Learning what you want to learn using programmable gradient information [J]. ArXiv preprint arXiv: 2402.13616, 2024.
- [39] WANG A, CHEN H, LIU L, et al. YOLOv10: Real-time end-to-end object detection [J]. ArXiv preprint arXiv:2405.14458, 2024.

## 作者简介

黎作鹏(通信作者),博士,副教授,主要研究方向为边缘计算、计算机视觉。

E-mail: lizuopeng@hebeu.edu.cn

刘佳祥,硕士研究生,主要研究方向为图像处理、计算机视觉。

E-mail: ljx823566386@163.com

张少文,硕士研究生,主要研究方向为计算机视觉。

E-mail: zhangsw0821@163.com