

基于 YOLOv8s 的 X 光违禁品识别^①



陈冠豪, 潘广贞

(中北大学 软件学院, 太原 030051)
通信作者: 潘广贞, E-mail: 228390888@qq.com

摘要: 快速增长的安检需求推动了智能安检技术的发展, 由于 X 光图片的特殊性, 导致小目标违禁品检测难度大, 本文提出一种改进 YOLOv8s 的违禁品识别网络进行解决. 首先引入 Focal L1 Loss 函数改进 *CIoU*, 优化预测框位置与长宽比, 以提高网络对于违禁品的标记能力. 将改进后的可变形卷积加入浅层主干网络中, 捕获不同方向上的违禁品特征信息. 在 SPPF 模块加入 LSKA 扩大网络感受野, 并使用 Swin-CS 模块捕获全局信息、补充维度交互, 最后使用 3 种注意力堆叠的注意力块进行处理, 提高了网络对小目标的敏感性. 改进后的网络在 SIXray 数据集上的平均精度均值达到 96.1%, 相比原有的 YOLOv8s 提高 5.4%, *mAP*₅₀₋₉₅ 达到 0.682, 提高 4.5%. 实验结果表明, 提出的模型能够准确给出预测框, 应对复杂场景中的违禁品检测, 证明了算法的有效性.

关键词: 安检; 违禁品; 感受野; 小目标检测

引用格式: 陈冠豪,潘广贞.基于 YOLOv8s 的 X 光违禁品识别.计算机系统应用. <http://www.c-s-a.org.cn/1003-3254/9696.html>

X-ray Contraband Identification Based on YOLOv8s

CHEN Guan-Hao, PAN Guang-Zhen
(School of Software, North University of China, Taiyuan 030051, China)

Abstract: The rapid growth of security inspection demand drives the development of intelligent security inspection technology. Due to the unique characteristics of X-ray images, detecting small contraband items is challenging. This study proposes an improved YOLOv8s network for contraband recognition to address this issue. Firstly, the Focal L1 Loss function is introduced to enhance *CIoU* and optimize the position and aspect ratio of prediction boxes to improve the network's ability to identify contraband items. Improved deformable convolution is added to the shallow backbone network to capture features of contraband items in different directions. LSKA is incorporated into the SPPF module to expand the network's receptive field, while the Swin-CS module captures global information and supplements dimensional interaction. Finally, three stacked attention blocks are used for processing, enhancing the network's sensitivity towards small targets. The improved network achieves an average precision mean of 96.1% on the SIXray dataset, a 5.4% improvement over YOLOv8s with *mAP*₅₀₋₉₅ reaching 0.682, a 4.5% increase. Experimental results indicate that the proposed model can accurately generate prediction boxes, effectively handle contraband detection in complex scenarios, and validate algorithm effectiveness.

Key words: security check; contraband; receptive field; small target detection

目前安检工作仍依赖人力, 容易发生漏检误检, 产生安全隐患或导致人员拥堵. 使用深度学习技术检测

违禁品, 能够客观可靠地辅助安检人员工作, 为人们带来更好的安全保障.

^① 基金项目: 山西省自然科学基金面上基金 (20210302124192)

收稿时间: 2024-04-29; 修改时间: 2024-06-04; 采用时间: 2024-06-28; csa 在线出版时间: 2024-10-25

检测图片中含有的违禁品并标识出其准确位置,属于计算机视觉中的目标检测领域,目前目标检测模型主要分为双阶段(two-stage)和单阶段(one-stage)两类.双阶段目标检测特点是先生成 region proposal 区域,随后使用卷积神经网络完成样本分类,代表算法有:R-CNN^[1]、Mask R-CNN^[2]、SPPNet^[3]、Fast R-CNN^[4]和 Faster R-CNN^[5]等.单阶段目标检测则直接通过深度卷积神经网络进行定位分类,因此可以在一个阶段完成目标类别以及位置坐标的生成,典型算法有 YOLO 系列^[6-8]、SSD^[9]等.

Ren 等人基于 YOLOv4 算法,将轻量级特征金字塔网络(LFPN)与卷积块注意力模块(CBAM)^[10]结合,捕获不同尺度特征,实时性得到增强^[11].李松等人通过引入 MoblieViTv3 额外分类器,同时对 MPConv 模块进行了改进,减少了模型参数量,提升了模型检测速度^[12].上述方法均考虑了违禁品检测中的实时性问题,但是对于违禁品图片中小目标违禁品的识别仍有欠缺,在各类目标检测中,小目标更容易在图片中出现紧凑收纳、与其他物品重叠等特殊情况,引起漏检.而 X 光违禁品图片生成依靠的是 X 光的透视特性,因此 X 光图片中的目标重叠情况要比普通目标检测情况更复杂,因此需要模型能够对小目标有更好的鲁棒性.郭守向等人构建了 YOLO-C 目标检测网络,引入特征增强模块(FAB)增强小目标的检测性能^[13].李文强等人提出一种自适应空间特征融合结构提高模型对小目标的检测能力^[14].上述方法一定程度增强了模型对小目标的识别能力,而在复杂背景下,小目标违禁品的识别检测依旧有一定难度,X 光违禁品图片常含有大量空白区域,这将影响训练时的权重分配,剔除无用背景并保留违禁品特征,实现识别效果提升.

X 光违禁品检测其独特检测难题主要体现在小目标违禁品检测难度大,重叠情况更复杂,而 X 光违禁品图片中的大量空白区域对违禁品的权重分配产生较大的影响,增加了小目标违禁品的检测难度,给传统目标检测带来了一定的挑战,为此,本文基于 YOLOv8s^[15]做出改进,保证检测速度的同时有效解决违禁品识别中复杂背景小目标难检测问题,减少漏检情况的发生.

1 YOLOv8 网络模型

相比其他 YOLO 模型, YOLOv8 拥有最新的网络架构. YOLOv8 模型为单阶段目标检测模型,检测速度

快,能够满足违禁品检测实时性需求,有 5 种不同大小的模型, YOLOv8n、s、m、l、x,本文以 YOLOv8s 作为基础模型进行改进, YOLOv8 网络结构主要由输入端(Input)、骨干网络(Backbone)、脖颈(Neck)部分以及预测端(Detect)组成.

输入端主要进行数据预处理. YOLOv8 在骨干网络采用了梯度流更丰富的 C2f 结构,并在最后使用 SPPF (spatial pyramid pooling) 模块减少图的相关重复特征的多次提取,加速候选框生成速度并提高生成质量. Neck 部分依旧使用 C2f 结构帮助网络学习相关特征,配合 Upsample 帮助生成不同大小候选框.解耦头预测端完成对检测物品种类和位置的输出.

2 YOLOv8 模型改进

本文对 YOLOv8s 做出的主要改进如下.

引入 Focal L1 Loss^[16]优化 $CIoU$, 提高预测框和目标框的重合度,保证预测框精度.

网络浅层加入 DCH 模块以使网络掌握不同方向上的违禁品特征,为模型带来不同方向上的违禁品特征,提高了模型的特征提取能力并减少了空白区域对模型训练的影响.

SPPF 上使用深度空洞可分离卷积,并使用两个级联的水平和垂直卷积核进行优化,降低了该模块的参数量同时完成图的相关重复特征的多次提取,充分利用违禁品特征信息.

在 Neck 部分加入 Swin-CS 模块,该模块融合 Swin Transformer^[17]和 CBAM 的思想,通过分组计算避免了计算量的增加,使该模块感受野扩大的同时补充了违禁品多维交互信息,增强对小目标违禁品的特征提取能力.最后使用 3 种注意力叠加对检测头进行改进,增强模型对小目标违禁品的检测能力.改进后的结构如图 1 所示.

2.1 使用 Focal L1 Loss 优化预测框

YOLOv8 使用的损失函数主要包括 3 部分:分类损失、回归损失和 DFL 损失.分类损失使用的是二元交叉熵损失,回归损失来度量模型对目标边界框的回归准确度,用来衡量预测边界框与真实框之间的差距. YOLOv8 中使用的回归损失为 $CIoU$, 如式 (1):

$$CIoU = IoU - \frac{\rho^2(b, b^{gt})}{c^2} - av \quad (1)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{\omega^{gt}}{h^{gt}} - \arctan \frac{\omega}{h} \right)^2 \quad (2)$$

$$\alpha = \frac{v}{(1 - IoU) + v} \quad (3)$$

其中, $IoU = |A \cap B| / |A \cup B|$, $\frac{\rho^2(b, b^{gt})}{c^2}$ 为惩罚项, αv 作为额外的惩罚项, 使预测框的长宽比更接近真实框, 避免了预测框和目标框重合产生损失函数退化, 识别效果变差.

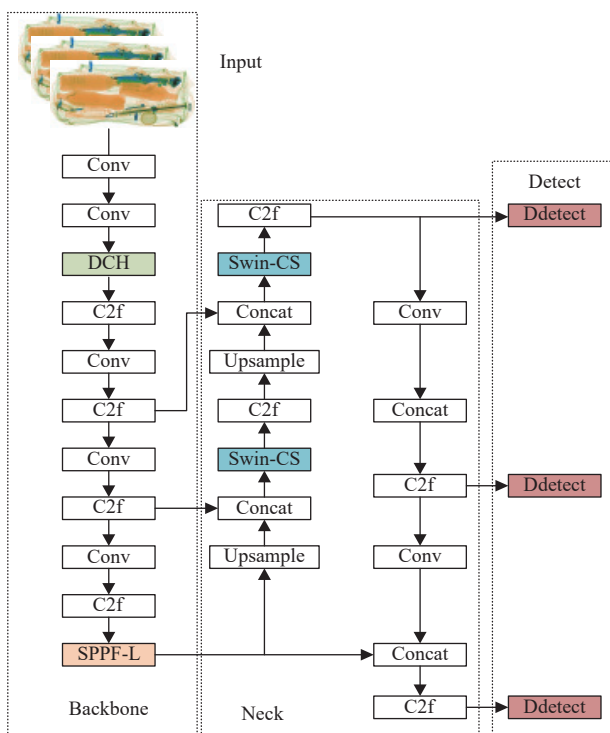


图 1 改进后的 YOLOv8s 网络结构图

数据在进行训练的时候, 由于不同数据类型的数量存在的差异可能导致模型训练的偏向性和性能下降, 因此 Zhang 等人^[16]借鉴 Focal Loss^[18]的思想提出使用 Focal L1 Loss 将高质量锚框和低质量锚框分开来解决这个问题, 并且整合论文中提出的 $EIoU$, 得到 Focal $EIoU$ Loss. $EIoU$ 以及 Focal $EIoU$ Loss 公式如式 (4) 和式 (6):

$$EIoU = IoU - \frac{\rho^2(b, b^{gt})}{c^2} - \Delta C_{\omega h} \quad (4)$$

$$L_{EIoU} = 1 - EIoU \quad (5)$$

$$\Delta C_{\omega h} = \frac{\rho^2(\omega, \omega^{gt})}{C_{\omega}^2} + \frac{\rho^2(h, h^{gt})}{C_h^2} \quad (6)$$

$$L_{Focal-EIoU} = IoU^{\gamma} L_{EIoU} \quad (7)$$

其中, γ 为超参数, 本文设置为 2, 负样本的 IoU 较小, 使用该参数可以降低负样本的 Loss 值, 以 Focal $EIoU$ Loss 为基础, Focal $CIoU$ 的公式为式 (8):

$$L_{Focal-CIoU} = IoU^{\gamma} L_{CIoU} \quad (8)$$

$$L_{CIoU} = 1 - CIoU \quad (9)$$

同样的思想也可以运用到其他损失函数如 $GIoU$ 、 $DIoU$ 上, 本文还加入了以预测框和真实框之间向量角度作为惩罚项的 $SIoU$ ^[19]进行了实验. $SIoU$ 公式如式 (10):

$$L_{SIoU} = 1 - IoU + \frac{\Delta + \Omega}{2} \quad (10)$$

其中, Δ 和 Ω 分别为距离损失和形状损失.

2.2 改进的可变形卷积模块

违禁品的种类多种多样, 不同违禁品形状以及在图片中的位置是更复杂多样的, 靠人眼识别容易发生误检漏检情况. 可变形卷积能够掌握不同方向上的信息, 将不同方向上的违禁品特征信息集中到一个卷积核中, 可以看作一种特殊的空洞卷积, 不会扩大卷积的大小, 但是依旧能够让神经元突破普通卷积核的限制. 其作用原理是在普通卷积的作用区域上加入可学习的参数 Δp_n , 权重系数 Δm_k 用以剔除无用背景, 最终扩散得到非网格形状的卷积核, 使网络学习到不同方向上的违禁品特征. 可变形卷积 v2 的具体计算公式如式 (11)^[20]所示:

$$y(p_0) = \sum_{k=1}^K w_k m_k x(p_0 + p_k + \Delta p_k) \quad (11)$$

安检机进行安检时, 如果物品较少, 那么传来的图片中会含有大量的空白背景区域, 模型不能在这些区域学习到有用信息, 感受野扩大后会使这些区域权重增加, 反而对模型训练产生影响, 因此本文对 DCH 模块进行重新设计, 增加跳跃连接淡化空白位置权重, 并使用 Hard Sigmoid 激活函数更快捕捉特征, 结构如图 2 所示.

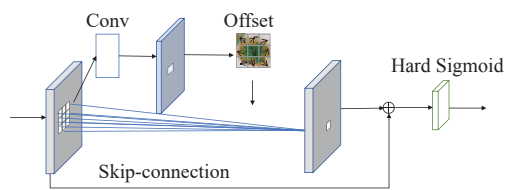


图 2 基于可变形卷积的 DCH 模块

2.3 融入 LSKA 结构的 SPPF 模块

感受野决定了神经元能够掌握的原始图像范围,影响着网络学习到更全局或更局部的信息.在违禁品识别过程中,模型通过一定大小的卷积核进行特征提取,这意味着在固定的算法中,模型的感受野将是受限的,模型在训练过程中可能只提取到了违禁品的部分特征信息,通过扩大网络感受野让神经元掌握到更全局的信息以及语义层次更高的特征.

SPPF 模块能够完成图的相关重复特征的多次提取,因此在该模块扩大感受野能够有效剔除更多无用背景特征. LKA (large kernel attention) 使用一种级联的深度卷积和空洞的深度卷积,将其融入 SPPF 模块中扩大模型感受野.图 3(a)展示了 LKA 结构,其中 k 是卷积核大小, d 是空洞率. (b) 为 LSKA^[21]结构,其使用对应的水平和垂直卷积核 DW-Conv (depthwise separable convolution) 和 DW-D-Conv (depthwise dilated separable convolution) 进行了替换,相关参数计算如式 (12)、式 (13):

$$Parm_{LKA} = (2d - 1)^2 \times C + \left[\frac{k}{d} \right]^2 \times C + C^2 \quad (12)$$

$$Parm_{LSKA} = (2d - 1) \times C \times 2 + \left[\frac{k}{d} \right] \times C \times 2 + C^2 \quad (13)$$

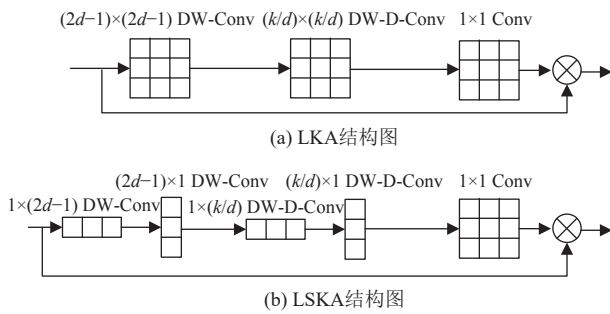


图 3 LKA 和 LSKA 结构图

比较式 (12)、式 (13), 容易得出 LSKA 相比 LKA 在 DW-Conv 部分节省 $(2d-1)/2$ 个参数, DW-D-Conv 节省 $1/2[k/d]$, 并将 ConvBNSiLU (Conv+BN+Sigmoid) 替换为速度更快的 ConvBNReLU (Conv+BN+ReLU), 提升检测速度, 改进后的 SPPF-L 模块如图 4.

2.4 优化小目标检测

复杂背景中的小目标经常受到背景的干扰,小目标在图片中所占比例小,信息量少,将背景的干扰降低并发掘更多小目标本身的信息有助于减少漏检,因此本文将 Swin Transformer 的滑动多头自注意力与 CBAM

注意力进行融合,使 Swin Transformer 能够处理违禁品图片中信道和通道信息,实现维度交互信息的提取利用,同时 Swin Transformer 能够提取全局特征,减少无关背景对小目标识别的影响.

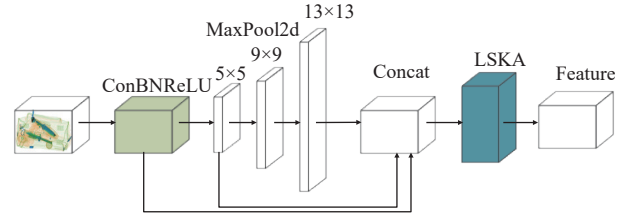


图 4 SPPF-L 结构示意图

YOLOv8 在 Neck 部分使用特征金字塔结构 FPN (feature pyramid network) 对不同层级的特征进行融合,生成 3 种不同大小的锚框来对违禁品进行定位.因此将设计的新模块融入 Neck 部分来提高模型对小目标信息的特征提取能力.

新的注意力模块称为 Swin-CS 模块,结构如图 5 所示. Swin Transformer 对 Transformer 计算复杂度大的问题进行了解决,将 token 进行分组计算以降低计算复杂度. Swin Transformer 由带 W-MSA (window multi-head self-attention module) 的多头自注意力模块和带 SW-MSA (shifted window multi-head self-attention module) 的多头自注意力模块组成,其中 W-MSA 完成分组工作,直接计算每个窗口中 token 的自注意力值. SW-MSA 通过滑动窗口的方式完成相邻 patch 的信息交互,有效减少特征信息的丢失. Channel 取自 CBAM 中的通道注意力模块,其用来压缩特征图空间维数,聚合空间信息,使模块关注通道信息, Spatial 则取自 CBAM 中的空间注意力模块,作用是对通道进行压缩,使模块关注空间信息.

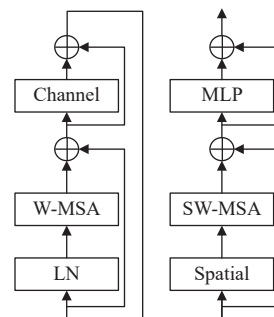


图 5 Swin-CS 模块结构

为了将 Neck 传来的带有不同维度交互信息的违禁品信息进行充分利用,本文将 DyHead^[22]融入检测头

部分进行处理, Dyhead 能够综合尺度信息、空间信息以及通道信息, 该模块由 3 部分组成: π_L 、 π_S 、 π_C . 其中 π_L 根据尺度语义重要性融合不同尺度信息, 平滑不同尺度特征, 将小目标的尺度信息提取出来并将其用

于 π_S 上, π_S 通过结合可变形卷积聚合空间信息, 获取到不同方向上的信息, 最后由 π_C 实现任务感知, 动态控制关键特征的使用. 改进后的 Detect 检测头结果如图 6 所示.

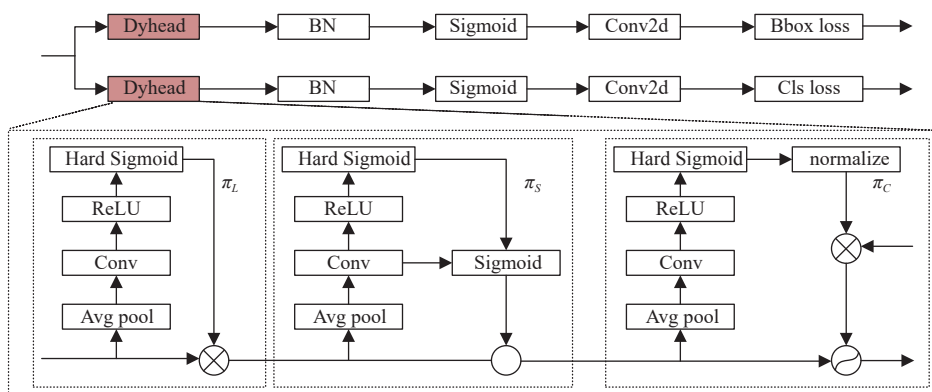


图 6 改进后的 Detect 检测头

3 实验

3.1 实验环境

本文基于 SIXray 数据集进行实验, SIXray 数据集由 Miao 等人^[23]收集, 共包含 1059231 张 X 光图像, 其中 8929 张图像包含违禁品, 共有 5 类违禁物品, 分别为 Gun、Knife、Wrench、Pliers、Scissors. 本文将包含违禁品的图像按照比例 7:2:1 的比例分为训练集、验证集、测试集.

实验环境: NVIDIA GeForce RTX3090. 编程语言为 Python3.8, Cuda 版本为 12.1, PyTorch 为 2.0.

超参数配置: 模型训练中, 目标的类别置信度阈值为 0.7, 初始学习率为 0.01, 权重衰减系数设置为 0.0005, 以防止数据过拟合. Batch size 设置为 16, 一共训练 100 个 epoch.

3.2 评价标准

准确率 (precision, P)、召回率 (recall, R) 和平均精度均值 (mean average precision, mAP) 通常用来对目标检测算法性能进行评价. 其中, 准确率 P 是指真正的正样本与模型识别出的正样本的比率, TP 表示模型正确预测违禁品的个数; FP 表示模型将并不存在的违禁品错误检测的个数. 计算公式如式 (14) 所示:

$$P = \frac{TP}{TP + FP} \times 100\% \quad (14)$$

召回率 R 是指在测试集所有真正的正样本中, 模型正确检测出违禁品的个数所占的比率, FN 表示没有

被检测到的个数, 计算公式如式 (15) 所示:

$$R = \frac{TP}{TP + FN} \times 100\% \quad (15)$$

mAP 值可以作为一个综合的参数来检测网络性能, $mAP50$ 指交并比 $IoU=50$ 时, 模型的类识别准确率, $mAP50-95$ 则更为严苛, 其计算方式为: 从 $mAP50$ 到 $mAP95$, 每间隔 5% 取一次 mAP , 对取得的 10 个值取平均, 反映了模型预测框的准确程度, 多用于一些准确性和安全性需求高的场景中, 因而能够反应违禁品识别模型的性能. 计算公式如式 (16) 所示:

$$mAP = \frac{1}{C} \sum_{K=1}^N P(K) \Delta R(K) \quad (16)$$

3.3 对比实验

3.3.1 损失函数改进对比

表 1 展示了引入 Focal L1 Loss 后不同损失函数的实验结果. 在 $mAP50$ 值中, Focal L1 $CIoU$ 与 Focal L1 $SIoU$ 持平, 但是 $mAP50-95$ 值中 Focal L1 $CIoU$ 损失函数提升最高, 为 3.1 个百分点, 综合来讲 Focal L1 $CIoU$ 能够得到比其他损失函数更好的回归, 而 Focal L1 $SIoU$ 或许也是优化损失函数的一个方向.

3.3.2 改进的可变形卷积实验

为研究 DCH 在网络能够发挥最大作用的位置, 本文在不同网络层数进行了相关对比实验, 结果如表 2, 其中 L3 表示在网络第 3 层使用 DCHv2 对卷积核进行改进, 实验发现 DCHv2 在浅层网络表现更好, 这说明

浅层网络的方向向量对于识别违禁品更有价值,提供的方向向量更有利于对违禁品进行检测定位,因此在后面的消融实验中在网络第3层进行改进.

表1 损失函数实验结果

模型	P	R	$mAP50$	$mAP50-95$
YOLOv8s	0.937	0.851	0.907	0.637
Focal GIoU	0.912	0.849	0.905	0.66
Focal DIoU	0.918	0.837	0.904	0.655
Focal $CloU$	0.933	0.846	0.91	0.668
Focal $EIoU$	0.925	0.83	0.904	0.662
Focal $SIoU$	0.93	0.837	0.91	0.656

表2 可变形卷积 v2 不同位置对比

模型	P	R	$mAP50$	$mAP50-95$
L3	0.93	0.852	0.917	0.676
L5	0.935	0.846	0.916	0.675
L7	0.898	0.843	0.91	0.654

3.3.3 SPPF 改进对比

改进之后的 SPPF 模块是否能够提高违禁品的检测效果,本文设计了对比实验进行分析,实验结果如表3所示.将 SPPF 改进之后, mAP 值提高了 1.4 个百分点, $mAP50-95$ 则提高了 3.6 个百分点,说明改进之后的算法能够有效利用扩大后的感受野信息,提取违禁品重要特征,更加准确地识别违禁品并进行匹配.

表3 改进 SPPF 实验结果

模型	P	R	$mAP50$	$mAP50-95$
YOLOv8s	0.937	0.851	0.907	0.637
改进SPPF	0.945	0.852	0.921	0.673

3.3.4 小目标检测实验

表4展示了优化小目标检测的实验结果,实验发现添加 Swin-CS 的模型能够很好捕获违禁品特征信息,准确率和召回率均有提升, mAP 和 $mAP50-95$ 也提升明显.另外本文添加了额外的小目标检测头实验进行对比,观察额外的小目标检测头是否优于本文检测头改进,表中数据可以看出额外的检测头确实提高了模型的检测性能, $mAP50-95$ 提升了 3.2 个百分点,但是相比本文方法低 0.4 个百分点, $mAP50$ 低 0.7 个百分点,因此相较于额外检测头,本文方法更佳.

表4 小目标检测对比实验

模型	P	R	$mAP50$	$mAP50-95$
YOLOv8s	0.937	0.851	0.907	0.637
Swin-CS	0.938	0.853	0.927	0.681
额外检测头	0.931	0.855	0.91	0.669
Dyhead	0.941	0.849	0.917	0.673

3.4 不同模型对比实验

为了能够体现本文改进算法的优越性,与其他深度学习模型在相同的硬件环境下进行了对比实验,实验结果如表5所示,改进后的算法在各方面均展现出优势,有较高的处理速度,能满足实时性需求,FPS 相比 YOLOv8s 提高 4.4,说明在 SPPF 中使用的深度可分离卷积能够有效提高模型的计算速度,分组计算也有效避免了模型计算量的增加,准确率、召回率和 mAP 均领先其他深度学习算法,意味着模型也能完成高准确率的检测.

表5 不同模型对比实验

模型	P	R	$mAP50$	FPS
Faster R-CNN	0.773	0.692	0.774	10.5
SSD	0.775	0.687	0.776	41.6
DenseNet	0.768	0.671	0.772	72.2
YOLOv5s	0.902	0.821	0.899	94.9
YOLOv6s	0.899	0.811	0.884	93.3
YOLOv5-X	—	—	0.904	98
YOLOv8s	0.937	0.851	0.907	105
Ours	0.954	0.868	0.961	109.4

3.5 消融实验

此节对本文提出的改进 YOLOv8s 模型进行消融实验以验证各模块的有效性,其结果如表6中所示,将损失函数替换为 Focal L1 $CloU$ (FC) 后, $mAP50-95$ 得到明显的提升,上升了 3.1 个百分点,DCHv2 利用不同方向上的违禁品特征信息 $mAP50$ 提升至 0.925. 使用 LSKA 增强 SPPF 特征提取后, $mAP50$ 继续提升 1.1 个百分点. Swin-CS 模块和 Dyhead 检测头最终使得 $mAP50-95$ 稳定在 0.682, $mAP50$ 提高至 0.961.

表6 消融实验

模型	P	R	$mAP50$	$mAP50-95$
YOLOv8s	0.937	0.851	0.907	0.637
FC	0.933	0.846	0.91	0.668
FC+DCH	0.935	0.847	0.925	0.668
FC+DCH+SPPF-L	0.938	0.852	0.936	0.674
Ours	0.954	0.868	0.961	0.682

综上所述,改进后的网络拥有较高的准确率和特征提取能力,在复杂场景下有较好的鲁棒性,能够满足安检中违禁物品识别应用的实时性和准确性需求.

3.6 改进效果展示

热力图可以观察到网络对存在违禁品区域的关注度程度,图7展示了改进前(左列)和改进后(右列)普

通样本和复杂背景下的热力图对比. 改进前的热力图中, 违禁品有很多部分没有很好地分配到权重. 改进后热力图显示的违禁品区域更完整.

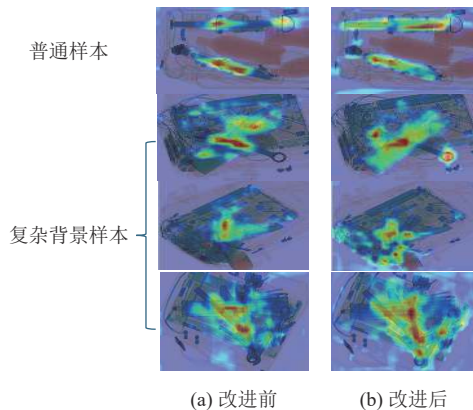


图7 改进后的热力图展示

图8展示了改进前(左图)和改进后(右图)模型对部分小目标的关注程度, 颜色越鲜艳代表关注程度越高. 可以观察到在复杂背景样本中, 小目标有明显的漏检情况, 图中存在的钳子、扳手没有进行权重分配, 而改进后模型对小目标的关注度明显提高, 给钳子和扳手分配了较高的权重.

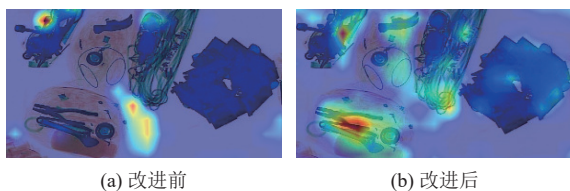


图8 小目标检测效果热力图

4 结束语

本文提出了一种改进YOLOv8s网络模型的违禁品识别网络, 针对小目标违禁品识别难的问题进行了改进, 使用Focal L1 Loss改进损失函数优化预测框. 改进后的DCH模块在网络浅层获取到不同方向上的违禁品关键特征信息, 再通过改进SPPF以及Swin-CS模块在网络的主干和Neck部分提取了全局信息, 剔除无用背景的同时, 使模型对小目标更敏感, 改进后的Dyhead检测头综合多方面信息对检测出的目标信息进行了输出. 最终使得准确率和召回率均提高1.7个百分点, $mAP50$ 提高了5.4个百分点, $mAP50-95$ 提高了4.5个百分点.

参考文献

- 1 Girshick R, Donahue J, Darrell T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014. 580–587.
- 2 He KM, Gkioxari G, Dollár P, *et al.* Mask R-CNN. Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 2980–2988.
- 3 He KM, Zhang XY, Ren SQ, *et al.* Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904–1916. [doi: [10.1109/TPAMI.2015.2389824](https://doi.org/10.1109/TPAMI.2015.2389824)]
- 4 Ren SQ, He KM, Girshick R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks. Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal: MIT Press, 2015. 91–99.
- 5 Girshick R. Fast R-CNN. Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015. 1440–1448.
- 6 Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Optimal speed and accuracy of object detection. arXiv:2004.10934, 2020.
- 7 Li CY, Li LL, Jiang HL, *et al.* YOLOv6: A single-stage object detection framework for industrial applications. arXiv: 2209.02976, 2022.
- 8 Wang CY, Bochkovskiy A, Liao HYM. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 7464–7475.
- 9 Liu W, Anguelov D, Erhan D, *et al.* SSD: Single shot multibox detector. Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 21–37.
- 10 Woo S, Park J, Lee JY, *et al.* CBAM: Convolutional block attention module. Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich: Springer, 2018. 3–19.
- 11 Ren Y, Zhang HG, Sun HX, *et al.* LightRay: Lightweight network for prohibited items detection in X-ray images during security inspection. Computers and Electrical Engineering, 2022, 103: 108283. [doi: [10.1016/j.compeleceng.2022.108283](https://doi.org/10.1016/j.compeleceng.2022.108283)]

- 12 李松, 亚森江·木沙. 改进 YOLOv7 的 X 射线图像违禁品实时检测. 计算机工程与应用, 2023, 59(12): 193–200.
- 13 郭守向, 张良. YOLO-C: 基于单阶段网络的 X 光图像违禁品检测. 激光与光电子学进展, 2021, 58(8): 0810003.
- 14 李文强, 陈莉, 谢旭, 等. 改进 YOLOv5 的 X 光图像违禁品检测算法. 计算机工程与应用, 2023, 59(16): 170–176.
- 15 Jocher G, Chaurasia A, Qiu J. Ultralytics YOLO (Version 8.0.0). <https://github.com/ultralytics/ultralytics>. [2023-12-14].
- 16 Zhang YF, Ren WQ, Zhang Z, *et al.* Focal and efficient IoU loss for accurate bounding box regression. Neurocomputing, 2022, 506: 146–157. [doi: 10.1016/j.neucom.2022.07.042]
- 17 Liu Z, Lin YT, Cao Y, *et al.* Swin Transformer: Hierarchical vision Transformer using shifted windows. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021. 9992–10002.
- 18 Lin TY, Goyal P, Girshick R, *et al.* Focal loss for dense object detection. Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 2990–3007.
- 19 Gevorgyan Z. SIOU loss: More powerful learning for bounding box regression. arXiv:2205.12740, 2022.
- 20 Zhu XZ, Hu H, Lin S, *et al.* Deformable ConvNets V2: More deformable, better results. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 9300–9308.
- 21 Lau KW, Po LM, Rehman YAU. Large separable kernel attention: Rethinking the large kernel attention design in CNN. Expert Systems with Applications, 2024, 236: 121352. [doi: 10.1016/j.eswa.2023.121352]
- 22 Dai XY, Chen YP, Xiao B, *et al.* Dynamic head: Unifying object detection heads with attentions. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 7369–7378.
- 23 Miao CJ, Xie LX, Wan F, *et al.* SIXray: A large-scale security inspection X-ray benchmark for prohibited item discovery in overlapping images. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 2114–2123.

(校对责编: 张重毅)