

# 一种改进 YOLOv5 的 X 光违禁品检测模型

董乙杉<sup>1</sup>, 李兆鑫<sup>1</sup>, 郭靖圆<sup>1</sup>, 陈天宇<sup>1</sup>, 卢树华<sup>1,2\*</sup>

<sup>1</sup>中国人民公安大学信息安全学院, 北京 102600;

<sup>2</sup>公安部安全防范技术与风险评估重点实验室, 北京 102600

**摘要** 针对 X 光行李图像安全检测过程中存在物品高度重叠遮挡及复杂背景干扰等问题, 提出了一种融合注意力机制、数据增强策略与加权边框融合算法的改进 YOLOv5 网络模型用于 X 光违禁品检测。模型在 Neck 部分引入卷积注意力模块加强网络对违禁品深层重要特征的提取, 抑制背景干扰; 训练阶段采用 Mixup 数据增强策略模拟带有高度重叠及遮挡物品的检测场景, 加强模型复杂样本的学习能力; 测试阶段采用加权边框融合算法对冗余预测框进行优化, 提高模型精准预测能力。所提模型在 3 个大型复杂数据集 SIXray、HiXray、OPIXray 进行了测试, 平均精度均值分别达到了 89.6%、83.1% 和 91.6%。结果表明: 所提模型能够有效提高 YOLOv5 检测复杂违禁品的能力, 与现有诸多先进算法相比, 具有较高的准确率和稳健性。

**关键词** X 光图像; 违禁品检测; 注意力机制; 加权边框融合

中图分类号 TP391.4

文献标志码 A

DOI: 10.3788/LOP212848

## Improved YOLOv5 Model for X-Ray Prohibited Item Detection

Dong Yishan<sup>1</sup>, Li Zhaoxin<sup>1</sup>, Guo Jingyuan<sup>1</sup>, Chen Tianyu<sup>1</sup>, Lu Shuhua<sup>1,2\*</sup>

<sup>1</sup>College of Information and Cyber Security, People's Public Security University of China, Beijing 102600, China;

<sup>2</sup>Key Laboratory of Security Technology and Risk Assessment Ministry of Public Security, Beijing 102600, China

**Abstract** An improved YOLOv5 network model is proposed to resolve high overlap, heavy occlusion, and complex background interference in X-ray luggage image security detection by introducing the convolutional block attention module, data enhancement strategy, and the weighted boxes fusion algorithm for X-ray prohibited item detection. The convolutional block attention module is introduced in the Neck to enhance the extraction of deep important features and suppress background interference of X-ray prohibited items features. The Mixup data augmentation strategy is employed during the training process to simulate the detection scene with high overlap and heavy occlusion items to strengthen the learning ability of the model for complex samples. During the testing process, the weighted boxes fusion algorithm is used to optimize the redundant prediction boxes to enhance its prediction accuracy. The proposed model is tested on three large-size complex datasets (SIXray, HiXray, and OPIXray), resulting in mean average precision values of 89.6%, 83.1%, and 91.6%, respectively. The results show that the proposed model can effectively improve the ability of YOLOv5 in detecting complex contrabands. The proposed model performs better than many current advanced algorithms, indicating its high accuracy and robustness.

**Key words** X-ray image; prohibited items detection; attention mechanism; weighted boxes fusion

## 1 引言

X 光图像安全检查在机场、车站和地铁等人群密集场所应用广泛, 是安全防范的最后一道防线, 具有重要的作用<sup>[1-3]</sup>。然而, 目前的安检工作主要依靠人工对 X 光行李图像进行违禁品筛查, 需要安检人员根据专

业经验和知识进行判断, 准确率为 80%~90%<sup>[4]</sup>, 具有一定的漏报率和误报率<sup>[5]</sup>, 特别是在遇到大客流的情况下, 长时间操作易使工作人员处于疲劳状态, 导致检测效率与准确率下降<sup>[6]</sup>。因此, 使用计算机辅助人工进行违禁品检测成为研究的热点, 具有良好的应用前景<sup>[7-11]</sup>。

收稿日期: 2021-11-01; 修回日期: 2021-12-04; 录用日期: 2021-12-21; 网络首发日期: 2021-12-30

基金项目: 中央高校基本科研业务经费重大项目(2021JKF102)、公安学科基础理论研究专项(2021XKZX08)

通信作者: \*lushuhua@ppsuc.edu.cn

近年来,随着深度学习在计算机视觉领域的不断发展,基于卷积神经网络(Convolutional neural networks, CNN)的X光违禁品检测方法研究取得了显著的成果<sup>[9-10,12-13]</sup>。与传统的基于K最近邻法(K-nearest neighbor, KNN)分类<sup>[14-15]</sup>或支持向量机(Support vector machine, SVM)分类算法<sup>[11,16]</sup>的X光行李违禁品检测方法相比,基于CNN的检测方法在检测准确度与可靠性等方面都有较高的提升。例如,Hassan等<sup>[17]</sup>提出的级联张量架构(Cascaded structure tensor framework, CSTF)通过迭代不同方向上X光图像违禁品的轮廓信息来提取网络感兴趣的候选区域,降低噪声对网络的干扰;针对违禁品类别不平衡问题,Miao等<sup>[18]</sup>设计的类平衡层次改进(Class-balanced hierarchical refinement, CHR)模型利用相邻尺寸的特征信息来反向过滤当前尺寸下的不相关特征信息,加强对X光行李图像中重叠物品的特征提取性能;此外,Shao等<sup>[19]</sup>运用前后景分离(Foreground and background separation, FBS)模型自适应地学习带有X光违禁品的前景特征,将样本图像中的前景目标与背景分离出来,使得网络能够专注对前景目标的学习;为解决违禁品的遮挡问题,Wei等<sup>[20]</sup>提出去遮挡注意力(De-occlusion attention module, DOAM)模块,从注意力的角度出发,通过引入两个子模块分别强调网络对X光图像违禁品边缘信息和材料信息的提取来提升对X光行李图像违禁品的检测性能。上述方法在较大程度上提高了检测算法的性能,为违禁物品的智能检测奠定了良好的基础,但接近现实场景的违禁品检测依然面临物品高度重叠、复杂背景干扰<sup>[21]</sup>以及目标弱小<sup>[22]</sup>等挑战,准确性和鲁棒性仍需进

一步提高。

为解决上述问题,受文献<sup>[23-25]</sup>的启发,本文以YOLOv5<sup>[26]</sup>网络为基础框架,引入卷积块注意力(Convolutional block attention module, CBAM)模块<sup>[23]</sup>提高网络对深层有效信息的关注,采用Mixup<sup>[24]</sup>策略模拟高度重叠遮挡目标场景,使用加权边框融合(Weighted boxes fusion, WBF)算法<sup>[25]</sup>降低预测框筛选错误率,提出一种融合CBAM模块、Mixup数据增强以及WBF等策略的改进X光违禁品检测模型。所提模型在3个大型公开数据集上进行了验证和测试,性能优于当前大多数主流算法,展现出较好的先进性。

## 2 违禁品检测模型

### 2.1 基本原理

所提模型以YOLOv5网络为基本框架,结构如图1(a)所示,主要包括backbone、neck和prediction等3部分。其中,backbone部分由focus层、4个Conv层、4个C3模块以及空间金字塔池化(Spatial pyramid pooling, SPP)层构成,主要提取多尺度特征信息。neck部分采用特征金字塔网络(Feature pyramid networks, FPN)与路径聚合网络(Path aggregation network, PAN)结构对提取的特征进行多尺度融合,引入CBAM模块对网络提取的3种不同尺度的特征图进行优化,使网络聚焦不同尺度的有效特征,抑制背景干扰,结构如图1(b)所示。prediction部分采用3个detect模块,用于检测不同大小的目标。此外,在训练阶段,采用Mixup数据增强策略进行训练,提高网络对遮挡和重叠违禁品目标的学习能力;在测试阶段,使用WBF算法对冗余预测框进行处理,提高模型对复杂目标精准预测能力。

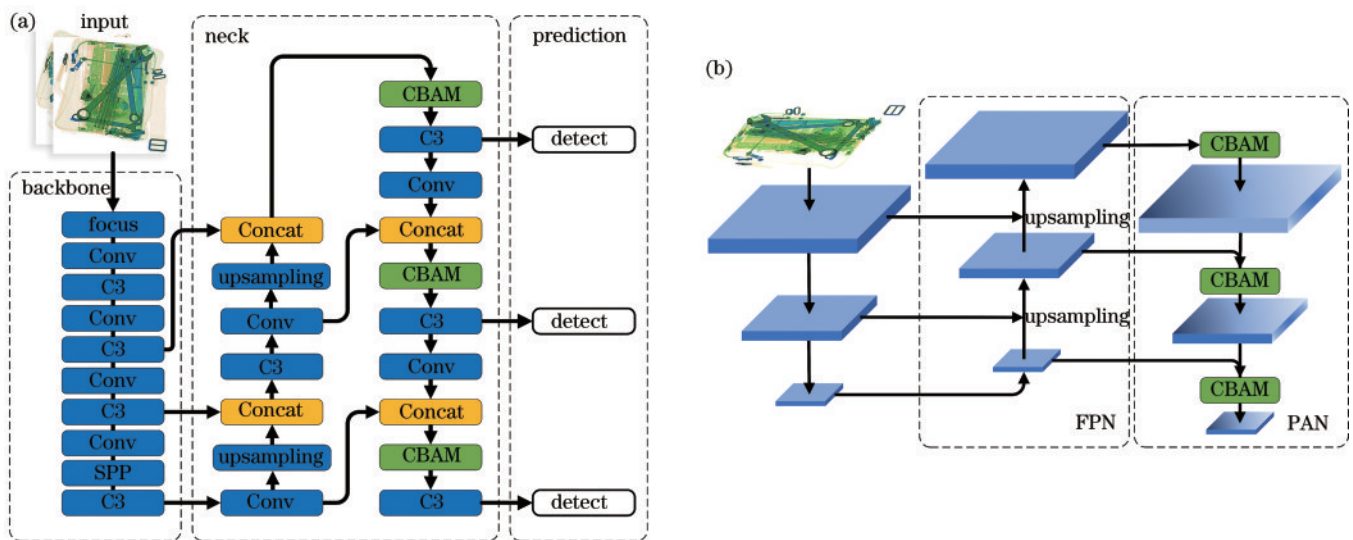


图1 网络结构图。(a)改进后的YOLOv5;(b)FPN与PAN

Fig. 1 Network structure. (a) Improved YOLOv5 model; (b) FPN and PAN

### 2.2 CBAM

YOLOv5网络仅采用卷积操作对数据特征进行

提取,缺乏有效的注意力机制来使网络关注更具重要性的特征。因此,引入CBAM模块对网络结构进行优

化。CBAM由通道注意力机制和空间注意力机制组合而成,如图2所示,能够在通道和空间2个独立维度

对输入特征进行压缩与重新加权,提高网络对于重要特征的关注度,抑制背景信息干扰。

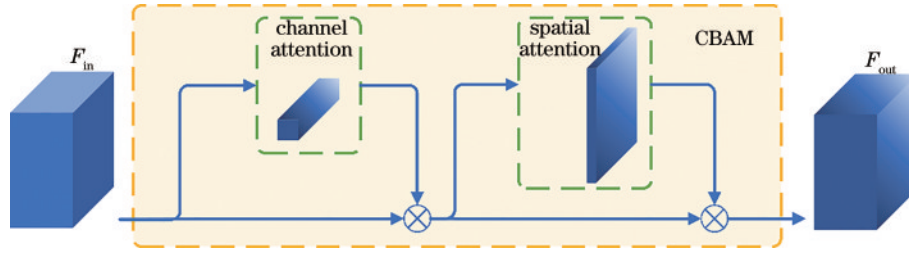


图2 CBAM注意力模块结构图

Fig. 2 Structure of CBAM attention module

给定一组输入特征图  $F_{in} \in \mathbb{R}^{C \times H \times W}$ , 其中  $C$  为输入特征图的通道数,  $H$  和  $W$  为特征图的高度和宽度。该特征图分别经过全局平均池化处理和全局最大池化处理得到2组处理后的特征图, 随后将2条分支经

过只含1个隐藏层的多层感知器, 得到2组尺度为  $\mathbb{R}^{C/r \times 1 \times 1}$  的特征向量, 其中  $r$  为隐藏层的压缩率。将2组特征向量相加, 然后使用 Sigmoid 函数激活得到输出特征向量:

$$F_{out}^c = \text{Sigmoid}(\text{MLP}(\text{AvgPool}(F_{in})) + \text{MLP}(\text{MaxPool}(F_{in}))) = \text{Sigmoid}(W_1(W_0(F_{avg})) + W_1(W_0(F_{max}))), \quad (1)$$

式中:  $F_{out}^c$  为经过通道注意力机制后的输出向量; Sigmoid 为激活函数; MLP 为多层感知器;  $W_0 \in \mathbb{R}^{C/r \times C}$  与  $W_1 \in \mathbb{R}^{C \times C/r}$  为共享的多层感知器权重; AvgPool 为全局平均池化操作; MaxPool 为全局最大池化操作;  $F_{avg}$  为经过全局平均池化操作后得到的特征向量;  $F_{max}$

为经过全局最大池化操作后得到的特征向量。

给定输入特征图  $F_{in} \in \mathbb{R}^{C \times H \times W}$ , 分别进行全局平均池化操作和全局最大池化操作, 得到2条特征向量并拼接, 经过1次  $7 \times 7$  的卷积操作, 最后通过 Sigmoid 函数激活并得到输出特征向量:

$$F_{out}^s = \text{Sigmoid}(f^{7 \times 7}([\text{AvgPool}(F_{in}); \text{MaxPool}(F_{in})])) = \text{Sigmoid}(f^{7 \times 7}[F_{avg}; F_{max}]), \quad (2)$$

式中:  $F_{out}^s$  为经过空间注意力机制后的输出向量,  $f^{7 \times 7}$  为使用卷积核大小为  $7 \times 7$  的卷积操作。

通过融合 CBAM 注意力机制, 对网络提取的特征图进行重构, 尽可能消除背景信息的干扰, 使得网络更加专注 X 光下各类违禁品的有效特征提取。

### 2.3 WBF

YOLOv5 网络采用非极大值抑制<sup>[27]</sup> (Non-maximum suppression, NMS) 算法对生成预测框进行筛选, 可能导致有效的预测框被删除, 降低检测效果。引入 WBF 算法, 其示意图如图3所示, 对生成预测框进行融合, 以提高预测的准确度, 减少重叠目标的漏检率, 使得预测边框更接近真实值 (Ground truth)。使用单个模型网络产生的多个预测框进行 WBF 操作, 具体过程如下:

1) 将网络生成的预测框进行归一化处理并放入单个列表 B 中, 按照置信度降序排列。

2) 建立空表 L 和 F。表 L 将所有预测框按类别进行分类并储存, 表 F 用于存储融合后的新边框。表 L 中的每个位置存储单个目标可能对应的所有预测框, 表 F 中的每个位置只存储 1 个来自于表 L 中由多个预测框融合后的新边框。

3) 遍历表 B 中每个预测框, 与表 F 中的预测框进行交并比 (Intersection-over-union, IoU) 计算, 如果该

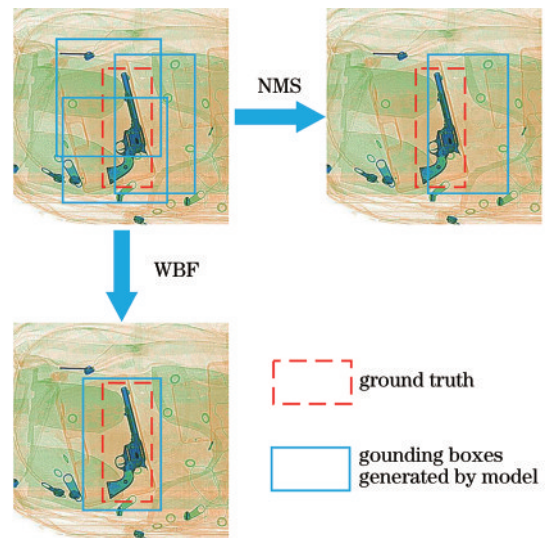


图3 WBF与NMS算法示意图

Fig. 3 Schematic diagrams of WBF and NMS algorithms

数值大于所设置的阈值, 即认为两框匹配。

4) 如果没有找到匹配的预测框, 则将该框添加到表 L 和表 F 的尾端并在下次循环中与表 B 中的预测框进行匹配。如果找到匹配的预测框, 将其按类别添加到表 L 对应的位置, 且该位置是表 F 中匹配框对应的索引值 (Index)。

5) 将表 L 中每个目标下的多个不同的预测框进行融合, 计算方式为

$$G = \max(G_i), i = 1, 2, \dots, N, \quad (3)$$

$$X_{1,2} = \sum_{i=1}^N G_i \times X_{1i,2i} / \sum_{i=1}^N G_i, \quad (4)$$

$$Y_{1,2} = \sum_{i=1}^N G_i \times Y_{1i,2i} / \sum_{i=1}^N G_i, \quad (5)$$

式中:  $N$  为表 L 中单个目标下的所有预测框个数;  $G$  为融合后预测框的置信度;  $X_{1,2} Y_{1,2}$  为融合后预测框坐标值;  $G_i$  为每个预测框的置信度;  $X_{1i,2i} Y_{1i,2i}$  为每个预测框的坐标。

6) 遍历完表 B 中的所有预测框后, 再对表 F 中的预测框按置信度分数进行降序排列并输出。

#### 2.4 数据增强策略

为提高网络对遮挡和重叠目标的分辨能力, 采用 Mixup 方法来模拟真实条件下的 X 光行李图像中遮挡重叠场景, 即在训练集中随机挑选 2 张图像, 将其像素值与标签值分别加权得到新图像与标签值, 再送入网络中进行训练, 训练过程中对 2 张图像中的标签分别计算损失并加权求和, 随后反向传播以更新网络参数。图 4 为 Mixup 数据增强实例图。在 Mixup 增强方法的基础上, 采用 Mosaic 数据增强方法, 将 4 张图像拼接为 1 张图像后送到网络中进行训练, 以减小计算量、加快训练速度。此外, 还对图像的亮度、对比度、饱和度、角度等参数进行随机改变, 减少过拟合现象。

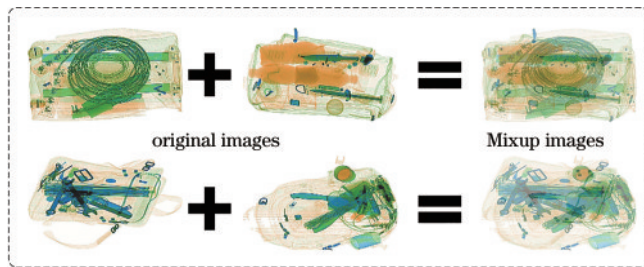


图 4 Mixup 数据增强实例  
Fig. 4 Example of Mixup data augmentation

## 3 实验与结果分析

### 3.1 数据集

SIXray: 数据集<sup>[18]</sup>总共包含 1059231 张 X 光行李图像, 包含枪、刀、扳手、钳子、剪刀、锤等 6 种类别的目标。数据集为模拟真实安检环境下违禁品出现的低概率性, 仅包含 8929 张含违禁品的图像 (占总数据集样本数量的 0.84%), 且只有 5 种类别的目标拥有标注信息。在实验中对含标注信息的 8929 张样本图像按 8:1:1 的比例分为训练集、验证集、测试集对模型进行训练和检测。

OPIXray: 数据集<sup>[20]</sup>包含 8885 张图像, 共有折叠刀、直刀、剪刀、美工刀、多用刀等 5 种类别的刀具, 实验中使用 6398 张图像作为训练集, 711 张图像作为验

证集, 1776 张图像作为测试集。

HiXray: 数据集<sup>[28]</sup>拥有 45364 张 X 光行李图像, 共 102928 个正样本标注信息, 包含菱形锂电池、圆柱形锂电池、液体、笔记本电脑、手机、平板电脑、化妆品、非金属类打火机等 8 种类别的目标。实验中将数据集分成 3 个子集, 分别为训练集 32665 张, 验证集 3630 张, 测试集 9069 张。

### 3.2 实验配置

实验在 Ubuntu 16.4 LTS 系统下进行, 网络框架使用 Pytorch 1.8 搭建, GPU 为 Nvidia RTX2080Ti。Batchsize 设置为 16, Epoch 设置为 300 次, 使用随机梯度下降 (Stochastic gradient descent, SGD) 法对网络参数进行调整。初始学习率为 0.01、权重衰减为 0.0005。Mixup 的重叠系数为 0.3, CBAM 压缩率  $r$  为 16, WBF 阈值  $R$  为 0.5。对图像预处理时将图像的宽和高均固定为 640。

### 3.3 评价指标

采用平均精度均值 (Mean average precision, mAP) 与平均精度 (Average precision, AP) 作为衡量预测结果的指标。mAP 与 AP 需要计算模型训练样本的准确度 ( $P_{\text{recision}}$ ) 和召回率 ( $R_{\text{ecall}}$ ), 其表达式为

$$P_{\text{recision}} = \frac{f_{\text{TP}}}{f_{\text{TP}} + f_{\text{TF}}}, \quad (6)$$

$$R_{\text{ecall}} = \frac{f_{\text{TP}}}{f_{\text{TP}} + f_{\text{TN}}}, \quad (7)$$

式中:  $f_{\text{TP}}$  为被正确识别的正样本;  $f_{\text{TF}}$  为被正确识别的负样本;  $f_{\text{TN}}$  为被错误识别的负样本。通过由  $P_{\text{recision}}$  和  $R_{\text{ecall}}$  所组成的准确度-召回率曲线 ( $P$ - $R$  曲线) 的下方面积可以计算出每种目标类别的平均精度:

$$x_{\text{AP}} = \int_0^1 P_{\text{recision}} dR_{\text{ecall}}, \quad (8)$$

式中:  $P_{\text{recision}}$  为  $P$ - $R$  曲线的纵标准准确度;  $R_{\text{ecall}}$  为  $P$ - $R$  曲线上的横坐标召回率。对所有类别的平均精度求和, 并除以总的类别数量即可得到平均精度均值, 其计算公式为

$$x_{\text{mAP}} = \frac{1}{n} \sum_{j=0}^n x_{\text{AP},j}, \quad (9)$$

式中:  $n$  为训练样本集类别总数;  $j$  为当前类别的编号。

### 3.4 结果分析

为检验模型的有效性, 在 SIXray 数据集上进行了训练与测试, 并用 YOLOv5s 和 YOLOv5m<sup>[26]</sup> 网络进行了对比实验, 结果如表 1 所示, 表中:  $G$ 、 $K$ 、 $W$ 、 $P$ 、 $S$  分别为“gun”“knife”“wrench”“pliers”“scissors”。相比于基础网络, 所提模型的 mAP 分别提升了 2.4 个百分点与 3.0 个百分点。两种模型改进前后对所有违禁品类别检测的 AP 变化如图 5 所示, 可以看出, 改进后的两种模型 5 种类别的平均精度均有提高, 表明所提基于 YOLOv5 网络模型的改进策略是有效的。

表 1 两种模型改进前后的性能指标

Table 1 Comparison indicators of the two models before and after improvement

Method	mAP / %	G / %	K / %	W / %	P / %	S / %	Size / MB
YOLOv5s	87.2	98.1	79.2	84.8	90.3	83.8	14.04
Ours(5s)	89.6	98.4	82.2	88.4	92.2	86.7	15.72
YOLOv5m	89.5	98.2	83.1	87.6	94.2	84.5	41.46
Ours(5m)	92.5	99.1	85.6	90.5	95.2	92.0	45.18

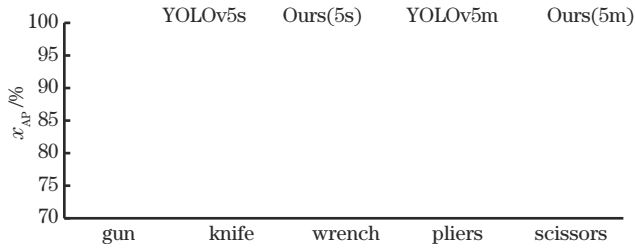


图 5 2 种模型改进前后在 SIXray 数据集的检测结果

Fig. 5 Detection results of the two models before and after improvement on the SIXray dataset

为验证所提模型的先进性,在 SIXray 数据集进行了性能测试,并与当前几种主流算法进行对比,结果如表 2 所示,在该数据集上,本文所提模型的 mAP 指标达到最高,为 89.6%。此外,为检验模型的泛化能力及应对复杂场景的能力,选取具有高度遮挡重叠违禁品的数据集 OPIXray 和接近现实场景的数据集 HiXray 与当前诸多先进算法进行对比实验, mAP 分别为 91.6% 和 83.1%,超过了大部分先进算法,结果如表 3 所示。进一步分析表 2 和表 3 可知,与 YOLOv3<sup>[29]</sup>、YOLOv4<sup>[30]</sup>、YOLOv5s<sup>[26]</sup> 经典模型相比,本文算法在 SIXray 数据集上的 mAP 分别有 10.4 个百分点、6.5 个百分点和 2.4 个百分点的提升;与 YOLO 系列的相关改进算法 YOLOv3+DOAM<sup>[20]</sup>、ASPP-YOLOv4<sup>[31]</sup>、YOLOv5s+DOAM<sup>[20]</sup> 相比,所提算法在准确性方面有一定的提高;此外,与 CHR<sup>[18]</sup>、FBS<sup>[19]</sup>、SSD<sup>[32]</sup>、FOCS<sup>[33]</sup> 等一些先进算法相比,本文算法表现也较为优异。综上,所提网络模型对于 X 光违禁品目标的识别更为精准,且具有良好的泛化性和稳健性。所提模型在 SIXray、OPIXray、HiXray 数据集上的混淆矩阵和对应 P-R 曲线如图 6 所示,可以看出,其对大部分类别的违禁品均具有较高识别准确率。图(b)中:FO 为折

表 2 SIXray 数据集结果对比

Table 2 Comparison results on the SIXray dataset

Method	mAP / %
YOLOv3 <sup>[29]</sup>	79.2
YOLOv4 <sup>[30]</sup>	83.1
ASPP-YOLOv4 <sup>[31]</sup>	85.2
SSD <sup>[32]</sup>	82.9
YOLOv5s <sup>[26]</sup>	87.2
YOLOv5s+Ours	89.6

表 3 OPIXray 与 HiXray 数据集结果对比

Table 3 Comparison results on the OPIXray and HiXray datasets

Method	OPIXray mAP / %	HiXray mAP / %
YOLOv3 <sup>[29]</sup>	78.2	-
YOLOv3+DOAM <sup>[20]</sup>	79.2	-
YOLOv4 <sup>[30]</sup>	78.9	-
CHR <sup>[18]</sup>	78.6	-
FBS <sup>[19]</sup>	81.7	-
SSD <sup>[32]</sup>	70.9	71.4
SSD+DOAM <sup>[20]</sup>	74.0	72.1
SSD+LIM <sup>[28]</sup>	74.6	73.1
FCOS <sup>[33]</sup>	82.0	75.7
FCOS+DOAM <sup>[20]</sup>	82.4	76.2
FCOS+LIM <sup>[28]</sup>	83.1	77.3
YOLOv5s <sup>[26]</sup>	87.8	81.7
YOLOv5s+DOAM <sup>[20]</sup>	88.0	82.2
YOLOv5s+LIM <sup>[28]</sup>	90.6	83.2
YOLOv5s+Ours	91.6	83.1

叠刀、ST 为直刀、SC 为剪刀、UT 为美工刀、MU 为多用刀;图(c)中:P1 为菱形锂电池、P2 为圆柱形锂电池、WA 为液体、LA 为笔记本电脑、MP 为手机、TA 为平板电脑、CO 为化妆品、NL 为非金属类打火机。

### 3.5 消融实验

为进一步探讨不同改进部分对网络性能的影响,分别以 YOLOv5s 和 YOLOv5m 为基线模型,在 SIXray 数据集的测试集上进行了消融实验,结果如表 4 所示。可见,采用 Mixup 数据增强方法和 WBF 算法均能提高基线模型的性能,并没有显著增加模型的大小,可归因于上述两种策略分别提升了网络对复杂目标的学习能力和精准预测能力。单纯引入注意力机制,基线模型性能提升不明显,而与 Mixup 数据增强方法和 WBF 算法融合能显著提高准确率,原因很可能是在推理阶段,物品因重叠产生的多个预测框被 NMS 算法剔除,导致网络的检测性能下降;但是采用 WBF 算法替代 NMS 算法后,预测阶段产生的目标框筛选错误率降低,对重叠预测框的筛选更精确,CBAM 使模型专注于有效特征学习,提升了性能。图 7 展示了采用 NMS 算法与 WBF 算法生成的预测框实例,可以看出,WBF 算法生成预测框的数量更准确。

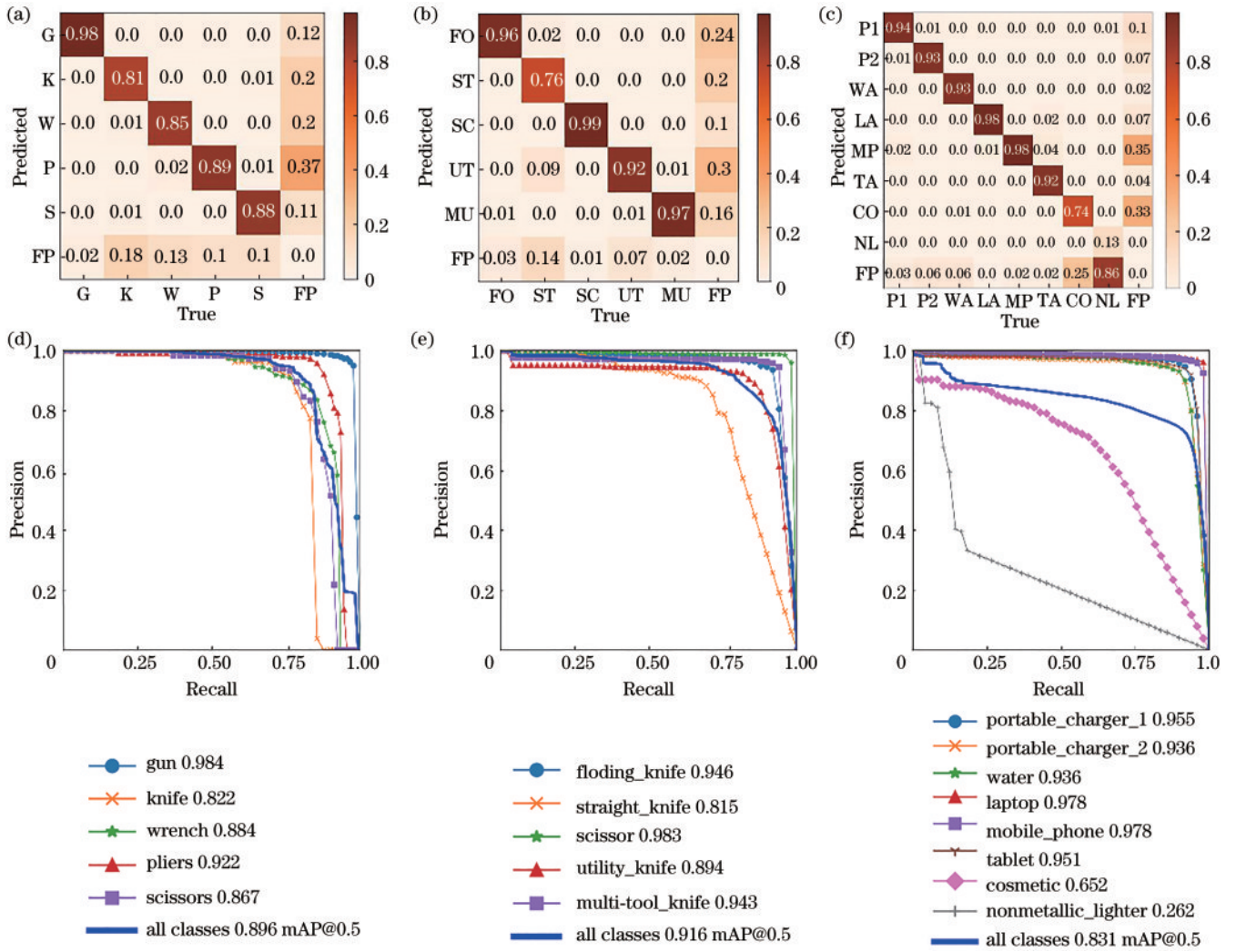


图 6 本文算法在 SIXray、OPIXray、HiXray 数据集的混淆矩阵与  $P$ - $R$  曲线,其中 FP 表示 background false positive。(a) SIXray 数据集混淆矩阵;(b) OPIXray 数据集混淆矩阵;(c) HiXray 数据集混淆矩阵;(d) SIXray 数据集  $P$ - $R$  曲线;(e) OPIXray 数据集  $P$ - $R$  曲线;(f) HiXray 数据集  $P$ - $R$  曲线

Fig. 6 Confusion matrix and  $P$ - $R$  curve of proposed method on SIXray, HiXray, OPIXray dataset. FP for background false positive. (a) Confusion matrix on SIXray; (b) confusion matrix on OPIXray; (c) confusion matrix on HiXray; (d)  $P$ - $R$  curves on SIXray; (e)  $P$ - $R$  curves on OPIXray; (f)  $P$ - $R$  curves on HiXray

表 4 SIXray 数据集上消融实验研究  
Table 4 Ablation study on the SIXray dataset

Component	YOLOv5m		YOLOv5s	
	mAP / %	Size / MB	mAP / %	SIZE / MB
Baseline	89.5	41.46	87.2	14.04
Baseline+CBAM	89.7	45.18	88.6	15.72
Baseline+Mixup	90.4	41.46	87.6	14.04
Baseline+Mixup+CBAM	90.5	45.18	88.0	15.72
Baseline+WBF	90.2	41.46	88.2	14.04
Baseline+WBF+CBAM	90.1	45.18	89.1	15.72
Baseline+WBF+Mixup	91.0	41.46	89.4	14.04
Baseline+WBF+Mixup+CBAM	92.5	45.18	89.6	15.72

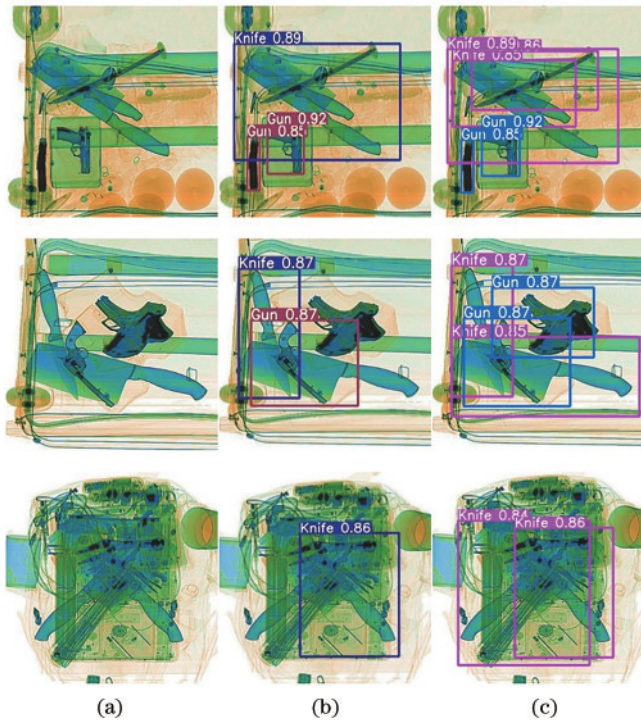


图7 预测边框实例,图片来自于SIXray数据集。(a)原始图像;  
(b)NMS算法生成的预测框;(c)WBF算法生成的预测框  
Fig. 7 Example of bounding boxes which are taken from SIXray  
dataset. (a) Original images ; (b) bounding boxes  
generated by NMS; (c) bounding boxes generated by WBF

## 4 结 论

以YOLOv5网络为基线模型,融合CBAM注意力机制、Mixup数据增强策略和WBF算法对网络模型进行优化。CBAM注意力机制可使网络更加专注于有效特征,抑制背景干扰;训练阶段采用Mixup数据增强策略,模拟重叠遮挡物品场景,提高模型复杂目标学习能力;测试阶段WBF算法能够提高预测边框的筛选准确度。在3个大型复杂数据集SIXray、HiXray和OPIXray的验证结果表明,改进后的YOLOv5网络模型优于大多数当前的X光违禁品图像检测先进算法,具有较好的准确性与稳健性。

## 参 考 文 献

- [1] Mery D, Saavedra D, Prasad M. X-ray baggage inspection with computer vision: a survey[J]. IEEE Access, 2020, 8: 145620-145633.
- [2] Zentai G. X-ray imaging for homeland security[J]. International Journal of Signal and Imaging Systems Engineering, 2010, 3(1): 13-20.
- [3] Akcay S, Breckon T. Towards automatic threat detection: a survey of advances of deep learning within X-ray security imaging[J]. Pattern Recognition, 2022, 122: 108245.
- [4] Michel S, Koller S M, de Ruiter J C, et al. Computer-based training increases efficiency in X-ray image interpretation by aviation security screeners[C]//2007

41st Annual IEEE International Carnahan Conference on Security Technology, October 8-11, 2007, Ottawa, ON, Canada. New York: IEEE Press, 2007: 201-206.

- [5] Mery D, Svec E, Arias M, et al. Modern computer vision techniques for X-ray testing in baggage inspection [J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2017, 47(4): 682-692.
- [6] 张友康, 苏志刚, 张海刚, 等. X光安检图像多尺度违禁品检测[J]. 信号处理, 2020, 36(7): 1096-1106. Zhang Y K, Su Z G, Zhang H G, et al. Multi-scale prohibited item detection in X-ray security image[J]. Journal of Signal Processing, 2020, 36(7): 1096-1106.
- [7] Baştan M, Yousefi M R, Breuel T M. Visual words on baggage X-ray images[M]//Real P, Diaz-Pernil D, Molina-Abril H, et al. Computer analysis of images and patterns. Lecture notes in computer science. Heidelberg: Springer, 2011: 360-368.
- [8] Turcsany D, Mouton A, Breckon T P. Improving feature-based object recognition for X-ray baggage security screening using primed visualwords[C]//2013 IEEE International Conference on Industrial Technology, February 25-28, 2013, Cape Town, South Africa. New York: IEEE Press, 2013: 1140-1145.
- [9] Akçay S, Kundegorski M E, Devereux M, et al. Transfer learning using convolutional neural networks for object classification within X-ray baggage security imagery [C]//2016 IEEE International Conference on Image Processing, September 25-28, 2016, Phoenix, AZ, USA. New York: IEEE Press, 2016: 1057-1061.
- [10] Akcay S, Kundegorski M E, Willcocks C G, et al. Using deep convolutional neural network architectures for object classification and detection within X-ray baggage security imagery[J]. IEEE Transactions on Information Forensics and Security, 2018, 13(9): 2203-2215.
- [11] Baştan M. Multi-view object detection in dual-energy X-ray images[J]. Machine Vision and Applications, 2015, 26(7): 1045-1060.
- [12] Akcay S, Breckon T P. An evaluation of region based object detection strategies within X-ray baggage security imagery[C]//2017 IEEE International Conference on Image Processing, September 17-20, 2017, Beijing, China. New York: IEEE Press, 2017: 1337-1341.
- [13] Liu J Y, Leng X X, Liu Y. Deep convolutional neural network based object detector for X-ray baggage security imagery[C]//2019 IEEE 31st International Conference on Tools with Artificial Intelligence, November 4-6, 2019, Portland, OR, USA. New York: IEEE Press, 2019: 1757-1761.
- [14] Mery D, Rizzo V, Zuccar I, et al. Object recognition in X-ray testing using an efficient search algorithm in multiple views[J]. Insight-Non-Destructive Testing and Condition Monitoring, 2017, 59(2): 85-92.
- [15] Mery D. Automated detection in complex objects using a tracking algorithm in multiple X-ray views[C]//CVPR 2011 WORKSHOPS, June 20-25, 2011, Colorado Springs, CO, USA. New York: IEEE Press, 2011: 41-48.

- [16] Bastan M, Byeon W, Breuel T. Object recognition in multi-view dual energy X-ray images[C]//Proceedings of the British Machine Vision Conference 2013, September 9-13, 2013, Bristol, UK. London: British Machine Vision Association, 2013.
- [17] Hassan T, Khan S H, Akcay S, et al. Cascaded structure tensor framework for robust identification of heavily occluded baggage items from multi-vendor X-ray scans[EB/OL]. (2019-12-09)[2021-10-31]. <https://arxiv.org/abs/1912.04251>.
- [18] Miao C J, Xie L X, Wan F, et al. SIXray: a large-scale security inspection X-ray benchmark for prohibited item discovery in overlapping images[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 2114-2123.
- [19] Shao F T, Liu J, Wu P, et al. Exploiting foreground and background separation for prohibited item detection in overlapping X-Ray images[J]. Pattern Recognition, 2022, 122: 108261.
- [20] Wei Y L, Tao R S, Wu Z J, et al. Occluded prohibited items detection: an X-ray security inspection benchmark and de-occlusion attention module[C]//MM '20: Proceedings of the 28th ACM International Conference on Multimedia, October 12-16, 2020, Seattle, WA, USA. New York: ACM Press, 2020: 138-146.
- [21] 郭瑞鸿, 张莉, 杨莹, 等. 基于改进 SSD 的 X 光图像管制刀具检测与识别[J]. 激光与光电子学进展, 2021, 58(4): 0404001.  
Guo R H, Zhang L, Yang Y, et al. X-ray image controlled knife detection and recognition based on improved SSD[J]. Laser & Optoelectronics Progress, 2021, 58(4): 0404001.
- [22] 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): 0810012.  
Wang Y X, Zhang L. Dangerous goods detection based on multi-scale feature fusion in security images[J]. Laser & Optoelectronics Progress, 2021, 58(8): 0810012.
- [23] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 3-19.
- [24] Zhang H Y, Cisse M, Dauphin Y N, et al. Mixup: beyond empirical risk minimization[EB/OL]. (2017-10-15)[2021-10-31]. <https://arxiv.org/abs/1710.09412>.
- [25] Solovyev R, Wang W M, Gabruseva T. Weighted boxes fusion: Ensembling boxes from different object detection models[J]. Image and Vision Computing, 2021, 107: 104117.
- [26] Glenn J, Alex S, Jirka B, et al. ultralytics/yolov5: v5.0-YOLOv5 -P6 1280 models, AWS, Supervise. ly and YouTube integrations[EB/OL]. (2020-05-18) [2021-10-31]. <https://github.com/ultralytics/yolov5>.
- [27] Neubeck A, Van Gool L. Efficient non-maximum suppression[C]//18th International Conference on Pattern Recognition (ICPR'06), August 20-24, 2006, Hong Kong, China. New York: IEEE Press, 2006: 850-855.
- [28] Tao R S, Wei Y L, Jiang X J, et al. Towards real-world X-ray security inspection: a high-quality benchmark and lateral inhibition module for prohibited items detection [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV), October 10-17, 2021, Montreal, QC, Canada. New York: IEEE Press, 2021: 10903-10912.
- [29] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08)[2021-10-31]. <https://arxiv.org/abs/1804.02767>.
- [30] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23) [2021-10-31]. <https://arxiv.org/abs/2004.10934>.
- [31] 吴海滨, 魏喜盈, 刘美红, 等. 结合空洞卷积和迁移学习改进 YOLOv4 的 X 光安检危险品检测[J]. 中国光学, 2021, 14(6): 1417-1425.  
Wu H B, Wei X Y, Liu M H, et al. Improved YOLOv4 for dangerous goods detection in X-ray inspection combined with atrous convolution and transfer learning [J]. Chinese Optics, 2021, 14(6): 1417-1425.
- [32] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 21-37.
- [33] Tian Z, Shen C H, Chen H, et al. FCOS: fully convolutional one-stage object detection[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 9626-9635.