

反向加权融合多尺度特征的 X 射线图像违禁品检测^{*}

马昌嵩¹ 裴晓芳^{1,2} 周磊¹ 周进¹ 杨继海¹

(1. 南京信息工程大学电子与信息工程学院 南京 210044; 2. 无锡学院电子信息工程学院 无锡 214105)

摘要:针对现阶段违禁品检测方法存在的混叠效应以及在类内变化显著的场景下检测精度较低等问题,提出一种反向加权融合多尺度特征的 X 射线图像违禁品检测算法,通过反向自适应地引导融合多尺度上下文特征来实现准确的违禁品目标检测。首先,使用多尺度场景感知模块获取从局部到全局的目标表征信息,帮助处理显著的类内变化。其次,利用反向加权融合结构采用特征引导加权的方式,高效融合蕴含丰富上下文特征的多级特征,缓解融合过程中易出现的混叠效应。最后,设计了一种 Focal-SIOU 损失函数,用于平衡不同质量违禁品目标预测框之间的贡献差异,并结合角度和边长损失进一步提升预测框的收敛速度和回归精度。本文方法在 SIXray、OPIXray、PIDray 等 3 个非常具有挑战性的基准数据集上进行了广泛的评测实验,平均精度均值(mAP)分别达到 93.2%、90.7%和 85.1%。实验结果充分表明,方法相比于最新方法性能更优,并且能够满足实时目标检测的实际应用需求。

关键词:违禁品检测;多尺度融合;上下文特征;损失函数

中图分类号: TP391.4 **文献标识码:** A **国家标准学科分类代码:** 520.20

Reverse weighted fusion of multi-scale features for prohibited object detection in X-ray images

Ma Changsong¹ Pei Xiaofang^{1,2} Zhou Lei¹ Zhou Jin¹ Yang Jihai¹

(1. School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China; 2. School of Electronic and Information Engineering, Wuxi University, Wuxi 214105, China)

Abstract: Aiming at the problems of aliasing effects and low detection accuracy in scenes with significant intra-class variations generally found in the existing prohibited object detection methods, this paper proposes a prohibited object detection algorithm for X-ray images with reverse weighted fusion of multi-scale features, so as to accurately detect the prohibited object by reverse adaptively guiding the fusion of multi-scale context features. First, a multi-scale scene perception module is used to obtain the object representation information from local and global, which helps to deal with significant intra-class variations. Second, by utilizing the reverse weighted fusion structure, the feature-guided weighting is employed to efficiently fuse multi-level features with rich context features, so as to alleviate the aliasing effects during the fusion process. Finally, a Focal-SIOU loss function is designed to balance the contribution differences between the predicted box of different quality for prohibited objects, and the convergence speed and regression accuracy of predicted box are further improved by combining the angle and side length losses. Extensive experiments were carried out on three very challenging benchmark datasets of SIXray, OPIXray, and PIDray by using the method proposed in this paper, and the mAP reached 93.2%, 90.7%, and 85.1% on the three datasets, respectively. The experimental results have fully demonstrated that our method is not only better than the state-of-the-art methods, but also can meet the practical application requirements of real-time object detection.

Keywords: prohibited object detection; multi-scale fusion; context feature; loss function

收稿日期:2023-10-18

^{*} 基金项目:国家自然科学基金青年基金(42205078)、高校哲学社会科学课题(2022SJYB0979)、苏高教会专项课题(2022JDKT138)、江苏职业教育研究课题(XHYBLX2023282)、无锡学院教改课题(JGZD202107)项目资助

0 引言

X射线安全检查具有实时成像和非侵入性等优点,是公共交通枢纽和边境检查站等常用的安全检查措施^[1]。近年来,应用基于计算机视觉的违禁品检测技术协助安全检查,已成为违禁品检测的主流方法。在违禁品检测中,能够准确、快速地定位、分类和识别出违禁品是一项极具挑战性的任务,其在安全检查、快递物流、仓储管理、医疗器械等领域具有广泛的应用价值。

目前,利用X射线安检仪确认包裹内是否藏有违禁品是最常用的安全检查方法。现代双能X射线安检设备激发X射线光束去穿透待检测物品,并接收穿透后的不同能量的X射线光子,得到原始的高能和低能X射线透视图像,再结合伪彩色映射算法给不同物质的等效原子序数赋予不同的颜色并显示为伪彩色图像^[2],最终由安检人员依据显示的颜色和轮廓信息进行检查。由于X射线图像的分析通常需要人工进行,这使得安全检查容易受到安检人员工作状态和经验的影响,从而导致漏检和误检等潜在的安全风险。针对这种情况,研究人员开发了多种技术手段,以辅助违禁品检测。然而,相较于自然场景下的图像,X射线图像有其特殊的特性,即物体纹理信息不足、高度重叠、背景杂乱、目标尺度变化大以及成像多变性等^[3]。因此,违禁品检测具有高度的复杂性和特殊性。长期以来,一直有大量致力于违禁品检测方法的研究。其中,早期方法主要是基于手工设计的特征,通过提取物体的形状、边缘、颜色等特征对违禁品进行分类。然而,这些特征区分违禁品与一般物品的能力十分有限,实际使用中泛化性较差,仅适用于一些特定的场景。

近年来,基于深度学习的X射线安检图像违禁品检测方法取得了突破性的进展。Shao等^[4]为解决X射线图像严重重叠的问题,提出基于前景与背景分离的检测框架;Miao等^[5]关注现有数据集的类不平衡问题,提出了一个CHR模型,该模型整合了多层次的视觉线索,并通过类平衡损失函数实现了类平衡;Wei等^[6]强调了违禁品的边缘信息和材料信息,设计了去遮挡注意力模块;Wu等^[7]将无预测检测器的概念引入违禁品检测任务,提出了基于尺度自适应中心的违禁品检测网络SA-CenterNet;Chang等^[8]统计分析了不同违禁品类别的物理尺寸分布,在训练阶段将物理尺寸约束形式表述为正则化项,设计了两阶段检测违禁品检测网络Xdnet。

虽然,上述方法从不同层面上极大提高了违禁品检测的性能,但仍然存在以下问题。

1)现有的违禁品检测方法只是简单地通过相加或拼接操作将主干网络产生的各级特征(高级语义信息与低级细节信息)结合起来,缺乏对不同特征贡献的考虑,容易产生混叠效应。

2)违禁品目标的类内变化显著也是导致检测性能下降的关键原因之一,因缺乏不同尺度的丰富上下文信息而

无法应对物体尺度、视角、风格的变化。

3)现有违禁品检测模型训练使用的损失函数无法应对不同质量预测框的贡献差异。

为了解决上述问题,本文提出一种反向加权融合多尺度特征的X射线图像违禁品检测算法(you only look once ray, YOLOray)。首先,在YOLOv5s的基础上重新设计了颈部特征融合结构,即反向加权特征融合(reverse weighted feature fusion, RWFF),RWFF包含了两个关键模块,即特征引导下采样模块(feature guided upsampling module, FGUM)和二阶自适应加权融合模块(two-stage adaptive weighted fusion module, TAWFM)。在模块层面上,FGUM通过构建浅层低级特征引导信息流,引导恢复高级特征精细的细节信息。TAWFM结合多尺度通道注意力思想,将局部通道上下文添加到全局通道上下文中,同时强调分布更全局的大物体和突出分布在局部的小物体。在结构层面上,RWFF改善了现有的全局双向融合路径以及等权重的传统拼接方式,弥合了不同层级特征之间的差距,更加高效地融合多尺度特征,有效避免混叠效应。其次,使用一个多尺度场景感知模块(multi-scale scene perception module, MSPM),MSPM通过一个可学习的全局上下文权重向量来调制多尺度上下文特征,感知复杂场景下丰富的上下文信息,以有效应对显著的类内变化。最后,设计了Focal-SIOU损失函数,该损失函数使用边长和角度作为惩罚项,并使用聚焦平衡函数改善高质量预测框在边界框优化过程中贡献较低问题,从而有利于目标边界框回归。

本文提出了一种新颖的X射线安检图像违禁品检测模型YOLOray,通过将MSPM、RWFF和Focal-SIOU损失函数三者相结合,准确、快速地定位和识别违禁品目标。

本文提出了FGUM和TAWFM,FGUM使用浅层低级特征引上来引导下采样过程,以恢复精细的细节信息;TAWFM聚集局部通道上下文和全局通道上下文,自适应地融合不同尺度的特征。

本文设计了一种适用于违禁品检测的Focal-SIOU损失函数,Focal-SIOU采用聚焦方式平衡不同质量预测框的贡献差异,并在回归计算中表征预测框的边长和角度损失来提高预测框的收敛速度和回归精度。

本文在3个X射线安检图像基准数据集上进行了广泛的实验,结果充分表明,所提出的YOLOray网络优于当前最先进的违禁品检测方法。

1 本文方法

1.1 概述

YOLOray网络结构如图1所示,该网络采用YOLOv5s作为基本框架。给定一个单一的RGB图像 $X \in \mathbf{R}^{3 \times 640 \times 640}$,将其送入CSPDarknet53骨干网络提取多级特征 $C_i, i = 1, \dots, 5$ 。然后,将最后3层特征分别送入3个MSPM,以捕获从局部到全局的多尺度上下文信息。

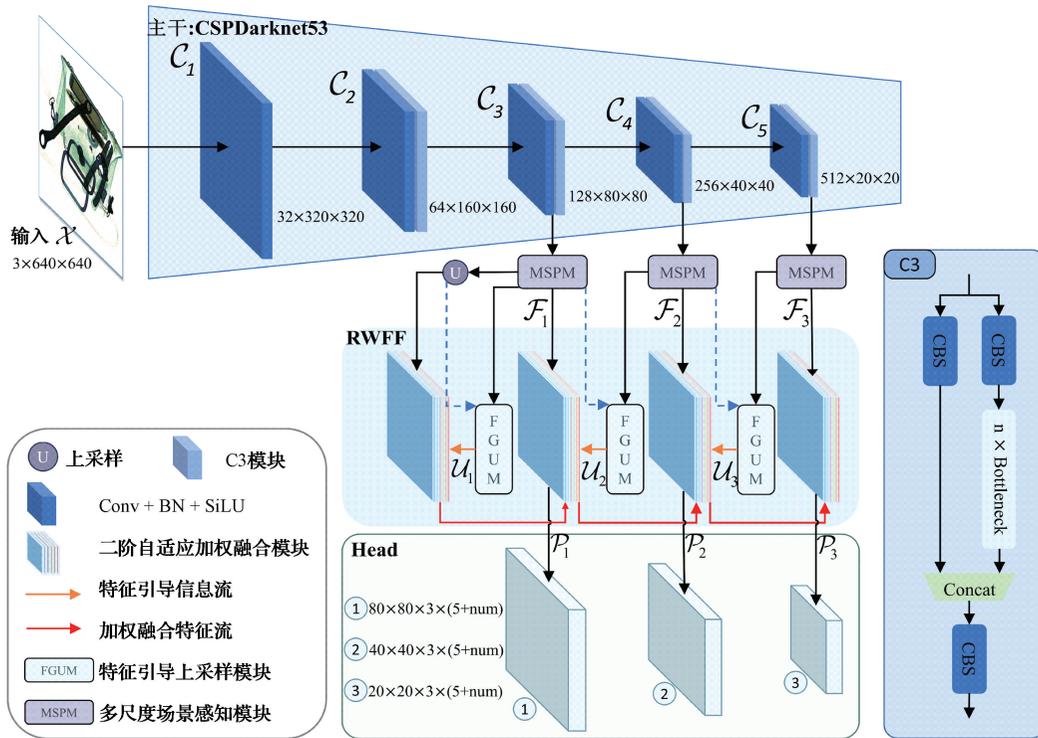


图1 YOLOray网络整体框架
Fig. 1 The overall framework of YOLOray network

接着,利用RWFF通过特征引导加权的方式,高效地整合多级特征,以提升不同尺度特征的代表能力。最后,由头部特征图 $P_i, i = 1, 2, 3$ 预测违禁品目标类别和位置,通过Focal-SIOU损失函数减轻不同质量预测框在边界框优化过程中的贡献不平衡问题,提高违禁品目标预测框的收敛速度和回归精度。

1.2 多尺度场景感知模块

X射线安检图像背景复杂,目标重叠严重。而且违禁品会以许多不同的尺度、视角、风格甚至子类型出现。这

些因素都会导致相当大的类内变化,进而增加识别的难度。为了解决这些问题,需要拓宽网络的感受野,以捕获不同尺度的上下文信息,从而聚集更多的鉴别性特征来帮助处理复杂的违禁品目标类内变化。为此,本文引入Liu等^[9]提出的MSPM, MSPM结构如图2所示。MSPM利用一个可学习的全局上下文权重向量来调制由具有不同局部感受野的池化操作生成的多尺度上下文特征。

MSPM包括3个关键部分,全局自适应感知、局部多尺度感知和全局到局部调制。输入特征 $C_i \in \mathbf{R}^{C \times H \times W}$ 经过降维后分别送入3个关键部分, C, H 和 W 分别代表通道数、

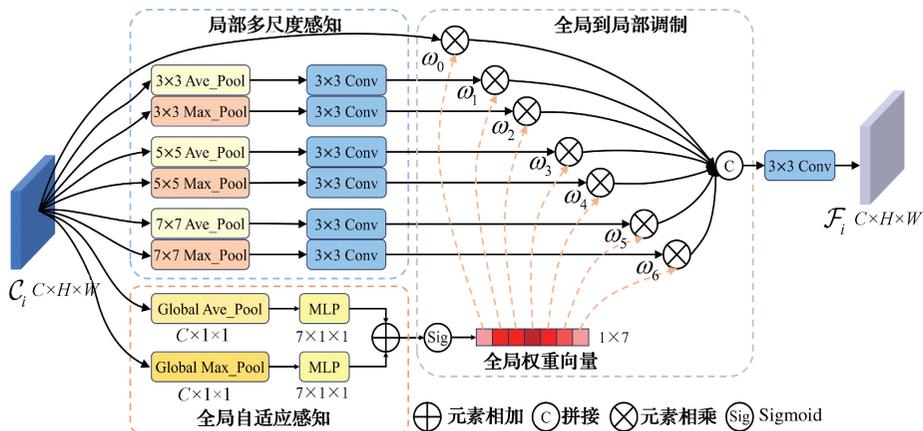


图2 MSPM结构
Fig. 2 Structure diagram of the MSPM

高度和宽度。其中,全局自适应感知部分通过全局平均池化操作和全局最大池化操作捕捉全局场景上下文信息,并获得全局权重向量。局部多尺度感知部分使用具有不同感受野的滤波器来获得丰富的局部多尺度上下文信息,其中感受野大小被设定为 $P \in \{3 \times 3, 5 \times 5, 7 \times 7\}$ 。全局到局部调制部分用于对局部多尺度特征进行调制,增强网络对局部特征变化的适应能力,使其能够应对类内变化等情况。

1.3 反向多尺度特征融合

现有的违禁品检测方法^[3,7-8,10-12]主要使用特征金字塔网络(feature pyramid networks, FPN)进行特征融合,以提升违禁品检测的精度和鲁棒性。然而,根据研究发现,使用 FPN 进行特征融合存在一个普遍的局限性,即跨尺度融合过程中因缺少合理的筛选或加权处理导致的混叠效应。虽然现有的方法,如 PAFPN^[13]、NAS-FPN^[14]、BiFPN^[15]等都可以在一定程度上缓解此问题,但这些方法会引入更长的融合路径,无法在精度和速度上取得平衡。针对上述问题,本文提出反向多尺度特征融合。

RWFF 由 FGUM 和 TAWFM 两部分组成。如图 1 所示,各层级间的 FGUM 和 TAWFM 组成了多级局部反向连接,多级局部反向连接可以模拟全局双向信息流的效果^[16],并且推理速度相比 PAFPN、BiFPN、NAS-FPN 等显著提高,多级局部反向连接以最小地效率损失双向集成特征并获得更优的性能。

1) 特征引导上采样模块

不同级别的特征通常具有不同的特点,高层特征通常嵌入了更多的语义知识,而浅层特征则具有更大的空间尺

寸,蕴含更丰富的细节信息。为了弥补上采样过程中的信息损失,受 Zong 等^[16]提出的局部反向连接启发,本文设计了一种轻量且高效的上采样模块,称为特征引导上采样模块。FGUM 结构如图 3 所示,使用浅层低级别特征来引导上采样过程,以恢复精细的细节信息。给定相对而言的高级特征 F_h 和低级特征 F_l ,首先,通过亚像素卷积将 F_h 上采样到和 F_l 一样的大小。其次,对上采样后的高级特征和原始低级特征应用拼接操作,以丰富特征表示。拼接后的特征图分别沿通道方向进行最大池化和平均池化操作,生成一组二维特征映射 $F_{max} \in \mathbf{R}^{1 \times H \times W}$ 和 $F_{ave} \in \mathbf{R}^{1 \times H \times W}$:

$$\begin{cases} F_{max} = \text{Max}(\text{Cat}[\text{SPC}_{r=2}(F_h); F_l]) \\ F_{ave} = \text{Ave}(\text{Cat}[\text{SPC}_{r=2}(F_h); F_l]) \end{cases} \quad (1)$$

式中: $\text{Max}(\cdot)$ 和 $\text{Ave}(\cdot)$ 分别为最大池化和平均池化操作; $\text{SPC}_{r=2}(\cdot)$ 表示放大系数为 2 的亚像素卷积上采样操作,通过亚像素卷积可以有效缓解因通道降维导致的信息损失。接着,再将特征映射沿着通道维度拼接,并使用一个 3×3 卷积整合拼接后的特征得到特征图 F' :

$$F' = \text{Conv}_{3 \times 3}(\text{Cat}[F_{avg}; F_{max}]) \quad (2)$$

然后,将卷积后的结果经过归一化和激活操作得到空间权重。最后,将空间权重和经过亚像素卷积输出的高级特征相乘,使较粗糙的高级特征受低级特征引导而恢复精细的细节信息,最终得到引导特征 U_i :

$$U_i = \text{Sig}(\delta(\mathbf{B}(\text{Conv}_{3 \times 3}(F')))) \otimes \text{SPC}_{r=2}(F_h) \quad (3)$$

式中: $\mathbf{B}(\cdot)$ 为批归一化; $\delta(\cdot)$ 为 ReLU 函数; $\text{Sig}(\cdot)$ 为 Sigmoid 函数。

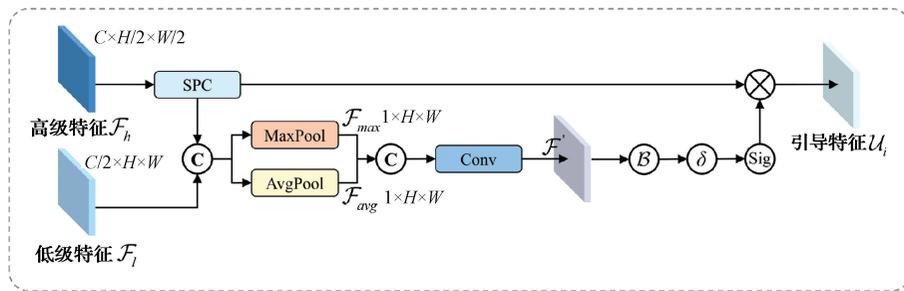


图 3 FGUM 结构

Fig. 3 Structure diagram of the FGUM

2) 二阶自适应加权融合模块

为对特征图进行合理的筛选和加权,避免混叠效应,一些研究^[17-18]尝试将 SENet^[19]和 ECANet^[20]引入 FPN。然而,根据研究发现,全局通道注意力机制(SENNet、ECA-Net)虽然考虑了特征融合中不同通道的贡献,在一定程度上增强了网络的表征能力,但其并不能很好地处理多尺度特征融合。受 Dai 等^[21]提出的多尺度通道注意力模块启发,并结合违禁品成像的特性,本文设计了 TAWFM。TAWFM 是一种根据输入特征调整融合阶段,并通过网络训练自适应地调节融合权重的特征加权融合模块。

TAWFM 将局部通道上下文添加到全局通道上下文中,从局部和全局两个角度自动学习到不同通道特征的重要程度,有助于网络根据不同通道的贡献度来增强对不同特征的利用和感知能力。

TAWFM 的结构如图 4 所示,本文以三输入融合为例,两输入融合参照一阶段自适应加权融合过程。给定经过 FGUM 处理的高级特征 $U_h \in \mathbf{R}^{C \times H \times W}$, 经过 MSPM 输出的低级特征 $F_l \in \mathbf{R}^{C \times H \times W}$, 以及前一阶段融合后的输出特征 $P_l \in \mathbf{R}^{C \times 2H \times 2W}$ 。首先,通过逐元素相加操作将两种特征 (U_h, F_l) 聚合,并使用一个 3×3 的卷积核进行卷积得

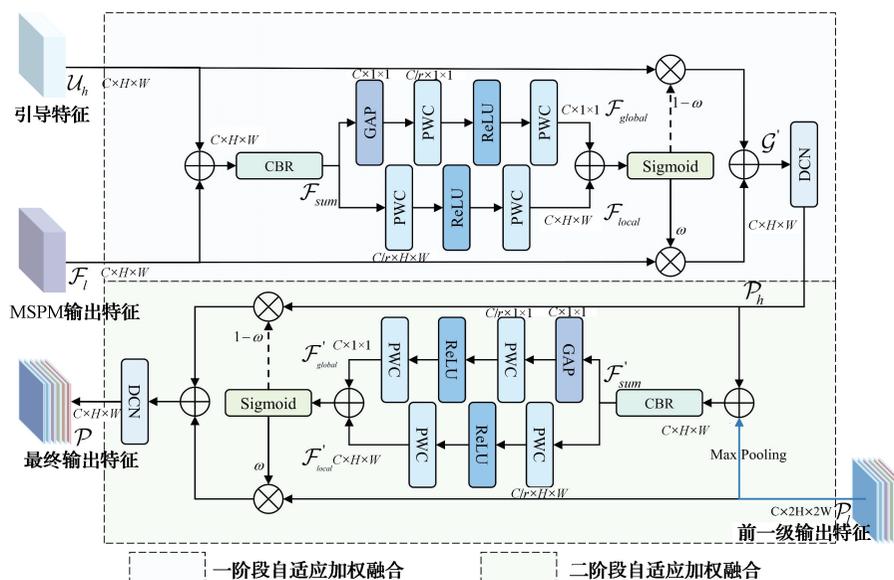


图4 TAWFM结构

Fig. 4 Structure diagram of the TAWFM

到初步细化特征 F_{sum} :

$$F_{sum} = \text{CBR}(U_h + F_l) \quad (4)$$

式中: $\text{CBR}(\cdot)$ 指卷积、批归一化和 ReLU 激活函数的组合操作。其次,聚合后的特征分别通过两个卷积分支,即局部卷积分支和全局卷积分支。局部卷积分支负责计算局部通道上下文特征,全局卷积分支负责计算全局通道上下文特征。对于局部卷积分支而言,特征 F_{sum} 通过一个由逐点卷积组成的瓶颈结构,得到上下文特征 F_{local} ,代表了每个空间位置上不同通道的重要程度。此外, *Bottleneck* 结构已被证明可以有效减少参数量^[22]。公式如下:

$$F_{local} = \mathbf{B}(\text{PWC}_2(\delta(\mathbf{B}(\text{PWC}_1(F_{sum})))))) \quad (5)$$

式中: $\text{PWC}_1(\cdot)$ 和 $\text{PWC}_2(\cdot)$ 指逐点卷积操作,卷积核参数分别为 $C \times C/r \times 1 \times 1$ 和 $C/r \times C \times 1 \times 1$, 本文将 r 设置为 4。 $F_{local} \in \mathbf{R}^{C \times H \times W}$ 的形状与输入特征相同,可以保留并突出低级特征中的细微细节。

对于全局卷积分支而言,特征 F_{sum} 首先通过全局平均池化得到特征层的通道向量 $X_i \in \mathbf{R}^{C \times 1 \times 1}$, 用于获取特征的全局表示,接着再通过瓶颈结构建模通道间的相关性。此外,使用 ReLU 函数可以使网络具有更多的非线性,更好地拟合通道间复杂的相关性。公式如下:

$$F_{global} = \mathbf{B}(\text{PWC}_2(\delta(\mathbf{B}(\text{PWC}_1(\text{GAP}(F_{sum})))))) \quad (6)$$

式中: $\text{GAP}(\cdot)$ 指全局平均池化操作。然后,通过广播式加法运算将两个上下文特征 F_{local} 和 F_{global} 进行融合,再由 Sigmoid 非线性函数将数值映射到 0~1 的范围内,得到基于全局和局部特征的通道权重 ω 。公式如下:

$$\omega = \text{Sig}(F_{local} \oplus F_{global}) \quad (7)$$

式中: \oplus 表示广播式加法操作。

然后,将原始特征 U_h 和 F_l 乘以逐通道权重来获得加

权特征图,完成在通道维度上对原始特征的重标定,并通过逐元素加法将两个特征图相加,以获得特征图 G' 。此外,将特征图通过可变形卷积层,进一步建模不同尺度的几何变换。公式如下:

$$P_h = \text{DCN}(((1-\omega) \otimes U_h) + (\omega \otimes F_l)) \quad (8)$$

式中: $\text{DCN}(\cdot)$ 指 ReLU 函数、可变形卷积 DCNv2 以及批归一化的组合操作,本文将 U_h 的权重设置为 $1-\omega$,这使得网络能够在两种不同的特征之间进行软选择或加权平均。

最后,将经过前一阶段融合后的输出特征 P_1 与一阶段输出特征 P_h 进行二阶段融合,依然使用上述方式构建,公式如下:

$$P = \varphi(P_h, \text{Max}_{2 \times 2}(P_1)) \quad (9)$$

式中: $\text{Max}_{2 \times 2}(\cdot)$ 指池化核大小为 2×2 的最大池化操作; $\varphi(\cdot)$ 表示二阶段融合操作。

1.4 Focal-SIOU 损失函数

YOLOv5 算法模型中采用 CIUO 作为边界框回归损失函数^[23]。CIUO 考虑了预测框 B 和真实框 B^{gt} 之间的重叠面积和中心距离,并将长宽比损失引入损失函数中,以提高回归精度。

然而,CIUO 中的长宽比只能反映预测框和真实框之间模糊的差异。例如,当预测框与真实框的宽高比呈线性关系时,宽高比惩罚项就失去了作用。因此,CIUO 损失可能会以不合理的方式优化相似度,从而影响预测框的回归。此外,在训练阶段,回归误差小的高质量预测框的数量远少于低质量预测框的数量,因此,低质量预测框对于边界框优化起了最大作用,然而,质量较差的预测框会产生过大的梯度影响训练过程,这将导致网络更加关注低质

量预测框的回归而忽略高质量预测框的回归。为了克服上述问题,受 Zhang 等^[24]和 Zhora 等^[25]工作的启发,本文为违禁品检测设计了 Focal-SIOU 损失函数,它将 CIOU 的长宽比损失项拆分成单独的宽度损失项 L_w 和高度损失项 L_H ,并额外增加了角度损失 L_{AandD} ,加速预测框的收敛,提高预测框的回归精度。各损失项如下所示:

$$\begin{cases} L_w = \frac{\rho^2(\omega, \omega^{gt})}{C_w^2} \\ L_H = \frac{\rho^2(h, h^{gt})}{C_h^2} \\ \Delta = \cos\left(2\left(\arcsin\left(\frac{P_h}{\sigma}\right) - \frac{\pi}{4}\right)\right) \\ L_{\text{AandD}} = \sum_{t=x,y} (1 - e^{-\tau \rho_t}) \\ \rho_x = \left(\frac{b_x^{gt} - b_x}{C_w}\right)^2, \rho_y = \left(\frac{b_y^{gt} - b_y}{C_h}\right)^2, \tau = 2 - \Delta \end{cases} \quad (10)$$

式中: ω 和 ω^{gt} 分别表示 B 和 B^{gt} 的宽度; h 和 h^{gt} 分别表示 B 和 B^{gt} 的高度; P_h 为 B 和 B^{gt} 中心点的高度差; σ 为 B 和 B^{gt} 中心点的距离; C_w 和 C_h 分别为包含 B 和 B^{gt} 的最小外接矩形的宽度和长度; (b_x^{gt}, b_y^{gt}) 为 B^{gt} 的中心坐标; (b_x, b_y) 为 B 的中心坐标。此外,在所有损失项之前添加一个聚焦平衡函数 F_{focal} ,用于控制对低质量预测框的抑制程度。最后, Focal-SIOU 损失函数如下所示:

$$\begin{cases} F_{\text{focal}} = IoU^\gamma \\ L_{\text{Focal-SIoU}} = F_{\text{focal}} \times (1 - IoU + L_{\text{AandD}} + L_w + L_H) \end{cases} \quad (11)$$

式中: γ 是一个超参数,本文设置为 0.5。

2 实验设置与结果分析

2.1 数据集

本文在 3 个公开数据集上对算法模型进行评估。SIXray^[5]是目前使用频率最高的违禁品检测数据集,包含 8 929 张 X 射线图像,数据集标注了 6 类常见违禁品,包括枪、刀、扳手、钳子、剪刀和锤子,训练集中包含 7 144 张图像,测试集中包含 1 785 张图像。OPIXray^[6]是专为安全检查中的遮挡违禁品检测而设计的数据集,包含 8 885 张 X 射线图像,数据集标注了 5 类日常生活中随处可见的刀具违禁品,训练集中包含了 7 109 张图像,测试集中包含 1 776 张图像。PIDray^[26]是在不同场景下收集的数据集,包含 47 677 张 X 射线图像,涵盖机场、地铁站和火车站,并使用 3 台来自不同制造商的安检仪来收集 X 射线数据。数据集中定义了 12 个违禁品类别,训练集中包含 29 457 张图像,测试集中包含 18 220 张图像。

2.2 评估方法

本文使用精确率 (precision, P)、召回率 (recall, R)、平均精度均值 (mean average precision, mAP)、参数量 (parameters, Params) 和浮点运算量等指标评估违禁品检测

算法模型的性能。其中, P 是衡量模型预测准确性的指标,表示某一类别预测目标中正确预测样本数占总预测样本数的比例。 R 是衡量模型查全能力的指标,表示在所有违禁品目标中,模型正确检测出违禁品的比例。 mAP 用以衡量所有类别的检测精度,由所有类别的平均精度 (average precision, AP) 求均值得到。 Params 为模型中的权重和偏置总参数量,浮点运算量为模型计算量指标,用以衡量模型计算复杂度。

$$P = \frac{T_p}{T_p + F_p} \quad (12)$$

$$R = \frac{T_p}{T_p + F_N} \quad (13)$$

$$AP = \int_0^1 P(R) dR \quad (14)$$

$$mAP = \frac{\sum_{i=1}^N AP_i}{N} \quad (15)$$

式中: T_p 指预测为违禁品实际也为违禁品的数量; F_p 指预测为违禁品但实际被误检的数量; F_N 代表违禁品未被检测出的数量; N 为检测的违禁品目标类别数。

2.3 实现细节

本文所有实验都基于 Pytorch 深度学习框架,实验使用的计算机 CPU 型号为 Intel(R) Core I7-13700KF, GPU 型号为 NVIDIA GeForce RTX 4070Ti。为了保证实验中算法模型对比的公平性,实验使用相同的超参数对不同算法模型进行训练和测试。在超参数中,初始学习率决定了模型在每次参数更新时的步长大小,动量参数用于控制梯度下降过程中的加速度,衰减系数用于控制模型的正则化,旨在减少模型的复杂性。本文根据大量实验所得经验,在训练阶段,所有模型均使用 SGD 优化器进行参数迭代更新,初始学习率为 1×10^{-3} ,动量参数为 0.937,衰减系数设置为 5×10^{-4} ,输入图像的尺寸设置为 640×640 ,批量大小为 20,训练周期设置为 300,此外,为使数据内容更加丰富,训练时采用了 Mosaic 数据增强方法。在测试阶段,本文不使用任何预处理手段,只是将输入图像的尺寸设置为 640×640 ,然后送入模型进行评估。

2.4 对比实验

为了全面评估本文模型性能,在上述参数配置下,将其与当前 4 种最先进的违禁品检测方法进行对比实验,涉及的方法包括 DOAM^[6]、EAOD-Net^[11]、POD^[27]、Light-Ray^[12]。其中,EAOD-Net 的实验结果收集自原始论文,其他 3 种违禁品检测方法的实验结果由本文根据公开的代码重新训练和测试得出。此外,本文还将提出的方法与现有的 3 种先进的常规目标检测方法进行对比实验,涉及的方法包括 Faster R-CNN^[28]、DetectoRS^[29]、YOLOv7^[30]。

1) 定量比较

利用 P 、 R 、mAP 对违禁品检测算法进行定量比较,定量比较结果如表 1 所示。其中,最佳结果用粗体表示,

表1 不同方法在 SIXray、OPIXray、PIDray 三个数据集上的定量比较

Table 1 Quantitative comparison of different methods on the SIXray, OPIXray and PIDray datasets

方法	年份	主干	SIXray			OPIXray			PIDray		
			P/%	R/%	mAP/%	P/%	R/%	mAP/%	P/%	R/%	mAP/%
Faster R-CNN	2016	ResNet-50	86.78	73.85	82.8	84.31	77.32	81.9	79.34	70.87	73.3
DetectoRS	2021	ResNet-50	86.18	74.02	83.1	82.03	78.43	82.0	81.33	70.59	76.5
Yolov7-tiny	2023	CSPDarknet-53	89.79	83.33	89.1	83.24	80.67	84.6	86.22	76.06	83.7
DOAM	2020	VGG16	78.42	71.20	77.9	80.79	70.90	74.0	77.28	64.76	70.3
EAOD-Net	2022	ResNeXt-50	87.08	77.39	85.9	86.09	80.14	85.8	—	—	—
LightRay	2022	MobileNetV3	93.58	76.30	87.3	86.91	69.87	78.4	79.50	66.52	71.4
POD	2023	ResNet-50	88.35	77.21	86.1	85.16	80.77	84.9	81.05	74.11	80.7
本文		CSPDarknet-53	94.76	87.28	93.2	91.92	87.65	90.7	88.74	78.91	85.1

次优结果用下划线表示。总的来看,本文提出的YOLOray在3个数据集上的性能指标均优于其他检测算法。

不同算法在SIXray数据集上的结果如表1所示,YOLOray算法各项性能指标均为最优, P 达到94.76%, R 达到87.28%, mAP 达到93.2%,相较于各项指标的次优结果,YOLOray算法分别提升了1.18%、3.95%和4.1%。此外,表2为不同算法在该数据集上不同类别违禁品的定量指标比较结果。可以看出,YOLOray算法在刀、扳手、钳子和剪刀4个类别上均取得了最佳的 mAP ,虽然在枪这一类别上未取得最佳结果,但与最优结果相比仅差0.1%。

表2 不同方法在 SIXray 数据集上 5 个类别的定量比较

Table 2 Quantitative comparison of different methods for five categories on the SIXray dataset

方法	mAP/%	枪	刀	扳手	钳子	剪刀
Faster R-CNN	82.8	88.3	78.8	80.1	80.3	86.3
DetectoRS	83.1	81.6	80.5	84.5	87.3	81.4
Yolov7-tiny	89.1	98.7	86.0	84.1	91.1	85.9
DOAM	77.9	81.4	75.9	76.7	79.5	76.0
EAOD-Net	85.9	89.3	81.3	82.3	88.4	88.3
LightRay	87.3	97.2	84.6	80.2	89.5	84.9
POD	86.1	88.9	82.5	83.8	87.6	87.7
本文	93.2	98.6	89.7	89.9	95.0	92.6

表1为不同算法在OPIXray数据集上的性能。其中,YOLOray算法取得了最优性能,相较于次优结果, P 提高了5.01%, R 提高了6.88%, mAP 提高了4.9%。此外,表3为不同算法在OPIXray数据集上不同类别违禁品的定量指标比较结果。可以看出,YOLOray算法在4个类别上达到最好的检测性能。在3个遮挡最严重的类别上,即直刀、折叠刀和美工刀,YOLOray算法明显优于其他7个现有的方法。结果表明,YOLOray算法在严重遮挡的数据集上仍然有效。

由表1可知,YOLOray算法在 mAP 指标方面相比于

表3 不同方法在 OPIXray 数据集上 5 个类别的定量比较

Table 3 Quantitative comparison of different methods for five categories on the OPIXray dataset

方法	mAP /%	直刀	折叠刀	剪刀	美工刀	多功能刀
Faster R-CNN	81.9	67.1	88.6	89.9	74.8	89.1
DetectoRS	82.0	64.0	88.5	89.9	81.0	86.6
Yolov7-tiny	84.6	59.1	92.5	98.7	79.8	93.1
DOAM	74.0	41.5	81.4	95.1	68.2	83.8
EAOD-Net	85.8	76.1	89.6	90.7	83.2	89.2
LightRay	78.4	56.2	79.6	96.6	70.7	88.9
POD	84.9	76.0	88.7	88.9	82.8	88.1
本文	90.7	79.3	94.2	98.0	86.5	95.3

其他7种检测方法有了明显的提高,改善幅度从1.4%~14.8%不等。此外,YOLOray算法在 P 和 R 指标方面也有显著的优势。然而,YOLOray算法在该数据集上的改进并不像在SIXray和OPIXray数据集上那样突出,这可能是由于PIDray数据集中的图像来自不同的机器采集,其中存在着人类无法感知的内生偏移^[31],这种偏移会严重影响检测的性能。

除了使用上述评价指标进行总体定量比较外,不同数据集上的精确率-召回率曲线如图5所示。曲线与 X 轴和 Y 轴所围成的面积记为平均精度,面积越大代表算法性能越好。从结果来看,与其他最先进的检测方法相比,YOLOray算法在3个数据集上进行X射线安检图像违禁品检测时拥有更大优势。

2) 定性比较

YOLOray算法与其他算法的定性比较结果如图6所示。根据视觉结果可以得出,YOLOray算法在处理违禁品检测的不同挑战因素方面可以取得更好的结果。

图6(a)比较了检测算法在小目标违禁品上的检测性能。通过对比观察发现,只有POD、YOLOv7-tiny和YOLOray算法检测到尺寸较小的目标,其中YOLOray

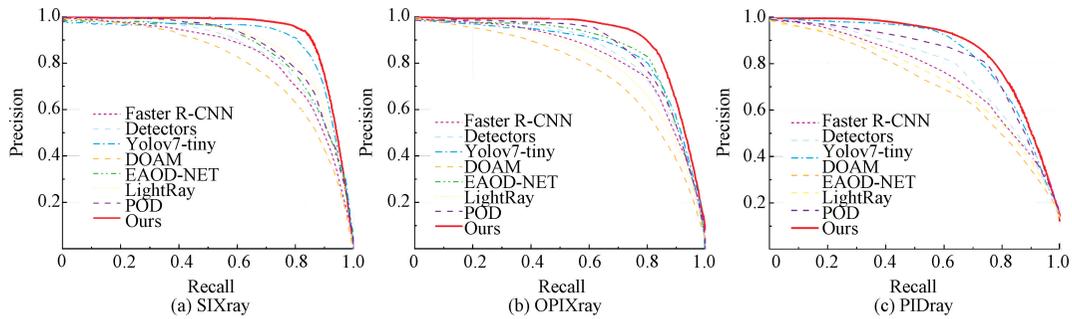


图 5 不同方法在 3 个数据集上的 P-R 曲线

Fig. 5 P-R curves of different methods on the three datasets



图 6 不同方法的定性比较结果

Fig. 6 Qualitative comparison results of different methods

算法的预测置信度最高;图 6(b)比较了检测算法在违禁品被严重遮挡情况下的检测性能。由于存在一些遮挡干扰,导致目标局部细节信息丢失,这使得部分算法发生漏检或是误检。然而,YOLOray 算法通过高效融合多尺度

上下文特征强化网络的表征能力,从而能够准确检测出违禁品目标,并取得最高的置信度;图 6(c)比较了检测算法在类内变化显著情况下的检测性能。其中大部分算法的检测效果欠佳,漏检情况较多,相比之下,YOLOray 算法在该情况下表现更优。

以上定量与定性比较实验,充分展现了本文算法的有效性和鲁棒性。这主要受益于 MSPM 能够有效地感知局部与全局上下文信息,从而有能力帮助处理违禁品类内变化。并且,RWFF 弥合了不同层级特征之间的差距,更加高效地融合了多尺度特征,从而提高违禁品检测器的性能。此外,Focal-SIOU 损失函数增强了对高质量预测框的关注,并引入了角度和边长损失,使得违禁品目标的预测框更加接近真实框。

3)模型参数数量和计算复杂度比较

为了研究模型参数数量和计算复杂度,本文对 8 种模型的 Params 和浮点量进行了对比,结果如表 4 所示。在 5 个专用于违禁品检测的方法中,YOLOray 算法的浮点量虽然略微高于 LightRay,但在 Params 指标上取得了最佳结果。这是由于 MSPM 在获取从局部到全局的目标特征信息过程中,输入通道首先被降维处理,然后采用了简单的池化、卷积、元素相乘等操作,并不涉及复杂的计算。同

表 4 模型参数数量和计算复杂度的比较

Table 4 Comparison of model parameters and computational complexity

方法	输入图 像尺寸	Params /($\times 10^6$)	浮点量 /GFLOPs
Faster R-CNN	1 333 \times 800	41.09	118.8
DetectoRS	1 333 \times 800	93.47	228.0
Yolov7-tiny	640 \times 640	6.23	13.9
DOAM	300 \times 300	24.30	32.8
EAOD-NET	1 333 \times 800	119.67	338.9
LightRay	640 \times 640	18.96	14.1
POD	1 333 \times 800	118.23	334.1
本文	640 \times 640	6.66	15.3

时,相比其他违禁品检测方法中使用的类 FPN 结构,RWFF 通过多级局部反向连接模拟全局双向信息流的效果,显著减少模型参数量和计算量。此外,YOLOray 算法与常规目标检测方法相比,本文模型的 Params 和浮点量也远低于其中的 Faster R-CNN 和 DetectoRS,仅比目前最轻量的 YOLOv7-tiny 略微高一些,但本文算法在精度指标上比 YOLOv7-tiny 表现更优。

2.5 消融实验

为检验 YOLOray 算法模型中每个关键组件的有效性,本文采用以 CSPDarknet-53 为主干的 YOLOv5s 作为本文的基线算法,并在 3 个安检数据集 SIXray、OPIXray、OPIXray 上进行了全面的消融实验。表 5 为本文模型在

3 个数据集上的消融结果。其中“√”表示添加相应的模块,加粗字体表示指标结果最优。

在本文模型中,MSPM 用于捕获主干网提取的特征中充足的场景上下文信息。从表 5 可以看出,增加 MSPM 后的模型各项指标均有显著提升,例如在 3 个数据集上 mAP 分别增加了 0.7%、1.2% 和 0.8%。这得益于 MSPM 通过多组池化操作获取多尺度上下文特征,并利用一个可学习的全局上下文权重向量来进行调制,通过 MSPM 可以弥补主干网络丢失的细节信息,挖掘出更多的鉴别性特征,使得模型能够应对安检数据集中严峻的类内变化,从而提高模型在安检数据集上的检测能力。

表 5 本文方法在 SIXray、OPIXray、PIDray 三个数据集上的消融实验

Table 5 Ablation experiments of the proposed method on the SIXray, OPIXray and PIDray datasets

MSPM	RWFF	Focal-SIOU	SIXray			OPIXray			PIDray			浮点数 /GFLOPs /($\times 10^6$)	Params /ms	Time
			P/%	R/%	mAP /%	P/%	R/%	mAP /%	P/%	R/%	mAP /%			
			93.32	85.76	91.6	88.93	85.22	87.9	86.63	76.69	83.1	16.0	7.03	2.5
√			94.18	85.72	92.3	91.31	85.76	89.1	87.01	77.29	83.9	17.3	7.45	2.8
	√		94.11	86.16	92.4	91.75	86.44	90.1	87.73	77.93	84.3	14.2	6.25	3.1
		√	93.89	86.13	92.2	89.54	86.81	88.2	86.97	76.86	83.6	16.0	7.03	2.5
√	√		94.26	87.25	92.8	91.38	87.47	90.5	88.38	78.59	84.7	15.3	6.66	3.5
		√	94.43	86.23	92.6	91.11	88.18	90.3	87.51	78.45	84.5	14.2	6.25	3.1
√		√	94.29	86.07	92.5	91.68	85.44	89.2	87.56	77.85	84.1	17.3	7.45	2.8
√	√	√	94.76	87.82	93.2	91.92	88.20	90.7	88.74	78.91	85.1	15.3	6.66	3.5

RWFF 用于更加高效地融合多尺度特征,旨在增强多个尺度上的语义表达和定位能力。通过表 5 消融实验比较可以得出,使用 RWFF 可以提高违禁品检测的性能,并显著降低模型的计算量和参数量。这得益于 RWFF 改善了现有的全局双向融合路径以及等权重的传统拼接方式,更加高效地融合了不同层级的特征,丰富了特征图的表现能力。

将图 6(a)的原图送入网络进行消融可视化验证,从消融可视化特征图前后对比如图 7 所示,可以看出,YOLOray 算法模型在使用 RWFF 后,有效减轻混叠效应的负面影响,给予了小目标更多的关注,这进一步证明本文提出的 RWFF 的有效性。

Focal-SIOU 加入了角度和边长惩罚项,并在损失函数之前添加了聚焦平衡函数,旨在加速预测框收敛的同时,解决边界框的回归贡献不平衡问题。从表 5 对比结果可以看出,使用 Focal-SIOU 后的网络模型,各项精度指标均有提升。此外,为了验证 Focal-SIOU 中 γ 的取值对模型精度的影响,本文通过设置不同的 γ 取值,在最具挑战性的数据集 SIXray 上进行了广泛实验,实验结果如表 6 所示,可以看出,将 γ 设置为 0.5 时效果更好,与基线模型相比,P、R 和 mAP 分别提高了 0.57%、0.37% 和 0.6%。

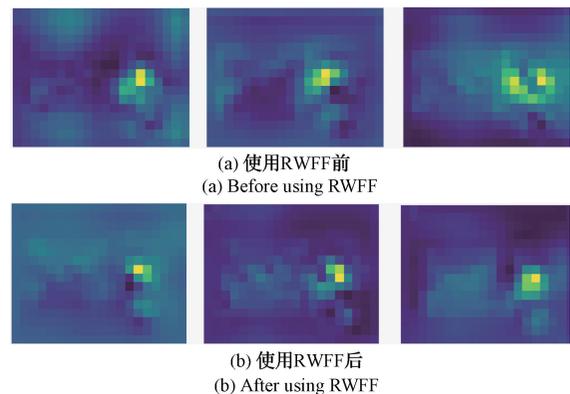


图 7 使用 RWFF 前后特征图的可视化比较

Fig. 7 Visualization comparison of feature maps before and after using RWFF

表 6 Focal-SIOU 中设置不同 γ 的实验结果

Table 6 Experimental results of setting different γ in Focal-SIOU

损失函数	γ	P/%	R/%	mAP/%
CIoU	—	93.32	85.76	91.6
Focal-SIOU	0.1	93.51	85.74	91.7
Focal-SIOU	1.0	93.43	86.03	91.9
Focal-SIOU	0.5	93.89	86.13	92.2

3 结论

本文结合 X 射线图像的特点以及实际应用需求设计了一种反向加权融合多尺度特征的 X 射线图像违禁品检测算法 YOLOray, 解决了现阶段违禁品检测方法中存在的问题。该算法主要包含 MSPM、RWFF 和 Focal-SIOU 3 个关键部分。其中, MSPM 能够获取丰富的多尺度上下文信息, 从而有效应对 X 射线安检图像中显著的类内变化; RWFF 则采用特征引导加权的方式高效融合蕴含丰富上下文特征的多级特征, 减轻了融合过程中的混叠效应; Focal-SIOU 损失函数可以有效平衡不同质量预测框的贡献差异, 并进一步提升预测框的收敛速度和回归精度。本文在 3 个极具挑战性的基准数据集上进行了广泛的评测实验, 并对比了最新的 7 种算法。实验结果充分证明了本文算法在 X 射线安检图像违禁品检测任务上具有一定优势。在未来工作中, 将继续致力于提高违禁品检测的性能, 并进一步探索在 X 射线图像场景下进行少样本目标检测的研究工作。

参考文献

- [1] NGUYEN H D, CAI R, ZHAO H, et al. Towards more efficient security inspection via deep learning: A task-driven X-ray image cropping scheme [J]. *Micromachines*, 2022, 13(4): 565.
- [2] ROGERS T W, JACCARD N, MORTON E J, et al. Automated X-ray image analysis for cargo security: Critical review and future promise [J]. *Journal of X-ray Science and Technology*, 2017, 25(1): 33-56.
- [3] 粟兴旺, 王晓明, 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.
- [4] SU X, WANG X M, HUANG J B, et al. Detection of contraband in X-ray security screening based on deformable convolution and attention mechanism [J]. *Electronic Measurement Technology*, 2023, 46(10): 98-108.
- [5] SHAO F, 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.
- [6] 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]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 2119-2128.
- [7] WEI Y, TAO R, WU Z, et al. Occluded prohibited items detection: An X-ray security inspection benchmark and deocclusion attention module [C]. *Proceedings of the 28th ACM International Conference on Multimedia*, 2020: 138-146.
- [8] WU J, SONG Y, ZHANG Y, et al. SA-CenterNet: Scale adaptive CenterNet for X-ray luggage image detection [C]. *26th International Conference on Pattern Recognition (ICPR)*. IEEE, 2022: 4857-4863.
- [9] CHANG A, ZHANG Y, ZHANG S, et al. Detecting prohibited objects with physical size constraint from cluttered X-ray baggage images [J]. *Knowledge-Based Systems*, 2022, 237: 107916.
- [10] LIU Y, ZHANG K, ZHAO Y, et al. Bi-RRNet: Bi-level recurrent refinement network for camouflaged object detection [J]. *Pattern Recognition*, 2023, 139: 109514.
- [11] XU L, JI Z, ZHOU X D, et al. Prohibited items detection on X-ray images with multi-task learning [C]. *Proceedings of the 8th International Conference on Computing and Artificial Intelligence*, 2022: 558-565.
- [12] MA C, ZHUO L, LI J, et al. EAOD-Net: Effective anomaly object detection networks for X-ray images [J]. *IET Image Processing*, 2022, 16(10): 2638-2651.
- [13] REN Y, ZHANG H, SUN H, et al. LightRay: Lightweight network for prohibited items detection in X-ray images during security inspection [J]. *Computers and Electrical Engineering*, 2022, 103: 108283.
- [14] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 8759-8768.
- [15] GHIASI G, LIN T Y, LE Q V. Nas-FPN: Learning scalable feature pyramid architecture for object detection [C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 7036-7045.
- [16] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection [C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 10781-10790.
- [17] ZONG Z, CAO Q, LENG B. RCNet: Reverse feature pyramid and cross-scale shift network for object detection [C]. *Proceedings of the 29th ACM International Conference on Multimedia*, 2021: 5637-5645.
- [18] 童小钟, 魏俊宇, 苏绍璟, 等. 融合注意力和多尺度特征的典型水面小目标检测 [J]. *仪器仪表学报*, 2023, 44(1): 212-222.
- [19] TONG X ZH, WEI J Y, SU SH J, et al. Fusion of

- attention and multiscale features for detection of typical small targets on water surface[J]. Chinese Journal of Scientific Instrument, 2023, 44(1): 212-222.
- [18] 张银胜,杨宇龙,吉茹,等.改进YOLOv5s的风力涡轮机表面缺陷检测[J].电子测量与仪器学报,2023,37(1):40-49.
ZHANG Y SH, YANG Y L, JI R, et al. Surface defect detection of wind turbine based on YOLOv5s[J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(1): 40-49.
- [19] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.
- [20] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11534-11542.
- [21] DAI Y, GIESEKE F, OEHMCKE S, et al. Attentional feature fusion [C]. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021: 3560-3569.
- [22] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [23] 赵子文,金永,陈友兴,等.基于改进YOLVOv5s的X射线图像粘接缺陷实时检测[J].国外电子测量技术,2023,42(4):181-186.
ZHAO Z W, JIN Y, CHEN Y X, et al. Real-time detection of bonding defects in X-ray images based on improved YOLVOv5s [J]. Foreign Electronic Measurement Technology, 2023, 42(4): 181-186.
- [24] ZHANG Y F, REN W, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression [J]. Neurocomputing, 2022, 506: 146-157.
- [25] GEVORGYAN Z. SIoU loss: More powerful learning for bounding box regression [J]. arXiv preprint arXiv:2205.12740, 2022.
- [26] WANG B, ZHANG L, WEN L, et al. Towards real-world prohibited item detection: A large-scale X-ray benchmark [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 5412-5421.
- [27] MA C, ZHUO L, LI J, et al. Occluded prohibited object detection in X-ray images with global context-aware multi-scale feature aggregation[J]. Neurocomputing, 2023, 519: 1-16.
- [28] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[C]. Conference and Workshop on Neural Information Processing Systems, 2016.
- [29] QIAO S, CHEN L C, YUILLE A. Detectors: Detecting objects with recursive feature pyramid and switchable atrous convolution[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 10213-10224.
- [30] 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, 2023: 7464-7475.
- [31] TAO R, LI H, WANG T, et al. Exploring endogenous shift for cross-domain detection: A large-scale benchmark and perturbation suppression network [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2022: 21157-21167.

作者简介

马昌嵩,硕士研究生,主要研究方向为深度学习、X射线违禁品检测等。

E-mail: 20211249618@nuist.edu.cn

裴晓芳(通信作者),硕士,副教授,主要研究方向为智能信号处理与应用、嵌入式系统设计等。

E-mail: xiaofangpei@163.com