



DOI: 10.12086/oee.2025.240286

CSTR: 32245.14.oee.2025.240286

自适应全景聚焦 X 射线图像 违禁品检测算法

崔丽群,杨莹莹*,金海波,吴正伟 辽宁工程技术大学软件学院,辽宁葫芦岛125105



摘要:针对 X 射线安检图像中样本重叠遮挡占比高、关键特征提取困难、背景噪声大导致的漏检和误检问题,提出 一种自适应全景聚焦 X 射线图像违禁品检测算法。首先,设计前景特征感知模块,通过强化前景目标的边缘结构和 纹理细节,精准区分违禁品和背景噪声,提高特征表达的准确性和完整性。然后,结合多分支结构和双重交叉注意力 机制构造多路径双维信息整合模块,优化通道和空间维度的特征交互与融合,加强关键特征的提取能力,有效抑制背 景干扰。最后,构建全景动态聚焦检测头,通过频率自适应空洞卷积实现感受野的动态调整,以适配小尺寸违禁品目 标的特征频率分布,增强模型对小目标的识别能力。在公开数据集 SIXray 和 OPIXray 上进行训练和测试, mAP@0.5 分别达到 93.3% 和 92.5%,优于其他对比算法。实验结果表明,该模型显著改善了 X 射线图像中违禁品 的漏检和误检情况,具有较高的准确性和鲁棒性。

关键词:X射线图像;违禁品检测;前景特征感知;多路径双维信息整合;频率自适应空洞卷积 中图分类号:TP391.4 文献标志码:A

崔丽群,杨莹莹,金海波,等. 自适应全景聚焦 X 射线图像违禁品检测算法 [J]. 光电工程,2025, **52**(4): 240286 Cui L Q, Yang Y Y, Jin H B, et al. Adaptive panoramic focusing X-ray image contraband detection algorithm[J]. *Opto-Electron Eng*, 2025, **52**(4): 240286

Adaptive panoramic focusing X-ray image contraband detection algorithm

Cui Liqun, Yang Yingying^{*}, Jin Haibo, Wu Zhengwei

Software College, Liaoning Technical University, Huludao, Liaoning 125105, China

Abstract: Aiming at the problem of leakage and misdetection caused by the high percentage of sample overlapping and occlusion, the difficulty of key feature extraction, and the large background noise in X-ray security images, an adaptive panoramic focusing X-ray image contraband detection algorithm is proposed. Firstly, the foreground feature awareness module is designed to accurately distinguish contraband and background noise by enhancing the edge structure and texture details of the foreground target to improve the accuracy and completeness of feature representation. Then, the multi-path two-dimensional information integration module is constructed by combining the multi-branch structure and dual cross attention mechanism to optimize the feature interaction and fusion in the channel and spatial dimensions, to strengthen the extraction capability of key features, and to effectively suppress the background interference. Finally, a panoramic dynamic focus detection head is constructed, which dynamically

收稿日期: 2024-12-05; 修回日期: 2025-02-16; 录用日期: 2025-02-17 基金项目: 国家自然科学基金 (62173171); 辽宁省高等学校基本科研项目 (LJKMZ20220699) *通信作者:杨莹莹, 1429337791@qq.com 版权所有©2025 中国科学院光电技术研究所

adjusts the receptive field through frequency adaptive dilated convolutions to accommodate the feature frequency distribution of small-sized contraband targets, thereby enhancing the model's ability to recognize small targets. Trained and tested on the public datasets SIXray and OPIXray, the mAP@0.5 reaches 93.3% and 92.5%, respectively, outperforming the other compared algorithms. The experimental results show that the proposed model significantly improves the leakage and false detection of contraband in X-ray images with high accuracy and robustness.

Keywords: X-ray images; contraband detection; foreground feature awareness; multi-path two-dimensional information integration; frequency adaptive dilated convolution

1 引 言

X射线图像违禁品检测在公共交通、物流运输、 海关检查等多个领域^[1-3]中发挥着重要作用,是图像 处理和目标检测中的关键技术之一,其核心任务是从 复杂背景环境中准确识别出违禁品的种类和位置信息, 确保人民生命财产安全和物品运输安全。与自然图像 不同,违禁品图像由X射线成像技术^[4]生成,能够清 晰捕捉目标的轮廓和内部结构信息,然而,由于X 射线成像的固有限制,易造成目标的颜色、纹理等细 节特征丢失。此外,X射线安检图像中的目标种类繁 多、形态各异,且常受到目标堆叠、遮挡、低对比度 和复杂背景等因素的影响,难以正确识别出目标,易 出现漏检和误检问题。因此,如何精准识别违禁品并 提高检测效率,已成为当前研究的重点和难点。

早期 X 射线图像违禁品自动识别技术主要基于 K 邻近算法^[5]、随机森林^[6]和支持向量机^[7]等传统机 器学习方法实现,因其检测精度低、可扩展性差、主 观依赖性强和参数调优困难等问题,现已被深度学习 智能识别算法所取代。基于深度学习的目标检测算法 具有自动特征学习、端到端处理、高检测精度以及应 对复杂场景和高维数据的强大处理能力等优势,在大 规模和复杂任务中表现优异,因此广泛应用于 X 射 线图像违禁品检测算法的研究中。现阶段,深度学习 目标检测算法根据是否生成候选区域分为单阶段算法 和两阶段算法两大类别。以 R-FCN^[8]、Faster R-CNN^[9] 等为代表的两阶段算法先生成候选区域,再进行精细 分类和定位,其检测精度较高,但模型结构复杂,训 练和推理时间较长,不适用于实时检测任务。代表性 的单阶段算法包括 SSD^[10]、RetinaNet^[11] 和 YOLO^[12-16] 系列算法,通过一个端到端的网络直接预测所有目标 的类别和位置,其模型结构简单、计算效率高、易部 署在嵌入式设备或资源受限的环境中,在X射线图

像违禁品检测领域具有更多优势。

近年来,X射线图像违禁品检测算法研究已经取 得了显著进展,众多高效检测算法不断被提出。Zhu 等[17] 提出了一种融合频率感知和全局与通道注意的 双流 Transformer 检测网络, 使网络能够同时捕捉图 像的空间特征和频率特征,实现浅层与深层特征的有 效融合,但模型的参数量和计算复杂度较高,不利于 部署在资源有限的设备上。Ahmed 等[18] 提出了一种 基于轮廓驱动学习的高分辨率网络,聚焦违禁品目标 的轮廓和边缘特征,并在低分辨率和高分辨率特征之 间实现信息交互,避免细节特征丢失,从而生成更为 丰富的特征表示,但模型易受到背景噪声的干扰,且 泛化能力不佳。董乙杉等^[19]在 YOLOv5 算法中结合 反向瓶颈结构和轻量化卷积注意力机制,有效提取了 违禁品目标的多样化特征,解决了违禁品目标因视角 和摆放位置变化导致的特征提取困难问题。Zhou 等^[20] 基于 YOLOv7 网络提出低参数特征和高密度特征聚 合模块, 使网络能够捕获更精细、更全面的特征, 以 提高复杂背景下小尺寸违禁品的检测精度,但在重叠 遮挡场景中存在漏检问题。Han 等^[21]基于 YOLOv8 模型提出了轻量级 SC-Lite 算法,使用 C2F FM (CSPNet faster convolution network module) 加速模块 来减少计算冗余和内存访问频率,并通过 Adaptation-BiFPN 模块和 LAMP 剪枝策略进一步降低模型体积 和参数量,但模型的检测精度有所降低。Han 等^[22] 在 YOLOv8 模型中引入可变形卷积和空间金字塔多 头注意力机制,动态调整感受野的位置和形状,以适 应形状复杂、尺寸和角度变化严重的违禁品目标,提 高了复杂场景下违禁品识别的准确性,但对于小尺寸 违禁品目标存在漏检问题。Wang 等^[23]在 YOLOv8n 基础上引入广义高效层聚合网络和高效多尺度注意力 机制,通过优化梯度传输路径和并行多尺度处理,增 强了网络对重叠遮挡目标的特征提取能力,但模型的

训练和推理速度较慢,难以满足实时应用场景的需求。 Wang 等^[24]在 YOLOv8 模型中引入自适应空间特征融 合机制和加权特征拼接算法,充分提取多尺度特征, 并结合 Soft-NMS 算法来提高遮挡场景下违禁品目标 的检测精度,但模型的泛化能力和鲁棒性较差。

针对上述问题,本文以 YOLOv8n 为基线模型, 提出一种自适应全景聚焦 X 射线图像违禁品检测算 法。首先,设计前景特征感知模块,突出前景目标的 关键特征,提高模型对违禁品目标的敏感度和关注度, 使模型能够精准区分违禁品和背景噪声。其次,设计 多路径双维信息整合模块,充分融合多尺度特征与全 局上下文信息,提高模型检测性能的同时减少模型计 算量。最后,构建全景动态聚焦检测头,增强模型对 小尺寸违禁品目标的识别与定位能力,有效缓解复杂 场景下小尺寸违禁品的漏检问题。

2 自适应全景聚焦 X 射线图像违禁品 检测

为解决 X 射线图像违禁品检测中目标重叠遮挡、 关键特征提取困难和小尺寸违禁品漏检等问题,本文 以 YOLOv8n 为基线模型,提出一种自适应全景聚焦 X 射线图像违禁品检测算法。本文算法的整体结构如 图 1 所示,其中 n'表示 C2f 中有 n'个 Bottleneck 单元。

所提算法在基线模型 YOLOv8n 的基础上做出了 三项改进:1)设计前景特征感知模块 (foreground feature awareness module, FFAM), 通过构建边缘感知 模块和纹理感知模块来强化模型对违禁品边缘结构和 纹理细节的提取能力,进一步增强模型聚焦能力与前 景特征表示质量,提高重叠遮挡场景下违禁品目标的 检测精度: 2) 设计多路径双维信息整合模块 (multipath two-dimensional information integration module, MPTI),结合多分支结构、双重交叉注意力机制与 FFAM 模块,在通道维度和空间维度上优化多尺度特 征的协同交互与深度融合,提高模型对关键特征的识 别能力与全局上下文信息的捕捉能力,有效抑制背景 冗余,加快模型推理速度;3)构建全景动态聚焦检测 头 (panoramic dynamic focus detection head, PDF Detect),通过频率自适应空洞卷积优化有效感受野聚 焦范围,增强模型对小目标的捕捉能力,显著提高复 杂场景下小尺寸违禁品的检测精度。



图 1 算法的整体结构 Fig. 1 Overall structure of algorithm

240286-3

2.1 前景特征感知模块

为解决在违禁品检测中因目标重叠遮挡和细节特 征缺失导致检测精度不佳的问题,本文提出前景特征 感知模块 (foreground feature awareness module, FFAM)。从边缘和纹理两个方面提取并强化违禁品目 标的外观特征,边缘感知模块旨在捕捉和凸显图像中 的边缘特征,帮助模型更加精确地识别目标的形状和 轮廓;纹理感知模块通过提取从粗略纹理到细微纹理 的多层次特征,构建更全面的特征表示,增强模型对 目标细节的表征能力。FFAM 的具体结构如图 2 所示, 主要由边缘感知模块 (edge aware module, EAM) 和纹 理感知模块 (texture aware module, TAM) 构成。

2.1.1 边缘感知模块

为解决因目标重叠遮挡和复杂背景导致的样本混 淆与边缘特征提取困难问题,提出了边缘感知模块。 该模块通过提取和强化违禁品目标的边缘结构,使前 景目标的边缘轮廓更加完整清晰,显著增强违禁品与 背景之间的视觉差异,有效增强了模型对违禁品边缘 特征的感知能力。

首先,使用 Sobel 算子^[25] 对输入图像 $X \in \mathbb{R}^{C \times H \times W}$ (C 表示特征图的通道数, $H \approx W$ 分别表示特征图的 高度和宽度)进行边缘检测,利用水平 Sobel 算子 X_h 和垂直 Sobel 算子 X_v 对其进行卷积运算,计算每个像 素点的梯度幅值,提取不同角度的边缘特征。生成的 边缘特征图 X_E 可表示为

$$X_{\rm E} = \sqrt{(X_{\rm h} * X)^2 + (X_{\rm v} * X)^2}$$
, (1)

式中:*表示卷积运算。

为进一步提高有效特征权重占比,降低背景噪声 干扰,设计了自适应权重学习块 (adaptive weight learning block, AWLB)。该模块通过全局平均池 化^[26](Global-AvgPool)、1×1Conv和 Sigmoid激活函 数^[27]对初始图像 *X*进行处理,以生成自适应权重。 自适应权重*W*;可表示为

 $W_i = \text{Sigmoid}(\text{Conv}(\text{Global-AvgPool}(X)))$. (2)

然后,通过点阵特征增强^[28](lattice feature enhanced, LFE)结构,利用自适应权重矩阵 W_1 和 W_2 对输入特征图 X 与边缘特征图 X_E 重新加权,进行特征增强与自适应融合,以提高违禁品目标边缘的完整性和连续性,使模型能够精准区分违禁品目标与背景区域,同时抑制无关背景的干扰。边缘感知模块输出的特征图 Y_1 可表示为

 $Y_1 = \operatorname{ReLU}(X \oplus X_E \otimes \phi(W_1)) + \operatorname{ReLU}(X_E \oplus X \otimes \phi(W_2)),$ (3)

式中: ReLU(·)表示非线性激活函数; ⊕表示逐元素 加法; ⊗表示逐元素乘法; *φ*(·)表示权重沿通道维度 的广播操作。

2.1.2 纹理感知模块

纹理感知模块主要由区域信息聚合 (region information aggregation, RIA)^[29] 与点阵特征增强 (lattice



图 2 前景特征感知模块 (FFAM) 的结构 Fig. 2 Structure of foreground feature awareness module (FFAM)

feature enhanced, LFE) 结构组成,旨在提取更加细致 且全面的纹理特征,有效缓解了样本重叠遮挡导致的 细节特征丢失问题。

首先,将边缘感知模块生成的特征图 $Y_1 \in \mathbb{R}^{C \times H \times W}$ 输入至区域信息聚合 RIA 模块中,利用多个不同大 小的池化核 $K = \{k_1, k_2, \cdots, k_n\}$ 对其进行多尺度平均池 化操作,从全局和局部视角捕捉不同尺度的纹理特征, 增强模型对多尺度目标的识别能力。通过池化操作, 特征图的尺寸被缩小至 $C \times (H/K) \times (W/K)$,随后利用 扩展操作将特征图恢复至原始尺寸 $C \times H \times W$ 。扩展 操作能够将局部特征信息传播至整个图像区域,在更 大范围内整合不同尺度的上下文信息,有助于准确捕 捉细节丰富且结构复杂的纹理特征。由平均池化和扩 展操作生成的临时特征图 T_1 可表示为

$$\boldsymbol{T}_{1ij}^{k} = \frac{\sum_{m=i-(i \bmod k)}^{i-(i \bmod k)+k} \sum_{n=j-(j \bmod k)}^{j-(j \bmod k)+k} \boldsymbol{Y}_{1mn}}{k^{2}}, \qquad (4)$$

式中: k表示当前操作所使用的池化核大小, $k \in K$; T_{1ij}^{k} 表示临时特征图 T_1 中第i行第j列的特征值; Y_{1nn} 表示输入特征图 Y_1 中第m行第n列的特征值。 $i-(i \mod k)$ 和 $j-(j \mod k)$ 用于确定池化窗口的起始 位置,并按照大小为 $k \times k$ 的网格进行滑动。将特征图 Y_1 和 T_1 在通道维度上进行拼接,生成一个通道数加倍 的特征图,其尺寸为 $2C \times H \times W$ 。然后通过卷积操作 对拼接后的特征图进行处理,其尺寸恢复为 $C \times H \times W$ 。由于聚合区域的尺度不同(k值各异), RIA 生成了一个特征图集合 $S = \{T_{2}^{k_1}, T_{2}^{k_2}, \dots, T_{2}^{k_n}\}$,在此 集合中,特征图 T_2 中的每个像素点均能感知其周围 $k \times k$ 大小的纹理特征。为了从特征图集合S中选择最 优特征图 Y_R 作为输出,RIA引入了门控卷积^[30](gated Conv),通过门控机制对特征图进行选择性过滤和优 化,保留关键纹理信息,去除冗余特征,进一步提高 检测准确性。所选特征图 Y_R 包含丰富的纹理细节,其 表达式为

$$Y_{\rm R} = G(S) \,, \tag{5}$$

式中: *G*(·)表示门控卷积; *S*表示特征图集合。RIA 模块的具体结构如图 3 所示。

然后,通过 Conv、批归一化 (batch normalization, BN)和 ReLU 激活函数^[31] 对边缘特征图 X_E 进行增强, 以增强模型对边缘特征的表达能力,生成更加完整且 清晰的边缘特征图 Y_E 。接着将特征图 Y_E 和 RIA 模块 生成的纹理特征图 Y_R 输入点阵特征增强结构,通过自 适应权重 W_3 和 W_4 来动态调整边缘特征和纹理特征在 融合过程中的比例,充分融合边缘与纹理信息,提高 前景目标的特征表示质量。纹理感知模块生成的纹理 增强特征图 Y_2 可表示为

$$Y_2 = \operatorname{ReLU}(Y_{\mathbb{R}} \oplus Y_{\mathbb{E}} \otimes \phi(W_3)) + \operatorname{ReLU}(Y_{\mathbb{E}} \oplus Y_{\mathbb{R}} \otimes \phi(W_4)).$$
(6)

最后,将纹理增强特征图 Y_2 通过 Conv、BN 和 Sigmoid 函数进行强化,生成注意力特征图 Y_A ,并利 用 Y_A 重新加权边缘增强特征图 Y_1 ,进一步细化和凸 显违禁品目标的外观特征,显著增强模型对前景目标 的感知能力。最终生成边缘清晰和纹理丰富的外观增 强特征图 Y_0 该过程可表示为



Fig. 3 Region information aggregation (RIA)

$$Y_{\rm A} = {\rm Sigmoid}({\rm BN}({\rm Conv}(Y_2))),$$
 (7)

$$Y = Y_{\rm A} \otimes Y_1 \ . \tag{8}$$

2.2 多路径双维信息整合

本文提出了一种多路径双维信息整合模块 (multipath two-dimensional information integration module, MPTI),充分利用通道维度和空间维度的特征信息, 增强模型对全局上下文信息和长程依赖关系的捕捉能 力。MPTI模块采用多路径结构,将输入特征划分为 多个并行路径,每条路径独立处理特定的特征子集。 在通道维度上,利用通道交叉注意力 (channel cross attention, CCA) 在不同特征通道间建立依赖关系,增 强模型对关键特征的聚焦能力;在空间维度上,通过 空间交叉注意力 (spatial cross attention, SCA)建立不 同位置间的长程关联,提取图像的空间上下文特征。 最后,通过残差连接和特征融合进一步整合各路径的 联合特征表示,在丰富上下文信息的同时有效抑制背 景冗余。MPTI模块的具体结构如图 4 所示。

受到 DaViT^[32] 和 DA-TransUNet^[33] 等相关研究的 启发,本文设计双重交叉注意力机制 (dual cross attention, DCA),并将其添加到 MPTI结构的 Bottleneck 单元中。双重交叉注意力 DCA 的总体结构如图 5(a) 所示。该模块通过顺序捕获多尺度特征的通道和空间依赖关系,实现多层次、多维度的全局特征交互,增强模型对关键特征的识别能力,有效抑制复杂背景干扰。首先,通过 Average pool、Projection和 Reshape 操作^[34],将特征图*X*_i转化为一系列能够有效保留多尺度特征的 Tokens *T*_i,该过程可表示为

$$T_i = \text{Reshape}(\text{Conv}(\text{AvgPool}(X_i))),$$
 (9)

式中: $T_i \in \mathbb{R}^{N \times C_i}$,其中 N 表示 Token 的数量, C_i 表示第 i 层特征图的通道数,i=1,2,...,n; Conv(·) 表示 Projection 操作; $X_i \in \mathbb{R}^{C_i \times \frac{H}{2^{i-1}} \times \frac{W}{2^{i-1}}}$ 表示主干网络输出的 多尺度特征图。而后将 Tokens T_i 顺序输入 CCA 和 SCA 模块进行处理,捕捉通道和空间维度的长程依 赖关系和全局上下文信息,并输出特征图 Y_i 。最后采 用层归一化^[35](layer norm)和 GeLU 激活函数^[36]进一 步优化模型的特征表达能力。

CCA 模块的具体结构如图 5(b) 所示。在 CCA 模块中,对输入的各个 Token T_i 进行层归一化 (layer norm)处理,并将归一化后的部分 Tokens T_i 在通道维度进行拼接,得到全局特征矩阵 Tokens T_c 。然后将





https://doi.org/10.12086/oee.2025.240286



图 5 双重交叉注意力机制 (DCA) 结构。(a) DCA 模块; (b) CCA 模块; (c) SCA 模块 Fig. 5 Structure of dual cross attention mechanism (DCA). (a) DCA module; (b) CCA module; (c) SCA module

拼接后的特征*T*。经过处理生成键(*K*)和值(*V*),其余 单独的*T*_i生成查询(*Q*),这种结构能够更广泛地捕捉 不同通道之间的交互信息,形成更丰富的特征表示, 帮助模型更有效地提取和融合关键信息。在生成查 询*Q*、键*K*和值*V*时,CCA模块采用深度卷积^[37]投 影(DWConv projection, DWP)代替传统注意力机制中 的 Projection操作。深度卷积用于投影的主要作用在 于保留通道内的局部信息,增强模型的细粒度特征表 达能力,同时减少不同通道间的冗余特征,降低计算 成本和内存消耗。此外,对*Q*、*K*、*V*进行 Reshape 操作来调整特征矩阵的形状,使其适配后续的矩阵运 算。通过 DWConv projection 和 Reshape 操作生成*Q*、 *K*、*V*的过程可表示为

$$\boldsymbol{Q} = \operatorname{Reshape}(\operatorname{DWP}(\boldsymbol{T}_i)), \qquad (10)$$

$$\boldsymbol{K} = \operatorname{Reshape}(\operatorname{DWP}(\boldsymbol{T}_{c})), \qquad (11)$$

$$V = \text{Reshape}(\text{DWP}(T_c)),$$
 (12)

式中: $Q \in \mathbb{R}^{N \times C_i}$, $K \in \mathbb{R}^{N \times C_c}$, $V \in \mathbb{R}^{N \times C_c}$, C_i 表示原始

输入特征 T_i 的通道数, C_c 表示合并特征 T_c 的通道数; DWP(·)表示 DWConv projection 操作。接着对矩阵 Q和 K进行点积运算,计算每对 Token 之间的相关 性,捕捉各通道之间的依赖关系。然后通过 Softmax 运算将相关性转化为权重分布,表示每个 Token 与其 他所有 Tokens 之间的关联强度;再将 Softmax 输出 的权重矩阵与 V的转置矩阵 V^T 进行点积运算,计算 全部 Tokens 的注意力分数,以确保每个 Token 在通 道维度上都具有全局视野,从而使模型能够更加精准 地捕捉跨通道的全局交互信息,有效促进通道间的信 息交流与融合。CCA 模块的输出特征 \overline{T}_i 可表示为

$$\bar{\boldsymbol{T}}_{i} = \operatorname{Soft} \max\left(\frac{\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{K}}{\sqrt{d_{\mathrm{k}}}}\right) \times \boldsymbol{V}^{\mathrm{T}}, \qquad (13)$$

式中: d_k 表示向量维度, $1/\sqrt{d_k}$ 表示缩放因子; $Q^T K$ 表示将矩阵Q转置为 $C \times N$,再与矩阵K进行点积运算。

SCA 模块的具体结构如图 5(c) 所示。SCA 模块

的输入是来自 CCA 模块的输出特征 $\overline{T}_i \in \mathbb{R}^{N \times C_i}$,首先 对所有 Tokens \overline{T}_i 进行 Layer Norm 处理,然后将部分 Tokens \overline{T}_i 沿通道维度拼接得到 \overline{T}_c 。与 CCA 模块不同, SCA 模块将拼接后的 \overline{T}_c 经过处理生成 Q 和 K,将单 独的 \overline{T}_i 生成 V,实现多尺度特征在空间和通道维度上 的特征交互与融合。通过 DWConv Projection 和 Reshape 操作生成 Q、K、V的过程可表示为

$$\boldsymbol{Q} = \operatorname{Reshape}(\operatorname{DWP}(\bar{\boldsymbol{T}}_{c})),$$
 (14)

$$\boldsymbol{K} = \text{Reshape}(\text{DWP}(\bar{\boldsymbol{T}}_{c})), \qquad (15)$$

$$V = \text{Reshape}(\text{DWP}(\bar{T}_i)), \qquad (16)$$

式中: $Q \in \mathbb{R}^{N \times C_c}$, $K \in \mathbb{R}^{N \times C_c}$, $V \in \mathbb{R}^{N \times C_i}$ 。然后,通过 点积运算计算矩阵 $Q \cap K$ 之间的相似性,以度量任 意两个空间位置之间的关联强度。接着,利用 Softmax 函数将相似性转化为空间注意力权重,捕捉 多层级和多尺度空间区域之间的依赖关系,使模型能 够聚焦于重要的空间位置。再将 Softmax 输出的空间 注意力权重矩阵与值矩阵 V进行点积运算,生成具备 全局空间依赖的特征表示,整合不同位置的空间信息, 从而实现对全局空间上下文的精准建模,有效增强模 型对复杂空间关系的理解能力。SCA 模块的输出特 征 Y_i 可表示为

$$Y_i = \operatorname{Soft} \max\left(\frac{QK^{\mathrm{T}}}{\sqrt{d_{\mathrm{k}}}}\right) \times V$$
, (17)

式中: *QK*^T表示 *Q* 与 *K* 的转置矩阵*K*^T进行点积运算。 最后,将输出结果*Y*,通过深度卷积投影到统一的通道 维度上,保证输出特征在通道维度的一致性,便于后 续模块处理。

2.3 全景动态聚焦检测头

空洞卷积 (Dilated convolution, DC)^[38] 通过在卷积 核相邻元素之间插入间隔来扩大感受野,在不增加计 算复杂度的情况下捕获更广泛的上下文信息。但仅使 用固定的空洞率会导致特征图出现网格效应^[39],细节 信息丢失,特别是在检测较小或复杂形态的目标时易 出现漏检问题,从而影响检测结果的完整性和准确性。 因此本文对空洞卷积进行两部分优化,构建频率自适 应空洞卷积^[40](frequency adaptive dilated convolution, FADC),结构如图 6 所示,由自适应卷积核 (adaptive kernel)、频率特征分解 (frequency feature decompose) 和自适应空洞率 (adaptive dilation rate) 三部分组成; 利用该卷积进一步优化检测头,得到全景动态聚焦检 测头 (panoramic dynamic focus detection head, PDF_Detect),结构如图7所示。

1) 自适应卷积核 (Adaptive kernel)。首先,利用 全局池化 (global pool) 提取特征图的全局特征,再通 过两个并行卷积路径 Conv-ReLU-Conv-Sigmoid 将全 局特征分解为低频特征和高频特征。低频特征有助于 提取全局上下文信息,高频特征有助于捕捉局部差异 和高频细节信息。然后,通过低频卷积核与高频卷积 核分别提取特征图的低频特征和高频特征,并利用低 频权重 W_L和高频权重 W_H对这两种特征进行加权融合, 从而得到自适应卷积核。此外,自适应卷积核通过引 入低频加权系数λ和高频加权系数λ_h,能够根据上下 文信息动态调整权重分布,增强卷积核对不同频率特 征的处理能力。自适应卷积核的权重 W[']可表示为

$$W' = \lambda_1 W_{\rm L} + \lambda_{\rm h} W_{\rm H} , \qquad (18)$$

$$\begin{cases} \boldsymbol{W}_{\mathrm{L}} = \frac{1}{K \times K} \sum_{i=1}^{K \times K} \boldsymbol{W}_{i} \\ \boldsymbol{W}_{\mathrm{H}} = \boldsymbol{W} - \boldsymbol{W}_{\mathrm{L}} \end{cases}$$
(19)

式中: $K \times K$ 表示卷积核的尺寸; $1/(K \times K)$ 表示对求 和结果取平均值; $\sum_{i=1}^{K \times K} W_i(i = 1, 2, \dots, K)$ 表示将卷 积核中所有 W_i 进行求和; W表示静态卷积核的权 重。

類率特征分解 (Frequency feature decompose)。
 首先,通过傅里叶变换 (discrete Fourier transform,
 DFT)^[41]将输入特征图*X* ∈ ℝ^{H×W}从空间域转换到频域,
 计算表达式为

$$X_{\rm F}(u,v) = \frac{1}{HW} \sum_{h=0}^{H-1} \sum_{w=0}^{W-1} X(h,w) e^{-2\pi j(uh+vw)} , \qquad (20)$$

$$\begin{cases} u \in \left\{-\frac{H}{2}, -\frac{H+1}{2}, \cdots, \frac{H-1}{2}\right\} \\ v \in \left\{-\frac{W}{2}, -\frac{W+1}{2}, \cdots, \frac{W-1}{2}\right\} \end{cases},$$
(21)

式中: $X_{\rm F} \in \mathbb{R}^{H \times W}$ 表示 DFT 的输出; $X_{\rm F}(u,v)$ 表示傅里 叶变换后的频率坐标; X(h,w)表示特征图 X 的坐标。 然后,将输入特征分解为 B 个频率带,覆盖从高频到 低频的不同频段特征,再利用二值掩码 M_b 提取每个 频率带中的关键信息。该过程可表示为

$$M_b(u,v) = \begin{cases} 1, \ \phi_b \leq \max(|u|,|v|) < \phi_{b+1} \\ 0, \text{ otherwise} \end{cases}, \quad (22)$$

式中: *B* 表示划分的频率带数量, 对应于 *B*+1 个预定 义的频率阈值集合 $\left\{0,\phi_1,\phi_2,\cdots,\phi_{B-1},\frac{1}{2}\right\}$; $b \in \{0,1,\cdots,B-1\}$ 表示第 *b* 个频率带, 每个频率带区间为[ϕ_b,ϕ_{b+1})。



图 6 频率自适应空洞卷积 (FADC) 结构 Fig. 6 Structure of frequency adaptive dilated convolution (FADC)



图 7 全景动态聚焦检测头 (PDF_Detect) 结构 Fig. 7 Structure of panoramic dynamic focus detection head (PDF_Detect)

最后,利用选择映射A_b(i, j)对不同频段的特征进行空间重加权,根据不同空间位置动态调整频率响应,平衡高频细节和低频全局特征。生成的频率平衡特征 *X*(i, j)可表示为

$$\hat{X}(i,j) = \sum_{b=0}^{B-1} A_b(i,j) X_b(i,j) , \qquad (23)$$

 $X_b = F^{-1}(M_b X_{\rm F}) , \qquad (24)$

式中: $X_b(i, j)$ 表示傅里叶逆变换后的频率坐标; F^{-1} 表示傅里叶逆变换; $A_b \in \mathbf{R}^{H \times W}$ 表示第 b个频率带的选择图 (selection map)。

3) 自适应空洞率 (Adaptive dilation rate)。根据特 征图的空间变化,自适应地为每个像素分配最优空洞 率,在扩大感受野的同时保留频率信息的完整性。在 低频区域增加空洞率以扩展感受野,在高频区域抑制 空洞率以减少频率信息丢失。输出特征图中位置 *p* 处 的像素值*Y*(*p*)可表示为

$$Y(p) = \sum_{i=1}^{K \times K} \boldsymbol{W}_i \times X(p + \Delta p_i \times \hat{D}(p)), \qquad (25)$$

式中: Δp_i 表示偏移量; $\hat{D}(p)$ 表示自适应空洞率,通 过具有参数 θ 的卷积层进行预测。每个像素的感受野 RF(p)与自适应空洞率 $\hat{D}(p)$ 相关,并通过高频功率 HP(p)衡量频率信息的丢失,生成预测参数 θ ,其过 程可以表示为

$$\theta = \max\left(\sum RF(p) - \sum HP(p)\right), \qquad (26)$$

$$RF(p) = (K-1) \times \hat{D}(p) + 1$$
, (27)

$$HP(p) = \sum_{H^+} \left| X_{\rm F}^{(p,s)}(u,v) \right|^2, \qquad (28)$$

式中: $H^*_{\hat{D}(p)}$ 表示因感受野限制无法捕获的高频分量集合; $X^{(p,s)}_{F}(u,v)$ 表示局部特征的傅里叶变换。为获取自适应空洞率 $\hat{D}(p)$ 的最佳分布,进一步优化参数以平衡有效带宽与感受野大小。参数 θ_{o} 的优化可以表示为

$$\theta_{o} = \max\left(\sum_{p \in HP^{-}} \hat{D}(p) - \sum_{p \in HP^{+}} \hat{D}(p)\right), \qquad (29)$$

式中: *HP*⁻表示高频功率较低的像素集合; *HP*⁺表示 高频功率较高的像素集合。

PDF_Detect 根据特征的频率分布自适应地选择 最优感受野,增强模型对多尺度、多频率特征的感知 能力,进一步增强模型对小目标的捕获与保留能力, 有效减少小尺寸违禁品在复杂场景下的漏检与误检 情况。

3 实验结果与分析

3.1 数据集和实验环境

SIXray 数据集^[42]广泛应用于国内外违禁品检测 研究领域,其图像背景复杂,能够高度还原真实安检 场景。该数据集包含 8929 张违禁品图像,按照 8:1:1 的比例随机划分为训练集、验证集和测试集,涵盖 5 种违禁品类别:gun (枪)、knife (刀具)、wrench (扳 手)、pilers (钳子)和 scissors (剪刀),分别简称为 GU、 KN、WR、PL 和 SC。

OPIXray 数据集^[29]由专业机场安检人员手工标注, 其图像中存在不同程度的遮挡问题。该数据集包含 8885 张违禁品图像,其中 7109 张用于模型训练, 1776 张用于测试。图像的检测类别为 straight knife (直刀)、folding knife (折叠刀)、scissors (剪刀)、 utility knife (美工刀)和 multi-tool knife (多用刀),分 别简称为 ST、FO、SC、UT和 MU。

https://doi.org/10.12086/oee.2025.240286

实验环境为 Ubuntu 20.04 操作系统,硬件设备 为 Intel Xeon Platinum 8358P 处理器和一块 24 GB 显 存的 NVIDIA GeForce RTX 3090 显卡。软件环境为 Python 3.8,深度学习框架为 Pytorch 1.11.0+CUDA 11.3。采用 YOLOv8n 作为基线模型,网络训练采用 SGD 优化器,初始学习率为 0.01,权重衰减为 0.0005,动量参数为 0.937, epoch 为 200, batch size 为 16。

3.2 评价指标

本实验采用准确率 (precision, *P*)、召回率 (recall, *R*)、平均精度 (average precision, AP)、每秒检测帧 数 (frames per second, FPS)、参数量 (params)、模型计 算量 (GFLOPs)、平均精度均值 mAP@0.5 (阈值 *IoU*=0.5 时的平均值)和 mAP@0.5 : 0.95 (阈值 *IoU* 从 0.5 到 0.95 之间,以 0.05 为步长,分别计算每个 阈值下的平均精度,并对这些精度取平均值)对模型 进行综合评价。*P*、*R*、*AP*、*mAP*和*FPS*的计算表达 式分别为

$$P = \frac{TP}{TP + FP} , \qquad (30)$$

$$R = \frac{TP}{TP + FN},$$
(31)

$$AP = \int_0^1 PR dR , \qquad (32)$$

$$mAP = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{1} PR dR$$
, (33)

$$FPS = \frac{Framenum}{ElapsedTime},$$
 (34)

式中:n表示违禁品的总类别数量;TP表示检测正确的目标数量;FP表示误检的目标数量;FN表示漏检的目标数量;Framenum表示总帧数;ElapsedTime表示总耗时。

3.3 消融实验

为验证各项改进模块的可行性与有效性,在保证 实验环境和参数设置一致的前提下,以YOLOv8n为 基线模型,在SIXray数据集上进行消融实验,实验 结果如表1所示。其中A、B、C分别表示前景特征 感知(FFAM)模块、多路径双维信息整合(MPTI)模

		Ta	ble 1	Ablation experiment results of the improved algorithm in the SIXray dataset								
No.	YOLOv8n	А	В	С	P /%	R /%	mAP@0.5/%	mAP@0.5 : 0.95/%	Params/M	GFLOPs		
1	\checkmark	×	×	×	91.3	84.2	89.7	66.1	3.01	8.1		
2	\checkmark	\checkmark	×	×	93.8	85.4	91.8	68.6	4.17	11.3		
3	\checkmark	×	\checkmark	×	92.0	84.0	90.6	66.2	2.48	6.2		
4	\checkmark	×	×	\checkmark	91.8	84.9	91.3	66.5	2.78	7.7		
5	\checkmark	\checkmark	\checkmark	×	94.1	86.3	92.2	69.4	2.86	7.8		
6	\checkmark	\checkmark	×	\checkmark	94.7	86.7	92.6	70.7	3.59	9.4		
7	\checkmark	×	\checkmark	\checkmark	94.3	87.3	91.9	68.6	2.65	6.8		
8	\checkmark	\checkmark	\checkmark	\checkmark	94.6	88.8	93.3	72.4	2.92	7.9		

表1 改进算法在 SIXray 数据集的消融实验结果

块和全景动态聚焦检测头 (PDF_Detect); "√" 表示使 用改进策略,"×"表示不使用。

表1表明,单独添加前景特征感知 (FFAM) 模块 时,准确率提高 2.5%, mAP@0.5 提高 2.1%, 说明 FFAM 能够增强模型对前景目标的聚焦能力与特征表 示质量,使模型更准确地识别和定位违禁品目标。仅 添加多路径双维信息整合 (MPTI) 模块时,准确率提 高 0.7%, mAP@0.5 提高 0.9%, 参数量减少 0.53 M, GFLOPs 降低 1.9, 说明 MPTI 通过优化通道和空间 维度的特征交互与融合,能够增强模型对全局上下文 信息和长程依赖关系的捕捉能力,同时有效减少了模 型的计算复杂度。仅添加全景动态聚焦检测头 (PDF Detect)时, 召回率提高 0.7%, mAP@0.5 提高 1.6%,参数量减少 0.23 M, GFLOPs 降低 0.4,说明 PDF Detect 能够根据特征的频率分布自适应地选择 最优感受野大小,从而有效减少背景噪声和计算冗余, 增强了模型对小尺寸违禁品目标的识别能力。

在引入 FFAM 的基础上添加 MPTI 模块,准确率 提高 2.8%, mAP@0.5 提升至 92.2%, 参数量减少

0.15 M, GFLOPs 降低 0.3, 进一步提高了模型的检 测精度和计算效率。同时引入 FFAM 和 PDF Detect 模块, 准确率提升至 94.7%, mAP@0.5 提高 2.9%, mAP@0.5:0.95 提高 4.6%, 有效提升了模型的检测 性能,降低漏检率。同时引入 MPTI 和 PDF Detect 模块,准确率提高 3.0%, mAP@0.5 提高 2.2%,参数 量减少 0.36 M, GFLOPs 降低 1.3, 在提高违禁品检 测精度的同时降低了模型的计算复杂度。同时添加 FFAM、MPTI和 PDF Detect 模块,即本文改进算法, 相比 YOLOv8n, 准确率提高 3.3%, 召回率提高 4.6%, mAP@0.5 达到 93.3%, mAP@0.5:0.95 达到 72.4%, 参数量减少 0.09 M, GFLOPs 降低 0.2。由数据分析 可知,本文改进模型在检测性能和计算效率上均得到 了显著提升,有效减少了 X 射线安检图像中违禁品 的漏检和误检问题,验证了改进模型的有效性。

为进一步评估改进模型对违禁品检测性能的影响, 本文在 SIXray 数据集上进行消融实验来对比各类别 的平均精度 AP, 实验结果如表 2 所示。数据显示, GU、KN、WR、PL和SC的AP值分别达到99.4%、

Table 2 Comparison results of the accuracy of ablation experiments by category										
Ne	YOLOv8n	A	В	С						
NO.					GU	KN	WR	PL	SC	- IIIAF@0.5/%
1	\checkmark	×	×	×	97.8	86.5	87.6	94.5	84.8	89.7
2	\checkmark	\checkmark	×	×	98.5	87.8	90.5	95.3	87.4	91.8
3	\checkmark	×	\checkmark	×	98.2	86.9	88.6	95.1	85.2	90.6
4	\checkmark	×	×	\checkmark	98.4	87.4	88.4	94.6	87.9	91.3
5	\checkmark	\checkmark	\checkmark	×	99.2	88.4	90.7	96.0	86.5	92.2
6	\checkmark	\checkmark	×	\checkmark	99.1	88.2	92.1	95.8	86.8	92.6
7	\checkmark	×	\checkmark	\checkmark	98.7	86.8	92.6	95.6	85.8	91.9
8	\checkmark	\checkmark	\checkmark	\checkmark	99.4	89.0	93.5	96.4	87.1	93.3

89.0%、93.5%、96.4%和87.1%,较基线模型分别提高了1.6%、2.5%、5.9%、1.9%和2.3%,提升效果明显,说明所提改进模型应用于X射线图像违禁品检测是准确且高效的。

3.4 对比实验

所提改进算法与近年主流算法 SSD^[10]、Faster R-CNN^[9]、YOLOv5n、YOLOv6^[43]、YOLOv7-Tiny^[44]、 YOLOv8n、YOLOv9^[45]和YOLOv10n^[46]在SIXray数 据集上进行性能测试,对比实验结果如表3所示。

与经典算法 SSD 和 Faster R-CNN 相比,改进算 法的 mAP@0.5 分别提高了 13.7% 和 7.8%, 每秒检测 帧率 (FPS) 分别提升了约2 倍和4倍, 在检测性能和 推理速度方面展现了明显优势;此外,改进算法在参 数量和计算量上显著降低,有效减少了计算冗余问题, 更适合部署在嵌入式设备上。与轻量化模型 YOLOv5n 相比,改进算法在略微增加参数量和计算 量的基础上, 召回率高出 10.1%, mAP@0.5高出 5.4%, FPS 高出 24.9; 对于 WR、PL 和 SC 三个类别 的检测精度提升效果最明显,其 AP 值分别提升 10.9%、5.9% 和 7.4%。与 YOLOv6 和 YOLOv7-Tiny 相比,改进算法的准确率分别高出 7.2% 和 8.0%,召 回率高出 6.7% 和 7.5%, mAP@0.5 高出 5.1% 和 6.6%,参数量分别减少了 1.31 M 和 3.1 M, GFLOPs 降低了 3.9 和 5.3, 改进算法在检测性能和计算效率 之间实现了良好的平衡。与 YOLOv9 和 YOLOv10n 相比,改进算法的准确率分别高出 3.2% 和 4.5%,召 回率高出 6.3% 和 3.1%, mAP@0.5 高出 3.2% 和 1.9%, FPS 高出 14.4 和 9.6, GFLOPs 分别减少了 0.3 和 0.5; 改进算法在不增加模型计算复杂度的情况 下,实现了更快的检测速度和更高的检测性能。综合 实验数据可知,所提算法在检测性能和计算效率方面 具有优越性,对X射线图像违禁品检测具有更强的 特征提取能力和更快的推理速度。

为进一步验证本文算法的优越性和泛化性,在 OPIXray数据集上对比各算法的检测性能,实验结果 如表 4 所示。由表 4 可知,所提算法在 P、R 和 mAP@0.5 指标上分别达到了 93.2%、89.4% 和 92.5%, FPS 值为123.8,且各类别违禁品目标的检测精度均优 于其他对比算法,综合检测性能表现最佳,表明所提 改进算法在违禁品检测任务中具有良好的鲁棒性和普 适性。

图 8 展示了改进模型与 YOLOv8n 算法在训练过 程中各评价指标的曲线对比情况。可直观地看出,改进 模型的 Precision、Recall、mAP@0.5 和 mAP@0.5: 0.95 等指标均优于基线模型 YOLOv8n,表明改进模 型具有更强的收敛能力和更稳定的检测性能,能够更 加准确高效地识别出违禁品目标。

F1 分数 (F₁) 是准确率和召回率的调和平均数, 综合反映模型的准确性和全面性。F1-curve 用于展示 F1 分数随置信度阈值变化的趋势,其取值范围在 [0,1] 之间,0表示模型性能最差,1为最佳表现。 从图 9 可以直观地看出,改进模型的各类别 F₁ 均高 于基线模型 YOLOv8n。具体而言,YOLOv8n 的平 均 F₁ 为 0.84,而改进模型的平均 F₁ 达到 0.91,表 明改进模型显著提高了各类别违禁品的检测精度,有 效减少了漏检和误检情况,充分验证了所提算法的有 效性。

所提算法在训练集 (train) 和验证集 (val) 上的边 界框损失 (box_loss)、分类损失 (cls_loss)、分布式焦

AP/% **R**/% mAP@0.5/% Model P/% FPS Params/M GFLOPs GU WR ΡL SC KN 79.6 60.4 26.28 62.7 SSD 91.6 74.8 69.8 80.9 83.4 83.5 70.8 Faster R-CNN 89.2 79.4 80.1 86.4 84.2 87.9 75.1 85.5 31 136.72 369.8 YOLOv5n 98.7 90.5 79.7 92.7 78.7 87.9 101 2.50 7.1 88.2 82.6 YOLOv6 97.1 83.7 87.0 92.2 81.1 87.4 82.1 88.2 121.6 4.23 11.8 YOLOv7-Tiny 97.7 82.6 85.0 89.8 78.7 86.6 81.3 86.7 94.5 6.02 13.2 YOLOv8n 97.8 86.5 87.6 94.5 84.8 91.3 84.2 89.7 108.6 3.01 8.1 YOLOv9 91.4 82.5 90.1 98.4 86.6 88.0 92.9 84.8 111.5 2.27 8.2 YOLOv10n 98.3 87.5 90.2 95.8 86.0 90.1 85.7 91.4 116.3 2.71 8.4 Ours 99.4 89.0 93.5 96.4 87.1 94.6 88.8 93.3 125.9 2.92 7.9

表 3 SIXray 数据集对比实验结果 Table 3 Comparison experiment results on the SIXray dataset

https://doi.org/10.12086/oee.2025.240286

Table 4 Comparison experiment results on OPIXray dataset											
Marial	AP/%						DI01	mAB@0.5/0%		Dereme/M	
Model	ST	FO	SC	UT	MU	- <i>F1%</i> 0	1170		FFS	Params/ivi	GFLOPS
SSD	33.5	73.4	89.5	64.1	80.8	70.5	62.8	65.7	53.2	26.28	62.7
Faster R-CNN	68.3	88.7	90.0	82.5	89.4	89.0	82.7	84.8	25.4	136.72	369.8
YOLOv5n	72.6	92.4	98.5	85.1	93.3	87.8	83.9	88.4	109.8	2.50	7.1
YOLOv6	78.6	92.0	98.3	87.1	94.6	89.2	86.1	90.1	119.5	4.23	11.8
YOLOv7-Tiny	65.7	91.6	97.9	84.0	84.0	90.7	81.4	86.4	102.9	6.02	13.2
YOLOv8n	76.9	94.3	98.0	85.3	93.8	90.5	85.5	89.7	116.8	3.01	8.1
YOLOv9	75.0	91.6	98.6	83.0	93.5	88.5	85.2	88.3	111.7	2.27	8.2
YOLOv10n	68.9	93.8	97.4	81.8	93.4	88.3	81.5	84.9	108.5	2.71	8.4
Ours	79.4	95.9	99.2	88.5	96.1	93.2	89.4	92.5	123.8	2.92	7.9

表4 OPIXray 数据集对比实验结果





点损失 (dfl_loss) 以及 Precision、Recall、mAP@0.5 和 mAP@0.5:0.95 的收敛曲线如图 10 所示。可直观 地看出,训练过程中的 loss 曲线呈显著下降趋势,表 明模型在不断减少预测误差;验证集的各项 loss 曲线 收敛至更低的值并趋于平稳,表明改进模型在边界框 预测和目标类别预测方面具有较强的鲁棒性与收敛性。 此外, Precision、Recall、mAP@0.5 和mAP@0.5:0.95 等指标均展现出明显的上升趋势,其曲线振荡幅度较 小、收敛速度较快,在达到较高数值后趋于稳定。整 体而言,所提算法的收敛曲线平滑,表现出良好的收 敛性能,进一步验证了其在违禁品检测任务中的有效 性与稳定性。



图 9 基线模型 YOLOv8n(左) 与改进模型 (右) 在 SIXray 数据集上的 F1-curve 对比图 Fig. 9 F1-curve comparison diagram between the baseline model YOLOv8n (left) and improved model (right) on the SIXray dataset





为直观地反映所提算法的检测效果,选取三组不 同场景的检测结果对比图,如图 11 所示。从图 11 第 一行的结果可以看出,在重叠遮挡场景中,所提算法 能够以较高的检测精度识别出更多的违禁品目标,表 明所提算法有效增强了模型对前景目标的感知能力, 能够更精准地识别和定位被遮挡的违禁品目标。从 图 11 第二行的结果可以看出,针对小目标样本,所 提算法的检测精度更高,表明所提算法通过在检测头 处构建频率自适应空洞卷积,显著增强了模型对小尺 寸违禁品目标的聚焦能力,有效减少了小尺寸违禁品 在复杂场景下的漏检与误检情况。从图 11 第三行的 结果可以看出,针对复杂背景下的多尺度目标,所提 算法能够更全面地检测出不同尺度的违禁品目标,具 有更高的识别准确率,表明所提算法能够准确地区分 违禁品目标与背景区域,有效抑制背景噪声。

综上所述,本文算法能够在多种复杂场景中精准 高效地识别出违禁品目标,表现出更强的泛化能力和 更稳定的检测性能,进一步验证了其在X射线图像 违禁品检测领域的实际应用价值和研究意义。

4 结 论

为解决 X 射线图像违禁品检测中存在的漏检和

https://doi.org/10.12086/oee.2025.240286



图 11 检测效果对比

Fig. 11 Comparison of detection results

误检问题,本文提出一种自适应全景聚焦 X 射线图 像违禁品检测算法。通过前景特征感知模块的创新设 计, 显著增强模型对前景目标的特征表达能力, 实现 对重叠遮挡场景中违禁品目标的精准识别。多路径双 维信息整合模块通过优化多尺度特征在通道维度和空 间维度的交互与整合,增强模型对关键特征的识别能 力,使模型能够提取更丰富、更全面的上下文信息。 全景动态聚焦检测头通过引入频率自适应空洞卷积和 动态聚焦机制,能够根据特征的频率分布自适应选择 最优感受野大小, 增强模型对小尺寸违禁品目标的聚 焦能力,有效改善了小尺寸违禁品在复杂场景下的漏 检与误检问题。实验结果表明,所提算法在 SIXrav 数据集和 OPIXray 数据集上的表现相比于其他算法具 有一定的优越性,满足 X 射线图像违禁品检测所需 的性能与要求。虽然所提算法已经取得了较好的检测 效果,但仍存在局限性,需要进一步优化和改进。在 未来的工作中,将针对色彩对比度低、类内差异大和 正负样本不均衡等问题进行深入研究,进一步提高模 型的检测性能和实际应用价值。

利益冲突:所有作者声明无利益冲突

参考文献

 Chang Q Q, Chen J M, Li W J. Dangerous goods detection technology based on X-ray images in urban rail transit security inspection[J]. Urban Mass Transit, 2022, 25(4): 205–209. 常青青, 陈嘉敏, 李维姣. 城市轨道交通安检中基于 X 射线图像的危险品识别技术研究[J]. 城市轨道交通研究, 2022, 25(4): 205-209.

[2] Chen Z Q, Zhang L, Jin X. Recent progress on X-ray security inspection technologies[J]. *Chin Sci Bull*, 2017, **62**(13): 1350-1364. 陈志强, 张丽, 金鑫. X 射线安全检查技术研究新进展[J]. 科学通

报, 2017, **62**(13): 1350-1364.

- [3] Akcay S, Breckon T. Towards automatic threat detection: a survey of advances of deep learning within X-ray security imaging[J]. *Pattern Recognit*, 2022, **122**: 108245.
- [4] Liang T F, Zhang N F, Zhang Y X, et al. Summary of research progress on application of prohibited item detection in X-ray images[J]. *Comput Eng Appl*, 2021, **57**(16): 74-82.
 梁添汾, 张南峰, 张艳喜, 等. 违禁品 X 光图像检测技术应用研究 进展综述[J]. 计算机工程与应用, 2021, **57**(16): 74-82.
- [5] Mery D, Riffo V, Zuccar I, et al. Object recognition in X-ray testing using an efficient search algorithm in multiple views[J]. *Insight-Non-Destr Test Cond Monit*, 2017, **59**(2): 85–92.
- [6] Wang Y, Zou W H, Yang X M, et al. X-ray image illegal object classification based on computer vision[J]. Chin J Liq Cryst Disp, 2017, 32(4): 287-293. 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.
- [7] Turcsany D, Mouton A, Breckon T P. Improving feature-based object recognition for X-ray baggage security screening using primed visualwords[C]//Proceedings of 2013 IEEE International Conference on Industrial Technology, 2013: 1140–1145. https://doi.org/10.1109/ICIT.2013.6505833.
- [8] Dai J F, Li Y, He K M, et al. R-FCN: object detection via regionbased fully convolutional networks[C]//Proceedings of the 30th International Conference on Neural Information Processing Systems, 2016: 379–387.
- [9] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards realtime object detection with region proposal networks[J]. IEEE Trans Pattern Anal Mach Intell, 2017, 39(6): 1137–1149.

- [10] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector[C]//Proceedings of the 14th European Conference on Computer Vision, 2016: 21–37. https://doi.org/10.1007/978-3-319-46448-0 2.
- [11] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[J]. *IEEE Trans Pattern Anal Mach Intell*, 2020, **42**(2): 318–327.
- [12] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779–788. https://doi.org/10.1109/CVPR.2016.91.
- [13] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017: 6517–6525. https://doi.org/10.1109/CVPR.2017.690.
- [14] Redmon J, Farhadi A. Yolov3: an incremental improvement[Z]. arXiv: 1804.02767, 2018. https://arxiv.org/abs/1804.02767.
- [15] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[Z]. arXiv: 2004.10934, 2020. https://arxiv.org/abs/2004.10934.
- [16] Ge Z, Liu S T, Wang F, et al. YOLOX: exceeding YOLO series in 2021[Z]. arXiv: 2107.08430, 2021. https://arxiv.org/abs/2107. 08430.
- [17] Zhu Z M, Zhu Y, Wang H R, et al. FDTNet: enhancing frequencyaware representation for prohibited object detection from X-ray images via dual-stream transformers[J]. *Eng Appl Artif Intell*, 2024, **133**: 108076.
- [18] Ahmed A, Velayudhan D, Hassan T, et al. Enhancing security in X-ray baggage scans: a contour-driven learning approach for abnormality classification and instance segmentation[J]. *Eng Appl Artif Intell*, 2024, **130**: 107639.
- [19] Dong Y S, Guo J Y, Li M Z, et al. X-ray prohibited items detection based on inverted bottleneck and light convolution block attention module[J]. *J Front Comput Sci Technol*, 2024, 18(5): 1259–1270.

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.

- [20] Zhou Y T, Cao K Y, Piao J C. Fine-YOLO: a simplified X-ray prohibited object detection network based on feature aggregation and normalized Wasserstein distance[J]. *Sensors* (*Basel*), 2024, 24(11): 3588.
- [21] Han L, Ma C H, Liu Y, et al. SC-Lite: an efficient lightweight model for real-time X-ray security check[J]. *IEEE Access*, 2024, 12: 103419–103432.
- [22] Han L, Ma C H, Liu Y, et al. SC-YOLOv8: a security check model for the inspection of prohibited items in X-ray images[J]. *Electronics*, 2023, **12**(20): 4208.
- [23] Wang A L, Yuan P F, Wu H B, et al. Improved YOLOv8 for dangerous goods detection in X-ray security images[J]. *Electronics*, 2024, **13**(16): 3238.
- [24] 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.
- [25] Chaple G, Daruwala R D. Design of Sobel operator based image edge detection algorithm on FPGA[C]//Proceedings of 2014 International Conference on Communication and Signal Processing, 2014: 788–792.

https://doi.org/10.1109/ICCSP.2014.6949951.

- [26] Lin M, Chen Q, Yan S C. Network in network[Z]. arXiv: 1312.4400, 2014. https://arxiv.org/abs/1312.4400.
- [27] Yin X Y, Goudriaan J, Lantinga E A, et al. A flexible sigmoid

function of determinate growth[J]. Ann Bot, 2003, **91**(3): 361-371.

- [28] Yang F H, Jiang R Q, Yan Y, et al. Dual-mode learning for multidataset X-ray security image detection[J]. *IEEE Trans Inf Foren Secur*, 2024, **19**: 3510–3524.
- [29] Wei Y L, Tao R S, Wu Z J, 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. https://doi.org/10.1145/3394171.3413828.
- [30] Guo S C, Jin Q Z, Wang H Z, et al. Learnable gated convolutional neural network for semantic segmentation in remote-sensing images[J]. *Remote Sens*, 2019, **11**(16): 1922.
- [31] Arora R, Basu A, Mianjy P, et al. Understanding deep neural networks with rectified linear units[C]//Proceedings of the 6th International Conference on Learning Representations, 2018.
- [32] Ding M Y, Xiao B, Codella N, et al. DaViT: dual attention vision transformers[C]//Proceedings of the 17th European Conference on Computer Vision, 2022: 74–92. https://doi.org/10.1007/978-3-031-20053-3_5.
- [33] Sun G Q, Pan Y Z, Kong W K, et al. DA-TransUNet: integrating spatial and channel dual attention with transformer U-net for medical image segmentation[J]. *Front Bioeng Biotechnol*, 2024, **12**: 1398237.
- [34] Wang H N, Cao P, Wang J Q, et al. UCTransNet: rethinking the skip connections in U-Net from a channel-wise perspective with transformer[C]//Proceedings of the 36th AAAI Conference on Artificial Intelligence, 2022: 2441–2449. https://doi.org/10.1609/aaai.v36i3.20144.
- [35] Lei Ba J, Kiros J R, Hinton G E. Layer normalization[Z]. arXiv: 1607.06450, 2016. https://arxiv.org/abs/1607.06450.
- [36] Hendrycks D, Gimpel K. Gaussian error linear units (GELUs)[Z]. arXiv: 1606.08415v4, 2023. https://arxiv.org/abs/1606.08415v4.
- [37] Chollet F. Xception: deep learning with depthwise separable convolutions[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1800–1807. https://doi.org/10.1109/CVPR.2017.195.
- [38] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions[C]//Proceedings of the 4th International Conference on Learning Representations, 2016.
- [39] Yu F, Koltun V, Funkhouser T. Dilated residual networks[C]// Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017: 636–644. https://doi.org/10.1109/CVPR.2017.75.
- [40] Chen L W, Gu L, Zheng D Z, et al. Frequency-adaptive dilated convolution for semantic segmentation[C]//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 3414–3425. https://doi.org/10.1109/CVPR52733.2024.00328.
- [41] Wahab M F, Gritti F, O'Haver T C. Discrete Fourier transform techniques for noise reduction and digital enhancement of analytical signals[J]. *TrAC Trends Anal Chem*, 2021, **143**: 116354.
- [42] 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]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 2114–2123. https://doi.org/10.1109/CVPR.2019.00222.
- [43] Li C Y, Li L L, Jiang H L, et al. YOLOv6: a single-stage object detection framework for industrial applications[Z]. arXiv:

https://doi.org/10.12086/oee.2025.240286

2209.02976, 2022. https://arxiv.org/abs/2209.02976.

- [44] Liu C, Hong Z Y, Yu W H, et al. An efficient helmet wearing detection method based on YOLOv7-tiny[C]//Proceedings of the 6th International Conference on Machine Learning and Machine Intelligence, 2023: 92–99. https://doi.org/10.1145/3635638.3635652.
- [45] Wang C Y, Yeh I H, Mark Liao H Y. YOLOv9: learning what

作者简介



崔丽群 (1969-), 女, 副教授, 硕士生导师, 主 要研究方向为智能数据处理、图像与视觉信息 计算等。 E-mail: 373604814@qq.com

【通信作者】杨莹莹 (1998-), 女,硕士研究生, 主要研究方向为图像与视觉信息计算。 E-mail: 1429337791@qq.com you want to learn using programmable gradient information[C]//Proceedings of the 18th European Conference on Computer Vision, 2024: 1–21. https://doi.org/10.1007/978-3-031-72751-1 1.

[46] Wang A, Chen H, Liu L H, et al. Yolov10: real-time end-to-end object detection[C]//Proceedings of the 38th Conference on Neural Information Processing Systems, 2024.



金海波(1983-),男,副教授,硕士生导师,主 要研究方向为计算机视觉、复杂系统可靠性分 析等。

E-mail: jinhaibo@lntu.edu.cn



吴正伟 (1998-), 男,硕士研究生,主要研究方 向为机器学习、图像与视觉信息计算。 E-mail: 1525545769@qq.com



Adaptive panoramic focusing X-ray image contraband detection algorithm

Cui Liqun, Yang Yingying^{*}, Jin Haibo, Wu Zhengwei



Overall model structure

Overview: X-ray image detection of prohibited items plays a crucial role in various fields, including public transportation, logistics, and customs inspection. It is a key technology in image processing and object detection, with the primary task of accurately identifying the category and location of prohibited items within complex background environments to ensure the safety of human life, property, and goods transportation. Unlike natural images, prohibited item images are generated using X-ray imaging technology, where the targets exhibit diverse categories and varying shapes. Moreover, these images are often affected by challenges such as target stacking, occlusion, low contrast, and complex backgrounds, making it difficult to accurately identify the correct targets, thereby leading to missed and false detections. Consequently, achieving precise identification of prohibited items and improving detection efficiency have become critical challenges and focal points in current research. To address the issues of target overlap and occlusion, difficulty in key feature extraction, and missed detection of small-sized contraband in X-ray images, this paper proposes an adaptive panoramic focus X-ray contraband detection algorithm based on the YOLOv8n model. This algorithm incorporates several novel components designed to enhance detection accuracy and efficiency. First, a foreground feature awareness module (FFAM) is proposed to significantly enhance the model's ability to represent the features of foreground targets, enabling accurate identification of contraband objects in overlapping and occluded scenes. Second, a multi-path two-dimensional information integration (MPTI) module is designed to enhance the model's ability to recognize key features by optimizing the interaction and integration of multi-scale features across both channel and spatial dimensions, enabling the extraction of more comprehensive and richer contextual information. Finally, a panoramic dynamic focus detection head (PDF_Detect) is introduced. By incorporating frequency-adaptive dilated convolutions and a dynamic focusing mechanism, the model can adaptively select the optimal receptive field size based on the frequency distribution of features. This enhances the model's ability to focus on small-sized contraband targets, effectively improving the detection of small targets and reducing both missed and false detections in complex scenes. Experiments were conducted on the public datasets SIXray and OPIXray. The experimental results show that the proposed method achieved mAP@0.5 values of 93.3% and 92.5%, representing improvements of 3.6% and 2.8% over the baseline model, respectively, and outperforming other comparative algorithms. These results demonstrate that the proposed algorithm significantly reduces missed and false detections of contraband in X-ray images, exhibiting high accuracy and robustness.

Cui L Q, Yang Y Y, Jin H B, et al. Adaptive panoramic focusing X-ray image contraband detection algorithm[J]. *Opto-Electron Eng*, 2025, **52**(4): 240286; DOI: 10.12086/oee.2025.240286

* E-mail: 1429337791@qq.com

Foundation item: National Natural Science Foundation of China (62173171), Basic Scientific Research Project of Liaoning Provincial Universities (LJKMZ20220699)

Software College, Liaoning Technical University, Huludao, Liaoning 125105, China