doi:10.12068/j.issn.1005-3026.2025.20230341

基于X-ray-RTDETR的X射线图像违禁品检测算法

李立振，马淑华，郭泽旭，车晓辰
（东北大学秦皇岛分校 控制工程学院，河北 秦皇岛 066004）

摘 要：针对X射线违禁品图像大小不一致、背景噪声高和尺度变化大导致检测精度低的问题，在RT-DETR-R18的基础上进行优化，提出了X射线图像违禁品检测算法X-ray-RTDETR. 该算法首先使用嵌入高效多尺度注意力的CSPRepResNet作为主干网络增强特征提取能力；其次，在主干网络输出的3个特征图之后引入简化的快速空间金字塔池化模块提高模型的鲁棒性和泛化能力；最后，将SPoolFormer编码器应用于语义概念更丰富的高级特征图进行尺度内特征交互. 实验结果表明，X-ray-RTDETR在PIDray测试集上检测精度达到了74.6%，比RT-DETR-R18提升了8.5%，参数量和浮点操作次数n_{FLOP}分别减少了1.67×10^{6}，2.24×10^{9}. 与当前最先进的同量级目标检测算法实验对比结果表明，X-ray-RTDETR不仅检测精度更高，而且参数量与n_{FLOP}也更少，同时推理速度在RTX2070 Max-Q GPU上达到了85.47帧/s.

关 键 词：违禁品检测；多尺度注意力；特征提取；金字塔池化；SPoolFormer编码器
中图分类号：TP 391.4 **文献标志码：**A **文章编号：**1005-3026(2025)06-0008-09

X-ray Image Prohibited Item Detection Algorithm Based on X-ray-RTDETR

LI Li-zhen, *MA Shu-hua*, *GUO Ze-xu*, *CHE Xiao-chen*
(School of Control Engineering, Northeastern University at Qinhuangdao, Qinhuangdao 066004, China. Corresponding author: LI Li-zhen, E-mail: lilizhen559@163.com)

Abstract: In response to the problem of low detection precision caused by inconsistent size, high background noise, and large-scale changes in X-ray image prohibited item, the optimization is performed based on RT-DETR-R18 and an X-ray image prohibited item detection algorithm named X-ray-RTDETR is proposed. Firstly, the algorithm employs CSPRepResNet embedded with efficient multi-scale attention as the backbone network to enhance feature extraction capabilities. Secondly, the simplified fast spatial pyramid pooling module is introduced after the three features maps output by the backbone network to improve the robustness and generalization ability of the model. Finally, the SPoolFormer encoder is applied to high-level feature maps with richer semantic concepts for intra-scale feature interaction. The experimental results show that the detection accuracy of X-ray-RTDETR achieves 74.6% on PIDray test set, surpassing RT-DETR-R18 by 8.5%, while reducing the number of parameters and n_{FLOP} by 1.67×10^{6} and 2.24×10^{9}, respectively. Compared to the state-of-the-art object detection algorithms at the same scale shows that X-ray-RTDETR not only has higher detection accuracy, but also has less number of parameters and n_{FLOP}. At the same time, its inference speed reaches 85.47 frames per second on RTX2070 Max-Q GPU.

Key words: prohibited item detection; multi-scale attention; feature extraction; pyramid pooling; SPoolFormer encoder

收稿日期：2023-12-25
基金项目：河北省自然科学基金资助项目(F2021501021).
作者简介：李立振(1996—)，男，山东聊城人，东北大学硕士研究生；马淑华(1967—)，女，河北秦皇岛人，东北大学秦皇岛分校教授.

近年来,随着我国交通行业的快速发展与人们出行频率的提高,机场、车站等交通枢纽的旅客行李安检任务越来越繁重.由于行李中的违禁物品遮挡严重、出行高峰时客流量大、安检员疲劳等因素,单纯依靠人工对X射线安检机生成的图像进行违禁品检测的方法极易造成漏检和误检,对旅客的出行安全造成影响.

基于深度学习的目标检测算法在工业领域得到了广泛应用,它凭借卷积神经网络强大的特征提取能力抽象出图像不同维度的特征并进行融合,能够有效地对图像进行分析和识别,所以研究人员开始设计基于深度学习的X射线违禁品检测算法,并取得了较好的检测结果.

基于深度学习的目标检测算法可概括为二阶段目标检测算法、一阶段目标检测算法和基于Transformer的目标检测算法.二阶段目标检测算法分为获取候选区域和进行分类识别两个阶段,以R-CNN[1-4]系列为代表,检测精度较高,但实时性较低.Gaus等[5]使用ResNet101作为主干网络的Faster R-CNN在X射线违禁品图像检测中获得了比Mask R-CNN和RetinaNet更高的检测精度;Ma等[6]以Cascade R-CNN为基准,通过引入动态可变形卷积在卷积参数中增加方向偏移和可调权重,使网络能够在X射线图像中很好地处理不同尺度和方向的各种物体;Liao等[7]通过在Faster R-CNN中引入形状引导的特征增强模块和违禁品感知模块,提高了在具有重叠噪声情况下的违禁品的检测精度.一阶段目标检测算法将目标检测定位为回归问题,可以同时预测位置和类别,以SSD[8],YOLO系列[9-13]为代表,检测速度较快,更适合部署在实时性要求较高的场景下.Wei等[14]基于空间金字塔池化和DIoU损失函数改进YOLOv3,获得了比SSD和SSD-ResNet50更先进的性能;Wang等[15]为了解决行李物品图像高度重叠和复杂背景的问题,提出了一种基于改进YOLOv5的X射线违禁品检测算法,该算法在OPIXray数据集上平均检测精度达到了91%,检测速度可达47帧/s;Liu等[16]以SSD-VGG16为基准网络,提出自动行李威胁检测网络ABTD-Net,在OPIXray和HiXray数据集上的大量实验证明了该方法的优越性.

研究人员已经使用二阶段算法和一阶段算法取得了一系列研究成果,尤其YOLO系列可以很好地实现精度与速度的平衡,成为工业部署的最佳选择,YOLOv8和YOLOv6 3.0在COCO数据集上的优异表现也为X射线违禁品检测提供了新的改进思路.经过多年的不断发展,锚框已经不再是制约YOLO检测器性能的因素,然而此类检测器会产生大量的冗余预测框,需要在后处理阶段使用非极大值抑制(non-maximum suppression, NMS)过滤.NMS的超参数和输入预测框的数量对检测器的准确性和速度有很大影响,这导致了YOLO检测器的性能瓶颈.

Carion等[17]首先提出了基于Transformer的端到端目标检测器DETR(detection transformer),后续出现一些变体,如Deformable-DETR[18],Conditional-DETR[19],DN-DETR[20],DINO[21]等对原始DETR进行了优化.DETR检测器无需设计锚框和NMS后处理,但此类检测器计算开销太大,不能作为实时检测器使用,详情见文献[22].RT-DETR是由百度飞桨团队提出的第一个实时DETR检测器,它不仅在精度和速度上优于当前的实时检测器,而且不需要后处理,推理速度保持稳定.因此,本文提出一种基于RT-DETR-R18的违禁品X射线检测算法X-ray-RTDETR,主要贡献如下:

1)针对违禁品被严重遮挡时背景噪声太大的问题,采用CSPRepResNet作为RT-DETR-R18的主干网络,增强对有效特征的提取能力,同时减少模型的参数量与计算量.

2)采用高效多尺度注意力取代通道注意力嵌入到CSPRepResNet中,进一步减少了参数,增强了有效特征的表达能力,提高了对违禁品的X射线检测精度.

3)采用3个简化的快速空间金字塔池化模块分别替代3个1×1卷积,将主干网络输出的3个不同尺度的特征图映射到同一通道数,使网络能够捕获更全局的特征以保留更多的特征信息,提高网络的鲁棒性.

4)提出一种SPoolFormer编码器模块代替原始Transformer编码器进行高级特征的尺度内特征交互,进一步提高违禁品的X射线检测精度.

1　RT-DETR-R18算法概述

RT-DETR-R18由主干网络ResNet18、高效混合编码器和带有辅助预测头的Transformer解码器组成.主干网络作为特征提取器,输出降采样倍率分别为8,16,32的3个尺度的特征图$\{S_3$,

$S_4, S_5\}$作为高效混合编码器的输入.高效混合编码器由基于注意力的尺度内特征交互(attention-based intra-scale feature interaction, AIFI)模块和基于CNN的跨尺度特征融合(CNN-based cross-scale feature fusion, CCFF)模块组成.

AIFI本质是一个Transformer编码器,由于缺乏语义概念,并且存在与高级特征交互重复和混淆的风险,低级特征的尺度内交互是效率低下且没必要的,所以AIFI只作用于特征图S_5.特征图S_5首先被展平成一维张量,经过AIFI后再还原回原来的维度,然后与S_3, S_4一起送入CCFF模块.

CCFF是一个路径聚合特征金字塔结构的跨尺度融合模块,它将3个尺度的特征图进行融合,然后转换为一个特征序列,随后使用不确定性最小化查询选择(uncertainty-minimal query selection)模块从高效混合编码器输出的特征序列中,选择固定数量的图像特征作为解码器的初始目标查询向量.最后,由带有辅助预测头的解码器迭代优化目标查询向量来生成预测框和置信度分数.

2 基于RT-DETR-R18改进的X-ray-RTDETR检测算法

X射线安检机生成的图像与自然图像不同,X射线图像成像颜色受物体密度的影响.当行李物品之间重叠严重时,非违禁品的成像就是严重的背景噪声.针对上述问题,本文主要改进了RT-DETR-R18的主干网络和高效混合编码器两部分:第一,以嵌入高效多尺度注意力的CSPRepResNet作为主干网络;第二,以简化的快速空间金字塔池化模块和SPoolFormer编码器增强高效混合编码器.改进后的网络在保证推理速度的同时实现了检测精度的大幅提升,整体网络结构如图1所示.

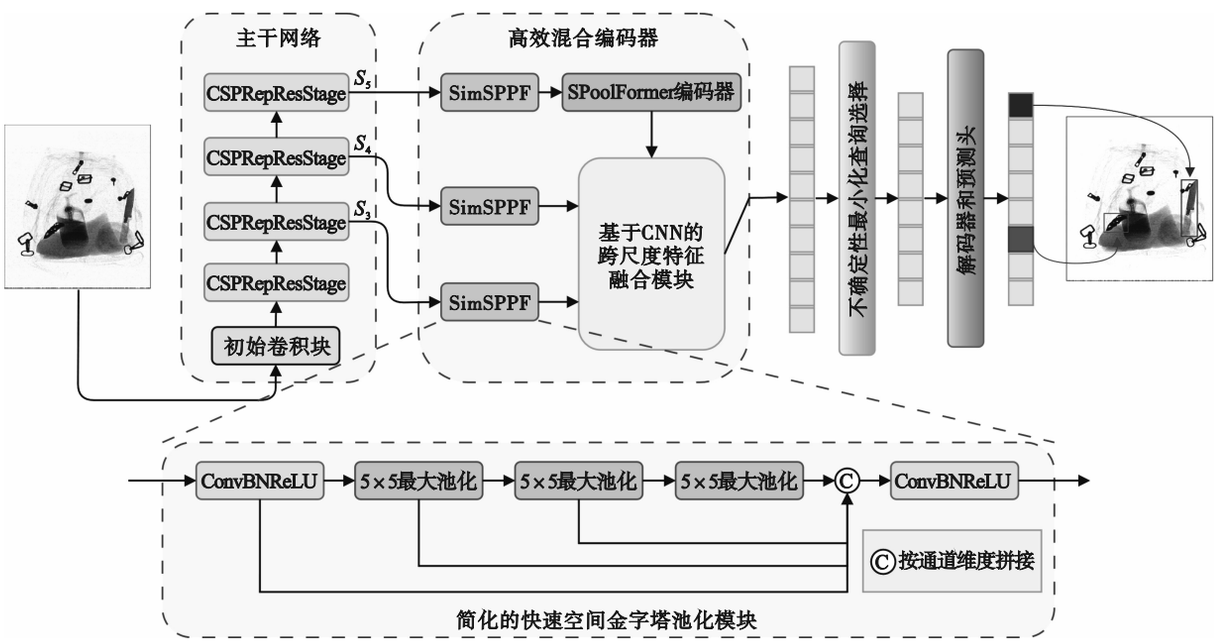

图1 X-ray-RTDETR网络结构

Fig. 1 The network architecture of X-ray-RTDETR

2.1 高效多尺度注意力

高效多尺度注意力(efficient multi-scale attention, EMA)[23]是一种新的跨空间学习的注意力机制,它能够在不降低通道维数的情况下学习有效的通道描述,并为高级特征图产生更好的像素级关注,EMA模块的整体结构如图2所示.

对于任意给定的输入特征图$\boldsymbol{X}\in\mathbf{R}^{N\times C\times H\times W}$,其中$H$和$W$分别表示特征图的空间维度,EMA将$\boldsymbol{X}$在通道维度$C$上划分为$G$组子特征图,划分后将组数$G$合并到批系数$N$,输入特征图的形状变为$[N\times G, C//G, H, W]$.而后,一方面,在1×1卷积分支上有2条平行支路分别沿2个空间方向对精确位置信息进行编码.第1条支路中采用沿水平方向的一维全局平均池化,如式(1)所示:

$$z_c(H)=\frac{1}{W}\sum_{i=1}^{W}x_c\ (H,i). \tag{1}$$

其中,x_c表示第c个通道的输入特征.第1条支路可以捕获水平方向上的长期依赖关系,并在垂直方向上保持精确的位置信息.第2条支路采用了沿垂直方向的一维全局平均池化,如式(2)所示:

$$z_c(W)=\frac{1}{H}\sum_{j=1}^{H}x_c\,(j,W). \tag{2}$$

第2条支路可以捕获垂直方向上的长期依赖关系,并在水平方向上保持精确的位置关系.将2条支路的输出与原始输入进行相乘以实现对原始输入特征的重新校准,使得EMA能够学习到低级特征图详细的特征表示.

另一方面,1×1卷积分支与3×3卷积分支的输出共同作为跨空间学习模块的输入,跨空间学习模块提供了一种在不同空间维度方向上的跨空间信息聚合方法,以实现更丰富的特征聚合.其中,二维全局平均池化操作被用来对全局空间信息进行编码,可以表示为

$$z_c=\frac{1}{H\times W}\sum_{j=1}^{H}\sum_{i=1}^{W}x_c\,(i,j). \tag{3}$$

EMA的最终输出被还原到与输入相同的形状,可以作为即插即用模块,灵活高效地嵌入到现有的网络架构中.

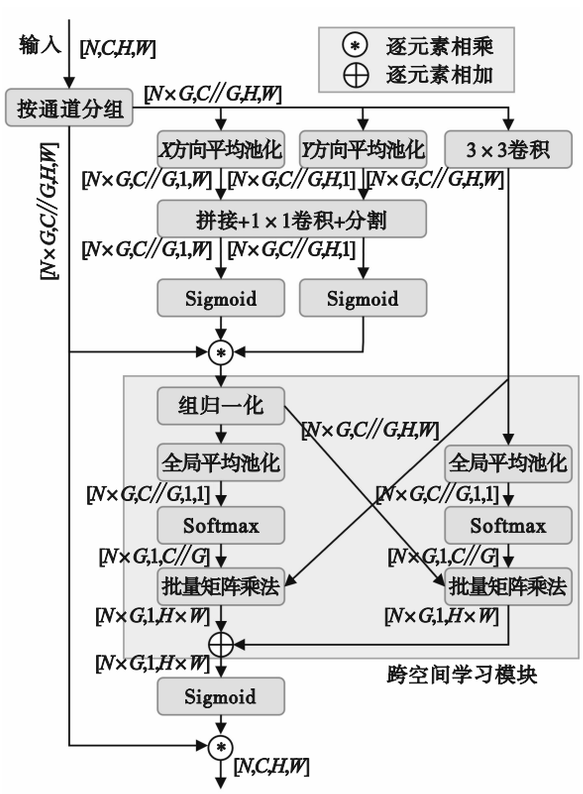

图2　EMA模块结构

Fig. 2　Structure of EMA module

2.2　嵌入EMA的CSPRepResNet主干网络

CSPRepResNet最早应用于PP-YOLOE[24]目标检测算法,整个网络包含5个阶段,每个阶段特征图的下采样率分别为2,4,8,16,32.第1个阶段为初始卷积块,包含3个卷积模块,每个卷积模块由3×3卷积、批归一化(batch normalization,BN)和Swish激活函数组成,作用是初步提取输入图像的特征,快速缩减图像的空间尺寸,同时增加通道数,以便更高效地提取高层次的语义信息.

4个后续阶段为跨阶段部分连接可重参数化残差阶段(cross-stage partially reparameterized residual stage, CSPRepResStage),其结构如图3所示,共包含N个结构可重参数化的残差块(reparameterized residual block, RepResBlock),同时在每个阶段中使用了跨阶段部分连接方式,以避免众多3×3卷积带来的大量参数和计算负担.高效压缩提取(effective squeeze and extraction, ESE)注意力被用来在每个阶段施加通道注意力.在本文中,主干网络的4个后续阶段中包含的RepResBlock个数N分别为2,4,4,2,注意力机制采用EMA代替ESE.

RepResBlock由可重参数化视觉几何组块(reparametrized visual geometry group block, RepVGGBlock)和残差连接构成,结构重参数化技术可以减少推理时的内存占用,提高推理速度,而且不影响推理性能.RepVGGBlock在训练时包含3×3卷积分支和1×1卷积分支,每个分支中都使用批归一化.在推理时利用原始3×3卷积核、1×1卷积核以及BN层的训练参数来重新构造一个3×3卷积核,这一过程提高了模型的推理效率,保证了模型的简洁性和高性能.

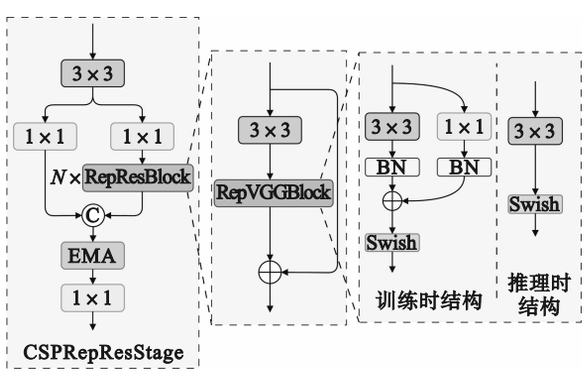

图3　CSPRepResStage结构

Fig. 3　Structure of CSPRepResStage

2.3　简化的快速空间金字塔池化结构

空间金字塔池化(spatial pyramid pooling, SPP)由He等[25]提出,它通过并联大小为5×5,9×9,13×13的最大池化层分别对输入特征图进行池化,然后将池化结果与输入进行拼接,从而将不同感受野的特征融合,有效提升了对不同尺寸目

标的检测能力.SimSPPF(simplified SPP-fast)[11]是一种简化的快速空间金字塔池化结构,其结构如图1所示,它采用串联3个尺寸为5×5的最大池化层来等效SPP的并联结构,采用ReLU替代SPPF中使用的SiLU激活函数,比SPP速度超过2.5倍,极大提高了计算效率.

RT-DETR主干网络输出的3个不同尺度的特征图进入高效混合编码器后,分别先使用一个1×1卷积将特征图的通道映射到一个较低的维度,以降低后续的计算成本.本文改用SimSPPF模块替代1×1卷积,在增加少量计算量的情况下,既能起到通道映射的作用,又能带来较高的精度收益.

2.4 SPoolFormer编码器

Transformer编码器由2个残差子块组成,第1个子块包含归一化层和在token之间混合信息的多头自注意力(multi-head self-attention,MHSA)模块,第2个子块包含归一化层和带有激活函数的多层感知机.Yu等[26]将注意力模块视为一个特定的令牌混合器(token mixer),进而将整个Transformer编码器抽象为一个通用的架构,命名为MetaFormer.通过实验证明,Transformer模型的成功很大程度上归功于MetaFormer结构.另外还提出了以简单池化操作作为token mixer的PoolFormer.在此基础上,本文提出以更为先进的软池化(SoftPool)[27]作为token mixer的SPoolFormer编码器,其结构如图4所示.

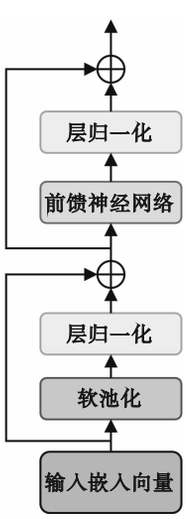

图4 SPoolFormer编码器

Fig. 4 SPoolFormer encoder

SoftPool基于softmax的加权思想,能够在保留更多原始特征的同时放大主要特征.具体地,对于给定的特征图$\boldsymbol{X}\in\mathbf{R}^{C\times H\times W}$中的局部区域$R$,为了简化表示,省略通道维度,假设池化核的大小为$k\times k$,$|R|=k^2$,局部区域$R$内的每个神经元$x_i$的权重$w_i$为

$$w_i=\frac{\mathrm{e}^{x_i}}{\sum_{j\in R}\mathrm{e}^{x_j}}. \tag{4}$$

那么局部区域R的SoftPool输出$\tilde{x}$是通过将R内所有神经元先加权再求和产生的,如式(5)所示:

$$\tilde{x}=\sum_{i\in R}w_i x_i. \tag{5}$$

平均池化为平均降低区域内所有神经元的效果,最大池化仅保留区域内值最高的神经元,具有丢失重要信息的风险.SoftPool介于两者之间,结合了两者的有利特性,既不丢弃占比较低的神经元,又使得占比较高的神经元更占主导地位.

3 实验及结果分析

3.1 数据集准备

本文基于PIDray数据集[28]进行实验,PIDray数据集使用来自不同制造商的3台安检机进行收集,因为不同机器生成的图像在目标和背景的大小和颜色上有一定差异,收集地点为机场、地铁站和火车站.数据集中共定义了12类违禁物品,分别为枪、刀、扳手、钳子、锤子、剪刀、手铐、棍棒、喷雾器、充电宝、打火机和子弹,总共包含47 677张图像,其中29 457张图像用于训练,18 220张图像用于测试.此外根据违禁品检测的难易程度,测试集又被划分为3个子集,即简单子集、困难子集和隐藏子集.简单子集指每张图像中仅包含一个违禁品,共9 482张;困难子集指每张图像中包含多个违禁品,共3 733张;隐藏子集指图像中包含故意隐藏的违禁品,共5 005张.图5是3个测试子集的标注框的宽、高与图片真实宽、高的占比的标注框数量.

本文采用线上数据增强策略,其中消融实验在训练过程中进行随机像素内容变换、随机填充、随机裁剪、随机翻转和多尺度训练,其中多尺度训练的大小为320,352,384,416,448,480,512,544和576.在对比实验中,其他先进算法的数据增强策略与原算法保持一致.在进行验证和推理时,图像的尺寸统一调整为480×480作为算法的输入.

3.2 评价指标

本文分别在3个测试子集和全部测试集上使

用平均精度(average precision, AP)指标评估模型的检测精度,以体现模型针对不同难度样本的检测能力.其中使用AP_{50}表示当交并比(intersection over union, IoU)等于0.5时12种违禁品的平均检测精度,使用AP表示IoU从0.5到0.95,以0.05为步长取值计算平均检测精度,取10次计算结果的平均值.

此外,本文还对比了模型的浮点操作次数n_{FLOP}、模型的参数量以及推理速度,以体现模型在实际部署和应用中的可行性.

3.3　消融实验

本文在PIDray数据集上以RT-DETR-R18作为基准模型设计消融实验,R18表示主干网络为ResNet18.在此基础上,将CSPRepResNet,EMA,SimSPPF,SPoolFormer编码器依次进行添加,表1展示了消融实验结果.表1数据表明,每项改进策略对于检测精度的提升都有不同程度的贡献,改进后的最终模型在全部测试集上的AP_{50}和AP较原始模型均提升了8.5%,且参数量减少了1.67×10^6,n_{FLOP}减少了2.24×10^9.

模型B将基准模型A中的主干网络由ResNet18替换为CSPRepResNet,大幅提升了在3个测试子集上的检测精度,提升最为明显的为隐藏子集,与模型A相比,AP_{50}和AP分别提升了12.5%,11.7%,这说明CSPRepResNet在提取背景复杂、噪声干扰大的隐藏违禁品方面具有较大的优势.在全部测试集上,模型B较模型A,AP_{50}和AP分别提升了5.9%,6.0%,精度提升占总提升量的70%,在所有改进中贡献最大,同时,参数量减少了1.76×10^6,n_{FLOP}减少了3.43×10^9.这些都得益于CSPRepResNet精妙的设计,CSPRepResNet不仅具有比ResNet18更深的网络层数来提取更具有表达能力的特征图,而且CSP连接方式与结构可重参数化残差块的使用,保证了参数量和计算量不会增加.

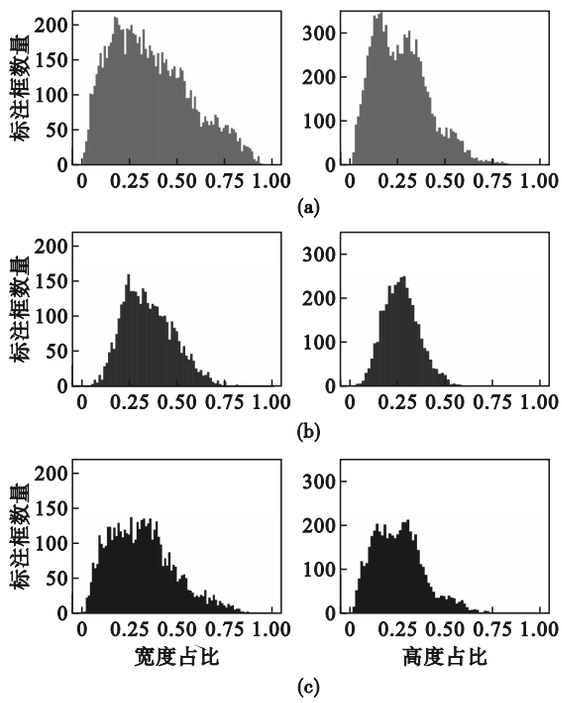

图5　3个测试子集的标注框的宽、高与图片真实宽、高的占比的标注框数量

Fig. 5　The quantity of annotated bounding boxes whose width and height ratios to actual image dimensions in three test subsets

(a)—简单子集;(b)—困难子集;(c)—隐藏子集.

表1　消融实验结果

Table 1　Results of ablation experiment

模型	基准模型及其改进	AP_{50} / %				AP / %				参数量×10^{-6}	n_{FLOP}×10^{-9}
		简单子集	困难子集	隐藏子集	全部	简单子集	困难子集	隐藏子集	全部		
A	RT-DETR-R18	86.2	86.9	61.3	78.1	76.8	72.7	48.9	66.1	20.09	29.03
B	A+(R18→CSPRepResNet)	89.2	89.0	73.8	84.0	80.5	75.1	60.6	72.1	18.33	25.60
C	B+(ESE→EMA)	90.0	90.2	75.6	85.3	81.1	75.9	61.2	72.8	17.95	25.76
D	C+(Conv(1×1)→SimSPPF)	91.0	90.5	76.6	86.0	82.4	76.5	62.6	73.8	18.69	26.79
E	D+(MHSA→SoftPool)	91.4	91.2	77.3	86.6	82.9	77.4	63.4	74.6	18.42	26.79

模型C在模型B的主干网络中嵌入EMA代替ESE注意力,与模型B相比,模型C的计算量n_{FLOP}增加0.16×10^9,但是AP_{50}和AP分别提升了1.3%和0.7%,并且参数量减少了0.38×10^6,每个测试子集的精度也都得到了一定提升.实验表明,EMA与ESE注意力相比,在几乎不带来额外计算开销的情况下,能更有效地突出特征图的重点特征,进一步提升主干网络的特征提取能力.

模型D在模型C的基础上,将S_3,S_4,S_5特征图后面用于通道映射的1×1卷积替换为SimSPPF模块,与模型C相比,对于AP_{50},简单子集与隐藏子集均提升1.0%,而困难子集仅提升了0.3%,对

于AP,简单子集与隐藏子集分别提升了1.3%和1.4%,困难子集仅提升了0.6%.由图5可以发现,简单子集和隐藏子集比困难子集含有更多较小的目标,这说明SimSPPF在模型D上对于提升小目标的检测能力更有效.

模型E在模型D的基础上将原始的Transformer编码器替换为SPoolFormer编码器,保持了n_{FLOP}不变,参数减少0.27×10^{6},在3个子测试集上的检测精度均获得小幅提升,在全部测试集上较模型D,AP_{50}和AP分别提升了0.6%和0.8%,分别达到了86.6%与74.6%.

图6给出了模型A与模型E的部分检测结果的可视化对比.如图6a所示,模型A对于正面或角度变化不大而成像的违禁品能够较准确地检测出来,但是违禁品图像的外观形状一旦发生较大的尺度变化,便极容易发生误检和漏检,比如剪刀被识别成了刀具、扳手未被检测出.另外,当违禁品被遮挡时,如行李包中物品杂乱或违禁品被使用线圈缠绕的情况,模型A便失去了检测效果,比如被电线缠绕的剪刀、锤子和子弹均未被检测出.经过改进的模型E(本文模型)能较好地应对上述问题,如图6b所示,不仅能够将上述复杂环境中的违禁品检测出来,而且具有较高的置信度分数,比模型A具有更强的鲁棒性,能够达到令人满意的检测效果.

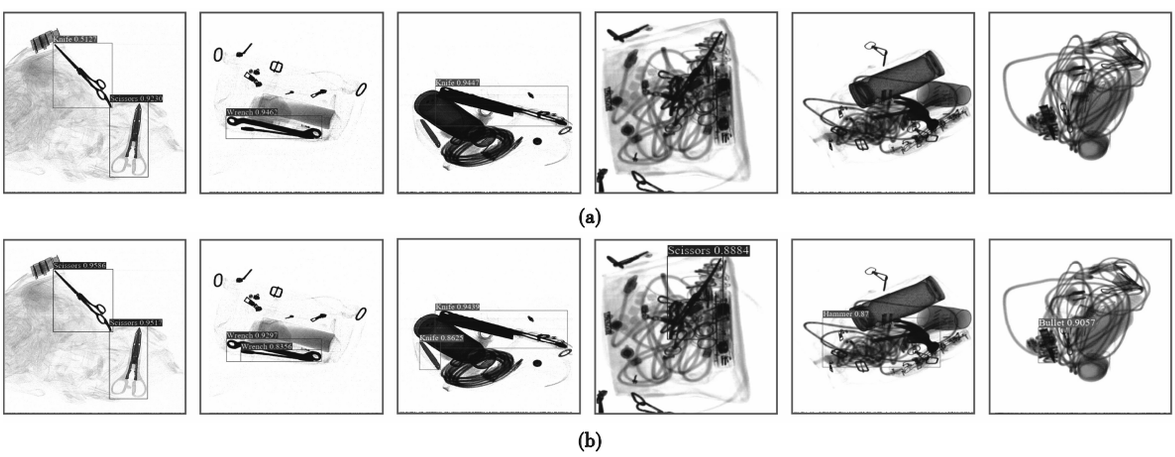

图6 模型A与改进的模型E的部分检测结果的可视化对比

Fig. 6 Visual comparison of partial detection results between Model A and improved Model E

(a)—模型A(RT-DETR-R18)部分检测结果;(b)—改进的模型E(X-ray-RTDETR)部分检测结果.

3.4 与其他SOTA检测器对比实验

为了进一步验证本文算法的优越性,将其在PIDray数据集上与当前先进的同量级目标检测算法进行了比较.其中推理速度测试环境基于Windows10操作系统,NVIDIA GeForce RTX2070 Max-Q GPU, CUDA 10.2和cuDNN 7.6.5.除X-ray-RTDETR外,其他检测器需要计入NMS处理时间,考虑到NMS的执行时间会受输入图像的影响,本文选择困难子集作为推理速度测试的基准数据集,然后计算每秒能够检测的图片数量来衡量推理速度.实验结果如表2所示.

表2 与其他先进检测算法实验结果对比

Table 2 Comparison of experimental results with other advanced detection algorithms

模型	AP_{50}/%				AP/%				参数量×10^{-6}	n_{FLOP}×10^{-9}	推理速度
	简单子集	困难子集	隐藏子集	全部	简单子集	困难子集	隐藏子集	全部			帧·s^{-1}
YOLOv5-m	83.4	86.0	61.4	76.9	69.2	64.8	44.5	59.5	20.92	27.15	54.95
YOLOv7-l	85.3	87.6	72.6	81.8	75.8	71.2	57.6	68.2	36.54	59.13	58.14
YOLOv8-m	88.0	88.1	73.7	83.3	79.8	75.8	61.7	72.4	25.85	44.44	59.17
PP-YOLOE-Plus-m	90.0	89.1	70.7	83.3	81.0	75.4	57.8	71.4	23.52	27.83	56.18
Gold-YOLO-m	87.2	89.6	72.2	83.0	77.3	73.5	57.1	69.3	41.28	49.12	79.36
YOLOv6-m 3.0	90.1	90.8	75.2	85.4	81.0	76.7	61.8	73.2	34.81	48.18	90.09
X-ray-RTDETR	91.4	91.2	77.3	86.6	82.9	77.4	63.4	74.6	18.42	26.79	85.47

对比实验结果可以发现,X-ray-RTDETR不仅检测精度高于一系列先进目标检测算法,而且参数量与n_{FLOP}也最小.YOLOv5-m与本文算法具有相近的参数量和n_{FLOP},但是其检测精度和推理速度表现均不佳.YOLOv7-l,YOLOv8-m和PP-YOLOE-Plus-m检测精度相对较高,但它们的推理速度仍小于60帧/s,与本文算法有大于25帧/s的差距.

其中YOLOv6-m 3.0与本文算法相比最具有竞争力,其在全部测试集上AP_{50}和AP分别达到了85.4%和73.2%,推理速度达到了90.09帧/s.本文算法的推理速度稍慢于YOLOv6-m 3.0,但也达到了85.47帧/s,而且本文算法的AP_{50}和AP分别比YOLOv6-m 3.0高1.2%和1.4%,参数量和n_{FLOP}分别减少了一半,而且本文算法不需要考虑NMS超参数对性能的影响.总的来说,针对X射线图像违禁品检测任务,与其他同量级的SOTA检测器相比,本文提出的X-ray-RTDETR的综合性能具有显著优越性.

4　结　　语

为了实现对具有复杂背景噪声的X射线违禁品图像实时、高精度的检测,本文通过使用嵌入EMA注意力的CSPRepResNet替换RT-DETR-R18的主干网络,引入SimSPPF和SPoolFormer编码器改进高效混合编码器模块,提出了基于X-ray-RTDETR的X射线图像违禁品检测算法.实验结果表明,X-ray-RTDETR显著提升了违禁品的检测精度,且推理速度完全满足实时性的要求,能够为整个检测系统的其他环节作出决策留有足够的冗余时间.算法的整体设计兼顾了检测精度与推理速度,综合性能超过其他诸多先进算法,具有较好的工程实用价值.

参考文献:

[1] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 580-587.

[2] Girshick R. Fast R-CNN[C]// 2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 1440-1448.

[3] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137-1149.

[4] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN[C]// 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 2980-2988.

[5] Gaus Y F A, Bhowmik N, Breckon T P. On the use of deep learning for the detection of firearms in X-ray baggage security imagery[C]// 2019 IEEE International Symposium on Technologies for Homeland Security (HST). Woburn: IEEE, 2019: 1-7.

[6] Ma C J, Zhuo L, Li J F, et al. Prohibited object detection in X-ray images with dynamic deformable convolution and adaptive IoU[C]// 2022 IEEE International Conference on Image Processing (ICIP). Bordeaux: IEEE, 2022: 3001-3005.

[7] Liao H Y, Huang B, Gao H X. Feature-aware prohibited items detection for X-ray images [C]// 2023 IEEE International Conference on Image Processing (ICIP). Kuala Lumpur: IEEE, 2023: 1040-1044.

[8] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector [C]// 2016 European Conference on Computer Vision (ECCV). Berlin: Springer, 2016: 21-37.

[9] Redmon J, Farhadi A. YOLOv3: an incremental improvement [EB/OL]. (2018-04-08) [2023-10-19]. https://arxiv.org/abs/1804.02767.

[10] Bochkovskiy A, Wang C Y, Liao H M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23) [2023-10-19]. https://arxiv.org/abs/2004.10934.

[11] Li C Y, Li L L, Jiang H L, et al. YOLOv6: a single-stage object detection framework for industrial applications [EB/OL]. (2022-09-07) [2023-10-19]. https://arxiv. org/abs/2209.02976.

[12] Li C Y, Li L L, Geng Y F, et al. YOLOv6 v3.0: a full-scale reloading [EB/OL]. (2023-01-13) [2023-10-19]. https://arxiv.org/abs/2301.05586.

[13] Wang C Y, Bochkovskiy A, Liao H M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [C]// 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver: IEEE, 2023: 7464-7475.

[14] Wei Y J, Dai C, Chen M S, et al. Prohibited items detection in X-ray images in YOLO network [C]// 2021 26th International Conference on Automation and Computing (ICAC). Portsmouth: IEEE, 2021: 1-6.

[15] Wang Z S, Zhang H Y, Lin Z B, et al. Prohibited items detection in baggage security based on improved YOLOv5 [C]// 2022 IEEE 2nd International Conference on Software Engineering and Artificial Intelligence (SEAI). Xiamen, 2022: 20-25.

[16] Liu W, Sun D G, Wang Y, et al. ABTD-Net: autonomous baggage threat detection networks for X-ray images [C]// 2023 IEEE International Conference on Multimedia and Expo (ICME). Brisbane: IEEE, 2023: 1229-1234.

[17] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with Transformers [C]// 2020 European Conference on Computer Vision (ECCV). Cham: Springer, 2020: 213-229.

[18] Zhu X Z, Su W J, Lu L W, et al. Deformable DETR: deformable Transformers for end-to-end object detection [EB/OL]. (2021-03-18) [2023-10-19]. https://arxiv.org/abs/2010.04159.

[19] Meng D P, Chen X K, Fan Z J, et al. Conditional DETR for fast training convergence [C]// 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2021: 3651-3660.

(下转第25页)