

DOI:10.19651/j.cnki.emt.2211230

基于可变形卷积与注意力机制的 X光安检违禁品检测*

粟兴旺 王小明 黄金玻 许茹玉 吴琳
(西华大学计算机与软件工程学院 成都 610039)

摘要: X光安检违禁品检测被广泛应用于维护公共交通安全和人身安全。针对X光图像中违禁品存在形状尺度多变、重叠遮挡严重等问题,提出一种结合可变形卷积与注意力机制改进的YOLOv5s模型用于违禁品检测。首先在主干网络中引入可变形卷积,通过学习采样偏移量来适应物体的不同形变,增强空间特征信息提取能力;其次利用混合卷积注意力模块加强模型对检测目标的感知能力,抑制无关背景干扰;然后构造通道引导的空洞空间金字塔模块,获取更加准确的全局上下文信息,提高模型对重叠遮挡目标的识别能力;最后采用CARAFE算子代替最近邻插值,在上采样过程中充分利用内容信息,提高模型检测精度。在SIXray_OD和OPIXray数据集上实验结果显示,所提出模型的mAP@0.5相较于原YOLOv5s分别提高了2.1%和1.8%,达到了90.6%和90.0%。与现有诸多先进算法相比,具有较好的检测精度与实时性。

关键词: X光图像;违禁品检测;可变形卷积;注意力机制;上采样

中图分类号: TP391.4 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Prohibited items detection based on deformable convolution and attention mechanism in X-ray security inspection

Su Xingwang Wang Xiaoming Huang Jinbo Xu Ruyu Wu Lin
(School of Computer and Software Engineer, Xihua University, Chengdu 610039, China)

Abstract: Prohibited items detection in X-ray security inspection is widely used to maintain public traffic safety and personal safety. In order to solve the problems of variable shape and scale, severe overlap and occlusion in X-ray images, an improved YOLOv5s model combining deformable convolution and attention mechanism is proposed for prohibited items detection. Firstly, deformable convolution is introduced into the backbone network to enhance spatial feature information extraction by learning sampling offsets to adapt to different deformations of objects. Secondly, the mixed convolution attention module is used to enhance the model's ability to perceive the detected target and suppress irrelevant background interference. Then a channel-guided atrous space pyramid module is constructed to obtain more accurate global contextual information and improve the model's ability to identify overlapping occlusion targets. Finally, the CARAFE operator is used to replace the nearest neighbor interpolation to make full use of the content information in the upsampling process and improve model's detection accuracy. The experimental results on the SIXray_OD and OPIXray datasets show that the model's mAP@0.5 is 2.1% and 1.8% higher than the original YOLOv5s, reaching 90.6% and 90.0%, respectively. Compared with many existing advanced algorithms, it has better detection accuracy and real-time performance.

Keywords: X-ray images; prohibited items detection; deformable convolution; attention mechanism; upsampling

0 引言

随着公共交通枢纽人群密度的不断增加,X光安检在

保障公共安全方面变得越来越重要。X光安检旨在使用扫描无损的识别行李中的物品,确保其中没有违禁品,以减少犯罪问题、恐怖主义、有害动植物入侵等危险^[1]。目前,

收稿日期:2022-08-30

* 基金项目:四川省自然科学基金(2022NSFSC0533)、西华大学研究生创新基金(yccj2021030)项目资助

这项工作主要依靠安检人员在短时间内观察X光图像来人工识别违禁品^[2]。然而行李中杂乱摆放的违禁品总是被其他物品遮挡,且只占整张X光图像中的小部分,给人工识别带来了很大困难。同时检查结果的好坏也受检查员经验与疲劳程度的影响,很容易由于人为原因造成违禁品的误检和漏检。因此,研究一种快速、准确、自动的检测方式尤为重要。

近年来,随着深度学习的快速发展,许多计算机视觉任务取得了巨大的进步,如人脸检测^[3]、行为识别^[4]、医学图像处理^[5]等。研究员们同样也将目光投到了X光安检违禁品检测领域。其中,Mery等^[6]评估了10种基于视觉池袋模型、稀疏表示、深度学习和经典模式识别方案的X光违禁品检测方法,发现深度学习方法效果最好。Miao等^[7]提出一种类平衡分层细化模型,使用不同尺度特征过滤无关信息并对违禁品进行识别分类。Wei等^[8]提出一种去遮挡注意力模块,利用违禁品的边缘信息和材料信息生成注意力图和特征图来进行检测。Gu等^[9]提出使用特征增强模块提高特征提取能力,同时利用多尺度融合方式得到更准确的感兴趣区域,提高了X光安检图像中违禁品检测的准确性和鲁棒性。上述方法在较大程度上提高了检测精度,为X光安检违禁品检测的发展奠定了基础,但真实场景下违禁品检测中目标形状尺度多变、重叠遮挡严重、复杂背景干扰等仍然是目前研究中急需解决的关键问题,尤其是X光安检违禁品图像本身特点带来的挑战依然十分突出,目前的检测精度仍然无法满足实际应用的要求。一方面,违禁品尺度和形状变化大,上下文信息分布不均匀,而常规卷积由于具有固定的采样位置,使其无法适应实际目标的感受野以及不能灵活处理上下文分布的高度变化,最终造成一些重要的上下文信息被忽略,削弱了特征提取能力;另一方面,同种材质的物品在图像中呈现出相同或近似的颜色,在重叠遮挡时极易造成目标与背景的混淆,甚至无法区分。当这些物品经过卷积层处理后,会得到相似的特征响应,造成识别与定位精度的降低。

针对上述问题,本文以YOLOv5s^[10]网络为基础框架,提出一种基于可变形卷积与注意力机制改进的X光安检违禁品检测模型。该模型能自适应学习不同尺度与形变目标的特征,同时增强目标特征在通道与空间维度上的表示能力,抑制冗余和混乱的背景干扰,突出目标与背景之间的区别,能在很大程度上改善因目标尺度多变、形状多样、遮挡重叠引起的检测精度低问题。主要创新点如下:

- 1) 在主干网络中引入可变形卷积(deformable convnets v2, DCNv2)^[11],通过自学习生成不规则卷积核来适应目标几何形状,增强模型对形状尺度多变目标的检测能力。
- 2) 设计通道注意力引导的空洞空间金字塔模块(channel attention-atrous spatial pyramid module, CA-ASPM)替换原有的快速空间金字塔池化模块(spatial pyramid pooling-fast, SPPF),减少池化层造成的信息损失,进一步扩大感受野并关注更重要的特征,聚合更多全局上下文信息,提高模型处理重叠遮挡目标的能力。
- 3) 利用一种新颖的内容感知特征重组算子(content-aware reassembly of features, CARAFE)^[12]替换颈部网络中的最近邻插值进行上采样,以深入探索多级特征的语义和空间信息,减少有用信息丢失。
- 4) 在网络中引入和差注意力机制(common-and-differential attention network, CDANet)^[13],削弱无用背景干扰,提升模型对目标重要特征的关注度。

1 YOLOv5 目标检测模型

YOLOv5 作为 YOLO 系列的较新版本,在精度与速度上都得到了很大的提升,是一种适合工业落地的模型。YOLOv5 自发布以来,更新了很多版本,本文选择最新的 6.1 版本为研究对象。YOLOv5 根据模型深度和参数量的不同可分为 5 种网络模型,基于模型大小与精度的考虑,本文选择 YOLOv5s 作为网络改进研究的基础模型。

YOLOv5s 包括 4 个部分:输入端、主干网络、颈部网络和预测端,如图 1 所示。

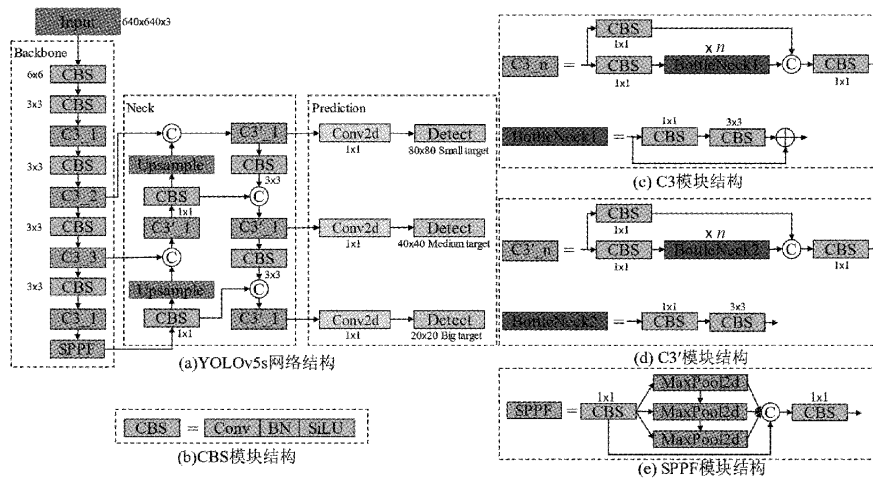


图1 YOLOv5s 网络结构

输入端包括数据增强、图像尺寸处理和自适应锚框计算。其中 mosaic 增强采用随机组合拼接 4 张图片的方式来丰富数据集中的背景与目标的数量,提高了模型的检测精度与鲁棒性;图像尺度处理自动将输入图片统一成固定大小;自适应锚框通过判断预测锚框与真实锚框的偏差来生成适合数据集目标的锚框参数。

主干网络包括 CBS(Conv-BN-SiLU)、C3、SPPF 模块。6.0 及以后的版本将原有的 Focus 模块用一个等效的 CBS 模块替换。CBS 模块由卷积层、批量归一化层和 SiLU 激活函数层组成,旨在提取特征;C3 中采用残差结构来减小模型参数并提取更细粒度的特征;SPPF 模块用于提取多尺度特征,是由 SPP 模块改进而来的,使每一次池化的输出作为下一次池化的输入,相对于 SPP 模块而言效果相同,速度更快。

颈部网络使用了 FPN+PAN 的特征金字塔结构进行多尺度融合。FPN 实现了语义信息从高到低的传递,而 PAN 实现了定位信息从低到高的传递。

预测端通过 3 个具有不同大小特征图的检测层检测不同大小的目标,生成原始图像中目标的预测边界框和类别,并进行标记。

2 改进 YOLOv5s 的 X 光违禁品检测模型

2.1 动机与整体网络模型结构

YOLOv5s 模型在目标检测中表现的很优秀,检测精度与时效性都达到了较高水平,但其是基于常规自然图像检测设计的,并未考虑到目标的几何多变性与背景复杂性。直接使用 YOLOv5s 检测 X 光违禁品会出现多形变目标与重叠遮挡严重目标检测精度低的现象,如图 2 所示。

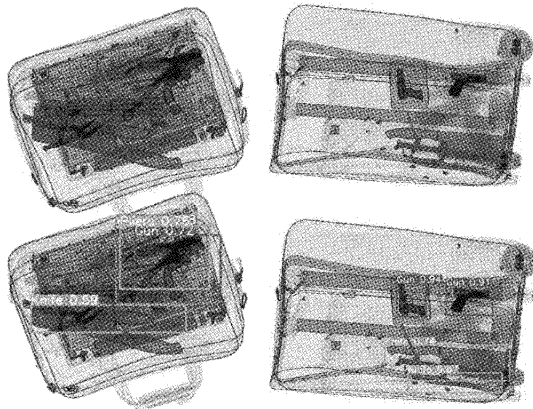


图 2 YOLOv5s 模型检测效果

为此,本文通过改进 YOLOv5s 网络,设计一种更加适合于 X 光违禁品检测的模型。整体模型如图 3 所示,改进部分由红线框出。首先,在主干网络中引入可变形卷积,设计 DBS(DCNv2-BN-SiLU)模块,自适应学习违禁品的几何多变性,增强模型对多形态违禁品目标特征的提取能力;其次,设计通道注意力引导的空洞空间金字塔模块

CA-ASPM 替换原有的 SPPF 模块,在减少信息损失的同时扩大感受野,聚合更多全局上下文信息,提高模型对多尺度和遮挡严重目标的处理能力;然后,使用 CARAFE 算子实现多级特征之间的上采样,充分利用多级特征的语义与空间信息;最后,引入和差注意力机制 CDANet,削弱无用背景的干扰,使模型更加关注违禁品目标。

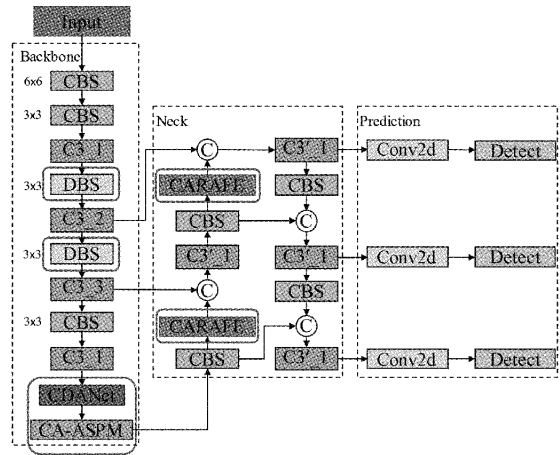


图 3 改进后的网络模型结构

2.2 可变形卷积模块设计

实际 X 光安检图像中,违禁品呈现出尺度多变、姿态各异、几何形变程度大等特点,而常规的 CNN 模型在提取特征时并没有适应几何形变的机制。最关键的原因在于普通的卷积运算本身具有固定的几何结构,并且由其堆积而成的卷积网络的几何结构也是固定的,因此削弱了网络模型对几何形变程度大的物体的识别能力。为了缓解这一问题,本文引入带有调制机制的可变形卷积 DCNv2 来更好的提取图像特征。通过在卷积核中的每个采样位置添加一个偏移变量,实现在当前位置附近的随机采样,从而不再局限于普通的卷积核在规则的网格点上采样。并且利用调制机制来调整不同空间位置输入特征的幅度,学习每一个采样点的权重,从而减小特征受到无关图像内容的影响,提高模型专注目标图像区域的能力。标准卷积与可变形卷积采样点的对比如图 4 所示。

可变形卷积公式如下:

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

式中: $x(p)$ 和 $y(p)$ 分别表示输入特征图 x 和输出特征图 y 中位置 p 处的特征; K 表示卷积核的采样位置个数; w_k 与 p_k 分别表示第 k 个采样位置的权重和预先指定的偏移量; Δp_k 和 Δm_k 分别表示第 k 个采样位置的可学习偏移量和调制标量; $p + p_k + \Delta p_k$ 通常是小数,与特征图上实际点不对应,需要通过双线性插值转换。

本文将上述可变形卷积引入到 YOLOv5s 的主干网络中,增加几何形变特征提取能力。由于 1×1 的可变形卷积不具有改变感受野的能力,且在计算采样点偏移量时不

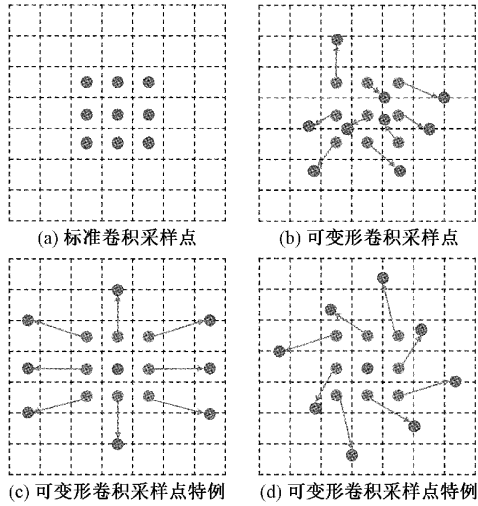


图4 标准卷积与可变形卷积采样对比

稳定,故本文只研究 3×3 的可变形卷积。与此同时,可变形卷积插入到网络的哪个位置并没有固定的理论说明,故本文分别将 CBS 模块与 C3 模块中的标准 3×3 卷积用同样大小的可变形卷积代替,形成模块 DBS 与 DC3,如图 5(a)、(b)所示。实现自学习生成形变卷积核来更好地提取适应违禁品目标形状的特征,从而提高检测精度,并在后续试验中讨论改进模块及其组合方式对模型的影响,结果如表 1 所示,选择最优方案。

2.3 CA-ASPM 模块设计

原始 YOLOv5s 模型中 SPPF 模块使用相同大小的最大池化层获取不同感受野特征,实现全局特征与局部特征的融合,丰富了特征表达能力,能在一定程度上缓解 X 光安检违禁品尺度多变以及重叠遮挡造成的检测精度低的问题。但其中的最大池化层在扩大感受野并保留显著特征的同时带来了局部信息丢失。基于这个问题,本文考虑使用空洞卷积来替换原始的最大池化层,实现在保持特征图分辨率不变和不增加卷积核参数的情况下,有效扩大感受野,同时对每一个像素点进行计算,减小信息损失。近年来有许多学者研究了空洞卷积,文献[14]提出空洞空间

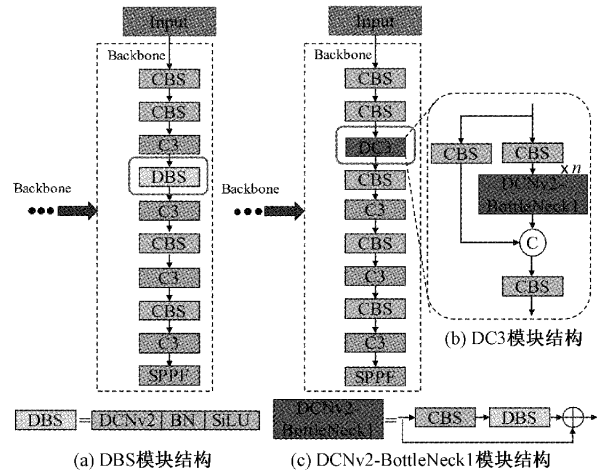


图5 DBS 与 DC3 的结构

金字塔池化(atrous spatial pyramid pooling, ASPP)结构,通过组合 3 个具有不同空洞率的平行空洞卷积,实现多个比例获取图像上下文信息和多尺度物体信息;文献[15]提出感受野模块(receptive field block, RFB),使用多个空洞卷积来模拟人类视觉的感受野,加强特征的提取能力。

ASPP 能在增大感受野的同时获取多尺度特征信息,但直接引入到模型主干网络中替换 SPPF 模块会存在两个问题:一方面,ASPP 中 3×3 卷积的空洞率分别为 6、12、18,相应的理论感受野大小分别为 13、25、37,相比于 YOLOv5s 主干网络顶层 20×20 大小的输出特征图来说并不合适,会引入大量噪声;另一方面,ASPP 只是同等对待各分支中不同空洞卷积层处理的结果,并没有区分融合后不同通道的重要程度。为解决上述问题,本文借鉴 SPPF 与 ASPP 的思想,使用与 SPPF 同样感受野大小的空洞卷积代替最大池化层,并使用通道注意力机制来引导特征融合,为不同通道分配不同的权重,增大重要特征通道的权重,缩小无关甚至有负面影响的通道权重。此外,由于空洞卷积中的空洞会导致关键局部信息的丢失,因此引入残差连接来保持原始特征图中丰富的局部细节信息,最终构造 CA-ASPM 模块,如图 6(c)所示。

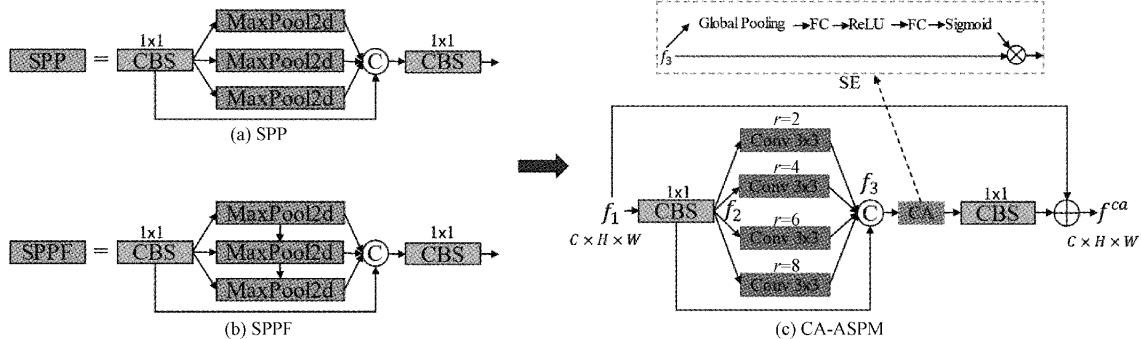


图6 SPP、SPPF、CA-ASPM 的结构对比

鉴于模型主干网络顶层特征图 20×20 的大小,本文

使用空洞率分别为 2, 4, 6, 8 的 3×3 空洞卷积,相应的感受

野大小为 5, 9, 13, 17, 其中空洞率为 8 的空洞卷积层能进一步扩大感受野, 获取更多全局上下文信息, 缓解目标遮挡问题。同时去掉 ASPP 原有用于解决空洞率太大, 容易丢失信息而添加的 Pooling 层, 以及变相继续增大空洞率的 1×1 卷积层。具体来说, 首先将从主干网络得到的输入特征 f_1 通过 1×1 的 CBS 模块进行通道数减半, 减少后续计算量后, 得到特征 f_2 。然后经过 4 个 3×3 空洞卷积处理, 与原 f_1 进行通道融合后, 得到多尺度融合特征 f_3 。接着使用通道注意力 SENet^[16] 对 f_3 进行通道过滤, 分配不同通道的权重。再经过 1×1 的 CBS 模块对通道进行调整, 恢复到原特征 f_1 的通道数。最后通过残差结构与原始特征图 f_1 相加融合, 保留原特征 f_1 中丰富的局部信息, 得到最后的输出特征 f^{ca} 。

2.4 CARAFE 上采样

原始 YOLOv5s 网络中通过 FPN 结构来完成语义信息从高到低的传递, 通过特征融合实现多尺度目标的检测。其中的多尺度特征融合采用最近邻插值上采样来完成特征图大小的统一, 这种方式会带来两种限制: 一方面,

插值上采样方式仅仅使用特征图的空间信息, 而不考虑其语义信息, 造成目标位置与噪声位置同时进行上采样; 另一方面, 插值上采样方式的感受野通常较小, 容易造成全局特征信息使用不足。这些限制不利于充分利用多级特征, 从而导致检测结果不理想。为了解决这些问题, 本文在模型中采用一种轻量且高效的上采样算子 CARAFE 替换最近邻插值, 通过内容感知的方式重组具有相似语义信息的特征点, 并在较大的感受野中聚合特征, 实现适应特定于实例内容的上采样, 提高检测精度。

CARAFE 上采样由两部分组成: 上采样内核预测模块和内容感知重组模块, 如图 7 所示。首先使用上采样内核预测模块进行通道压缩, 减少输入特征通道数, 再将压缩后的特征映射进行内容编码并根据每个目标位置的内容预测一个重组内核, 最后使用内容感知重组模块将重组内核与原始特征图对应区域作点积, 完成上采样。CARAFE 上采样使得局部相关区域内的信息得到了更好的关注, 使上采样后的特征图具有更有效的语义信息, 从而能够在 FPN + PAN 结构的特征信息传递的过程中达到更好的效果。

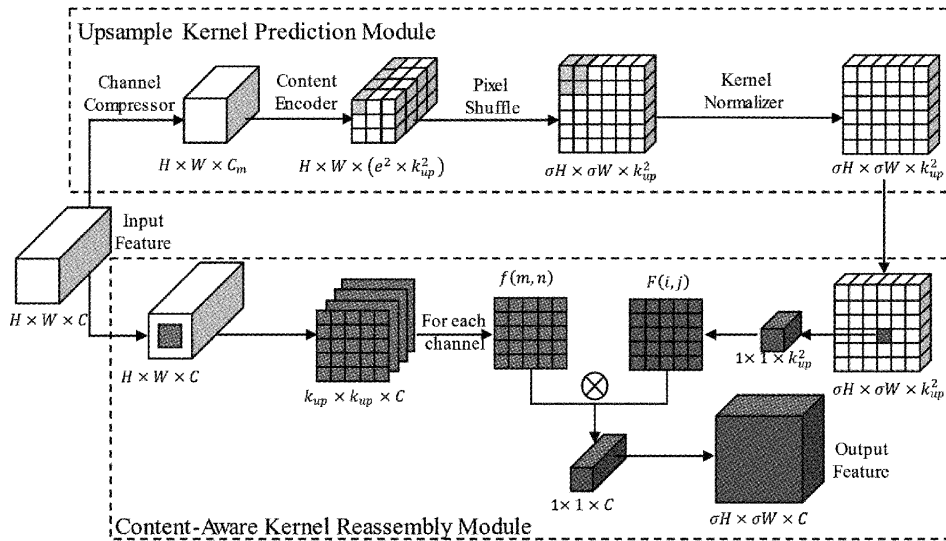


图 7 CARAFE 上采样算子工作原理

2.5 融合注意力模块

由于 X 光安检机的成像特点, 图像中的物体呈现半透明的状态, 且失去了纹理和光泽特性。在高度重叠遮挡的行李物品中, 伴随同类材料物品的重叠, 违禁品与背景很难区分。经过卷积网络提取特征后, 会带来大量冗余信息, 导致目标边界的错误定位, 造成检测精度丢失。基于这个问题, 本文引入一种和差注意力机制 CDANet 来增大违禁品目标的特征权重, 减小无关物体以及背景的干扰。

CDANet 由通道注意力模块和空间注意力模块并行组成, 两个互补的注意力模块分别沿着通道和空间维度来细化输入特征, 并都使用 CD(common-and-differential) 操作来处理 1×1 卷积降维后的特征, 达到初步抑制冗余和混

淆的背景干扰, 刻画物体和背景之间边界的突变和细节特征, 进一步强调对目标的响应并精确定位的作用。其中的通道注意力模块利用输入特征中各个通道之间的相关性, 使用加权乘法来聚合所有通道的特征, 计算相应通道权重, 突出更加重要的特征通道。空间注意力模块则是利用空间上下文信息的依赖性, 通过加权乘法获取空间注意力图来细化空间位置特征, 获取更加精确的位置信息。各注意力子模块结构如图 8 所示。

和差注意力 CDANet 能在很大程度上削弱无用背景干扰, 缓解由于冗余和混乱的背景所导致的目标边界错误定位问题。但将其插入到网络的哪一部分并没有明确的理论说法。因此本文基于 2.2~2.4 节改进后的模型

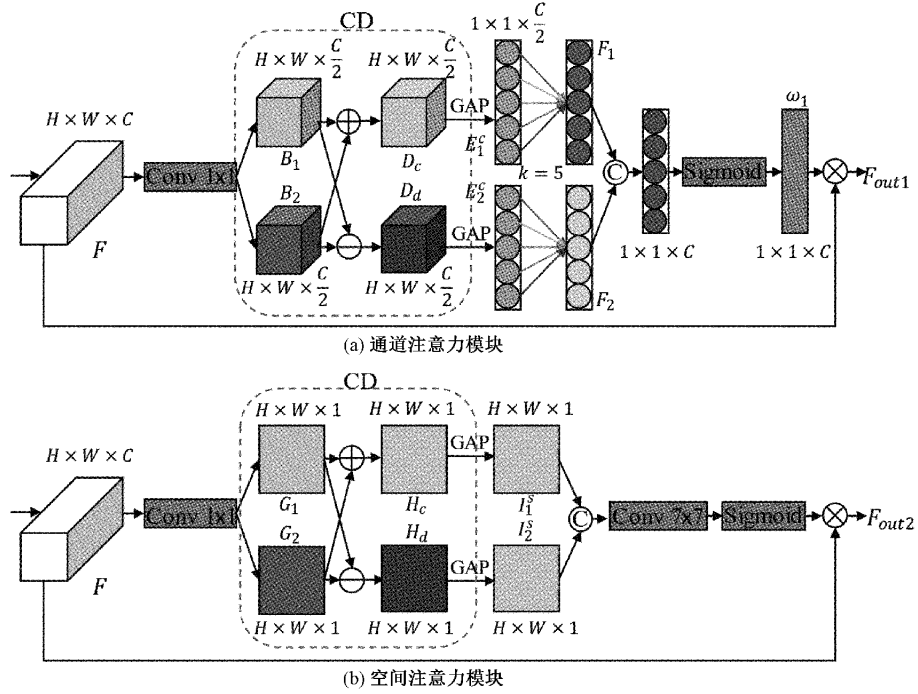


图8 CDANet中的注意力子模块

YOLOv5s-DAC (DBS + CA-ASPM + CARAFE), 将 CDANet 模块分别插入到模型的主干网络、颈部网络、预测端, 通过实验探究这 3 种使用位置对检测精度的影响。具体的, 第 1 种在主干网络末尾 C3 模块与 CA-ASPM 模块

之间添加; 第 2 种在颈部网络的所有 C3 模块中添加; 第 3 种在 3 个预测层的头部添加。分别命名为 YOLOv5s-DAC-A、YOLOv5s-DAC-B、YOLOv5s-DAC-C, 具体位置如图 9 所示。

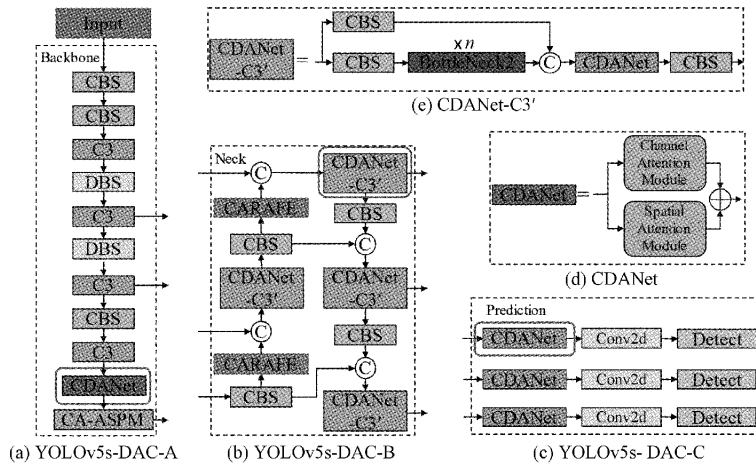


图9 CDANet 在 YOLOv5s-DAC 中的插入位置

3 实验与分析

3.1 实验环境

本文实验在 Ubuntu 18.04.6 LTS 系统下完成。所采用的 GPU 为 NVIDIA TITAN X(Pascal), 显存为 12 GB, CPU 为 Intel(R) Xeon(R) E5-2620 v4。深度学习框架为 Pytorch1.8.1, 编程语言为 Python3.7.13, 同时安装了 CUDA10.1 和 cuDNN7.6.5 以支持 GPU 的使用。

3.2 实验数据集及设置

SIXray_OD(SIXray_Object Detection); 其源于用于分类的大规模 X 光安全检查数据集 SIXray^[7]。SIXray 数据集收集了 1 059 231 张双能 X 光行李图片, 其中存在违禁品的有 8 929 张, 共 6 个类别: 枪、刀、扳手、钳子、剪刀、锤子。由于锤子样本数量远远低于其他 5 类, 不足以对检测模型进行全面训练, 故本文进行数据筛选, 保留其余 5 类共 8 718 张图片重新进行人工标注, 制作成实验所需的数

据集 SIXray_OD。其中包含枪的图片有 2 936 张、刀 1 561 张、扳手 2 266 张、钳子 3 951 张、剪刀 1 159 张。本文设置训练集、验证集、测试集的比例为 8 : 1 : 1。

OPIXray^[8]数据集包含 8 885 张专业安检人员手动标注的 X 光安检图片,其中违禁品共 5 类,分别为折叠刀、直刀、剪刀、工具刀、多功能刀。训练集、验证集、测试集的比例同样为 8 : 1 : 1。

3.3 实验参数设置

实验设置基本采用 YOLOv5s 官方推荐的参数设置,在数据预处理阶段采用 mosaic 数据增强,使用自适应锚框计算,输入图像尺寸为 640×640,训练阶段采用随机梯度下降策略 SGD 优化网络参数,初始学习率设为 0.01,学习率动量为 0.937,权重衰减为 0.000 5, batchsize 为 16, CPU 线程数为 8, epoch 为 200。

3.4 性能评价指标

本文使用平均精确率 AP(average precision)、平均精确率均值 mAP(mean average precision)、召回率 Recall、占用存储空间大小、模型参数量、浮点运算次数 GFLOPs (giga floating-point operations per second)、帧率 (frames per second, FPS) 作为 X 光安检图像违禁品检测结果的评价指标。其中 AP 是以精确率 P (precision) 和召回率 R (recall) 分别作为纵、横坐标构成 P-R 曲线围成的面积,表达式如下:

$$Precision = \frac{N_{TP}}{N_{TP} + N_{FP}} \quad (2)$$

$$Recall = \frac{N_{TP}}{N_{TP} + N_{FN}} \quad (3)$$

$$AP = \int_0^1 P(R) d(R) \quad (4)$$

式中: N_{TP} 表示预测正确的正样本数; N_{FP} 表示本来为负样本,但被错误预测为正例的样本数; N_{FN} 表示本来是正样本,但被错误预测为负例的样本数。

mAP 是每个目标类别的 AP 值求和,再按照类别数量进行算数平均后的精确率,表达式如下:

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (5)$$

式中: N 表示总的类别数。

3.5 实验结果分析

1) 可变形卷积模块实验分析

为了验证可变形卷积在模型中的有效性以及探究具体的使用位置,本文以原始 YOLOv5s 模型为基准,将 3×3 的可变形卷积分别替换主干网络中 CBS 模块与 C3 模块中的 3×3 标准卷积,形成了 2.2 节中的 2 个新模块 DBS 与 DC3,并讨论两个模块各自的使用个数以及组合方式对整体模型的影响,在 SIXray_OD 数据集上进行对比实验,实验结果如表 1 所示。

表 1 可变形卷积模块验证试验

算法模型	体积/MB	参数量/M	计算量/GFLOPs	mAP@0.5/%	FPS
YOLOv5s	14.4	7.03	16.0	88.5	102
YOLOv5s-DBS-1	14.4	7.04	16.3	88.7(+0.2)	88
YOLOv5s-DBS-2	14.4	7.05	16.2	89.7(+1.2)	94
YOLOv5s-DBS-3	14.4	7.06	16.1	89.4(+0.9)	98
YOLOv5s-DBS-4	14.5	7.10	16.1	88.8(+0.3)	102
YOLOv5s-DBS-2,3	14.5	7.08	16.3	89.8(+1.3)	95
YOLOv5s-DBS-2,3,4	14.6	7.14	16.4	88.6(+0.1)	93
YOLOv5s-DBS-all	14.6	7.15	16.8	89.5(+1.0)	83
YOLOv5s-DC3-1	14.4	7.04	16.4	88.6(+0.1)	90
YOLOv5s-DC3-2	14.4	7.06	16.4	88.6(+0.1)	92
YOLOv5s-DC3-3	14.6	7.13	16.3	89.4(+0.9)	93
YOLOv5s-DC3-4	14.5	7.10	16.1	88.7(+0.2)	100
YOLOv5s-DC3-all	14.8	7.23	17.2	89.3(+0.8)	77
YOLOv5s-DBS-2,3+DC3-3	14.7	7.17	16.7	89.5(+1.0)	87

表 1 中 YOLOv5s-DBS- i 表示将原主干网络中第 i 个 3×3 大小的 CBS 模块用 DBS 模块替换。同理, YOLOv5s-DC3- i 表示将原主干网络中第 i 个 C3 模块用 DC3 模块替换。实验结果表明,在模型主干网络的不同位置使用可变形卷积,模型的检测精度都能得到不同程度提升,证明了 DBS 模块与 DC3 模块的有效性,说明可变形卷积能够显著

增强对形状变化大的违禁品的特征提取能力,提高检测精度。其中 YOLOv5s-DBS-2,3 模型(简称 YOLOv5s-D)即使用 DBS 模块替换主干网络中第 2 和第 3 个 3×3 大小的 CBS 模块后达到了最高检测精度,在参数量只增长 0.05 M,计算量只增大 0.3GFLOPs 的情况下, mAP@0.5 达到了 89.8%,相比于 baseline 提高了 1.3%,并且 FPS 下降很小。

同时发现可变形卷积使用个数不是越多越好,且使用在中层特征层时效果更好,这可能是由于可变形卷积同时需要定位信息与语义信息,而这两种信息在中层特征层中比较均衡。

2)CA-ASPM 模块实验分析

为了验证 CA-ASPM 模块的有效性,本文以原始

YOLOv5s 模型为基准,将主干网络末尾的 SPPF 模块替换为 CA-ASPM 模块,其余部分保持不变。改进后的模型命名为 YOLOv5s-A,并将其与原 YOLOv5s 模型以及使用 ASPP 模块的 YOLOv5s 模型在 SIXray_OD 数据集上进行对比实验,实验结果如表 2 所示。

表 2 CA-ASPM 模块验证实验

算法模型	体积/MB	参数量/M	计算量/GFLOPs	mAP@0.5/%	FPS
YOLOv5s-SPPF	14.4	7.03	16.0	88.5	102
YOLOv5s-SPP	14.4	7.03	16.0	88.5	102
YOLOv5s-ASPP	18.3	9.00	17.6	88.8(+0.3)	100
YOLOv5s-CA-ASPM	19.8	9.73	18.1	89.5(+1.0)	98

实验结果表明,本文设计的 CA-ASPM 模块在少量增加体积、参数量、计算量以及 FPS 轻微下降的情况下,提高了 1% 的检测精度,证明了该模块的有效性。说明 CA-ASPM 模块能在减少信息损失的同时,进一步扩大感受野并关注更重要的特征,达到聚合更多全局上下文信息并提高模型处理重叠遮挡目标能力的作用。

3)上采样方式实验分析

为了验证 CARAFE 上采样方式在模型中的有效性,同时也为了同其他上采样方式进行全面比较,本文以 YOLOv5s 模型为基准,将模型中原始的最近邻插值上采样改为 CARAFE 上采样以及双线性插值和反卷积两种常用的上采样方式,其余部分保持不变。将这些改动了的模型与原始 YOLOv5s 模型在 SIXray_OD 数据集上进行对比实验,实验结果如表 3 所示。

表 3 上采样方式验证实验

算法模型	体积/MB	参数量/M	计算量/GFLOPs	mAP@0.5/%	FPS
YOLOv5s	14.4	7.03	16.0	88.5	102
YOLOv5s-BILINEAR	14.4	7.03	16.0	88.7(+0.2)	102
YOLOv5s-TRANPOSED	14.4	7.04	16.0	88.8(+0.3)	65
YOLOv5s-CARAFE	14.7	7.17	16.5	89.1(+0.6)	93

实验结果表明,CARAFE 上采样相比于其他上采样,虽然带来了少量参数量与计算量,但却能获得最高精度,得到了 0.6% 的提升,证明了 CARAFE 上采样能在保持轻量化的同时能更好的聚合特征信息,完成适应于目标内容的上采样。

4)注意力模块实验分析

为了验证在模型中融合 CDANet 注意力模块的有效性以及探究其具体的使用位置,本文以经过上述方法改进后的模型 YOLOv5s-DAC(DBS+CA-ASPM+CARAFE)为基准,将 CDANet 模块分别插入到主干网络、颈部网络、预测端,形成了如图 9 中的 3 种模型,然后在 SIXray_OD 数据集上进行实验对比,结果如表 4 所示。

表 4 注意力模块验证实验

算法模型	体积/MB	参数量/M	计算量/GFLOPs	mAP@0.5/%	FPS
YOLOv5s-DAC	20.1	9.91	18.9	90.3	87
YOLOv5s-DAC-A	21.5	10.04	19.1	90.6(+0.3)	79
YOLOv5s-DAC-B	23.0	12.70	19.7	89.7(-0.6)	74
YOLOv5s-DAC-C	20.5	10.08	19.3	89.6(-0.7)	77

实验结果表明,只有 YOLOv5s-DAC-A 模型即把 CDANet 注意力模块插入到主干网络的 CA-ASPM 模块前的方式才能提高检测精度,如图 9(a)所示。在参数量只增长 0.13 M,计算量只增大 0.2GFLOPs 的情况下,mAP@0.5 上涨了 0.3%,证明了在此处使用注意力模块的有效性。

5)消融实验

为了验证本文所提出的 4 种改进方法对模型性能的影响,即 D(DBS)、A(CA-ASPM)、C(CARAFE)、M(CDANet),以原始 YOLOv5s 模型为基准,在 SIXray_OD 数据集上进行消融实验,实验结果如表 5 所示。其中 G、K、W、P、S 分别代表 SIXray_OD 数据集中的枪、刀、扳手、钳子、剪刀。

表 5 模型在 SIXray_OD 数据集上的消融实验结果

算法模型	AP/%					mAP@ 0.5/%	mAP@ 0.5;0.95/%	体积/ MB	参数量/ M	计算量/ GFLOPs	FPS	训练 时间/h
	G	K	W	P	S							
YOLOv5s (baseline)	97.9	83.2	87.1	92.2	82.3	88.5	60.5	14.4	7.03	16.0	102	6.3
YOLOv5s-D	98.3	84.6	87.8	92.6	85.8	89.8(+1.3)	62.6	14.5	7.08	16.3	95	7.7
YOLOv5s-A	97.8	84.8	86.7	92.7	85.6	89.5(+1.0)	61.4	19.8	9.73	18.1	98	7.0
YOLOv5s-C	98.2	84.8	86.8	91.8	84.0	89.1(+0.6)	61.4	14.7	7.17	16.5	93	6.9
YOLOv5s-M	98.2	84.2	87.4	92.2	84.1	89.2(+0.7)	61.0	15.8	7.73	16.2	98	6.4
YOLOv5s-DA	98.1	86.0	87.9	93.3	85.4	90.1(+1.6)	62.5	19.9	9.78	18.4	88	8.9
YOLOv5s-DAC	98.4	86.3	87.8	93.6	85.6	90.3(+1.8)	63.1	20.1	9.91	18.9	87	9.1
YOLOv5s-DACM	98.5	86.4	88.1	94.0	86.1	90.6(+2.1)	63.3	21.5	10.04	19.1	79	9.2

实验结果表明,本文所提出的 4 种改进方法相比于原始 YOLOv5s 模型而言,检测精度都得到了不同程度的提升,其中使用可变形卷积带来的提升最大,mAP@0.5 和 mAP@0.5;0.95 分别提高了 1.3% 和 2.1%,剪刀这类违禁品检测精度上升的尤为明显,AP 提高了 3.5%,充分证明了 DBS 模块能对尺寸、形状变化大的目标提取有效特征。后续结合 CA-ASPM 模块、CARAFE 上采样、CDANet 混合注意力模块,在带来少量参数量与计算量的同时,模型检测精度得到进一步提升,最终的 YOLOv5s-DACM 模型相对于原模型来说每一类违禁品的检测精度都上升明显,且 mAP@0.5 和 mAP@0.5;0.95 分别提高了 2.1% 和 2.8%,虽然检测速度有所下降,但仍然能满足实际工程应用的需求。

6) 对比实验

为了进一步验证本文所提出的最终模型 YOLOv5s-DACM 的先进性与泛化性,分别在 SIXray_OD 和 OPIXray 数据集上,将其与一些目前主流的通用目标检测模型^[17-22]以及一些改进的违禁品检测模型^[8,23-25]进行对比,结果如表 6 所示。实验结果表明,本文提出的模型相较于主流的通用目标检测模型以及一些改进的违禁品检测算法更具有优势,在两个 X 光违禁品数据集上分别都达到了 90.6% 和 90.0% 的最高精度,且模型大小以及检测速度好于大部分算法,充分证明的本文模型的优越性。为更直观的评价算法改进前后的效果,分别将改进前后算法在两个数据集上进行可视化对比,如图 10 所示。从中可以看出在复杂的 X 光行李背包背景下,本文提出的算法能有效提升重叠遮挡目标、形状尺度变化大的目标的检测效果。

表 6 不同检测模型性能对比

算法模型	体积/MB	SIXray_OD		OPIXray	
		mAP@0.5/%	FPS	mAP@0.5/%	FPS
Faster R-CNN ^[17]	159.5	83.9	21	81.9	24
SSD ^[18]	200.1	74.4	54	70.9	60
RetinaNet ^[19]	80.0	73.7	48	73.4	55
CenterNet ^[20]	124	82.5	27	74.1	30
YOLOv3 ^[21]	123.5	79.2	56	78.2	64
YOLOv4 ^[22]	100.6	83.1	64	78.9	75
YOLOv5s ^[10]	14.4	88.5	102	88.2	117
YOLOv5m ^[10]	42.2	90.0	58	89.6	65
LightRay ^[23]	87.3	87.3	42	—	—
EAOD-Net ^[24]	—	85.9	—	85.8	—
SSD+DOAM ^[8]	—	—	—	74.6	—
FCOS+LIM ^[25]	—	—	—	83.1	—
本文	21.5	90.6	79	90.0	89

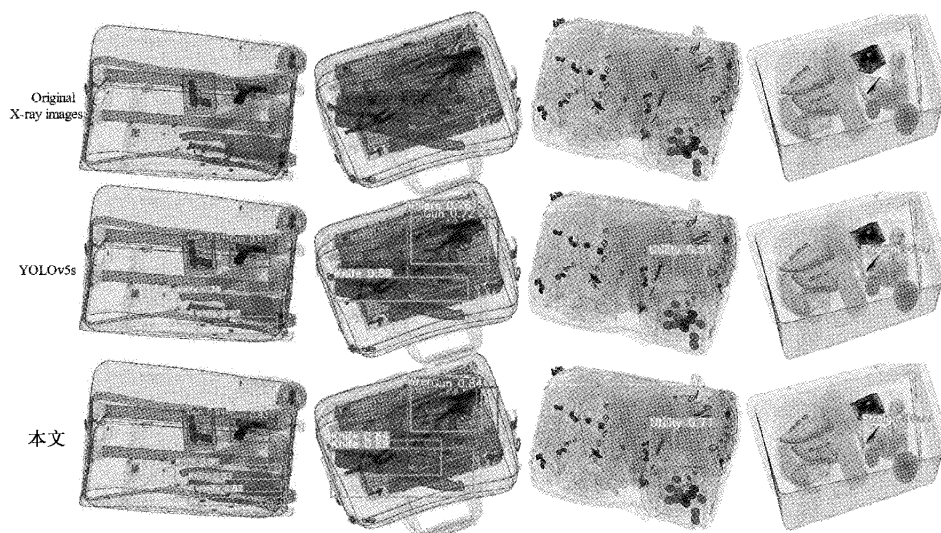


图10 算法改进前后的检测效果对比

4 结 论

针对X光安检图像中违禁品尺度多变、形状各异、重叠遮挡严重的问题,本文提出一种改进的违禁品检测模型。在YOLOv5s网络模型的基础上,使用可变形卷积、和差混合注意力机制、CA-ASPM模块、CARAFE上采样算子来联合优化整体模型。在两个大型的X光安检数据集SIXray_OD、OPIXray上实验,结果表明改进后的模型优于诸多主流通用检测算法和改进的X光违禁品检测算法,具有较高的检测精度与检测速度,更适合X光违禁品检测。下一步工作是在保证检测精度与速度不变的前提下,使用轻量化方法降低参数量与计算量,便于嵌入式落地。

参考文献

- [1] ZENTAI G. X-ray imaging for homeland security[J]. International Journal of Signal and Imaging Systems Engineering, 2010, 3(1): 13-20.
- [2] 邢琛聪, 吕淑静, 吕岳. 基于小样本学习的X光图像违禁物品识别[J]. 电子测量与仪器学报, 2021, 35(3): 204-210.
- [3] 许春和, 宋领赞. 基于时序融合策略的深度卷积神经网络动态表情识别[J]. 国外电子测量技术, 2021, 40(10): 151-157.
- [4] 罗旭飞, 崔敏, 张鹏. 基于骨骼的双支融合模型的人体行为识别[J]. 电子测量技术, 2022, 45(11): 140-146.
- [5] 于宁波, 刘嘉男, 高丽, 等. 基于深度学习的膝关节MR图像自动分割方法[J]. 仪器仪表学报, 2020, 41(6): 140-149.
- [6] 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.
- [7] 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 the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 2119-2128.
- [8] 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]. Proceedings of the 28th ACM International Conference on Multimedia, 2020: 138-146.
- [9] GU B Z, GE R J, CHEN Y, et al. Automatic and robust object detection in X-ray baggage inspection using deep convolutional neural networks[J]. IEEE Transactions on Industrial Electronics, 2021, 68(10): 10248-10257.
- [10] ULTRALYTICS. YOLOv5 [DB/OL]. (2020-06-26) [2022-08-26]. <https://github.com/ultralytics/YOLOv5>.
- [11] ZHU X Z, HU H, LIN S, et al. Deformable convnets v2: More deformable, better results[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 9308-9316.
- [12] WANG J Q, CHEN K, XU R, et al. CARAFE: Content-aware reassembly of features[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 3007-3016.
- [13] WANG Y, LI Y, GUO X H, et al. CDANet: Common-and-differential attention network for object

- detection and instance segmentation [J]. Pattern Recognition Letters, 2022, 158: 48-54.
- [14] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.
- [15] LIU S T, HUANG D, WANG Y H. Receptive field block net for accurate and fast object detection [C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 385-400.
- [16] 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.
- [17] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.
- [18] LIU W, ANQUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. European Conference on Computer Vision. Springer, Cham, 2016: 21-37.
- [19] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980-2988.
- [20] ZHOU X Y, WANG D Q, KRÄHENBÜHL P. Objects as po-ints [EB/OL]. (2019-04-08) [2022-08-26]. <https://arxiv.org/abs/1904.07850>.
- [21] REDMON J, FARHADI A. YOLOv3: An incremental improvement [EB/OL]. (2018-04-08) [2022-08-26]. <https://arxiv.org/abs/1804.02767>.
- [22] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection [EB/OL]. (2020-04-23) [2022-08-26]. <https://arxiv.org/abs/2004.10934>.
- [23] REN Y, ZHANG H G, SUN H X, et al. LightRay: Lightweight network for prohibited items detection in X-ray images during security inspection [J]. Computers and Electrical Engineering, 2022, 103: 108283.
- [24] MA C J, ZHUO L, LI J F, et al. EAOD-Net: Effective anomaly object detection networks for X-ray images [J]. IET Image Processing, 2022, 16(10): 2638-2651.
- [25] 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]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 10923-10932.

作者简介

粟兴旺, 硕士研究生, 主要研究方面为图像处理、计算机视觉。

E-mail: 1361429236@qq.com

王晓明(通信作者), 博士, 副教授, 主要研究方向为模式识别、机器学习、图像处理、计算机视觉。

E-mail: wangxmwm@163.com

黄金玻, 硕士研究生, 主要研究方向为机器学习、模式识别。

许茹玉, 硕士研究生, 主要研究方向为计算机视觉、图像处理。

吴琳, 硕士研究生, 主要研究方向为计算机视觉、图像处理。