

Yolo-C: 基于单阶段网络的 X 光图像违禁品检测

郭守向, 张良*

中国民航大学天津市智能信号与图像处理重点实验室, 天津 300300

摘要 为了检测 X 光图像中的违禁物品, 提出一种基于深度学习的单阶段双网络目标检测算法。在单阶段目标检测网络 YOLOv3 的基础上, 结合复合骨干网络的思想, 构建了 Yolo-C 目标检测网络。Yolo-C 的骨干网 DarkNet-C 由辅助骨干网络 DarkNet-A 和引导骨干网络 DarkNet-L 组成。DarkNet-A 中的各个特征层与 DarkNet-L 中对应的上一层级进行特征级联, 然后向下一层级传播, 最终得到表征图像信息的特征图。为提升对小目标的检测性能, 引入特征增强模块(FAB)。对级联后的特征图进行特征融合, 以增强特征的非线性表达能力, 达到特征平滑的目的。此外, 采用迁移学习和数据增强的方法训练网络, 提升了网络的鲁棒性。该算法在 SIXray_OD 数据集上平均精度均值(mAP)达到了 73.68%, 检测速度达 $40 \text{ frame} \cdot \text{s}^{-1}$ 。实验结果表明, Yolo-C 在检测 X 光图像领域, 有效提高了对多类违禁物品的检测精度, 且满足检测的实时性要求。

关键词 图像处理; 违禁品检测; 单阶段双网络; Yolo-C; 特征增强模块; 迁移学习

中图分类号 TP391

文献标志码 A

doi: 10.3788/LOP202158.0810003

Yolo-C: One-Stage Network for Prohibited Items Detection Within X-Ray Images

Guo Shouxiang, Zhang Liang*

Tianjin Key Laboratory of Advanced Signal & Image Processing, Civil Aviation University of China,
Tianjin 300300, China

Abstract To detect prohibited items in X-ray images, this study proposed a one-stage dual-network object detection algorithm based on deep learning. Based on the one-stage object detection algorithm YOLOv3 and combined with the idea of a composite backbone network, a Yolo-C object detection network is developed. The backbone of Yolo-C (DarkNet-C) consists of an assistant backbone network (Darknet-A) and a lead backbone network (Darknet-L). Each feature layer of the DarkNet-A is cascaded by feature with the upper feature level corresponding to DarkNet-L and then propagated to the next feature level. Finally, a feature map representing image information is obtained. The feature enhancement block (FAB) is introduced to improve detection performance of small object. Feature fusion is performed on the cascaded feature maps to enhance the nonlinear expression ability of features and achieve the purpose of feature smoothing. Besides, transfer learning and data enhancement was adopted to train the network and improve its robustness. The mAP in the SIXray_OD dataset is 73.68%, and detection speed is $40 \text{ frame} \cdot \text{s}^{-1}$. In the X-ray image detection field, Yolo-C has effectively improved the detection accuracy of different prohibited items and met the real-time requirements of detection.

Key words image processing; prohibited items detection; one-stage dual-network; Yolo-C; feature enhancement block; transfer learning

OCIS codes 100.2980; 100.3008; 100.4996; 110.2960; 110.7440

收稿日期: 2020-08-17; 修回日期: 2020-09-03; 录用日期: 2020-09-09

基金项目: 国家自然科学基金(61179045)、民航安全能力建设项目(20600523)

* E-mail: l-zhang@cauc.edu.cn

1 引言

如今, X 光安检仪已被广泛应用于世界上大大小小的安检场所。由于 X 光的透视性和假彩性, 在对旅客行李进行安全检查时, 不需要行李开箱即可进行检查, 这为安检人员提供了便利, 同时大大加快了安检速度。然而随着全球化和电子商务的发展, 旅客运输量和客货运输量不断增多, 安检人员的安检压力越来越大。近年来, 有些研究人员将太赫兹技术应用于隐藏违禁品的检测^[1], 但该技术由于目前无法对违禁品准确标注而未得到广泛的应用。针对该热点问题, 人们开始依靠计算机辅助或代替人工检测, 以加快过检速度、减轻安检人员的工作压力。

Singh 等^[2]回顾了信号处理和模式识别在分析安全设备产生的数据和图像方面的应用。Mery 等^[3]制作了共 19407 幅 X 光图像的公共数据集 GDXray, 包含铸件、焊缝以及行李等 5 大类安检图像, 用于安检人员训练、图像分析和计算机视觉算法研究。Roomi 等^[4]采用模糊 K 近邻(Fuzzy KNN)算法识别行李中的手枪, 采用图像分割法、应用特征提取方法(包括形状上下文描述符和 Zernike 矩)从图像中提取感兴趣的物体, 最后使用模糊 KNN 算法判断图像中是否含有手枪。Kundegorski 等^[5-6]根据视觉词袋(BOVW)的方法对行李的 X 光图像进行图像分类, 先通过一系列特征点检测器和描述符得到图像中违禁物品的特征, 形成“词袋”, 再由支持向量机(SVM)和随机森林算法对“词袋”里的特征进行分类。上述的传统特征提取方法, 检测精度较低, 检测速度慢。

近年来, 深度学习越来越受到人们关注, 尤其自从 AlexNet^[7] 出现起, 卷积神经网络迅猛发展。Jaccard 等^[8]利用 VGG Net^[9]对货物的 X 光图像进行分类, 完成了对违禁物品的分类任务, 而且根据图像的乘法特性实现了数据增强。但是该算法结构简单, 对于复杂背景的 X 光图像, 分类精度大大降低。Akçay 等^[10-11]采用迁移学习的方式, 探讨了基于 AlexNet 神经网络的图像分类方法和基于滑动窗口和基于区域的目标检测方法, 研究了 R-CNN^[12]和 R-FCN^[13]目标检测网络对 X 光图像进行违禁品筛查的适用性。Cañizares 等^[14]提出一种形式化的分布式检测框架——FORTIFIER。该方法集成了多种分布式图像检测算法, 采用有限状态机(FSM)组成网络框架, 用于检测行李中的违

禁物品。该网络结构设计复杂, 识别违禁品的类别数量受到 FSM 数量的限制, 难以与实际安检场景相结合。

Akçay 等^[15]考虑到绝大部分安检图像都是正常 X 光图像(不含有违禁品), 即异常图像(含有违禁品)较少, 提出一种新的异常检测模型——GANormaly, 利用条件生成对抗网络, 联合学习高维图像空间的生成和潜在空间的推理, 使用大量未含有违禁品的 X 光图像训练网络模型, 然后用异常图像进行异常检测。该网络模型虽然得到较好的实验结果, 但实际安检时行李的外观、大小多种多样, 且行李内物品随意摆放, 网络很难提取到正常图像的特征, 从而导致检测结果具有较高的误检率。Gaus 等^[16]使用一种双卷积神经网络结构, 用于复杂 X 光图像中的异常检测。该网络先检测图像中可能含有的目标, 再经过异常检测模型判断检测结果是否正确。这种目标检测与异常检测相结合的网络虽然检测精度有所提高, 但是检测的速度十分缓慢。Miao 等^[17]基于实际安检场合采集的数据, 制作了含有 1059231 幅 X 光图像数据集 SIXray, 并且采用 ResNet50^[18]骨干网络, 针对样本类别数量不均衡的问题提出类分层细化(CHR)算法, 实现了对图像的多标签分类。多标签分类任务只能检测出 X 光图像中是否含有违禁品, 不能标出违禁品的位置。

综上所述, 考虑到 X 光图像假彩性、重叠性造成的图像混乱、物体难以检测的问题, 本文在单阶段检测网络 Yolov3 的基础上, 结合复合骨干网的思想, 提出了 Yolo-C 目标检测网络, 以满足检测准确性和实时性要求。

2 单阶段双网络算法(Yolo-C)

2.1 Yolov3

Yolov3^[19]是典型的单阶段检测网络。该网络实现了端对端训练, 首先将图像经过 DarkNet-53 骨干网提取特征, 根据类似特征金字塔(FPN)^[20]的多尺度思想, 将提取的特征分为三个不同尺寸的特征图, 实现了对不同大小目标的检测。此外, Yolov3 网络根据 Faster R-CNN 中 anchor 的思想, 通过预先设好的不同形状、尺寸的锚框生成候选框, 然后对候选框内的物体进行分类和定位, 提高了目标的定位精度, 最后对得到的检测框利用非极大值抑制算法除去重复的检测框。由于 Yolov3 有较高的检测性能且网络结构简单, 很多研究人员对其进行了改进和优化。李成跃等^[21]对 Yolov3

进行剪枝,得到轻量化检测网络——DS-YOLO,使得检测速度更快。郭进祥等^[22]利用空洞卷积和线性衰减置信度对 YOLOv3 进行优化,提升了网络对小目标检测效果。图 1 为 YOLOv3 网络结构图,图中 NMS 表示非极大值抑制算法,用以删除重复的检测框。

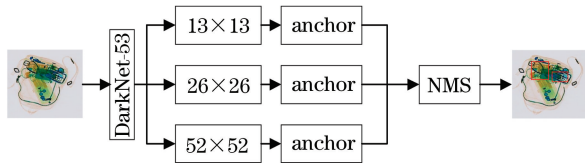


图 1 YOLOv3 网络结构图

Fig. 1 YOLOv3 network structure

2.2 复合骨干网

骨干网对检测效果来说十分重要,一个好的骨干网可有效提高目标的检测精度。但设计一个好的骨干网并非易事,Liu 等^[23]利用 ResNet101 作为基础骨干网,结合 FPN 和 Cascade R-CNN^[24]两阶段检测网络,到复合骨干网络——CBNet,该骨干网络将现有的骨干网络进行 n 层级联,形成具有 n 个骨干网络的复合骨干网络,其中第 n 层骨干网被称为引导骨干网(Lead Backbone),第 $1, 2, \dots, n-1$ 层骨干网被称为辅助骨干网(Assistant Backbone),通过相邻两个骨干网络跨层或同层的特征级联,向后传递图像特征信息,最后经过引导骨干网将特征信息传递到检测器进行检测。CBNet 在目标检测和实例分割领域取得了很好的成果,但是其复杂的网络结构使得检测速度受到极大的限制,仅仅达到 $5 \sim 8 \text{ frame} \cdot \text{s}^{-1}$ 。图 2 为复合骨干网络结构图。

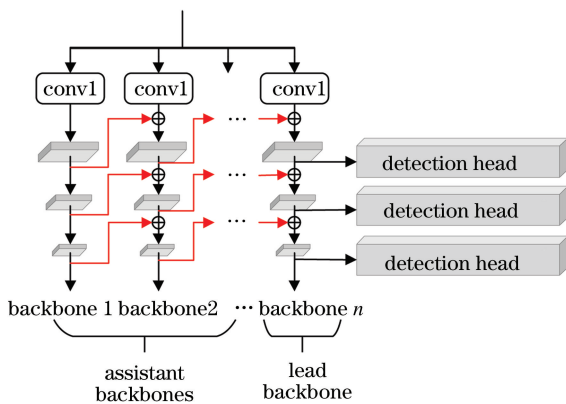


图 2 复合骨干网络结构图

Fig. 2 Composite backbone network structure

2.3 Yolo-C

X 光图像不同于自然图像,旅客行李中物品混叠,由于 X 光具有穿透性质,产生的图像十分混乱。而且违禁物品的类内差异大,增大了检测难度。利用已知的目标检测网络对 X 光图像提取特征,并不能获得精确的特征信息。针对 X 光图像的特殊性,为了满足检测的准确性和实时性的要求,笔者根据复合网络的思想,在单阶段检测网络 YOLOv3 的基础上提出单阶段双网络检测模型 Yolo-C,网络各层的参数见表 1,表中:DBL 由卷积层、BN 层和 Leaky 激活层组成;Res n ($n = 1, 2, 3 \dots$) 为 n 个残差块单元串联。

表 1 Yolo-C 参数

Table 1 Parameters of Yolo-C

Type	Layer	Filter number	Size	Output
DBL	Conv	32	3×3	416×416
Res1	Conv2	64	$3 \times 3, 1 \times 1$	208×208
—	Upsample	32	3×3	416×416
Res1'	Conv2	64	$3 \times 3, 1 \times 1$	208×208
Res2	Conv4	128	$3 \times 3, 1 \times 1$	104×104
—	Upsample	64	3×3	208×208
Res2'	Conv4	128	$3 \times 3, 1 \times 1$	104×104
Res8	Conv16	256	$3 \times 3, 1 \times 1$	52×52
—	Upsample	128	3×3	104×104
Res8'	Conv16	256	$3 \times 3, 1 \times 1$	52×52
Res8	Conv16	512	$3 \times 3, 1 \times 1$	26×26
—	Upsample	256	3×3	52×52
Res8'	Conv16	512	$3 \times 3, 1 \times 1$	26×26
Res4	Conv8	1024	$3 \times 3, 1 \times 1$	13×13
—	Upsample	512	3×3	26×26
Res4'	Conv8	1024	$3 \times 3, 1 \times 1$	13×13
DBL	Conv	30	3×3	13×13
Head	—	—	—	—
FAB	Upsample	—	$3 \times 3, 1 \times 1$	26×26
DBL	Conv	30	3×3	26×26
Head	—	—	—	—
FAB	Upsample	—	$3 \times 3, 1 \times 1$	52×52
DBL	Conv	30	3×3	52×52
Head	—	—	—	—

2.3.1 DarkNet-C

Yolo-C 以 DarkNet-53 作为基础骨干网, 提出双层复合骨干网 DarkNet-C, DarkNet-53 骨干网共 53 层, 使用了类似于 ResNet 的残差单元, 且用卷积层代替池化层进行下采样, 实现全卷积网络。DarkNet-C 由两个 DarkNet-53 网络组成, 分别为辅助骨干网 DarkNet-A 和引导骨干网 DarkNet-L。DarkNet-A 各层特征信息分别传送至 DarkNet-L 对应层的上一层, 经过特征级联后向后传播, 最后得到能够更好地表征 X 光图像的特征图。

2.3.2 特征增强模块

考虑到有些体积小的违禁物品如打火机、剪刀、折叠刀等, 在 X 光图像中难以发现, Yolo-C 采用多尺度的思想, 实现了对不同尺寸目标的检测。为了提升对超小目标的检测性能(美容剪刀、打火机), 在图像经过 DarkNet-C 骨干网后, 分别在中、小目标检测通道添加特征增强模块。该模块由 6 层卷积层组成, 首先将低层特征经过步长为 1 的卷积层, 增强了特征的非线性表达能力, 然后添加降维卷积层压缩参数量, 以确保模型检测速度。对于高层特征, 也是先经过降维卷积层, 再将融合后的特征经过平滑层, 此方式可以消除特征融合造成的混叠效应。图 3 为特征增强模块结构图。

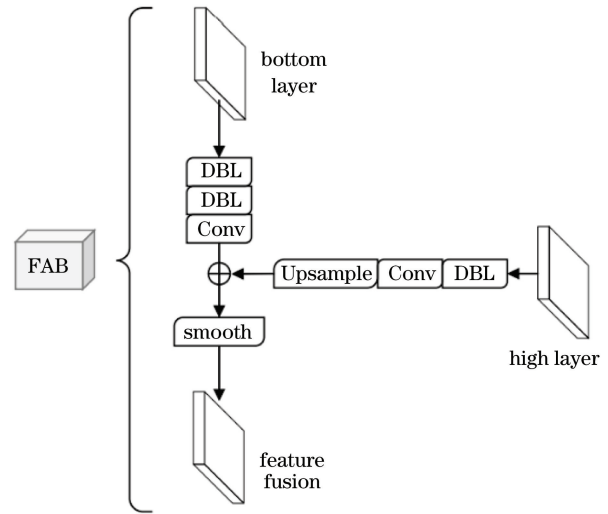


图 3 特征增强模块

Fig. 3 Feature augment block

2.3.3 损失函数

图像经过卷积层提取的特征被划分为 13×13 , 26×26 和 52×52 的网格, 不同大小的特征对应不同的 anchor, 每一个网格经过检测单元都可以得到该网格内检测框的位置及类别, 然后经过损失函数与真实值做比较, 得到边框损失。Yolo-C 的损失函数由边框损失 L_{coord} 、置信度损失 L_{con} 和类别损失 L_{classes} 加权得到。损失函数可表示为

$$L_{\text{Yolo-C}} = L_{\text{coord}} - L_{\text{con}} - L_{\text{classes}}, \quad (1)$$

$$L_{\text{coord}} = \lambda_{\text{coord}} \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{\text{obj}} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] + \lambda_{\text{coord}} \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{\text{obj}} (2 - w_i \times h_i) [(\omega_i - \hat{\omega}_i)^2 + (h_i - \hat{h}_i)^2], \quad (2)$$

$$L_{\text{con}} = \lambda_{\text{obj}} \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{\text{obj}} [\hat{C}_i \log C_i + (1 - \hat{C}_i) \log (1 - C_i)] + \lambda_{\text{noobj}} \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{\text{noobj}} [\hat{C}_i \log C_i + (1 - \hat{C}_i) \log (1 - C_i)], \quad (3)$$

$$L_{\text{classes}} = \sum_{i=0}^{K \times K} I_{ij}^{\text{obj}} \sum_{c \in N} \{ \hat{p}_i(c) \log [p_i(c)] + [1 - \hat{p}_i(c)] \log [1 - p_i(c)] \}, \quad (4)$$

式中: (x_i, y_i) 和 (\hat{x}_i, \hat{y}_i) 分别为检测框的中心坐标和真实框的中心坐标; w_i 和 h_i 表示检测框的宽和高; \hat{w}_i 和 \hat{h}_i 为真实框的宽和高; λ_{coord} , λ_{obj} 和 λ_{noobj} 分别表示检测框以及框内有目标与无目标时对应的权值; I_{ij}^{obj} 和 I_{ij}^{noobj} 表示第 i 个网格的第 j 个 anchor 是否含有目标, 如果含有目标, 则 I_{ij}^{obj} 为 1, I_{ij}^{noobj} 为 0, 反之, I_{ij}^{obj} 为 0, I_{ij}^{noobj} 为 1; $K \times K$

为网格大小; M 为 anchor 数目; \hat{C}_i 和 C_i 表示第 i 个网格内有目标与无目标的置信度; c 表示第 c 个类别; N 为总类别数; $p(\cdot)$ 为当前类别的类别概率。

首先采用均方误差分别得到边框的中心坐标和尺寸损失, 从而求得 L_{coord} 。在计算边框尺寸损失时, $w_i \times h_i$ 越小, 表示检测框的面积越小, 与

anchor 比较发现,检测框与真实框的重叠比率 (IoU) 也就会越小,故检测框的面积可以用来弱化边界框尺寸对损失值的影响。 L_{con} 和 $L_{classes}$ 可由二值交叉熵的形式计算得到。

2.3.4 网络复杂度分析

神经网络模型的复杂度分为时间复杂度 T 和空间复杂度 Ω 。时间复杂度即模型的运算次数,用浮点运算数 (FLOPs) 衡量, FLOPs 值越小,说明计算速度越快。时间复杂度过高会导致模型训练和预测耗费大量时间,其计算表达式为

$$T \sim O(\sum_{l=1}^D S_l^2 \cdot Q_l^2 \cdot H_{l-1} \cdot H_l), \quad (5)$$

式中, S 表示每个卷积层输出特征图的边长, Q 为每个卷积核的边长, D 为网络深度, l 指神经网络第 l 个卷积层, H_l 为当前卷积层的卷积核数量。空间复杂度由总参数量和各层输出特征图两部分组成。其用参数量 Params 来衡量,它决定了模型的参数量,模型的参数越多,训练模型所需的数据量就越大,其计算表达式为

$$\Omega \sim O(\sum_{l=1}^D Q_l^2 \cdot H_{l-1} \cdot H_l + \sum_{l=1}^D S^2 \cdot H_l). \quad (6)$$

表 2 为网络修改前后的复杂度分析,对比 YOLOv3 可见,特征增强模块对网络复杂度的影响较小,而 DarkNet-C 骨干网包含两个基础骨干网,复杂度变化较大,使得 YOLO-C 拥有更强大的学习能

力,同时一定程度上降低了检测速度。

表 2 不同网络复杂度分析

Table 2 Analysis of different network complexity

Model	Backbone	FAB	FLOPs	Params / 10^6
Yolov3	DarkNet-53		32.77	61.55
Ours	DarkNet-53	✓	35.58	64.65
	DarkNet-C	✓	61.85	105.93

2.3.5 迁移学习

由于 X 光图像数据较少, YOLO-C 采用迁移学习的方法训练网络。使用 DarkNet-53 骨干网在 COCO 数据集上训练的权重作为 YOLOv3-C 基础骨干网的预训练权重。其他网络部分采用权重随机初始化的方式,设权重 W 呈均值 $\mu = 0$ 和方差 $\delta = 0.01$ 的高斯分布,偏置 b 设为 0。这种半迁移学习的初始化方法,使网络能够在前期训练过程中更容易得到图像中违禁物品的特征信息。

图 4 为 YOLO-C 的整体网络结构图, YOLO-C 采用双层 DarkNet-53 网络,两层网络相邻层级的特征作跨层连接,即将前一个骨干网的输出特征作为后一个骨干网输入特征的一部分,由引导骨干网输出图像特征。然后由多尺度方法将特征分为 13×13 , 26×26 和 52×52 三个不同尺寸的特征,最后将三个特征分别送入检测单元,实现不同尺寸目标的检测。

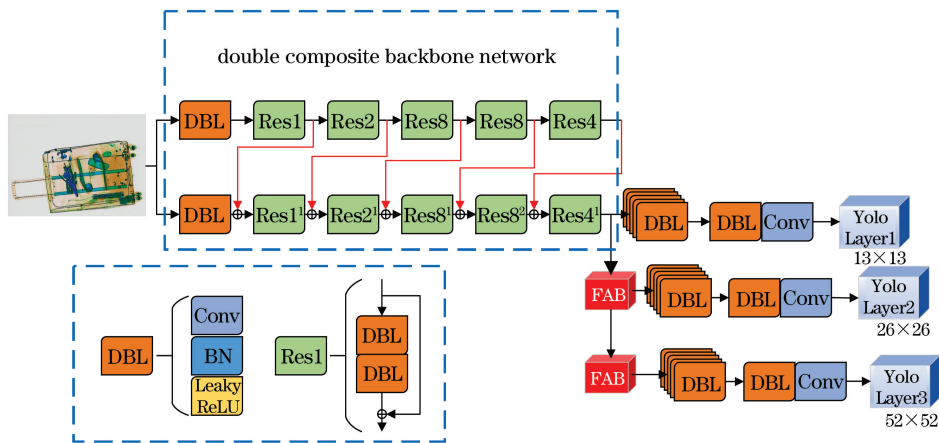


图 4 单阶段双网络目标检测算法 YOLO-C 网络结构图

Fig. 4 YOLO-C network structure of one-stage dual-network object detection algorithm

3 数据集

旅客行李的 X 光图像涉及旅客的隐私问题,所以公开的 X 光图像数据集少之又少,该领域研究人员大都是使用自制的 X 光图像数据集。目前已知

的有 GDXray 数据集和 SIXray 数据集。Mery 等^[3]制作的 GDXray 数据集包含 19407 幅由单能量安检仪拍摄的 X 光图像,如图 5 所示。该数据集分为铸件图像、焊接图像、行李图像、自然图像和场景图像 5 大部分。SIXray 数据集是由 Miao 等^[17]制作的大

规模 X 光图像数据集,该数据集含有 1059231 幅由双能量安检仪拍摄的 X 光图像,这些图像全部来自实际安检场所,如图 6 所示。SIXray 数据集含违禁品的图像 8929 幅,其中含有枪的图像 3131 幅、刀 1943 幅、扳手 2199 幅、钳子 3961 幅、剪刀 983 幅和锤子 60 幅,违禁物品展示如图 7 所示。

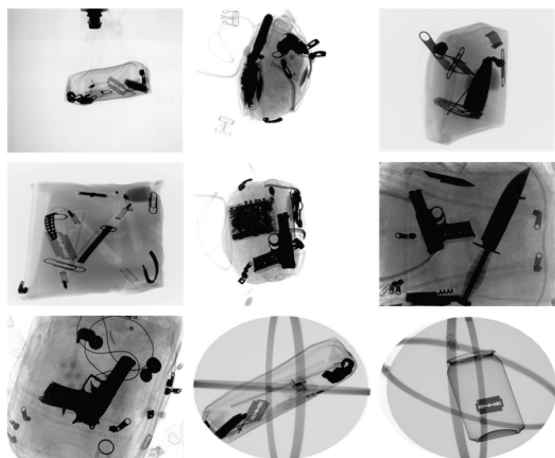


图 5 GDxray 数据集随机选取的 X 光图像

Fig. 5 Random X-ray images of GDxray dataset

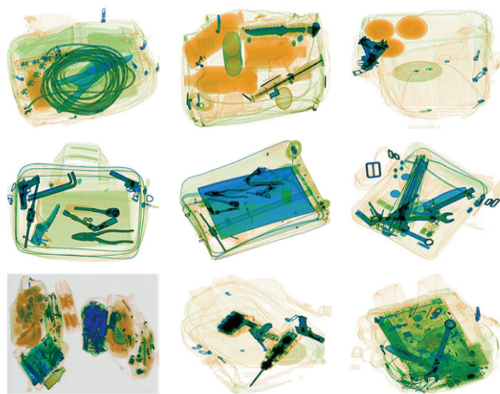
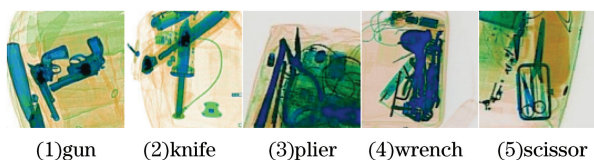


图 6 SIXray 数据集随机选取的 X 光图像

Fig. 6 Random X-ray images of SIXray dataset



(1)gun (2)knife (3)plier (4)wrench (5)scissor

图 7 违禁物品样例展示

Fig. 7 Sample display of prohibited items

GDxray 数据集中的图像呈灰白色,图像中物品单一、重复,且背景简单,拍摄图像的单能量安检仪已被淘汰,故不再对该数据集进行研究。SIXray 数据集中的图像是在实际安检场所拍摄的,所采用的双能量安检仪如今仍是主流的货物行李安检仪器,在机场、火车站以及地铁站等广泛使用,故

SIXray 数据集具有很大的研究价值,满足研究要求。但该数据集是基于图像分类任务制作的,不能直接用于目标检测。所以根据 SIXray 数据集制作了用于目标检测的数据集 SIXray_OD。

SIXray_OD 数据集是对 SIXray 数据集中包含违禁品的图像添加位置标签生成的,严格按照 COCO 数据集的标签格式,利用框图软件手动框出违禁品的位置并标注其类别,最终生成每幅图像的标签文件。SIXray_OD 数据集包含枪、刀、钳子、扳手和剪刀 5 类违禁物品(含锤子的图像数量太少本文不用于研究)。为了提高模型的泛化能力,通过反转、旋转、平移、亮度以及对比度调节等数据增强方法,将数据集扩大为 69744 幅 X 光图像集,包含枪的图像有 23488 幅、刀 12488 幅、钳子 18128 幅、扳手 31608 幅和剪刀 9272 幅,最终将 SIXray_OD 按照 4 : 1 的比例分为训练集和测试集。

4 实验结果

4.1 训练环境

训练网络采用的工作站操作系统为 Ubuntu16.04,该系统配置有 2 块 RTX 2080 ti GPU 显卡。使用 Pytorch 深度学习框架,整个网络使用随机梯度下降法(SGD)对网络参数进行优化。网络训练过程中,设置复合骨干网部分的初始学习率为 10^{-3} ,网络其他部分的初始学习率为 10^{-2} ,设置权重衰减为 10^{-5} ,batchsize 为 16,网络模型经过 100 轮迭代训练,设置非极大值抑制的阈值为 0.5。Yolo-C 网络和对照网络训练的对比如图 8 所示。

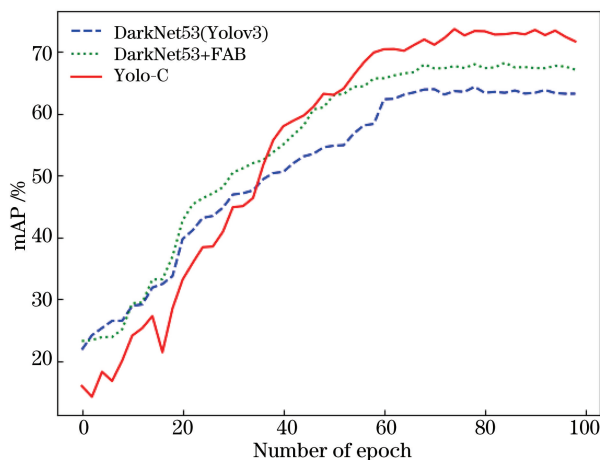


图 8 网络训练过程对比

Fig. 8 Comparison graph of network training process

4.2 评估标准

在目标检测中,IoU 的定义与计算示意图如

图 9 所示,其中 IoU 用来表示预测框与真实框之间的重叠度。通常 IoU 阈值设置为 0.5。当 IoU 大于 0.5 时,认为预测框含有目标, N_{TP} 表示判断正确的预测框数量, N_{FP} 表示判断错误的预测框数量;当 IoU 小于 0.5 时,则认为预测框不含有目标, N_{FN} 表示判断错误的预测框数量,反之, N_{TN} 表示判断正确的预测框数量。由这 4 项可求得训练样本的精度 P 和召回率 R ,表达式为

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}}, \quad (7)$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}}。 \quad (8)$$

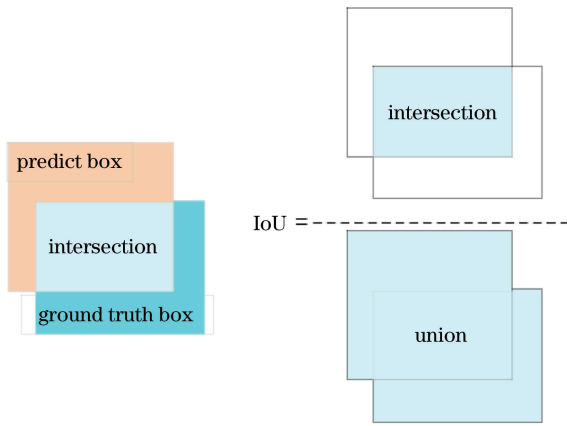


图 9 IoU 定义及计算示意图

Fig. 9 IoU definition and calculation diagram

表 3 在 SIXray_OD 数据集上对 Yolo-C 网络进行消融实验

Table 3 Ablation experiments on the Yolo-C network based on SIXray_OD dataset

No.	Model	Backbone	FAB	AP for gun / %	AP for knife / %	AP for pliers / %	AP for wrench / %	AP for scissor / %	mAP / %	Detection rate / (frame · s ⁻¹)
1	SSD	ResNet50		85.69	72.44	51.21	60.62	41.99	62.39	56
2	FASF	ResNet101		82.77	70.23	48.6	57.7	38.7	59.6	48
3	Yolov3	DarkNet-53		88.67	76.53	52.49	61.41	42.6	64.34	57
4	Faster-RCNN	ResNet101		93.83	83.74	58.85	71.58	54.1	72.18	10
5		DarkNet-53	✓	90.1	79.57	55.67	64.81	51.00	68.23	55
6	Ours	DarkNet-C		91.6	82.1	58.7	69.7	54.4	71.10	42
7		DarkNet-C	✓	93.14	83.12	60.18	73.82	58.14	73.68	40

通过对表 3 提到的目标检测网络进行训练与评估,由实验 1, 2, 3, 4 可以得出,这 4 种检测网络对图像中枪和刀的检测精度较高,但是由于钳子和扳手的类内差异大、相互重叠,以及剪刀的目标小、易被遮挡,这 4 种网络对这 3 种违禁品检测精度偏低。实验 5, 6, 7 表明,Yolo-C 对于复杂的 X 光图像有着更加强大的学习能力,对违禁品的检测精度更高。

通过设置目标的置信度阈值,得到数组满足条件的精度和召回率。这些精度和召回率组成的精度-召回率曲线(P - R 曲线)下的面积就代表训练样本中各类目标的平均精度(AP, X_{AP}), AP 的计算公式为

$$X_{AP} = \frac{1}{11} \sum_{R=0}^1 \text{pinterp}(R) \text{pinterp}(R) = \max_{R: R \geq R}(\bar{R}), R=0,1,2,\dots,10, \quad (9)$$

式中: $\text{pinterp}(R)$ 为 P - R 曲线的平滑函数, \bar{R} 是指当前 R 值右侧精度最大的召回率。平均精度均值(mAP, X_{mAP})则指所有类别的平均精度求和除以类别数目,计算公式为

$$X_{mAP} = \frac{1}{N} \sum_{i=0}^N X_{AP_i}。 \quad (10)$$

目标检测网络的检测效果常以平均精度均值作为评估指标。

4.3 消融实验

在 SIXray_OD 数据集上对 Yolo-C 进行消融实验,比较 Yolo-C 和几种典型的目标检测网络,可以发现 Yolo-C 目标检测网络在 X 光图像检测领域具有优越性。针对选取的目标检测网络,在相同的训练环境下,设置 batchsize 为 8,训练轮次 100 轮,对 SIXray_OD 数据集进行训练与评估。且通过消融实验评估了 DarkNet-C 和 FAB 对检测网络的影响,实验结果如表 3 所示。

在不同的骨干网上添加 FAB 模块训练网络,通过与无 FAB 模块的相应网络比较,发现 FAB 模块可以有效提高对小目标的精测精度。表 3 中实验 3, 5 表明:在 DarkNet-53 骨干网络下,添加 FAB 模块的网络,mAP 提高 3.89 个百分点,且对于各个类别的 AP 均有提高;对于小目标,如扳手和剪刀,两者的 AP 分别提高 3.4 和 8.4 个百分点。实验 6,7

表明,在 DarkNet-C 骨干网的加持下,添加 FAB 模块的实验 7 检测精度仍得到了提高。相对于实验 6,各个类别的 AP 分别提高 1.54, 1.02, 1.48, 4.12 和 3.74 个百分点。

复合骨干网是 Yolo-C 网络的重要组成部分。将表 2 对照组与实验 5, 6 作比较,可以直观地看出复合骨干网 DarkNet-C 的检测效果,实验 7 在 DarkNet-C 加持下,mAP 达到 73.68%,检测精度取得新高。正如 1.2 节详细阐述,利用复合骨干网的方法能够从混乱、重叠的 X 光图像中更加准确地提取出目标特征。

DarkNet-C 骨干网采用两个相同的基础骨干网,能够实现两层基础骨干网的参数共享,利于网络实现实时性检测。且 FAB 模块相对于网络的检测速度几乎没有影响。在 2 块 RTX 2080 ti GPU 显卡,CUDA 10.1 以及 batchsize 为 1 的实验环境下进行检测速度的测试,Yolo-C 网络在检测性能最佳时,检测速度能够达到 $40 \text{ frame} \cdot \text{s}^{-1}$ 。为了验证 Yolo-C 在实际安检环境下的工作效率,笔者采用安天下 TH5030 型号安检机,模拟实时过检环境。设置安检机传送带的传动速度为 0.3 m/s ,通过手动放置行李包裹,将得到的 X 光图像实时传送至网络

模型。实验证明,Yolo-C 单阶段双网络的检测速度与行李过检速度能够同步、实时、准确地获得过检行李的检测结果。

受到硬件条件的限制,训练网络时,batchsize 大小最大只能设为 8,若能继续扩大 batchsize,则可进一步提高 Yolo-C 的检测精度。在 X 光图像检测领域,部分双阶段目标检测网络取得了更好的检测结果,但检测速度约为 $10 \text{ frame} \cdot \text{s}^{-1}$,难以满足旅客过检的实时性要求。此外,由于 X 光图像的保密性,在 SIXray 数据集出现之前,研究人员大都使用自制的 X 光图像数据集,取得了较好的研究成果,但自制的数据集图像的复杂度以及真实性较低,针对真实的过检行李 X 光图像,并不能体现其算法模型在 X 光图像检测领域的适用性^[9-10]。

4.4 结果展示

为了更加具体地展示 Yolo-C 网络检测 X 光图像的效果,笔者对实验 3, 5 和 7 的样本检测结果可视化。如图 10 所示,将样本分为 6 组,分别为未含有违禁品图像、清晰图像、模糊图像、含有小目标图像、目标重叠图像以及多目标图像。事实证明,Yolo-C 能够更精确地提取图像特征,检测具有挑战性的目标,具有较高的目标检测精度。

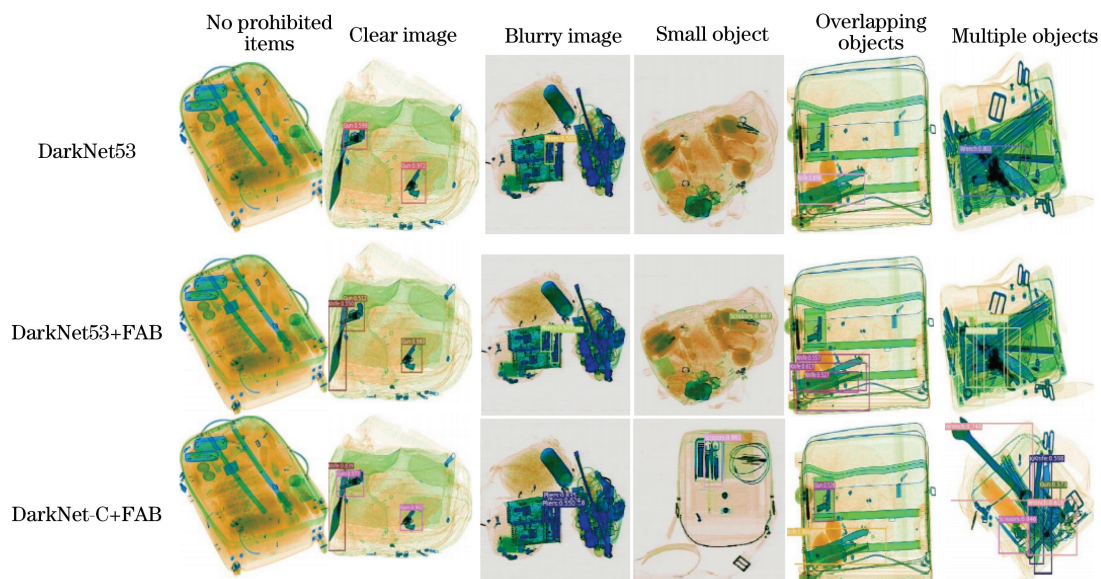


图 10 实验 3, 5 和 7 的检测结果

Fig. 10 Detection results of experiments 3, 5 and 7

5 结 论

提出一种新的目标检测网络 Yolo-C,创新地将复合骨干网 DarkNet-C 和特征增强模块 FAB 结合。在检测 X 光图像领域,该网络有效提高了

对多类违禁物品的检测精度,同时满足检测的实时性要求。Yolo-C 检测网络结合单阶段网络检测速度快和双阶段网络检测精度高的优点,在 SIXray_OD 数据集上,最终检测结果中 mAP 达到了 73.68%,检测速度为 $40 \text{ frame} \cdot \text{s}^{-1}$,能够

出色地完成 X 光图像中违禁物品的检测。这表明 Yolo-C 网络适用于 X 光图像违禁品检测领域,对公共场所的安全筛查有十分重要的研究价值。

虽然 Yolo-C 网络能够出色地完成对 SIXray_OD 数据集的检测任务,但是面对复杂多变的安检环境,针对重叠物体的检测,X 光因其具有穿透作用,易使图像中的物品相互重叠,影响检测精确度。后期需考虑提出一种新的非极大值抑制算法来解决该问题,这将是我们的下一步的研究内容。

参 考 文 献

- [1] Li T, Zhang L, He J A, et al. Rapid online identification of hazardous substances in mail using terahertz technology [J]. *Laser & Optoelectronics Progress*, 2019, 56(23): 233001.
李涛, 张良, 何建安, 等. 基于太赫兹技术在线快速识别邮件隐匿危险品 [J]. *激光与光电子学进展*, 2019, 56(23): 233001.
- [2] Singh S, Singh M. Explosives detection systems (EDS) for aviation security [J]. *Signal Processing*, 2003, 83(1): 31-55.
- [3] Mery D. Computer vision technology for X-ray testing [J]. *Insight - Non-Destructive Testing and Condition Monitoring*, 2014, 56(3): 147-155.
- [4] Roomi M M. Detection of concealed weapons in X-ray images using fuzzy K-NN [J]. *International Journal of Computer Science, Engineering and Information Technology*, 2012, 2(2): 187-196.
- [5] Kundegorski M E, Akcay S, Devereux M, et al. On using feature descriptors as visual words for object detection within X-ray baggage security screening [C] // 7th International Conference on Imaging for Crime Detection and Prevention (ICDP 2016), Madrid, Spain. London: Institution of Engineering and Technology, 2016: 1-6.
- [6] Pinol D C, Reyes E J M. Automatic handgun detection in X-ray images using bag of words model with selective search [EB/OL]. (2019-03-04) [2020-08-02]. <https://arxiv.org/abs/1903.01322>.
- [7] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [J]. *Communications of the ACM*, 2017, 60(6): 84-90.
- [8] Jaccard N, Rogers T W, Morton E J, et al. Using deep learning on X-ray images to detect threats [D]. London: University College London, 2016.
- [9] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [EB/OL]. (2015-04-10) [2020-08-02]. <https://arxiv.org/abs/1409.1556v6>.
- [10] Akcay S, Kundegorski M E, Devereux M, et al. Transfer learning using convolutional neural networks for object classification within X-ray baggage security imagery [C] // 2016 IEEE International Conference on Image Processing (ICIP), September 25-28, 2016, Phoenix, AZ, USA. New York: IEEE Press, 2016: 1057-1061.
- [11] Akcay S, Kundegorski M E, Willcocks C G, et al. Using deep convolutional neural network architectures for object classification and detection within X-ray baggage security imagery [J]. *IEEE Transactions on Information Forensics and Security*, 2018, 13(9): 2203-2215.
- [12] 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, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 580-587.
- [13] Dai J F, Li Y, He K M, et al. R-FCN: object detection via region-based fully convolutional networks [EB/OL]. (2016-06-21) [2020-08-02]. <https://arxiv.org/abs/1605.06409v2>.
- [14] Canizares P C, Merayo M G, Núñez A. FORTIFIER: a formal distributed framework to improve the detection of threatening objects in baggage [J]. *Journal of Information and Telecommunication*, 2018, 2(1): 2-18.
- [15] Akcay S, Atapour-Abarghouei A, Breckon T P. GANomaly: semi-supervised anomaly detection via adversarial training [C] // Jawahar C, Li H, Mori G, et al. *Computer Vision - ACCV 2018*, 2018, 11363: 622-637. DOI:10.1007/978-3-030-20893-6_39.
- [16] Gaus Y F A, Bhowmik N, Akçay S, et al. Evaluation of a dual convolutional neural network architecture for object-wise anomaly detection in cluttered X-ray security imagery [C] // 2019 International Joint Conference on Neural Networks (IJCNN), July 14-19, 2019, Budapest, Hungary. New York: IEEE Press, 2019: 1-8.
- [17] Miao C J, Xie L X, Wan F, et al. SIXray: a large-scale security inspection X-ray benchmark for prohibited item discovery in overlapping images [C] // 2019 IEEE/CVF Conference on Computer Vision and

- Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 2114-2123.
- [18] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [19] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08)[2020-08-02]. <https://arxiv.org/abs/1804.02767>.
- [20] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 936-944.
- [21] Li C Y, Yao J M, Lin Z X, et al. Object detection method based on improved YOLO lightweight network [J]. Laser & Optoelectronics Progress, 2020, 57(14): 141003.
- 李成跃, 姚剑敏, 林志贤, 等. 基于改进 YOLO 轻量化网络的目标检测方法[J]. 激光与光电子学进展, 2020, 57(14): 141003.
- [22] Guo J X, Liu L B, Xu F, et al. Airport scene aircraft detection method based on YOLO_v3 [J]. Laser & Optoelectronics Progress, 2019, 56(19): 191003.
- 郭进祥, 刘立波, 徐峰, 等. 基于 YOLO_v3 的机场场面飞机检测方法[J]. 激光与光电子学进展, 2019, 56(19): 191003.
- [23] Liu Y D, Wang Y T, Wang S W, et al. CBNet: a novel composite backbone network architecture for object detection [EB/OL]. (2019-09-09)[2020-08-02]. <https://arxiv.org/abs/1909.03625>.
- [24] Cai Z W, Vasconcelos N. Cascade R-CNN: delving into high quality object detection [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 6154-6162.