

优化 YOLOv4 算法的安检 X 光图像检测网络^①



杨嘉诚, 黄佳慧, 韩永麟, 王 萍, 李晓辉

(长安大学 电子与控制工程学院, 西安 710021)

通讯作者: 杨嘉诚, E-mail: jagger.yang@foxmail.com

摘 要: 针对目前安检场景中违禁品种类复杂、人工检查效率低易出错等问题, 文章提出一种名为 Res152-YOLO 的网络架构, 该架构基于 YOLOv4 (You Only Look Once) 优化目标检测网络. 为提高对 X 光图像中危险品的检测精度, Res152-YOLO 采用 ResNet-152 网络代替原 YOLOv4 中的 CSPDarknet-53 网络, 将改进后的 ResNet 残差网络与 YOLOv4 网络连接. 实验中利用 YOLOv4、Res152-YOLO 等一系列网络在同一数据集上进行对比实验, 分别比较上述网络的损失曲线、对各类危险品的检测性能以及各网络的总体性能. 结果表明, Res152-YOLO 网络在以上实验中性能优于原 YOLOv4 网络, 并且满足安检设备的帧率要求. 改进后的网络有效提高了安检的准确率, 能够消除一定的安全隐患.

关键词: YOLOv4; 残差网络; 深度学习; 卷积神经网络; 数据增强; 安检 X 光; 违禁品识别

引用格式: 杨嘉诚, 黄佳慧, 韩永麟, 王萍, 李晓辉. 优化 YOLOv4 算法的安检 X 光图像检测网络. 计算机系统应用, 2021, 30(12): 116-122. <http://www.c-s-a.org.cn/1003-3254/8221.html>

X-Ray Security Image Detection Network Based on Optimized YOLOv4

YANG Jia-Cheng, HUANG Jia-Hui, HAN Yong-Lin, WANG Ping, LI Xiao-Hui

(School of Electronics and Control Engineering, Chang'an University, Xi'an 710021, China)

Abstract: Given the diverse prohibited varieties in current security inspection scenes and low-efficiency error-prone manual inspections, this study proposes network architecture, Res152-YOLO, based on the YOLOv4 optimized target detection network. Res152-YOLO uses the ResNet-152 network to replace the CSPDarknet-53 network in the original YOLOv4 and connects the improved ResNet to the YOLOv4 network to enhance the detection accuracy of dangerous goods in X-ray images. A series of networks such as YOLOv4 and Res152-YOLO are used to conduct comparative experiments on the same data set to compare the loss curves of the above-mentioned networks, the detection results for various dangerous goods, and the overall performance of each network. The results show that the improved network can improve the accuracy of security detection and eliminate potential safety hazards.

Key words: YOLOv4; ResNet; deep learning; Convolutional Neural Network (CNN); data augmentation; X-ray security; contraband identification

随着社会的发展, 人们的出行方式变得越来越多样化. 与此同时, 违禁品种类也变的越来越复杂. 因此, 安检则是当今社会必不可少的一项工作. 安检应用于飞机、火车、地铁等场所, 在人们的生活中起着重要

的作用^[1].

X 光安检仪是安检过程中最常用的设备, 该设备检测图像具有以下特点: (1) 多尺度性: 安检违禁品种类很多, 形状各不相同, 且同一类别的违禁品在外观上

① 黄佳慧和韩永麟为共同第二作者

收稿时间: 2021-03-09; 修改时间: 2021-03-31; 采用时间: 2021-04-06

也可能存在很大差异。(2) 杂乱性: 安检图像的一块区域中会有很多物品混合在一起, 会产生严重干扰。

目前, X 光安检仪与人工检查相配合是大部分场所的安检手段。这种模式下安检质量很大程度上决定于安检员的工作状态。人工检查受环境因素影响大, 导致效率较低, 很可能出现漏检、误检等问题^[2,3]。

要想解决上述问题, 需要有一个自动的检测模式。目前, 深度学习是一种较好的选择。本文主要针对 X 光安检图像, 将深度学习模型 YOLOv4 (You Only Look Once) 与 ResNet 残差网络相融合, 提出了一种新型网络架构 Res152-YOLO 以优化对安检过程中违禁品的检测效果。

实验结果表明, Res152-YOLO 网络的性能相较于 YOLOv4 原网络有了较显著的提升, 更适用于违禁品检测。

本文主要贡献有:

(1) 针对 ResNet 网络进行优化和激活函数的更新; 通过适当降低帧率、增加网络层数以获得更佳效果; 采用了创新式结构融合算法。

(2) 针对 X 光图像进行数据增强处理, 突出物品特征; 制作后端接口软件, 使检测操作一体化。

(3) 进行了多种相关变体网络的对比实验。

1 相关工作

1.1 基于深度学习的目标检测算法

深度学习提出了一种让计算机自动学习出模式特征的方法, 并将特征学习融入到了建立模型的过程中, 从而减少了人为设计特征造成的不完备性, 深度学习的网络层数较深, 有利于多尺度违禁物品的识别。

基于深度学习的目标检测算法主要分为基于区域建议的方法和无区域建议的方法两类。

(1) 基于区域建议的方法: 2014 年, Girshick 等人提出的 RCNN 模型^[4], 比传统的方法有了明显的改进; 同年, He 等人针对 RCNN 要求输入图像块尺寸固定以及速度慢等问题提出了空间金字塔池化模型 SPP^[5]; 2015 年, Girshick、Ren 等人先后提出了 Fast RCNN^[6]、Faster RCNN^[7], 该类方法检测精度高, 但是检测速度较慢。

(2) 无区域建议的方法: 2015 年 He 等人针对梯度弥散的问题, 提出了深度残差网络 ResNet^[8]; 2019 年发布的 YOLOv4^[9] 相较于以往版本, 进一步在检测性能上有了提升。之后, 在 YOLOv4 的基础上产生了一些改进的算法, 例如: 2020 年 11 月蒋镛圻等人

结合 scSE 注意力机制和 SPP, 设计了一种 scSE-IYOLOv4 算法^[10], 其能够显著提升小目标的检测精度。

1.2 YOLOv4 基础网络

YOLOv4 的网络结构分为: (1) Input 输入端: 使用 Mosaic 及 CutMix 数据增强以增加输入图像可变性, 丰富图像特征信息, 所设计的目标检测模型可以获得更高的鲁棒性。(2) Backbone 主干网络: 提取输入图像的特征, 并通过卷积层进行 5 次下采样实现特征图的降维。该部分的改进在于融合了多种新方式。(3) Neck 部分: 由 YOLOv3 采用的特征金字塔网改为加入空间金字塔池化层的路径聚合网络, 改善了浅层特征丢失的问题。(4) Prediction 部分: 实现了对小、中、大型目标的检测, 并在原图像上对检测物进行锚框标注。

1.3 数据集

目前, 常见的红外 X 光图像数据集有两类: (1) GDxray^[11]: 数据集仅包含简单背景中的灰度图像, 包含来自 5 个不同子集的 19407 个 X 射线样本, 此数据集内容相对比较简单。(2) SIXray^[12]: 数据集由中国科学院大学模式识别与智能系统开发实验室构建, 由 1059231 张 X 射线图像组成, 其中 5 类 8929 个违禁品是标注的, 此数据集内容比较复杂, 较有挑战性。SIXray 数据集部分图片如图 1 所示。



图 1 SIXray 数据集部分图片

2 网络设计

2.1 输入数据增强

X 射线安检系统根据物品的物质类别对安检图像进行颜色设定, 把属于有机物的物体颜色设定为橙色, 把无机物设定为蓝色, 把混合物设定为绿色。针对 X 光

数据集图像普遍的特性:颜色昏暗、图像噪声大,存在物品遮挡带来的干扰^[13],本文统一对输入数据进行增强处理.该部分具体实现过程如下:

(1) 比例校正:由于检测网络只能接受正方形图像,这会强制大多数 X 光数据集图片产生拉伸形变.因此,本文对图像周围进行灰度填充,使图像大小标准化.

(2) 亮度与色度增强:物品颜色较深,包裹颜色较浅.经亮度与色度增强后,包裹亮度增强,趋近于白色,轮廓及细节得以弱化,而物品仍保有深色,便可有效区分包裹背景和物体.

(3) 对比度增强:为了解决图像中物品重叠、遮挡所造成的干扰问题,进行了对比度增强处理,提高重叠部分物品亮度,同时降低被遮挡违禁品亮度,使其在图像中显示更加清晰.如图 2 所示.



图 2 数据增强对比图

2.2 网络结构优化

在 YOLOv4 的主干特征提取网络中,采用由一系列残差网络结构组成的 CSPDarknet-53 网络^[14].该网络采用了较低层数提取特征同时保证了检测速度与精度,能够应用于实时传输的视频流处理.然而,目前大多数红外安检设备具有图像存储功能,可以进行任意图像处理、连续保存工作.若将视频处理转化为截取图像处理会大大减少算法网络对于检测速度的要求,故可适当的减少检测帧率,这也使得架构中可以扩充原网络的层数. YOLOv4 及优化后结构如图 3 所示.

2.2.1 ResNet-152 网络搭建

本文针对 X 光安检场景提出了一种改进后的 Res152-YOLO 网络结构. Res152-YOLO 网络结构舍去了原 YOLOv4 中的 CSPDarknet-53 浅层网络,代替为 ResNet-152 深层网络. ResNet-152 是具有 152 层的 ResNet 残差网络,可以看作是残差单元 (RU) 的堆栈.该堆栈包含 3 个卷积层:第一个 1×1 卷积层仅具有 64 个特征图,它充当了瓶颈层;然后是一个 3×3 卷积层,具有 64 个特征图;其次是另一个 1×1 卷积层,具有 256 个特征图来恢复原始深度. ResNet-152 包含

3 个此类 RU,可输出 256 个特征图,然后是 512 个特征图的 8 个 RU,具有 1024 个特征图的 36 个 RU,最后是具有 2048 个特征图的 3 个 RU.

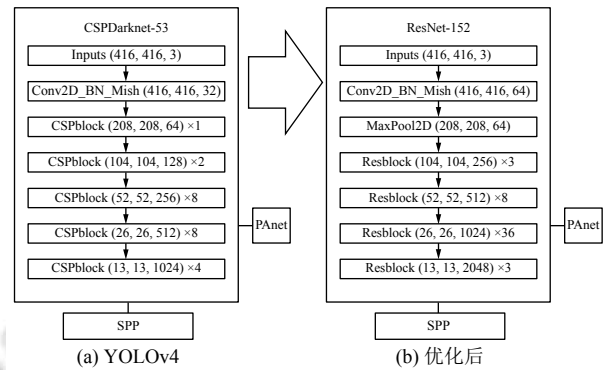


图 3 YOLOv4 及优化后结构图

2.2.2 ResNet-152 网络优化

ResNet-152 网络最初应用于图像分类而非物体检测任务,相较于 Inception、VGG、Densenet 等分类网络, ResNet-152 是其中图片分类任务中表现最好的神经网络之一,并可继续延伸至更高层数^[15,16].因此本文选择该网络作为 YOLOv4 的主干特征提取网络.但是,对于 X 光图像这类检测任务,需要将 ResNet-152 进行修改使其能够与 YOLOv4 的后端结构进行衔接.

因为原 ResNet-152 网络的后端紧密连接层用于执行分类任务,故删除网络末尾的 average pool 层、flatten 层、Dense 层,只保留输入端以及骨干卷积层部分,并根据需要调整卷积层的部分参数设置.将原 ResNet 网络的激活函数 ReLU 改为 Mish 函数,因为 Mish 在训练稳定性和准确性上均优于原激活函数.

2.2.3 ResNet-152 网络与 YOLOv4 相融合

首先,将 ResNet 卷积网络与 YOLOv4 后端网络连接.当输入为 416×416 的图像时,最开始的输出为 256 个特征图的 3 个 RU 层不变,随后将 8 个 RU 的输出送至 PANet 高层进行特征提取,36 个 RU 的输出送至 PANet 中层,将输出为 2048 个特征图的 3 个 RU 进行 3 次卷积后送至 SPP 结构,分别利用 4 个不同尺度的最大池化进行处理,最大池化的池化核大小分别为 13×13 、 9×9 、 5×5 、 1×1 .此结构能极大地增加感受野,分离出显著的特性.后将 SPP 结构的输出送至 PANet 底层.至此 3 个不同特征图的输出已全部连接 PANet 特征金字塔的结构. PANet 利用该特征金字塔从下到

上以及从上到下进行特征的反复提取,送至 YoloHead 结构。

其次,利用 YoloHead 对获得的特征进行预测并解码预测结果,从而得到预测框的具体位置,再将预测框绘制在原图上。

Res152-YOLO 网络相较于 YOLOv4 网络具有更深的网络层,使其在理论上可以获得更好的效果:对 X 光图像中违禁品的检测精度更高。同时,Res152-YOLO 网络也会因为残差块的增加而减慢单张图像的检测速度。结合前文所述,在 X 光检测场景中更注重精度性能,故可以接受检测速度下降带来的负面效果,Res152-YOLO 仍具有可行性。

2.3 建立扩展网络

除 ResNet-152 外,ResNet 还有其他的多种变体

网络如:ResNet-18、ResNet-34 等^[17,18]。本文尝试将 ResNet 一些变体同样替换掉 YOLOv4 原主干特征提取网络,目的是将这些新网络的性能进行横向对比。

本文分别构建了不同层数的 ResNet 网络与 YOLOv4 网络的其他结合,分别命名为 Res50-YOLO 与 Res101-YOLO,它们的前端残差块层数分别为 50、101,在各自网络中每个残差模块的 RU 的个数分别为 3、4、6、3 与 3、4、23、3。这两种网络的大体结构与 Res152-YOLO 相近,区别仅在于 RU 个数的不同导致网络深度小于后者。在 YOLOv4 原网络、Res50-YOLO、Res101-YOLO、Res152-YOLO 四个网络中均加入了 Mosaic 数据增强、Label Smoothing 平滑以及余弦退火衰减算法用于优化学习率的调节。总网络结构见图 4 所示。

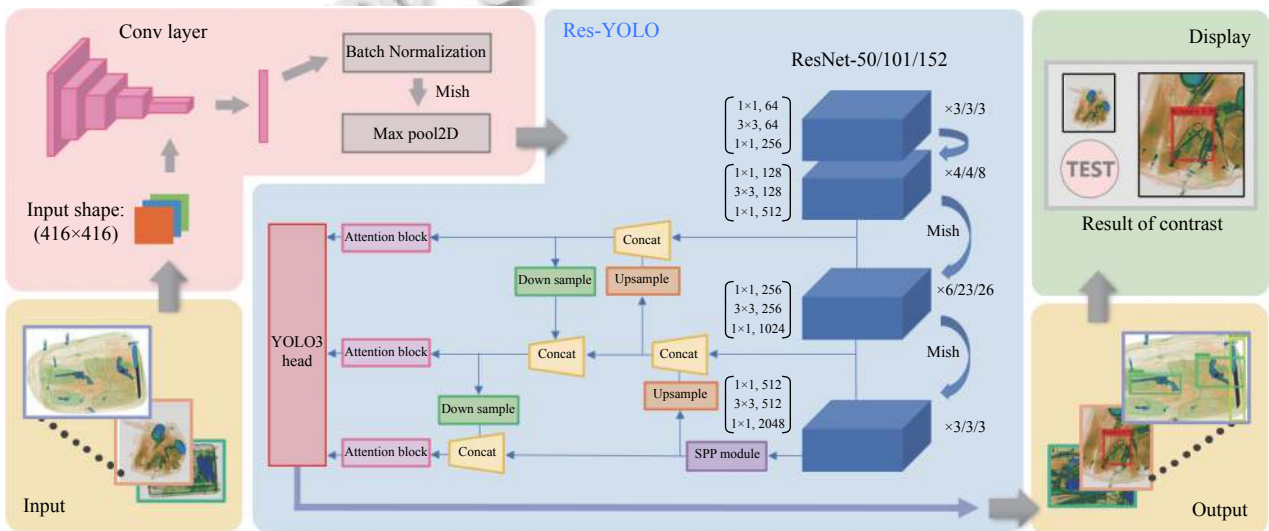


图4 网络总结构图

2.4 后端接口搭建

该部分利用软件形式搭建后端,使检测结果可视化以实现以下目的:(1)便于操作者使用和检测:用户可添加自定义 X 光图像进行检测,同时还可以在文件路径栏和当前状态栏获取文件路径信息和软件运行状态。(2)输入输出图像实时对比:能够同时查看检测前图像、数据增强后的输出图像及二者对比效果,使检测结果更加直观。界面如图 5 所示。

3 实验测试与结果

本节将针对 YOLOv4 原网络、Res50-YOLO、Res101-YOLO、Res152-YOLO 进行对比实验,分别比

较四者的损失曲线、检测各违禁品类别的性能以及网络的总体检测性能。最后,本文将测试 Res152-YOLO 与 YOLOv4 网络的运行帧率,测试其是否能满足安检设备运行条件。

实验环境:本实验运行平台为 Linux;显卡型号 GeForce GTX 2080Ti (显存 11 GB);电脑内存为 31 GB。

3.1 实验数据准备

3.1.1 数据集选择

- (1) 数据集名称: SIXray.
- (2) 违禁品类别:小刀、手枪、扳手、钳子、剪刀共 5 类。
- (3) 数据集特点:图像较为复杂,包含物体重叠、

噪声等图像干扰。

3.1.2 超参数设置

- (1) Batch 批量大小: 8.
- (2) Epoch 训练轮数: 100.

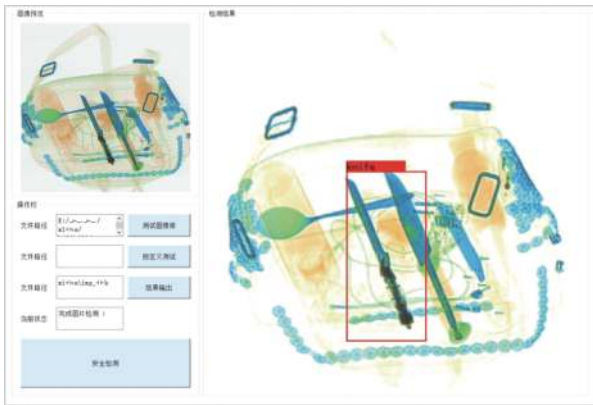


图5 后端接口界面

- (3) Learning rate 初始学习率: 0.0001.

3.2 网络训练实验

训练并记录 YOLOv4 原网络、Res50-YOLO、Res101-YOLO、Res152-YOLO 四个网络在训练轮次中的训练损失函数曲线以及验证损失函数曲线。损失函数 loss 采用原 YOLOv4 中的 CIoU loss 函数。曲线如图 6 所示。

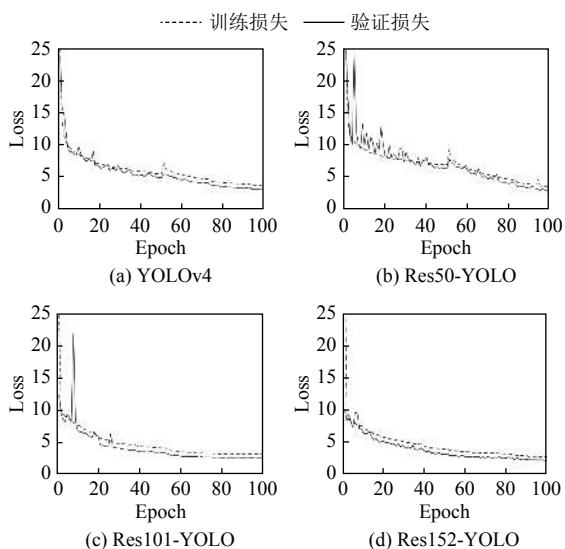


图6 各网络 loss 函数损失

首先,从图 6 中曲线中可以看出,所有网络的损失函数值总体趋于下降趋势,且每个网络的训练损失与

验证损失曲线基本吻合。这说明没有产生梯度爆炸、过拟合等问题,所有网络都具有一定的非线性表达能力,可以拟合复杂的特征输入。

其次,在 100 次迭代过程中,Res101-YOLO、Res152-YOLO 模型的损失函数值明显低于其它两个模型,且后者在最终训练轮次达到了 Loss 值的最低点。这说明 Res152-YOLO 的检测效果在数据集的图像上最接近于真实目标。

3.3 性能对比实验

3.3.1 违禁品检测结果

在训练结束后,测试 YOLOv4 原网络、Res50-YOLO、Res101-YOLO、Res152-YOLO 分别针对 5 个类别违禁品的性能测试指标:精度、召回率、AP 值。检测结果分别如表 1、表 2、表 3 所示。

表1 安检违禁品的检测精度 (%)

违禁品	原YOLOv4	Res152-YOLO	Res101-YOLO	Res50-YOLO
钳子	90.7	96.1	94.3	90.8
剪刀	93.7	97.3	98.2	98.3
小刀	90.7	91.5	93.1	92.9
手枪	97.5	98.1	98.9	97.7
扳手	86.9	92.8	89.1	94.1

表2 召回率 (%)

违禁品	原YOLOv4	Res152-YOLO	Res101-YOLO	Res50-YOLO
钳子	44.5	67.1	54.2	50.8
剪刀	39.4	67.8	52.2	50.8
小刀	53.5	59.6	53.0	55.4
手枪	78.3	86.1	79.8	76.2
扳手	27.6	55.0	42.6	33.5

表3 AP (%)

违禁品	原YOLOv4	Res152-YOLO	Res101-YOLO	Res50-YOLO
钳子	75.3	84.6	78.8	74.6
剪刀	68.7	81.5	78.5	74.8
小刀	72.9	74.0	74.4	72.1
手枪	95.1	93.5	92.2	94.5
扳手	53.0	73.8	63.4	58.3

根据实验得到以下结论: (1) 精度: Res101-YOLO 网络与 Res152-YOLO 网络的检测精度大致相同,其中 Res152-YOLO 网络的对于 5 类违禁品的检测精度都大于 90%, 对手枪、钳子、剪刀这 3 类违禁品的检测精度达到了 95% 以上。(2) 召回率: 召回率是所有网络普遍较弱的性能, YOLOv4 原网络对于扳手的召回率低至 27.6%, 这说明各网络对于各违禁品难以查全。

Res152-YOLO的各类违禁品召回率能保持在50.0%以上,其中对于扳手的召回率55.0%相较于YOLOv4提高了99.2%。(3) AP值: Res152-YOLO网络的各AP值均大于70%,其中钳子、剪刀、扳手分别高出YOLOv4原网络12.9%, 18.6%, 39.2%,进一步说明了Res152-YOLO网络的优越性。

3.3.2 网络总体性能对比

将4个网络中对所有违禁品的AP值进行平均化后得到mAP,以反应每个网络的总体检测性能。

从对比实验图7中可以看出, Res152-YOLO的mAP值最大,其值为81.52%,其次是Res101-YOLO值为77.49%。Res50-YOLO与YOLOv4网络的mAP相差不多,分别为74.91%和73.03%。相较于YOLOv4原网络, Res152-YOLO的mAP提高了11.62%。这说明在总体的检测性能上, Res152-YOLO网络的安检违禁品检测能力超越了YOLOv4网络。

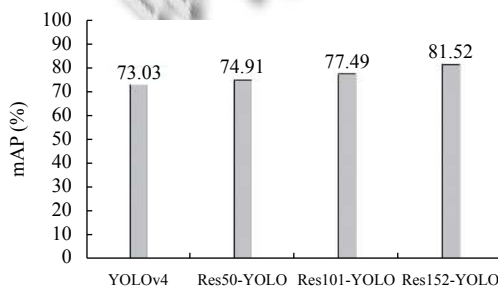


图7 各网络 mAP 值

3.4 帧率测试实验

对网络 Res152-YOLO 与 YOLOv4 分别进行帧率测试,测试结果如表4。

表4 帧率测试

网络	检测速度(f/s)
YOLOv4	62.26
Res152-YOLO	40.09

根据表格得出, Res152-YOLO 的检测速度弱于 YOLOv4, 其值为 40.09 f/s。由于一般红外安检设备的 30 至 60 f/s, 且违禁品检测方法并非实时对象追踪, 故 Res152-YOLO 网络检测速度仍能满足多数安检场景。

4 结语

本文基于 X 光安检设备特性的考虑, 在原 YOLOv4 网络的基础上改进其网络结构, 进行了数据增强并构建了 Res152-YOLO 等一系列残差融合网络。根据多组

对比实验, 本文得到结论: Res152-YOLO 网络的总体性能较 YOLOv4 提升了近 11%, 且在部分违禁品如扳手、剪刀上效果远超 YOLOv4。综上所述, Res152-YOLO 网络在安检场景下满足帧数的要求, 可以更精确的识别违禁品并具备更高的工作效率, 有助于公共场所的安全防范。

参考文献

- 1 Mery D. Automated detection in complex objects using a tracking algorithm in multiple X-ray views. CVPR 2011 Workshops. Colorado Springs: IEEE, 2011. 41–48.
- 2 Cover T, Hart P. Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 1967, 13(1): 21–27. [doi: 10.1109/TIT.1967.1053964]
- 3 Mery D, Mondragon G, Riffo V, et al. Detection of regular objects in baggage using multiple X-ray views. Insight-Non-Destructive Testing and Condition Monitoring, 2013, 55(1): 16–20. [doi: 10.1784/insi.2012.55.1.16]
- 4 Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014. 580–587.
- 5 He KM, Zhang XY, Ren SQ, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904–1916. [doi: 10.1109/TPAMI.2015.2389824]
- 6 Girshick R. Fast R-CNN. Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015. 1440–1448.
- 7 Ren SQ, He KM, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137–1149. [doi: 10.1109/TPAMI.2016.2577031]
- 8 He KM, Zhang XY, Ren SQ, et al. Deep residual learning for image recognition. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 770–778.
- 9 Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Opstimal speed and accuracy of object detection. arXiv: 2004.10934, 2020.
- 10 蒋榕圻, 彭月平, 谢文宣, 等. 嵌入 scSE 模块的改进 YOLOv4 小目标检测算法. 图学学报, 2021, 42(4): 546–555.

- 11 Mery D, Riffó V, Zscherpel U, *et al.* GDXray: The database of X-ray images for nondestructive testing. *Journal of Nondestructive Evaluation*, 2015, 34(4): 42. [doi: [10.1007/s10921-015-0315-7](https://doi.org/10.1007/s10921-015-0315-7)]
- 12 Miao CJ, Xie LX, Wan F, *et al.* SIXray: A large-scale security inspection X-ray benchmark for prohibited item discovery in overlapping images. *Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019. 2119–2128.
- 13 张友康, 苏志刚, 张海刚, 等. X光安检图像多尺度违禁品检测. *信号处理*, 2020, 36(7): 1096–1106.
- 14 徐子睿, 刘猛, 谈雅婷. 基于YOLOv4的车辆检测与流量统计研究. *现代信息科技*, 2020, 4(15): 98–100, 103.
- 15 徐晨玉. 基于YOLOv3的危险物品检测模型研究与改进 [硕士学位论文]. 兰州: 兰州大学, 2019.
- 16 Zhou JY, Zhao YM. Application of convolution neural network in image classification and object detection. *Computer Engineering and Applications*, 2017, 53(13): 34–41.
- 17 李彦冬, 郝宗波, 雷航. 卷积神经网络研究综述. *计算机应用*, 2016, 36(9): 2508–2515, 2565. [doi: [10.11772/j.issn.1001-9081.2016.09.2508](https://doi.org/10.11772/j.issn.1001-9081.2016.09.2508)]
- 18 张慧, 王坤峰, 王飞跃. 深度学习在目标视觉检测中的应用进展与展望. *自动化学报*, 2017, 43(8): 1289–1305.