# 基于深度学习的 X 光图像违禁品检测算法

## YOLO 系列算法研究进展

JidaDia0 114514114514

2025 年 1 月 16 日

**摘要**

随着深度学习技术的发展，X 光图像违禁品检测在公共安全领域中扮演着越来越重要的角色。YOLO（You Only Look Once）系列算法以其卓越的实时性和准确性，在目标检测领域取得了显著成就。本综述旨在全面分析 YOLO 系列算法在 X 光图像违禁品检测中的应用、进展和挑战，探讨其在公共安全领域中的潜力和未来发展方向。

## 1 引言

阿巴巴巴巴爸爸巴巴爸爸巴巴爸爸巴巴爸爸啊爸爸巴巴爸爸吧

## 2 研究背景

### 2.1 X 光图像

X 射线成像是一种高效的无损检测技术，在行李安检中被广泛用于检测隐藏的危险物品。其成像原理基于 X 射线管发射的高能电离射线穿透扫描对象，根据物体材料的密度和厚度导致射线信号的衰减（如图1A 所示）。衰减的程度可以用公式 $I_x = I_0 e^{-\mu x}$ 表示，其中 $I_x$ 是穿透材料后的射线强度，$I_0$ 为初始射线强度，$\mu$ 为线性衰减系数，$x$ 为材料厚度 [1]。材料的密度越高，衰减越显著，因此在图像中表现为强度较低的区域（如图1B 所示）。这种物理特性使 X 射线成像在检测高密度危险物品（如金属武器和爆炸物）时尤为有效，同时也能够区分低密度有机材料（如塑料和液体）。

在安检应用中，目前市场上最常见的是二维 X 射线图像技术，因其设备相对低成本且能够快速生成行李内容的内部结构信息。二维成像通过单能级或多能级 X 射线进行扫描，单能级成像利用固定能量的射线生成灰度图像，能够提供物体基本的密度信息。然而，单能级技术在复杂行李场景中可能难以清晰区分多种材料，例如金属、塑料和有机物的混合情况。为应对这一问题，现代 X 射线机广泛采用多能级成像技术，通过不同能量级别的 X 射线生成多张图像，结合查找表（Look-Up Table，LUT）方法，将不同材料的密度和有效原子序数（$Z_{eff}$）转换为伪彩色图像（如图1B 所示）[2] 。在伪彩色图像中，高密度金属可能被标记为蓝色，而低密度有机材料则呈现为橙色或红色，从而显著增强了危险物品的可识别性。这种技术特别适用于检测复杂行李中的危险品，例如将枪支、刀具与普通物品区分开来，同时减少误报。

此外，多视角成像技术进一步提升了二维 X 射线成像的检测能力。多视角技术通过从不同角

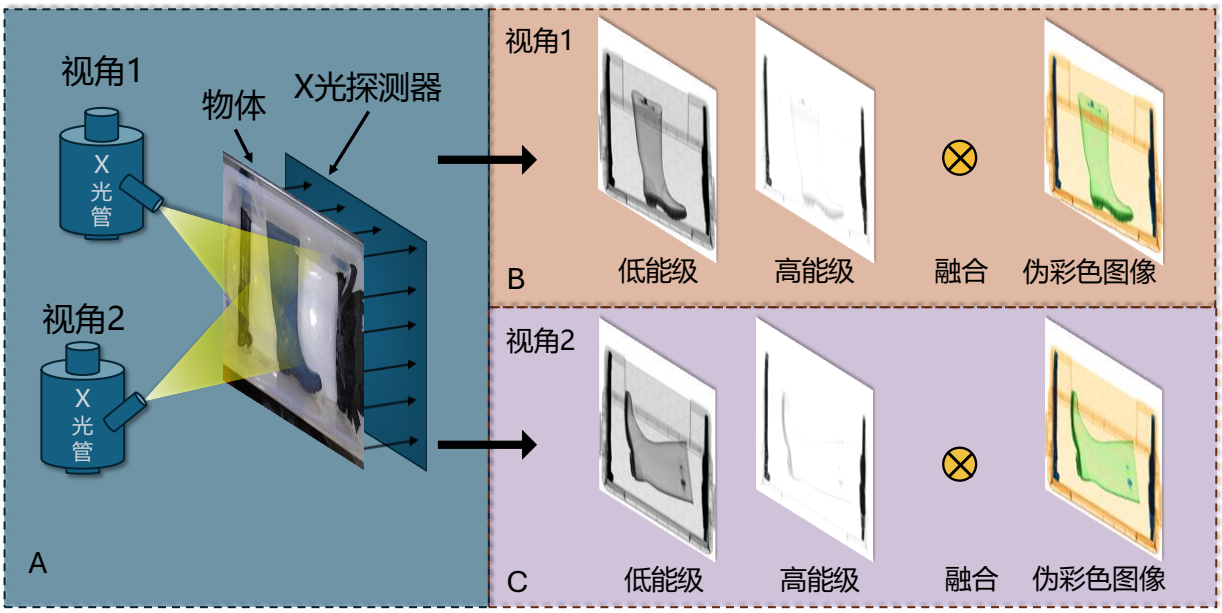


图 1: X 光图像成像原理

度对行李进行扫描，生成多张图像（如图1所示），从而提供更丰富的空间信息，使隐藏在复杂结构中的危险物品（如刀具、金属爆炸物等）更加突出，同时提高了对多层结构的解析能力 [1]。尽管多视角成像需要更多的数据处理能力，但在实际安检中已被广泛使用，其与多能级技术结合，显著提高了检测效果。

### 2.2 传统方法检测

在深度学习技术尚未普及之前，传统的 X 光图像违禁品检测方法在相关研究中占据重要地位。这些方法主要涵盖图像增强、威胁图像投影（Threat Image Projection, TIP）、目标分类、目标检测和图像分割等技术领域，依托计算机视觉与传统机器学习手段，旨在提升 X 光图像的可读性和检测性能，为安检工作提供基础性技术支持。这些传统方法的研究成果在一定程度上为后续深度学习技术的发展奠定了理论和方法论基础。

图像增强技术是传统 X 光安检检测的关键环节，其主要目的是通过预处理提升图像质量，使操作者或算法更容易识别潜在威胁目标。常见的技术包括低能和高能 X 光图像的融合，以利用不同能量水平的射线穿透特性提升目标物体的显著性。此外，伪彩色技术被广泛应用于灰度图像的可视化处理，通过将灰度值映射到颜色空间显著增强目标的可视性 [2–6]。这些增强方法虽然能够提高安检人员对目标物体的检测效率，但它们仍然受到算法设计复杂度和对环境变化适应能力的限制。

威胁图像投影（TIP）是一种用于生成合成数据集的重要传统技术，通过将二值威胁物体掩码投影到无威胁的 X 光图像上生成包含威胁目标的合成图像。TIP 技术的优势在于，其通过仿射变换、对数变换等操作生成了多样化的合成数据，极大地丰富了数据集的规模和多样性。这不仅可以用来训练机器学习模型，还在培训人类筛查人员的过程中得到了有效应用 [7–11]。然而，需要指出的是，TIP 生成的合成图像可能在纹理细节和真实威胁物体分布方面与实际情况存在

偏差，这种限制可能会对模型在实际应用中的泛化性能造成一定影响。

在目标分类任务中，传统方法主要依赖于视觉词袋（Bag of Visual Words, BoVW）模型，该方法通过特征提取与聚类构建视觉词典，再利用支持向量机（SVM）或随机森林（RF）等分类器对目标物体进行分类 [12–15]。此外，稀疏表示技术通过稀疏编码的方式提取图像中的高维特征，其分类性能在部分研究中表现出优势 [16–20]。例如，在 GDXray 数据集上的研究显示，基于稀疏表示和 k-NN 分类器的传统方法能够取得与深度学习模型相近的性能（94.7% 对比 96.3%）[21]。然而，这些方法的计算复杂度较高，且对大规模、多样化数据集的适应能力有限，难以在实际安检场景中取代深度学习方法。

在目标检测方面，传统方法同样以 BoVW 框架为主导 [22–24]，并结合稀疏强度域图像描述符（SPIN）[25] 等特征提取技术，利用 SVM 分类器实现了较高的检测精度。此外，多视角图像的引入显著增强了检测的鲁棒性，特别是在物体发生旋转或部分遮挡的情况下，多视角数据能够提供更多的形态和纹理信息，从而提升检测性能 [22, 26–28]。然而，这些方法在复杂场景中往往需要依赖大量的先验知识或人工设计特征，这在一定程度上限制了其应用的广泛性和自动化程度。

传统图像分割技术最初主要依赖于固定阈值法以分离目标与背景 [29, 30]，这种方法虽然计算简单但对噪声和光照变化的敏感性较高。随后，基于图论的分割方法逐渐兴起，如模糊相似距离、谱聚类和变分分割技术等 [31–33]，在小型数据集上的表现较为优异。然而，这些方法在应对复杂结构的目标分割以及大规模数据处理时，计算效率和鲁棒性方面仍存在不足。因此，它们难以直接应用于实际安检场景，特别是在要求实时性的任务中表现较为乏力。

总体而言，传统 X 光图像违禁品检测方法通过一系列图像处理与机器学习技术，为 X 光图像的解读和威胁物体检测提供了重要的理论支持与实践经验。这些方法不仅在提高检测精度方面发挥了积极作用，还为后续深度学习技术的发展奠定了基础。然而，这些传统方法在应对复杂场景、多样化数据以及大规模处理任务时存在明显的局限性，包括特征提取的自动化程度不足、泛化性能有限和对高效计算的依赖性较高。随着深度学习技术的快速发展，研究逐渐向更加智能化和高效的解决方案转变。这种转变不仅在检测精度和效率上带来了显著提升，同时也为 X 光图像违禁品检测领域开辟了新的研究方向与应用场景。

### 2.3 深度学习方法检测

随着深度学习技术的迅猛发展，其在 X 光图像违禁品检测中的应用研究不断涌现，与传统方法相比，深度学习方法展现出显著的性能优势与应用潜力。在早期的工作中，akcay 等人 [34] 首次将卷积神经网络（CNN）应用于 X 光图像分类任务，并利用迁移学习技术验证了其在数据有限场景下的有效性。在枪支与非枪支的二分类任务中，通过逐层冻结预训练的 AlexNet 层的权重的方式发现，即使在网络所有权重层冻结的情况下，CNN 依然显著优于基于视觉词袋（BoVW）方法的支持向量机（SVM）和随机森林（RF）。更进一步的实验则表明，CNN 在多类别分类问题中同样具有较高的应用前景，尤其是在类别间具有较高相似性的情况下。

基于深度学习的目标检测技术也被应用于 X 光图像违禁品检测中。早期基于深度学习的目标检测方法多以多阶段检测为主。以 [35] 为例，其通过滑动窗口将 UCL TIP 数据集的 X 光图像分割成多个小块，分别提取多种特征（如强度特征、oBIF 特征 [36]、PHOW 特征 [37] 以及 CNN 特征），并使用不同分类器（如 SVM、

表 1: X 光图像违禁品检测技术概述

| 类别 | 解决的问题 | 具体方法 | 引用 |
|---|---|---|---|
| 传统方法 | 提升图像质量，增强目标显著性 | X 光图像融合（低能、高能）；伪彩色技术 | [2–6] |
| | 数据集不足，丰富训练数据 | 威胁图像投影 (Threat Image Projection, TIP) | [7–11] |
| | 分类 X 光图像中的威胁目标 | 视觉词袋 (BoVW) 模型结合机器学习算法 | [12–15] |
| | 提高分类性能 | 稀疏表示技术结合 k-NN 分类器 | [16–20] |
| | 检测威胁物体位置 | 稀疏强度域图像描述符 (SPIN)、多视角检测 | [22–26] |
| | 提取目标区域与背景 | 固定阈值法；基于图论的分割 | [29–33] |
| 深度学习方法 | 解决小数据集场景下的性能问题 | 卷积神经网络 (CNN) ；迁移学习 | [34] |
| | 检测威胁物体位置 | 多阶段检测；基于 VGG 特征（滑窗 + 多分类器） | [35–37] |
| | 扩展标注数据集的规模与多样性 | 生成式对抗网络 (GAN) 结合对抗训练策略 | [39, 40] |
| | 类别不平衡问题 | 跨层特征融合模型 (CHR) | [41] |
| | 威胁目标定位，解决目标遮挡问题 | 即插即用模块结合边缘与材料信息强化注意力机制 | [42] |
| | 提高模型跨数据集泛化能力 | 在不同扫描仪数据上训练和验证深度学习模型 | [43, 44] |
| | 提高威胁目标检测的精度 | 多视角输入检测；改进 Faster RCNN 的多视图池化 | [45]，[46] |

RF 和 CNN）进行检测，结果显示基于 VGG-18 提取的特征（CNN 特征）配合 RF 分类器取得了最高性能（FPR：0.22%）。同时，Jaccard 等人 [38] 通过比较不同输入类型对 CNN 性能的影响发现，双通道输入（包含原始图像及其对数变换图像）训练的 VGG-19 模型在分类任务中优于单通道输入模型（AUC：97%，FPR：6%）。

随着生成式对抗网络 GAN 在图像生成领域的成熟与应用，Zhao 等人 [39] 提出了一种结合 GAN 的三阶段算法，用于扩展标注数据集的多样性和规模。该方法首先通过前景物体的角度信息对图像进行初步分类，随后利用 GAN 生成高质量的 X 光图像，并结合对抗训练策略进一步提升生成样本的质量。后续研究 [40] 在此基础上改进了 GAN 的训练机制，生成的图像在视觉质量上显著优于之前的方法。这种基于生成模型的策略不仅在小数据集场景中提高了模型的泛化能力，同时也为后续的检测任务提供了更加丰富的训练数据。

为了解决 X 光图像违禁品检测任务中常见的类别不平衡问题，Miao 等人 [41] 提出了一个基于跨层特征融合的模型（CHR），通过将连续层的特征进行连接和冗余信息去除，优化模型对复杂背景和少见类别的检测性能。使用 SIXray 数据集训练 ResNet-101 模型时，提出的方法实现了 mAP 提升 2.13% 的显著改进。类似研究 [42] 则进一步通过引入即插即用模块，利用边缘和材料信息强化注意力机制，从而显著提升了模型对检测目标的定位能力。

在跨数据集泛化能力的研究中，Caldwell 等人 [43] 探索了深度学习模型在不同扫描仪产生的数据间的迁移性能。实验表明，由于扫描仪参数的未知性以及数据集分布差异，深度学习模型在未见数据上的表现仍具有较大挑战。进一步研究 [44] 则通过在 DBF3 和 SIXray 数据集上训练和验证模型，定量分析了 CNN 模型的泛化能力，揭示了领域转移对检测性能的显著影响。

多视图 X 光图像的使用也在近年来成为提升检测性能的重要方向。研究 [45] 通过结合多视角检测结果，验证了多视图输入在提高模型精度上的效果（R-FCN 与 ResNet-101 结合时，多视图 mAP 为 0.938，而单视图仅为 0.798）。此外，工作 [46] 改进了 Faster RCNN 的结构，通过多视图池化层提取三维特征，并生成三维候选区域以进行后续检测，实验结果表明多视图方法较单视图有显著性能提升（mAP：95.56% vs. 91.23%）。

现有的 X 光图像违禁品检测算法整理如表1所示。从表中可以看出，已经有许多研究者们已经从多个角度对 X 光图像中的违禁品检测任务进行了广泛探索，其中，基于深度学习的 X 光图像违禁品检测方法更是在分类、检测及生成数据扩展等多个方向均取得了显著的研究进展。然而，尽管方法种类繁多，YOLO 算法凭借其端到端设计和出色的实时性，已经在安检检测任务中脱颖而出，成为该领域的重要研究方向。后续章节将重点讨论 YOLO 在 X 光图像违禁品检测中的相关研究及其应用前景。

### 2.4 YOLO 目标检测模型各版本的发展历程

## 3 YOLO 系列模型在 X 光图像违禁品检测中的应用

自 2016 年 YOLO（You Only Look Once）目标检测算法问世以来，该系列算法在不到十年的时间内经历了快速的迭代与演进。尤其是在应对复杂背景、小目标检测等挑战性场景方面，其表现取得了显著提升。这些技术进步使得 YOLO 系列算法逐渐成为 X 光图像违禁品检测这一关键领域的重要技术工具，为安全检查的效率与精准度提供了强有力的技术支撑。

图2概述了 YOLO 系列模型从 YOLOv1 到 YOLOv11 的发展历程、发布时间及其作者。在众多版本中，YOLOv1 至 YOLOv8，以及 YOLOX

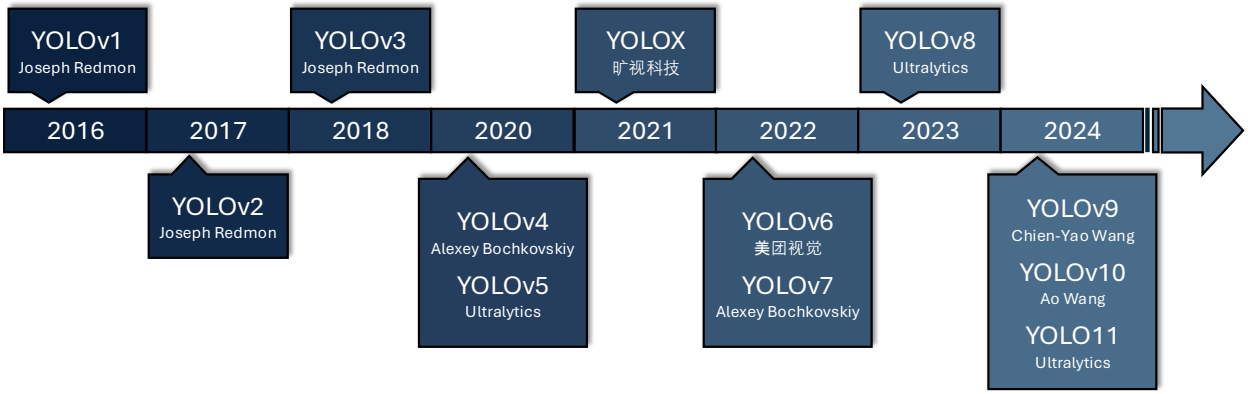


图 2: YOLO 系列模型发展历程

等代表性模型凭借卓越的实时性能、高检测精度和轻量化设计，在 X 光图像违禁品检测领域得到了广泛应用。因此，本节将重点聚焦于 YOLOv1 至 YOLOv8 及 YOLOX 的关键改进方向与核心技术特点，清晰勾勒出 YOLO 系列算法在目标检测领域的技术演化脉络。随后，还将进一步探讨这些版本在 X 光图像违禁品检测中的具体应用实践与表现。

### 3.0.1 YOLOv1

YOLO（You Only Look Once）[47] 是由 Redmon 等人于 2016 年提出的一种一阶段目标检测算法，它在目标检测领域具有重要的里程碑意义。与传统的两阶段检测模型（如 R-CNN [48–50] 系列）相比，YOLO 基于端到端的回归策略，直接预测目标边界框的位置和类别概率，从而实现了目标检测任务的统一建模和高效处理。这种创新设计显著简化了检测流程并提升了运行速度。

如图3所示，YOLOv1 首次提出了将输入图像划分为 $S \times S$ 的网格的设计思路。每个网格单元负责预测 $B$ 个边界框及其置信度分数，并输出 $C$ 个类别的概率分布。具体而言，每个边界框的预测由 5 个参数组成：$(P_c, b_x, b_y, b_w, b_h)$，其中，$P_c$ 表示边界框的置信度分数，即目标是否存在及边界框预测的准确程度；$b_x$ 和 $b_y$ 是边界框中心点相对于网格单元的偏移量；$b_w$ 和 $b_h$ 分别表示边界框的宽度和高度。最终，YOLO 的输出可以被表示为一个维度为 $S \times S \times (B \times 5 + C)$ 的张量，其中 5 包括 4 个边界框坐标（$b_x$, $b_y$, $b_w$, $b_h$）和 1 个边界框置信度分数 $P_c$，$C$ 是类别数。对于每个网格单元，YOLOv1 只预测一个边界框，因此每个网格单元的输出维度为 $5 + C$。

YOLOv1 的骨干网络基于 GoogLeNet [51] 的设计思路，包含 24 个卷积层和 2 个全连接层，并引入了 $1 \times 1$ 卷积以减少计算复杂度和参数量。该模型首先在 ImageNet 数据集上进行预训练，以获取通用的特征表示；随后在 PASCAL VOC 数据集上进行微调与验证。在 PASCAL VOC 2007 数据集上，YOLOv1 实现了 63.4% 的平均精度（AP），展现出在检测速度上的显著优势。

然而，YOLOv1 的设计也存在一定的局限性。由于采用固定网格划分以及每个网格单元固定数量的边界框，模型在检测密集目标和小目标时性能受限，容易出现漏检和定位不准确的问题。此外，为了减少边界框冗余，YOLOv1 引入了非极大值抑制 [52]（Non-Maximum Suppression, NMS），用于筛选与真实框最接近且置信度最高的预测边界框。

YOLOv1 的提出开创了实时目标检测的先

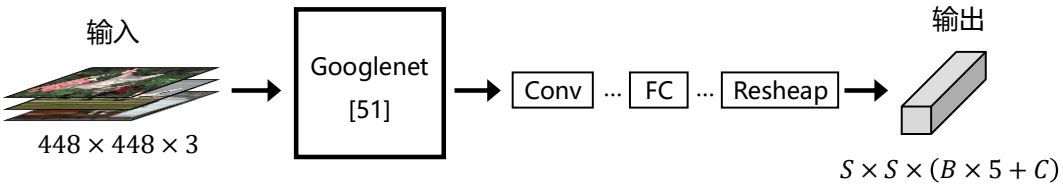


图 3: YOLOv1 网络结构

河，其端到端设计和统一的处理流程显著简化了目标检测过程，同时在速度和效率上取得了革命性进展。这一创新奠定了后续 YOLO 系列模型发展的坚实基础，为目标检测领域的研究和应用开辟了新的方向。

### 3.0.2 YOLOv2

YOLOv2 [53] 是由 Joseph Redmon 和 Ali Farhadi 于 2017 年提出的目标检测算法，它在 YOLOv1 的基础上进行了多项关键改进，实现了在保持实时检测性能的同时显著提升检测精度。在网络结构方面，YOLOv2 采用了更为轻量化且高效的骨干网络 Darknet-19，该网络包含 19 个卷积层和 5 个最大池化层，并通过 $1 \times 1$ 卷积层有效减少了参数量，同时增强了特征提取能力。此外，YOLOv2 改进了卷积层模块 CBL，使用步幅为 2 的卷积层替代部分池化层，从而在降低特征图空间分辨率的同时保留了更多的细节信息，缓解了池化操作的低层次细节丢失问题。

相比 YOLOv1，YOLOv2 的改进可以归纳为以下几个核心方面，这些改进共同推动了模型性能的显著提升。首先，YOLOv2 引入了锚框机制，通过预定义一组固定高宽比的锚框（Anchor），模型在训练过程中学习锚框与真实框之间的偏移量，从而直接预测边界框的位置、尺寸和类别分布。锚框机制显著提升了模型在检测密集目标和小目标时的表现和边界框预测的精度。

其次，YOLOv2 在每一层网络中加入了批量归一化（Batch Normalization），这不仅加速了模型的收敛速度，还显著降低了过拟合风险，提升了模型的鲁棒性。第三，YOLOv2 采用了多尺度训练（Multi-scale Training）策略，在训练过程中动态调整输入图像的分辨率（范围为 $320 \times 320$ 至 $608 \times 608$，每 10 个 epoch 随机更改一次尺寸），从而使模型能够学习不同分辨率下的特征，增强了对目标尺度变化的适应能力。这一策略使得 YOLOv2 能够在不同输入尺度下对目标进行检测，增加了检测小目标和大目标的能力。

如 图4所 示，YOLOv2 的输出尺度和 YOLOv1 有所不同。YOLOv2 使用了多个尺度的特征图来进行目标检测，特别是在使用更高分辨率输入时，它会生成多个不同大小的特征图，每个特征图负责不同大小的目标检测。输出的维度可以表示为 $S \times S \times (B \times 5 + C)$，其中 $S$ 是网格的大小，$B$ 是每个网格单元的锚框数量，$C$ 是类别数。相比 YOLOv1，YOLOv2 的输出更加精细，能够处理不同尺度的目标。

为了进一步提高模型对小目标的检测性能，YOLOv2 引入了多项优化机制。其一，通过特征融合（Passthrough 层），即后续版本中被称为特征金字塔网络（FPN）的早期形式，YOLOv2 将浅层高分辨率特征与深层语义特征通过通道拼接的方式融合，生成的特征图包含更多细节信息，从而提高了模型对小目标的检测能力。其二，YOLOv2 在训练数据中应用了 K-means 聚类算法以优化锚框大小，使锚框分布更加贴合数据的实际分布特性，从而提升了模型的边界框预测质量，并减少了对无关框的学习偏差。

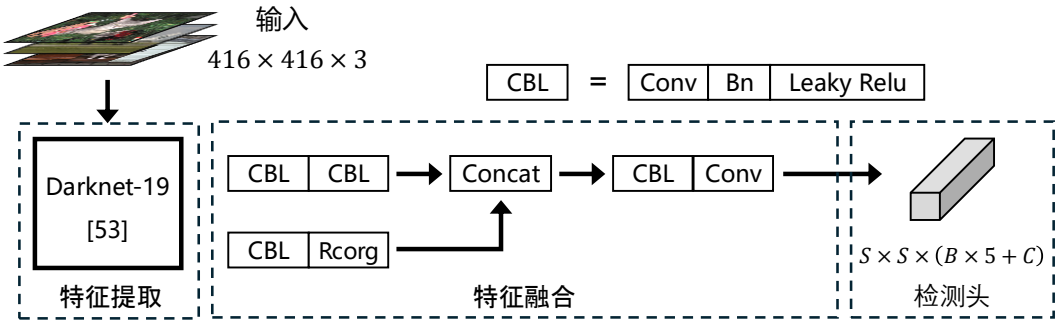


图 4: YOLOv2 网络结构

### 3.0.3 YOLOv3

YOLOv3 [54] 是由 Joseph Redmon 和 Ali Farhadi 于 2018 年提出的一种目标检测算法，通过多项改进进一步增强了模型的检测能力，同时延续了其在实时检测中的优势。YOLOv3 引入了全新的骨干网络 Darknet-53，该网络包含 53 个卷积层，并借鉴了 ResNet [55] 的残差连接（Residual Connection）设计，从根本上缓解了深层网络中的梯度消失问题。这一设计显著提升了模型的特征提取能力，使得 YOLOv3 在 ImageNet 数据集上的分类性能与 ResNet-152 相当，但计算速度却快近两倍，充分体现了其在高效性和性能平衡上的优越性。

在目标检测任务中，如图5所示，YOLOv3 的改进主要体现在多尺度检测、锚框机制优化和分类方法的调整上。为提升对不同尺度目标（特别是小目标）的检测能力，YOLOv3 采用了类似于特征金字塔网络（Feature Pyramid Network, FPN）[56] 的多尺度检测机制。模型通过从不同分辨率的特征图上进行检测，分别在 13×13、26×26 和 52×52 的特征图上预测目标边界框。这种分层检测策略使得每个分辨率的特征图能够专注于特定大小目标的检测：低分辨率特征图（13×13）负责大目标的检测，中分辨率特征图（26×26）检测中等目标，而高分辨率特征图（52×52）则专注于小目标。这一设计极大地提高了模型对不同尺度目标的适应性，尤其是在复杂场景中检测小目标的能力。

在输入尺度方面，YOLOv3 支持多种输入分辨率（例如 416 × 416 或 608 × 608），通过多尺度训练方法（每隔 10 个 epoch 随机改变输入图像的尺寸），增强了模型对不同大小目标的适应能力。通过这种方式，YOLOv3 能够在不同尺度的输入图像上进行有效的检测，确保在各种实际应用场景中的高效性和准确性。

在锚框机制方面，YOLOv3 延续了使用锚框（Anchors）作为边界框先验的设计，并通过 K-means 聚类对训练数据生成更合理的锚框尺寸。相比之前的版本，YOLOv3 在每种尺度的特征图上由每个网格单元预测三个边界框，这种改进进一步提升了目标尺度的建模精度。每个锚框的预测包含 4 个边界框坐标和 1 个置信度分数（表示目标是否存在），并且每个边界框有 $C$ 个类别的概率。

同时，在类别预测上，YOLOv3 放弃了传统的 Softmax 激活函数，转而采用独立的逻辑回归方法。这种方法不仅支持多标签分类（例如一个目标可以同时属于“动物”和“宠物”两类），还为每个边界框预测独立的类别概率，提高了分类的灵活性和对复杂场景的适应能力。此外，为了进一步提升检测性能，YOLOv3 支持引入空间金字塔池化（Spatial Pyramid Pooling, SPP）[57] 模

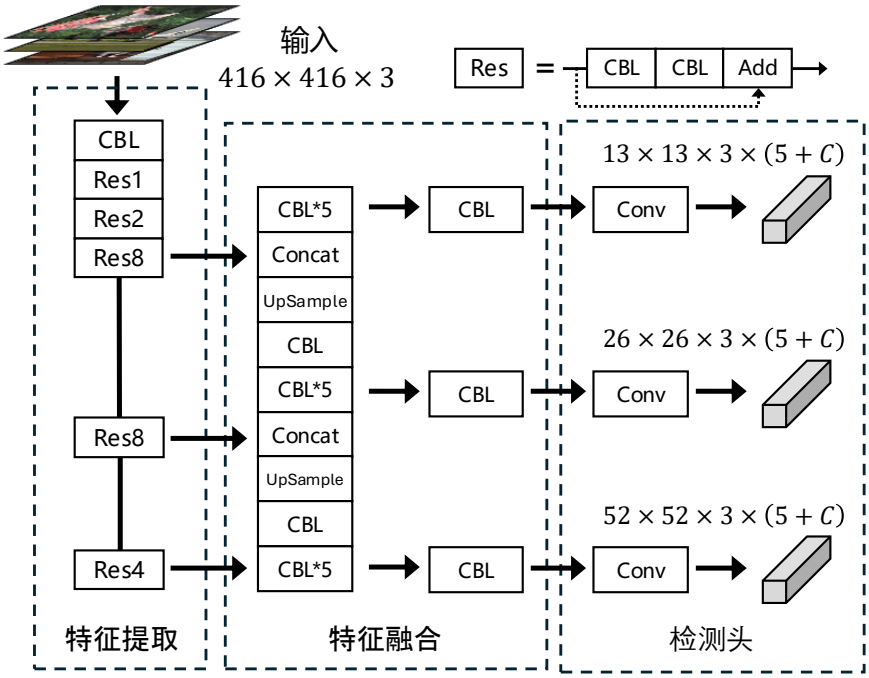


图 5: YOLOv3 网络结构

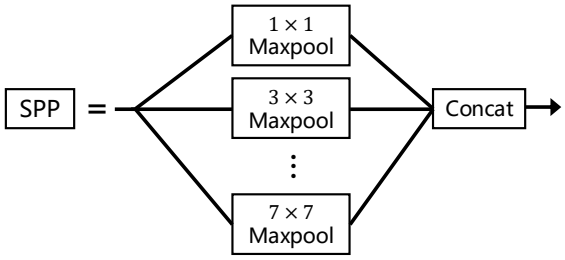


图 6: 空间金字塔池化 SPP

块，如图6所示，通过整合不同感受野的池化特征（例如 $1 \times 1$、$3 \times 3$、$5 \times 5$ 和 $7 \times 7$），显著扩大了感受野范围，增强了模型对全局和局部特征的捕捉能力。

#### 3.0.4 YOLOv4

YOLOv4 [58] 是由 Alexey Bochkovskiy 等人于 2020 年提出的一种目标检测算法，它通过整合多种优化策略，在保持实时检测性能的同时显著提升了检测精度。YOLOv4 的架构进行了全面的设计优化，包括特征提取、特征融合和检测头三大模块。在特征提取阶段，YOLOv4 引入了 CSPDarknet-53 作为骨干网络。该网络基于 YOLOv3 的 Darknet-53 结构，结合了跨阶段部分连接（CSPNet）的设计，通过跨层连接有效减少梯度信息的冗余，提升了模型的学习效率，同时降低了计算复杂度和参数量，而不损失检测精度。此外，YOLOv4 采用了 Mish 激活函数，这是一种平滑的非线性激活函数，相较于 ReLU 和 Leaky ReLU，它在梯度流动和深层网络的稳定性上表现更为优异，从而进一步增强了特征提取能力。

在特征融合阶段，YOLOv4 结合了多项先进设计，显著提升了对不同尺度目标的检测能力。首先，通过改进版的空间金字塔池化（SPP）模块，模型通过多尺度池化操作（例如 $1 \times 1$、$5 \times 5$、$9 \times 9$ 和 $13 \times 13$ 的池化核），在不增加计算成本的前提下扩大了感受野范围，并增强了特征图中多尺度信息的表达能力。这种设计使得

模型能够更精准地捕捉不同大小目标的轮廓特征。此外，YOLOv4 采用了路径聚合网络（Path Aggregation Network, PANet）[59]，进一步优化了特征金字塔结构中不同分辨率特征之间的信息传递，尤其是在浅层高分辨率特征与深层低分辨率特征之间的融合上表现突出，从而提升了模型对小目标检测的性能以及复杂场景下的适应性。

在模型的训练阶段，YOLOv4 集成了多种创新的优化策略，以增强训练的稳定性并提高检测性能。其中，Mosaic 数据增强是一项显著提升训练样本多样性的技术，通过随机将四张图像拼接为一张，模型在小目标检测和复杂背景适应性上表现更加出色。通过引入自对抗训练（Self-Adversarial Training, SAT）策略，即在训练过程中生成对抗样本，使模型在面对扰动或噪声时具有更高的鲁棒性。YOLOv4 还优化了边界框回归的损失函数，使用 CIoU（Complete IoU）损失，综合考虑了边界框之间的距离、重叠度和长宽比一致性，从而显著提升了边界框定位的精度。为了减少模型的过拟合问题，YOLOv4 在训练中加入了类别标签平滑（Label Smoothing）和 DropBlock 正则化策略。前者通过对类别标签分布的平滑处理，降低了模型对训练数据的过度拟合倾向；后者则通过在特征图上随机遮挡部分区域，增强了模型的泛化能力。此外，YOLOv4 还首次在目标检测中采用了遗传算法进行超参数优化，通过自动搜索找到最佳的模型配置参数，进一步提升了性能。

#### 3.0.5 YOLOv5

YOLOv5 [60] 是由 Ultralytics 公司于 2020 年发布的目标检测算法，与之前的 YOLO 系列相比，其在易用性和部署性能方面实现了显著突破。作为 YOLO 系列的延续与优化，YOLOv5 通过多项创新设计进一步提升了检测精度、运行效率和适配能力。其中，引入的 Focus 模块、C3 模块、SPPF 模块和 AutoAnchor 技术，在不同层面优化了模型结构、锚框生成和多尺度目标的检测能力，使其成为一款更高效且易于部署的目标检测框架。

在网络结构方面，YOLOv5 通过关键模块的改进进一步增强了模型性能。首先，Focus 模块通过对输入特征图进行切片操作实现了无损下采样，不仅大幅提高了下采样效率，还保留了更多空间细节信息，从而增强了对小目标的检测能力。其次，YOLOv5 的 C3 模块是对 YOLOv4 中 CSP 模块的优化，通过跨层部分连接减少梯度冗余的同时，在不显著增加参数量的前提下有效提升了特征表达能力。此外，模型还引入了 SPPF 模块（快速空间金字塔池化），通过整合不同尺度的特征图来增强对多尺度目标的表达能力，尤其是在复杂场景中有效捕捉多目标信息。

在锚框生成方面，YOLOv5 创新性地采用了 AutoAnchor 技术，能够根据训练数据集的特性和配置动态优化锚框大小。这种动态调整不仅简化了锚框参数设置的流程，还提升了模型对遮挡目标的检测精度和召回率。此外，YOLOv5 提供了全面的数据增强策略，包括 Mosaic 数据增强、复制粘贴（Copy-Paste）、MixUp 、随机仿射变换和颜色变换等，大幅提升了训练样本的多样性和模型的泛化能力，使其在复杂场景下表现更加鲁棒。

YOLOv5 的显著特点之一是模型版本的多样性。它提供了五种不同规模的版本，分别为 YOLOv5n（纳米）、YOLOv5s（小型）、YOLOv5m（中型）、YOLOv5l（大型）和 YOLOv5x（超大型），以适应不同计算资源和应用需求。轻量化版本（如 YOLOv5n 和 YOLOv5s）因其极低的计算需求，非常适合部署在边缘设备或移动端环境中，而高性能版本（如 YOLOv5x）则在检测精度上表现突出。与 YOLOv4 相比，YOLOv5 显示出显著的训练速度优势，同时通过解决网格

敏感性问题，提高了梯度稳定性，进一步增强了模型的训练效率和推理性能。

#### 3.0.6 YOLOX

YOLOX（You Only Look Once Extended）[61] 是由旷视科技于 2021 年发布的一种目标检测模型，相较于传统的 YOLO 系列，YOLOX 在架构设计上进行了深度改进，显著提升了检测精度和训练效率，并简化了模型的实现。作为对 YOLOv3 的延续与优化，YOLOX 引入了多项创新技术，包括无锚框架（Anchor-free Design）、解耦头（Decoupled Head）、SimOTA 标签分配、多正样本策略以及高级数据增强技术，实现了在检测速度与精度之间的优异平衡，奠定了其在目标检测领域的重要地位。

YOLOX 的核心改进体现在以下几个方面，首先，无锚框架设计摒弃了从 YOLOv2 起使用的锚框机制，直接预测目标的网格偏移、宽度和高度，从而简化了模型的结构和训练流程。这种无锚方案受到无锚目标检测器（如 FCOS）的启发，避免了锚框匹配过程中的超参数干扰，同时提升了模型的效率和检测性能。在 MS COCO 数据集上的实验表明，无锚框架设计使 YOLOX 的平均精度（AP）提高了 0.9 个百分点，并降低了训练复杂度。

此外，如图7所示，YOLOX 采用了解耦头（Decoupled Head）的检测结构，将分类任务和边界框回归任务分离为两个独立的子网络。这一改进解决了任务耦合带来的特征共享冲突问题，使模型在收敛速度和检测性能上均有所提升。与传统的检测头设计相比，解耦头设计进一步提高了模型的平均精度，使 AP 增加了 1.1 个百分点。同时，YOLOX 提出了基于最优传输理论（Optimal Transport, OT）的标签分配策略 SimOTA ，从全局视角优化了正负样本的分配规则。该策略不仅有效缓解了样本不平衡和标签分配的歧义性，还显著提升了目标特征与样本匹配的质量，使 AP 增加了 2.3 个百分点。

在无锚框架下，YOLOX 进一步引入了多正样本策略，通过中心采样机制，将每个预测网格的中心区域标记为正样本，从而缓解了正负样本数量的不平衡问题。这一策略显著提升了模型的检测能力，使 AP 增加了 2.1 个百分点。同时，YOLOX 在训练阶段结合了多种高级数据增强技术，包括 Mosaic 数据增强和 MixUp ，大幅扩展了训练样本的多样性。这些技术使得模型即使在没有 ImageNet 预训练的情况下，也能够取得优异的性能，并通过提升泛化能力使 AP 增加了 2.4 个百分点，进一步巩固了 YOLOX 在目标检测任务中的优势。

在性能评估中，YOLOX-L 在 MS COCO 测试-dev 2017 数据集上实现了 50.1% 的平均精度（AP），表现出了卓越的检测能力。与传统锚框机制的模型相比，YOLOX 在复杂场景和大规模数据集中的表现尤为突出，同时在训练效率和模型精度方面实现了显著突破。这种高效的无锚设计和创新的优化策略，使 YOLOX 成为新一代目标检测模型的典范，在学术研究与工业应用中均展现了强大的潜力。

#### 3.0.7 YOLOv6

YOLOv6 [62] 是由美团视觉 AI 部门于 2022 年发布的一款目标检测模型，旨在平衡检测精度与速度，同时满足工业应用中对高效部署的需求。作为 YOLO 系列的最新发展之一，YOLOv6 在继承经典设计理念的基础上，结合多项前沿技术，对模型架构、训练策略和推理效率进行了全面优化，展现出卓越的性能和广泛的适应性。

在网络架构设计上，YOLOv6 引入了基于 RepVGG [63] 的高效模块 EfficientRep ，通过灵活的模块设计针对不同规模的模型进行了优化。小型模型采用了 RepBlock ，而大型模型则引入

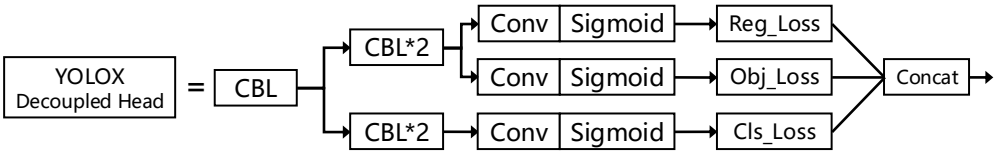


图 7: YOLOX 解耦头

CSPStackRep 模块，显著增强了特征提取能力并提升了计算的并行性。在颈部结构上，YOLOv6 使用增强型路径聚合网络（Enhanced PAN），结合 RepBlock 构建了 RepPAN 模块，从而进一步提升了特征融合效率，使模型在处理复杂场景时表现尤为出色。同时，YOLOv6 的检测头设计为高效解耦头（Efficient Decoupled Head），通过分类任务与边界框回归任务的分离有效缓解了两者之间的特征冲突问题。此外，混合通道策略的引入进一步减少了计算量与推理延迟，为实现实时检测提供了可靠的技术支撑。

YOLOv6 在模型设计中摒弃了传统的锚框机制，转而采用无锚框架，通过直接预测边界框参数简化了模型结构并提升了训练效率。这一无锚设计结合了任务对齐学习（Task-Aligned One-Stage Object Detection, TOOD）[64] 策略，动态优化标签分配，以生成更多高质量的正样本，提高了检测精度。在损失函数方面，YOLOv6 使用 VariFocal Loss 优化分类任务，并通过 SIoU/GIoU 损失函数进一步提升了边界框回归的准确性。

在模型压缩与加速方面，YOLOv6 进行了系统性的优化。通过引入 RepOptimizer 和结合通道蒸馏的量化方法，模型在推理速度与精度之间实现了良好的平衡。同时，自蒸馏策略进一步降低了推理成本，使得模型的部署效率显著提升。为适应多样化的硬件资源和应用需求，YOLOv6 提供了多个版本，从轻量化的 YOLOv6-N（纳米）到高性能的 YOLOv6-L6（超大型），涵盖了从边缘设备到高性能服务器的各种场景需求。

凭借高效的架构设计和灵活的版本选择，YOLOv6 在工业场景中展现了广泛的应用潜力。它不仅在物流分拣、视频监控等领域得到大规模应用，还为高效目标检测提供了新的思路和发展方向。其创新设计不仅推动了目标检测领域的技术进步，也为后续研究与实践奠定了坚实的理论与技术基础。

#### 3.0.8 YOLOv7

YOLOv7 [65] 是由 YOLOv4 和 YOLOR 的研究团队于 2022 年 7 月提出的一种高效目标检测模型，其核心目标是实现检测精度与推理效率的高度平衡。与 YOLOv4 类似，YOLOv7 支持在 MS COCO 数据集上从头开始训练，无需预训练的骨干网络。这一设计不仅简化了模型的训练流程，还显著增强了其在处理多样化数据分布时的适应能力，进一步拓展了模型的实用性。

YOLOv7 的核心架构引入了扩展高效层聚合网络（Extended Efficient Layer Aggregation Network, E-ELAN），这一模块在原有 ELAN 模块的基础上进行了重要改进。通过优化最短和最长梯度路径，E-ELAN 显著提升了深层特征的学习能力，同时加速了模型的收敛过程。其设计通过扩展、洗牌和合并基数等操作，在保持原始梯度路径完整性的同时有效整合了不同组间的特征，从而进一步增强了网络的表达能力。此外，YOLOv7 针对深度模型中输入与输出通道比例变化可能对硬件效率产生的影响，提出了一种基于连接的复合缩放策略。该策略在调整网络宽度和深度时，能够保持网络结构的特性，从而优化

了模型的硬件推理性能，为实际应用中的部署提供了可靠支持。

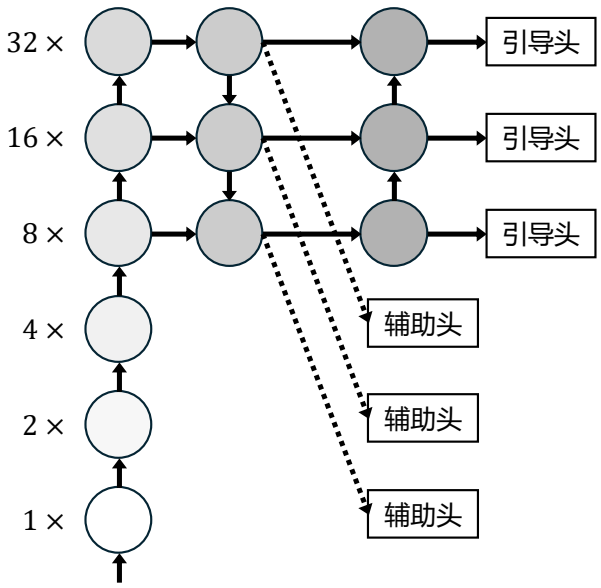


图 8: 辅助检测头

为了进一步提升训练效率和检测精度，YOLOv7 在训练过程中引入了多个工具包（Bag-of-Freebies），显著优化了模型性能。首先，受 YOLOv6 中重参数化卷积（RepConv）思想的启发，YOLOv7 提出了改进版的重参数化卷积 RepConvN，通过移除恒等连接避免了残差连接和密集连接的破坏，从而增强了卷积操作的适配性。其次，如图8所示，YOLOv7 在训练阶段引入了辅助检测头，采用粗略的标签分配机制为主检测头提供辅助学习信号，这种策略有效提升了训练精度，同时不影响模型的推理速度。此外，YOLOv7 采用了批量归一化重新参数化策略，将推理阶段的批量均值和方差整合至卷积层权重中，从而显著提升了推理效率。训练过程中，YOLOv7 还结合了隐式知识增强模型的表示能力，并通过指数移动平均（EMA）技术进一步优化了模型的推理效果和检测精度。

#### 3.0.9 YOLOv8

YOLOv8 [66] 是由 Ultralytics 团队于 2023 年 1 月发布的最新目标检测模型。在继承 YOLO 系列优秀特性的基础上，YOLOv8 通过多项改进显著提升了检测精度和灵活性，同时保持了高效的推理速度。其主干网络基于 CSPDarknet-53，并将传统的 C3 模块替换为具有更多残差连接的 C2f 模块（一种跨阶段部分瓶颈模块）。这一设计优化了梯度流和特征提取能力，在几乎不增加计算成本的情况下，显著提高了模型对小目标和复杂场景的适应能力。此外，YOLOv8 延续使用了 SPPF（快速空间金字塔池化）模块，通过整合多尺度信息，进一步增强了特征表达能力。

在检测头部分，YOLOv8 采用了解耦头（Decoupled Head）结构，将分类任务与边界框定位任务分离，允许两者独立优化，从而显著提升了检测精度。与此同时，YOLOv8 摒弃了传统的锚框机制，转而采用无锚点设计（Anchor-free Design），从根本上简化了训练与推理流程，同时减少了参数量与计算复杂度。在损失函数方面，YOLOv8 引入了分布焦点损失（DFL）和 CIoU 损失，并结合任务对齐分配器优化了正负样本分配策略。这些改进显著增强了目标分类和边界框预测的性能，使模型在精度和效率上达到了更高的平衡。

在训练策略方面，YOLOv8 同样进行了优化。例如，尽管马赛克数据增强在训练早期能够显著提高模型性能，但在训练后期可能对精度产生负面影响。针对这一问题，YOLOv8 采用了动态增强策略，在训练的最后 10 个 epoch 阶段关闭马赛克增强，从而提升了模型在实际应用中的稳定性和精度。此外，为满足不同应用场景和计算资源的需求，YOLOv8 提供了五种规模的模型：N（纳米）、S（小型）、M（中型）、L（大型）和 X（超大型），从轻量化版本到高性能版本均有覆盖，能够灵活适配多样化的应用场景。

### 3.1 评价指标

### 3.2 数据集介绍

### 3.3 YOLO 各版本在 X 光图像违禁品检测中的应用

## 参考文献


[1] Domingo Mery and Christian Pieringer. *Computer vision for x-ray testing: Imaging, systems, image databases, and algorithms*. Springer Nature, 2020.

[2] Besma Abidi, Yue Zheng, Andrei Gribok, and Mongi Abidi. Screener evaluation of pseudo-colored single energy x-ray luggage images. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops*, pages 35–35. IEEE, 2005.

[3] Zhiyu Chen, Yue Zheng, Besma R Abidi, David L Page, and Mongi A Abidi. A combinational approach to the fusion, denoising and enhancement of dual-energy x-ray luggage images. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops*, pages 2–2. IEEE, 2005.

[4] Besma Abidi, Jimin Liang, Mark Mitckes, and Mongi Abidi. Improving the detection of low-density weapons in x-ray luggage scans using image enhancement and novel scene-decluttering techniques. *Journal of electronic imaging*, 13(3):523–538, 2004.

[5] Maneesha Singh and Sameer Singh. Optimizing image enhancement for screening luggage at airports. In *CIHSPS 2005. Proceedings of the 2005 IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety, 2005.*, pages 131–136. IEEE, 2005.

[6] Jer Chan, Paul Evans, and Xun Wang. Enhanced color coding scheme for kinetic depth effect x-ray (kdex) imaging. In *44th Annual 2010 IEEE International Carnahan Conference on Security Technology*, pages 155–160. IEEE, 2010.

[7] Thomas W Rogers, Nicolas Jaccard, Emmanouil D Protonotarios, James Ollier, Edward J Morton, and Lewis D Griffin. Threat image projection (tip) into x-ray images of cargo containers for training humans and machines. In *2016 IEEE International Carnahan Conference on Security Technology (ICCST)*, pages 1–7. IEEE, 2016.

[8] Mark Mitckes. Threat image projection–an overview. *Imaging, Robotics, and Intelligent Systems Laboratory Dept. of Electrical and Computer Engineering. The University of Tennessee*, 2003.

[9] Domingo Mery and Aggelos K Katsaggelos. A logarithmic x-ray imaging model for baggage inspection: Simulation and object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 57–65, 2017.

[10] Victoria Cutler and Susan Paddock. Use of threat image projection (tip) to enhance security performance. In *43rd Annual 2009 International Carnahan Conference on Security Technology*, pages 46–51. IEEE, 2009.

[11] Neelanjan Bhowmik, Qian Wang, Yona Falinie A Gaus, Marcin Szarek, and Toby P Breckon. Evaluating convolutional neural networks for prohibited item detection using real and synthetically composited x-ray imagery. *171,000 190M*, page 15.

[12] Muhammet Baştan, Mohammad Reza Yousefi, and Thomas M Breuel. Visual words on baggage x-ray images. In *Computer Analysis of Images and Patterns: 14th International Conference, CAIP 2011, Seville, Spain, August 29-31, 2011, Proceedings, Part I*, pages 360–368. Springer, 2011.

[13] John A Hartigan and Manchek A Wong. Algorithm as 136: A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics)*, 28(1):100–108, 1979.

[14] Leo Breiman. Random forests. *Machine learning*, 45:5–32, 2001.

[15] Marti A. Hearst, Susan T Dumais, Edgar Osuna, John Platt, and Bernhard Scholkopf. Support vector machines. *IEEE Intelligent Systems and their applications*, 13(4):18–28, 1998.

[16] Mikolaj E Kundegorski, Samet Akçay, Michael Devereux, Andre Mouton, and Toby P Breckon. On using feature descriptors as visual words for object detection within x-ray baggage security screening. 2016.

[17] Domingo Mery, Erick Svec, and Marco Arias. Object recognition in baggage inspection using adaptive sparse representations of x-ray images. In *Image and Video Technology: 7th Pacific-Rim Symposium, PSIVT 2015, Auckland, New Zealand, November 25-27, 2015, Revised Selected Papers 7*, pages 709–720. Springer, 2016.

[18] Jian Zhang, Li Zhang, Ziran Zhao, Yaohong Liu, Jianping Gu, Qiang Li, and Duokun Zhang. Joint shape and texture based x-ray cargo image classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 266–273, 2014.

[19] Nicolas Jaccard, Thomas W Rogers, and Lewis D Griffin. Automated detection of cars in transmission x-ray images of freight containers. In *2014 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 387–392. IEEE, 2014.

[20] Ning Zhang and Jinfu Zhu. A study of x-ray machine image local semantic features extraction model based on bag-of-words for airport security. *International journal on smart sensing and intelligent systems*, 8(1):45–64, 2015.

[21] Domingo Mery, German Mondragon, Vladimir Riffo, and Irene Zuccar. Detection of regular objects in baggage using multiple x-ray views. *Insight-Non-Destructive Testing and Condition Monitoring*, 55(1):16–20, 2013.

[22] Muhammet Baştan. Multi-view object detection in dual-energy x-ray images. *Machine Vision and Applications*, 26(7):1045–1060, 2015.

[23] Muhammet Bastan, Wonmin Byeon, Thomas M Breuel, et al. Object recognition in multi-view dual energy x-ray images. In *BMVC*, volume 1, page 11, 2013.

[24] Ludwig Schmidt-Hackenberg, Mohammad Reza Yousefi, and Thomas M Breuel. Visual cortex inspired features for object detection in x-ray images. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pages 2573–2576. IEEE, 2012.

[25] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. A sparse texture representation using local affine regions. *IEEE transactions on pattern analysis and machine intelligence*, 27(8):1265–1278, 2005.

[26] Thorsten Franzel, Uwe Schmidt, and Stefan Roth. Object detection in multi-view x-ray images. In *Joint DAGM (German association for pattern recognition) and OAGM symposium*, pages 144–154. Springer, 2012.

[27] M Stefan and A Schwaninger. Human-machine interaction in x-ray screening. In *Proceedings of the Carnahan Conference on Security Technology*, volume 41, pages 13–19, 2007.

[28] Claudia Christina Von Bastian, Adrian Schwaninger, and Stefan Michel. Do multi-view x-ray systems improve x-ray image interpretation in airport security screening? 2010.

[29] R Paranjape, M Sluser, and E Runtz. Segmentation of handguns in dual energy x-ray imagery of passenger carry-on baggage. In *Conference Proceedings. IEEE Canadian Conference on Electrical and Computer Engineering (Cat. No. 98TH8341)*, volume 1, pages 377–380. IEEE, 1998.

[30] M Sluser and Raman Paranjape. Model-based probabilistic relaxation segmentation applied to threat detection in airport x-ray imagery. In *Engineering Solutions for the Next Millennium. 1999 IEEE Canadian Conference on Electrical and Computer Engineering (Cat. No. 99TH8411)*, volume 2, pages 720–726. IEEE, 1999.

[31] Jianli Ding, Yuanxiang Li, Xing Xu, and Lingling Wang. X-ray image segmentation by attribute relational graph matching. In *2006 8th international Conference on Signal Processing*, volume 2. IEEE, 2006.

[32] Lingling Wang, Yuanxiang Li, Jianli Ding, and Kangshun Li. Structural x-ray image segmentation for threat detection by attribute relational graph matching. In *2005 International Conference on Neural Networks and Brain*, volume 2, pages 1206–1211. IEEE, 2005.

[33] Noeleene Mallia-Parfitt and Georgios Giasemidis. Graph clustering and variational image segmentation for automated firearm detection in x-ray images. *IET Image Processing*, 13(7):1105–1114, 2019.

[34] Samet Akçay, Mikolaj E Kundegorski, Michael Devereux, and Toby P Breckon. Transfer learning using convolutional neural networks for object classification within x-ray baggage security imagery. In *2016 IEEE*

*International Conference on Image Processing (ICIP)*, pages 1057–1061. IEEE, 2016.

[35] Nicolas Jaccard, Thomas W Rogers, Edward J Morton, and Lewis D Griffin. Detection of concealed cars in complex cargo x-ray imagery using deep learning. *Journal of X-ray Science and Technology*, 25(3):323–339, 2017.

[36] Lewis D Griffin, Martin Lillholm, Mike Crosier, and Justus Van Sande. Basic image features (bifs) arising from approximate symmetry type. In *Scale Space and Variational Methods in Computer Vision: Second International Conference, SSVM 2009, Voss, Norway, June 1-5, 2009. Proceedings 2*, pages 343–355. Springer, 2009.

[37] Anna Bosch, Andrew Zisserman, and Xavier Munoz. Representing shape with a spatial pyramid kernel. In *Proceedings of the 6th ACM international conference on Image and video retrieval*, pages 401–408, 2007.

[38] Nicolas Jaccard, Thomas W Rogers, Edward J Morton, and Lewis D Griffin. Automated detection of smuggled high-risk security threats using deep learning. In *7th International Conference on Imaging for Crime Detection and Prevention (ICDP 2016)*, pages 1–6. IET, 2016.

[39] Zihao Zhao, Haigang Zhang, and Jinfeng Yang. A gan-based image generation method for x-ray security prohibited items. In *Pattern Recognition and Computer Vision: First Chinese Conference, PRCV 2018, Guangzhou, China, November 23-26, 2018, Proceedings, Part I 1*, pages 420–430. Springer, 2018.

[40] Jinfeng Yang, Zihao Zhao, Haigang Zhang, and Yihua Shi. Data augmentation for x-ray prohibited item images using generative adversarial networks. *IEEE Access*, 7:28894–28902, 2019.

[41] Caijing Miao, Lingxi Xie, Fang Wan, Chi Su, Hongye Liu, Jianbin Jiao, and Qixiang Ye. Sixray: A large-scale security inspection x-ray benchmark for prohibited item discovery in overlapping images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2119–2128, 2019.

[42] Yanlu Wei, Renshuai Tao, Zhangjie Wu, Yuqing Ma, Libo Zhang, and Xianglong Liu. Occluded prohibited items detection: An x-ray security inspection benchmark and de-occlusion attention module. In *Proceedings of the 28th ACM international conference on multimedia*, pages 138–146, 2020.

[43] Matthew Caldwell, M Ransley, Thomas W Rogers, and Lewis D Griffin. Transferring x-ray based automated threat detection between scanners with different energies and resolution. In *Counterterrorism, Crime Fighting, Forensics, and Surveillance Technologies*, volume 10441, pages 130–139. SPIE, 2017.

[44] Yona Falinie A Gaus, Neelanjan Bhowmik, Samet Akcay, and Toby Breckon. Evaluating the transferability and adversarial discrimination of convolutional neural net-

works for threat object detection and classification within x-ray security imagery. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pages 420–425. IEEE, 2019.

[45] Kevin J Liang, Geert Heilmann, Christopher Gregory, Souleymane O Diallo, David Carlson, Gregory P Spell, John B Sigman, Kris Roe, and Lawrence Carin. Automatic threat recognition of prohibited items at aviation checkpoint with x-ray imaging: a deep learning approach. In *Anomaly Detection and Imaging with X-Rays (ADIX) III*, volume 10632, page 1063203. SPIE, 2018.

[46] Jan-Martin O Steitz, Faraz Saeedan, and Stefan Roth. Multi-view x-ray r-cnn. In *German Conference on Pattern Recognition*, pages 153–168. Springer, 2018.

[47] J Redmon. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.

[48] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.

[49] Ross Girshick. Fast r-cnn. *arXiv preprint arXiv:1504.08083*, 2015.

[50] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence*, 39(6):1137–1149, 2016.

[51] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[52] Alexander Neubeck and Luc Van Gool. Efficient non-maximum suppression. In *18th international conference on pattern recognition (ICPR'06)*, volume 3, pages 850–855. IEEE, 2006.

[53] Joseph Redmon and Ali Farhadi. Yolo9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7263–7271, 2017.

[54] Ali Farhadi and Joseph Redmon. Yolov3: An incremental improvement. In *Computer vision and pattern recognition*, volume 1804, pages 1–6. Springer Berlin/Heidelberg, Germany, 2018.

[55] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[56] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125, 2017.

[57] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 37(9):1904–1916, 2015.

[58] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*, 2020.

[59] Shu Liu, Lu Qi, Haifang Qin, Jianping Shi, and Jiaya Jia. Path aggregation network for instance segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8759–8768, 2018.

[60] Glenn Jocher. Yolov5 by ultralytics. `https://github.com/ultralytics/yolov5`, 2020.

[61] Z Ge. Yolox: Exceeding yolo series in 2021. *arXiv preprint arXiv:2107.08430*, 2021.

[62] Chuyi Li, Lulu Li, Hongliang Jiang, Kaiheng Weng, Yifei Geng, Liang Li, Zaidan Ke, Qingyuan Li, Meng Cheng, Weiqiang Nie, et al. Yolov6: A single-stage object detection framework for industrial applications. *arXiv preprint arXiv:2209.02976*, 2022.

[63] Xiaohan Ding, Xiangyu Zhang, Ningning Ma, Jungong Han, Guiguang Ding, and Jian Sun. Repvgg: Making vgg-style convnets great again. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13733–13742, 2021.

[64] Chengjian Feng, Yujie Zhong, Yu Gao, Matthew R Scott, and Weilin Huang. Tood: Task-aligned one-stage object detection. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3490–3499. IEEE Computer Society, 2021.

[65] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7464–7475, 2023.

[66] Jing Qiu Jonathan Choa Glenn Jocher, Ayush Chaurasia. Ultralytics yolov8: Cutting-edge, real-time object detection. `https://github.com/ultralytics/ultralytics`, 2023.