

融合双向路由注意力的多尺度X光违禁品检测

王若璇, 李野, 赵鹏

长春理工大学物理学院, 吉林 长春

收稿日期: 2024年2月16日; 录用日期: 2024年3月18日; 发布日期: 2024年3月26日

摘要

针对违禁品检测中存在的复杂背景干扰、物体间的重叠遮挡和多尺度变化问题, 提出一种基于改进YOLOv7的X射线违禁品目标检测算法。首先, 在主干中引入MBConv, 以更有效的捕获全局信息; 其次在特征融合网络中加入RFE模块, 以增加特征图的感受野, 从而提高违禁品多尺度检测的准确性。并设计出一种ELAN-BiF模块, 用于抑制复杂背景干扰, 使网络提取不同尺度的物品特征; 为了提高小目标物体的检测精度, 增加了一个微小物体检测头; 最后, 结合CARAFE上采样和Mish激活函数来提高网络对重叠和遮挡对象的识别能力, 并提升在正负样本不平衡情况下的检测能力。结果表明, 改进后的模型在SIXray_OOD数据集上进行测试, 该方法map达到了95.2%, 比原模型提高4.9%, 比其他主流检测模型在违禁品检测任务上具有更好的优越性。

关键词

X射线图像, 双向路由注意力, 小目标检测层, YOLOv7

Multiscale X-Ray Contraband Detection Incorporating Bidirectional Routing Attention

Ruoxuan Wang, Ye Li, Peng Zhao

College of Physics, Changchun University of Science and Technology, Changchun Jilin

Received: Feb. 16th, 2024; accepted: Mar. 18th, 2024; published: Mar. 26th, 2024

Abstract

Aiming at the problems of complex background interference, overlapping occlusion between objects and multi-scale change in contraband detection, an X-ray contraband target detection algo-

文章引用: 王若璇, 李野, 赵鹏. 融合双向路由注意力的多尺度 X 光违禁品检测[J]. 计算机科学与应用, 2024, 14(3): 78-95. DOI: 10.12677/csa.2024.143060

rithm based on improved YOLOv7 was proposed. Firstly, MBConv is introduced into the backbone to capture the global information more efficiently; secondly, an RFE module is added into the feature fusion network to increase the receptive field of the feature map, so as to improve the accuracy of contraband multi-scale detection. And an ELAN-BiF module is designed to suppress the complex background interference so that the network extracts the features of items at different scales; In order to improve the detection accuracy of small target objects, a small object detection head has been added; finally, CARAFE up-sampling and Mish activation function are combined to improve the network's ability to recognize overlapping and occluded objects and enhance the detection ability in the case of positive and negative sample imbalance situation. The results show that the improved model is tested on the SIXray_OOD dataset, and the method achieves a map of 95.2%, which is 4.9% better than the original model, and has a better superiority than other mainstream detection models in the contra-band detection task.

Keywords

X-Ray Images, Bidirectional Routing Attention, Small Target Detection Layer, YOLOv7

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在火车站、体育馆、地铁、机场等公共场所中，使用 X 射线安检机，可以有效避免枪支、管制刀具等违禁品对人员和公共财产的威胁与干扰[1]。但目前大多 X 射线安检机主要依靠人工视觉从拍摄的图像中识别违禁品，存在人为疏忽而导致漏检和错检的问题。

X 射线安检图像是通过目标检测算法来判断是否存在违禁品，并在安检图像中标记出它的位置和类别[2]。然而，X 射线安检图像有以下特点：1) 复杂背景：由于 X 射线成像与物体的密度和厚度有关，安检设备根据物质的性质呈现不同的颜色，当背景与违禁品厚度和密度相近时，会干扰违禁品的特征学习；2) 物体间重叠遮挡：物品在 X 射线投影下的形状会严重畸变，被检物品的随机放置会导致物体间的互相遮挡，增加违禁品识别的困难[3]；3) 违禁品具有多尺度变化：同类物品在图片中会呈现不同的尺度，不同位置的镜头监测到的目标也会呈现出尺度差异。违禁品检测安检图像如图 1 所示。

近年来，随着深度学习在计算机视觉领域发展迅速，目标检测算法在 X 射线安检图像违禁品识别中发挥了重要作用，有效提高了违禁品识别的准确率。Akay [4]等人首次将卷积神经网络应用于 X 射线安检图像，通过迁移学习对 X 光行李图像进行分类，再对特定物品——手枪进行检测，但识别类别单一；Gao 等[5]提出了一种双卷积神经网络架构，用于复杂 X 射线安检图像中的自动异常检测。Ge 等[6]人提出解耦检测头和无锚框的 YOLOX 框架，虽定位多尺度小目标准确，但泛化性不足；Zhang Y [7]等人在 FSSD 检测模型上添加扩张卷积并引入残差连接以避免过拟合现象，但检测精度仍有提升空间；Zhu X [8]等人提出基于注意力机制的多尺度检测网络(AMOD-Net)，通过融入通道注意力、构建深度特征融合结构，来解决 X 光违禁品图像中的重叠遮挡问题，但多尺度检测网络导致特征模糊，从而影响检测精度：

上述研究成果极大推动了深度学习在 X 光违禁品检测中的应用，提高了违禁品检测的准确性。但 X 射线安检图像仍存在违禁品的形状多尺度、重叠遮挡严重、复杂背景干扰等急需解决的关键问题，目前的检测精度和速度仍然无法满足实际应用的要求。

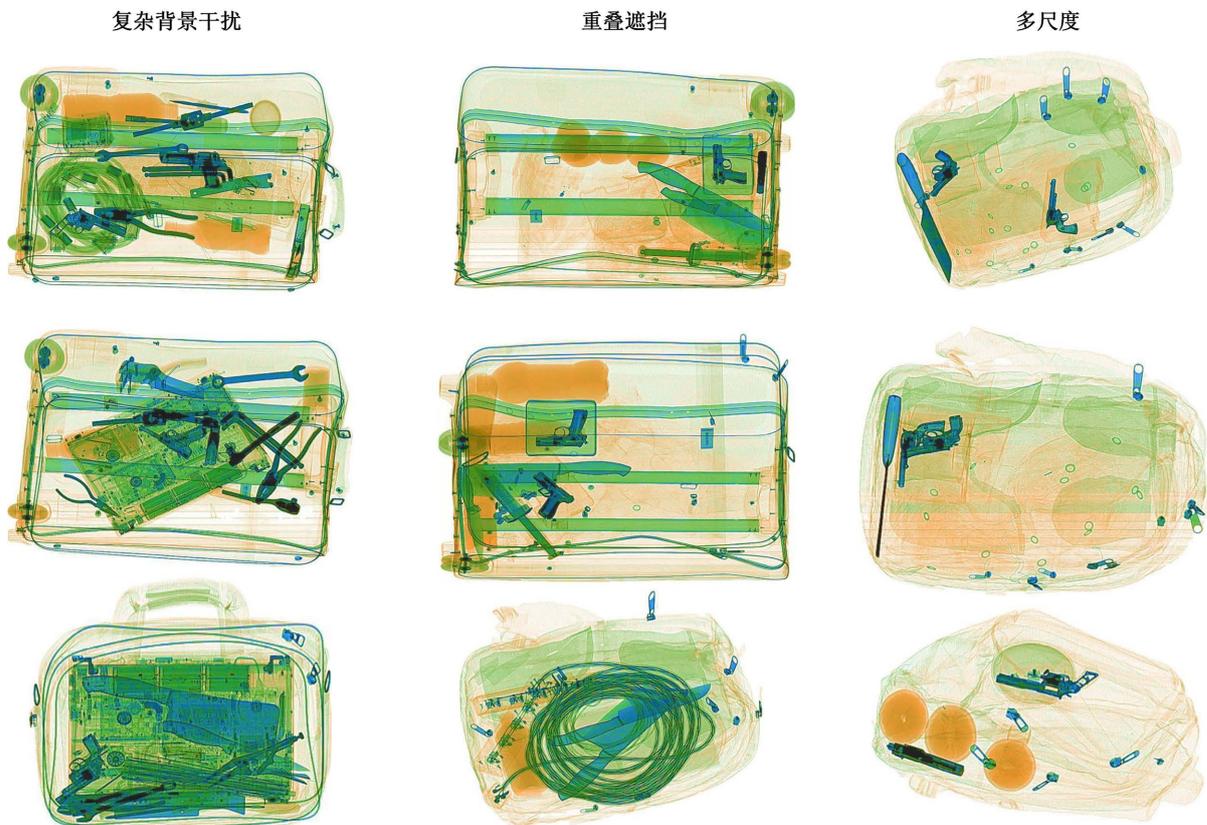


Figure 1. Sample of X-ray image contraband screening

图 1. X 射线图像违禁品安检样例

针对上述问题，本文基于 YOLOv7 [9]模型来进行改进，提出了 MBRC-YOLOv7 算法，进一步优化了检测的精度和速度。主要贡献如下：

1) 针对违禁品的多尺度性变换，通过在主干网络中将 ELAN (Efficient Layer Aggregation Networks) 结构替换成 MBConv (Mobile Inverted Residual Bottleneck Block)结构，对输入的特征进行全局建模，提高模型对违禁品关键特征的提取能力；使用 CARAFE (Content-Aware ReAssembly of FEatures)实现上采样，充分利用多级特征的语义与空间信息；

2) 针对违禁品检测中的复杂背景干扰，在 Neck 中将 ELAN-H (Enhanced Local Attention Network with Hierarchical Structure)中融合双层路由注意力机制模块，抑制复杂背景干扰，使网络提取不同尺度的物品特征；并加入 RFE (Recursive Feature Elimination)模块，以扩大图的感受野，并减少上采样中有效信息的丢失，丰富图像的语义信息；

3) 针对物体间的重叠遮挡，在 Head 中加入一个由低层、高分辨率的特征图生成的小目标检测层，提高对小目标物体感知能力；引入 Mish 激活函数来提高网络识别重叠和遮挡对象的能力。

2. X 光违禁品图像检测模型

2.1. 网络模型

针对违禁品检测实时性需求，主要考虑检测速度更快的一阶段检测器，因此，本文以 YOLOv7 为基础进行改进，YOLOv7 主要由 Backbone、Neck 和 Head 三部分构成。其中 Backbone 由多个卷积层、ELAN、MPCov (MaxPooling and Conv)等模块组成，用于图像特征的提取[10]。ELAN 模块是一个高效的层聚合

网络，可在不破坏原有梯度的情况下增强网络的学习能力。MPCConv 将 MaxPool 层添加到 CBS 层，形成上下分支来增强网络的特征提取能力。Neck 通过 PANet (Path Aggregation Network)和 SPPSCPC (Spatial Pyramid Pooling Cross Stage Partial Concat)结构，通过对主干部分获得的三个有效特征层进行上采样和下采样来实现特征融合。在 Head 中，利用 REPCov 结构调整 PAFPN 输出特征的图像通道数，最后通过卷积完成预测和输出。改进后的模型为 MBRC-YOLOv7，其网络结构如图 2 所示。

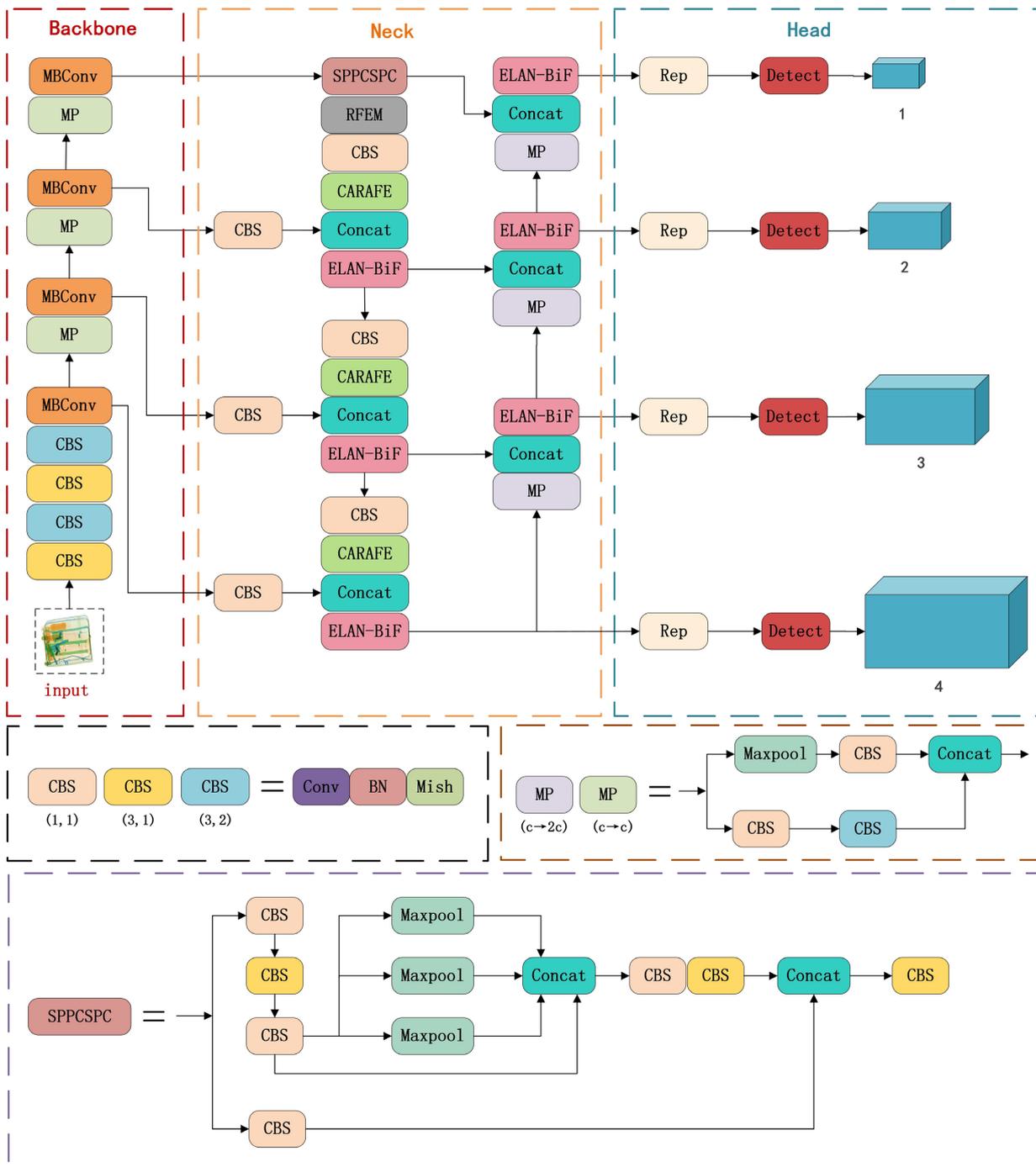


Figure 2. MBRC-YOLOv7 network structure diagram
图 2. MBRC-YOLOv7 网络结构图

2.2. 特征提取网络

为了提高对违禁品多尺度信息的提取能力，在主干网络中引入移动反向残差瓶颈卷积模块 **MBConv** [11]来代替 **ELAN** 模块。**MBConv** 是一种轻量级的卷积，在保持较高精度的同时，还可以减少计算量和内存消耗。它采用多个跨层连接和 **SE** 注意力机制来增加模型的鲁棒性和泛化能力，能够有效提取多尺度空间特征信息，其结构如图 3 所示。

MBConv 首先使用 1×1 卷积核对输入特征图进行通道数的扩张，通过扩大特征的通道数量，有助于提取物品的多尺度特征。然后使用深度可分离卷积进行特征提取和通道转换。在深度卷积阶段，使用 3×3 卷积核对输入的每个通道单独进行卷积，并为每个通道生成一个特征图。在逐点卷积阶段，使用 1×1 的卷积核对深度卷积的输出进行压缩，将通道数调整为所需的数量，并通过 **SE** (**Squeeze-and-Excitation Networks**)注意力机制来增强网络对不同通道特征的感知能力。最后将重新加权的特征图与输入特征图进行结合，得到具有通道注意力的特征图。此外，**MBConv** 还引入了残差连接，将输入特征图与深度可分离卷积的输出相加，促进了底层特征的直接传递，解决深层网络训练中的梯度消失问题，使得改进后的网络用更少的参数获取到更好的效果。

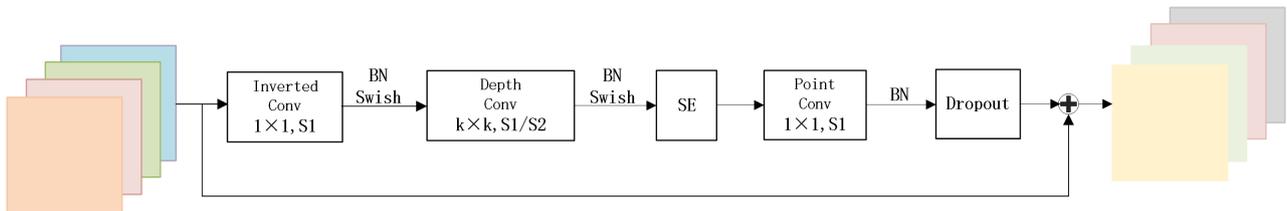


Figure 3. MBConv structure
图 3. MBConv 结构

2.3. 双层路由注意力机制

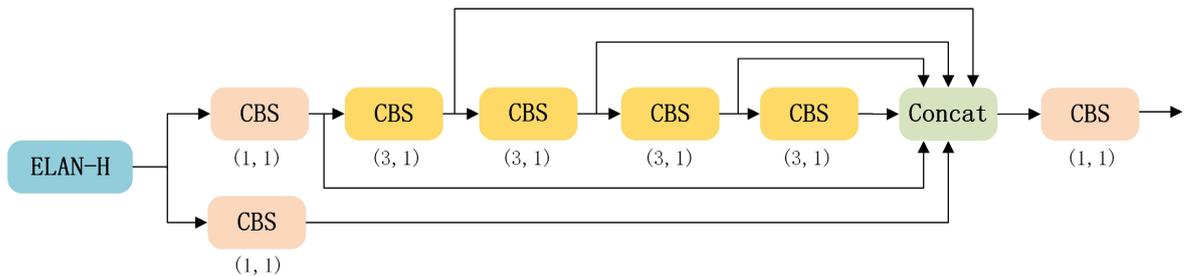


Figure 4. Original ELAN-H structure (1, 1) CBS
图 4. 原 ELAN-H 结构

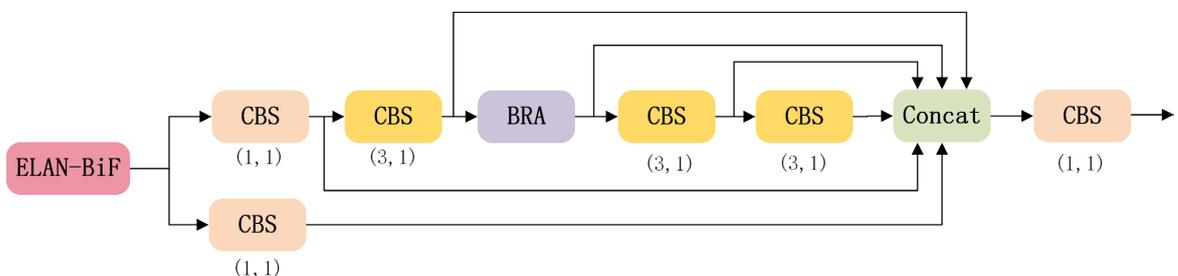


Figure 5. Structure of ELAN-BiF
图 5. ELAN-BiF 结构

由于 X 光安检机检测的图像背景复杂且物品间高度重叠,许多检测模型抑制背景信息的能力较差。为了使模型更多地关注输入特征中的关键信息,本文设计出一种双层路由注意力机制[12] (Bi-Level Routing Attention, BRA)与 ELAN-H 融合 的模块——ELAN-BiF。BRA 通过将局部和全局级别的注意力机制进行叠加和融合,从而实现多层级的特征交互,有助于提高网络对多尺度特征的感知和表征能力,并减少了计算和存储消耗。ELAN-H 原始结构和改进后的 ELAN-BiF 结构如图 4、图 5 所示。

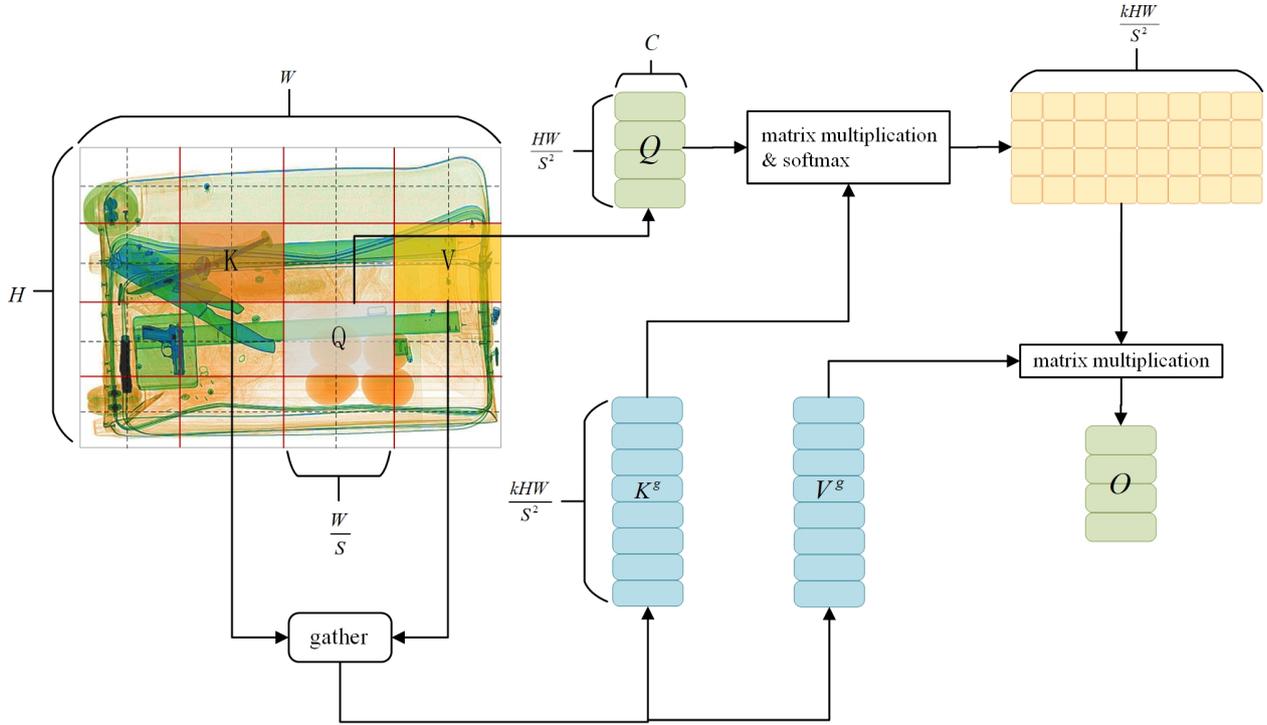


Figure 6. Structure of the Bi-Level Routing Attention

图 6. 双层路由注意力机制结构

如图 6 所示,输入特征图为 $X \in \mathbb{R}^{H \times W \times C}$, 首先被分割成 $S \times S$ 个子区域, 每个区域包含 $\frac{H \times W}{S^2}$ 个特征向量。通过改变 X 的形状, 得到维度为 $X^r \in \mathbb{R}^{\frac{S^2 \times H \times W}{S^2} \times C}$ 的特征图。再对特征向量进行线性变换, 得到三个矩阵 Q 、 K 和 V 。其计算公式如(1)所示。

$$Q = X^r W^Q, K = X^r W^K, V = X^r W^V \quad (1)$$

然后, 通过构造有向图来定位给定部分的相关区域, 从而获得区域与区域之间的注意力关系。首先, 分别计算每个划分区域中的 Q 和 K 的平均值, 得到 Q^r 和 $K^r \in \mathbb{R}^{S^2 \times C}$, 根据公式(2)来计算区域之间语义相关性的邻接矩阵 A^r , 再将 A^r 中不相关的标记过滤掉, 保留 k 个具有高相关性的查询区域来获得路由索引矩阵 $I^r \in \mathbb{N}^{S^2 \times k}$, 如公式(3)所示。

$$A^r = Q^r (K^r)^T \quad (2)$$

$$I^r = \text{topkIndex}(A^r) \quad (3)$$

最后, 收集 k 个区域内所有 K 和 V 的张量, 通过公式(4)进行张量计算得到 K^g 和 V^g ; 再对 K^g 和 V^g 进行注意力处理, 并添加局部上下文增强项 $LE(V)$, 得到输出张量 O 。其公式如式(5)所示:

$$K^s = \text{gather}(K, I^r), V^s = \text{gather}(V, I^r) \tag{4}$$

$$O = \text{Attention}(Q, K^s, V^s) + \text{LE}(V) \tag{5}$$

上层路由和下层路由的结合，通过建模特征依赖关系和融合上下文信息来提高特征表示，从而能够提高违禁品分类准确率，更好地区分不同违禁品类别之间的特征差异。

2.4. CARAFE 上采样

原始 YOLOv7 网络中的多尺度特征融合采用最近邻插值来完成上采样，仅是在原始图像的像素之间插入新元素以增加像素数量，因此，得到的特征图具有较小的感知范围，被检测物体的特征信息容易丢失。为了解决上采样中语义相关性不足的问题，采用一种轻量级且高效的上采样算子 CARAFE [13]。该方法根据输入特征内容预测上采样内核的大小，同时引入少量的参数。利用预测结果指导上采样过程，可以在上采样操作时增大感受野，并充分利用特征的语义信息提取目标特征。

CARAFE 上采样由两部分组成，如图 7 所示。一部分为内核预测模块，生成用于重组计算的核的权重；而内容感知重组模块是根据计算出的权重对特征进行重组。

首先通过 1×1 卷积层将 $H \times W \times C$ 特征图的通道数压缩为 C_m ，以减少上采样操作的计算量。再利用内核预测模块中的卷积层，输入通道数为 C_m 的特征，并将输出通道数改为 $e^2 \times k_{up}^2$ (其中 k_{up}^2 表示预测上采样卷积核的尺寸)。然后沿通道维度进行扩展，使其形状变为 $\sigma H \times \sigma W \times k_{up}^2$ ，再通过 Softmax 函数进行归一化，得到上采样卷积核的加权和为 1。最后，在内容感知重组模块中，将 $H \times W \times C$ 形状的特征图与预测的上采样卷积核进行点乘运算，获得 $\sigma H \times \sigma W \times C$ 形状的上采样输出结果。

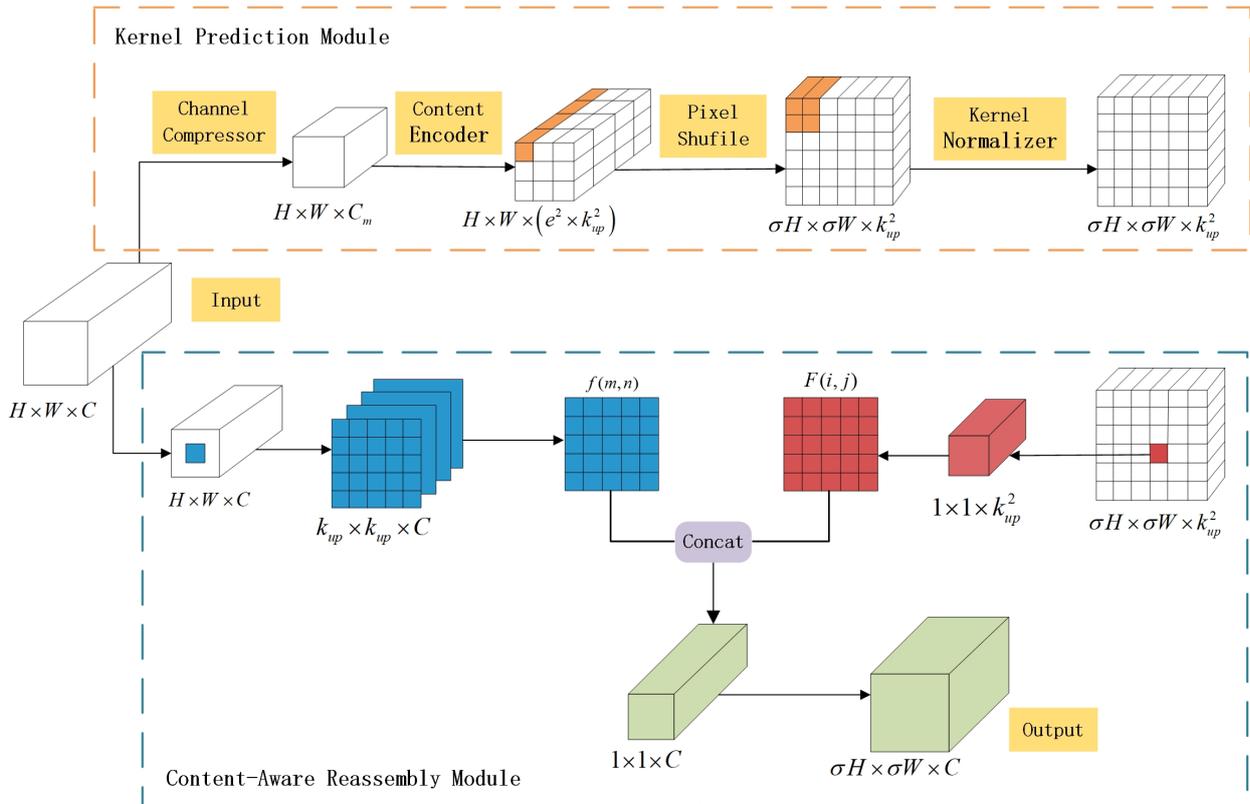


Figure 7. CARAFE schematic diagram.

图 7. CARAFE 原理图

2.5. Scale-Aware RFE Model

针对随着网络模型的深入，原网络感受野变窄，导致小目标漏检率较高的问题，在特征融合区域引入 Scale-Aware RFE [14]模块，以提高网络模型对图像中小物体的感受野。RFE 通过使用四个不同扩张卷积率的分支，每个分支都共享权重，既降低了过拟合的风险又可以充分利用每个样本特征。

如图 8 所示，RFE 由多分支结构和聚集加权层两个部分构成，多分支结构使用三个不同的分支来扩张卷积率，每个分支都使用固定核大小的 3×3 卷积。然后使用残差网络连接，来解决训练过程中的梯度爆炸和消失问题。聚集加权层用于收集来自不同分支的信息并对每个分支的特征进行加权。这种加权操作可以平衡分支之间的表示，以增强整体特征的表达能力。

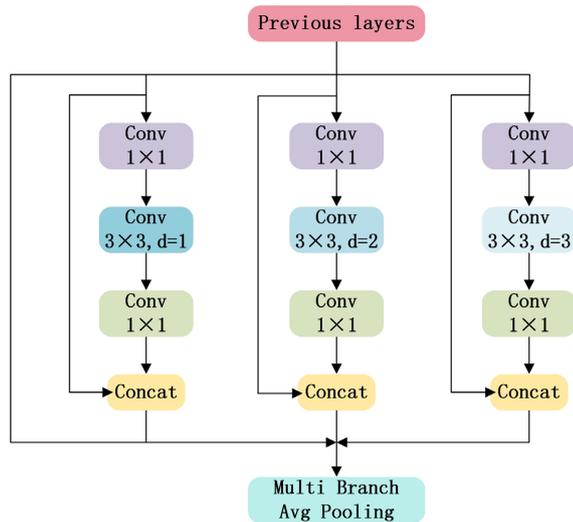


Figure 8. Architecture diagram of RFE

图 8. RFE 的架构图

2.6. 激活函数 Mish

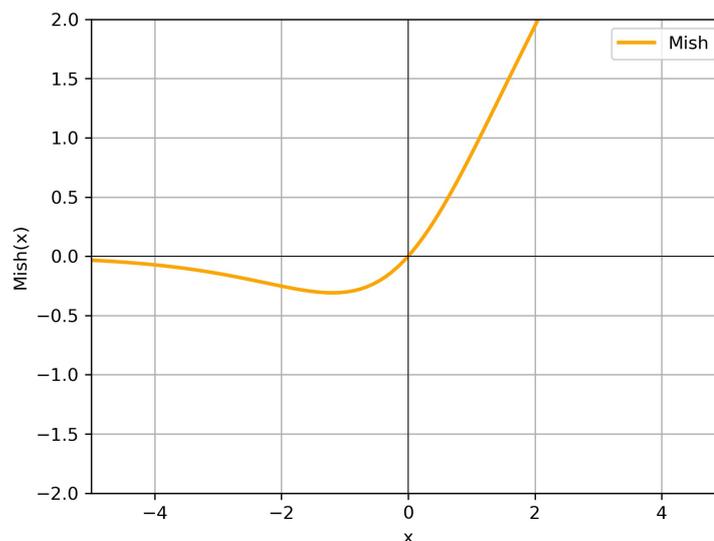


Figure 9. Mish activation function

图 9. Mish 激活函数

Mish [15]是一种非线性激活函数,它具有无上界特点,从而避免了由于封顶而导致训练结果的饱和。而有下界的属性,不仅可以加快训练过程,还有助于实现强正则化效果。Mish 函数公式如(6)所示。

$$\text{Mish}(x) = x \cdot \tanh(\ln(1 + e^x)) \quad (6)$$

由图 9 Mish 函数图可以看出,输入值在接近零的区域几乎保持线性,有助于减少梯度消失和爆炸的问题;而在远离零的区域产生非线性响应,这种非线性增强的特性可以帮助卷积神经网络捕捉更多复杂的特征。Mish 函数在接近零的区域表现更线性,并且在非零区域的曲线更平滑,这种平滑的激活函数允许更准确的数据进入神经网络,从而提高了检测的准确度。因此,Mish 在处理违禁品检测中的重叠遮挡和正负样本不平衡方面具有优越的能力。

2.7. 小目标检测层

为增加复杂背景下小目标特征的捕获量,本文在 Neck 增加了一个由低层、高分辨率的特征图生成的小目标检测层 P2,以及相对应尺度的检测头,如图 2 所示。P2 包含着大量小目标以及微小目标的细节信息,在依次传递并融合特征的过程中,能给深层特征图提供更充足的小目标信息,进而提高复杂背景下小目标检测能力。

通过将四种不同尺度的特征层融合,可以获取丰富的小目标纹理和细节特征,再结合新增的小目标检测头,能够高效提升复杂背景下小目标的检测率和准确率。

3. 实验

3.1. 实验参数设置

本文的算法是在 Windows10 系统下进行,使用 PyTorch 1.8.1 搭建网络框架,在 NVIDIA GeForce RTX 3090 GPU 进行训练和验证。在训练阶段使用部分预训练的 YOLOv7 模型。模型在数据集上进行 300 次 epoch 的训练,使用 SGD 优化器进行训练,初始学习率为 0.01。输入图像大小为 512×512 像素,批处理大小为 16,权重衰减系数为 $5e-4$,采用余弦退火算法(Cosine An-nealing)调整学习率,每个基线网络架构都使用相同的优化方案进行训练。

3.2. 数据集及预处理

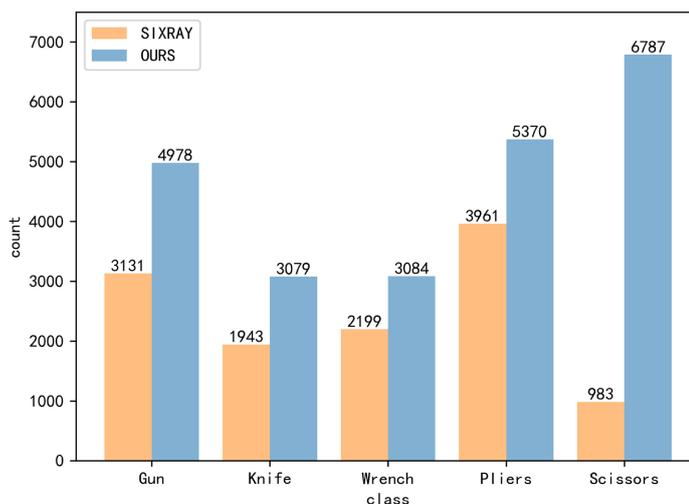


Figure 10. Comparison of SIXray_OOD and SIXray datasets
图 10. SIXray_OOD 与 SIXray 数据集对比图

实验中, 考虑实际安检场景违禁品图像数据的不平衡问题, 本文对 SIXray 数据集[16]进行预处理, 构建了一个为 SIXray_OOD 的新数据集, 如图 10 所示。其中包含了共计 9893 张图像, 六种不同类别的违禁品, 分别是枪、刀、扳手、钳子、剪刀、锤子。由于锤子样本数量明显低于其他五个类别, 故本文保留其余 5 类重新进行人工标注, 并对筛选后的数据集通过旋转和缩放、锐化、添加噪声等措施制作成实验所需的数据集 SIXray_OOD, 有助于降低模型过拟合的风险, 并提高模型的泛化能力。其中枪的图片有 4978 张、刀 3079 张、扳手 3084 张、钳子 5370 张、剪刀 6787 张。本文设置训练集、验证集、测试集的比例为 8:1:1。

3.3. 评价指标

实验采用准确率(Precision)、召回率(Recall)、平均准确率 AP (Average Precision)、平均准确率均值 mAP、类别识别平均准确率均值 mAP@0.5(mean Average Precision, IoU 阈值取大于 0.5)、FPS (Frames per second)作为评价指标, 计算方式为:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (7)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (8)$$

$$\text{AP} = \int_0^1 p(r) dr \quad (9)$$

$$m\text{AP} = \frac{1}{n} \sum_{i=1}^n \text{AP}_i \quad (10)$$

$$\text{FPS} = \frac{1}{t} \quad (11)$$

其中, TP 为被正确识别为阳性的正样本数量, FP 为被错误识别为阳性的负样本数量, FN 为被错误识别为阴性的正样本数量, n 为类别总数, i 为当前类别的编号。P-R 曲线由准确率和召回率组成, 其曲线下方的面积为 AP。mAP@0.5 (IoU 阈值取 0.5)是所有类别的识别平均准确率 AP 的平均值。FPS 表示平均每秒检测到的图像数量。

3.4. 结果分析

3.4.1. 消融实验

为了验证的每个改进部分对网络性能都有较好效果, 通过逐步添加改进的模块, 在 SIXray_OOD 数据集上设计 12 组消融实验, 分别验证各改进点的有效性, “√”表示某一组实验应用的改进点, 各项评价指标见表 1。

实验结果表明, 第一组为原算法的实验结果, mAP@0.5 值为 90.3%, FPS 为 86, 作为评价指标基线。第二组为在骨干网络加入 MBCConv 后, 模型的评价指标 mAP@0.5 增加了 0.9%, 证明了 MBCConv 模块可以提升模型的违禁品检测能力。第三组只在 Neck 部分引入双层路由注意力 BRA, mAP@0.5 达到 92.3, 推理速度也有所提升, 说明 BRA 利用瓶颈结构和通道注意力机制, 可以更好的捕获目标特征; 第四组将 MBCConv 与 BRA 结合在一起进行训练, 虽然推理速度有所降低, 但相比较第二组实验平均准确率提高了 1.2%; CARAFE 上采样通过自适应权重和空间变换, 可以更好的利用低分辨率特征图信息。第五组只引入 CARAFE 上采样, 检测精度有小幅上升, 推理速度提升到 89; 第六组将 MBCConv、BRA 与 CARAFE 相融合, 检测精度提升 2.5%。RFE 通过多分支结构和聚集加权层来扩张

卷积,以增大感受野,来提升违禁品中小目标检测的准确率。第七组通过加入 RFE 增大特征图感受野,检测精度提升 1.1%; Mish 具有梯度平滑性与非线性的特性,有助于减轻训练过程中梯度消失问题。第九组把激活函数替换为 Mish,在保证没有增加参数数量和计算量的条件下,加快网络模型收敛,同时 mAP@0.5 值提升 0.3%。第十二组是在头部加入大尺度特征层集合所有改进点后, mAP@0.5 值提升了 4.9%。

Table 1. Ablation experiments with different modules on the SIXray_OOD dataset

表 1. 不同模块在 SIXray_OOD 数据集上的消融实验

Number	MBCConv	BRA	CARAFE	RFE	Mish	P2	mAP@0.5(%)	FPS (ms)
P1	-	-	-	-	-	-	90.3	86
P2	√	-	-	-	-	-	91.2	92
P3	-	√	-	-	-	-	92.1	90
P4	√	√	-	-	-	-	92.4	91
P5	-	-	√	-	-	-	91.6	89
P6	√	√	√	-	-	-	92.8	93
P7	-	-	-	√	-	-	91.4	90
P8	√	√	√	√	-	-	93.7	94
P9	-	-	-	-	√	-	90.6	87
P10	√	√	√	√	√	-	94.1	93
P11	-	-	-	-	-	√	92	90
P12	√	√	√	√	√	√	95.2	92

3.4.2. 对比实验

为了验证本文模型 MBRC-YOLOv7 的有效性,本文选取 One-Stage 检测算法中 SSD [17]、YOLOv5s [18]、DETR [19]、YOLOv7、GF-YOLOv7 [20]以及 Two-Stage 中的 Faster RCNN [21]、Mask R-CNNR [22] 模型,总共 7 种目标检测模型在 SIXray_OOD 数据集上进行对比试验。如表 2 所示,所提出的算法的平均检测精度达到了 95.2%的准确率,明显高于其他算法,枪支、刀、扳手、钳子、剪刀的检测精度也高于其他主流检测算法。与 YOLOv7 相比,枪支 AP@50 提高了 4%,刀 AP@50 提高了 10%,扳手 AP@50 提高了 10%,钳子 AP@50 提高了 10%,剪刀 AP@50 提高了 10%。

综上可知,本文改进的 MBRC-YOLOv7 模型,能在保证平均检测精度最高的同时,推理速度也相对较快,证明了 MBRC-YOLOv7 算法相比其他算法具有高精度和快速检测的优势。

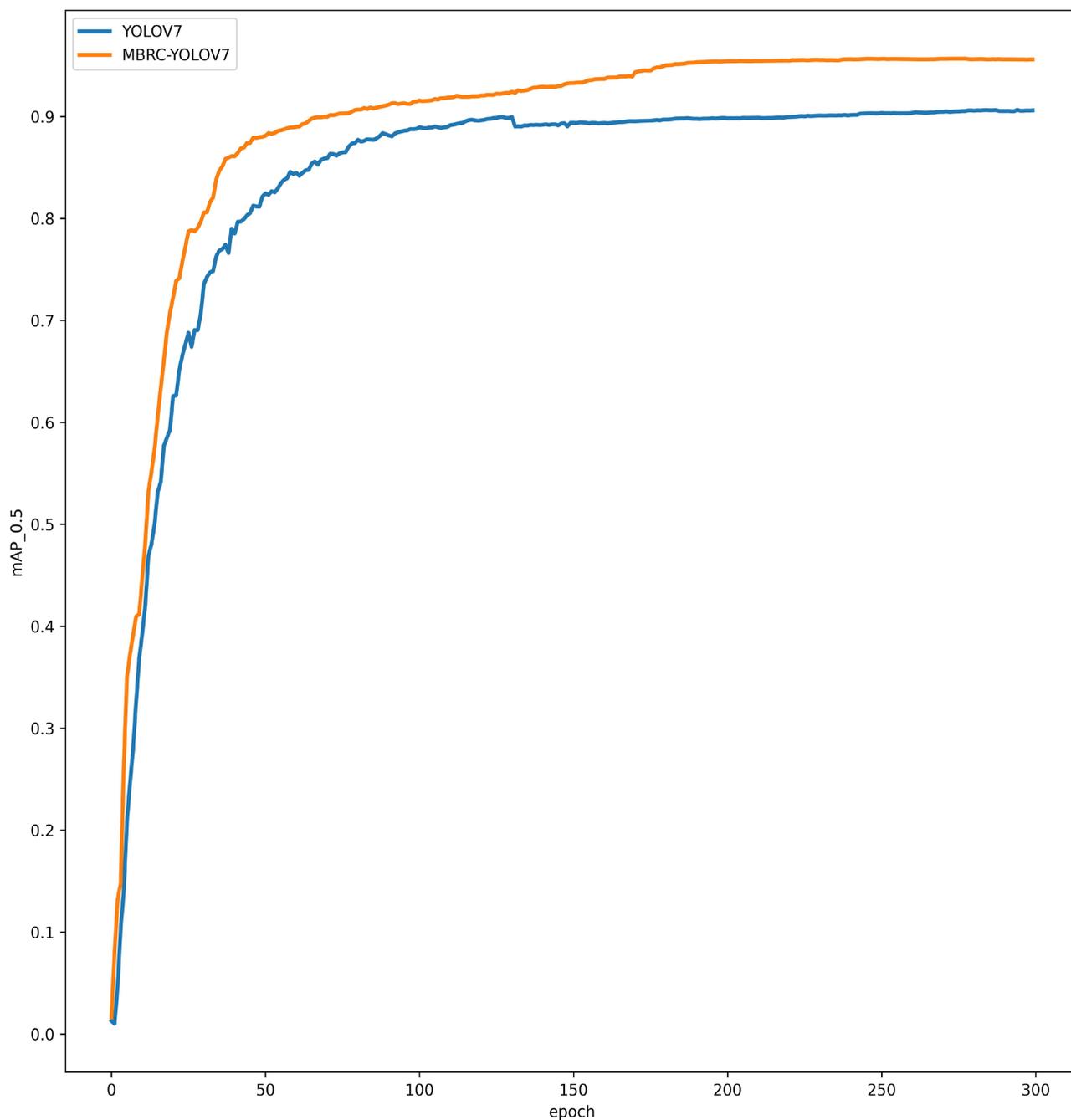
Table 2. Comparison of detection accuracy of different models on SIXray_OOD dataset

表 2. 不同模型在 SIXray_OOD 数据集的检测准确率对比

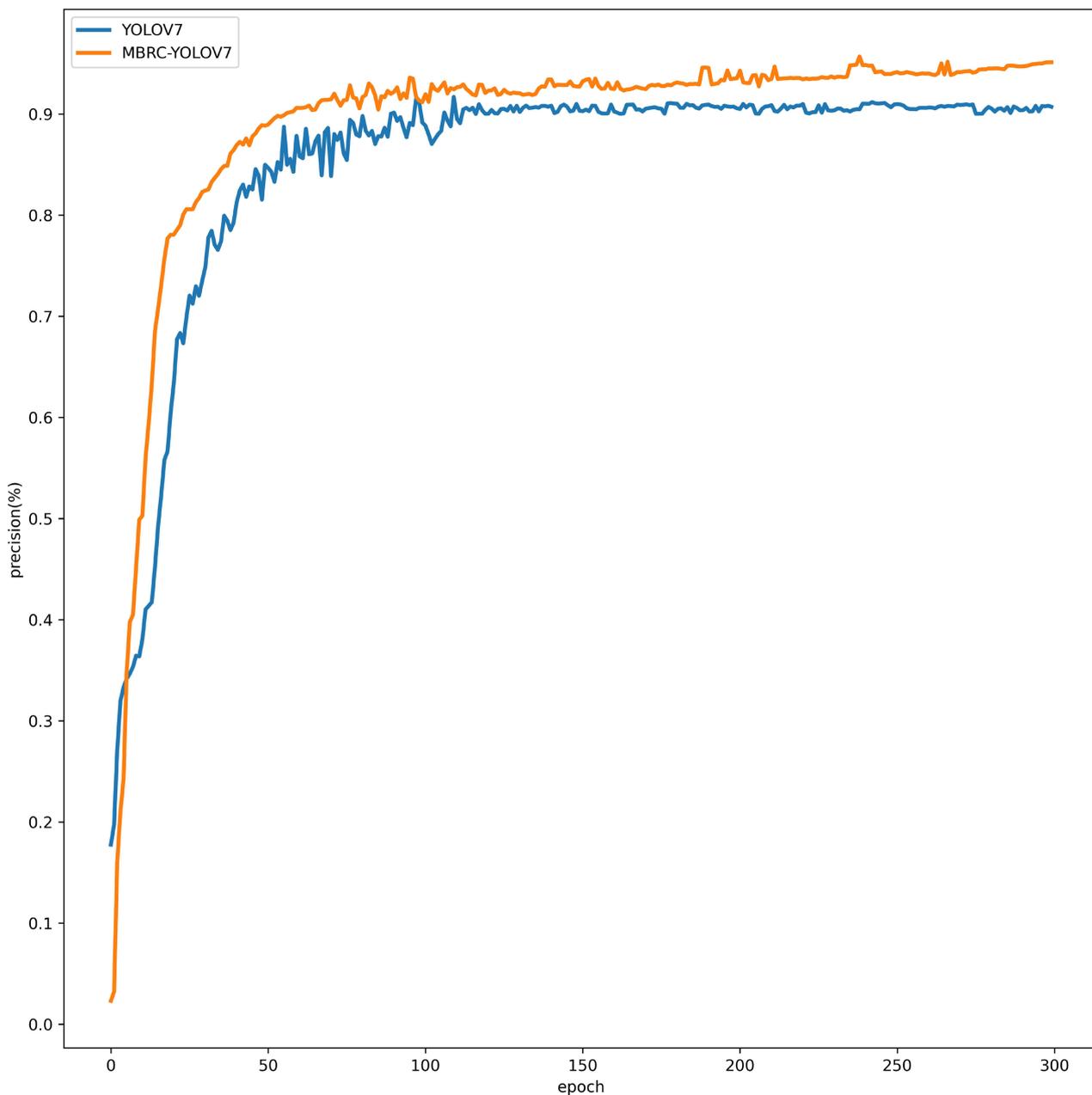
Number	AP@0.5(%)					mAP@0.5	FPS (ms)
	MBCConv	BRA	CARAFE	RFE	Mish		
Faster R-CNN	78	64	57	70	71	75.6	18
SSD	79	62	60	74	70	77.5	44
Mask R-CNN	80	68	78	75	69	81.3	52

续表

YOLOv5s	85	83	80	82	84	87.4	100
DETR	92	82	78	81	86	87.9	65
YOLOv7	94	83	84	88	87	90.3	86
GF-YOLOv7	95	86	86	89	89	91.3	90
MBRC-YOLOv7	98	93	90	94	93	95.2	92



(a) mAP@0.5



(b) precision

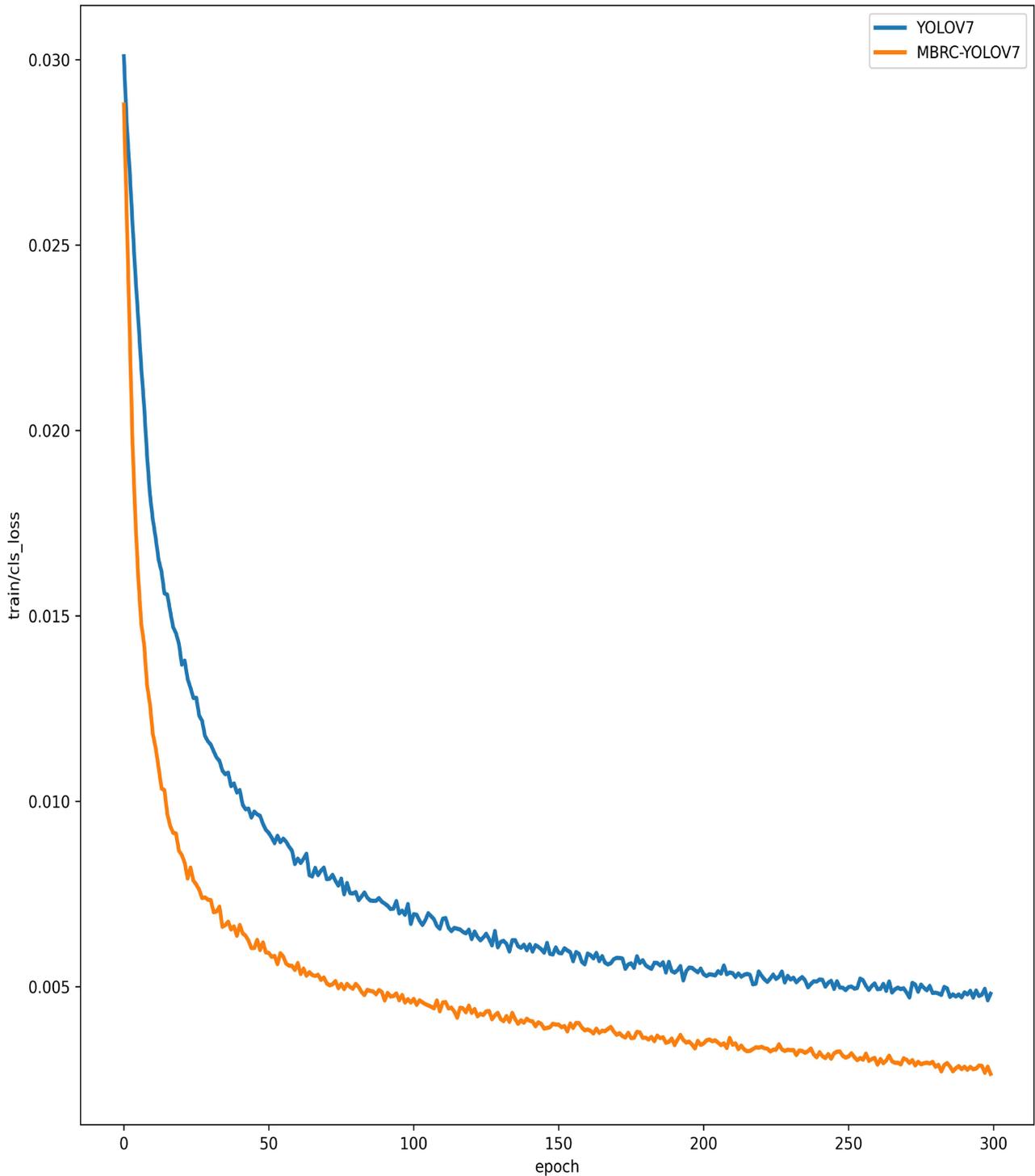
Figure 11. Comparison of MBRC-YOLOv7 training process

图 11. MBRC-YOLOv7 训练过程对比

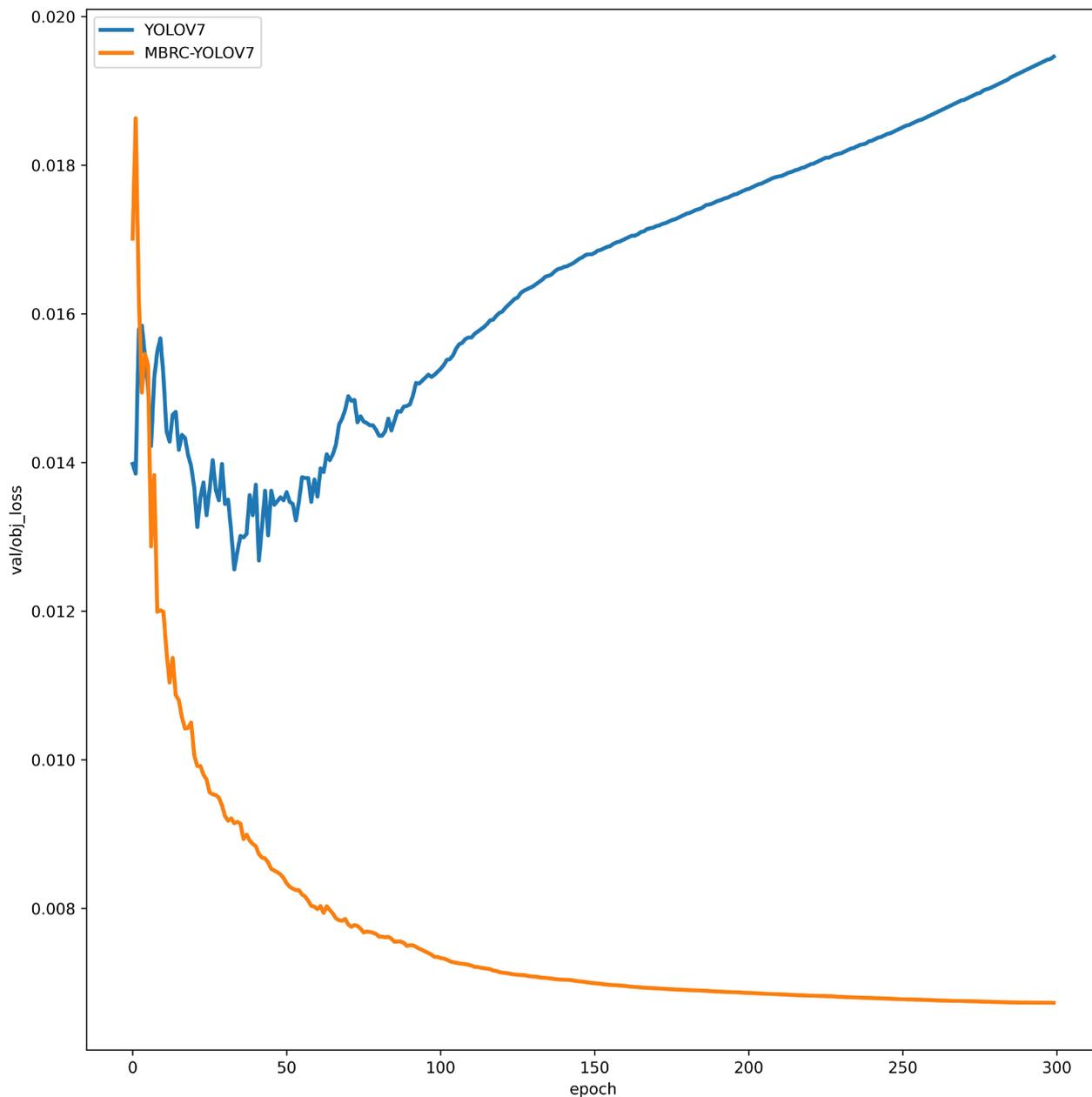
图 11 显示了本研究中提出的网络模型与原模型对数据集 SIXray_OOD 的训练过程对比。图 13(a)显示,在完成 300 轮的训练后,所提出模型平均识别准确率 mAP_0.5 达到了 95.2%,相比原模型提高了 4.9%;图 13(b)显示,经过 300 轮的训练后,所提出模型识别精度达到了 95.5%。

图 12 显示了本研究中提出的网络模型与原模型对数据集 SIXray_OOD 的训练过程损失函数对比。图 12(a)显示,经过 300 轮的训练后,所提模型训练集分类损失函数均值稳定在 0.0026,这表明模型在数据集 SIXray_OOD 中已经达到了稳定的最佳训练状态图;图 12(b)显示,经过 300 轮的训练后,所提

模型验证集损失函数逐渐减小，均值稳定在 0.006，而原模型验证集损失函数由于正负样本分布不均衡随着训练次数的递增而增大。与原算法相比，本研究提出的模型具有明显的优势，可以更准确地识别检测违禁品。



(a) 训练集损失函数



(b) 验证集损失函数

Figure 12. MBRC-YOLOv7 training process loss function comparison

图 12. MBRC-YOLOv7 训练过程损失函数对比

3.4.3. 可视化分析

为了验证改进模型对违禁品的检测性能，选取在重叠遮挡、背景杂乱、多尺度和普通样本四种不同的场景测试了改进后的模型。图 13 为重叠遮挡样本检测结果，原网络中只检测出一把小刀，而改进后网络 3 把重叠的刀子均被检测出来，精度分别为 75% 和 87%，说明优化后的模型可以有效减少重叠遮挡场景中漏检的情况。图 14 为背景杂乱场景下的识别结果，改进后网络检测到原网络中漏检的刀子，精度为 96%，说明优化后的模型可以减少在背景杂乱场景下的漏检情况。如图 15 所示，对于多尺度样本，改进

后的模型对不同方向放置的刀子检测精度提高了 14%，说明改进后网络可以减少违禁品多尺度变化引起的漏检情况。如图 16 所示，对于普通样本，网络改进前后都检测出所有的违禁品，但改进后网络检测精度更高。综上所述，改进后的模型可以有效减少由重叠遮挡、背景杂乱、物体多尺度变换引起的违禁品漏检情况，同时也提升了类别检测精度。

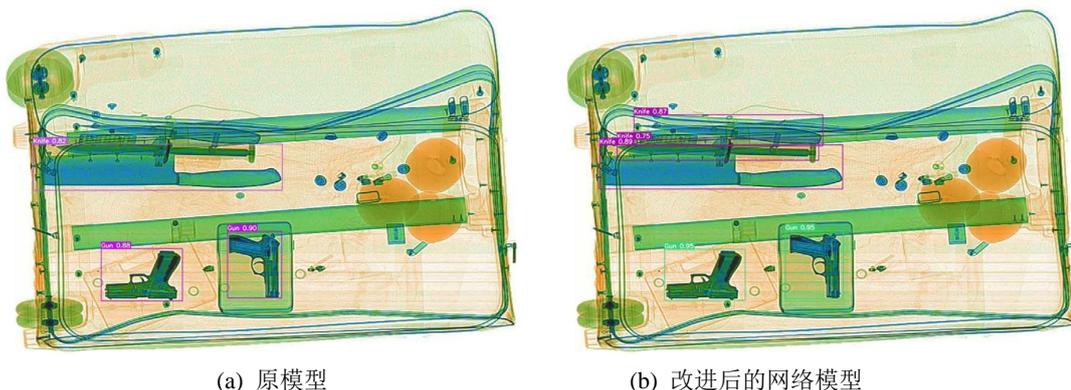


Figure 13. Overlapping Masked Sample Detection Visualization Results

图 13. 重叠遮挡样本检测可视化结果

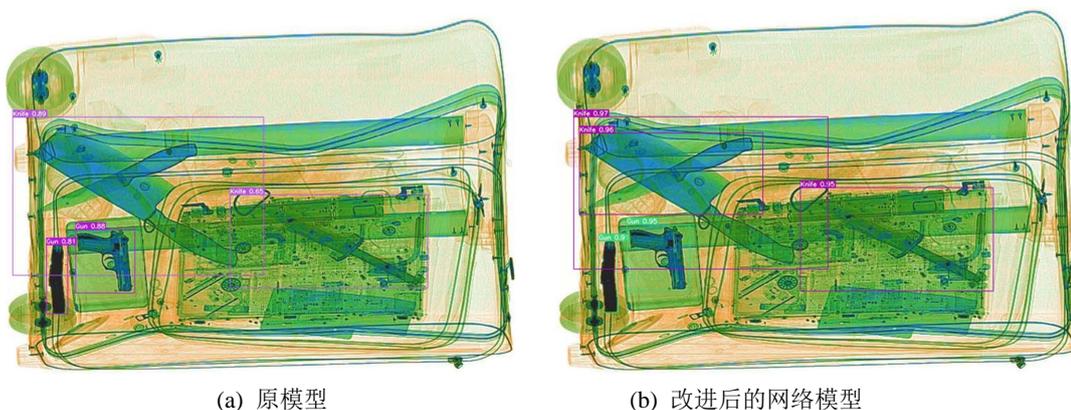


Figure 14. Visualization results of background cluttered sample detection

图 14. 背景杂乱样本检测可视化结果

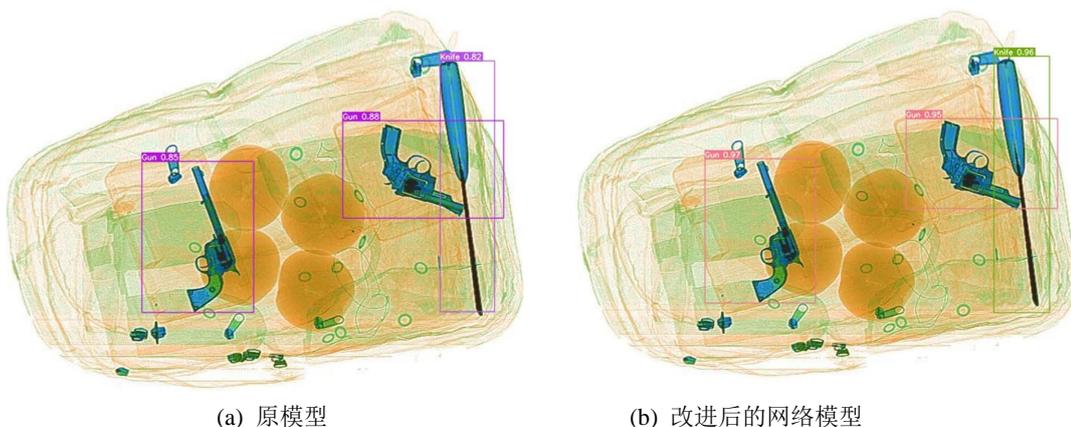


Figure 15. Visualization results of multiscale change sample detection

图 15. 多尺度变化样本检测可视化结果



Figure 16. Visualization of the results of testing common samples
图 16. 普通样本检测可视化结果

4. 总结

针对 X 射线图像违禁品检测中存在的难点, 本文基于 YOLOv7 提出 MBRC-YOLOv7 多尺度小目标检测算法。将主干网络中 ELAN 结构替换为 MBCConv 来加强违禁品关键特征提取能力, 在 Neck 中融合 BRA 注意力机制提取不同尺度的违禁品特征, 减少复杂背景干扰; 添加小目标检测层提高小目标检出率和准确率; 引入了 RFE 模块, 增加特征图的感受野, 提高了违禁品多尺度检测和识别的准确性, 最后采用了 CARAFE 上采样和 Mish 激活函数来提高网络对重叠和遮挡对象的识别能力, 进一步提升算法在正负样本与难易样本失衡情况下的检测性能。实验结果表明, 本文算法在 SIXray_OOD 数据集中对 X 射线违禁品的检测效果优于其他算法, 有效提升了违禁品检测精度与速度。为实际应用提供了一定的参考价值, 但在轻量化方面仍有进步空间, 需继续研究探索。

基金项目

国家自然科学基金(U2141239)。

参考文献

- [1] 张珂, 张良. 复杂背景下多尺度 X 光违禁品检测[J]. 激光与光电子学进展, 2021(22): 58.
- [2] Michel, S., Koller, S.M., de Ruiter, J.C., Moerland, R., Hogervorst, M. and Schwaninger, A. (2007) Computer-Based Training Increases Efficiency in X-Ray Image Interpretation by Aviation Security Screeners. 2007 41st Annual IEEE International Carnahan Conference on Security Technology, Ottawa, 8-11 October 2007, 201-206. <https://doi.org/10.1109/CCST.2007.4373490>
- [3] Zhang, K. and Zhang, L. (2021) Multi-Scale Detection for X-Ray Prohibited Items in Complex Background. *Laser & Optoelectronics Progress*, **58**, Article ID: 2210002. <https://doi.org/10.3788/LOP202158.2210002>
- [4] Akcay, S., Kundegorski, M.E., Willcocks, C.G. and Breckon, T.P. (2018) Using Deep Convolutional Neural Network Architectures for Object Classification and Detection within X-Ray Baggage Security Imagery. *IEEE Transactions on Information Forensics and Security*, **13**, 2203-2215. <https://doi.org/10.1109/TIFS.2018.2812196>
- [5] Gaus, Y.F.A., Bhowmik, N., Akcay, S. and Breckon, T. (2019) Evaluating the Transferability and Adversarial Discrimination of Convolutional Neural Networks for Threat Object Detection and Classification within X-Ray Security Imagery. 2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA), Boca Raton, 16-19 December 2019, 420-425. <https://doi.org/10.1109/ICMLA.2019.00079>
- [6] Ge, Z., Liu, S., Wang, F., Li, Z. and Sun, J. (2021) Yolox: Exceeding Yolo Series in 2021.
- [7] Zhang, Y.T., Zhang, H.G., Zhao, T.F. and Yang, J.F. (2020) Automatic Detection of Prohibited Items with Small Size

- in X-Ray Images. *Optoelectronics Letters*, **16**, 313-317. <https://doi.org/10.1007/s11801-020-9118-x>
- [8] Zhu, X., Zhang, J., Chen, X., Li, D., Wang, Y. and Zheng, M. (2021) AMOD-Net: Attention-Based Multi-Scale Object Detection Network for X-Ray Baggage Security Inspection. *Proceedings of the 2021 5th International Conference on Computer Science and Artificial Intelligence*, Beijing, 4-6 December 2021, 27-32. <https://doi.org/10.1145/3507548.3507552>
- [9] Wang, C.Y., Bochkovskiy, A. and Liao, H.Y.M. (2022) YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, 17-24 June 2023, 7464-7475. <https://doi.org/10.1109/CVPR52729.2023.00721>
- [10] Song, L.I. and Musa, Y. (2023) Improved YOLOv7 X-Ray Image Real-Time Detection of Prohibited Items. *Journal of Computer Engineering & Applications*, **59**, 193-200.
- [11] Tan, M. and Le, Q. (2019) Efficientnet: Rethinking Model Scaling for Convolutional Neural Networks. *International Conference on Machine Learning*, Long Beach, 9-15 June 2019, 6105-6114.
- [12] Zhu, L., Wang, X., Ke, Z., Zhang, W. and Lau, R.W. (2023) BiFormer: Vision Transformer with Bi-Level Routing Attention. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Vancouver, 17-24 June 2023, 10323-10333. <https://doi.org/10.1109/CVPR52729.2023.00995>
- [13] Wang, J., Chen, K., Xu, R., Liu, Z., Loy, C.C. and Lin, D. (2019) Carafe: Content-Aware Reassembly of Features. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Seoul, 27 October-2 November 2019, 3007-3016. <https://doi.org/10.1109/ICCV.2019.00310>
- [14] Yu, Z., Huang, H., Chen, W., Su, Y., Liu, Y. and Wang, X. (2022) Yolo-facev2: A Scale and Occlusion Aware Face Detector.
- [15] Mishra, D. (2020) Mish: A Self Regularized Non-Monotonic Activation Function. arXiv:1908.08681.
- [16] Miao, C., Su, C., Wan, F., Liu, H., Jiao, J., Xie, L., et al. (2019) SIXray: A Large-Scale Security Inspection X-Ray Benchmark for Prohibited Item Discovery in Overlapping Images. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 15-20 June 2019, 2114-2123. <https://doi.org/10.1109/CVPR.2019.00222>
- [17] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y. and Berg, A.C. (2016) Ssd: Single Shot Multibox Detector. *Computer Vision-ECCV 2016: 14th European Conference*, Amsterdam, 11-14 October 2016, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2
- [18] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A. and Zagoruyko, S. (2020) End-to-End Object Detection with Transformers. In: *European Conference on Computer Vision*, Springer International Publishing, Cham, 213-229. https://doi.org/10.1007/978-3-030-58452-8_13
- [19] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A. and Zagoruyko, S. (2020) End-to-End Object Detection with Transformers. In: Vedaldi, A., Bischof, H., Brox, T. and Frahm, J.M., Eds., *Computer Vision—ECCV 2020, Lecture Notes in Computer Science*, Vol. 12346, Springer, Cham, 213-229. https://doi.org/10.1007/978-3-030-58452-8_13
- [20] Wang, Y., Zhang, X., Li, L., Wang, L., Zhou, Z. and Zhang, P. (2023) An Improved YOLOv7 Model Based on Visual Attention Fusion: Application to the Recognition of Bouncing Locks in Substation Power Cabinets. *Applied Sciences*, **13**, Article No. 6817. <https://doi.org/10.3390/app13116817>
- [21] Ren, S., He, K., Girshick, R. and Sun, J. (2015) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 1137-1149.
- [22] He, K., Gkioxari, G., Dollár, P. and Girshick, R. (2017) Mask R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*, Venice, 22-29 October 2017, 2961-2969. <https://doi.org/10.1109/ICCV.2017.322>