基于深度学习的X射线违禁品检测系统的设计与实现

摘要：随着我国经济的高速发展，交通量急剧增加，公共安全问题日益凸显，安检需求愈发迫切。目前的安检主要依靠安检人员用肉眼观察X光安检机扫描得到图像，判断行李中是否夹带违禁品，这种检查方式存在漏检、错检等隐患。因此本文提出了一种基于深度学习的违禁品检测方案。

首先，下载公开的违禁品数据集SIXray作为基准数据集，对其进行预处理和数据增强后随机划分为训练集、验证集以及测试集，形成了共计5000张违禁品图像的数据集SIXray-s。其次，以RT-DETR模型作为基准，使用FasterNet-T0作为其主干特征提取网络；采用BiFPN4结构改进CCFF的特征融合结构；将Conv3XCNCSPELAN4添加到BIFPN4结构中作为融合单元；使用Slide Varifocal Loss分类损失函数优化样本权重分配。此外，设计了模块对比实验、模型消融实验，验证了改进模块的优越性和有效性。同时横向对比了其他常见的目标检测模型，证实了本文模型在精度和效率上的卓越性能。接着，通过混淆矩阵、损失值等指标佐证模型的出色性能。最后，本文使用PyQt5框架以改进后的RT-DETR模型为基础，搭建了一个本地的违禁品检测程序，用户可以在本地客户端运行该程序，便捷高效的进行违禁品检测。

实验结果显示，改进模型与原始模型相比参数量下降58.3%，计算量下降65%，FPS提高32.8%，mAP也有所提升。同时优于其它传统的目标检测模型。本研究提供了一个便捷高效的违禁品检测方案，为公共安全做出贡献。

关键词：深度学习；RT-DETR；目标检测；轻量化；违禁品

目 录

[1 绪论 1](#_Toc14245)

[1.1 研究背景和意义 1](#_Toc17548)

[1.2 国内外研究现状 1](#_Toc4935)

[1.3 论文的研究内容 2](#_Toc24496)

[1.4 论文的结构安排 2](#_Toc17874)

[2 理论基础 4](#_Toc4548)

[2.1 深度学习技术 4](#_Toc20887)

[2.2 目标检测算法 4](#_Toc2839)

[2.3 Transformer架构 4](#_Toc2016)

[3 基于深度学习的违禁品检测模型 6](#_Toc6422)

[3.1 RT-DETR模型 6](#_Toc1137)

[3.2 模型改进方案 7](#_Toc1118)

[3.2.1 FasterNet主干特征提取网络 7](#_Toc8651)

[3.2.2 Bifpn4特征融合模块 9](#_Toc2139)

[3.2.3 Conv3XCNCSPELAN4重参数化模块 10](#_Toc5332)

[3.2.4 Slide Varifocal Loss分类损失函数 11](#_Toc30939)

[3.3 构建完整的改进模型 12](#_Toc4766)

[4 实验测试与分析 13](#_Toc3391)

[4.1 环境配置 13](#_Toc14672)

[4.2 数据集 13](#_Toc17967)

[4.2.1 常见的违禁品数据集 13](#_Toc32138)

[4.2.2 数据增强 14](#_Toc18944)

[4.2.3 SIXray-s数据集 16](#_Toc10171)

[4.3 评价指标 18](#_Toc13607)

[4.4 模块对比实验 19](#_Toc2361)

[4.4.1 数据增强前后对比实验 19](#_Toc1906)

[4.4.2 主干特征提取网络对比实验 19](#_Toc3878)

[4.4.3 特征融合单元对比实验 21](#_Toc23520)

[4.5 改进消融实验（纵向对比） 21](#_Toc18784)

[4.6 模型对比实验（横向对比） 23](#_Toc18503)

[4.7 最终模型评估图 24](#_Toc28175)

[4.7.1 混淆矩阵 24](#_Toc27439)

[4.7.2 损失值变化 25](#_Toc26260)

[4.7.3 mAP变化 26](#_Toc8589)

[4.7.4 评价指标曲线图 26](#_Toc6361)

[4.7.5 热力图 28](#_Toc27455)

[5 违禁品检测系统 29](#_Toc30814)

[5.1 Ultralytics 29](#_Toc31864)

[5.2 PyQt5 29](#_Toc1145)

[5.3 功能与界面 29](#_Toc11132)

[6 总结与展望 32](#_Toc17809)

[6.1 总结 32](#_Toc9570)

[6.2 展望 32](#_Toc10609)

[参考文献 33](#_Toc27056)

# 

# 1 绪论

## 1.1 研究背景和意义

近年来，随着我国经济的高速发展和交通基础设施的完善，高铁、机场、地铁等公共场所的客流量呈现爆发式增长。据交通运输部统计，2024年我国全年客运量将达到645亿人次，日均1.8亿人次。在这样的背景下，公共安全面临严峻挑战，尤其是携带刀具、易燃易爆物品、枪支等违禁品进入公共场所的潜在风险持续增加。传统的安检流程依赖人工肉眼观察X光安检机扫描得到的图像[1]，存在效率低、漏检率与错检率高、人力成本高昂等问题。

为了解决上述问题，本文提出一种基于深度学习的X射线违禁品检测方案。本方案使用轻量化改进的RT-DETR[2]模型对违禁品进行检测，以目标检测算法为核心，搭建了便捷高效的X射线违禁品检测系统，实现了违禁品的实时端到端检测，有效提升了安检效率和准确率，降低了人力成本，为公共安全提供了有力保障。

## 1.2 国内外研究现状

近年来，随着公共安全需求的提升，违禁品检测吸引了广泛的研究。国内外学者们纷纷投入算法的创新研发，旨在提高违禁品检测的准确性和效率。

向丹[3]（2020）在YOLOv3的基础上进一步提出YOLOv3+。网络结构方面，对YOLOv3原有的Darknet-53网络进行改进，得到五组不同大小的特征图并采用多尺度融合的方式构成特征金字塔进行X光违禁品检测；损失函数方面，采用TSE损失函数，它可随着误差的变化而调整权重，避免出现梯度消失的问题。该模型在提高精度的同时，可实现实时检测。

李舒婷[4]（2021）针对X光图像重叠透明、杂乱无章的的特点，提出基于全卷积网络的X光图像违禁物品检测算法。使用全卷积网络作为基础网络，以逐像素预测的方式进行目标检测，可以避免以往基于锚框的传统算法在训练过程中复杂的IoU计算和匹配问题，提高了模型计算效率。

Kumar[5]等人（2022）改进了Faster-RCNN[6]网络，并提出了一种多通道区域建议网络（MCRPN）。利用视觉语义中不同层次的卷积特征互补实现多层特征提取，融合VGG16高级层的语义信息和低级层的边缘特征。为了构建多尺度违禁品检测网络，将多尺度候选目标区域映射到相应的特征图；引入扩张卷积，并设计了多分支扩张卷积模块（DCM）以增加感受野，从而增强不同尺度的特征。在自建数据集SIXray OD上，改进算法的平均检测准确率为84.69%，较原网络提高了6.28%。

Yu[7]等人（2023）使用可变形卷积重新设计YOLOv4的路径聚合网络，可变形卷积可以根据检测到目标的形状灵活地改变其感受野。当PANet[8]模块中高层次信息和低层次信息融合时，采用可变形卷积对特征进行对齐，可以有效提高检测精度。然后，引入Focal-EIOU损失函数，可以解决CIOU损失函数在处理低质量样品时容易引起严重损失值振荡的问题。最后，利用Soft-NMS改进YOLOv4的非极大值抑制，在SIXray[9]数据集上mAP达到85.51%。

Chen[10]等人（2024）提出了一种改进的YOLOv8网络模型。在输入阶段，采用Mixup数据增强方法来增强模型学习检测被遮挡和隐藏的违禁物品的能力。将Biformer双层路由注意力模块引入骨干网络，以增强模型对前景信息的关注，提高其识别能力。将C2f\_f模块整合到主干网络和Neck部分，以提高模型的特征表示能力并增强整体网络性能。在PIDray的Easy、Hard和Hidden数据集上mAP@达到85.2%、86.2%和68.3%，分别提高了2.3%、1.4%和2.2%。

综上所述，国内外关于使用深度学习进行违禁品检测的研究已取得一定的进展，但在精度和检测效率上还有提升空间。

## 1.3 论文的研究内容

本文提出一种基于深度学习的X射线违禁品识别方法，旨在实现对违禁品的快速准确识别。

1. 首先，下载公开的违禁品数据集SIXray作为基准数据集并对其进行预处理得到原始数据。接着使用Python脚本对图像进行了6种数据增强操作，扩充图像样本。最后随机划分为训练集、验证集以及测试集，形成了共计5000张违禁品图像的数据集SIXray-s。
2. 本文以RT-DETR模型作为基准，对其Backbone、CCFF等部分进行进一步改进，使其更加轻量化。设计了模块对比实验和模型消融实验，验证了改进模块的优越性和有效性。横向对比了其他常见的目标检测模型，证实了本文模型在精度和效率上的竞争力。通过混淆矩阵、损失值、mAP值、热力图等指标佐证了改进模型的卓越性能。
3. 本文使用PyQt5框架以改进后的RT-DETR模型为基础，搭建了一个本地的违禁品检测程序，用户可以在本地客户端运行该程序，从而便捷高效的完成违禁品检测。

## 1.4 论文的结构安排

本文的其余章节结构如下。

第二章简要概述了深度学习技术、目标检测算法以及Transformer[11]架构。

第三章首先重点介绍了本文的基准模型RT-DETR，并探讨了使用FasterNet-T0[12]作为主干特征提取网络，使用BIFPN4结构改进CCFF模块，引入Conv3XCNCSPELAN4模块以及引入Slide Varifocal Loss分类损失函数的改进方案。最后介绍了完整的改进模型。

第四章首先介绍了实验的超参数设置以及环境配置，同时说明了违禁品数据集的选择、数据增强方法与划分比例。然后，设计了模块对比实验，证明改进模块与其他同类型模块相比较的优越性；消融实验纵向对比模型改进前后的性能，验证改进模块的有效性。接着，通过横向对比实验，评估了改进模型与其他常见目标检测模型的性能差异，进一步验证了改进模型在精度和效率上的优势。通过分析混淆矩阵、损失值等指标佐证了改进模型的卓越性能。

第五章介绍了搭建系统使用的Python库，并说明了系统的功能。

第六章对本文进行总结和展望。

# 2 理论基础

## 2.1 深度学习技术

深度学习是一种基于深层神经网络的机器学习方法，其核心理念是通过多层非线性变换，从原始数据中自动学习并提取高级特征。与传统的手工设计特征不同，深度学习模型通过训练数据自动发现数据中的内在规律和层次化结构，从而完成复杂的预测和分类任务。深度学习强调端到端的学习过程，即从原始输入到预测输出的映射由模型自动完成，无需人工干预。这种自动化特征学习的能力使得深度学习模型在处理大规模、高维数据时具有显著优势。

## 2.2 目标检测算法

目标检测是计算机视觉领域的基础任务之一，其目的是对目标物体进行识别和定位。传统的目标检测方法依赖手工设计特征与滑动窗口策略，典型代表包括：HOG（方向梯度直方图）[13]+SVM[14]以及DPM（可变形部件模型）[15]。然而，传统方法受限于特征表达能力与计算效率，难以应对复杂场景下的多尺度、遮挡目标检测需求。深度学习能够将图像中的像素数据逐步转化为更高层次、更抽象的特征。因此，深度学习所提取的特征相较于传统方法，在表达能力和鲁棒性方面更为出色。这一特点使得深度学习在计算机视觉领域展现出不可替代的优越性。随着深度学习技术的发展，基于卷积神经网络和Transformer的目标检测算法开始占据主导地位。目标检测算法分类如图2.1所示。

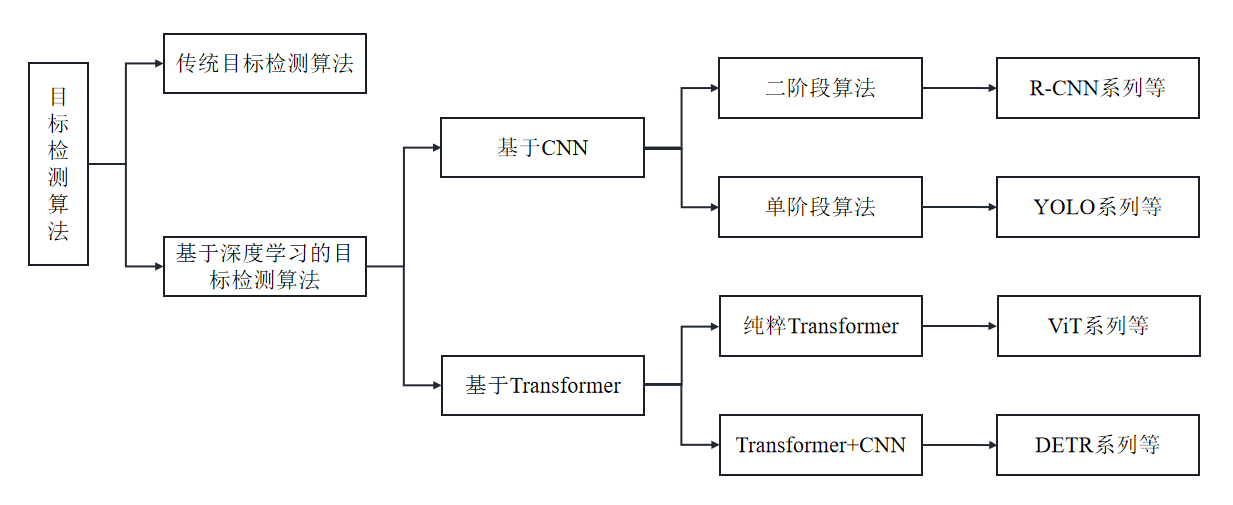


图2.1 目标检测算法分类

## 2.3 Transformer架构

Transformer是一类基于自注意力机制的深度神经网络，其由编码器和解码器两部分组成。在编码器的各层级中，包括多头自注意力模块和前馈神经网络，用于捕捉输入序列中的依赖关系。解码器每一层还包括编码—解码自注意力模块，用于生成输出序列。自注意力机制使Transformer能够捕捉长距离依赖关系。Transformer最初用于自然语言处理（NLP）任务，如机器翻译和语言建模。在计算机视觉任务中，Transformer相较于卷积神经网络具有可以理解图像整体语义擅长捕捉复杂图像特征以及无需NMS等后处理步骤的优势。近年来，随着DETR[16]系列与ViT[17]系列的诞生，Transformer开始在计算机视觉领域崭露头角。Transformer的模型架构如图2.2所示。

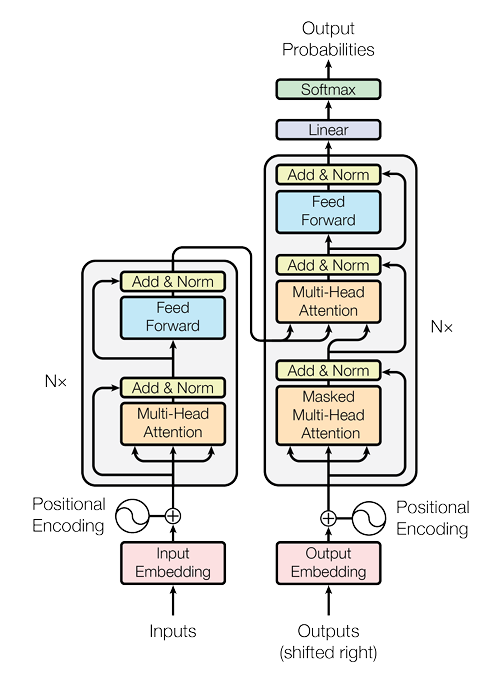


图2.2 Transformer的模型架构

# 3 基于深度学习的违禁品检测模型

## 3.1 RT-DETR模型

实时目标检测（Real-Time Object Detection）过去一直由YOLO[18]系列模型主导。YOLO检测器需要NMS后处理，其通常难以优化且不够鲁棒，因此检测器的速度存在延迟，且无法实现端到端检测。2020年DETR算法诞生，DETR是第一个基于Transformer的端到端目标检测算法，但是DETR的计算成本较高，尽管后续的优化算法不断加快收敛速度，提升推理速度，但仍然无法满足实时检测要求。

RT-DETR是百度于2023年提出的第一个实时端到端目标检测器。RT-DETR是DETR系列的延续，其不仅在精度和速度方面优于当前的实时检测器，而且不需要后处理过程，标志着目标检测算法的新高度。RT-DETR适合应用于需要实时处理的场景中，如智能安检、自动驾驶等。

现有的多尺度Transformer编码器普遍会同时进行尺度内和尺度间特征交互，导致模型产生较大的计算损耗。为了减少计算消耗，RT-DETR作者提出了高效混合编码器，通过独立进行尺度内和尺度间特征交互，来缩短序列长度，可以同时在速度和精度上带来正向收益。

查询选择是从编码器输出的特征序列中选择特征作为目标查询，其经过解码器后由预测头映射为置信度和边界框。目前的DETR变体都是直接根据特征的分类分数进行选择。然而，分类得分高的预测框并不一定是和GT最接近的框，这种选择机制会导致高分类分数低IoU的框被选中，而低分类分数高IoU的框被丢弃，对检测器的性能产生不利影响。为解决这一问题，RT-DETR作者提出了Uncertainty-Minimal Query Selection，通过在训练期间约束检测器对高IoU的特征产生高分类分数，对低IoU的特征产生低分类分数。从而使得模型根据分类分数选择的top-K特征对应的预测框也能具有高IoU分数。

此外RT-DETR可以通过调整模块的数量以及深度和宽度实现缩放，同时通过在推理时使用不同层数的解码器，可以实现无需执行额外的模型再训练流程便可直接调整推理速度。这种双重可调节特性使RT-DETR可以适用于不同场景下的实时检测需求。

RT-DETR由主干特征提取网络、高效混合编码器（Efficient Hybrid Encoder）和解码器组成。

RT-DETR采用经典的ResNet[19]作为主干特征提取网络，负责提取图像的多层次特征。编码器以主干特征提取网络输出的S3、S4、S5特征作为输入，通过基于注意力的尺度内特征交互（AIFI）和基于神经网络的跨尺度特征融合（CCFF），将多尺度特征转化为特征序列。然后，Uncertainty-Minimal Query Selection选择一定的编码器特征作为解码器的初始对象查询。最后，具有辅助预测头的解码器迭代优化对象查询以生成类别和框。

RT-DETR模型的结构如图3.1所示。

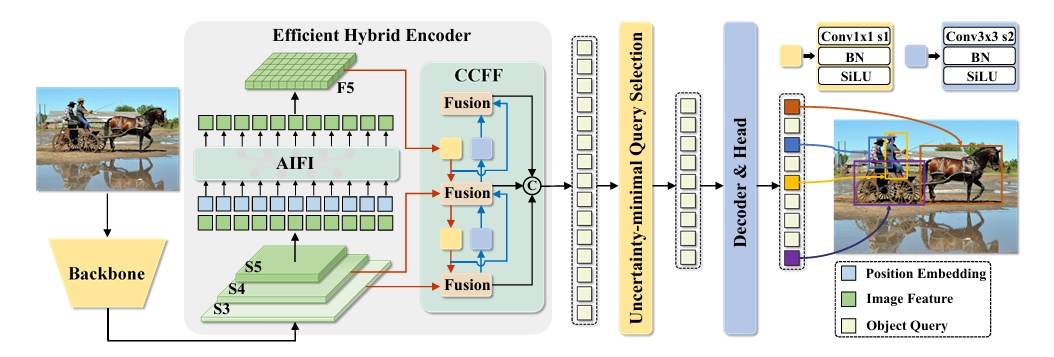


图3.1 RT-DETR模型结构图

## 3.2 模型改进方案

### 3.2.1 FasterNet主干特征提取网络

随着目标检测算法的广泛应用，人们对算法的实时性和准确性要求越来越高。如何在不损失精度的前提下降低算法的延迟成为关键的挑战。算法的延迟与算法的FLOPs以及内存带宽有关。MobileNet[20]、ShuffleNet[21]和GhostNet[22]等利用DWConv（深度卷积）[23]和GConv（组卷积）[24]来提取空间特征，大大降低了算法的计算量，但是算法对内存的频繁访问导致延迟仍然较高。而且这些网络中通常伴随着额外的数据操作，这些操作的运行时间同样不容忽视。

在FasterNet中作者提出了PConv，其通过仅对输入特征图的部分通道进行空间特征提取，其余通道保持不变，可以减少冗余计算和内存访问，提升模型效率，同时保持了特征提取能力。

DWConv是Conv的一种流行变体，已被广泛采用为许多神经网络的关键模块。DWConv可以降低FLOPs，但直接替换常规卷积会导致精度下降。对于输入，DWConv使用的卷积核，输出，在操作时DWConv通常需要增加通道数c到（）来补偿精度的下降。但额外的通道会导致内存访问增加：

 （3.1）

普通卷积的内存访问为：

 （3.2）

PConv的通道数为（），内存访问仅为：

 （3.3）

比DWConv和普通卷积的内存访问都要低。同时普通卷积的FLOPs为：

 （3.4）

PConv的FLOPs为：

 （3.5）

同样要比普通卷积的FLOPs小，证明了PConv的优越性。普通卷积、DWConv与PConv的示意图如图3.2所示。

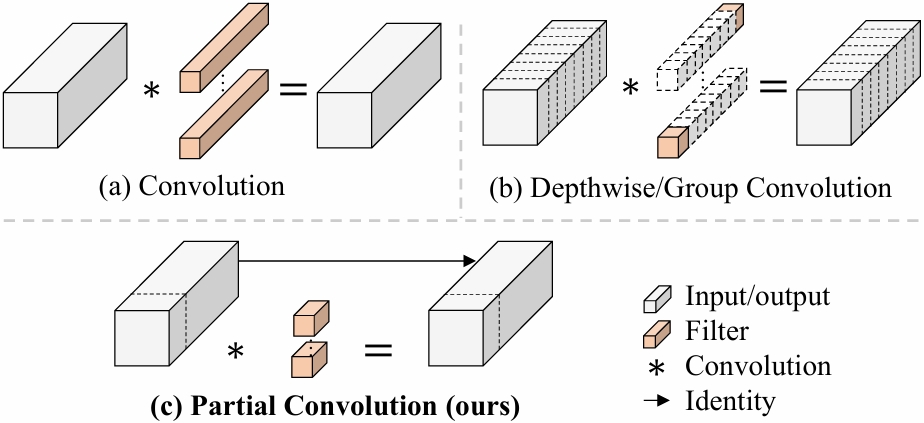


图3.2 普通卷积、DWConv与PConv的示意图

为了充分有效地利用所有通道的信息，PConv后通常连接逐点卷积。FasterNet作者基于PConv与逐点卷积提出了FasterNet，其对计算机视觉任务非常有效，可以显著提升模型的推理速度，同时保持较高的准确率。FasterNet结构图如图3.3所示。

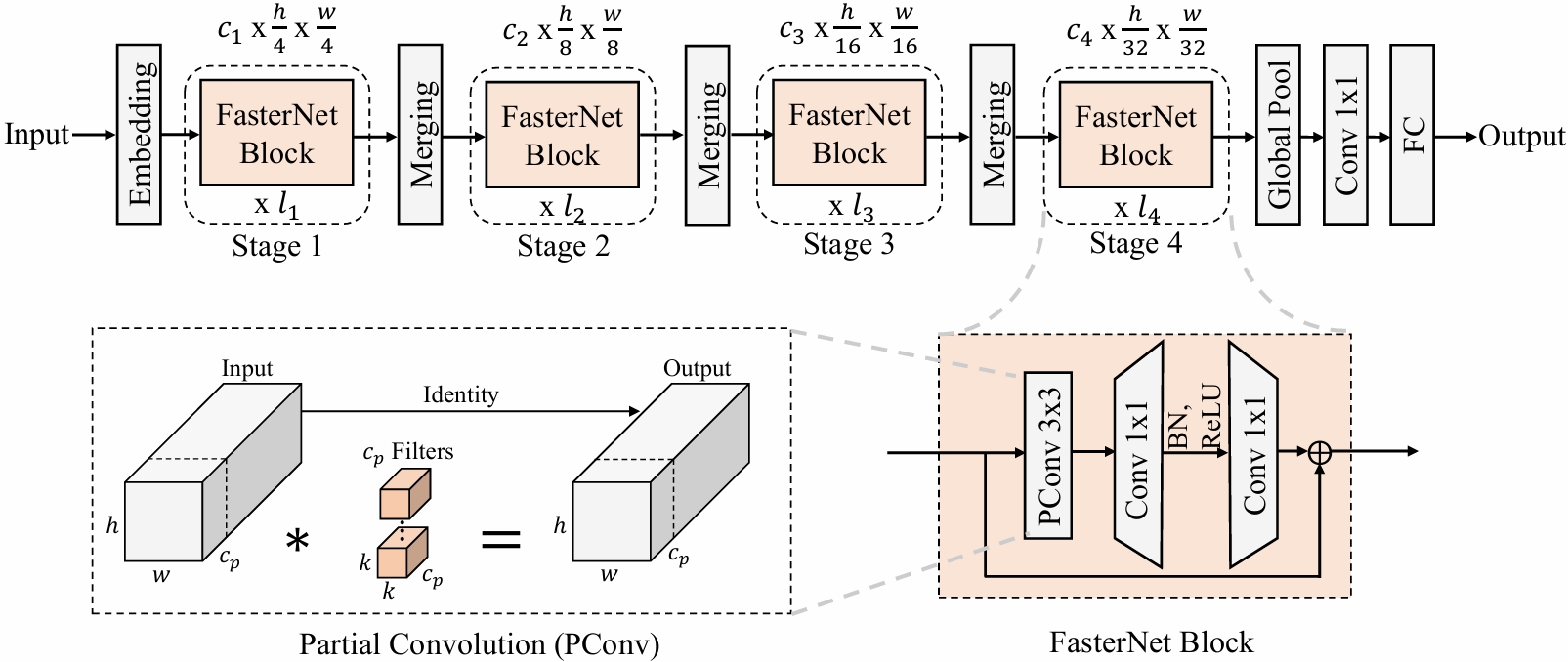


图3.3 FasterNet结构图

### 3.2.2 Bifpn4特征融合模块

RT-DETR的跨尺度特征融合模块（CCFF）由自顶向下的路径、自底向上的路径和横向连接三部分组成，是经典的PAFPN（Path Aggregation Feature Pyramid Network）[8]结构。其存在特征层级有限：仅融合P3-P5三个层级，低层特征（如P2）未被利用，限制了小目标的细节捕捉能力；固定权重融合：不同层级的特征通过简单拼接或相加融合，未引入自适应权重机制，导致噪声特征干扰（如堆叠遮挡场景）的不足。

BiFPN（Bidirectional Feature Pyramid Network）[25]是一种高效的特征金字塔网络结构，专为优化多尺度目标检测任务而设计。它通过双向跨尺度连接和加权特征融合机制，显著提升了模块的特征融合能力。原始BiFPN结构图如图3.4所示。

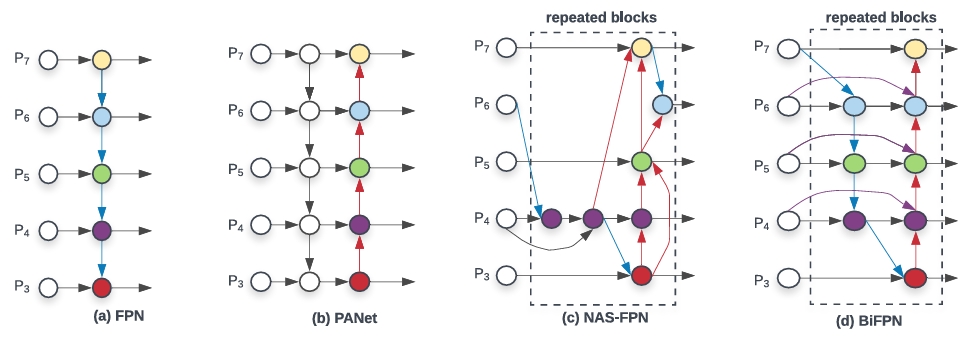


图3.4 原始BiFPN结构图

但是原始的BiFPN结构复杂，计算开销大。本文受BiFPN的启发，提出BiFPN4结构，改进CCFF的特征融合结构。BiFPN4保留了加权特征融合机制，通过优化节点连接策略和通道压缩在保留多尺度特征融合能力的同时，显著降低计算冗余。BiFPN4结构图如图3.5所示。

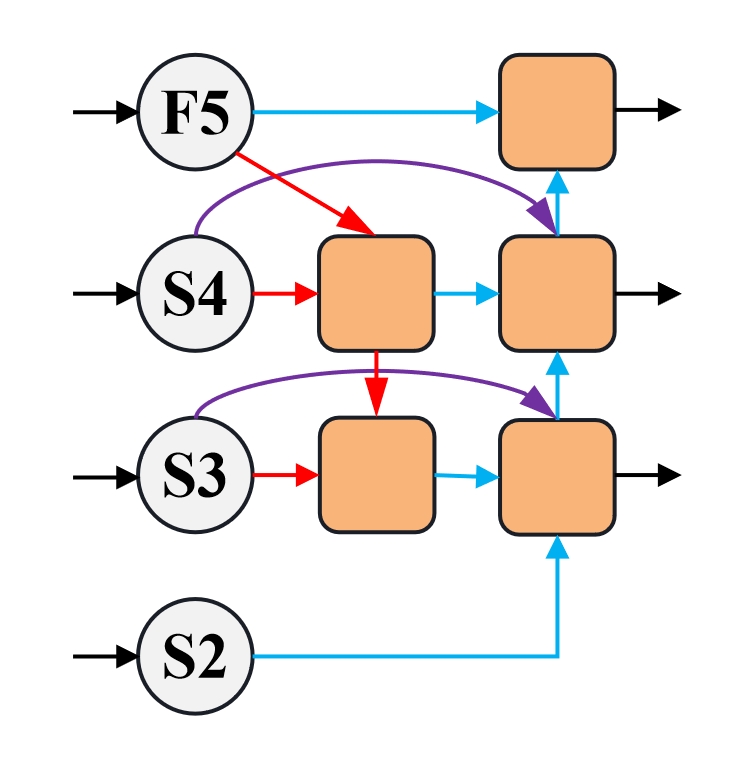


图3.5 BiFPN4结构图

### 3.2.3 Conv3XCNCSPELAN4重参数化模块

CSPNet（Cross Stage Partial Network）[26]是一种通过优化梯度传播路径和计算冗余来提升网络效率的架构设计，其核心思想是通过特征图分割与梯度流截断实现轻量化与性能平衡。RT-DETR中的特征融合单元RepC3模块由跨阶段部分网络（CSP）与重复堆叠的RepConv[27]结构构成。RepC3的固定通道分配策略会影响融合适应性，无法增强关键特征。RepC3的结构图如图3.6所示。

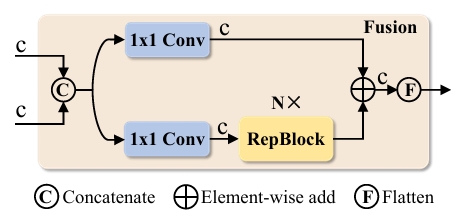


图3.6 RepC3结构图

ELAN（Efficient Layer Aggregation Network）[28]是一种基于梯度路径优化的特征融合架构，旨在通过多分支并行计算提升模型的特征表达能力。YOLOv9[29]作者通过将CSPNet与ELAN神经网络架构相结合，创新性地提出了广义高效层聚合网络（GELAN）。该架构通过引入CSPNet的设计思想，实现了对原始ELAN结构的泛化改进：将原本仅通过堆叠普通卷积层构建的ELAN架构，扩展为能够灵活整合任意计算模块的通用化框架。这种结构创新使GELAN在保持较高推理速度的同时，显著提升了模型的轻量化水平和检测准确性，实现了计算效率与模型性能的平衡优化。

重参数化是一种通过结构等效变换优化神经网络的方法，核心思想是将训练时的复杂结构（如多分支、冗余操作）转换为推理时的简化结构（如单路径），从而在不增加推理计算量的情况下提升模型性能，RepConv是最简单的重参数化模块。YOLOv9中作者将RepConv模块加入到GELAN结构中形成RepNCSPELAN4模块。

RepNCSPELAN4中的CSP通道分割和重参数化技术降低了模型的计算量，同时ELAN技术通过并行卷积分支提取多尺度特征，有效捕获不同大小的特征，提升遮挡场景下的检测精度，增强了模块对复杂场景的适应性。

然而，其核心组件RepConv的并行异构分支结构在训练时引入了梯度竞争问题，导致特征融合效率受限。针对X射线安检图像中违禁品堆叠的特性，本文提出采用串联式重参数化模块Conv3XC[30]替代RepConv，形成Conv3XCNCSPELAN4模块，替换RT-DETR原始的特征融合单元RepC3。

Conv3XC通过深度特征变换序列与跳跃连接（Skip Connection）协同优化，在减少参数量的同时增强局部细节保留能力，从而更适配高密度遮挡场景下的特征提取需求。

Conv3XC由三部分组成：跳跃连接的1×1卷积，保留原始特征的低阶语义信息；串联的1×1→3×3→1×1卷积序列，通过逐层通道压缩与空间特征提取实现高效上下文建模；动态权重融合机制，在推理阶段将多阶段卷积核自适应合并为单一等效3×3卷积核。相较于RepConv的并行异构分支，Conv3XC的串联结构显著降低了训练时的内存访问不连续性，并通过跳跃连接缓解深度特征变换导致的信息损失。Conv3XCNCSPELAN4结构图如图3.7所示。

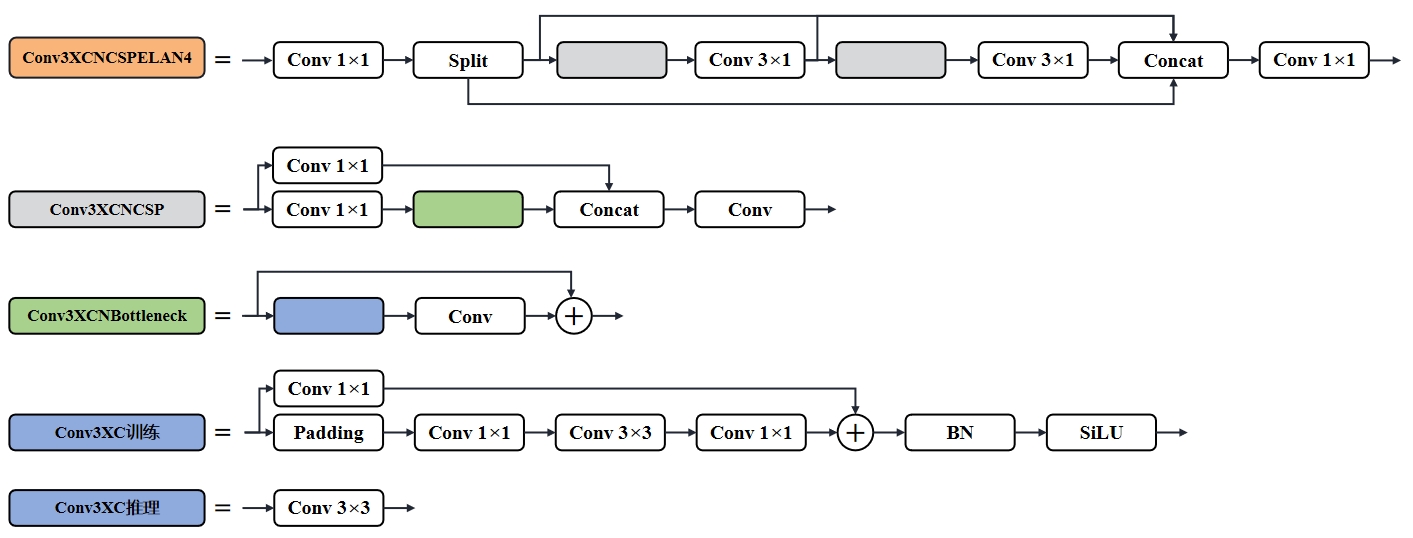


图3.7 Conv3XCNCSPELAN4结构图

### 3.2.4 Slide Varifocal Loss分类损失函数

RT-DETR中损失函数是分类损失函数和边界框回归损失函数的组合。其中分类损失函数为Varifocal Loss[31]。样本质量差异是目前分类损失函数面临的一大挑战，为了优化这一问题，本文使用样本加权函数改进Varifocal Loss形成Slide Varifocal Loss分类损失函数。

低质量样本的精度难以提升，过高质量的样本提升空间有限，中等质量样本精度的提升性价比更高。Slide Varifocal Loss在Varifocal Loss的基础上可以区分出不同质量的样本并通过分段调制权重分别设置高、中、低质量样本的权重。实现对低质量和过高质量样本的抑制，将训练提升的重点集中在中等质量样本上。

具体来说，Slide Varifocal Loss分类损失函数将所有边界框的IoU值的平均值作为阈值µ，区分出正负样本，之后给边界附近的样本分配更高的权重，样本加权函数可以表示为：

 （3.6）

样本权重分布如图3.8所示。

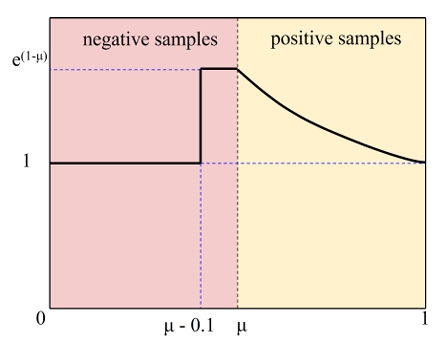


图3.8 样本权重分布[32]

## 3.3 构建完整的改进模型

本文首先将RT-DETR模型中的主干特征提取网络替换为FasterNet-T0，以PConv为单位对特征图的部分通道进行特征提取，在维持精度的同时减少冗余计算和内存访问提高违禁品检测效率。然后采用BiFPN4结构改进CCFF的特征融合结构，优化多尺度特征融合路径，提升特征融合能力。之后将Conv3XCNCSPELAN4添加到BIFPN4结构中，作为特征融合单元，通过并行卷积提取图像的多尺度特征，并利用重参数化技术进一步降低模型计算量。最后使用Slide Varifocal Loss分类损失函数优化样本权重分配，重点关注提升潜力更大的违禁品目标，提高整体检测精度。通过这一系列改进，模型可以在保持准确性和鲁棒性的同时，显著降低模型计算量、参数量和延时，实现高效的违禁品检测。

最终改进模型结构如图3.9所示。

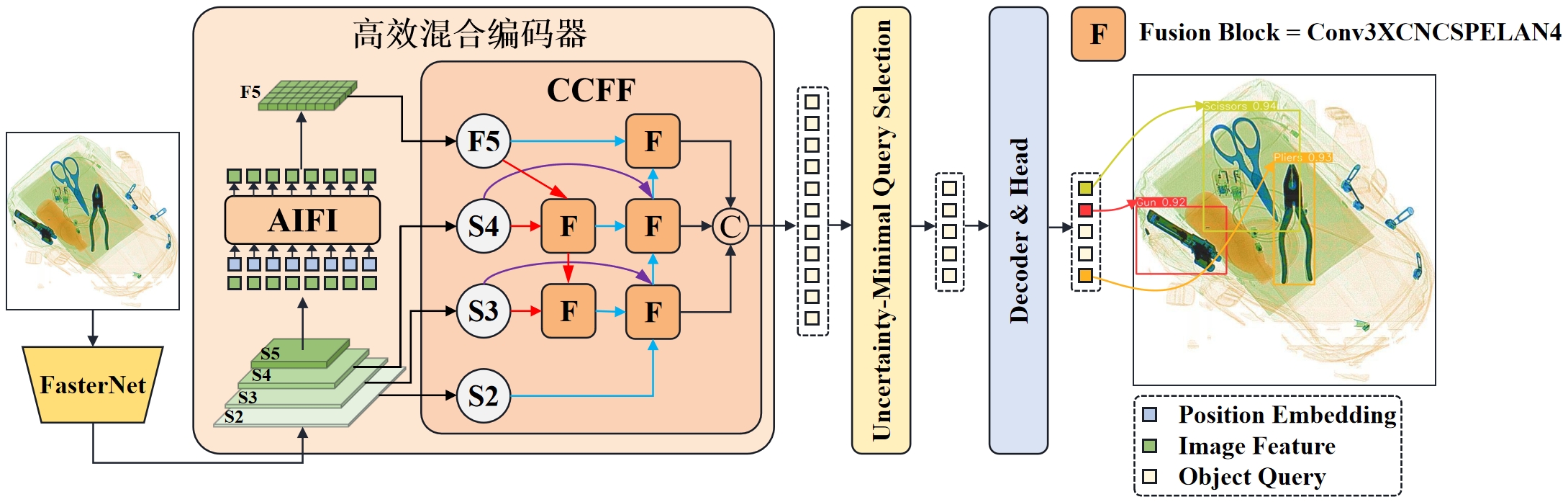


图3.9 最终改进模型结构

# 4 实验测试与分析

## 4.1 环境配置

本文实验在固定环境中进行，实验环境配置如下：

模型超参数设置为：预设分辨率为640×640像素，模型训练迭代周期数设定为350，每个批次处理的样本数量为16，数据加载时的工作线程数为16，使用AdamW优化器，初始学习率设定为0.0001。

本文所涉及实验采用的环境配置如表4.1所示。

表4.1 环境配置

|  |  |
| --- | --- |
| Configuration Items | Parameters |
| System environment | Ubuntu 20.04.5 |
| CPU | Intel(R) Xeon(R) Platinum 8163 CPU @ 2.50GHz |
| GPU | NVIDIA V100-16GB |
| Framework | Pytorch 1.12.1+cu113 |
| Cuda | 11.3 |

## 4.2 数据集

### 4.2.1 常见的违禁品数据集

常用的违禁品数据集如下：

GDXray[33]：是一个由1.9万张X射线图像组成的数据集，有五个组别，包括了金属刀具、食品等多种类型的物体，每张图片具有物体类别、物体位置等标注信息。此数据集中的图像为灰阶图像，生活中常见的安检图像为彩色，与现实场景相距甚远，同时该数据发布时间较早，图像质量不够高，物体背景简单，目标物体中既有危险物品又有合规物品重点不够突出，故未在本文中使用。

OPIXray[34]：是由北京航空航天大学在2020年发布的具有8885张图像，包含有折叠刀、直刀、剪刀、美工刀、多功能刀共5种类别的刀具，由国际机场的专业检查员手动标注，并根据物体的遮挡程度分为了三个类别。但是该数据集只涉及违禁刀具目标，没有包含其他违禁物品，种类单一，故未在本文中使用。

SIXray：是由中国科学院大学在2019年发布的由1,059,231张X射线图像组成的违禁物品检测数据集。其中8,929张图像包含了枪支、刀具、扳手、钳子、剪刀、锤子共6类生活中常见的违禁物品，锤子只有60个样本且没有标注信息，所以实际带标注信息的违禁物品为5种。这些图像从多个地铁站收集而来，具有违禁物品与合规物品混合、物体重叠、尺寸不同、角度不同等现实中客观存在的检测困难。SIXray数据集是常用的X射线安检违禁物品检测数据集，故本文以此为原始数据集。

### 4.2.2 数据增强

在真实的安检场景中使用目标检测算法检测违禁品时，主要面临以下三个关键挑战：

（1）违禁品角度多样性与完整性缺失

X射线安检图像中，违禁品（如刀具、枪支）可能因行李摆放姿态或遮挡呈现任意角度或局部残缺。检测模型对旋转目标的敏感性不足，易导致漏检或误检。为此，本文在SIXray数据集上引入以下数据增强策略：

随机水平/垂直翻转：模拟行李不同放置方向，强制模型学习对称性不变特征。

中心裁剪：截取图像中心区域，模拟遮挡场景，提升模型对残缺目标的识别能力。

（2）图像对比度动态变化

X射线成像设备因硬件差异或行李密度不均（如金属与衣物混装）可能导致成像对比度剧烈波动。低对比度图像中违禁品边缘模糊，易与背景混淆。本文通过以下方法增强数据对比度多样性：

随机对比度调整：对输入图像应用对比度系数，提高或降低像素数值，模拟不同穿透力下的成像效果。代码如下：

def random\_contrast(self, img, lower=0.5, upper=1.5, p=1):

# -------------------------------------

# 随机增强对比度

# -------------------------------------

if np.random.random() < p:

alpha=np.random.uniform(lower, upper)

img\*=alpha

img=img.clamp(min=0, max=1.0)

return img

（3）噪声干扰与图像退化

安检设备噪声（如光子计数噪声）、运动模糊或低分辨率成像会导致违禁品细节丢失。本文通过模拟真实退化过程增强数据：

高斯噪声注入：添加均值为0、标准差为0.1的高斯噪声，模拟传感器噪声。代码如下：

def add\_gasuss\_noise(self, img, mean=0, std=0.1):

noise=torch.normal(mean,std,img.shape)

img+=noise

img=img.clamp(min=0, max=1.0)

return img

高斯模糊与锐化：应用核大小为9的高斯模糊模拟运动模糊，随后通过锐化（拉普拉斯算子）恢复部分边缘信息，迫使模型同时学习抗模糊与细节捕捉能力。代码如下：

def rh(input\_folder,output\_folder):

if not os.path.exists(output\_folder):

os.makedirs(output\_folder)

for filename in os.listdir(input\_folder):

if filename.endswith('.jpg') or filename.endswith('.png'):

input\_path = os.path.join(input\_folder, filename)

output\_filename = "rh\_" + filename

output\_path = os.path.join(output\_folder, output\_filename)

image = cv2.imread(input\_path)

# 高斯模糊参数

gaussian\_blur\_kernel\_size = (9,9)

gaussian\_blur\_sigma\_x = 0

# gaussian\_blur\_kernel\_size：这个参数是高斯模糊的核大小，它定义了模糊效果的强度。核的大小越大，模糊效果越明显。

# gaussian\_blur\_sigma\_x：这是高斯模糊的标准差参数。它控制了高斯核的分布范围。较大的标准差会导致更广泛的模糊效果。如果设为0，OpenCV会根据核的大小自动计算标准差。

# 高斯模糊

blurred\_image = cv2.GaussianBlur(image, gaussian\_blur\_kernel\_size, gaussian\_blur\_sigma\_x)

# 拉普拉斯锐化

laplacian\_image = cv2.Laplacian(blurred\_image, cv2.CV\_64F)

laplacian\_image = cv2.convertScaleAbs(laplacian\_image)

# 合并原始图像和锐化图像以保持彩色

sharpened\_color\_image = cv2.addWeighted(image, 1, laplacian\_image, -1, 0)

cv2.imwrite(output\_path, sharpened\_color\_image)

print(f'Processed {filename} and saved to {output\_path}')

print('Processing completed.')

数据增强过程中，使用脚本从所有样本中随机选择图片进行处理，被一种数据增强方式处理过的图片会回到样本中有可能被继续选中进行另一种数据增强方式的处理。通过这种方式，图像经过多重增强，其多样性和丰富度可以得到进一步提升。数据增强效果图如图4.1所示。

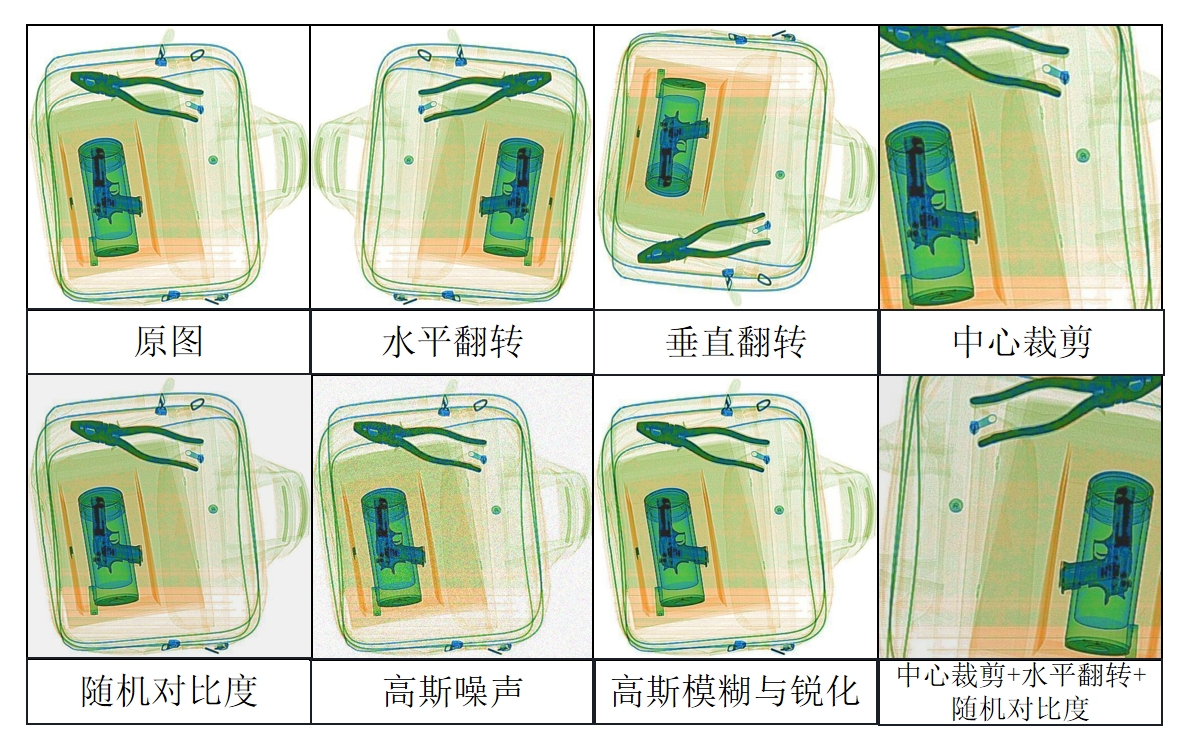


图4.1 数据增强效果图

### 4.2.3 SIXray-s数据集

本次实验从SIXray数据集含有的8929张有违禁品的图像中随机标记4000张图像，分析后发现剪刀一类数量太少，加上了剩余图像中只有剪刀一类违禁品的图像，并对图像进行了上述六种方式的数据增强得到5000张图像。这5000张经过处理的图像组成了具有枪支、刀具、扳手、钳子、剪刀五类违禁品、各种类样本数量更加平衡、图像特点更加多样的X射线违禁品数据集SIXray-s。SIXray数据集与SIXray-s数据集的各类别占比如图4.2所示。

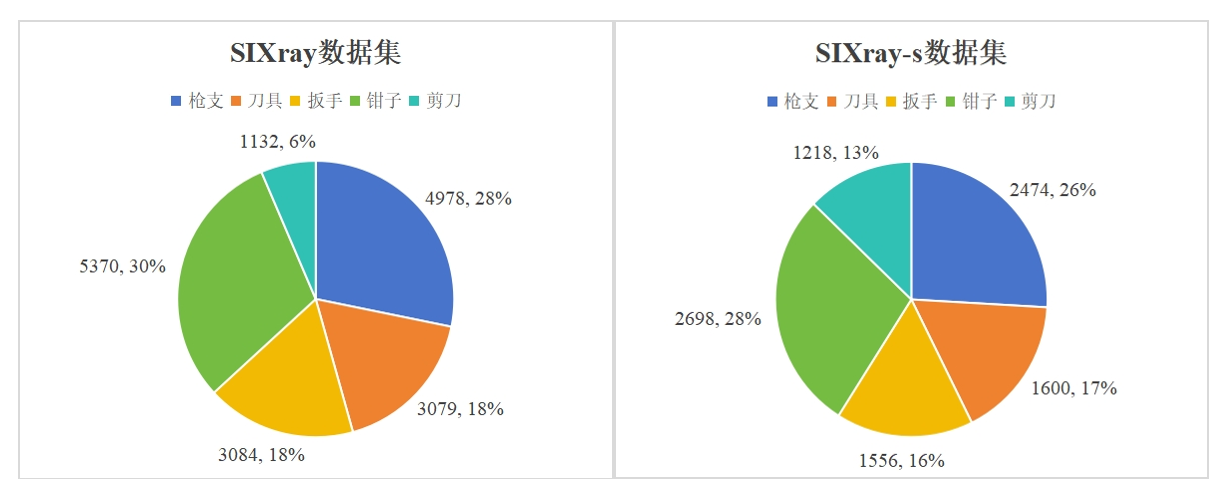


图4.2 SIXray数据集与SIXray-s数据集的各类别占比

在目标检测任务中，训练集是模型学习目标特征与定位能力的核心基础，通过大量标注数据驱动模型参数优化，迫使模型学习旋转不变性、抗噪声干扰等鲁棒特征；验证集则在训练过程中独立于训练数据，用于动态调整超参数（如学习率、锚框尺寸），同时通过监控验证集上的性能指标（如mAP@0.5）实现早停机制，防止模型因过度拟合训练数据中的噪声或特定样本而丧失泛化能力；测试集作为最终评估的“终极考场”，严格隔离于训练和调优过程，用于客观衡量模型在未知数据上的实际表现。三者共同构成模型开发的闭环流程——训练集提供知识输入，验证集指导优化方向，测试集确保性能可信。

将SIXray-s数据集以7：1：2的比例划分训练集、验证集和测试集，确保各阶段数据分布均匀，便于模型在不同阶段的学习与评估。本文所有实验均使用SIXray-s数据集进行。SIXray-s数据集的各类别分布如表4.2所示。SIXray-s数据集中训练集的标签情况如图4.3所示。

表4.2 SIXray-s数据集的各类别分布

|  |  |  |  |
| --- | --- | --- | --- |
| Categories | Training Set | Validation Set | Test Set |
| 枪支 | 1747 | 219 | 508 |
| 刀具 | 1140 | 143 | 317 |
| 扳手 | 1072 | 165 | 319 |
| 钳子 | 1893 | 292 | 513 |
| 剪刀 | 864 | 125 | 229 |

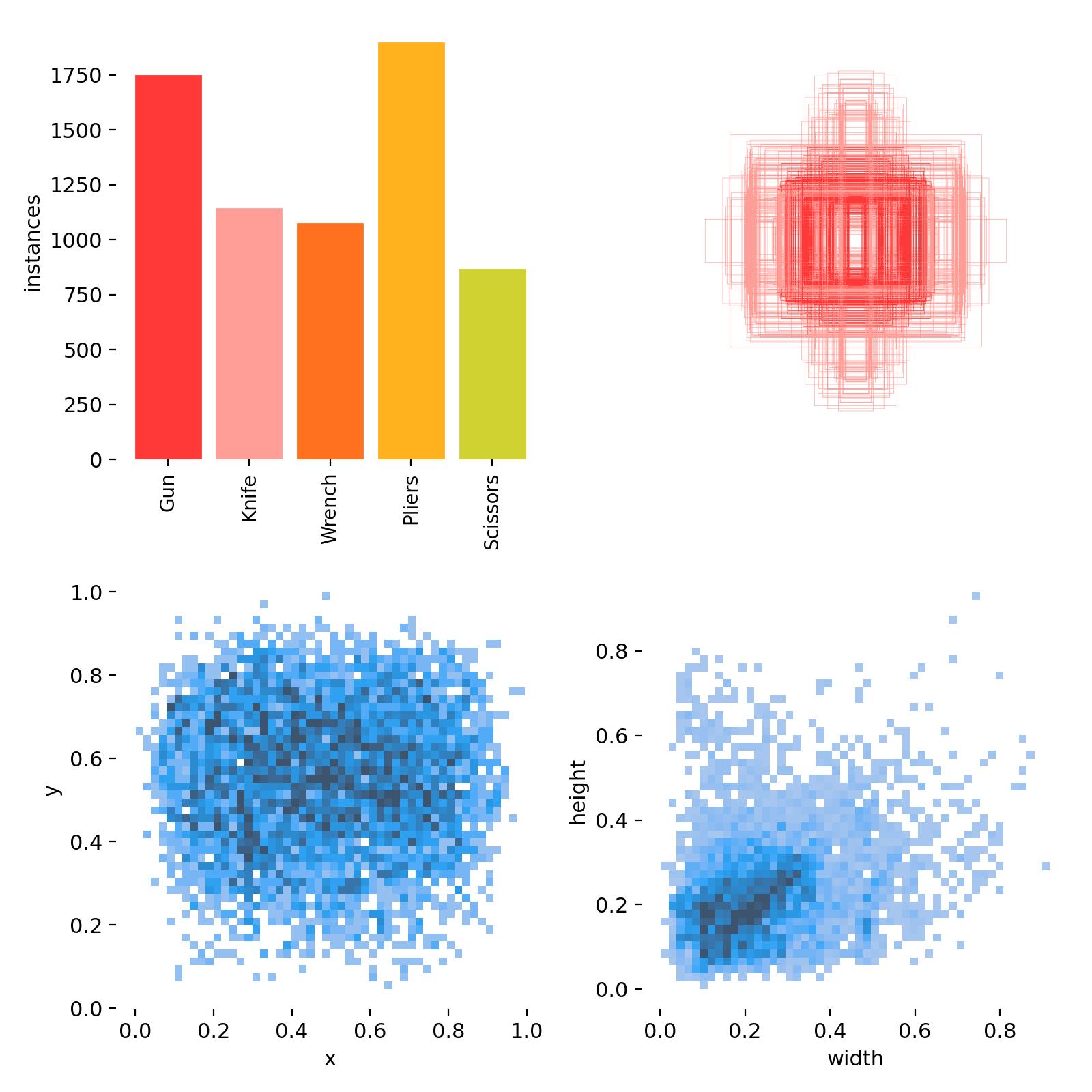


图4.3 SIXray-s数据集中训练集的标签情况

## 4.3 评价指标

为了分析改进前后模型的性能，本文以mAP@0.5、mAP@0.5:0.95、参数量Parameters（M）、计算量FLOPs（G）和FPS作为评价指标。常用指标如表4.3所示。

表4.3 常用指标

|  |  |  |
| --- | --- | --- |
| 指标 | 公式 | 意义 |
| 精确率 |  | 在被所有预测为正的样本中实际为正类的比例 |
| 召回率 |  | 在实际为正的样本中被预测为正类的比例 |
| AP值 |  | 以精确率和召回率分别作为纵、横坐标构成P-R曲线围成的面积 |
| mAP值 |  | AP的平均值 |

其中，TP为正类中被判断为正类的样本数，FN为正类中被判为负类的样本数，FP为负类中被判断为正类的样本数。使用IoU来划分正负样本，当IoU值大于阈值时，为正类样本，反之，为负类样本。N表示总的类别数。

mAP@0.5表示在IoU阈值为0.5时AP的平均值。

mAP@0.5:0.95表示在不同的IoU阈值（范围从0.50到0.95）下计算的mAP的平均值。它全面反映了模型在不同检测难度下的表现。

参数量和计算量用于衡量模型的复杂程度。

FPS为模型每秒可以处理的帧数，用于衡量检测速度，本文中的FPS取Batch size为16时测量的结果。

通过这些指标的综合评估，能够全面了解模型在精度、参数量、计算度和检测速度方面的表现。

4.4 模块对比实验

### 4.4.1 数据增强前后对比实验

本文针对真实安检场景中面临的违禁品角度多样性与完整性缺失、图像对比度动态变化、噪声干扰与图像退化的挑战，对数据集进行了水平翻转、垂直翻转、中心裁剪、随机对比度、高斯噪声、高斯模糊与锐化的数据增强操作。

本节将RT-DETR算法在SIXray-s数据集数据增强前和数据增强后分别进行了实验，数据增强前后RT-DETR的性能对比如表4.4所示。实验结果表明使用数据增强后数据集训练模型的mAP@0.5可达0.899，mAP@0.5:0.95可达0.659，较使用数据增强前数据集训练的模型均提高0.7个百分点。

这表明，通过针对性数据增强，本文构建了更贴近真实安检场景的训练集，显著提升了模型对角度变化、对比度波动与噪声干扰的鲁棒性，为复杂场景下的实时违禁品检测奠定了数据基础。

本节实验数据有力地证明了数据增强的必要性。

表4.4 数据增强前后RT-DETR的性能对比

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dataset | mAP@0.5 | mAP@0.5:0.95 | Params(M) | FLOPs(G) | FPS |
| SIXray-s数据集（数据增强前） | 0.892 | 0.652 | 19.9 | 57 | 199 |
| SIXray-s数据集（数据增强后） | 0.899 | 0.659 | 19.9 | 57 | 198 |

### 4.4.2 主干特征提取网络对比实验

本文选择FasterNet-T0替代RT-DETR中的原始主干特征提取网络ResNet-18。为了验证FasterNet-T0的效果，额外选择了其他几种经典的主干特征提取网络进行对比。主干特征提取网络替换性能对比结果如表4.5所示。实验结果表明，FasterNet-T0的参数量为10.8M，仅为ResNet-18（19.9M）的54.3%；计算量为28.5G，较ResNet-18（57G）减少50%；FPS为267，较ResNet-18（198）提升34.8%；mAP@0.5较ResNet-18降低了0.1个百分点。在对比模型中，FasterNet-T0的参数量和计算量处于较低水平，略高于EfficientViT[35]和MobileNet V3-Large[36]；但mAP和FPS处于较高水平，显著高于前两者。

替换模型主干特征提取网络性能对比图如图4.4所示。此图以FPS为横轴，mAP@0.5为纵轴，标记尺寸反映模型参数量，将主干特征提取网络对比实验中涉及的共七个模型的性能绘制到图中。位置越接近右上角、尺寸越小的模型性能越好。从图中可以直观的看出使用以FasterNet-T0作为主干特征提取网络的模型的位置最接近右上角，且尺寸在所有模型中处于较低水平。这表明在七种主干特征提取网络中FasterNet-T0的综合性能最好，在精度、效率之间实现了平衡。

本节实验数据有力地证明了FasterNet-T0作为主干特征提取网络的优越性。

表4.5 主干特征提取网络替换性能对比

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Backbone | mAP@0.5 | mAP@0.5:0.95 | Params(M) | FLOPs(G) | FPS |
| EfficientViT | 0.876 | 0.644 | 10.7 | 27.2 | 216 |
| ConvNeXt V2[37] | 0.88 | 0.626 | 12.3 | 31.9 | 165 |
| MobileNet V3-Large | 0.871 | 0.648 | 11.4 | 27.1 | 200 |
| MobileNet V4-Conv-Small[38] | 0.885 | 0.646 | 11.3 | 39.5 | 275 |
| StarNet-S1[39] | 0.872 | 0.645 | 11.2 | 29.7 | 205 |
| ResNet-18（原始） | 0.899 | 0.659 | 19.9 | 57 | 198 |
| FasterNet-T0（改进） | 0.898 | 0.644 | 10.8 | 28.5 | 267 |

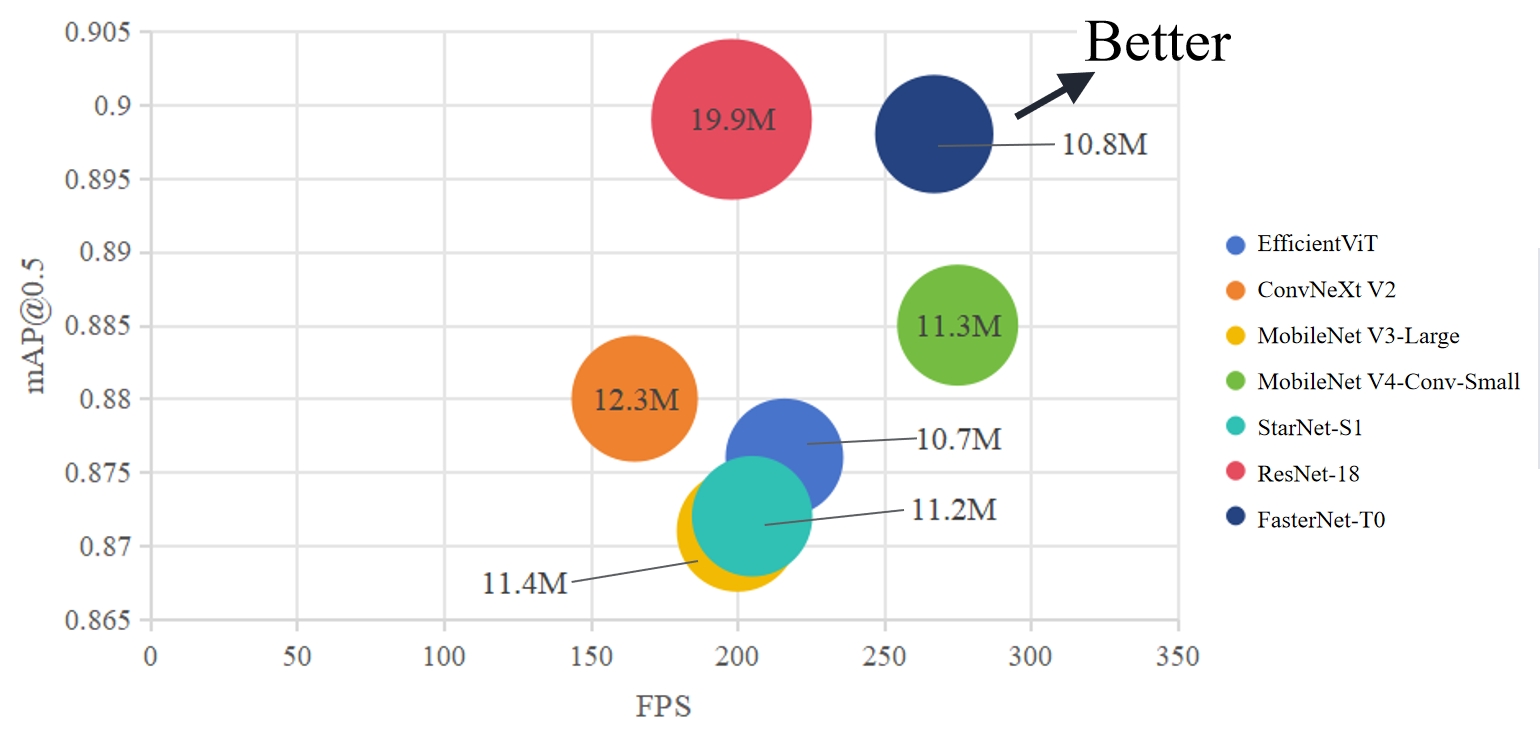


图4.4 替换模型主干特征提取网络性能对比图

### 4.4.3 特征融合单元对比实验

DiverseBranchBlock[40]是常用的重参数化模块，为验证Conv3XCNCSPELAN4模块的效果，本文在RepNCSPELAN4与RepC3中引入DiverseBranchBlock模块，形成DBBNCSPELAN4与DBBC3，一起训练进行性能对比。以FasterNet-T0作为主干特征提取网络的基础上，特征融合单元替换性能对比结果如表4.6所示。实验结果表明，在对比模型中Conv3XCNCSPELAN4的精度最高，mAP@0.5为0.898与RepC3持平，同时较RepNCSPELAN4提升1.2个百分点；mAP@0.5:0.95为0.666较RepC3提高了2.2个百分点，同时较RepNCSPELAN4提升3.9个百分点；参数量为9.4M，较RepC3（10.8M）降低13%；计算量为21.5G，较RepC3（28.5G）降低24.6%；FPS为272，较RepC3（267）也有所提升。在对比模型中，Conv3XCNCSPELAN4在的精度最高，同时保证了轻量化水平。

本节实验数据有力地证明了Conv3XCNCSPELAN4作为特征融合单元的优越性。

表4.6 特征融合单元替换性能对比

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Block | mAP@0.5 | mAP@0.5:0.95 | Params(M) | FLOPs(G) | FPS |
| RepNCSPELAN4 | 0.886 | 0.627 | 9.4 | 21.5 | 281 |
| DBBNCSPELAN4 | 0.884 | 0.651 | 9.4 | 21.5 | 279 |
| DBBC3 | 0.89 | 0.662 | 10.8 | 28.5 | 267 |
| Conv3XCC3 | 0.894 | 0.666 | 10.8 | 28.5 | 260 |
| RepC3（原始） | 0.898 | 0.644 | 10.8 | 28.5 | 267 |
| Conv3XCNCSPELAN4（改进） | 0.898 | 0.666 | 9.4 | 21.5 | 272 |

## 4.5 改进消融实验（纵向对比）

消融实验（Ablation Study）是一种在科学研究和深度学习中广泛使用的实验方法，主要用于评估一个系统、模型或理论的组成部分对整体的影响。消融实验通过逐步增加或删除系统的部分，观察系统性能的变化，从而深入理解各个组成部分的贡献。为了验证改进模块对违禁品检测的有效性，在测试集上进行了下面六组实验。

改进消融实验性能对比结果如表4.7所示。实验结果表明，将RT-DETR的主干特征提取网络从ResNet-18改为以PConv为基础的FasterNet-T0后，较模型A的参数量下降了45.7%、计算量下降了50%、FPS提升了34.8%，表明PConv可以减少冗余计算和内存访问，提升模型效率。

以FasterNet-T0作为主干特征提取网络的基础上，将模型的CCFF模块改为通道数为128的BIFPN4结构后，较模型B的参数量进一步下降25%、计算量下降34.4%、FPS提升13.5%，证明BIFPN4优化了节点连接策略和通道压缩的双向跨尺度连接和加权特征融合机制可以在低参数、低计算的基础上保持特征融合能力。

以FasterNet-T0作为主干特征提取网络的基础上将模型的融合块从RepC3改为Conv3XCNCSPELAN4后，较模型B的mAP@0.5保持不变，mAP@0.5:0.95提升了2.2个百分点，同时参数量下降了13%、计算量下降了24.6%、FPS也有所提升。同时，在模型C的基础上引入Conv3XCNCSPELAN4后，mAP@0.5提升0.6个百分点，mAP@0.5:0.95提升4.2个百分点，弥补了模型C在轻量化之后精度上的损失。实验数据证明结合了CSPNet与ELAN结构，并引入Conv3XC的Conv3XCNCSPELAN4模块，可以在降低模型的参数量、计算量，提高模型FPS的同时通过并行卷积分支提取多尺度特征，有效捕获不同大小的特征，提升违禁品的检测精度。

在模型E的基础上引入Slide Varifocal Loss分类损失函数后，模型mAP@0.5提升0.7个百分点，证明Slide Varifocal Loss分类损失函数可以提高模型整体的检测精度。

消融实验模型性能对比图如图4.5所示。此图以FPS为横轴，mAP@0.5为纵轴，标记尺寸反映模型参数量，将消融实验中涉及的A、B、C、D、E、F六个模型的性能绘制到图中。位置越接近右上角、尺寸越小的模型性能越好。从图中可以直观的看出每次添加模块后模型性能的变化。模型A为最原始模型，模型F为最终模型，从模型A到模型F，位置向右上角靠近，同时尺寸变小。图中模型F的位置最接近右上角，且尺寸在所有模型中处于较低水平。这些都证明最终模型F是最优的，每个改进模块都是有效且不可或缺的。

本节实验数据有力地证明了改进模块的有效性。

表4.7 改进消融实验性能对比

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ID | FasterNet-T0 | BIFPN4 | Conv3XCNCSPELAN4 | Slide Varifocal Loss | mAP@0.5 | mAP@0.5:0.95 | Params(M) | FLOPs(G) | FPS |
| A | × | × | × | × | 0.899 | 0.659 | 19.9 | 57 | 198 |
| B | √ | × | × | × | 0.898 | 0.644 | 10.8 | 28.5 | 267 |
| C | √ | √ | × | × | 0.889 | 0.621 | 8.1 | 18.7 | 303 |
| D | √ | × | √ | × | 0.898 | 0.666 | 9.4 | 21.5 | 272 |
| E | √ | √ | √ | × | 0.895 | 0.663 | 8.3 | 20 | 263 |
| F | √ | √ | √ | √ | 0.902 | 0.661 | 8.3 | 20 | 263 |

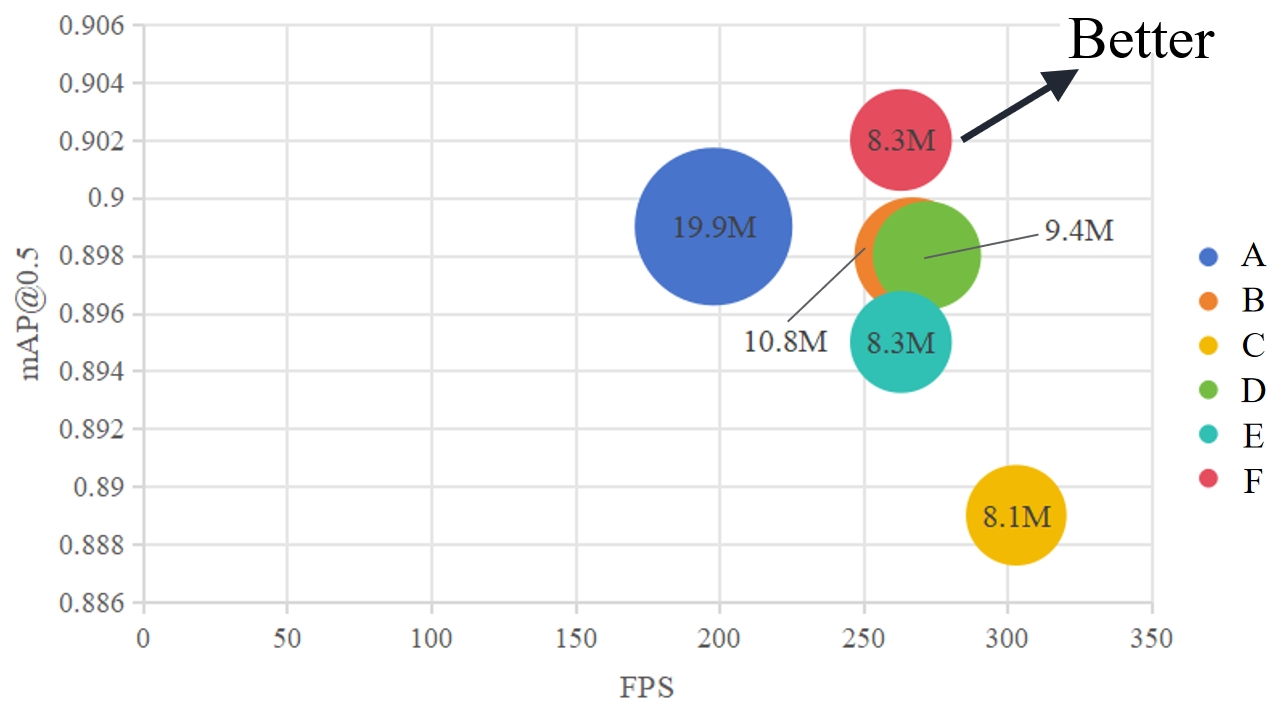


图4.5 消融实验模型性能对比图

## 4.6 模型对比实验（横向对比）

目标检测是基础的计算机视觉任务，目标检测算法种类繁多。为了证明本文改进模型的优越性，本节加入了多种经典的目标检测算法进行横向对比。

模型对比实验性能对比如表4.8所示。实验结果表明，本文改进模型可以在低参数量、低计算量的条件下，实现高mAP和高FPS。同时改进模型具有无需后处理，可以实现实时端到端检测的优势。这些都证明了本文改进模型与当前主流算法相比在实际应用中，特别是在违禁品检测等需要实时性能场景中具备显著的优势。

本节实验数据有力地证明了本文模型的优越性。

表4.8 模型对比实验性能对比

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | mAP@0.5 | mAP@0.5:0.95 | Params(M) | FLOPs(G) | FPS |
| Conditional DETR(ResNet-18)[41] | 0.781 | 0.415 | 30.7 | 44.7 | × |
| Deformable DETR(ResNet-18)[42] | 0.819 | 0.519 | 23.5 | 119 | × |
| Dino(ResNet-18)[43] | 0.88 | 0.606 | 31 | 187 | × |
| RTMDet-m[44] | 0.878 | 0.576 | 24.7 | 39.1 | × |
| Faster-RCNN(ResNet-18) | 0.881 | 0.529 | 28.3 | 129 | × |
| YOLOv5m | 0.897 | 0.652 | 25 | 64 | 210 |
| YOLOv8m | 0.892 | 0.654 | 25.8 | 78.7 | 184 |
| RT-DETR(ResNet-18) | 0.899 | 0.659 | 19.9 | 57 | 198 |
| 本文模型 | 0.902 | 0.661 | 8.3 | 20 | 263 |

## 4.7 最终模型评估图

### 4.7.1 混淆矩阵

在目标检测任务中，混淆矩阵（Confusion Matrix）是评估模型性能的重要工具。它展示了预测结果与真实标签之间的关系，通过混淆矩阵可以直观地了解模型在各个类别上的表现，从而帮助分析模型的分类性能。

模型A和模型E在测试集上的混淆矩阵如图4.6所示。该混淆矩阵呈现了模型对各品种违禁品的预测结果，矩阵中对角线上为各个类别中预测正确的比例，纵向为模型对违禁品种类的预测结果，横向为违禁品的实际种类。

最终模型的混淆矩阵如图4.7所示。最终模型以Slide Varifocal Loss作为分类损失函数，通过对比图4.6与图4.7可知，最终模型的混淆矩阵中处于对角线位置的值与模型E相比提升明显，特别是实际为刀具被最终模型预测为刀具的比例为0.89，较模型E（0.71）提高25.4%，这证明引入样本权重函数的Slide Varifocal Loss分类损失函数可以提升模型的分类性能。同时，最终模型的混淆矩阵中处于对角线位置的值与模型A和模型E相比均为最高值，这展示了最终模型的优秀表现。

本节混淆矩阵数据有力地证明了Slide Varifocal Loss分类损失函数的有效性以及最终模型分类性能的优越性。



图4.6 模型A和模型E的混淆矩阵

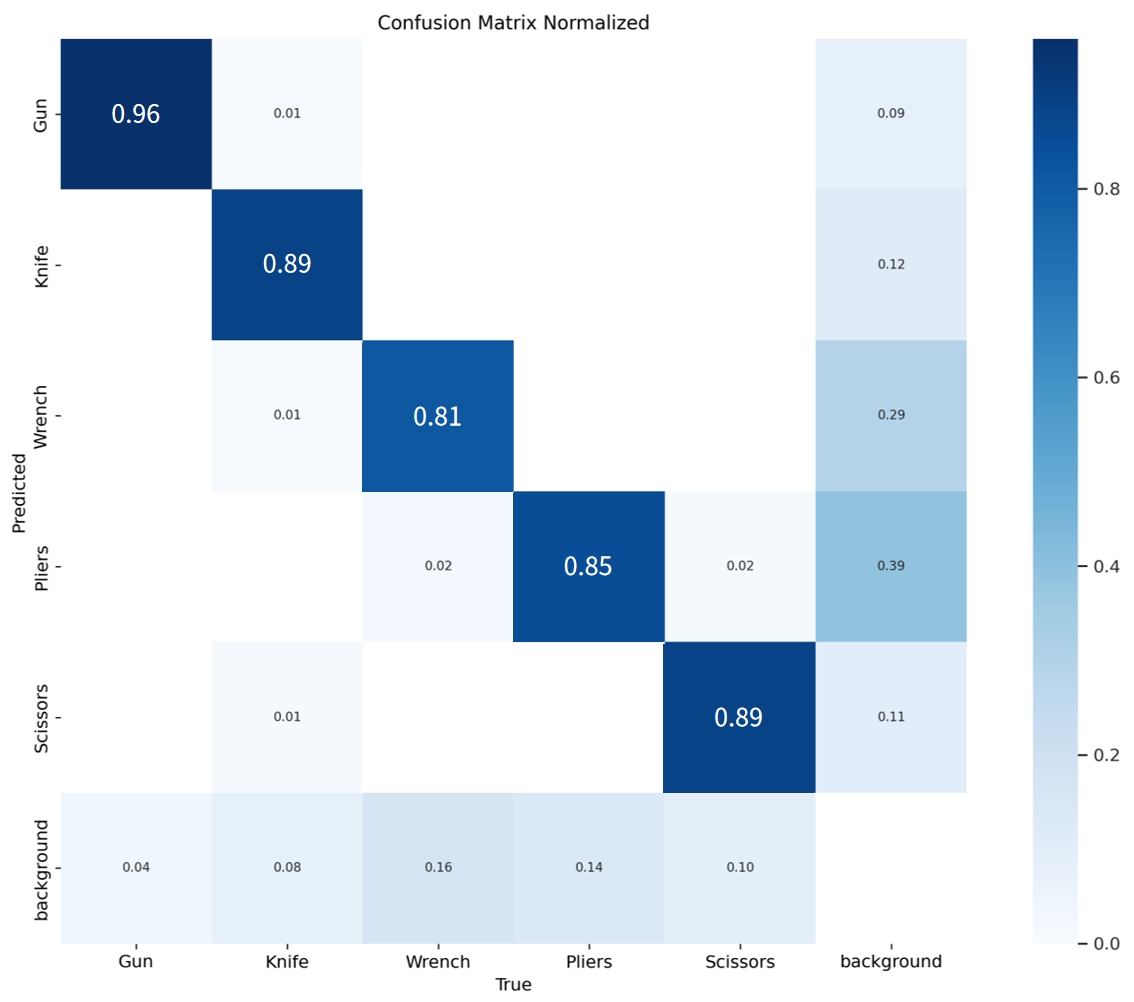


图4.7 最终模型的混淆矩阵

### 4.7.2 损失值变化

在模型训练过程中，损失值的变化可以反映出模型的收敛情况和优化效果。本节使用两种损失值来评估最终模型的训练效果。定位损失（giou\_loss）用于衡量预测框与真实框之间的误差，数值越小表示定位准确度越高。分类损失（cls\_loss）则用于评估预测类别与实际类别之间的准确性，数值越小表示分类结果越准确。

最终模型训练损失变化如图4.8所示。从图中可以看出最终模型在训练集和验证集上的损失值在初始阶段迅速下降，随后趋于平稳，表明模型在训练过程中逐渐学习到有效特征，最终性能已接近其最优性能。

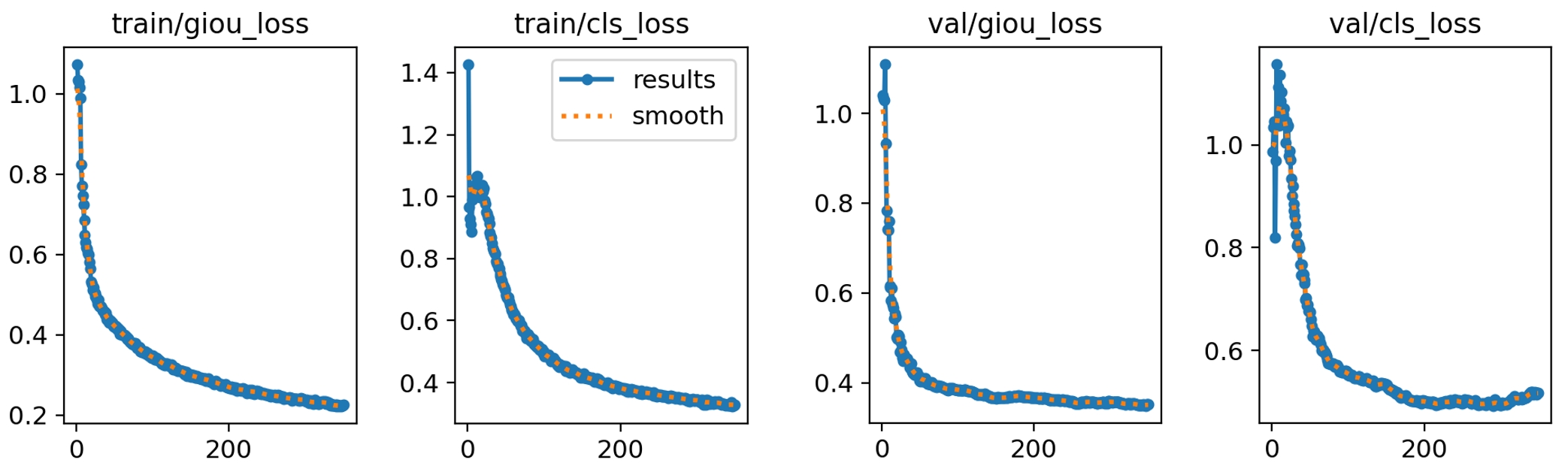


图4.8 最终模型训练损失变化

### 4.7.3 mAP变化

在目标检测任务中，mAP（mean Average Precision）是衡量模型性能的核心指标。它综合了分类精度和召回率，能够全面反映模型在不同类别上的检测能力。随着训练轮数（epoch）的增加，mAP@0.5和mAP@0.5:0.95逐渐提高并最终趋于稳定，训练所得模型已接近其最优性能，最终模型训练mAP变化如图4.9所示。

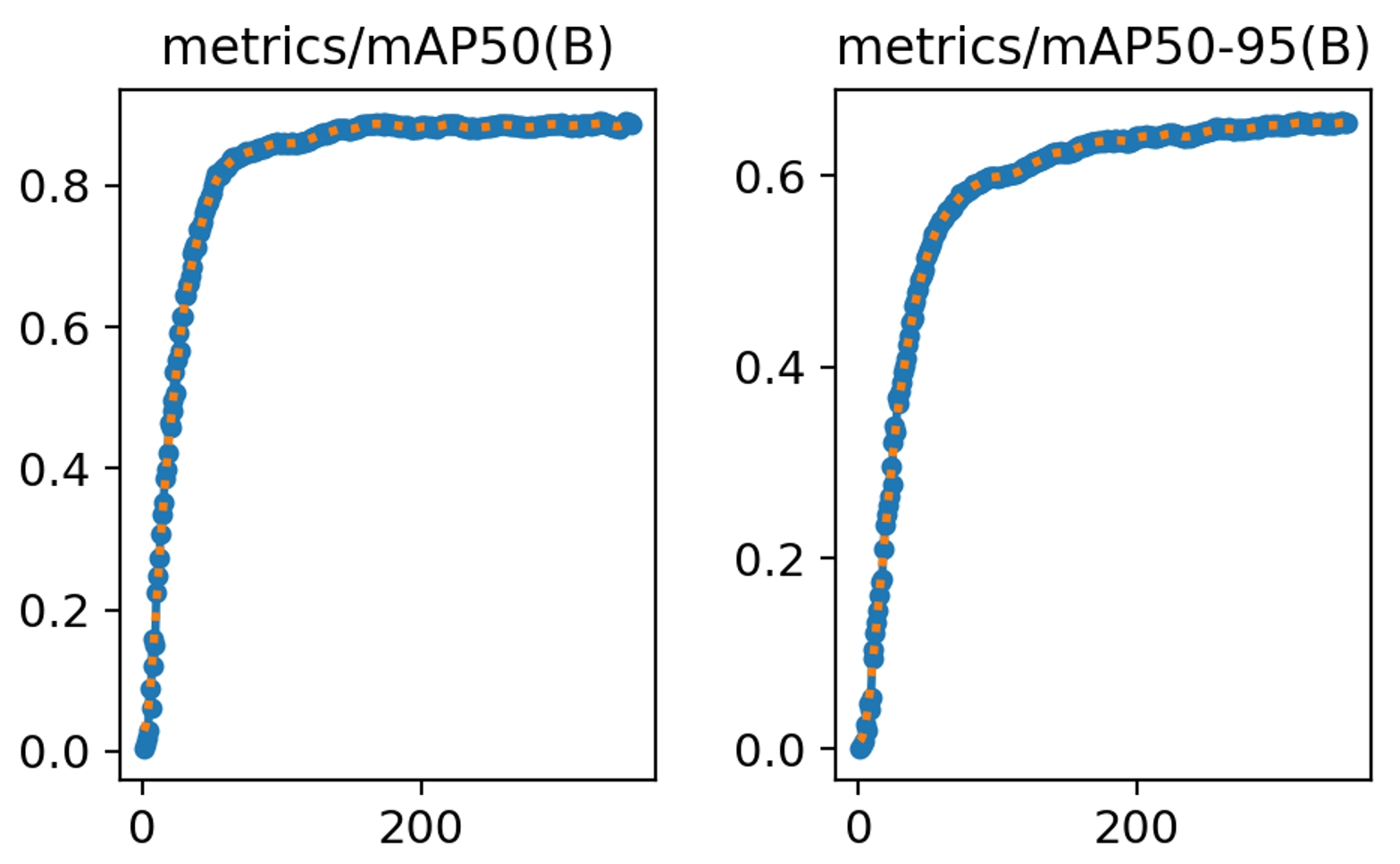


图4.9 最终模型训练mAP变化

### 4.7.4 评价指标曲线图

在评估模型的性能时，Precision-Confidence曲线、Recall-Confidence曲线以及F1\_Confidence曲线也是常用的工具，本节采用以上三种曲线联合分析模型的多维度性能。曲线分别如图4.10-4.12所示。

从图中可以看出，精确率随着置信度的增大而增大，这是因为高置信度意味着模型只保留更加确信的预测框，低置信度的预测（可能是噪声或模糊目标）被过滤，使得FP减小，进而导致精确率增大；召回率随着置信度的增大而减小，这是因为高置信度会使一些正类因置信度低于阈值而被丢弃，使得FN增加，进而导致召回率减小；

而F1值随着置信度的增大先增大后减小，是因为F1值是精确率和召回率的调和平均数。当置信度增大时，一开始精确率的提升比召回率的下降更快，导致F1值上升。但到了某个阈值后，召回率的下降速度超过了精确率的提升速度，F1值就会开始下降。所以F1值呈现先升后降的趋势。从图4.12中可知，在置信度为0.501时F1值取得最大值0.87。

本节曲线数据说明在使用最终模型进行违禁品检测的过程中，置信度阈值设置为0.5时，模型能够实现最优的检测效果。

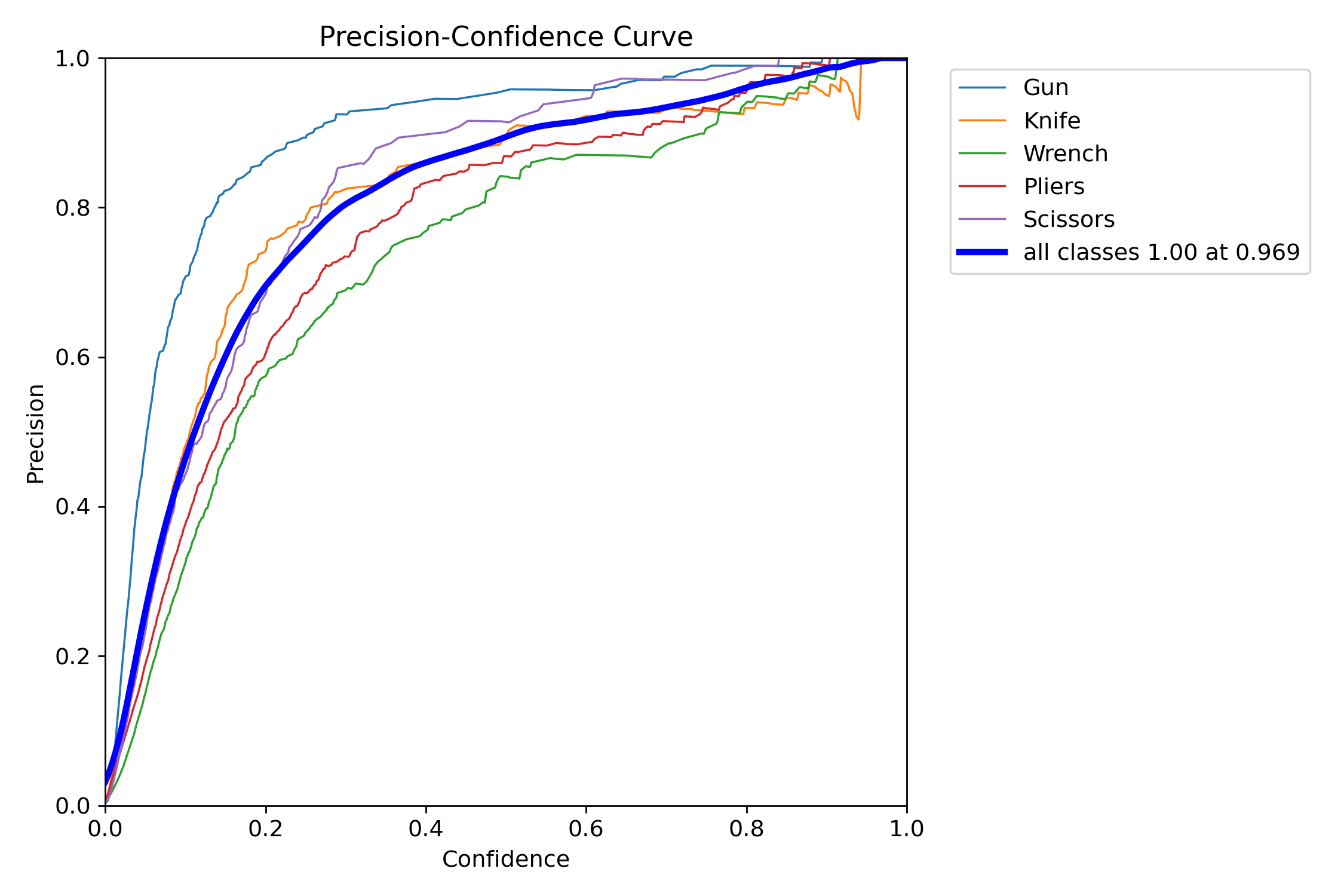


图4.10 Precision-Confidence曲线图

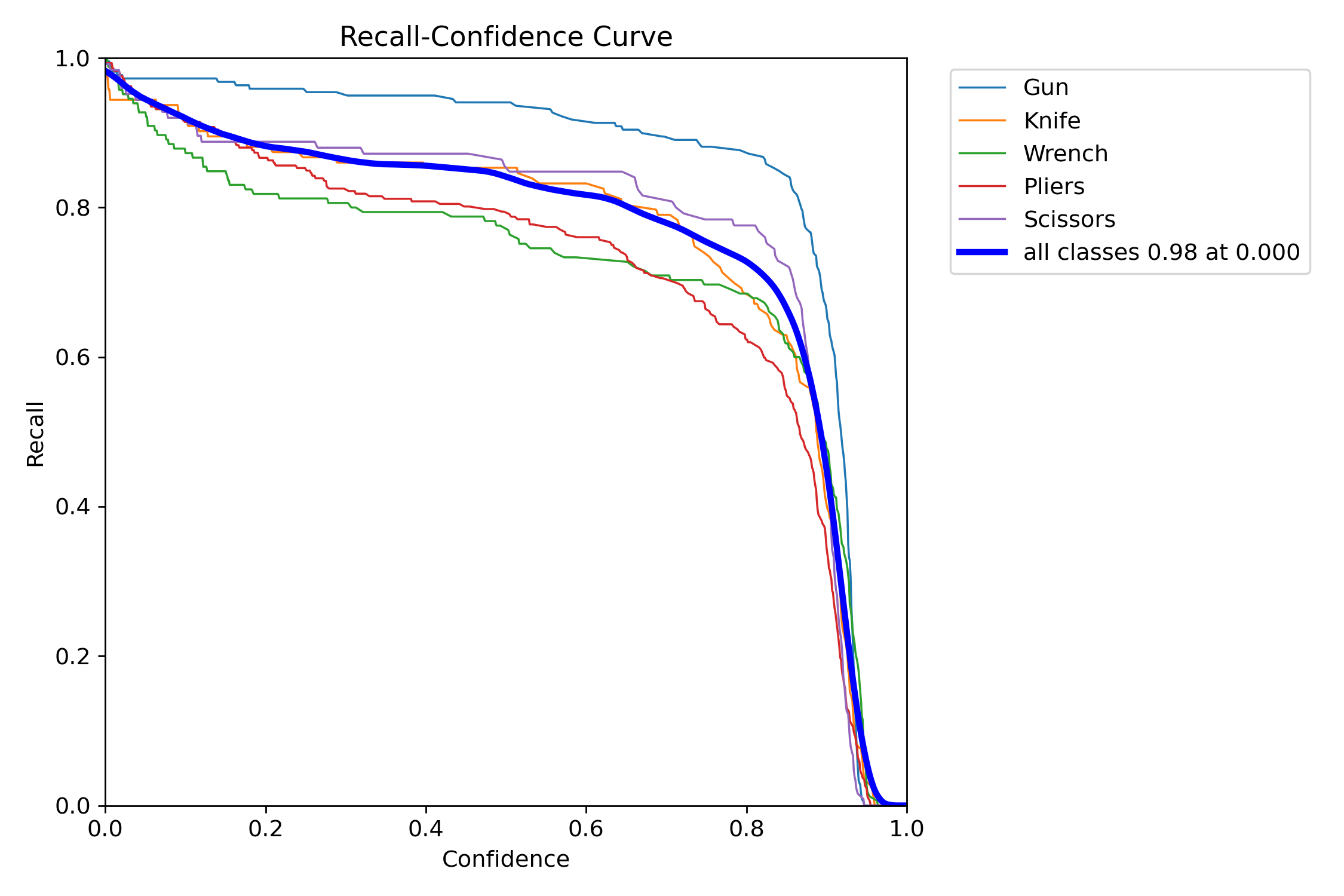


图4.11 Recall-Confidence曲线图

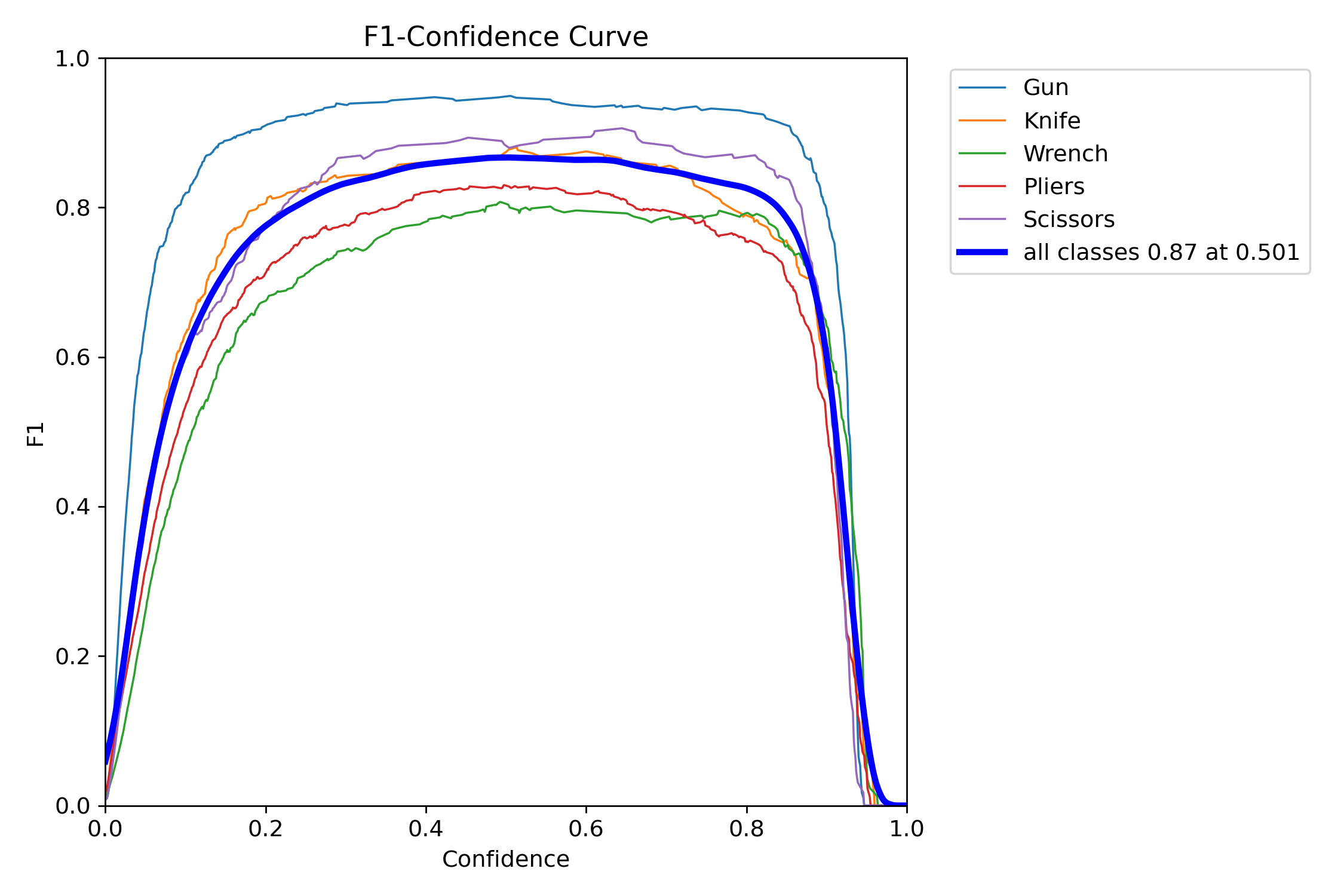


图4.12 F1-Confidence曲线图

### 4.7.5 热力图

GradCAM++（Gradient-weighted Class Activation Mapping++）[45]是一种改进的类激活映射技术，用于可视化模型在目标检测或分类任务中的关注区域。它通过分析模型梯度信息，生成热力图（Heatmap）来指示输入图像中对模型决策起关键作用的区域。

本节利用GradCAM++技术生成了最终模型的类激活热力图，以可视化该模型对违禁品图像的关注区域。最终模型热力图如图4.13所示。图中展示了违禁品的原始图像以及对应的热力图。从图中可以看出，热力图高响应区域与违禁品位置重合，表明模型能够正确捕捉违禁品的关键特征。

本节热力图证明最终模型能够实现对违禁品的有效检测。

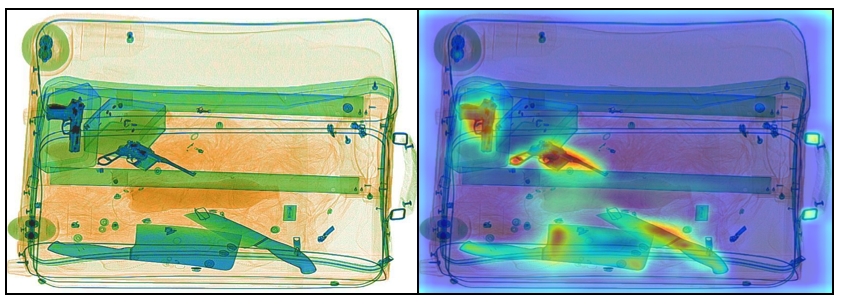


图4.13 最终模型热力图

# 5 违禁品检测系统

## 5.1 Ultralytics

Ultralytics是一个基于PyTorch构建的开源计算机视觉工具库，专注于为深度学习模型的开发、训练与部署提供高效且易用的解决方案。该库以简化复杂视觉任务的工程化实现为核心目标，重点支持目标检测、图像分类及图像分割等任务，其核心模型体系涵盖经典的YOLO系列（如YOLOv5、YOLOv8）以及RT-DETR。

通过模块化设计，Ultralytics将数据处理、模型优化与部署流程深度整合：内置的data模块提供数据集加载（BaseDataset）、预处理与增强方法（Augment），支持Mosaic增强、MixUp等策略以提升模型泛化能力；命令行工具链则允许开发者通过yolo detect train等指令快速完成模型训练、验证及超参数调优。此外，该库集成混合精度训练、模型量化技术（如TensorRT/ONNX导出）及Triton推理服务器适配功能，显著优化了从研发到生产环境的迁移效率。

## 5.2 PyQt5

PyQt5是基于Qt框架的Python跨平台GUI开发库，通过封装Qt的C++接口为Python模块，实现高效桌面应用开发。相较于Tkinter等基础工具包，PyQt5在复杂界面响应速度与功能扩展性上具有显著优势，尤其适用于交互界面开发。该框架将Qt生态的完整性和Python语法简洁性结合，是桌面应用开发的主流选择之一。

5.3 功能与界面

为了方便用户进行违禁品检测，本文应用‌Ultralytics库和PyQt5库，设计了一套便捷高效的违禁品检测系统。系统界面如图5.1所示。

图5.1 系统界面图

本系统支持模型、参数、运行环境的自由配置。用户在使用本系统前需要先选择Yaml配置文件。在配置文件中用户可以进行检测模型权重、交互比阈值、置信度阈值、系统运行环境、检测类别名称，图片尺寸、模型类型的自由设置。系统为用户提供了充分的自由度，通过不同的配置，系统可以适用于不同的使用场景。Yaml配置文件示例如图5.2所示。



图5.2 Yaml配置文件示例

本系统支持四种检测方式，分别为图片、文件夹、视频和摄像头检测。图片检测为对单一图片进行违禁品检测；文件夹检测支持选择图片文件夹地址，对整个文件夹中的图片进行批量检测；视频检测支持用户导入MP4视频文件，系统逐帧分析并实时显示检测结果；摄像头检测则连接摄像头设备，实时捕捉画面并进行违禁品识别。检测完成后会在右侧显示带有违禁品标签框、类别以及置信度的检测结果。同时会显示文件地址和各种类违禁品的数量。其中图片、视频、摄像头检测效果图如图5.3-5.5所示。

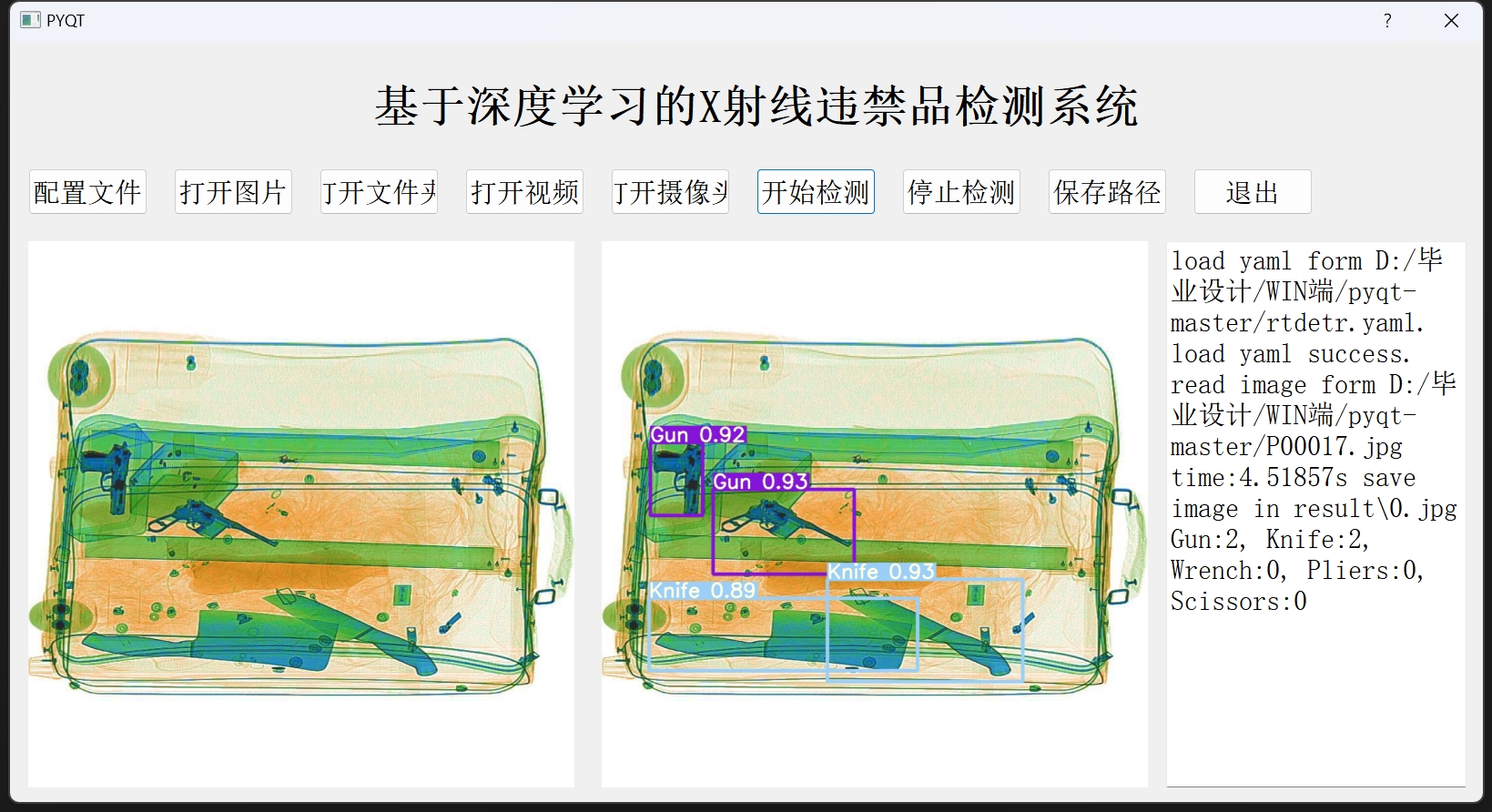


图5.3 图片检测效果图

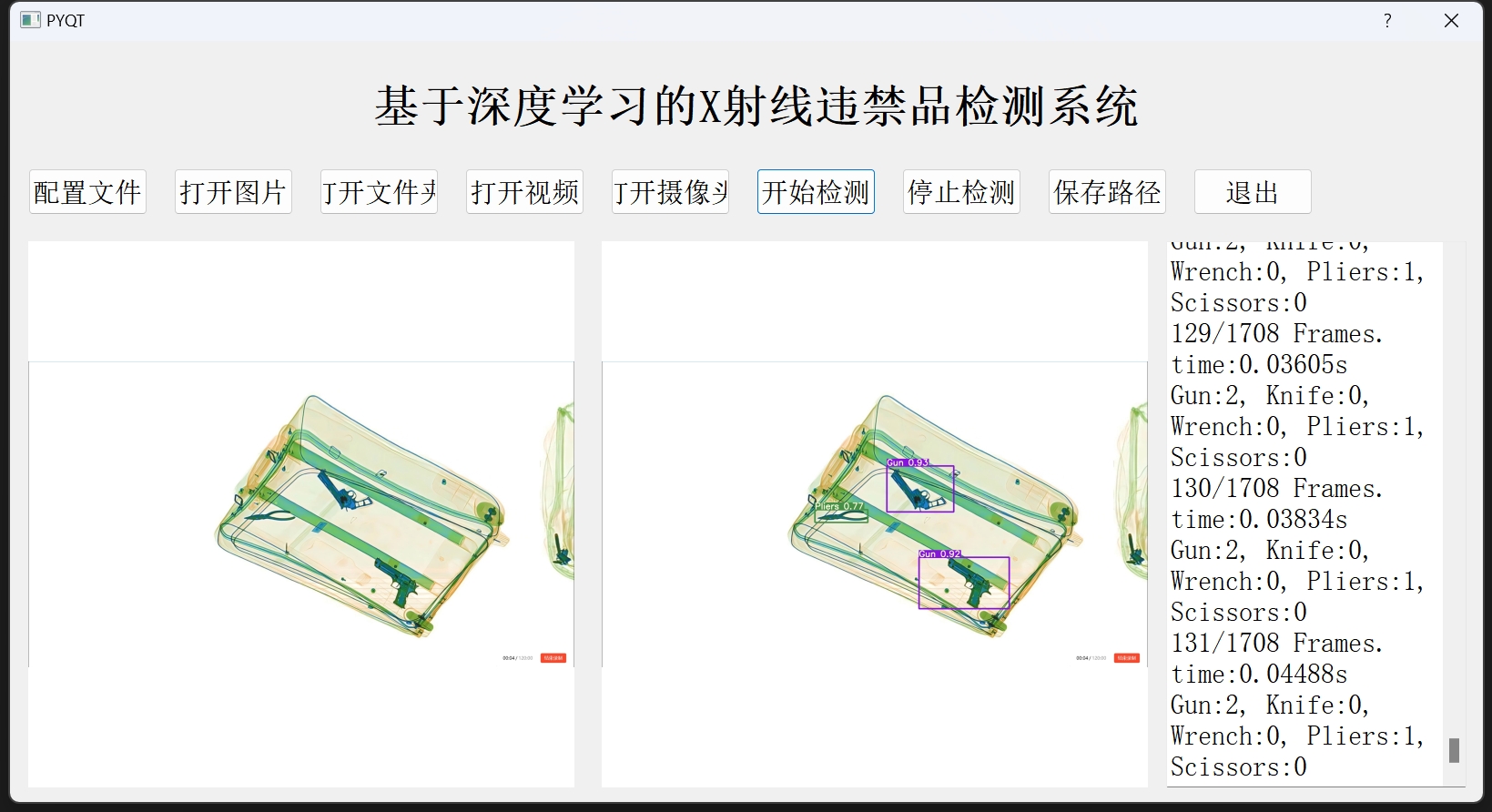


图5.4 视频检测效果图

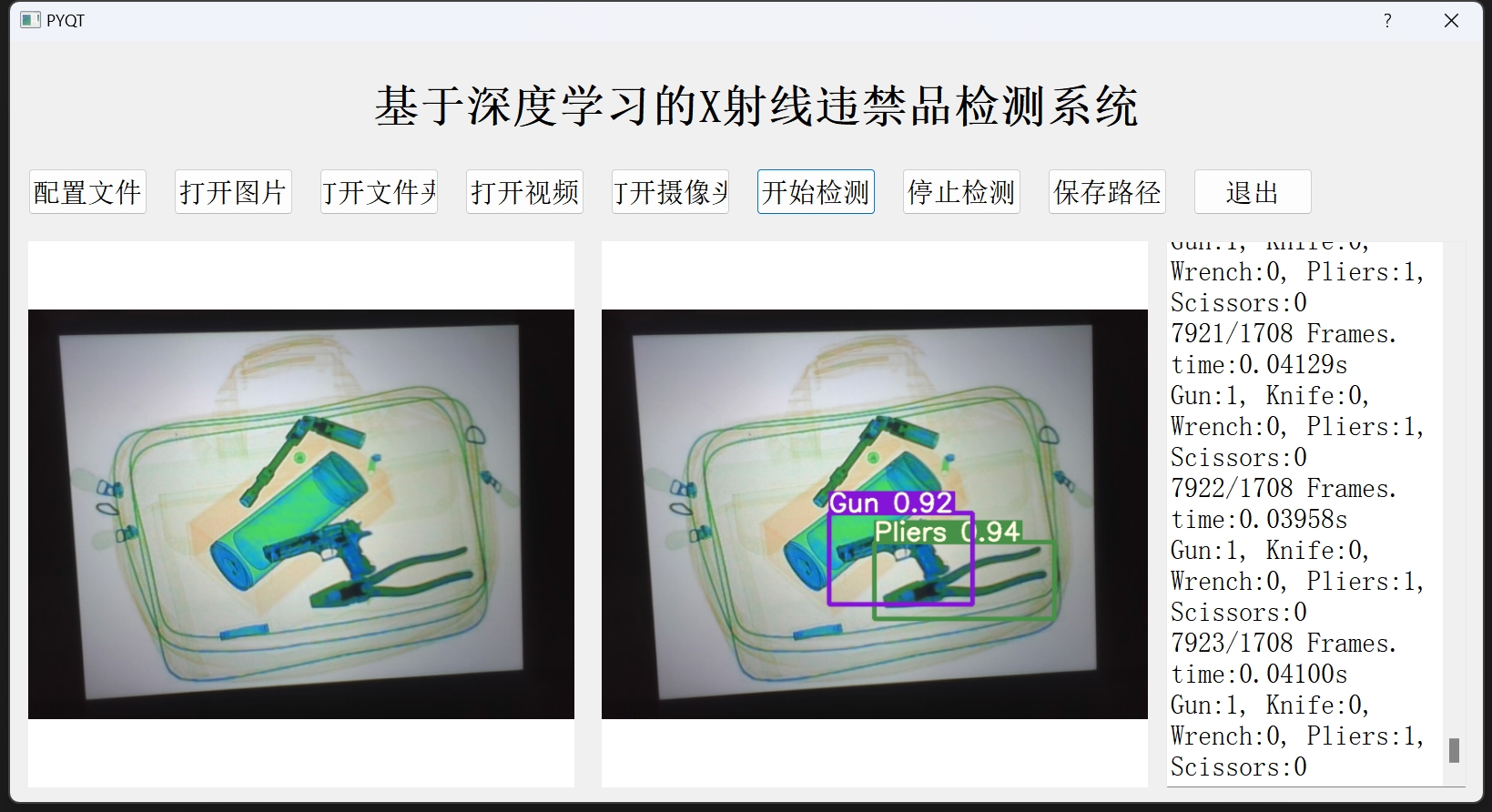


图5.5 摄像头检测效果图

系统还具备结果保存功能。在进行图片、文件夹、视频检测后，系统会将结果图片和视频保存到用户指定的地址，便于用户存档和后续分析。结果保存如图5.6所示。

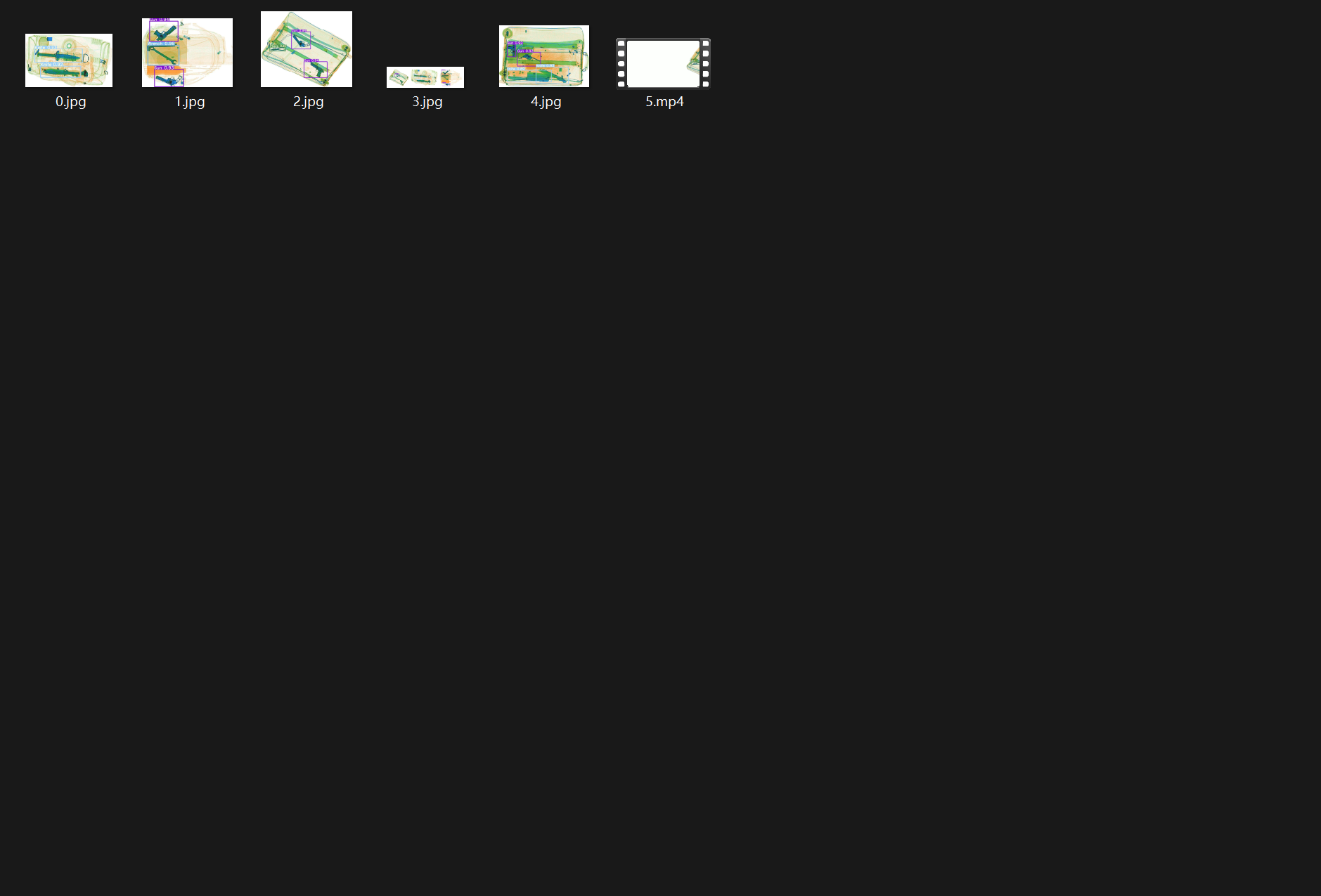


图5.6 结果保存

# 6 总结与展望

## 6.1 总结

随着我国经济的迅速发展，交通量急剧增加，公共安全问题日益凸显，安检需求愈发迫切。传统的人工检查方式存在漏检、错检的隐患。因此本文提供了一种基于深度学习的违禁品检测方案，现将研究内容总结如下：

首先，下载公开的违禁品数据集SIXray作为基准数据集对其进行预处理和数据增强后随机划分为训练集、验证集以及测试集，形成了共计5000张违禁品图像的数据集SIXray-s。

其次，以RT-DETR模型作为基准，对其Backbone、CCFF等部分进行进一步改进，使其更加轻量化。设计了模块对比实验、模型消融实验，验证了改进模块的优越性和有效性。同时对比了其他常见的目标检测模型，证实了本文模型在精度和效率上的卓越性能。

接着，通过混淆矩阵、F1分数等指标详细评估模型性能，结果表明改进模型在各类违禁品检测中均表现出色。

最后，本文使用PyQt5框架以改进后的RT-DETR模型为基础，搭建了一个本地的违禁品检测程序，用户可以在本地客户端运行该程序，便捷高效的进行违禁品检测。

## 6.2 展望

本文主要依赖于图像数据进行违禁品检测，由于时间和设备的局限性，本文仅使用共5000张图片的数据集，对常见的五类违禁品进行检测，可以考虑扩充更大、种类更丰富违禁品数据集。未来还可以和有关部门合作，在真实的安检场景中进行实际应用。

本文实现的违禁品检测系统旨在通过深度学习技术提升安检效率，未来可以代替传统的人工检查方式，为公共安全做出贡献。

# 参考文献

1. 王燕青,刘媛媛,徐晶. 基于生理—心理的机场安检人员疲劳风险评估[J]. 综合运输, 2016, 38(07): 49-55.
2. Zhao Yian,Lv Wenyu,Xu Shangliang, et al. DETRs Beat YOLOs on Real-time Object Detection[C]//IEEE: IEEE, 2024: 16965-16974.
3. 向丹.. 基于深度学习的安防图像目标识别技术研究[D]. 中国石油大学(北京), 2020
4. 李舒婷.. 基于全卷积网络的X光图像违禁物品检测[D]. 中国民航大学, 2021
5. Kumar R.-Senthil,A. Syed-Musthafa,Balaji A., et al. Recursive CNN Model to Detect Anomaly Detection in X-Ray Security Image[C]//IEEE: IEEE, 2022: 742-747.
6. Ren S , He K , Girshick R , et al.Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J].IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137-1149.DOI:10.1109/TPAMI.2016.2577031.
7. Xiaoyu Yu,Wenjun Yuan,Aili Wang. X-ray Security Inspection Image Dangerous Goods Detection Algorithm Based on Improved YOLOv4[J]. Mdpi Ag, 2023, 12(12): 2644.
8. Liu S , Qi L , Qin H , et al.Path Aggregation Network for Instance Segmentation[J].IEEE, 2018.DOI:10.1109/CVPR.2018.00913.
9. Miao Caijing,Xie Lingxi,Wan Fang, et al. SIXray: A Large-Scale Security Inspection X-Ray Benchmark for Prohibited Item Discovery in Overlapping Images[C]//IEEE: IEEE, 2019: 2114-2123.
10. Chen Zhaoyuan,Xu Yan,Li Fang, et al. Improved YOLOv8 Detection Algorithm of Contraband in X-Ray Security Inspection Image[C]//IEEE: IEEE, 2024: 1-6.
11. Vaswani A , Shazeer N , Parmar N , et al.Attention Is All You Need[J].arXiv, 2017.DOI:10.48550/arXiv.1706.03762.
12. Chen Jierun,Kao Shiu-hong,He Hao, et al. Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks[C]//IEEE: IEEE, 2023: 12021-12031.
13. Dalal N.,Triggs B.. Histograms of Oriented Gradients for Human Detection[C]//IEEE: IEEE: 886-893.
14. Wang Qiyu. Support Vector Machine Algorithm in Machine Learning[C]//IEEE: IEEE, 2022: 750-756.
15. Felzenszwalb P-F,Girshick R-B,Mcallester D, et al. Object Detection with Discriminatively Trained Part-Based Models[J]. Institute of Electrical and Electronics Engineers (IEEE), 2010, 32(9): 1627-1645.
16. Carion N , Massa F , Synnaeve G , et al.End-to-End Object Detection with Transformers[C]//2020.DOI:10.1007/978-3-030-58452-8\_13.
17. Dosovitskiy A , Beyer L , Kolesnikov A , et al.An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale[C]//International Conference on Learning Representations.2021.
18. Redmon J , Divvala S , Girshick R , et al.You Only Look Once: Unified, Real-Time Object Detection[C]//Computer Vision & Pattern Recognition.IEEE, 2016.DOI:10.1109/CVPR.2016.91.
19. He Kaiming,Zhang Xiangyu,Ren Shaoqing, et al. Deep Residual Learning for Image Recognition[C]//IEEE: IEEE, 2016.
20. Howard A G , Zhu M , Chen B , et al.MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications[J]. 2017.DOI:10.48550/arXiv.1704.04861.
21. Zhang Xiangyu,Zhou Xinyu,Lin Mengxiao, et al. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices[C]//IEEE: IEEE, 2018.
22. Han Kai,Wang Yunhe,Tian Qi, et al. GhostNet: More Features From Cheap Operations[C]//IEEE: IEEE, 2020: 1577-1586.
23. Chollet Francois. Xception: Deep Learning with Depthwise Separable Convolutions[C]//IEEE: IEEE, 2017: 1800-1807.
24. Krizhevsky A , Sutskever I , Hinton G .ImageNet Classification with Deep Convolutional Neural Networks[C]//NIPS.Curran Associates Inc. 2012.DOI:10.1145/3065386.
25. Tan Mingxing,Pang Ruoming,Le Quoc-V.. EfficientDet: Scalable and Efficient Object Detection[C]//IEEE: IEEE, 2020.
26. Wang Chien-Yao,Liao Hong-Yuan-Mark,Wu Yueh-Hua, et al. CSPNet: A New Backbone that can Enhance Learning Capability of CNN[C]//IEEE: IEEE, 2020.
27. Ding Xiaohan,Zhang Xiangyu,Ma Ningning, et al. RepVGG: Making VGG-style ConvNets Great Again[C]//IEEE: IEEE, 2021: 13728-13737.
28. Zhang X , Zeng H , Guo S , et al.Efficient Long-Range Attention Network for Image Super-resolution[J]. 2022.DOI:10.48550/arXiv.2203.06697.
29. Wang C Y , Yeh I H , Mark Liao H Y .YOLOv9: Learning What You Want toLearn Using Programmable Gradient Information[C]//European Conference on Computer Vision.Springer, Cham, 2025.DOI:10.1007/978-3-031-72751-1\_1.
30. Wan Cheng,Yu Hongyuan,Li Zhiqi, et al. Swift Parameter-free Attention Network for Efficient Super-Resolution[C]//IEEE: IEEE, 2024: 6246-6256.
31. Zhang Haoyang,Wang Ying,Dayoub Feras, et al. VarifocalNet: An IoU-aware Dense Object Detector[C]//IEEE: IEEE, 2021.
32. Ziping Yu,Hongbo Huang,Weijun Chen, et al. YOLO-FaceV2: A scale and occlusion aware face detector[J]. Elsevier Bv, 2024, 155: 110714.
33. Domingo Mery,Vladimir Riffo,Uwe Zscherpel, et al. GDXray: The Database of X-ray Images for Nondestructive Testing[J]. Springer Science and Business Media Llc, 2015, 34(4).
34. Wei Yanlu,Tao Renshuai,Wu Zhangjie, et al. Occluded Prohibited Items Detection: An X-ray Security Inspection Benchmark and De-occlusion Attention Module[C]//ACM, New York, NY, USA: ACM, 2020.
35. Liu Xinyu,Peng Houwen,Zheng Ningxin, et al. EfficientViT: Memory Efficient Vision Transformer with Cascaded Group Attention[C]//IEEE: IEEE, 2023: 14420-14430.
36. Howard Andrew,Sandler Mark,Chen Bo, et al. Searching for MobileNetV3[C]//IEEE: IEEE, 2019: 1314-1324.
37. Woo Sanghyun,Debnath Shoubhik,Hu Ronghang, et al. ConvNeXt V2: Co-designing and Scaling ConvNets with Masked Autoencoders[C]//IEEE: IEEE, 2023.
38. Qin D , Leichner C , Delakis M , et al.MobileNetV4 -- Universal Models for the Mobile Ecosystem[J]. 2024.
39. Ma Xu,Dai Xiyang,Bai Yue, et al. Rewrite the Stars[C]//IEEE: IEEE, 2024: 5694-5703.
40. Ding Xiaohan,Zhang Xiangyu,Han Jungong, et al. Diverse Branch Block: Building a Convolution as an Inception-like Unit[C]//IEEE: IEEE, 2021: 10881-10890.
41. Meng Depu,Chen Xiaokang,Fan Zejia, et al. Conditional DETR for Fast Training Convergence[C]//IEEE: IEEE, 2021: 3631-3640.
42. Zhu X , Su W , Lu L , et al.Deformable DETR: Deformable Transformers for End-to-End Object Detection[C]//International Conference on Learning Representations.2021.
43. Zhang H , Li F , Liu S , et al.DINO: DETR with Improved DeNoising Anchor Boxes for End-to-End Object Detection[J].arXiv e-prints, 2022.DOI:10.48550/arXiv.2203.03605.
44. Lyu C , Zhang W , Huang H , et al.RTMDet: An Empirical Study of Designing Real-Time Object Detectors[J]. 2022.DOI:10.48550/arXiv.2212.07784.
45. Chattopadhyay A , Sarkar A , Howlader P , et al.Grad-CAM++: Generalized Gradient-based Visual Explanations for Deep Convolutional Networks[J]. 2017.DOI:10.48550/arXiv.1710.11063.