面向安检违禁品检测的改进 RT-DETR 模型

张自勖[1] 罗 婷[1,2*] 董佳鑫[1] 马根炜[3,4]
ZHANG Zixu LUO Ting DONG Jiaxin MA Genwei

摘 要 通过 X 射线透视原理对行李和包裹扫描，已经成为公共场所安全检查广泛采用的方式，传统人工识别 X 射线图像中违禁品的方法工作量巨大且费时费力，文章以 RT-DETR 网络为基线模型，提出了一种改进的目标检测算法，以提高违禁品检测的准确率。首先，将 RT-DETR 的骨干网络替换为 RMT 轻量化特征提取主干，增强网络在复杂场景中精准定位能力，同时有效减少模型参数量；其次，在 AIFI 模块中引入 HiLo 注意力机制，该结构能够有效区分高频和低频特征，适应违禁品图像中较小细节和全局形状的处理，从而提高模型的检测准确性与效率；随后，在 CCFM 特征融合部分加入 CA 注意力机制，通过对特征通道进行加权，强化了关键特征的表示能力，提升了模型在处理复杂背景和模糊边缘时的表现。实验结果表明，所提网络在 PIDray 数据集中的 mAP@0.5 提升了 2.6%，尤其是遮挡测试集中的 mAP@0.5 提升了 7.1%，说明改进的 RMT 轻量化特征提取主干、HiLo 注意力以及 CA 模块成功实现了违禁品检测精度的提升，验证了所设计算法的优越性和鲁棒性，为未来的智能检测技术提供了新的思路。

关键词 X 射线安检；违禁品检测；轻量化主干网络；注意力机制；遮挡鲁棒性

doi：10.3969/j.issn.1672-9528.2025.12.006

0 引言

随着社会的不断进步和人们对公共安全的日益关注，安检成为维护社会稳定和人身安全不可或缺的环节。在众多安检技术中，X 射线安检因其高效、无损、非侵入的特点成为广泛应用的手段之一。X 射线安检技术在机场、车站等公共场所发挥着至关重要的作用[1]，通过对行李快速扫描，可以辅助安检人员识别出刀、枪、子弹、扳手、打火机等具有危害运输安全的违禁物品，确保了这些场所的安全。

但随着我国高铁、飞机等大型综合运输系统的广泛应用，长时间的高强度工作会使安检人员专注度降低，容易导致漏检与误检的情况发生[2]。在这种背景下，开发一种自动、准确且高效的违禁品检测算法，不仅可以提升安检效率，减少行李安检运输时在中转站的堆积，还能有效降低人为因素导致的检查失误，保障人身和财产安全。

早期的 X 射线违禁品检测主要依赖于传统机器学习方法实现目标的分类与识别。典型算法包括支持向量机[3]、随机森林[4]以及视觉词袋模型[5]等。这类方法普遍依赖人工设计的特征，难以充分挖掘数据中蕴含的深层信息，且其性能易受超参数设置的影响，在面对复杂背景干扰、多目标共存及小目标检测等挑战时，表现往往不尽如人意。

近年来，深度学习技术取得了突破性进展[6]，基于卷积神经网络的目标检测算法逐渐成为主流。其中，两阶段检测器如 Faster R-CNN[7] 通过生成候选区域并对每个区域进行精细分类与回归，实现了较高的检测精度；而单阶段检测器如 YOLO 系列则将检测任务视为回归问题，可以在图像中预测边界框与类别概率，凭借其单次前向传播即可完成预测的特性，大幅提升了检测速度。在安检违禁品检测这一特定领域，众多研究者基于 YOLO 架构进行了一系列改进：Tan 等[8]在 YOLOv5s 中引入可变形卷积与注意力机制，以增强对小目标的检测能力；左景等[9]针对 X 光图像中目标重叠、特征提取困难及背景复杂等问题，提出了结合多分支轻量化卷积与注意力机制的检测模型；董佳鑫等[10]则通过为 YOLOv8s 融入动态卷积、加权双向特征金字塔及全局注意力机制，强化了模型对不同尺度目标的特征提取能力；张良等[11]在 YOLOv5s 的主干网络中嵌入 Transformer 模块，利用其强大的全局建模能力弥补卷积操作的局部性缺陷，并设计感受野自适应融合模块以优化多尺度特征利用，从而缓解复杂背景

1. 中国人民公安大学信息网络安全学院 北京 100038
2. 中国人民公安大学公安大数据战略研究中心 北京 100038
3. 首都师范大学交叉科学研究院 北京 100048
4. 首都师范大学北京国家应用数学中心 北京 100048
[基金项目] 中央高校基本科研业务费专项资金资助（2025JKF01ZK02）

与目标尺度多变对检测性能的影响。

尽管 YOLO 系列模型在检测任务中表现出色，但其本质上仍受限于 CNN 架构。这些模型在实际部署时依赖非极大值抑制等后处理操作，难以实现完全端到端的实时检测。此外，CNN 结构更擅长捕获局部特征，而在建模长距离依赖于多尺度上下文信息方面存在不足，制约了其在复杂安检场景下的性能上限。

为了应对这些挑战，部分研究者开始尝试将 Transformer 结构引入目标检测领域，DETR（Detection Transformer）[12] 是 Carion 等于 2020 年提出的一种新的目标检测网络。DETR 将目标检测任务转变成一个序列到序列的问题，将通过 CNN 提取的特征图像作为 Transformer 中编码器的输入，通过自注意力机制捕捉特征图中目标的全局信息，再通过解码器将编码器的输出进行解码，预测生成图像中目标的类别和边界框。DETR 模型摒弃了传统检测流程中人工设定的锚点和 NMS 组件，采用二分匹配机制直接预测出对应的目标集合，简化检测流程。DETR 模型具备许多优点，但在实际应用过程中仍然面临模型训练收敛效率不高以及查询机制优化复杂等关键性挑战。针对这些问题，研究者们提出了多种改进方案。Deformable-DETR[13] 通过改进注意力机制的计算效率，加快了多尺度特征的训练收敛速度；Conditional DETR[14] 通过条件空间查询，使模型更快学习位置信息；Anchor DETR[15] 引入瞄点生成查询，减少随机初始化查询的优化难度；Group DETR[16] 采用分组一对多分配策，同样有助于模型效率的提升；结合 DN-DETR[17] 的去噪与 DAB-DETR[18] 的动态锚框，进一步优化训练稳定性。近期百度团队提出了 RT-DETR（real-time detection transformer）[19] 模型， RT-DETR 进一步优化了特征提取能力，保持了端到端的设计，同时去除了传统方法中非极大值抑制（NMS）的后处理步骤。这一改进不仅提升了检测的实时性，同时也增强了检测精度，使得 RT-DETR 在各种应用场景中表现更加出色。孙嘉傲 [20] 基于 RT-DETR 模型提出梯度流优化的重参数骨干网络，实现多分支训练与单分支推理的双优势结合，加速网络收敛提升违禁品特征提取性能，同时设计了池化混合解码器 HP-Decoder 聚合违禁品特征图向量的有效信息，精确捕获目标特征，以避免相似背景、色彩信息缺乏与违禁品重叠的影响。然而上述方法的模型参数量比较大，限制了应用发展，而且在检测遮挡违禁品时仍存在精度较低，推理速度较慢的问题。

基于上述问题，为了提升模型应对复杂场景的能力，缓解 X 光违禁品图像中目标重叠问题，本文以 RT-DETR 为基线模型，提出一种新的 X 射线违禁品检测模型。本文工作如下：

（1）以 RT-DETR 网络为基础，对其骨干网络进行优化，替换为轻量级特征提取主干（retentive networks meet vision transformers, RMT），减少了模型的参数量，增强了网络对图像中违禁品位置的精准定位能力。

（2）在尺度内特征交互模块（intra-scale feature interaction, AIFI）模块中引入了高低频注意力机制（HiLo），有助于模型对图像中的高频和低频特征的提取，在违禁品有遮挡或其他边缘不清晰的情况下仍能保持较高的检测准度。

（3）在跨尺度特征融合模块（cross-scale feature fusion module, CCFM）特征融合部分，加入了坐标注意力机制（coordinate attention, CA），实现对特征通道的加权，强化了模型对关键特征的表示能力，使得模型在复杂背景和模糊边缘处理时表现更为出色，提升了模型的检测性能。

1 RT-DETR 模型

RT-DETR 采用 ResNet 作为主干网络，从输入图像中提取多尺度特征。主干网络包括卷积层、最大池化以及由残差模块组成的深层特征提取模块，用于捕捉多尺度特征。该模型的编码器采用了一种精心设计的混合架构，核心由两个功能互补的模块构成：AIFI 模块负责在单一尺度上执行自注意力计算，专注于高级语义信息的提取并有效控制计算开销；CCFM 模块则负责跨尺度特征融合，通过卷积层与 RepC3 组件的协同工作，采用逐元素相加的方式整合不同层级的特征，确保融合过程中特征信息的完整性。在解码器设计上，RT-DETR 引入了一种基于交并比的查询选择策略。该机制通过量化编码器输出特征在分类置信度与定位精度上的联合不确定性，为解码器筛选出高质量的初始查询向量，显著提升了目标边界框的预测准确性。作为一款端到端检测器，RT-DETR 摒弃了传统检测流程中常见的非极大值抑制操作与预定义锚框机制，转而采用一种精简的后处理方案来保证最终预测框的质量。相较于 Faster R-CNN 和 YOLO 等需要复杂区域提议或后处理的架构，RT-DETR 简化了检测流程，显著减少了推理延迟，使其在对实时性要求严苛的 X 射线违禁品检测场景中具有显著优势。然而，将 RT-DETR 直接应用于安检场景仍面临特定挑战：一是其主干网络针对自然图像设计，对 X 射线图像独特的材质表征与纹理特征的提取能力有待加强；二是 AIFI 模块聚焦深层特征的策略，可能导致子弹等小尺寸违禁品在浅层细节信息利用不足，影响其检测精度。

为适应不同的计算资源与性能需求，RT-DETR 提供了多种规模的模型变体。基于 ResNet 的版本包括参数量与深度递增的 R18、R34、R50 和 R101 等；基于 HGNet 的版本则提供了 L 与 X 等更大规模的配置。其中，轻量化版本 RT-DETR-R18 通过网络深度的优化，在保持检测精度的同时进一

步提升了推理速度，非常适合计算资源有限的场景，因此本研究采用 R18 网络模型作为基准模型展开研究。

2 改进 RT-DETR 模块

为了更好地应对 X 射线违禁品检测任务的需求，本文选用 RT-DETR-R18 作为基础模型，将原始主干网络替换为 RMT 轻量级特征提取主干，减少模型的参数量，增强了网络对图像中违禁品位置的精准定位能力。同时，在 AIFI 模块中引入了高低频注意力机制帮助模型对 X 射线图像中违禁品边缘轮廓等高频信息的提取。在 CCFM 特征融合部分，加入了坐标注意力机制，实现对特征通道的加权，强化了模型对关键特征的表示能力。使得模型在复杂背景和违禁品遮挡问题处理时表现更为出色，提升模型的检测性能。改进的模型结构如图 1 所示。

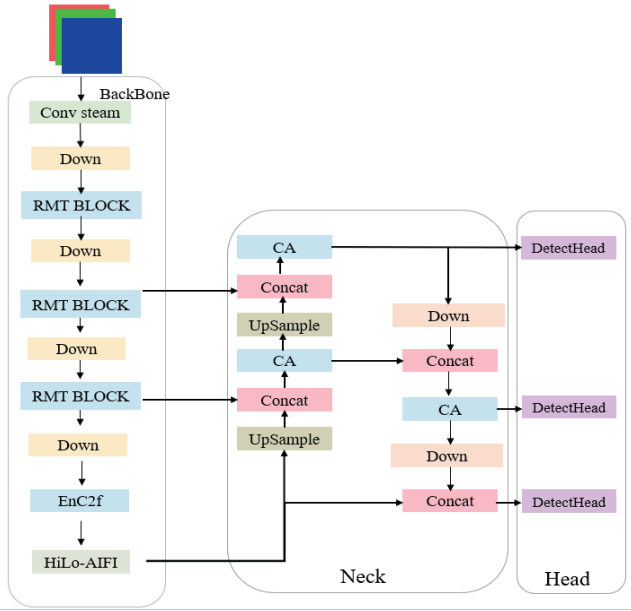


图 1 改进后的网络结构

2.1 RMT 特征提取主干

在 X 射线违禁品图像检测中，模型的准确性和效率至关重要。违禁品图像经常存在重叠物体和复杂轮廓等影响检测精度的问题，为了解决这一问题，增强算法对违禁品图像的特征提取能力，采用 RMT 特征提取主干改进 RT-DETR 模型的特征提取框架。

RMT 是一种将 RetNet 的核心思想迁移并适配于视觉任务的创新架构，巧妙融合了 RetNet 的序列建模优势与 Transformer 的全局交互能力。该模型的关键创新在于为视觉模型注入了与距离相关的空间先验约束。具体而言，RMT 定义了一个基于曼哈顿距离（manhattan self-attention, MaSA）的空间衰减矩阵，并据此构建了全新的曼哈顿自注意力模块。该模块通过度量图像中像素对之间的二维空间距离，自适应地调整注意力权重。为降低模型的计算复杂度，MaSA 模块被解耦为分别沿图像高度和宽度两个轴向的顺序计算过程。将复杂的二维全局注意力计算转化为两个一维操作的组合，在基本保留完整空间感受野的同时，显著降低了算法的总体计算复杂度。RMT 主干的整体结构如图 2 所示。

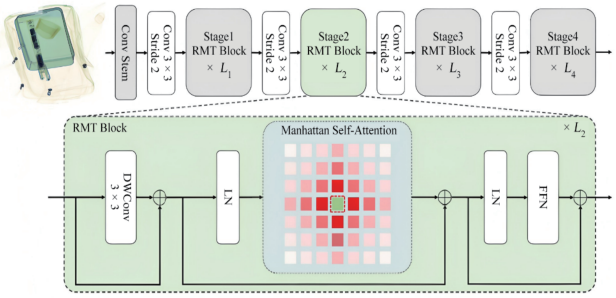


图 2 RMT 主干网络

其中 MaSA 模块的计算公式为：

$$\mathrm{MaSA}(\boldsymbol{X}) = (\mathrm{Softmax}(\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}) \odot \boldsymbol{D}^{2d})\boldsymbol{V} \tag{1}$$

$$\boldsymbol{D}^{2d}_{nm} = \gamma^{|x_n - x_m| + |y_n - y_m|} \tag{2}$$

式中：$\boldsymbol{X}$ 为输入序列；⊙为 Hadamard 积；$\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 分别为 Query、Key 和 Value 向量；x_n 和 y_n 为第 n 个 token 的二维空间坐标；γ 为预设的常数尺度因子；衰减矩阵$\boldsymbol{D}^{2d}_{nm}$的核心作用在于引入空间先验，其元素值由对应像素对之间的二维曼哈顿距离决定。为优化计算效率，MaSA 模块被设计为沿图像的水平和垂直轴进行分解，具体计算公式为：

$$\mathrm{Attn}_H = \mathrm{Softmax}(\boldsymbol{Q}_H \boldsymbol{K}_H^{\mathrm{T}}) \odot \boldsymbol{D}^H \tag{3}$$

$$\mathrm{Attn}_W = \mathrm{Softmax}(\boldsymbol{Q}_W K_W^{\mathrm{T}}) \odot \boldsymbol{D}^W \tag{4}$$

$$\mathrm{MaSA}(\boldsymbol{X}) = \mathrm{Attn}_H(\mathrm{Attn}_W \boldsymbol{V})^{\mathrm{T}} \tag{5}$$

式中：Attn_H 和 Attn_W 分别代表行和列方向上的注意力机制。具体来说，分别计算图像水平和垂直方向的注意力分数。然后，将一维双向衰减矩阵应用于这些注意力权重。

基于 MaSA 的分解，每个 token 的感受域形状如图 3 所示，与完整 MaSA 的感受域形状相同。

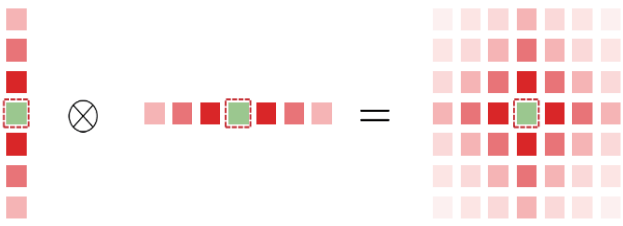


图 3 卷积核形状

如图 3 所示，上述分解策略有效保留了模型中的显式空间先验。在本研究中，采用 RMT-T 架构作为特征提取网络，该模型共包含 4 个阶段，各阶段所堆叠的 RMTBlock 数量依

次为 2、2、8 和 2。

2.2 AIFI-HiLo 模块

内尺度特征交互模块（AIFI）的构建基础是经典 Transformer 编码器中的多头自注意力（MSA）机制。该机制通过并行地在线性变换后计算多个注意力头，使模型能够从不同的表示子空间捕获多样化的特征依赖关系，从而显著增强了模型的表达能力和并行处理效率。然而，标准的 MSA 在处理图像时，倾向于对所有图像块施以同等权重的全局注意力，这种平等处理模式未能区分特征在频率上的固有差异，致使模型在面对高分辨率图像时承受巨大的计算负荷。

在违禁品检测中，图像往往展现出高度的复杂性和多样性，包含多种颜色、形状和纹理特征。这些特征在高分辨率图像中尤为明显，但由于存在不同的基础频率，使得标准的 MSA 对其处理不够有效。因此，为了解决这个问题，AIFI 模块引入了 HiLo 注意力结构如图 4 所示。该结构有效区分高频和低频特征，针对违禁品图像中较小的细节和全局形状进行相应的处理，从而保持模型效率，提高检测准确性。

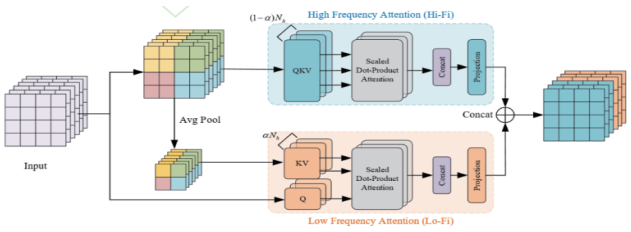


图 4 HiLo 高低频分离机制

HiLo注意力将多头自注意力分为高频注意力部分（Hi-Fi）和低频注意力部分（Lo-Fi）。高频注意力通过局部自注意力和高分辨率特征图编码高频交互，关注图像中的细节如边缘与纹理，对于识别违禁品的细微变化至关重要，该路径通过限制感受野，避免长程依赖计算，从而实现了高效的特征转换。低频注意力则通过全局注意力与下采样特征来处理低频交互，旨在减轻全局自注意力的计算负担，以提高编码效率。

高频注意力专注于精细特征的捕获，利用 2×2 的窗口自注意力机制，通过非重叠窗口的分区，显著降低计算复杂性。低频部分在每个窗口上采用平均池化获取低频信息，再利用全局注意力捕获丰富的低频特征，如形状和颜色分布，这些特征在识别违禁品时也非常重要。最终，来自高频与低频路径的特征图被有效融合，形成 AIFI-HiLo 模块的最终输出。得益于这种将复杂计算分解为两个异构分支的策略，每个分支的计算开销都显著低于标准的全局多头自注意力。因此，该框架不仅整体计算复杂度更低，能够充分利用 GPU 的并行计算能力以实现高吞吐量，还通过互补的频率感知机制，增强了对不同属性目标特征的捕捉能力。

2.3 坐标注意力机制

本文在特征融合模块（CCFM）中引入坐标注意力机制。对比多头自注意力机制，在进行全局依赖捕获过程中需要计算全局空间范围注意力权重矩阵，涉及查询 $\boldsymbol{Q}$、键 $\boldsymbol{K}$ 和值 $\boldsymbol{V}$ 的多次矩阵乘法操作。其计算复杂度与输入特征图空间尺寸呈平方级别增长。CA 注意力通过一维全局池化将特征图空间信息分解为垂直和水平方向，分别计算注意力权重，将复杂度从 $O(N^2)$ 降低至 $O(2N)$，显著减少了计算开销。

坐标注意力机制实现局部细节保留与全局上下文建模完美融合。在一维池化过程中，垂直方向池化核沿列方向时聚合全局特征，并保留水平方向位置信息。同理，水平方向池化核沿行方向实现全局特征聚合，保留垂直方向位置信息。CA 在未被池化方向有效地保留空间位置信息，精准捕获长距离依赖关系，建模目标的空间结构与位置信息，提升目标定位精度。对比多头自注意力机制，实现所有位置加权聚合，仅关注全局依赖关系，导致模型无法有效保留空间方向上显著位置信息，影响目标精确定位。此外，多头自注意力机制注意力基于全局，容易忽略局部细节信息，特别是在小目标或密集目标场景中，极易导致信息丢失。

CA 坐标注意力机制的结构设计如图 5 所示。具体而言，对于给定的输入特征图，首先在垂直方向应用宽度为 1 的池化核进行编码，以捕获特征图在高度方向上的长程依赖；同时，在水平方向应用高度为 1 的池化核进行编码，以捕获宽度方向上的上下文信息。通过此方式，可获得第 c 个通道在高度 h 处的全局特征表示，其计算过程如式（4）所示。其中给定输入 $\boldsymbol{X}\in\mathbb{R}^{H\times W\times C}$，水平方向池化核为 $(1, W)$、垂直方向池化核为 $(H, 1)$ 分别沿垂直坐标和水平坐标依次对各通道编码，第 c 个通道在高度 h 处的输出为：

$$Z_c^h(h)=\frac{1}{W}\sum_{0\le i\le W}x_c(h,i) \tag{6}$$

第 c 个通道在宽度为 w 处的输出为：

$$Z_c^h(w)=\frac{1}{H}\sum_{0\le j\le H}x_c(j,w) \tag{7}$$

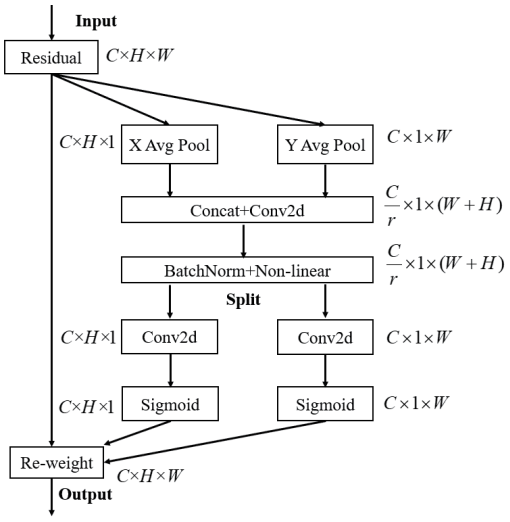


图 5 CA 坐标自注意力机制网络结构

针对宽度和高度两个方向上获得全局感受域特征图进行拼接，并通过共享 1×1 卷积模块对特征进行处理，以将通道维度降至原来的 C/r。将经过批量归一化处理的特征图 F_1 送入 Sigmoid 激活函数，继而得到形如 $\frac{C}{r}\times 1\times(W+H)$ 尺寸的特征图 $\boldsymbol{f}$。对应公式为：

$$\boldsymbol{f}=\delta\left(F_1\left(\left[z^h,z^w\right]\right)\right) \tag{8}$$

沿着空间维度对特征图 $\boldsymbol{f}$ 进行分解，划分为 $\boldsymbol{f}^h\in\mathbb{R}^{C/r\times H\times 1}$ 和 $\boldsymbol{f}^w\in\mathbb{R}^{C/r\times 1\times W}$，分别用 1×1 卷积核沿空间维度进行升维操作，并结合 sigmoid 激活函数得到最后的注意力向量 $\boldsymbol{g}^h\in\mathbb{R}^{C\times H\times 1}$ 与 $\boldsymbol{g}^w\in\mathbb{R}^{C\times 1\times W}$。最后在原始特征图上通过乘法加权计算，得到宽度和高度方向上带有注意力权重的特征图。

$$y_c(i,j)=x_c(i,j)\times\boldsymbol{g}_c^h(i)\times\boldsymbol{g}_c^w(j) \tag{9}$$

在违禁品检测任务中，考虑到 X 射线违禁品图像的特点，将坐标注意力模块集成到 CCFM（channel compound feature mapping）的上采样模块中，从而显著提升模型的性能。

X 射线图像通常具有图像中的物体对比度较低，且可能存在多种重叠物体的特点。其次，由于不同材料在 X 射线下的透过率不同，导致图像中物体的形态和边缘往往不清晰。此外，违禁品往往呈现出复杂的形状与不同的遮挡形式，给检测带来了额外的困难。CA 注意力模块通过对特征通道的加权，有效强化了物品边缘轮廓等关键特征的表示，提升了模型在处理复杂背景和模糊边缘时的能力。CA 机制能够捕捉到物体的几何结构和位置关系，从而更好地识别和分类违禁品的形态。此外，该注意力机制有效滤除了冗余信息，保证了检测系统在特征提取时的高效性。整合 CA 模块后的 RT-DETR 模型在违禁品检测任务中提高了检测精度，还增强了复杂场景下物体的识别能力，使得检测系统在实际应用中更加可靠和准确。

3 实验与结果分析

3.1 实验配置与训练策略

实验采用 PyTorch1.7 深度学习框架，编程语言为 Python3.8，在 Ubuntu22.04 系统上运行，GPU 为 NVIDIA GTX 3090，服务器内存为 32 GB，显存为 24 GB，输入图片大小为 640 px×640 px，初始学习率为 0.01，循环学习率为 0.01，权重衰减系数为 0.001，Batchsize 设置为 16，Epoch 设置为 200。

3.2 评价指标

为系统评估所提模型的综合性能，本研究综合采用精确率、召回率、mAP@0.5、模型参数量及单张图像推理时间等多项指标进行全面衡量。各指标的具体计算公式为：

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}} \tag{10}$$

$$\text{Recall}=\frac{\text{TP}}{\text{TP}+\text{FN}} \tag{11}$$

式中：TP 表示正确预测的正例数量；FP 表示被错误预测为正例的负例数量；FN 表示被错误判断为负例的正例数量。

F_1 分数是用于评估模型整体有效性的关键，其基于精确率和召回率调和平均加以计算。F_1 分数计算过程为：

$$F_1\text{-score}=2\cdot\frac{\text{Precision}\cdot\text{Recall}}{\text{Precision}+\text{Recall}} \tag{12}$$

模型检测性能的评估首先依赖于平均精度 AP，它通过计算 Precision-Recall 曲线（其中 Recall 为 X 轴，Precision 为 Y 轴）下的面积来量化模型对单一类别的识别精度。为评估模型的整体表现，引入 mAP 指标，即对所有类别的 AP 值进行平均，能综合反映网络在不同类别上的泛化性能。AP 与 mAP 的计算公式为：

$$\text{AP}=\int_0^1 P(R)\mathrm{d}R \tag{13}$$

$$\text{mAP}=\frac{1}{N}\sum_{i=1}^{N}\int_0^1 P\mathrm{d}R \tag{14}$$

作为衡量模型复杂度一项重要指标，参数量 Params 代表模型学习参数规模。参数规模的扩充虽可潜在提升模型的表达能力，但也相应增大了其计算与存储成本。理想情况下，应设计紧凑高效的轻量化模型，使其在参数量显著减少的同时，依然能够维持卓越的检测性能。

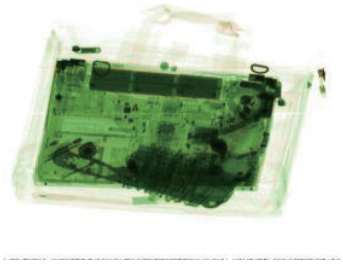
3.3 数据集介绍

为了提升违禁品检测方法的性能，一些研究人员通过数据增强和标注优化等方式构建了适用于违禁品检测的数据集。GDXray 公共数据集 [21] 包含 3 种违禁物品：枪、飞镖和剃须刀片。但由于几乎没有复杂的背景和重叠，很容易识别或检测该数据集中的目标。与 GDXray 相比，OPIXray[22] 的背景复杂、数据重叠，但图像数量和禁用项数量仍然不足。SIXray[23] 的大规模安全检查标准中，包含 1 059 231 张带有图像级注释的 X 射线图像，但是数据集中包含违禁物品的图像较少（只有 8 929 张照片，约 0.84%）。此外，该数据集包含 6 个类别的违禁物品：枪、刀、扳手、钳子、剪刀、锤子，其中锤子未被标注。

PIDray 数据集 [24] 中含有 47 677 张图片和 12 类违禁物品，包含枪、刀、扳手、钳子、剪刀、锤子、手铐、警棍、喷雾器、充电宝、打火机和子弹，是现有最大的 X 射线违禁物品检测数据集。该数据集还针对违禁品检测环境的复杂程度进行了更细的划分，包括 Easy、Hard 和 Hidden，分别代表测试集中的图像只包含一个违禁品、图像同时包含多个违禁品和图

像包含故意隐藏的违禁品，同实际场景十分相似。实验时将数据集分成两部分：29 457 张图像用于训练，按照 4:1 划分训练集和验证集；剩余 18 220 张图像为实验测试集。图 6 展示部分样本实例。

（1）Easy　　（2）Hard



（3）Hidden

图 6 数据集样例

3.4 对比实验

为验证模型先进性，本文在 PIDray 数据集 Easy、Hard、Hidden 三种类别下进行测试，同时与当前主流目标检测模型 Faster R-CNN、RT-DETR、YOLOv8n、YOLOv7[25]、YOLOv9c[26] 等进行了对比分析，实验结果如表 1 所示。

表 1 PIDray 数据集对比实验

Models	AP%			Params/10^6	FPS/(帧·s^{-1})
	Easy	Hard	Hidden		
FasterR-CNN	67.1	60.2	48.2	28.3	29
YOLOv7	77.4	75.2	55.7	103.3	10
YOLOv8n	83.5	80.7	60.3	8.1	26
YOLO-MS	80.6	78.9	57.7	8.24	24
YOLOv9c	80.1	78.6	58.0	101.7	13
YOLOv10n	84.8	81.9	61.3	8.4	30
RT-DETR	85.3	81.2	62.7	42.8	47
Our	**86.9**	**83.0**	**69.8**	**40.7**	**48**

结果表明，所提改进网络在 PIDray 数据集上实现了最佳的检测性能，其 mAP 比最先进的 RT-DETR 实时检测模型提升了 2.6%。与现有的单阶段违禁品检测网络，如 YOLO 系列模型相比，所提网络在较少的训练迭代次数中展现出优异的检测表现。尤其针对 PIDray 数据集中容器类违禁品所面临的背景相似性和色彩信息单一的问题，RT-DETR 已经超越了 YOLOv8 检测效果，而所提网络则进一步提升了检测效果。

此外，在 PIDray 数据集中，存在的违禁品重叠和遮挡问题的实验结果表明，所提网络在 PIDray 遮挡测试集中的 mAP 提升达 7.1%，超过了总体的提升水平 3%，充分说明 DETR 系列在遮挡重叠违禁品检测中的显著改善。

值得一提的是，在确保神经网络容量充足且特征提取能力较强的前提下，数据量越丰富、数据质量越高（不包含过多噪点和相似数据），检测效果越佳，且在大型数据集上的预训练模型权重能够很好地迁移到其他数据集，反映出模型的真实性能。PIDray 数据集的规模大、数据分布全面，几乎涵盖了日常安检情况，使得训练结果能够更真实地反映模型的泛化性与总体检测效能。改进模型在仅以 40.7×10^6 的参数量下取得较优的检测结果，并且推理速度优于基线模型，以上结果充分证明了所提网络在违禁品检测任务中具备高效和卓越的性能。

3.5 消融实验

为验证所提方法有效性并探究 RMT 特征提取主干，AIFI-HiLo 模块以及添加坐标注意力机制下的跨尺度特征融合模块对模型准确率和效率的影响，本文在 PIDray 数据集上进行消融实验，实验结果如表 2 所示。

表 2 PIDray 数据集上的消融实验研究

RMT	HiLo	CA	mAP@0.5/%	Params/10^6
			62.7	42.8
√			63.6	40.5
	√		68.7	42.9
		√	65.6	43.0
√	√	√	69.8	40.7

实验结果表明，采用 RMT 作为特征提取主干，可以提供有效的特征表示，但仅依靠该模块的 mAP 仅为 63.6，参数量为 40.5×10^6。引入 AIFI-HiLo 模块后，模型的 mAP 显著提升至 68.7%，这一改进的主要原因在于 AIFI-HiLo 模块通过增强信息融合和特征自适应能力，成功提升了模型在复杂场景中的检测能力。当原模型结合 CA 模块时，模型表现出更优越的性能，mAP 达到了 65.6%，显示出 CA 模块通过捕捉通道间的语义信息，有效提高了特征的表达力。尽管在不加入 AIFI-HiLo 的情况下，CA 模块仍能带来一定提升。在同时加入 RMT，AIFI-HiLo，CA 后，mAP 达到最高的 69.8%，参数量保持在 40.7×10^6，验证了 3 个模块的良好互补性以及对于提升目标检测性能的关键作用 3 个模块的结合不仅增强了模型的特征提取能力，也提高了对违禁品检测中复

杂场景和遮挡情况的适应。

3.6 可视化分析

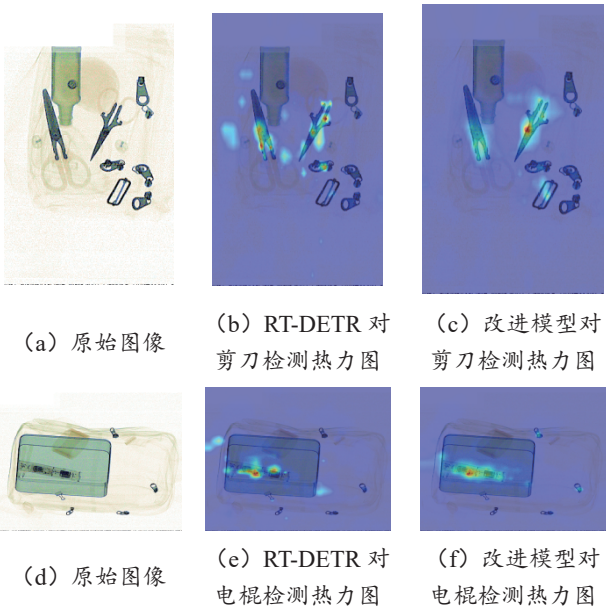
为验证所提出模型的先进性，实验在 PIDray 数据集下的部分违禁品实例的检测效果如图 7 所示。可以看出，改进模型可以准确地识别出 X 光下行李物品中的违禁品，对违禁品的位置和类别进行标注的同时给出置信度。

（a）PIDray 数据集违禁品标注位置

（b）改进模型在 PIDray 数据集上的检测效果

图 7 违禁品检测效果图

为了更加直观地评估改进后模型的检测效果，采用 Grad-CAM 可视化方法分析模型特征图在预测中的作用，通过热力图揭示模型对违禁品图像不同区域的关注程度，从而提升了模型的可解释性。如图 8 所示，选取剪刀、电棍两种常见的违禁品图像，图（a）和图（d）为原始违禁品图像，图（b）和图（e）为基线模型 RT-DETR 的检测效果，图（c）和图（f）为改进后模型的检测效果。可以观察到，本文提出的方法在违禁品目标捕捉方面表现出明显的优势。这种优势主要源自我们对特征提取和特征融合能力的改进。相比于 RT-DETR 模型，本文方法的热力图展示出更加集中和明确的激活区域，表现为颜色更亮的区域，不仅表明模型的关注点更加聚焦于违禁品目标位置，而且这些激活区域与真实目标的空间范围高度吻合，精准地贴合违禁品物品的形状，说明本文模型对违禁目标边缘特征的捕捉更加敏锐，能够更好地识别出局部细微的特征差异。此外，在检测过程中，模型能够有效忽略背景干扰等其他非目标区域，很好地避免了漏检与误检情况的发生。而从 RT-DETR 模型的热力特征图中可以看出，由于判断不够准确，其激活区域显得相对分散和广泛，包含了大量的非目标区域，这不仅降低了检测的准确性，还导致模型在多目标、遮挡复杂环境中的表现不够理想。通过对比分析，本文提出的模型在特征聚焦、精确性和可靠性等方面均具备明显的优势，在实际应用中对违禁品检测具有优越的性能。



（a）原始图像　（b）RT-DETR 对剪刀检测热力图　（c）改进模型对剪刀检测热力图

（d）原始图像　（e）RT-DETR 对电棍检测热力图　（f）改进模型对电棍检测热力图

图 8 改进模型与 RT-DETR 在 PIDray 数据集上的特征可视化对比

4 结论与讨论

本文以 RT-DETR 网络为基线模型，首先将 RT-DETR 的骨干网络替换为 RMT 轻量级特征提取主干，有效减少了模型的参数量，并且有助于网络对图像中违禁品位置精准定位。其次，在 AIFI 模块引入了 HiLo 注意力结构。该结构有效区分高频和低频特征，针对违禁品图像中较小的细节和全局形状进行相应的处理，从而保持模型效率，提高检测准确性。之后，在 CCFM 特征融合部分加入 CA 注意力机制，通过对特征通道的加权，有效强化了关键特征的表示，提升了模型在处理复杂背景和模糊边缘时的能力。由于替换的 RMT 是轻量化模块，并且 AIFI-HiLo 相较于 AIFI 参数量基本维持不变，最终维持了违禁品测出率与模型轻量化的平衡。所提改进后网络在 PIDray 数据集上实现了最佳的检测性能，其 mAP 比最先进的 RT-DETR 实时检测模型提升了 2.6%，在 PIDray 遮挡测试集中的 mAP 提升达 7.1%，超过了总体的提升水平 3%，在遮挡重叠违禁品检测中的显著改善表现明显好于其他主流算法，验证了所设计算法的优越性和鲁棒性。未来的研究将转向多视角 X 射线三维重建技术，进一步提升

对隐蔽违禁品的识别能力，同时探索轻量化模型在边缘计算设备上的实时部署应用，推动智能安检技术向更高效、更精准的方向发展。

参考文献：

[1] WELLS K, BRADLEY D A. A review of X-ray explosives detection techniques for checked baggage[J].Applied radiation and isotopes,2012,70(8) :1729-1746.

[2] 梁添汾 , 张南峰 , 张艳喜 , 等 . 违禁品 X 光图像检测技术应用研究进展综述 [J]. 计算机工程与应用 , 2021, 57(16): 74-82.

[3] BASTAN M, YOUSEFI M R, BREUEL T M. Visual words on baggage X-ray images[C]//Computer Analysis of Images and Patterns. Berlin:Springer, 2011: 360-368.

[4] 王宇 , 邹文辉 , 杨晓敏 , 等 . 基于计算机视觉的 X 射线图像异物分类研究 [J]. 液晶与显示 ,2017,32(4):287-293.

[5] TURCSANY D, MOUTON A, BRECKON T P. Improving feature-based object recognition for X-ray baggage security screening using primed visualwords[C]//2013 IEEE International conference on industrial technology (ICIT), Piscataway: IEEE, 2013: 1140-1145.

[6] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2015: 1-9.

[7] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection withregion proposal networks [J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 39(6): 1137-1149.

[8] TAN M X, PANG R M, LE Q V. Efficientdet: scalable and efficient object detection [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Piscataway: IEEE, 2020: 10781-10790.

[9] 左景 , 石洋宇 , 卢树华 . 基于轻量化卷积和 SCAM 改进的 X 光违禁品检测 [J]. 计算机科学与探索 , 2025, 19(6): 1598-1610.

[10] 董佳鑫 , 罗婷 , 李根 , 等 . 基于改进 YOLOv8s 的 X 射线安检图像违禁品检测方法 [J]. 激光与光电子学进展 , 2024, 61(22): 268-279.

[11] 张良 , 薛志诚 . 基于自适应多尺度特征融合的 X 光违禁品检测 [J]. 信号处理 , 2024, 40(4): 789-800.

[12] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]//Computer Vision – ECCV 2020. Berlin: Springer, 2020:213-229.

[13]ZHU X Z, SU W J, LU L W, et al. Deformable DETR: deformable transformers for end-to-end object detection[EB/OL].(2021-03-18)[2025-07-01].https://doi.org/10.48550/arXiv.2010.04159.

[14]MENG D P, CHEN X K, FAN Z J, et al. Conditional DETR for fast training convergence[C]//Proceedings of the IEEE/CVF Inter-national Conference on Computer Vision. Piscataway: IEEE, 2021: 3631-3640.

[15]WANG Y M, ZHANG X Y, YANG T, et al. Anchor DETR: query design for transformer-based detector[C]//Proceedings of theAAAI Conference on Artificial Intelligence.Palo Alto: AAAI Press, 2022: 2567-2575.

[16]CHEN Q, CHEN X K, ZENG G, et al. Group DETR: fast training convergence with decoupled one-to-many label assignment[EB/OL].2022[2025-07-01].https://www.semanticscholar.org/paper/Group-DETR:-Fast-Training-Convergence-with-Label-Chen-Chen/0c85e977a35d1e49e9a54e4f1047d69f21f71408.

[17]LI F, ZHANG H, LIU S L, et al. DN-DETR: accelerate DETR training by introducing query denoising[J].IEEE transactions on pattern analysis and machine intelligence,2023,46(4): 2239-2251.

[18]LIU S L, LI F, ZHANG H, et al. DAB-DETR: dynamic anchor boxes are better queries for DETR[EB/OL]. (2022-03-30) [2024-07-01].https://doi.org/10.48550/arXiv.2201.12329.

[19] ZHAO Y A, LÜ W Y, XU S L, et al. DETRs beat YOLOs on real-time object detection[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Piscataway: IEEE, 2024: 16965-16974.

[20] 孙嘉傲 . 基于改进 YOLO 与 DETR 的 X 光违禁品检测研究 [D]. 北京 : 中国人民公安大学 ,2024.

[21]MERY D, RIFFO V, ZSCHERPEL U, et al. GDXray: the database of X-ray images for nondestructive testing[J].Journal of nondestructive evaluation, 2015, 34(4): 42.

[22] WEI Y L, TAO R S, WU Z J, et al. Occluded prohibited items detection: an X-ray security inspection benchmarkand de-occlusion attention module[C]//Proceedings of the 28th ACM International Conference on Multimedia,NewYork: ACM, 2020: 138-146.

[23] MIAO C J, XIE L X, WAN F, et al. SiXray: a large-scale security inspection X-ray benchmark for prohibited item discovery in overlapping images[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2019: 2119-2128.

基于 LightGBM 的酒驾换驾风险预测

赵瑞阳[1] 刘聪慧[1]
ZHAO Ruiyang LIU Conghui

摘 要 在车险理赔场景中，酒驾换驾案件因肇事者虚报出险时间、伪造现场等行为，导致调查效率低且损失较大，亟需科学的风险预测模型以提升管控效能。针对此问题，文章提出一种基于机器学习的酒驾换驾风险预测模型。研究通过分析 2024 年 4 万多条已决赔案数据（含 172 个字段），将风险识别转化为二分类任务（0= 无风险，1= 有风险）。针对正负样本分布极不平衡（约 345:1）的挑战，通过特征工程构建了“报案 - 出险时间间隔”“驾驶员与被保险人身份一致性”关键衍生特征，并创新性地利用朴素贝叶斯算法预测驾驶员性别，以补全关键信息维度。为确保模型鲁棒性，采用了数据标准化、5 折交叉验证及早停策略进行优化。在 GBDT、XGBOOST、LightGBM 三种集成学习模型的对比实验中，LightGBM 表现最佳，其准确率（Accuracy）为 84.56%，精确率（Precision）为 94.73%，召回率（Recall）为 89.82%，F_1 分数（F_1-score）为 0.922 0，显著优于对比模型及传统人工筛查方法。该模型可有效识别高风险案件，为车险理赔环节降低调查成本、提升风险管理效率提供数据驱动的解决方案，对打击保险欺诈具有重要的现实意义和应用价值。

关键词 酒驾换驾；风险预测；机器学习；特征工程；LightGBM；理赔管控；数据挖掘；模型评估

doi：10.3969/j.issn.1672-9528.2025.12.007

0 引言

酒驾换驾是指肇事者在酒后驾驶车辆发生交通事故后，为逃避法律责任和保险拒赔，通过更换驾驶员、虚报出险时间、伪造事故现场等方式进行保险欺诈的行为[1]。这类行为在车险理赔场景中属于典型的保险欺诈范畴，不仅严重违反道路交通法规，还会导致保险公司面临巨大的理赔损失与调查压力。从行业现状来看，酒驾换驾案件具有“损失金额大、调查时效要求高、现场勘查前难以察觉”的显著特点，如某分公司数据显示，2024 年其受理的车险理赔案件中，共发现 204 起酒驾换驾疑点案件，其中仅 5 起最终因证据确凿成功拒赔，涉及拒赔金额达 6.8 万元。

1. 中国太平洋财产保险股份有限公司营运山东分中心
山东济南 250000

[24] WANG B Y, ZHANG L B, WEN L Y, et al. Towards real world prohibited item detection: a large-scale X-ray benchmark[C] // 2021 IEEE/CVF International Conferenceon Computer Vision (ICCV). Piscataway:IEEE,2021:5392-5401.

[25] WEN L H, JO K H. Three-attention mechanisms for one-stage 3-D object detection based on LiDAR and camera[J].IEEE transactions on industrial informatics,2021, 17(10):6655-6663.

[26]WANG C Y, YEH I H, MARK LIAO H Y. YOLOv9:Learning what you want to learn using programmable gradient information[C]//European conference oncomputer vision. Berlin:Springer, 2025: 1-21.

【作者简介】

张自勖（2000—），男，山东济南人，硕士研究生，研究方向：计算机视觉。

罗婷（1989—），通信作者（email:luoting@ppsuc.edu.cn），女，山东聊城人，博士研究生，讲师、硕士生导师，研究方向：X 射线 CT 成像与应用、人工智能等。

董佳鑫（1999—），男，安徽涡阳人，硕士研究生，研究方向：计算机视觉。

马根炜（1992—），男，江西吉安人，博士研究生，研究方向：CT 理论与成像技术、深度学习人工智能等。

（收稿日期：2025-10-10 修回日期：2025-12-12）