

基于混合注意力和任务对齐的违禁品检测

朱 鹏¹

ZHU Peng

摘 要

为解决 X 光安检中物品检测精度较低、易误检和漏检的问题,提出一种基于混合注意力和任务对齐的多尺度特征融合检测网络——TAFNet。该算法在 TOOD 模型基础上进行了改进。首先,在主干网络中引入混合注意力模块,将坐标、通道和空间三种注意力机制的信息融合,提升网络的特征表达能力;其次,在网络尾部引入 SPPCSPC 模块,以提取多尺度特征并为遮挡区域提供全局上下文信息;最后,设计多尺度特征融合金字塔,对高维语义信息进行解码,并与低维特征进行多尺度融合,生成丰富语义和细节信息的特征图,从而提升网络的尺度感知能力。在 SIXray 数据集上进行实验,结果表明,该方法的平均检测精度分别达到了 92.59%,较原始算法提高了 3.99%,并优于多种现有先进算法,为 X 光违禁物品检测提供了一种高效且可靠的解决方案。

关键词

违禁物品检测;注意力机制;任务对齐;特征解码;多尺度特征融合

doi: 10.3969/j.issn.1672-9528.2025.05.011

0 引言

鉴于航空、铁路及城市轨道交通对于运输安全的高度需求,X 光行李安全检查系统已成为不可或缺的防线,显著提升了公共交通安全防护水平。

人工检查的方式无法满足对检查速度和精度的要求。为进一步提高检测精度和效率,众多学者使用深度学习的方式检测 X 光图像中的违禁物品。李晨等人^[1]提出特征增强融合策略,使用坐标注意力机制(coordinate attention, CA)^[2]增强主干网络特征提取能力并降低背景噪声干扰。游玺等人^[3]提出融合空间注意力的自适应检测方法,使用自适应注意力模块抑制复杂背景干扰,并提出多尺度自适应候选区生成网络,来提高网络的召回率。Wang 等人^[4]提出材料感知跨通道交互关注模块(MCIA),利用 X 光图形的材料信息来处理违禁物品和非违禁物品之间的遮挡问题。Zhao 等人^[5]以 ATSS^[6]为基准网络,提出一种新的标签感知机制来解决对象重叠问题。成浪等人^[7]基于逐层渐进式搜索策略和归一化指标性能评估策略,根据 X 射线图像违禁品数据集自适应的搜索设计最佳骨干网组件。王清等人^[8]为平衡检测精度和速度,提出结构重参数化方法,优化骨干网,提高模型检测速度。

目前的算法仍存在误检和漏检的情况,检测精度不高,本文选用 TOOD^[9]作为基准网络,从特征提取、注意力机制和多尺度特征融合方面,提出一种基于混合注意力和任务对齐的多尺度特征融合检测网络,用于解决违禁物品遮挡

严重、尺度变化大的问题,提高检测精度。在 SIXray 数据集^[10]进行验证,并与其他先进算法进行对比,验证本文方法的有效性。

1 本文方法

TAFNet 主要由 3 个部分组成,分别是特征提取网络、混合注意力机制和多尺度特征融合金字塔,整体框图如图 1 所示。

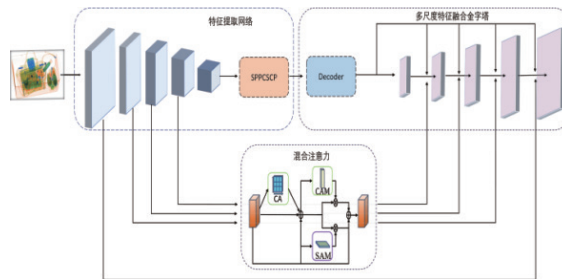


图 1 TAFNet 整体网络结构图

为解决 X 光图像中违禁品遮挡严重的问题,提出混合注意力模块(HAM),融合坐标、通道和空间三个层级的注意力特征,以增强模型对遮挡目标的感知能力;为应对 X 光图像中违禁品尺度差异显著、形状变化复杂等问题,引入 SPPCSPC 模块,加强模型对多尺度上下文信息的建模能力;并设计多尺度融合特征金字塔(multi-scale feature fusion pyramid, MFFP),将提取的高维语义信息与低维语义信息进行多尺度融合,得到兼具丰富细节与语义的综合特征图,最后将得到的特征图输入 T-Head 进行检测。

1.1 TOOD 算法框架

单阶段检测器通常通过分别优化目标分类和定位这两个子任务来实现目标检测，然而该优化方式可能导致两个任务的预测结果存在空间错位，从而影响检测效果。TOOD 使用任务对齐头（task-aligned head, T-Head），更好地对齐目标分类和定位两个任务。

T-Head 和 TAP（task-aligned predictor）如图 2、3 所示，T-Head 先将 FPN 输出的语义特征，输入任务特征提取模块，算出任务交互特征，然后将这些交互特征输入两个并行的 TAP 结构，分别完成分类和定位预测。之后，运用任务对齐学习策略 TAL，计算出一个用于衡量分类、定位任务对齐程度的对齐指标。在反向传播阶段，T-Head 依据这一对齐指标，自动调整分类概率与定位预测，以此优化两项任务间的协调性。

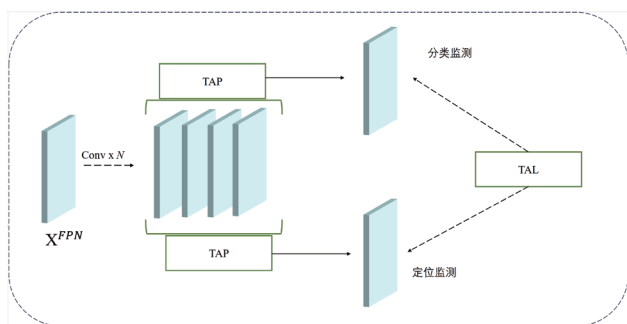


图 2 T-Head 结构图

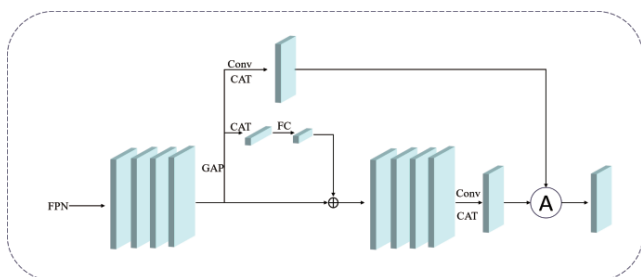


图 3 TAP 结构图

1.2 混合注意力模块

混合注意力模块如图 4 所示，由坐标注意力模块 CA、通道注意力模块（channel attention module, CAM）和空间注意力模块（spatial attention module, SAM）融合而成。坐标注意力聚焦于垂直与水平方向的远程依赖关系，精确保留了位置信息。然而，其单独操作在建模特征图通道间的整体相关性和空间位置的整体相关性上存在局限性，将坐标注意力、通道注意力和空间注意力机制联合使用，通过整合各自的优势，可以实现对特征图的更全面、深入的建模。具体来说，过程如下：在坐标信息嵌入阶段，坐标注意力机制通过沿水平和垂直方向的池化操作，有效地将输入特征图 X 的通道信息编码为两个方向上的空间特征表示。

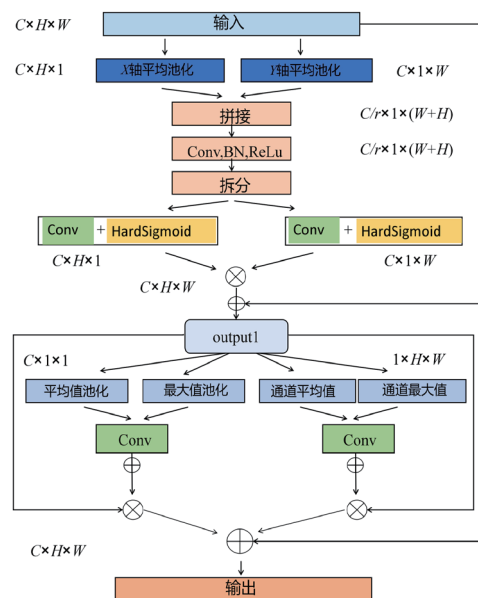


图 4 混合注意力机制

高度为 h 的垂直坐标第 c 通道的输出表示为：

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} x_c(h, i) \quad (1)$$

宽度为 w 的水平坐标第 c 通道的输出表示为：

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq i < H} x_c(i, w) \quad (2)$$

式（1）和式（2）的两种变换分别沿垂直和水平两个空间方向聚合特征。

坐标注意力生成过程中，将水平和垂直坐标嵌入得到的特征图进行拼接，随后通过 1×1 卷积变换函数 F_1 对拼接后的特征进行融合和变换，得到包含丰富空间信息的特征表示。变换公式为：

$$f = \delta(F_1([z^h, z^w])) \quad (3)$$

式中： $[\cdot]$ 为沿空间维数的 concat 操作； δ 为非线性激活函数； f 为对空间信息在水平方向和垂直方向进行编码的中间特征映射。

然后将中间特征映射 f 沿着空间维数分解为两个单独的张量：

$$\begin{cases} f^h \in \mathbf{R}^{C/r \times H} \\ f^w \in \mathbf{R}^{C/r \times W} \end{cases} \quad (4)$$

利用两个 1×1 卷积变换 F_h 和 F_w 分别将 f^h 和 f^w 变换为具有相同通道数的张量，得到：

$$\begin{cases} g^h = \sigma(F_h(f^h)) \\ g^w = \sigma(F_w(f^w)) \end{cases} \quad (5)$$

式中： σ 为 Hardsigmoid 激活函数，较 Sigmoid 函数，其计算更加简单，速度更快。

最后通过残差链接弥补局部丢失，坐标注意力模块的输出：

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) + x_c(i, j) \quad (6)$$

得到坐标注意力模块的输出之后，将输出的特征图并行地输入通道注意力模块和空间注意力模块。通道注意力模块最后输出表示为：

$$y_{CA} = X_{input} \times (g_{Avg} + g_{Max}) \quad (7)$$

式中： g_{Avg} 和 g_{Max} 分别为平均池化和最大池化的值。

空间注意力模块最后输出表示为：

$$y_{SA} = X_{input} \times (k_{Avg} + k_{Max}) \quad (8)$$

式中： k_{Avg} 和 k_{Max} 分别为通道平均值和通道最大值。

经过残差链接，混合注意力模块最后输出为：

$$output = y_{SA} + y_{CA} + x \quad (9)$$

1.3 多尺度特征融合金字塔

在违禁品检测任务中，目标通常具有尺度变化大、形态复杂且结构不统一等特性，增加了检测的复杂性。为此，本文设计多尺度特征融合金字塔，以提升模型对各类尺度目标的检测性能。其结构如图 5 所示，Decoder 模块负责对输入的高维度语义信息进行解码，之后把解码后的高维语义信息，与 Stage2、Stage3、Stage4 输入的低维语义信息，在特征金字塔进行融合，实现了高维语义信息和低维语义信息的互补，能够增强网络对前景目标的识别及定位能力。

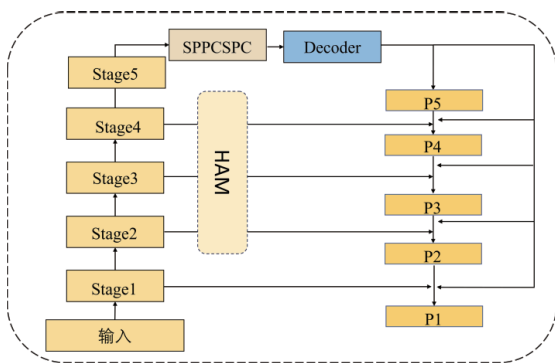


图 5 多尺度特征融合金字塔

特征解码模块如图 6 所示。特征解码模块包含一个输入和 A、B、C 和 D 四个输出支路，TCBR 为反卷积加归一化加 ReLU 激活函数。

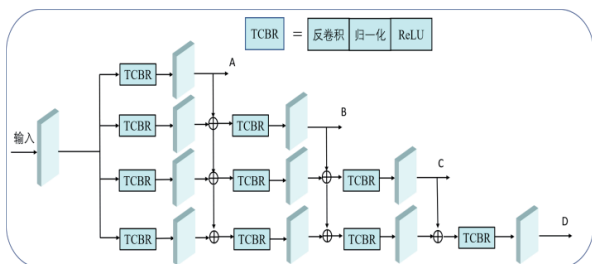


图 6 特征解码模块

具体来说，采用密集连接思想，将每一层的 TCBR 输出作为下层 TCBR 输入，增强局部信息融合，增强对小目标检测性能。

2 实验结果与分析

2.1 实验设置

本文实验环境：Windows10 操作系统，CPU 为 I7-4770，显卡型号为 NVIDIA GTX 1660，编程语言为 Python 3.8，深度学习框架为 PyTorch1.12，其余部分实验参数如表 1 所示。

表 1 实验参数设置

实验参数	设定值
Optimizer	Adam
Workers	8
Epoch	100
Batchsize	8

SIXray 数据集包含违禁物品的 X 光图像共计 8 929 张，数据集所包含的违禁物品分为五类：刀、枪、扳手、钳子和剪刀。

采用的技术指标包括：平均精度（average precision, AP）、平均精度均值（mean average precision, MAP）、参数量。

2.2 消融实验

为验证本文设计各个模块对模型性能提升效果，对各个模块进行了单独实验。以 TOOD（ResNet50+FPN）为基准模型，在 SIXray 数据集所得实验数据如表 2 所示。

表 2 组合不同模块实验结果对比

HAM	SPPCSCP	MFFP	mAP/%	参数量 /MB
			88.60	37.83
✓			90.24	38.19
✓	✓		91.17	65.21
✓	✓	✓	92.59	67.22

由表 2 可得，各个模块添加后在 mAP 上均有提升，嵌入 HAM 的 mAP 提高 1.64%，再加入 SPPCSCP 后 mAP 提高 0.93%，改进特征金字塔后 mAP 提高 1.42%，结合 3 个模块后的模型比基准模型提高 3.99%。以上结果表明本算法各个模块的有效性。

2.3 对比试验

为验证本文模型的优越性和泛化能力，以 TOOD 为基准模型，在 SIXray 和 HiXray 数据集上与诸多先进算法进行对比。实验结果如表 3 所示，“—”表示原文献没有提到此项指标。对比实验的对比算法包括单阶段算法 ATSS 和双阶段算法 DEYO^[11]，YOLO 系列先进算法 YOLOv7^[12]；MCIA-FPN 和 POD-F-R^[13] 是 X 光违禁物品检测领域先进算法。

表 3 SIXray 数据集结果对比

检测模型	Backbone	mAP/%	参数量 /MB
TOOD	ResNet50	85.32	37.83
ATSS	ResNet50	85.20	37.28
DEYO	ResNet50	87.90	136.30
YOLOv7	CSPDarknet53	89.02	37.62
POD-F-R	ResNet50	86.1	118.23
MCIA-FPN	ResNet50	83.70	—
TAFNet	ResNet50	92.59	67.22

由表 3 可知, 本文提出的模型 mAP 达到 92.59%, 与一般目标检测算法和违禁物品检测先进算法相比都取得更加优异的结果, 与 X 光违禁物品检测先进算法相比表现也十分优异。

图 7 展示了在不同场景和同一目标下的可视化检测结果, 对模型性能进行了直观对比分析。结果显示, 所提出的模型在遮挡环境和多尺度目标检测方面表现出更强的鲁棒性, 有效提升了对尺度变化大、形态各异违禁物品的检测能力。综合对比结果, 充分验证了本文模型在检测精度与效率方面的显著优势。

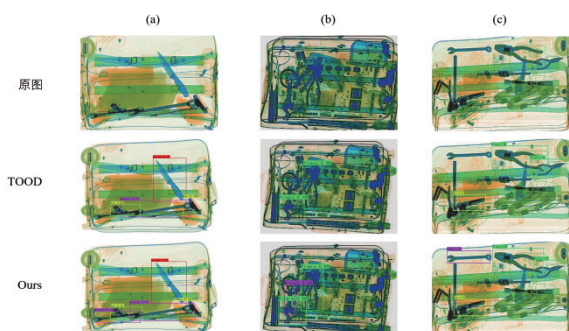


图 7 可视化结果对比

3 结论

本文提出了一种基于混合注意力和任务对齐的多尺度特征融合检测网络。针对违禁物品间的遮挡问题, 设计了混合注意力模块, 有效缓解特征丢失现象, 提升了网络的细节特征提取能力; 针对违禁物品尺度差异大的挑战, 构建多尺度特征融合金字塔, 以增强网络对小目标的检测能力。基于 SIXray 数据集的实验结果表明, 与经典模型及其他先进的 X 光违禁品检测算法相比, 所提出的模型在检测精度方面表现更为优异, 能够有效降低漏检率和误检率。

参考文献:

[1] 李晨, 张辉, 张邹铨, 等. 融合多尺度特征与全局上下文信息的 X 光违禁物品检测[J]. 中国图象图形学报, 2022, 27(10): 3043-3057.
[2] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition

(CVPR). Piscataway:IEEE,2021: 13708-13717.
[3] 游望, 侯进, 任东升, 等. 融合空间注意力的自适应安检违禁品检测方法[J]. 计算机工程与应用, 2023, 59(21):176-186.
[4] WANG M, DU H Q, MEI W B, et al. Material-aware cross-channel interaction attention (MCIA) for occluded prohibited item detection[J]. The visual computer, 2023, 39: 2865-2877.
[5] ZHAO C R, ZHU L, DOU S G, et al. Detecting overlapped objects in X-Ray security imagery by a label-aware mechanism[J].IEEE transactions on information forensics and security, 2022, 17: 998-1009.
[6] ZHANG S F, CHI C, YAO Y Q, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 9756-9765.
[7] 成浪, 敬超, 陈文鹏. 基于神经网络架构搜索的 X 射线图像违禁品检测算法[J]. 科学技术与工程, 2024, 24(2):665-675.
[8] 王清, 黄细霞, 阮慧. 结构重参数化与任务对齐的 X 光违禁品目标检测[J]. 制造业自动化, 2024, 46(10):7-16.
[9] FENG C J, ZHONG Y J, GAO Y, et al. TOOD: task-aligned one-stage object detection[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway:IEEE, 2021: 3490-3499.
[10] 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.
[11] OU-YANG H D. DEYO: DETR with YOLO for step-by-step object detection[EB/OL].(2023-06-16)[2024-05-26].https://doi.org/10.48550/arXiv.2211.06588.
[12] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR).Piscataway: IEEE, 2023: 7464-7475.
[13] MA C J, ZHUO L, LI J F, et al. Occluded prohibited object detection in X-ray images with global context-aware multi-scale feature aggregation[J]. Neurocomputing, 2023, 519: 1-16.

【作者简介】

朱鹏(1999—), 男, 湖北咸宁人, 硕士研究生, 研究方向: 计算机视觉、目标检测。

(收稿日期: 2025-04-11 修回日期: 2025-05-06)