E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035
中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS
©中国图象图形学报版权所有

中图法分类号:TP391.41;TP18　　**文献标识码**:A　　**文章编号**:1006-8961(2026)10-3636-17

论文引用格式:Han P, Bai H F and Luo S Y. 2026. Integrating 3D geometric constraints into RT-DETR for dual-view X-ray prohibited item detection network. Journal of Image and Graphics, 31(10):3636-3652(韩萍, 白海峰, 罗思宇. 2026. 融合三维几何约束的RT-DETR双视角X光违禁品检测网络. 中国图象图形学报, 31(10):3636-3652)[DOI:10.11834/jig.250597]

融合三维几何约束的RT-DETR双视角X光违禁品检测网络

韩萍*,白海峰,罗思宇

中国民航大学电子信息与自动化学院,天津　300300

摘　要:**目的**　基于双视角X光图像的违禁品检测技术因其在检测性能和检测成本间的平衡优势,在安检领域获得了广泛关注。现有双视角X光图像的违禁品智能检测技术研究主要聚焦于利用双视角间的特征融合提升检测性能,但模型在物品分布杂乱环境下的检测能力有限,且对不同大小违禁品检测的泛化性能不足。**方法**　为此,设计了融合三维几何约束的RT-DETR(real-time detection Transformer)双通道改进网络。在编码器部分,设计了双视角多尺寸特征加权融合模块,通过多尺寸窗口计算交叉注意力融合双视角特征,增强模型对不同尺寸违禁品特征的感知能力;在解码器部分,设计了三维锚框引导定位融合模块,生成三维锚框坐标与包含双视角特征的查询向量,为解码器提供明确的空间位置引导,有效校正定位偏差。在网络训练阶段,设计了双视角一致性数据增强策略,对双视角图像施加一致的几何变换,在保持双视角空间对应关系的前提下提升了网络对物品分布杂乱环境的泛化能力。**结果**　在DvXray数据集上的对比实验表明,提出的模型在俯视角和侧视角的平均精度均值(mean average precision,mAP)分别达到93.4%和83.9%;双视角联合统计结果中,P-R曲线(precision-recall curve)下面积为84.9%,F1分数最大值为83.5%。相较基线模型平均精度均值分别提高2.1%和6.8%,P-R曲线下面积和F1分数提升5.7%和5.2%。**结论**　实验结果表明,所提方法通过多尺寸特征融合与三维几何约束的协同作用,克服了单一视角的局限,提升了复杂安检场景下的违禁品检测精度。

关键词:双视角X光图像;违禁品检测;RT-DETR;特征融合;三维几何约束

Integrating 3D geometric constraints into RT-DETR for dual-view X-ray prohibited item detection network

Han Ping*, Bai Haifeng, Luo Siyu

College of Electronic Information and Automation, Civil Aviation University of China, Tianjin　300300, China

Abstract: **Objective** The growing demand for security screening in public spaces necessitates more efficient and reliable X-ray baggage inspection systems. Current manual interpretation of X-ray images at stations, airports, and stadium venues risks missing prohibited items due to operator fatigue. This method also requires considerable human resources, leading to high operational costs. Intelligent X-ray image detection technology, using computer-assisted verification, enhances security check reliability, improves efficiency, and reduces operational costs. Intelligent X-ray detection methods fall into three

收稿日期:2025-12-05;**修回日期**:2026-01-22;**预印本日期**:2026-01-29

*通信作者:韩萍　hanpingcauc@163.com

基金项目:中国民航安全能力建设基金(KJZ49420250165)

Supported by:Security Capacity Building Project of Civil Aviation Administration of China(KJZ49420250165)

categories based on imaging approaches: single view, dual view, and CT scanning. Dual-view systems, adding an orthogonal X-ray image to complement single-view inspection, offer improved detection performance by revealing occluded details. This approach achieves an optimal trade-off between detection capability and cost compared with CT scanners. Research on dual-view X-ray prohibited item detection has focused on leveraging the spatial correspondence between views to enhance model performance. Some studies have adopted a primary-auxiliary view strategy. In this strategy, the auxiliary view enhances feature representation in the primary view, somewhat improving detection accuracy. However, this asymmetric fusion approach only considers predictions from the primary view, which can potentially compromise accuracy in complex scenarios where prohibited items are poorly angled or heavily occluded in the primary view. Other researchers have developed symmetric multiscale architectures to boost detection performance through deep dual-view information fusion. However, these methods employ single-column feature fusion, limiting the exploration of potential correlations between feature columns at different scales. Such a limitation may affect the model's ability to recognize prohibited items of varying sizes and its overall robustness. The challenge remains in developing more effective dual-view feature fusion strategies. These strategies are needed to address complex screening scenarios and improve detection performance across different sizes of prohibited items. **Method** To tackle this challenge, this study presents a dual-view prohibited item detection model, extending the RT-DETR single-channel architecture to a dual-channel network. In the encoder, features are fed into a dual-view multiscale weighted fusion module. This module enhances feature extraction for different-sized prohibited items through multisize sliding windows and establishes spatial correspondence between dual-view features using a cross-attention mechanism. A gated feature fusion mechanism then integrates dual-channel features to strengthen the encoder's semantic representation. In the decoder, we introduce a 3D anchor-guided localization fusion module to mitigate occlusion. The core of this approach is a supervised fusion stage where the module dynamically generates a set of 3D anchor boxes and query vectors, guided by a 3D bounding box loss function. This encourages the model to learn a geometrically consistent representation, improving localization accuracy via geometric constraints. Furthermore, a synchronous dual-view data augmentation strategy was designed; by applying consistent geometric transformations to both views, this strategy enriches sample diversity while maintaining their spatial correspondence, further improving model generalization in cluttered environments. **Result** To validate the proposed method's effectiveness, experiments are conducted using ResNet-50 as the backbone network. The study compares the model against two mainstream detection approaches: Transformer-based models ((denosing DETR (DN-DETR), real-time detection Transformer (RT-DETR)) and YOLO models (YOLO11-X). Additionally, recent dual-view X-ray prohibited item detection methods YOLO_multi and Trans2Ray are included for comparison. All evaluations are performed on the DvXray dataset. Results show remarkable improvements in mAP_{50} scores. For the overhead view (OL) view, the proposed method achieves improvements of 2.2%, 12.2%, 2.1%, 7.6%, and 16.7% compared with YOLO11, DN-DETR, RT-DETR, Trans2Ray, and YOLO_multi, respectively. For the side view (SD) view, improvements are 5.7%, 18.7%, 6.8%, 25.1%, and 17.0%, respectively. Analysis of individual category performance across 15 prohibited items reveals that the proposed method achieves superior detection in 12 categories for the OL view and 12 categories for the SD view. Under dual perspective joint statistical results, the method demonstrates optimal performance with an area under the P-R curve of 84.9% and a maximum F1 score of 83.5%. For visual analysis, detection results from three representative image sets were examined. The first set comprises contraband under challenging conditions characterized by cluttered backgrounds and rotation, the second set evaluates the model's proficiency in detecting small objects, occluded backgrounds, and two contraband instances, and the final set consists of negative samples devoid of prohibited items. These visualizations demonstrate that our proposed model achieves optimal performance. Next, we visualized the attention mechanisms within the encoder. The corresponding visualizations for our proposed model were clearly superior to the baseline model. The experimental outcomes validate the effectiveness of the proposed methodology and its components through comprehensive ablation studies. These results demonstrate that the proposed dual-view detection framework effectively improves prohibited item detection performance in X-ray security screening applications. **Conclusion** This study proposes an improved detection network based on the RT-DETR architecture to enhance contraband detection accuracy in X-ray images, particularly in cluttered environments. The model incorporates a dual-view multiscale weighted fusion module with a multiscale window cross-attention mechanism, strengthening recognition capability for vari-

ably sized contraband. A 3D anchor-guided localization fusion module further imposes geometric constraints, providing explicit positional guidance to enhance the decoder's localization accuracy. Additionally, our dual-view joint decision-making strategy enables comprehensive evaluation of detection performance across dual-view data. Experiments demonstrate the model's competitive performance and robustness in dual-view X-ray contraband detection. However, the model faces persistent challenges including architectural complexity, excessive parameters, and time-consuming training, primarily from computational overhead induced by the dual-channel structure. Future work will explore lightweight dual-view fusion methods and develop efficient collaborative optimization algorithms to balance computational efficiency with detection performance.

Key words: dual-view X-ray image; prohibited item detection; real-time detection Transformer (RT-DETR); feature fusion; 3D geometric constraint

0 引　言

随着社会发展,公共场所对安全检测的需求持续增长。当前,车站、机场和体育馆等公共场所的X光行李安检主要依赖人工判读,这不仅存在因工作量大导致安检人员注意力下降而漏检违禁品的风险(Chavaillaz等,2019),而且需要投入大量人力,导致运营成本较高。应用X光图像智能检测技术辅助人工核查,既可提升安检可靠性,又能提高效率并降低运营成本(Huegli等,2020)。

现有X光图像的违禁品智能检测方法按照X光成像方式可分为3类:单视角、双视角和CT(computed tomography)安检机图像违禁品检测。其中,双视角X光成像系统通过增加一个与单视角正交的X光图像(Ma等,2024),有效补充被遮挡细节,获得比单视角更好的检测效果。同时,其成本低于CT安检机,在检测性能和成本之间实现了良好平衡。双视角X光图像的空间对应关系如图1所示,由X光俯视图和侧视图组成,在三维笛卡儿坐标系中存在共X轴的空间几何对应关系。成像物体俯视图和侧视图的边界框可组成三维边界框。

当前,针对双视角X光图像的违禁品检测研究主要分为经典方法与深度学习方法两类。在经典方法领域,Mery等人(2013)提出了二阶段模型,先提取单视角图像关键点,再利用空间对应关系将关键点投影至三维空间聚类,最终映射回二维视角完成分类。Baştan(2015)则设计了双视角互校正机制,利用几何一致性抑制单视角的误检率。随着深度学习的发展,Steitz等人(2018)基于Faster R-CNN(faster region-based convolutional neural network)框架,设计了基于几何关系的多视图池化层,将多视图

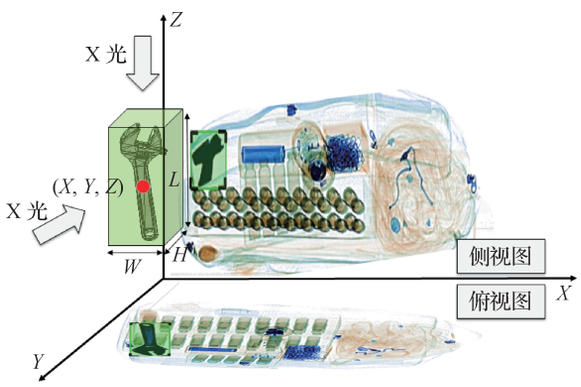


图1　三维笛卡尔坐标系下双视角X光违禁品图像

Fig. 1　Dual-view X-ray images of prohibited items in a three-dimensional Cartesian coordinate system

特征聚合至三维空间以实现检测。Isaac-Medina等人(2021)基于YOLOv3(you only look once)引入极线几何约束,通过视角间的几何关联抑制虚警。然而,上述方法缺乏对双视角特征在语义层的交互与融合,忽视了视角间特征表示的相互促进与增强,限制了模型在复杂场景下对特征的挖掘能力。

近年来的研究表明,在高光谱的光谱特征与合成孔径雷达(synthetic aperture radar,SAR)数据融合(金学鹏 等,2025)、可见光图像与红外图像的融合(魏思和杨文璐,2025),以及RGB图像与深度图像的融合(宋霄罡 等,2025)等任务中,多源特征融合均被验证为克服感知局限、提升鲁棒性和模型检测性能的有效手段。

与这一技术发展趋势相一致,针对双视角X光违禁品检测任务,为进一步挖掘双视角特征间的关联信息,现有研究在融合架构的设计上形成两种主流策略。

第1种策略是主—辅视角非对称融合策略,即利用辅助视角增强主视角的特征表达。张海刚等人提出Dualray模型(Wu等,2022),采用双视角单通道

架构，结合通道注意力与空间注意力机制实现侧视角信息向主视角的有效迁移；随后，该团队发布Trans2Ray模型（Meng等，2024），利用全局交叉注意力区分背景干扰，结合局部特征选择模块增强违禁品表征；此外，该团队还基于DINO（DETR with improved denoising anchor boxes）架构设计了双视图检测模型（Sun等，2025），开发垂直方向信息补偿模块以应对极端视角下的特征缺失。同期北京交通大学陶仁帅团队（Tao等，2025）提出专家模型架构，通过主辅视图协同检测提升困难样本识别精度。然而，这种非对称的融合方式仅关注主视角的预测，未能充分利用辅助视角的独立检测价值，在主视角违禁品角度不佳或者遮挡严重等复杂安检实际应用场景下，会影响违禁品检测的准确率。

第2种策略是双视角对称特征融合策略，旨在构建对等的双通道架构以挖掘视角间关联。Isaac-Medina等人（2022）引入Transformer架构，利用交叉注意力机制实现了视角间特征的互补。马博文团队提出AHCR（adaptive hierarchical cross refinement）框架（Ma等，2024），构建多尺度双视角融合模块以辅助分析。但该方法采用单列特征融合，限制了对不同尺度列特征之间潜在关联的探索。针对这一问题，该团队设计了MACA（material-aware coordinate-aligned attention）方法（Jia等，2025），通过多尺寸池化核聚合单个视图的上下文信息，以增强对不同尺寸物体的识别能力，并使用哈达玛乘积进行双视角的特征交互。然而，该方法在生成引导信号时，将特征图在垂直空间维度上压缩至一维，这一操作模糊了同一水平位置上不同物体的空间分布，因空间上下文的丢失而引入特征歧义，从而削弱了双视角特征的融合效果。综上所述，如何在保留关键空间信息的前提下，实现更鲁棒的特征对齐与融合，并充分利用双视角固有的几何先验来约束定位，仍是当前研究需要关注的问题。

针对上述问题，本文提出一种面向双视角X光图像的违禁品检测网络。通过设计双视角交互机制对RT-DETR（real-time detection Transformer）框架（Zhao等，2024）进行改进，结合特征融合与几何定位的联合优化以有效处理来自双视角的语义与空间信息，其主要贡献如下：

1）为增强编码器对不同尺寸违禁品的检测能力，本文设计了双视角多尺寸特征加权融合模块。该模块首先引入侧重保留垂直空间结构的多尺寸滑窗机制，在适应不同违禁品尺寸的同时有效剔除背景噪声。在此基础上，利用交叉注意力机制实现双视角特征的对齐，自适应地聚合互补信息。最后，结合自适应加权与门控融合机制，动态调节融合比例以解决特征间的语义冲突，从而优化了编码器在复杂场景下的特征提取与泛化能力。

2）针对物体堆叠与遮挡导致的特征模糊与定位误差问题，本文设计了三维锚框引导定位融合模块。该模块首先构建双视角全局上下文记忆，在此基础上推断目标的三维空间位置，生成三维锚框及其对应的查询向量。这些信息被同时作为几何先验与语义先验注入解码器，为后续检测提供明确的位置引导与特征初始化。该机制通过联合优化几何与语义信息，有效校正了复杂遮挡环境下的定位偏差，并增强了特征的判别能力。

3）针对常规数据增强会破坏双视角图像间空间对应关系的问题，本文设计了一种双视角一致性数据增强策略。通过对图像对施加一致的几何变换，保持空间关系的同时提升了模型在物品分布杂乱环境下的泛化能力与检测鲁棒性，降低了过拟合风险。

4）为验证所提方法的有效性，本文在DvXray数据集上，将其与通用目标检测模型及近期相关的双视角X光违禁品检测模型进行了全面对比。结果显示，该方法在俯视角和侧视角的mAP_{50}分别达到93.4%和83.9%。双视角联合统计结果的P-R曲线（precision-recall curve）下面积为84.9%，最大F1分数达到83.5%。相较基线模型平均精度均值分别提高2.1%和6.8%，P-R曲线下面积和F1分数提升5.7%和5.2%。上述指标在对比实验中均表现最优，实验结果证明了本文方法的有效性。

1 本文方法

1.1 总体框架

本文选取RT-DETR作为基线，如图2所示。这一选择首先考虑到Transformer架构具备全局感受野，能有效捕捉违禁品的全局上下文信息。其次，其无NMS（non-maximum suppression）范式更适于处理密集堆叠的X光图像。此外，该模型在保持高精度的同时推理速度优于同等量级YOLO，符合安检系

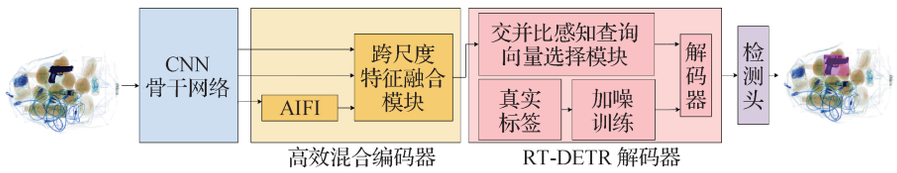


图2　RT-DETR架构

Fig. 2　Architecture of RT-DETR

统对实时性的需求。既往研究中，胡佳乐等人(2024)、于梦源和刘向阳(2026)、金涛和胡配雨(2025)分别在小目标检测、跨模态融合及安检遮挡场景下证实了该架构的有效性，共同佐证了将其作为双视角X光违禁品检测基线的合理性。

在此基础上，如图3所示，本文将RT-DETR从处理单视角扩展至双视角，以俯视角(overhead view，OL)和侧视角(side view，SD)的图形作为双路输入，通过构建双视角多尺寸特征融合模块和三维锚框引导定位融合模块，提出了一种融合三维几何约束的RT-DETR双视角违禁品检测模型。该模型由卷积神经网络(convolutional neural network，CNN)骨干网络、基于双视角多尺寸特征融合的增强编码器、基于三维锚框引导定位融合的增强解码器及检测头4部分构成。

1.2　基于双视角多尺寸特征融合的编码器增强

为增强编码器对不同尺寸违禁品的检测能力，在原编码器基础上，利用双视角X光图像空间对应关系，本文设计了双视角多尺寸特征加权融合模块，如图4所示，包含4个部分，即多尺寸滑窗机制、交叉注意力机制、多尺寸特征加权融合模块和门控特征融合机制。

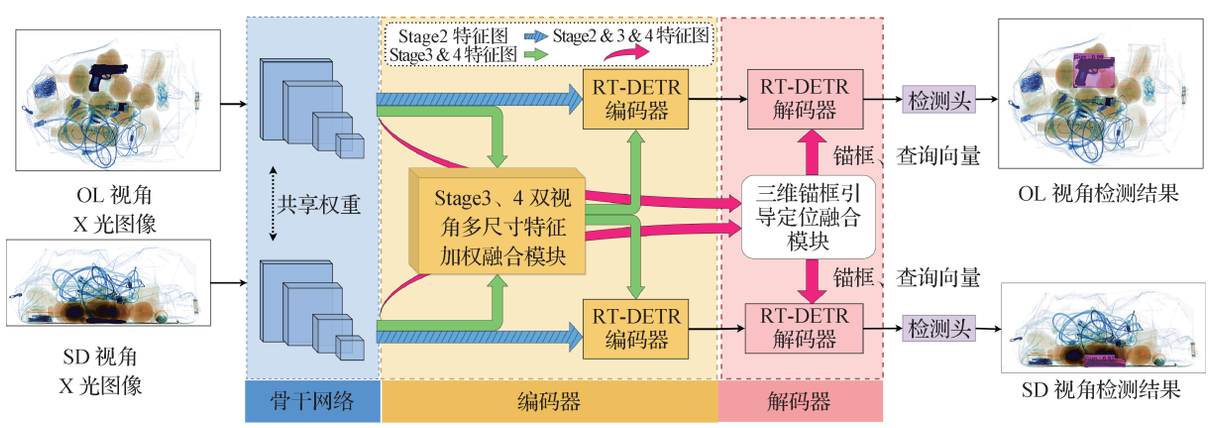


图3　模型总体框架

Fig. 3　Overall framework of the model

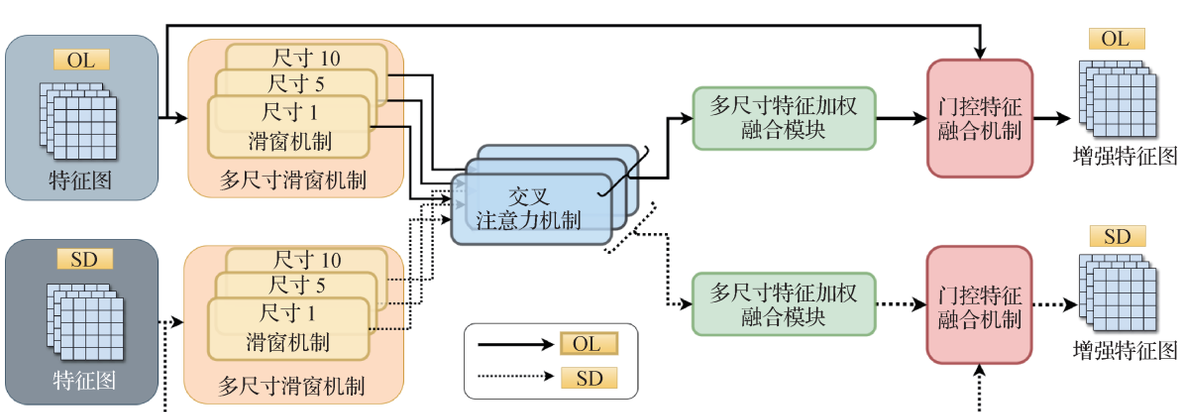


图4　双视角多尺寸特征加权融合模块

Fig. 4　Dual-view multiscale feature weighted fusion module

1.2.1 多尺寸滑窗机制

为了在计算双视角特征相关性时，避免因违禁品与背景特征的错误融合而引入噪声，同时有效应对X光安检中违禁品种类繁多、尺寸各异的挑战，本文设计了一种多尺寸滑动窗口机制。该机制独立地应用于每个视角的特征图，具体结构如图5所示。

该机制设计侧重于保留特征的垂直空间结构。为兼顾检测性能与计算效率，选取 $S=(1,5,10)$ 3种大小的窗口，对特征图 $\boldsymbol{f}_{\text{org}}$ 进行多尺寸滑动窗口截取，以适应不同尺寸违禁品的融合需求，从而有效地截取出违禁品特征，排除背景特征的干扰。随后，为进一步挖掘多尺寸滑动窗口所得张量包含的信息，借鉴平均去噪的思想，计算窗口平均特征以获取更鲁棒的特征表示 $\boldsymbol{f}_{W_i}^{S}$。其中 S 代表不同窗口大小，$\boldsymbol{W}_i$ 为特征图的第 i 列。这种在水平维度上进行局部压缩的处理，使得传递给后续交叉注意力模块的特征向量，依然保有丰富的垂直空间信息，有助于缓解因在交互前压缩垂直维度引入的特征歧义。

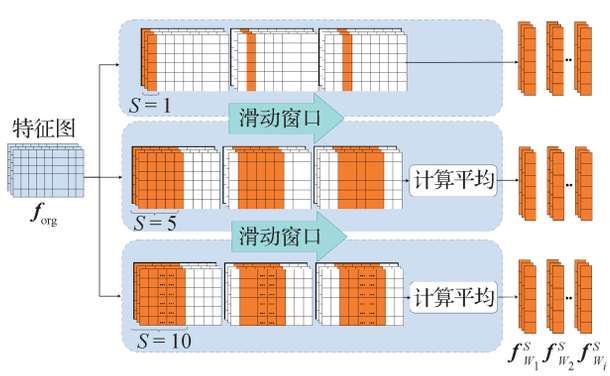


图5 多尺寸滑窗机制

Fig. 5 Multisize sliding window mechanism

1.2.2 交叉注意力机制

在通过滑动窗口提取出多尺寸局部特征后，为实现双视角特征的融合，本文引入了交叉注意力机制，以OL视角为例，如图6所示。将双视角不同窗口尺寸提取的局部特征 $\boldsymbol{f}_{W_i}^{S}$ 分别记为 $\boldsymbol{f}_{W_i_\text{OL}}^{S}$ 与 $\boldsymbol{f}_{W_i_\text{SD}}^{S}$，并分尺寸独立地构建注意力交互。这种处理方式确保了双视角特征的对齐严格限制在相同的感受野内，有效规避了跨尺寸特征间的错误关联，从而保证了特征交互在同尺寸下的语义一致性。

具体计算中，首先，由自身特征生成的查询向量(query)与另一视角的键向量(key)计算点积，从而衡量双视角间的特征相似性，构建双视角间的语义

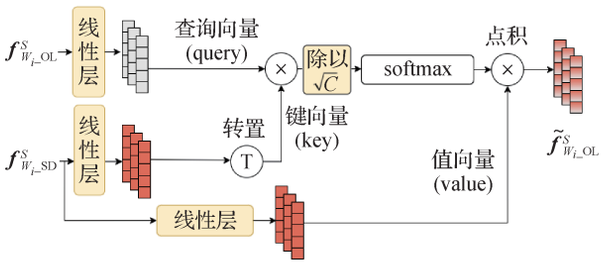


图6 交叉注意力机制计算过程

Fig. 6 Calculation process of cross attention

关联。为缓解点积结果过大导致的softmax函数输出饱和进而引起梯度不稳的问题，将其除以缩放因子 $\sqrt{C}$ 进行归一化。最终，将该权重与另一视角的值向量(value)相乘，模型能够自适应地将另一视角中高相关性的互补特征动态聚合到本视角中，并得到融合后的特征 $\tilde{\boldsymbol{f}}_{W_i_\text{OL}}^{S}$，同理可得 $\tilde{\boldsymbol{f}}_{W_i_\text{SD}}^{S}$。由于双视角分支的后续处理逻辑对称，为简化表述，下文将统一用 $\tilde{\boldsymbol{f}}_{W_i}^{S}$ 表示。这种基于内容的交互方式使模型能够聚焦于双视角图像中的语义一致区域，有效抑制背景干扰的同时提升特征提取的鲁棒性。

1.2.3 多尺寸特征加权融合模块

为保留各尺寸特有的对齐信息，本文设计了多尺寸特征加权融合模块，对不同尺寸的特征进行自适应融合，为每个尺寸的特征图动态地生成一个重要性权重并加权求和，具体为

$$\boldsymbol{f}_{\text{fuse}}^{W_i}=\sum_{S=1,5,10}\left(\frac{\sum_{h=0}^{H}\tilde{\boldsymbol{f}}_{W_i_h}^{S}}{H}*\tilde{\boldsymbol{f}}_{W_i}^{S}\right) \tag{1}$$

式中，$\boldsymbol{f}_{\text{fuse}}^{W_i}$ 表示多尺寸融合特征 $\boldsymbol{f}_{\text{fuse}}$ 的第 i 列特征，$\tilde{\boldsymbol{f}}_{W_i_h}^{S}$ 代表交叉注意力增强后的特征 $\tilde{\boldsymbol{f}}_{W_i}^{S}$ 在垂直高度 H 上第 h 行的特征值。首先，对每一个尺寸的特征图独立执行全局平均池化操作，压缩空间信息以生成能够量化该尺寸重要性的动态权重。通过将这些权重与它们对应的多尺寸融合特征进行加权求和，模型能够根据当前输入图像的内容，自动判断并强化包含丰富违禁品信息的关键尺寸，避免了采用固定权重分配可能导致的对关键尺寸信息的抑制问题，有效提升了模型对不同尺寸目标的表达能力。

1.2.4 门控特征融合机制

为实现融合特征与原始特征的有效整合，同时避免直接相加或拼接可能导致的细节丢失与语义冲

突，本文设计了如图7所示的门控特征融合机制。

首先，该机制采用了按列处理的策略，而非将特征图整体拼接后进行全局变换。这种设计旨在为每一列特征独立分配融合权重，以适应不同列之间差异化的融合需求，从而避免全局统一处理导致无法兼顾局部特性的问题，确保了特征变换在精准的局部感受野内进行。在此基础上，为了提升模型对违禁品特征的敏感度，模块引入通道注意力机制对拼接后的第i列原始特征$\boldsymbol{f}_{\text{org}}^{W_i}$与多尺寸融合特征$\boldsymbol{f}_{\text{fuse}}^{W_i}$进行预处理增强，突出关键通道的响应。处理后的特征进入双分支路径进行处理。其中，门控函数分支旨在生成自适应的在[0,1]区间内的权重系数γ；而与之并行的特征混合函数分支则负责对原始特征与融合特征进行非线性的初步重组与变换，这种变换能够在保留违禁品纹理细节的同时，有效地适配融合特征的分布。最后，通过门控权重系数γ，对特征混合函数分支输出的重组特征与原始特征进行加权融合。这种设计不仅解决了特征间的语义冲突问题，而且确保了RT-DETR编码器能够获得兼具丰富上下文信息与清晰的原始细节特征表达。

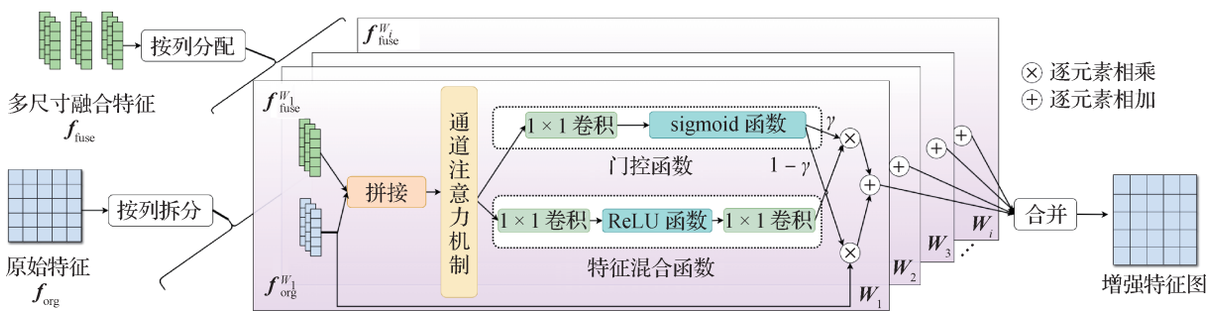


图7　门控特征融合机制

Fig. 7　Gated feature fusion mechanism

1.3　基于三维锚框引导定位融合的解码器增强

针对双视角X光图像检测中因物体堆叠和遮挡导致的定位不准问题，本文利用双视角的几何约束，设计了三维锚框引导定位融合模块，如图8所示。需明确的是，本文所述的三维并非指基于体素或点云的三维重建，而是指利用双视角几何关系推断出的目标三维空间位置。该模块通过构建双视角全局上下文记忆以生成三维锚框与特征查询，并将其作为几何先验与语义先验注入解码器，从而提供明确的空间位置引导，有效校正复杂场景下的定位偏差。

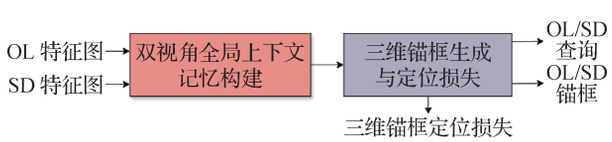


图8　三维锚框引导定位融合模块

Fig. 8　3D anchor-guided localization fusion module

1.3.1　双视角全局上下文记忆构建

为建立双视角间的语义关联，本文首先构建了基于Transformer的双视角全局上下文记忆模块。首先，为在二维特征向一维序列转化的过程中保留空间几何信息，本文引入了正弦位置编码$\boldsymbol{Pos}$，并将其注入到序列化后的特征中。随后，将俯视角与侧视角的特征序列进行拼接，构建包含双视角信息的联合特征序列。最后，利用Transformer编码器的自注意力机制，对双视角特征进行全局交互建模。该过程可以描述为

$$\boldsymbol{M}_{\text{global}} = T\left(Concat\left(F\left(\boldsymbol{f}_{\text{OL}}\right),F\left(\boldsymbol{f}_{\text{SD}}\right)\right) + \boldsymbol{Pos}\right) \tag{2}$$

式中，$F(\cdot)$表示对输入的双视角特征图$\boldsymbol{f}_{\text{OL}}$与$\boldsymbol{f}_{\text{SD}}$的展平操作，$Concat(\cdot)$表示拼接操作，$T(\cdot)$表示多层Transformer编码器的非线性映射。借助自注意力机制的全局建模能力生成的统一记忆表征$\boldsymbol{M}_{\text{global}}$，不仅保留了每个视角的局部细节，更有效整合了两个视角的互补信息，为后续生成符合几何约束的三维锚框提供了丰富上下文的特征基础。

1.3.2　三维锚框生成与定位损失构建

为在解码器中注入明确的几何先验，增强双视角特征之间的几何约束，本文借鉴DAB-DETR（Liu等，2022）的思想，将二维锚框扩展为三维锚框，其形式化表示为(x,y,z,w,h,l)，x,y,z分别代表锚框的归一化三维中心坐标，w,h,l代表长宽高。该三维表示方法利用几何先验，对双视角图像中X坐标轴与长

度W维度的一致性施加了有效约束。一方面有效约束了查询向量包含的位置特征信息，另一方面为RT-DETR解码器提供了高质量的初始值。这不仅加快了模型收敛速度，还削弱了查询向量随机初始化的负面影响，使模型能更精确地定位违禁品。

为实现前述三维锚框的设计理念，本文建立了从生成筛选到损失监督的完整机制。在三维锚框生成阶段，结合特征图尺寸动态生成密集的三维锚框，并根据$\boldsymbol{M}_{\text{global}}$计算出三维锚框的预测类别得分与锚框偏移量，筛选出置信度排名前M的候选查询向量和对应三维锚框。为了适配解码器的交互机制，这些高质量的三维锚框被投影至各视角的二维视图平面，作为空间参考点与对应的查询向量一并传递给后续的RT-DETR解码器。通过这一过程，模型在解码的起点便获得了三维空间的先验知识，从而能够利用几何信息校正二维平面上的定位误差。

最后，在模型监督阶段，本文设计了针对三维锚框的损失函数，以对候选框的输出进行监督。本文基于匈牙利匹配算法（Kuhn，1955），对置信度排名前M的每组预测类别$\hat{\boldsymbol{P}}$和三维锚框$\boldsymbol{B}_{\text{3D}}$，与标签类别$\boldsymbol{Cls}$和真值框$\boldsymbol{TB}_{\text{3D}}$寻找全局最优二分匹配$\delta$，其匹配成本矩阵$\boldsymbol{C}$为分类成本与回归成本的加权，每个元素$C_{ij}$代表第$i$个预测与第$j$个真值间的匹配代价，计算式为

$$C_{ij}=C_{\text{cls}}\left(\hat{\boldsymbol{P}}_i,\boldsymbol{Cls}_j\right)+C_{\text{L1}}\left(\boldsymbol{B}_{\text{3D_}i},\boldsymbol{TB}_{\text{3D_}j}\right) \tag{3}$$

式中，$C_{\text{cls}}(\cdot)$为Focal Loss，用以计算分类成本，$C_{\text{L1}}(\cdot)$为锚框与真值框间L1距离损失。根据匈牙利匹配算法，在代价取得最小时确定最优匹配关系δ，最终损失函数L_{3d}定义为所有匹配对的分类损失与回归损失的加权和，并由正样本总数N_{pos}进行归一化，该损失可表示为

$$L_{\text{3d}}=\frac{\sum_{\delta}\left[\alpha\cdot L_{\text{focal}}\left(\hat{\boldsymbol{P}}_{\delta},\boldsymbol{Cls}_{\delta}\right)+\beta\cdot L_{\text{L1}}\left(\boldsymbol{B}_{\delta},\boldsymbol{TB}_{\delta}\right)\right]}{N_{\text{pos}}} \tag{4}$$

式中，L_{focal}和L_{L1}分别代表Focal Loss和L1损失，α和β为平衡两类任务的超参数，考虑到遮挡场景下几何定位的收敛难度，将权重设定为$\alpha=2$，$\beta=5$，以强化三维几何约束的引导作用。通过三维锚框生成和三维定位损失监督，模型被引导学习到双视角内在的几何对应关系，不仅有效规避了由物体遮挡等因素引发的复杂环境干扰，更实现了高效、鲁棒的目标定位。

1.4 双视角一致性数据增强

数据增强是提升模型泛化能力的有效技术，但其在双视角检测任务的应用中，若对两个视角独立执行随机变换，会破坏其中固有的几何对应关系，进而损害依赖于此的特征融合模块和三维锚框引导模块。这一挑战已导致现有方法（Jia等，2025）被迫在数据多样性与空间对应关系中做出妥协，即牺牲泛化性以保留几何先验。

为解决此问题，本研究提出了一种双视角一致性数据增强策略。该策略的核心是用一个共享的随机决策源取代独立的变换，在样本间保持随机数据增强的同时，在视角间保持一致。具体决策源生成过程为

$$S_{\text{sync}}=(i\times C)+e+K \tag{5}$$

式中，S_{sync}代表同步种子，i代表双视角对应的索引，e表示训练周期数，C和K为常数，以保证不同索引下的随机种子不同。通过同步种子初始化随机数生成器，以保证双视角一致的随机决策。

该策略与双视角多尺寸特征加权融合模块和三维锚框引导定位融合模块形成了有效的协同设计。所生成的训练数据，既保证了样本的多样性，又维持了双视角间的几何一致性，促使模块学习到更泛化的对齐能力，更鲁棒地处理视角间的几何变化。

2 实验及分析

在本节中，首先，与多个主流目标检测网络和近期发布的双视角X光图像目标检测模型进行效果对比。之后，通过消融实验、计算效率分析和可视化，阐释本文设计网络的有效性。

2.1 数据与参数设置

限于当前可查询的公开资料，双视角X光违禁品检测研究可用的数据集的数量有限。如表1所示，除DvXray和LDXray数据集开源外，其他具备标注的多视角数据集均为私有数据集。开源数据集中，仅DvXray数据集包含两个视角的标注，LDXray数据集仅对一个视角进行标注。因此本文实验使用DvXray数据集。共包含15类违禁品，如图9所示。正样本和负样本的数量分别为22 000幅和10 000幅。将正负样本混合，训练与测试比例为4∶1。

对于目标检测评估指标，使用COCO（common objects in context）（Lin等，2014）评估指标和本文提

表1　双视角和多视角违禁品安检数据集

Table 1　Dual-view and multi-view datasets for prohibited items security inspection

名称	年份	种类	正/负样本数	视角	是否公开
LDXray(Tao等,2025)	2025	12	293 994/0	2	是
DvXray(Ma等,2024)	2024	15	10 000/22 000	2	是
Dualray(Sun等,2025)	2025	14	24 584/0	2	否
Dbf4(Isaac-Medina等,2021)	2021	4	10 112/0	4	否

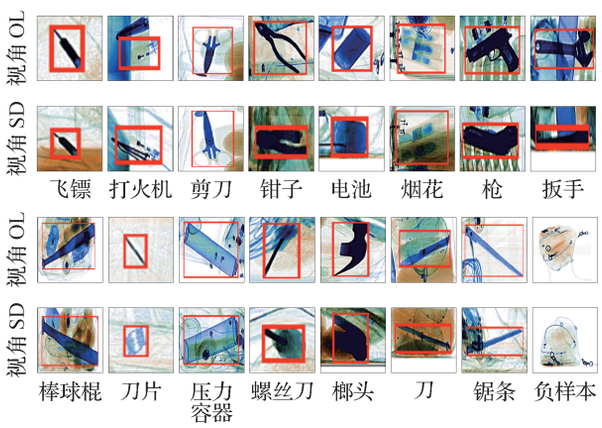


图9　DvXray数据集展示

Fig. 9　DvXray dataset demonstration

出的双视角联合判断指标。其中,AP_{50}和mAP_{50}分别反映了单类别及全类别在交并比阈值为0.5时的平均精度。此外,对双视角输出结果进行联合统计,具体逻辑如图10所示。只有双视角均成功检测到目标,为真正例(true positive,TP),否则为假负例(false negative,FN);任一视角出现虚警即记为假正例(false positive,FP)。基于此标准,选取[0,0.01,0.02,0.05,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1.0]作为置信度阈值进行遍历统计,计算出对应的查准率(precision,P)、查全率(recall,R)和F1分数。根据P-R曲线下面积、P-R曲线平衡点和F1曲线最大值,作为双视角联合统计结果。

所有模型均基于AutoDL算力云平台的1张VGPU-32 GB显卡。本文模型学习率为1×10^{-5},批数为8,训练72轮。本文采用在线数据增强策略。在此机制中,增强概率p定义了样本接受变换的概率,本文实验设定为$p=0.5$,其决定了每个训练周期内参与增强的样本占总训练样本的期望比例。

2.2　对比实验

为验证本文方法的有效性,实验将本文方法与目标检测领域的两类主流通用模型进行了对比,包括基于Transformer架构的DN-DETR、RT-DETR,以及YOLO系列中的YOLO11-X。为进一步验证模型的泛化能力,将近年来公开发表的双视角X光图像违禁品检测方法YOLO_multi(Isaac-Medina等,

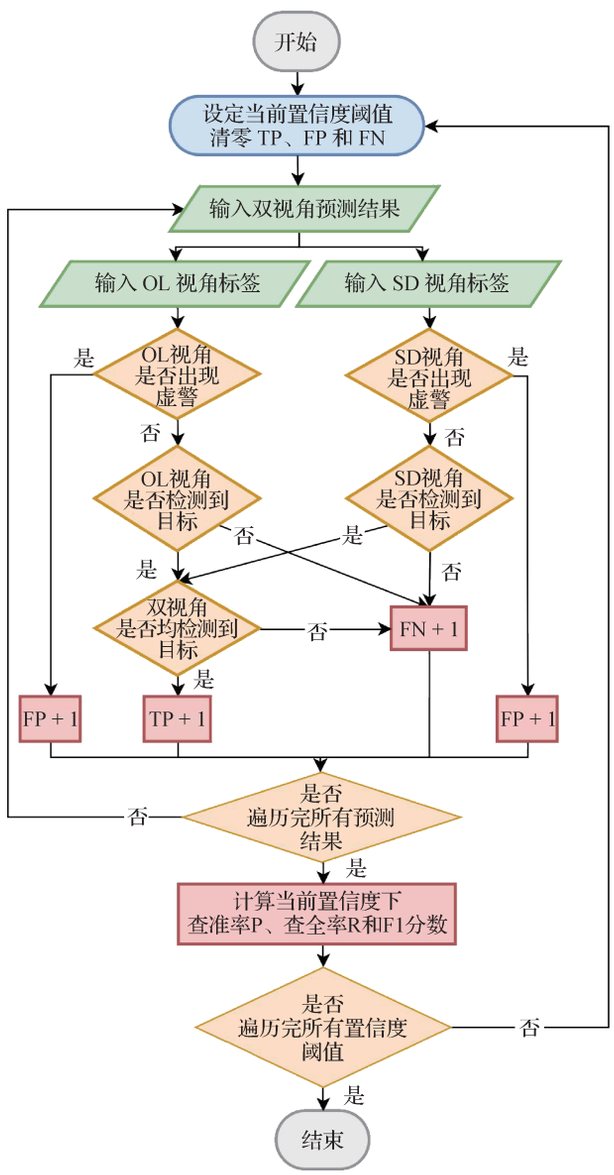


图10　双视角联合统计逻辑

Fig. 10　Dual perspective joint statistical logic

2021）和Trans2Ray（Meng等，2024）作为同方向比较。对于DvXray数据集提出的方法AHCR（Ma等，2024），由于其主要针对分类任务，其定位结果依赖于弱监督热力图生成，缺乏边界框回归分支，评估指标不同，难以进行统一对比，因此未将其纳入本节实验。对比结果如表2和表3所示。

从表2和表3可以看出，本文方法相较于其他几种目标检测算法具有良好的检测性能。从全类平均精度（mAP_{50}）看，OL视角相比于通用目标检测算法YOLO11、DN-DETR和RT-DETR和双视角目标检测模型Trans2Ray、YOLO_multi分别提升了2.2%，12.2%，2.1%，7.6%，16.7%；SD视角提升了5.7%，18.7%，6.8%，25.1%，17.0%。从单一类别AP_{50}上分析，在15种违禁品中，本文方法在OL视角有12种检测性能占优，SD视角有12种检测性能占优。以上结果表明，双视角多尺寸特征加权融合模块和三维锚框引导定位融合模块分别强化了特征层面的互补表达与解码层面的几何定位精度，提升了网络对不同尺寸违禁品检测的泛化性。

为评估本文方法的先进性，与本领域较新的检测模型（Jia等，2025）进行比较，如表4所示。本文方法在mAP_{OL}和mAP_{SD}的各项指标上得分均高于最新模型。在$mAP_{50:95}$指标上，本文方法在双视角分别高出10.7%和11.4%。与此同时，本文方法的帧率达到了91.0帧/s，高于对比模型。这些数据表明，本文方法在检测精度和推理速度上均表现出优势。

本文方法对于SD视角的检测能力提升明显，相比于基线模型RT-DETR，mAP_{50}提升了6.8%。对于

表2　OL视角对比实验结果

Table 2　Comparison of experimental results from the OL view

/%

模型	mAP_{50}	AP_{50}														
		枪	刀	扳手	钳子	剪刀	打火机	电池	棒球棍	刀片	锯条	烟花	榔头	螺丝刀	飞镖	压力容器
YOLO11（Ultralytics，2025）	91.2	**99.2**	87.3	97.6	98.5	73.7	82.1	97.2	99.5	**77.4**	86.7	91.6	99.1	94.6	85.8	97.0
DN-DETR（Li等，2022）	81.2	96.7	73.8	96.7	92.5	66.4	64.8	96.8	97.8	48.5	80.9	78.5	96.8	72.5	67.1	88.8
RT-DETR（Zhao等，2024）	91.3	98.7	93.5	97.5	**99.1**	76.6	79.7	98.9	**100.0**	68.1	87.4	92.8	99.8	94.8	86.7	96.2
Trans2Ray（Meng等，2024）	85.8	96.5	81.6	92.1	96.9	76.3	74.4	92.5	98.0	56.1	80.9	81.0	97.0	94.7	71.8	92.9
YOLO_multi（Isaac-Medina等，2021）	76.7	58.8	69.7	87.5	91.5	55.3	50.6	87.0	**100.0**	67.4	69.7	77.2	86.7	80.0	78.4	82.1
本文	**93.4**	99.1	**94.5**	**99.5**	**99.1**	**78.8**	**83.0**	**99.8**	98.6	76.5	**93.5**	**93.3**	**100.0**	**95.3**	**90.3**	**99.9**

注：加粗字体表示各列最优结果。

表3　SD视角对比实验结果

Table 3　Comparison of experimental results from the SD view

/%

模型	mAP_{50}	AP_{50}														
		枪	刀	扳手	钳子	剪刀	打火机	电池	棒球棍	刀片	锯条	烟花	榔头	螺丝刀	飞镖	压力容器
YOLO11（Ultralytics，2025）	78.2	90.3	55.3	88.4	86.7	53.0	60.8	95.7	99.5	57.2	63.8	**89.2**	97.1	72.2	70.8	93.3
DN-DETR（Li等，2022）	65.2	88.3	34.3	86.5	80.8	46.1	42.9	83.0	96.2	33.2	46.8	74.7	95.2	43.6	36.3	90.3
RT-DETR（Zhao等，2024）	77.1	92.4	61.5	91.2	87.4	43.3	55.9	94.2	**100.0**	55.3	59.3	84.1	**98.8**	71.6	66.2	**96.0**
Trans2Ray（Meng等，2024）	58.8	85.2	38.2	72.2	73.4	39.4	22.7	77.5	98.0	20.9	38.0	78.3	88.2	50.6	9.4	89.7
YOLO_multi（Isaac-Medina等，2021）	66.9	52.6	51.8	85.9	85.0	33.0	42.7	86.0	98.0	41.8	52.0	75.6	79.0	61.5	68.4	83.0
本文	**83.9**	**97.1**	**68.1**	**96.9**	**95.5**	**68.2**	**66.0**	**97.9**	**100.0**	**62.0**	**71.1**	85.3	98.3	**78.0**	**77.6**	95.9

注：加粗字体表示各列最优结果。

表4　与最新模型的性能对比

Table 4　Comparison with the latest model

模型	$mAP_{OL\ 50}$/%	$mAP_{OL\ 50:95}$/%	$mAP_{SD\ 50}$/%	$mAP_{SD\ 50:95}$/%	FPS/(帧/s)
Jia等人(2025)	92.1	67.6	78.2	51.7	14.6
本文	**93.4**	**78.3**	**83.9**	**63.1**	**91.0**

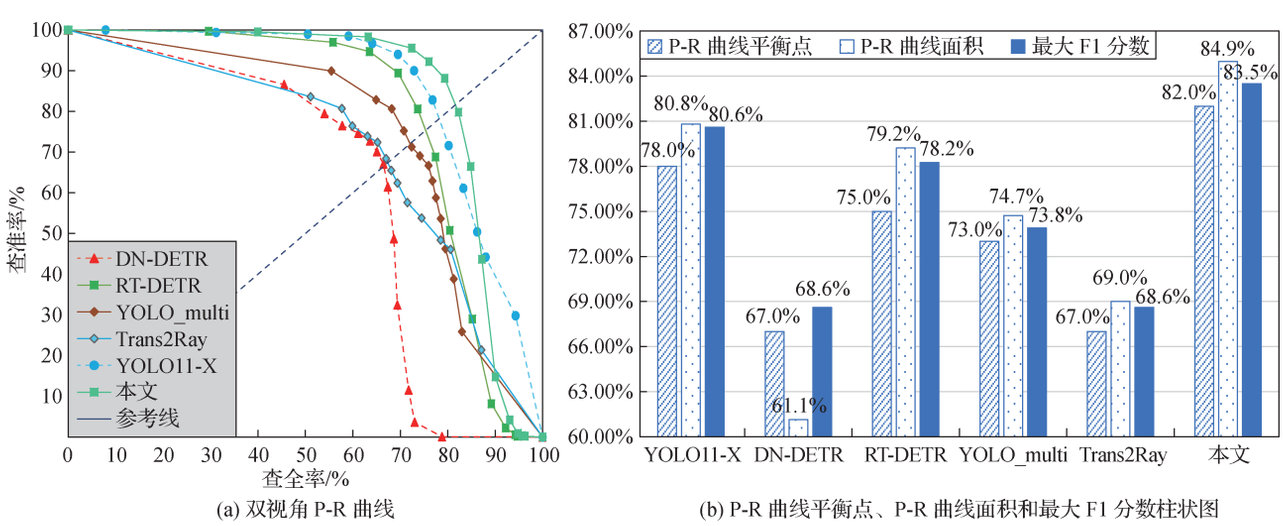
注：加粗字体表示各列最优结果。

SD视角相比OL视角提升更为明显的现象，本文结合mAP_{50}的计算原理分析如下，由于SD视角遮挡多，其特征质量受到一定影响，导致许多真正例的置信度得分偏低。这使得它们在排名列表中与假正例检测框相混淆，获得了较低的初始AP值。对此，一方面，通过本文设计的特征融合模块，引入了OL视角中的违禁品特征，有效增强了SD视角特征的判别力，使原有置信度低的真正例的得分获得了提升。这种提升使它们在排名中高于更多的假正例，因此SD视角下的mAP_{50}值提升幅度相比更大。另一方面，三维锚框引导定位融合模块引入了几何约束。在SD视角视觉信息遮挡严重时，基于双视角几何一致性生成的三维锚框能够提供明确的位置引导，不仅校正了定位偏差，更为特征的准确提取与判别提供了可靠的空间基准。相对而言，OL视角遮挡较少，基线特征质量本身较高，其大多数真正例的置信度已排在多数假正例之前。特征补充带来的提升虽然也提高了置信度，但在全局排名中的相对位次变化相对有限，因此mAP_{50}提升幅度不如SD视角。因此，尽管两个视角都受益于特征融合与几何约束的协同作用，但由于基线特征的差异，最终在数值上表现为SD视角的提升幅度更为明显。

总体而言，模型的整体判别能力得到了有效增强，如图11中双视角联合统计结果所示。从图11(a)的P-R曲线可以看出，本文方法的平衡点优于对比模型，表明其具有更好的综合检测性能。图11(b)所示的P-R曲线下面积和F1分数柱状图表明，本文方法P-R曲线下面积为84.9%，相较对比模型分别提高了4.1%、23.8%、5.7%、10.2%和15.9%；本文方法取得83.5%的最高F1分数，高于基线模型5.2%。结果表明，本文方法在保持检测精度的同时进一步降低了虚警率，提升了违禁品在双视角下同时被检出的正确率，较好地满足了实际安检场景对检测模型的性能要求。

2.3　消融实验

为了进一步探究本文方法的有效性，本节设计了消融实验，分析不同模块组件对模型性能的影响，具体结果如表5—表7和图12所示。



(a) 双视角P-R曲线　　(b) P-R曲线平衡点、P-R曲线面积和最大F1分数柱状图

图11　双视角联合统计结果

Fig. 11　Dual-view joint statistical results ((a) dual-view precision-recall curves; (b) bar chart of break even point, area under the precision-recall curve and maximum F1-score)

为验证本文提出的一致性数据增强策略的有效性，并探究数据增强强度对模型性能的影响，设计了多组消融实验，结果如表5所示。

表5　不同数据增强策略的性能对比

Table 5　Ablation study on different data augmentation strategies

模型	增强概率p	增强策略	视角OL/%	视角SD/%
基线	0.5	独立数据增强	91.3	77.1
本文	0.5	独立数据增强	78.3	63.4
本文	0	无增强	90.7	78.9
本文	0.5	一致性数据增强	**93.4**	**83.9**
本文	1.0	一致性数据增强	77.8	62.5

注：加粗字体表示各列最优结果。

首先，分析独立数据增强策略的局限性。在相同增强概率下，当本文模型与基线模型均采用独立数据增强时，本文模型性能出现明显下滑，OL视角mAP_{50}从91.3%跌至78.3%，SD视角从77.1%跌至63.4%。印证了独立的数据增强破坏了双视角几何对应关系，从而在训练中引入了误导性的噪声，不仅干扰了编码器中融合特征的学习，并且导致解码器的三维锚框引导机制失效。其次，验证一致性数据增强策略的有效性。保持$p=0.5$不变，将本文模型的增强策略改为一致性数据增强策略后，模型性能达到了最优水平，OL视角从78.3%提升至93.4%，SD视角从63.4%提升至83.9%。证明了一致性增强策略通过保留几何先验，有效保障了双视角多尺寸特征加权融合模块与三维锚框引导定位融合模块的训练稳定性。最后，分析增强概率p的影响。在采用一致性增强策略的前提下，性能随p变化呈现先升后降的趋势，相较于无增强的基础表现，适度的增强在引入数据多样性的同时维持了特征稳定性，使模型达到最优。而当p增至1.0时，即所有样本均增强时，性能衰退。表明过度的增强会破坏数据分布，从而损害模型的收敛能力。

为验证不同模块对于实验结果的影响，表6展示了以基线模型为参照的消融实验结果。

其在OL与SD两个视角上的mAP_{50}分别为91.3%和77.1%。在基线模型基础上，当仅集成了双视角多尺寸特征加权融合模块时，模型性能得到了提升，OL视角提升1.4%，SD视角提升了8.0%。这一现象表明，该模块利用多尺寸滑窗机制，有效增强了模型对不同尺度目标的感知，能够自适应地为双视角提供不同感受野大小的融合特征信息，从而改善了双视角的识别能力。同时，单独引入三维锚框引导定位融合模块亦对两个视角均有助益，OL视角提升1.7%，SD视角提升3.7%，验证了引入三维锚框作为几何先验的有效性，表明通过施加约束，模型能够利用生成的三维锚框提供明确的位置引导，不仅辅助解码器校正了遮挡场景下的定位偏差，更为模糊特征的提取提供了空间基准。当两个模块协同工作时，模型在OL视角上达到了93.4%的最佳性能，同时SD视角mAP_{50}达到83.9%。综合来看，特征层面的多尺寸感知与解码层面的几何约束具有互补性，二者协同作用有效提升了模型检测精度。

表6　不同模块的消融结果

Table 6　Ablation results of different modules

/%

模型	视角OL	视角SD
基线模型	91.3	77.1
+ 双视角多尺寸特征加权融合模块	92.7	**85.1**
+ 三维锚框引导定位融合模块	93.0	80.8
+ 双视角多尺寸特征加权融合模块 & 三维锚框引导定位融合模块	**93.4**	83.9

注：加粗字体表示各列最优结果；&表示同时添加多个模块。

然而，一个值得深入探究的现象是，两个模块协同工作时，模型在SD视角上的增益略低于单独添加双视角多尺寸特征加权融合模块的增益。

为了探究其根本原因，本文对融合模块的窗口尺寸进行了进一步的消融，实验均在开启三维锚框引导定位融合模块的基础上进行，结果如表7所示。表7的数据揭示了两个视角对融合窗口大小的不同需求。对于SD视角，尺寸为1的窗口取得了84.3%的最佳性能，而随着窗口尺寸增大到10，其性能反而下降至80.2%。与此相反，OL视角的性能随着窗口尺寸的增大而平稳提升，在窗口尺寸为10时达到93.0%的峰值。这一趋势表明，SD视角由于其自身特征信息不足，在一定程度上依赖与OL视角间的高精度、细粒度的特征对齐，而这正是尺寸为1的窗口所提供的。更大的窗口如尺寸10，在进行窗口平均

时会引入相邻的背景特征，从而稀释了SD视角所需的精确对齐信号，导致性能下降。相反，信息丰富的OL视角其基线已达91.3%，它的主要挑战并非识别模糊特征，而是在违禁品与复杂背景物体堆叠时进行确认与区分。尺寸为10的窗口为其提供了更广阔的双视角感受野，使其能够聚合更大区域的特征信息以辅助判别，性能也随之提升至93.0%。最终，本文采用多尺寸联合的策略。如表7最后一行所示，该策略在OL视角取得了最高精度，超越了单一大窗口的表现，说明多尺寸的引入进一步丰富了特征表达。同时在SD视角上达到了83.9%，虽然略低于单窗口1的极值，但优于其余单一窗口表现。这种对感受野需求的差异也解释了表6的性能权衡，三维锚框引导定位融合模块引入的全局先验在满足OL视角对大感受野偏好的同时，在一定程度上抑制了SD视角对局部细粒度特征的灵活表达，导致其性能从85.1%的峰值回落。综上所述，双视角多尺寸特征加权融合模块中采用的多尺寸窗口设计，有益于模型自适应地平衡并满足两个视角在大感受野与局部特征感知间的需求。

表7　多尺寸滑窗机制中的窗口尺寸影响

Table 7　Influence of window size in the multiscale sliding window mechanism

/%

窗口尺寸	视角OL	视角SD
1	92.4	**84.3**
5	92.6	81.5
10	93.0	80.2
1,5,10	**93.4**	83.9

注：加粗字体表示各列最优结果。

为探究式(4)中分类损失权重α和定位损失权重β对模型性能的影响并确定最佳权重，本文采用控制变量法进行了敏感性分析，结果如图12所示。

观察各子图可知，敏感性曲线均呈现出先升后降的单峰趋势，并在$\alpha=2$，$\beta=5$时达到全局最优。具体而言，对于OL视角，β的变化引起的性能波动大于α，表明OL视角对定位损失权重β的变化更为敏感。这是因为OL视角成像相对清晰，其性能瓶颈主要在于物体堆叠导致的定位模糊。因此通过设定β大于α以加强定位约束，能有效利用三维锚框引导模型实现更精确的定位。相反，当调节α时，对SD视角的性能影响更大，说明SD视角对分类损失权重α更为敏感。这是由于SD视角受遮挡与背景干扰严重，包含更多难以分类的样本，特征判别性较弱。增大α能够强化Focal Loss对难样本的挖掘，驱动模型学习更具判别性的局部特征。但是当α过大时，分类损失的主导可能导致梯度优化方向偏离，抑制模型收敛。因此，合理设置α值以平衡分类损失，有利于提升SD视角的检测精度。

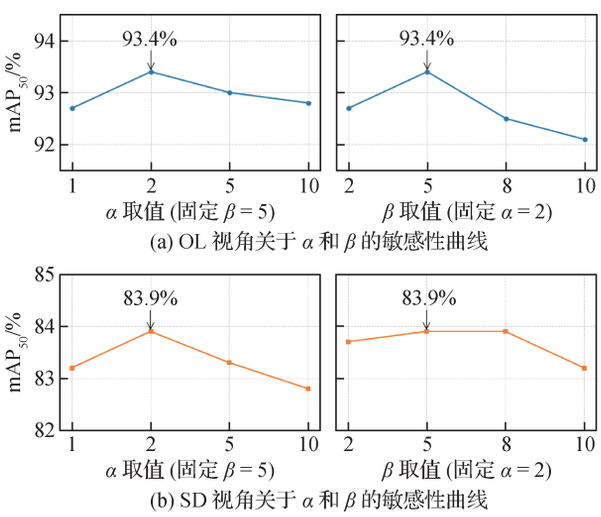


图12　超参数α和β在双视角下的敏感性分析

Fig. 12　Sensitivity analysis of hyperparameters α and β across dual views ((a) sensitivity curves of α and β for the OL view; (b) sensitivity curves of α and β for the SD view)

2.4　计算效率分析

在同一硬件环境下，对本文方法、RT-DETR与YOLO11-X的参数量(Params)、浮点运算量(floating point operations，FLOPs)及帧率(frames per second，FPS)进行了统计比较，具体结果如表8所示。观察表8，相较于基线模型，本文方法的参数量与计算量

表8　不同模型计算效率对比

Table 8　Comparison of computational efficiency across different models

模型	mAP_{50}/%		参数量/M	浮点运算量/G	FPS/(帧/s)
	OL	SD			
YOLO11-X	91.2	78.2	56.84	195.50	111.11
RT-DETR-r50	91.3	77.1	**40.89**	**135.86**	**136.15**
本文	**93.4**	**83.9**	43.90	142.61	91.03

注：加粗字体表示各列最优结果。

略有提升，推理速度为91.03帧/s。这一现象源于模型双视角多尺寸特征加权融合模块与三维锚框引导定位融合模块的引入，通过适度增加计算负担，实现了检测效能的增幅。相较于模型YOLO11-X，本文方法不仅实现了更高的检测精度，OL和SD视角的mAP_{50}提升了2.2%和5.7%，同时，参数量与FLOPs更低，展现了更优的计算效率。

表9进一步量化了各个模块的开销。双视角多尺寸特征加权融合模块是主要的计算开销来源，其引入的3.96 G浮点运算量将帧率从136.15降低至102.42。这一下降的主要原因是该模块为实现多尺寸融合，在不同尺寸的窗口上均执行了计算交叉注意力的操作，但也正是该模块在上一节中证实为SD视角带来了8.0%性能提升。相比之下，三维锚框引导定位融合模块的设计则更为轻量，其FLOPs开销仅为2.57 G，对帧率的影响相对可控。此外，考虑到实际安检场景中传送带速度受国标（国家市场监督管理总局和中国国家标准化管理委员会，2018）限制，91帧/s的推理速度满足实时检测要求，表明本文以适度增加计算负载换取精度提升的策略是合理的，以可接受的计算成本，有效改善了两个视角性能，并促进了双视角整体性能的均衡。

表9　不同模块对计算效率的影响

Table 9　Comparison of computational efficiency across different modules

模型	参数量/M	浮点运算量/G	FPS/(帧/s)
RT-DETR-r50	**40.89**	**135.86**	**136.15**
+ 双视角多尺寸特征加权融合模块	41.99	139.82	102.42
+ 三维锚框引导定位融合模块	42.81	138.43	125.80
+ 双视角多尺寸特征加权融合模块 & 三维锚框引导定位融合模块	43.90	142.61	91.03

注：加粗字体表示各列最优结果；&表示同时添加多个模块。

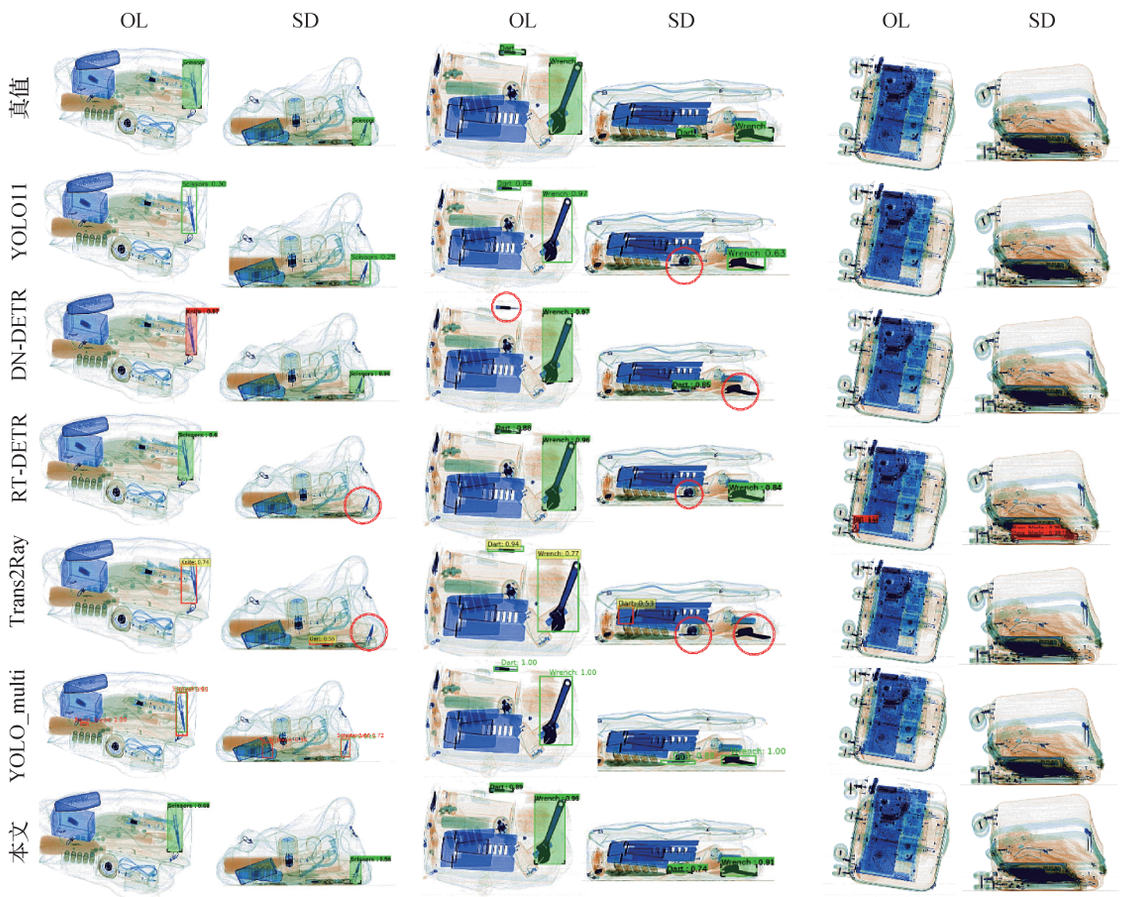
2.5　可视化分析

为更直观地观察模型的检测效果，在测试集中选取3组代表性的图像并预测出违禁品的定位框。具体模型可视化结果和定位框如图13所示。其中绿色框代表正确检测，红色框代表虚警，红色空心圆指示漏检目标。

图13(a)展示了模型对于复杂背景和形状旋转的违禁品的识别能力，图中违禁品剪刀在SD视角旋转至特征较难提取的角度，且OL和SD视角均存在大量杂物。观察发现仅本文方法和YOLO11模型实现零虚警和零漏警。DN-DETR模型在SD视角可以正确检出，但是在OL视角下错误识别为刀。RT-DETR模型虽然在OL视角可以正确检出，但是在SD视角下出现了漏警，未能正确检出违禁品。对于双视角模型YOLO_multi，虽然能正确检出双视角中的违禁品，但是虚警率较高；双视角模型Trans2Ray在OL视角和SD视角均存在虚警和漏警。分析结果表明，本文提出的双视角多尺寸特征加权融合模块与三维锚框引导定位融合模块实现了双视角特征的有效互补，利用清晰视角获取的高质量特征信息，引导杂乱环境下受限视角的识别，不仅提升了检测准确率，而且降低了虚警率。

图13(b)展示了模型对于小物体、遮挡背景和两种违禁品的识别能力，图中违禁品类别为飞镖和扳手。观察结果，在OL视角，所有模型对扳手的检测均表现良好。本文方法和YOLO_multi通过双视角间的空间先验有效判读违禁品，在两视角均正确检出两种违禁品。其余各模型均存在不同程度漏警。分析结果表明，本文提出的双视角多尺寸特征加权融合模块通过多尺寸滑窗机制增强了特征提取能力。该模块动态调整关注区域大小，提升了模型对不同尺寸违禁品的检测适应性。同时，三维锚框引导定位融合模块发挥了辅助作用，利用几何约束引导解码器关注小物体的位置，有效防止了特征微弱的小物体被背景噪声淹没，从而实现了准确识别。

图13(c)为负样本组，即无违禁品。从OL视角可以明显观察到笔记本电脑，在SD视角中对其他物体造成了遮挡，容易引起虚警。在置信度阈值均设置为0.2的情况下，可以观察到，基线模型RT-DETR由于笔记本电脑的遮挡，在双视角中误检了多个违禁品。分析结果表明，本文提出的三维锚框引导定位融合模块施加几何约束，有效抑制了解码器对背



(a) 复杂背景样本　(b) 双违禁品和小尺寸违禁品样本　(c) 负样本

图 13　各模型检测结果可视化

Fig. 13　Visualization of detection results for different models ((a) complex background samples; (b) samples with two prohibited items and small-sized prohibited items; (c) negative samples)

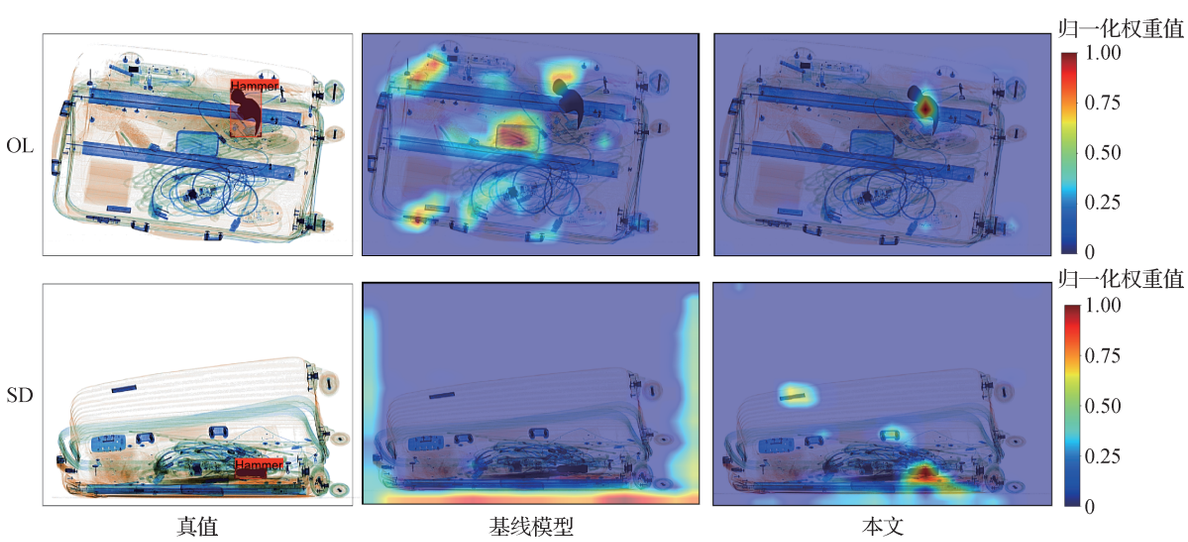


图 14　基线模型与改进模型编码器特征图可视化对比

Fig. 14　Comparative visualization of encoder feature maps between baseline and improved models

景干扰的响应，降低了检测中的虚警率。

为验证双视角多尺寸特征加权融合模块对编码器全局特征融合的优化效果，对 RT-DETR 编码器经过 AIFI 模块后的特征图可视化，如图 14 所示，图中

违禁品为锤子，伪彩色图像红色框内为违禁品，热力图示数显示归一化关注权重大小，其中红色区域表示较高权重值，蓝色区域表示较低权重值。对比基线模型和本文方法的权重可视化，可以发现，在OL视角中，加入双视角多尺寸特征加权融合模块后，编码器的注意力更集中关注违禁品区域，在SD视角中，本文方法相比基线模型对于违禁品的关注更加清晰准确。表明通过双视角多尺寸特征加权融合模块进行双视角特征融合后，编码器利用融合信息有效补充特征图的违禁品注意力权重，提升编码器对于违禁品特征的感知。进而提升复杂背景下违禁品检测性能。

3 结 论

本文设计了一种融合三维几何约束的RT-DETR双视角违禁品检测网络，以应对违禁品尺寸多变及分布杂乱的挑战，提高检测精度。通过设计双视角多尺寸特征加权融合模块，采用多尺寸窗口交叉注意力机制，增强了模型对不同尺寸违禁品的识别能力。同时，提出三维锚框引导定位融合模块，通过引入三维几何先验并施加几何约束，为解码器提供了明确的位置引导，提升定位精度。此外，本文所设计的双视角一致性数据增强策略，通过对图像对施加同步变换，在保持其空间对应关系的同时提升了模型的泛化能力与鲁棒性。实验结果显示，本文方法提升了双视角的检测精度，特别是SD视角相比OL视角获得了更明显的性能提升，验证了模型利用优势视角补偿弱视角特征以克服复杂遮挡的有效性。但是，当前模型仍存在架构复杂、参数量大和训练耗时等问题，这主要由双通道结构引入的额外计算开销导致。未来研究将重点探索轻量化的双视角特征融合方法，开发高效的双视角协同优化算法，以实现计算效率与检测性能的平衡。

参考文献(References)

Baştan M. 2015. Multi-view object detection in dual-energy X-ray images. Machine Vision and Applications, 26(7): 1045-1060 [DOI: 10.1007/s00138-015-0706-x]

Chavaillaz A, Schwaninger A, Michel S and Sauer J. 2019. Expertise, automation and trust in X-ray screening of cabin baggage. Frontiers in Psychology, 10: #256 [DOI: 10.3389/fpsyg.2019.00256]

Hu J L, Zhou M, and Shen F. 2024. Improved detection algorithm of RTDETR for UAV small target. Computer Engineering and Applications, 60(20): 198-206 (胡佳乐, 周敏, 申飞. 2024. 面向无人机小目标的RTDETR改进检测算法. 计算机工程与应用, 60(20): 198-206) [DOI: 10.3778/j.issn.1002-8331.2404-0114]

Huegli D, Merks S and Schwaninger A. 2020. Automation reliability, human-machine system performance, and operator compliance: a study with airport security screeners supported by automated explosives detection systems for cabin baggage screening. Applied Ergonomics, 86: #103094 [DOI: 10.1016/j.apergo.2020.103094]

Isaac-Medina B K S, Willcocks C G and Breckon T P. 2021. Multi-view object detection using epipolar constraints within cluttered X-ray security imagery//Proceedings of the 25th International Conference on Pattern Recognition (ICPR). Milan, Italy: IEEE: 9889-9896 [DOI: 10.1109/ICPR48806.2021.9413007]

Isaac-Medina B K S, Willcocks C G and Breckon T P. 2022. Multi-view vision transformers for object detection//Proceedings of the 26th International Conference on Pattern Recognition (ICPR). Montreal, Canada: IEEE: 4678-4684 [DOI: 10.1109/ICPR56361.2022.9956443]

Jia Z H, Ma B W and Chen D Y. 2025. Delving into end-to-end dual-view prohibited item detection for security inspection system. Computers, Materials and Continua, 85(2): 2873-2891 [DOI: 10.32604/cmc.2025.067460]

Jin T and Hu P Y. 2025. HK-DETR: improved knife-holding dangerous behavior detection algorithm based on RT-DETR. Journal of Image and Graphics, 30(4): 1027-1040 (金涛, 胡配雨. 2025. 改进实时目标检测Transformer的持刀危险行为检测算法. 中国图象图形学报, 30(4): 1027-1040) [DOI: 10.11834/jig.240295]

Jin X P, Gao F, Shi X C and Dong J Y. 2025. Gated cross-modal aggregation network for multi-source remote sensing data classification. Journal of Image and Graphics, 30(3): 883-894 (金学鹏, 高峰, 石晓晨, 董军宇. 2025. 针对多源遥感图像分类的门控跨模态聚合网络. 中国图象图形学报, 30(3): 883-894) [DOI: 10.11834/jig.240359]

Kuhn H W. 1955. The Hungarian method for the assignment problem. Naval Research Logistics Quarterly, 2(1/2): 83-97 [DOI: 10.1002/nav.3800020109]

Lin T Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al. 2014. Microsoft COCO: common objects in context//Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland: Springer: 740-755 [DOI: 10.1007/978-3-319-10602-1_48]

Li F, Zhang H, Liu S L, Guo J, Ni L M and Zhang L. 2022. Dn-detr: Accelerate detr training by introducing query denoising//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 13619-13627 [DOI: 10.1109/CVPR52688.2022.01325]

Liu S L, Li F, Zhang H, Yang X, Qi X B, Su H, et al. 2022. DAB-DETr: dynamic anchor boxes are better queries for DETR//Proceedings of the 10th International Conference on Learning Representations. [s.l.]: ICLR: 1-20

Ma B W, Jia T, Li M Y, Wu S S, Wang H and Chen D Y. 2024. Toward dual-view X-ray baggage inspection: a large-scale benchmark and adaptive hierarchical cross refinement for prohibited item discovery. IEEE Transactions on Information Forensics and Security, 19: 3866-3878 [DOI: 10.1109/tifs.2024.3372797]

Meng X L, Feng H, Ren Y, Zhang H G, Zou W D and Ouyang X Y. 2024. Transformer-based dual-view X-ray security inspection image analysis. Engineering Applications of Artificial Intelligence, 138: #109382 [DOI: 10.1016/j.engappai.2024.109382]

Mery D, Riffo V, Zuccar I and Pieringer C. 2013. Automated X-ray object recognition using an efficient search algorithm in multiple views//Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Portland, USA: IEEE: 368-374 [DOI: 10.1109/CVPRW.2013.62]

Song X G, Tan Y P, Guo F Q, Lu X F and Hei X H. 2025. Cross-modal feature fusion and detail-enhanced RGB-D salient object detection. Journal of Image and Graphics, 30(12): 3838-3854 (宋霄罡, 谭裕平, 郭富强, 鲁晓锋, 黑新宏. 2025. 跨模态特征融合与细节信息增强的RGB-D显著目标检测. 中国图象图形学报, 30(12): 3838-3854) [DOI: 10.11834/jig.240653]

State Administration for Market Regulation, Standardization Administration of the People's Republic of China. 2018. GB 15208.2-2018 Micro-dose X-ray security inspection system—part 2: transmission baggage security inspection system. Beijing: Standards Press of China: 4 (国家市场监督管理总局, 中国国家标准化管理委员会. 2018. GB 15208.2-2018 微剂量X射线安全检查设备 第2部分: 透射式行包安全检查设备. 北京: 中国标准出版社: 4)

Steitz J M O, Saeedan F and Roth S. 2018. Multi-view X-ray R-CNN//Proceedings of the 40th German Conference on Pattern Recognition. Stuttgart, German: Springer: 153-168 [DOI: 10.1007/978-3-030-12939-2_12]

Sun B Y, Zhang W Y, Chen Q X and Zhang H G. 2025. Detection of prohibited items in dual-view X-ray security inspection images based on DINO model//Proceedings of the 8th International Conference on Computer Science and Artificial Intelligence. Beijing, China: ACM: 55-61 [DOI: 10.1145/3709026.3709065]

Tao R S, Wang H Y, Guo Y Z, Chen H R, Zhang L, Liu X L, et al. 2025. Dual-view X-ray detection: can AI detect prohibited items from dual-view X-ray images like humans?//Proceedings of 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 10338-10347 [DOI: 10.1109/CVPR52734.2025.00967]

Ultralytics. 2025. YOLOv11: real-time object detection model [EB/OL]. [2025-04-02]. https://github.com/ultralytics/ultralytics.

Wei S and Yang W L. 2025. Visible-infrared person re-identification algorithm integrating structural and visual features. Journal of Image and Graphics, 30(10): 3335-3345 (魏思, 杨文璐. 2025. 融合结构与视觉特征的可见光—红外行人重识别. 中国图象图形学报, 30(10): 3335-3345) [DOI: 10.11834/jig.240600]

Wu M D, Yi F F, Zhang H G, Ouyang X Y and Yang J F. 2022. Dual-ray: dual-view X-ray security inspection benchmark and fusion detection framework//Proceedings of the 5th Chinese Conference on Pattern Recognition and Computer Vision. Shenzhen, China: Springer: 721-734 [DOI: 10.1007/978-3-031-18916-6_57]

Yu M Y and Liu X Y. 2026. Ship detection method based on multimodal visible and infrared image fusion. Computer Engineering, 52(6): 278-287 (于梦源, 刘向阳. 2026. 基于多模态可见光和红外图像融合的船舶检测方法. 计算机工程, 52(6): 278-287) [DOI: 10.19678/j.issn.1000-3428.0070436]

Zhao Y A, Lyu W Y, Xu S L, Wei J M, Wang G Z, Dang Q Q, et al. 2024. DETRs beat YOLOs on real-time object detection//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 16965-16974 [DOI: 10.1109/CVPR52733.2024.01605]

作者简介

韩萍, 女, 教授, 主要研究方向为智能图像处理。
E-mail: hanpingcauc@163.com
白海峰, 男, 硕士研究生, 主要研究方向为图像处理和深度学习。E-mail: whitebhf@163.com
罗思宇, 男, 硕士研究生, 主要研究方向为图像处理和深度学习。E-mail: syluofoden47@163.com