中国光学（中英文） Chinese Optics 核心期刊

结合YOLOv11改进的双视角X光图像违禁品检测

吴海滨 刘文柏 袁鹏飞 王爱丽

Improved prohibited item detection in double-view X-ray images combined with YOLOv11

WU Hai-bin, LIU Wen-bai, YUAN Peng-fei, WANG Ai-li

引用本文:
吴海滨, 刘文柏, 袁鹏飞, 等. 结合YOLOv11改进的双视角X光图像违禁品检测[J]. 中国光学, 2026, 19(6): 1−14. doi: 10.3724/CO.2026−0062
WU Hai-bin, LIU Wen-bai, YUAN Peng-fei, *et al*. Improved prohibited item detection in double-view X-ray images combined with YOLOv11[J]. *Chinese Optics*, 2026, 19(6): 1-14. doi: 10.3724/CO.2026-0062

在线阅读 View online: https://doi.org/10.3724/CO.2026−0062

您可能感兴趣的其他文章

Articles you may be interested in

基于跨域交互注意力和对比学习引导的红外与可见光图像融合

Infrared and visible image fusion guided by cross−domain interactive attention and contrastive learning

中国光学（中英文）. 2025, 18(2): 317 https://doi.org/10.37188/CO.2024−0147

轻量型YOLOv5s车载红外图像目标检测

Lightweight YOLOv5s vehicle infrared image target detection

中国光学（中英文）. 2023, 16(5): 1045 https://doi.org/10.37188/CO.2022−0254

冷轧钢表面与内部缺陷检测研究

Detection of surface and internal defects in cold rolled steel

中国光学（中英文）. 2024, 17(4): 823 https://doi.org/10.37188/CO.2023−0189

多尺度注意力融合的图像超分辨率重建

Image super−resolution reconstruction with multi−scale attention fusion

中国光学（中英文）. 2023, 16(5): 1034 https://doi.org/10.37188/CO.2023−0020

基于图像块分解融合的水下标定图像增强

Underwater calibration image enhancement based on image block decomposition and fusion

中国光学（中英文）. 2024, 17(4): 810 https://doi.org/10.37188/CO.2023−0218

面向地物混杂背景的偏振光谱图像融合方法

Polarization spectral image fusion method for hybrid backgrounds of ground objects

中国光学（中英文）. 2024, 17(5): 1098 https://doi.org/10.37188/CO.2023−0185

文章编号 2097-1842(2026)06-0001-14

结合 YOLOv11 改进的双视角 X 光图像违禁品检测

吴海滨, 刘文柏, 袁鹏飞, 王爱丽*
（哈尔滨理工大学 黑龙江省激光光谱技术及应用重点实验室, 黑龙江 哈尔滨 150080）

摘要: 针对现有双视角 X 光安检图像违禁品检测方法在跨视角特征融合过程中自适应性不足、互补信息利用不充分的问题, 本文提出一种结合 YOLOv11 改进的双视角融合检测方法（Dual View Fusion combined with YOLOv11, DVF-YOLOv11）。该算法采用参数共享的双分支 YOLOv11 骨干网络分别提取俯视图与侧视图的多尺度特征; 设计跨视角注意力融合模块（Cross-View Attention Fusion, CVAF）, 通过通道注意力与空间注意力的级联机制实现双视角特征的自适应增强; 采用自适应权重预测网络动态调整各视角融合权重, 结合通道压缩卷积形成双路融合策略; 设计由特征保留损失、互补性损失和权重平衡损失组成的联合损失函数引导融合学习。在 DvXray 数据集上, 本文方法的 mAP50 达到 94.02%, mAP50-95 达到 79.41%, 较俯视图单视角分别提升 2.99% 和 5.29%。实验结果表明本文方法能够提升双视角 X 光安检图像中违禁品检测的精度与鲁棒性。

关 键 词: X 光安检; YOLOv11; 双视角融合; 自适应权重融合
中图分类号: TP391.4 **文献标志码:** A **doi:** 10.3724/CO.2026-0062 **CSTR:** 32171.14.CO.2026-0062

Improved prohibited item detection in double-view X-ray images combined with YOLOv11

WU Hai-bin, LIU Wen-bai, YUAN Peng-fei, WANG Ai-li*
(*Heilongjiang Province Key Laboratory of Laser Spectroscopy Technology and Application*, *Harbin University of Science and Technology*, *Harbin* 150080, *China*)
* *Corresponding author*, *E-mail*: *aili*925@*hrbust.edu.cn*

Abstract: To address the issues of insufficient adaptability and inadequate utilization of complementary information in the cross-view feature fusion process of existing prohibited item detection methods for dual-view X-ray security inspection image, we propose an improved dual-view fusion detection method combined with YOLOv11 (Dual View Fusion combined with YOLOv11, DVF-YOLOv11). The proposed method employs a parameter-shared dual-branch YOLOv11 backbone network to extract multi-scale features from the overlook-view and side-view images, respectively. A Cross-View Attention Fusion (CVAF) module is designed to adaptively enhance dual-view features through a cascaded mechanism of channel attention and spatial attention. An adaptive weight prediction network is introduced to dynamically adjust the fusion weights of each view, and is combined with channel compression convolution to form a dual-path fusion strategy. A

收稿日期: xxxx-xx-xx; **修订日期:** xxxx-xx-xx
基金项目: 黑龙江省自然科学基金项目（No. LH2023F034）
Supported by Natural Science Foundation of Heilongjiang Province of China (No. LH2023F034)

joint loss function composed of feature preservation loss, complementarity loss, and weight balance loss is further designed to guide the fusion learning process. On the DvXray dataset, the proposed method achieves an mAP50 of 94.02% and an mAP50-95 of 79.41%, improving by 2.99% and 5.29%, respectively, over the single overlook-view baseline. Experimental results demonstrate that the proposed method improves the accuracy and robustness of prohibited item detection in dual-view X-ray security inspection images.

Key words: X-ray security inspection; YOLOv11; dual-view fusion; adaptive weight fusion

1 引 言

随着公共交通运输规模的持续增长，X 光安检技术在机场、火车站、地铁站等公共场所的行李安全检查中发挥着日益重要的作用。然而，X 光安检图像具有对比度低、物品透明叠加、背景干扰复杂等特点，使得违禁品检测面临物品重叠遮挡、小目标漏检等挑战。近年来，基于深度学习的方法在图像检测、重建与分析领域取得了显著进展[1-2]；推动了 X 光安检图像违禁品检测的发展，有效提升了违禁品自动识别的效率与准确性[3]。

在单视角 X 光安检方面，研究人员从多个角度对检测算法进行了改进。Wei 等[4] 针对物品重叠遮挡导致的目标特征退化问题，提出了去遮挡注意力模块（De-occlusion Attention Module, DOAM），通过融合边缘信息和区域信息生成注意力图，增强了对遮挡目标的检测能力。Tao 等[5]受人眼视觉抑制机制启发，设计了侧抑制模块（Lateral Inhibition Module, LIM），通过双向传播抑制噪声并强化目标边界表达，在 HiXray 数据集上取得了优于主流检测方法的精度。Zhu 等[6] 提出了基于双流 Transformer 的频率感知检测网络（Frequency-aware Dual-stream Transformers, FDTNet），利用频域信息增强违禁品的特征表示，有效改善了背景杂乱和目标遮挡条件下的检测效果。刘建军等[7] 提出了融合 Swin-Transformer 和多尺度可变形注意力的 YOLO-STM 模型，通过全局上下文感知能力提升了安检图像中多尺度违禁品的识别精度。YOLO 系列单阶段检测算法经过不断发展与完善，凭借速度与精度的良好均衡在安检场景中获得了广泛应用[8]，其中 YOLOv11 引入了 C3k2 和 C2PSA 等模块，在多尺度特征提取与上下文信息提取方面表现出更强的能力[9]。以上研究有效推动了单视角 X 光安检技术的发展，但由于 X 光沿固定方向穿透行李成像时，不同密度和厚度的物品在投影平面上会发生重叠，部分目标的关键形态信息可能因遮挡而弱化，这些问题在单一视角条件下难以完全消除。

随着双视角 X 光安检设备在实际安检场景中的逐步普及，引入双视角互补信息成为进一步提升检测性能的有效途径。双视角 X 射线安检设备通常从俯视（Overlook, OL）和侧视（Side, SD）两个相互正交的方向对行李进行透射成像，分别获取物品的水平投影和垂直投影信息，从而为同一目标提供两个互补的观测角度。

在双视角 X 光检测方面，已有研究人员围绕双视角信息融合开展了一系列探索。Steitz 等[10] 提出多视角 R-CNN，利用 X 光成像设备的已知几何参数将双视角二维特征映射至公共三维特征空间进行融合检测，相比单视角基线将 mAP 从 91.2% 提升至 95.6%，但其依赖 Faster R-CNN 两阶段检测框架，推理效率受限。Tuli 等[11] 系统比较了单视角、双视角和多视角设置下的检测性能差异，验证了双视角融合带来的性能增益。Wu 等[12] 构建了 Dualray 数据集并提出基于 YOLOv4 的双视角融合检测框架，利用通道注意力和空间注意力对辅助视角特征进行筛选后与主视角特征进行拼接，初步验证了单阶段检测模型在双视角场景中的适用性。Meng 等[13] 基于 DETR 框架设计了 Trans2Ray 模型，利用 Transformer 的注意力机制分别在全局和局部层面实现双视角特征的交互与融合，较单视角 DETR 基线提升了 1.8%。Hong 等[14] 提出双视角注意力引导网络，通过频域交互模块、双视角层次增强模块和卷积引导融合模块实现系统化的跨视角特征交互，但其将任务定义为多标签分类问题而非目标检测问题。Tao 等[15] 构建了包含 35 万实例的大规模 LDXray 数据集并提出辅助视角增强网络，以主辅视角协同的方式模拟安检员的双视角判图逻辑，在 LDXray 数据集上取得了优于单视角基线的检测精度。以

上研究表明双视角融合能够有效提升检测精度，但现有方法在双视角特征融合的自适应性方面仍存在改进空间，部分方法采用固定的特征拼接或求和策略，未能根据输入内容动态调节两个视角的贡献权重；部分方法虽引入了注意力机制，但仅从通道或空间单一维度进行特征增强，缺乏对两个维度的联合处理，且融合过程中缺少针对性的损失函数加以引导，导致不同视角的互补信息未能被充分利用。

因此，针对现有双视角 X 光安检图像违禁品检测方法在跨视角特征融合过程中自适应性不足、难以充分利用不同视角互补信息的问题，本文进一步提出了结合 YOLOv11 改进的双视角信息融合的违禁品检测算法 DVF-YOLOv11（Dual View Fusion combined with YOLOv11），并设计了跨视角注意力融合（Cross-View Attention Fusion, CVAF）模块。通过通道注意力与空间注意力的级联机制实现双视角特征的自适应增强，并采用自适应权重预测网络动态调节各视角的融合比重，结合通道缩减卷积构成双路径融合策略，进一步设计了包含特征保留损失、互补性损失和权重均衡损失的联合损失函数。有效整合了双视角互补信息，缓解因目标遮挡和投影形态退化导致的漏检问题，进一步提升了 X 光安检图像的违禁品检测精度。

2　DVF-YOLOv11 模型

本文构建的 DVF-YOLOv11 整体网络结构如图 1（彩图见期刊电子版）所示。该模型以 YOLOv11 为骨干，采用双分支并行结构分别对俯视图和侧视图进行多尺度特征提取，并通过 CVAF 模块实现双视角特征的自适应交互与融合，最终由双路检测头输出检测结果。

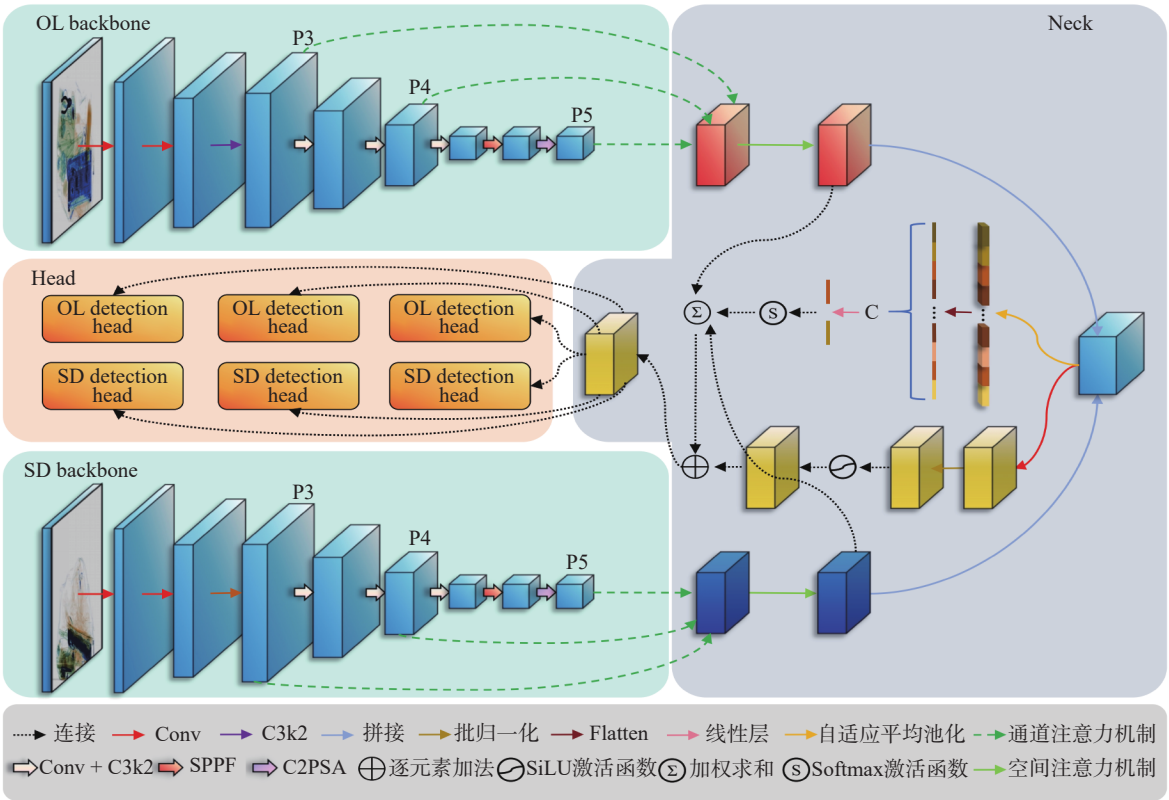


图 1　DVF-YOLOv11 网络结构图

Fig. 1　Overall architecture of DVF-YOLOv11

2.1　双分支特征提取模块

双分支特征提取模块负责从俯视图和侧视图中提取多尺度特征表示。由于两个视角的成像对象相同，仅是观察角度不同，本文采用参数共享的设计策略，即两个分支使用相同结构和相同预训练权重的骨干网络。图 1 中 Neck 部分仅以单尺

度融合过程进行示意，实际网络在 P3、P4、P5 三个特征尺度上分别执行对应的融合操作，并独立完成各尺度特征的交互与加权。

骨干网络采用 YOLOv11 的特征提取结构，C3k2（Cross Stage Partial with kernel size 2）模块作为骨干网络的核心构建单元，能够在保持计算效率的同时提取丰富的特征信息；SPPF（Spatial Pyramid Pooling-Fast）模块位于骨干网络末端，通过多尺度池化操作扩大感受野，增强网络对不同尺度目标的感知能力。骨干网络在三个关键层输出 P3、P4、P5 三个尺度的特征图，构成多尺度特征金字塔，不同尺度特征包含由浅至深的细节信息与语义信息。对于俯视图 $\boldsymbol{I}_{\mathrm{OL}}$ 和侧视图 $\boldsymbol{I}_{\mathrm{SD}}$，特征提取过程如下：

$$\boldsymbol{F}_{\mathrm{OL}}^{(l)} = \Phi_{\mathrm{backbone}}(\boldsymbol{I}_{\mathrm{OL}}), l \in \{3,4,5\} \quad , \qquad (1)$$

$$\boldsymbol{F}_{\mathrm{SD}}^{(l)} = \Phi_{\mathrm{backbone}}(\boldsymbol{I}_{\mathrm{SD}}), l \in \{3,4,5\} \quad , \qquad (2)$$

其中：Φ_{backbone} 表示共享的骨干网络，$\boldsymbol{F}_{\mathrm{OL}}^{(l)}$ 和 $\boldsymbol{F}_{\mathrm{SD}}^{(l)}$ 分别表示俯视图和侧视图在第 l 个尺度上的特征图。

采用预训练权重冻结策略的原因是预训练模型已在大规模数据集上学习到丰富的视觉特征，冻结骨干网络可以减少可训练参数量，降低过拟合风险，同时加快训练收敛速度。

2.2 跨视角注意力融合模块

CVAF 模块结构如图 2（彩图见期刊电子版）所示。该模块对双视角特征进行通道注意力和空间注意力增强，再通过自适应权重预测与通道压缩卷积的双路策略完成融合。

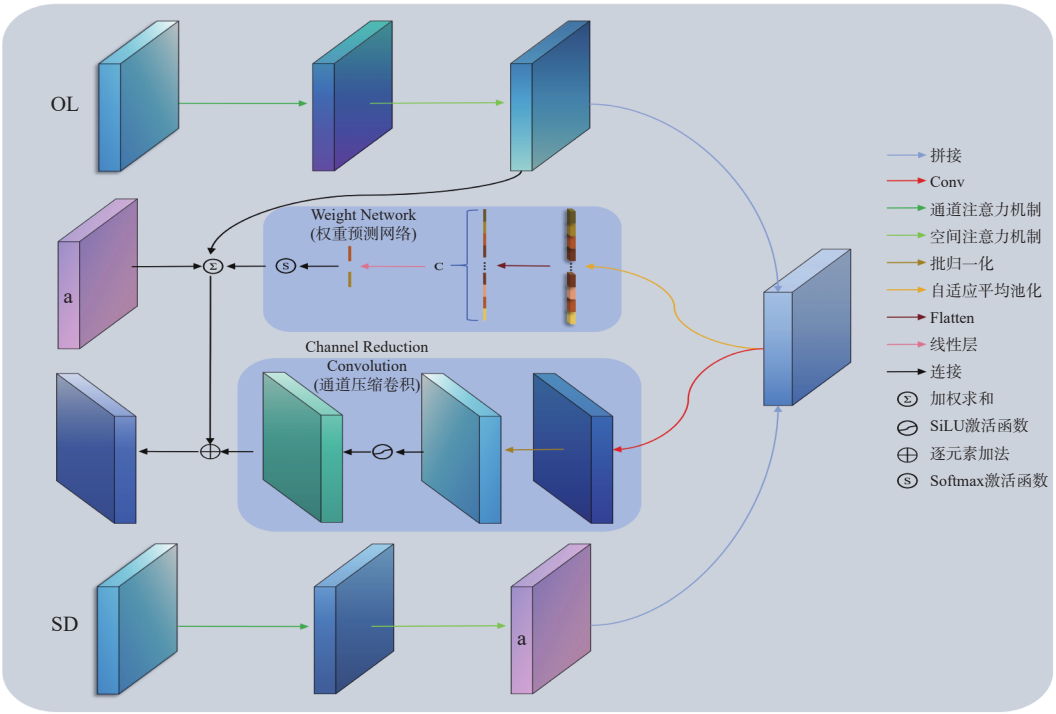


图 2 CVAF 模块结构图

Fig. 2 Structural diagram of the CVAF module

2.2.1 通道注意力机制

通道注意力机制通过学习各特征通道的重要性权重，使网络能够自适应地关注与违禁品检测相关的判别性特征通道，其结构如图 3（彩图见期刊电子版）所示。在 X 光安检图像中，不同的特征通道通常编码不同类型的视觉模式，如边缘、纹理、形状等。部分通道对金属物体的检测更为敏感，而另一些通道更擅长捕捉有机物的特征。

对输入特征图 $\boldsymbol{F} \in \mathbb{R}^{C\times H\times W}$ 进行全局平均池化，将每个通道的空间信息压缩为标量，进而得到通道特征向量 $\boldsymbol{z} \in \mathbb{R}^{C\times 1\times 1}$：

$$z_c = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} F_c(i,j) \quad , \qquad (3)$$

其中：$F_c(i,j)$ 表示输入特征图 F 在第 c 个通道、空间位置 (i,j) 处的特征值，z_c 表示第 c 个通道经全

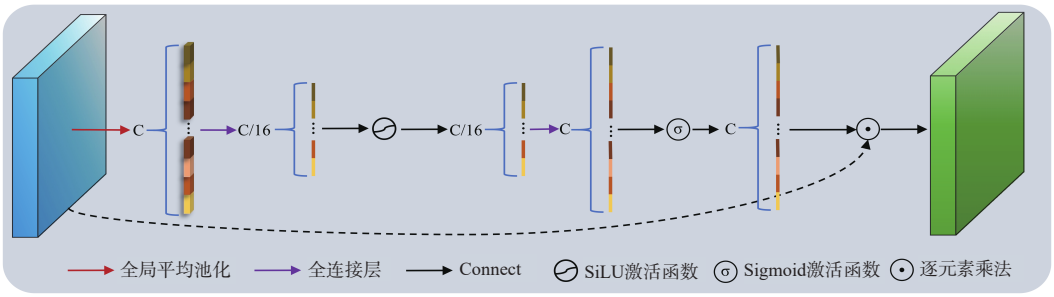


图 3　通道注意力机制结构图

Fig. 3　Structural diagram of the channel attention mechanism

局平均池化后的特征标量。

然后, 通道特征向量经过一个轻量级的多层感知机(Multi-Layer Perceptron, MLP)学习各通道的重要性权重。为减少参数量和计算开销, MLP 采用瓶颈结构, 先将通道数由 C 压缩至 C/16, 经 SiLU 激活后再恢复至原始通道数 C, 最终通过 Sigmoid 函数归一化输出通道权重向量:

$$\boldsymbol{s}=\sigma(\boldsymbol{W}_2\cdot\delta(\boldsymbol{W}_1\cdot\boldsymbol{z}))\quad,\qquad(4)$$

其中: $\boldsymbol{W}_1\in\mathbb{R}^{(C/r)\times C}$ 和 $\boldsymbol{W}_2\in\mathbb{R}^{C\times(C/r)}$ 为全连接层参数, δ 表示 SiLU 激活函数, σ 表示 Sigmoid 激活函数, $\boldsymbol{s}\in\mathbb{R}^{C\times1\times1}$ 为学习到的通道权重向量。

最后将学习到的通道权重与原始特征图进行逐元素乘法, 实现特征的自适应重标定:

$$\tilde{\boldsymbol{F}}=\boldsymbol{F}\odot\boldsymbol{s}\quad,\qquad(5)$$

其中: ⊙表示逐元素乘法, $\tilde{\boldsymbol{F}}$ 为增强后的特征图。

2.2.2　空间注意力机制

空间注意力机制通过学习特征图各空间位置的重要性权重, 引导网络聚焦于目标所在区域。其结构如图 4(彩图见期刊电子版)所示。在 X 光安检图像中, 违禁品通常只占图像的一小部分区域, 大量背景区域包含无关物品的特征, 这些背景特征可能对检测产生干扰。空间注意力机制通过学习空间权重图, 能够自动突出目标区域的特征响应, 抑制背景区域的干扰。

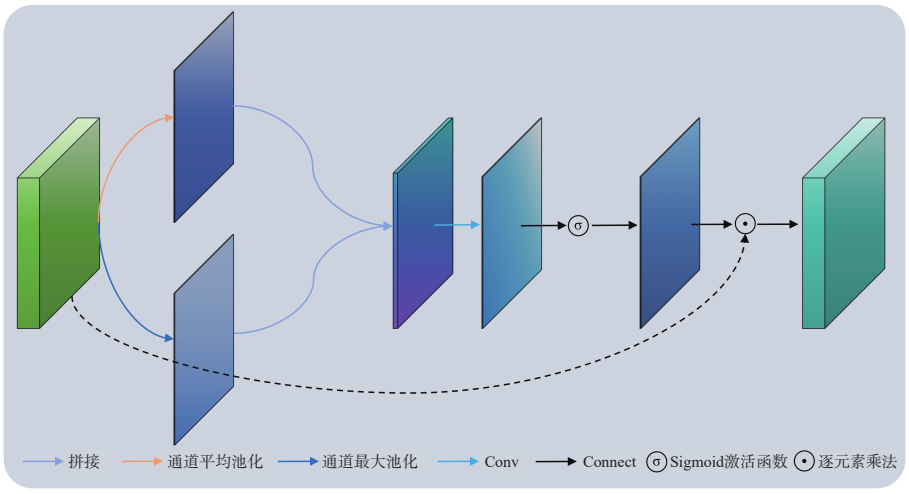


图 4　空间注意力机制结构图

Fig. 4　Structural diagram of the spatial attention mechanism

对输入特征图 $\boldsymbol{F}\in\mathbb{R}^{C\times H\times W}$ 沿通道维度分别进行平均池化和最大池化操作, 生成两个单通道的空间特征图:

$$\boldsymbol{F}_{\text{avg}}=\frac{1}{C}\sum_{c=1}^{C}\boldsymbol{F}_c\in\mathbb{R}^{1\times H\times W}\quad,\qquad(6)$$

$$\boldsymbol{F}_{\max}=\max_c\boldsymbol{F}_c\in\mathbb{R}^{1\times H\times W}\quad,\qquad(7)$$

平均池化能够反映各空间位置的整体激活强度, 而最大池化能够捕捉各位置重要的特征响应。两者结合提供了更全面的空间信息表达。

然后将两个空间特征图沿通道维度拼接, 通

过一个 3×3 卷积层学习空间注意力图，得到：

$$\boldsymbol{M}=\sigma\left(Conv_{3\times3}([\boldsymbol{F}_{\text{avg}};\boldsymbol{F}_{\max}])\right) \quad , \tag{8}$$

其中：[;]表示通道维度的拼接操作，$Conv_{3\times3}$为 3×3 卷积层（输入通道数为 2，输出通道数为 1），σ表示 Sigmoid 激活函数，$\boldsymbol{M}\in\mathbb{R}^{1\times H\times W}$为学习到的空间注意力图。

最后将空间注意力图与原始特征图进行逐元素乘法，得到：

$$\hat{\boldsymbol{F}}=\boldsymbol{F}\odot\boldsymbol{M} \quad , \tag{9}$$

其中：$\hat{\boldsymbol{F}}$为空间增强后的特征图。

在 CVAF 模块中，通道注意力和空间注意力采用级联方式对双视角特征进行处理。对于俯视图特征$\boldsymbol{F}_{\text{OL}}$和侧视图特征$\boldsymbol{F}_{\text{SD}}$，处理过程如下：

$$\boldsymbol{F}'_{\text{OL}}=SA(CA(\boldsymbol{F}_{\text{OL}})) \quad , \tag{10}$$

$$\boldsymbol{F}'_{\text{SD}}=SA(CA(\boldsymbol{F}_{\text{SD}})) \quad , \tag{11}$$

其中：$CA(\cdot)$表示通道注意力操作，$SA(\cdot)$表示空间注意力操作，$\boldsymbol{F}'_{\text{OL}}$和$\boldsymbol{F}'_{\text{SD}}$为增强后的双视角特征。

通道注意力在全局范围内筛选重要的特征通道，过滤掉无关的特征响应；空间注意力在此基础上聚焦于目标所在的空间区域。两者协同作用，能够有效提升特征的判别性。

2.2.3 自适应权重预测机制

在实际安检场景中，两个视角对目标检测的贡献程度随目标类别与成像条件的不同而动态变化。对于扁平状刀具类违禁品，俯视图能够完整呈现刀刃轮廓，包含更丰富的形态信息；对于立式压力容器类物品，侧视图能够清晰展现其筒状结构，具有更强的判别性；对于被周围物品部分遮挡的目标，受遮挡影响较小的视角则更为可靠。

因此，固定的视角融合权重难以适应上述多变的检测场景，需要根据输入内容自适应地调整各视角的贡献权重。本文设计了一个权重预测网络，能根据当前输入特征自适应地学习融合权重。

将增强后的俯视图特征$\boldsymbol{F}'_{\text{OL}}$和侧视图特征$\boldsymbol{F}'_{\text{SD}}$沿通道维度拼接，得到 2C 通道的特征图：

$$\boldsymbol{F}_{\text{concat}}=[\boldsymbol{F}'_{\text{OL}};\boldsymbol{F}'_{\text{SD}}]\in\mathbb{R}^{2C\times H\times W} \quad , \tag{12}$$

然后对拼接特征进行全局平均池化，将空间维度压缩至单个标量，获取全局特征向量：

$$\boldsymbol{g}=\text{GAP}(\boldsymbol{F}_{\text{concat}})\in\mathbb{R}^{2C} \quad . \tag{13}$$

进一步通过 Flatten 操作展平后，经全连接层和 Softmax 激活函数预测归一化融合权重：

$$[\alpha,\beta]=\text{Softmax}(\boldsymbol{W}_w\cdot\boldsymbol{g}) \quad , \tag{14}$$

其中：$\boldsymbol{W}_w\in\mathbb{R}^{2\times2C}$为全连接层参数，$\alpha$和$\beta$分别为俯视图和侧视图的融合权重，满足$\alpha+\beta=1$。Softmax 归一化确保了两个权重的和为 1。当网络判断俯视图特征更可靠时，会增大α的值，反之则增大β的值。同时，由于权重和恒为 1，避免了融合后特征尺度的变化。

2.2.4 通道压缩卷积

为充分利用双视角的互补信息，本文采用加权融合与通道压缩卷积相结合的双路融合策略，利用预测的权重对双视角特征进行加权求和：

$$\boldsymbol{F}_{\text{weighted}}=\alpha\cdot\boldsymbol{F}'_{\text{OL}}+\beta\cdot\boldsymbol{F}'_{\text{SD}} \quad , \tag{15}$$

通道压缩卷积路径对拼接特征通过 1×1 卷积进行通道压缩和特征交互：

$$\boldsymbol{F}_{\text{conv}}=\text{SiLU}(\text{BN}(\text{Conv}_{1\times1}(\boldsymbol{F}_{\text{concat}}))) \quad , \tag{16}$$

其中：$\text{Conv}_{1\times1}$将通道数从 2C 压缩至 C，BN 表示批量归一化，后接 SiLU 激活函数。

将两路融合结果相加得到融合特征，并通过残差连接保留原始视角信息：

$$\boldsymbol{F}_{\text{fused}}=\boldsymbol{F}_{\text{weighted}}+\boldsymbol{F}_{\text{conv}} \quad , \tag{17}$$

$$\boldsymbol{F}_{\text{out}}^{\text{OL}}=\text{Head}_{\text{OL}}(\boldsymbol{F}_{\text{fused}})+0.5\cdot\boldsymbol{F}_{\text{OL}} \quad , \tag{18}$$

$$\boldsymbol{F}_{\text{out}}^{\text{SD}}=\text{Head}_{\text{SD}}(\boldsymbol{F}_{\text{fused}})+0.5\cdot\boldsymbol{F}_{\text{SD}} \quad , \tag{19}$$

其中：Head_{OL}和Head_{SD}分别为俯视图和侧视图的输出头，每个输出头由 3×3 卷积、批量归一化和 SiLU 激活函数组成，残差系数 0.5 用于平衡融合特征与原始特征的贡献。

上述融合操作在 P3、P4、P5 三个特征尺度上独立进行：

$$\{\boldsymbol{F}_{\text{out}}^{\text{OL},(l)},\boldsymbol{F}_{\text{out}}^{\text{SD},(l)},\boldsymbol{w}^{(l)}\}=\text{FusionBlock}^{(l)}\left(\boldsymbol{F}_{\text{OL}}^{(l)},\boldsymbol{F}_{\text{SD}}^{(l)}\right) \quad , \tag{20}$$

其中：$l\in\{3,4,5\}$，$\text{FusionBlock}^{(l)}$表示第l个尺度的融合块，$\boldsymbol{w}^{(l)}=[\alpha^{(l)},\beta^{(l)}]$为该尺度的融合权重。

各尺度独立预测融合权重，以适应不同层级特征中两个视角相对重要性的差异。

2.3　联合损失函数

本文设计了由特征保留损失、互补性损失和权重平衡损失组成的联合损失函数, 以引导网络学习有效的融合策略。特征保留损失可以确保融合过程中不丢失原始视角的有用信息。对于每个尺度l的特征, 特征保留损失见式(21):

$$\mathcal{L}_{\text{preserve}}^{(l)}=\frac{1}{2}\left(\| \boldsymbol{F}_{\text{out}}^{\text{OL},(l)}-\boldsymbol{F}_{\text{OL}}^{(l)} \|_2^2+\| \boldsymbol{F}_{\text{out}}^{\text{SD},(l)}-\boldsymbol{F}_{\text{SD}}^{(l)} \|_2^2\right) \quad , \tag{21}$$

其中: $\|\cdot\|_2$表示 L2 范数。该损失项鼓励融合输出保留原始特征的主要信息, 防止融合过程中因过度变换而丢失判别性特征。

多尺度特征保留损失为各尺度损失的加权和:

$$\mathcal{L}_{\text{preserve}}=\sum_{l=3}^{5}\lambda_l\cdot\mathcal{L}_{\text{presarve}}^{(l)} \quad , \tag{22}$$

其中λ_l为各尺度的权重系数。考虑到深层特征包含更丰富的语义信息, 本文设置$\lambda_3=0.25$、$\lambda_4=0.5$、$\lambda_5=1.0$, 对深层特征给予更高的保留权重。

互补性损失可以使网络充分利用双视角的互补信息, 而非简单地复制单一视角的特征。互补性损失通过约束双视角融合输出之间的差异来实现, 如下:

$$\mathcal{L}_{\text{comp}}^{(l)}=\max\left(0,\tau-\| \boldsymbol{F}_{\text{out}}^{\text{OL},(l)}-\boldsymbol{F}_{\text{out}}^{\text{SD},(l)} \|_2^2\right) \quad , \tag{23}$$

其中τ为差异阈值。当双视角输出的差异小于τ时, 损失为正值, 促使网络增大两者的差异; 当差异大于τ时, 损失为零, 允许网络自由学习。

权重平衡损失防止自适应权重出现极端分配, 将权重向均匀分布方向正则化:

$$\mathcal{L}_{\text{balance}}^{(1)}=\left(\alpha^{(l)}-0.5\right)^2+\left(\beta^{(l)}-0.5\right)^2 \quad . \tag{24}$$

由于$\alpha+\beta=1$, 上式可简化为:

$$\mathcal{L}_{\text{balance}}^{(1)}=2\cdot\left(\alpha^{(l)}-0.5\right)^2 \quad . \tag{25}$$

该损失项将权重向均匀分布方向正则化。多尺度权重平衡损失, 如下:

$$\mathcal{L}_{\text{blance}}=\sum_{l=3}^{5}\mathcal{L}_{\text{baltance}}^{(l)} \quad , \tag{26}$$

权重平衡损失提供一个软约束, 防止训练初期的极端权重分配。随着训练的进行, 网络会根据实际输入自适应地调整权重, 在整体约束下学习到合理的权重分布。联合损失函数定义见式(27):

$$\mathcal{L}_{\text{total}}=\mathcal{L}_{\text{det}}+\lambda_1\cdot\mathcal{L}_{\text{preserve}}+\lambda_2\cdot\mathcal{L}_{\text{comp}}+\lambda_3\cdot\mathcal{L}_{\text{balance}} \quad , \tag{27}$$

其中: $\mathcal{L}_{\text{det}}$为目标检测损失, 包含分类损失、边界框回归损失和置信度损失, λ_1、λ_2、λ_3为各辅助损失项的权重系数, 各损失项的作用如表 1 所示。

表 1　各损失项的作用

Tab. 1　The role of each loss item

损失项	符号	作用	权重
检测损失	$\mathcal{L}_{\text{det}}$	监督目标检测任务	1.0
特征保留损失	$\mathcal{L}_{\text{preserve}}$	保持融合特征与原始特征的一致性	λ_1
互补性损失	$\mathcal{L}_{\text{comp}}$	促进双视角特征的差异性与互补性	λ_2
权重平衡损失	$\mathcal{L}_{\text{balance}}$	防止权重分配的极端化	λ_3

在本文中, 经过超参数搜索, 各权重系数设置分别为$\lambda_1=0.1$、$\lambda_2=0.01$、$\lambda_3=0.1$。确保了检测损失占主导地位, 同时辅助损失项提供适当的正则化作用。

3　实验结果及分析

为验证所提出的双视角融合检测算法的有效性, 本节在 DvXray 数据集上开展系统实验, 包括不同模型与不同融合策略的对比实验、各模块贡献的消融实验以及检测结果的可视化分析。

3.1　数据集描述

本文实验在 DvXray 数据集[16]上进行。DvXray 由 Ma 等于 2024 年发布, 是目前首个可公开获取的大规模双视角 X 光安检数据集。其包含 32 000 幅图像, 即 16 000 对双视角图像, 其中 5 000 对为包含违禁品的正样本, 11 000 对为不包含违禁品的负样本, 涵盖枪支、刀具、扳手、打火机、电池等 15 类常见违禁品, 所有图像均为 PNG 格式, 分辨率为 800 pixel×600 pixel, 如图 5(彩图见期刊电子版)所示。

数据集融合了真实场景与模拟环境, 约 66% 的样本采集自真实地铁站, 其余 34% 通过人工设计复杂重叠的行李并经 X 光安检机扫描获得。在标注方面, 训练集提供图像级标签, 测试集额外提供了边界框级标注, 并特别标注了因旋转或遮挡而仅在单一视角下可见的较困难样本。训练集与测试集按照 8∶2 划分。



图 5 DvXray 中的 X 光图像示例

Fig. 5 X-ray images in DvXray

3.2 对比实验及分析

为全面评估本文提出的双视角融合检测算法的性能，表 2 对比了不同模型在单视角与双视角配置下的检测效果，展示了 YOLOv8[17]、YOLOv10[18]、YOLOv11、YOLOv12[19] 和 YOLOv-13[20] 在 DvXray 数据集上的检测性能对比结果。

表 2 不同模型检测性能对比

Tab. 2 Comparison of detection performance under different models

模型	视角	P(%)	R(%)	F_1(%)	mAP50(%)	mAP50-95(%)	Params(M)	GFLOPs	FPS
YOLOv8	OL	92.03	83.52	87.56	89.63	72.48	3.01	8.1	98.3
	SD	84.31	74.21	78.93	79.28	56.83	3.01	8.1	98.9
	Dual	94.27	86.58	90.27	92.41	77.03	4.75	17.0	45.2
YOLOv10	OL	91.42	82.76	86.87	88.91	71.58	2.27	6.5	75.8
	SD	83.68	73.47	78.26	78.62	55.93	2.27	6.5	76.3
	Dual	93.71	85.86	89.61	91.92	76.04	4.01	13.8	35.6
YOLOv12	OL	92.82	84.23	88.32	90.53	73.51	2.51	5.8	73.8
	SD	85.14	74.92	79.71	80.02	57.86	2.51	5.8	74.5
	Dual	95.08	87.42	91.07	93.21	78.23	4.25	12.4	35.0
YOLOv13	OL	90.73	82.13	86.22	88.17	70.68	2.45	6.2	77.1
	SD	83.04	72.71	77.56	77.93	55.12	2.45	6.2	77.8
	Dual	93.02	85.17	88.91	91.23	75.14	4.19	13.2	36.5
YOLOv11	OL	93.28	84.51	88.69	91.03	74.12	2.58	6.3	107.3
	SD	85.47	75.62	80.27	80.81	59.07	2.58	6.3	108.1
	Dual	95.91	88.32	91.93	94.02	79.41	4.32	13.4	49.6

可以看出, 各模型在不同视角配置下呈现出以下规律, OL 单视角整体优于 SD 单视角, 而在结合本文提出的 CVAF 融合方法后, 各模型在 OL 单视角基础上均获得了稳定提升。以 mAP50 为例, YOLOv8、YOLOv10、YOLOv11、YOLOv12 和 YOLOv13 分别提升 2.78%、3.01%、2.99%、2.68% 和 3.06%, 说明本文方法对不同骨干网络均具有良好的普适性与鲁棒性。主要是由于 OL 视角虽具备较强的平面轮廓表达能力, 但难以反映目标在高度与厚度方向上的结构信息; SD 视角则能够提供侧向轮廓与深度方向补充, 本文方法通过 CVAF 模块有效弥补了单一视角在遮挡场景下的信息缺失。

本文方法不仅提高了整体检测精度, 也同步改善了精确率、召回率和定位质量。以 YOLOv11 为例, 精确率由 93.28% 提升至 95.91%, 召回率由 84.51% 提升至 88.32%; 同时, 各模型的 mAP50-95 提升幅度均大于 mAP50, 其中 YOLOv8、YOLOv10、YOLOv11、YOLOv12 和 YOLOv13 分别提升 4.55%、4.46%、5.29%、4.72% 和 4.46%。这表明 CVAF 模块不仅能够增强对目标区域的判别能力, 还能更有效地利用双视角互补信息提升边界框的定位精度。在计算效率方面, 双视角配置带来了一定的参数量和计算量增长, 但主要来源于双分支骨干网络的前向传播, 融合模块本身增加的开销相对有限。结果表明, 本文方法在不同模型上均表现出较好的普适性, 并在检测性能与计算开销之间取得了较为合理的平衡。

为了分析不同融合方法的性能差异, 表 3 对比了不同特征融合策略在双视角违禁品检测任务上的性能差异。由于现有代表性双视角融合方法在任务定义与数据集上与本文存在差异, DVNet、Trans2Ray 所用的 Dualray 数据集尚未公开, DAGNet、AHCR 在 DvXray 上将任务建模为多标签分类, 难以与本文检测任务进行端到端指标对比。表 3 在相同骨干网络、相同数据集、相同检测设置下, 对比了若干通用特征融合策略, 包括特征拼接、逐元素相加、通道注意力、卷积块注意力等。可以看出, 加入融合方法与注意力机制后, 模型检测性能均得到不同程度提升, 整体精度水平有所提高。本文方法表现相对较优, 精确率达到 95.91%, 召回率为 88.32%, F1 分数为 91.93%。

表 3　不同融合方法的性能对比

Tab. 3　Performance comparison of different fusion methods

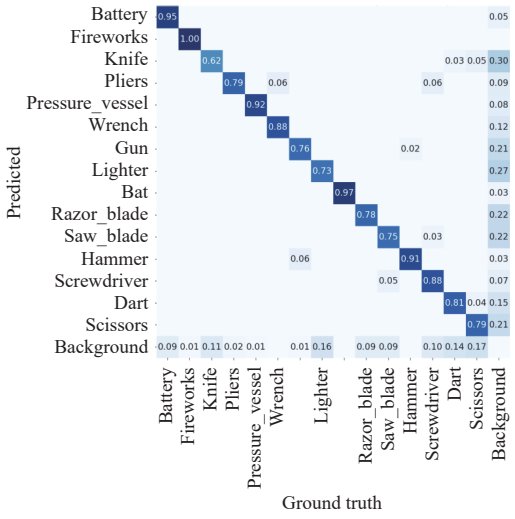
方法	精确率 (%)	召回率 (%)	F1分数 (%)	mAP50 (%)	mAP50-95 (%)
特征拼接	94.13	85.27	89.48	92.61	76.38
特征相加	94.42	84.53	89.21	92.14	75.23
SE-Net融合[21]	95.08	86.57	90.68	93.42	77.61
CBAM融合[22]	95.31	86.24	90.54	93.47	78.18
ECA-Net融合[23]	94.93	87.38	91.02	93.83	78.71
本文方法	95.91	88.32	91.93	94.02	79.41

在检测精度指标方面, mAP50 提升至 94.02%, mAP50-95 提升至 79.41%, 均为各方法中的最高值。结果表明, 本文提出的 CVAF 模块能够更有效地整合双视角互补信息并抑制冗余干扰, 从而在严格 IoU 阈值下仍保持更高的定位精度与检测鲁棒性。

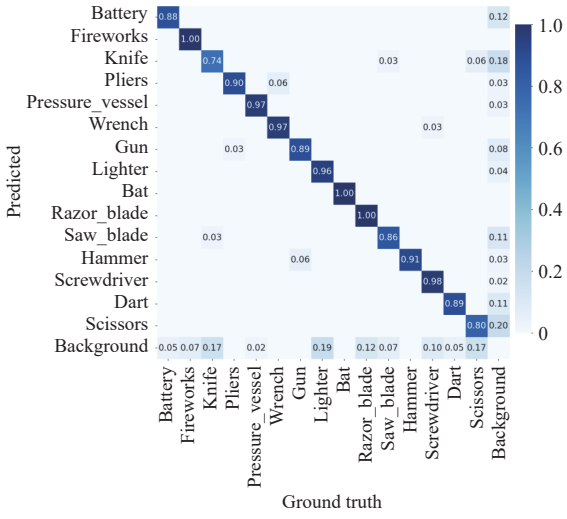
相较现有双视角融合方法, 本文方法中的通道与空间注意力的级联设计实现了双维度协同增强, 自适应权重预测网络在三个尺度上独立预测融合权重, 使融合策略可随输入动态调节, 克服了固定权重或单维度注意力的局限。同时三项辅助损失对融合过程施加显式约束, 避免融合层退化为单一视角的复制或简单平均。本文方法也存在适用边界, 其适用于两视角均能稳定成像、夹角为正交的标准设备, 当某一视角失效时融合增益会显著衰减。

如图 6(彩图见期刊电子版)所示, 归一化混淆矩阵分别展示了 SD 单视角、OL 单视角和双视角三种配置下的检测效果, 直观地反映了各方法在不同类别上的检测表现。由图 6(c) 可以看出, 在 Knife、Pliers、Gun 等关键违禁品类别上, 融合后的精确率均达到 82% 以上。融合方法的非对角线区域整体较浅, 表明类间混淆得到有效抑制。其中 SD 单视角将 30% 的 Background 误检为 Knife, 而融合方法将此误检率降低至 14%。对于 Scissors、Dart 等小目标, SD 单视角的精确率分别为 79% 和 81%, 而融合方法提升至 84% 和 93%, 提升幅度较大。

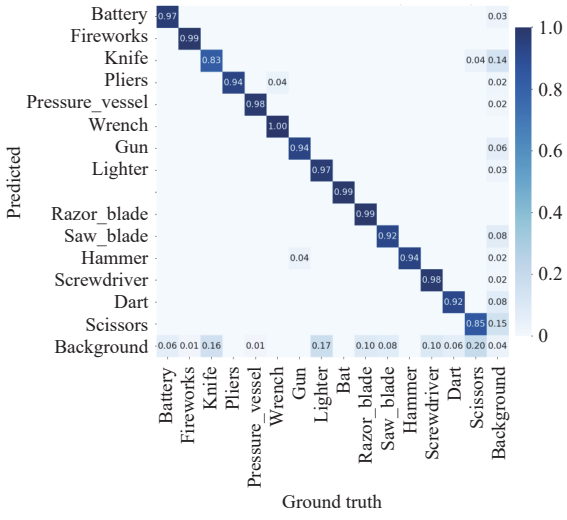
这些对比结果充分说明: 双视角融合检测算法通过整合互补信息, 既增强了模型对各类违禁品的判别能力, 又有效降低了类间混淆和背景误检, 从而实现了检测性能的提升, 证明了双视角融合能够学习到更具判别性的特征表示。



(a) SD视角混淆矩阵
(a) SD perspective confusion matrix



(b) OL视角混淆矩阵
(b) OL perspective confusion matrix



(c) 双视角混淆矩阵
(c) Dual-view confusion matrix

图 6 检测结果的混淆矩阵

Fig. 6 Confusion matrix of detection results

3.3 消融实验及分析

为验证 CVAF 模块中各组件的有效性，本文设计了消融实验，分别评估通道注意力（Channel Attention, CA）、空间注意力（Spatial Attention, SA）、权重预测网络和通道压缩卷积 4 个组件对检测性能的贡献。表 4 展示了全部 16 种组合的消融实验结果。A、B、C、D 分别代表 CA、SA、权重预测网络和通道压缩卷积 4 个组件。

表 4 消融实验结果

Tab. 4 Results of the ablation experiments

基线模型	A	B	C	D	精确率(%)	召回率(%)	mAP50 (%)	mAP50 −95(%)
√					93.51	84.82	91.27	74.53
√	√				94.03	85.38	91.82	75.47
√		√			93.76	85.63	91.57	75.68
√			√		93.68	85.14	91.46	75.03
√				√	93.57	84.93	91.38	74.72
√	√	√			94.52	86.41	92.53	76.87
√	√		√		94.18	85.82	92.14	76.12
√	√			√	94.07	85.57	91.93	75.83
√		√	√		94.03	85.96	92.04	76.28
√		√		√	93.91	85.78	91.82	76.01
√			√	√	93.82	85.31	91.63	75.27
√	√	√	√		95.37	87.72	93.58	78.63
√	√	√		√	95.21	87.48	93.42	78.27
√	√		√	√	94.68	86.53	92.71	77.14
√		√	√	√	94.47	86.58	92.63	77.23
√	√	√	√	√	95.91	88.32	94.02	79.41

可以看出，CA 单独加入后精确率提升 0.52%、mAP50 提升 0.55%，有效增强了通道维度的语义判别能力，使模型能够自适应地聚焦于与目标类别强相关的特征通道。SA 单独加入后召回率提升 0.81%、mAP50-95 提升 1.15%，通过融合平均池化与最大池化两种空间统计信息生成注意力图，强化了模型对目标空间位置的感知能力，在高 IoU 阈值下的定位精度改善较为明显。权重预测网络加入后 mAP50-95 提升 0.50%，能够根据输入特征动态生成融合权重，增强了模型对不同视角特征差异的自适应调节能力。通道压缩卷积加入后通过降低特征维度、消除冗余信息，可以

提升有效特征的利用率。改进后的模型, 精确率、召回率分别达到 95.91% 和 88.32%, mAP50 与 mAP50-95 分别达到 94.02% 和 79.41%, 较基线提升 2.75% 和 4.88%。上述结果表明 4 个模块分别从通道增强、空间聚焦、动态权重分配和特征压缩 4 个维度协同优化了特征融合过程, 提升了模型的检测性能。

图 7(彩图见期刊电子版)展示了 4 种模型配置在训练过程中的 mAP50 变化曲线。可以看出, 基线模型收敛至 91.27% 后趋于平缓, 加入 CA 与 SA 后 mAP50 提升至 92.53%, 进一步加入权重预测网络后达到 93.58%, 采用本文方法时, 收敛值最高达 94.02%, 曲线已基本趋于平稳。

3.4　可视化分析

为直观展示双视角融合检测算法的实际检测效果, 本节从双视角形态信息获取、形态判别能力和视角检测盲区互补 3 个维度选取检测示例进行可视化分析, 检测结果分别如图 8~图 10(彩图见期刊电子版)所示。

在双视角形态信息获取方面, 图 8(a)~8(d) 展示了刀具类违禁品的检测情况, OL 视角能够清晰捕捉到刀刃的平面轮廓, SD 视角则展现了刀身的侧面厚度信息, 两个视角的融合使得模型能够从不同角度全面理解目标形态, 两个视角均成功检测并输出较高置信度。

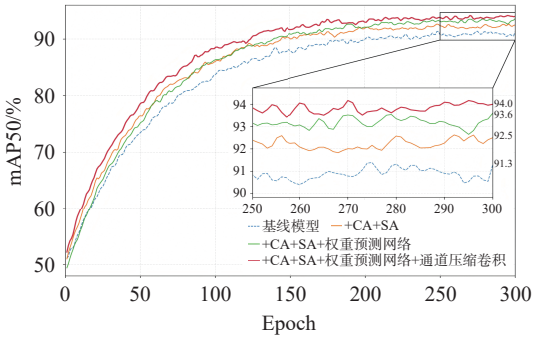


图 7　消融实验训练过程可视化结果

Fig. 7　Visualization results of ablation experiment training process

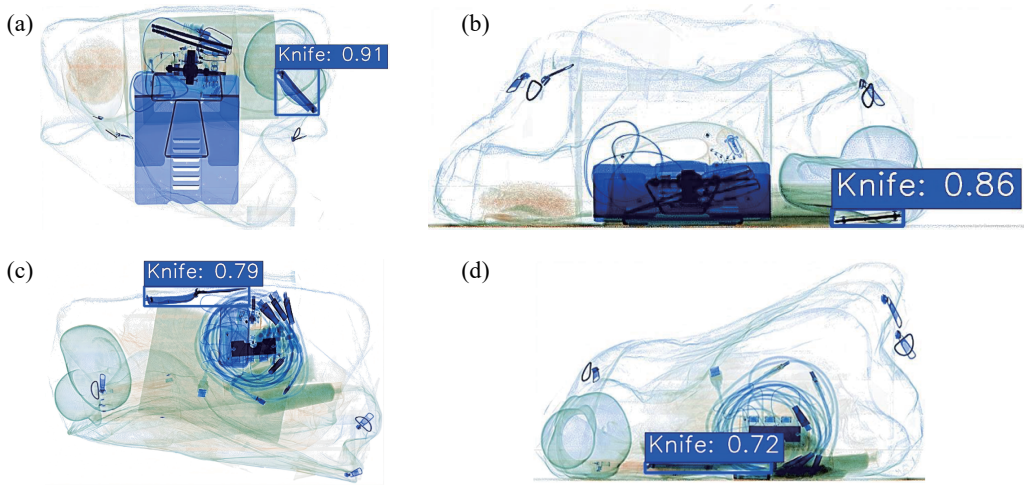
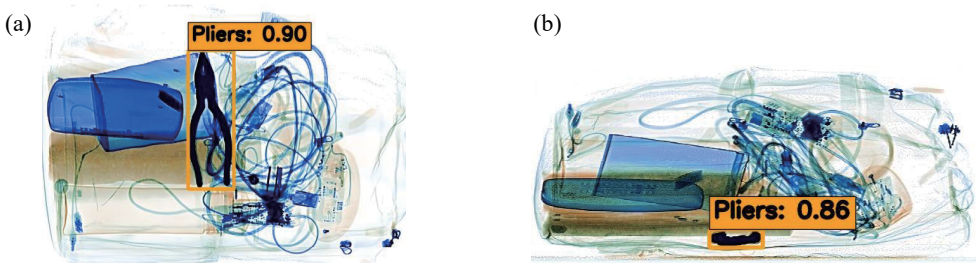


图 8　双视角形态信息获取示例。(a), (c)为 OL 视角检测结果; (b), (d)为 SD 视角检测结果

Fig. 8　Examples of dual-view morphological information acquisition. (a), (c) OL perspective detection result; (b), (d) SD perspective detection result



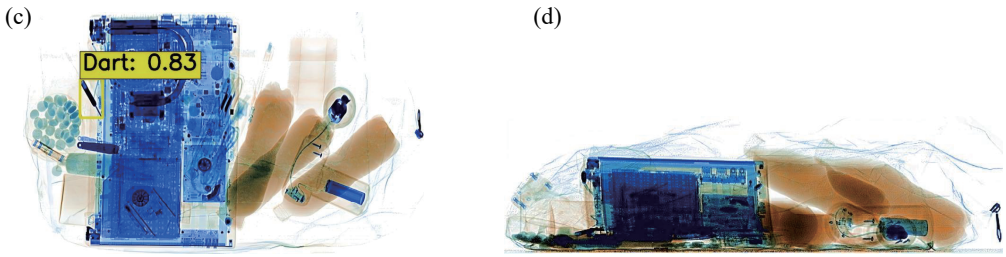


图 9 形态判别能力检测示例。(a), (c)为 OL 视角检测结果; (b), (d)为 SD 视角检测结果

Fig. 9 Examples of morphological discrimination detection. (a), (c) OL perspective detection result; (b), (d) SD perspective detection result

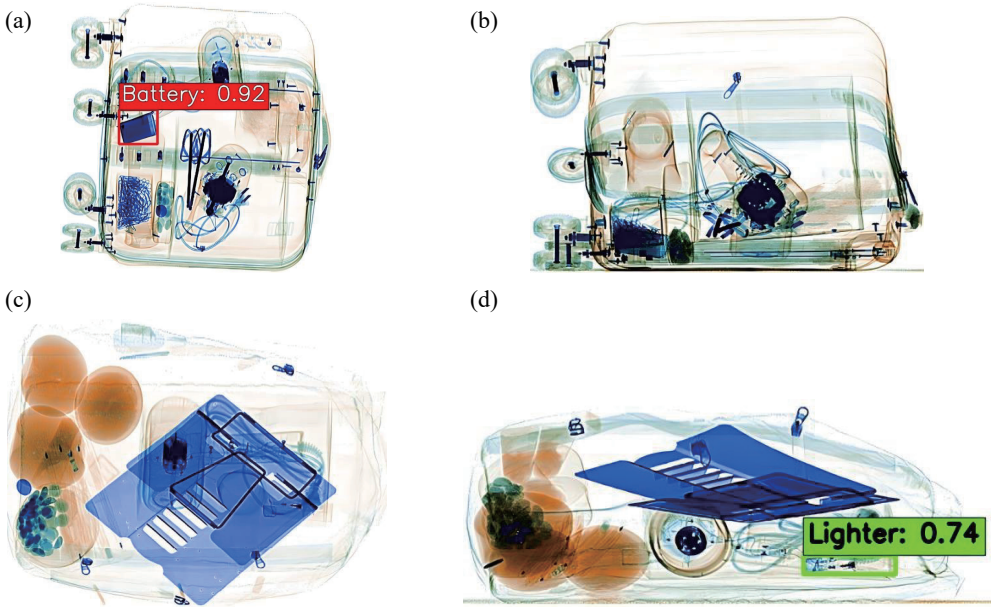


图 10 视角检测盲区互补示例。(a), (c)为 OL 视角检测结果; (b), (d)为 SD 视角检测结果

Fig. 10 Examples of complementary blind zones between two views. (a), (c) OL perspective detection result; (b), (d) SD perspective detection result

在形态判别能力方面，图 9(a) 和 9(b) 展示了钳子类金属工具的检测情况。从图 9(a) 可观察到，钳子的两个把手与周围线缆交织，轮廓不完整，但仍被成功检测；从图 9(b) 可观察到，钳子的关键判别特征在侧面投影中有所退化，但 SD 视角同样被成功检测，是由于双视角融合检测算法通过整合两个视角的互补信息，进一步增强了对该类形态复杂目标的检测可靠性。图 9(c) 和 9(d) 展示了飞镖类小目标的检测情况，飞镖形态细小，图 9(c) 展示了 OL 视角凭借平面轮廓成功检测，图 9(d) 展示了 SD 视角中飞镖侧面投影几乎退化为点状，未能成功检测。双视角融合检测算法通过自适应权重预测机制，在 SD 视角检测能力不足时自动增大 OL 视角的融合权重，确保了对形态判别困难目标的准确检测。

在遮挡鲁棒性方面，图 10(a) 和 10(b) 展示了复杂堆叠场景下电池的检测情况，图 10(a) 中物品分布相对分散，电池轮廓清晰可见，成功检测；而图 10(b) 中物品相互重叠，特征叠加导致目标分离困难，未能成功检测。图 10(c) 和 10(d) 中的打火机情况相反，图 10(c) 中打火机被其他物品遮挡，未能检测到，而图 10(d) 中其侧面轮廓清晰可见，被成功检测并标注。两个示例共同说明不同目标在不同视角下的可检测性存在差异，双视角融合通过整合两个视角的互补信息，确保了在不同遮挡场景下均能实现对目标的有效检测。

检测结果表明，CVAF 模块通过通道注意力和空间注意力的级联设计有效增强了双视角特征的判别性，缓解了因遮挡导致的漏检问题；自适应权重预测机制在不同视角检测能力存在差异时动

态调整融合权重, 弥补了固定融合策略对复杂场景适应性不足的缺陷。

4　结　论

本文提出了双视角融合检测方法 DVF-YOLOv11, 该方法包含双分支特征提取、CVAF 和自适应权重预测机制 3 个核心组件。CVAF 模块通过通道注意力与空间注意力的级联设计对双视角特征进行自适应增强, 自适应权重预测机制结合加权融合与通道压缩卷积的双路策略动态调整视角融合权重。在 DvXray 数据集上的实验结果表明, 该方法 mAP50 达到 94.02%, mAP50-95 达到 79.41%, 较俯视图单视角分别提升 2.99% 和 5.29%; 精确率达到 95.91%, 召回率达到 88.32%, 较俯视图单视角分别提升 2.63% 和 3.81%。未来工作可从以下 3 个方面进一步展开: (1)在模型轻量化方面, 探索知识蒸馏、模型剪枝与通道稀疏化策略, 在保持双视角融合精度的同时降低参数量与推理时间, 以满足实际安检设备的实时部署需求; (2)在数据层面, 构建覆盖更多违禁品类别与更复杂遮挡场景的双视角数据集, 并研究半监督与自监督学习方法, 缓解大规模标注数据获取困难的问题; (3)在任务扩展方面, 探索面向 X 光安检场景的开放世界目标检测方法, 提升模型对未知类别违禁品的发现与持续学习能力, 推动违禁品检测由固定类别识别向动态类别认知演进。

参考文献:

[1] 林俊豪, 张云飞, 陈少伟, 等. 无监督掩码循环对抗网络实现细胞虚拟染色[J]. 中国光学 (中英文), 2026, 19(4): 787-800.
LIN J H, ZHANG Y F, CHEN SH W, *et al*. . Unsupervised masked cycle-adversarial network for cellular virtual staining[J]. *Chinese Optics*, 2026, 19(4): 787-800. (in Chinese).

[2] 汪建民, 赵浩冰, 王轲, 等. 无人机飞行单光子动态成像中姿态补偿及重建方法[J]. 中国光学 (中英文), 2026, 19(3): 605-618.
WANG J M, ZHAO H B, WANG K, *et al*.. Attitude compensation and reconstruction methods for single-photon dynamic imaging during UAV flight[J]. *Chinese Optics*, 2026, 19(3): 605-618.

[3] XU Y, ZHANG Q Y, SU Q, *et al*.. PIXDet: prohibited item detection in X-ray image based on whole-process feature fusion and local-global semantic dependency interaction[J]. *IEEE Transactions on Instrumentation and Measurement*, 2023, 72: 5032917.

[4] WEI Y L, TAO R SH, WU ZH J, *et al*. . Occluded prohibited items detection: an X-ray security inspection benchmark and de-occlusion attention module[C]. *Proceedings of the 28th ACM International Conference on Multimedia*, ACM, 2020: 138-146.

[5] TAO R SH, WEI Y L, JIANG X J, *et al*. . Towards real-world X-ray security inspection: a high-quality benchmark and lateral inhibition module for prohibited items detection[C]. *2021 IEEE/CVF International Conference on Computer Vision*, IEEE, 2021: 10923-10932.

[6] ZHU Z M, ZHU Y, WANG H R, *et al*.. FDTNet: enhancing frequency-aware representation for prohibited object detection from X-ray images via dual-stream transformers[J]. *Engineering Applications of Artificial Intelligence*, 2024, 133: 108076.

[7] 刘建军, 冯沛, 廖威, 等. YOLO-STM: 基于 Swin-Transformer 与 MSDA 的 X 光安检图像危险品识别网络[J]. 中国体视学与图像分析, 2024, 29(3): 230-241.
LIU J J, FENG P, LIAO W, *et al*.. YOLO-STM: a network model for identifying prohibited items in X-ray security inspection images based on Swin-Transformer and MSDA[J]. *Chinese Journal of Stereology and Image Analysis*, 2024, 29(3): 230-241. (in Chinese).

[8] REDMON J, DIVVALA S, GIRSHICK R, *et al*. . You only look once: unified, real-time object detection[C]. *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2016: 779-788.

[9] KHANAM R, HUSSAIN M. YOLOv11: an overview of the key architectural enhancements[J]. arXiv preprint arXiv: 2410.17725, 2024.

[10] STEITZ J M O, SAEEDAN F, ROTH S. Multi-view X-ray R-CNN[C]. *Proceedings of the 40th German Conference on*

Pattern Recognition, Springer, 2019: 153-168.

[11] TULI A, BOHRA R, MOGHE T, *et al.* . Automatic threat detection in single, stereo (two) and multi view X-ray images[C]. *Proceedings of 2020 IEEE 17th India Council International Conference*, IEEE, 2020: 1-7.

[12] WU M D, YI F F, ZHANG H G, *et al.* . Dualray: dual-view X-ray security inspection benchmark and fusion detection framework[C]. *Proceedings of the 5th Chinese Conference on Pattern Recognition and Computer Vision*, Springer, 2022: 721-734.

[13] MENG X L, FENG H, REN Y, *et al.*. Transformer-based dual-view X-ray security inspection image analysis[J]. *Engineering Applications of Artificial Intelligence*, 2024, 138: 109382.

[14] HONG S L, ZHOU Y Z, XU W C. DAGNet: a dual-view attention-guided network for efficient X-ray security inspection[C]. *Proceedings of 2025 International Joint Conference on Neural Networks*, IEEE, 2025: 1-8.

[15] TAO R SH, WANG H Y, GUO Y ZH, *et al.* . Dual-view X-ray detection: can AI detect prohibited items from dual-view X-ray images like humans?[C]. *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, 2025: 10338-10347.

[16] MA B W, JIA T, LI M Y, *et al.*. Toward dual-view X-ray baggage inspection: a large-scale benchmark and adaptive hierarchical cross refinement for prohibited item discovery[J]. *IEEE Transactions on Information Forensics and Security*, 2024, 19: 3866-3878.

[17] VARGHESE R, SAMBATH M. YOLOv8: a novel object detection algorithm with enhanced performance and robustness[C]. *Proceedings of the 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems*, IEEE, 2024: 1-6.

[18] WANG A, CHEN H, LIU L H, *et al.* . YOLOv10: real-time end-to-end object detection[C]. *Proceedings of the 38th International Conference on Neural Information Processing Systems*, Curran Associates Inc. , 2024: 3429.

[19] TIAN Y J, YE Q X, DOERMANN D. YOLOv12: attention-centric real-time object detectors[J]. arXiv preprint arXiv: 2502.12524, 2025.

[20] LEI M Q, LI S Q, WU Y H, *et al.* . YOLOv13: real-time object detection with hypergraph-enhanced adaptive visual perception[J]. arXiv preprint arXiv: 2506.17733, 2025.

[21] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. *Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, 2018: 7132-7141.

[22] WOO S, PARK J, LEE J Y, *et al.* . CBAM: convolutional block attention module[C]. *Proceedings of the 15th European Conference on Computer Vision*, Springer, 2018: 3-19.

[23] WANG Q L, WU B G, ZHU P F, *et al.* . ECA-Net: efficient channel attention for deep convolutional neural networks[C]. *Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, 2020: 11531-11539.

作者简介：



吴海滨 (1977—)，男，上海人，博士，教授，博士生导师。2000 年于哈尔滨工业大学获得学士学位，2002 年于哈尔滨工业大学获得硕士学位，2008 年于哈尔滨理工大学获得博士学位，现为哈尔滨理工大学教授，主要研究方向为计算机视觉、虚拟现实、遥感图像处理。E-mail: woo@hrbust.edu.cn



王爱丽 (1979—)，女，天津人，博士，教授，硕士生导师。2002 年于哈尔滨工业大学获得学士学位，2004 年于哈尔滨工业大学获得硕士学位，2008 年于哈尔滨工业大学获得博士学位，现为哈尔滨理工大学教授，主要研究方向为遥感图像处理。E-mail: aili925@hrbust.edu.cn