http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0705

面向飞机舱门检测的改进 RT-DETR 模型

张长勇[1,*]，张轩铖[1]，乔亚海[1]，李铮[2]

(1. 中国民航大学 电子信息与自动化学院，天津 300300；2. 中国民航大学 计算机科学与技术学院，天津 300300)

摘 要：飞机舱门检测对于实现机场特种车辆自动靠机至关重要，针对因天气光照变化、机身涂装干扰及设备限制而导致的检测精度低、模型部署困难等问题，提出一种基于改进实时检测 Transformer (RT-DETR) 的飞机舱门检测模型。选用轻量级的残差神经网络 ResNet18 作为主干特征提取基准网络；通过引入上下文引导模块，优化主干网络中的 BasicBlock，增强对上下文信息的利用，降低计算资源的消耗；在颈部编码网络，将高低频 (HiLo) 注意力机制与内尺度特征交互 (AIFI) 模块相结合，形成 AIFI-HiLo 模块，增强特征表达，降低漏检率和误检率；采用二次改进后的高级筛选特征金字塔网络 (HS-FPN) 替换跨尺度特征融合模块 (CCFM)，增强模型对多尺度特征的融合能力。实验结果表明：所提改进模型的平均精度值达到 94.9%，对比基准模型提高了 1.8%，参数量、浮点运算量和模型体积分别降低了 41.3%、40% 和 41.6%；改进模型具有体积小、准确率高的特点，更适用于机场特种车辆搭载设备。

关 键 词：飞机舱门检测；机场特种车辆；实时检测 Transformer；上下文引导模块；高低频注意力；高级筛选特征融合金字塔

中图分类号：TP391.4; V35

文献标志码：A **文章编号**：1001-5965(2026)11-3790-11

在民航业追求高质量发展的时代背景下，“四型机场”——即平安、绿色、智慧、人文机场的建设理念，为未来中国机场的发展绘制了宏伟蓝图，其中，建设智慧机场不仅是技术创新与应用的集中展现，更是提升运营效率、保障安全服务、优化旅客体验的重要途径。在智慧机场的建设框架中，实现各种机场特种车辆的自动靠机是不可或缺的一环，特别是针对需要精准对接飞机舱门的特种车辆，如旅客登机梯、行动不便旅客登机车、集装货物装载机、散装货物装载机、航空食品车、航空垃圾接收车等。对于这些特种车辆来说，准确快速地获取飞机舱门位置是进行靠机作业的前提。因此，迫切需要开发高效、可靠的手段来完成飞机舱门检测任务。

历年来，相关学者和业界专家们积极运用先进的传感器技术与计算机领域的前沿成果，围绕飞机舱门检测开展了一系列研究。邓览[1]在舱门检测领域引入了双目视觉与图像处理技术，通过精确的边缘检测算法，实现了视觉方法在舱门识别中的初步应用；陈鸿顺等[2]采用高精度激光轮廓仪检测舱门的竖直边缘和下边缘，从而确定舱门位置，利用更为可靠的深度信息提高了舱门检测的精度；Casado Merino 等[3]借鉴人脸识别技术的思路，通过识别舱门周围独特的舷窗特征，为舱门检测提供辅助信息，提升了对不同舱门进行检测的普适性和可靠性；Andonovski 等[4]同样聚焦于利用与舱门相关的元素来辅助检测飞机舱门，但不仅仅局限于舷窗，还考虑了把手、文字标识、脚踏板、箭头指示等多种特征，并提出了一种逻辑算法，从而综合多种特征信息，最终确认舱门的准确位置；叶润等[5]则利用单次多框检测器(single shot multibox detector, SSD)

收稿日期：2024-09-26；**录用日期**：2024-10-25；**网络出版时间**：2024-11-20 18:09
网络出版地址：link.cnki.net/urlid/11.2625.V.20241120.1428.001
基金项目：民航首台（套）重点项目 (3122023PY04)
*通信作者. E-mail：cyzhang@cauc.edu.cn

引用格式：张长勇，张轩铖，乔亚海，等. 面向飞机舱门检测的改进 RT-DETR 模型 [J]. 北京航空航天大学学报，2026，52（11）：3790-3800.
Zhang C Y，Zhang X C，Qiao Y H，et al. Improved RT-DETR model for aircraft door detection [J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52（11）：3790-3800 （in Chinese）.

目标检测算法, 结合轻量级的 MobileNet 特征提取网络, 构造了一种基于 MobileNet-SSD 的舱门检测模型, 能够在保证检测性能的同时, 降低计算复杂度; 谢玉森[6]将相机-激光雷达-惯性测量单元(inertial measurement unit, IMU)传感器组合, 通过提取图像特征、融合深度信息及捕捉动态变化, 运用同时定位与建图(simultaneous localization and mapping, SLAM)技术实现了在复杂和动态环境下的舱门跟踪检测, 为舱门检测技术的发展提供了新的思路与方法。

在机场实际环境下, 采取多传感器融合的方式能够得到较为精确的检测结果, 但其集成度不高, 不易搭载至机场特种车辆上, 且使用激光雷达这类主动式传感器, 可能存在干扰机场其他设备正常运作的风险, 相较之下, 基于深度学习的视觉目标检测方法凭借其简单实用、精准高效、非侵入式的特性, 无疑是更加适合舱门检测的优选方案。然而, 将视觉方法应用于实际场景时, 还需要克服 2 方面的问题: 一方面, 飞机舱门检测系统需要利用边缘端设备作为载体, 但高精度的目标检测模型往往拥有复杂的网络结构和巨大的参数计算量, 导致其难以部署到计算资源和存储空间均有限的智能边缘设备中; 另一方面, 轻量化模型会牺牲部分准确度, 在天气光照变化的影响下, 难以保证模型对不同大小舱门的检测精确性, 可能出现错检、漏检的情况。因此, 只有设计出检测性能较高且模型大小适中的飞机舱门检测模型才能满足机场特种车辆使用需求。

综上, 本文提出一种基于改进实时检测 Transformer(real-time detection Transformer, RT-DETR)的飞机舱门检测模型。该模型主要检测舱门的形状、轮廓及舱门与机身相对位置等特征, 并针对这些特征易受环境干扰、模型不易部署的问题, 引入轻量级的上下文引导模块[7], 对原始主干网络中的 BasicBlock 模块重新设计, 增强模型对上下文信息的利用, 减少模型的参数量和计算量; 在颈部编码网络部分融入高低频(high-low frequency, HiLo)注意力机制, 强化内尺度特征交互(intra-scale feature interaction, AIFI)模块特征捕捉能力; 同时, 采用新的高级筛选特征金字塔网络(high-level screening feature pyramid network, HS-FPN)替换跨尺度特征融合模块(cross-scale feature fusion module, CCFM), 并借鉴坐标注意力(coord attention, CA)针对 HS-FPN 独特的筛选机制进行二次改进(记为 CA-HS-FPN), 提升模型对检测目标特征的表征能力和融合效率。利用实验验证网络改进的有效性, 力求在降低模型参数与计算量的同时, 提高模型的检测精度。

1　RT-DETR 模型结构

RT-DETR[8]是基于 DETR 的实时端到端目标检测器。DETR[9]采用 Transformer 架构, 通过自注意力机制在整个图像中捕捉长程依赖关系, 实现端到端的目标检测。RT-DETR 在此基础上, 通过将多尺度特征交互分解为内尺度交互和跨尺度融合的两步操作, 显著减少了计算资源消耗, 避免了传统卷积神经网络方法中复杂的候选区域生成和非极大值抑制(non-maximum suppression, NMS)等后处理步骤, 相较于主流的更快的基于区域的卷积神经网络(faster region-based convolutional neural network, Faster R-CNN)、YOLO 等方法, 在速度和精度上均有显著提升, 更适用于实时应用场景。在 RT-DETR 目标检测模型中, 官方共给出以 ResNet 为主干的 RT-DETR-R18、RT-DETR-R34、RT-DETR-R50、RT-DETR-R101 和以 HGNet 为主干的 RT-DETR-L、RT-DETR-X 等 12 个版本, 考虑到机场特种车辆对于飞机舱门检测模型的实际需求, 本文选择网络深度与检测精度较为适中的 RT-DETR-R18 作为基础模型, 其具体网络结构主要包含主干网络、高效混合编码器、不确定性-最小查询选择机制、搭载辅助预测头的解码器 4 个部分, 如图 1 所示。

主干网络采用 ResNet-18[10]实现特征提取, 构成部分有卷积层、最大池化层(MaxPool)及具有残差结构的 BasicBlock。卷积层用于初步提取图像特征, 通过卷积、批量标准化和激活函数来增强特征表达能力; 最大池化层则用于下采样, 减少特征图的尺寸; 在深层特征提取阶段, 网络使用多个 BasicBlock 来进一步提取复杂的图像特征, 每个阶段的 BasicBlock 具有不同的输出通道数, 从而能够捕捉多尺度的特征。

高效混合编码器以主干网络最后 3 个阶段的输出特征 $\boldsymbol{S}_3$、$\boldsymbol{S}_4$、$\boldsymbol{S}_5$ 作为输入, 利用 AIFI 和 CCFM 将多尺度特征转换为图像特征序列。其中, AIFI 只在高级特征 $\boldsymbol{S}_5$ 上执行内尺度交互, 大大减少了计算冗余; CCFM 将由卷积层和 RepC3 组成的融合块插入到相邻特征的融合路径中, 对不同路径输出通过逐元素加法融合, 确保融合过程中的特征信息不会丢失。编码器计算过程为

$$\boldsymbol{Q}=\boldsymbol{K}=\boldsymbol{V}=\mathrm{Flatten}(\boldsymbol{S}_5) \tag{1}$$

$$\boldsymbol{F}_5=\mathrm{Reshape}(\mathrm{Aifi}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})) \tag{2}$$

$$\boldsymbol{O}=\mathrm{Ccfm}(\{\boldsymbol{S}_3,\boldsymbol{S}_4,\boldsymbol{F}_5\}) \tag{3}$$

式中: $\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 分别表示不同的数据矩阵或向量,

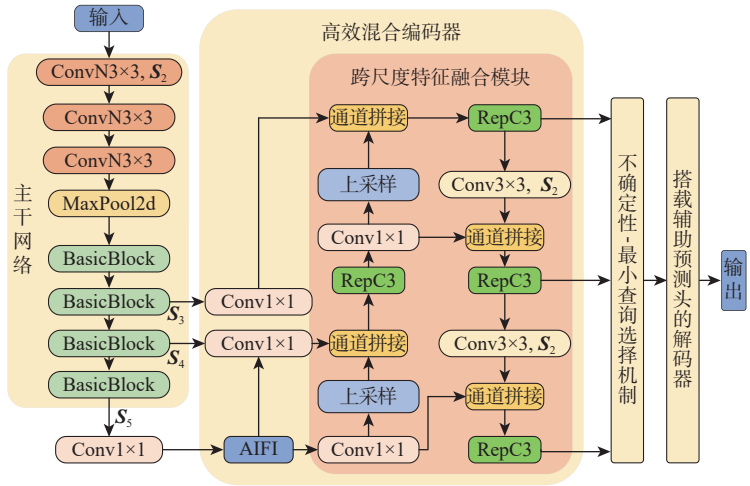


图 1 RT-DETR 模型结构

Fig. 1 RT-DETR model structure

$\boldsymbol{Q}$(查询)包含了当前需要关注或查询的信息，K(键)提供了用于与 $\boldsymbol{Q}$ 进行比较的基准信息，V(值)包含了实际的信息内容；Flatten 表示将 S_5 特征展平并重新排列，以适应 Transformer 的输入格式；Reshape 表示将处理后的特征图恢复为原始的空间维度和通道数；Aifi 表示内尺度特征交互操作；Ccfm 表示跨尺度特征融合操作。

不确定性-最小查询选择机制特别定义了编码器输出特征在分类与定位预测上的联合不确定性，并将该不确定性集成到损失函数中用以指导模型的优化过程，从而为解码器提供具有更高分类得分和更高交并比(intersection over union, IoU)[11]得分的初始查询，搭载辅助预测头的解码器会对初始查询进一步迭代优化，最终输出精确的预测框及与之对应的高置信度评分。联合不确定性 U 和损失函数 L 分别表示为

$$U(\hat{\boldsymbol{x}}) = \|\boldsymbol{P}(\hat{\boldsymbol{x}}) - \boldsymbol{C}(\hat{\boldsymbol{x}})\| \qquad \hat{\boldsymbol{x}} \in \mathbf{R}^D \tag{4}$$

$$L(\hat{x},\hat{y},y) = L_{\text{box}}(\hat{b},b) + L_{\text{cls}}(U(\hat{x}),\hat{c},c) \tag{5}$$

式中：$\hat{\boldsymbol{x}}$ 为编码器特征；$\boldsymbol{P}$ 为位置预测；$\boldsymbol{C}$ 为分类预测；$\hat{y} = \{\hat{c},\hat{b}\}$ 和 y 分别为预测值和真值，$\hat{c}$ 和 $\hat{b}$ 分别表示类别框和边界框，b 表示标注框真实值，c 表示类别真实值。

2 检测模型改进设计

针对因天气光照变化、机身涂装干扰及设备限制而导致的舱门检测精度低、模型部署困难等问题，通过引入上下文引导模块，优化主干网络中的 BasicBlock，增强对上下文信息的利用，降低计算资源的消耗；在颈部编码网络，将 HiLo 注意力机制与 AIFI 相结合，形成 AIFI-HiLo 模块，增强特征表达，降低漏检率和误检率；采用 CA-HS-FPN 替换 CCFM，增强模型对多尺度特征的融合能力。最终得到优化后的飞机舱门检测模型，其整体结构如图 2 所示。

2.1 特征提取网络改进

在 RT-DETR-R18 主干网络中，BasicBlock 作为基础构建模块，旨在高效提取并传递输入特征图的信息，以支持后续复杂任务的处理。为了从来自上一层网络的输入特征图中提取出丰富的局部特征，

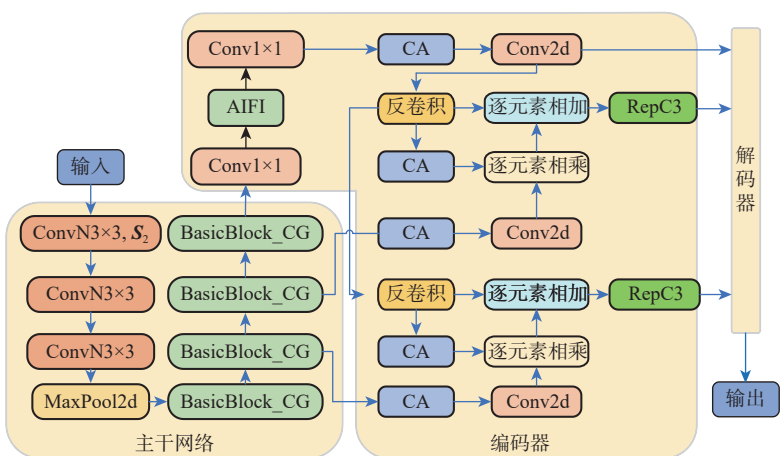


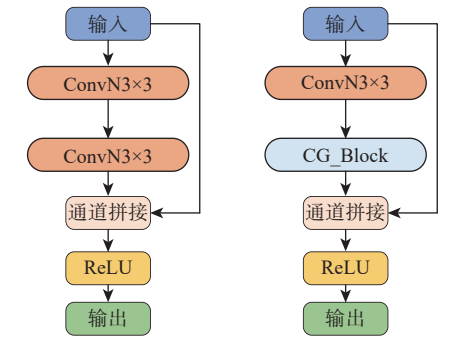
图 2 改进后的飞机舱门检测模型结构

Fig. 2 Improved aircraft door detection model structure

BasicBlock 内部部署了 2 个连续的卷积规范层。每个卷积规范层均集成了卷积层(Conv)、批量归一化(BN)及 ReLU[12] 激活函数, 能够有效地提高模型的性能和训练速度。此外, BasicBlock 采用了残差连接设计, 其将模块的输入直接加至第 2 个卷积规范层的输出上, 形成了一条“捷径”, 使得输入信息能够无损地跨越多个层级, 直接参与到后续特征的构建中, 有效缓解了梯度消失问题, 提高了网络的训练稳定性。

然而, BasicBlock 主要聚焦于局部特征的提取与传递, 而对全局上下文信息的利用相对不足[13]。上下文信息对于理解飞机舱门在图像中的位置关系至关重要, 其能够提供额外的线索来帮助模型更准确地识别和定位舱门。为了提高模型对全局上下文信息的利用率, 如图 3 所示, 引入上下文引导模块 CG_Block 替换 BasicBlock 中的第 2 个卷积规范层, 构成新的 BasicBlock_CG 模块, 可以在提高计算效率的同时, 更好地捕捉长距离依赖关系, 增强模型利用周围信息来推断被遮挡或噪声干扰部分的能力, 以应对图像中如飞机机身涂装鲜艳、恶劣天气影响、光照反射影响、出现遮挡物等复杂场景, 使模型在各种情况下都能够保持稳定的飞机舱门检测性能。

上下文引导模块的基本原理是模拟人类视觉系统依赖上下文信息来辨识物体, 通过捕获局部特征、周围上下文和全局上下文, 并将这些信息融合起来, 以提高特征提取的准确性。模块结构如图 4 所示。其中, 局部特征提取器使用标准 3×3 卷积层, 从 8 个相邻特征向量中学习局部特征, 这些局部特征随后会与周围上下文特征结合, 形成网络对图像各个区域的全面理解。周围上下文特征提取



(a) BasicBlock 基础结构　(b) BasicBlock_CG 基础结构

图 3　改进前后 BasicBlock 对比

Fig. 3　Comparison of BasicBlock before and after improvement

器与局部特征提取器同时进行工作, 通过空洞/膨胀卷积层增加感受野的大小, 从而能够在更大的区域内捕获丰富的上下文信息, 帮助模型更好地理解复杂场景。局部特征提取器与周围上下文特征提取器均采用了通道卷积, 消除了跨通道的计算成本, 能够大大减小模型的参数量及计算量, 以方便其部署至计算资源有限的设备。联合特征提取器通过连接层、BN 及参数化 ReLU(PReLU)操作整合提取出局部特征和周围上下文特征, 使得网络在局部精细度和全局上下文间达到平衡, 从而提高目标检测的准确性; 全局特征提取器使用全局平均池化层(global average pooling, GAP)聚合全局信息, 并通过全连接层(fully connected, FC)进一步细化全局上下文特征, 全局上下文被视为加权向量, 逐通道地重新标定联合特征, 强调有用的特征部分, 抑制不重要的特征部分, 以增强联合特征提取器学习到的特征。CG_Block 内部也采用了全局残差学习, 添加了从输入到全局特征提取器的连接, 有助于学习高度复杂的特征。

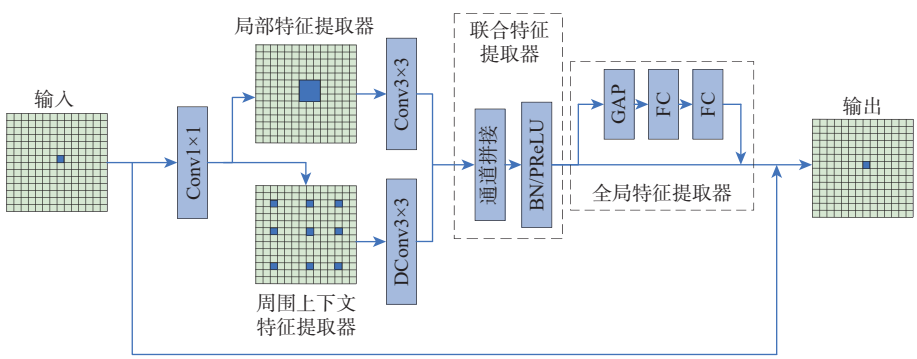


图 4　CG_Block 结构

Fig. 4　CG_Block structure

2.2　AIFI-HiLo 模块设计

AIFI 与传统的 Transformer 编码器设计类似, 都离不开多头自注意力(multi-head self-attention, MSA)机制[14]。MSA 通过并行处理和从不同表示子空间中捕捉信息, 每个头都能独立地进行 Q(查询)、K(键)、V(值)的线性变换和注意力计算, 提高了模型的并行性和表达能力, 从而能够更有效地捕获输入数据中的多种依赖关系。然而, 标准多头自注意

力层往往在所有图像块上强制应用相同的全局注意力，忽略了特征存在不同的基础频率，导致模型在高分辨率图像上需要消耗大量的计算资源。对此，在AIFI中引入HiLo注意力[15]可以有效解决该问题。HiLo注意力的结构如图5所示。图中，N_h为该层中的自我关注头的总数，a为高频或低频磁头的分频比。

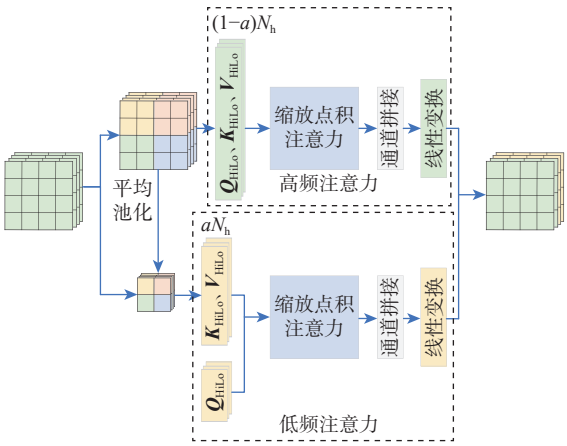


图5 HiLo注意力结构

Fig. 5 HiLo attention structure

HiLo注意力将MSA按照特征高低频划分为2个部分：高频注意力部分(high-frequency attention, Hi-Fi)通过局部自注意力和高分辨率特征图对高频交互进行编码，减少了长距离依赖的处理需求，对于实现快速且高效的视觉变换器至关重要；低频注意力部分(low-frequency attention, Lo-Fi)通过全局注意力和下采样特征对低频交互进行编码，减少了全局自注意力的计算负担，从而大大提高编码效率。具体来说，HiLo注意力将MSA中相同数量的磁头分为2组，分频比为a，其中，$(1-a)N_h$个磁头用于高频注意力，其他N_h个用于低频注意力，这种分配策略确保了不同类型的特征得到适当的关注和处理。高频注意力部分专注于图像中的高频细节，如边缘、纹理等，通过局部窗口自注意力(在2×2的窗口中进行)捕获细粒度高频，并采用简单的非重叠窗口分区，减少了需要计算的自注意力的范围和冗余计算，大大节省了计算复杂性。低频注意力部分接收每个窗口经平均池化后获得的低频，将平均池化后的特征图投影成K(键)和V(值)，并保留来自原始特征图的Q(查询)，随即通过全局注意力来捕获特征图中丰富的低频信息，如形状、颜色分布等，其主要作用于下采样后的特征图上，可以减少处理的特征数量，进一步提高效率。最终，将改进的高频、低频特征整合起来，作为AIFI-HiLo模块的输出。由于每个注意力的复杂性都低于标准MSA，因此，AIFI-HiLo的整个框架保证了低复杂性，并确保了GPU上的高吞吐量，同时增强了模型捕获目标特征的能力。AIFI-HiLo计算过程为

$$\boldsymbol{Q}_{\mathrm{HiLo}}=\boldsymbol{K}_{\mathrm{HiLo}}=\boldsymbol{V}_{\mathrm{HiLo}}=\mathrm{Flatten}(\boldsymbol{S}_5) \tag{6}$$

$$\mathrm{AIFI\text{-}HiLo}(\boldsymbol{x})=[\mathrm{HiFi}(\boldsymbol{x}),\mathrm{LoFi}(\boldsymbol{x})] \tag{7}$$

$$\boldsymbol{F}_5=\mathrm{Reshape}(\mathrm{AIFI\text{-}HiLo}(\boldsymbol{Q}_{\mathrm{HiLo}},\boldsymbol{K}_{\mathrm{HiLo}},\boldsymbol{V}_{\mathrm{HiLo}})) \tag{8}$$

2.3 CA-HS-FPN

2.3.1 HS-FPN

HS-FPN[16]的结构如图6所示，包括特征选择和特征融合2个主要组成部分。

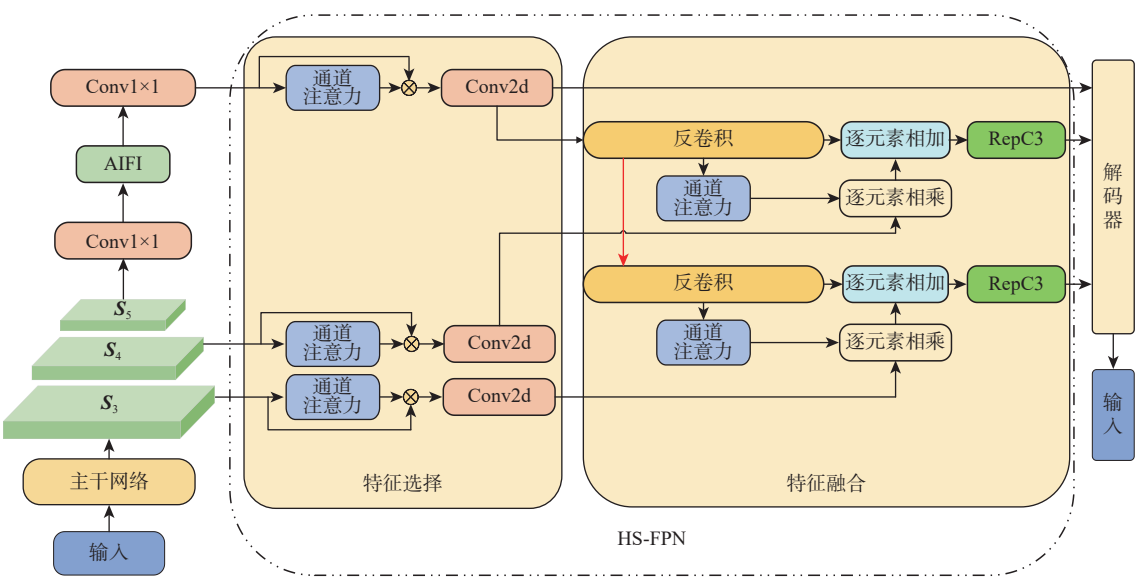


图6 HS-FPN结构

Fig. 6 HS-FPN architecture

在特征选择过程中，通道注意力[17]模块首先采用全局平均池化和全局最大池化对输入特征图进行处理，分别计算每个通道的平均值和最大值。然后，使用Sigmoid激活函数来确定每个通道的权重

值, 从而得到每个通道的权重。最后, 通过将权重信息与相应比例的特征图相乘来生成过滤后的特征图, 实现对不同尺度特征的筛选。最大池化的主要目的是从每个通道中提取最相关的数据, 而平均池化旨在统一地从特征图中获取所有数据, 从而最大限度地减少过度损失, 结合这些池化方法有助于从每个通道中提取最具代表性的信息, 同时最大限度地减少信息损失。

在特征融合模块中, 为了充分利用高级特征包含的丰富语义信息和低尺度特征提供的精确目标位置, HS-FPN 使用了选择性特征融合(selective feature fusion, SFF)模块。该机制不同于传统的融合方法, 不是将上采样的高级特征和低尺度特征像素值进行简单求和, 而是通过使用高级特征作为权重来过滤低尺度特征中包含的必要语义信息, 再将过滤后的低尺度特征与高级特征协同集成, 大大减少了融合过程中的信息冗余, 增强了模型的特征表达能力。SFF 结构如图 7 所示。图中: $\boldsymbol{P}_i$ 为低层原始输入特征图, $\boldsymbol{N}_i$ 为 SFF 融合输出后的新特征图。

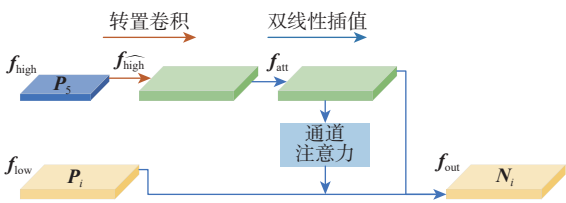


图 7　SFF 结构

Fig. 7　SFF structure

当给定一个输入高级特征 $\boldsymbol{f}_{\text{high}} \in \mathbf{R}^{C\times H\times W}$ 和一个输入低尺度特征 $\boldsymbol{f}_{\text{low}} \in \mathbf{R}^{C\times H\times W_1}$ 时, 高级特征首先使用一个步幅大小为 2、卷积核大小为 3×3 的转置卷积(T-Conv)进行扩展, 得到特征大小 $\boldsymbol{f}_{\widehat{\text{high}}} \in \mathbf{R}^{C\times 2H\times 2W}$。其次, 为统一高级特征和低尺度特征的维度, 使用双线性插值来向上或向下采样高级特征, 得到特征 $\boldsymbol{f}_{\text{att}} \in \mathbf{R}^{C\times H_1\times W_1}$。然后, 使用通道注意力模块将高级特征转换为相应的注意力权重, 在获得具有一致维度的特征时过滤低级特征。最后, 将过滤后的低尺度特征与高级特征融合, 以增强模型的特征表示, 并得到 $\boldsymbol{f}_{\text{out}} \in \mathbf{R}^{C\times H_1\times W_1}$。其中, C 为通道数, H 为高级特征图的高度, W 为高级特征图的宽度, W_1 为低尺度特征图的宽度, H_1 为将高级特征进行双线性插值处理后的特征图高度。在图像采样过程中, 通过采用转置卷积和双线性插值相结合的方法, 有效增强了模型的检测能力。SFF 选择特征融合过程为

$$\boldsymbol{f}_{\text{att}} = \text{BL}\left(\text{T-Conv}\left(\boldsymbol{f}_{\text{high}}\right)\right) \tag{9}$$

$$\boldsymbol{f}_{\text{out}} = \boldsymbol{f}_{\text{low}}\text{CA}\left(\boldsymbol{f}_{\text{att}}\right) + \boldsymbol{f}_{\text{att}} \tag{10}$$

式中: BL 表示双线性插值; CH 表示通道注意力。

2.3.2　坐标注意力机制

获取飞机舱门的位置和形状是完成飞机舱门检测任务的关键因素, 考虑到目标图像在实际情况中可能会出现不同尺度、受天气光照影响发生形变及遮挡等多种问题, 改进模型利用坐标注意力[18]替换 HS-FPN 中原始的通道注意力模块, 通过在特征图上引入空间坐标信息, 实现了对舱门位置的细粒度权重调整, 能够更好地捕捉舱门的精确位置和轮廓, 同时自适应地处理不同视角和光照条件下的变化, 有效整合舱门与周围环境信息, 增强 HS-FPN 的特征选择和 SFF 模块过滤低尺度特征的能力, 进而提高检测精度。此外, 相比于通道注意力及其他复杂的注意力机制, 坐标注意力整体采用轻量级设计, 易于集成至现有网络, 不会显著增加模型的复杂性和计算负担, 更为适用飞机舱门检测的现场部署应用场景。

坐标注意力模块结构如图 8 所示, 主要通过 2 个步骤来捕获通道间的依赖关系和位置信息。首先, 沿着一个空间维度(如宽度或高度)收集特征。然后, 沿着另一个维度聚合这些特征。具体而言, 坐标注意力模块使用尺寸为 $(H,1)$ 和 $(1,W)$ 池化核分别沿着输入的水平方向和垂直方向对每个通道进行编码, 高度为 h 时的第 c 个通道输出 $Z_c^h(h)$ 可表示为

$$\boldsymbol{Z}_c^h(h) = \frac{1}{W}\sum_{0\leqslant i\leqslant W}\boldsymbol{X}_c(h,i) \tag{11}$$

同理, 宽度为 w 时的第 c 个通道的输出 $Z_c^w(w)$ 可表示为

$$\boldsymbol{Z}_c^w(w) = \frac{1}{H}\sum_{0\leqslant j\leqslant H}\boldsymbol{X}_c(j,w) \tag{12}$$

式中: $\boldsymbol{X}_c(h,i)$ 为输入特征图沿水平方向的输入; $\boldsymbol{X}_c(j,w)$ 为输入特征图沿垂直方向的输入。

具有特定方向的特征图编码为 2 个保存位置信息的特征图, 位置信息指特征图沿水平和垂直方向上提取的信息, 之后使用 1×1 卷积对其进行变换操作:

$$\boldsymbol{f} = \delta\left(F_1\left[\boldsymbol{z}^h, \boldsymbol{z}^w\right]\right) \tag{13}$$

$$\boldsymbol{g}^h = \sigma\left(\boldsymbol{F}_h\left(\boldsymbol{f}^h\right)\right) \tag{14}$$

$$\boldsymbol{g}^w = \sigma\left(\boldsymbol{F}_w\left(\boldsymbol{f}^w\right)\right) \tag{15}$$

式中: F_1 表示卷积; δ 表示 Sigmoid 激活函数。水平和垂直方向上的空间信息包含进中间特征图 $\boldsymbol{f}$ 中, $\boldsymbol{f}^h \in \mathbf{R}^{C/r\times H}$ 和 $\boldsymbol{f}^w \in \mathbf{R}^{C/r\times W}$ 是将 $\boldsymbol{f}$ 沿水平和垂直方向拆分而来的 2 个特征图, 注意力权重 $\boldsymbol{g}^h$ 和 $\boldsymbol{g}^w$ 是经过 $\boldsymbol{f}^h$ 和 $\boldsymbol{f}^w$ 卷积将 $\boldsymbol{f}^h$ 和 $\boldsymbol{f}^w$ 这 2 个特征图的通道数转化

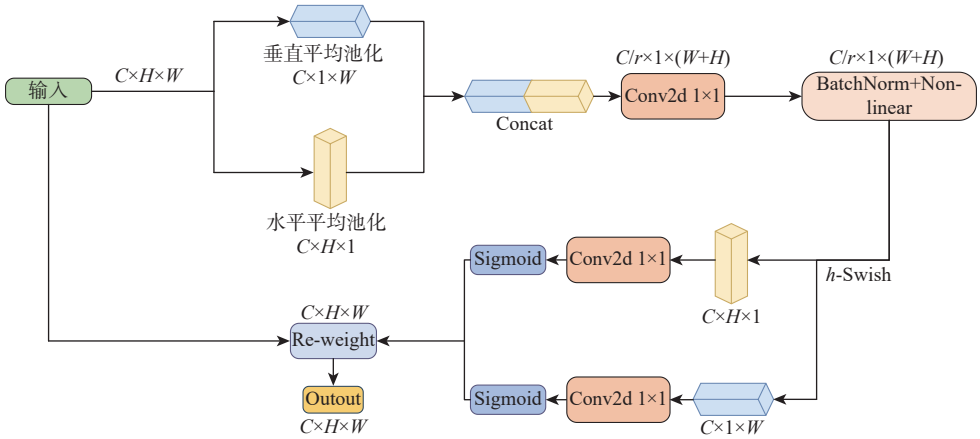


图 8 坐标注意力模块结构

Fig. 8 Structural of coord attention module

而来的，$\boldsymbol{g}^h$ 和 $\boldsymbol{g}^w$ 的通道数和输入特征图相同。

最后，利用乘法将权重 $\boldsymbol{g}^h$ 和 $\boldsymbol{g}^w$ 应用于输入特征图，得到注意力模块输出：

$$\boldsymbol{y}_c(i,j)=\boldsymbol{x}_c(i,j)\cdot\boldsymbol{g}_c^h(i)\cdot\boldsymbol{g}_c^w(i) \tag{16}$$

式中：$\boldsymbol{y}_c(i,j)$为第 c 个通道的输出；$\boldsymbol{x}_c(i,j)$为第 c 个通道的输入；$\boldsymbol{g}_c^h(i)$为第 c 个通道沿水平方向的权重；$\boldsymbol{g}_c^w(i)$为第 c 个通道沿垂直方向的权重。

3 实验结果与分析

3.1 实验环境

实验环境配置如下：基于 PyTorch 2.3.0 深度学习框架，系统环境为 Linux 操作系统（Ubuntu 22.04.4LTS），CPU 型号为 Intel® Xeon(R) Platinum 8256 CPU @ 3.80 GHz × 16，图像处理器（Graphics processing unit，GPU）型号为 NVIDIA Corporation GA100，40 GB 显存，CUDA 版本为 12.1。模型超参数设置如下：训练迭代周期数 epoch=150，各批次处理的样本数量 batch size=8，输入图像尺寸统一为 640 像素×640 像素，用于数据加载的线程数量为 4，使用 AdamW 作为优化器，网络初始学习率为 0.000 1，权重衰减系数为 0.000 1，其他超参数均保持默认值。

3.2 数据集搭建

由于目前没有针对飞机舱门的公共数据集，本文使用的数据集为自建数据集，图片来源主要是互联网、国内某机场及专用 3D 模拟软件，共 2 167 张图像，同一张图像上可能包含多种舱门，考虑到图像数量和雨雪等恶劣条件下的样本较少，经过旋转、水平翻转、亮度调节、对比度调节、增加噪声等数据增强技术，将图像扩增到 10 835 张，保证数据集的多样性，提高模型的鲁棒性。该数据集主要包含客舱门与货舱门 2 种典型飞机舱门目标，其中，客舱门目标为 14 145 个，货舱门目标为 7 360 个。数据集图像标注软件采用 LabelImg 工具，标注为 VOC 格式的标签文件，训练集、测试集和验证集的划分比例为 7∶2∶1。部分数据集图像如图 9 所示。

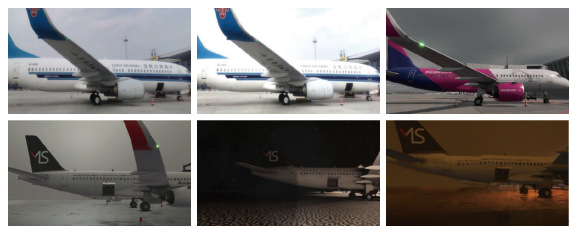


图 9 不同情况下的数据图像

Fig. 9 Images of data in different situations

3.3 评价指标

本文实验采用目标检测任务中常用的精度评估指标 mAP50 和 mAP50-90。mAP 是多个类别平均准确率（average precision，AP）的平均值，mAP 的数值越大，说明此检测模型的性能越优，是非常重要的评价指标。模型复杂度上选用参数量、浮点运算量及模型体积 3 种指标。其中，模型体积表示训练结束后生成权重文件的大小。比较不同模型之间的计算复杂度大小，参数量、浮点运算量、模型体积越大，模型越复杂，同时能够“吞吐”的数据量也越多，能完成复杂任务，但对硬件计算资源有更大要求；较小的参数量、浮点运算量和模型体积则表示任务或模型的计算需求较低，可能更加高效或轻量。特种车辆上外接设备大多为工控机、单片机等，硬件计算水平不高，因此，低参数量、低浮点运算量、模型体积小的检测模型更适合在特种车辆设备中使用。

3.4 消融实验

模型结构框架的改进基于 RT-DETR 模型，为验证各改进的有效性，在自建飞机舱门数据集上对检测精度与模型复杂度进行消融实验，实验结果如表 1 所示。由实验结果可知，加入 CG_Block 设计之后，模型的浮点运算量从 5.69×10^{10} 次降低到

4.29×10^{10} 次, 体积从 38.5 MB 减少到 27.5 MB, mAP50 从 93.1% 提升到 94.5%, 表明 CG_Block 能够在减少计算复杂度和模型大小的同时提高检测性能。单独添加 AIFI-HiLo 时, 模型的浮点运算量和体积几乎不变, 但 mAP50 从 93.1% 提升到 94.7%, 表示 AIFI-HiLo 能够在不改变模型复杂度的基础上显著提升检测精度。单独使用 CA-HS-FPN 网络结构替换 CCFM 模块时, 模型大小虽然只减小到了 33.4 MB, 但检测精度提升明显, 表示其同样能够在参数量和浮点运算量有所减少的情况下提升检测精度。当同时添加 2 个改进时(No.5、No.6、No.7), 可以看到, 模型仍能保留 CG_Block 和 CA-HS-FPN 降低计算复杂度的优点, 同时将检测精度保持或提升至较高水平。将 3 种改进方法结合起来, 模型的整体表现达到最优, 相较于原始的 RT-DETR 模型在参数量、浮点运算量、模型体积上分别减少了 41.3%、40%、41.6%, 在检测精度上提升了 1.8%, 表明改进后的飞机舱门检测模型在保持高检测率的同时, 实现了模型的轻量化, 使模型更契合机场特种车辆搭载舱门检测设备要求。

表 1　消融实验结果

Table 1　Ablation experiments results

No.	CG_Block	AIFI-HiLo	CA-HS-FPN	参数量	浮点运算量/10^{10}次	模型体积/MB	mAP50/%
1				19 874 328	5.69	38.5	93.1
2	√			14 066 768	4.29	27.5	94.5
3		√		19 748 888	5.70	38.5	94.7
4			√	17 406 152	4.79	33.4	94.7
5	√	√		13 941 328	4.29	27.3	94.9
6	√		√	12 275 712	3.92	23.9	94.3
7		√	√	17 280 712	4.80	33.2	94.4
8	√	√	√	11 676 160	3.41	22.5	94.9

3.5　对比实验

为验证本文改进模型在飞机舱门检测中具有优异的性能, 使用自建飞机舱门数据集, 挑选当今主流目标检测模型进行对比实验, 涵盖 YOLO 系列中模型较小的 YOLOv5m[19]、YOLOv6s[20]、YOLOv7-tiny[21]、YOLOv8s、YOLOv8m 版本。评价指标选取参数量、浮点运算量、模型体积、mAP50 及 mAP50-90 进行模型性能评估, 对比结果如表 2 所示。根据对比结果可以看出, 本文改进模型在多个关键指标上均展现出卓越的性能。与当前流行的 YOLO 系列模型相比, 改进模型的参数量、浮点运算量、模型体积分别为 11 676 160、3.41×10^{10} 次、22.5 MB, 较 YOLOv5m、YOLOv8m 大大减少, 与 YOLOv6s、YOLOv8s[22] 相近, 但高于 YOLOv7-tiny, 处于适中水平, 不会过于庞大以致于计算资源消耗过高, 也不会过于精简而牺牲太多性能, 实现了良好的平衡, 既降低了对硬件资源的需求, 又确保了高效的推理速度, 非常适合资源受限或需要快速响应的应用场景。与较为新颖的模型 RT-DETR-ELA-HS-FPN 和 RT-DETR-CAA-HS-FPN 相比, 本文改进模型不仅模型体积更小, 而且检测精度较高, 说明了利用坐标注意力机制进行优化的思路尽管较为常见, 但用在飞机舱门检测场景中更具备适用性和有效性, 有利于工程应用的落实。在平均精度方面, 本文改进模型不仅在常用的 mAP50 评估标准下达到了 94.9% 的高水平, 与 YOLOv8m 等先进模型相媲美, 甚至在更为严格的 mAP50-90 评估标准下也展现出了出色的性能, 达到了 71.5%, 充分证明了本文改进模型在复杂场景下的鲁棒性和准确性。由此可知, 本文改进模型通过优化网络结构和参数配置, 成功地在保持高效计算和较小体积的同时, 实现了高精度的飞机舱门检测性能, 为深度学习在飞机舱门检测领域的应用提供了新的思路和解决方案。

3.6　检测结果

本文通过分析 RT-DETR 基准模型与改进模型的测试效果图来评估其性能, 相关测试效果如图 10 和图 11 所示。从结果中可以看出, 改进后的检测模型通过捕获舱门的形状、轮廓及舱门与机身相对位置等特征, 在不同环境条件下均表现出较为稳定的检测效果, 能够准确检测出飞机处于 20 m 远时, 在雨雪雾天气、强弱光照等情况下飞机舱门的位置及类型, 置信度较高, 预测框能完整包裹住舱门和检测出未完全打开的舱门, 具体如图 10 所示。

相比之下, 基准网络在光线暗淡时易出现未检测到舱门轮廓特征的情况(无论舱门是否关闭), 即使检测到了舱门, 也存在置信度较低的问题, 如图 11(a)和图 11(b)所示; 在舱门边界受飞机涂装影响过大时往往会检测不到关闭状态的舱门轮廓, 如图 11(c)所示; 当图片中背景线条过于复杂时, 线条构成的轮廓特征易导致模型错检其为舱门轮廓特

表2 对比实验结果

Table 2 Comparative experiments results

模型	参数量	浮点运算量/10^{10}次	模型体积/MB	mAP50/%	mAP50-90/%
YOLOv5m	25 046 374	6.40	48.2	94.6	70.4
YOLOv6s	16 297 814	4.40	31.3	93.7	67.4
YOLOv7-tiny	6 017 694	1.32	12.3	91.4	60.9
YOLOv8s	11 126 358	2.84	21.5	94.7	68.9
YOLOv8m	25 840 918	7.87	49.6	94.8	71.2
RT-DETR-ELA-HS-FPN	18 465 296	5.48	36.8	94.5	70.3
RT-DETR-CAA-HS-FPN	18 258 072	5.34	36.1	94.2	66.8
本文改进模型	11 676 160	3.41	22.5	94.9	71.5

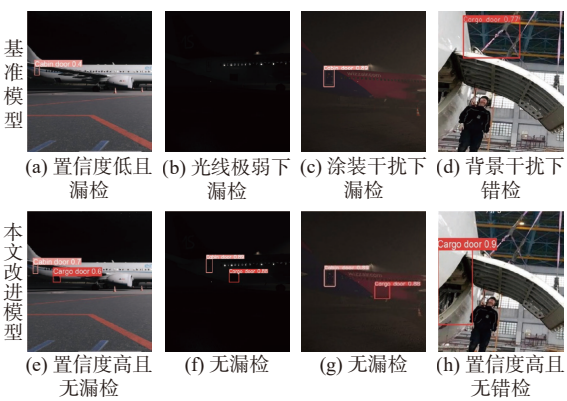


(a) 雨天 (b) 雪天 (c) 雾天

(d) 夜晚 (e) 日间 (f) 正午

图10 不同环境条件下改进后模型的检测效果

Fig. 10 Detection effect of the improved model under different environmental conditions



(a) 置信度低且漏检 (b) 光线极弱下漏检 (c) 涂装干扰下漏检 (d) 背景干扰下错检

(e) 置信度高且无漏检 (f) 无漏检 (g) 无漏检 (h) 置信度高且无错检

图11 改进后的RT-DETR模型检测效果

Fig. 11 Detection effect of improved RT-DETR model

征，如图11(d)所示。这些极端环境下的检测效果，对于准确度要求极高的飞机舱门检测任务来说至关重要，改进后的模型在同样的环境下展现出了更具优势的检测性能，如图11(e)～图11(h)所示，不仅避免了错检漏检，而且保证了较高的置信度，更适合完成复杂条件下的飞机舱门检测。

4 结 论

1）本文模型可实现较为优异的检测性能，应对包含舱门过小、机身涂装影响过大、雨雪恶劣天气等复杂条件的数据集，在mAP50评估标准下达到94.9%的高水平，在更为严格的mAP50-90标准下也能够达到71.5%，为深度学习在飞机舱门视觉检测领域的应用提供了研究思路与可行性验证。

2）本文模型在稳定提升检测准确率同时，相较于原始模型将参数量、浮点运算量、模型体积分别降低至11 676 160、3.41×10^{10}次和22.5 MB，实现了模型的参数量、浮点运算量和模型体积大幅度降低，减轻了模型对硬件计算资源的要求，更符合模型在机场特种车辆应用场景中的实际部署需求。

为使本文改进模型后续能够更好地在现实场景中应用，仍需在数据集中不断扩充实际采集的各种复杂情况下的高分辨率舱门图像，并将模型部署进特种车辆系统中进一步实测验证。

参考文献（References）

[1] 邓览. 基于计算机视觉的登机桥全自动接机系统设计[J]. 科技创新导报, 2012, 9(13): 43-46.
Deng L. Design of automatic pick-up system for boarding bridge based on computer vision[J]. Science and Technology Innovation Herald, 2012, 9(13): 43-46(in Chinese).

[2] 陈鸿顺, 卓志辉, 陈汉樑. 自动乘客登机桥对接系统: CN201680060383.9[P]. 2018-06-08.
Chen H S, Zhuo Z H, Chen H L. Automated passenger boarding bridge docking system: CN201680060383.9[P]. 2018-06-08(in Chinese).

[3] Casado Merino J M, Belmecho A M, Freez Castro A, 等. 用于将乘客登机桥自动对接至飞机的方法: CN201980036068.6[P]. 2024-05-14.
Casado Merino J M, Belmecho A M, Freez Castro A, et al. Method for automatic docking of passenger boarding bridge to aircraft: CN201980036068.6[P]. 2024-05-14(in Chinese).

[4] Andonovski B, Wang J L. Computer vision system for cabin door detection and location[C]//Proceedings of the 2018 15th International Conference on Control, Automation, Robotics and Vision. Piscataway: IEEE Press, 2018: 1537-1542.

[5] 叶润, 张成, 李旭, 等. 基于计算机视觉的客机舱门识别与定位技

术研究[J]. 计算机测量与控制, 2021, 29(3): 224-229.
Ye R, Zhang C, Li X, et al. Research on recognition and positioning technology of passenger aircraft door based on computer vision[J]. Computer Measurement & Control, 2021, 29(3): 224-229 (in Chinese).

[6] 谢玉森. 面向登机廊桥自动导引的多源异构传感器数据融合和实时跟踪导航[D]. 北京: 北京信息科技大学, 2023.
Xie Y S. Multi-source heterogeneous sensor data fusion and real-time tracking navigation for automatic guidance of boarding corridor bridge[D]. Beijing: Beijing Information Science & Technology University, 2023(in Chinese).

[7] Wu T Y, Tang S, Zhang R, et al. CGNet: a light-weight context guided network for semantic segmentation[J]. IEEE Transactions on Image Processing, 2021, 30: 1169-1179.

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

[9] Zong Z F, Song G L, Liu Y. DETRs with collaborative hybrid assignments training[C]//Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2023: 6725-6735.

[10] Li Y, Ibanez-Guzman J. Lidar for autonomous driving: the principles, challenges, and trends for automotive lidar and perception systems[J]. IEEE Signal Processing Magazine, 2020, 37(4): 50-61.

[11] Rezatofighi H, Tsoi N, Gwak J, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 658-666.

[12] Agarap A F M. Deep learning using rectified linear units (ReLU) [EB/OL]. (2026-04-14)[2026-09-01]. https://arxiv.org/abs/1803.08375.

[13] Wang S, Jiang H P, Yang J X, et al. Lightweight tomato ripeness detection algorithm based on the improved RT-DETR[J]. Frontiers in Plant Science, 2024, 15: 1415297.

[14] Messaoud W, Trabelsi R, Cabani A, et al. Multi-head self attention for enhanced object detection in the maritime domain[C]//Proceedings of the 2023 International Conference on Cyberworlds. Piscataway: IEEE Press, 2023: 179-184.

[15] Pan Z Z, Cai J F, Zhuang B H. Fast vision transformers with HiLo attention[C]//Proceedings of the Advances in Neural Information Processing Systems 35. San Diego: Neural Information Processing Systems Foundation, Inc., 2022: 14541-14554.

[16] Chen Y F, Zhang C Y, Chen B, et al. Accurate leukocyte detection based on deformable-DETR and multi-level feature fusion for aiding diagnosis of blood diseases[J]. Computers in Biology and Medicine, 2024, 170: 107917.

[17] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 7132-7141.

[18] Hou Q B, Zhou D Q, Feng J S. Coordinate attention for efficient mobile network design[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 13708-13717.

[19] Chen J, Bao E, Pan J Y. Classification and positioning of circuit board components based on improved YOLOv5[J]. Procedia Computer Science, 2022, 208: 613-626.

[20] Li C Y, Li L L, Jiang H L, et al. YOLOv6: a single-stage object detection framework for industrial applications[EB/OL]. (2022-09-07)[2026-09-01]. https://arxiv.org/abs/2209.02976.

[21] Wang C Y, Bochkovskiy A, Liao H M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2023: 7464-7475.

[22] Varghese R M S. 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. Piscataway: IEEE Press, 2024: 1-6.

Improved RT-DETR model for aircraft door detection

ZHANG Changyong[1,*], ZHANG Xuancheng[1], QIAO Yahai[1], LI Zheng[2]

(1. College of Electronic Information and Automation, Civil Aviation University of China, Tianjin 300300, China;
2. College of Computer Science and Technology, Civil Aviation University of China, Tianjin 300300, China)

Abstract: Aircraft door detection is crucial for enabling the automatic docking of airport special vehicles. To address the issues of low detection accuracy and difficult model deployment caused by changes in weather and lighting conditions, interference from body coating, and equipment limitations, an improved real-time detection Transformer RT-DETR based aircraft door detection model is proposed, using a lightweight residual neural network ResNet18 as the backbone feature extraction benchmark network. By introducing the context guided block, the BasicBlock in the backbone network is optimized to enhance the utilization of contextual information and reduce the consumption of computing resources. The AIFI-HiLo module is built in the neck encoder network by incorporating the high-low frequency (HiLo) attention into an intra-scale feature interaction (AIFI) module. This innovation strengthens feature representation, effectively reducing missed and false detections. Replace cross-scale feature fusion module (CCFM) with a two-stage improved high-level screening feature pyramid network (HS-FPN) to enhance the model's ability to fuse multiscale features. The experimental results showed that the average accuracy of the improved model reached 94.9%, which was 1.8% higher than the benchmark model. The parameter count, floating point operation count, and memory usage were reduced by 41.2%, 40% and 41.5%, respectively. The improved model has the characteristics of small size and high accuracy, making it more suitable for airport special vehicles to carry equipment.

Keywords: aircraft door detection; airport special vehicles; real-time detection Transformer; context guidance block; high-low frequency attention; high-level screening feature pyramid network

Received: 2024-09-26; **Accepted**: 2024-10-25; **Published Online**: 2024-11-20 18:09
URL: link.cnki.net/urlid/11.2625.V.20241120.1428.001
Foundation item: The First Civil Aviation Key Project (3122023PY04)
* **Corresponding author**. E-mail: cyzhang@cauc.edu.cn