

FDH-DETR 工况场景工人行为及火灾检测算法 *

董润华¹ 常青² 孔鹏伟³ 王耀力²

(1. 太原理工大学集成电路学院 太原 030600; 2. 太原理工大学电子信息工程学院 太原 030600;
3. 山西众鑫乐食汇食品有限公司 临汾 041000)

摘要: 针对工厂时刻面临的安全生产问题,例如厂区车间内严禁烟火、需时刻关注工作人员的行为安全、恶劣工况场景下工人是否佩戴口罩等,提出了一个基于 RT-DETR 改进的工人行为及火灾检测算法 FDH-DETR。首先,通过 Deep Faster 特征深度融合模块与 FasterNet 的融合,减少了算法的参数量和计算量;其次,通过 DRBC3 模块大小卷积核转换机制,减少了模型的推理成本;最后,通过 HiLo-AIFI 高低频尺度内特征交互模块,增强了对高低频特征的提取能力。实验结果表明,改进后的算法平均准确度达到了 93.8%,参数量减少了 31.6%,计算量减少了 61.4%,FPS 达到了 150 fps,并在真实工况场景下进行推理实验,验证了算法的有效性。

关键词: RT-DETR; 工人行为检测; 注意力机制; 模型轻量化; 目标检测; 特征融合; 火灾检测

中图分类号: TP391.4; TN919.8 **文献标识码:** A **国家标准学科分类代码:** 520.60

FDH-DETR worker behavior and fire detection algorithm in working condition

Dong Runhua¹ Chang Qing² Kong Pengwei³ Wang Yaoli²

(1. School of Electronic Information and Optical Engineering, Taiyuan University of Technology, Taiyuan 030600, China;
2. School of Electronic Information Engineering, Taiyuan University of Technology, Taiyuan 030600, China;
3. Shanxi Zhongxin Leshihui Food Co., Ltd., Linfen 041000, China)

Abstract: Regarding the perennial safety production problems that factories constantly encounter, such as the strict prohibition of smoke and fire in the workshop area, the need for constant attention to the behavioral safety of workers, and whether workers wear masks in adverse working condition scenarios, an improved worker behavior and fire detection algorithm FDH-DETR based on RT-DETR was proposed. Firstly, through the fusion of the Deep Faster feature depth fusion module and FasterNet, the number of parameters and the amount of computation of the algorithm were reduced. Secondly, through the DRBC3 module size convolution kernel conversion mechanism, the inference cost of the model was decreased. Finally, through the HiLo-AIFI high-low frequency scale within-feature interaction module, the extraction ability of high-low frequency features was enhanced. Experimental results indicate that the improved algorithm achieved an average accuracy of 93.8%, a reduction of 31.6% in parameters, a reduction of 61.4% in computation, and an FPS of 150 frames per second. Inference experiments were conducted in real working condition scenarios, verifying the effectiveness of the algorithm.

Keywords: RT-DETR; worker behavior detection; attention mechanism; model lightweighting; object detection; feature fusion; fire detection

0 引言

随着计算机视觉技术的发展,视频监控技术已逐渐走向了智能化。一些可以检测人员状态的智能化监控设备,可以对异常情况进行实时监测提醒,降低了人工成本^[1]。

智能监控技术可以用来记录监控视频、检测人员的行为、检测火灾等。在智能监控设备中主要的技术是目标检测算法,因其要满足实时检测且需部署在性能受限小型设备上,故一些实时性较差或不够轻量化的目标检测算法就不能满足要求^[2]。在保持较高准确率的前提下,实时性和易部署

性是智能监控目标检测算法中两个最大的挑战^[3]。

目前,主流的实时目标检测算法有 YOLO(you only look once)系列和百度飞桨团队提出的 RT-DETR^[4](real-time detection transformer)。RT-DETR 相比于传统的目标检测算法在准确性上有一定的优势,但也存在一些缺点。RT-DETR 采用 Transformer^[5]结构,参数量较大,模型不够轻量化,这使得其在性能受限的小型终端设备上,性能会受到影响。

李海川等^[6]通过在 YOLOv8n 的主干网络中引入 3D 空洞卷积、SE(Squeeze & Excitation)注意力机制和可变形卷积提高模型对矿工行为的检测精度;刘恒等^[7]使用添加通道重排机制的 C2fg 模块代替原本 C2f 模块、添加小目标检测层等方法对 YOLOv8 进行改进,解决了由于摄像头安装位置较高,对小目标检测精度低的问题;罗国富等^[8]对 YOLOv5s 特征融合网络以及输出层进行删除,并进行结构化剪枝和知识蒸馏,对模型进行了轻量化处理,满足了智能车间工人不安全行为的实时性检测和边缘端部署的需求,但其检测精度受到较大影响;何赟泽等^[9]通过引入相对位置编码和高斯分布权重,并采用空间调制共同注意力机制的方法改进 DETR,使网络解码器更加高效地识别行为特征,提高对车间人员行为的自动化监测水平;吴晓宁等^[10]对 RT-DETR 进行改进,采用 Adown 子采样模块对原卷积模块进行改进,提高了检测精度,减少了参数值的个数。在骨干网络上使用 LSKA 注意力模块进一步提高了检测精度;徐仟祥等^[11]通过添加额外的检测层来增加 RT-DETR 网络对小目标检测的特征提取能力,并在多尺度特征融合阶段使用了改进的 BIFPN 结构以提高模型信息交互的能力;沈烨等^[12]在 RT-DETR 主干网络中引入 FADC 模块,引入小目标检测层,采用 DySample 上采样算子,在满足一定实时性的同时,提升了算法对小目标物体的检测精度。王文杰等^[13]采用 FasterNet 部分卷积替换 RT-DETR 主干网络的传统卷积,有效提升了模型的运算速度,并引入 Faster_Rep 融合特征模块,保留更多有效特征并减少计算量,提高了算法检测的实时性,但使其检测精度受到一定的限制。

上述部分学者使用不同的方法来提高算法检测的准确度,但大多都增加了算法的参数量和计算量,这就增加了算法的推理成本,会使检测速度和实时性受到较大影响。还有部分学者对算法进行了轻量化的处理,增强了算法的检测速度和实时性,但使算法的检测准确度受到不同程度的限制。只有沈烨等在满足一定实时性的同时提升了对小目标的检测精度,但其算法的 FPS 较低,当部署到性能受限的小型终端设备或服务器上时,实时性会受到较大影响。

针对以上问题,选择以 RT-DETR 作为基准模型进行改进,将改进后的模型命名为 FDH-DETR。通过对比实验说明,对比于其他主流算法,改进的算法在保持准确度较高的同时,拥有较低的参数量和计算量,拥有较高的 FPS,很大程度上满足了实时检测的要求,即对硬件设备的要求较低,使

其易于部署在性能受限的小型终端设备上。并在真实工况场景下进行推理实验,体现了改进算法的优势和有效性。

1 RT-DETR 算法

RT-DETR 由百度飞桨团队提出,它将两种经典算法 Transformer 和 DETR(detection transformer)结合到了一起。Transformer 是一种用于序列建模的神经网络架构,最初是应用于自然语言处理,近几年来,基于 Vision Transformer 架构的目标检测算法在计算机视觉任务中取得了显著进展,通过利用自注意力机制捕获视觉图像中的各种依赖关系,这使得 Transformer 能够有效处理并提取图像中的特征信息,实现诸如图像分类、目标检测和图像分割等任务。

RT-DETR 由主干网络、高效混合编码器和带辅助预测头的变压器解码器组成,模型结构如图 1 所示。

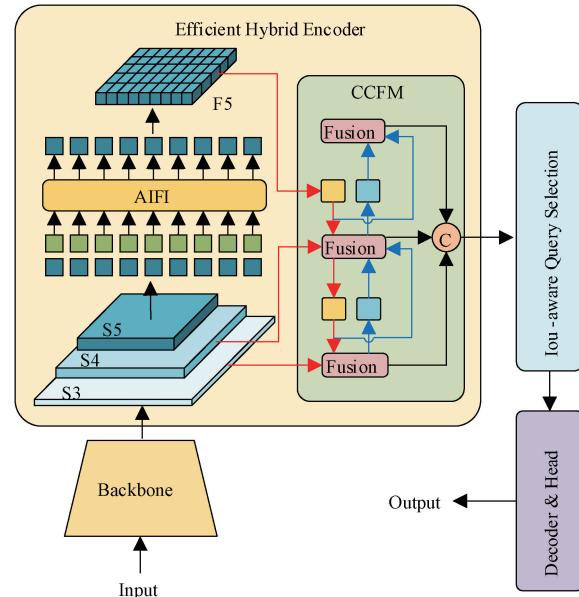


图 1 RT-DETR 结构

Fig. 1 RT-DETR architecture

高效混合编码器由两个模块组成,即基于注意力的尺度内特征交互模块 (attention-based intra-scale feature interaction, AIFI) 和基于卷积神经网络 (convolutional neural network, CNN) 的跨尺度特征融合模块 (CNN-based cross-scale feature-fusion module, CCFM)。AIFI 是一个标准的 Transformer 的 Encoder 层,包含多头自注意力和前馈神经网络 FFN(feed forward networks, FFN)。而 CCFM 其实就是全局特征金字塔 (panoptic feature pyramid network, PAFPN) 结构:首先用若干层 1×1 卷积将所有的特征的通道数都映射至同一数目,随后再进行 Top-down 和 Bottom-up 两部分的特征融合。首先,将二维的 S_5 特征拉成向量,然后交给 AIFI 模块处理,其数学过程就是多头自注意力与 FFN 前馈神经网络。随后,将输出调整回二维,记作

F_5 ,再使用CCFM模块去完成跨尺度的特征融合,最后将融合后的多尺度特征拼接在一起,交给后续的网络去处理,得到最终的输出。

2 RT-DETR 算法的改进

2.1 Deep Faster 模块

在FasterNet^[14]轻量级网络中,部分卷积(partial conv, PConv)可以减少计算冗余和对内存访问的频率,它利用特征映射中的冗余性,对部分输入通道应用普通卷积Conv,其余的通道保持不变,PConv结构如图2右上角所示。在时,PConv的FLOPs为,仅为普通卷积的1/16,内存访问量为,只有普通卷积的1/4。PConv比Conv具有更低的FLOPS和内存访问量,更好地利用了设备的计算能力。

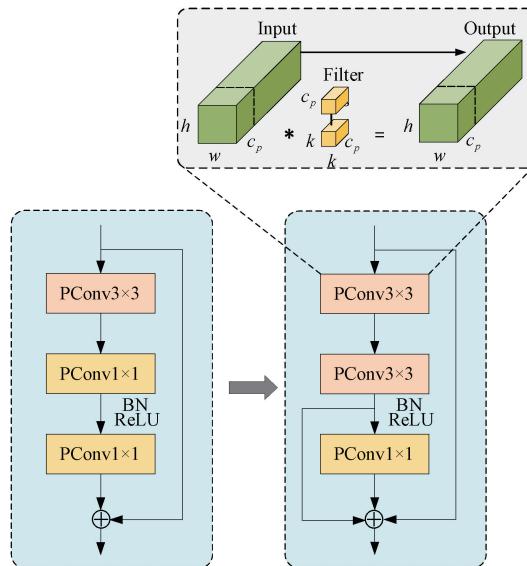


图 2 Deep Faster 模块

Fig. 2 Deep Faster module

对FasterNet进行了深入的分析,FasterNet的成功得益于PConv的成功设计,但是在FasterNet网络中的主要模块FasterNet Block仅使用了一个PConv,随后使用两个1×1卷积来减少参数量并加速计算,这样不能很好地捕捉空间维度上的特征关系,限制模型学习更复杂的特征^[15],并且1×1卷积只侧重于单个像素周围的计算,无法充分考虑到更大范围内的局部特征关系。

因此,对FasterNet Block进行了改进,并将其定义为Deep Faster Block,结构如图2右侧所示。首先,将FasterNet Block中的第2个1×1卷积换成了3×3的部分卷积PConv,使模型能够捕获更大范围的特征信息,提高网络对图像结构的理解能力。并且可以与前一个3×3卷积联动,减少模型的参数数量,提高了模型的泛化能力和训练效率。此外,还在第2个卷积后面加入了残差连接,使用残差连接可以跨层传递特征信息,保留原始特征,增加网络的深度,进一步提高了模型的准确度和泛化能力。

将改进后的FasterNet引入到RT-DETR主干网络中,在不牺牲较多性能的基础上,使模型的计算量和参数量显著减少,加快了推理速度。

2.2 DRBC3 模块

忽略输入的像素相当于在卷积核中插入额外的零项,因此具有较小卷积核的卷积扩展层可以等效地转换为稀疏的具有较大卷积核的非扩展层(即膨胀率 $r=1$)。设 k 为扩展层的卷积核大小,通过插入零项,对应的非扩展层的卷积核大小为 $(k-1)r+1$ 。此外,由前一个卷积核 $W \in R^{k \times k}$ 到后一个卷积核 $W' \in R^{((K-1)r+1) \times (k-1)r+1}$ 的这种变换通过步长为 r 的转置卷积和单位核 $I \in R^{1 \times 1}$ 来实现,转换公式为:

$$W' = \text{conv_transpose2d}(W, I, \text{stride} = r) \quad (1)$$

对任意 $W \in R^{k \times k}$ 和输入通道,对 W 的膨胀卷积(膨胀率 r)总是与对的非膨胀卷积产生相同的结果。

膨胀重参数化模块(dilated reparam block,DRB)基于这种等价变换^[16],它使用一个非扩展的小核和多个扩展的小核层来增强一个非扩展的大核转换层,结构如图3所示。其超参数包括大卷积核的大小 K 、并行转换层大小 k 和膨胀率 r 。图3为4个并行层的情况,用 $K=9, r=(1, 2, 3, 4), k=(5, 5, 3, 3)$ 表示。DRB中除了大核卷积以外,还使用了并行的膨胀卷积(dilated convolution, DConv),利用结构重参数化的思想,整个模块可以等价转换为一个大核卷积。为了将DRB转换成用于推理的大核转换层,首先将每个批归一化层(batch normalization, BN)合并到前面的转换层中,将每个 $r>1$ 的层根据式(1)进行转换,并将所有得到的核加起来,使用适当的零填充。例如将图3中 $k=3, r=3$ 的层转换为稀疏的 7×7 的核,并将其添加到 9×9 的核中,每边都有一个零像素填充。

使用DRB模块对RT-DETR中的RepC3模块进行改进,结构如图4所示,对大核卷积层进行重参数化,以提高性能而不需要额外增加推理成本。DRB模块使用了膨胀卷积,这种卷积操作可以扩大感受野,在不增加参数量的情况下增加网络的感知范围,从而更好地捕获输入数据的全局信息,并且在一定程度上减轻了过拟合的风险。DRB模块可以通过调整扩张率来处理不同尺度的特征,从而更好地适应多尺度的输入数据。

2.3 HiLo-AIFI 模块

HiLo(high-low frequency)注意力机制^[17]在注意力层对特征映射中的高频和低频分别进行处理,HiLo将多头自注意力层中相同数量的Head分成两组,分割比为 α ,其中 $(1-\alpha)N_h$ 个Head用于Hi-Fi,另外 αN_h 个Head用于Lo-Fi。低频注意分支(Lo-Fi)是捕获输入图像特征的全局依赖性,它不需要高分辨率的特征映射,但是需要对全局进行关注。高频注意分支(Hi-Fi)是捕获输入图像特征精细的局部依赖关系,需要高分辨率的特征图,可以通过局部注意来完成。HiLo的输出是将Hi-Fi和Lo-Fi两个输出结果拼接起来,公式为:

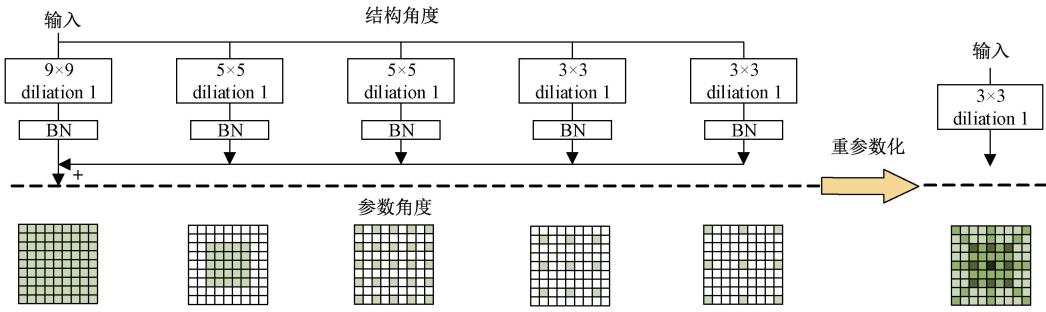


图 3 DRB 模块

Fig. 3 DRB module

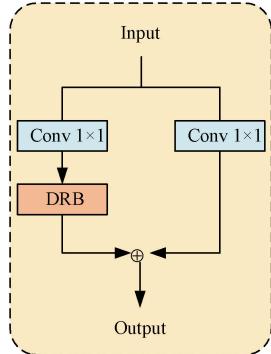


图 4 DRBC3 模块

Fig. 4 DRBC3 module

$$HiLo(X) = [Hi - Fi(X); Lo - Fi(X)] \quad (2)$$

在一般情况下,假设 Hi-Fi 和 Lo-Fi 具有相同数量的头(即 $\alpha = 0.5$),并且特征映射具有相同的宽度和高度。

Hi-Fi 和 Lo-Fi 的计算成本分别为 $\frac{4}{7}ND^2$ 和

$\left(\frac{3}{4} + \frac{1}{S^2}\right)ND^2 + \frac{1}{S^2}N^2D$ 其中, N 表示 token 的数量, D 表示隐藏的维度, s 表示窗口的大小。而多头自注意力 (multi-head self-attention, MSA) 的计算成本为 $4ND^2 +$

$2N^2D$, 远高于 HiLo 注意力的计算成本, 在不同分辨率下 HiLo 和 MSA 的 FLOPs 对比如图 5 所示。

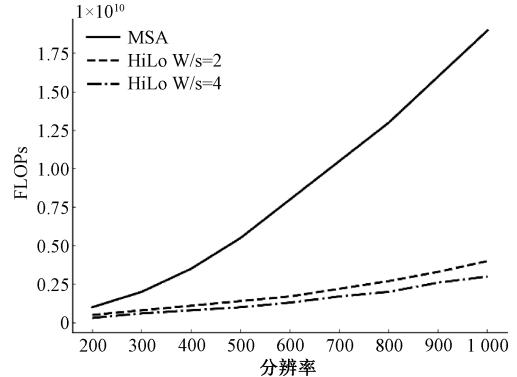


图 5 HiLo 和 MSA 的 FLOPs 对比

Fig. 5 Comparison of FLOPs between HiLo and MSA

如图 5 所示,在较小的输入图像分辨率和较小的 s 值(例如 $s=2$)下,Hi-Fi 和 Lo-Fi 都具有相当的效率。然而,在更高的分辨率下,低保真将导致巨大的计算成本。在这种情况下,稍微增加 s (例如 $s=4$)有助于 Lo-Fi 在保持精度的同时获得更好的效率。更大的窗口尺寸也有助于 HiLo 的整体框架在高分辨率图像上减少更多的 FLOPs。

HiLo 注意力机制如图 6 所示。对于高频注意力,高频

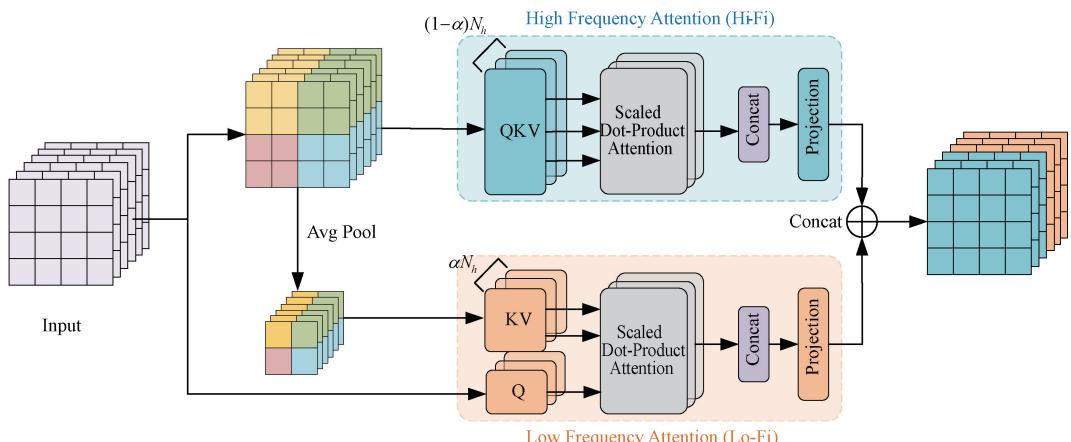


图 6 HiLo 注意力机制

Fig. 6 HiLo attention mechanism

编码对象的局部细节在特征映射上应用全局关注可能是冗余的,并且计算成本很高。HiLo 设计了 Hi-Fi 来捕捉具有局部窗口自注意力的细粒度高频特征,从而减少了计算复杂度。对于低频注意力,Lo-Fi 首先对每个窗口进行平均池化,得到输入图像特征中的低频特征,然后将平均池化后的特征映射投影到键 K 和值 V 中,以模拟输入特征映射中的每个查询位置与每个窗口中平均池化的低频键和值之间的关系。Lo-Fi 中的 Q 仍然来自原始特征图,然后应用标准注意捕获特征图中的低频信息。最后,将高低频特征连接起来,并将结果传递给后续层。

将 RT-DETR 原来 AIFI 模块中的多头自注意力模块换成 HiLo 注意力模块,改进后的结构如图 7 所示,充分利用图像的高频和低频特征,降低了计算的复杂度并且提高了检测性能。

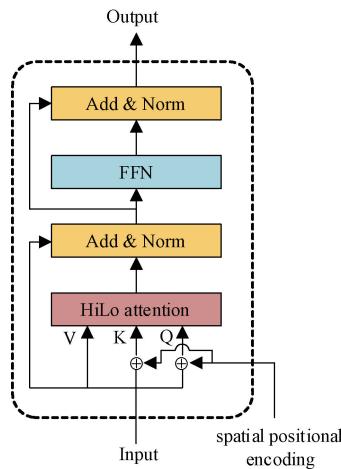


图 7 HiLo-AIFI 模块

Fig. 7 HiLo-AIFI module

3 数据集与实验准备

3.1 数据集

本研究实验使用组合数据集,共计 11 795 张图片,其中 1/3 取自于公共数据集 UR Fall Detection Dataset,还有 1/3 采集于山西众鑫乐食汇食品有限公司的工厂车间,另 1/3 通过网络爬虫获取。图片尺寸为 640×480 ,使用 Labelimg 工具对图片进行标注,数据集划分为 8 个类别,分别为 bending、stand、fall down、fire、smoke、mask、no mask、false mask,其中 bending、stand、fall down 为第 I 大类,用来检测工人的正常和异常行为,如判断是否发生跌倒;fire、smoke 为第 II 大类,用来检测工厂是否出现火灾和烟雾;mask、no mask、false mask 为第 III 大类,用来检测工人是否正确佩戴口罩。

数据集类别及数据分布如图 8 所示。将训练集、验证集和测试集按 7:2:1 的比例进行划分,即训练集包含 8 256 张图片,验证集包含 2 359 张图片,测试集包含 1 180 张图片。

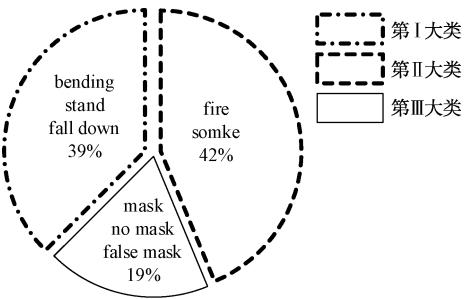


图 8 数据集类别分布

Fig. 8 Distribution of dataset categories

3.2 实验环境与训练参数

为了避免不同硬件环境对算法性能评估的影响,所有实验均在相同的计算机环境下进行。具体环境配置如表 1 所示。

表 1 实验环境配置

Table 1 Experimental environment configuration

项目	配置
操作系统	Windows 11 专业版
CPU	Intel(R) Xeon(R) Silver 4214R CPU@2.40 GHz
GPU	NVIDIA RTX A4000 16 G
python	3.9.18
CUDA	11.7
CUDNN	8.9.5
Pytorch	1.13.1

在训练过程中,调整输入图片尺寸为 640×640 ,batchsize 设置为 8,线程数设置为 8,使用 AdamW 优化器,初始学习率为 0.000 1,权重衰减系数为 0.000 1 共训练 200 个 epoch。另外,为了使每次实验结果具有可复现性,固定随机种子 seed 为 2。

3.3 模型评估指标

实验使用的评估指标有:准确率(precision, P)、召回率(recall, R)、平均准确率(mean average precision, mAP)、模型参数量(Params)和浮点计算量(GFLOPs),还有一个用于评估实时目标检测器性能的指标:每秒传输帧数(FPS)。上述评价指标部分公式表达式为:

$$precision = \frac{TP}{TP + FP} \quad (3)$$

$$recall = \frac{TP}{TP + FN} \quad (4)$$

其中,TP 表示被正确识别的正样本数,FP 表示错误识别的负样本数,TN 表示正确识别的负样本数,FN 表示错误识别的正样本数。

n 个类别平均准确率计算公式为:

$$mAP = \frac{1}{n} \frac{1}{n} \sum_{i=1}^n \int_0^1 precision(reall) d(reall) \quad (5)$$

FPS 表示模型每秒能处理的图片数量,用来对模型的实时性进行评估,公式为:

$$FPS = \frac{\text{sum of pictures}}{\text{sum of time}} \quad (6)$$

模型参数量(Params)和浮点计算量(GFLOPs)这 3 个指标用来对模型的轻量化进行评估,越小越好,以便部署在性能受限小型设备上。

4 实验与结果

RT-DETR 共有 7 个版本,分别为 R18、R34、R50m、R50、R101、HGNetv2-L、HGNetv2-L, 其中 R 表示

ResNet, HGNetv2 为百度飞桨团队自研的主干网络,其是在 HGNet^[18]的基础上进行改进。为满足本文实时检测和轻量化部署的要求,选择参数量和计算量最少、FPS 最高的 RT-DETR-R18 为基准模型进行改进实验。

4.1 与 RT-DETR 及其改进算法比较

为了证明模型改进后检测的提升效果,将其与基准模型 RT-DETR 和基于 RT-DETR 改进的 FSSD-DETR、RT-DETR-Faster 算法进行对比实验,实验结果如表 2 所示。

表 2 与 RT-DETR 及其改进算法对比实验结果

Table 2 Comparison of experimental results with RT-DETR and its improved algorithms

算法	P/%	R/%	mAP50/%	Params/10 ⁶	GFLOPs	FPS
RT-DETR	94.1	88.9	92.0	19.9	57.0	117
FSSD-DETR	95.4	89.0	92.7	21.3	63.2	49.9
RT-DETR-Faster	88.6	85.2	85.3	10.2	15.0	162
本文	96.2	89.1	93.8	13.6	22.0	150

改进后的模型对比于 RT-DETR-R18 基准模型,平均准确度提高了 1.8%,参数量减少了 31.6%,计算量减少了 61.4%,FPS 提升了 33,达到了 150 fps;对比于 FSSD-DETR 算法,本研究改进算法的平均准确度提高了 1.1%,参数量约为其 1/2,计算量约为其 1/3,FPS 比其约高 90 fps;对比于 RT-DETR-Faster 算法,本研究改进算法的参数量、计算量约比其多 20%,FPS 比其低 12 fps,但检测准确度比起高 7.5%。

由实验结果分析可知,与本研究改进算法相比较,FSSD-DETR 算法的检测准确度与其大致达到同一水平,但 FPS 仅为其 1/3,这将使其检测实时性受到较其低

8.5%,这将使其检测的精度受到较大负面影响,会造成更多的漏检和误检。只有本研究改进算法在检测准确度和检测实时性上得到了较好的权衡。

为了更直观的描述所改进模型的提升效果,对 RT-DETR 基准模型和改进后的模型进行推理实验,图 9 展示了推理实验对比结果,在第 1 张图、第 2 张图、第 3 张图和第 5 张图中,改进算法把原算法漏检的跌倒人员、未戴口罩人员和着火点检测了出来。在第 4 张图中,原算法将空中的云误检成了烟雾,而改进后的算法则进行了很好的区分。由此可见,改进算法比原算法在实际应用时可以达到更高的检测准确度,取得更好的检测效果。



(a) RT-DETR 原始算法
(a) RT-DETR original algorithm



(b) 改进后的
(b) Improved

图 9 推理实验结果对比

Fig. 9 Comparison of inference experiment results

4.2 消融实验

为验证本研究提出的每种改进策略的有效性,以 RT-

DETR 为基准模型,对每种改进策略进行验证。消融实验结果如表 3 所示。

表 3 消融实验结果

Table 3 Ablation study results

实验组别	FasterNet	Deep Faster	DRB	HiLo-AIFI	P/%	R/%	mAP50/%	Params/10 ⁶	GFLOPs	FPS
1	—	—	—	—	94.1	88.9	92.0	19.9	57.0	117
2	√	—	—	—	93.6	87.8	91.2	14.3	37.3	131
3	√	√	—	—	94.2	89.0	92.1	15.4	30.6	136
4	√	√	√	—	94.7	89.2	92.5	13.7	21.9	148
5	√	√	√	√	96.2	89.1	93.8	13.6	22.0	150

表 3 中的实验结果表明,对比于 RT-DETR 基准模型,所提出的每种改进策略都有不同程度的检测性能的提升。轻量化网络 FasterNet 的引入,使模型的参数量减少了 28.1%,计算量减少了 34.5%,但检测准确度略有下降。Deep Faster 模块的添加使模型仅增加了 1.1 M 的参数量,却减少了 6.7 GFLOPs 的计算量,并且使模型的检测性能达到和基准模型相同的水平。DRBC3 模块的添加将模型的检测准确度提高 0.5%,平均准确度提高 0.4%,并且使模型的参数量和计算量进一步减少。最后,HiLo-AIFI 模块的应用将模型的检测准确度和平均准确度分别提高 1.5% 和 1.3%,模型参数量和计算量基本保持不变。

综上所述,本文提出每种改进策略都使模型的检测性能有不同程度的提升,利用 Deep Faster 特征深度融合模块对 FasterNet 进行改进,改进后的 FasterNet 减少了

模型的参数量和计算量,提高了模型的检测速度。DRBC3 模块通过大小卷积核转换机制,减少了模型的推理成本。HiLo-AIFI 高低频尺度内特征交互模块,增强网络对高低频特征的提取能力,进一步提高了模型检测的准确度。

4.3 对比实验

为了证明本研究提出的改进算法的优越性和有效性,将其与目前主流的目标检测算法进行对比实验。目前的主流算法分为一阶段算法和两阶段算法,一阶段主流算法如 YOLO 系列和两阶段主流算法如 RCNN (region convolutional neural network) 系列。考虑到两阶段算法的检测实时性较差,不适合本研究实时检测的背景,故选择与实时性较好的 YOLO 系列目标检测算法进行对比实验,选取 YOLOv5s、YOLOv5m、YOLOv7-tiny、YOLOv7 和 YOLOv8s 进行对比实验,对比实验结果如表 4 所示。

表 4 不同算法对比实验结果

Table 4 Comparison of experimental results with different algorithms

算法	P/%	R/%	mAP50/%	Params/10 ⁶	GFLOPs	FPS
YOLOv5s	91.8	85.0	90.3	7.0	15.8	212
YOLOv5m	92.1	85.3	90.5	20.8	47.9	134
YOLOv7tiny	87.8	88.9	91.2	6.0	13.1	128
YOLOv7	93.7	88.5	92.5	36.5	103.3	62
YOLOv8s	92.1	86.6	92.2	11.2	28.6	190
RT-DETR	94.1	88.9	92.0	20.0	57.0	117
本文	96.2	88.1	93.8	13.6	22.0	150

从表 4 的实验结果可以看出,改进后的算法在准确度和平均准确度都取得了最好的结果,分别为 96.2% 和 93.8%。虽然 YOLOv7tiny 的参数量和计算量最低,但是对比于改进模型,准确度降低了 8.4%,平均准确度降低了 2.6%,FPS 降低了 32 fps。YOLOv8s 虽取得了最佳的 FPS,达到了 190 fps,但是其准确度比改进模型低 4.1%,召回率低 1.5%,平均准确度低 1.6%,在实际应用时可能会出现漏检、误检的情况。

综上,对比于其他主流算法,改进的算法在保持较高准确度的同时,拥有较低的参数量和计算量,拥有较高的 FPS,检测精度和实时性得到了较好的权衡。

4.4 真实场景推理实验

为验证算法在实际应用时检测的有效性和实时性,将算法部署在工厂本地的服务器上,使用多个车间的多路视频监控作为数据源,将多路数据源同步进行实时的检测,并将检测结果实时输出到监控室的大屏幕中,检测效

果如图 10 所示。算法将每个车间的员工的状态行为都检

测了出来,取得了良好的检测效果,达到了预期的要求。



图 10 工厂实时检测效果

Fig. 10 Real-time detection results in factory

为验证算法在真实工况场景下的检测效果,取山西众鑫乐食汇食品有限公司的工厂车间现场照片进行推理实验,结果如图 11 所示,算法将工人的大部分行为都检测了出来,特别是对于没有正确佩戴口罩的异常行为,也达到了预期的检测效果。

RT-DETR 改进的工人行为及火灾检测算法 FDH-DETR。Deep Faster 特征深度融合模块,减少了模型的参数量和计算量,提高了检测速度。利用 DRB 模块对 RepC3 模块进行重新设计设计,通过大小卷积核转换机制,减少了模型的推理成本。HiLo-AIFI 高低频尺度内特征交互模块,增强网络对高低频特征的提取能力,提高了检测的准确度。改进后的算法准确度达到了 96.2%,对比原始算法,参数量减少了 31.6%,计算量减少了 61.4%,FPS 达到了 150 fps。

改进的算法在实际应用时取得的良好的检测效果,并且算法的轻量化使之能部署在性能受限的小型终端设备上,达到了预期的要求。下一步的任务是解决算法对小目标检测能力不足的问题。

参考文献

- [1] 李金成,闫睿懿,代雪晶. 基于改进注意力机制与 VGG-BiLSTM 的暴力行为检测[J]. 现代电子技术, 2024, 47(21): 131-138.
LI J CH, YAN R AO, DAI X J. Violence behavior detection based on improved attention mechanism and VGG-BiLSTM[J]. Modern Electronic Technology, 2024, 47(21): 131-138.
- [2] 叶彦斐,胡龙葵,张成龙. 基于改进 YOLOv8n-Pose 的轨道作业人员跨轨安全动作识别[J]. 国外电子测量技术, 2024, 43(8): 181-188.
YE Y F, HU L G, ZH CH L. Cross-track safety action recognition of track workers based on improved YOLOv8n-pose[J]. Foreign Electronic Measurement Technology, 2024, 43(8): 181-188.
- [3] 张化龙. 人工智能在安全监控领域的应用与发展[J]. 中国信息界, 2024(1): 235-238.

图 11 现场推理实验结果

Fig. 11 On-site inference experiment results

5 结论

针对工厂时刻面临的安全生产问题,提出了一个基于

- ZHANG H L. Application and development of artificial intelligence in the field of security monitoring [J]. China Information Society, 2024(1):235-238.
- [4] ZHAO Y, LYU W Y, XU SH L, et al. Detrs beat yolos on real-time object detection [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024:16965-16974.
- [5] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers [C]. European Conference on Computer Vision, Cham: Springer International Publishing, 2020: 213-229.
- [6] 李海川,贺星亮,贾仟国,等.基于S3DD-YOLOv8n的矿工行为检测算法[J].矿业安全与环保,2024,51(5):96-104.
- LI H CH, HE X L, JIA Q G, et al. A behavior detection algorithm for mining workers based on S3DD-YOLOv8n[J]. Mining Safety and Environmental Protection, 2024, 51(5): 96-104.
- [7] 刘恒,林虹宇,吴涛.卷烟厂卷包车间工人违规作业行为检测方法[J].计算机科学,2024,51(S1):553-560.
- LIU H, LIN H Y, WU T. Method for detecting illegal operations by workers in the cigarette packing workshop of a tobacco factory[J]. Computer Science, 2024,51(S1):553-560.
- [8] 罗国富,王源,李浩,等.基于改进YOLOv5s的智能车间工人不安全行为实时检测方法[J].计算机集成制造系统,2024,30(5):1610-1619.
- LUO G F, WANG Y, LI H, et al. Real-time detection method of workers' unsafe behavior in intelligent workshop based on improved YOLOv5s[J]. Computer Integrated Manufacturing Systems, 2019,30(5):1610-1619.
- [9] 何贊泽,谯灵俊,王洪金,等.基于改进DETR的智慧车间人员典型行为识别算法[J].电子测量与仪器学报,2024,38(9):76-84.
- HE Y Z, QIAO L J, WANG H J, et al. Typical behavior recognition algorithm of intelligent workshop personnel based on improved DETR [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(9): 76-84.
- [10] 吴晓宁,司占军.基于改进RT-DETR的火灾探测模型(英文)[J].印刷与数字媒体技术研究,2024(4):107-114.
- WU X N, SI ZH J. Fire detection model based on improved RT-DETR (English) [J]. Research on Printing and Digital Media Technology, 2024(4): 107-114.
- [11] 徐仟祥,曾勇,卢倩,等.基于改进RT-DETR的车门内拉手表面缺陷检测方法[J].电子测量技术,2024,47(18): 172-181.
- XU Q X, ZENG Y, LU Q, et al. A method for surface defect detection of car door handle based on improved RT-DETR [J]. Electronic Measurement Technology, 2024, 47(18): 172-181.
- [12] 沈烨,王泉,吴昊. FSSD-DETR自动驾驶场景实时目标检测算法[J].电子测量技术,2024,47(14):88-95.
- SHEN Y, WANG Q, WU H. Real-time target detection algorithm for FSSD-DETR automatic driving scenario [J]. Electronic Measurement Technology, 2024, 47(14): 88-95.
- [13] 王文杰,陈伟,路锦通,等.基于RT-DETR-Faster的苹果采摘机器人实时目标检测算法[J].自动化与仪表,2024,39(7):57-62.
- WANG W J, CHEN W, LU J T, et al. Real-time target detection algorithm for apple picking robot based on RT-DETR-Faster [J]. Automation and Instrumentation, 2024, 39(7): 57-62.
- [14] CHEN J R, KAO S, HE H, et al. Run, don't walk: Chasing higher FLOPS for faster neural net-works [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 12021-12031.
- [15] WANG G, CHEN Y, AN P, et al. UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios [J]. Sensors, 2023, 23(16): 7190.
- [16] DING X H, ZHANG Y Y, GE Y X, et al. UniRepLKNet: A universal perception large-kernel ConvNet for audio video point cloud time-series and image recognition [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024:5513-5524.
- [17] PAN Z ZH, CAI J F, ZHUANG B H. Fast vision transformers with hilo attention [J]. Advances in Neural Information Processing Systems, 2022, 35: 14541-14554.
- [18] CHEN J T, LEI B W, SONG Q Y, et al. A hierarchical graph network for 3d object detection on point clouds[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 392-401.

作者简介

- 董润华,硕士研究生,主要研究方向为深度学习、目标检测。
E-mail: 1303364890@qq.com
- 常青(通信作者),博士,副教授,主要研究方向为嵌入式系统、信息系统设计。
E-mail: 13103251868@163.com
- 孔鹏伟,工程师,主要研究方向为信息安全、大数据。
E-mail: 1433305379@qq.com
- 王耀力,博士,副教授,主要研究方向为机器视觉、计算智能与最优化建模。
E-mail: gyx6686360@163.com