

基于动静结合互学习的预制梁工序检测方法

冯晓飞¹, 谢诚², 张秀振¹, 董仕奎¹, 陈军胜¹, 叶舒³, 钟欣^{3*}

(1. 山东东方路桥建设有限公司技术创新中心, 山东 济南 250101;

2. 山东省交通规划设计院集团有限公司科技研发中心, 山东 济南 250101;

3. 武汉理工大学计算机与人工智能学院, 湖北 武汉 430070)

摘要: 针对预制梁场位置偏远、场景复杂、光线不足, 导致数据采集困难、背景干扰、画质受损等问题, 提出一种基于动静结合互学习的预制梁工序检测方法。在单阶段目标检测模型上建立互学习框架, 分别使用数据扩增技术在空间和时间上对样本干扰的能力, 构造动静结合的双分支子网络, 在网络中引入基于归一化的注意力通道子模块动态地调整通道权重, 以适应真实场景下的环境光照复杂性和噪声干扰随机性。为充分发挥两支子网络各自的优势, 利用目标检测模型真实值的预测边界框不唯一的特性, 提出正样本对齐策略, 实现边界框数量及表征分布的双重对齐。构建一个基于真实场景的预制梁工序数据集, 在自制数据集上的实验结果表明, 该方法的精确率和均值平均精度分别达到了 97.2% 和 97.7%, 推理速度达到了 78 帧/s, 在满足工业落地应用需求的同时, 为预制梁工序检测识别问题提供了一种有效且可靠的解决方案。

关键词: 动静结合; 候选框互学习; 正样本对齐; 工序检测识别; 目标检测

中图分类号: TP391.41

文献标志码: A

DOI: 10.19678/j.issn.1000-3428.0069033

Detection Method of Precast Beam Process Based on Dynamic-Static Fusion Mutual Learning

FENG Xiaofei¹, XIE Cheng², ZHANG Xiuzhen¹, DONG Shikui¹, CHEN Junsheng¹, YE Shu³, ZHONG Xian^{3*}

(1. Technical Innovation Center, Shandong Dongfang Highway and Bridge Construction Co., Ltd., Jinan 250101, Shandong, China;

2. Technology Research and Development Center, Shandong Provincial Communications Planning and Design Institute Group Co., Ltd., Jinan 250101, Shandong, China;

3. School of Computer Science and Artificial Intelligence, Wuhan University of Technology, Wuhan 430070, Hubei, China)

【Abstract】 The remote location of precast beam yards, complex scenes, difficulties in data collection due to poor lighting, background interference, and degraded image quality, all create challenges for precast beam processing. This study proposes a dynamic-static mutual learning detection method for precast beam processing. The study establishes a mutual learning framework on a single-stage object-detection model. It uses data augmentation techniques to enhance the ability of a model to manage sample spatial and temporal interference, constructing a dual-branch subnetwork that combines dynamic and static features. Simultaneously, a normalization-based attention channel submodule is introduced into the network to dynamically adjust the channel weights. Through these techniques, the model becomes more adaptable to the complexity of environmental lighting and the randomness of noise interference in real scenes. To fully leverage the respective advantages of the two subnetworks, the study also proposes a positive sample alignment strategy, leveraging the inherent nonunique characteristics of a single real value's predicted bounding box in the object detection model. Consequently, a dual alignment is achieved, addressing both the quantity and distribution of bounding boxes. A precast beam process dataset based on real scenarios is constructed and used to validate the effectiveness of the proposed method. The precision and mean average precision reach 97.2% and 97.7%, respectively, at an inference speed of 78 frame/s, which meets industrial application demands and offers an effective solution for precast beam process detection and recognition.

【Key words】 dynamic-static fusion; candidate box mutual learning; positive sample alignment; process detection and recognition; object detection

0 引言

预制梁工序检测识别是指利用计算机视觉和机

器学习多媒体技术, 通过摄像头画面自动化地检测预制梁台座并识别和监测预制梁生产过程中的各个工序, 帮助管理人员准确地监测和记录工序的执行

收稿日期: 2023-12-15 修回日期: 2024-02-18

基金项目: 国家自然科学基金(62271361); 山东省交通运输科技计划项目(2022B45)。

通信作者 E-mail: *zhongx@whut.edu.cn

情况,从而确保工序的规范性和质量控制。目前,大多数预制梁场采用人工巡查的方式对梁体的所处工序进行记录并检查预制梁的质量,这种方法效率低且容易出错。习近平总书记在党的二十大报告中强调:“坚持把发展经济的着力点放在实体经济上,推进新型工业化”,以及“促进数字经济和实体经济深度融合”。基于真实场景的工序检测识别算法不仅能够提高预制梁生产过程的自动化程度,还能够提高工序识别的效率和准确性,从而增强梁场质量控制和工程管理能力。因此,对复杂环境鲁棒性强、准确率高的预制梁工序检测识别算法的研究具有重要意义。

在实际工业环境中,一个摄像头与其视野范围内的台座数量可能为一对一、一对多的关系,工序之间 T 梁的形态差别有大有小,因此预制梁工序识别任务本质上是一个基于目标检测的分类任务,且分类难度较高。近年来,基于深度学习的目标检测算法在自动驾驶、医疗、工业检测等领域取得了显著进展,并得到了广泛的应用。现阶段主流的目标检测算法分为双阶段和单阶段两大类。双阶段检测算法将检测过程划分为候选框生成和预测框确定两个阶段,然而由于处理大量的候选框速度较慢。针对这一问题,以 YOLO^[1]为代表的单阶段目标检测算法应运而生,直接预测目标对象的类别概率和位置。相较于双阶段算法,单阶段算法极大地减少了运算成本,加快了检测速度,为实时应用提供了可行的解决方案。

尽管预制梁在建筑领域中得到广泛应用,然而预制梁工序检测识别在视觉领域尚处于研究的初始阶段,相关公开研究工作较少,技术还不够成熟,有待进一步的探究。在对预制梁工序检测识别任务的探索中,发现现有研究在对台座的识别过程中存在一些不足之处:未充分考虑光照变化会引起图像亮度、对比度和色彩的变化,同时忽略了物体遮挡和遮光等因素可能对目标检测和识别造成的影响。因此,针对复杂场景下的预制梁台座检测与工序识别问题,有待有针对性地解决。此外,目前没有公开可直接使用的预制梁或梁场数据集。梁场所处位置偏远且工序多样,不同角度和视角的覆盖及工人操作的实时性等因素进一步加大了数据采集困难的挑战。由于预制梁工序的多样性和复杂性,确保获取具有代表性的数据集并进行准确标注也是一个复杂的任务。而基于真实场景的数据集通常包含多种环境条件、光照变化等多样性因素,造成的噪声和干扰会将模型的注意力从关键特征转移到不相关的信息

上,导致模型性能下降。

针对上述真实场景样本的环境光照复杂性和噪声干扰随机性问题,本文提出一种基于动静结合互学习的预制梁工序检测方法。受 ZHANG 等^[2]在分类任务上深度互学习(DML)成功的启发,本文在目标检测模型上构建动静结合的双分支互学习框架:静态学习网络分支在空间维度上模拟真实场景的环境光照复杂性;动态学习网络分支在时间维度上模拟真实场景中噪声干扰的随机性;两个网络分支互相交换学习经验,从而提升预制梁工序识别模型对不同环境的适应性、泛化能力及工序识别的准确度和精度。

本文的贡献主要有以下 4 点:

1)新数据集。构建一个面向预制梁工序检测识别任务的全新数据集,克服了复杂工况和天气环境挑战,实现了高质量的数据收集和标注,为算法研究提供了标准化和可靠的数据基础,推动了相关领域的工业智能化发展。

2)新任务。引入了来源于实际工业需求的预制梁工序检测识别任务,涵盖了在全天候复杂场景中多样的随机性噪声等实际挑战,为理解和解决在建筑工程实践中的关键问题提供了独特的视角。

3)新思路。为工业场景任务提出了新的工作方法,设计动静结合的检测互学习框架,利用正样本对齐策略实现分布表征双重对齐,进而通过散度约束静态学习网络和动态学习网络,使模型在面对静态和动态情境时都能取得更优越的性能。

4)可落地应用。本文模型具备卓越的实时检测能力和优异的泛化性能,在山东多个梁场工程实践中展现出其有效性,与物联网技术共同实现了工业智能化。

1 相关工作

以 YOLO 系列目标检测网络为基础,针对各类需求进行了一系列改进工作^[3]。在工业领域常被应用于缺陷检测^[4-5]、火灾检测等任务,而在其他领域也常被应用于定位识别任务^[6]。季娟娟等^[7]通过优化多支路上采样结构和引入 ELAN-RepConv 结构,改进了多尺度特征融合模块,从而提升微小缺陷的识别准确率。CAO 等^[8]在 YOLOR 网络颈部的 ELAN 模块中添加了卷积注意力机制(CBAM)以扩大模型的感受野,增强其对火灾图像通道和空间信息的注意力。黄小平等^[9]采用了坐标注意力机制(CANet)增强网络的信息传递和特征提取。郑道

能^[10]将空间金字塔池化(SPP)网络以及挤压和激励(SE)模块引入 YOLOv3-tiny 以提升煤矸石识别模型的速度和准确性。王金平等^[11]采用双流网络结构,将 YOLOv4 与基于掩码区域的卷积神经网络(Mask R-CNN)作为两个分支,分别获取人员位置信息及危险区域的语义信息,从而有效地判断施工人员是否位于危险区域。这些改进方法丰富了 YOLO 系列的应用领域,提高了其检测性能和实用性。

在面对训练数据多样性不足和复杂场景挑战等问题时,一些学者尝试将最初设计用于分类任务的互学习方法迁移到目标检测任务上,并成功地解决了这些挑战。ZHANG 等^[12]为解决正样本过拟合问题,在视频异常检测任务中引入互学习框架,使用边界感知的模拟损失作为两个生成器之间的约束。WU 等^[13]提出了包含 3 个连续层子网络的互学习模块,在一定程度上对杂乱的背景边缘进行了抑制。但以上用于检测的互学习在框架中使用的交互损失函数均由范数组成,用于测量预测值和真实值之间的差异,无须考虑样本数量和表征的一致性。此外,范数损失函数具备对称性,当多个子网络结构相同时,以一致的程度对各个子网络进行约束。传统用于分类模型的互学习框架中的交互损失函数为 Kullback-Leibler(KL)损失,它衡量的是从一个分布到另一个分布的“信息增益”,当子网络结构有所区分时,KL 损失能不同程度地对不同子网络施加约束,灵活利用同伴网络的优势,充分挖掘每个子网络的潜在贡献。因此,对于本文动静结合的互学习框架,KL 散度可以更大程度地发挥网络的互补性。

2 预制梁工序数据集

由于预制梁生产场地通常位于偏远地区,其环境条件的复杂性以及光线条件的限制为数据采集工作带来显著挑战,因此目前还没有公开的预制梁工序数据集,但标准化的数据采集对于模型训练的重要性不可忽视。为了更好地适应预制梁场景的视觉识别任务,本文构建了一个基于真实场景的预制梁工序数据集 Precast Beam Process (WUT-PBP)。

2.1 考虑因素

为确保数据采集和标注的规范性,在图片内容、分辨率等方面应保持一致性。在采集过程中,本文充分考虑真实的应用场景,致力于覆盖更多不同的场景,以提高数据的多样性。考虑的主要因素包括:

1)数据规模。现有制梁工程通常由多个梁场合作完成,每个梁场设置的摄像头数量非常多,一个摄像头包含的台座数量不一。因此,数据集数据样本的数量应尽可能大。

2)摄像头环境。当前国内梁场的摄像头在室内和室外环境中都有应用,光照和天气条件可能对摄像头成像产生显著影响,构建时应涵盖尽量多的复杂环境。

3)样本表示。在本预制梁场监控场景中,预制梁工序的持续时长存在显著的差异,从短暂的 1~2 h 到较长的 3~4 天不等,但在短时间的变化微乎其微,因此本数据集以小时单位为间隔采集数据。

2.2 数据采集与分布

WUT-PBP 数据集采集于山东某预制梁场 30 余台实时工业摄像头,每个摄像头正对一个或两个台座,涵盖钢筋绑扎、模板工程、混凝土浇筑、梁体养生、预应力工程、压浆施工、封锚工程等预制梁主要施工工序。在构建本数据集时,充分考虑不同的光照强度(白天、黑夜)和天气状况(晴天、阴天、雨天等),标注工具为 labelme,经过筛选、分类和梁场专业质检员反复检查,每类的样本数量得到了较好的均衡。预制梁工序的全过程以及各类的样本数量如图 1 所示。

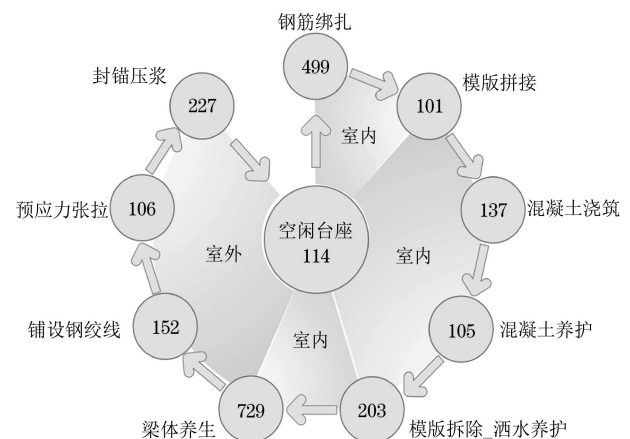


图 1 各类图像数量分布

Fig.1 Distribution of various types of images

数据集在数据规模方面与现有的主流公开工业数据集相当。表 1 展示了本文 WUT-PBP 数据集与当前主流公开工业数据集在多个维度上的比较情况,包括东北大学发布的带钢表面缺陷数据集、Garbage In Images(GINI)^[14]、磁瓦缺陷数据集^[15]、DeepPCB^[16]。本文的 WUT-PBP 数据集在数量上与主流工业数据集相当,并且具有较高的分辨率,同时包含了更多样本退化因素。

表 1 不同数据集统计信息对比

Table 1 Comparison of statistic information of different datasets

数据集	类别/个	数量/个	分辨率/像素	背景	退化
表面缺陷数据集	3	1 800	200×200	×	×
GINI ^[14]	2	2 561	—	✓	×
磁瓦缺陷数据集 ^[15]	6	1 344	—	×	×
DeepPCB ^[16]	6	3 000	640×640	×	×
WUT-PBP(本文)	10	2 373	1 920×1 080	✓	✓

2.3 数据集特点

WUT-PBP 数据集是一个基于真实场景且具有显著挑战性的预制梁工序数据集。该数据集的采集工作全面覆盖了全天不同时间段,保证了时间维度的广泛性和数据的多样性。此外,它还包含了在多种气象条件下的场景数据,如晴朗、多云以及雨雪等复杂气候状态,确保了环境因素的全面考量。这一数据集的复杂性为研究和优化预制梁工序算法提供了一个真实且多变的测试环境。

此外,本数据集涵盖了 10 个不同的类别,这些类别不仅涵盖了细粒度的差异,还展示了显著的粗粒度特征差异。其中,同一阶段的工序具有较小的差异,难以区分;而不同阶段的工序之间存在显著的差异,易于区别。这种多层次的类别划分对于网络的学习能力提出了更高层次的挑战。

3 动静结合的互学习检测方法

3.1 概要

为了提高预制梁检测和工序识别的准确率、提高模型面对复杂多样场景的鲁棒性,本文在目标检测模型上构建互学习框架,引入 KL 散度损失,同时设计正样本对齐策略,以实现两支子网络的预测分布的计算与交互,如图 2 所示。在测试和推理阶段,本文选取表征能力更强的动态学习网络分支作为最终精度高、性能好、适用于实际梁场环境并可以直接部署的预制梁工序检测识别算法模型。

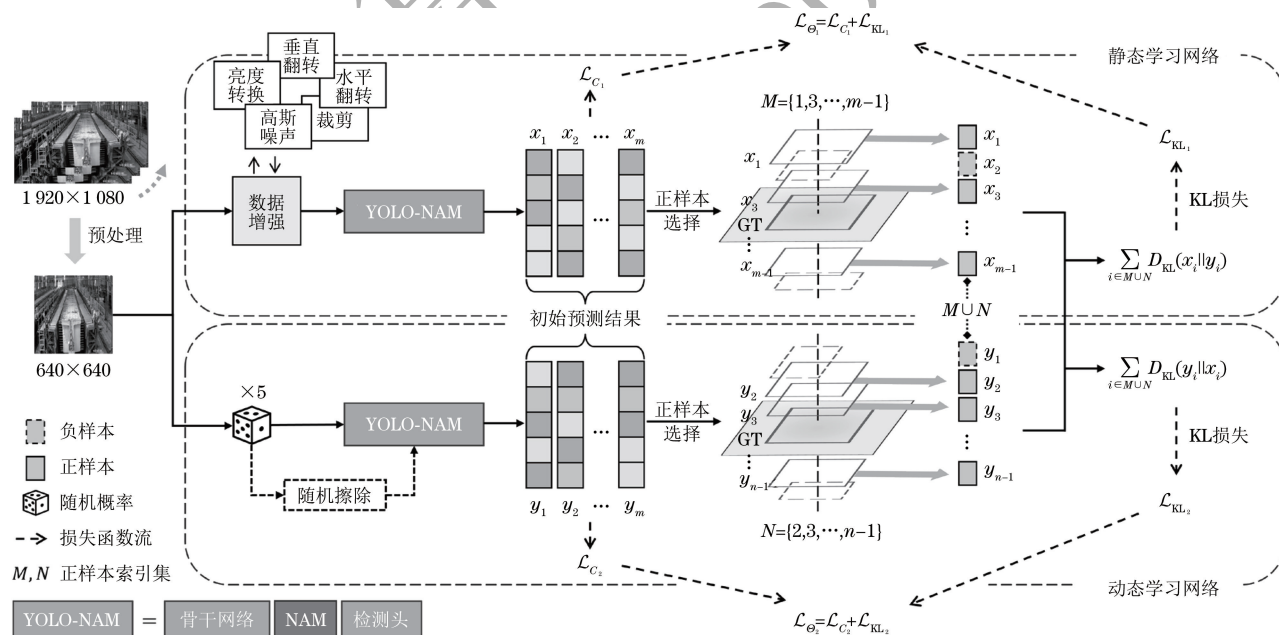


图 2 算法框架图

Fig.2 Algorithm framework diagram

3.2 YOLO-NAM

由于国内预制梁场的建设因地制宜、布局不一,预制梁工序检测识别算法的对象可能涉及一个台座或多个台座,因此目标检测模型成为首选解决方案。在实际工程应用中,通常更倾向于采用经过验证且稳定的解决方案,以确保在不同场景下的可靠性和有效性。本文选择 YOLOv7^[17] 为基线模型。YOLOv7 网络模型通过特征提取网络提取样本的外观特征,检测头来预测检测框及其类别。

为了降低复杂环境造成的噪声干扰对模型性能

的影响,本文在特征提取网络中引入基于归一化的注意力模块(NAM)^[18]。本文仅采用 NAM 的通道注意力子模块,称为 NAM_C。在 NAM_C 中, F 表示原始输入特征图,通过计算特征图在通道维度上的均值和标准差,得到每个通道的重要程度。由于批量归一化(BN)层缩放因子 σ 反映通道变化大小,借由缩放因子调整通道特征的权重,以增强重要通道的表示能力,得到改善后的特征图 M ,如式(1)、式(2)所示:

$$M = \text{Sigmoid}(W \cdot B_{\text{BN}}(F)) \quad (1)$$

$$W = \frac{\sigma}{\sum \sigma} \quad (2)$$

为了不引入过多计算复杂度,有效地增强网络的特征提取能力,本文将 NAM_C 放置在特征提取网络的后部,将经过通道注意力子模块处理后的特

征与输入特征相乘,作为特征提取网络最终输出的特征然后送入检测头进行预测,将这样的检测模型命名为 YOLO-NAM,如图 3 所示。通过这样的注意力机制,模型能够自适应地选择和加权特征图中的重要信息,降低了复杂背景的干扰程度。

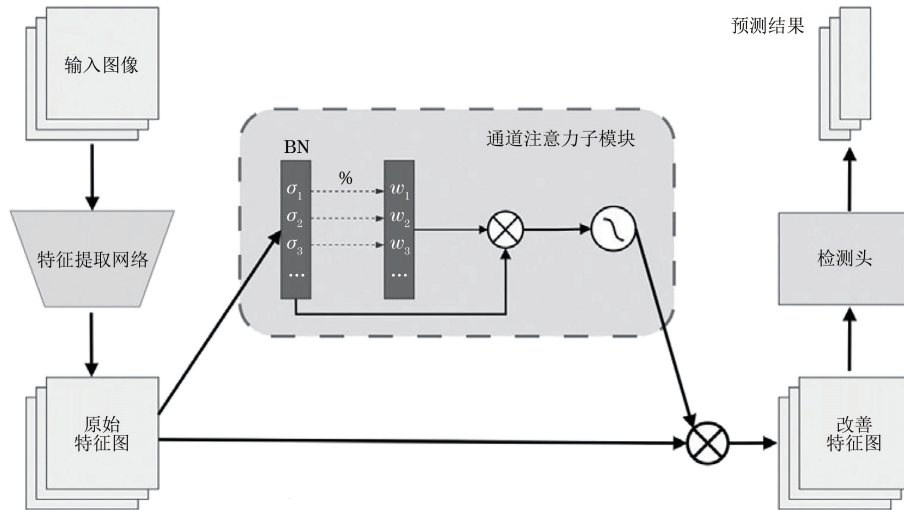


图 3 YOLO-NAM 结构

Fig.3 YOLO-NAM structure

3.3 静态学习网络

本文构造了一个数据增强模块,包括水平翻转、垂直翻转、裁剪、随机亮度转换、添加高斯噪声这 5 种数据增强方法,其中后 2 种数据增强技术尤其模拟了真实环境下由天气、时间等因素引起的不利光线条件造成的干扰,以扩充训练数据的样本空间,有助于模型更好地学习并适应不同的复杂场景,部分增强方法的效果如图 4 所示。

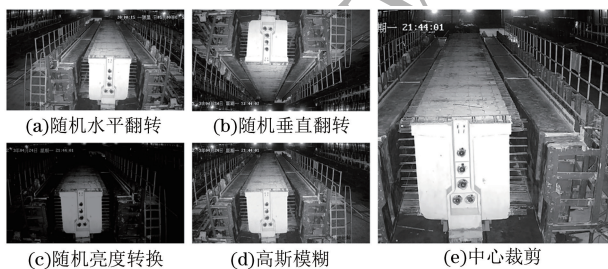


图 4 数据增强效果示例

Fig.4 Examples of data augmentation effects

将数据增强模块应用于静态学习网络来生成丰富多样的训练样本,模拟真实场景的环境光照复杂性,其增强图片作为检测模型的输入。通过多样的数据增强方法,静态学习网络在训练过程中能够更好地适应不同尺度、形状和光照条件下的目标。

3.4 动态学习网络

静态学习网络分支在空间维度上对样本进行永

久增强,每个轮次接收相同的图片,学习静态的、多样的特征;而动态学习网络分支的随机擦除嵌入在网络中,每个轮次重新对原图片进行双重随机擦除(RE)^[19]技术,动态地调整网络学习的方向。模型通过生成随机数来决定是否执行,概率为 50%。若执行,则在图像内随机地选择随机大小和随机长宽比的矩形区域并填充。

通过随机遮挡图像中的部分区域,模拟真实世界中由于遮挡、噪声等因素引起的信息缺失,在训练过程中引入这种随机性,模型能够更鲁棒地学习并适应不同的图像变体,提高其泛化能力,防止模型对于特定的局部信息过于依赖,提高模型对于整体图像的理解和表达能力,从而提升模型在应对真实场景中不同工业环境变化和不确定性的能力。

3.5 动静结合的互学习框架

在本文设计的动静结合的互学习框架中,静态学习网络模拟了常见场景,而动态学习网络则更好地捕捉了真实场景中的变化和不确定性;静态数据增强有助于模型更好地学习基本特征,而动态学习网络则能防止模型过度适应特定场景。为了更好地发挥两个网络分支的优势,本文采用互学习策略优化网络性能,使模型能够更全面地适应各种场景,平衡对常见和变化场景的有效学习。

在训练阶段,每个网络不仅受到真实标签的监

约束,还借鉴了同伴网络的学习经验,以进一步提升网络的泛化能力。互学习的核心思想在于,在两个分支网络中引入交互损失函数,该损失函数与 YOLOv7 中约束检测精度和分类准确度的损失函数并存。本文交互损失函数采用 KL 散度损失,用于度量两个网络分类预测结果分布之间的差异,并通过最小化两者参数之间的差异来实现目标。对于一个预测结果,从静态学习网络 Θ_1 的预测概率分布 p_1 到动态学习网络 Θ_2 的预测概率分布 p_2 的 KL 散度损失 $\mathcal{L}_{KL_1}$ 和 $\mathcal{L}_{KL_2}$ 的计算如式(3)、式(4)所示:

$$\mathcal{L}_{KL_1} = D_{KL}(p_2 \parallel p_1) = \sum_{i=1}^N \sum_{m=1}^M p_1^m(x_i) \log_a \frac{p_1^m(z_i)}{p_2^m(z_i)} \quad (3)$$

$$\mathcal{L}_{KL_2} = D_{KL}(p_1 \parallel p_2) = \sum_{i=1}^N \sum_{m=1}^M p_2^m(x_i) \log_a \frac{p_2^m(z_i)}{p_1^m(z_i)} \quad (4)$$

将 KL 损失加到每个网络原本的损失函数 $\mathcal{L}_c$ 上,网络 Θ_1 和网络 Θ_2 最终的损失函数如式(5)、式(6)所示:

$$\mathcal{L}_{\Theta_1} = \mathcal{L}_{c_1} + \mathcal{L}_{KL_1} \quad (5)$$

$$\mathcal{L}_{\Theta_2} = \mathcal{L}_{c_2} + \mathcal{L}_{KL_2} \quad (6)$$

式中: $\mathcal{L}_c$ 表示原网络模型中约束检测精度和分类准确度的损失函数。

3.6 正样本对齐策略

本文将深度互学习策略应用于单阶段目标检测模型,不同于全局表征的分类模型,目标检测模型丰富的局部表征造成了不同模型输出的预测结果数量的不唯一性,为了使两个网络分支的预测信息能够通过 KL 散度损失函数对预测分布相互约束,更综合地利用各网络分支的优势,提出正样本对齐策略,使两模型输出的预测结果达到边界框数量及表征分布的双重对齐,实现目标检测模型的候选框互学习。

在 YOLOv7 中,针对每个目标对象,筛选出交并比(IoU)最高的多个预测结果,并生成一组正样本索引。为了更有效地利用静态学习网络分支和动态网络分支各自的优势,本文将训练数据中多样性较大的静态学习网络分支与表征能力更强的动态学习网络分支的正样本索引取并集,重新映射两支网络的样本集,进而计算 KL 散度损失,实现双向约束的检测互学习。

对于同一个真实样本,设静态学习网络 Θ_1 的所有预测结果为一个集合 $X = \{x_1, x_2, \dots, x_m\}$,其中被计算为正样本的索引为 $M = \{m_1, m_2, \dots, m_i\}$,

动态学习网络 Θ_2 的所有预测结果为一个集合 $Y = \{y_1, y_2, \dots, y_n\}$,其中被计算为正样本的索引表示为 $N = \{n_1, n_2, \dots, n_j\}$ 。那么最终两个网络分支的正样本索引的计算如式(7)所示:

$$Q = \{q_i | q_i \in M \cup N\} \quad (7)$$

4 实验结果与分析

4.1 数据集

目前,国内外尚未公开具有代表性的预制梁工序数据集。为了验证本文提出的预制梁工序检测识别算法的有效性,本文采用自制的基于真实场景的预制梁工序数据集 WUT-PBP,该数据集包含 10 个类别,按照 8:2 的比例随机地将数据划分为训练集和测试集,其中,训练数据 1 899 张图片,测试数据 474 张图片。

数据集部分类别样本如图 5 所示。其中,第 1 行中的图像呈现清晰状态且光线充足,有助于网络进行有效的特征学习,第 2 行中的图像由于夜间光照不均匀,出现了噪点和眩光现象。此外,由于照明不足,图像中的颜色对比度降低,使得物体边缘不够清晰,进而影响了网络对关键特征的提取。这些不利条件使得数据集具有挑战性,要求模型能够在不同的光照和环境条件下实现鲁棒的工序识别,这对于实际应用中的预制梁生产具有重要意义。

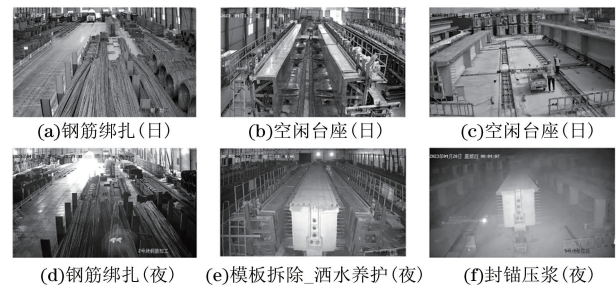


图 5 数据集示例

Fig.5 Dataset examples

4.2 评价指标

本实验采用精确率(Precision)、均值平均精度(mAP)作为模型的评价指标。精确率反映了真正的正样本在所有预测为正样本的样本中的百分比,是模型分类性能的重要度量。高精确率表示模型能够有效地过滤出真正的目标对象,降低误报率,提高分类的可信度。mAP 综合考虑模型在多个类别上的检测准确性,通过计算每个类别的平均精度(AP),然后对所有类别的 AP 进行平均。特别地, $mAP@0.5$ 表示在 IoU 阈值为 0.5 时的平均 mAP,在衡量目标检测模型在物体定位方面的性能时更为具体。

4.3 实验设置

本实验在 Linux 操作系统环境下, GPU 采用 2 张显存为 16 GB 的 Nvidia Tesla V100, 使用 CUDA 11.4 计算架构, 通过 PyTorch 深度学习框架来实现模型的搭建、训练及验证。本实验使用基线模型 YOLOv7 的默认参数设置, 输入的样本图像大小为 640×640 像素, 初始学习率设置为 0.01, 训练批次大小为 32。

随机擦除模块借鉴原论文^[19]的初始参数, 擦除的概率值 p 设置为 0.5, 填充区域的面积在整张图片中的占比设置为 0.02~0.2, 擦除块中的填充数值设置为公共数据集 ImageNet 的样本像素归一化的平均值[0.491 4, 0.482 2, 0.446 5]。

4.4 对比实验

表 2 中展示了本文方法与基线模型在自制数据集 WUT-PBP 的测试集上的性能表现, 包括不同类别的精确率。从结果可以观察到, 基线模型在大多数类别上取得了高于 90.0% 的精确率, 凸显数据集的高质量 and 标注的严格准确性。相较于基线模型, 本文方法在绝大多数类别上取得了更高的精确率, 彰显了其有效性和卓越性。特别值得注意的是, “预应力张拉”类由原来的 85.5% 提高到了 97.7%, 显著提高了该类的预测准确率。对于“钢筋绑扎”类, 精确率相较基线模型下降了 0.3 百分点, 可能的原因是该类工序在室内进行, 且光线来源单一、易出现眩光等图片对比度较大的情况。本文模型专注于提高对不同亮度的整体适应能力, 可能导致在处理具有极端对比度的室内场景时出现了一些性能下降。但总体而言, 本文方法在全体样本上的精确率从 95.0% 提升至 97.2%, 这表明在大多数情况下本文

表 2 模型验证结果

Table 2 Model vasslidation results %

类别	精确率	
	基线模型	本文方法
空闲台座	100.0	99.8
模板拼接	97.7	98.8
钢筋绑扎	96.1	95.8
混凝土浇筑	94.5	97.7
混凝土养护	85.9	87.7
模板拆除_洒水养护	96.7	100.0
梁体养生	99.8	99.8
铺设钢绞线	100.0	100.0
预应力张拉	85.5	97.7
封锚压浆	93.5	94.6
全体样本	95.0	97.2

方法都能够有效地提高预制梁工序检测识别的准确性, 尽管对于室内环境下的某些特定工序可能存在轻微的性能下降, 但整体影响并不明显。

为了进一步验证本算法的有效性和优越性, 本文将自制 WUT-PBP 数据集的标注格式转换为 COCO 格式, 重新训练了 SSD^[20]、DDOD^[21]、TOOD^[22]、RTMDet-s^[23]、DiffusionDet^[24]、DDQ^[25] 模型, 取与 $mAP@0.5$ 含义相同的评价指标 $AP^{IoU_{th}=0.5}$, 并计算各模型的参数量和推理速度, 如表 3 所示(加粗数值为最优值, 下划线数值为次优值)。其中, SSD 模型输入尺寸为 300×300 像素, 在 $mAP@0.5$ 方面为 79.2%, FPS 为 44 帧/s, 表现相对较好。DDQ、DiffusionDet、DDOD 和 TOOD 在 $mAP@0.5$ 指标上分别达到了 97.5%、96.6%、95.6%、93.8%, 但在 FPS 上略低于 SSD300, 更适合对准确性要求更高、而不需要高帧率的应用场景。RTMDet-s 是一个轻量化模型, 模型参数量仅为 8.860×10^6 , 在 $mAP@0.5$ 上比 YOLOv7 高了 0.3 百分点, 在准确性和模型复杂度方面均取得了良好的平衡, 但 FPS 略低于 YOLOv7。本文方法达到了最高的 $mAP@0.5$ (97.7%), 在 FPS 方面也表现出色, 仅次于 YOLOv7 的 84 帧/s, 模型参数量与其他模型相当, 且相比于基线模型, 参数量增加幅度较小, 表明本文方法在保持高准确性和实时性的同时, 有效地控制了模型复杂度, 适用于工业场景下的预制梁工序检测。

表 3 不同模型在 WUT-PBP 数据集上复现结果对比

Table 3 Reproduction results comparison of different models on WUT-PBP dataset

模型	$mAP@0.5/\%$	FPS/(帧 $\cdot s^{-1}$)	Params/ 10^6
SSD300 ^[20]	79.2	44	36.038
DDOD ^[21]	95.6	18	32.219
TOOD ^[22]	93.8	16	32.039
RTMDet-s ^[23]	97.2	60	8.860
DiffusionDet ^[24]	96.6	30	82.295
DDQ ^[25]	<u>97.5</u>	24	48.703
YOLOv7 ^[15]	96.9	84	36.530
本文方法	97.7	<u>78</u>	36.532

4.5 消融验证实验

表 4 展示了数据增强技术和注意力模块在基线模型上的表现, 其中, F 表示随机水平翻转和随机垂直翻转, B 表示随机亮度转换, G 表示添加高斯噪声, E 表示在动态学习网络中加入随机擦除模块, 静态学习网络和动态学习网络为添加了注意力模块后

单独训练的结果。实验结果表明,在精确率指标上,随机翻转和 5 种数据增强方法等全部应用方法表现出明显的改进,分别提高了 0.8 和 1.8 个百分点。在 $mAP@0.5$ 指标上,随机翻转和随机擦除相较基线模型分别提高了 0.6 和 0.5 个百分点,能够有效改善模型的检测性能。添加高斯噪声的方法在 $mAP@0.5$ 指标上有所下降,本文认为添加高斯噪声会导致图像中的细节丢失和特征模糊化,由于不同的模型可能会提取不同类型的特征或关注不同的图像部分,通过互学习的训练策略,两个网络分支互相学习,帮助对方捕获更多的信息和多样性特征,从而提高性能。

表 4 消融实验结果

Table 4 Ablation experiment results

模型	精确率	$mAP@0.5$
基线模型	95.0	96.9
基线模型+F	95.8	97.5
基线模型+B	95.4	97.2
基线模型+G	95.2	95.5
基线模型+E	95.4	97.4
静态学习网络	96.8	96.8
动态学习网络	96.0	96.8

为了验证 NAM 模块的有效性以及在网络中的最佳嵌入方式,以动态学习网络为例,以多种方式加入注意力模块,同时将 NAM_C 模块与全局注意力机制(GAM)^[26]和 SE 模块^[27]的效果进行对比,如图 6 所示。对于注意力模块的嵌入方式,3 种注意力模块在特征提取网络后部嵌入与在特征提取网络所有最大池化(MP)层之前加入相比,效果普遍较高,将 NAM_C 放置在特征提取网络后部的表现最佳,精确率相比于基线模型+E 提高了 0.6 个百分点。

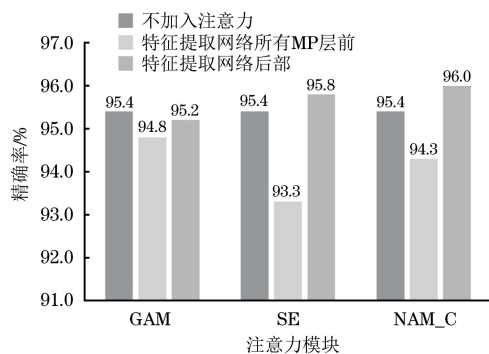


图 6 注意力模块验证实验

Fig.6 Validation experiment of attention module

4.6 互学习与正样本对齐验证实验

为验证本文设计的互学习框架与正样本对齐方式的有效性,本节首先采用独立训练的方式,分别训

练静态和动态学习网络分支,然后应用动静结合的互学习框架,采用两种正样本对齐方式,联合训练两个网络分支。如表 5 所示,交集在表示两个网络分支的正样本索引取交集作为最终的正样本索引,并集表示两个网络分支的正样本索引取并集作为最终的正样本索引。

表 5 互学习策略验证实验

Table 5 Validation experiment of

mutual learning strategy

互学习	网络分支	精确率	$mAP@0.5$
交集	静态分支	96.8	96.8
	动态分支	96.0	96.8
	静态分支	96.8	97.4
	动态分支	96.1	97.7
并集	静态分支	96.1	97.3
	动态分支	97.2	97.7

从表 5 可以看出,经过深度互学习,两个网络分支在检测性能上都胜过它们各自独立训练的同行模型,对于取并集的正样本对齐策略中的动态学习网络分支,不仅在精确率方面取得了最高水平,而且在检测性能上也取得了显著的提升。深度互学习的引入有效地提升了模型的整体性能,尤其是通过对动态学习网络的精细调整,使其在应对复杂场景和变化性情境中表现出色。

对于正样本对齐策略,可以验证当设计两个侧重能力不同的网络分支时,取并集有助于综合利用它们的优势。

4.7 可视化效果

为了更直观地显示本文方法的有效性,本节展示了热力图和检测框的可视化效果,如图 7 所示。在这些示例中,第 1 列和第 2 列的样本工序都为“混凝土浇筑”,然而,基线模型将它们错误地归类为“模板拆除/洒水养护”和“混凝土养护”类别,而本文方法则准确地识别到了它们的真实工序。此外,第 3 列样本中的“铺设钢绞线”也被错误地识别为“封锚压浆”。这些结果表明,在良好的光线条件下,基线模型对工序的正确识别仍然存在较高的误差率。相比之下,本文方法能够准确识别 T 梁台座上各工序的阶段。通过热力分布可以观察到,在光线不佳的第 4 列样本中,基线模型的关注区域较为模糊,未能准确地捕获所有的台座,并未成功检测到最左侧的台座。相比之下,本文模型的关注区域更为准确且全面,成功检测并正确地识别了全部 4 个台座的工序,展现出了卓越的检测和识别性能。

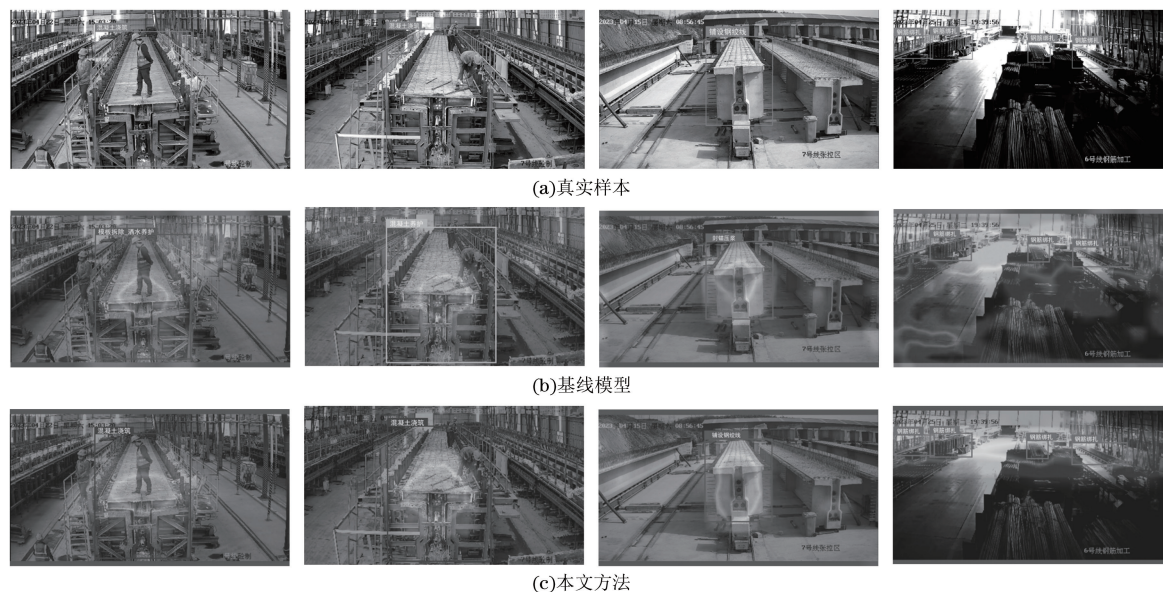


图 7 热力图和检测框的可视化效果

Fig. 7 Visualization effect of thermal maps and detection boxes

5 结束语

本文针对预制梁工序检测识别问题,构建了一个基于真实场景的高质量标注的预制梁工序数据集 WUT-PBP,提出一种基于动静结合的预制梁工序检测方法。在单阶段目标检测模型上构建了一个动静结合的互学习框架,静态学习网络分支专注于模拟真实场景在空间维度上的环境光照复杂性,而动态学习网络分支则致力于在时间维度上模拟真实场景中的噪声干扰随机性。通过正样本对齐策略的应用,实现边界框数量和表征分布的有效对齐,利用动态学习网络和静态学习网络的预测分布约束来相互学习、优化参数,真正综合利用了两个网络的优势,从而提高了目标识别的准确率。在自制的预制梁工序数据集 WUT-PBP 上进行的实验结果表明,本文方法在保证实时性的情况下精确率和 mAP@0.5 指标上达到了 97.2% 和 97.7%,相比于主流目标检测算法展现了其优异的性能。下一步将探索更多的正样本对齐策略,实现可适应多种应用场景的候选框互学习,进一步研究和开发能够更好地处理高度动态和复杂背景的模式。

参考文献

- [1] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Washington D. C., USA: IEEE Press, 2016: 779-788.
- [2] ZHANG Y, XIANG T, HOSPEDALES T M, et al. Deep mutual learning[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition, Washington D. C., USA: IEEE Press, 2018: 4320-4328.
- [3] 王琳毅,白静,李文静,等. YOLO 系列目标检测算法研究进展[J]. 计算机工程与应用, 2023, 59(14): 15-29. WANG L Y, BAI J, LI W J, et al. Research progress of YOLO series target detection algorithms [J]. Computer Engineering and Applications, 2023, 59(14): 15-29. (in Chinese)
- [4] 梁华刚,赵慧霞,刘丽华,等. 结合跨层特征融合与级联检测器的防震锤缺陷检测[J]. 中国图象图形学报, 2023, 28(11): 3485-3496. LIANG H G, ZHAO H X, LIU L H, et al. Combining cross-layer feature fusion with cascade detectors for anti-vibration hammer defects detection[J]. Journal of Image and Graphics, 2023, 28(11): 3485-3496. (in Chinese)
- [5] 阎光伟,周香君,焦润海,等. 融合先验信息和特征约束的杆塔螺栓缺陷检测[J]. 中国图象图形学报, 2023, 28(11): 3497-3508. YANG W, ZHOU X J, JIAO R H, et al. Defect detection of tower bolts by fusion of priori information and feature constraints [J]. Journal of Image and Graphics, 2023, 28(11): 3497-3508. (in Chinese)
- [6] 闵巍庆,刘林虎,刘宇昕,等. 食品图像识别方法综述[J]. 计算机学报, 2022, 45(3): 542-566. MIN W Q, LIU L H, LIU Y X, et al. A survey on food image recognition[J]. Chinese Journal of Computers, 2022, 45(3): 542-566. (in Chinese)
- [7] 李娟娟,王佳,陈亚杰,等. 基于改进 YOLO v4 的热轧带钢表面缺陷检测[J]. 计算机工程与设计, 2023, 44(9): 2786-2793. JI J J, WANG J, CHEN Y J, et al. Surface defect detection of hot rolled strip based on improved YOLO v4 [J]. Computer Engineering and Design, 2023, 44(9): 2786-2793. (in Chinese)
- [8] CAO X H, SU Y X, GENG X, et al. YOLO-SF: YOLO for fire segmentation detection[J]. IEEE Access, 2023, 11: 111079-111092.
- [9] 黄小平,冯涛,郭阳阳,等. 基于改进 YOLO v5s 的轻量级奶牛体况评分方法[J]. 农业机械学报, 2023, 54(6): 287-296. HUANG X P, FENG T, GUO Y Y, et al. Lightweight dairy cow body condition scoring method based on improved YOLO v5s[J]. Transactions of the Chinese Society for Agricultural

- Machinery, 2023, 54(6): 287-296. (in Chinese)
- [10] 郑道能. 一种改进的 tiny YOLO v3 煤矸石快速识别模型[J]. 工矿自动化, 2023, 49(4): 113-119.
- ZHENG D N. An improved tiny YOLO v3 rapid recognition model for coal-gangue[J]. Industry and Mine Automation, 2023, 49(4): 113-119. (in Chinese)
- [11] 王金平, 刘春芽, 郑凤飞, 等. 基于目标检测与语义分割智慧工地施工人员安全状况判断研究[J]. 公路, 2023, 68(4): 282-288.
- WANG J P, LIU C Y, ZHENG F F, et al. Research on safety status judgment of smart site construction workers based on target detection and semantic segmentation[J]. Highway, 2023, 68(4): 282-288. (in Chinese)
- [12] ZHANG Y, FANG X, LI F, et al. Mutual learning inspired prediction network for video anomaly detection[C]//Proceedings of Conference on Pattern Recognition and Computer Vision. Berlin, Germany: Springer, 2022: 578-593.
- [13] WU R M, FENG M Y, GUAN W L, et al. A mutual learning method for salient object detection with intertwined multi-supervision[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2019: 8142-8151.
- [14] MITTAL G, YAGNIK K B, GARG M, et al. SpotGarbage: smartphone app to detect garbage using deep learning[C]//Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. New York, USA: ACM Press, 2016: 940-945.
- [15] HUANG Y B, QIU C Y, GUO Y, et al. Surface defect saliency of magnetic tile[C]//Proceedings of the 14th IEEE International Conference on Automation Science and Engineering. Washington D. C., USA: IEEE Press, 2018: 612-617.
- [16] TANG S L, HE F, HUANG X L, et al. Online PCB defect detector on a new PCB defect dataset[EB/OL]. [2023-11-10]. <https://arxiv.org/abs/1902.06197v1>.
- [17] 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 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2023: 7464-7475.
- [18] LIU Y C, SHAO Z R, TENG Y Y, et al. NAM: normalization-based attention module[EB/OL]. [2023-11-10]. <https://arxiv.org/abs/2111.12419v1>.
- [19] ZHONG Z, ZHENG L, KANG G L, et al. Random erasing data augmentation[J]. Artificial Intelligence, 2020, 34(7): 13001-13008.
- [20] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2016: 21-37.
- [21] CHEN Z H, YANG C, LI Q F, et al. Disentangle your dense object detector[C]//Proceedings of the 29th ACM International Conference on Multimedia. New York, USA: ACM Press, 2021: 4939-4948.
- [22] FENG C J, ZHONG Y J, GAO Y, et al. TOOD: task-aligned one-stage object detection[C]//Proceedings of IEEE/CVF International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2021: 3490-3499.
- [23] LYU C Q, ZHANG W W, HUANG H A, et al. RTMDet: an empirical study of designing real-time object detectors[EB/OL]. [2023-11-10]. <https://arxiv.org/abs/2212.07784v2>.
- [24] CHEN S F, SUN P Z, SONG Y B, et al. DiffusionDet: diffusion model for object detection[C]//Proceedings of IEEE/CVF International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2023: 19773-19786.
- [25] ZHANG S L, WANG X J, WANG J Q, et al. Dense distinct query for end-to-end object detection[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2023: 7329-7338.
- [26] LIU Y C, SHAO Z R, HOFFMANN N. Global attention mechanism: retain information to enhance channel-spatial interactions[EB/OL]. [2023-11-10]. <https://arxiv.org/abs/2112.05561v1>.
- [27] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press 2018: 7132-7141.

编辑 索书志