中图法分类号:TP18;TP391　　**文献标识码**:A　　**文章编号**:1006-8961(2026)10-3671-20

论文引用格式:Wang Y, Zhang M J, Cao J J, Cheng H J, Xue X S and Liu W X. 2026. Cross-domain feature alignment and two-stage enhancement for low-light object detection. Journal of Image and Graphics, 31(10):3671-3690(王洋, 张美璟, 曹菁菁, 程红举, 薛醒思, 刘文犀. 2026. 跨域特征对齐与两阶段增强的低光目标检测方法. 中国图象图形学报, 31(10):3671-3690)[DOI:10.11834/jig.250588]

跨域特征对齐与两阶段增强的低光目标检测方法

王洋[1],张美璟[2*],曹菁菁[3],程红举[5],薛醒思[4],刘文犀[5]

1. 福建农林大学机电工程学院,福州　350002; 2. 福建警察学院计算机与信息安全管理系,福州　350007;
3. 武汉理工大学交通与物流工程学院,武汉　430070; 4. 福建理工大学计算机科学与数学学院,福州　350118;
5. 福州大学计算机与大数据学院,福州　350108

摘　要:**目的**　解决低光环境下图像噪声、对比度不足及细节退化导致的传统目标检测性能下滑问题,缓解高质量标注低光图像稀缺瓶颈,挖掘大规模自然图像数据集价值以提供高效训练支撑。**方法**　提出一种创新的低光目标检测框架,以正常光照图像预训练模型为基础构建核心训练范式。首先,为弥合正常光照与低光图像间的域分布鸿沟,设计光照不变特征对齐模块(illumination-invariant feature alignment,IIFA),通过全局光照特征对齐与光照无关特征提取的协同机制实现跨域特征融合,并嵌入自适应样本优先级处理机制以适配域间分布差异;其次,针对低光场景下小型及远距离目标因细节退化、对比度不足导致的检测难题,提出一种基于两阶段增强策略的局部特征强化注意力模块(two-stage feature enhancement,TSFE),实现对关键目标特征的精准增强。**结果**　在低光检测基准ExDark数据集上的实验结果表明,所提模型性能优于直接低光训练、正常光预训练后低光微调等现有方法,尤其在小型目标检测任务中表现突出。具体指标如下:IIFA跨域适应场景下平均精度均值(mean average precision,mAP)达65.74%,帧率(frames per second,FPS)达57.03帧/s;所提模型mAP达66.83%,精确率(precision,P)达83.73%,召回率(recall,R)达54.78%,同时保持48.81帧/s的FPS,综合性能与检测效率优势显著。**结论**　本框架有效弥合跨域差异,提升小型及远距离目标检测能力,显著改善低光检测性能,提供了高效解决方案。

关键词:目标检测;域自适应;跨域融合;图像增强;低光

Cross-domain feature alignment and two-stage enhancement for low-light object detection

Wang Yang[1], Zhang Meijing[2*], Cao Jingjing[3], Cheng Hongju[5], Xue Xingsi[4], Liu Wenxi[5]

1. College of Mechanical and Electrical Engineering, Fujian Agriculture and Forestry University, Fuzhou　350002, China;
2. Department of Computer and Information Security Management, Fujian Police College, Fuzhou　350007, China;
3. School of Transportation and Logistics Engineering, Wuhan University of Technology, Wuhan　430070, China;
4. School of Computer Science and Mathematics, Fujian University of Technology, Fuzhou　350118, China;
5. College of Computer and Data Science, Fuzhou University, Fuzhou　350108, China

Abstract: **Objective**　Object detection, a foundational task in computer vision, has attained remarkable performance under well-illuminated conditions with the advancement of deep learning techniques. However, its practical deployment is

收稿日期:2025-11-21;**修回日期**:2026-02-10;**预印本日期**:2026-02-17

*通信作者:张美璟　zhangmeijing@fjpsc.edu.cn

基金项目:国家自然科学基金项目(62072110,62372111);福建省科技兴警科研专项项目(2024YZ040001)

Supported by: National Natural Science Foundation of China (62072110, 62372111); Policing Science and Technology of Fujian Province (2024YZ040001)

severely impeded by low-light environments, which are ubiquitous in real-world scenarios including nighttime surveillance, autonomous driving during dusk, and low-illumination indoor monitoring. Traditional object detection methods exhibit substantial performance degradation under low-light conditions, primarily stemming from three core challenges: excessive image noise induced by insufficient photon capture of imaging sensors, reduced contrast between foreground objects and backgrounds that obscures object contours, and severe loss of fine-grained details (e. g. , texture and edge information), which are all critical for accurate object localization and classification. Meanwhile, the development of low-light object detection techniques is further constrained by the scarcity of high-quality annotated low-light datasets. Annotating low-light images demands professional expertise to distinguish genuine object features from noise interference, resulting in high annotation costs and limited dataset scales. By contrast, large-scale natural image datasets (e. g. , COCO and ImageNet) with high-quality annotations are readily accessible, encompassing diverse object categories and scene variations. Consequently, this study aims to address the performance degradation of traditional object detection methods in low-light environments, alleviate the bottleneck posed by scarce high-quality annotated low-light images, and fully exploit the value of large-scale natural image datasets to provide efficient and reliable training support for low-light object detection. **Method** An innovative low-light object detection framework, which adopts a core training paradigm based on normal-light pretrained models to maximize the utilization of existing high-quality natural image resources, is proposed to achieve the aforementioned objectives. The framework consists of two key modules designed to address domain differences and small/distant object detection challenges, respectively. First, to bridge the substantial domain gap between normal-light and low-light images, characterized by distinct illumination distributions and feature representations, an illumination-invariant feature alignment (IIFA) module is developed. This module implements cross-domain feature fusion through a twofold strategy: global illumination feature alignment and illumination-independent feature extraction. The global illumination alignment component normalizes the illumination statistics (e. g. , brightness and contrast) of normal-light and low-light features to a unified space, reducing the impact of illumination intensity differences. The illumination-independent feature extraction component employs an attention mechanism to focus on object structural features (e. g. , shape and contour) that are less affected by illumination changes, while suppressing illumination-related noise features. Additionally, an adaptive sample priority processing mechanism is embedded in the IIFA module to adjust the weight of training samples based on their domain adaptation difficulty dynamically. Samples with larger domain distribution differences (e. g. , low-light images with extreme noise) are assigned higher weights during training, enabling the model to prioritize learning from challenging samples and adapt more effectively to interdomain distribution differences. Second, aiming at the prominent challenge of detecting small or distant objects in low-light environments where degraded details and low contrast make it difficult for the model to capture discriminative features, a two-stage feature enhancement (TSFE) module is proposed. The first stage of the module performs multiscale feature fusion to aggregate low-level detail features and high-level semantic features of small/distant objects, compensating for the detail loss caused by low illumination. The second stage applies a local feature reinforcement attention mechanism to highlight the key feature regions of small/distant objects (e. g. , object heads and limbs for pedestrian detection) while suppressing background noise. This two-stage enhancement strategy ensures that the model can accurately locate and identify small or distant objects even in low-light conditions by sequentially enhancing feature completeness and discriminability. **Result** Comprehensive experiments are conducted on the ExDark dataset, a widely recognized low-light object detection benchmark that contains 8 856 low-light images covering 12 common object categories (e. g. , person, car, and bicycle) and various low-light scenarios (e. g. , night streets and dim rooms). The proposed model is compared with two mainstream low-light object detection methods: direct training on low-light datasets and normal-light pretraining followed by low-light fine-tuning. The evaluation metrics include mean average precision (mAP) for detection accuracy, precision (P) and recall (R) for category-specific performance, and frames per second (FPS) for detection efficiency. Experimental results demonstrate that the proposed model outperforms the comparison methods in all evaluation metrics, with particularly outstanding performance in small object detection. Specific metrics are as follows: In cross-domain adaptation scenarios (using normal-light pretrained models without fine-tuning on low-light data), the model achieves an mAP of 65. 74% and an FPS of 57. 03 frame/s, outperforming the direct low-light training method and maintaining comparable efficiency. When fine-tuned on the ExDark dataset, the proposed model achieves an mAP of 66. 83%, a

precision of 83.73%, and a recall of 54.78%, while maintaining an FPS of 48.81 frame/s, verifying the effectiveness of the IIFA and TSFE modules. The FPS of 48.81 also meets the real-time detection requirements of practical applications such as surveillance and autonomous driving. **Conclusion** This framework effectively bridges cross-domain differences between normal-light and low-light images, enhances the detection capability for small and distant objects, and remarkably improves overall low-light object detection performance, providing an efficient solution to low-light detection challenges.

Key words: object detection; domain adaptive; cross-domain fusion; image enhancement; low-light

0 引　言

目标检测作为计算机视觉的基础任务，在自动驾驶、监控系统和工业质检等领域应用广泛（Badue等，2021；Shen等，2025）。然而现有目标检测方法在低光环境下表现显著下降，这主要归因于图像噪声放大、对比度降低及光照分布不均等因素（Pan等，2024；Xu等，2025a）。研究表明，这些因素会导致目标可见性降低，检测精度大幅下降（Shen等，2025）。主流目标检测任务大致可划分为两类：单阶段检测器（如SSD（single shot multibox detector）（Cui等，2022）、YOLO（you only look once）（Redmon和Farhadi，2018）以及FCOS（fully convolutional one-stage object detection）（Dong等，2017））与双阶段检测器（如Faster R-CNN（faster region-based convolutional neural network）（Ren等，2017）与R-FCN（region-based fully convolutional network）（Dai等，2016））。其中，单阶段目标检测器通过单步预测直接输出目标边界框和类别标签，双阶段目标检测器则先生成区域建议集，再经分类与边界框回归完成优化。近年来，基于Transformer架构的DETR（detection transformer）（Carion等，2020）及其衍生方法（Meng等，2021；Liu等，2022；Li等，2023）相继提出，用于提升通用目标检测的综合性能。与之不同，低光目标检测器是针对低光照场景专门设计的检测器，其架构多源于通用目标检测方法，常采用双阶段设计思路，同步生成目标边界框与分类置信度。目前，该领域的研究重点主要集中在模型架构优化、锚点采样与匹配策略改进以及特征增强技术研发的3个方面。

当前，主流低光目标检测的研究（Wang和Liu，2021；Xu等，2021；Du等，2024）主要分为两类：一类在大规模低光真实数据集上训练（Wang和Liu，2021）；另一类在正常光照图像上完成预训练，再基于低光数据进行微调。然而，低光图像的收集与标注过程面临诸多困难，同时获取大规模、高质量的低光数据集也存在较大挑战。此外，仅在正常光图像上训练的检测器在低光条件下往往表现欠佳（Li等，2022a），其具体问题主要表现为信噪比下降和特征分布偏移。另一条研究路径则是借助直方图均衡化、对比度拉伸和色彩还原等图像增强技术，以实现低光图像质量的提升（Zhang等，2023）。尽管此类方法可通过增强图像亮度和对比度提升目标物体的可见度，但往往难以充分保留图像细节特征，进而导致其在小型目标或远距离物体检测任务时效果欠佳。潘晓英等人（2023）系统总结了小目标检测方法，指出小目标因像素占比小、语义信息少、易受复杂场景干扰等问题一直是目标检测领域的难点，并从数据增强、超分辨率以及多尺度特征融合等方面梳理了现有解决方案。近年来，相关工作主要围绕图像增强、专用检测架构及领域自适应技术三大方向展开。传统方法（Zhang等，2021；Dai等，2024）依赖Retinex（Land和McCann，1971）理论或直方图均衡化等技术实现低光图像增强，但这类方法往往易引入噪声和伪影（罗强 等，2018）。马龙等人（2022）在低光照图像增强综述中系统梳理了该领域研究现状，指出传统方法受限于数据分布且泛化能力有限。随着深度学习技术的快速发展，基于神经网络的低光图像增强方法（如Zero-DCE（zero-reference deep curve estination）（Guo等，2020）、KinD（Li等，2019））已取得了突破性进展。Cai等人（2023）基于Retinex理论提出了Retinexformer，通过光照引导的Transformer建模不同光照区域的非局部交互，在多个基准数据集上显著超越现有方法，并在ExDark低光目标检测任务上验证了其实际应用价值。Hashmi等人（2023）提出的FeatEnHancer模块，通过多头注意力机制层次化结合多尺度特征，确保网络能够提取更具代表性和判别行的增强特征。赵明华等人（2024）提出基于照度与场景纹理注意力图的低光图

像增强算法，利用最小通道约束图进行照度和纹理注意力估计，设计全局与局部相结合的增强模块，有效解决了局部欠增强和过增强问题。张奇等人(2026)提出基于HVI色彩空间的门控卷积双分支增强网络，通过将亮度信息与色彩结构信息分离处理，实现了色彩保真度与亮度伪影抑制的平衡。Zhang等人(2024)提出的ISP-Teacher，从相机图像信号处理(image signal processing，ISP)成像过程角度重新审视教师—学生架构以解决暗光目标检测问题，设计了与相机传感器ISP流程一致的日到夜转换模块(ISP-based day-to-night Transformer module，ISP-DTM)，并通过解耦正则化避免ISP退化任务与检测任务的冲突，实现跨域迁移。Du等人(2024)提出的DAI-Net(day- night adaptation interaction network)，实现了零样本日夜域自适应，无需真实低光数据即可将检测器从正常光照场景泛化到低光场景。然而，低光图像的收集与标注过程面临诸多困难，获取大规模高质量的低光数据集存在较大挑战，仅在正常光图像上训练的检测器在低光条件下往往表现欠佳。同时，Nightowls(Neumann等，2019)、ExDark(Loh和Chan，2019)、DARKFACE(Wang等，2023)及NOD(novel object detection)(Morawski等，2021)等低光数据集也已相继构建。然而，这些数据集的规模仍不及正常环境下的数据集。

域自适应技术通过利用标注充分的源域数据辅助目标域模型训练，从而缓解跨域分布差异带来的性能下降，为解决低光场景数据稀缺问题提供了有效途径。早期研究中，Saito等人(2019)提出的SWDA(strong-weak distribution alignment)网络，通过在局部特征层采用强对齐、全局特征层采用弱对齐的对齐策略实现自适应目标检测。Hsu等人(2020)提出的渐进式域自适应框架PDA(progressive domain adaptation)，通过构建中间域将大的域差距分解为多个较小的适应子任务，逐步缩小源域与目标域之间的分布差距。He等人(2022)提出TDD(target-perceived dual-branch distillation)网络，通过目标感知的双分支蒸馏架构实现跨域目标检测，利用知识蒸馏机制有效迁移源域检测知识。近年来，无监督域自适应方法取得了长足进步。Kennerley等人(2023)提出的两阶段一致性无监督域自适应网络2PCNet，在第1阶段中使用教师网络生成高置信度边界框，并将其加入学生网络的候选区域中；在第2阶段中教师网络对这些候选区域进行重新评估，生成包含高/低置信度的伪标签，并反馈给学生网络进行训练。Zhou等人(2023a)从控制理论中稳定性的概念出发，将不同域之间图像和区域的差异视为扰动，引入网络稳定性分析框架，通过分析不同干扰下的稳定性实现域自适应检测。针对域移位问题，Li等人(2022b)提出了逐步域自适应YOLO框架(S-DAYOLO)，通过构建辅助域桥接域差距并结合域自适应YOLO实现自动驾驶场景的跨域目标检测。Zhou等人(2023a)提出的SSDA-YOLO，采用基于Mean Teacher的半监督训练框架结合域自适应技术，将YOLOv5与域自适应相结合实现高效的跨域目标检测。在无源域自适应研究领域中，Varailhon等人(2025)提出的SF-YOLO，通过teacher-student框架结合目标域特定数据增强策略进行模型优化，实现了无需访问源域数据的YOLO目标检测域自适应。针对现有方法在特征对齐时对不同区域重要性考虑不足的问题，He等人(2025)设计了预测差异反馈实例对齐模块，自适应地为教师—学生检测差异更大的实例分配更高权重，并且开发了基于不确定性的前景目标图像对齐模块，自适应地优先处理前景目标区域。Zheng等人(2025)提出的双概率对齐(dual probabilistic alignment，DPA)框架，通过将域概率建模为高斯分布，在全局层面对齐域私有类别、在实例层面对齐域共享类别，从而统一解决开集、部分集和闭集场景下的通用域自适应目标检测问题。这些域自适应方法为弥合正常光照与低光环境之间的分布差异提供了重要技术支撑。

注意力机制(贾可心 等，2022)通过自适应地聚焦于图像中的关键区域，已成为提升目标检测性能的重要技术手段。SE(squeeze-and-excitation)注意力模块(Hu等，2018)通过全局平均池化和全连接层建模通道间依赖关系，在此基础上，CBAM(convolutional block attention module)模块(Woo等，2018)引入空间注意力分支实现通道与空间的联合建模。CA(coordinate attention)模块(Hou等，2021)通过水平和垂直方向的池化操作精确编码空间位置信息实现将位置信息嵌入通道注意力。何伟和潘晨(2022)提出的AGNet(attention-guided network)网络，结合通道注意力和空间注意力机制有选择地聚合深层和浅层的特征信息，实现了显著性目标检测中不同层次特征的有效传递和聚合。Dong等人(2025)提出

的AugDETR网络，通过设计的混合注意力编码器扩大可变形编码器的感受野并引入全局上下文特征，并且自适应利用多层编码器进行更具判别性的目标特征提取。在分析了CA模块中批归一化泛化能力不足和通道降维的负面影响后，Xu等人（2025b）提出了ELA（efficient local attention）模块，采用一维卷积和组归一化高效编码位置特征，在目标检测任务上优于SE、CBAM和CA等经典模块。TA-YOLO网络（Li等，2024）中构建的多头通道和空间Trans-Attention模块，从通道和空间维度分别进行远程像素交互完成注意力特征捕获，有效解决了遥感图像复杂背景下的小目标检测问题。这些注意力机制的发展为低光场景下的特征增强提供了丰富的技术基础。

现有目标检测方法在正常光照下表现优异，但在低光环境中面临两大核心挑战。正常光照与低光环境间存在显著领域偏移，导致模型难以直接迁移。现有域适应方法多停留在像素级光照调整，忽视深层特征空间的分布差异，在源域与目标域的语义一致性处理上缺乏系统设计，易引发类别错位。此外，低光环境下对比度下降、细节丢失，小型及远距离目标检测尤为困难。现有特征增强局限于单一维度改进，如仅强化通道或空间注意力，缺乏多维度协同建模。而且固定增强策略无法适配不同尺度目标需求，难以同时满足小型目标的局部纹理捕捉与远距离目标的上下文语义理解。

为解决上述问题，本文提出了一种基于跨域对齐与细节增强的低光目标检测算法。在跨域适配上，提出光照不变特征对齐模块，通过最大均值差异实现多尺度特征分布对齐，引入基于语义亲和度的对比学习提取光照不变特征，设计动态类别对齐损失自适应调整样本权重，从分布、特征和类别3个层面协同弥合域差距。在特征判别上，提出两阶段特征增强模块，第1阶段通过通道—空间双向交互实现多维度协同增强，第2阶段整合动态残差、自适应门控、空间注意力与多尺度增强，构建多路径并行的上下文感知机制，使模型灵活适配从局部纹理到宏观上下文的不同尺度需求。

本文的主要创新点和工作总结如下：1）提出光照不变特征对齐模块（illumination-invariant feature alignment，IIFA），融合跨域特征对齐、光照不变表征学习与动态类别对齐的三要素混合优化策略，通过最大均值差异实现多尺度特征分布对齐，引入基于语义亲和度的对比学习提取域不变特征，并设计类别动态对齐损失函数自适应调整样本权重，从分布、特征和类别3个层面协同弥合跨域差距。2）提出两阶段特征增强模块（two-stage feature enhancement，TSFE），第1阶段通过通道—空间双向交互网络实现多维度协同特征增强；第2阶段整合动态残差块、自适应注意力门控、高效空间注意力及多尺度特征增强4个组件，构建多路径并行与多注意力协调的上下文感知机制，提升小型及远距离目标的特征判别力。3）在ExDark数据集上的实验验证了各模块的有效性及协同增益，所提模型在小型目标检测任务中达到前沿性能。

1 本文方法

1.1 模型概述

本研究提出一种基于迁移学习的低光场景的弱监督目标检测框架，充分利用正常光照条件下大规模数据集的先验知识，通过跨域特征对齐实现向低光场景的有效迁移。如图1所示，该框架以低光场景目标域图像$\boldsymbol{I}_{\text{target}}$和随机选取的正常光照源域图像$\boldsymbol{I}_{\text{source}}$为输入进行训练。其核心包含光照不变特征对齐模块和两阶段特征增强模块两个关键组件。前者负责弥合正常光照与低光环境之间的域分布差异；后者针对低光场景中小型及远距离目标的检测难题进行特征强化。

在训练阶段，源域图像和目标域图像分别经过共享参数的CSPNet（cross stage partial network）主干网络进行特征提取，再通过特征金字塔网络（feature pyramid network，FPN）获取多尺度特征表示（黄强等，2023）。源域图像的多尺度特征经过检测头生成预测结果，与源域标注标签计算检测损失L_{DET}，为目标检测模型提供监督信息。与此同时，目标域图像通过光照不变特征对齐模块的域适应单元，与随机选取的源域参考图像进行特征交互，该模块通过3个域适应损失函数实现特征空间的对齐：跨域特征对齐损失L_{MMD}用于缩小两域全局特征的分布差异；对比学习损失L_{Contrast}通过正负样本对的语义关联学习光照不变的特征表征；动态类别对齐损失L_{DCA}则通过自适应加权机制校准类别不平衡导致的跨域偏移。上述损失函数的联合优化使模型能够在弱监

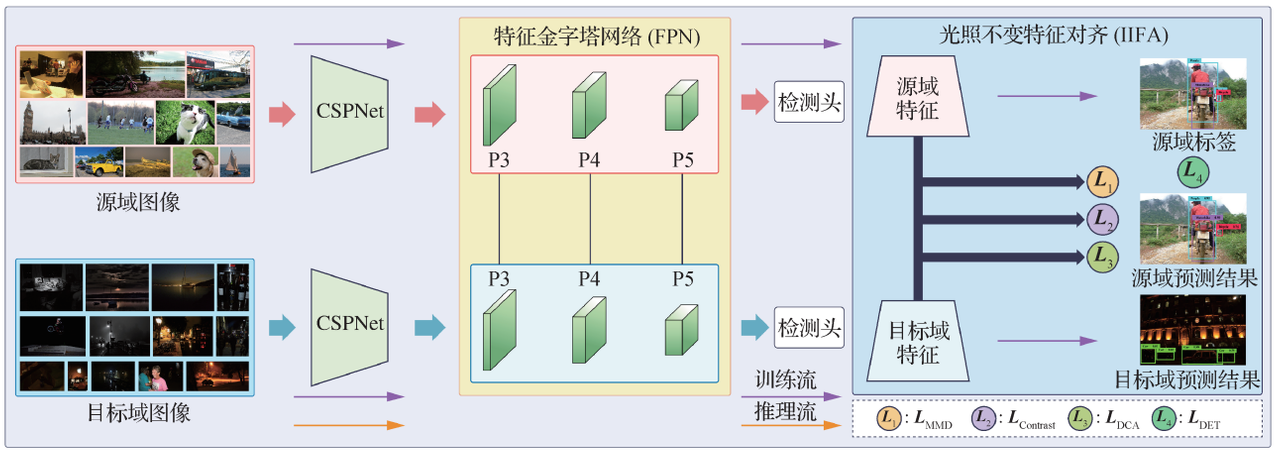


图 1　系统架构全景图

Fig. 1　Overview of our framework

督设置下实现跨域特征表征的对齐。

在推理阶段，仅输入目标域低光图像，经主干网络和特征金字塔网络提取多尺度特征后，通过两阶段特征增强模块进行特征强化，最终由检测头输出预测结果。由于训练阶段已完成跨域特征对齐，推理时无需源域图像参与，从而保证了实际部署的高效性。下文将详细阐述各核心模块的技术实现细节。

1.2　光照不变特征对齐

尽管自然图像数据集为目标检测模型提供了丰富的训练资源，但标注完备的低光数据仍然稀缺。这种稀缺性进一步被正常光照与低光环境之间的显著领域差异而加剧——当直接将正常光照场景下训练所得的模型迁移到低光场景时，光照强度、图像对比度及噪声分布模式的固有差异会导致模型性能显著下降。为有效弥合这一跨域差异，本文提出一种IIFA模块，专门用于低光场景与正常光照场景的跨域适应任务。该模块通过分离光照相关伪影与物体判别性特征，实现源域与目标域多尺度特征的精细对齐，从而使模型在不同光照条件下均能保持稳定的特征迁移能力。

为实现上述目标，该模块采用融合三要素的混合优化策略：1）跨域特征对齐技术（cross-domain feature alignment，CDFA），通过缩小源域与目标域全局特征的分布距离，降低跨域分布偏移影响；2）光照不变表征学习（illumination-invariant feature representation，IIFR），强化特征对光照变化的鲁棒性，使同类目标在不同光照条件下形成统一的特征表征；3）动态类别对齐机制（dynamic category alignment，DCA），通过动态调整类别权重分配，有效解决跨域场景下类别不平衡问题。

1.2.1　跨域特征对齐

充分利用正常光照条件下积累的先验知识，核心环节之一是实现目标场景（即低光场景）与正常光照场景的光照条件对齐。

为此，本文致力于最小化源图像$\boldsymbol{I}_{\text{src}}$与目标图像$\boldsymbol{I}_{\text{tar}}$分布之间的域差异。具体而言，采用最大均值差异（maximum mean discrepancy，MMD）（Mohamadi等，2024）作为跨域特征对齐度量方法。通过FPN网络从$\boldsymbol{I}_{\text{src}}$和$\boldsymbol{I}_{\text{tar}}$中提取层次化多尺度特征。鉴于FPN的中间输出特征包含不同层级的语义信息，本文选取3个代表性层的特征，分别标记为$\{\boldsymbol{X}'_{\text{src}}, \boldsymbol{X}''_{\text{src}}, \boldsymbol{X}_{\text{src}}\}$和$\{\boldsymbol{X}'_{\text{tar}}, \boldsymbol{X}''_{\text{tar}}, \boldsymbol{X}_{\text{tar}}\}$，对应第3、4和最后一个卷积块。浅层特征主要编码边缘、纹理等低级视觉细节，而深层特征则蕴含更丰富的目标语义表征信息。为最小化跨域特征差异，同时缓解光照条件变化与场景内容差异引发的域偏移问题，本文将模型训练目标具体化为

$$L_{\text{MMD}}\left(\boldsymbol{X}'_{\text{src}}, \boldsymbol{X}'_{\text{tar}}\right) = \left\| \phi\left(\boldsymbol{X}'_{\text{src}}\right) - \phi\left(\boldsymbol{X}'_{\text{tar}}\right) \right\|^2 \tag{1}$$

$$L_{\text{MMD}}\left(\boldsymbol{X}_{\text{src}}, \boldsymbol{X}_{\text{tar}}\right) = \left\| \phi\left(\boldsymbol{X}_{\text{src}}\right) - \phi\left(\boldsymbol{X}_{\text{tar}}\right) \right\|^2 \tag{2}$$

式中，$\phi(g)$表示将特征映射到再生核希尔伯特空间（reproducing kernel Hilbert space，RKHS）（Park等，2019）的函数。通过最小化该损失函数L_{MMD}，能够有效缩小源域与目标域的特征分布差异，促使两域特征分布趋于一致，进而提升跨域特征的可迁移性。

1.2.2　光照不变表征学习

为进一步强化领域无关特征的表征能力，本文

引入对比学习框架构建对比损失函数。具体而言，通过采样源域图像与目标域图像对，并结合共享语义分析技术实现正负样本的有效区分。考虑到单幅图像通常包含多个不同语义标签的对象，本文通过量化标签共现频率来度量图像对的语义亲和度：语义高度重合的图像对被定义为正样本，语义差异显著的图像对则被定义为负样本。具体地，两幅图像的语义亲和度通过标签集合的 Jaccard similarity（Chen 等，2019）进行计算，具体为

$$\boldsymbol{S}_{i,j}=\frac{\boldsymbol{L}_{\mathrm{src}}^{i}\cap\boldsymbol{L}_{\mathrm{tar}}^{j}}{\left|\boldsymbol{L}_{\mathrm{src}}^{i}\right|\cup\left|\boldsymbol{L}_{\mathrm{tar}}^{j}\right|} \tag{3}$$

式中，$\boldsymbol{L}_{\mathrm{src}}^{i}$ 表示第 i 个源域图像的语义标签集合；$\boldsymbol{L}_{\mathrm{tar}}^{j}$ 表示第 j 个目标域图像的语义标签集合；对于每个源域样本 $\boldsymbol{I}_{\mathrm{src}}$，从目标域中随机选取一个语义亲和性较高的图像 $\boldsymbol{I}_{\mathrm{tar}}^{+}$ 作为正样本对 $\{\boldsymbol{I}_{\mathrm{src}},\boldsymbol{I}_{\mathrm{tar}}^{+}\}$，反之，从目标域中随机选取一个语义亲和性较低的图像样本 $\boldsymbol{I}_{\mathrm{tar}}^{-}$ 作为负样本对 $\{\boldsymbol{I}_{\mathrm{src}},\boldsymbol{I}_{\mathrm{tar}}^{-}\}$。因此，目标函数可定义为

$$L_{\mathrm{Contrast}}=-\log\frac{\exp\left(sim\left(\boldsymbol{X}_{\mathrm{src}},\boldsymbol{X}_{\mathrm{tar}}^{+}\right)/\tau\right)}{\exp\left(sim\left(\boldsymbol{X}_{\mathrm{src}},\boldsymbol{X}_{\mathrm{tar}}^{+}/\tau\right)+\sum_{c=1}^{C}\exp\left(sim\left(\boldsymbol{X}_{\mathrm{src}},\boldsymbol{X}_{\mathrm{tar}}^{-}\right)/\tau\right)\right.} \tag{4}$$

式中，$\boldsymbol{X}_{\mathrm{src}}$，$\boldsymbol{X}_{\mathrm{tar}}^{+}$ 和 $\boldsymbol{X}_{\mathrm{tar}}^{-}$ 分别代表源域样本（$\boldsymbol{I}_{\mathrm{src}}$）、正样本（$\boldsymbol{I}_{\mathrm{tar}}^{+}$）对中目标域样本和负样本（$\boldsymbol{I}_{\mathrm{tar}}^{-}$）对中目标域样本的特征向量；$sim(\cdot,\cdot)$ 用于衡量两个特征向量间的余弦相似度（刘爱琴 等，2013）；τ 是相似度评分调节参数（Chen 等，2019；Wang 和 Liu，2021）。通过该对比学习机制进一步强化跨域特征对齐，模型能够有效学习到鲁棒性更强的域不变特征表征，为低光场景目标检测提供可靠的特征支撑。

1.2.3　动态类别对齐

此外，本文引入对抗学习机制，基于图像纹理细节的风格特征，将源域图像特征映射到目标域特征空间，实现跨域特征的精准匹配与转换。然而，训练样本集中不同类别目标的数量存在差异，导致类别不平衡问题成为影响跨域对齐的关键挑战。为此，本文设计了类别动态对齐损失函数，通过自适应加权机制，对包含少数类目标的样本赋予更高权重，优先优化此类样本的特征对齐效果。基于源域特征 $\boldsymbol{X}_{\mathrm{src}}$ 和目标域特征 $\boldsymbol{X}_{\mathrm{tar}}$，对抗损失函数可定义为

$$L_{\mathrm{DCA}}\left(\boldsymbol{X}_{\mathrm{src}},\boldsymbol{X}_{\mathrm{tar}}\right)=\boldsymbol{E}\left[\log\boldsymbol{D}\left(\boldsymbol{X}_{\mathrm{src}}\right)+\boldsymbol{E}\left[\omega\log\left(1-\boldsymbol{D}\left(\boldsymbol{G}\left(\boldsymbol{X}_{\mathrm{tar}}\right)\right)\right)\right]\right] \tag{5}$$

式中，ω 表示目标域图像的动态权重；$\boldsymbol{G}(\cdot)$ 为生成器，负责将源域局部特征转换为符合目标域风格的特征；$\boldsymbol{D}(\cdot)$ 为判别器，用于区分输入特征源自目标域还是生成器生成的转换特征。在实验实现中，本文采用 U-Net（Yan 等，2019）架构构建生成器 $\boldsymbol{G}(\cdot)$，并选用 PatchGAN（Schönfeld 等，2020）作为判别器 $\boldsymbol{D}(\cdot)$。

为实现动态权重 ω 的自适应调整，本文引入多标签分类器 $\boldsymbol{F}_{\mathrm{MLC}}$ 评估目标域样本中各类别的跨域适应难度（Xiao 和 Zhang，2021）。具体而言，首先利用源域标注样本训练该多标签分类器，随后将训练完成的分类器应用于目标域样本的适应难度评估。其输出结果为目标域样本经源域至目标域跨域适应后，各候选类别的识别概率分布。基于此，本文构建类别动态对齐损失函数，具体为

$$L_{\mathrm{MLC}}=-\sum_{c=1}^{C}\left[\boldsymbol{y}_{\mathrm{src}}^{c}\log\left(\hat{\boldsymbol{y}}_{\mathrm{src}}^{c}\right)+\left(1-\boldsymbol{y}^{c}\right)\log\left(1-\hat{\boldsymbol{y}}_{\mathrm{src}}^{c}\right)\right] \tag{6}$$

式中，C 代表对象类总数；$\boldsymbol{y}_{\mathrm{src}}^{c}\left(\boldsymbol{y}_{\mathrm{src}}^{c}=\{0,1\}\right)$ 表示类别 c 是否存在于源域图像中；$\hat{\boldsymbol{y}}_{\mathrm{src}}^{c}$ 表示类别 c 存在于源域图像的概率，该概率由多标签分类器 $\boldsymbol{F}_{\mathrm{MLC}}$（刘爱琴 等，2013）估算得出，即 $\hat{\boldsymbol{y}}_{\mathrm{src}}^{c}=\boldsymbol{F}_{\mathrm{MLC}}(\boldsymbol{X}_{\mathrm{src}};c)$。

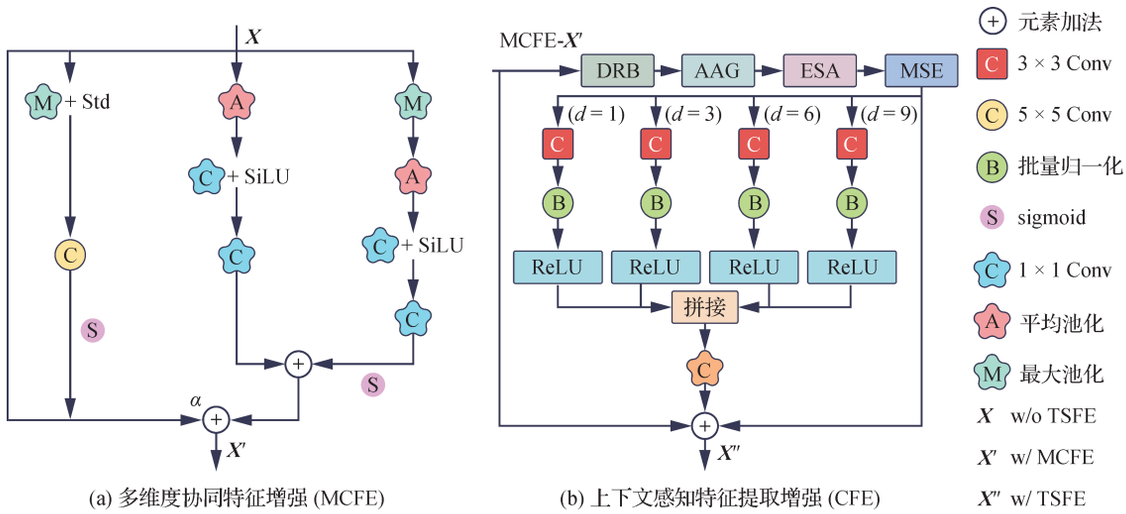
通过训练完成的多标签分类器 $\boldsymbol{F}_{\mathrm{MLC}}$（Long 等，2015），本文可在模型训练过程中，依据目标函数的优化需求动态调整各样本的权重系数，从而针对性缓解类别不平衡问题。

$$\omega=\theta\left(\frac{1}{C'}\sum_{c=1}^{C}\boldsymbol{I}(\hat{\boldsymbol{y}}_{\mathrm{tar}}^{c}>\epsilon)\hat{\boldsymbol{y}}_{\mathrm{tar}}^{c}+1\right)+(1-\theta)\exp\left(1-\frac{N_{\mathrm{tar}}^{c}}{N_{\mathrm{tar}}}\right)$$
$$\text{s.t. } C'=\sum_{c=1}^{C}\boldsymbol{I}(\hat{y}_{\mathrm{tar}}^{c}>\epsilon) \tag{7}$$

式中，θ 为平衡系数，用于调节损失项的权重占比；$\boldsymbol{I}(\cdot)$ 为指示函数；ϵ 为高置信度目标检测的置信阈值；$\hat{\boldsymbol{y}}_{\mathrm{tar}}^{c}$ 表示目标域图像中类别 c 检测概率的估计值，即 $\hat{\boldsymbol{y}}_{\mathrm{tar}}^{c}=\boldsymbol{F}_{\mathrm{MLC}}\left(\boldsymbol{X}_{\mathrm{src}};\ c\right)$；$N_{\mathrm{tar}}$ 代表目标域图像样本总数；N_{tar}^{c} 则是 c 类的图像数量。通过联合最小化局部对抗损失 $L_{\mathrm{DCA}}(\cdot,\cdot)$ 与动态权重 ω（式（5）），能够进一步缩小源域与目标域的特征分布差异，促使两域特征分布趋于一致，提升跨域适应性能。

1.3 两阶段特征增强

为解决低光环境下远处模糊背景和小型目标检测难度大的难题，本文提出一种TSFE模块，其网络架构图如图2所示。该模块通过两大核心创新设计显著提升低光场景的特征提取性能：1）多维度协同特征增强模块（multi-column feature enhancement module，MCFE），通过对空间维度、通道维度及语义维度的分层强化与协调优化，有效提升目标特征的区分度；2）上下文感知特征提取增强（context-aware feature enhancement，CFE），在精准保留目标关键结构信息的前提下，自适应抑制背景噪声干扰，强化有效特征表征。实验结果验证，相较于现有主流方法，所提两阶段特征增强模块在低光场景目标检测任务中表现出显著的性能优势。



(a) 多维度协同特征增强 (MCFE)　　(b) 上下文感知特征提取增强 (CFE)

图2　两阶段特征增强模块

Fig. 2　Two-stage feature enhancement module ((a)MCFE; (b) CFE)

1.3.1　多维度协同特征增强

为解决低光环境下特征表示能力不足的难题，本文提出一种MCFE模块，该模块基于注意力机制的构建，能够协同整合通道间与空间维度的特征的依赖关系。具体而言，通过跨通道相关性建模、空间依赖捕捉以及自适应特征重校准三大核心操作，实现对低光场景全方位特征增强。如图2(a)所示，MCFE模块采用残差方式处理输入特征图 $\boldsymbol{X} \in \mathbf{R}^{C \times H \times W}$，具体流程为

$$\boldsymbol{Y} = \boldsymbol{X} + \alpha_1 \cdot \boldsymbol{F}(\boldsymbol{X}) \tag{8}$$

式中，$\boldsymbol{X}, \boldsymbol{Y} \in \mathbf{R}^{C \times H \times W}$ 分别代表输入和输出的特征图，$\alpha_1 \in [0, 1]$ 是可学习的缩放参数，用于自适应地调整特征的重校准的强度；$\boldsymbol{F}(\boldsymbol{X})$ 代表跨通道相关性建模和空间依赖性捕捉的联合操作，其具体技术实现细节将在下文展开详述。

1）跨通道相关性（cross-channel correlation）。该模块采用双分支并行处理技术，通过全面分析特征图各通道间的内在关联性，精准捕捉对低光目标检测至关重要的关键语义特征，其具体实现方式为

$$\boldsymbol{M}_{\text{channel}}(\boldsymbol{X}) = \sigma\left(\boldsymbol{W}_2 \cdot GELU\left(\boldsymbol{W}_1 \cdot \boldsymbol{Z}\right)\right) \tag{9}$$

式中，$\boldsymbol{Z} \in \mathbf{R}^{C \times 1 \times 1}$ 通过全局平均池化获取；$\boldsymbol{W}_1 \in \mathbf{R}^{\frac{C}{16} \times C}$ 和 $\boldsymbol{W}_2 \in \mathbf{R}^{C \times \frac{C}{16}}$ 为低秩变换矩阵；$\sigma(\cdot)$ 表示sigmoid激活函数；$GELU(\cdot)$ 为高斯误差线性单元（Gaussian error linear unit，GELU）。并行设置的最大池化分支采用相同操作流程。

2）空间依赖模块（spatial dependency）。该模块通过融合通道维度的标准差与最大值两种统计特征，强化低光照区域中目标物体的特征表征，进而提升此类区域的目标检测性能。其数学表达式为

$$\boldsymbol{M}_{\text{spatial}}(\boldsymbol{X}) = \sigma\left(Conv_{5\times5}\left(Concat\left[\boldsymbol{M}_{\max}, \boldsymbol{M}_{\text{std}}\right]\right)\right) \tag{10}$$

式中，$\boldsymbol{M}_{\max}, \boldsymbol{M}_{\text{std}} \in \mathbf{R}^{1 \times H \times W}$ 分别表示通道的最大值和标准差；$Conv_{5\times5}$ 表示一个5×5的卷积层。

3）自适应特征重校准（adaptive feature recalibration）。本文采用残差学习机制，通过可学习参数 α_1 将跨通道相关性建模与空间依赖捕捉的输出特征整合至原始输入特征中，实现对低光场景特征的自适应重校准。其计算式为

$$\boldsymbol{Y}=\boldsymbol{X}+\alpha_1\cdot\boldsymbol{F}(\boldsymbol{X})=\boldsymbol{X}+\alpha_1\cdot\boldsymbol{M}_{\text{channel}}(\boldsymbol{X})\odot\boldsymbol{M}_{\text{spatial}}(\boldsymbol{X})\odot\boldsymbol{X}\tag{11}$$

式中，⊙表示逐元素乘法运算。

总体而言，MCFE模块有效提升了模型在低光环境下目标检测性能，尤其在小型目标与远距离目标的特征提取与识别任务中表现更为突出。

1.3.2 上下文感知特征提取

CFE模块通过多路径并行、多尺度融合和多注意力机制协调设计，实现了对语义特征图显著增强。该架构在保证计算效率的前提下，有效提升了模型对多尺度目标的特征表征能力与检测性能。CFE模块整合了4个核心组件：动态残差块（dynamic residual block，DRB）、自适应注意力门控（adaptive attention gate，AAG）、高效空间注意力（efficient spatial attention，ESA）和多尺度特征增强（multi-scale feature enhancement，MSE）。CFE模块的整体架构及各组成组件的连接关系如图2(b)所示，各组件的网络架构图如图3所示。

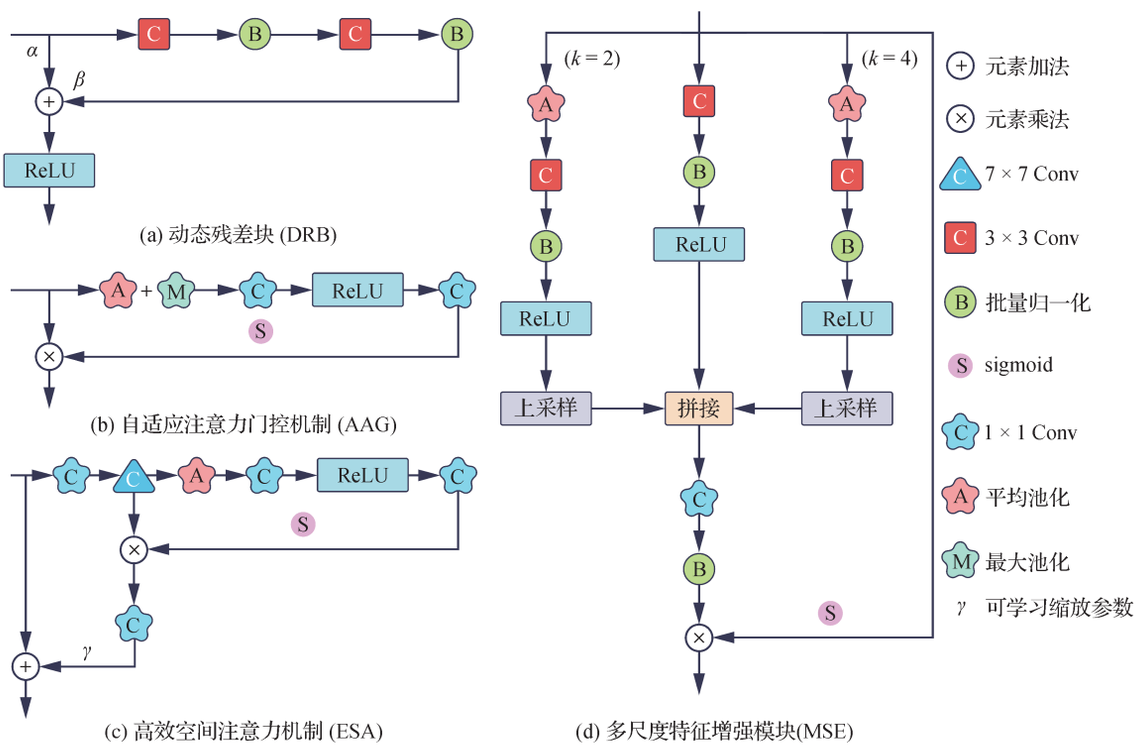


图3 上下文感知特征提取模块

Fig. 3 Context-aware feature enhancement module ((a) dynamic residual block (DRB); (b) adaptive attention gate (AAG); (c) efficient spatial attention (ESA); (d) multi scale feature enhancement (MSE))

1)动态残差块。如图3(a)所示，首先在模型中引入动态残差块，该模块通过双3×3卷积层与可学习的动态残差连接对输入特征图$\boldsymbol{X}$进行特征转换与增强，其具体处理方式为

$$\boldsymbol{Y}=ReLU\left(\alpha\cdot\boldsymbol{X}+\beta\cdot BN\left(Conv_{3\times3}\left(BN\left(Conv_{3\times3}(\boldsymbol{X})\right)\right)\right)\right)\tag{12}$$

式中，$\boldsymbol{Y}$表示输出特征图；α和β是控制残差权重的可训练参数；BN(batch normalization)代表批量归一化。与传统的残差块不同，DRB通过自适应调节特征提取与原始特征保留的权重比例，有效缓解了深度网络训练中的梯度消失问题，同时实现了动态化的特征信息流传递，提升了特征表征的灵活性。

2)自适应注意力门控。如图3(b)所示，AAG模块在结构设计上借鉴常规残差块的核心思想，通过在侧路径中引入平均池化与最大池化并行操作，一方面高效捕捉全局上下文信息，另一方面有效强化特征区域间的空间关联，为低光场景目标特征的精准定位提供支撑。

3)高效空间注意力模块。如图3(c)所示，ESA采用“通道缩减、深度可分离卷积、通道注意力增强”的串联结构：通过1×1卷积层实现通道维度缩减，降低计算复杂度；借助7×7卷积核的深度可分离卷积捕捉局部空间特征；结合通道注意力机制，实现对

长距离空间依赖关系的高效建模，在保证计算效率的同时提升特征表征的针对性。

4）多尺度特征增强。如图3（d）所示，MSE模块基于金字塔分解策略提取多尺度特征：设置两个平均池化分支（空间分辨率分别为输入特征图的1/2和1/4），各分支通过3×3卷积完成局部特征增强；低分辨率分支经双线性上采样恢复至原始特征图尺寸；3个分支通过通道拼接与1×1卷积实现特征交互融合；最后通过sigmoid激活函数生成自适应权重，对原始特征图进行动态增强。该设计有效平衡了局部细节信息与全局语义特征，显著提升了低光场景下特征表征的完整性与判别力。

5）多尺度感受野。CFE模块通过引入孔率分别为{1, 3, 6, 9}的并行扩张卷积支路，实现对多尺度感受野的精准捕捉。将DRB、AAG、ESA及MSE模块输出的中间特征图与原始输入特征图进行融合，通过多特征互补优化，最终输出增强后的特征图，为后续目标检测提供全面的特征支撑。

1.4 检测头

本文采用解耦检测头架构进行目标检测，通过将分类任务和定位任务分离处理，有效缓解了两类任务在特征学习中的冲突问题。

检测网络多尺度特征金字塔网络输出的3个尺度特征图分别为$\{P_3, P_4, P_5\}$，对应的下采样倍率为{8, 16, 32}，对于每个尺度的特征图，检测头采用并行的双分支结构，其中分类分支通过两个连续的3×3卷积层提取分类相关特征，最终输出维度为$H \times W \times C$（C为类别数量）；回归分支同样采用两个3×3卷积层，输出边界框回归参数。本文采用基于分布的边界框回归表示方法，将边界框的连续回归问题转化为离散概率分布的学习问题，具体为

$$\hat{y} = \sum_{i=0}^{n} P(y_i) \cdot y_i \tag{13}$$

式中，$P(y_i)$表示边界框边界位于离散位置y_i的概率，n为离散化的区间数量。

1.5 损失函数

本文采用目标检测任务中的损失函数框架，并结合提出的光照不变特征对齐模块构建完整的损失函数。

检测损失函数（Yaseen，2024）为

$$L_{\mathrm{DET}} = \lambda_{\mathrm{cls}} L_{\mathrm{cls}} + \lambda_{\mathrm{box}} L_{\mathrm{box}} + \lambda_{\mathrm{obj}} L_{\mathrm{obj}} \tag{14}$$

式中，L_{cls}、L_{box}和L_{obj}分别为分类损失、边界框回归损失和目标性损失，λ_{cls}、λ_{box}和λ_{obj}分别为各损失的权重系数。

结合式（2）、式（4）和式（5），完整的训练损失函数定义为

$$L_{\mathrm{total}} = L_{\mathrm{DET}} + L_{\mathrm{MMD}} + L_{\mathrm{Contrast}} + L_{\mathrm{DCA}} \tag{15}$$

2 实验结果

2.1 数据集

COCO2017（common objects context）（Lin 等，2014）是微软公司推出的广泛应用于目标检测任务的基准数据集，包含海量自然场景图像。该数据集涵盖80个常见目标类别，图像均取自真实日常场景，虽然包含多种光照条件，但整体以明亮均匀的光照环境为主。依托多样性的场景和高质量的标注信息，COCO2017成为理想的源域数据集，能够为模型训练提供稳定且可靠的监督信号。然而，该数据集的光照条件分布与低光场景存在显著差异，这为跨域适应任务带来了挑战。在本文实验中，使用该数据集中的86 456幅图像作为源域训练数据。

ExDark（Loh 和 Chan，2019）是一个专为低光环境下的物体检测设计的数据集，包含12类物体：自行车、瓶子、摩托车、猫、狗、公交车、汽车、人物、桌子、船、椅子和杯子。这些图像主要拍摄于夜间或光线昏暗的环境，具有亮度低、噪点高等特征。此外，数据对比度不足。ExDark数据集包含5 963幅训练图像和2 893幅测试图像，这些图像作为目标域样本。由于光照条件与COCO2017存在显著差异，ExDark成为评估模型在低光场景下泛化能力的有效目标域数据集。鉴于ExDark包含的类别数量少于COCO2017，实验中评估了两个数据集共有的12个类别。

2.2 指标

为全面评估所提方法在低光目标检测任务中的性能，本研究采用该领域广泛认可的经典评估指标，精确率（precision，P）、召回率（recall，R）、平均精度（average precision，AP）和平均精度均值（mean average precision，mAP）。这些指标基于主流评估方案构建，可有效量化模型在低光环境下的检测精度与鲁棒性，为所提方法有效性验证提供客观可靠的评价依据。

2.3 实验细节

实验中，源域数据集采用COCO2017，目标域数据集选用 ExDark。模型训练基于单张 24 GB 显存的RTX3090显卡，预训练阶段加载COCO2017 数据集上的预训练权重，批量大小设为 10，总训练迭代次数为100次。实验数据由 ExDark 与COCO2017 数据集的图像共同构成，其中ExDark 数据集按7∶3 的比例划分为训练子集与测试子集。模型初始学习率设定为0.001；针对模块特有的超参数，光照对齐损失的权重系数设为0.3，动态样本优先级的判定阈值设为 0.6。

2.4 对比实验

为验证本文方法性能，分别与以下基线策略进行对比实验：基于CSPDarkNet的方法（S-DAYOLO（Li等，2022b）、SSDA-YOLO（semi-supervised domain adaptative yolo）（Zhou 等，2023a）、SFDA（source-free domain adaptation）（Varailhon 等，2025））；基于 VGG（Visual Geometry Group）的方法（NSA-UDA（network stability analysis for unsupervised domain adaptive detection）（Zhou 等，2023b）和 TDD（He 等，2022））；以及基于ResNet（residual neural network）主干网络的方法（PDA（Hsu 等，2020）、DAOD（domain adaptive object detection）（He 等，2025）、SWDA（Saito 等，2019）、DPA（Zheng 等，2025））。实验通过平均精度均值（mAP）、精确率（P）、召回率（R）和帧率（frames per second，FPS）等指标进行综合评估，实验结果如表1所示。

现有主流跨域目标检测方法仍存在明显局限性：SFDA 因缺乏源域数据锚点，导致细粒度跨域对齐不充分，且存在知识遗忘风险；NSA-UDA 的噪声抑制策略过于盲目，易忽略局部特征差异；早期方法DAOD 存在跨域对齐粒度较粗、动态适应性不足的问题；DPA 的阶段划分依赖主观设定，易产生累积误差，这些缺陷均限制了其跨域特征对齐精度与场景适配能力。本文方法采用IIFA，可有效过滤光照差异带来的干扰，显著提升模型在不同光照条件及跨域场景下的泛化能力与鲁棒性；同时整合TSFE，通过强化远距离与小型目标的显著区域权重，有效缓解背景噪声引发的特征模糊问题，大幅提升弱特征目标的识别性能。实验结果表明，所提方法的P与R均接近最优基准，FPS达到48.41帧/s；如表1所示，其在mAP指标上实现66.83%的最优性能，相较于传统基线方法均具有显著优势。特别地，本文提出的IIFA方法，不仅在mAP指标上超越前述所有域自适应方法，其FPS值更是达到57.03帧/s，位列所有对比方法之首，验证了该领域自适应策略的有效性与优越性。

为进一步验证MCFE与CFE模块的优越性，本研究通过控制变量实验展开分析：在保持其他实验条件一致的前提下，将TSFE网络中的MCFE与CFE

表1 与当前先进域适应方法的对比实验

Table 1 Comparison with the state-of-the-art domain adaptation methods

方法	backbone	mAP/%	精确率/%	召回率/%	FPS/(帧/s)
TDD（He等，2022）	VGG-16	47.28	72.15	47.67	10.74
NSA-UDA（Zhou等，2023b）	VGG-16	55.70	78.36	49.28	8.38
PDA（Hsu等，2020）	ResNet-50	49.73	74.69	47.48	12.62
DAOD（He等，2025）	ResNet-50	64.87	**83.78**	52.16	16.15
SWDA（Saito等，2019）	ResNet-101	52.66	79.47	48.30	11.60
DPA（Zheng等，2025）	ResNet-101	65.19	82.65	54.92	14.83
S-DAYOLO（Li等，2022b）	CSPDarkNet	55.23	77.71	50.40	37.32
SSDA-YOLO（Zhou等，2023a）	CSPDarkNet	54.69	78.50	49.92	34.49
SFDA（Varailhon等，2025）	CSPDarkNet	64.98	81.69	**55.73**	31.81
IIFA	CSPDarkNet	65.74	81.47	53.20	**57.03**
本文	CSPDarkNet	**66.83**	83.73	54.78	48.41

注：加粗字体表示各列最优结果。

模块替换为现有主流注意力机制及多尺度融合机制，对比结果如表2和表3所示。实验表明，当检测网络集成MCFE模块时，能够在保持参数与计算量最小化的同时，实现最高检测精度；而集成CFE模块的检测网络虽然取得最优检测精度，但其参数量（Params）与计算开销FLOPs（floating point operations）均高于SPP（spatial pyramid pooling）（He等，2015）、PPM（pyramid pooling module）（Zhao等，2017）和ASPP（atrous spatial pyramid pooling）（Chen等，2017）等方法。这是因为CFE模块通过对全局与局部上下文特征的统计分析生成精细权重掩膜，并引入结构感知约束机制，在自适应抑制背景噪声的同时精准保留目标结构信息。具体而言，上下文统计过程与权重掩膜生成需依赖多层卷积与注意力操作，而结构感知约束机制也会引入额外计算成本；此外，特征维度匹配所需的过渡层，以及两模块级联时的特征重复增强处理，进一步导致参数与计算量开销的增加。

表4和表5进一步展示了本文方法与先进方法的对比。根据表5的定义，目标尺度按实例在图像

表2　MCFE与现有先进注意力机制的对比结果

Table 2　Comparison results of MCFE with existing advanced attention mechanisms

网络结构	mAP/%	精确率/%	召回率/%	参数量/M	FLOPs/G	FPS/(帧/s)
SE（Hu等，2018）	65.82	81.57	**54.91**	122.39	384.26	**51.28**
CBAM（Woo等，2018）	65.16	82.35	53.91	**122.04**	384.26	48.79
CA（Hou等，2021）	64.39	81.70	54.02	122.09	384.28	43.48
MCFE	**66.83**	**83.73**	54.78	**122.04**	**384.25**	48.81

注：加粗字体表示各列最优结果。

表3　CFE与现有先进多尺度融合机制的对比结果

Table 3　CFE and comparison results with existing advanced multi-scale fusion mechanisms

网络结构	mAP/%	精确率/%	召回率/%	参数量/M	FLOPs/G	FPS/(帧/s)
SPP（He等，2015）	63.73	80.54	52.81	71.94	269.27	**59.56**
PPM（Zhao等，2017）	62.96	79.47	53.69	**69.63**	**261.19**	54.39
ASPP（Chen等，2017）	64.40	82.69	52.12	72.17	269.66	52.69
CFE	**66.83**	**83.73**	**54.78**	122.04	384.25	48.81

注：加粗字体表示各列最优结果。

表4　与前沿方法在各类别上结果对比（AP）

Table 4　Comparison of category results with state-of-the-art methods（AP）

/%

方法	类别											
	Bus	Car	Dog	Person	Table	Cat	Boat	Chair	Cup	Motor	Bike	Bottle
NSA-UDA（Zhou等，2023b）	74.39	61.34	73.84	68.10	11.57	63.49	32.60	53.72	53.35	54.77	62.95	58.32
DAOD（He等，2025）	87.46	76.02	82.32	76.06	19.35	**73.66**	46.39	56.26	59.34	62.50	73.38	65.70
DPA（Zheng等，2025）	88.65	76.98	**83.36**	76.52	21.88	72.58	47.52	54.40	58.49	61.22	74.29	66.41
SFDA（Varailhon等，2025）	87.71	75.95	81.49	76.59	20.76	71.71	46.38	55.86	58.06	63.49	73.81	67.95
IIFA	90.07	76.89	82.87	77.23	22.87	71.44	48.23	57.38	58.14	62.97	72.31	68.42
本文	**91.06**	**77.01**	83.09	**77.68**	**23.53**	72.89	**49.22**	**58.33**	**59.73**	**63.82**	**76.65**	**69.00**

注：加粗字体表示各列最优结果。

中的相对面积划分：小型目标的边界框尺寸小于32 × 32像素；中型目标的边界框尺寸介于32 × 32像素至96 × 96像素之间；大型目标的边界框尺寸大于96 × 96像素。从表4和表5可以看出，除猫和狗两个类别外，本文方法在其他所有类别的AP值均显著优于对比方法，并且在小目标上本文的检测结果明显优于其他算法。

表5　与前沿方法在各目标尺寸上的结果对比（AP）

Table 5　Comparison of object size results with state-of-the-art methods（AP）

/%

方法	目标尺寸		
	小	中	大
NSA-UDA（Zhou等，2023b）	1.27	11.45	31.46
DAOD（He等，2025）	1.49	13.95	35.79
DPA（Zheng等，2025）	1.50	15.07	**42.16**
SFDA（Varailhon等，2025）	1.58	15.13	39.85
IIFA	1.69	15.10	40.78
本文	**2.48**	**16.44**	41.97

注：加粗字体表示各列最优结果。

图4可视化结果直观表明，在不同光照条件下，本文方法的检测精度显著优于现有最先进方法。此外，漏检率和误检率均大幅降低。特别是在极低光照场景（< 1 lux）下，该方法能有效提升远距离目标检测的准确度。分析其核心机制：通过有效提取光照不变特征，并结合两阶段特征增强模块的协同效应，实现了小型目标的精准检测，展现出卓越的鲁棒性。此外，该方案在保持高精度的同时，展现出良好的实时应用潜力，为低光照环境下的视觉感知任务提供了创新解决方案。

2.5　消融研究

为验证所提方法中各模块的有效性，通过一系列消融实验分析其对整体性能的影响。具体而言，通过逐步增加或替换关键模块量化各组件的贡献度，对比不同配置下的模型性能差异，从而明确各组件的独立作用；同时，进一步探索模型结构与超参数的优化空间，为最终方案的合理性提供实验支撑。

2.5.1　跨域特征对齐的有效性

表6和图5分别是本文各模块消融实验和组件模块有效性的实验结果。



(a) 原图　(b) 真值　(c) NSA-UDA　(d) DAOD　(e) DPA　(f) SFDA　(g) 本文

图4　低光照环境下不同方法可视化效果对比

Fig. 4　Comparing the visualization of different methods in low illumination

((a) original images; (b) ground truth; (c) NSA-UDA; (d) DAOD; (e) DPA; (f) SFDA; (g) ours)

与基线方法相比，CDFA模块的mAP从62.07%提升至63.19%。同时，P和R分别达到79.97%和49.00%，FPS达到62.33帧/s。实现了性能与效率的良好平衡。其核心优势源于跨域特征对齐机制与多级优化策略的协同作用：该机制能够系统性抑制光照差异带来的干扰，通过让模型自适应学习不同光照场景的特征分布规律，强化跨域场景下的特征一致性。这一结果充分证明，CDFA模块能显著提

表6　各模块消融实验

Table 6　Module ablation research

网络结构					评价指标			
CDFA	IIFR	DCA	MCFE	CFE	mAP/%	精确率/%	召回率/%	FPS/(帧/s)
–	–	–	–	–	62.07	78.81	47.73	54.91
√	–	–	–	–	63.19	79.97	49.00	62.33
–	√	–	–	–	63.44	80.75	50.05	71.94
–	–	√	–	–	63.58	81.05	49.28	60.98
√	√	–	–	–	64.45	80.36	51.30	**75.03**
√	–	√	–	–	64.64	80.79	51.29	62.31
–	√	√	–	–	64.34	80.79	52.40	62.96
√	√	√	–	–	65.74	81.47	53.20	57.03
√	√	√	√	–	66.16	82.54	53.18	49.25
√	√	√	–	√	65.92	82.03	54.48	42.15
√	√	√	√	√	**66.83**	**83.73**	**54.78**	48.41

注：加粗字体表示各列最优结果，“√”表示采用，“–”表示未采用。

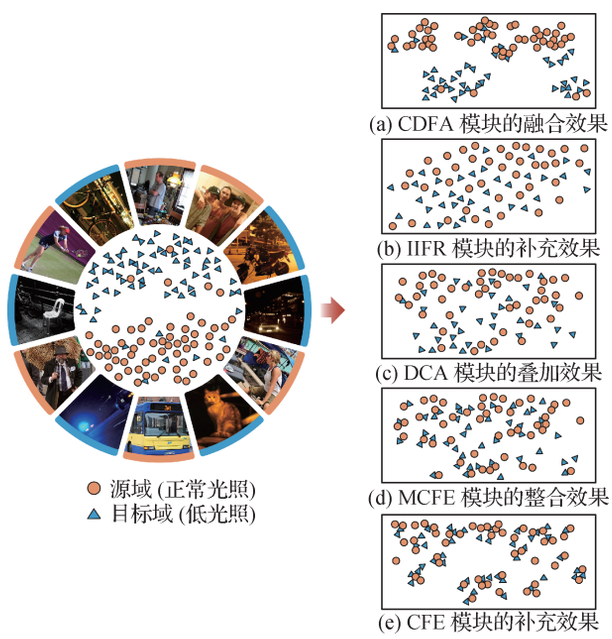


图5　本文框架各组件模块的有效性

Fig. 5　Effects of the components in our framework ((a) integration effect of the CDFA module; (b) supplementary effect of the IIFR module; (c) addition effect of the DCA module; (d) integration effect of the MCFE module; (e) supplementary effect of the CFE module)

升模型在不同光照条件下的鲁棒性，这对提升模型在实际应用中的泛化能力意义重大。如图5(a)所示，引入CDFA后，模型性能得到显著提升。该模型能有效提升对复杂光照环境的适应能力，从而确保在低光环境下仍能保持基础检测性能。

2.5.2　光照不变表征学习的有效性

由表6可知，加入IIFR模块后，模型的mAP从63.19%提升至64.45%，同时P达到80.36%，R保持在51.30%，FPS稳定在75.03帧/s。这一结果充分证明，IIFR模块通过构建光照不变特征空间，有效解耦亮度变化与目标本质特征之间的耦合关系，使模型在特征学习阶段即可精准聚焦目标的内在属性，降低光照干扰对检测性能的影响。如图5(b)所示，IIFR模块的引入显著增强了模型对复杂光照环境的适应能力，同时确保其在低光条件下仍能保持稳定的基础检测性能。

2.5.3　动态类别对齐的有效性

由表6可知，加入DCA模块后，mAP从64.45%提升至65.74%，同时P达到81.47%，R达到53.20%，检测性能持续优化。其核心机制在于，DCA能够根据不同目标类别的特征分布特性，动态调整跨域对齐权重与特征匹配策略：在交叉光照场景中，优先保障同一类别特征的一致性表征，同时自适应抑制类别间的干扰信息。这一设计表明，DCA在强化类别特异性检测性能、提升整体检测效果方面具备显著优势。如图5(c)所示，通过引入类别感知对齐策略，模型的跨类别检测能力得到显著增强，

尤其在光照条件剧烈变化的场景中表现更为突出。

此外，本文对 3 种损失函数 L_{MMD}、L_{Contrast} 和 L_{DCA} 进行了消融实验。如表 7 所示，每种损失函数都发挥着重要作用：MMD 损失通过最小化不同域间的分布差异，有效提升了跨场景特征的统计一致性，为特征对齐奠定了坚实基础；Contrast 损失通过强化相似特征的聚合与差异特征的分散，显著增强了特征的判别能力；而 DCA 损失则专注于类别级别的精细对齐，精准修正了同类特征的偏差。三者从分布对齐、特征区分、类别校准 3 个维度协同优化特征空间，共同推动模型性能提升。如图 6 所示，IIFA 的 3 个核心模块协同优化了检测性能，有效降低了误报率和漏检率：CDFA 能够整体上缓解域偏移，提升特征对齐和整体可靠性；IIFR 保留了区分性特征以缓解光照干扰（减少误报）；DCA 增强了跨域类别一致性（降低罕见/模糊目标的漏检率），显著提升了特征对齐效果和模型可靠性。

2.5.4 多维度协同特征增强的有效性

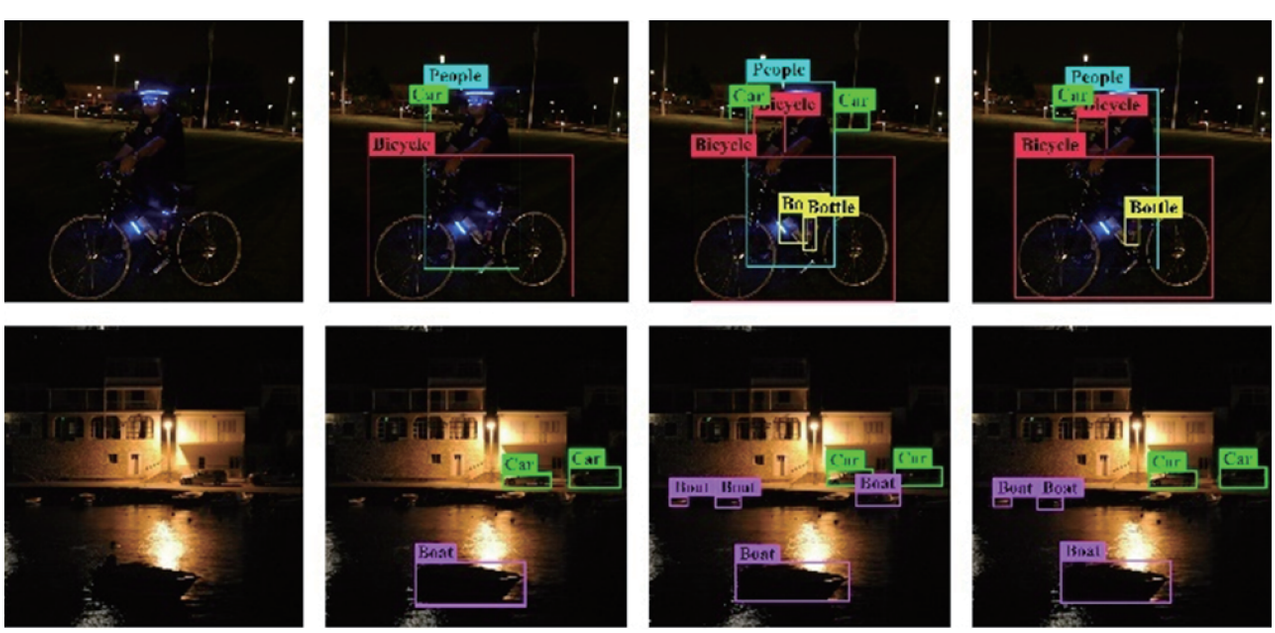
由表 6 可知，加入 MCFE 后，模型的 mAP 从 65.74% 提升至 66.16%，P 达到 82.54%，R 维持在 53.18%。这一性能提升得益于 MCFE 模块构建的通道与空间双向交互网络，其多维协同增强机制具备双重核心作用：通道注意力分支通过自适应权重分配，强化关键语义通道的特征响应，同时抑制冗余通道的噪声干扰；空间注意力分支采用多尺度感受野融合策略，精准捕捉目标区域的空间上下文关联，对易受光照变化影响的边缘细节特征进行针对性增强。图 5(d)呈现了加入 MCFE 模块前后的特征可视化对比结果，直观印证了该模块对特征表征质量的提升效果。如图 7 所示，热图中的红色区域表示感

表 7　损失函数消融研究

Table 7　Ablation study on the loss function

L_{DET}	L_{MMD}	L_{Contrast}	L_{DCA}	mAP/%	精确率/%	召回率/%	FPS/(帧/s)
√	–	–	–	63.04	79.86	49.15	43.73
√	√	–	–	64.09	81.68	51.46	48.10
√	√	√	–	65.38	82.59	53.04	**56.24**
√	√	√	√	**66.83**	**83.73**	**54.78**	48.41

注：加粗字体表示各列最优结果，“√”表示采用，“–”表示未采用。



(a) 原图　(b) 没有 IIFA 模块　(c) 有 IIFA 模块　(d) 真值

图 6　加入 IIFA 前后的结果可视化，IIFA 包含 CDFA、IIFR 和 DCA

Fig. 6　Visualization of the results with and without IIFA that includes CDFA, IIFR and DCA

((a) original images; (b) without IIFA module; (c) with IIFA module; (d) ground truth)

兴趣区域(目标区域)。在没有MCFE的情况下,某些图像中的红色区域部分覆盖了背景区域;同时,小而远的目标显示出高漏检率和误报率,并且物体之间的边界不清晰。相比之下,在整合MCFE之后,红色区域更集中在目标物体上,提高了小而远目标的检测率,并使物体之间的边界更加清晰。此外,红色区域覆盖了更多的目标物体,范围更广。显然,整合MCFE使网络能够更精确地定位感兴趣的目标。这些结果共同表明,多维协同增强机制能够全面捕捉特征的通道相关性和空间依赖性,从而显著提高了网络的表征能力,并为优化目标检测任务中的性能提供了有效解决方案。

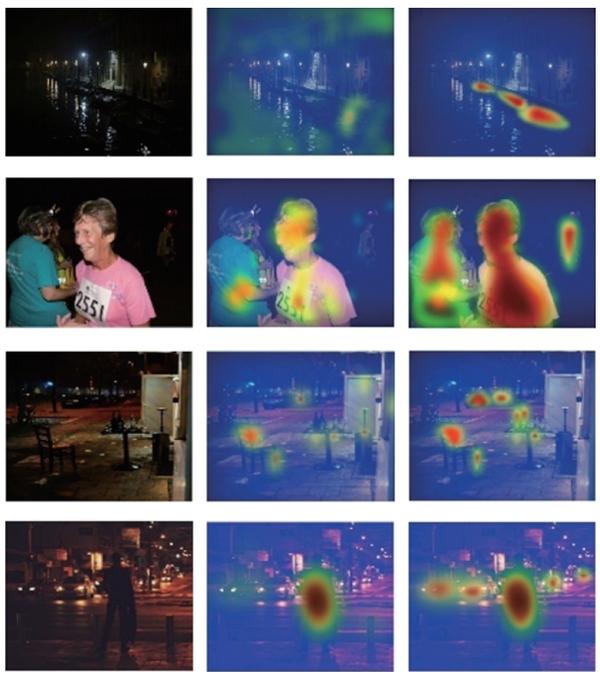
2.5.5 上下文感知特征增强的有效性

由表6可知,加入CFE后,其mAP从66.16%提升至66.83%。模型P达到83.73%,R提升至54.78%。实现了检测性能的最优突破。这一提升的核心原因在于,CFE模块能够在低光环境下显著提升目标特征的信噪比,同时强化特征的判别能力与鲁棒性,使模型更高效地感知并表征易被噪声淹没的小型、远距离目标特征。此外,该模块通过增强特征的空间一致性与上下文关联性,进一步优化了多目标场景中特征的可区分性,有效减少了目标间的相互干扰。如图5(e)所示,可视化结果直观印证了模型在低光条件下的多目标检测能力显著提升,尤其优化了小型及远距离目标的检测性能,这也使得该配置成为应对复杂低光环境目标检测的最优方案。

表8实验结果表明,MCFE与CFE模块在小型、远距离目标检测任务中表现突出,针对小型、远距离目标的检测难题,本文提出TSFE双阶段特征增强模块。如图8可视化结果所示,该模块在提升此类目标检测精度方面成效显著,为低光场景下多尺度目标检测提供了高效解决方案。

3 结 论

在低光照环境下,传统目标检测方法常因噪声干扰、对比度不足和细节丢失等问题导致性能显著下降,而现有相关方法又受限于低光照场景高质量标注数据的稀缺性。本文提出的检测框架通过充分利用常规光照条件下的预训练资源,有效破解了这一难题:特征对齐模块通过全局光照对齐、光照无关特征学习和自适应优先级处理,有效缓解跨域差异;同时,基于注意力机制的双阶段增强模块通过强化小尺寸远距离物体的局部特征,弥补低光环境下的对比度衰减与细节损失。大量实验验证了其前沿性能,尤其在小与远目标检测方面,平均精度均值达到66.83%,精确率达到83.73%,召回率达到54.78%,最大帧率高达75.03帧/s,综合性能显著优于现有算法。然而,该方法仍存在一定局限性:1)对非光照相关的跨域变化适应能力较弱,易受此类领域差异干扰,进而影响特征表征的一致性;2)双阶段特征增强



(a) 原图 (b) 没有MCFE模块 (c) 有MCFE模块

图7 添加MCFE前后热力图

Fig. 7 Heat map before and after adding MCFE ((a) original images; (b) without MCFE module; (c) with MCFE module)

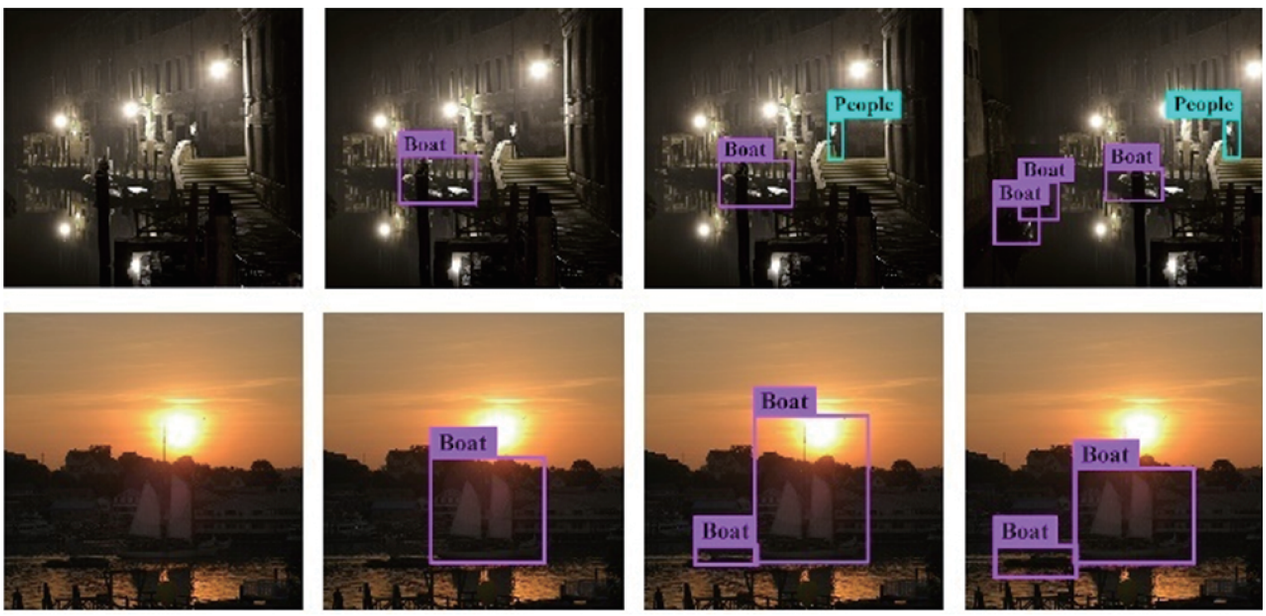
表8 MCFE和CFE检测小型、中型和大型物体的AP效果

Table 8 Detection of AP effects by MCFE and CFE for small, medium, and large objects

/%

网络结构		目标尺寸		
MCFE	CFE	小	中	大
√	–	1.85	15.36	41.36
–	√	1.77	15.56	41.78
√	√	**2.48**	**16.44**	**41.97**

注:加粗字体表示各列最优结果,"√"表示采用,"–"表示未采用。



(a) 原图　(b) 没有 TSFE 模块　(c) 有 TSFE 模块　(d) 真值

图 8　加入TSFE前后的结果可视化(TSFE包含MCFE和CFE)

Fig. 8　Visualization of the results with and without TSFE that includes MCFE and CFE

((a) original images; (b) without TSFE module; (c) with TSFE module; (d) ground truth)

模块虽然能强化目标显著区域,但过度依赖初始定位的精准性,且易受低光环境下复杂域变化的影响,导致对模糊目标的特征增强效果不理想。

参考文献(References)

Badue C, Guidolini R, Carneiro R V, Azevedo P, Cardoso V B, Forechi A, et al. 2021. Self-driving cars: a survey. Expert Systems with Applications, 165: #113816 [DOI: 10.1016/j.eswa.2020.113816]

Cai Y H, Bian H, Lin J, Wang H Q, Timofte R and Zhang Y L. 2023. Retinexformer: one-stage retinex-based transformer for low-light image enhancement//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 12470-12479 [DOI: 10.1109/ICCV51070.2023.01149]

Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A and Zagoruyko S. 2020. End-to-end object detection with transformers//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 213-229 [DOI: 10.1007/978-3-030-58452-8_13]

Chen L C, Papandreou G, Schroff F and Adam H. 2017. Rethinking atrous convolution for semantic image segmentation [EB/OL]. [2025-11-21]. https://arxiv.org/pdf/1706.05587.pdf

Chen T S, Xu M X, Hui X L, Wu H F and Lin L. 2019. Learning semantic-specific graph representation for multi-label image recognition//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 522-531 [DOI: 10.1109/ICCV.2019.00061]

Cui L Q, Yu C H, Jiang S, Sun X Y, Peng R L, Lundgren J E, et al. 2022. A new approach for determining GND and SSD densities based on indentation size effect: an application to additive-manufactured Hastelloy X. Journal of Materials Science and Technology, 96: 295-307 [DOI: 10.1016/j.jmst.2021.05.005]

Dai H, Zhang B, Zhao H, Su Y Q, Zhang X M, Li B P, et al. 2024. An algorithm for enhancing low-light remote-sensing images based on zero-reference deep curve estimation (Zero-DCE)//Proceedings of SPIE 13239, Optoelectronic Imaging and Multimedia Technology XI. Nantong, China: SPIE: 75-84 [DOI: 10.1117/12.3036367]

Dai J F, Li Y, He K M and Sun J. 2016. R-FCN: object detection via region-based fully convolutional networks//Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain: Curran Associates Inc.: 379-387

Dong J P, Lin Y T, Li C, Zhou S P and Zheng N N. 2025. AugDETR: improving multi-scale learning for detection transformer//Proceedings of the 18th European Conference on Computer Vision. Milan, Italy: Springer: 238-255 [DOI: 10.1007/978-3-031-72691-0_14]

Dong X L, Gao S M, Tian J J and Yao Y. 2017. Multipoint fiber-optic laser - ultrasound generation along a fiber based on the core-offset splicing of fibers. Photonics Research, 5(4): 287-292 [DOI: 10.1364/PRJ.5.000287]

Du Z P, Shit M and Deng J K. 2024. Boosting object detection with zero-shot day-night domain adaptation//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 12666-12676 [DOI: 10.1109/CVPR52733.2024.01204]

Guo C L, Li C Y, Guo J C, Loy C C, Hou J H, Kwong S, et al. 2020.

Zero-reference deep curve estimation for low-light image enhancement//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 1777-1786 [DOI: 10.1109/CVPR42600.2020.00185]

Hashmi K A, Kallempudi G, Stricker D and Afzal M Z. 2023. FeatEnHancer: enhancing hierarchical features for object detection and beyond under low-light vision//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 6702-6712 [DOI: 10.1109/ICCV51070.2023.00619]

He K M, Zhang X Y, Ren S Q and Sun J. 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(9): 1904-1916 [DOI: 10.1109/TPAMI.2015.2389824]

He M Z, Wang Y L, Wu J X, Wang Y R, Li H Q and Li B. 2022. Cross domain object detection by target-perceived dual branch distillation//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 9560-9570 [DOI: 10.1109/CVPR52688.2022.00935]

He W and Pan C. 2022. The salient object detection based on attention-guided network. Journal of Image and Graphics, 27(4): 1176-1190 (何伟, 潘晨. 2022. 注意力引导网络的显著性目标检测. 中国图象图形学报, 27(4): 1176-1190) [DOI: 10.11834/jig.200658]

He X Y, Li X H and Guo X J. 2025. Differential alignment for domain adaptive object detection//Proceedings of the 39th AAAI Conference on Artificial Intelligence. Philadelphia, USA: AAAI: 17150-17158 [DOI: 10.1609/aaai.v39i16.33885]

Hou Q B, Zhou D Q and Feng J S. 2021. Coordinate attention for efficient mobile network design//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 13708-13717 [DOI: 10.1109/CVPR46437.2021.01350]

Hsu H K, Yao C H, Tsai Y H, Hung W C, Tseng H Y and Singh M. 2020. Progressive domain adaptation for object detection//Proceedings of 2020 IEEE Winter Conference on Applications of Computer Vision. Snowmass, USA: IEEE: 738-746 [DOI: 10.1109/WACV45572.2020.9093358]

Hu J, Shen L and Sun G. 2018. Squeeze-and-excitation networks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 7132-7141 [DOI: 10.1109/CVPR.2018.00745]

Huang Q, Pan Q and Tian N L. 2023. A position-weighted feature pyramid network for object detection. Computer Simulation, 40(6): 192-196 (黄强, 潘晴, 田妮莉. 2023. 一种用于目标检测的位置加权特征金字塔网络. 计算机仿真, 40(6): 192-196) [DOI: 10.3969/j.issn.1006-9348.2023.06.034]

Jia K X, Ma Z H, Zhu R and Li Y G. 2022. Attention-mechanism-based light single shot multiBox detector modelling improvement for small object detection on the sea surface. Journal of Image and Graphics, 27(4): 1161-1175 (贾可心, 马正华, 朱蓉, 李永刚. 2022. 注意力机制改进轻量SSD模型的海面小目标检测. 中国图象图形学报, 27(4): 1161-1175) [DOI: 10.11834/jig.200517]

Kennerley M, Wang J G, Veeravalli B and Tan R T. 2023. 2PCNet: two-phase consistency training for day-to-night unsupervised domain adaptive object detection//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 11484-11493 [DOI: 10.1109/CVPR52729.2023.01105]

Land E H and McCann J J. 1971. Lightness and retinex theory. Journal of the Optical Society of America, 61(1): 1-11 [DOI: 10.1364/JOSA.61.000001]

Li C Y, Guo C L, Han L H, Jiang J, Cheng M M, Gu J W, et al. 2022a. Low-light image and video enhancement using deep learning: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(12): 9396-9416 [DOI: 10.1109/TPAMI.2021.3126387]

Li F, Zeng A L, Liu S L, Zhang H, Li H Y, Zhang L, et al. 2023. Lite DETR: an interleaved multi-scale encoder for efficient DETR//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 18558-18567 [DOI: 10.1109/CVPR52729.2023.01780]

Li G F, Ji Z F, Qu X D, Zhou R and Cao D P. 2022b. Cross-domain object detection for autonomous driving: a stepwise domain adaptative YOLO approach. IEEE Transactions on Intelligent Vehicles, 7(3): 603-615 [DOI: 10.1109/TIV.2022.3165353]

Li H H, Zeng X S, Li Y Z, Zhou S D and Wang J. 2019. Convolutional neural networks based indoor Wi-Fi localization with a novel kind of CSI images. China Communications, 16(9): 250-260 [DOI: 10.23919/JCC.2019.09.019]

Li M Z, Chen Y L, Zhang T and Huang W. 2024. TA-YOLO: a lightweight small object detection model based on multi-dimensional trans-attention module for remote sensing images. Complex and Intelligent Systems, 10(4): 5459-5473 [DOI: 10.1007/s40747-024-01448-6]

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

Liu A Q, Zhang J F and Xun Y L. 2013. Outlier iteration algorithm based on large entropy vary and cosine similarity. Journal of Chinese Computer Systems, 34(7): 1518-1521 (刘爱琴, 张继福, 荀亚玲. 2013. 基于大熵值变化区域和余弦相似度的离群迭代算法. 小型微型计算机系统, 34(7): 1518-1521) [DOI: 10.3969/j.issn.1000-1220.2013.07.012]

Liu S L, Li F, Zhang H, Yang X, Qi X B, Su H, et al. 2022. DAB-DETR: dynamic anchor boxes are better queries for DETR [EB/OL]. [2025-11-21]. https://arxiv.org/pdf/2201.12329.pdf

Loh Y P and Chan C S. 2019. Getting to know low-light images with the exclusively dark dataset. Computer Vision and Image Understanding, 178: 30-42 [DOI: 10.1016/j.cviu.2018.10.010]

Long M S, Cao Y, Wang J M and Jordan M I. 2015. Learning transferable features with deep adaptation networks//Proceedings of the 32nd International Conference on Machine Learning. Lille, France: JMLR.org

Luo Q, Wu J F, Yu H, Sun J W and Zhang M L. 2018. Saliency detection algorithm based on improved histogram equalization. Journal of Chinese Computer Systems, 39(5): 1092-1096 (罗强, 吴俊峰, 于红, 孙建伟, 张美玲. 2018. 一种基于改进直方图均衡化的显著图提取算法. 小型微型计算机系统, 39(5): 1092-1096) [DOI: 10.3969/j.issn.1000-1220.2018.05.040]

Ma L, Ma T Y and Liu R S. 2022. The review of low-light image enhancement. Journal of Image and Graphics, 27(5): 1392-1409 (马龙, 马腾宇, 刘日升. 2022. 低光照图像增强算法综述. 中国图象图形学报, 27(5): 1392-1409) [DOI: 10.11834/jig.210852]

Meng D P, Chen X K, Fan Z J, Zeng G, Li H Q, Yuan Y H, et al. 2021. Conditional DETR for fast training convergence//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 3631-3640 [DOI: 10.1109/ICCV48922.2021.00363]

Mohamadi H, Keyvanrad M A and Mohammadi M R. 2024. Feature based methods in domain adaptation for object detection: a review paper [EB/OL]. [2025-11-21]. https://arxiv.org/pdf/2412.17325.pdf

Morawski I, Chen Y A, Lin Y S and Hsu W H. 2021. NOD: taking a closer look at detection under extreme low-light conditions with night object detection dataset//Proceedings of the 32nd British Machine Vision Conference. [s.l.]: BMVA

Neumann L, Karg M, Zhang S S, Scharfenberger C, Piegert E, Mistr S, et al. 2019. NightOwls: a pedestrians at night dataset//Proceedings of the 14th Asian Conference on Computer Vision. Perth, Australia: Springer: 691-705 [DOI: 10.1007/978-3-030-20887-5_43]

Pan J W, Liu X M, Bai Y C, Zhai D M, Jiang J J and Zhao D B. 2024. Illumination-aware low-light image enhancement with transformer and auto-knee curve. ACM Transactions on Multimedia Computing, Communications and Applications, 20(8): #254 [DOI: 10.1145/3664653]

Pan X Y, Jia N X, Mu Y Z and Gao X R. 2023. Survey of small object detection. Journal of Image and Graphics, 28(9): 2587-2615 (潘晓英, 贾凝心, 穆元震, 高炫蓉. 2023. 小目标检测研究综述. 中国图象图形学报, 28(9): 2587-2615) [DOI: 10.11834/jig.220455]

Park H, Ju M, Moon S and Yoo C D. 2019. Unsupervised domain adaptation for object detection using distribution matching in various feature level//Proceedings of the 17th International Workshop on Digital Watermarking. Jeju Island, Korea(South): Springer: 363-372 [DOI: 10.1007/978-3-030-11389-6_27]

Redmon J and Farhadi A. 2018. YOLOv3: an incremental improvement [EB/OL]. [2018-04-08]. https://arxiv.org/pdf/1804.02767.pdf

Ren S Q, He K M, Girshick R and Sun J. 2017. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6): 1137-1149 [DOI: 10.1109/TPAMI.2016.2577031]

Saito K, Ushiku Y, Harada T and Saenko K. 2019. Strong-weak distribution alignment for adaptive object detection//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 6949-6958 [DOI: 10.1109/CVPR.2019.00712]

Schönfeld E, Schiele B and Khoreva A. 2020. A U-Net based discriminator for generative adversarial networks//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 8204-8213 [DOI: 10.1109/CVPR42600.2020.00823]

Shen X Y, Li H B, Li Y Q and Zhang W M. 2025. LDWLE: self-supervised driven low-light object detection framework. Complex and Intelligent Systems, 11(1): #82 [DOI: 10.1007/s40747-024-01681-z]

Varailhon S, Aminbeidokhti M, Pedersoli M and Granger E. 2025. Source-free domain adaptation for YOLO object detection//Proceedings of 2025 European Conference on Computer Vision. Milan, Italy: Springer: 218-235 [DOI: 10.1007/978-3-031-91672-4_14]

Wang F and Liu H P. 2021. Understanding the behaviour of contrastive loss//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 2495-2504 [DOI: 10.1109/CVPR46437.2021.00252]

Wang W J, Wang X H, Yang W H and Liu J Y. 2023. Unsupervised face detection in the dark. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(1): 1250-1266 [DOI: 10.1109/TPAMI.2022.3152562]

Woo S, Park J, Lee J Y and Kweon I S. 2018. CBAM: convolutional block attention module//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 3-19 [DOI: 10.1007/978-3-030-01234-2_1]

Xiao N and Zhang L. 2021. Dynamic weighted learning for unsupervised domain adaptation//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 15237-15246 [DOI: 10.1109/CVPR46437.2021.01499]

Xu R, Niu Y Z, Li Y Z, Xu H B, Liu W X and Chen Y Z. 2025a. URWKV: unified RWKV model with multi-state perspective for low-light image restoration//Proceedings of 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 21267-21276 [DOI: 10.1109/CVPR52734.2025.01981]

Xu W, Wan Y and Zhao W N. 2025b. ELA: efficient location attention for deep convolution neural networks. Journal of Real-Time Image Processing, 22(4): #140 [DOI: 10.1007/s11554-025-01719-6]

Xu X, Wang S Q, Wang Z, Zhang X L and Hu R M. 2021. Exploring image enhancement for salient object detection in low light images. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 17(1s): #8 [DOI: 10.1145/3414839]

Yan W J, Wang Y Y, Gu S J, Huang L, Yan F H, Xia L M, et al. 2019. The domain shift problem of medical image segmentation and vendor-adaptation by Unet-GAN//Proceedings of the 22nd International Conference on Medical Image Computing and Computer-Assisted Intervention. Shenzhen, China: Springer: 623-631 [DOI: 10.1007/978-3-030-32245-8_69]

Yaseen M. 2024. What is YOLOv8: an in-depth exploration of the internal features of the next-generation object detector [EB/OL]. [2025-11-21]. https://arxiv.org/pdf/2408.15857.pdf

Zhang B, Guo Y C, Yang R Z, Zhang Z H, Xie J Y, Suo J L, et al. 2023. DarkVision: a benchmark for low-light image/video perception [EB/OL]. [2025-11-21]. https://arxiv.org/pdf/2301.06269.pdf

Zhang Q, Miao Z, Wang J B, Ji W Y, Bi X H and Lei X Z. 2026. Dual-branch feature enhancement and calibration network for low-light images. Journal of Image and Graphics, 31(8): 2882-2896 (张奇, 苗壮, 王家宝, 纪文宇, 毕翔鹤, 雷小舟. 2026. 低光照图像的双分支特征增强与校准网络. 中国图象图形学报, 31(8): 2882-2896 [DOI: 10.11834/jig.250530]

Zhang R K, Guo L Q, Huang S Y and Wen B H. 2021. ReLLIE: deep reinforcement learning for customized low-light image enhancement//Proceedings of the 29th ACM International Conference on Multimedia. [s. l.]: ACM: 2429-2437 [DOI: 10.1145/3474085.3475410]

Zhang Y, Zhang Y Q, Zhang Z A, Zhang M, Tian R and Ding M L. 2024. ISP-Teacher: image signal process with disentanglement regularization for unsupervised domain adaptive dark object detection//Proceedings of the 38th AAAI Conference on Artificial Intelligence. Vancouver, Canada: AAAI: 7387-7395 [DOI: 10.1609/aaai.v38i7.28569]

Zhao H S, Shi J P, Qi X J, Wang X G and Jia J Y. 2017. Pyramid scene parsing network//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 6230-6239 [DOI: 10.1109/CVPR.2017.660]

Zhao M H, Wen Y C, Du S L, Hu J, Shi C and Li P. 2024. Low-light image enhancement algorithm based on illumination and scene texture attention map. Journal of Image and Graphics, 29(4): 862-874 (赵明华, 汶怡春, 都双丽, 胡静, 石程, 李鹏. 2024. 基于照度与场景纹理注意力图的低光图像增强. 中国图象图形学报, 29(4): 862-874) [DOI: 10.11834/jig.230271]

Zheng Y F, Wu J L, Li W Y and Chen Z. 2025. Universal domain adaptive object detection via dual probabilistic alignment//Proceedings of the 39th AAAI Conference on Artificial Intelligence. Philadelphia, USA: AAAI: 10644-10652 [DOI: 10.1609/aaai.v39i10.33156]

Zhou H Y, Jiang F and Lu H T. 2023a. SSDA-YOLO: semi-supervised domain adaptive YOLO for cross-domain object detection. Computer Vision and Image Understanding, 229: #103649 [DOI: 10.1016/j.cviu.2023.103649]

Zhou W Z, Fan H, Luo T J and Zhang L B. 2023b. Unsupervised domain adaptive detection with network stability analysis//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 6963-6972 [DOI: 10.1109/ICCV51070.2023.00643]

作者简介

王洋,女,硕士研究生,主要研究方向为计算机视觉低光目标检测。E-mail:2937331762@qq.com

张美璟,通信作者,男,副教授,主要研究方向为人工智能、警务大数据分析与挖掘。E-mail: zhangmeijing@fjpsc.edu.cn

曹菁菁,女,副教授,主要研究方向为图像目标检测与语义分割。E-mail: bettycao@whut.edu.cn

程红举,男,教授,主要研究方向为机器学习、多模态、数字孪生和区块链。E-mail: cscheng@fzu.edu.cn

薛醒思,男,教授,主要研究方向为自然语言处理、异构大数据集成、知识网谱和语义网。E-mail: jack8375@gmail.com

刘文犀,男,教授,主要研究方向为计算机视觉和人工智能领域。E-mail: wenxi.liu@hotmail.com