

图像增强与特征自适应联合学习的低光图像目标检测方法

乔成平^{①②} 金佳翌^{①②} 张俊超^{*①②} 朱政亮^③ 曹祥旭^①

^①(中南大学自动化学院 长沙 410083)

^②(光电智能测控湖南省重点实验室 长沙 410083)

^③(集美大学海洋信息工程学院 厦门 361021)

摘要: 针对低光照图像目标特征弱、检测精度不足等问题, 该文提出了一种基于图像增强与特征自适应联合学习的目标检测模型, 该模型采用串联结构, 将有监督的图像增强模块与YOLOv5目标检测模块相结合, 以端到端的方式实现低光照图像目标检测。首先, 利用正常光数据集生成匹配的正常光与低光图像对, 实现数据集增强, 并据此指导图像增强模块的学习; 其次, 联合图像增强损失、特征匹配损失和目标检测损失, 从像素级和特征级两个层面优化目标检测结果; 最后, 基于真实低光照数据集进行模型参数的优化和微调。实验结果表明, 该方法在仅使用真实正常光数据集训练的情况下, 在LLVIP和Polar3000低光照数据集上的检测精度分别达到79.5%和85.7%, 进一步在真实低光照数据集上微调后, 检测精度分别提升至91.7%和92.3%, 显著优于主流的低光照图像目标检测方法, 并在ExDark和DarkFace的泛化实验中取得最佳检测效果。此外, 该方法在提升检测精度的同时, 仅带来2.5%的参数增加, 具有良好的实时检测性能。

关键词: 目标检测; 低照度图像; 特征匹配; 图像增强网络

中图分类号: TN9211.7; TN247; TP391

文献标识码: A

文章编号: 1009-5896(2025)10-3929-12

DOI: 10.11999/JEIT250302

CSTR: 32379.14.JEIT250302

1 引言

低光环境下的目标检测是计算机视觉领域的重要研究方向, 广泛应用于自动驾驶、安防监控和医学影像等领域。在正常光照条件下, 目标检测技术已经日渐成熟, 出现了一系列同时具备精度和速度的深度学习算法。然而, 在低光照条件下, 图像存在亮度低、对比度弱和受噪声干扰严重等问题, 导致目标物体难以清晰呈现, 出现边缘模糊、颜色信息缺失等问题, 同时收集和标注低光照数据集既困难又昂贵, 这些都是低光照目标检测技术面临的挑战。

针对低光照图像的目标检测, 最朴素的思想是直接进行图像增强。近年来, 基于深度学习的增强方法(如RetinexNet^[1], Zero-DCE^[2])在低光照图像增强领域取得了显著进展, 可以有效提升图像亮度和恢复图像颜色。Luo等人^[3]在2025年提出的无监

督低光图像增强方法ULIE-SPL利用自适应学习策略, 通过逐步引入困难样本来改善网络训练, 使得在缺乏成对数据监督的情况下, 能够实现亮度提升和细节恢复, Wang等人^[4]设计的RFFNet通过多尺度不一致性校准和双域自适应融合模块, 有效抑制了噪声, 同时保护了图像纹理信息。然而, 直接将增强图像用于检测模型常因优化目标不一致而影响性能, 图像增强的主要目标是提升低光照图像的整体质量, 而目标检测更关注精准识别和定位目标。这两个方向在优化重点上存在差异, 因此通过联合学习的方法实现图像增强与目标检测的协同^[5,6]。

为解决“先增强后检测”方法中存在的任务不匹配问题, 研究者提出了图像增强与目标检测的端到端模型。Qin等人^[7]提出的DEYOLO和Yin等人^[8]提出的PEYOLO采用了拉普拉斯金字塔结构对图像高低频增强, 并与YOLOv3^[9]模型串联训练。然而, 这类方法存在训练时间长、增强中间结果缺乏损失约束等问题, 效率低且不稳定。针对上述问题, 本文提出一种图像增强与特征自适应联合学习的低光图像目标检测方法, 该方法包括一个图像增强子网络, 与YOLOv5以端到端的方式训练。其中, 图像增强模块为有监督的轻型模块, 负责将输入的低光照图像恢复到正常照度, 特征自适应损失用于约束图像增强后的特征部分, 引导优化过程朝着有利于提升目标检测精度的方向进行。本文的工作和创新总结如下:

收稿日期: 2025-04-25; 改回日期: 2025-08-20; 网络出版: 2025-08-28

*通信作者: 张俊超 junchaozhang@csu.edu.cn

基金项目: 国家自然科学基金(62105372), 自动目标识别全国重点实验室基础研究基金(WDZC20255290209), 煤炭智能开采与岩层控制全国重点实验室开放基金(SKLIS202404)

Foundation Items: The National Natural Science Foundation of China (62105372), The Fundamental Research Foundation of National Key Laboratory of Automatic Target Recognition (WDZC 20255290209), The Open Funding of State Key Laboratory of Intelligent Coal Mining and Strata Control (SKLIS202404)

(1)设计一种图像数据增强模型,将正常光照图像用数据增强模块合成低光照图像,并将正常光图像和低光图像进行配对,实现数据集增强,并据此指导图像增强模块的学习。

(2)提出一种图像增强与特征自适应联合学习的低光图像目标检测方法,先对低光照图像用有监督的图像增强模块进行像素级增强,再对增强后的结果在特征提取阶段用匹配损失约束实现特征级自适应匹配,以实现低光图像目标检测准确率的提升^[10]。

(3)以端到端的方式将图像增强模块与目标检测模块进行级联训练,显著提升方法的泛化性能。提出的方法在LLVIP^[11]和Polar3000^[12]数据集上进行系统性的实验,获得了较先进的SOTA(State-Of-The-Art)结果,具有更优的检测性能,并在ExDark和DarkFace的泛化实验中取得最佳检测效果。

2 相关工作

本文所提的低光目标检测方法是一种低光照图像增强和目标检测级联的方法,本质上是由一个有监督的图像增强模块和目标检测模块级联构成的端到端网络,因此这里介绍低光照图像增强和目标检测的相关工作,并回顾一些为本文提供灵感的算法。

2.1 图像增强

低光照图像增强方法分为传统方法和深度学习方法。传统方法如直方图均衡化和Retinex^[13]理论,前者通过调整图像灰度来提升对比度,但可能导致噪声的放大。Retinex理论将图像分解为反射分量和光照分量以实现增强,其中光照分量用于增强光照强度,反射分量则用于去噪和增强梯度信息,但与目标检测模块级联会大幅增加模型参数量,不利于实时处理。Guo等人^[2]提出Zero-DCE算法通过精确的像素曲线估计来增强图像的照度,该模型较小且无监督,具有很强的实时性。但由于其无监督的特点,增强后的图像具有不可控性,不适合与目标检测模型级联。Wang等人^[5]提出的Deep-UPE则设计了一个网络,用于估计曝光不足图像到光照图的映射,并引入了一个包含重建损失、平滑损失和颜色损失的全新损失函数,以此来约束增强结果与真实正常光图像之间的差异。

2.2 低光照目标检测

近几年,低照度场景下的目标检测受到关注。Loh等人^[14]研究低照度条件对目标检测任务的影响,并创建了ExDark的低照度数据集。Qin等人^[7]基于拉普拉斯特征金字塔提出DEYOLO,通过注意力机制和高低频增强实现图像的增强,Yin等人^[8]通过全局和边缘增强改进拉普拉斯金字塔,提出了

PEYOLO。另外,基于注意力机制的即插即用模块FeatEnhancer^[15]被用于目标检测,与YOLOv5^[6]结合应用于低光检测为FEYOLO。Cui等人^[16]提出了一种多任务自动编码转换的MAET模型,以自监督的方式对图像照度退化转换进行编码和解码,从而学习到图像内在的视觉结构。Liu等人^[10]的IA-YOLO引入可微图像处理模块DIP,实现图像自适应增强。DAI-Net^[17]借助MAET中的低光合成模块,结合Retinex模型,获取光照不变的特征,构建了最新的低光目标检测模型。当前的低光目标检测方法主要集中在增强输入图像或提取域不变信息上,但这些方法未能有效结合图像增强与特征学习。为此,本文提出一种级联方法,可同时对图像进行像素级增强和特征自适应学习。

3 文中算法

3.1 算法框架

本文采用图像增强串联目标检测的端到端方法,通过像素级图像增强和特征匹配损失联合优化,提升低光照图像的检测性能。

如图1所示,整个网络以串联的图像增强模块和目标检测模块为基础,是一个端到端的模型。其中的目标检测模块以YOLOv5为基本架构,包括Backbone, Neck和Head 3个部分。Backbone为特征提取模块,用于提取不同尺度的特征; Neck为特征融合模块,具体为FPN(Feature Pyramid Network)+PAN(Path Aggregation Network)结构,FPN主要通过自顶向下特征融合来提升特征层次之间的信息共享能力,PAN在FPN基础上增加自底向上的路径,进一步增强了全局和局部特征的表达能力; Head为检测头,负责输出最终的检测结果。图1底部左右子图分别表示数据增强流程和网络子模块细节。数据增强过程为N张正常光图像通过暗光合成通道Dark ISP生成低光图像,正常光图像与对应低光图像MixUp后得到低光训练集,正常光图像与自身MixUp后得到正常光训练集。

受文献[5]启发,本文的图像增强模块如图2所示,它以Retinex模型为基础,该模型基于基本的物理假设:

$$\mathbf{I} = \mathbf{L} * \mathbf{R} \quad (1)$$

其中, \mathbf{I} 表示原始低光照图像; \mathbf{L} 表示图像照度,本文为图像增强网络预测值; \mathbf{R} 表示图像反射分量,即希望得到的增强图像。为确保模型轻量化并提高处理速度,本文引入双边学习^[18-24](Deep Bilateral Learning),采用低分辨率和全分辨率两条路径进

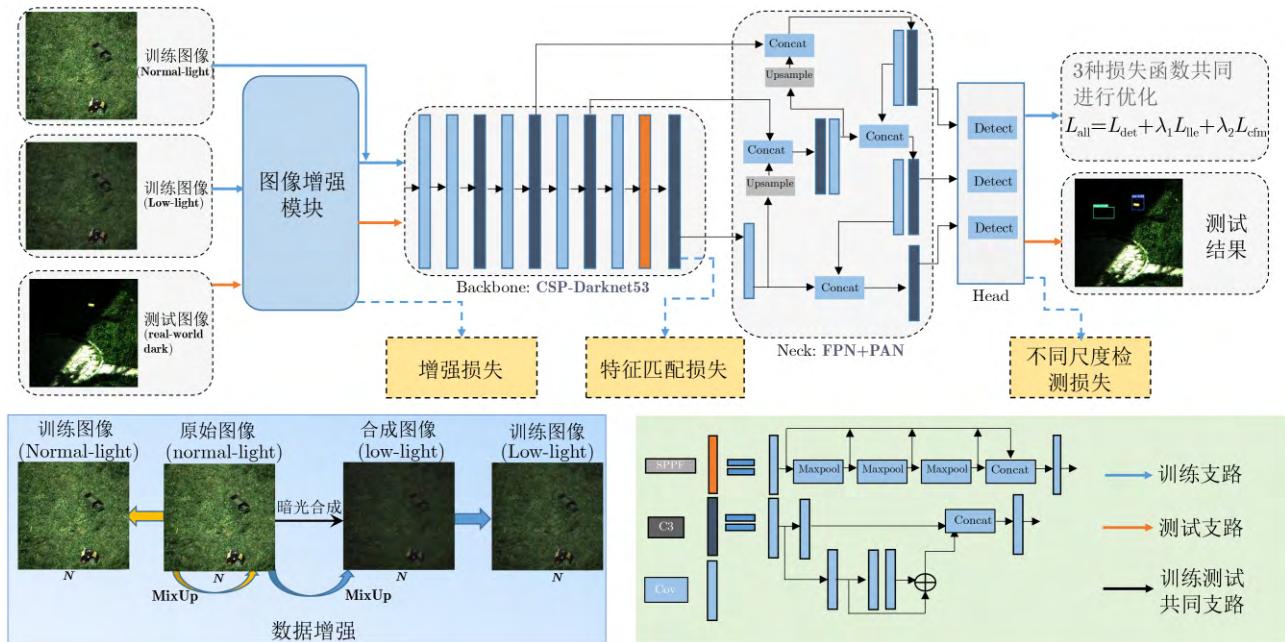


图1 网络整体结构图

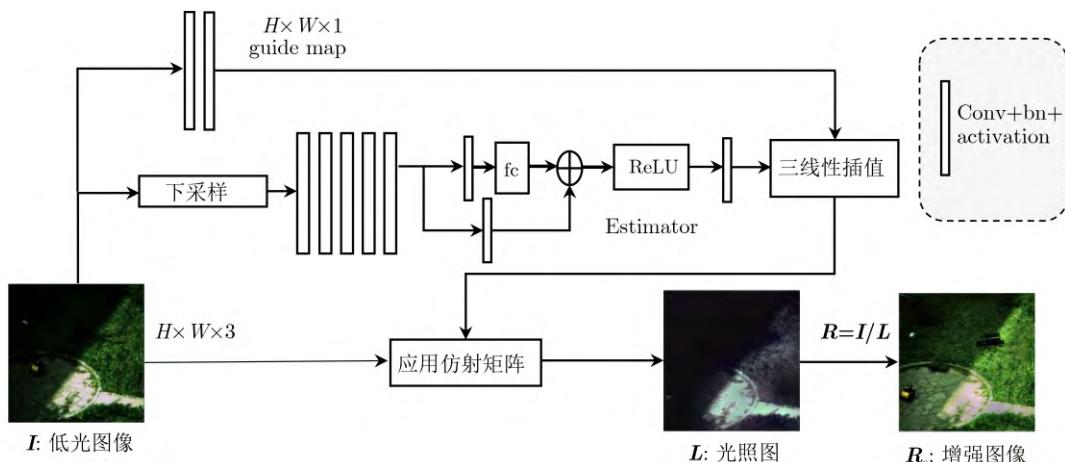


图2 图像增强模块

行优化。低分辨率路径通过下采样后得到 256×256 分辨率图像,再经过卷积层提取特征,通过系数评估器模块Estimator从低分辨率路径提取低分辨率图像照度。而全分辨率路径学习单通道的全分辨率引导图,并在相应的3D坐标上对低分辨率双边网格进行线性插值,以生成最终的全分辨率图像照度,从而得到增强图像。大部分计算在低分辨率路径进行,因此这是一个轻量级的有监督图像增强网络,训练时以图1中正常光图像(Normal-light)作为监督约束,以其合成的低光图像(Low-light)为输入,通过在颜色和平滑度等方面约束增强图像与正常光图像的一致性来训练网络,损失函数部分的详细内容将在后续进行介绍。

在传统有监督学习框架下,低光照增强模型虽可通过像素级重建约束改善图像亮度,但由于优化

目标与检测网络的语义特征提取存在偏差,增强图像经Backbone网络提取的深层特征仍与正常光图像存在显著分布差异。这种域间特征分布不匹配现象会削弱检测网络对有效语义特征的表征能力,最终降低目标检测精度。针对低光照增强图像与真实正常光图像间在深层语义特征上的差异,本文提出基于全局统计建模的Kullback-Leibler散度约束方法,以实现特征自适应学习。具体而言,在网络的特征对齐阶段,对Backbone网络输出的最后一层高层语义特征进行空间维度的平均化处理(Spatial Global Average Pooling),提取其通道级统计向量,分别表征增强图像特征空间 F_{enh} 和正常光图像特征空间 F_{real} 的全局分布特性。通过计算两者之间的KL散度来构建特征匹配损失,从而约束它们的一致性。

3.2 图像数据增强

本文采用暗光合成方法用正常光照图像合成低光图像，并与相应的正常光图像配对形成图像对。在训练阶段，使用配对的数据集进行训练，再在真实低光图像上进行finetune，可以有效提高低光检测的精度。参照文献[14]，通过考虑物理噪声模型和图像信号处理(Image Signal Processing, ISP)来模拟真实照明退化过程，在图1中用Dark ISP表示。

如图1所示的数据增强过程，用 N 张正常光图像 \mathbf{I}_n (normal-light)合成 N 张相对应的低光图像 \mathbf{I}_l (low-light)，基于MixUp将两种图像按比例随机混合，得到 N 张低光训练图像 \mathbf{I}_L (Low-light)，同时，对正常光图像自身进行MixUp得到 N 张正常光训练图像 \mathbf{I}_N (Normal-light)，这一操作不仅扩充了训练数据的多样性，还强制增强模块学习光照条件连续变化的映射关系，从而缓解过拟合并提升对复杂光照的适应能力。MixUp具体实现如式(2)所示，Beta(α, α)表示beta分布，其中参数 $\alpha=0.4$ 。上述过程可以用下式表示

$$\mathbf{I}_{\text{mixup}} = \theta \mathbf{I}_A + (1 - \theta) \mathbf{I}_B, \theta \sim \text{Beta}(\alpha, \alpha) \quad (2)$$

$$\mathbf{I}_l = \text{DarkISP}(\mathbf{I}_n) \quad (3)$$

$$\mathbf{I}_L = \text{MixUp}(\mathbf{I}_l, \mathbf{I}_n), \mathbf{I}_N = \text{MixUp}(\mathbf{I}_n, \mathbf{I}_n) \quad (4)$$

3.3 损失函数

由于本文的训练范式为增强和检测的端到端训练，本文的损失函数包括增强损失、匹配损失和检测损失，其表达式为

$$L_{\text{all}} = \lambda_1 L_{\text{lle}} + \lambda_2 L_{\text{cfm}} + L_{\text{det}} \quad (5)$$

式中， L_{lle} 为图像增强损失， L_{cfm} 为特征匹配损失， L_{det} 为检测损失。 λ_1 和 λ_2 为平衡损失函数引入的权重因子，根据实验结果设定为0.3。

低光图像增强损失 L_{lle} 表示低光照图像增强后的结果与真实正常光照图像之间的损失，它的计算如式(6)

$$L_{\text{lle}} = L_r(\hat{\mathbf{R}}_e, \mathbf{R}) + \lambda_3 L_s \left(\frac{\mathbf{I}}{\hat{\mathbf{R}}_e} \right) + \lambda_4 L_c(\hat{\mathbf{R}}_e, \mathbf{R}) \quad (6)$$

其中， $\hat{\mathbf{R}}_e$ 表示低光图像增强后的结果， \mathbf{R} 表示对应的真实正常光照图像， \mathbf{I} 表示原始的低光图像。根据参考文献[4]， L_r 表示图像重建损失， L_s 用来增强图像的平滑度， L_c 用于控制低光图像增强结果中的颜色偏差。权重因子 $\lambda_3=2$ ， $\lambda_4=1$ 。

L_{cfm} 损失表示低光图像增强后的结果与真实正常光照图像在检测阶段提取到的特征之间的匹配损失，使增强图像在特征层面渐进接近正常光图像。它的计算公式如式(7)

$$L_{\text{cfm}} = \text{KL}(\mathbf{F}_{\text{real}} \parallel \mathbf{F}_{\text{enh}}) + \text{KL}(\mathbf{F}_{\text{enh}} \parallel \mathbf{F}_{\text{real}}) \quad (7)$$

式(7)用于计算低光图像增强后与真实正常光照图像的特征之间的KL散度，其中 \mathbf{F}_{real} 表示正常光图像直接输入Backbone提取的特征空间平均化结果， \mathbf{F}_{enh} 表示配对的模型增强的低光图像经Backbone提取的对应结果。特征匹配损失函数约束增强图像与正常光图像的特征在深层语义空间中的分布距离。因此，特征匹配损失总体为：基于低光照增强图像与对应正常光图像间存在的语义一致性先验条件，将增强图像与配对的正常光图像一同输入YOLOv5模型，经过Backbone提取高层语义特征后，使用本文提出的特征匹配损失约束两种图像特征的统计分布差异，能够实现增强图像特征向正常光特征空间的渐进式自适应迁移。

L_{det} 损失表示目标检测模块输出的边框预测值与真实值之间的计算损失，根据YOLOv5的设计，它的计算公式如式(8)

$$L_{\text{det}} = \beta_1 L_i(\hat{y}_{\text{box}}, y_{\text{box}}) + \beta_2 \text{CE}(\hat{y}_{\text{cls}}, y_{\text{cls}}) + \beta_3 \text{CE}(\hat{y}_{\text{conf}}, y_{\text{conf}}) \quad (8)$$

$L_{\text{ciou}}(\hat{y}_{\text{box}}, y_{\text{box}})$ 用来约束预测框和真实框的距离， $\text{CE}(\hat{y}_{\text{cls}}, y_{\text{cls}})$ 和 $\text{CE}(\hat{y}_{\text{conf}}, y_{\text{conf}})$ 表示用交叉熵损失来约束预测的类别和置信度与真实值的距离。权重因子 $\beta_1=0.05$ ， $\beta_2=0.50$ ， $\beta_3=0.10$ 。

4 实验

4.1 实验数据

本文在LLVIP和Polar3000数据集上进行了测试，在ExDark和DarkFace数据集进行了泛化实验，进行了训练数据集包括和不包括低光照图像的两种实验，实验数据如下。

Winderperson→LLVIP: LLVIP数据集是一个贴近自动驾驶场景的包含道路的低光照数据集，包括30976张图像(15488对，即可见光+红外)。这里只用可见光部分数据15488张图像进行实验，其中12025张用于finetune，3463张用于测试。用另一个正常光照的道路数据集Winderperson进行训练，它包括13382张图像，其中训练集为8000张图像，使用3.2节中的暗光合成方法合成对应低光图像形成配对数据集，最终得到训练数据集。

Polar3000: 针对白昼场景的测试研究，本文选用了Polar3000数据集作为评估基准，该数据集包括car, tank, UAV和truck 4个类别。该数据集中的图像均在强光照环境下采集，呈现出大面积的过曝区域与阴影区域，无人机UAV的亮度略大于过曝区域，其中阴影区域的检测为本文研究的低光照图像目标检测任务。使用正常光图像进行训练，包括

2376张图像, 用912张图像finetune, 用190张包含阴影区域的图像进行测试。

Widerface→DarkFace: DarkFace数据集为专门构建的低光照人脸检测数据集, 图像为夜间、微光等昏暗场景, 使用其5 400张图像用于finetune, 600张图像用于测试。Widerface包含丰富的人脸图像, 原始图像采集于多种日常场景, 光照充足。使用其12 880张图像用于训练, 使用3.2节的暗光合成方法合成对应低光图像形成配对数据集, 最终得到训练数据集。

ExDark: 是一个专为低光照物体检测与识别设计的数据集, 包括室内、街道、海上等环境。数据集中包含7 363张图像, 涵盖12个类别(如车和人等), 均采集于低照度场景, 目标细节模糊。使用其5 896张图像finetune, 1 467张图像测试。

4.2 实验环境

本文在Windows环境下RTX4090显卡, Python3.8和Torch2.1.0+cu121的特定环境下对模型进行了训练和测试。参照YOLOv5的实验设定, 本文使用随机梯度下降法对整个模型进行优化, 学习率为0.01, 权重衰减设置为0.0005, 动量设置为0.937。为了对模型进行3个周期的预热, 我们使用了One-CycleLR学习率调整策略, 初始学习率为0.003, 然后进行了总共50个周期的训练, 最后一个周期的学习率调整为0.002, 并且本文还使用了YOLO11模型对比, 所有参数设置和YOLOv5一致。对于文中用于比较的模型, 我们使用相应作者提供的训练方法和参数进行训练。

4.3 实验设置

将本文提出的方法与范式(a)中先图像增强再检测的方法进行对比, 涵盖了DeepUPE^[5], DUAL^[18], LIME^[19], ZeroDCE^[2], RetinexNet^[1], SNRALLE^[20], PairLIE^[21], NeRCo^[22]和EMNet^[23]等, 也与最近的低光照目标检测方法MAET^[16], PEYOLO^[8], DEYOLO^[7]和FEYOLO^[15]等进行了比较。测试数据集为LLVIP和Polar3000。为了展示本文方法的泛化性能, LLVIP数据集使用L-type的YOLOv5模型, 并使用L-type的YOLO11模型进行对比, Polar3000数据集使用S-type的YOLOv5模型, 用S-type的YOLO11模型进行对比。然后根据是否使用真实低光照数据集进行训练, 将实验设置划分为A组和B组, 其中B组实验中先增强后检测的方法使用了YOLOv5预训练权重。A组实验为零样本自适应实验(Zero-shot Adaptation), 即仅使用正常光照图像进行训练, 并在低光照图像上测试检测精度。该实验旨在评估模型的泛化能力, 以及在缺少真实低

光照图像数据时对低光环境的检测性能; B组实验采用完全监督学习(Fully Supervised), 训练数据包含真实低光照图像。具体而言, 模型先使用正常光照图像进行训练, 再利用低光照图像进行微调, 以提升不同模型对低光照图像的检测能力。

4.4 定性分析

图3展示了基于Polar3000数据集的A组实验下使用不同模型的检测效果图, 直接对2376张图像使用YOLOv5模型训练后, 测试集结果如图3(a)所示, 由于训练场景和测试场景的不同, 检测精度较低, 无法识别出藏在阴影区域tank样本, 且将过曝区域大量明亮场景识别为同样明亮的无人机UAV, 这是因为训练图像样本中不存在阴影区域和明亮区域, 所以YOLOv5模型在这类场景下检测率低, YOLO11模型虽然检测到黑暗中坦克目标, 但仍将过曝区域大量明亮场景识别为同样明亮的无人机UAV。DUAL^[18], LIME^[19], DeepUPE^[5]和ZeroDCE^[2]直接对低光照图像进行增强, 再使用基于2376张正常光图像训练的YOLOv5模型进行检测, 通过目标检测结果图可以看出, 图像整体亮度提升, 但噪声增加, 图像细节存在部分丢失, 且过曝区域存在过度增强现象。这类先增强后检测的方法目标检测精度虽然有所提升, 但仍然大量存在将过曝区域识别为无人机UAV的现象。这是因为图像增强方法虽然增加了光照强度, 但难以协同阴影区域和过曝区域的增强效果, 使过曝区域光照进一步增强, 因此过曝区域识别为无人机UAV的现象难以消除。而使用低光照目标检测方法MAET时, 由于没有对图像进行增强, 过曝光区域没有被大量检测为无人机, 但阴影区域仍然存在误识别, 表明MAET方法在正常光图像训练低光图像测试的情况下, 从正常光到低光的泛化能力较弱。而本文方法结合YOLOv5和YOLO11均实现了较高的检测精度, 能够同时检测出阴影区域和过曝区域的目标, 这是由于本文方法中数据增强、图像增强模块和特征匹配损失的联合作用, 正常光数据集经过暗光合成使训练数据集中同时存在正常光图像和低光照图像, 低光照图像输入到本文端到端的模型后, 经过正常光图像和图像增强模块的作用, 隐式地增强了光照且有效抑制噪声放大效应, 基于KL散度的特征匹配损失引导增强图像深层语义特征向正常光特征空间渐进对齐, 二者的协同优化实现高精度目标检测。

图4展示了基于LLVIP数据集的B组实验下的检测效果图, 在这种全监督的训练模式下, 可以使用YOLOv5自带的预训练权重以最大化检测能力。方法DUAL^[18], LIME^[19], DeepUPE^[5]和ZeroDCE^[2]

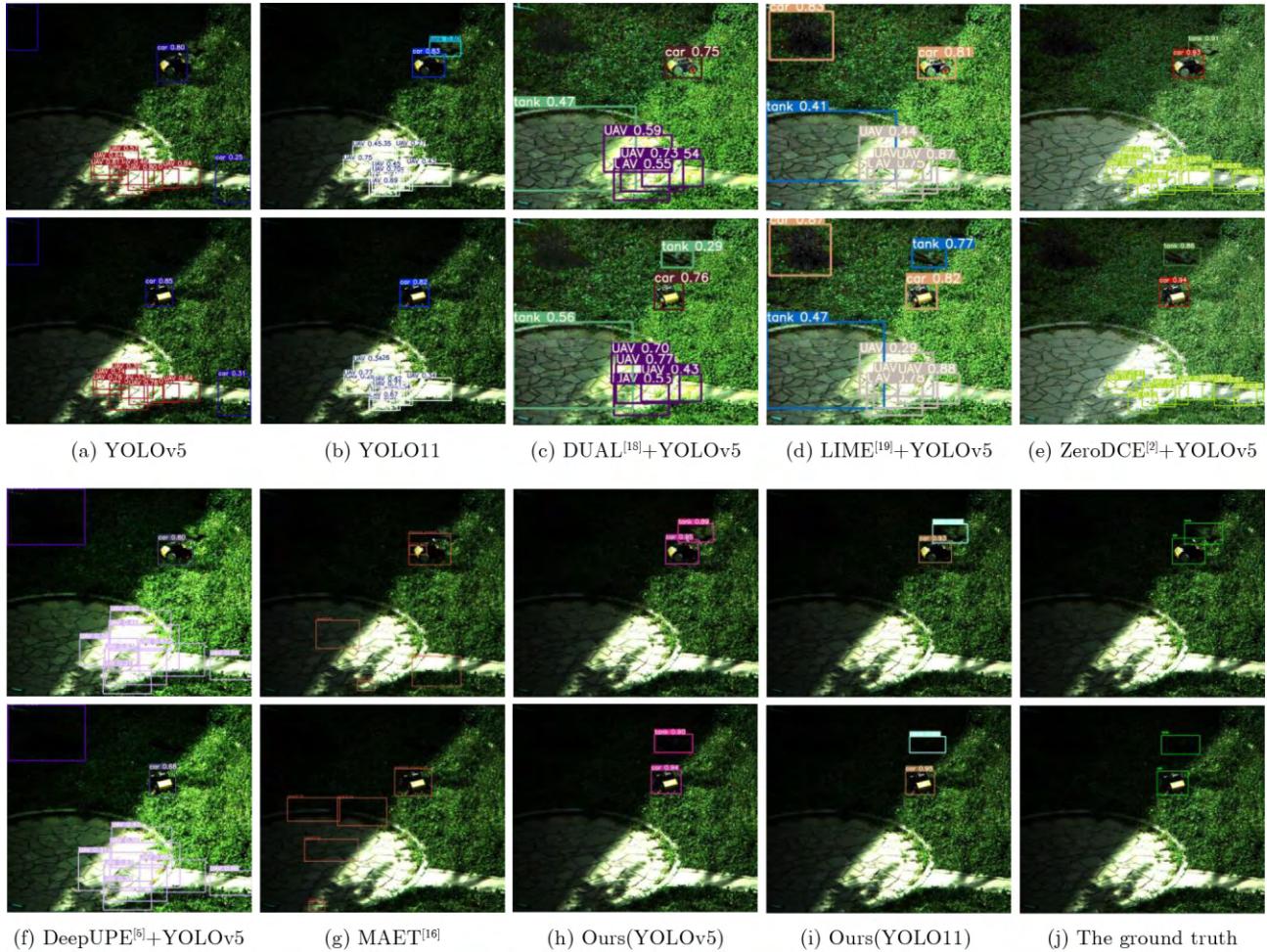


图3 不同方法在Polar3000数据集上的A组实验测试结果

直接对LLVIP低光图像进行增强，使用13 382张Widerperson正常光数据集对YOLOv5模型进行训练，然后在12 025张LLVIP图像微调使达到最佳目标检测能力。通过检测结果图看出，使用预训练模型和微调后，原YOLOv5模型基本能检测出此光照条件下的行人，但是存在检测置信度不高的问题，相较于YOLOv5模型，YOLO11模型在低照度环境中的目标检测能力有所提升，先增强后检测的方法虽提升图像亮度，但易引入噪声和信息丢失，导致检测精度下降，其中DUAL^[18]和LIME^[19]图像噪点多，视觉效果差，存在较多将背景检测为行人的现象。MAET对检测到的行人置信度高，但存在部分漏检，且误检率高，存在将交通信号灯和反光的垃圾箱检测为行人的现象，说明其在复杂低光环境下性能受限。而本文的方法通过数据增强、图像增强模块和特征匹配损失的联合作用，在训练过程中混合了低光和正常光目标，能够检测不同光照场景，在相同条件下无漏检和误检，检测框与真实框的匹配程度最高，且检测置信度较高，结合YOLOv5和YOLO11均取得良好的结果。综上，本文方法不仅

具备零样本自适应能力，而且用真实低光图像微调后，在低光环境下能实现高精度检测。

图4展示了基于DackFace数据集的B组实验下的检测效果图，在这种光照较暗且在变化、人脸密集且为小目标的情形，目标检测模型的检测精度大多处于较低的水平。从图中可以看出，尽管图像增强方法提升了亮度，但易破坏纹理并引入噪声，如图4(c)中DUAL方法所示。如图4(e)所示，ZeroDCE图像增强方法图像增强效果较好，但图像还是存在噪声干扰，影响目标检测结果，而MAET方法应对小目标人脸检测置信度较低。本文方法结合YOLOv5和YOLO11后，目标检测的漏检率和误检率均有所下降，对于容易检测出的目标检测置信度有所提升，证明了本文通过数据增强、图像增强模块和特征匹配损失的联合作用实现低光图像特征向正常光图像特征对齐的方法，应对光照变化的小目标具有泛化性。

4.5 定量分析

本文通过在LLVIP和Polar3000两个数据集上进行实验，将所提出的方法分别与YOLOv5结合其

他图像增强方法以及一些先进的低光照目标检测方法进行了检测精度的对比分析,并在DarkFace和ExDark数据集进行了泛化实验。

如表1所示, A组实验为在Winderperson数据集上训练, LLVIP数据集上测试的结果,使用的图像增强模型有LIME^[19], NeRCo^[22]和EMNet^[23]。在使用真实正常光图像训练的模式下,对低光照图像

进行增强能提升检测精度,但效果有限,这是因为图像增强提升了光照强度,但是增强给图像带来了噪声,所以效果有限,MAET和PEYOLO等模型检测精度较低,主要原因在于它们仅在正常光照条件下进行训练,缺乏对低光图像的学习。因此,这些模型在面对光照条件变化时缺乏从正常光照到低光环境的泛化能力。相比之下,本文提出的图像增

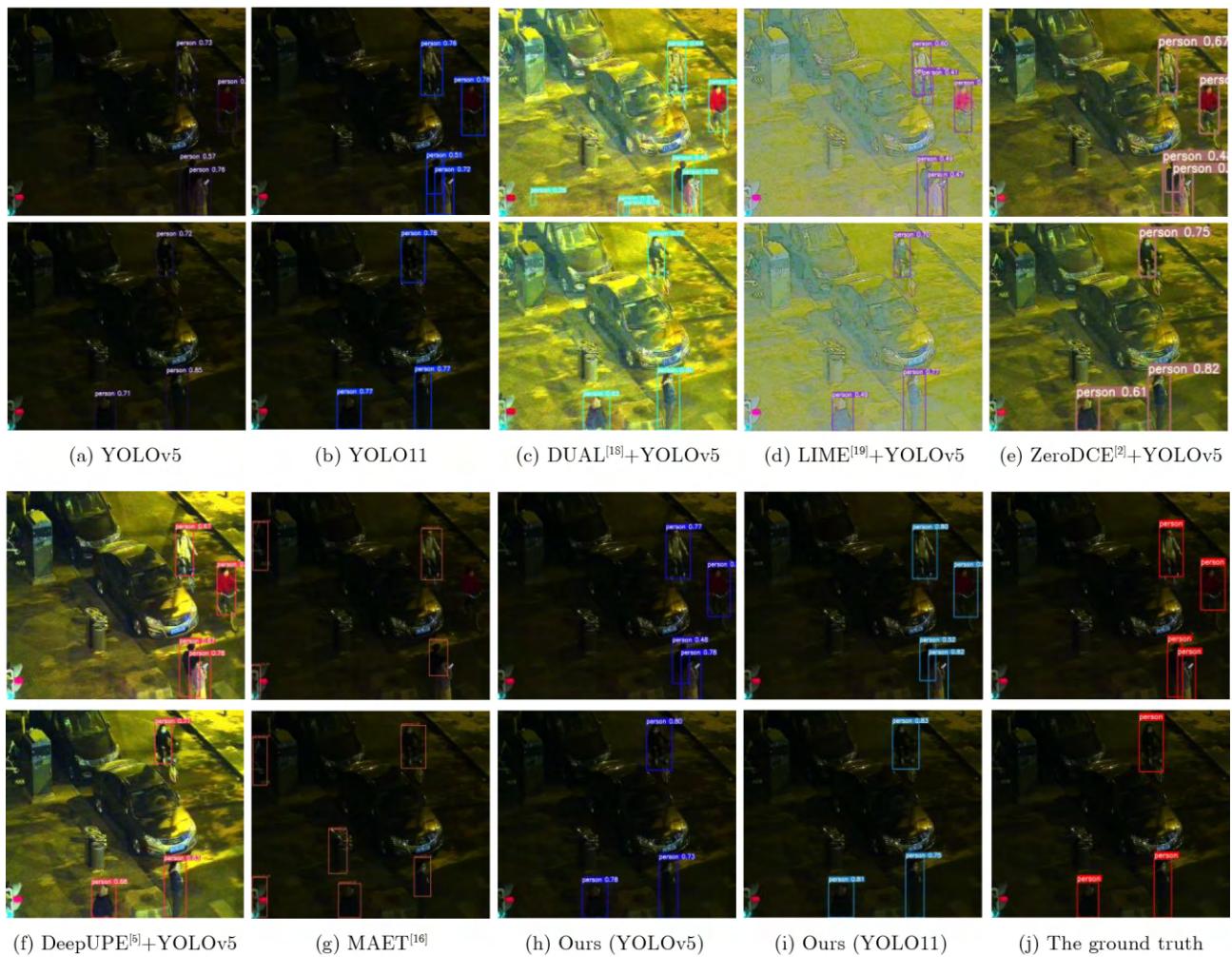


图4 不同方法在LLVIP数据集上的B组的实验测试结果

表1 Winderperson→LLVIP的对比实验结果(%)

A组实验			B组实验		
方法	mAP50	mAP75	方法	mAP50	mAP75
YOLOv5	55.6	10.2	YOLOv5	82.1	37.9
+LIME ^[19]	59.1	11.8	+EMNet	74.6	31.8
+NeRCo ^[22]	60.5	12.9	+NeRCo	75.2	33.4
+EMNet ^[23]	62.9	13.5	+LIME	75.1	35.2
MAET ^[16] (YOLOv3)	24.5	3.1	MAET(YOLOv3)	62.1	12.9
FEYOLO ^[15] (YOLOv5)	64.6	14.1	PEYOLO(YOLOv5)	84.1	36.8
Ours(YOLOv5)	79.5	20.3	Ours(YOLOv5)	91.7	53.1
Ours(YOLO11)	83.5	24.1	Ours(YOLO11)	93.8	59.1

强与特征自适应联合学习的模型结合YOLOv5和YOLO11后以79.5%和83.5%的mAP值实现了较高的检测精度,说明在没有真实低光照图像作为训练数据时,本方法具有良好的泛化能力。在B组实验中,YOLOv5结合图像增强的方法使用YOLOv5自带的预训练权重。在这种情况下,对LLVIP进行图像增强会降低检测精度,这是因为YOLOv5预训练时使用大量图像数据训练,具备良好的泛化能力,此时,图像增强带来的照度提升不足以弥补其带来的噪声和信息损失。本文提出的方法、MAET和PEYOLO在LLVIP上进行finetune后精度均有所提升,其中本文方法结合YOLOv5和YOLO11后以91.7%和93.8%的mAP值实现了最高的检测精度,这是因为引入图像增强模块和特征匹配损失后,模型在处理真实低光图像时能够隐式提升图像照度,同时有效抑制噪声,使得在进入Backbone进行特征提取时,其特征分布更趋近于正常光图像,从而提升下游任务性能。

如表2所示, A组实验为在Polar3000数据集正常光图像上训练,在190张含阴影图像上测试的结果,当仅用YOLOv5从头开始训练时, mAP值仅为5.7%,表明模型在这种光照条件下几乎无法检测到图像中的物体,使用YOLO11时, mAP值提升

至25.1%,仍是较低水平。对这190张含阴影的图像用不同方法增强后,检测精度虽有提升,但仍不理想,主要是因为图过曝区域光照过度增强,大量背景被误识别为明亮无人机UAV,这与图1中定性分析中结果相互印证。相比PEYOLO和MAET等方法,本文方法以85.7%的检测精度达到了最高,结合YOLO11进一步提升至87.9%。B组实验YOLOv5结合图像增强的方法且使用YOLOv5自带的预训练权重,其他低光照目标检测方法使用912张含阴影区域图像微调,先增强后检测的实验结果与表2中B组实验结果相似,在此情况下,图像增强带来的照度提升不足以弥补其带来的噪声和信息损失。本文方法以92.3%达到最高检测精度,结合YOLO11进一步提升至93.7%,进一步验证了其在低光目标检测中的优势。

如表3所示, A组实验为在12880张正常光人脸数据集Widerface上训练,在低光照数据集DarkFace的600张图像上测试结果,由于DarkFace图像亮度处于较低水平, YOLOv5模型检测面临困难,采用不同图像增强方法后,检测精度有所提升,但图像增强方法给图像带来不同程度噪声,如图5所示, DUAL方法所带来噪声最多,因此检测效果不佳。而本文在A和B两组实验分别表现出最高的检

表2 Polar3000的对比实验结果(%)

A组实验			B组实验		
方法	mAP50	mAP75	方法	mAP50	mAP75
YOLOv5	5.7	2.7	YOLOv5	82.1	37.9
+DeepUPE ^[5]	10.8	2.8	+ZeroDCE	75.1	38.9
+ZeroDCE ^[2]	20.6	9.0	+LIME	77.1	43.8
+LIME ^[19]	34.9	21.5	+DeepUPE	80.5	46.3
PEYOLO ^[8] (YOLOv5)	53.9	38.9	PEYOLO(YOLOv5)	75.1	49.3
MAET ^[16] (YOLOv3)	58.9	16.1	MAET(YOLOv3)	85.7	1.0
本文方法(YOLOv5)	85.7	72.0	本文方法(YOLOv5)	92.3	76.1
本文方法(YOLO11)	87.9	74.1	本文方法(YOLO11)	93.7	77.4

表3 Widerface→DarkFace的对比实验结果(%)

A组实验			B组实验		
方法	mAP50	mAP75	方法	mAP50	mAP75
YOLOv5	11.5	0.66	YOLOv5	40.8	8.33
+DeepUPE ^[5]	17.5	0.59	+PairLIE	42.0	8.30
+PairLIE ^[21]	21.8	0.92	+DeepUPE	42.1	8.28
+EMNet ^[23]	22.2	1.30	+EMNet	43.9	8.81
PEYOLO ^[8] (YOLOv5)	10.2	0.68	PEYOLO(YOLOv5)	38.9	7.92
MAET ^[16] (YOLOv3)	14.9	1.12	MAET(YOLOv3)	35.7	7.52
本文方法(YOLOv5)	29.3	1.63	本文方法(YOLOv5)	55.5	16.6

测性能, mAP值分别为33.6%和60.1%, 高于图像增强方法和其他低光目标检测方法。实验证明, 在这中光照低且目标较小, 人脸分布密集的情况下, PEYOLO和MAET的检测精度均较低, 表明本文方法在低光照环境下对小目标具有较强的检测能力, 证明了本文方法的泛化性能。

为进一步验证本文方法的泛化能力, 本文进行了交叉数据集验证实验。首先使用Winderperson→LLVIP实验中训练的模型, 即在8 000张正常光行人数据集Winderperson进行训练后在LLVIP数据集测试的模型权重, 因为Winderperson和LLVIP均为行人数据集, 所以对ExDark数据集中609张People类别图像测试, 单独检测People这一类别检测精度, 以验证在正常光数据集Winderperson训练好的不同目标检测模型在低光数据集ExDark上的泛化能力。本文方法在此场景得到最佳检测精度66.3%, 结果如表4所示。

此外, 本文对参数量和检测速度进行了比较, 实验中使用的YOLOv5模型均为L-type。PEYOLO引入了拉普拉斯金字塔, 虽然参数量增加量较小, 但高低频处理方式导致图像处理速度慢。FEYOLO在YOLOv5前引入了增强模块FeatEnHancer, 用于增强特征层次结构, 其引入了注意力机制, 卷积层多, 浮点运算量大。本文方法与YOLOv5基准模型相比, 增加的模块主要是图像增强模块, 但由于其主要计算在低分辨率路径, 因此参数量和计算量较小, 如表5所示, 参数量仅增加了2.5%, 在计算复杂度GFLOPs方面同样表现出色, 在图像大小均设置为640×640时, 检测速度在Polar3000数据集上仅比基准模型慢5.1ms。表明其在保持高精度的同时, 也具备良好的实时性。

4.6 消融实验

如表6所示, 展示了Winderperson→LLVIP消融实验结果。本文将所提方法划分为4个部分: 数



图5 不同方法在DarkFace数据集上的B组的实验测试结果

表4 Exdark数据集People类别对比实验结果(%)

方法	mAP50	mAP75	方法	mAP50	mAP75
YOLOv5	41.8	5.2	MAET ^[16] (YOLOv3)	23.1	4.1
+DeepUPE ^[5]	47.5	6.9	PEYOLO ^[10] (YOLOv5)	46.8	8.2
+EMNet ^[23]	49.8	7.1	FEYOLO ^[15] (YOLOv5)	55.3	11
+ZeroDCE ^[2]	51.2	7.4	本文方法(YOLOv5)	66.3	14.9

表5 模型复杂度比较

方法	参数量	浮点运算次数(GFLOPs)	推理时间(每张图像)
YOLOv5	47,025,981	115.3	14.0 ms
PEYOLO ¹ (YOLOv5)	47,117,184	137.7	49.0 ms
FEYOLO ¹ (YOLOv5)	47,165,381	202.0	26.2 ms
本文方法(YOLOv5)	48,245,503	141.5	19.1 ms

表6 消融实验结果(%)

任务	数据增强	增强损失	匹配损失			是否微调	mAP50	mAP75
			L_{cfm}	L_2	L_{ef}			
Winderperson →LLVIP	×	×	×	×	×	×	55.6	10.2
	√	×	×	×	×	×	71.2	19.1
	√	√	×	×	×	×	73.1	19.7
	√	√	×	√	×	×	73.5	19.2
	√	√	×	×	√	×	75.6	19.7
	√	√	√	×	×	×	79.5	20.3
	√	×	√	×	×	×	77.1	19.6
	√	√	×	×	×	√	89.1	52.2
	√	√	√	×	×	√	91.7	53.1

据增强、增强损失、特征匹配损失以及低光图像数据微调。

在数据增强部分的实验中,设置了两组对照方案:第1组仅使用正常光照图像进行训练;第2组则采用正常光照图像与合成低光照图像配对,并结合MixUp方法进行训练,结果显示, mAP值从55.6%提升至71.2%,这表明数据增强对提升低光照图像检测精度具有显著作用。当在模型中加入图像增强损失模块时, mAP值从71.2%提升至73.1%;在此基础上,进一步加入特征匹配损失后, mAP值提升至79.5%。为了更充分体现特征匹配损失的作用, 使用欧氏距离L2替代特征匹配损失约束正常光图像特征和增强图像特征 F_{enh} 之间距离, map值为73.5%,如果在匹配损失中添加约束增强图像与低光图像损失的部分

$$L_{\text{ef}} = \text{KL}(\mathbf{F}_{\text{real}} \parallel \mathbf{F}_{\text{enh}}) + \text{KL}(\mathbf{F}_{\text{enh}} \parallel \mathbf{F}_{\text{real}}) + \text{KL}(\mathbf{F}_{\text{low}} \parallel \mathbf{F}_{\text{enh}}) + \text{KL}(\mathbf{F}_{\text{enh}} \parallel \mathbf{F}_{\text{low}}) \quad (9)$$

mAP值为75.6%,这是因为低光图像特征分布改变了增强图像特征空间分布,只用正常光图像特征约束效果更好。加入匹配损失而不使用图像增强损失模块时, mAP值为77.1%;这些结果分别验证了图像增强模块和特征匹配的作用。为了更深入地验证特征匹配的有效性,本文还在LLVIP数据集上进行了微调实验。结果显示,加入匹配损失后,模型的检测精度mAP值从89.1%提升至91.7%,这一结果充分证明了特征匹配对提升低光图像目标检测的作用。

5 结论

本文提出一种基于图像增强与特征自适应联合学习的低光照目标检测方法,通过图像增强损失、特征匹配损失和目标检测损失进行联合优化。该方法能够以端到端的框架提升低光照图像的光照强度,保留图像细节,并使增强后的图像特征与正常光照图像的特征高度匹配,从而提高检测效果。在LLVIP和Polar3000数据集上的实验表明,本文方法表现优异,并在Exdark和DarkFace的泛化实验中取得最佳检测效果。为了进一步提高在极暗环境下的目标检测精度,未来的工作将聚焦于可见光与红外图像的多模态数据下的目标检测任务。

参 考 文 献

- [1] WEI Chen, WANG Wenjing, YANG Wenhan, et al. Deep retinex decomposition for low-light enhancement[C]. Proceedings of British Machine Vision Conference, Newcastle, UK, 2018: 155.
- [2] GUO Chunle, LI Chongyi, GUO Jichang, et al. Zero-reference deep curve estimation for low-light image enhancement[C]. Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020: 1777–1786. doi: 10.1109/CVPR42600.2020.00185.
- [3] LUO Yu, CHEN Xuanrong, LING Jie, et al. Unsupervised low-light image enhancement with self-paced learning[J]. IEEE Transactions on Multimedia, 2025, 27: 1808–1820. doi: 10.1109/TMM.2024.3521752.

- [4] WANG Qiang, CUI Yuning, LI Yawen, *et al.* RFFNet: Towards robust and flexible fusion for low-light image denoising[C]. Proceedings of the 32nd ACM International Conference on Multimedia, Melbourne, Australia, 2024: 836–845. doi: 10.1145/3664647.3680675.
- [5] WANG Ruixing, ZHANG Qing, FU C W, *et al.* Underexposed photo enhancement using deep illumination estimation[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 6842–6850. doi: 10.1109/CVPR.2019.00701.
- [6] JOCHER G, STOKEN A, BOROVEC J, *et al.* ultralytics/yolov5: v5.0-YOLOv5-P6 1280 models, AWS, supervise.ly and YouTube integrations[EB/OL]. <https://zenodo.org/records/4679653>, 2021.
- [7] QIN Qingpao, CHANG Kan, HUANG Mengyuan, *et al.* DENet: Detection-driven enhancement network for object detection under adverse weather conditions[C]. Proceedings of the 16th Asian Conference on Computer Vision, Macao, China, 2022: 491–507. doi: 10.1007/978-3-031-26313-2_30.
- [8] YIN Xiangchen, YU Zhenda, FEI Zetao, *et al.* PE-YOLO: Pyramid enhancement network for dark object detection[C]. Proceedings of the 32nd International Conference on Artificial Neural Networks on Artificial Neural Networks and Machine Learning, Heraklion, Crete, Greece, 2023: 163–174. doi: 10.1007/978-3-031-44195-0_14.
- [9] REDMON J and FARHADI A. YOLOv3: An incremental improvement[J]. arXiv preprint arXiv: 1804.02767, 2018. doi: 10.48550/arXiv.1804.02767.
- [10] LIU Wenyu, REN Gaofeng, YU Runsheng, *et al.* Image-adaptive YOLO for object detection in adverse weather conditions[C]. Proceedings of the 36th AAAI Conference on Artificial Intelligence, 2022: 1792–1800. doi: 10.1609/aaai.v36i2.20072.
- [11] JIA Xinyu, ZHU Chuang, LI Minzhen, *et al.* LLVIP: A visible-infrared paired dataset for low-light vision[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, Montreal, Canada, 2021: 3489–3497. doi: 10.1109/ICCVW54120.2021.00389.
- [12] FENG Bin, XIAO Jinpei, ZHANG Junchao, *et al.* Color-polarization synergistic target detection method considering shadow interference[J]. *Defence Technology*, 2024, 37: 50–61. doi: 10.1016/j.dt.2024.01.007.
- [13] WANG Shuhang and LUO Gang. Naturalness preserved image enhancement using *a priori* multi-layer lightness statistics[J]. *IEEE Transactions on Image Processing*, 2018, 27(2): 938–948. doi: 10.1109/TIP.2017.2771449.
- [14] LOH Y P and CHAN C S. Getting to know low-light images with the exclusively dark dataset[J]. *Computer Vision and Image Understanding*, 2019, 178: 30–42. doi: 10.1016/j.cviu.2018.10.010.
- [15] HASHMI K A, KALLEMPUDI G, STRICKER D, *et al.* FeatEnHancer: Enhancing hierarchical features for object detection and beyond under low-light vision[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2023: 6702–6712. doi: 10.1109/ICCV51070.2023.00619.
- [16] CUI Ziteng, QI Guojun, GU Lin, *et al.* Multitask AET with orthogonal tangent regularity for dark object detection[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, Canada, 2021: 2533–2542. doi: 10.1109/ICCV48922.2021.00255.
- [17] DU Zhipeng, SHI Miaojing, and DENG Jiankang. Boosting object detection with zero-shot day-night domain adaptation[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2024: 12666–12676. doi: 10.1109/CVPR52733.2024.01204.
- [18] ZHANG Qing, NIE Yongwei, and ZHENG Weishi. Dual illumination estimation for robust exposure correction[J]. *Computer Graphics Forum*, 2019, 38(7): 243–252. doi: 10.1111/cgf.13833.
- [19] GUO Xiaojie, LI Yu, and LING Haibin. LIME: Low-light image enhancement via illumination map estimation[J]. *IEEE Transactions on Image Processing*, 2017, 26(2): 982–993. doi: 10.1109/TIP.2016.2639450.
- [20] XU Xiaogang, WANG Ruixing, FU C W, *et al.* SNR-aware low-light image enhancement[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, USA, 2022: 17693–17703. doi: 10.1109/CVPR52688.2022.01719.
- [21] FU Zhenqi, YANG Yan, TU Xiaotong, *et al.* Learning a simple low-light image enhancer from paired low-light instances[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, Canada, 2023: 22252–22261. doi: 10.1109/CVPR52729.2023.02131.
- [22] YANG Shuzhou, DING Moxuan, WU Yanmin, *et al.* Implicit neural representation for cooperative low-light image enhancement[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2023: 12872–12881. doi: 10.1109/ICCV51070.2023.01187.
- [23] YE Dongjie, NI Zhangkai, YANG Wenhan, *et al.* Glow in the dark: Low-light image enhancement with external memory[J]. *IEEE Transactions on Multimedia*, 2024, 26: 2148–2163. doi: 10.1109/TMM.2023.3293736.
- [24] GHARBI M, CHEN Jiawen, BARRON J T, *et al.* Deep bilateral learning for real-time image enhancement[J]. *ACM Transactions on Graphics (TOG)*, 2017, 36(4): 118.

乔成平: 男, 硕士生, 研究方向为微光场景下目标检测。

金佳莹: 男, 硕士生, 研究方向为基于偏振视觉的目标检测。

张俊超: 男, 副教授/博导, 研究方向为光电信息处理、图像处理、模式识别和机器学习。

朱政亮: 男, 副教授, 研究方向为水声信息处理、光电信息处理。

曹祥旭: 男, 硕士生, 研究方向为医学图像处理。

责任编辑: 陈 倩

Low-Light Object Detection via Joint Image Enhancement and Feature Adaptation

QIAO Chengping^{①②} JIN Jiakun^{①②} ZHANG Junchao^{①②} ZHU Zhengliang^③
CAO Xiangxu^①

^①(School of Automation, Central South University, Changsha 410083, China)

^②(Hunan Provincial Key Laboratory of Optic-Electronic Intelligent Measurement and Control, Changsha 410083, China)

^③(College of Marine Information Engineering, Jimei University, Xiamen 361021, China)

Abstract:

Objective Object detection has advanced significantly under normal lighting conditions, supported by numerous high-accuracy, high-speed deep learning algorithms. However, in low-light environments, images exhibit reduced brightness, weak contrast, and severe noise interference, leading to blurred object edges and loss of color information, which substantially degrades detection accuracy. To address this challenge, this study proposes an end-to-end low-light object detection algorithm that balances detection accuracy with real-time performance. Specifically, an end-to-end network is designed to enhance feature quality and improve detection accuracy in real time under low-light conditions.

Methods To improve object detection performance under low-light conditions while maintaining detection accuracy and real-time processing, this study proposes an end-to-end low-light image object detection method. Detection accuracy is enhanced through joint learning of image enhancement and feature adaptation, with the overall network structure shown in Fig. 1. First, a data augmentation module synthesizes low-light images from normal-light images. The paired normal-light and low-light images are mixed using the MixUp function provided by YOLOv5 to generate the final low-light images. These synthesized images are input into the low-light image enhancement module. In parallel, the matched normal-light images are provided as supervision to train the image enhancement network. Subsequently, both the enhanced low-light images and the corresponding normal-light images are fed into the object detection module. After processing through the YOLOv5 backbone, a matching loss is computed to guide feature adaptation.

Result and Discussions The experiments are conducted primarily on the Polar3000 and LLVIP datasets. Fig. 3 presents the detection results obtained using YOLOv5 with different image enhancement methods applied to the Polar3000 dataset. Most existing methods tend to misclassify overexposed regions as bright Unmanned Aerial Vehicles (UAVs). In contrast, the proposed method demonstrates accurate object detection in low-light conditions without misidentifying overexposed areas as UAVs (Fig. 3). Furthermore, the detection performance of the proposed method, termed MAET, is compared with that of a standalone YOLOv5 model. Quantitative experiments show that the proposed method outperforms both image-enhancement-first detection pipelines and recent low-light object detection methods across both experimental groups A and B, irrespective of low-light fine-tuning. On the LLVIP dataset, the proposed method achieves a detection accuracy of 91.7% (Table 1), while on the Polar3000 dataset, it achieves 92.3% (Table 2). The model also demonstrates superior generalization performance on the ExDark and DarkFace datasets (Tables 4 and 3). Additionally, compared to the baseline YOLOv5 model, the proposed method increases parameter size by only 2.5% while maintaining real-time detection speed (Table 5).

Conclusions This study proposes a low-light object detection method based on joint learning of image enhancement and feature adaptation. The approach simultaneously optimizes image enhancement loss, feature matching loss, and object detection loss within an end-to-end framework. It improves image illumination, preserves fine details, and aligns the features of enhanced images with those acquired under normal lighting conditions, enabling high-precision object detection in low-light environments. Comparative experiments on the LLVIP and Polar3000 datasets demonstrate that the proposed method achieves improved detection accuracy while maintaining real-time performance. Furthermore, the method achieves the best generalization results on the ExDark and DarkFace datasets. Future work will explore low-light object detection based on multimodal data fusion of visible and infrared images to further enhance detection performance in extremely dark conditions.

Key words: Object detection; Low-light images; Feature matching; Image enhancement network