

基于ICFIE-YOLO的低照度图像目标检测方法

秦嘉奇, 江泽涛*, 雷晓春

(桂林电子科技大学广西图像图形与智能处理重点实验室, 广西桂林 541004)

摘要: 低照度环境下获取的图像往往亮度低、对比度低、光照不均匀,从而造成图像特征变弱及模糊难于提取,同时在有限提取的特征中也存在大量噪声信息,导致目标难于检测识别,因而现有低照度目标检测成果极少. 针对低照度目标特征难于提取及特征空间噪声大的问题,本文提出一种基于光照矫正与特征交互增强(Illumination Correction and Feature Interacted Enhancement, ICFIE-YOLO)网络的低照度目标检测方法. 该方法首先利用提出的ICFIE-YOLO内部多尺度光照矫正网络(Multi Scale Illumination Correction Network, MSICN)对低照度图像进行光照矫正,突出隐藏在图像背景中目标的模糊特征,使特征提取模块能更好地提取到目标特征;其次,为充分利用有效特征信息,过滤特征图中的噪声干扰,提出特征交互增强(Feature Interacted Enhancement, FIE)检测头,通过特征注意力交互方式实现特征增强,建立低照度图像中各个区域特征之间的空间关联和语义关联,从而抑制噪声对有效特征的干扰,实现降噪效果;最后,在增强特征及去除噪声的基础上用改进的检测头实现高精度目标检测. 在ExDark和DarkFace数据集上的实验表明,所提方法较主流目标检测方法mAP提高2.1个百分点以上,较现有低照度目标检测方法召回率提高4.2个百分点以上,同时召回率较基线模型提高了2.6个百分点,所提方法具有较好的泛化性.

关键词: 目标检测;低照度;光照矫正;特征去噪;特征增强

基金项目: 国家自然科学基金(No.62172118);广西自然科学基金(No.2021GXNSFDA196002);广西图像图形智能处理重点实验室项目(No.GIIP2302, No.GIIP2303, No.GIIP2304);研究生创新基金(No.2024YCXB09, No.2024YCX039)

中图分类号: TP391

文献标识码: A

文章编号: 0372-2112(2025)02-0514-13

电子学报URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20240648

Low Illumination Image Object Detection Method Based on ICFIE-YOLO

QIN Jia-qi, JIANG Ze-tao*, LEI Xiao-chun

(Guangxi Key Laboratory of Image and Graphic Intelligent Processing, Guilin University of Electronic Technology, Guilin, Guangxi 541004, China)

Abstract: Images obtained in low light environments often have low brightness, low contrast, and uneven lighting, resulting in weakened and blurred image features that are difficult to extract. At the same time, there is also a large amount of noise information in the limited extracted features, making it difficult to detect and recognize objects. Therefore, there are very few existing low light object detection results. This paper proposes a low illumination object detection method based on the Illumination Correction and Feature Interaction Enhancement (ICFIE-YOLO) network to address the difficulties in extracting features from low illumination objects and the large noise in the feature space. This method first utilizes the proposed ICFIE-YOLO internal Multi Scale Illumination Correction Network (MSICN) to correct low illumination images, highlighting the blurry features of objects hidden in the image's background, and enabling the feature extraction module to better extract object features. Secondly, to fully utilize effective feature information and filter out noise interference in feature maps, a Feature Interacted Enhancement (FIE) detection head is proposed. Through feature attention interaction, feature enhancement is achieved, establishing spatial and semantic correlations between features in different regions of low illumination images, thereby suppressing the interference of noise on effective features and achieving feature enhancement. Finally, on the basis of enhancing features and removing noise, an improved detection head is used to achieve high-precision object detection. Experiments on the ExDark and DarkFace datasets show that the proposed Method improves mAP by over

2.1 percentage points compared to mainstream object detection models, increases recall by over 4.2 percentage points compared to existing low light object detection Methods, and improves recall by 2.6 percentage points compared to baseline models. The proposed Method has good generalization performance.

Key words: object detection; low illumination; light correction; feature denoising; feature enhancement

Foundation Item(s): National Natural Science Foundation of China (No.62172118); Natural Science Foundation of Guangxi Province (No.2021GXNSFDA196002); Guangxi Key Laboratory of Image and Graphic Intelligent Processing (No.GIIP2302, No.GIIP2303, No.GIIP2304); Innovation Project of Guangxi Graduate Education (No.2024YCXB09, No.2024YCX039)

1 引言

目标检测的任务是找出图像中所有感兴趣的物体,确定它们的类别和位置是计算机视觉领域的核心问题之一。由于各类物体有不同的外观、形状和姿态,加上成像时距离、光照、遮挡等因素的干扰,目标检测一直是计算机视觉领域具有挑战性的问题。在该领域,使用深度学习方法进行目标检测已经成为主流。具有代表性的是两阶段目标检测算法 Faster-RCNN 及其衍生模型^[1-5],以及单阶段的 YOLO、SSD 目标检测框架及其衍生框架^[6-10]。随着 Transformer 结构的提出和发展,人们开始探索将此结构用于目标检测任务中^[11]。

现有的目标检测方法取得了较好的效果并得到广泛应用,但现有的方法大都基于正常照度环境进行目标检测,而低照度环境下的目标检测研究成果却很少,将现有的正常照度环境下的目标检测方法移植到低照度环境中效果不理想。现实生活中具有很多低照度环境下进行目标检测的实际需要,如夜间监控、夜间无人机侦察与攻击、夜间汽车自动驾驶等工作场景,因此低照度场景下的目标检测具有重要的实用价值。

在基于深度学习的低照度目标检测方法中,可分为先对图像增强再进行检测以及级联训练增强网络和检测器两类方法,如图 1(a)和图 1(b)所示。为实现低照度图像目标检测,Loh 等人^[12]数字上标提出了低照度图像数据集 ExDark,并指出提高低照度图像目标检测性能的两个方向是去噪以及提升亮度。在此基础上,Yuan 等人^[13]提出面向目标检测的即插即用低光图像增强方法 LLIE,实验结果表明该方法对低照度图像目标检测性能有一定提升,但是其增强方式过于简单,没有考虑到环境光的复杂性,因此提升有限。Xue 等人^[14]在 DETR 检测器前加入多尺度金字塔增强网络 MPE,提升低照度图像目标检测的性能。上述方法将低照度目标检测拆分成增强和检测两个步骤,增强模型与检测模型独立训练,因此无法保证增强结果有助于检测网络,很难获得较好的性能。级联训练增强网络和检测模型的方法在一定程度上保证了图像的增强有利于目标检测精度的提高。Chen 等人^[15]提出 YOLO in the Dark 通过域适应的方法来合并多个模型,使得新的模

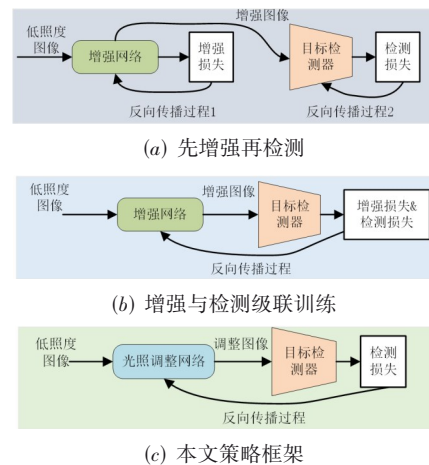


图1 低照度目标检测不同框架策略示意图

型可以适应低照度图像上的目标检测任务,但域适应方法仍然面临成对数据集获取困难的问题。Yin 等人^[16]提出 PE-YOLO,通过在检测器之前加入金字塔增强网络并与检测器级联训练,在低照度数据集上取得了较好的效果;Cui 等人^[17]利用多任务自动编码变换对低照度图像的光照变换进行编码和解码,辅助特征提取网络学习内在的低照度图像退化过程,从而提升检测器的性能;江泽涛等人^[18]首先利用一个像素映射提升低照度图像的显著性,然后对主干网络的输出特征进行增强,最终检测器取得了较好的检测性能。上述方法利用增强模型和检测模型的损失函数相加共同训练,在一定程度上保证了图像增强有助于目标检测精度的提高,但是总体损失函数中仍然存在图像增强的部分损失,导致增强网络的优化方向并不完全有助于目标检测任务,部分基于重建的增强网络会导致待检测图像的信息丢失,检测性能提升有限。为解决上述问题,本文所提低照度目标检测方法使用与上述方法不同的框架策略,如图 1(c)所示。为约束图像增强网络对低照度图像的增强过程,防止欠拟合导致对图像进行错误的处理,对后续检测产生不利影响,因此图 1(b)方法需要加入图像增强损失进行约束。但若是加上增强损失,则根据损失函数的选取和设计,可能需要成对的数据集,并且数据集分布和损失函数都会影响检测器的优化,导致检测方法泛化性较差;根据反向传播优化权重的

方式,图像增强网络会受到检测损失和增强损失两者的调整,造成检测网络和图像增强网络优化方向不一致,不利于检测任务的性能提升.图1(c)方法将图像光照调整网络和检测器集成在一起,光照调整网络使用检测损失进行优化,其优化方向完全由检测器决定,统一了检测器和光照调整网络的优化方向.由于不需要加入增强损失约束,因此图1(c)方法降低了对数据集的依赖,具有更好的泛化性.但是,为了解决光照调整网络的优化问题,防止无增强损失约束导致欠拟合,光照调整网络的结构和计算过程需要特别设计,并使用基于检测损失对齐的训练方法进行训练.除上述3种模型策略外,江泽涛等人^[19]将低照度目标检测视为一种困难目标检测任务,将注意力机制、多尺度特征感知引入正常照度目标检测框架中,提高原有目标检测框架对特征的表达能力,从而提高目标检测精度.

受局部光照影响,低照度图像的特征提取要比正常照度图像困难,因此本文在低照度目标检测器中设计了一个利用检测损失进行训练且能够无参考调整图像光照的结构.另外,由于低照度图像中包含大量的背景信息,会导致提取到的特征空间中存在大量无用背景特征和噪声特征,严重干扰目标检测模型对目标分类和定位,因此在特征输出阶段引入特征交互注意力机制以对有效特征进行增强.

综上所述,本文提出ICFIE-YOLO目标检测方法.本文方法具有以下特点:

(1)为解决低照度图像特征模糊难以提取和特征空间噪声较多导致检测精度不高的问题,提出基于多尺度光照矫正和特征交互增强的ICFIE-YOLO低照度图像目标检测方法.该方法首先利用提出的多尺度光照矫正网络(Multi Scale Illumination Correction Network, MSICN)实现对低照度图像的光照矫正以利于目标特征提取过程;然后使用特征交互增强(Feature Interacted Enhancement, FIE)检测头滤除特征图中的噪声,实现特征清洗和增强,最终达到较高的检测精度.在ExDark数据集和DarkFace数据集上的实验表明,所提方法在检测精度和召回率上较其他现有方法均具有更好表现,能够有效提高低照度图像目标检测性能.

(2)MSICN对低照度图像成像时全局光照弱、局部光源不均匀等不利光照条件进行光照矫正.图像经过光照矫正后,图像中目标的边缘、色彩、纹理得到增强,突出原本隐藏在背景中的模糊目标,避免有用目标特征消失在黑暗的背景中,图像中的目标特征能够更容易被目标检测器的主干网络提取,解决低照度环境带来的特征提取困难问题.结合基于检测损失对齐的训练方式,统一MSICN与检测器的优化方向,最大可能减少对目标检测不利的调整,保证矫正后图像像素对于

目标检测器的有效性.

(3)为充分利用有效特征信息,过滤特征中的噪声,提出特征交互增强FIE. FIE工作在检测器的特征编码端,利用交互注意力机制分别对图像特征的空间上下文关联和语义关联进行建模表达,从而达到抑制背景和噪声特征、强化目标特征的目的.

2 相关工作

2.1 目标检测中引入图像增强方法

为实现高质量的低照度目标检测方法,一个典型的思路是将图像先进行增强,再送入目标检测模型得到检测结果.这一思路可以灵活运用现有的图像增强方法和目标检测方法,在一定程度上解决低照度图像整体亮度不足、光照不均匀、对比度低、信噪比低等问题.目前有很多低照度图像增强算法被提出.早期一般使用基于直方图均衡化的方法,该方法通过扩大图像通道像素值之间的差值提升图像对比度;基于色彩恒常(Retinex)理论,学者提出MSRCR(Multi-Scale Retinex with Color Restoration)方法,该方法利用多尺度高斯核卷积计算光照分量,从而解出原始图像^[20];基于大气散射模型理论,Dong等人^[21]直接用去雾算法对低光照图像的反转图像处理,经过去雾后再取反得到低照度增强结果.由于上述传统图像增强方法均基于人为设计的模型对增强参数进行估计,并不能很好拟合低照度图像到正常照度图像的映射关系,加上深度学习的快速发展,很快就有学者将神经网络用于低照度图像增强.基于深度学习的图像增强将图像增强视为从低照度图像到正常照度图像的一种映射学习任务,通过对卷积神经网络(Convolutional Neural Network, CNN)进行合理建模就能得到低照度图像增强网络^[22-24];利用CNN对Retinex理论中的未知参数进行估计,则可以得到基于Retinex理论的低照度图像增强模型^[25,26];Jiang等人^[27,28]则通过深度学习模型估计图像增强所需的曲线映射参数,提出无须成对图像的图像增强方法.

现有低照度图像增强算法较好实现了低照度图像到正常光照图像的转换,但在“先增强,再检测”的低照度目标检测框架中无法保证图像增强的形式有利于后续的目标检测任务,增强任务和检测任务的优化方向并不一致.尽管提升图像亮度确实一定程度上可以提升低照度图像目标检测效果,但是这种调整应当是以提升目标检测精度为目的的调整,而以增强为目的的低照度图像调整会导致增强后的图像出现色彩畸变、目标特征变形,从而导致待检测图像特征信息出现损坏,被同步放大的噪声和背景信息也会对目标检测任务造成干扰,所以直接将增强后的图像送入目标检测网络很难取得较好的检测精度,文献[16]的研究证

实了此观点。

2.2 目标检测框架

随着深度学习的发展,基于深度学习的目标检测方法因其优异的性能已经得到广泛的使用。所有基于深度学习的目标检测方法均可划分为特征提取和分类回归两个部分。如图 2 所示,首先利用特征提取网络提取图像特征,然后再经过编码输出分类和位置信息,其中特征提取器、编码输出均使用 CNN 或者 MLP 实现。

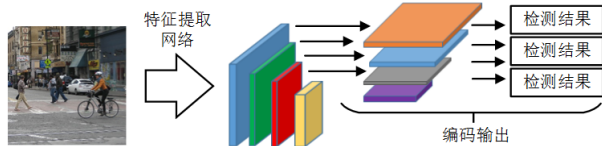


图2 目标检测通用框架

目前经典的目标检测方法主要包括单阶段和两阶段方法。单阶段目标检测方法将特征提取、目标定位、分类集成在一个神经网络模型上,其速度比两阶段的方法快,但是在精度上略微欠佳。两阶段目标检测方法在第一阶段使用一个区域推荐网络(Region Proposal Network, RPN)得到可能存在目标的区域,然后再将此区域的特征送入分类和回归网络得到检测结果,这导致两阶段目标检测方法优劣势恰好与单阶段方法相反,速度较慢,但是精度较高。两阶段目标检测在曾经的一段时间内精度普遍优于单阶段算法,但是随着网络模型结构的优化和多尺度特征金字塔结构的提出与优化^[29],目前单阶段目标检测方法在速度、精度上都比两阶段目标检测方法有优势。单阶段目标检测方法主要思想是首先利用深度网络提取图像特征,然后使用深度神经网络将特征编码为目标类别、目标大小和位置、目标置信度等信息,再据此设计损失函数进行模型训练。单阶段目标检测方法将特征提取、目标分类、位置定位等过程设计成一个整体,中间无须其他人工干预,因此可以实现端到端的训练和预测。

对于不同的目标检测方法,其主要区别是特征提取部分结构不一样或者输出编码部分不一样,有大量的改进模型均基于上述两点思想。上述模型及其衍生、改进模型已经被应用于人脸检测、自动驾驶、智能监控、医学图像检测等领域,目标检测方法及相关应用逐渐成熟,但面对低照度图像目标检测仍然存在许多挑战。

3 本文方法

3.1 ICFIE-YOLO 总体设计

由于低照度图像目标特征难于提取且提取到的特

征中存在大量噪声,因此正常照度目标检测方法在低照度图像上达不到理想精度。为实现更高效的低照度目标检测器,本文提出一种 ICFIE-YOLO 低照度目标检测方法。ICFIE-YOLO 使用 YOLOv7 主干网络作为特征提取网络,通过引入多尺度光照矫正网络(MSICN)和特征交互增强(FIE)检测头,实现更加高效的特征提取和特征编码,其结构如图 3 所示。

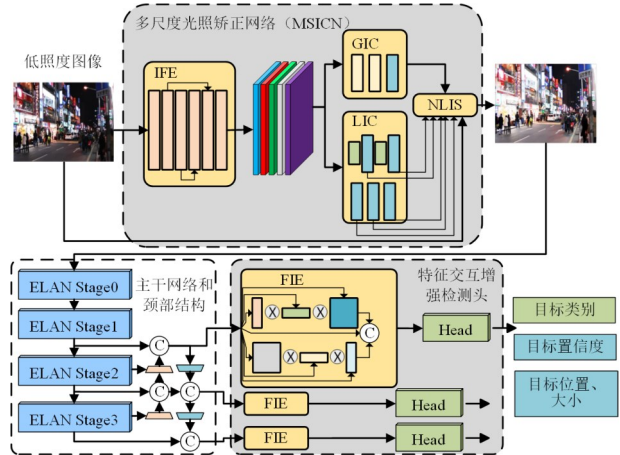


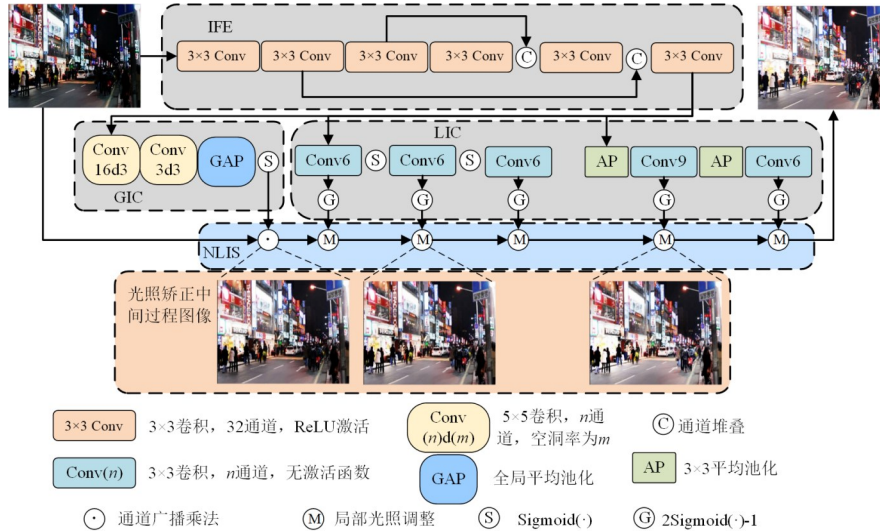
图3 ICFIE-YOLO 整体结构

针对低照度图像中目标和背景模糊造成特征难以提取以及图像特征中含有大量噪声的现状,使目标检测模型的主干网络更容易提取低照度图像中的特征,设计 MSICN 对低照度图像进行光照矫正。MSICN 网络包含光照特征提取模块 IFE (Illumination Feature Extraction)、全局光照矫正模块 GIC (Global Illumination Correction)、局部光照矫正模块 LIC (Local Illumination Correction) 和非线性光照堆叠 NLIS (Non-Linear Illumination Stacking)。由于一幅低照度图像中存在一个较弱的全局光源和较少的、影响范围不同的局部光源,因此 IFE 模块通过多层级、多尺度卷积操作提取低照度图像的光照信息,然后通过 GIC 和 LIC 利用卷积感受野堆叠增长的性质将此光照信息转换为具有多尺度性质的光照矫正系数矩阵。光照矫正系数矩阵指示了图像中全局光照和局部光照的矫正关系,NLIS 通过堆叠的乘法操作将光照矫正系数矩阵应用到低照度图像中,实现图像全局光照和多个局部光照的调整。利用基于检测损失的对齐训练方式仅使用目标检测损失对 MSICN 中的参数进行调整,补偿低照度图像特征提取过程中暗处目标损失的特征,从而使目标检测模型主干网络更容易提取到有效目标特征。低照度图像中还存在较多噪声,为抑制特征图中噪声对目标特征造成的影响,提高检测模型对低照度图像中目标的召回率和分类精度,在检测模型特征输出部分设计特征交互增强检测头 FIE。FIE 使用特征的内部关联建

立低照度图像各个区域中目标特征之间的空间关联和语义关联,强化对检测结果具有更多正面影响的特征,同时过滤特征中的噪声,减少低照度区域特征造成的干扰。

3.2 多尺度光照矫正网络

MSICN 对低照度图像光照进行矫正,使黑暗背景中的目标凸显出来,从而解决低照度图像特征难以提取的问题。MSICN 结构如图 4 所示。



MSICN 使用基于检测损失的对齐训练保证优化方向与检测器一致,因此能够根据检测器的需要对低照度图像进行调整,调整的目的在于提高检测精度。由于低照度图像存在全局光照不足和局部光照不均的问题,MSICN 需要对图像的全局光照和局部光照做调整。其中,GIC 负责生成全局光照矫正所需的参数,LIC 生成局部光照矫正所需的参数。为了让 GIC 和 LIC 输出的光照矫正参数作用到低照度图像中,并且保证输出的图像像素具有有效性,光照矫正图像输出最终由 NLIS 使用光照矫正参数对低照度图像经过若干次堆叠的光照曲线调整实现。如图 4 中光照矫正中间过程图像所示,GIC 输出的参数实现对图像全局光照和色调的矫正,LIC 输出的参数则对图像中光照不均匀的部分进行调整,最终得到有助于后续检测任务的图像。

为给光照矫正过程提供包含多个尺度的公用特征,同时考虑到光照特征信息所需感受野较大,IFE 使用少量小尺寸卷积多层次递进提取图像的光照特征。IFE 由 6 个卷积核大小为 3×3 、步幅为 1 的卷积层堆叠而成,并使用 ReLU 激活函数保证输出非负。连续的卷积层可以提供感受野从小到大的特征图,对应不同尺度的光照特征,同时为了保证后续计算的多尺度特性,IFE 还将具有不同感受野的不同层级的多个特征图进行通道维度拼接,最终得到 32 通道的光照特征。

GIC 模块通过对彩色图像 RGB 的 3 个通道分别做曲线变换,调整低照度图像得到矫正图像。3 个通道分别利用不同的全局光照系数做曲线变换实现全局光照

矫正。因为低照度环境下环境光照中 R、G、B 颜色分量是不均衡的,为了将全局光照矫正为对目标检测影响最小的白色,对于 R、G、B 这三个通道要做不同的调整。由于是对全局光照进行修正,因此该单元利用大尺寸的空洞卷积来获取更大的感受野,同时考虑到全局光照的拟合曲线不会像局部光照那样过于复杂,所以只使用一个浅层卷积网络来拟合全局光照的调整。GIC 使用光照特征作为输入,经过 2 个 5×5 大小空洞率为 3 的卷积层得到全局光照特征。全局光照特征经过一个全局平均池化和 Sigmoid 函数输出为全局光照矫正系数 k ,其形状为 $1 \times 1 \times 3$,且其值在 $(0, 1)$ 之间。

GIC 提供的参数仅能矫正全局光照,图像仍然存在局部亮度过高或者过低、对比度过低的问题。因此,LIC 利用光照特征对图像局部光照进行估计,其输出是一个与输入图像同维度的局部光照矫正矩阵。在具有多个局部光源的情况下,多个不同尺度、亮度的局部光照不能像全局光照那样使用一个二次曲线拟合得到,而是需要更加高阶的曲线进行拟合。考虑到局部光照估计的结果是多个不同尺度、亮度的光源照射估计的叠加,且局部光照存在尺度、亮度两个维度上的不同,因此 LIC 通过多尺度卷积结构拟合不同范围大小的局部光源对图像造成的影响,同时通过多个卷积组的叠加来拟合多个不同亮度但大小相同的局部光源。LIC 分为非下采样光照矫正和下采样光照矫正两个部分,非下采样光照矫正使用连续 3×3 卷积对局部光照调整系数进行计算,可以拟合作用范围较小的局部光源;而下采

非光照矫正则用于拟合作用范围较大的局部光源,因此在模型更深层次以获取更大的感受野.考虑到大范围局部光源之间可能具有相关性,因此利用池化核为3、步幅为1的平均池化模拟光源之间的叠加.为了保证局部光照矫正能够拟合复杂的局部光照矫正过程所需参数,LIC一共输出30通道的局部光照矫正系数,这些光照矫正系数分为10组,通过迭代运算相当于给每个通道都使用二次曲线调整10次,光照矫正系数值域为(-1,1),可以拟合出几乎覆盖值域(0,1)的所有调整曲线.局部光照矫正系数经过激活函数“2Sigmoid(x)-1”输出,保证其值在(-1,1)之间,从而保证后续非线性光照堆叠计算的有效性.

NLIS模块中没有可训练参数,其作用是将LIC和GIC中计算得到的光照矫正系数应用到输入图像中,实现最终的光照矫正.NLIS的工作原理是将光照调整系数设置为一个像素映射曲线的参数,且该像素映射曲线需要满足如下要求:

- (1)为保证应用该映射后图像像素值之间的大小关系不变,因此该曲线在[0,1]区间必须是单调递增的;
- (2)要提升图像整体亮度,该曲线在区间[0,1]必须处于直线 $y=x$ 的上方;
- (3)为保证图像像素值的有效性,该曲线的值域应为[0,1].

为满足上述要求,本文使用一个经过原点的二次曲线来表示像素映射曲线.因此,GIC输出的全局光照矫正计算过程可由式(1)表示.

$$I_{g_out} = I_{in} \cdot (-k \cdot I_{in} + (1+k)), \quad k \in (0,1) \quad (1)$$

其中, I_{g_out} 表示输出图像; I_{in} 表示输入图像; k 是有3个元素的一维向量,表示图像3个通道的全局光照矫正系数. k 是经过Sigmoid激活函数输出,因此 $k \in (0,1)$,对于 $I_{in} \in (0,1)$.式(1)是一个经过原点和点(1,1)的二次曲线,且曲线在(0,1)上均处于直线 $y=x$ 上方,因此可以提高输入图像的通道灰度值,同时保证像素值仍然处于[0,1]之间.由于不同图像和通道所需要增强的程度不同,即需要不同的增强曲线,神经网络模型可以通过调整 k 的值得到不同的增强曲线.

对于局部光照调整系数,NLIS将LIC中得到的10组系数用于计算局部光照矫正,其计算过程和计算公式如式(2).

$$\begin{cases} I_{out1} = I_{g_out} \cdot (k_1 \cdot I_{g_out} + (1-k_1)) \\ I_{out2} = I_{out1} \cdot (k_2 \cdot I_{out1} + (1-k_2)) \\ \dots\dots\dots \\ I_{outn} = I_{out(n-1)} \cdot (k_n \cdot I_{out(n-1)} + (1-k_n)) \end{cases}, \quad n = 10 \quad (2)$$

其中, $I_{out(i)}$ 表示LIC模块第*i*个局部光照矫正单元的输入;该模块的输入是GIC的输出 I_{g_out} ; k_i 表示由网络参

数估计得到的局部光照矫正系数,是与输入图像同维度的矩阵.局部光照存在需要增强或者减弱两种情况,因此 k_i 的值需要规定化到(-1,1).本文使用激活函数“2Sigmoid(x)-1”输出,显然 $k_i \in (-1,1)$,当 $k_i > 0$ 时起到增强局部光照作用,当 $k_i < 0$ 时起到减弱局部光照作用.由于局部光照较全局光照复杂,使用一个二次曲线拟合局部光照的影响是不够的,因此MSICN将10组局部光照矫正参数以迭代方式进行堆叠.此种方式可以拟合出一个高幂次的调整曲线以应对复杂的局部光照矫正.

在目标检测任务中,目标检测的损失函数是衡量网络工作性能的唯一量指标,因此MSICN中的参数均由检测损失反向传播得到.MSICN保证了输出图像的有效性(即输出值在0~1之间),同时给目标检测网络提供自适应光照调整的能力,有效提高低照度图像特征提取效率,从而提升低照度目标检测的性能.

3.3 特征交互增强模块

尽管MSICN在一定程度上缓解低照度图像特征提取困难的问题,但是低照度图像中提取到的特征仍然不可避免地存在噪声干扰,这会导致目标特征混杂在噪声特征中,最终导致目标检测精度下降.目标特征不够显著还会导致目标分类错误.为解决这一问题,本文提出特征交互增强检测头FIE.受自注意力机制能够建立长程依赖关系的启发,FIE通过建立特征图中通道之间、像素之间的交互注意力关联,将这种关联以注意力机制的形式保存,从而获得低照度图像各个区域中目标特征之间的空间关联和语义关联.FIE结构如图5所示.为捕获特征图不同像素之间的特征关联以及同一像素内部的语义关联,FIE具有空间特征交互增强和通道特征交互增强两个分支.

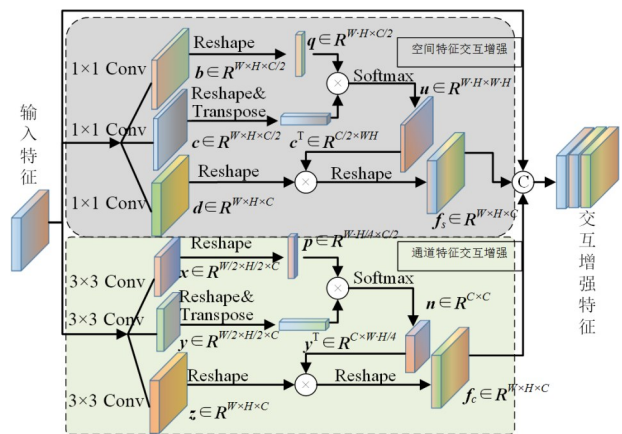


图5 FIE结构图

特征图上的每一个像素都是对邻近区域中目标特征的一种表述,在低照度环境下,目标特征、噪声特征和背景特征混淆在一起,对目标检测器的最终输出造

成不利影响. 空间特征交互增强分支对特征图上的像素进行建模, 通过建立特征图像素之间的相关性, 增强有用特征并弱化背景和噪声特征, 从而实现特征增强. 在空间特征交互增强分支中, 通过计算特征图中像素之间的关联性, 得到描述像素之间关联的特征图, 使隐藏在背景和噪声中的特征得到增强, 从而提高目标检测的性能. 如图4所示, 空间特征交互增强分支将输入特征图经过3个 1×1 的卷积得到输出**b**、**c**、**d**, 其中**b**、**c**的输出通道数为输入通道数的一半, **d**的通道数与输入一致. 接下来, 将**b**逐像素展开得到**q**, **c**展开后并转置得到**c^T**, **q**与**c^T**相乘并经过Softmax函数得到特征图**u**. 特征图**u** $\in \mathbf{R}^{W \times H \times W \times H}$ 对输入特征中每个像素之间的关系进行描述, 通过Softmax过滤掉背景和噪声对目标特征造成的干扰. 为将空间特征关联编码进特征图中, 将**u**与**d**逐像素展开后的矩阵相乘并调整维度得到**f_s** $\in \mathbf{R}^{W \times H \times C}$. 该过程如式(3)所示.

$$\mathbf{f}_s = \text{Re}(\text{Soft max}(\mathbf{q} \otimes \mathbf{c}^T) \otimes \text{Flattening}(\mathbf{d})) \quad (3)$$

其中, \otimes 表示矩阵乘法, $\text{Flattening}(\cdot)$ 表示逐像素展开, $\text{Re}(\cdot)$ 表示将维度调整到与输入特征图一致. 特征图**f_s**对原始特征中的空间上下文信息进行编码, 能够对重要目标特征区域进行增强, 并且抑制背景和噪声带来的干扰特征.

特征图中每个通道都描述了目标特征的一个方面, 通道中含有越多的有效特征用于描述目标, 那么目标检测器就能够达到越好的性能. 通道特征交互增强分支通过对特征图中的通道之间进行建模, 从而建立特征通道之间的语义关联, 以此达到特征增强的目的. 该分支的工作原理与空间特征增强分支类似, 首先利用3个并行的 3×3 卷积对特征图尺寸进行压缩, 得到特征图**x**、**y**、**z**; 然后将**y**逐像素展开并转置与**x**逐像素展开的矩阵**p**相乘, 再经过Softmax函数得到特征图**n**. 特征图**n** $\in \mathbf{R}^{C \times C}$ 包含了输入特征图中**C**个通道之间的关联, 然后将此关联通过矩阵乘法的形式编码到输出特征中, 该过程如公式(4)所述.

$$\mathbf{f}_c = \text{Re}(\text{Flattening}(\mathbf{z}) \otimes \text{Soft max}(\mathbf{y}^T \otimes \mathbf{p})) \quad (4)$$

由于**b**和**c**的通道数决定了参与计算像素关联程度的向量长度, 参与计算的向量长度越小则该向量中每个值对像素关联的影响越大, 因此通过通道数减半的形式筛选出输入特征中较为重要的部分, 得到更为显著的空间像素关联描述. 在空间交互特征增强分支中, 计算空间交互特征的特征图**b**和**c**的通道数为输入特征通道数的一半, 其目的是强制对输入特征进行像素上的筛选, 从而达到抑制背景和噪声特征的作用. 同样, **x**和**y**的长度和宽度决定参与计算通道关联程度的向量长度, 为了实现特征通道的筛选, 抑制语义信息较

弱的无效特征, 通道交互特征增强分支中的特征图**x**和**y**的长宽均为输入特征图的1/2. 为合并空间交互特征**f_s**和通道交互特征**f_c**, FIE最终将输入特征**f**与**f_s**、**f_c**沿通道维度拼接得到交互增强特征.

3.4 基于检测损失的对齐训练

为保证MSICN与目标检测器优化方向一致, MSICN应当使用检测损失作为其参数调整依据, 因此提出使用基于检测损失的对齐训练方法对MSICN进行训练. 该方法通过正常照度数据集和低照度数据集图像(不要求成对)在目标检测器上的检测损失进行同方向优化训练, 使MSICN学习到两者之间的光照矫正关系, 如图6所示.

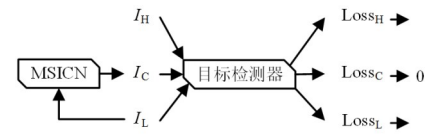


图6 基于检测损失的对齐训练示意图

基于检测损失的对齐训练过程如下. 首先在正常照度数据集上(例如MS COCO)训练得到一个目标检测器. 该检测器能够很好地提取正常照度图像的特征并将特征解码为目标检测输出, 记该检测器在上的损失值为Loss_H. 由图6可知, 若将低照度图像送入该目标检测器, 由于低照度图像低质特性, 在正常照度目标检测器上得到的损失值Loss_L将会变大. 将低照度图像经过MSICN处理后得到光照矫正图像, 将送入目标检测器后, 其损失值应当也大于Loss_H, 但是MSICN中存在可学习参数, 若将目标检测器部分的参数固定与MSICN进行联合训练, 最终三者损失值关系应当为Loss_L > Loss_C > Loss_H. 由于对应正常照度图像无法获取, 在实际计算中无法得到Loss_H, 因此无法直接通过优化损失值的距离对MSICN进行优化. 但是上述损失均向检测损失趋于0的方向进行优化, 此时固定目标检测器的参数仅仅使用Loss_C优化MSICN参数会使MSICN将低照度图像向有利于正常照度目标检测器特征提取和解码的方向进行优化, 从而提升低照度目标检测的精度. 这是MSICN与一般图像增强网络最大的不同, 它以目标检测为优化目标, 并非提升图像质量.

MSICN与目标检测网络作为一个整体共同训练, 由于低照度目标检测数据集中图像数量、实例数量较少, 因此目标检测的主干网络一般使用在大数据集上预训练好的参数. 在MSICN学习到从低照度图像到正常照度图像的潜在光照矫正关系后, 再解冻目标检测器部分的参数, 对全模型参数进行微调训练.

4 实验分析

4.1 实验环境基本参数

本文实验在通用低照度目标检测数据集 ExDark^[12]和低照度人脸检测数据集 DarkFace^[30]上进行. ExDark 数据集有 12 个类别的 7 363 张低光照图像. 实验中将数据集以 8:2 的比例划分成训练集和测试集, 即 5 890 张图像作为训练集, 1 473 张图像作为测试集; DarkFace 数据集包含了 6 000 张图像, 由于图像总量较小, 为保证训练效果, 将其中 5 400 张作为训练集, 600 张图像作为测试集. 实验模型均使用 Tensorflow 深度学习框架搭建, 运行在具有 11 GB 显存的 GTX1080Ti 图形处理器上.

为验证本文方法的有效性, 将本文与目前主流的目标检测方法以及低照度目标检测方法进行对比. 由于数据集训练图像较少, 因此所有方法的特征提取网络均使用 MS COCO 数据集上预训练得到的参数作为初始化参数. 在训练过程中, 先加载预训练参数并冻结特征提取部分网络参数, 所有模型训练均使用 Adam 优化器, 学习率初始值为 10^{-2} , 学习率最低为 10^{-5} .

4.2 ICFIE-YOLO 的检测性能

本文在 ExDark 数据集上实验验证 ICFIE-YOLO 目标检测方法的有效性, 将其与主流的正常光照目标检测方法、低照度目标检测方法以及结合了新近优秀图像增强算法的 YOLOv7 检测器 3 类方法进行对比. 其中, Retinexformer^[31]、HVI-CIDNet^[32]、NeRCO^[33] 都是近年来在图像增强任务上获得较好效果的图像增强方法, 按照先增强再检测的策略得到 Retinexformer+YOLOv7、HVICIDNet+YOLOv7、NeRCO+YOLOv7 方法. 由于上述方法都需要明暗成对的数据集, 因此无法在目标检测数据集上以联合检测和增强的形式进行组合. Zero-DCE^[28]是近年来较为优秀的无参考低光图像检测方法, 将其与 YOLOv7 检测器分别以先增强再检测和联合增强与检测的方法进行组合得到 Zero-DCE+YOLOv7 和 Zero-DCE-YOLOv7 方法. 除 EfficientDET 使用 512 像素×512 像素图像作为输入, 其余方法输入图像大小均为 416 像素×416 像素. 表 1 所示为本文方法与上述其他方法在 Exdark 数据集上各类别平均精度 (AP) 和 mAP 的比较结果.

表 1 在 ExDark 数据集上不同目标检测算法精度比较

单位: %

算法	类别												mAP
	单车	船	瓶子	巴士	汽车	猫	椅子	杯子	狗	摩托	人	桌子	
YOLOv4 ^[8]	78.8	67.3	71.3	88.6	75.1	60.2	65.7	60.9	72.7	71.2	73.5	54.2	70.0
YOLOv5_m ^[34]	82.0	70.1	72.3	81.7	80.8	67.4	75.7	82.8	77.0	78.5	83.7	59.5	76.0
YOLOX_m ^[7]	82.3	71.5	70.5	90.1	77.7	74.8	70.5	72.0	75.8	75.8	77.2	64.5	75.2
YOLOv7 ^[35]	84.7	76.5	74.4	93.2	81.6	81.4	75.5	79.8	87.5	82.3	81.1	64.4	80.2
EfficientDET_d0 ^[36]	77.2	71.9	57.9	90.1	76.8	72.7	64.2	65.5	74.7	74.1	70.1	52.5	70.6
SSD ^[9]	77.7	66.2	52.8	87.4	72.6	70.7	56.8	56.1	76.9	75.0	63.6	50.7	67.2
Centernet ^[37]	77.9	61.7	53.6	84.1	64.5	67.2	56.5	43.3	75.6	67.7	59.5	43.2	62.9
FCOS ^[38]	76.5	68.0	66.1	88.7	69.5	71.9	67.0	61.5	75.2	68.9	69.0	54.6	69.7
PE-YOLO ^[16]	84.7	79.2	79.3	92.5	83.9	71.5	71.7	79.7	79.7	77.3	81.8	55.3	78.1
MAET ^[17]	81.3	71.6	74.5	89.7	82.1	69.5	65.5	72.6	75.4	72.7	77.4	53.3	73.8
文献[18](YOLOv5_m)	86.0	75.0	76.0	88.0	84.0	75.0	73.0	71.0	76.0	82.0	85.0	62.0	77.8
SAM-MSFF ^[19]	82.7	77.8	65.1	92.8	85.2	77.4	69.0	70.2	78.7	81.1	82.5	62.1	77.1
RetinexFormer+YOLOv7	85.0	75.7	74.9	95.2	80.6	79.0	77.7	82.4	83.2	83.4	81.3	64.4	80.2
HVICIDNet+YOLOv7	85.6	74.3	74.8	93.5	81.8	74.7	76.4	78.8	86.0	80.8	82.1	66.3	79.7
NeRCO+YOLOv7	73.5	63.3	67.1	85.3	70.4	69.1	66.7	69.8	73.4	68.6	70.5	56.6	69.52
Zero-DCE+YOLOv7	85.4	73.6	75.6	93.8	81.2	78.1	76.0	81.6	85.1	79.8	82.0	64.2	79.7
Zero-DCE-YOLOv7	84.4	76.5	76.6	92.9	80.3	80.7	76.7	77.4	85.6	81.5	81.2	65.5	79.9
ICFIE-YOLO	87.5	77.6	77.2	94.4	83.8	81.1	78.4	85.1	88.0	83.2	83.1	68.2	82.3

由表 1 可以看出, 主流正常照度下的目标检测方法在低照度数据集上难以获得理想性能, 尽管 YOLOv5、YOLOv7 和 YOLOX 检测器通过优化特征提取网络和检测头的形式取得了一定程度上的精度提升, 但效果与低照度目标检测方法还是存在差距. 本文所提 ICFIE-YOLO 通过优化检测图像质量和特征增强, 在大多数类

别的检测精度上都具有优势, 多数类别平均精度在所有比较的方法中都取得最优或次优结果, 各类别平均精度达到 82.3%. 文献[17]、文献[18]方法将图像矫正网络与目标检测网络进行联合训练, 但是文献[17]方法建立在假设的图像退化模型上, 这与真实低照度图像目标分布不一致, 导致其性能提升有限; 文献[18]使

用像素高阶映射模块来解决低照度图像特征不显著的问题,但是该模块仍然使用独立的损失函数,因此尽管将损失函数加入到目标检测的损失函数一同进行优化,缓解了优化方向不一致的问题,但仍然不能完全解决优化方向不一致的问题,在低照度目标检测数据集上的效果提升有限. PE-YOLO^[16]方法通过一个级联训练的图像增强网络提升低照度图像检测精度,增强网络中使用了拉普拉斯分解和重建,而重建过程可能导致图像信息损失. RetinexFormer方法在低光图像增强数据集 LOL 上获得的 PSNR 和 SSIM 指标(越高越好)分别是 25.16 和 0.845,而 NeRCo 和 Zero-DCE 在同样的指标上取得的值分别是 19.00 和 0.536、14.08 和 0.561,因此在图像增强任务上,前者显著强于后面两者. 但 RetinexFormer+YOLOv7 的检测性能与 Zero-DCE+YOLOv7 差距并不大,NeRCo+YOLOv7 却取得了相对最差的结果. 这是因为“先增强,后检测”的策略中,图像增强的结果不是以目标检测精度为导向的,这种增强对于低照度目标检测而言具有更大的不确定性. NeRCo+YOLOv7 的结果也证明了不适合目标检测的图像增强会导致图像中的目标特征畸形,导致后续检测性能下降. 从上述实验结果可以看出,一方面,检测结果及增强方法与低照度图像的增强结果好坏没有直接关系;另一方面,增强方法的性能很大一部分取决于训练数据集,但是目标检测环境下无法保证被检测的低照度图像分布与增强模型的训练集一致,因此这种策略效果十分有限. “联合增强和检测”的策略一定程度上使增强方法对图像的增强有利于后续目标检测,因此 Zero-DCE-YOLOv7 的性能略优于 Zero-DCE+YOLOv7,但是损失函数的优化方向仍然有部分倾向于图像增强,这就导致了检测效果受限.

为更加直观对比本文方法与主流目标检测方法检测效果的差异,图 7 展示了在 ExDark 数据集上检测结果的可视化对比. 可以看出,由于低照度图像质量退化的影响,一般目标检测方法对于较暗区域的目标无法检出. 具体来说,由于一些目标在图像中较暗的区域,这些目标与昏暗的背景几乎融合在一起,导致特征提取存在困难,检测器难以区分背景和特征,因此存在较多的目标漏检. 本文所提出的方法首先通过光照矫正使目标从背景中凸显出来,背景中的目标特征提取更加顺利,使得许多在暗处原本无法检出的目标被检出;然后通过特征交互增强强化提取到的目标特征并滤除背景和噪声特征,提供给检测头的特征更加清晰、有效,使目标分类更加准确. 图 7 中最后一行右下角的桌子与背景几乎融为一体,仅有 ICFIE-YOLO 能够将其检测出来,本文方法确实能够更加高效地检出黑暗中的目标.

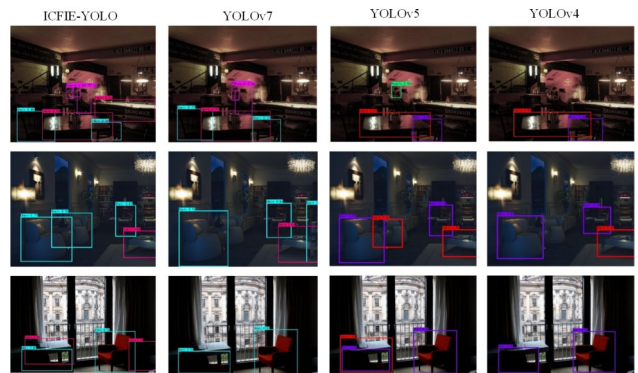


图 7 本文方法与其他主流目标检测方法在 ExDark 数据集上可视化检测结果

为验证 ICFIE-YOLO 的泛化性,将所提方法在 DarkFace 数据集上与性能较好的目标检测方法的平均精度(AP)、50%目标阈值下的准确率(P)以及召回率(R)进行对比,结果如表 2 所示.

表 2 在 Darkface 数据集上不同目标检测算法性能比较 单位:%

方法	AP	P	R
YOLOv7	52.19	89.79	33.48
HVICIDNet+YOLOv7	50.02	86.56	28.56
RetinexFormer+YOLOv7	50.82	88.96	27.57
Zero-DCE+YOLOv7	43.27	82.72	29.05
Zero-DCE-YOLOv7	48.79	91.17	28.12
文献[18](YOLOv5_m)	51.25	84.70	37.47
ICFIE-YOLO	55.94	91.53	37.41

DarkFace 数据集中的图像光照较 ExDark 更弱,并且具有更多的小尺寸目标. 由表 2 可见,所提 ICFIE-YOLO 在 DarkFace 数据集上各指标均取得较好效果,平均精度达到 55.94%,显著优于其他方法. 由于 DarkFace 数据集图像更加劣质,其数据分布与 HVICIDNet 和 RetinexFormer 方法的训练数据差异更大,因此采用“先增强后检测”的策略组合检测性能下降较 ExDark 数据集更加严重,Zero-DCE 同样难以处理照度更低的图像,因此检测效果也出现了较大下降. Zero-DCE-YOLOv7 采用“联合增强和检测”的策略,增强结果对于目标检测而言相对有利,因此其性能稍有提升. 文献[18]也采用“联合增强和检测”的策略,其中的长距离特征捕获模块使其在召回率上取得了不错的效果,但是由于增强网络对目标特征造成了不利影响,因此准确率并不理想.

4.3 消融实验

为验证 ICFIE-YOLO 方法中所提 MSICN 和 FIE 结构对低照度图像目标检测的作用和有效性,对模块进行消融实验,除了针对所提 MSICN 和 FIE 进行消融外,还对 MSICN 内部的全局光照矫正 GIC 和局部光照矫正

LIC进行了消融,结果如表3所示.

表3 MSICN和FIE的有效性验证

基线模型	MSICN		FIE	mAP/%	Recall@0.5/%
	GIC	LIC			
√				80.2	67.3
√	√	√		81.4	69.8
√	√	√	√	82.3	69.9
√			√	81.8	66.4
√		√	√	81.9	69.0
√	√		√	81.5	69.2

表3中基线模型使用YOLOv7-l,在加入MSICN后检测mAP增加1.2个百分点,引入FIE后mAP再提高0.9个百分点.在加入MSICN后,检测器在50%置信度下的召回率提升到69.8%.这表明在单独加入MSICN后,模型对于目标的检出率和分类精度都有所提高,本来难以检出的目标在经过MSICN调整后被检出,低照度图像经过MSICN调整后检测器也能够提取到更加丰富的特征,因此mAP也得到提升.再加入FIE模块mAP得到进一步提升,但召回率仅有小幅度提升,这是因为FIE的目的是过滤主干网络提取的特征中的噪声,使得检测头能够获取到高质量特征并且进行有效表达,从而增加检测头对目标类别预测的准确率,因此mAP得到进一步提升而召回率小幅提升.当基线模型仅仅加入FIE时也能够提升mAP,但是召回率却略有下降.本文认为造成此现象的原因是低照度图像未经过光照调整,导致部分隐藏在背景中的目标特征提取不足,尽管特征经过了FIE进行增强,但仅仅是增强了图像中那些被成功提取到特征的目标的特征,而特征提取不足的

目标被漏检,因而单独使用FIE时召回率没有提升. MSICN能够解决FIE计算过程中目标特征质量不佳的问题,MSICN内部的GIC和LIC对低照度图像光照进行调整,有效帮助后续特征提取网络更好地对目标特征进行提取,因此单独加入GIC和LIC都能够提升目标检测的召回率.

另外,为验证MSICN对图像光照调整的处理效果,将原图像与MSICN、仅使用LIC、仅使用GIC以及RetinexFormer输出的图像进行对比,结果如图8所示. MSICN的目的是根据检测网络的需求调整被检测图像中的光照,使图像的特征更容易被检测网络提取,尽可能减少目标特征信息丢失,因此对图像的调整并不如RetinexFormer等低照度图像增强网络那样明显.但是由图8可以看出,MSICN确实一定程度上提升了图像整体亮度 and 对比度,对于图像目标存在的区域光照调整更为自然,并且没有造成图像失真. RetinexFormer增强后的图像在视觉效果上好于MSICN,但是图像部分区域存在失真和大量噪声,这会导致目标特征变形,影响检测效果.因此,MSICN确实提高了目标检测器的检测性能,达到了其设计目的.

最后,为验证FIE对特征的增强作用,将FIE模块输出的特征图与不带有FIE模块输出的特征图使用热力图的形式进行可视化.图9展示了特征图的可视化效果,图中颜色越红的区域表示特征图中的目标置信度越高.可见,FIE模块可以使输出特征更加准确地聚焦在目标上,使目标定位更加准确,原本由于置信度较低导致漏检的目标也得以检出,最终提升检测模型的准确率.

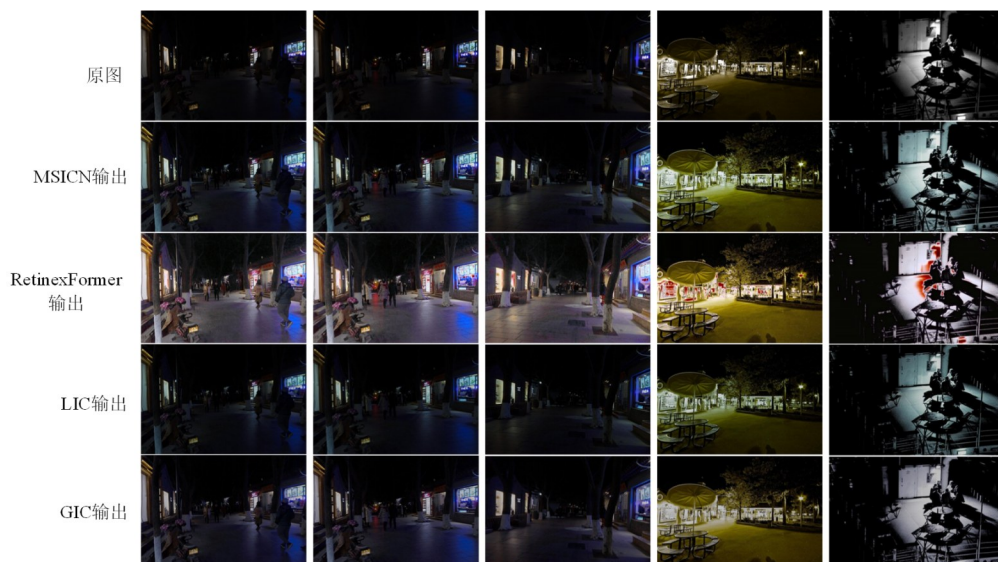


图8 MSICN、LIC、GIC和RetinexFormer输出图像与原图像

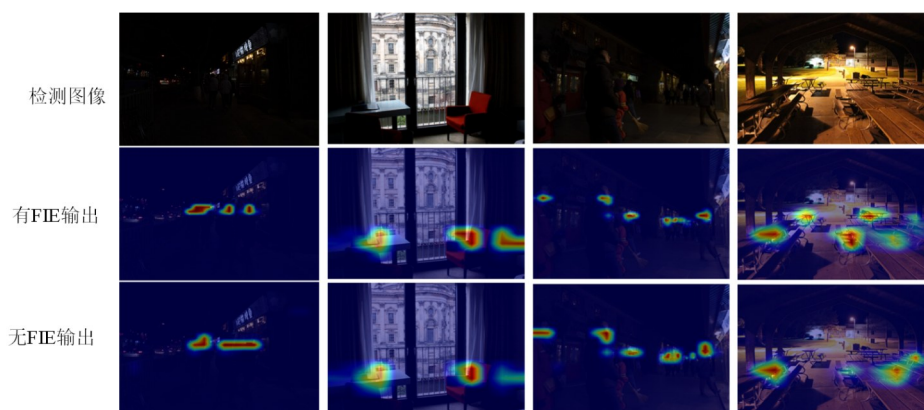


图9 有无FIE模块的检测器输出特征可视化对比

4.4 性能代价分析

ICFIE-YOLO结合了MSICN和FIE能够更进一步提升检测方法的检测性能,但会造成一定的内存和计算开销,为量化分析参数量和推理速度代价,将ICFIE-YOLO的参数量和推理时间与基线模型进行对比,结果如表4所示.

表4 ICFIE-YOLO时间、空间代价分析

方法/模块	输入尺寸	推理时间/ms	可训练参数量
YOLOv7	640	51	37.3×10^6
YOLOv7	416	29	37.3×10^6
ICFIE-YOLO	640	68	75.7×10^6
ICFIE-YOLO	416	40	75.7×10^6
MSICN	416/640	—	0.85×10^6
FIE	416/640	—	38.2×10^6

以上结果均在 GTX1080Ti 图形处理器上计算得到,检测方法均使用12分类输出.由于MSICN和FIE模块的参数量均由卷积层产生,因此参数量仅受输入通道数影响,其中MSICN输入通道数为3,FIE接受YOLOv7的3个特征输出层作为输入,通道数分别为256、512和1024.由表4中的实验结果可见,FIE中需要依赖卷积生成交互特征,因此导致参数量较大,达到了 38.2×10^6 ,而MSICN仅需要 0.85×10^6 参数.ICFIE-YOLO在416像素 \times 416像素的图片输入和640像素 \times 640像素的图片输入时检测推理时间较YOLOv7分别增加了11ms和17ms,这对于整个目标检测模型的计算速度而言是可以接受的.虽然MSICN和IFE导致模型的参数量和推理时间有所增加,但是模型的mAP、准确率和召回率均有明显特征.

5 结论

为解决低照度图像中全局光照不足或局部光照不均匀带来的特征难以提取、特征空间噪声大导致目标检测困难的问题,本文提出基于多尺度光照矫正和特

征交互增强的ICFIE-YOLO低照度目标检测方法.所提方法首先利用一个多尺度光照矫正网络(MSICN)结合基于检测损失的对齐训练对低照度图像光照进行矫正,使低照度图像中的目标特征能更容易被主干网络提取;然后,提出特征交互增强(FIE)结构,建立特征在空间和通道上的关联关系,从而对特征空间中的目标特征进行增强并过滤背景和噪声特征,最终提高低照度目标检测精度.在ExDark和DarkFace数据集上的实验结果表明,相较于主流目标检测方法和现有低照度目标检测方法,本文方法在低照度环境下的目标检测任务上具有更好的检测性能.考虑到目前低照度目标检测数据集数据量少,成对的正常照度-低照度图像目标检测数据集难以获取,本文所提ICFIE-YOLO在无成对低照度图像数据集的情况下实现了较高精度的目标检测,因此具有较好的潜在应用价值.下一步将进一步优化所提算法,尝试提升更低照度图像的目标检测精度,在提升检测精度的同时实现检测模型轻量化,提升检测速度.

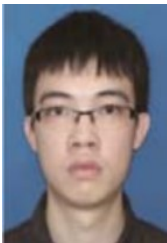
参考文献

- [1] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [2] KIM H, LEE E C, SEO Y, et al. Character detection in animated movies using multi-style adaptation and visual attention[J]. IEEE Transactions on Multimedia, 2020, 23: 1990-2004.
- [3] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[EB/OL]. (2018-01-24)[2024-07-09]. <https://arxiv.org/abs/1703.06870>.
- [4] ZHU Y S, ZHAO C Y, WANG J Q, et al. CoupleNet: Coupling global structure with local parts for object detection[C]//2017 IEEE International Conference on Com-

- puter Vision (ICCV). Piscataway: IEEE, 2017: 4124-4134.
- [5] WANG X L, SHRIVASTAVA A, GUPTA A. A-fast-RCNN: Hard positive generation *via* adversary for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2017: 2606-2615.
- [6] REDMON J, FARHADI A. YOLOv3: An incremental improvement[EB/OL]. (2018-04-08)[2024-07-09]. <https://arxiv.org/abs/1804.02767v1>.
- [7] GE Z, LIU S T, WANG F, et al. YOLOX: Exceeding YOLO series in 2021[EB/OL]. (2021-08-06) [2024-07-09]. <https://arxiv.org/abs/2107.08430v2>.
- [8] BOCHKOVSKIY A, WANG C Y, LIAO H M. YOLOv4: Optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2024-07-09]. <https://arxiv.org/abs/2004.10934v1>.
- [9] PAN H D, JIANG J, CHEN G F. TDFSSD: Top-down feature fusion single shot MultiBox detector[J]. Signal Processing: Image Communication, 2020, 89: 115987.
- [10] CUI L S, MA R, LV P, et al. MDSSD: Multi-scale deconvolutional single shot detector for small objects[J]. Science China Information Sciences, 2020, 63(2): 120113.
- [11] VAIDWAN H, SETH N, PARIHAR A S, et al. A study on transformer-based object detection[C]//2021 International Conference on Intelligent Technologies (CONIT). Piscataway: IEEE, 2021: 25-27.
- [12] LOH Y P, CHAN C S. Getting to know low-light images with the Exclusively Dark dataset[J]. Computer Vision and Image Understanding, 2019, 178: 30-42.
- [13] YUAN J J, HU Y L, SUN Y F, et al. A plug-and-play image enhancement model for end-to-end object detection in low-light condition[J]. Multimedia Systems, 2024, 30(1): 27.
- [14] XUE R, DUAN J L, DU Z W. MPE-DETR: A multiscale pyramid enhancement network for object detection in low-light images[J]. Image and Vision Computing, 2024, 150: 105202.
- [15] CHEN C, CHEN Q F, XU J, et al. Learning to see in the dark[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 3291-3300.
- [16] YIN X C, YU Z D, FEI Z T, et al. PE-YOLO: Pyramid enhancement network for dark object detection[M]//Artificial Neural Networks and Machine Learning - ICANN 2023. Cham: Springer, 2023: 163-174.
- [17] CUI Z T, QI G J, GU L, et al. Multitask AET with orthogonal tangent regularity for dark object detection[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2021: 2533-2542.
- [18] 江泽涛, 翟丰硕, 钱艺, 等. 结合特征增强和多尺度感受野的低照度目标检测[J]. 计算机研究与发展, 2023, 60(4): 903-915.
JIANG Z T, ZHAI F S, QIAN Y, et al. Low illumination object detection combined with feature enhancement and multi-scale receptive field[J]. Journal of Computer Research and Development, 2023, 60(4): 903-915. (in Chinese)
- [19] 江泽涛, 李慧, 雷晓春, 等. 一种基于SAM-MSFF网络的低照度目标检测方法[J]. 电子学报, 2024, 52(1): 81-93.
JIANG Z T, LI H, LEI X C, et al. A low-light object detection method based on SAM-MSFF network[J]. Acta Electronica Sinica, 2024, 52(1): 81-93. (in Chinese)
- [20] PARTHASARATHY S, SANKARAN P. Fusion based multi scale RETINEX with color restoration for image enhancement[C]//2012 International Conference on Computer Communication and Informatics. Piscataway: IEEE, 2012: 1-7.
- [21] DONG X, WANG G, PANG Y, et al. Fast efficient algorithm for enhancement of low lighting video[C]//2011 IEEE International Conference on Multimedia and Expo. Piscataway: IEEE, 2011: 1-6.
- [22] 江泽涛, 覃露露. 一种基于U-Net生成对抗网络的低照度图像增强方法[J]. 电子学报, 2020, 48(2): 258-264.
JIANG Z T, QIN L L. Low-light image enhancement method based on U-Net generative adversarial network[J]. Acta Electronica Sinica, 2020, 48(2): 258-264. (in Chinese)
- [23] LU B B, PANG Z B, GU Y N, et al. Channel splitting attention network for low-light image enhancement[J]. IET Image Processing, 2022, 16(5): 1403-1414.
- [24] 江泽涛, 钱艺, 伍旭, 等. 一种基于ARD-GAN的低照度图像增强方法[J]. 电子学报, 2021, 49(11): 2160-2165.
JIANG Z T, QIAN Y, WU X, et al. Low-light image enhancement method based on ARD-GAN[J]. Acta Electronica Sinica, 2021, 49(11): 2160-2165. (in Chinese)
- [25] 江泽涛, 覃露露, 秦嘉奇, 等. 一种基于MDARNet的低照度图像增强方法[J]. 软件学报, 2021, 32(12): 3977-3991.
JIANG Z T, QIN L L, QIN J Q, et al. Low-light image enhancement method based on MDARNet[J]. Journal of Software, 2021, 32(12): 3977-3991. (in Chinese)
- [26] SHANG X K, AN N, ZHANG S M, et al. Toward robust and efficient low-light image enhancement: Progressive attentive retinex architecture search[J]. Tsinghua Science

- and Technology, 2023, 28(3): 580-594.
- [27] JIANG Y F, GONG X Y, LIU D, et al. EnlightenGAN: Deep light enhancement without paired supervision[J]. IEEE Transactions on Image Processing, 2021, 30: 2340-2349.
- [28] GUO C L, LI C Y, GUO J C, et al. Zero-reference deep curve estimation for low-light image enhancement[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 1777-1786.
- [29] 陈科圻, 朱志亮, 邓小明, 等. 多尺度目标检测的深度学习研究综述[J]. 软件学报, 2021, 32(4): 1201-1227.
CHEN K Q, ZHU Z L, DENG X M, et al. Deep learning for multi-scale object detection: A survey[J]. Journal of Software, 2021, 32(4): 1201-1227. (in Chinese)
- [30] YANG W H, YUAN Y, REN W Q, et al. Advancing image understanding in poor visibility environments: A collective benchmark study[J]. IEEE Transactions on Image Processing, 2020, 29: 5737-5752.
- [31] CAI Y H, BIAN H, LIN J, et al. Retinexformer: One-stage retinex-based transformer for low-light image enhancement[C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2023: 12470-12479.
- [32] YAN Q S, FENG Y X, ZHANG C, et al. You only need one color space: An efficient network for low-light image enhancement[EB/OL]. (2024-06-17)[2024-07-09]. <https://arxiv.org/abs/2402.05809v3>.
- [33] YANG S Z, DING M X, WU Y M, et al. Implicit neural representation for cooperative low-light image enhancement[C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2023: 12872-12881.
- [34] JOCHER G, NISHIMURA K, MINEEVA T. YOLOv5[EB/OL]. (2022-11-12)[2024-07-09]. <https://github.com/ultralytics/yolov5>.
- [35] WANG C Y, BOCHKOVSKIY A, LIAO H M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[EB/OL]. (2022-07-06)[2024-07-09]. <https://arxiv.org/abs/2207.02696v1>.
- [36] TAN M X, PANG R M, LE Q V. EfficientDet: Scalable and efficient object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 10781-10790.
- [37] DUAN K W, BAI S, XIE L X, et al. CenterNet: Keypoint triplets for object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 6568-6577.
- [38] TIAN Z, SHEN C H, CHEN H, et al. FCOS: Fully convolutional one-stage object detection[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2019: 9627-9636.

作者简介



秦嘉奇 男, 博士研究生, 讲师, 系统架构设计师. 主要研究领域为计算机视觉.
E-mail: 18878396109@163.com



江泽涛 男, 博士, 教授, 博士生导师. 主要研究领域为深度学习、计算机视觉.
E-mail: zetaojiang@guet.com



雷晓春 女, 博士, 硕士生导师. 主要研究领域为计算机视觉、深度学习.
E-mail: 22704797@qq.com