

基于改进 YOLOv5 的低光照图像目标检测

舒子婷^{1,2}, 张泽斌^{1,2}, 宋尧哲^{1,2}, 吴萌萌^{1,2}, 袁晓兵^{1*}

¹中国科学院上海微系统与信息技术研究所微系统技术重点实验室, 上海 201800;

²中国科学院大学, 北京 100049

摘要 针对低光照环境下现有的目标检测算法普遍存在检测精度较低的问题, 提出一种改进 YOLOv5 的双通道低光照图像目标检测算法(YOLOv5_DC)。首先, 通过伽马变换和叠加高斯噪声的方法合成低光照图像, 扩充数据集, 提高模型的泛化能力; 其次, 提出特征增强模块, 引入通道注意力机制, 融合增强图像和原始图像的低级特征, 抑制噪声特征的影响, 改善网络的特征提取能力; 最后, 在颈部网络中加入特征定位模块, 增加特征图在目标区域的响应值, 使网络更关注目标区域, 提高网络的检测能力。实验结果表明: 所提 YOLOv5_DC 算法实现了更高的检测精度, 在低光照图像目标检测数据集 ExDark* 上的平均精度均值(mAP)@0.5 达 71.85%, 较原始的 YOLOv5 算法, 提高了 1.28 个百分点。

关键词 目标检测; 低光照图像; YOLOv5 网络; 图像增强; 注意力机制

中图分类号 TP391.4

文献标志码 A

DOI: 10.3788/LOP212965

Low-Light Image Object Detection Based on Improved YOLOv5 Algorithm

Shu Ziting^{1,2}, Zhang Zebin^{1,2}, Song Yaozhe^{1,2}, Wu Mengmeng^{1,2}, Yuan Xiaobing^{1*}

¹Key Laboratory of Microsystem Technology, Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai 201800, China;

²University of Chinese Academy of Sciences, Beijing 100049, China

Abstract Aiming at the low detection accuracy of existing object detection algorithms in a low-light environment, a dual-channel low-light image object detection algorithm called YOLOv5_DC according to an enhanced YOLOv5 algorithm is suggested. First, we synthesize low-light images using Gamma transformation and superimposing Gaussian noise to expand the dataset and promote the network's generalization. Second, a feature enhancement module is proposed. The channel attention method is used to integrate the low-level characteristics of the improved image and the original image to decrease the effect of noisy features and increase the network's feature extraction capabilities. Finally, a feature location module is added to the neck network to boost the response value of the feature map in the target area, allowing the network to focus more on the target area and improve the network detection capabilities. The experimental results show that the proposed YOLOv5_DC algorithm achieves higher detection accuracy. On the low-light object detection dataset known as ExDark*, the mean average precision (mAP) @0.5 of the proposed algorithm reaches 71.85%, which is 1.28 percentage points higher than the original YOLOv5 algorithm.

Key words object detection; low-light image; YOLOv5 network; image enhancement; attention mechanism

1 引言

作为计算机视觉领域的基本任务之一, 目标检测被广泛应用于遥感监测、视频监控、自动驾驶等领域。摄像系统捕捉到的高质量、干净清晰的图像可以帮助检测系统得到精确的检测结果。然而, 在实际应用中, 检测过程会受到不同光照的影响。低光照情况下, 成像设备受到光线不足的影响, 捕获的图像质量退化, 有

颜色失真、对比度下降、信噪比低的缺点, 严重影响目标检测的精度, 给目标检测任务带来极大的挑战。

针对低光照环境下检测精度较低的问题, 一些研究改善硬件设备, 例如使用红外传感器和热成像的成像设备, 但是这种方法成本较高, 无法广泛应用。因此, 更多的研究工作希望提出一种有效的检测算法, 改善低光照环境下的检测性能。随着深度学习技术的发展, 基于深度神经网络的目标检测算法已经成为研究的热点, 主要

收稿日期: 2021-11-16; 修回日期: 2021-12-02; 录用日期: 2021-12-24; 网络首发日期: 2022-01-09

基金项目: 微系统技术重点实验室基金(6142804190304)

通信作者: *yuanxb@mail.sim.ac.cn

分为:以 RCNN^[1]、Fast RCNN^[2]和 Faster RCNN^[3]为代表的双阶段目标检测算法;以 YOLO^[4]、SSD^[5]为代表的单阶段目标检测算法。但常见的目标检测算法往往是针对高质量图像的,没有考虑图像降质的影响,对低光照环境下拍摄的图像的检测性能大幅度下降。

为了减少低光照对目标检测的负面影响,一些低光照图像目标检测算法研究通过改善检测网络的结构,使网络可以从低光照图像中提取更充足的信息,以提高目标检测性能。例如文献[6]在检测网络 RFB-Net^[7]的基础上引入特征金字塔和上下文融合模块,提高网络对低光照图像的特征提取能力,但是当低光照图像严重曝光不足时,检测目标和背景难以分辨,这类方法依旧难以提取足够的视觉特征,无法有效提升检测性能。光照强度是影响目标检测性能的关键因素,因此,更多的研究工作希望通过图像增强技术改善低光照图像的质量,减少低光照的影响。将图像增强作为目标检测的预处理步骤也是目前大多数低光照图像目标检测研究的主要方法。例如:文献[8]提出一种深度卷积对抗生成网络与 Faster RCNN 结合的低光照图像目标检测算法,该算法通过对抗生成网络复原低光照图像,生成类似于正常光照图像的图像,并在 Faster RCNN 中加入特征融合和多尺度池化,以得到更高精度的检测结果;文献[9]采用低分辨率、低光照图像增强网络,减少计算量,生成图像通过超分辨率网络后再

传输到 EfficientDet 检测网络中,得到检测结果;文献[10]从特征检索的角度考虑,提出一种更适用于目标检测的低光照图像增强网络,其与 SSD 检测网络结合,提高了检测精度。但是低光照图像增强算法在增强过程中可能受噪声和伪影的影响,因此无法确保增强算法一定有利于提高检测性能。

本文提出了一种改进 YOLOv5 的双通道低光照图像目标检测算法(YOLOv5_DC)。将低光照图像和增强图像作为模型的双通道输入,在模型骨干网络特征提取阶段添加特征增强模块,通过引入通道注意力机制,融合双输入的低级特征,并抑制噪声特征通道,增强对低级特征的表达能,提高检测和定位的精度;在 YOLOv5 的颈部网络加入特征定位结构,增加目标区域的特征响应值,让网络更关注目标所在的区域,提高检测的精度。实验结果表明,YOLOv5_DC 对低光照图像的检测性能高于其他算法,证明了改进算法的有效性。

2 YOLOv5 算法原理

YOLOv5 是 YOLO 系列中最新被提出的一种目标检测算法,在检测速度和精度上实现了更好的平衡,也是目前常用的目标检测算法之一。它的原理是:将目标检测任务视为定位和分类的回归问题,通过一个端到端的卷积神经网络对输入图像进行直接预测,得到目标的定位和分类概率。YOLOv5 的网络结构如图 1

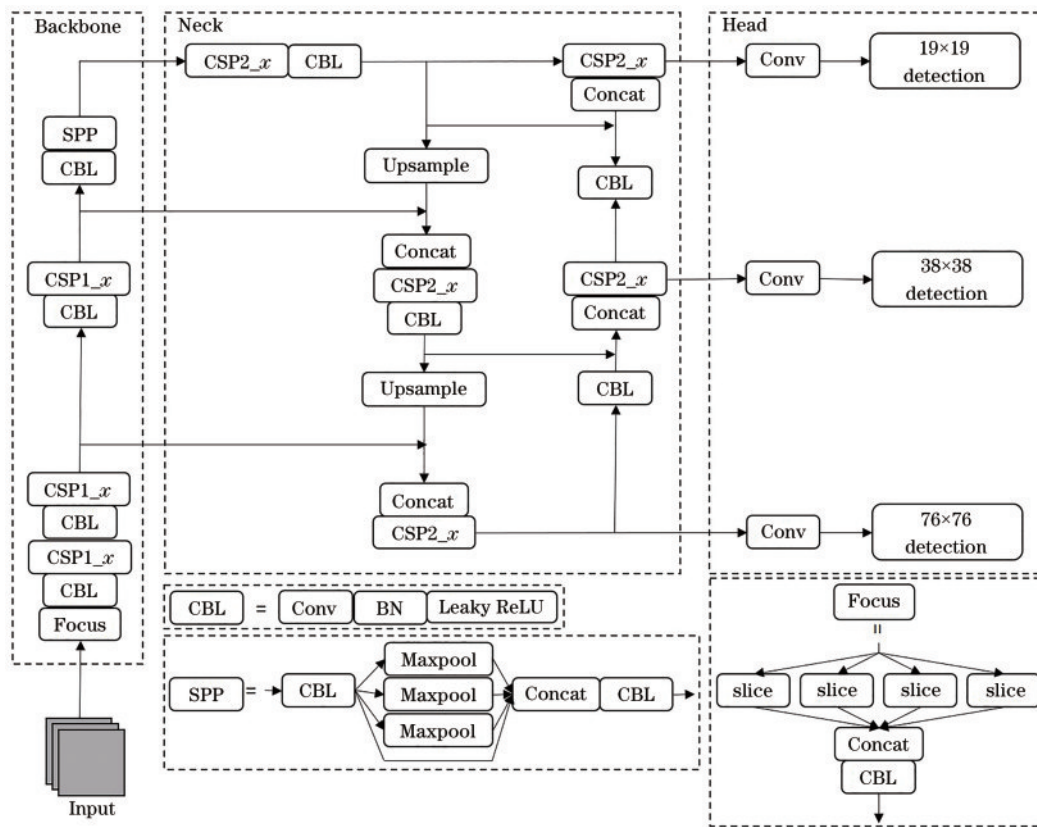


图 1 YOLOv5 网络框架

Fig. 1 YOLOv5 network structure

所示,主要包含 4 部分,分别为输入端(Input)、骨干(Backbone)网络、颈部(Neck)网络和检测头(Head)模块。

YOLOv5 的输入端具有自适应图像缩放功能,将输入图片缩放到需要的尺寸,并加入自适应锚框计算功能,用于计算不同数据集的最佳锚框值。为了丰富数据集背景,加快网络的收敛速度,YOLOv5 在输入端依旧使用 YOLOv4^[11]采用的 Mosaic 数据增强方法对输入进行增强^[12]。骨干网络主要用于提取输入图像的特征,为了提高训练速度,YOLOv5 在骨干网络中加入 Focus 结构,对输入图像进行切片,减少网络的参数量。YOLOv5 的颈部网络位于骨干网络和检测头模块之间,颈部网络采用 path aggregation network (PAN) 结构,对骨干网络提取的特征进行多尺度融合,增强对特征的表达能力。在检测头模块,网络利用颈部网络得到的特征进行多尺度检测,用于获得图片的检测结果。

3 问题的提出

为了探究低光照图像增强算法对目标检测算法的影响,选用 3 种不同的低光照图像增强算法进行实验,分别为 LIME^[13]、EnlightenGAN^[14]和 DCE_ZERO^[15]。使用增强算法对低光照图像数据集 ExDark^[16]进行处理,得到对应的 3 个增强数据集,增强结果示例如图 2 所示。对比发现:相较低光照图像,增强图像的整体亮度



图 2 低光照图像和增强图像。(a) ExDark 数据集; (b) LIME; (c) EnlightenGAN; (d) DCE_ZERO

Fig. 2 Low-light image and enhanced images. (a) ExDark dataset; (b) LIME; (c) EnlightenGAN; (d) DEC_ZERO

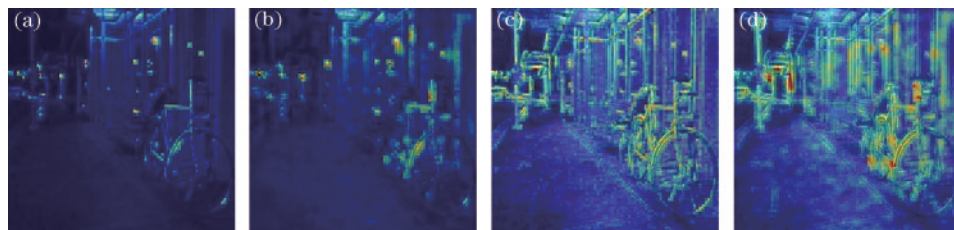


图 3 低光照图像和增强图像的特征可视化结果

Fig. 3 Feature visualization results of low-light images and enhanced images

更高,待检测目标更明显,但是增强算法在提高图像对比度的同时,也放大了图像噪声,使图像噪声难以忽略。

采用 YOLOv5 作为检测模型,对增强算法得到的数据集和原始数据集 ExDark 进行检测,检测结果如表 1 所示,其中 mAP@0.5:0.95 为在不同 IoU 阈值(从 0.5 到 0.95,步长为 0.05)下的平均检测精度,mAP@0.5 为交并比(IoU)阈值设置为 0.5 时算法对测试数据集中多类目标的平均检测精度。

表 1 不同数据集的检测结果

Table 1 Detection results of different datasets

Dataset	mAP@0.5:0.95	mAP@0.5
Dataset obtained by LIME	0.3758	0.6406
Dataset obtained by EnlightenGAN	0.3897	0.6577
Dataset obtained by DCE_ZERO	0.3915	0.6620
ExDark	0.3941	0.6605

从表 1 可以看出:较 ExDark 数据集,YOLOv5 在使用 LIME 和 EnlightenGAN 算法增强的数据集上的检测精度有所下降;在使用 DCE_ZERO 算法增强的数据集上的 mAP@0.5 指标提高 0.15 个百分点。低光照图像增强算法没有得到预期的结果,检测性能提升微弱,甚至出现下降的现象。这一现象证明:直接将增强图像输入到检测网络中进行检测不能有效提升网络的检测性能,增强算法在增强低光照图像的过程中受噪声和伪影的影响,导致目标检测性能下降。

4 改进的 YOLOv5 算法

4.1 增强算法对特征的影响

将在 COCO 数据集训练得到的 YOLOv5 预训练模型作为检测模型,对低光照图像和利用 DCE_ZERO 算法处理得到的增强图像进行检测。为了比较低光照图像和增强图像的低级特征和高级语义特征,对 Backbone 网络的特征提取阶段的第 4 层特征和输入到检测模块中进行 76×76 尺度检测的第 17 层特征进行可视化。特征可视化结果示例如图 3 所示,其中图 3(a)和图 3(b)分别表示低光照图像的第 4 层和第 17 层特征的可视化结果,图 3(c)和图 3(d)分别表示增强图像的第 4 层和第 17 层特征的可视化结果。

对比图 3(a)和图 3(c)低级特征可视化结果发现,

由于增强算法改善了原始低光照图像的光照分布,提高了图像的对比度,检测模型可以从增强图像上提取更多的低级特征,如边缘、形状等信息;对比图 3(b)和图 3(d)高级语义特征可视化结果发现,低光照图像特征在目标所在位置的高响应区域较少,网络没有充分提取目标区域特征,相比之下,增强图像特征在目标区域的响应值更高,网络提取到更多的目标特征。但是从图 3 可以看出,相较于低光照图像,增强图像包含更丰富的低级特征和高级语义特征的同时,也受到更多噪声特征和背景特征的干扰。

4.2 基于通道注意力机制的特征增强模块

YOLOv5 算法采用 PAN 结构作为颈部特征融合模块的主要框架,相较于 FPN 结构, PAN 结构增加了自下而上的特征融合过程,将高分辨率的低级特征传

递到低分辨率的高级语义特征,使融合后的特征包含更丰富的位置信息,提高网络多尺度检测的定位能力和检测精度。但是低光照图像对比度低,YOLOv5 无法从图像中提取足够的低级特征,影响检测和定位的准确性。增强图像可以为网络提供足够的特征信息,但在增强过程中会引入更大的图像噪声,导致图像特征受到更多噪声特征的影响,因此无法直接用网络对增强图像进行检测。

针对上述问题,本文借鉴 SKNet 网络^[17]的设计思想,提出了一个基于通道注意力机制的特征增强模块 (FE_CBM),学习特征图中每个通道的重要性,为每个特征通道添加一个权重,使网络自适应地对低光照图像特征和增强图像特征进行融合,获取更丰富的低级特征信息。FE_CBM 的结构如图 4 所示。

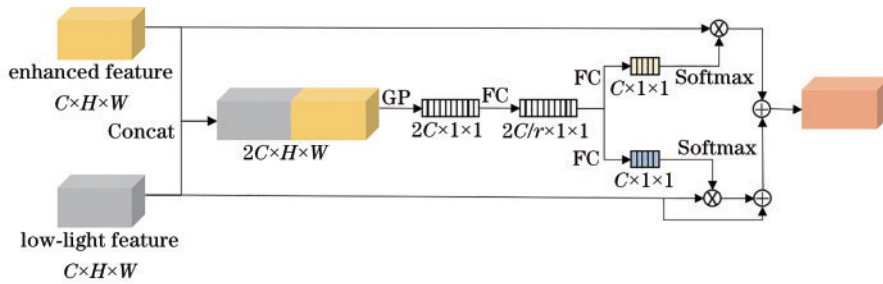


图 4 基于通道注意力机制的特征增强模块

Fig. 4 Feature enhancement module based on channel attention mechanism

FE_CBM 将低光照图像特征和增强图像特征作为模块的双通道输入,考虑到噪声特征的影响,输入特征逐像素相加将导致低光照图像特征受增强图像噪声特征的干扰,因此 FE_CBM 对输入特征进行拼接,得到融合后特征 $U \in \mathbf{R}^{2C \times H \times W}$,实现信息融合。为了在通道维度上对特征进行重标定,采用全局平均池化方法对空间维度的特征信息进行编码,将每一个二维的特征通道编码为一个参数,用于表征特征通道的全局信息。输出的通道向量 $M \in \mathbf{R}^{2C \times 1 \times 1}$ 与融合特征维度一致,计算公式可表示为

$$M(i, j) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W U(i, j), \quad (1)$$

式中: H 和 W 表示特征的尺寸; i 和 j 均表示特征的空间位置。为了学习特征通道之间的相关性, FE_CBM 将通道向量 M 输入两个耦合的双层全连接分支,得到两个权重向量,计算公式可表示为

$$\mathbf{Z}_s = g_1(\mathbf{M}, \mathbf{W}) = \sigma[\mathbf{W}_s \delta(\mathbf{W}\mathbf{M})], \quad (2)$$

$$\mathbf{Z}_x = g_2(\mathbf{M}, \mathbf{W}) = \sigma[\mathbf{W}_x \delta(\mathbf{W}\mathbf{M})], \quad (3)$$

式中: $\mathbf{Z}_s \in \mathbf{R}^{C \times 1 \times 1}$ 和 $\mathbf{Z}_x \in \mathbf{R}^{C \times 1 \times 1}$ 表示输出的两个权重向量; \mathbf{W} 为第一个全连接层的参数,维度为 $C/\gamma \times C$, γ 为缩放因子,用于降低向量维度,缩小计算量; \mathbf{W}_s 和 \mathbf{W}_x 分别为两个全连接分支中第二层全连接层的参数,维度为 $C \times C/\gamma$,用于生成对应输入特征的

权重向量; δ 表示 ReLU 层; σ 表示 Sigmoid 层。

FE_CBM 使用生成的两个权重向量分别对双通道输入特征通道进行加权重标定,抑制噪声特征通道,降低噪声的影响,使网络更关注目标特征信息所在的通道,提高对低级特征融合的表达能。最后,为了防止特征重标定后梯度弥散和网络退化,在低光照图像分支加入残差连接。

4.3 特征定位模块

低光照图像中前景目标隐匿在背景环境中, YOLOv5 网络难以检测目标所在区域,只能提取少量的目标区域特征信息;而在增强图像上,网络可以获取更多目标区域的特征响应,但同时网络也从增强图像上提取到更多易于混淆的背景特征,导致网络对增强图像的检测效果不如低光照图像。为了提高网络的检测能力,本文提出了一个特征定位 (FL) 模块,学习增强图像特征与低光照图像特征的相关性,使网络关注到目标区域所在的位置,降低背景特征噪声的影响,提高网络在目标区域的特征响应。

Non-local network^[18] 提出了 non-local 模块,计算输入特征图的某个位置与其他所有位置的相关性,并将相关性作为一个权重向量,得到特征图加权求和值,并将其表示为输出特征图在该位置的响应。借鉴 non-local 模块的设计思想, FL 模块通过计算增强图像特征上每个位置与低光照图像特征所有位置的相

关性,学习增强图像特征中目标所在区域的位置,让目标检测模块更关注该区域的特征信息,降低背景特征的干扰。FL 模块结构如图 5 所示,计算相关性的表达式为

$$f(s_i \cdot x_j) = \phi(s_i)^T \varphi(x_j), \quad (4)$$

式中:输入分别为增强图像特征 s 和低光照图像特征

x ; ϕ 和 φ 分别将特征值转换到嵌入空间,计算相关性。FL 模块的输出可表示为

$$y_i = \sum_{v_j} \text{Soft max} [f(s_i \cdot x_j)] g(s_j) + x_i, \quad (5)$$

式中: y 为输出特征; g 为一个映射函数,将一个点映射为一个向量,视为该点的特征。

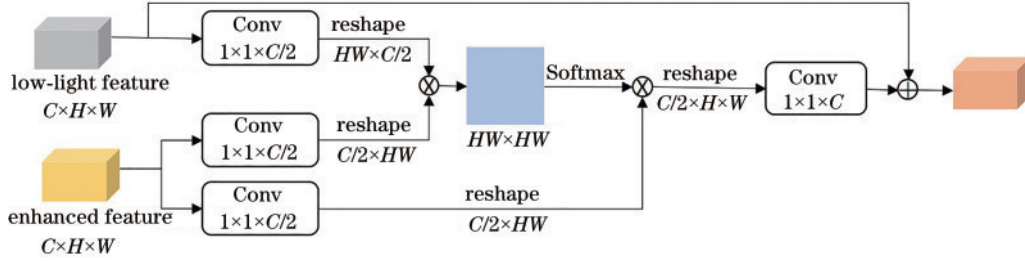


图 5 特征定位模块

Fig. 5 Feature location module

4.4 整体网络框架

本文提出了一种基于改进 YOLOv5 的双通道低光照目标检测算法,建立了 YOLOv5_DC 网络,如图 6 所示。该网络同时将原始图像和增强图像作为输入,在 Backbone 网络特征提取阶段加入特征增强模块,改善提取的低级特征信息,增强特征表达能力。为了提高网络对目标区域特征的响应值,YOLOv5_DC 在颈部特征融合阶段加入特征定位模块,使网络学习目标

特征所在的区域,提高网络对该区域的特征响应。YOLOv5_DC 没有直接使用增强图像的特征,而是通过深度学习学习增强图像有用的特征信息,改善低光照图像特征,然后输入到检测模块中得到检测结果。这不仅避免了增强操作导致图像噪声特征增加对检测性能产生影响的问题,又合理地利用增强图像的特征改善网络特征提取模块的性能,由此提高了网络的检测能力。

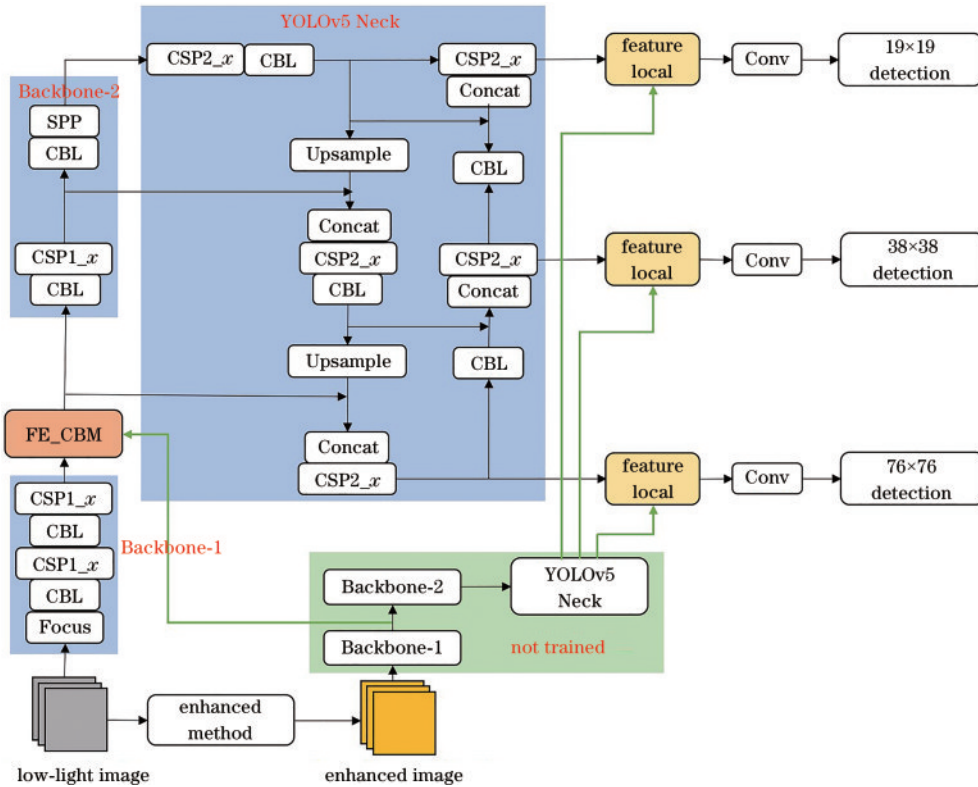


图 6 YOLOv5_DC 整体网络框架

Fig. 6 YOLOv5_DC overall network structure

5 实验及结果分析

实验使用的硬件配置为 Intel XeonE5-2620 V4、Nvidia GeForce GTX 1080Ti 显卡、Linux 操作系统, 软件环境为 CUDA11.1.0、Cudnn8.0, 深度学习框架为 Pytorch。训练过程采用 SGD 优化器, 初始学习率为 0.015, 动量大小为 0.937, batchsize 设为 8, 迭代次数为 280, 采用 Mosaic 数据增强方法。在训练过程中, 提取增强图像特征的 YOLOv5 网络参数不参与训练, 并且为了使模型的训练效果更稳定, 在训练前对提取低光照图像特征的骨干网络参数进行冻结, 直到 120 epoch, 将参数解冻, 参与网络训练。

5.1 低光照图像目标检测数据集

为了验证所提算法在低光照环境中的目标检测性能, 实验采用低光照图像目标检测数据集 ExDark。ExDark 数据集一共包含 12 类物体和 7363 张真实低光照环境下采集的图像, 其中 4800 张图像用于训练, 2563 张图像用于测试。考虑到训练数据集太小, 影响模型性能, 因此对 ExDark 数据集进行扩充, 以提高模型的泛化能力。

目前公开的低光照图像数据集大多数来自图像增强领域, 图像没有检测标注信息, 因此希望从目标检测数据集中选取图像, 通过人工合成的方法生成低光照图像, 对数据集进行扩充。PASCAL VOC 数据集^[19]是常用的目标检测数据集, 每张图像都含有检测标注信息, 但是 PASCAL VOC 数据集中包含的大多数图像为正常光照图像, 因此实验选取 PASCAL VOC 数据集中 4800 张图像, 对图像进行伽马变换和高斯噪声叠加, 生成低光照图像, 用以扩充 ExDark 数据集, 得到实验数据集 ExDark^{*}。变换结果示例如图 7 所示, 合成低光照图像的数学公式可表示为



图 7 正常光照图像和合成的低光照图像。(a) 正常光照图像; (b) 合成低光照图像

Fig.7 Normal illumination image and synthetic low-light image. (a) Normal illumination image; (b) synthetic low-light image

$$I_{low} = \beta I_{in}^{\gamma} + N_{\sigma}, \quad (6)$$

式中: I_{low} 为合成的低光照图像; I_{in} 为输入图像; N_{σ} 表示叠加的高斯噪声; β 和 γ 为伽马变换的系数。

使用 ExDark^{*} 数据集训练和验证检测模型。训练阶段, 模型在由 4800 张真实低光照图像和 4800 张合成低光照图像混合组成的数据集上训练, 测试阶段, 模型在 2563 张真实低光照图像上测试, 用于评估真实低光照环境下检测模型的性能。

5.2 实验结果与性能分析

采用平均精度均值 (mAP) 来衡量算法的性能。为了更好地评估模型的准确度, 选取 mAP@0.5:0.95 指标, 计算模型在多个 IoU 阈值的 mAP。

为了验证改进算法 YOLOv5_DC 的有效性, 使用 ExDark 数据集对 YOLOv5_DC 和现有的低光照图像目标检测算法进行训练和测试, 检测结果如表 2 所示。从表 2 可以看出: 与 EnlightenGAN 增强算法结合, 改进算法 YOLOv5_DC 的检测性能最佳, 相较于 YOLOv5 算法, mAP@0.5 提高 3.67 个百分点, mAP@0.5:0.95 提升 2.21 个百分点; 与低光照图像目标检测算法 RFB-Dark 相比, mAP@0.5 提高 4.22 个百分点, mAP@0.5:0.95 提升 6.32 个百分点; 与 CycleGAN-ResNet 相比, mAP@0.5:0.95 提高了 9.22 个百分点, 验证了改进算法的有效性。

表 2 不同目标检测算法在 ExDark 数据集上的性能对比
Table 2 Performance comparison of different object detection algorithms on ExDark dataset

Algorithm	mAP@0.5:0.95	mAP@0.5
YOLOv5	0.3941	0.6605
LIME+YOLOv5	0.3758	0.6406
DCE_ZERO+YOLOv5	0.3915	0.6620
EnlightenGAN+YOLOv5	0.3897	0.6577
RFB-Dark ^[6]	0.3530	0.6550
CycleGAN-ResNet ^[10]	0.3240	
LIME+YOLOv5_DC(ours)	0.4126	0.6934
DCE_ZERO+YOLOv5_DC(ours)	0.4152	0.6946
EnlightenGAN+YOLOv5_DC(ours)	0.4162	0.6972

为了评估合成低光照图像对改进算法的影响, 使用 ExDark^{*} 数据集进行实验。表 3 为联合不同的低光照图像增强算法 DCE_ZERO、EnlightenGAN 的情况下, YOLOv5_DC 与其他目标检测算法 Faster-RCNN、RFB-Net、YOLOv5 对 ExDark^{*} 数据集的检测结果。

对比检测算法在 ExDark 数据集和 ExDark^{*} 数据集上检测结果, 可以发现: 当使用 ExDark^{*} 数据集进行实验, YOLOv5 算法的检测精度得到明显提升, mAP@0.5 和 mAP@0.5:0.95 分别提高了 4.52 个百分点、3.19 个百分点, 改进算法 YOLOv5_DC 的 mAP@0.5 提高约 2 个百分点, mAP@0.5:0.95 提高约 1.5 个百

分点,证明了合成低光照图像有利于提高模型的性能。从表 3 可以看出:直接用目标检测算法对 ExDark^{*}数据集进行检测时,相较于 Faster RCNN 和 RFB-Net 检测算法,YOLOv5 的检测效果最好;与 DCE_ZERO 和 EnlightenGAN 增强算法联合后,RFB-Net 和 YOLOv5 模型对增强图像进行检测,检测精度下降明显;Faster

RCNN 与 DCE_ZERO 算法联合,检测精度得到少许提升;所提改进算法 YOLOv5_DC 与不同的增强算法结合,都可以实现检测精度的提升,其中联合 EnlightenGAN 算法对精度提升更有效,较原始 YOLOv5 算法,mAP@0.5 提高 1.28 个百分点,mAP@0.5:0.95 提升 0.94 个百分点。

表 3 不同目标检测算法在 ExDark^{*}数据集上的性能对比Table 3 Performance comparison of different object detection algorithms on ExDark^{*} dataset

Detection algorithm	Image size	mAP@0.5:0.95	mAP@0.5
Faster RCNN	600×1000	0.2491	0.5438
DCE_ZERO + Faster RCNN	600×1000	0.2527	0.5453
EnlightenGAN + Faster RCNN	600×1000	0.2456	0.5322
RFB-Net	300	0.3651	0.6484
DCE_ZERO + RFB-Net	300	0.3455	0.6201
EnlightenGAN + RFB-Net	300	0.3275	0.5941
YOLOv5	640	0.4260	0.7057
DCE_ZERO + YOLOv5	640	0.4240	0.7029
EnlightenGAN + YOLOv5	640	0.4237	0.7011
DCE_ZERO + YOLOv5_DC(ours)	640	0.4301	0.7176
EnlightenGAN + YOLOv5_DC(ours)	640	0.4354	0.7185

YOLOv5_DC 采用低光照图像和增强图像作为模型的双通道输入,通过特征增强模块和特征定位模型,学习增强图像特征有用的信息,提高模型的检测能力。为了验证增强特征对提高模型性能的有效性,采用低光照图像作为 YOLOv5_DC 的双通道输入,与采用增强图像作为输入进行实验,检测结果如表 4 所示。对比可知,当双通道输入包含增强图像,YOLOv5_DC 的检测精度更高,这也说明了增强图像为网络提供额外的信息,有助于提高模型检测的性能。

表 4 不同输入情况下 YOLOv5_DC 的检测性能

Table 4 Detection performance of YOLOv5_DC under different input conditions

Input	Dataset	mAP@0.5:0.95	mAP@0.5
low-light image+low-light image	ExDark [*]	0.4248	0.7077
low-light image+DCE_ZERO	ExDark [*]	0.4301	0.7176
low-light image+EnlightenGAN	ExDark [*]	0.4354	0.7185

为了探究 YOLOv5_DC 中不同模块对实验结果的影响,选取 EnlightenGAN 作为增强算法,在 ExDark^{*}数据集上进行消融实验,实验结果如表 5 所示,其中“√”表示在实验中添加该模块。

表 5 在 ExDark^{*}数据集上的消融实验结果Table 5 Results of ablation experimental on ExDark^{*} dataset

FE_CBM	FL module	mAP@0.5:0.95	mAP@0.5
		0.4237	0.7011
√		0.4306	0.7120
√	√	0.4351	0.7185

从表 5 可以看出:在 YOLOv5 骨干网络低级特征提取阶段加入 FE_CBM,融合增强特征和原始特征,网络获得更丰富的低级特征信息,相较于原始 YOLOv5 模型,加入 FE_CBM 的改进算法实现了更好的检测能力,mAP@0.5 由 70.11% 升至 71.2%,mAP@0.5:0.95 提升 0.69 个百分点;在此基础上,在 YOLOv5 颈部网络特征融合阶段加入 FL 模块,提高特征在目标区域的响应值,加强网络对目标区域的关注度,进一步提高了改进算法的检测性能,mAP@0.5 达 71.85%,mAP@0.5:0.95 提升 1.14 个百分点。结果证明了 FE_CBM 和 FL 模块的有效性。

6 结 论

针对低光照情况下目标检测任务性能下降的问题,对 YOLOv5 模型进行改进,提出了 YOLOv5_DC 模型。因为直接将低光照增强算法得到的增强图像输入检测算法无法提高算法的检测性能,YOLOv5_DC 模型将低光照图像和增强图像作为模型的双通道输入,首先通过特征增强模块,网络自适应地融合增强特征和低光照图像特征,提高对低级特征的提取能力;其次在颈部特征融合模块加入特征定位模块,学习增强图像特征与低光照图像特征的相关性,增加特征图在目标区域的响应值,使网络更关注目标区域;最后考虑到 ExDark 数据集包含的图片数量较少,通过伽马变换和叠加高斯噪声的方法合成低光照图像,对训练数据集进行补充,提高模型的泛化能力。实验结果表明,与其他的目标检测算法相比,YOLOv5_DC 模型在低光照图像上实现了更高的检测精度。

参 考 文 献

- [1] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//CVPR '14: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: ACM Press, 2014: 580-587.
- [2] Girshick R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1440-1448.
- [3] 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.
- [4] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.
- [5] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 21-37.
- [6] Xiao Y X, Jiang A W, Ye J H, et al. Making of night vision: object detection under low-illumination[J]. IEEE Access, 2020, 8: 123075-123086.
- [7] Liu S T, Huang D, Wang Y H. Receptive field block net for accurate and fast object detection[EB/OL]. (2017-11-21)[2021-05-06]. <https://arxiv.org/abs/1711.07767>.
- [8] Wang K, Liu M Z. Object recognition at night scene based on DCGAN and faster R-CNN[J]. IEEE Access, 2020, 8: 193168-193182.
- [9] Tao Q Y, Ren K, Feng B, et al. An accurate low-light object detection method based on pyramid networks[J]. Proceedings of SPIE, 2020, 11550: 1155015.
- [10] Chen W, Shah T. Exploring low-light object detection techniques[EB/OL]. (2021-07-30)[2021-08-09]. <https://arxiv.org/abs/2107.14382>.
- [11] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2021-05-04]. <https://arxiv.org/abs/2004.10934>.
- [12] 马永康, 刘华, 凌成星, 等. 基于改进YOLOv5的红树林单木目标检测研究[J]. 激光与光电子学进展, 2022, 59(18): 1828003.
- Ma Y K, Liu H, Ling C X, et al. Research on mangrove single wood object detection based on improved YOLOv5 [J]. Laser & Optoelectronics Progress, 2022, 59(18): 1828003.
- [13] Guo X J, Li Y, Ling H B. LIME: low-light image enhancement via illumination map estimation[J]. IEEE Transactions on Image Processing, 2017, 26(2): 982-993.
- [14] 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.
- [15] 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), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 1777-1786.
- [16] 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.
- [17] Li X, Wang W H, Hu X L, et al. Selective kernel networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 510-519.
- [18] Wang X L, Girshick R, Gupta A, et al. Non-local neural networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 7794-7803.
- [19] Everingham M, Gool L, Williams C K I, et al. The pascal visual object classes (VOC) challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303-338.