一种基于Night-YOLOX的低照度目标检测方法

江泽涛,施道权,雷晓春,何玉婷,李 慧,周永刚
(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: 由于在低照度场景下获取的图像具有亮度弱、对比度低、噪声多和细节丢失等特点,使用现有的检测模型对低照度图像进行目标检测会出现定位不准确和分类错误,从而导致最终的检测精度偏低. 针对以上现象,本文提出了一种基于Night-YOLOX的低照度目标检测方法. 该方法首先设计了一个低级特征聚集模块(Low-level Feature Gathering Module,LFGM)与主干网络合并. 在低照度场景下捕获更多有效的低级特征有利于定位目标,该模块通过聚集浅层特征图中具有判别性的低级特征并送入高级特征图和深层卷积阶段中,以补偿在对低照度图像进行特征提取过程中边缘、轮廓和纹理等低级特征的缺失. 然后,设计了一种注意力引导块(Attention Guidance Block,AGB)嵌入检测模型的颈部结构,从而减少低照度图像中噪声干扰的影响,引导检测模型推断出特征图中完整的对象区域范围并提取更多有用的对象特征信息,以提高目标分类的准确性. 最后,在真实低照度图像数据集ExDark上进行实验,结果表明所提出的Night-YOLOX相比于其它主流的目标检测方法,在低照度场景下具有更好的检测性能.

关键词: 目标检测;低照度图像;低级特征;注意力机制;YOLOX

基金项目: 国家自然科学基金(No.62172118);广西自然科学基金重点项目(No.2021GXNSFDA196002);广西图像图形智能处理重点实验项目(No.GIIP2203,No.GIIP2204);广西研究生教育创新计划(No.YCB2021070,No.YCBZ2018052,No.YCSW2022269,No.2021YCXS071)

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112(2023)10-2821-10

电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20221396

A Low-Illumination Object Detection Method Based on Night-YOLOX

JIANG Ze-tao, SHI Dao-quan, LEI Xiao-chun, HE Yu-ting, LI Hui, ZHOU Yong-gang
(*Guangxi Key Laboratory of Image and Graphic Intelligent Processing, Guilin University of Electronic Technology, Guilin, Guangxi* 541004, *China*)

Abstract: Images captured in low-illumination environments often have many quality problems, such as weak brightness, low contrast, much noise, and detail loss. These problems will lead to inaccurate localization and object classification errors when using the existing object detection models to detect low-light images, resulting in low detection accuracy. Aiming at the above phenomena, this paper proposes a low-illumination object detection method called Night-YOLOX. First, the low-level feature gathering module (LFGM) is designed to be incorporated into the backbone. Capturing more effective low-level features in low-illumination scenes is beneficial to locating objects. The LFGM aggregates more discriminative low-level features from the shallow feature maps and feeds them into the high-level feature maps and the deep convolution stages, so as to compensate for the loss of low-level edge, contour, and texture features during feature extraction in low-light images. Then, the attention guidance block (AGB) is designed to be embedded in the neck of the detection model. The AGB reduces the influence of noise interference in low-light images, guides the detection model to infer the complete object regions and extract more useful object feature information, so as to improve the accuracy of object classification. Finally, experiments are conducted on the real low-light image dataset ExDark. The experimental results show that compared with other mainstream object detection methods, the proposed Night-YOLOX has better detection performance in low-illumination scenarios.

Key words: object detection; low-light images; low-level features; attention mechanism; YOLOX

Foundation Item(s): National Natural Science Foundation of China (No.62172118); Key Projects of Guangxi Natural Science Foundation (No.2021GXNSFDA196002); Guangxi Key Laboratory of Image and Graphic Intelligent Processing (No.GIIP2203, No.GIIP2204); Innovation Project of Guangxi Graduate Education (No.YCB2021070, No.YCBZ2018052, No.YCSW2022269, No.2021YCXS071)

收稿日期:2022-12-25;修回日期:2023-04-10;责任编辑:朱梅玉

1 引言

目标检测是计算机视觉领域的研究热点之一,在行人检测、车辆追踪和人脸识别等实际应用场景中有着广泛的应用. 目标检测的任务是在给定的输入图像中绘制边界框来定位所有实例对象并识别每一个目标对象的类别. 然而,目标检测受成像条件和环境光照影响较大[1],对低照度图像进行检测仍然是一项具有挑战性的任务. 在光照不足的条件下拍摄的图像会不可避免地产生一系列质量退化问题,例如对比度低、亮度弱、噪声多以及细节丢失等. 在对低照度图像进行目标检测时,现有的目标检测方法难以准确地定位目标并提取有效的特征信息来识别对象类别,最终导致误检或漏检的现象. 低照度目标检测在现实生活中具有非常广泛的应用前景,例如夜间视频监控、夜间自动驾驶、无人机夜间侦察等. 因此,探索并研究低照度目标检测方法是非常有意义的.

基于深度学习的目标检测方法主要分为两类:双阶段目标检测和单阶段目标检测. 双阶段目标检测算法首先通过一个区域提议网络生成一组对象候选框,然后基于这些候选框进行回归和分类. 近几年提出的Dynamic R-CNN[2]、Sparse R-CNN[3]和DetectoRS[4]等检测模型都是优秀的双阶段目标检测方法. 单阶段目标检测算法是基于密集的点或锚框直接预测特征图中对象的类别和位置. YOLO系列的目标检测模型,例如YOLOv4[5]、YOLOX[6]等,凭借其出色的检测性能和效率,已经广泛应用于实践中,是单阶段目标检测的代表性方法. 这些现有的目标检测模型对正常照度图像进行目标检测时取得了较好的效果,但是对低照度图像进行目标检测时效果则很差.

一方面,由于低照度图像中的目标对象往往隐藏在黑暗区域中,只显示部分特征. 因此,主干网络只能提取到目标对象的少部分纹理、边缘和轮廓等低级特征,并且这些微弱、稀少的低级特征很容易又因为卷积操作而被融合到黑色背景中[7],尤其是在主干网络的更深层卷积阶段中,这些有价值的低级特征更容易丢失. 低级特征的缺失会影响低照度场景下的目标定位. 另一方面,由于低照度图像中的噪声干扰以及目标对象不可避免地会被黑暗背景遮挡或覆盖,检测模型难以判断特征图中完整的对象区域范围大小,也难以专注于相应的目标区域去学习足够的对象特征信息. 这样会造成目标分类错误,影响最终的低照度目标检测性能.

针对以上问题,本文提出了一种基于Night-YOLOX的低照度目标检测方法. 该方法首先设计了一个低级特征聚集模块(Low-level Feature Gathering Module, LFGM)以补偿在对低照度图像进行特征提取过程中低级特征的损失. 通常,在主干网络的浅层卷积阶段输出的特征图中包含较多的边缘、轮廓和纹理等低级特征. 将低级特征聚集模块与主干网络合并,对输出的浅层特征图进行融合并利用一组特定的池化层来捕获更具判别性的低级特征. 这些有效的低级特征会被送到高级特征图和深层卷积阶段中进行特征补偿. 然后,设计了一种注意力引导块(Attention Guidance Block, AGB)嵌入检测模型的颈部结构,引导检测模型推断出特征图中完整的对象区域范围并提取更多有用的对象特征信息,以提高目标分类的准确性. 该注意力引导块由区域捕获分支和特征采集分支两个主要部分组成. 区域捕获分支用于探索对象特征区域的范围大小,而特征采集分支旨在捕获更多有用的对象特征信息. 将两个分支的输出结果进行结合生成注意力权重,对特征图进行特征加权以进一步增强特征表示能力并减少噪声干扰、忽视黑暗背景等不相关的信息. 在真实低照度图像数据集ExDark[8]上进行实验,结果表明与其他主流的目标检测方法相比,本文提出的方法在低照度场景下具有更好的检测精度和检测效果.

2 相关工作

2.1 YOLOX目标检测模型

YOLOX[6]是非常出色的单阶段目标检测模型,在正常照明场景下进行目标检测能够取得很好的检测效果. YOLOX整体结构可分为四个部分:输入、主干网络、颈部结构和检测头. 为了适应工业实践的不同需求,通过同时调整模型的深度depth和宽度width这两个参数,可以产生四种型号的YOLOX模型,即YOLOX-s、YOLOX-m、YOLOX-l和YOLOX-x. 这四种型号的YOLOX模型的检测精度依次提高,但所需的参数开销也随之增加.

2.2 注意力机制

注意力机制模拟了人类的视觉系统,可以在复杂的场景中聚焦重要的区域并忽略无关信息[9]. 注意力机制会根据输入特征图中各部分区域对输出结果的影响程度分配不同的权重以突出需要强调的特征信息,其结果通常都是以概率图或者概率特征向量的形式展示,并与原始输入的特征图进行相乘完成自适应特征加权[10]. 近年来,注意力机制在计算机视觉任务中得到了广泛应用. 添加注意力机制能够选择性地强调显著特征并抑制不相关的背景信息,使得检测网络可以更好地捕获和利用目标对象的特征[11].

3 本文方法

3.1 Night-YOLOX设计思想

本文所提出的Night-YOLOX是在设计两个关键模

块的基础上再与基本框架YOLOX[6]相结合,从而提高低照度场景下的目标检测性能. Night-YOLOX模型结构如图1所示. 首先,为了提高检测模型对低照度图像中目标对象的定位能力,设计了一个低级特征聚集模块(Low-level Feature Gathering Module, LFGM)与主干网络合并来共同训练. 在低照度场景下捕获更多有效的低级特征有利于定位目标[12],低级特征聚集模块通过将主干网络的浅层卷积阶段(即Stage1、Stage2和Stage3)输出的特征图(即$C1$、$C2$和$C3$)进行融合,并利用一组特定的池化层来提取更具判别性的低级特征. 这些有效的低级特征会被送到高级特征图中以及深层卷积阶段中(Stage4和Stage5),从而补偿对低照度图像进行特征提取过程中边缘、轮廓和纹理等低级特征的损失. 然后,为了提高检测模型对低照度图像中目标对象的分类能力,设计一种注意力引导块(Attention Guidance Block, AGB)嵌入检测模型的颈部结构中. 通过添加注意力引导块,生成注意力权重对特征图进行特征加权,进一步增强特征图的特征表示能力并减少噪声干扰、忽略黑暗背景等不相关信息. 由此引导检测模型推断出特征图中完整的对象区域范围,并捕获相应区域内的对象特征信息,以提高检测模型识别目标类别的准确性.

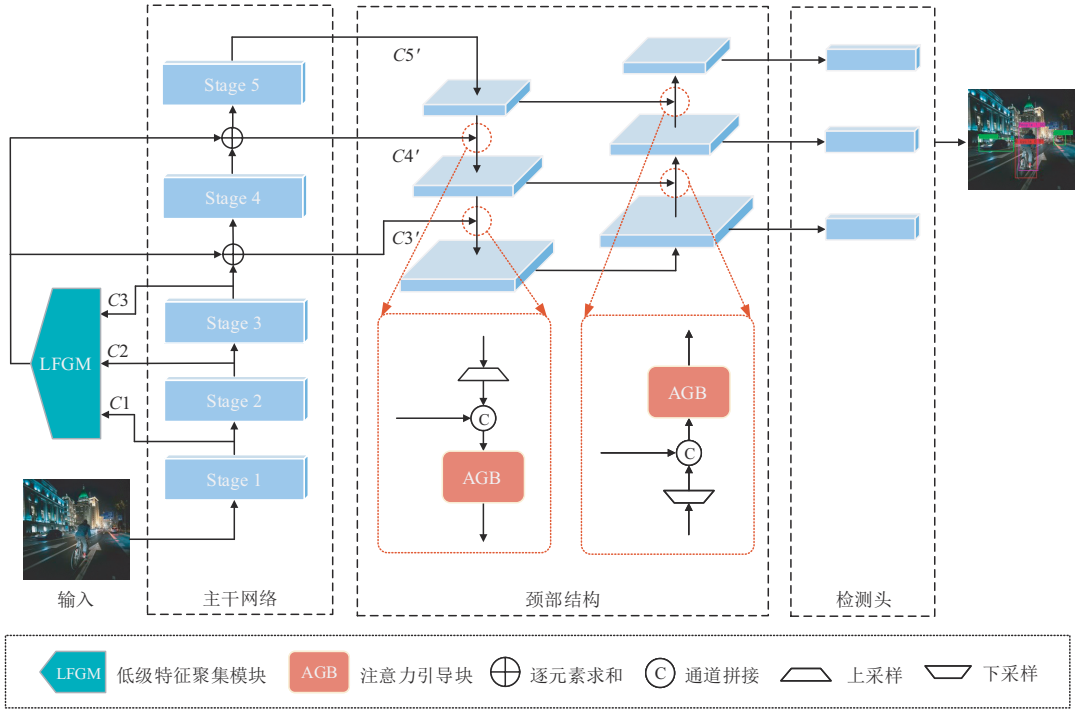

图1　Night-YOLOX模型结构图

3.2　低级特征聚集模块

受照明不均的影响,低照度图像中存在很多黑暗区域,目标对象往往隐藏在其中. 在这种情况下,主干网络能够提取到的目标对象纹理、边缘和轮廓等低级特征就非常少,并且提取到的低级特征在更深层卷积阶段中很容易受到卷积操作的影响而被融合到黑色背景中[7]. 为了解决这个问题,本文设计了一个低级特征聚集模块与主干网络合并,通过充分利用浅层卷积阶段输出的特征图中所包含的低级特征,以补偿高级特征图和深层卷积阶段中丢失的低级特征.

低级特征聚集模块的结构如图2所示. 主干网络由5个卷积阶段组成,使用Cl来表示第l个卷积阶段输出的特征图,其中$l\in[1,5]$. 低级特征聚集模块首先将浅层卷积阶段输出的三个特征图(即$C1$、$C2$和$C3$)作为输入,并按顺序进行特征融合生成一个聚合特征图$\boldsymbol{G}\in\mathbf{R}^{C\times H\times W}$,使其能够保留更多的低级特征信息. 融合过程可用式(1)表示:

$$\boldsymbol{G}=f_3\left(f_3(C1)\oplus C2\right)\oplus C3 \tag{1}$$

其中,f_3表示3×3卷积层,用于进行特征图分辨率大小和通道数的匹配,$\oplus$表示逐元素求和操作.

然后,聚合特征图$\boldsymbol{G}$被送入一个池化组处理. 该池化组旨在通过各种池化层来捕获聚合特征图$\boldsymbol{G}$中不同范围的上下文信息,以提取更具判别性的低级特征. 池化组由四个平行路径组成,分别是:1×W条带池化层、H×1条带池化层、S×S空间池化层和残差连接. 对于长宽分别为H和W的聚合特征图$\boldsymbol{G}$,引入池化范围为$(1,W)$和$(H,1)$的条带池化层. 条带池化层对特征图所有行或列中的特征值求平均,能够沿垂直或水平方向对特征图进

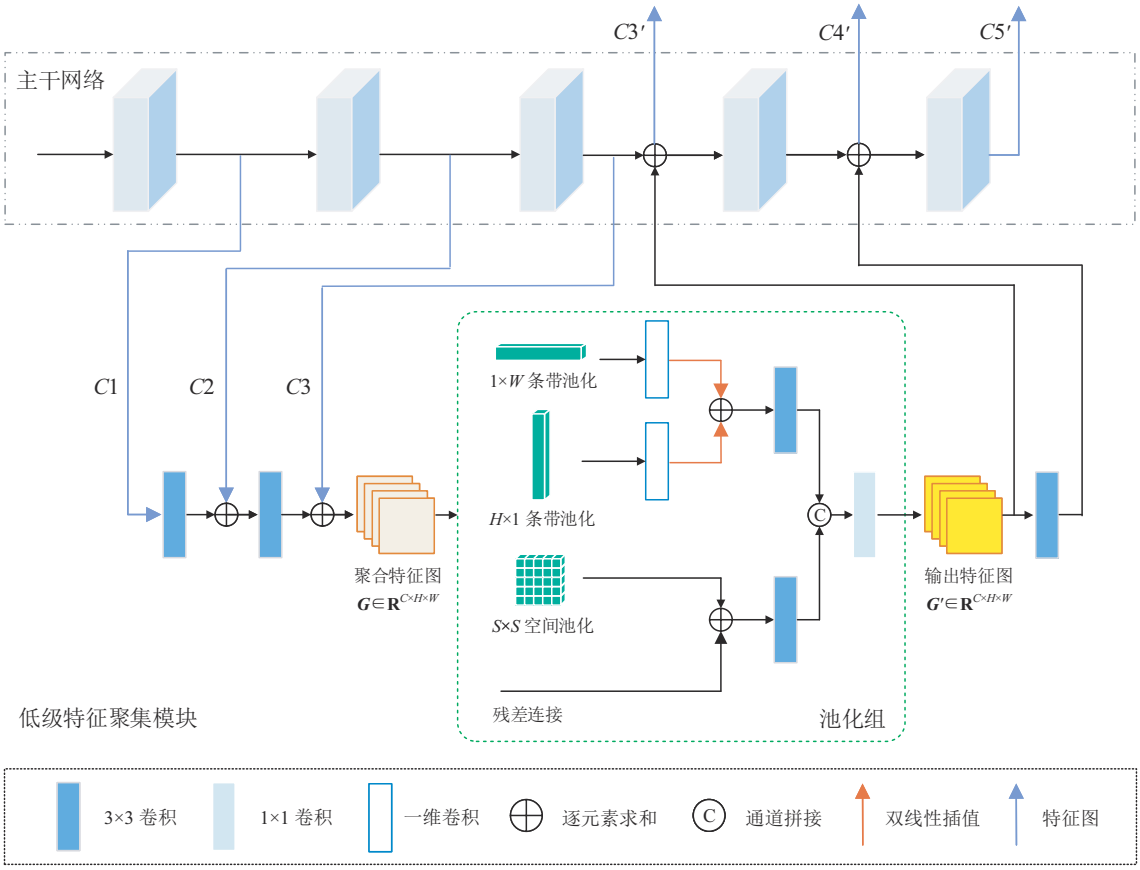

图2 低级特征聚集模块(LFGM)结构图

行压缩并编码特征信息. 利用两个条带池化层分别沿着垂直和水平的空间维度在离散分布的对象特征区域之间建立长距离依赖关系并进行编码,侧重于在全局上捕获目标对象的边缘和轮廓低级特征信息. 1×W和H×1条带池化操作可分别用式(2)和式(3)表示:

$$y_w=\frac{1}{H}\sum_{0\leqslant i<H}\boldsymbol{G}(i,W) \tag{2}$$

$$y_h=\frac{1}{W}\sum_{0\leqslant j<W}\boldsymbol{G}(H,j) \tag{3}$$

其中,$y_w\in\mathbf{R}^{C\times 1\times W}$以及$y_h\in\mathbf{R}^{C\times H\times 1}$分别表示通过1×$W$和$H$×1条带池化操作输出的特征张量.

条带池化层后跟着一个一维卷积层,用于整合特征张量内部相邻的特征信息. 接着通过双线性插值操作恢复两个特征张量y_w和y_h的空间大小,并通过逐元素求和操作将它们融合,生成一个包含着丰富的边缘、轮廓低级特征信息的特征张量$\boldsymbol{z}_1$. 该过程可用式(4)表示:

$$\boldsymbol{z}_1=F_{\mathrm{ex}}\left(f'\left(y_w\right)\right)\oplus F_{\mathrm{ex}}\left(f'\left(y_h\right)\right) \tag{4}$$

其中,F_{ex}表示双线性插值操作,f'表示卷积核大小为3的一维卷积层,⊕表示逐元素求和.

另外,引入池化范围是S×S的空间池化层,是为了通过方形的池化窗口来检测分布紧密的目标对象的特征区域,侧重于在局部上捕获目标对象的纹理低级特征信息. 根据4.3.1节图4中的实验结果,本文设置S=5. 第四个平行路径是保留了聚合特征图$\boldsymbol{G}$中原始空间信息的残差连接. 将第三和第四个平行路径的输出进行融合,生成一个包含着丰富的纹理低级特征信息的特征张量$\boldsymbol{z}_2$. 该过程可用式(5)表示:

$$\boldsymbol{z}_2=P_s(\boldsymbol{G})\oplus\boldsymbol{G} \tag{5}$$

其中,P_s表示S×S空间池化层,⊕表示逐元素求和.

在获得特征张量$\boldsymbol{z}_1$和$\boldsymbol{z}_2$后,使用3×3卷积层来提取捕获到的低级特征,并通过拼接操作以生成包含更多更具判别性的低级特征信息的特征图$\boldsymbol{G}'\in\mathbf{R}^{C\times H\times W}$. 该过程可用式(6)表示:

$$\boldsymbol{G}'=f_1\left(\left[f_3\left(\boldsymbol{z}_1\right);f_3\left(\boldsymbol{z}_2\right)\right]\right) \tag{6}$$

其中,f_1表示1×1卷积层,f_3表示3×3卷积层,[;]表示沿通道维度的拼接操作.

最后,将特征图$\boldsymbol{G}'$与主干网络的深层卷积阶段输出的特征图$C3$和$C4$融合,并将特征融合结果送入后续的卷积阶段,以补偿低级特征的损失. 由此生成的特征图$C3'$、$C4'$和$C5'$可以保留更多有效的边缘、轮廓和纹理等低级特征信息,从而有利于提高检测模型的目标定位能力. 生成特征图$C3'$、$C4'$和$C5'$的过程可分别用

式(7)~(9)表示:

$$C3'=C3\oplus \boldsymbol{G}' \tag{7}$$

$$C4'=f_3(\boldsymbol{G}')\oplus F^4_{\text{conv}}(C3') \tag{8}$$

$$C5'=F^5_{\text{conv}}(C4') \tag{9}$$

其中,f_3表示3×3卷积层,F^l_{conv}表示第l个卷积阶段,⊕表示逐元素求和.

3.3　注意力引导块

低照度图像不可避免的存在噪声干扰,在这种情况下检测模型难以推断出完整的对象特征区域,并且被黑暗背景遮挡或覆盖的区域内的对象特征信息往往容易被忽略. 为了解决这个问题,本文设计了一种嵌入在检测模型颈部结构中的注意力引导块,由此引导检测模型推断出完整的对象区域范围,并提取相应区域内的对象特征信息,以提高检测模型识别目标类别的能力. 注意力引导块的结构如图3所示,它由两个主要部分组成:区域捕获分支和特征采集分支.

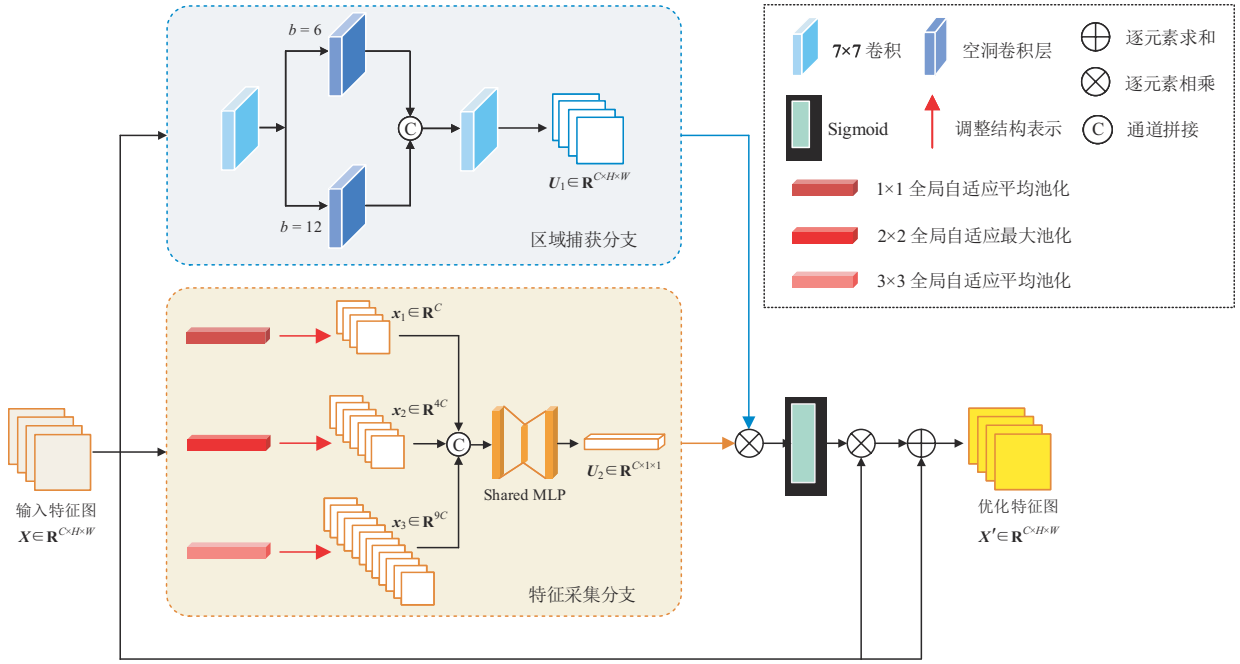

图3　注意力引导块(AGB)结构图

区域捕获分支旨在探索对象特征区域的范围大小. 在该分支中,首先使用一个卷积核较大的卷积层(即7×7卷积层)进行局部特征提取,对输入特征图$\boldsymbol{X}\in\mathbf{R}^{C\times H\times W}$中需要突出显示的特征区域信息进行学习. 接着在7×7卷积层之后平行放置两个空洞卷积层. 根据4.3.2节表1中的实验结果,两个空洞卷积层的扩张率分别设置为6和12. 通过精心设计的扩张率数值有效地控制了感受野的大小. 将两个空洞卷积层的输出结果进行拼接操作,便于聚合更广范围的上下文信息,进一步增强检测模型探索目标对象更广范围的特征区域的能力. 最后,利用一个7×7卷积层对聚合的全局上下文信息进行编码以捕获对象区域,生成一个区域注意映射图$\boldsymbol{U}_1\in\mathbf{R}^{C\times H\times W}$. 区域注意映射图$\boldsymbol{U}_1$的计算如式(10)所示:

$$\boldsymbol{U}_1=f_7\left(\left[F^6_{\text{atr}}\left(f_7(\boldsymbol{X})\right);F^{12}_{\text{atr}}\left(f_7(\boldsymbol{X})\right)\right]\right) \tag{10}$$

其中,[;]表示通道拼接操作,F^b_{atr}表示扩张率为b的空洞卷积层,f_7表示7×7卷积层.

特征采集分支旨在通过跨维(即通道、空间宽度和空间高度)相互作用来提取目标对象更多的特征信息. 在该分支中,首先横向添加三种全局自适应池化操作将输入特征图$\boldsymbol{X}$的全局空间特征信息压缩到三个不同范围尺寸的特征张量中. 具体实现过程是,添加一个1×1全局自适应平均池化操作实现压缩全局空间信息;添加一个尺寸较大的3×3全局自适应平均池化操作以保留更多的特征表示;添加一个2×2的全局自适应最大池化操作来关注特征图的结构化信息,并平衡其它两个尺寸全局自适应平均池化获得的特征表示. 接着,通过重新调整结构表示的操作(resize operation)将生成的三个特征张量的结构表示从张量转变为向量,使空间维度上压缩的特征信息完成跨维交互与通道维度上保留的对象特征相结合,由此生成三个一维特征向量,分别记为$\boldsymbol{x}_1\in\mathbf{R}^C$、$\boldsymbol{x}_2\in\mathbf{R}^{4C}$、$\boldsymbol{x}_3\in\mathbf{R}^{9C}$. 将这三个一维特征向量进行拼接,可以获得一个聚合了丰富的跨维交互特征信息的一维特征向量$\boldsymbol{X}_c\in\mathbf{R}^{14C}$. 该过程可以用式(11)表示:

$$\boldsymbol{X}_c=\left[F_{\text{re}}\left(P^1_{\text{avg}}(\boldsymbol{X})\right);F_{\text{re}}\left(P^2_{\text{max}}(\boldsymbol{X})\right);F_{\text{re}}\left(P^3_{\text{avg}}(\boldsymbol{X})\right)\right] \tag{11}$$

其中，P^n_{avg}表示$n\times n$全局自适应平均池化操作，$P^n_{\max}$表示$n\times n$全局自适应最大池化操作，F_{re}表示重新调整操作(resize operation)，[；；]表示通道拼接操作.

然后，特征向量$\boldsymbol{X}_c$被转发到多层感知机(Multi-Layer Perceptron，MLP)中进行特征编码，生成一个特征描述符$\boldsymbol{U}_2\in\mathbf{R}^{C\times1\times1}$. 多层感知机由两个全连接层和一个非线性激活函数ReLU组成，为了减少参数开销，降维比例r设置为32. 特征描述符$\boldsymbol{U}_2$的计算如式(12)所示：

$$\boldsymbol{U}_2=\text{MLP}\left(\boldsymbol{X}_c\right)=F_1\left(\delta\left(F_0\left(\boldsymbol{X}_c\right)\right)\right) \tag{12}$$

其中，$F_0\in\mathbf{R}^{C/r\times C}$和$F_1\in\mathbf{R}^{C\times C/r}$表示两个不同的全连接层，$\delta$表示ReLU激活函数.

最后，通过逐元素相乘将区域注意映射图$\boldsymbol{U}_1$和特征描述符$\boldsymbol{U}_2$结合，并采用Sigmoid激活函数将输出结果归一化至(0,1)范围，生成注意力权重$\boldsymbol{M}$. 对原始输入特征图$\boldsymbol{X}$进行特征加权以实现目标特征的自适应优化并突出显示完整的对象区域，同时减少噪声干扰、忽视黑暗背景等不相关的信息. 整个过程可由式(13)和式(14)表示：

$$\boldsymbol{M}=\sigma\left(\boldsymbol{U}_1\otimes\boldsymbol{U}_2\right) \tag{13}$$

$$\boldsymbol{X}'=\boldsymbol{X}\oplus\left(\boldsymbol{X}\otimes\boldsymbol{M}\right) \tag{14}$$

其中，$\otimes$表示逐元素相乘，σ表示Sigmoid激活函数，$\oplus$表示逐元素求和，$\boldsymbol{X}'$表示最终输出的优化特征图.

4 实验及结果分析

4.1 数据集处理及评价指标

本文在真实低照度图像数据集ExDark[8]上进行实验. ExDark数据集包含12个对象类别，分别是：自行车、船、瓶子、公交车、汽车、猫、椅子、杯子、狗、摩托车、人和桌子，每张低照度图像都配有对应的对象级注释标签文件. 本文首先将ExDark数据集以8∶2的比例随机划分为训练-验证集和测试集，然后将训练-验证集以9∶1的比例随机划分为训练集和验证集.

根据标准评估指标，本文使用平均均值精度(mean Average Precision，mAP)来衡量目标检测模型的性能. 在与其它目标检测方法进行对比实验时，本文还列出了数据集中每个对象类别的检测精度，该检测精度使用平均精度(Average Precision，AP)指标来衡量.

4.2 实现细节

本文基于Windows Server 2016操作系统使用NVIDIA Tesla P40 GPU对模型进行训练. 所提出的方法使用随机梯度下降(Stochastic Gradient Descent，SGD)进行50个epoch的训练，权值衰减和动量分别设置为0.000 5和0.9. 输入的训练图像的尺寸大小会被调整为608×608. 在训练过程中，批次大小设置为8，学习率的初始值为0.001. 本文使用余弦退火策略[13]来调整学习速率，最大迭代次数(*T*_max)的值设置为5，最小学习速率的值(eta_min)设置为0.000 01. 使用在ImageNet[14]上预训练的权重来初始化主干网络. 与YOLOX[6]相似，通过同时调整深度depth和宽度width两个参数可以生成s、m、l和x四个型号大小的Night-YOLOX模型，其中，depth=｛‘s’：0.33，‘m’：0.67，‘l’：1.00，‘x’：1.33｝，width=｛‘s’：0.50，‘m’：0.75，‘l’：1.00，‘x’：1.25｝. 在测试阶段，阈值为0.5的非极大值抑制(Non-Maximum Suppression，NMS)将分别应用于每个对象类别.

4.3 消融实验

4.3.1 空间池化层核大小*S*的设置

图4展示了设置空间池化层的核大小S为不同的值时，低级特征聚集模块对提高检测模型性能的影响. 可以看到，当S=5时，可以获得最佳的mAP，在s、m、l和x四个型号的基本框架YOLOX[6]上分别实现了2.8%、2.6%、1.9%和2.3%的mAP提升. 虽然当S=3和S=7时都实现了检测精度的提升，但是均比不上S=5时带来的性能提升，而且S=7时，获得的检测精度提升最小. 根据实验结果，本文认为如果内核S太小，则可以聚合的特征信息比较稀少；但是内核S太大时，可能会包含来自不相关区域里的无用信息. 所以，本文选择5×5空间池化层作为低级特征聚集模块中的默认参数设置.

4.3.2 空洞卷积的扩张率*b*设置

如表1所示，本节验证了注意力引导块中两个空洞卷积层的扩张率的值对提高检测性能的影响. 实验使用了基本框架YOLOX-l. 为了控制不同范围的感受野，将两个空洞卷积层的扩张率设置为逐渐增加. 当两个空洞卷积层的扩张率分别设置为6和12时，实现了最佳的目标检测精度，达到了75.1% mAP. 然而，当b=3，6或b=12，18时，只是稍微提升了一点检测性能，mAP提升不到1%. 根据实验结果，本文认为当设置b=3，6时，由于感受野范围受限，检测模型无法捕获更广范围的上下文信息；当b=12，18时，过大的感受野范围容易包含不相关的黑色背景，干扰检测模型推断目标对象的特征区域大小. 所以，本文使用b=6，12作为默认参数设置，让检测模型在有效的感受野范围内更好地捕获目标对象的特征区域，从而提高目标检测精度.

表1 空洞卷积的扩张率b对检测性能的影响

方法	参数量/M	mAP/%
YOLOX-l	54.2	72.2
YOLOX-l + AGB (b=3, 6)	66.9	72.8
YOLOX-l + AGB (b=6, 12)	66.9	**75.1**
YOLOX-l + AGB (b =12, 18)	66.9	72.9

4.3.3 注意力引导块的有效性

将注意力引导块嵌入到基本框架上进行训练和测

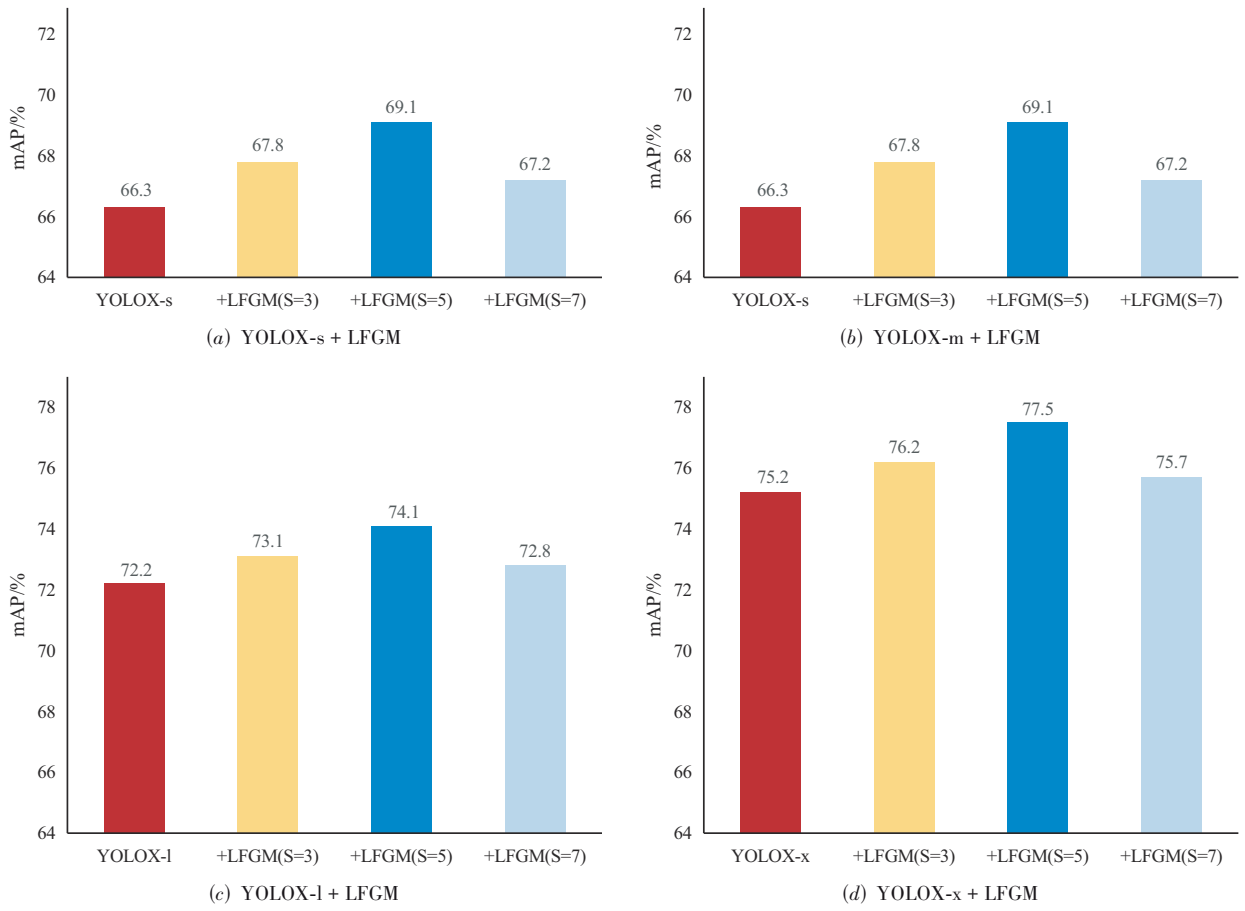

图4　空间池化层核大小S的值对检测精度的影响

试,以评估该模块的有效性. 实验使用了基本框架YOLOX-x. 图5展示了特征图的热力图可视化效果,与基本框架相比,添加注意力引导块后,检测模型可以更准确地聚焦在目标对象上,捕获的特征区域的范围大小可以包含更完整的目标对象.

4.3.4　两个模块结合的效果

在基本框架YOLOX[6]上同时应用低级特征聚集模块和注意力引导块,实验结果如表2所示. 可以看出,在可承受的参数开销范围内结合两个模块,能够更进一步提升检测模型的检测精度. 根据实验结果,结合两个模块后,虽然计算量和参数量有了一定程度的增加,但是mAP得到了明显提升.

4.4　与其它注意力模块的对比

为了验证注意力引导块相比其它注意力模块,例如

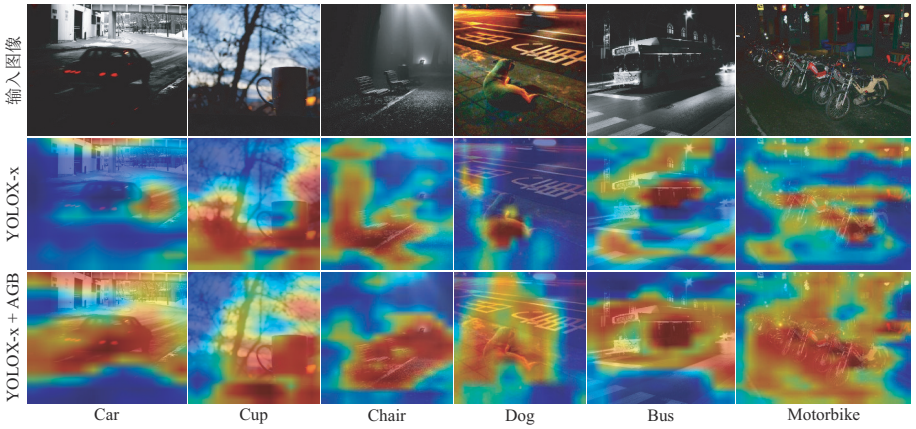

图5　添加注意力引导块(AGB)前/后特征图的热力图可视化对比

表2 结合两个模块对检测性能的影响

方法	FLOPs/G	参数量/M	mAP/%
YOLOX-s	12.03	8.9	66.3
YOLOX-s + LFGM + AGB	15.43	12.6	70.4
YOLOX-m	33.18	25.3	70.3
YOLOX-m+ LFGM + AGB	40.82	33.5	74.1
YOLOX-l	70.11	54.2	72.2
YOLOX-l + LFGM + AGB	83.68	68.8	76.2
YOLOX-x	127.06	99.0	75.2
YOLOX-x + LFGM + AGB	148.26	121.9	**79.3**

挤压激励(Squeeze and Excitation, SE)[15]、瓶颈注意力模块(Bottleneck Attention Module, BAM)[16]和坐标注意力(Coordinate Attention, CA)[17]在低照度场景中的优势,本节进行了对比实验. 实验使用基本框架YOLOX-x,实验结果如表3所示. 当使用所提出的注意力引导块时,检测模型在大多数对象类别中获得了最佳AP分数. 模型的检测精度从75.2% mAP上升到78.0% mAP,提高了2.8%. 相比之下,当其它注意力模块嵌入检测模型时,检测精度的提升是有限的,只提高了不到0.7% mAP. 本文认为原因在于前人的工作主要是针对正常照明图像,在设计注意力模块时忽略了低照度图像可见度低和曝光不足等问题. 因此,现有的注意力模块难以引导模型在低照度图像中提取更多有效的特征信息.

表3 嵌入不同注意力模块对检测性能的影响

方法	Bicycle	Boat	Bottle	Bus	Car	Cat	Chair	Cup	Dog	Mbike	People	Table	mAP/%
YOLOX-x	88.4	72.8	64.9	**93.5**	78.8	70.6	68.9	76.2	77.6	76.2	77.3	56.9	75.2
YOLOX-x + SE[15]	87.9	72.1	74.6	90.6	76.1	77.8	69.2	**77.9**	75.6	74.1	77.7	54.8	75.7
YOLOX-x + BAM[16]	90.9	73.3	74.2	87.5	76.2	74.3	66.8	76.2	76.8	**79.1**	76.7	58.6	75.9
YOLOX-x + CA[17]	88.3	68.8	70.6	91.9	**78.8**	76.7	66.1	77.2	77.7	78.9	77.4	56.0	75.7
YOLOX-x +AGB (Ours)	**91.5**	**73.8**	**75.1**	93.3	78.4	**82.2**	**69.7**	76.3	**80.2**	76.4	**78.7**	**60.7**	**78.0**

4.5 在ExDark数据集上的实验结果

本节使用真实低照度图像数据集ExDark[8]来评估所提出的Night-YOLOX的检测性能,并将其与其它主流的目标检测方法进行比较. 在实验中,所有的目标检测方法都在ExDark数据集上进行了训练和测试. 由于低照度图像存在光照分布不均和能见度低等问题,现有的目标检测方法的检测精度并不理想. 实验结果如表4所示,除了YOLOv4[5]、DetectoRS[4]、GFLv2[18]和YOLOX[6]之外,大多数目标检测方法的mAP得分均低于70%. 相比之下,本文所提出的Night-YOLOX的检测性能远远优于其它方法. 特别是,Night-YOLOX-x模型达到了最高的79.3% mAP,与相应的基本框架YOLOX-x相比,mAP提升了大约4%. 此外,表4中还展示了每个对象类别的检测精度,通过AP指标来衡量. 与其它目标检测方法相比,本文所提出的Night-YOLOX在大多数对象类别中都实现了最佳的AP分数.

4.6 低照度目标检测的可视化效果

图6展示了本文所提出的Night-YOLOX和其它主流的目标检测方法在ExDark[8]数据集上的检测结果可视化的对比. 可以观察到,本文所提出的Night-YOLOX可以准确地检测到低照度图像中的目标对象,而其它检测方法可能存在误检或漏检的现象. 具体来说,有些

表4 本文方法与其它主流的目标检测方法在ExDark数据集上检测精度的对比 单位:%

方法	Bicycle	Boat	Bottle	Bus	Car	Cat	Chair	Cup	Dog	Mbike	People	Table	mAP
Dynamic R-CNN[2]	80.1	59.4	62.0	87.2	66.5	64.6	52.0	66.0	73.3	65.7	68.0	45.8	65.9
EfficientDet-D0[19]	79.6	51.5	52.9	71.6	59.2	51.5	55.4	60.5	71.8	53.2	60.8	36.6	58.7
SABL[20]	78.8	64.3	63.4	87.8	69.9	69.1	51.7	65.3	71.7	68.2	68.6	51.0	58.7
YOLOv4[5]	83.3	54.4	71.4	93.5	54.7	**83.5**	**71.5**	77.3	61.9	69.3	71.3	59.2	70.9
VarifocalNet[21]	77.8	61.1	60.6	87.5	64.0	60.2	48.2	62.5	73.3	56.7	65.9	46.4	63.7
Sparse R-CNN[3]	81.8	63.2	66.5	87.3	65.3	71.4	58.8	69.7	**83.0**	67.7	65.0	49.3	69.1
YOLOF[22]	78.6	61.6	62.2	90.4	60.9	71.4	55.3	67.1	74.8	64.4	62.3	49.3	66.5
DetectoRS[4]	83.6	67.5	67.6	89.5	70.7	71.0	61.6	69.2	75.8	74.1	71.7	54.6	71.4
GFLv2[18]	82.6	66.7	71.7	88.0	73.2	67.7	55.8	72.2	74.2	65.7	72.4	49.8	70.0
FCOS[23]	81.8	66.3	65.7	86.4	67.1	71.7	58.4	71.4	80.6	63.2	67.6	51.0	69.3
FreeAnchor[24]	78.5	56.9	60.3	85.4	64.7	58.9	45.9	61.2	68.4	58.7	67.6	40.8	62.3
YOLOX-x[6]	88.4	72.8	64.9	93.5	78.8	70.6	68.9	76.2	77.6	76.2	77.3	56.9	75.2
Night-YOLOX-s	86.3	65.3	68.6	88.5	71.6	64.8	57.3	71.8	73.3	71.8	72.7	52.5	70.4
Night-YOLOX-m	83.9	70.4	73.2	93.5	75.3	80.0	65.3	77.0	76.2	69.8	73.1	51.9	74.1
Night-YOLOX-l	89.3	69.3	72.6	92.0	78.8	78.8	68.4	78.1	77.4	77.5	77.6	54.5	76.2
Night-YOLOX-x	**89.7**	**75.4**	**73.8**	**96.0**	**79.4**	80.2	71.4	**79.6**	82.0	**81.4**	**79.9**	**62.5**	**79.3**

目标对象可能会隐藏在图像的较暗角落里,但是本文所提出的方法可以通过绘制适当大小的边界框来定位它们. 此外,一些低照度图像中包含了各种类别的对象,但本文所提出的方法能够在低照度条件下正确辨别所有的对象类别. 尽管某些对象被其他物体遮挡或者仅显示部分特征,但仍然可以正常检测到它们.

图6　本文方法与其它主流的目标检测方法在ExDark数据集上检测结果可视化的对比

5　结束语

本文提出了一种基于Night-YOLOX的低照度目标检测方法,该方法提出并设计了低级特征聚集模块和注意力引导块两个关键模块. 实现步骤是先设计一个低级特征聚集模块与主干网络合并,以补偿在对低照度图像进行特征提取过程中边缘、轮廓和纹理等低级特征的损失,从而更好地在低照度场景下定位目标对象. 然后,设计一种注意力引导块嵌入检测模型的颈部结构中,以减少低照度图像中的噪声干扰,引导检测模型推断出特征图中完整的对象区域范围大小,并捕获相应区域内的对象特征信息,从而提高低照度环境下目标分类的准确性. 最后,在真实低照度图像数据集ExDark上进行实验验证,结果表明相比于其他主流的目标检测方法,所提出的Night-YOLOX在低照度环境下具有更好的检测精度和检测效果. 后续的工作将针对该模型的轻量化展开进一步研究,从而更容易部署在对时间具有敏感性的实际应用中.

参考文献

[1] HUANG Y, JIANG Z, LAN R, et al. Infrared image super-resolution via transfer learning and PSRGAN[J]. IEEE Signal Processing Letters, 2021, 28: 982-986.

[2] ZHANG H, CHANG H, MA B, et al. Dynamic R-CNN: Towards high quality object detection via dynamic training [C]//European Conference on Computer Vision. Berlin: Springer, 2020: 260-275.

[3] SUN P, ZHANG R, JIANG Y, et al. Sparse R-CNN: End-to-end object detection with learnable proposals[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 14454-14463.

[4] QIAO S, CHEN L C, YUILLE A. Detectors: Detecting objects with recursive feature pyramid and switchable atrous convolution[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 10213-10224.

[5] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOV4: Optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2022-12-05]. https://arxiv.org/abs/2004.10934.

[6] GE Z, LIU S, WANG F, et al. YOLOX: Exceeding YOLO series in 2021[EB/OL]. (2021-07-28)[2022-12-05]. https://arxiv.org/abs/2107.08430.

[7] WANG W, YANG W, LIU J. Hla-face: Joint high-low adaptation for low light face detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 16195-16204.

[8] 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.

[9] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Attention is all you need[EB/OL]. (2017-06-12)[2022-12-05]. https://arxiv.org/abs/1706.03762.
[10] ZHANG H, GOODFELLOW I, METAXAS D, et al. Self-attention generative adversarial networks[C]//International Conference on Machine Learning. New York: PMLR, 2019: 7354-7363.
[11] GUO M H, XU T X, LIU J J, et al. Attention mechanisms in computer vision: A survey[J]. Computational Visual Media, 2022, 8: 331-368.
[12] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 8759-8768.
[13] LOSHCHILOV I, HUTTER F. Sgdr: Stochastic gradient descent with warm restarts[EB/OL]. (2016-08-13)[2022-12-05]. https://arxiv.org/abs/1608.03983v3.
[14] DENG J, DONG W, SOCHER R, et al. Imagenet: A large-scale hierarchical image database[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2009: 248-255.
[15] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 7132-7141.
[16] PARK J, WOO S, LEE J Y, et al. Bam: Bottleneck attention module[EB/OL]. (2018-07-17)[2022-12-05]. https://arxiv.org/abs/1807.06514v1.
[17] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 13713-13722.
[18] LI X, WANG W, HU X, et al. Generalized focal loss v2: Learning reliable localization quality estimation for dense object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 11632-11641.
[19] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 10781-10790.
[20] WANG J, ZHANG W, CAO Y, et al. Side-aware boundary localization for more precise object detection[C]//European Conference on Computer Vision. Berlin: Springer, 2020: 403-419.
[21] ZHANG H, WANG Y, DAYOUB F, et al. Varifocalnet: An iou-aware dense object detector[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 8514-8523.
[22] CHEN Q, WANG Y, YANG T, et al. You only look one-level feature[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 13039-13048.
[23] TIAN Z, SHEN C, CHEN H, et al. Fcos: A simple and strong anchor-free object detector[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44 (4): 1922-1933.
[24] ZHANG X, WAN F, LIU C, et al. Learning to match anchors for visual object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(6): 3096-3109.

作者简介

江泽涛 男,1961年出生,江西九江人. 桂林电子科技大学计算机与信息安全学院教授、博士生导师. 主要研究方向为图像处理、计算机视觉.
E-mail: zetaojiang@guet.edu.cn

施道权(通讯作者) 男,1998年出生,广西横县人. 桂林电子科技大学计算机与信息安全学院硕士研究生. 主要研究方向为图像处理、计算机视觉.
E-mail: sdaoquan@qq.com

雷晓春 女,1981年出生,广西南宁人. 桂林电子科技大学计算机与信息安全学院高级实验师、硕士生导师. 主要研究方向为深度学习、计算机视觉.
E-mail: glleixiaochun@qq.com