

# 一种基于SAM-MSFF网络的低照度目标检测方法

江泽涛<sup>1</sup>, 李 慧<sup>1\*</sup>, 雷晓春<sup>1</sup>, 朱玲红<sup>2</sup>, 施道权<sup>1</sup>, 翟丰硕<sup>1</sup>

(1. 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; 2. 南昌航空大学, 江西南昌 330063)

**摘要:** 由于低照度图像具有对比度低、细节丢失严重、噪声大等缺点, 现有的目标检测算法对低照度图像的检测效果不理想. 为此, 本文提出一种结合空间感知注意力机制和多尺度特征融合 (Spatial-aware Attention Mechanism and Multi-Scale Feature Fusion, SAM-MSFF) 的低照度目标检测方法. 该方法首先通过多尺度交互内存金字塔融合多尺度特征, 增强低照度图像特征中的有效信息, 并设置内存向量存储样本的特征, 捕获样本之间的潜在关联性; 然后, 引入空间感知注意力机制获取特征在空间域的长距离上下文信息和局部信息, 从而增强低照度图像中的目标特征, 抑制背景信息和噪声的干扰; 最后, 利用多感受野增强模块扩张特征的感受野, 对具有不同感受野的特征进行分组重加权计算, 使检测网络根据输入的多尺度信息自适应地调整感受野的大小. 在ExDark数据集上进行实验, 本文方法的平均精度 (mean Average Precision, mAP) 达到 77.04%, 比现有的主流目标检测方法提高 2.6%~14.34%.

**关键词:** 低照度图像; 目标检测; 空间感知注意力机制; 多尺度特征融合; 多感受野增强模块

**基金项目:** 国家自然科学基金 (No.62172118, No.61876049); 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.GIIP2006, No.GIIP2007, No.GIIP2008); 广西研究生教育创新计划项目 (No.YCB2021070, No.YCBZ2018052, No.YCSW2022269, No.2021YCXSO71)

中图分类号: TP319

文献标识码: A

文章编号: 0372-2112(2024)01-0081-13

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20220666

## A Low-Light Object Detection Method Based on SAM-MSFF Network

JIANG Ze-tao<sup>1</sup>, LI Hui<sup>1\*</sup>, LEI Xiao-chun<sup>1</sup>, ZHU Ling-hong<sup>2</sup>, Shi Dao-quan<sup>1</sup>, ZHAI Feng-shuo<sup>1</sup>

(1. The Key Laboratory of Image and Graphic Intelligent Processing in Guangxi, Guilin University of Electronic Technology, Guilin, Guangxi 541004, China;

2. Nanchang Hangkong University, Nanchang, Jiangxi 330063, China)

**Abstract:** The existing object detection methods are insufficient for low-light images due to their intrinsic property such as low contrast, detail loss and high noise. To solve this problem, a low-light object detection method that combines spatial-aware attention mechanism with multi-scale feature fusion (SAM-MSFF) is proposed. Firstly, multi-scale features are fused by multi-scale interactive memory pyramid to enhance effective information under low-illumination condition, and features of memory vector storage samples are set to capture potential correlation between samples. Then, a spatial-aware attention mechanism is introduced to obtain long-distance context information and local information of features in spatial domain, thereby enhancing the object features in low-light images and suppressing the interference of background information and noise. Finally, multiple receptive field enhancement module is used to expand receptive field of the features, and the features with different receptive fields are grouped and re-weighted, so that detection network can adaptively adjust the size of receptive field according to input multi-scale information. Experimental results on the ExDark dataset show that mAP (mean Average Precision) of the proposed method reaches 77.04%, which is 2.6%~14.34% higher than existing mainstream object detection methods.

**Key words:** low-light images; object detection; spatial-aware attention mechanism; multi-scale feature fusion; multiple receptive field enhancement module

Foundation Item(s): National Nature Science Foundation of China (No.62172118, No.61876049); Nature Science Foundation of Guangxi (No.2021GXNSFDA196002); Guangxi Key Laboratory of Image and Graphic Intelligent Processing under Grants (No.GIIP2006, No.GIIP2007, No.GIIP2008); Innovation Project of Guangxi Graduate Education under Grants (No.YCB2021070, No.YCBZ2018052, No.YCSW2022269, No.2021YCXS071)

## 1 引言

目标检测是众多视觉任务的基础,被广泛应用到各种实际应用中,如自动驾驶<sup>[1]</sup>、工业检测<sup>[2]</sup>、无人机侦察<sup>[3]</sup>以及军事打击<sup>[4]</sup>等。在照度低、曝光不足的环境中拍摄出的图像会出现多种退化现象,其成像质量下降表现为亮度低、对比度低、图像噪声增多、细节丢失严重等。图像的退化不仅影响人类对图像的主观视觉体验,而且会降低目标检测器的性能。这可能为夜间目标识别、低照度环境下的工业检测等领域带来潜在的风险。

随着目标检测的发展,目标检测器主要分为双阶段目标检测器和单阶段目标检测器。但是这些检测器主要用来检测正常照度图像,而对低照度图像检测效果不理想。双阶段目标检测器以 Faster R-CNN<sup>[5]</sup> 为代表,对于 Faster R-CNN,在第一阶段,将主干网络提取到的特征图送入 RPN (Region Proposal Network) 中, RPN 生成大量的候选区域;在第二阶段,通过 RoI (Region of Interest) 操作从每个候选区域中提取特征,用于分类和边界框回归任务。单阶段目标检测器以 YOLO 系列<sup>[6-10]</sup> 为代表, YOLOv1 将输入图像划分为  $S \times S$  的网格,对每个网格预测  $N$  个边界框; YOLOv2-YOLOv4 使用 K-means 聚类方法生成锚框,然后根据锚框来预测边界框相对于先验框的偏移量; YOLOv5 采用 Mosaic 数据增强、自适应图片缩放等机制对输入的图像进行预处理,在训练过程中自适应地计算最佳锚框值。YOLOX 采用无锚机制进行预测,并将分类和定位任务解耦。为了满足不同的任务需求, YOLOX 分为 YOLOX-S, YOLOX-M, YOLOX-L, YOLOX-X 这 4 个不同规模的版本,其中, YOLOX-X 版本的参数量最大,对图像的检测精度最高。由于双阶段检测器产生大量的候选框,并且在第二阶段需要对候选框进行二次修正,所以双阶段检测器检测精度高,但是检测速度较慢。单阶段检测器省略了产生候选框的过程,节约了时间成本。因此,单阶段目标检测器速度快,但是精度较低。在近几年的发展中,单阶段目标检测器可以很好地权衡精度和速度,在工业中被大量使用。

针对现有目标检测器对低照度图像检测效果不佳的问题, Wang 等<sup>[11]</sup> 提出一种正常照度图像和低照度图像联合适应的框架,通过渐进式域适应来减小正常照度图像和低照度图像之间的语义差距。但是该方法只针对低照度场景中的人脸检测,缺少泛化性,而且检测精度较低。此外,该方法的网络结构过于复杂,不能进

行端到端的训练。Miao 等<sup>[12]</sup> 采用先增强后检测的两步策略,首先利用 MSR (Multi-Scale Retinex) 算法增强夜间图像,提升图像的对比度和细节信息,然后将图像送入 YOLOv3 网络中进行检测。该方法无法满足实时检测的要求,不能对探测器的性能带来实质性的改善。之后, Xiao 等<sup>[13]</sup> 在 RFB-Net 的基础上,提出一种具有特殊设计的特征金字塔网络和上下文融合网络的夜视探测器。该检测器虽然能够满足实时监测的要求,但是对低照度图像的检测精度仍有待提高。目前,低照度目标检测的研究成果较少,具有很大的研究空间。

本文针对低照度图像存在对比度低、细节丢失严重、噪声大等缺点,导致现有的目标检测器对低照度图像检测效果极差的现象,提出一种结合空间感知注意力机制和多尺度特征融合 (Spatial-aware Attention Mechanism and Multi-Scale Feature Fusion, SAM-MSFF) 的低照度目标检测方法。其主要贡献如下。

(1) 提出多尺度交互内存金字塔结构。通过融合多层次多尺度的特征来增强低照度图像特征中的有效信息,有利于后续模块能够提取到关键的特征信息;并且设置内存向量存储样本的特征,建立不同样本之间的潜在关联性,从而抽象出属于同一类别物体之间相似的高级特征,提升模型的泛化能力。

(2) 提出空间感知注意力机制。通过建立部分像素之间的关系来获取低照度图像特征中的长距离上下文信息,利用池化和卷积操作来提取低照度图像特征中的局部信息;然后将长距离上下文信息和局部信息编码到低照度图像特征中,从而突出低照度图像中的目标物体,增强低照度图像中的目标物体和背景区域的可辨别性。

(3) 提出多感受野增强模块,该模块采用具有不同扩张率的空洞卷积分组并行提取低照度图像特征;然后为每组特征赋予不同的权重,使网络根据目标物体的尺度自适应地调整感受野的大小;最后将包含不同尺度信息的特征进行融合,提高对目标识别的效率。

实验结果表明,该方法能够进行端到端训练并且有效提升目标检测器对低照度图像的目标检测性能。

## 2 相关工作

为了增强目标检测器对低照度图像特征的提取能

力,本文主要从低照度图像特征增强和低照度图像特征融合两方面进行改进.使用注意力机制可以增强低照度图像中的目标特征,减少背景噪音的干扰.空洞卷积在扩大感受野的同时保持特征图的大小不变,能够减少特征信息丢失.因此,本文在特征融合模块中采用空洞卷积来扩张感受野,融合不同感受野的特征,从而自适应地检测低照度图像中不同大小的物体.

## 2.1 注意力机制

注意力机制在很多视觉任务中起着重要作用,如目标检测、图像分割等.注意力机制通过在特征图中添加权重,从而突出特征中的重要信息,使神经网络专注于感兴趣的区域.近年来,注意力机制蓬勃发展,衍生了很多注意力方法.SENet(Squeeze-and-Excitation Networks)<sup>[14]</sup>将二维的特征图压缩成一维的向量,从而高效地构建通道之间的依赖关系.CBAM(Convolutional Block Attention Module)<sup>[15]</sup>由通道注意力机制和空间注意力机制两部分组成,分别从通道和空间两个维度对特征进行优化.非局部注意力机制<sup>[16]</sup>通过建立特征图中每对像素之间的关系来捕获长距离依赖关系.

## 2.2 空洞卷积

当识别尺寸较大的物体时,特征层需要更大的感受野.通常情况下,我们通过增加卷积层或下采样层来增大特征层的感受野.但是,一方面,简单增加卷积层的数量会引入更多的参数,增加计算成本,并且过深的网络由于优化困难会导致过拟合;另一方面,增加下采样层会减小特征层的大小,从而导致信息丢失.因此,如何在保持高分辨率特征图的同时构建一个能够实现大感受野的模型一直是目标检测的关键问题.空洞卷积可以在增加感受野的同时保持特征图的尺寸不变.

与正常卷积不同,空洞卷积引入了一个被称为“扩张率  $r$ ”的超参数.该参数定义了在处理数据时卷积核各值之间的间距,空洞卷积在普通卷积的卷积核中的非零值之间插入  $r-1$  个零值.空洞卷积的感受野计算公式为

$$k' = k + (k-1) \times (r-1) \quad (1)$$

其中,  $k$  为空洞卷积核的大小;  $k'$  为与空洞卷积感受野等效的普通卷积的卷积核大小.空洞卷积后特征图大小的计算公式为

$$f = \left\lfloor \frac{i + 2p - k - (k-1) \times (r-1)}{s} \right\rfloor + 1 \quad (2)$$

其中,  $i$  为原始特征图的大小;  $p$  为空洞卷积的填充数;  $s$  为空洞卷积的步长;  $f$  为特征图的大小.

## 3 本文方法

### 3.1 SAM-MSFF 网络框架

YOLOv4对正常照度图像目标检测精度较高,而对低照度图像目标检测效果很差.其根本原因是低照度图像对比度低、信噪比低、图像信息损失严重,导致YOLOv4无法提取到丰富的特征信息,从而影响了目标检测效果.为了更好地提取到低照度下的图像目标特征信息,本文提出一种SAM-MSFF的网络框架.该框架以YOLOv4为基础,保留YOLOv4的骨干网络CSPDarknet(Cross Stage Partial Darknet)53<sup>[17]</sup>、空间金字塔池化模块<sup>[18]</sup>(Spatial Pyramid Pooling, SPP)和检测头3个部分.在此基础上,增加多尺度交互内存金字塔(Multi-scale Interactive memory Pyramid, MIP)、空间感知注意力机制(Spatial-aware Attention Mechanism, SAM)和多感受野增强模块(Multiple Receptive field Enhancement Module, MREM),其网络结构如图1所示.

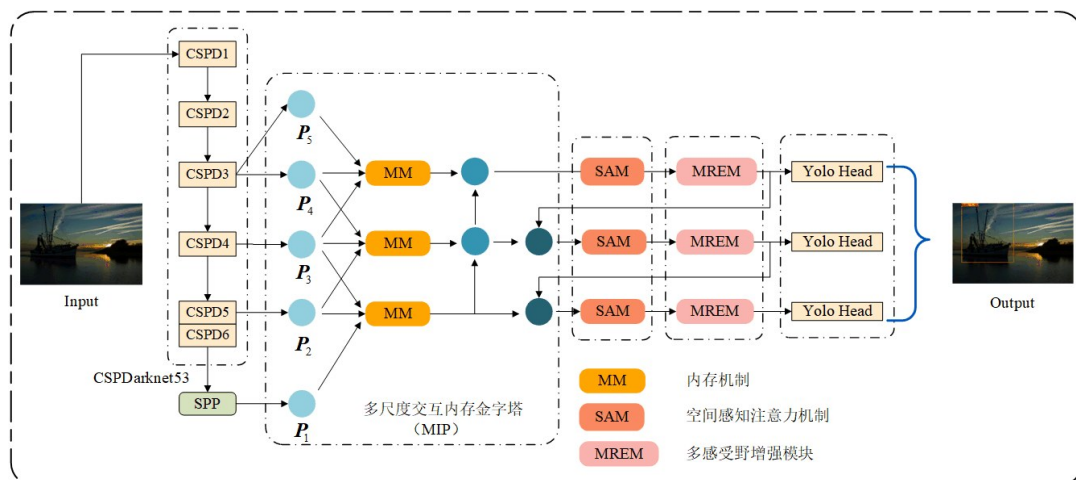


图1 SAM-MSFF网络框架

首先,将低照度图像送入CSPDarknet53网络和SPP网络中提取特征;其次,通过多尺度交互内存金字塔网

络对提取到的特征进行多层次多尺度融合,增强低照度图像特征的代表能力,并设置内存向量保存样本的

特征,建立样本之间的关系;然后,利用空间感知注意力机制同时捕获特征长距离依赖关系和局部依赖关系,通过结合长距离信息和局部信息,有效抑制背景信息和噪声的干扰,从而突出低照度图像特征中的目标区域;接着,引入多感受野增强模块自适应调整特征的感受野,提升低照度目标检测器对不同尺度物体的检测能力;最后,将特征送入检测头,由检测头执行对目标的定位和完成目标检测任务。

### 3.2 多尺度交互内存金字塔

由于低照度图像细节丢失严重,CSPDarknet53网络提取到的特征包含的有效信息较少.为了增强低照度图像特征的代表能力,本文提出一种多尺度交互内存金字塔模块.目前,常用的特征金字塔网络包括FPN(Feature Pyramid Networks)<sup>[19]</sup>,PANet(Path Aggregation Network)<sup>[20]</sup>,BiFPN(Bidirectional Feature Pyramid Network)<sup>[21]</sup>等.它们在目标检测任务中,取得了较好的效果.但这些特征金字塔网络只是对不同层次的特征进行简单的融合,无法获取更深层次的特征表示.与之前提出的特征金字塔网络不同,多尺度交互内存金字塔网络通过融合多层次多尺度的特征来加强层间特征信息交流,增强低照度图像特征的代表能力,并且设置内存向量存储样本的特征,捕获不同样本之间的相关性,从而抽象出同一类别物体之间相似的高级特征,提升模型的泛化能力.

本文的多尺度交互内存金字塔结构如图1所示.它将骨干网络的输出表示为 $\{P_1, P_2, P_3, P_4\}$ ,相对于输

入图像的步长为 $\{32, 16, 8, 4\}$ , $P_5$ 是对 $P_4$ 上采样得到的特征图.本文分别将 $\{P_1, P_2, P_3\}$ , $\{P_2, P_3, P_4\}$ 和 $\{P_3, P_4, P_5\}$ 送入内存机制中.内存机制能够融合相邻层的多尺度特征,并学习整个数据集最具区别性的特征,达到增强低照度图像特征中有效信息的目的.内存机制的结构如图2所示.以 $\{P_3, P_4, P_5\}$ 为例,首先,对 $P_3$ 和 $P_5$ 分别进行上采样和下采样,使它们的特征图尺寸和通道数与 $P_4$ 相同.然后,将 $\{P_3, P_4, P_5\}$ 在通道维度上进行拼接得到特征图 $F$ ,其中, $P_3$ 为浅层特征包含更多的细节信息, $P_4$ 为中间特征既包含细节信息又包含语义信息, $P_5$ 为顶层特征可以挖掘更丰富的语义信息.将3层特征进行融合,使语义信息和细节信息相互补充,显著提高模型的特征表达能力.然后,通过 $1 \times 1$ 的卷积降低 $F$ 的通道数,得到特征 $F'$ .调整 $F'$ 的维度,将 $F'$ 重塑为 $A \in \mathbb{R}^{N \times C}$ , $N$ 为特征图像素的个数, $C$ 为特征图的通道数.全连接层 $M_1 \in \mathbb{R}^{d \times C}$ 和 $M_2 \in \mathbb{R}^{d \times C}$ 是2个可学习、可共享的内存单元,它们能够存储样本特征,隐含地构建所有样本的相关性.将 $F'$ 送入全连接层 $M_1$ ,然后对特征图进行Softmax计算得到 $O_{(i,j)}$ , $O_{(i,j)}$ 表示第 $i$ 个像素与全连接层 $M_1$ 第 $j$ 行的相似性.根据得到的相似特征图 $O_{(i,j)}$ 来更新存储在全连接层 $M_2$ 的输入特征,输出特征图 $E$ ,计算公式如下:

$$O_{(i,j)} = \frac{\exp(F'M_1^T)}{\sum_{i=1}^d \exp(F'M_1^T)} \quad (3)$$

$$E = O_{(i,j)} M_2 \quad (4)$$

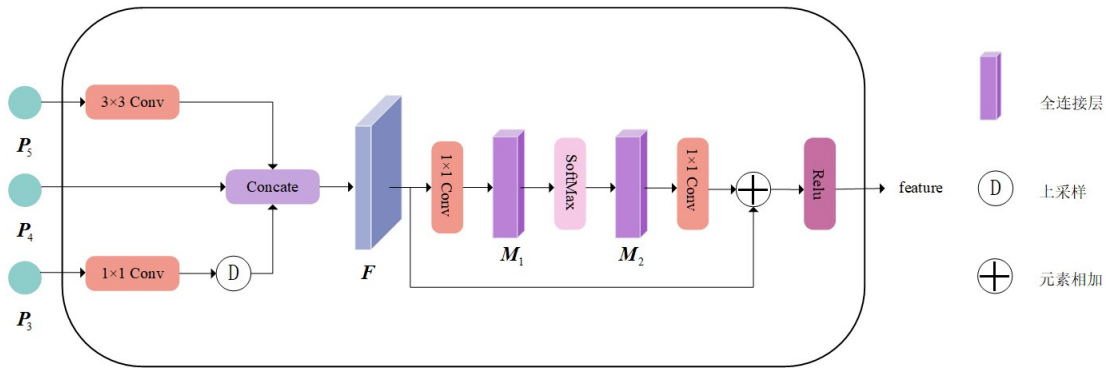


图2 内存机制

通过 $1 \times 1$ 的卷积增加特征图 $E$ 的通道数,并与特征 $F$ 进行逐元素相加,最后使用ReLU函数进行激活.内存机制输出的3个特征图包含不同的特征信息,浅层特征分辨率高,包含更多位置信息.但是由于经过的卷积层较少,其包含的语义信息较少.高层特征具有更强的语义信息,但是分辨率低,对位置信息的感知能力较差.为了使低照度图像特征包含丰富的位置信息和语义信息,将内存机制输出的特征通过自上而下的路径

进行上采样和自下而上的路径进行下采样,然后通过横向连接合并相同尺寸大小的自底向上得到的特征图和自顶向下得到的特征图,将位置信息强、语义信息弱的特征图和位置信息弱、语义信息强的特征图相融合,使每个层次的特征更具鲁棒性.

### 3.3 空间感知注意力机制

由于低照度图像对比度低、噪声大,神经网络提取到的低照度图像特征中的背景信息和前景信息区分度

不高,并且包含较多噪声.为了解决这一问题,本文提出了空间感知注意力机制.现有的注意力机制在目标检测中取得较好的效果,但是依然存在一些缺点,如:SENet只考虑通道间的信息编码,而忽略了位置的重要性;CBAM通过引入大尺寸核卷积来获取空间信息,而卷积只能捕获局部的依赖关系,不能建立对视觉任务至关重要的长距离依赖关系;非局部注意力机制需要建立特征图中每对像素之间的关系,计算量较大.不同于之前的工作,空间感知注意力机制能够同时建立局部依赖关系和长距离依赖关系,增强低照度图像的特征的上下文信息,进而更好地分离低照度图像中的前景特征和背景特征.

在特征图中,像素和像素之间不是孤立存在的,同一目标区域中的像素和像素之间存在联系.低照度图像特征受背景信息和噪声的干扰,导致前景区域中部分像素比较模糊,容易与噪声和背景混淆.因此,为了区分前景区域和背景区域,需要聚合表示相似特征的

像素点.空间感知注意力机制首先通过计算部分像素与像素之间的相似性,建立长距离依赖关系,从而有选择性地聚集相似的特征,然后通过池化和卷积操作提取局部特征,建立局部依赖关系,最后将长距离上下文信息和局部信息编码到低照度图像特征中,增强表示属于同一目标像素之间的联系,从而突出低照度图像特征的前景区域.

空间感知注意力机制的结构如图3所示.在第一阶段,首先,分别在水平和垂直方向对特征  $F \in \mathbb{R}^{C \times H \times W}$  进行平均池化操作,生成特征图  $Q$  和  $K$ ,表达式为

$$Q_c(h) = \frac{1}{W} \sum_{0 \leq i < W} F_c(h, i) \quad (5)$$

$$K_c(w) = \frac{1}{H} \sum_{0 \leq j < H} F_c(j, w) \quad (6)$$

其中,  $Q_c(h)$  表示特征  $Q$  在像素点  $(c, h)$  处的值,  $K_c(w)$  表示特征  $K$  在像素点  $(c, w)$  处的值.

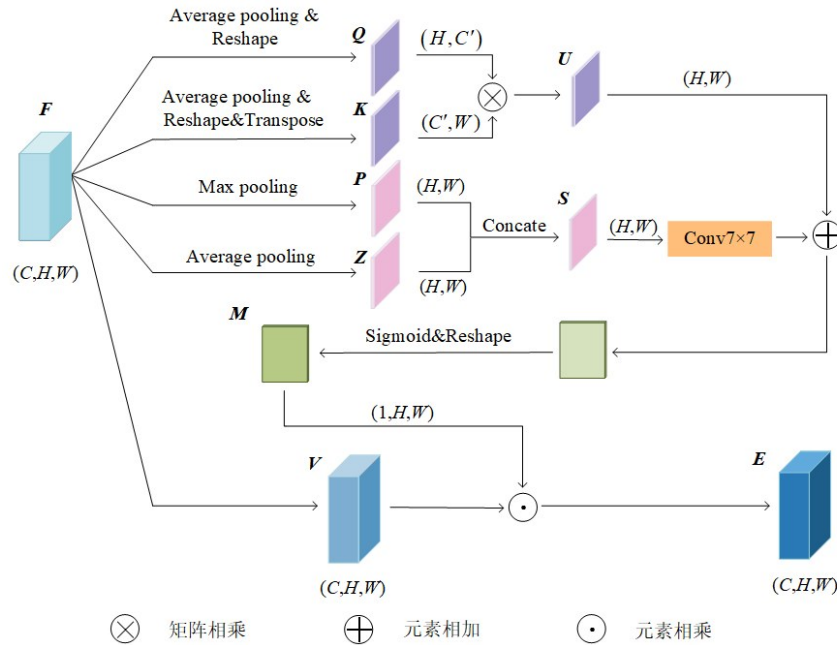


图3 空间感知注意力机制

其次,调整  $Q$  和  $K$  的维度,使  $Q \in \mathbb{R}^{C' \times H}$ ,  $K \in \mathbb{R}^{C' \times W}$ . 然后,将  $Q$  的转置和  $K$  相乘,得到特征图  $U$ ,  $U \in \mathbb{R}^{H \times W}$ , 计算公式为

$$U = Q^T K \quad (7)$$

特征图  $U$  在部分像素之间建立关系,避免在相距较远的像素之间建立大多数不必要的连接.与需要建立每对像素之间关系的非局部注意力机制相比,该模块的计算量较小.在第二阶段,首先,沿着通道方向使用平均池化和最大池化操作来聚合特征中的空间信息,生成两个包含不同空间上下文信息的特征图.其

次,对得到的两个特征图进行 Concat 操作.然后,使用一个  $7 \times 7$  的卷积来进一步提取特征,得到特征图  $S$ . 特征图  $S$  可以有效地凸显信息丰富的区域,公式表示为  $S = f^{7 \times 7}(\text{Cat}[\text{AvgPool}(f^{1 \times 1}(F)) + \text{MaxPool}(f^{1 \times 1}(F))])$  (8)

将特征图  $U$  和  $S$  逐元素相加并对相加后的特征进行 Sigmoid 运算来生成上下文注意力特征图  $M$ . 特征图  $M$  编码了上下文信息,能够突出重要的特征区域并且抑制不重要的特征区域.最后,对上下文注意力特征图  $M$  和特征  $V$  进行逐元素相乘,得到特征  $E$ , 计算公式为

$$M = \frac{1}{1 + e^{-(s+t)}} \quad (9)$$

$$E_i = M \odot V_i \quad i = 1, 2, 3, \dots, C \quad (10)$$

其中,  $E_i$  和  $V_i$  中的下标  $i$  表示特征中沿通道维度方向第  $i$  个特征图,  $\odot$  为逐元素相乘.

在第一阶段, 通过建立部分像素之间的关系来捕获长距离上下文信息; 在第二阶段, 利用池化和卷积操作来获取局部信息. 空间感知注意力机制把长距离上下文信息和卷积提取到的局部信息编码到低照度图像特征中, 从而突出低照度图像特征的前景信息, 使网络模型忽略低照度图像特征中的背景信息和噪声, 更加关注感兴趣的目标区域.

### 3.4 多感受野增强模块

由于 YOLOv4 算法只使用单一尺度的卷积核来提取特征, 限制了每个特征层的感受野大小, 在低照度图像中存在不同尺寸的目标时检测效果不佳. 在图像分割任务中, 为了更好地对图像中每个像素点进行分类, 通常采用 ASPP (Atrous Spatial Pyramid Pooling)<sup>[22]</sup> 网络来增强特征的感受野, 从而更好地适应不同尺寸的目标. 然而, ASPP 网络对感受野的自适应调整能力差, 并且其设置的扩张率不适用于目标检测网络. 因此, 本文借鉴语义分割任务中常用的 ASPP 结构, 提出了多感受野增强模块. 该模块采用具有不同扩张率的空洞卷积分组并行提取低照度图像特征, 然后进行加权计算, 为每组特征赋予不同的权重, 使网络能够根据目标物体

的尺寸动态选择相适应的感受野大小, 最后将具有不同感受野的特征图在通道的维度上进行拼接, 融合后得到的特征具有不同大小的感受野. 当低照度图像存在不同尺寸的目标时, 可以提取图像中不同尺寸目标的特征, 进而提升低照度目标检测器对低照度图像中不同大小目标的检测能力. 考虑到多尺度交互内存金字塔的 3 个分支输出的特征图尺寸分别为  $26 \times 26$ ,  $52 \times 52$  和  $104 \times 104$ , 尺寸较小的特征图已经具备较大的感受野, 不适合使用扩张率较大的空洞卷积来提取特征. 所以, 本文对多尺度交互内存金字塔的 3 个分支输出的特征图设置不同的扩张率组合, 分别为  $R_1 = (1, 2)$ ,  $R_2 = (1, 2, 5)$ ,  $R_3 = (1, 2, 5, 8)$ , 3 个扩张率组合对应多感受野增强模块的分支数分别为 2, 3 和 4.

多感受野增强模块的结构如图 4 所示. 以  $R_1 = (1, 2)$  为例, 首先, 采用卷积核大小为  $3 \times 3$ , 扩张率为 1 和扩张率为 2 的两组空洞卷积并行提取特征  $F \in \mathbb{R}^{C \times H \times W}$ , 得到两组特征的通道数变为原来的  $1/2$ . 其次, 对 2 组特征分别使用平均池化操作来编码通道信息, 得到对应的通道权重向量, 公式表示为

$$a_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W F_1^c(i, j) \quad (11)$$

$$b_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W F_2^c(i, j) \quad (12)$$

其中,  $a_c$  表示向量  $a$  中第  $c$  个元素的值,  $b_c$  表示向量  $b$  中第  $c$  个元素的值.

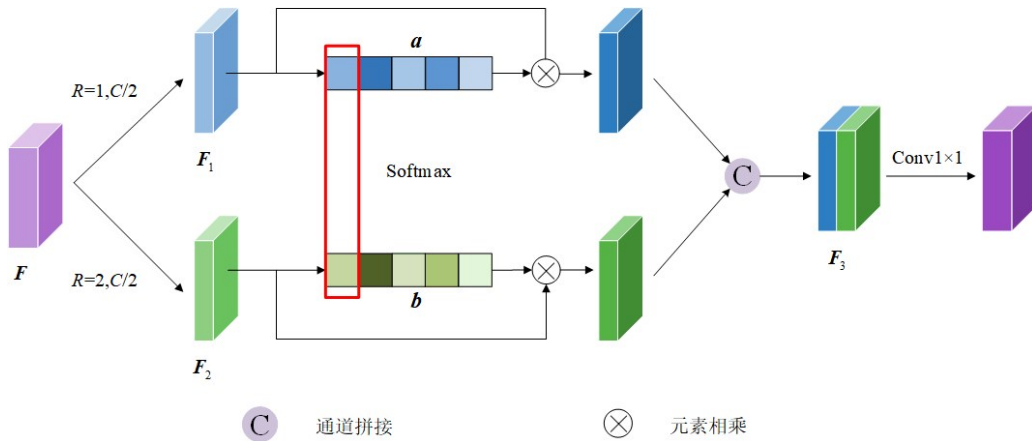


图 4 多感受野增强模块

通过 Softmax 函数重新标定 2 个分支的通道权重, 使网络在检测尺寸较大的物体时, 对扩张率大的分支分配较大的权重. 在检测尺寸较小的物体时, 对扩张率小的分支分配较大的权重, 从而使网络动态的选择感受野. 计算公式为

$$a'_c = \frac{e^{a_c}}{e^{a_c} + e^{b_c}} \quad (13)$$

$$b'_c = \frac{e^{b_c}}{e^{a_c} + e^{b_c}} \quad (14)$$

$$a'_c + b'_c = 1 \quad (15)$$

然后, 让重标定的通道权重与对应分支的特征图相乘, 将得到的特征按通道拼接, 计算公式为

$$F_3 = \text{Cat}(a'F_1 + b'F_2) \quad (16)$$

最后,用 $1 \times 1$ 的卷积调整特征的通道数.上述描述为当模块只包含2个分支时的情况,通过扩展公式,可以比较容易推导出多分支的情况.

低照度图像中包含不同尺寸的目标物体,检测不同尺寸的物体需要不同大小的感受野.使用单一尺寸卷积核提取到的低照度图像特征的感受野是固定的,导致网络无法准确识别不同尺寸的物体.多感受野增强模块利用不同扩张率的空洞卷积提取低照度图像特征,得到的每组特征具有不同大小的感受野,对具有不同感受野的特征进行加权运算,使网络能够根据不同尺寸的目标物体设计出不同大小的感受野,将加权后的特征融合,达到增强感受野的目的.

## 4 实验部分

### 4.1 实验设置细节

本实验的运行环境硬件配置为1块Tesla P40显卡,软件系统为Windows 10.实验程序在PyTorch 1.2深度学习框架下使用python编程语言实现.算法使用YOLOv4在COCO数据集中训练好的权重进行预训练,训练的迭代次数为65轮,每轮的batchsize设置为4.在前25轮的训练过程中冻结骨干网络,在后40轮的训练过程中解冻骨干网络.本文使用的是ExDark<sup>[23]</sup>数据集,对应12个类别对象,按照4:1的比例随机划分数据集,得到5892张训练集图片和1473张测试集图片,训练和测试时输入图像的尺寸大小均为 $416 \times 416$ 像素.本文使用在IoU=0.5条件下的AP(Average Precision)和mAP(mean Average Precision)以及FPS(Frames Per Second)作为检测器的评价指标.

### 4.2 消融实验

为进一步评估本文方法的性能,验证本文方法中各策略的有效性,对模型进行消融实验.表1展示了模型消融实验的结果,粗体为最优结果. Baseline使用的是YOLOv4网络.仅将YOLOv4的PANet替换成MIMP,检测精度提升1.96%;仅引入SAM,检测精度提升2.48%;仅引入MREM,检测精度提升2.64%.这表明,将每个模块单独放入检测器中都会提升检测器的检测性.当引入MIMP和SAM时,检测精度提升3.26%;当引入MIMP和MREM时,检测精度提升3.36%;当引入SAM和MREM时,检测精度提升3.47%.这表明,将模块两两组合能进一步提升检测器的检测性能.当同时引入3个模块时,检测精度提升4.32%.由表1可以看出,本文采用的各策略对目标检测性能均有一定的提升.

为了得到一组最适合的空洞卷积扩张率,本文进行不同空洞卷积扩张率组合的对比实验.实验结果如表2所示,粗体为最优结果.不同空洞卷积扩张率的组

表1 本文算法有效性的验证

网络结构				mAP/%
Baseline	MIMP	SAM	MREM	
√				72.72
√	√			74.68
√		√		75.20
√			√	75.36
√	√	√		75.98
√	√		√	76.08
√		√	√	76.19
√	√	√	√	<b>77.04</b>

合能够影响低照度图像检测器对目标物体的检测效果.当 $R_1=(1, 2)$ ,  $R_2=(1, 2, 5)$ ,  $R_3=(1, 2, 5, 8)$ 时,检测器的检测精度最高.因此,本文采用该扩张率组合作为多感受野增强模块的扩张率组合.

表2 不同扩张率组合下的mAP

扩张率	mAP/%
$R_1=(1, 2, 5, 8)$ , $R_2=(1, 2, 5, 8)$ , $R_3=(1, 2, 5, 8)$	76.48
$R_1=(1, 2, 5)$ , $R_2=(1, 2, 5)$ , $R_3=(1, 2, 5)$	76.01
$R_1=(1, 2)$ , $R_2=(1, 2, 5)$ , $R_3=(1, 2, 5, 8)$	<b>77.04</b>
$R_1=(1, 3)$ , $R_2=(1, 3, 7)$ , $R_3=(1, 3, 7, 9)$	76.42

图5是本文算法在加入空间感知注意力机制之前和加入空间感知注意力机制之后的特征可视化结果.如图5所示,热力图中红色区域覆盖的是感兴趣的目标区域.加入上下文注意力模块之前,部分图像中的红色区域覆盖到背景区域中,没有清晰地划分出物体和物体之间的界限.加入上下文注意力模块后,红色区域更加聚焦于目标物体上,使物体之间具有清晰的界限.此外,红色区域覆盖的目标物体更多并且对目标物体的覆盖范围更广.显然,加入空间感知注意力机制使网络对感兴趣的目标定位更精确.

为了进一步验证SAM的优越性,让SAM与现有的先进注意力机制进行比较.本文将SAM-MSFF网络中的SAM替换成现有注意力机制,其他实验条件保持不变.注意力机制的比较结果如表3所示,粗体为最优结果.从表3可以看出,当检测网络加入SAM时,检测精度最高,参数量最小.SAM的计算量要稍大于SENet和CBAM,这是因为SAM为了建立图像上的长距离依赖关系使用了矩阵乘法运算,从而增加了计算量.SAM的计算量要明显小于Non-local,这是由于Non-local通过计算任意2个像素位置之间的关系来获取长距离依赖,而SAM使用池化操作在水平和垂直方向对特征进行压缩,通过建立部分像素之间的联系来获取长距离依赖,因此减少了计算量.

表4所示为本文提出的MIMP与现有先进特征金字塔网络在检测精度上的比较结果,粗体为最优结果.

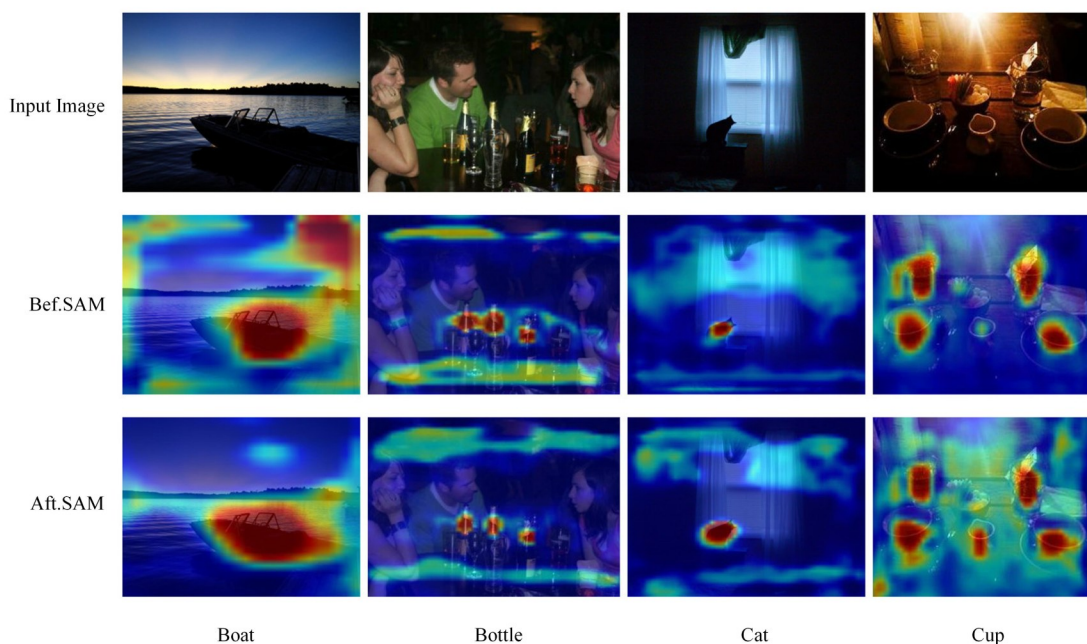


图5 加入SAM之前和之后的特征可视化结果

表3 SAM与现有先进注意力机制比较结果

算法	mAP/%	Params/M	GFLOPs
SENet	76.38	108.9	126.82
CBAM	75.86	108.9	<b>126.82</b>
Non-local	75.75	109.3	128.04
SAM	<b>77.04</b>	<b>108.7</b>	127.13

本文将SAM-MSFF网络中的MIMP替换成现有的特征金字塔网络,其他实验条件保持不变.从表4可以看出,检测网络加入MIMP时取得了最高的检测精度,这说明MIMP充分融合了高层特征的语义信息和低层特征的细节与位置信息,并通过内存机制挖掘出样本的潜在特征,使模型能够增强低照度图像特征的有效信息,提升模型的检测精度.

表4 MIMP与现有先进特征金字塔网络比较结果

算法	mAP/%
FPN	68.90
PANet	76.19
BiFPN	76.16
MIMP	<b>77.04</b>

### 4.3 对比实验

表5所示为本文算法与目前主流的目标检测算法在ExDark数据集上检测精度和检测速度的对比结果,粗体为最优结果.从表5可以看出,在检测精度方面,本文算法的mAP为77.04%,比单阶段目标检测算法YOLOX-X提高了2.6%.在检测速度方面,与双阶段目标检测器和部分单阶段目标检测器相比,本文算法要

明显优于它们.但是与YOLO系列算法相比,本文算法的检测速度有所下降.这是由于为了提升检测精度,本文算法向基础网络中添加模块来增强对低照度图像特征的提取能力,从而影响了检测速度,但依然满足实时性的要求.

表6所示为本文算法与目前主流的目标检测算法在ExDark数据集上各个类别的AP比较结果,粗体为最优结果.从表6中可以看出,和其他目标检测算法相比,本文算法除了自行车、瓶子和猫这3个类别的AP没有达到最优,其他类别的AP都具有显著优势.

为了验证骨干网络和图像尺寸对模型的影响,我们通过实验对不同设置的模型进行复杂度分析.实验结果如图6所示.在平均精度方面,随着图像输入尺寸的增大,模型的平均精度逐渐增加.当输入图像的尺寸保持相同、骨干网络为CSPDarknet53时,模型的平均精度最高,最高精度为77.68%.在参数量方面,当模型输入不同尺寸的图像时,模型的参数量保持不变;当输入图像的尺寸保持相同,骨干网络为Modified CSP v5时,模型的参数量最小,最小值为101.3 M.在计算量方面,随着图像输入尺寸的增大,模型的计算量逐渐增大.当输入图像的尺寸保持相同,骨干网络为ResNet-50时,模型的计算量最小,最小值为125.42 GFLOPs.在检测速度方面,随着图像输入尺寸的增大,模型的检测速度逐渐降低,当输入图像的尺寸保持相同,骨干网络为ResNet-50时,模型的检测速度最快,最快速度为19.25 FPS.

为了证明在多尺度交互内存金字塔中融合多尺度

表 5 不同算法在 ExDark 数据集上的 mAP 比较

算法	基础网络	输入尺寸	mAP/%	FPS
Cascade R-CNN <sup>[24]</sup>	ResNet-101	416×416	68.50	4.50
Grid R-CNN <sup>[25]</sup>	ResNet-101	416×416	63.00	5.20
Libra R-CNN <sup>[26]</sup>	ResNet-101	416×416	71.30	3.60
Sparse R-CNN <sup>[27]</sup>	ResNet-101	416×416	65.90	3.80
FCOS <sup>[28]</sup>	ResNet-101	416×416	65.70	7.50
FreeAnchor <sup>[29]</sup>	ResNet-101	416×416	62.70	6.20
PISA <sup>[30]</sup>	ResNet-101	416×416	66.60	6.50
PAA <sup>[31]</sup>	ResNet-101	416×416	65.50	6.10
FSAF <sup>[32]</sup>	ResNeXt-101	416×416	66.90	6.40
VFNet <sup>[33]</sup>	ResNeXt-101	416×416	70.00	3.30
EfficientDet	EfficientNet-D2	416×416	68.68	5.66
YOLOv3	DarkNet-53	416×416	68.23	16.11
YOLOv4	DarkNet-53	416×416	72.72	12.76
YOLOF <sup>[34]</sup>	ResNet-50	416×416	64.50	<b>17.10</b>
YOLOX-X	Modified CSP v5	416×416	74.44	10.72
SAM-MSFF	DarkNet-53	416×416	<b>77.04</b>	10.18

表 6 不同算法在 ExDark 数据集上各个类别的 AP 比较

单位: %

算法	类别											
	自行车	船	瓶子	公交车	小汽车	猫	椅子	杯子	狗	电动车	人	桌子
Cascade R-CNN	72.10	67.00	57.80	84.10	72.90	77.40	57.90	55.70	78.60	70.80	71.60	52.60
Grid R-CNN	70.60	60.60	54.00	79.40	69.70	71.10	53.00	51.60	67.00	63.60	69.20	46.50
Libra R-CNN	78.30	69.40	64.70	86.90	75.20	77.60	60.60	61.10	76.10	76.00	74.40	55.70
Sparse R-CNN	76.30	58.70	51.30	84.30	69.20	71.00	55.10	57.60	76.80	68.50	67.60	54.90
FCOS	72.40	63.30	56.30	83.50	68.50	<b>78.60</b>	58.10	52.40	74.50	66.70	67.40	46.90
FreeAnchor	70.70	59.40	52.00	81.20	69.10	70.30	52.10	49.30	70.80	63.50	68.10	45.60
PISA	73.70	63.40	56.80	83.20	71.20	76.50	55.60	53.00	75.10	70.10	69.60	51.50
PAA	72.70	62.90	51.50	85.40	70.00	76.50	56.20	52.70	74.90	66.50	67.30	49.90
FSAF	74.90	61.60	58.60	80.40	70.90	76.50	60.00	56.20	74.60	68.50	71.80	49.20
VFNet	76.70	69.50	59.60	84.80	73.00	78.30	61.80	58.10	79.00	72.10	71.30	55.90
EfficientDet	80.04	69.17	53.44	86.73	81.29	72.13	57.25	63.80	70.66	70.09	69.74	49.90
YOLOv3	67.34	65.68	64.64	90.92	77.00	67.54	59.47	59.15	59.53	73.06	74.58	59.87
YOLOv4	81.88	67.70	<b>66.56</b>	89.97	80.55	70.54	61.47	66.99	75.04	78.13	80.80	53.05
YOLOF	74.50	63.40	47.00	85.30	63.40	77.80	59.00	49.10	77.90	66.10	59.90	50.60
YOLOX-X	<b>82.74</b>	75.61	66.11	90.43	81.30	69.32	66.35	70.05	74.62	78.70	80.00	58.04
SAM-MSFF	82.65	<b>77.80</b>	65.08	<b>92.81</b>	<b>85.20</b>	77.37	<b>68.92</b>	<b>70.17</b>	<b>78.73</b>	<b>81.14</b>	<b>82.50</b>	<b>62.10</b>

特征能够增强低照度图像中的有效信息,将各层特征进行可视化,特征可视化后的结果如图 7 所示.在浅层特征中,目标物体的边缘、纹理以及细节都比较清晰,包含大量细粒度信息.在中层特征中,物体的内部细节变得模糊.在深层特征中,物体的细节信息损失严重,只能分辨出大致的轮廓.此时,物体特征变得更加抽象化,蕴含丰富的语义信息.将 3 层特征进行融合,得到增强后特征.增强后的特征吸收了各层特征中的优势,同时含有丰富的细节信息和语义信息,从而增强低照度特征中的有效信息.

为了验证空间感知注意机制能够有效地增强低照度图像特征中的目标特征,并抑制背景信息的干扰,本文对通过注意力机制增强前后的特征进行可视化,特征可视化结果如图 8 所示,特征中的目标物体用红色框标注.与增强前的特征相比,增强后的特征中目标物体的边缘和内部细节更加清晰,背景的噪声更少,充分证明了空间感知注意力机制的有效性.

图 9 所示为本文算法与目前主流算法在 ExDark 数据集上的部分检测结果.本文选取了 5 张不同光照条件下的图像,图像中包含不同尺寸的物体.从图 9 可以

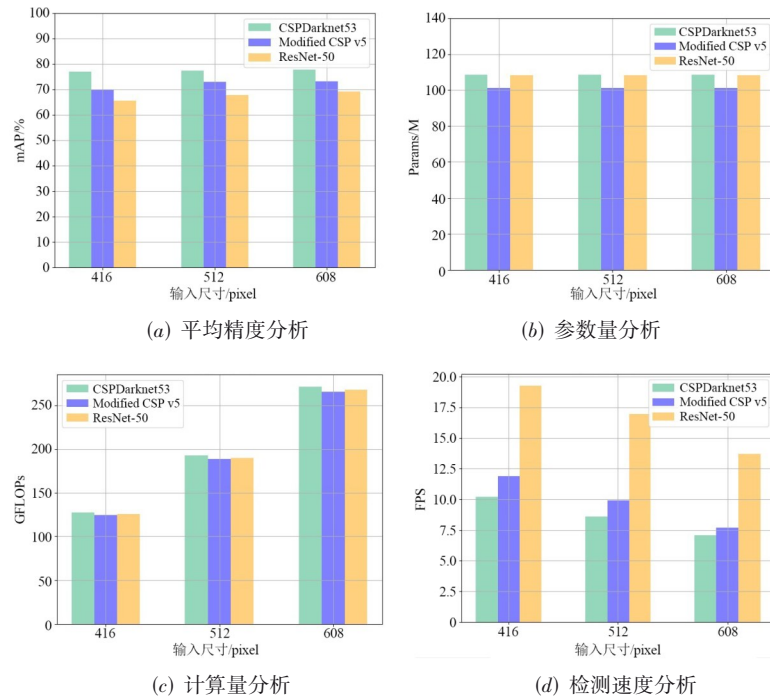


图6 算法复杂度

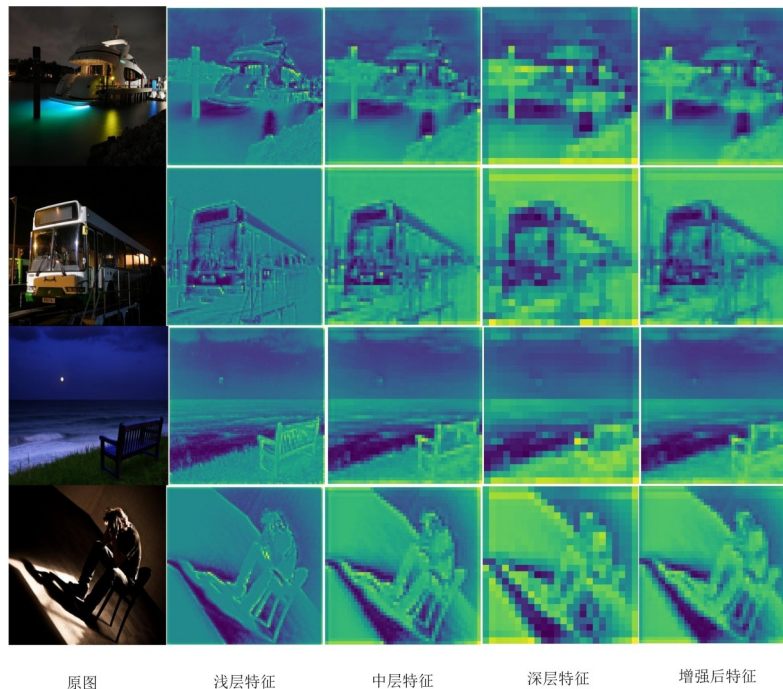


图7 多尺度特征融合时各层特征可视化结果

看出, SAM-MSFF 网络模型能够准确识别不同照度下图像中不同尺寸的物体, 达到较好的检测效果. 与目前的主流算法相比, 本文的检测模型测效果最好, 其他的检测模型存在更多漏检和误检的情况.

表7为不同算法在PASCAL VOC数据集上的比较结果, 粗体为最优结果. 所有算法使用VOC2007和

VOC2012的训练集进行训练, 并使用VOC2007的测试集进行测试. 从表7可以看出, 在正常照度的数据集下, SAM-MSFF的mAP优于其他算法. 这说明本文算法不仅适用于低照度场景下的目标检测, 而且在正常照度场景下也能取得较好的检测效果, 充分体现了本文算法具有较强的稳定性和泛化性.



图8 增强前后的特征可视化结果

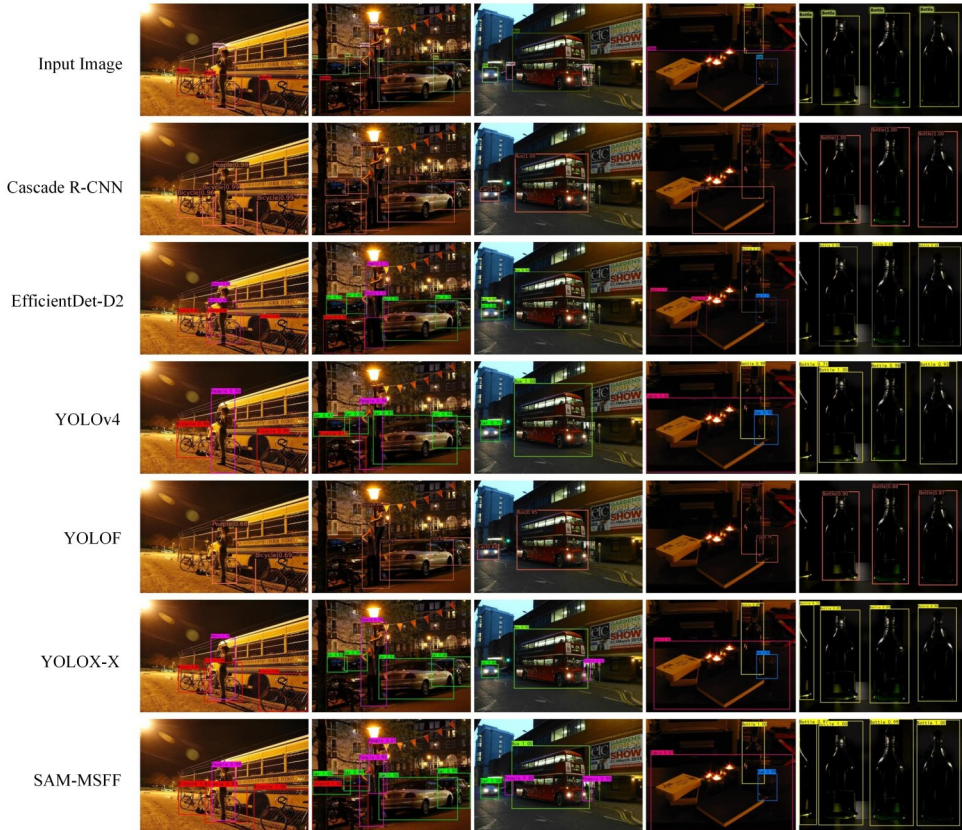


图9 不同算法在ExDark数据集上部分检测结果

表7 不同算法在PASCAL VOC数据集上的mAP比较

算法	mAP/%
EfficientDet-D2	68.77
YOLOv3	77.98
YOLOv4	82.23
YOLOX-X	83.42
SAM-MSFF	<b>84.04</b>

## 5 结语

针对低照度图像存在对比度低、细节丢失、噪声多等缺点,导致现有目标检测器对低照度图像检测效果不理想的问题,本文提出了一种结合空间感知注意力机制和多尺度特征融合的低照度下目标检测方法.通过引入多尺度交互内存金字塔网络融合多层次多尺度

特征并且捕获样本之间的相关性,增加低照度图像特征的利用率;通过引入空间感知注意力机制,同时获取低照度图像特征在空间域的长距离上下文信息和局部信息,增强低照度图像目标物体的特征;通过引入多感受野增强模块,使检测器能够更好地检测不同尺寸的目标.实验结果表明,本文算法有效地提升了检测器对低照度图像的检测精度.与目前主流的算法相比,本文算法在目标检测准确率方面具有明显的优势.未来将进一步研究本文算法的优化方法<sup>[35-39]</sup>,在提升检测精度的同时,提升算法的检测速度.

#### 参考文献

- [1] ZHOU Y, WEN S J, WANG D L, et al. Object detection in autonomous driving scenarios based on an improved Faster-RCNN[J]. *Applied Sciences*, 2021, 11(24): 11630.
- [2] GORSCHLÜTER F, ROJTBERG P, PÖLLABAUER T. A survey of 6D object detection based on 3D models for industrial applications[J]. *Journal of Imaging*, 2022, 8(3): 53.
- [3] HU Y Y, WU X J, ZHENG G D, et al. Object detection of UAV for anti-UAV based on improved YOLO v3[C]//2019 Chinese Control Conference (CCC). Piscataway: IEEE, 2019: 8386-8390.
- [4] ZHANG Y E, YE L, FANG L, et al. Benchmarking the robustness of object detection based on near-real military scenes[J]. *Wireless Communications and Mobile Computing*, 2022, 2022: 5884625.
- [5] 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.
- [6] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection [EB/OL]. (2015-06-08)[2022-06-09]. <https://arxiv.org/abs/1506.02640>.
- [7] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2017: 6517-6525.
- [8] REDMON J, FARHADI A. YOLOv3: An incremental improvement[EB/OL]. (2018-04-08)[2022-06-09]. <https://arxiv.org/abs/1804.02767>.
- [9] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection [EB/OL]. (2020-04-23)[2022-06-09]. <https://arxiv.org/abs/2004.10934>.
- [10] GE Z, LIU S T, WANG F, et al. YOLOX: Exceeding YOLO series in 2021[EB/OL]. (2021-08-06)[2022-06-09]. <https://arxiv.org/abs/2107.08430>.
- [11] WANG W J, YANG W H, LIU J Y. HLA-face: Joint high-low adaptation for low light face detection[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 16190-16199.
- [12] MIAO Y, LIU F, HOU T, et al. A nighttime vehicle detection method based on YOLO v3[C]//2020 Chinese Automation Congress (CAC). Piscataway: IEEE, 2021: 6617-6621.
- [13] 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.
- [14] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 7132-7141.
- [15] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]//European Conference on Computer Vision. Cham: Springer, 2018: 3-19.
- [16] WANG X L, GIRSHICK R, GUPTA A, et al. Non-local neural networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 7794-7803.
- [17] WANG C Y, MARK LIAO H Y, WU Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE, 2020: 1571-1580.
- [18] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1904-1916.
- [19] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2017: 936-944.
- [20] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 8759-8768.
- [21] 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: 10778-10787.

- [22] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.
- [23] 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.
- [24] CAI Z W, VASCONCELOS N. Cascade R-CNN: Delving into high quality object detection[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 6154-6162.
- [25] LU X, LI B Y, YUE Y X, et al. Grid R-CNN[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 7355-7364.
- [26] PANG J M, CHEN K, SHI J P, et al. Libra R-CNN: Towards balanced learning for object detection[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 821-830.
- [27] SUN P Z, ZHANG R F, JIANG Y, et al. Sparse R-CNN: End-to-end object detection with learnable proposals[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 14449-14458.
- [28] 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, 2020: 9626-9635.
- [29] ZHANG X S, WAN F, LIU C, et al. FreeAnchor: Learning to match anchors for visual object detection [EB/OL]. (2019-09-05)[2022-06-09]. <https://arxiv.org/abs/1909.02466>.
- [30] CAO Y H, CHEN K, LOY C C, et al. Prime sample attention in object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 11580-11588.
- [31] KIM K, LEE H S. Probabilistic anchor assignment with IoU prediction for object detection[C]//European Conference on Computer Vision. Cham: Springer, 2020: 355-371.
- [32] ZHU C C, HE Y H, SAVVIDES M. Feature selective anchor-free module for single-shot object detection[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 840-849.
- [33] ZHANG H Y, WANG Y, DAYOUB F, et al. Varifocal-Net: An IoU-aware dense object detector[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 8510-8519.
- [34] CHEN Q, WANG Y M, YANG T, et al. You only look one-level feature[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 13034-13043.
- [35] HUANG Y S, JIANG Z T, LAN R S, et al. Infrared image super-resolution via transfer learning and PSRGAN [J]. IEEE Signal Processing Letters, 2021, 28: 982-986.
- [36] HUANG Y S, JIANG Z T, WANG Q Z, et al. Infrared image super-resolution via heterogeneous convolutional WGAN[C]//Pacific Rim International Conference on Artificial Intelligence. Cham: Springer, 2021: 461-472.
- [37] HUANG Y S, WANG Q Z, OMACHI S. Rethinking degradation: Radiograph super-resolution via AID-SRGAN [EB/OL]. (2022-08-05)[2022-10-17]. <https://arxiv.org/abs/2208.03008>.
- [38] JIANG Z T, PI K, HUANG Y S, et al. Difference value network for image super-resolution[J]. IEEE Signal Processing Letters, 2021, 28: 1070-1074.
- [39] 权宇, 李志欣, 张灿龙, 等. 融合深度扩张网络和轻量化网络的目标检测模型[J]. 电子学报, 2020, 48(2): 390-397.
- QUAN Y, LI Z X, ZHANG C L, et al. Fusing deep dilated convolutions network and light-weight network for object detection[J]. Acta Electronica Sinica, 2020, 48(2): 390-397. (in Chinese)

#### 作者简介



江泽涛 男, 1961年生, 江西九江人. 桂林电子科技大学教授, 博士生导师. 主要研究方向为图像处理、计算机视觉.



李慧 女, 1997年生, 辽宁大连人. 桂林电子科技大学硕士研究生. 主要研究方向为目标检测.

E-mail: 1023941087@qq.com