第 60 卷 第 6 期/2023 年 3 月/激光与光电子学进展 研究论文

激光与光电子学进展 先进成像

基于自适应特征增强的小目标检测网络

吴萌萌[1,2]，张泽斌[1]，宋尧哲[1,2]，舒子婷[1,2]，李宝清[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 由于尺寸有限、外观和几何线索较少以及缺少大规模小目标数据集，从图像中检测小目标仍然是计算机视觉领域中一个具有挑战性的难题。针对这个问题，提出一种自适应特征增强的目标检测网络（YOLO-AFENet）来改善小目标的检测精度。首先，通过引入特征融合因子，设计改进的自适应双向特征融合模块，充分利用各个尺度的特征图，提高网络的特征表达能力；其次，结合网络自身的特点，提出空间注意力生成模块，通过学习图像中感兴趣区域的位置信息以提高网络的特征定位能力。在UAVDT数据集上实验结果表明：所提YOLO-AFENet的平均精度（AP）比改进前的YOLOv5提高了6.3个百分点，同时也优于其他目标检测网络。

关键词 图像处理；目标检测；小目标；多尺度特征融合；空间注意力；UAVDT

中图分类号 TP391.4 **文献标志码** A **DOI**：10.3788/LOP213048

Small-Target Detection Network Based on Adaptive Feature Enhancement

Wu Mengmeng[1,2], Zhang Zebin[1], Song Yaozhe[1,2], Shu Ziting[1,2], Li Baoqing[1*]

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

[2]*University of Chinese Academy of Sciences, Beijing* 100049, *China*

Abstract Small-target detection from images remains a challenge in the field of computer vision because of the limited size, small appearance and geometric clues, and lack of large-scale small-target datasets. To solve this issue, an adaptive feature-enhanced target detection network called YOLO-AFENet is proposed to improve the accuracy of small-target detection. First, by introducing the feature fusion factor, an improved adaptive bidirectional feature fusion module is designed using feature maps of various scales to improve the network's feature expression ability. Second, combined with the network characteristics, a spatial attention generation module is proposed to improve the network's feature localization ability by identifying the location information of the region of interest in the image. The experimental results of the UAVDT dataset show that YOLO-AFENet has a 6.3 percentage points higher average accuracy compared with YOLOv5 and is better than other target-detection networks.

Key words image processing; object detection; small object; multi-scale feature fusion; spatial attention; UAVDT

1 引 言

目标检测技术是实例分割、图像字幕、目标跟踪等许多计算机视觉任务的基础，在视频监控、自动驾驶、医疗诊断等领域得到了广泛的应用[1]。深度学习对目标检测技术的发展作出了显著的贡献，现有的目标检测技术多基于深度学习技术[2]，可划分为两类：1）两阶段目标检测算法（two-stage），如R-CNN系列[3-5]，具有准确率高的优势；2）单阶段目标检测算法（one-stage），如YOLO系列[6-9]、SSD系列[10-12]等，检测速度相对更快。

基于深度学习的目标检测算法，克服了传统算法手工特征提取适应性差的缺点，取得了快速发展。然而，小目标检测作为目标检测任务的子类，仍然是计算机视觉领域的困难和挑战。MS COCO数据集[13]中将像素数小于32×32的物体定义为小目标。与中、大目标相比，小目标难以检测的原因主要有两个方面：1）小目标的像素较少，可用于检测的特征不足；2）小目标可能出现在图像的任何位置，存在高密度和遮挡问题[14]。为解决小目标难以检测的问题，Lin等[15]提出特征金

收稿日期：2021-11-24；**修回日期**：2021-12-25；**录用日期**：2022-01-14；**网络首发日期**：2022-01-24

通信作者：*sinoiot@mail.sim.ac.cn

字塔网络(FPN),首次构造多尺度特征融合结构,小目标检测效果得到明显提升。Liu等[16-17]相继优化FPN结构,提出双向特征金字塔结构。汪亚妮等[18]在SSD模型中引入注意力机制以提取目标潜在的位置。刘鑫等[19]将相邻特征层融合之后再进行更深层的特征融合,以生成语义信息更加丰富的特征图,提高SSD模型的检测精度。这些方法大多是在PASCAL VOC和MS COCO等通用目标检测数据集上进行实验的,在小目标场景中的检测效果尚未验证。此外,这些网络结构检测精度的提升是以实时性下降为代价的。因此,需要针对小目标的像素少、特征表示弱的特点设计一个准确率高、实时性好的模型来解决小目标检测难的问题。

本文以YOLOv5为基础,提出一种自适应特征增强的目标检测算法(YOLO-AFENet)。首先对原网络中的多尺度特征融合模块进行优化,去除单输入节点,引入跳跃连接以及自适应融合因子的思想,充分利用特征图之间的关联,改善多尺度特征融合的效果。其次,利用多尺度特征融合模块和特征提取网络自然形成的U型结构,设计了一个简单而有效的解码器模块来预测输入图像中感兴趣区域(ROI)的位置,并将其作为空间注意力增强原检测特征图中ROI的响应,改善目标检测的定位精度。在UAVDT数据集上的实验结果表明,所提算法的平均精度(AP)相比于YOLOv5提高了6.3个百分点。

2　YOLOv5基本原理

YOLOv5是YOLO系列的第5代目标检测算法,包含了YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x等多种不同规模的网络。相比于YOLOv4,YOLOv5引入focus模块,在训练网络之前引入自适应锚框计算,具有权重文件小、训练时间短和推理速度更快的特点,其整体结构如图1所示。

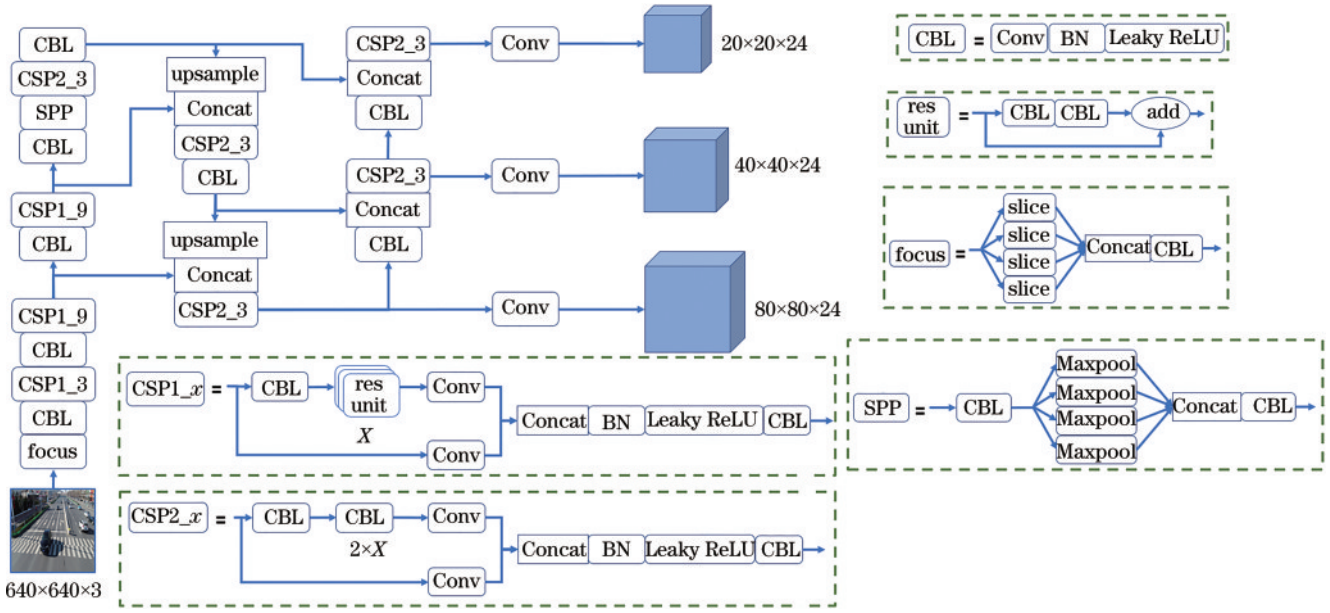

图1　YOLOv5l网络细节

Fig. 1　Details of YOLOv5l network

YOLOv5包括输入(input)、特征提取网络(backbone)、多尺度特征融合模块(neck)和检测头(head)等4个部分。YOLOv5的输入部分主要包含数据增强和自适应锚框计算2个部分。数据增强包含输入图像尺寸调整、色彩空间变换和mosaic数据增强。自适应锚框计算首先计算默认anchor的最大可能召回率(BPR),如果BPR小于98%,就利用K-means和遗传算法更新anchor。在特征提取网络中,YOLOv5新增了focus模块,用于减少网络的参数量和计算复杂度、提高网络的推理速度。YOLOv5采用path aggregation network(PANet)结构作为多尺度特征融合模块,具有自顶向下、自底向上的特点,卷积部分采用了与主干网络不同的CSP结构,具有更好的多样性和鲁棒性。检测头负责对目标进行定位和识别,并输出最终的检测结果。

3　YOLO-AFENet

YOLOv5目标检测算法因其在模型大小和检测速度方面表现优秀而备受关注。同时,YOLOv5在PASCAL VOC和COCO数据集上的实验证明了该网络在检测精度方面的优势。但YOLOv5的研究内容是通用目标检测,在网络设计时并未过多考虑到小目标的特点,因此需对网络进行改进和优化以适用于小目标场景。

3.1　网络整体结构

所提YOLO-AFENet由特征提取网络、改进的自适应多尺度融合网络、空间注意力生成网络(SAGN)、检测头等4部分组成,网络的整体结构如图2所示。

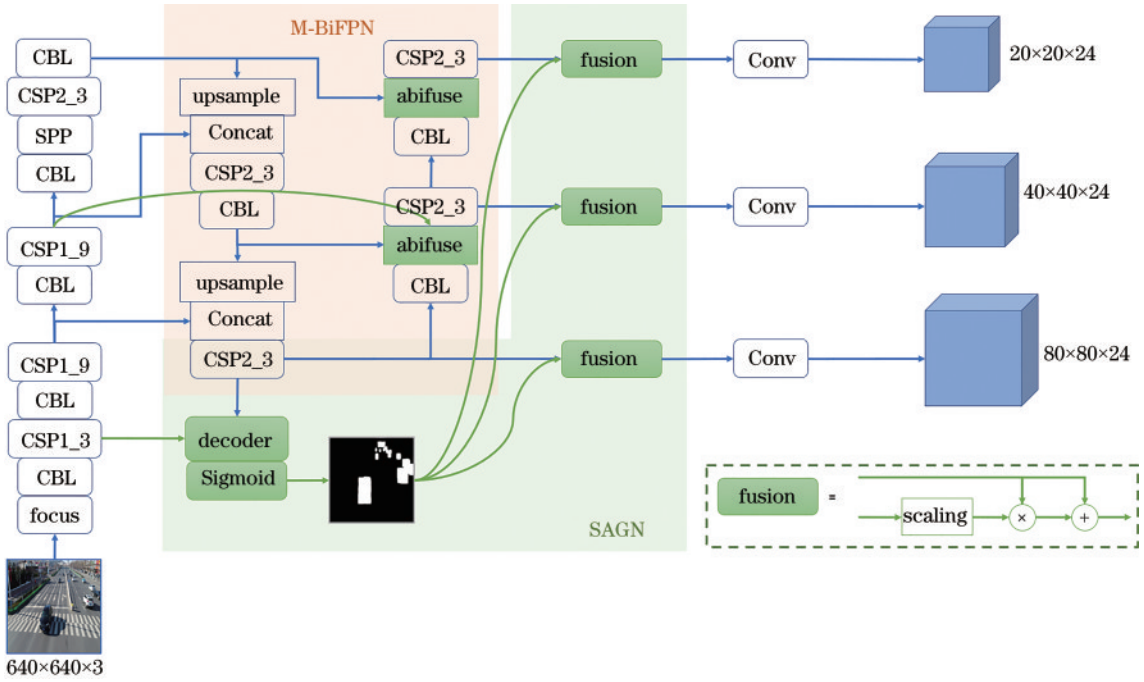

图 2　YOLO-AFENet结构

Fig. 2　Structure of YOLO-AFENet

改进的自适应双向特征金字塔网络(M-BiFPN)在YOLOv5的基础上引入跳跃链接和自适应特征融合因子,充分利用骨干网络提取的特征信息,可有效提高网络的特征表示能力。空间注意力生成网络利用原网络中的特征金字塔层次,通过预测输入图片中ROI的位置以增强检测网络的目标定位能力,改善目标检测效果。

3.2　M-BiFPN模块

位于网络深层的特征图具有大感受野、强语义信息、强鲁棒性的特点,适用于分类任务,但分辨率低、丢失的细节信息较多。相反,浅层网络感受野小,特征图分辨率更高,具有更多的细节信息,更有利于目标的定位,但缺乏语义信息。目标检测任务不仅需要对目标进行定位还需要对检测到的目标进行分类,这就要求用于检测层输入的特征图同时具备丰富的细节特征和语义特征。YOLOv5中采用PANet结构来实现多尺度特征融合模块,结构如图3(a)所示。该结构通过双向融合深层特征图的语义信息和浅层特征图的定位信息提高了检测精度,但引入了更多的参数和计算量,降低了模型的效率,且PANet简单地认为来自不同阶段的特征图对最终的融合结果的贡献都是一样的,特征融合效果有待进一步提升。为提高模型效率和融合效果,本研究在PANet的基础上融合BiFPN的思想设计了改进的M-BiFPN,具体结构如图3(b)所示。其中,P_i表示backbone中第i级的特征图,O_i代表neck模块的输出特征图,F_i和N_i代表neck层生成的中间特征图。

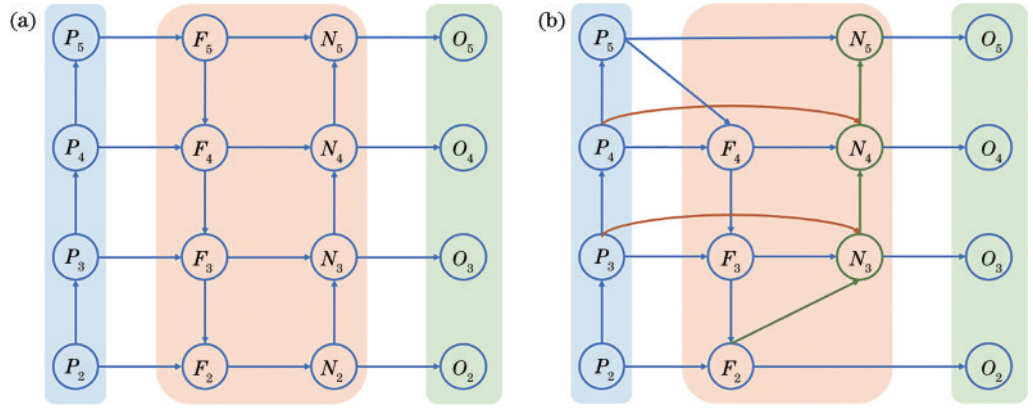

图 3　PANet和M-BiFPN结构。(a) PANet;(b) M-BiFPN

Fig. 3　Structures of PANet and M-BiFPN. (a) PANet; (b) M-BiFPN

M-BiFPN模块具有自顶向下、自底向上、跳跃连接和自适应融合的特点。自顶向下的路径用于向浅层网络传递具有强语义信息的深层特征,提高网络的目标识别能力。自底向上的路径通过向上传递浅层特征的细

节信息，增强整个特征层次的定位能力。由于单输入节点没有涉及任何的特征融合过程，M-BiFPN移除了这些节点，简化了多尺度特征融合结构。此外，为融合更多特征信息，添加了一条输入到输出的跳跃连接。在与输出直接相关的自底向上融合路径中，每一个节点都有多个输入节点，这些节点来自网络的不同阶段，对最终融合结果的影响是不相等的。为改善多特征融合结果，为每一个输入特征设置了一个可学习的特征融合因子，让网络学习每个输入特征的重要性。与BiFPN不同的是，在自顶向下的融合过程中，采用通道级联的方式完成特征融合操作，不仅可以保留更多的特征信息，还可以简化融合计算。M-BiFPN的融合过程可描述为

$$F_i = \mathrm{Conv}\left\{\mathrm{Concat}\left[P_i,\ \mathrm{resize}\left(F_{i+1}\right)\right]\right\}, \tag{1}$$

$$N_i = \mathrm{Conv}\left[\frac{\omega_1 \times P_i + \omega_2 \times F_i + \omega_3 \times \mathrm{resize}\left(N_{i-1}\right)}{\omega_1 + \omega_2 + \omega_3 + \varepsilon}\right], \tag{2}$$

$$O_i = \mathrm{Conv}\left(N_i\right), \tag{3}$$

式中：ω_i是可学习的特征融合因子，代表对应特征的重要性，值越大，特征对最终的融合结果影响越重大；ε是一个远小于1的数字，防止出现分母为0的情况；Conv通常是特征处理相关的一系列卷积操作；Concat表示两个特征图在维度方向上拼接；resize操作用于调整特征图大小，以适应当前特征层的大小。

3.3　SAGN模块

受启发于人眼的注意机制，本研究提出了SAGN模块，其核心思想是利用M-BiFPN中自顶向下的信息传递过程，结合图像分割网络U-Net[20]的思想设计了一个区分目标和背景的语义分割网络，并且将该网络的输出作为空间注意力图反馈至检测层，以达到提高原网络关于图像中感兴趣区域的特征响应的目的，其结构如图4所示。

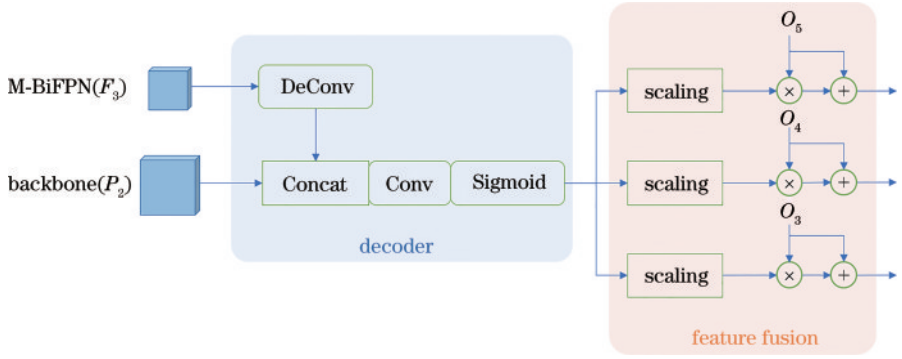

图4　SAGN模块的结构

Fig. 4　Structure of SAGN module

SAGN包含1个解码器和1个特征融合模块。其中，解码器、特征提取网络和M-BiFPN中自顶向下路径共同构成了U-Net的U型结构，用于预测图片中所有目标的空间位置，即图像中的ROI。特征融合模块是1个简单而有效的自注意力机制，用于融合解码器模块生成的前景/背景分割图与原网络中的特征图。具体做法是先将分割图缩放至原特征图的空间维度，然后在空间维度方向上将分割图与特征图按元素相乘，以削减网络对ROI之外区域的响应，降低不相关区域的误报概率。SAGN模块采用监督学习的方式优化网络中的参数，因此实验数据集需要为该模块提供ROI标注信息。考虑到SAGN的任务是提供1个空间注意力图用于增强目标检测网络中的特征图，因此该模块不需要生成非常精细的分割图，只需要学习到目标的大概位置即可。为简化数据集的加载及处理过程，本研究直接使用数据集中的目标位置信息生成ROI的标注信息，效果如图5所示，其中左图为原始图片，右图为该图片对应的ROI标注图。

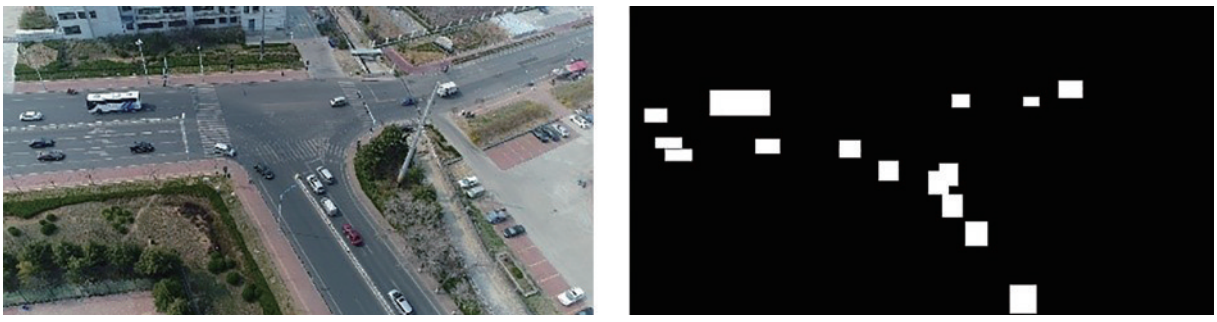

图5　输入图片和ROI标注图

Fig. 5　Input image and ROI label

3.4 损失函数

YOLO-AFENet的损失函数由4部分组成，分别是位置损失、置信度损失、分类损失和注意力损失：

$$L_{\text{total}}=\alpha L_{\text{CIoU}}+\beta\left(L_{\text{obj}}+L_{\text{SAGN}}\right)+\gamma L_{\text{class}}, \quad (4)$$

式中：α、β和γ是对应损失函数的权重，取值分别为0.05、0.07和0.03。

式(4)中的L_{SAGN}表示注意力损失，用于优化网络中SAGN模块的参数。SAGN模块用于生成前景/背景空间注意力图，可将其看作二类图像分割问题，所以损失函数采用二元交叉熵损失函数：

$$L_{\text{SAGN}}=-\left[y_i\log\hat{y}_i+(1-y_i)\times\log(1-\hat{y}_i)\right], \quad (5)$$

式中：y_i表示样本i的真实标签；$\hat{y}_i$是其对应的网络预测标签。

对于置信度损失和分类损失，置信度预测可以看作是二分类问题，损失函数采用二元交叉熵损失函数，计算公式同式(5)。类别预测可以看作是多分类问题，所以损失函数采用多元交叉熵损失函数：

$$L_{\text{cls}}=-\sum_{i=1}^{K}y_i\log\hat{y}_i。 \quad (6)$$

关于位置损失，本文采用complete intersection over union(CIoU)损失[21]作为位置损失函数。IoU描述两个框的重叠区域大小，是目标检测中常用的评估指标，但在IoU相等的情况下无法准确描述两个框的重合度，如两个框不相交时，IoU始终为0，就无法判断两个框的距离。CIoU可以很好地解决IoU相等的问题，它不仅关注重叠区域和中心点的距离，还考虑到两个检测框中心点重合的情况，可以更全面地描述预测框A和真实框B的关系：

$$L_{\text{CIoU}}=1-R_{\text{IoU}}+\frac{\rho^2(A,B)}{c^2}+\alpha\nu, \quad (7)$$

$$R_{\text{IoU}}=\frac{A\cap B}{A\cup B}, \quad (8)$$

$$\alpha=\frac{\nu}{(1-R_{\text{IoU}})+\nu}, \quad (9)$$

$$\nu=\frac{4}{\pi^2}\left(\arctan\frac{w^{\text{gt}}}{h^{\text{gt}}}-\arctan\frac{w}{h}\right)^2, \quad (10)$$

式中：$\rho(A,B)$表示预测框A与真实框B中心点坐标的欧氏距离；c代表能同时包含预测框和真实框的最小闭包区域的对角线距离；α是一个平衡参数，不参与梯度的计算；ν是用来衡量长宽比一致性的参数；w^{gt}、h^{gt}代表真实标注框的宽和高；w、h表示预测框的宽和高。

4 实验结果与分析

4.1 实验数据集

考虑到小目标广泛存在于无人机航拍场景中，并且车辆在日常生活中经常出现，为验证所提方法的有效性和可靠性，本实验采用2018年欧洲计算机视觉会议上由Du等[22]提出的数据集UAVDT。该数据集的检测目标可分为3类，分别是汽车、货车和公交车，由无人机在各种复杂的场景中拍摄，包含较多常见场景，如广场、高速公路、路口等。数据集包含50个视频片段，其中30个视频序列用于训练(有23829张图片)，剩余20个视频序列用于测试(有16580张图片)，每一帧的图像分辨率是1024×540。此外数据集还提供了天气、视角以及高度等属性标注。数据集中小目标、中等目标、大目标的数量关系如表1所示。

表1 数据集目标大小分布

Table 1 Size Distribution of object in dataset

Object	Small object (size≤32×32)	Medium object (32×32≤size≤96×96)	Large object (size≥96×96)	Total
Number	494117(61.9%)	290826(36.4%)	13852(1.7%)	798795(100%)

4.2 实验环境及参数设置

本实验平台操作系统为Ubuntu 20.04，CPU为Intel(R) Xeon(R) E5-2620 v4，GPU为Nvidia GeForce GTX 1080Ti，实验仿真使用PyTorch深度学习框架，开发环境为Python 3.9，PyTorch 1.9.0，CUDA 11.0。

本实验使用的训练集和测试集均采用UAVDT数据集官方说明文档中指定的图片，网络输入图片的分辨率大小设置为640×640。模型训练采用的优化算法是随机梯度下降法(SGD)，训练轮数为120，batch size设为32，初始学习率为0.0032，使用余弦退火策略调整学习率，动量因子为0.843，其中warm up阶段学习率使用线性调整策略，动量因子为0.5。

4.3 评估指标

使用UAVDT的训练集训练网络模型，以UAVDT中的测试集评估模型性能评估，使用MS COCO数据集的评估标准AP和每秒处理图片帧数(FPS)作为模型性能评估指标。AP衡量模型的查全率和查准率，值越高模型检测精度越高。FPS衡量模型的检测速度，值越大表示算法实时性越好，硬件对FPS有一定的影响。受到硬件的限制，在各种算法的对比实验中，部分实验结果未标明FPS。

4.4 实验结果

使用UAVDT数据集对所提改进网络进行训练和测试，图6为AP、AP_{50}曲线图。与改进前的YOLOv5相比，YOLO-AFENet的AP_{50}提高了6.3个百分点，AP提高了2.7个百分点。图7为两个网络的可视化检测效果对比，改进后的算法可以改善漏检、误检问题以及极端环境下的目标检测效果。

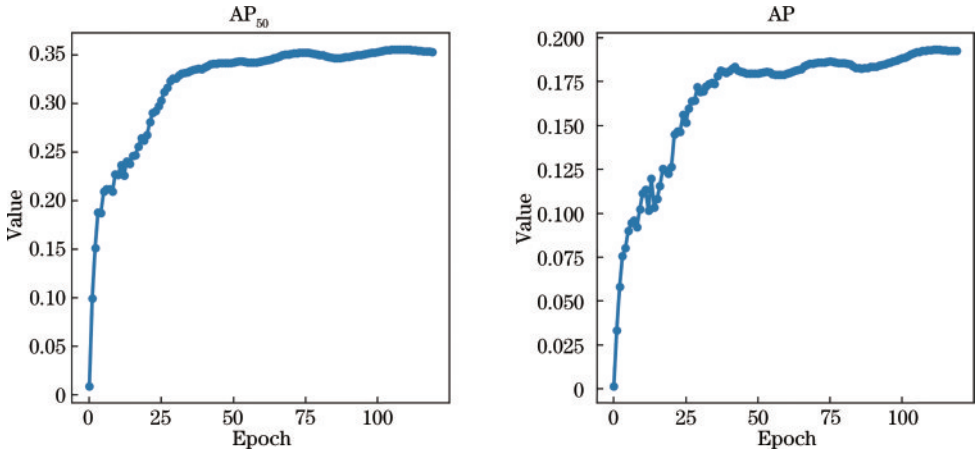

图 6　YOLO-AFENet 的 AP 曲线图

Fig. 6　AP curves of YOLO-AFENet

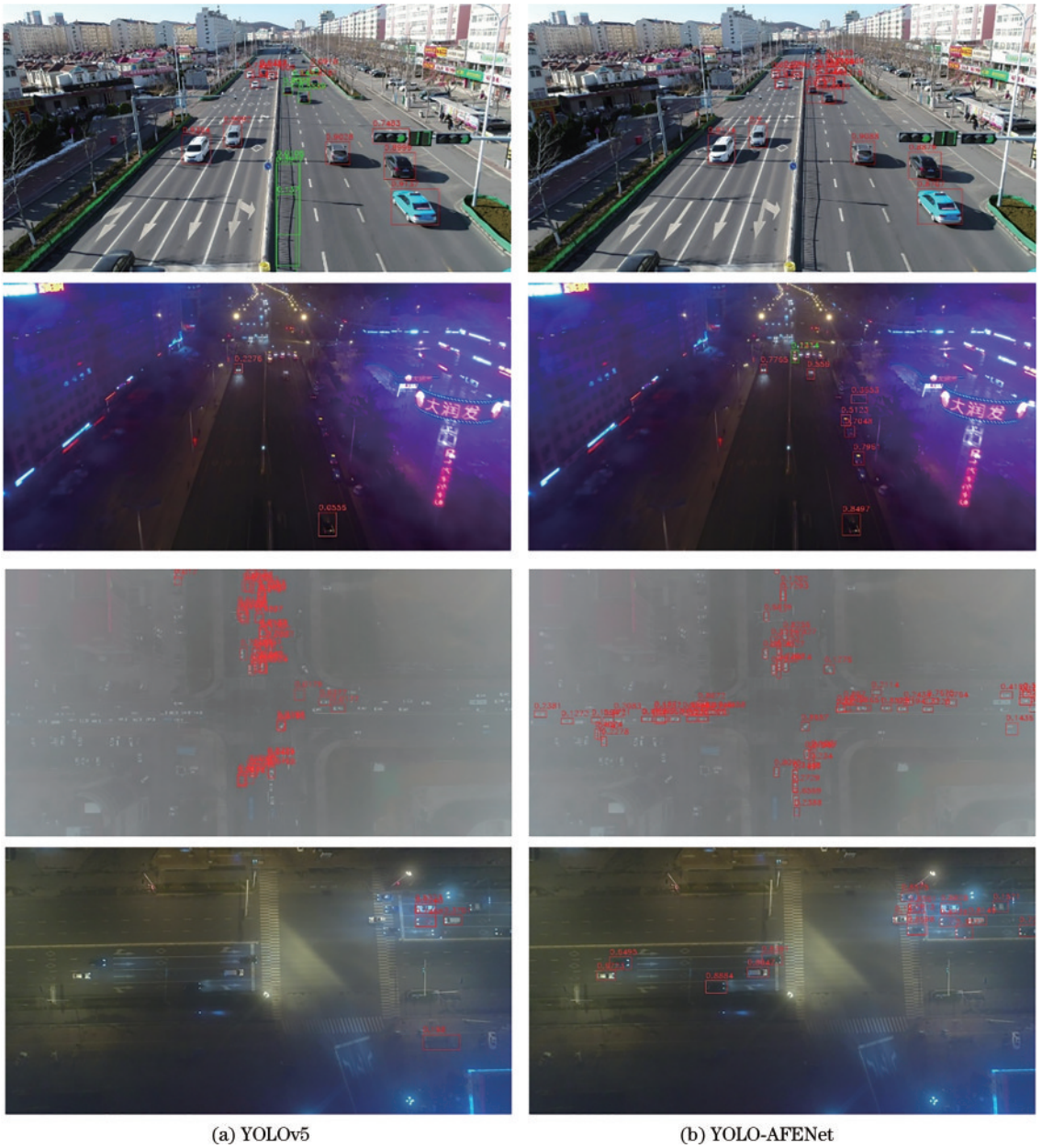

图 7　YOLO-AFENet 与 YOLOv5 在测试集上的检测结果对比。(a) YOLOv5;(b) YOLO-AFENet

Fig. 7　Comparison of detection results of YOLO-AFENet and YOLOv5 on test dataset. (a) YOLOv5; (b) YOLO-AFENet

为进一步验证YOLO-AFENet在精度方面的提升效果，对比了最近几年公开发表的其他目标检测算法在UAVDT数据集上的检测效果，表2给出了所提算法与现有的目标检测网络算法的对比结果，表格中其他网络的实验结果均来自相关论文。与两阶段目标检测算法Faster-RCNN+FPN相比，YOLO-AFENet的AP_{50}、AP分别提高了12.2个百分点、8.4个百分点，与单阶段目标检测算法SSD相比，所提算法的AP_{50}、AP分别提高了14.2个百分点、10.1个百分点，与ClusDet、DMNet相比，所提算法的AP_{50}分别提高了9.1个百分点、11.0个百分点，AP提高了5.7个百分点、4.7个百分点。在检测速度上，所提算法略低于YOLOv5，但高于其他算法。整体而言，所提算法对于小目标的检测效果优于其他算法，验证了所提算法的有效性。

为了验证YOLO-AFENet中各个模块对检测结果的影响，在UAVDT数据集上进行了消融实验，以YOLOv5目标检测网络为消融实验的baseline，实验结果如表3所示。改进的各个模块对于网络的检测精度都有一定的提升，M-BiFPN考虑到特征融合阶段各特征图对最终结果的影响各不相同，使每个特征图的作用最大化，提高了特征图的信息表示能力，与YOLOv5的检测结果相比，AP_{50}提升了6.3个百分点。SAGN利用YOLOv5具有的U型结构的特点，通过有监督的学习方式预先预测了感兴趣区域的位置，并通过特征融合的方式将其反馈给检测特征图，进一步增强了网络对目标的定位能力，将AP_{50}由34.47%提高至35.6%。这两个模块在一定程度上增加了模型的参数量，因此在检测速度上都有所下降，但下降的幅度较小。

表2 不同目标检测算法在UAVDT数据集上的结果对比

Table 2 Comparison of results of different object detection algorithms on UAVDT dataset

Detector	Input	backbone	AP /%	AP_{50} /%	AP_{75} /%	AP_S /%	AP_M /%	AP_L /%	FPS
R-FCN[23]	600×1000	ResNet-101	7.0	17.5	3.9	4.4	14.7	12.1	9
SSD	512×512	VGG16	9.3	21.4	6.7	7.1	17.1	12.0	22
RON[24]			5.0	15.9	1.7	2.9	12.7	11.2	
FRCNN	600×1000	VGG16	5.8	17.4	2.5	3.8	12.3	9.4	7
FRCNN+FPN		ResNet50	11.0	23.4	8.4	8.1	20.2	26.5	18
ClusDet[25]	600×1000	ResNet50	13.7	26.5	12.5	9.1	25.1	31.2	
DMNet[26]	600×1000	ResNet50	14.7	24.6	16.3	9.3	26.2	35.2	
YOLOv5	640×640	CSPDarkNet53	16.7	29.3	16.7	10.2	25.5	33.0	**55.65**
YOLO-AFENet	640×640	CSPDarkNet53	**19.4**	**35.6**	**17.6**	**11.9**	**28.6**	**35.3**	52.86

表3 在UAVDT数据集上的消融实验

Table 3 Ablation experiment on UAVDT dataset

Baseline YOLOv5	With M-BiFPN	With SAGN	AP /%	AP_{50} /%	AP_{75} /%	AP_S /%	AP_M /%	AP_L /%	FPS
√			16.7	29.3	16.7	10.2	25.5	33.0	**55.65**
√	√		17.5	34.47	17.0	10.6	27.7	34.5	54.21
√	√	√	**19.4**	**35.6**	**17.6**	**11.9**	**28.6**	**35.3**	52.86

5 结 论

针对小目标具有像素信息少，特征表示弱等特点，提出一种基于YOLOv5的自适应特征增强的目标检测算法。该算法首先引入可学习的特征融合因子衡量来自不同阶段的特征图对最终结果的重要性，使网络能够更充分地利用各个阶段的特征信息，提高检测层的输入特征图的特征表示能力。然后引入一个空间注意力生成模块，在检测之前预先预测感兴趣区域的位置，并将其反馈给目标检测分支，进一步提升网络的目标定位能力。实验结果表明，与现有的算法相比，所提小目标检测算法在保证检测速度没有明显下降的同时明显提高了检测精度。但是该算法还存在一些可完善的地方，如损失函数未考虑到数据集长尾分布的特点、对于大目标的检测性能提升不明显等，因此后续可以进一步优化网络设计，全面提升网络的检测性能。

参 考 文 献

[1] Zou Z X, Shi Z W, Guo Y H, et al. Object detection in 20 years: a survey[EB/OL]. (2019-05-16)[2021-11-15]. https://arxiv.org/abs/1905.05055

[2] Jiao L C, Zhang F, Liu F, et al. A survey of deep learning-based object detection[J]. IEEE Access, 2019, 7: 128837-128868.

[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] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN [C]//2017 IEEE International Conference on Computer

Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2980-2988.

[5] 王凤随, 王启胜, 陈金刚, 等. 基于注意力机制和Soft-NMS的改进Faster R-CNN目标检测算法[J]. 激光与光电子学进展, 2021, 58(24): 2420001.
Wang F S, Wang Q S, Chen J G, et al. Improved Faster R-CNN target detection algorithm based on attention mechanism and Soft-NMS[J]. Laser & Optoelectronics Progress, 2021, 58(24): 2420001.

[6] 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, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.

[7] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08)[2021-02-03]. https://arxiv.org/abs/1804.02767.

[8] Bochkovskiy A, Wang C, Liao H. Yolov4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2021-02-05]. https://arxiv.org/abs/2004.10934.

[9] 刘峰, 郭猛, 王向军. 基于跨尺度融合的卷积神经网络小目标检测[J]. 激光与光电子学进展, 2021, 58(6): 0610012.
Liu F, Guo M, Wang X J. Small target detection based on cross-scale fusion convolution neural network[J]. Laser & Optoelectronics Progress, 2021, 58(6): 0610012.

[10] 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.

[11] Fu C Y, Liu W, Ranga A, et al. DSSD: deconvolutional single shot detector[EB/OL]. (2017-01-23)[2021-11-15]. https://arxiv.org/abs/1701.06659

[12] 耿鹏志, 杨智雄, 张家钧, 等. 基于SSD的行人鞋子检测算法[J]. 激光与光电子学进展, 2021, 58(6): 0610009.
Geng P Z, Yang Z X, Zhang J J, et al. Pedestrian shoes detection algorithm based on SSD[J]. Laser & Optoelectronics Progress, 2021, 58(6): 0610009.

[13] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: common objects in context[M]//Fleet D, Pajdla T, Schiele B, et al. Computer vision-ECCV 2014. Lecture notes in computer science. Cham: Springer, 2014, 8693: 740-755.

[14] Chen G, Wang H T, Chen K, et al. A survey of the four Pillars for small object detection: multiscale representation, contextual information, super-resolution, and region proposal[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2022, 52(2): 936-953.

[15] 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, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 936-944.

[16] 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, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 8759-8768.

[17] 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), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 10778-10787.

[18] 汪亚妮, 汪西莉. 基于注意力和特征融合的遥感图像目标检测模型[J]. 激光与光电子学进展, 2021, 58(2): 0228003.
Wang Y N, Wang X L. Remote sensing image target detection model based on attention and feature fusion[J]. Laser & Optoelectronics Progress, 2021, 58(2): 0228003.

[19] 刘鑫, 陈思溢, 陈小龙, 等. 基于深度学习的深层次多尺度特征融合目标检测算法[J]. 激光与光电子学进展, 2021, 58(12): 1210029.
Liu X, Chen S Y, Chen X L, et al. Deep multi-scale feature fusion target detection algorithm based on deep learning[J]. Laser & Optoelectronics Progress, 2021, 58(12): 1210029.

[20] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation[M]//Navab N, Hornegger J, Wells W M, et al. Medical image computing and computer-assisted intervention-MICCAI 2015. Lecture notes in computer science. Cham: Springer, 2015, 9351: 234-241.

[21] Zheng Z H, Wang P, Liu W, et al. Distance-IoU loss: faster and better learning for bounding box regression[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.

[22] Du D W, Qi Y K, Yu H Y, et al. The unmanned aerial vehicle benchmark: object detection and tracking[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11214: 375-391.

[23] Dai J F, Li Y, He K M, et al. R-FCN: object detection via region-based fully convolutional networks[C]//NIPS'16: Proceedings of the 30th International Conference on Neural Information Processing Systems, December 5-10, 2016, Barcelona, Spain. New York: ACM Press, 2016: 379-387.

[24] Kong T, Sun F C, Yao A B, et al. RON: reverse connection with objectness prior networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 5244-5252.

[25] Yang F, Fan H, Chu P, et al. Clustered object detection in aerial images[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 8310-8319.

[26] Li C L, Yang T, Zhu S J, et al. Density map guided object detection in aerial images[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 14-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 737-746.