DOI:10.19651/j.cnki.emt.2417035

基于航拍图像的自适应感知目标检测网络*

袁玲玲[1,2] 陈春梅[1,2] 朱天鑫[1,2] 邓 豪[1,2] 刘桂华[1,2]
(1.西南科技大学信息工程学院 绵阳 621000;2.特殊环境机器人技术四川省重点实验室 绵阳 621000)

摘 要:由于无人机拍摄高度和角度的多样性,其图像往往呈现背景复杂且小目标居多的特征,这导致了相关检测算法性能较差。针对此问题,本文提出了一种基于自适应感知网络的航拍图像车辆检测方法,旨在从提高车辆特征显著度和改善特征信息损失两个方面来提升小目标的检测性能。首先,为了提取更高效的特征表征,提出了自适应感知特征提取模块,该模块通过捕捉长程依赖关系和更强的几何特征表示,能够自适应地对物体的形状进行建模。其次,为了减少下采样和连续池化造成的信息损失,设计了双分支空间感知下采样模块,该模块混合不同通道的特征图,以最大限度地保留小目标特征信息。然后,在特征融合网络中,引入了具有丰富空间信息的浅层特征图,以增强小目标的检测能力。最后,设计了新的动态回归损失函数 DEIoU,该函数引入惩罚项来度量真实框与检测框之间横纵比的相关性,从而进一步提高网络的预测精度。在 Visdrone 数据集上的实验结果表明,所提方法的平均精度均值 mAP 达到了 70%,推理速度达到了 99.26 fps,实现了较好的速度与精度的平衡,并且所提方法在 UCAS-AOD 数据集上取得了最佳的检测精度,具有较强的泛化能力。

关键词:无人机;目标检测;自适应感知特征提取;特征融合网络;双分支空间感知下采样
中图分类号:TP391.4;TN06 **文献标识码**:A **国家标准学科分类代码**:520.6040

Adaptive perception object detection network based on aerial photography

Yuan Lingling[1,2] Chen Chunmei[1,2] Zhu Tianxin[1,2] Deng Hao[1,2] Liu Guihua[1,2]
(1. School of Information Engineering, Southwest University of Science and Technology, Mianyang 621000, China;
2. Robot Technology Used for Special Environment Key Laboratory of Sichuan Province, Mianyang 621000, China)

Abstract: Due to the diverse height and angle of drone shots, the images often have complex backgrounds and mainly feature small targets. As a result, the performance of detection algorithms for these images is often poor. To address this issue, this paper presents a vehicle detection method for aerial images using an adaptive perception network. The goal is to improve the detection of small targets by focusing on two aspects: enhancing the saliency of vehicle features and improving the preservation of feature information. First, an adaptive perception feature extraction module is proposed to extract a more efficient feature representation. This module captures long-range dependencies and stronger geometric feature representations to adaptively model the shape of objects. Second, a dual-branch spatial perception downsampling module is introduced to mitigate information loss caused by down-sampling and continuous pooling. This module combines feature maps of different channels to maximize the retention of small target feature information. Next, the feature fusion network incorporates shallow feature maps with rich spatial information and adds detection heads to enhance the detection capability of small targets. Finally, a new dynamic regression loss function, DEloU, is designed. This function includes a penalty term to measure the correlation between the aspect ratio of the ground truth box and the detection box, further improving the prediction accuracy of the network. Experimental results on the Visdrone dataset show that the proposed method achieves an average precision (mAP) of 69.9% and an inference speed of 99.26 fps, indicating a good balance between speed and accuracy. Moreover, the proposed method has achieved the best detection accuracy on the UCAS-AOD dataset and has strong generalization ability.

Keywords: unmanned aerial vehicle (UAV); object detection; adaptive perception feature extraction; feature fusion network; dual-branch spatial-aware downsampling

0 引 言

随着计算机视觉的快速发展,基于无人机(UAV)航拍图像的目标检测在智能交通领域受到了广泛关注,如行人检测[1]、车牌识别[2]、驾驶员行为监测[3]、交通事故检测[4]等关键应用。与传统的静态地面监控摄像头相比,UAV 航

收稿日期:2024-10-04
* 基金项目:中央军委装备发展部项目(23ZG8102)、校博士基金(20zx7123)项目资助

拍图像提供了高度的机动性和更广阔的视野[5]。因此，在特殊路段和意外交通情况下，它往往充当动态辅助视角，为交通流特征分析、交通管理与控制、道路安全评价等研究提供了有力的数据支持，并显著提升城市交通安全和管理效率。但与标准图像相比，UAV 航拍图像的复杂背景、小目标的密集性以及像素点的稀缺性，为目标检测带来了前所未有的挑战。一方面，复杂的背景中视觉杂波丰富，与车辆目标的视觉特性相似，导致目标感知精度下降；另一方面，密集的小目标由于像素点组成较少，有效特征信息稀缺。

国内外研究人员对目标检测网络进行了广泛的研究，包括传统的图像处理方法[6-7]、基于卷积神经网络(convolutional neural network，CNN)的方法[8-10]、基于视觉 Transformer (vision transformer，ViT)的方法[11]。传统的目标检测算法通过手工设计的特征和滑动窗口进行检测。然而，手工设计特征效率低，泛化能力差；滑动窗口冗余，计算开销大。基于 CNN 的目标检测算法分为以 Faster R-CNN[12]为代表的两阶段方法和以 SSD[13]和 YOLO 系列[14-16]为代表的一阶段方法。两阶段算法先生成候选区域，然后检测候选区域内的对象，但速度相对较慢；一阶段算法则通过回归和分类直接预测边界框和类别，但精度相对较低。基于 ViT 的目标检测方法通过自注意力机制捕捉图像的长距离依赖关系和多尺度特征，提高了目标检测的准确性和有效性，但这种高性能伴随着高计算复杂度和对训练资源的高要求。

基于航拍图像的小目标检测，更多研究人员采用的是具有检测速度快和较高精度的一阶段算法进行无人机航拍图像的检测，并通过增强模型提取特征的能力、制定高效的特征融合策略和改进下采样方法等提高对航拍图像小目标检测的性能。Huang 等[17]提出了特征引导增强模块，通过设计两个非线性学习算子，在训练时引导提取更具判别力的特征。Li 等[18]提出了多尺度融合的通道注意力模型，将通道注意力机制与空间金字塔池化相结合，以融合不同尺度的通道特征，提高目标检测精度。翁俊辉等[19]通过改进下采样卷积核与在最大池化分支嵌入上下文信息以减少特征损失。Hao 等[20]设计了基于空洞卷积的多尺度特征重用模块，解决了骨干网络中下采样时小目标特征信息丢失的问题。童小钟等[21]通过多尺度策略融合全局先验、语义信息和位置信息来基于生成对抗网络的方法通过提高小目标的分辨率来增强网络的特征表示能力。Qi 等[22]提出了一个快速多尺度融合模块，采用两个快速上采样算子来统一和融合多尺度特征图，该模块缓解了语义信息和空间信息融合不足的问题。Zeng 等[23]采用不同扩张率的膨胀卷积和跳跃连接将浅层网络中的小目标坐标信息与深层网络中的上下文信息相结合来提高小目标的检测能力。

上述方法在一定程度上提高了检测性能，然而仍存在目标特征显著度低，有效特征的提取不足；小目标特征信息损失严重，小目标检出率低；简单样本与困难样本未能平衡的问题。为了解决上述问题，本文提出了一种自适应感知网络(adaptive perception network，APNet)，旨在通过增强目标显著性、减少特征信息损失和提高小目标的检测率来提高车辆检测的准确性，开展如下工作：

1)针对无人机航拍图像中目标视觉显著性较低的问题，提出了自适应感知特征提取模块(adaptive perception feature extraction module，APFE)，以增强特征表示能力，帮助网络更好的关注目标特征信息。该模块利用可变形卷积和注意力机制，能够自适应地捕捉目标的纹理变化和局部结构。

2)针对目标在连续卷积与池化操作中位置和细节信息损失的问题，提出了双分支空间感知下采样模块(dual branch spatial sensing downsampling module，DSD)。该模块结合了注意力机制和双路下采样结构，最大限度地保留了目标特征信息。

3)为了增强小目标的检测率，将含有更多空间信息、更高分辨率、对小目标更敏感的浅层特征图 C2 输入到特征融合网络中，并集成了一个感知小目标位置信息的检测头。

4)提出了动态损失函数(dynamic and efficient intersection over union，DEIoU)，增加了一个损失项来评估真值和预测值的纵横比，以提高小目标的定位精度和预测的收敛速率。

1　本文方法

针对无人机航拍图像背景复杂、小目标多的特点，本文提出了 APNet，其整体架构如图 1 所示，由输入、主干、颈部和检测头组成。首先，将分辨率为 640×640 的无人机航拍图像输入到由自适应感知特征提取模块和双分支空间感知下采样模块组成的主干特征提取网络中，分别得到四层不同尺寸的特征图{C2，C3，C4，C5}。然后，将浅层的具有高分辨率的特征图 C2 引入到特征融合网络中。接着，通过跨层连接结构将具有语义特征的特征图{P2，P3，P4，P5}以自上而下的方式传输，将具有位置特征的特征图{N2，N3，N4，N5}以自下而上的方式传输，实现不同层次的特征图融合。最后，将融合后的特征输入到检测头中，以获得物体的类别和位置，同时应用 DEIoU 损失函数计算预测的边界框的回归损失。

1.1　自适应感知特征提取模块

YOLOv7-tiny 主干特征提取网中高效层聚合网络(efficient layer aggregation network，ELAN)是核心的特征提取模块，通过多分支结构与注意力机制捕获图片中长距离依赖关系。但由于航拍视角的独特性，在复杂的背景下，ELAN 模块无法提取足够有效的特征来准确检测较小的车辆目标。

本文提出了 APFE 模块处理这个问题，通过动态调整卷积核的形状以适应输入数据的特征，并动态调整特征的权重以增强模型对重要特征信息的关注，从而提高算法的

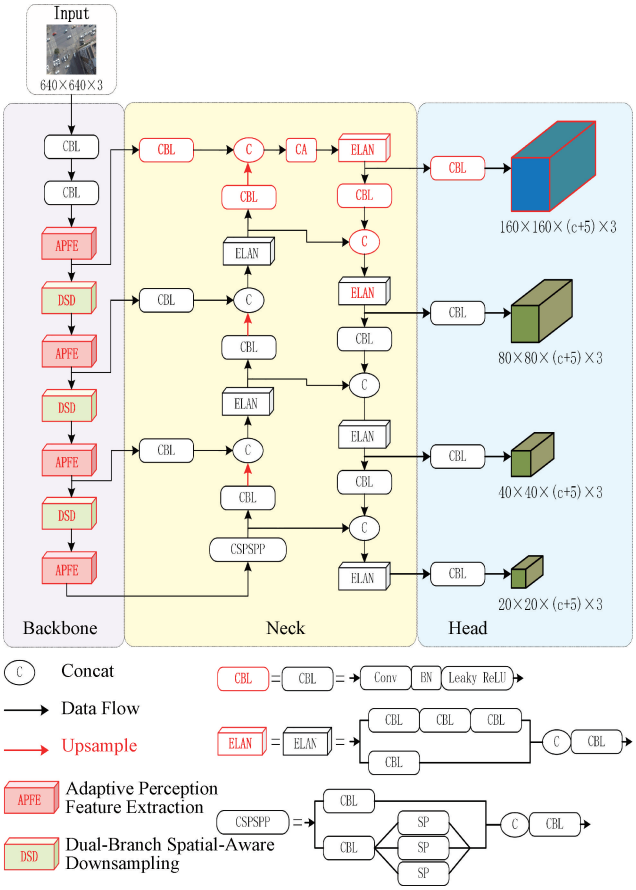


图 1 APNet 网络结构

Fig. 1 Structure of APNet

性能和鲁棒性。具体结构如图 2 所示,首先在输入特征 $X \in \mathbb{R}^{H\times W\times C}$ 上执行两次 1×1 卷积,再将输出 $X_1 \in \mathbb{R}^{H\times W\times \frac{C}{2}}$ 的通道数减少到输入的一半。然后,在 X_1 上执行两个 3×3 可变形卷积[24],并计算每个位置的位置偏移量,然后基于这些偏移量在每个位置周围生成采样点,通过双线性插值得到采样点处的特征值。接下来,在这些采样点上进行动态卷积操作,获得输出特征 $X_2 \in \mathbb{R}^{H\times W\times \frac{C}{2}}$。最后,通过元素求和来融合前两步得到的特征,并采用全局平均池化和 1×1 卷积处理融合特征图以获得局部区域特征信息,整个过程如下:

$$F_{pool} = FC(FC(P_{avg}(concat(X_1, X_2)))) \in R^{1\times 1\times C} \quad (1)$$

其中,$P_{avg}(*)$ 表示全局平均池化,$FC(*)$ 表示全连接层。第 1 个全连接层采用 GELU 激活函数,而第 2 个全连接层采用 Sigmoid 激活函数。最后,对权重图 F_{pool} 和融合特征图进行逐元素乘积操作。

1.2 双分支空间感知下采样模块

对于目标检测,通过最大池化进行下采样是简单且有效的,它可以减少特征图的大小和计算量,并保留主要特征,有助于快速识别目标。但 Yolov7-tiny 仅使用步长为 2,卷积核大小为 2×2 最大池化进行下采样,容易导致小目标信息严重丢失,进而造成小目标识别不准确,降低模型的鲁棒性,影响模型性能。

本文提出了 DSD 模块,能够抑制噪声的同时有效地捕获空间信息,如图 3 所示。在 DSD 模块中,采用两种不同的下采样方式来捕获和保留不同的特征信息,两者相结合

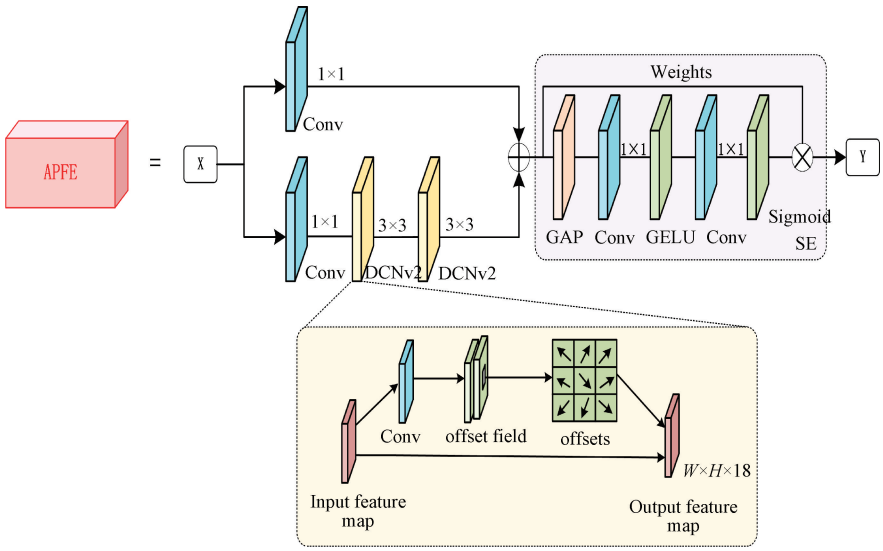


图 2 APFE 结构

Fig. 2 Structure of APFE

可以减少单一方法可能丢失的信息,让输出特征图尽可能多的保留原始图像信息,从而提供更全面的特征表示。Path 1 由步长为 1,卷积核大小为 1×1 卷积和步长为 2,卷积核大小为 3×3 卷积组成,该分支侧重于细节特征的捕捉;Path 2 由步长为 2,卷积核大小为 2×2 最大池化和步长为 1,卷积核大小为 1×1 卷积组成,该分支则侧重于更

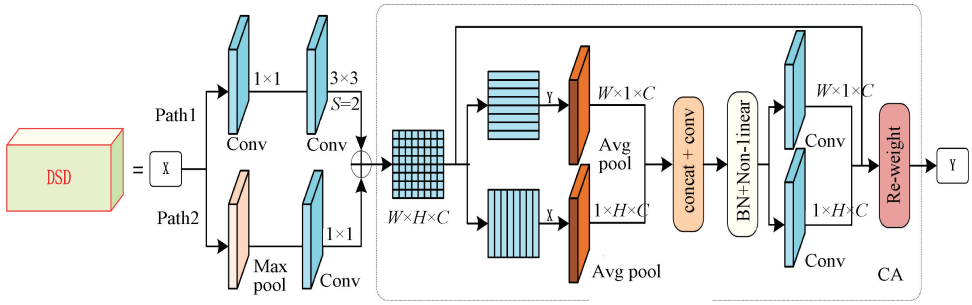


图 3　DSD 结构

Fig. 3　Structure of DSD

抽象的高层特征的捕捉。然后在 Path 1 和 Path 2 的输出特征图上应用坐标注意力机制[25]（coordinate attention，CA）。CA 注意力机制首先对输入特征 f 沿水平和垂直方向进行平均池化，分别计算其在宽度和高度方向上所有像素的均值，得到大小为（$1\times H\times C$）和（$W\times 1\times C$）的特征图。再将上层特征图通过一个全连接层和非线性函数 h_swish 进行变换，以生成空间注意力权重，然后对变换后特征图的每个位置进行 softmax 操作，得到水平方向和垂直方向的注意力权重。最后将原始输入特征图分别和水平方向的注意力权重、垂直方向的注意力权重逐元素相乘法并相加，其过程可表示为：

$$CA(f)=f\cdot Attn_w+f\cdot Attn_h \tag{2}$$

$$Attn_w=Act(FC(GAP_w))$$
$$Attn_h=Act(FC(GAP_h)) \tag{3}$$

其中，f 为输入特征，$Attn_w$ 和 $Attn_h$ 为归一化的注意力权重，$Act(\cdot)$ 为非线性函数，$FC(\cdot)$ 为全连接层，$GAP(\cdot)$ 为全局平均池化。

1.3　小目标检测层

YOLOv7-tiny 算法输出分辨率为 20×20、40×40、80×80 三个不同尺寸的特征图以检测不同尺寸的物体。这 3 个特征图感受野大，语义信息丰富，但缺乏清晰的局部细节特征，更适合检测大中型物体。为了增加网络对小目标关注的同时保留浅层重要的位置信息，本文在原网络的基础上增添了一个分辨率为 160×160 的带注意力机制的检测头。

改进后的网络结构如图 4 所示，特征金字塔网络以骨干网络的 C2、C3、C4、C5 结构输出的特征图为输入，与路径聚合网络（path aggregation network，PANet）中的特征图进行自上向下的融合。融合从语义信息最丰富的最高层，即第 5 层开始，将上采样的特征图添加到第 4 层特征图的对应位置。随后，将融合后的特征图与第 3 层进行相同的操作，不断迭代这个过程，直到第 2 层已经合并了所有上层信息。然后，将特征金字塔网络（feature pyramid network，FPN）中的特征图 P2、P3、P4、P5 作为输入，自下而上地与 PANet 中的特征图进行融合。在从 P5 到 N4 的融合过程中，引入同时关注通道和空间维度的 CA 注意力机制，该机制有助于网络加强对纹理细节的关注，从而增强对小尺寸目标或具有细微特征的目标的检测精度，减少漏检的数量。

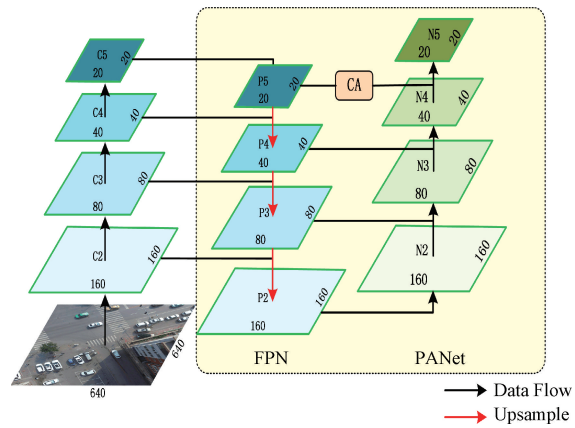


图 4　改进后颈部特征融合网络

Fig. 4　The improved neck feature fusion network

1.4　动态损失函数

在模型训练中，模型会偏向于数量多且和真实框重叠度低的困难样本，而忽略与真实框重叠度高的简单样本，则造成简单样本和困难样本之间不平衡的问题。YOLOv7-tiny 中使用完全交并比（complete intersection over union，CIoU）作为损失函数，并没有直接解决这个问题。为了解决这一挑战，提出了 DEIoU，它是对智能损失（wise intersection over union，WIoU）[26]的改进。DEIoU 在加速模型收敛的同时提高了预测框的准确性。WIoU 定义如下：

$$L_{WIoU}=R_{WIoU}L_{IoU} \tag{4}$$

$$R_{WIoU}=\exp\left(\frac{(x-x_{gt})^2+(y-y_{gt})^2}{(W_g^2+H_g^2)^*}\right) \tag{5}$$

其中，$R_{WIoU}\in[1,e)$，显著放大普通质量锚框 L_{IoU}，而 $L_{IoU}\in[0,1]$，则减少了优质锚框 L_{IoU}，并在锚框与目标框重叠良好时，显著降低其对中心点距离的关注。为了避免 R_{WIoU} 生成阻碍收敛的梯度，W_g 和 H_g 应从计算图中分离出来。

鉴于 WIoU 中动态处理 IoU 损失函数的思想,本文提出了动态损失函数计算准则,定义为:

$$L_{DEIoU} = R_{WIoU}L_{IoU} + \frac{1}{3}\left(\frac{\rho^2(b,b_{gt})}{c^2} + \frac{\rho^2(w,w_{gt})}{c_w^2} + \frac{\rho^2(h,h_{gt})}{c_h^2}\right) \tag{6}$$

其中,C_w 和 C_h 包含两个框的最小边界框的宽度和高度。DEIoU 通过动态调整边界框的回归损失,改善了传统 IoU 的相交和并集,同时削弱了对距离和纵横比等几何测量的惩罚,这样可以更好地考虑预测框和真实框之间的 IoU、位置、大小和形状等多种因素,从而提高目标检测的准确性。DEIoU 通过将损失函数的第 2、第 3 和第 4 项乘以 $\frac{1}{3}$,以较低的水平干预模型对几何测量(如距离和长宽比)的惩罚,增强模型的泛化能力。

2 实验验证

2.1 数据集和实验设置

1) 数据集

VisDrone 数据集[27]由天津大学收集,使用不同的无人机平台在不同的场景、天气条件和光照情况下获得,极具挑战性。该数据集共有 6 471 张训练集图像,548 张验证集图像和 1 610 张测试集图像,涵盖了 10 个不同的类别,但本文的主要关注点是车辆目标,所以仅使用相关的 4 个类别:汽车、货车、公交车和卡车,进行训练和评估,并采用 Mosaic 数据增强技术,对图像进行随机压缩、裁剪、重新排列和拼接。在本文实验中,按照 6∶2∶2划分为训练集、验证集和测试集。

UCAS-AOD 数据集由中国科学院大学于 2014 年发布,共 1 510 张图像分为飞机和车辆 2 个类别,其中飞机有 1 000 张图像,车辆有 510 张图像。在本文实验中,按照 6∶2∶2划分为训练集、验证集和测试集。

2) 实验设置

所有实验都在 Ubuntu 18.04.5 LTS 的服务器上进行。该服务器配备了 AMD 3700X CPU 和一块 8 GB 内存的 NVIDIA RTX2070Super GPU。具体的实验环境包括 Python 3.8.17、PyTorch 1.13.1、torchvision 0.14.1、OpenCV 4.8.0 和 CUDA 11.3。训练过程中使用随机梯度下降(SGD)优化器。所有实验的权重衰减和动量分别设置为 0.000 5 和 0.937,初始学习率设置为 0.01,批量大小设置为 8。

2.2 评价指标

采用准确率(precision, P)、召回率(recall, R)、平均精度(average precision, AP)和平均准确率(mean average precision, mAP)作为评价指标。公式如下:

$$Precision = \frac{TP}{TP + FN} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

其中,TP 为正确识别为正样本的数量,FP 为错误识别为正样本的数量,FN 为错误识别为负样本的数量。

$$AP = \int_0^1 P(R)\mathrm{d}R \tag{9}$$

$$mAP = \sum_{i=1}^{N}\frac{AP_i}{N} \tag{10}$$

其中,AP 为精度-查全率曲线的面积。AP 的值越高,检测性能越好。mAP 是 AP 所有类别的平均值,N 是类别的数量。

2.3 实验结果与分析

1) 对比实验

为了验证本文算法在复杂场景中检测小目标的有效性,在增强的 VisDrone 数据集上,将改进后的算法模型 APNet 与当前主流目标检测算法进行性能对比,实验数据如表 1 所示。

表 1 VisDrone 数据集上不同目标检测算法对比

Table 1 Comparison between the different target detection algorithms on the VisDrone dataset

算法	Precision ↑	Recall ↑	mAP@0.5 ↑	Params/M ↓	GFLOPs/G ↓
Faster RCNN	0.50	0.53	0.48	136.77	369.72
YOLOv3-tiny	0.42	0.46	0.43	8.27	12.90
YOLOv5n	0.62	0.42	0.47	**1.68**	**4.10**
YOLOv5s	0.74	0.52	0.60	6.69	15.80
YOLOv5m	**0.78**	0.59	0.670	19.89	47.90
YOLOv7-tiny	0.73	0.57	0.64	5.74	13.00
YOLOv8n	0.71	0.50	0.57	2.87	8.20
Gold-YOLO	0.76	0.59	0.663	5.35	11.90
RT-DETR	**0.78**	0.60	0.689	42.80	130.50
APNet(本文)	0.77	**0.61**	**0.70**	6.01	14.70

表 1 给出了不同目标检测算法在 VisDrone 数据集上的精度和计算资源消耗实验结果。本文算法的 Precision、Recall 和 mAP 值分别为 0.78、0.62 和 0.70。与模型大小相近的算法 YOLOv3-tiny、YOLOv5s、YOLOv7-tiny、

Gold-YOLO 相比，APNet 检测精度分别领先了 27%、10%、6%、3.7%，其精度和召回率也远高于前述算法。与模型大小最小的两种算法 YOLOv5n、YOLOv8n 相比，本文的精度、召回率和 mAP 有明显的提升。与模型较大的算法 Faster RCNN、YOLOv5m、RT-DETR[28] 相比，本文算法的召回率和 mAP 均最好，在精度上与最佳算法的精度相差不大，且本文的模型大小，复杂度远小于上述算法。

实验结果表明，本文算法在不同尺度和复杂背景的航拍图像上取得了最高的精度，且在模型复杂性和检测性能之间取得了最佳的平衡，如图 5 所示，其参数量和 GFLOPs 值分别为 6.01 M 和 14.7 G，鉴于大部分算法的相关数值集中于较小范围内，图中纵轴的斜线代表了数据截断操作。此操作能够更清晰地呈现数据的分布与变化趋势。在精度至关重要但计算资源有限的场景中，本文方法提供了一个有效的解决方案，在确保相对较低的模型复杂度的同时兼具良好的检测性能，这充分证明了其有效性和实用性。

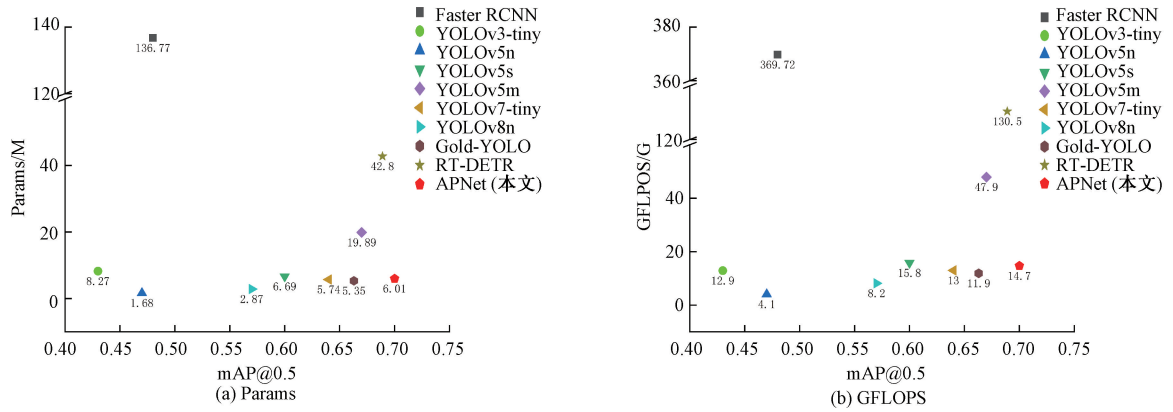


图 5　不同算法间精度与参数量、GFLOPS 的对比

Fig. 5　Comparison of mAP@0.5, Parameters and GFLOPS of different detectors

为了进一步验证本文方法的有效性，在 UCAS-AOD 数据集中与其他目标检测方法进行了比较。结果如表 2 所示，粗体表示最佳值。从表 2 中可以观察到，本文提出的算法模型 APNet 的 mAP 达到了 0.932，超过了其他方法。与检测精度排名第二的方法 RT-DETR 相比，本文方法高出 0.5%。本文方法在飞机检测结果方面是最好的，比排名第二的算法高出 0.2%；在汽车检测中排名第二，比排名第一的 YOLOv5s-CSL 方法落后 1.4%。

表 2　在 UCAS-AOD 数据集上目标检测结果

Table 2　Detection results on UCAS-AOD dataset

算法	Plane	Car	mAP@0.5
Faster RCNN	0.898	0.867	0.886
YOLOv3	0.871	**0.807**	0.839
YOLOv5s-CSL	0.898	0.887	0.893
YOLOv7tiny	0.984	0.843	0.913
YOLOv8n	0.978	0.863	0.92
Gold-YOLO	0.986	0.857	0.922
RT-DETR	0.989	0.864	0.927
APNet（本文）	**0.991**	0.873	**0.932**

此外，模型处理的 FPS 数量也是重要的性能指标，更快的处理速度使模型能够部署在嵌入式设备和移动设备中。从表 3 中可以看出，本文算法模型 FPS 为 99.26，虽然速度上不及 Gold-YOLO[29]，但是在检测精度上取得最好。RT-DERT 参数量与计算复杂度太高，不满足实时性需求，检测性能也远不如本文算法。本文所提出的算法检测精度与速度高且比较均衡，可以实现实时车辆准确检测，能满足实际应用需求。

表 3　本文方法与其他方法速度比较

Table 3　Speed comparison of our method with other methods

算法	mAP@0.5 ↑	帧率/(fps) ↑
Faster RCNN	0.48	13.02
YOLOv3-tiny	0.43	153.85
YOLOv5n	0.47	158.10
YOLOv5s	0.60	147.06
YOLOv5m	0.67	99.01
YOLOv7-tiny	0.64	190.02
YOLOv8n	0.57	120.48
Gold-YOLO	0.663	226.43
RT-DETR	0.689	18.4
APNet(本文)	0.70	99.26

2）消融实验

在 VisDrone 数据集上进行消融实验，验证各改进模块对原始网络模型的优化效果。表 4 给出了消融实验的结果。首先，在 YOLOv7-tiny 模型中引入 APFE、DSD、小目标检测层和 DEIoU 后，与原模型相比，在精度、召回率和 mAP 值上都有一定的提高。A 组：原始 YOLOv7-tiny 网络；B 组：在 A 组的基础上增加一个小目标检测层，精度、

召回率和 mAP 值分别为 0.730、0.591 和 0.658。与基线相比,精度下降了 0.1%,召回率提高了 2.4%,mAP 提高了 2.1%,表明通过整合浅层位置信息和深层语义信息,可以更准确地定位和识别小物体,从而减少漏检。C 组:在 B 组的基础上应用 DEIoU 损失函数,precision、recall 和 mAP 值分别为 0.742、0.590 和 0.667。与 B 组相比,准确率提高了 1.2%,召回率降低了 0.1%,mAP 提高了 0.9%。D 组:在 C 组的基础上,采用 APFE 模块,precision、recall、mAP 值分别为 0.754、0.615 和 0.687。与 C 组相比精度提高了 1.2%,召回率提高了 2.5%,mAP 值提高了 2.0%,表明 APFE 可以提取更有效的特征信息,抗干扰能力强。E 组:在 D 组的基础上,加入 DSD 模块,得到的精度、召回率和 mAP 值分别为 0.763、0.62 和 0.691。与 D 组相比,精度提高了 1.4%,召回率降低了 0.1%,mAP 提高了 1.2%,减少了下采样中的信息丢失,为后续的特征提取和融合奠定了基础。这说明每个模块都不同程度地增强了原模型的整体性能,充分证明了本文方法在提高目标检测精度方面的有效性。

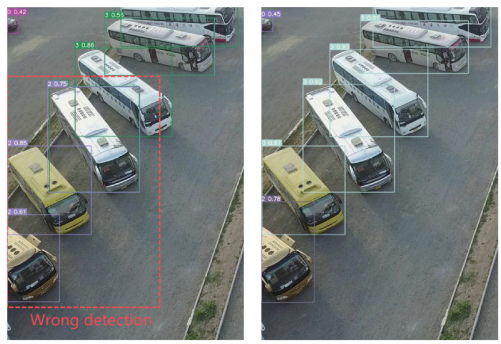
表 4 APNet 在 VisDrone 数据集上的消融实验

Table 4 Ablation experiments of APNet on the VisDrone dataset

编号	基线	s-head	DEIoU	APFE	DSD	Precision	Recall	mAP@0.5
A	√					0.731	0.567	0.637
B	√	√				0.730	0.591	0.658
C	√	√	√			0.742	0.590	0.667
D	√	√	√	√		0.754	0.615	0.687
E	√	√	√	√	√	0.768	0.614	0.699

3)与基线方法的可视化对比

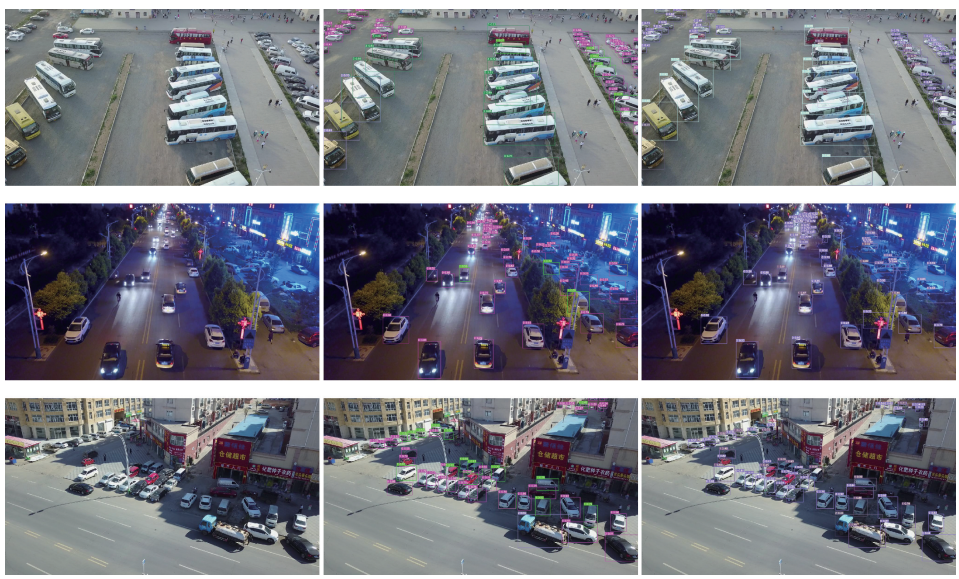
将基线 YOLOv7-tiny 和 APNet 在 VisDrone 数据集上的检测结果进行了视觉比较。如图 6 所示,图中标出了基线方法检测到的假阳性的可视化结果,用红色边框表示。改进后的模型的结果比原算法的结果准确程度有了实质性的提高,在密集目标和复杂环境下,漏检率和误检率都大幅降低。图 7(b)为基线的测试结果,图 7(c)为改进后模型的测试结果。图中,数字 0 表示“car”,1 表示“van”,2 表示 “truck”,3 表示“bus”。



(a) YOLOv7-tiny (b) APNet

图 6 VisDrone 数据集中的误检可视化结果

Fig. 6 Visualization results of false positives in VisDrone dataset



(a) Input　　(b) YOLOv7-tiny　　(c) APNet

图 7　VisDrone 数据集的可视化结果

Fig. 7　Visual results of the VisDrone dataset

3　结　　论

本文针对航拍图像中以小型车辆目标为主、背景复杂导致的检测精度低的问题，提出了端到端的自适应感知目标检测网络 APNet。通过设计自适应感知特征提取模块和双分支空间感知下采样模块，将其嵌入到 YOLOv7-tiny 骨干网络中，使网络能够兼顾全局与局部信息，保留更多的有效特征。同时，还设计了小目标检测层，通过保留更多来自浅层特征层的细粒度信息，增强通道相关性，以提高小目标的检测率和准确率。此外，设计了一个动态损失函数 DEIoU，提升算法的收敛速度和目标的定位精度。在 VisDrone 车辆数据集上进行实验，结果表明，在模型参数量为 6.01 M 的情况下，mAP 达到了 70%；相比基准算法，检测精度提升了 6%。同时，在 UCAS-AOD 数据集上的泛化实验表明，APNet 检测精度 mAP 相比于基准算法分别提升了 1.9%。本文方法在满足实时性的前提下取得了最佳的检测精度。

参考文献

[1] HSU W Y, LIN W Y. Ratio-and-scale-aware YOLO for pedestrian detection[J]. IEEE Transactions on Image Processing, 2020, 30: 934-947.

[2] ALGHYALINE S. Real-time Jordanian license plate recognition using deep learning[J]. Journal of King Saud University-Computer and Information Sciences, 2022, 34(6): 2601-2609.

[3] MA Y F, XIE ZH P, CHEN SH Y, et al. Real-time detection of abnormal driving behavior based on long short-term memory network and regression residuals[J]. Transportation Research Part C: Emerging Technologies, 2023, 146: 103983.

[4] FANG J W, QIAO J H, XUE J R, et al. Vision-based traffic accident detection and anticipation: A survey[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 34(4): 1983-1999.

[5] LIU W, WANG M L, ZHANG SH CH, et al. Research on vehicle target detection technology based on UAV aerial images[C]. 2022 IEEE International Conference on Mechatronics and Automation, 2022: 412-416.

[6] ZHANG N J, TIAN Y H, JIA H X. Research and application of insulator detection method based on information fusion of inspection characteristic quantity[J]. DEStech Transactions on Engineering and Technology Research, 2016, 5: 395-399.

[7] 肖锋，卢浩，张文娟，等. 基于旋转等变卷积的航拍红外图像目标识别算法[J]. 兵工学报，2024，45(8)：2817-2827.
XIAO F, LU H, ZHANG W J, et al. Aerial infrared image target recognition algorithm based on rotation equivariant convolution[J]. Acta Armamentarii, 2024, 45(8): 2817-2827.

[8] 赵亮，刘世鹏. 全局与局部图像特征自适应融合的小目标检测算法[J]. 控制与决策，2023，38(4)：935-943.
ZHAO L, LIU SH P. Small object detection algorithm based on adaptive fusion of global and local image features[J]. Control and Decision, 2023, 38(4): 935-943.

[9] YANG R J, LI W F, SHANG X N, et al. KPE-YOLOv5: An improved small target detection algorithm based on YOLOv5[J]. Electronics, 2023, 12(4): 817.

[10] LIU H Y, DUAN X H, CHEN H N, et al. DBF-YOLO: UAV small targets detection based on shallow feature fusion[J]. IEEJ Transactions on Electrical and Electronic Engineering, 2023, 18(4): 605-612.

[11] ZHANG J, TANG Y L, LUO Y S, et al. Effective grasp detection method based on Swin transformer[J]. Journal of Electronic Imaging, 2024, 33(3): 033008.

[12] REN SH 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.

[13] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]. Computer Vision-ECCV 2016: 14th European Conference, 2016: 21-37.

[14] 王铖浩，骆忠强，漆梓渊. 基于改进 YOLOv5 的变压器

漏油检测[J]. 国外电子测量技术,2023,42(9):169-176.
WANG CH H, LUO ZH Q, QI Z Y. Transformer oil leakage detection based on improved YOLOv5[J]. Foreign Electronic Measurement Technology, 2023, 42(9):169-176.

[15] LI CH Y, LI L L, JIANG H L, et al. YOLOv6: A single-stage object detection framework for industrial applications[J]. ArXiv preprint arXiv:2209.02976, 2022.

[16] WANG CH Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464-7475.

[17] HUANG SH Q, REN SH SH, WU W, et al. Discriminative features enhancement for low-altitude UAV object detection[J]. Pattern Recognition, 2024, 147: 110041.

[18] LI ZH K, LIU X L, ZHAO Y, et al. A lightweight multi-scale aggregated model for detecting aerial images captured by UAVs[J]. Journal of Visual Communication and Image Representation, 2021, 77: 103058.

[19] 翁俊辉,成乐,黄曼莉,等. 基于 CS-YOLOv5s 的无人机航拍图像小目标检测[J]. 电子测量技术,2024,47(7):157-162.
WENG J H, CHENG L, HUANG M L, et al. Small target detection for UAV aerial images based on CS-YOLOv5s[J]. Electronic Measurement Technology, 2024,47(7):157-162.

[20] HAO SH, ZHAO Q L, MA X, et al. YOLO-MSFR: real-time natural disaster victim detection based on improved YOLOv5 network[J]. Journal of Real-Time Image Processing, 2024, 21(1): 7.

[21] 童小钟,魏俊宇,苏绍璟,等. 融合注意力和多尺度特征的典型水面小目标检测[J]. 仪器仪表学报,2023,44(1):212-222.
TONG X ZH, WEI J Y, SU SH J, et al. Typical small target detection on water surfaces fusing attention and multi-scale features[J]. Chinese Journal of Scientific Instrument,2023,44(1):212-222.

[22] QI G Q, ZHANG Y CH, WANG K P, et al. Small object detection method based on adaptive spatial parallel convolution and fast multi-scale fusion[J]. Remote Sensing, 2022, 14(2): 420.

[23] ZENG N Y, WU P SH, WANG Z D, et al. A small-sized object detection oriented multi-scale feature fusion approach with application to defect detection[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 1-14.

[24] WANG W H, DAI J F, CHEN ZH, et al. Internimage: Exploring large-scale vision foundation models with deformable convolutions[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 14408-14419.

[25] HOU Q B, ZHOU D Q, FENG J SH. Coordinate attention for efficient mobile network design[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13713-13722.

[26] TONG Z J, CHEN Y H, XU Z W, et al. Wise-IoU: Bounding box regression loss with dynamic focusing mechanism[J]. ArXiv preprint arXiv: 2301.10051, 2023.

[27] DU D W, ZHU P F, WEN L Y, et al. VisDrone-DET2019: The vision meets drone object detection in image challenge results[C]. IEEE/CVF International Conference on Computer Vision Workshops, 2019: 213-226.

[28] ZHAO Y A, LYU W Y, XU SH L, et al. Detrs beat YOLOs on real-time object detection[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 16965-16974.

[29] WANG CH CH, HE W, NIE Y, et al. Gold-YOLO: Efficient object detector via gather-and-distribute mechanism[C]. Advances in Neural Information Processing Systems, 2023: 51094-51112.

作者简介

袁玲玲,硕士研究生,主要研究方向为航拍目标检测、目标跟踪、零样本学习。

E-mail: yuanlingling1108@163.com

陈春梅(通信作者),副教授,主要研究方向为图像处理、机器学习、物联网技术。

E-mail: ccm@mails.swust.edu.cn

朱天鑫,硕士研究生,主要研究方向为低光图像增强。

E-mail: 2837864977@qq.com

邓豪,博士,主要研究方向为图像去噪、目标检测与跟踪。

E-mail: 452194404@qq.com

刘桂华,教授,主要研究方向为机器人场景智能感知、图像处理、FPGA 集成电路设计。

E-mail: liuguihua@swust.edu.cn