

中图法分类号: TP391.41; V275.1 文献标识码: A 文章编号: 1006-8961(2025)01-0240-14
论文引用格式: Xu F L, Xiong B S, Ou Q F, Yu L and Rao Z B. 2025. Fusion of multi-feature and self-attention for rotating target detection of outdoor helicopter blades. Journal of Image and Graphics, 30(01):0240-0253(徐飞龙, 熊邦书, 欧巧凤, 余磊, 饶智博. 2025. 融合多特征与自注意力的室外直升机桨叶旋转目标检测. 中国图象图形学报, 30(01):0240-0253)[DOI:10.11834/jig.230693]

融合多特征与自注意力的室外直升机桨叶 旋转目标检测

徐飞龙, 熊邦书*, 欧巧凤, 余磊, 饶智博
南昌航空大学图像处理与模式识别省重点实验室, 南昌 330063

摘要: 目的 桨叶运动参数是直升机设计到生产的重要指标, 传统的视觉测量方法直接应用于室外环境下, 由于受复杂光照背景影响, 存在找不到桨叶区域、不能进行准确测量的问题。据此, 本文提出一种融合多特征与自注意力的旋转目标检测器(fusion multi-feature and self-attention rotating detector, FMSA-RD)。**方法** 首先, 针对YOLOv5s (you only look once)特征提取能力不足和冗余问题, 在主干网络中设计了更为有效的多特征提取和融合模块, 结合不同时刻位置与尺度下的特征信息以提高网络对室外桨叶的检测精度; 并去掉部分无关卷积层以简化模块结构参数。其次, 融合多头自注意力机制与CSP(crossstage partial convolution)瓶颈结构, 整合全局信息以抑制室外复杂光照背景干扰。最后, 引入倾斜交并比(skew intersection over union, SKEWIOU)损失和角度损失, 改进损失函数, 进一步提升桨叶检测精度。**结果** 本文进行了多组对比实验, 分别在自制的室外直升机桨叶数据集和公共数据集DOTA-v1.0(dataset for object detection in aerial images)上进行验证, 对比基线YOLOv5s目标检测网络, 本文模型平均精度均值(mean average precision, mAP)分别提高6.6%和12.8%, 帧速率(frames per second, FPS)分别提高21.8%和47.7%。**结论** 本文设计的旋转目标检测模型, 提升了室外复杂光照背景下桨叶的检测精度和速度。
关键词: 室外直升机桨叶; 旋转目标检测; 多特征; 多头自注意力机制(MHSA); 损失函数

Fusion of multi-feature and self-attention for rotating target detection of outdoor helicopter blades

Xu Feilong, Xiong Bangshu*, Ou Qiaofeng, Yu Lei, Rao Zhibo
Provincial Key Laboratory of Image Processing and Pattern Recognition, Nanchang Hangkong University, Nanchang 330063, China

Abstract: Objective The motion parameters of helicopter rotor blades include flapping angle, lead-lag angle, twist angle, and coning angle, which provide an important basis for rotor structure design, upper and lower limit block design of hub, and blade load design. They are important parameters that need to be measured in ground tests before rotorcraft certification and helicopter flight tests. The traditional visual measurement method for rotor blade motion parameters has achieved good results in indoor wind tunnel environments. However, under the influence of complex outdoor backgrounds, some problems exist, such as the inability to detect the rotor blades from the image and measure the parameters accurately.

收稿日期: 2023-10-07; 修回日期: 2024-05-13; 预印本日期: 2024-05-20
* 通信作者: 熊邦书 xiongbs@126.com
基金项目: 国家自然科学基金项目(62473187, 61866027); 江西省重点研发计划资助(20212BBE53017); 江西省研究生创新专项资金项目(YC2022-043)
Supported by: National Natural Science Foundation of China (62473187, 61866027); Key R&D Program of Jiangxi Province, China (20212BBE53017); Innovation Fund Designated for Graduate Students of Jiangxi Province(YC2022-043)

Unlike indoor environments, the outdoors have complex lighting conditions such as different seasons, weather, times, and lighting directions, as well as different sky and background environments. Under these complex background interferences, the features of the rotor blades are weakened, making it difficult to accurately locate the position of the rotor blades. Deep learning is a mainstream object detection method, and how to design deep learning models to enhance the target features of rotor blades and reduce the interference of complex backgrounds is a major challenge. In this paper, on the basis of the network structure of the YOLOv5s, a prediction angle branch is added, and a rotation object detector (FMSA-RD) that fuses multiple features and self-attention is proposed to facilitate the detection of outdoor helicopter rotor blades. **Method** First, the FMSA-RD model improves the C3 module used for feature extraction in YOLOv5s by adding multiple shortcuts. The improved feature extraction module is called convolution five (C5), which completes the feature extraction by fusing local features from different positions, thereby reducing the network structure while maintaining the feature extraction ability. Specifically, C5 replaces the BottleNeck module in C3 with two 3×3 convolution kernels to avoid the additional overhead caused by using multiple BottleNeck modules and increase the receptive field of the convolution layer feature map. Increasing the number of convolution layers does not necessarily lead to optimized parameters and may cause gradient divergence and network degradation. C5 adds shortcut branches to three main convolution layers to effectively avoid the accumulation of useless parameters and extract feature information from different positions. Secondly, the multi-feature fusion spatial pyramid pooling cross stage partial fast tiny module enhances the ability to fuse graphic features at different scales. This module uses a block merging method, using multiple serial 5×5 MaxPool layers to extract four different receptive fields, which improves detection accuracy while keeping the model lightweight. Then, to address the weak ability of convolutional neural network (CNN) structures to connect global features, the B3 module is designed to improve the extraction ability of global features by combining the multi-head self-attention mechanism of Transformer with the crossstage partial convolution (CSP) bottleneck structure, which suppresses the influence of complex outdoor rotor blade backgrounds. Finally, skew intersection over union (SKEWIOU) tilt intersection ratio loss and angle loss are introduced to improve the loss function and further enhance the accuracy of blade detection. **Result** Our experiments were conducted on a self-made outdoor helicopter rotor blade dataset, a self-made outdoor simulated blade dataset, and the public dataset DOTA-v1.0 for training and validation. The self-made outdoor rotor blade dataset contains 3 878 images, which were randomly divided into training, testing, and validation sets in a 7:2:1 ratio. Our FMSA-RD was compared with mainstream horizontal and rotational models such as RetinaNet, FCOS, YOLOv5s, YOLOv6s, YOLOv7 tiny, CenterRot, FAB + DRB + CRB, H2RBox, and R3Det. Experimental results show that our method achieves an average detection accuracy of 98.5% and 110.5 frames per second. On the basis of the comparison experiments using a self-made outdoor blade dataset, the analysis indicates that the mean average precision (mAP) of FMSA-RD is 14.1%, 7.8%, 6.6%, 3.2%, 3.9%, 3.0%, 3.1%, 2.3%, and 4.2% higher than those of RetinaNet, FCOS, YOLOv5, YOLOv6s, YOLOv7 tiny, CenterRot, FAB + DRB + CRB, H2RBox, and R3Det, respectively. The public dataset DOTA contains 2 806 remote sensing images with a resolution of 800×800 , covering various scene types such as cities, industrial areas, buildings, and roads. This comparative experiment aims to verify the generalization ability of the FMSA-RD network. We chose mainstream rotating object detection models for comparative experiments. On the self-made outdoor simulated paddle dataset, the data in the morning and noon are used as the training set, and the data at night are used as the validation set. Experiments show that FMSA-RD has low computational complexity, high detection accuracy, and good generalization ability, making it suitable for different scenarios and environments. **Conclusion** Our FMSA-RD can reduce complexity while integrating local feature information from different positions, suppressing complex background noise interference. The fusion of different scale features improves the accuracy of blade detection. The fusion of self-attention mechanism extracts global information and distinguishes blades without circular markers, achieving accurate detection of complex backgrounds and high aspect ratio blades while reducing model parameters and improving detection accuracy.

Key words: outdoor helicopter rotor blade; rotating target detection; multi-feature; multi-head self-attention mechanism (MHSA); loss function

0 引言

直升机桨叶运动参数测量是旋翼研制工作的重要过程(Zuo等, 2021)。基于立体视觉技术的非接触式测量法已经广泛应用于直升机桨叶运动参数测量,并取得了较好的效果(欧巧凤等, 2021)。该方法首先通过立体视觉技术检测出桨叶上圆形标记点,然后计算标记点的三维信息,最终解算得到挥舞角、摆振角和扭转角等运动参数。目前,直升机桨叶运动参数测量主要基于室内旋翼塔和风洞环境。尽管传统图像处理方法可以定位圆形标记点(Li等, 2022; Kim等, 2023),然而在将这些方法直接应用于室外铁鸟台和直升机地面试验时,由于室外复杂光照背景和桨叶高速旋转的影响,存在找不到桨叶位置的问题,从而无法定位圆形标记点,完成参数测量。据此,本文提出基于视觉摄影的室外直升机桨叶运动参数测量方法,首先从图像中检测出桨叶旋转目标,再通过立体视觉技术计算粘贴在桨叶上标记点的三维坐标,进而解算得到挥舞角、摆振角和扭转角等运动参数。这就要求旋转目标检测网络需要具备高准确性和实时性,能够实时可靠地检测桨叶的位置。

旋转目标检测主要分为多阶段检测器和单阶段检测器两种类型。

多阶段检测器 R2CNN (rotational region convolutional neural network) (Jiang等, 2017) 基于 Faster R-CNN (Ren等, 2017) 网络结构,在二阶段预测层中添加一个回归通道,以实现旋转检测。RoI Transformer (Ding等, 2019) 提出对旋转 RP (region proposal) 进行二次精修的三阶段网络,通过减少低质量 RP,以提升检测精度。Oriented R-CNN (Xie等, 2021) 提出高质量的旋转 RPN (region proposal network),以提升旋转目标检测精度。ReDet (Han等, 2021) 提出 ReResNet (rotation equivariant resnet),通过提取丰富的旋转特征,提高检测精度。Point RCNN (Zhou和Yu, 2022) 根据第一阶段分割的结果来产生候选框,优化候选框的质量,提高检测精度。上述多阶段旋转目标检测器虽然精度高,但是存在冗余卷积运算和大量耗时 RP 问题,直接运用于高速旋转的桨叶环境,无法满足桨叶检测的实时性。

单阶段检测器 R3Det (Yang等, 2021) 以 Reti-

naNet (Lin等, 2017) 为基础,在一个阶段内完成旋转目标检测,有效缩短了推理时间。然而由于从粗粒度到细粒度的直接回归,存在特征提取及融合效果不佳问题,导致检测准确率较低 (Qian等, 2022; Cheng等, 2022)。杜艳玲等人 (2022) 提出通过集成分类与回归两个子网络,对原检测器进行解耦,使两类任务互不影响,可以有效提升对旋转目标的检测精度,但模型检测速度下降。TricubeNet (Kim等, 2022) 采用基于热图的检测过程,能够减少网络设计中锚盒带来的计算和超参数的数量,这样可以简化模型的设计和训练过程,提高了检测的速度。虽然上述单阶段旋转目标检测器相对多阶段检测速度更快,但是直接应用于室外复杂光照背景环境,存在检测精度与速度无法平衡的问题。

针对室外复杂光照背景和桨叶高速旋转的环境,本文基于 YOLOv5s (Ultralytics, 2020) 的网络结构基础,提出一种融合多特征与自注意力的旋转目标检测器 FMSA-RD (fusion multi-feature and self-attention rotating detector),主要有以下 4 部分贡献: 1) 为了准确提取桨叶的多特征,同时满足高速旋转的需求。设计 C5 (convolution five) 模块,融合桨叶不同时刻位置的局部特征,同时减少模型的冗余参数。2) 为了应对复杂的室外环境中桨叶目标位置和大小变化,设计 SPPFCSP (spatial pyramid pooling crossstage partial fast tiny) 模块,融合多个不同尺度的空间特征,包括不同大小的感受野和不同分辨率的特征图。3) 为了提高模型对光照变化的适应性,设计 B3 (CSP bottleneck with three convolutions) 模块,结合 Transformer 的多头注意力机制与 CSP 瓶颈结构,更加关注于室外复杂光照特性。4) 为了解决水平检测器 IoU (intersection over union) 指标无法正确评估旋转框之间的交并比问题,引入 SKEWIOU (skew IoU) 倾斜交并比改进位置损失。使用 $Smooth_{L1}$ 损失函数作为角度损失函数,具有较强的非线性学习能力,进一步提升了桨叶的检测精度。

1 FMSA-RD

1.1 整体结构

FMSA-RD 的整体结构如图 1 所示,分为骨干网络、融合网络和输出网络。C5_1 和 C5_2 是 C5 模块卷积步幅为 1 和 2 的不同结构。C5_1 用于骨干网络

的特征提取, C5_2用于融合网络的下采样。CSPSPFT和B3分别是多特征融合和引入自注意力机制的CSP模块。CSPSPFT模块用于获取全局的特征信息, B3模块用于捕获特征图的重要上下文信息。FMSA-RD与水平检测器相比, 多了一个回归角度参数, 因此, 在输出网络不同尺寸锚框中, 输出深度分别是预设锚框 (x, y, w, h, θ) 、桨叶类别和置信度

参数。

1.2 多特征模块

1.2.1 多特征提取C5模块

C3模块由YOLOv5s网络提出, 主要用于提取骨干网络的特征信息, 其网络结构如图2(a)所示。C3模块包含卷积层和BottleNeck两部分, 借鉴CSPNet (crosstage partial network) (Wang等, 2020)的训练

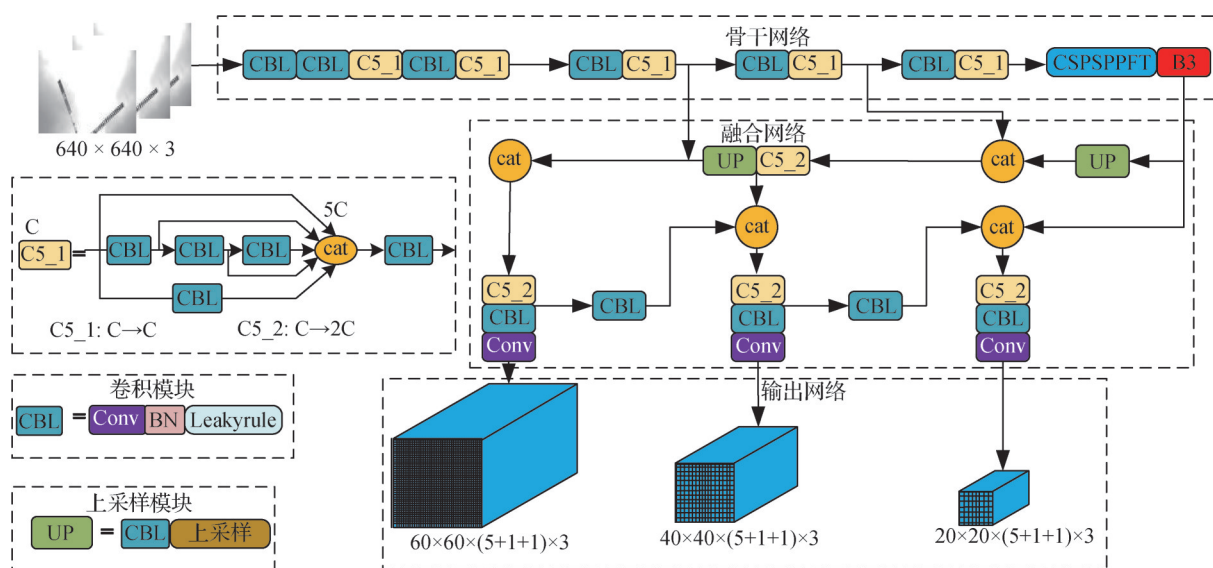


图1 FMSA-RD网络结构

Fig. 1 Structure of FMSA-RD network

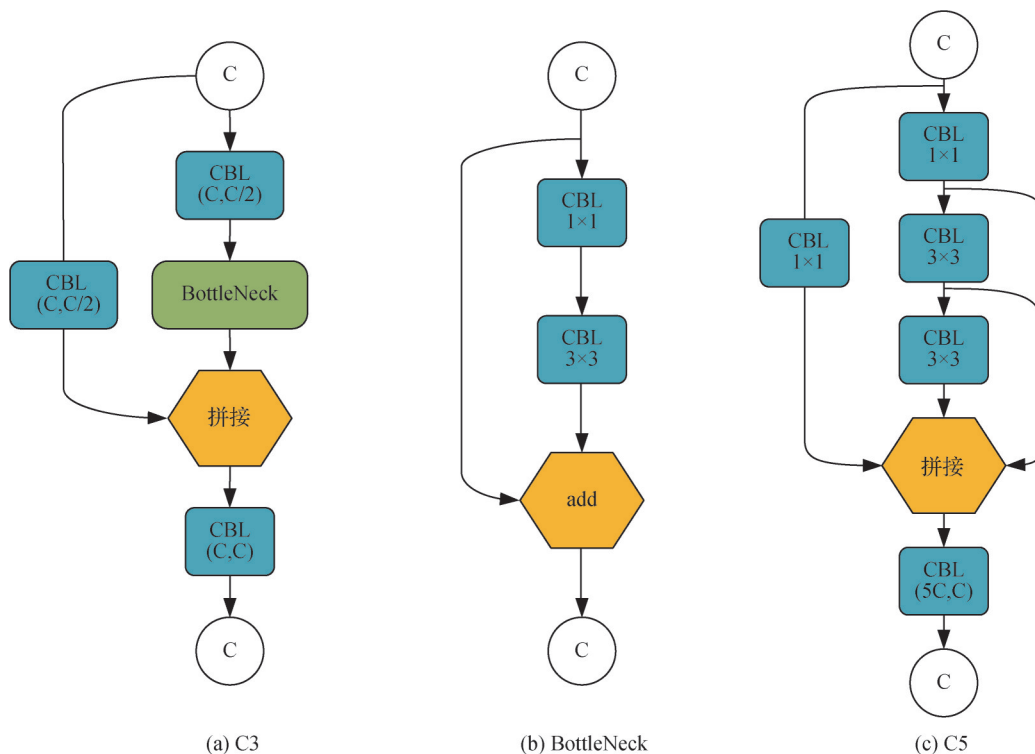


图2 3种特征提取模块结构

Fig. 2 Structure of three feature extraction modules ((a) C3; (b) BottleNeck; (c) C5)

策略,使用两条梯度路径,以减少多个卷积层对重复梯度信息的学习。BottleNeck 是 ResNet (He 等, 2016)的一种简化,其具体网络结构如图 2(b)所示。BottleNeck 首先使用 1×1 的卷积层,然后是 3×3 的卷积层,最后与捷径分支相加。相比于 ResNet, BottleNeck 省略了一个用于升维的 1×1 卷积层,直接使用 3×3 的卷积层进行升维操作。C3 模块采用多次叠加的 BottleNeck 模块,提取不同的特征信息,同时也导致卷积层的冗余,降低学习效率。

C5 是 C3 的改进模块,通过减少网络结构的复杂性,同时结合不同的捷径位置,提取多特征,其网络结构如图 2(c)所示。C5 模块使用两个 3×3 的卷积核来代替 BottleNeck 模块,避免使用多个 BottleNeck 模块带来的额外开销,并且通过增大卷积层特征图的感受野,可以更好地捕捉特征的上下文信息。由于增加卷积层数并不总是学习到更优化的参数,反而可能导致梯度发散和网络退化的问题(黄健宸等, 2022)。因此, C5 模块在 3 个主要卷积层上增加了捷径分支,有效避免了无用参数的累加,并能够提取不同位置下的特征信息。相比原本的 C3 模块而言, C5 具备高效的多特征提取能力。

1. 2. 2 多特征融合 SPPFCSPSPT 模块

SPPF (spatial pyramid pooling fast) 由 YOLOv5s 网络提出,用于融合不同特征图信息。SPPF 通过串行多个相同大小的 MaxPool 层,对输入进行处理,融合不同大小的感受野,其网络结构如图 3(a)所示。SPPF 模块的设计初衷是为了提高速度,通过串行多个 5×5 大小的 MaxPool 层替换原本并行的 9×9 和 13×13 大小的 MaxPool 层,以实现显著的速度提升。然而,在复杂的室外直升机桨叶环境中, SPPF 的融合特征能力表现不佳,导致桨叶检测的平均精度较低。

SPPFCSPSPT 是参考 SPPCSPC (spatial pyramid pooling crossstage partial convolution)(Wang 等, 2023)的设计思路,将 CSPNet 网络结构应用于 SPPF。SPPF 是通过最大池化来获取不同的感受野,从而增加感受野范围,最后融合多个尺度的语义信息,提升模型的检测平均精度。CSPNet 则是采用分块归并方式,轻量化的同时提高检测精度。本文结合两者的思想,提出高效的 SPPFCSPSPT 模块,其网络结构如图 3(b)所示。首先, SPPFCSPSPT 使用 1×1 卷积层将输入特征分为两部分。其次,对其中一个部分进行

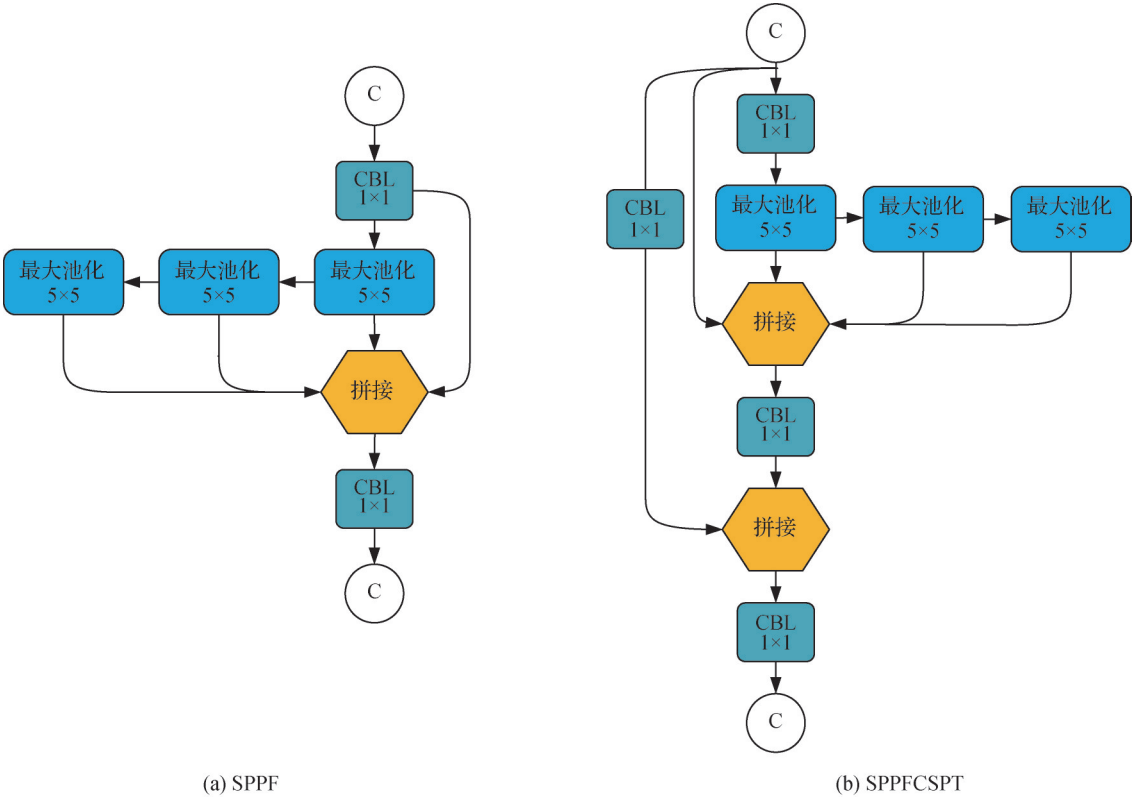


图3 SPPF 与 SPPFCSPSPT 结构
Fig. 3 Structure of SPPF and SPPFCSPSPT ((a) SPPF; (b) SPPFCSPSPT)

常规卷积处理,同时对另一个部分使用多个串行 5×5 大小的 MaxPool 层,提取 4 个不同尺度的感受野。最后,将两个部分提取的特征信息进行融合。通过这种方式,更好地捕捉室外复杂光照背景下直升机桨叶的语义信息,并融合不同尺度下的多特征信息,从而提升桨叶的平均精度。

1.3 融合自注意力机制 B3 模块

多头自注意力机制 (multi-head self-attention mechanism, MHSA) (Vaswani 等, 2017) 可以视为将一个查询向量映射到一系列键值对上,得到该查询向量与所有键的关系,并根据这些关系计算出相应的输出,其具体结构如图 4(a) 所示。首先,假定原有输入维度 d_m , 给定查询 Q 、键 K 和值 V 的维度为 d_k, d_k, d_v 。每个注意力头 $H_i (i = 1, \dots, n)$ 的计算为

$$H_i = F(W_i^{(q)}Q, W_i^{(k)}K, W_i^{(v)}V) \quad (1)$$

式中, $W_i^{(q)}, W_i^{(k)}, W_i^{(v)}$ 为可学习的参数, 维度分别是 $d_m \times d_k, d_m \times d_k, d_m \times d_v$ 。 F 表示将查询向量和键向量进行线性变换后相加, 然后使用 softmax 函数将其归一化。为了将多头输出映射回原始维度, 需要对

拼接后的输出进行一次线性变换, 得到最终的输出 Z , 维度为 d_m , 计算为

$$Z = \text{Concat}(H_1 + H_2 + \dots + H_n)W_o \quad (2)$$

式中, Concat 为拼接, W_o 为可学习的参数, 维度为 $d_m \times hd_v$ 。 $H_i (i = 1, \dots, n)$ 的维度为 d_v 。 MHSA 学习图像中不同位置之间的全局关系, 通过计算注意力得分, 将不同位置之间的相关性进行建模, 更好地捕捉到特征图中的语义信息。

B3 模块是融合 MHSA 与 CSPNet 瓶颈结构, 其网络结构如图 4(b) 所示。首先, 使用一个 1×1 卷积, 改变输入特征图的通道数, 降低网络的计算量。其次, 通过 MHSA 学习序列中不同位置之间的全局关系。然后, 使用另一个 1×1 卷积将输入特征图分成两个分支, 直接与另一部分相加, 帮助特征的传递和融合。最后, 使用 1×1 卷积进行通道数变换, 以适应下一层的输入要求。 B3 模块通过融合 MHSA 的全局关系学习能力和 CSPNet 的特征传递和融合能力, 能够更好地适应不同光照条件下的图像特征, 提升模型对光照变化的适应性。

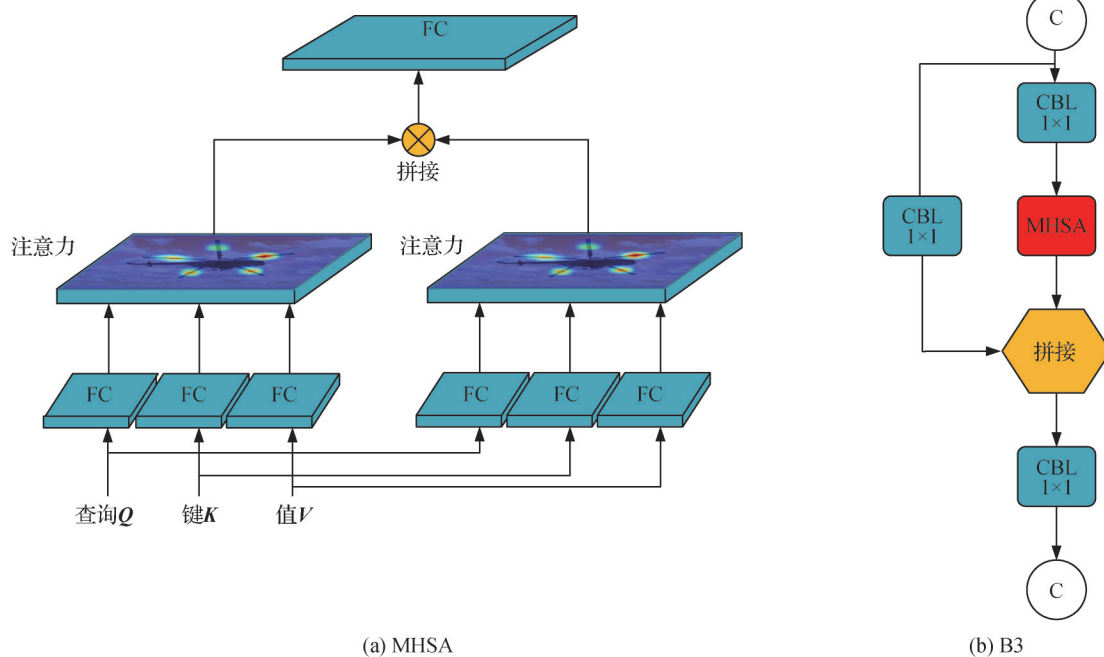


图4 MHSA与B3结构

Fig. 4 Structure of MHSA and B3 ((a) MHSA; (b) B3)

1.4 损失函数改进

传统的交并比 (IOU) 指标是计算两个水平矩形框的重叠面积与联合面积的比值, 衡量框之间的重合程度。但对于旋转目标, 框的形状不再是水平矩

形, 而是旋转的矩形。因此, 本文提出一种改进的位置损失函数, 引入 SKEWIOU 改进 CIOU (Zheng 等, 2020) 损失函数。在计算 SKEWIOU 时, 关键步骤是计算交集面积 (Ma 等, 2018), 即将交集区域划分成

若干个三角形,再对这些三角形的面积求和。旋转矩形框 R_1, R_2 的位置示意图如图 5 所示, $SKEWIOU$ 就是 R_1, R_2 交集面积与并集面积的比值,其中, p_{si} 为 R_1, R_2 的交点, p_{vi} 为 R_1, R_2 落在对方区域内的顶点。

改进的位置损失函数包含倾斜交并比损失和角度损失两部分。其中,倾斜交并比损失保留了 CIOU 损失的优势,对边界框的重合度、中心距离以及宽高比例等多重因素进行综合考虑,确保边界框的精准回归。改进后的倾斜交并比损失函数计算为

$$l_{SKEWIOU} = SKEWIOU - \frac{\rho^2(b, b^{gt})}{c^2} - \beta v \quad (3)$$

$$\beta = \frac{v}{1 - SKEWIOU + v} \quad (4)$$

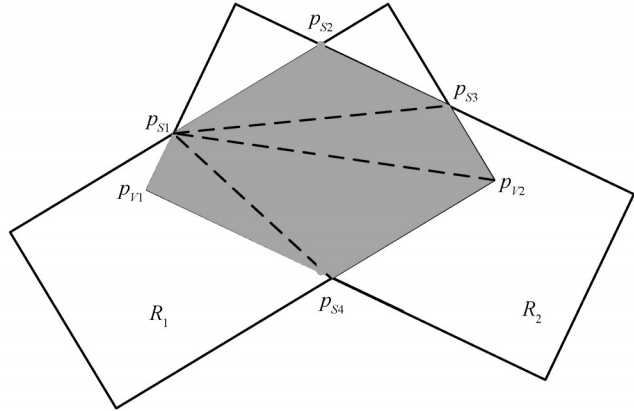
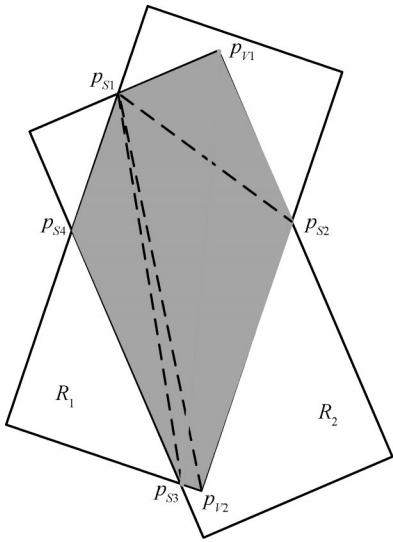


图5 旋转矩形框 R_1, R_2 的位置示意图

Fig. 5 Schematic diagram of the position of rotating rectangular boxes R_1 and R_2

2 实验与结果分析

2.1 数据集

本文实验使用了3个数据集:自制室外桨叶数据集、自制室外仿真桨叶数据集和公共数据集 DOTA-v1.0(dataset for object detection in aerial images)。部分数据集图像如图 6—图 8 所示。3 个数据集分别为:

1)自制室外桨叶数据集。是在高达 25 m 的旋翼台上,使用 2 台 BASLER acA2000-340 km 高速工业相机对真实运动桨叶进行采集。相机曝光时间为 50 μ s,分辨率为 2 040 $\times$ 1 024 像素。将相机放置在旋翼正下方,对旋转桨叶进行逐帧拍摄,共生成

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right) \quad (5)$$

式中, $SKEWIOU$ 为真实框与预测框的倾斜交并比, $\rho^2(b, b^{gt})$ 是预测框中心点到真实框中心点的欧氏距离, c 表示包含预测框和真实框的最小外界矩形的对角线长度, v 为衡量预测框和真实框长宽比差异性的参数, u 越小表明两者的长宽比例越接近, β 为权衡系数。

采用 Smooth L1 损失函数计算角度损失,计算为

$$l_\theta = \text{Smooth}_{L1}(\theta) = \begin{cases} 0.5\theta^2 & |\theta| \leq 1 \\ |\theta| - 0.5 & \text{其他} \end{cases}$$

式中, θ 为弧度制角度。

了 3 878 幅图像。

2)自制室外仿真桨叶数据集。是由虚幻引擎 4 制作,创建不同的天气条件,设置与真实环境相同相机参数,逐帧渲染图像序列,共生成 5 000 幅图像。

3)DOTA-v1.0(Xia 等, 2018)。是一个大规模遥感图像数据集,拥有 2 806 幅分辨率为 800 $\times$ 800 像素的遥感图像,包含多种各种尺度、形状和旋转角度的目标,涵盖城市、工业区、建筑物和道路等多种场景类型。本文使用 DOTA-v1.0 数据集,验证 FMSA-RD 模型在其他旋转目标数据集的泛化能力。

2.2 实验设置及评价指标

2.2.1 实验参数

实验操作系统为 Ubuntu16.04, CPU 是 Intel Core i7-9700K, GPU 使用 NVIDIA GTX 1080TI 11 GB,



图6 自制室外桨叶数据集示例图像

Fig. 6 Sample images of self-made outdoor blade dataset



图7 自制室外仿真桨叶数据集示例图像

Fig. 7 Sample images of a homemade outdoor simulated paddle dataset

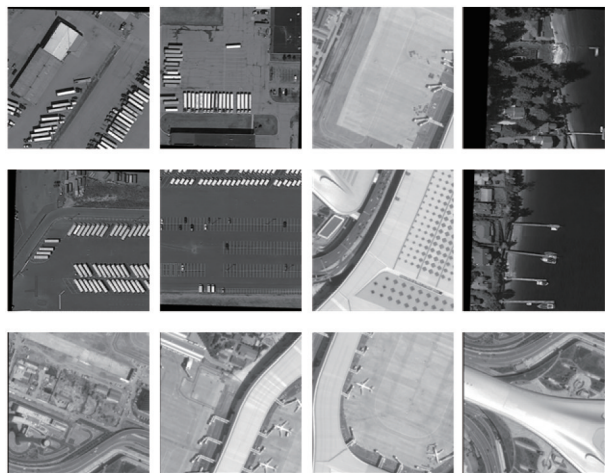


图8 DOTA-v1.0数据集示例图像

Fig. 8 Example images of the DOTA-v1.0 dataset

FMSA-RD网络模型是在YOLOv5开源代码v6.0版本上进行修改实现。模型训练应用上下翻转、左右翻转、马赛克等数据增强方法,重要实验参数如表1所示,其中Adam(adaptive moment estimation)是自适应矩估计优化器。在训练自制室外直升机桨叶数据集和DOTA-v1.0时均不加载预训练模型,保持实验

的一致性。

表1 重要实验参数设置

Table 1 Important experimental parameter setting

实验参数	取值
输入尺寸/像素	1 024 × 1 024
训练批次	300
优化器	Adam
学习率	0.01
上下翻转概率	0.5
左右翻转概率	0.5
马赛克概率	0.75

2.2.2 评价指标

为了验证FMSA-RD模型检测性能,选取了以下评估指标:

1)平均精度均值(mean average precision, mAP)。是衡量算法检测精度的指标,数值越大说明精度越高。

2)帧速率(frames per second, FPS)。是衡量算法运行速度的指标,数值越大说明速度越快。

3)每秒浮点运算次数(floating point operations per second, FLOPs)。是计算过程中所需要执行的浮点运算次数,用于评估旋转目标检测算法的复杂度性能。

2.3 FMSA-RD网络对比实验

为验证FMSA-RD网络的有效性,本文在自行构建的数据集上进行消融实验、对比实验和公共数据集DOTA-v1.0的对比实验。

2.3.1 消融实验

为分析所提出模块性能,进行消融实验。实验结果如表2所示。单独加入C5模块后,在mAP降低少量情况下,FPS提高24.5%,同时模型复杂度降低,这说明C5模块结合不同的捷径位置,提取多特征策略在提高图像处理速度方面非常有效。尽管在加入C5模块后可能会出现一些精度损失(mAP降低),但在相同的时间内,它可以处理更多的图像帧(FPS提高),这对于要求实时性能的桨叶应用场景非常有益。单独加入SPPFCST模块后,mAP提高4.7%,FPS下降3.0%。

如图9所示,SPPFCST模块在增加模型的目标检测准确度方面表现良好,将CSPNet网络结构应用

表 2 消融实验结果对比
Table 2 Comparison of ablation experiment results

模型	C5	SPPFCSP	B3	改进损失函数	mAP@0.5/%	FPS/(帧/s)	FLOPs/G
YOLOv5s	×	×	×	×	92.2	90.9	15.7
模型 1	√	×	×	×	90.1	113.2	12.3
模型 2	×	√	×	×	96.5	88.2	16.1
模型 3	×	×	√	×	95.6	89.4	15.6
模型 4	×	×	×	√	97.3	91.2	15.3
FMSA-RD	√	√	√	√	98.5	110.5	12.6

注:√为模型中包含该模块,×为模型中不包含该模块。

于SPPF可以有效提高精度。尽管在加入SPPFCSP模块后,处理每秒的图像帧数有所下降,但获得了更高的目标检测准确性。单独加入B3模块后,mAP提高3.7%。如图10和图11所示,B3模块的设计通过融合了MHSA的全局关系学习能力与CSPNet的特

征传递和融合能力,能够更好地适应不同光照条件下的图像特征。这种融合模块的引入,能够使模型更加关注复杂光照背景和带有圆形标记点的桨叶,从而提高了模型的mAP指标,使其更适用于光照变化较大的场景。单独加入改进损失函数后,mAP提

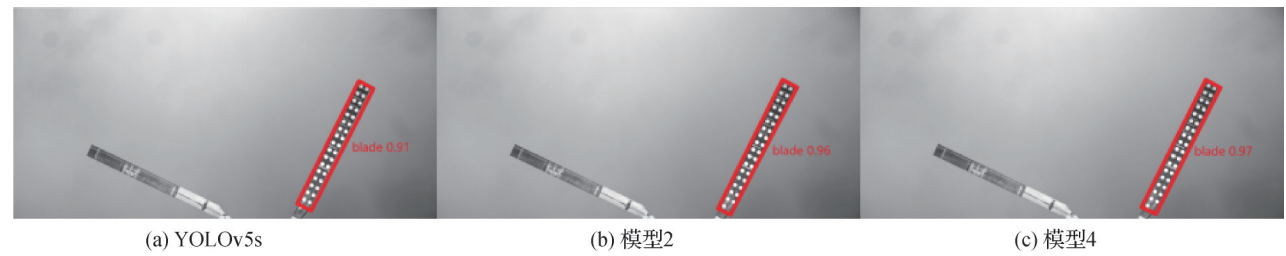


图 9 YOLOv5s、模型 2 和模型 4 效果对比图
Fig. 9 Effect comparison chart of YOLOv5s, model 2 and model 4 ((a) YOLOv5s; (b) model 2; (c) model 4)



图 10 模型 3 光照变化的效果图
Fig. 10 Rendering of lighting changes in model 3

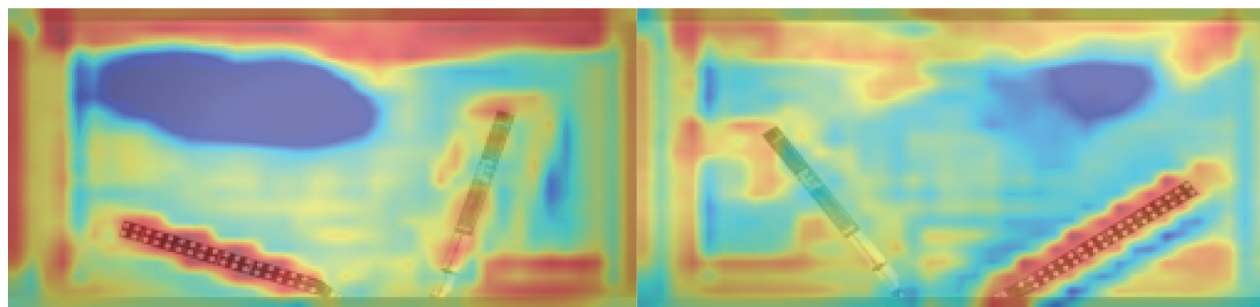


图11 模型3热力图

Fig. 11 Heat maps of the model three

高5.5%。如图9所示,改进损失函数对边界框的重合度、中心距离以及宽高比例进行综合考虑,会提高角度准确率,进而提升桨叶准确度。同时加入3个改进模块,mAP提高6.8%,FPS提升21.6%。消融实验结果表明,通过对模型的不同部分进行改进,优化了模型的整体性能。

2.3.2 对比实验

为了对比FMSA-RD网络与主流水平模型和旋转模型在自制室外桨叶数据集上的性能,设计对比实

验。对比水平模型包括RetinaNet、FCOS、YOLOv5s、YOLOv6s、YOLOv7 tiny,旋转模型包括CenterRot、FAB + DRB + CRB、H2RBox、R3Det。实验结果如表3所示。从表中可以看出,RetinaNet、FCOS、CenterRot、FAB + DRB + CRB、H2RBox、R3Det的FPS较低,说明不适合运用于高速旋转的桨叶环境。FMSA-RD网络在自制的桨叶数据集较对比算法,具有最高的mAP和FPS。实验结果表明,本文提出的FMSA-RD网络,能够有效提升桨叶的检测精度和检测速度。

表3 不同旋转目标检测算法和水平目标检测算法在自制室外桨叶数据集上的对比检测结果

Table 3 Comparison of detection results among different rotating target detection algorithms and horizontal target detection algorithms on self-made outdoor blade dataset

类型	模型	尺寸/像素	mAP@0.5/%	FPS/(帧/s)
H	RetinaNet(Lin等,2017)	1 024 × 1 024	86.3	40.5
H	FCOS(Tian等,2020)	1 024 × 1 024	91.4	53.2
H	YOLOv5s(Wang等,2020)	1 024 × 1 024	92.4	90.7
H	YOLOv6s(Li等,2022)	1 024 × 1 024	95.4	80.2
H	YOLOv7 tiny(Wang等,2023)	1 024 × 1 024	94.8	100.8
R	CenterRot(Wang等,2021)	1 024 × 1 024	95.6	38.5
R	FAB + DRB + CRB(Yuan等,2022)	1 024 × 1 024	95.5	37.6
R	H2RBox(Yang等,2022)	1 024 × 1 024	96.3	68.4
R	R3Det(Yang等,2021)	1 024 × 1 024	94.5	32.3
R	FMSA-RD(本文)	1 024 × 1 024	98.5	110.5

注:加粗字体表示各列最优结果,“H”表示水平目标检测算法,“R”表示旋转目标检测算法。

为验证FMSA-RD网络检测效果,选取一幅代表性桨叶图像,与上述模型检测效果进行对比。检测效果如图12所示,水平模型的检测区域与真实桨叶存在较大冗余。FMSA-RD网络检测出的桨叶区域是完整包括圆形标记,且与原标签最相近。因此,FMSA-RD网络具有更好的检测效果。

为验证FMSA-RD网络具有泛化性,在相同的公共数据集DOTA-v1.0上与主流模型进行对比实验,同时在自制室外仿真桨叶数据集上,将早上和中午的数据作为训练集、夜晚的数据作为验证集。各模型在mAP、FPS性能指标的对比结果如表4所示。由表4可知,FMSA-RD网络较对比模型,具有最高的

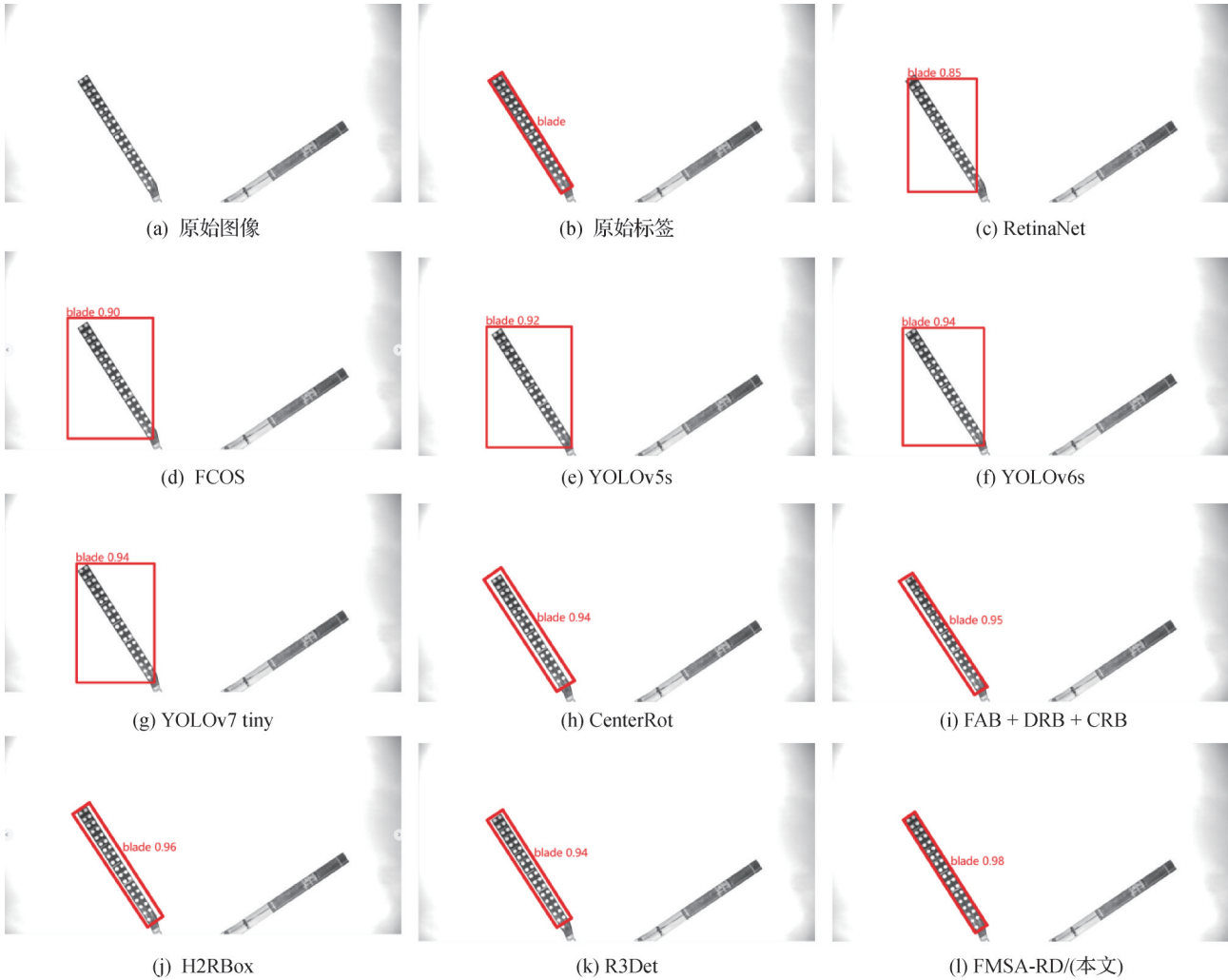


图 12 不同检测算法在自制室外桨叶数据集上的检测结果

Fig. 12 The detection results of different detection algorithms on self-made outdoor blade dataset
((a) original image; (b) original label; (c) Retina Net; (d) FCOS; (e) YOLOv5s; (f) YOLOv6s; (g) YOLOv7 tiny; (h) CenterRot; (i) FAB + DRB + CRB; (j) H2RBox; (k) R3Det; (l) FMSA-RD (ours))

表 4 不同模型在DOTA-v1.0数据集上的对比检测结果

Table 4 Comparative detection results of different models on the DOTA-v1.0 dataset

类型	模型	尺寸/像素	mAP/%	FPS/(帧/s)
H	RetinaNet(Lin 等, 2017)	1 024 × 1 024	67.2	25.4
H	FCOS(Tian 等, 2020)	1 024 × 1 024	73.1	29.5
H	YOLOv5s(Wang 等, 2020)	1 024 × 1 024	68.5	25.6
H	YOLOv6s(Li 等, 2022)	1 024 × 1 024	73.3	28.2
H	YOLOv7 tiny(Wang 等, 2023)	1 024 × 1 024	72.2	30.9
R	CenterRot(Wang 等, 2021)	1 024 × 1 024	73.3	25.4
R	FAB + DRB + CRB(Yuan 等, 2022)	1 024 × 1 024	74.3	24.4
R	H2RBox(Yang 等, 2022)	1 024 × 1 024	73.4	31.6
R	R3Det(Yang 等, 2021)	1 024 × 1 024	72.7	20.4
R	FMSA-RD(本文)	1 024 × 1 024	77.3	37.8

注:加粗字体表示各列最优结果,“H”表示水平目标检测算法,“R”表示旋转目标检测算法。

mAP 和 FPS。如图 13 所示,通过在夜晚数据上进行验证,可以评估出模型在光照条件较暗的环境下也能准确定位桨叶位置。实验结果表明,

FMSA-RD 网络具有较好的检测速度和较高的检测精度,同时具有良好的泛化能力,适用于不同的场景和环境。



图 13 FMSA-RD 方法在自制室外仿真桨叶夜晚验证集上的检测结果

Fig. 13 Detection results of FMSA-RD method on night validation set of self-made outdoor simulated blade dataset

3 结 论

针对室外直升机桨叶无法快速精确检测问题,本文提出融合多特征与自注意力的旋转目标检测网络(FMSA-RD)。在网络中,设计多特征提取 C5 模块,通过提取不同位置的局部特征信息和增添捷径分支,提高桨叶检测速度的同时保证了检测精度;设计多特征融合 SPPFCSP 模块,通过多路分支融合尺度特征,提高桨叶检测精度;设计融合自注意力机制 B3 模块,通过提取全局信息,提高室外复杂光照适应性;引入 SKEWIOU 倾斜交并比,改进位置损失,进一步提高桨叶检测精度。实验结果表明, FMSA-RD 模型对自制室外桨叶数据集的 mAP 为 98.5%,FPS 为 110.5 帧/s,能够快速精确检测到室外

直升机桨叶。

FMSA-RD 网络模型实现了室外单一桨叶定位,但是面对不同桨叶的泛化问题还需要进一步研究。此外,在直升机桨叶图像采集环境有限情况下,如何进一步结合生成式对抗网络,利用已有的少量数据集,生成更多的训练样本也是下一步的研究工作。

参考文献(References)

- Cheng G, Wang J B, Li K, Xie X X, Lang C B, Yao Y Q and Han J W. 2022. Anchor-free oriented proposal generator for object detection. IEEE Transactions on Geoscience and Remote Sensing, 60: #5625411 [DOI: 10.1109/TGRS.2022.3183022]
- Ding J, Xue N, Long Y, Xia G S and Lu Q K. 2019. Learning ROI Transformer for oriented object detection in aerial images//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pat-

- tern Recognition. Long Beach, USA: IEEE: 2844-2853 [DOI: 10.1109/CVPR.2019.00296]
- Du Y L, Liu Q Q, Wang L L, Xu X, Wei Q M and Song W. 2022. Multi-scale rotating anchor mechanism based automatic detection of ocean mesoscale eddy. *Journal of Image and Graphics*, 27(10): 3092-3101 (杜艳玲, 刘倩倩, 王丽丽, 徐鑫, 魏泉苗, 宋巍. 2022. 融合多尺度旋转锚机制的海洋中尺度涡自动检测. *中国图象图形学报*, 27(10): 3092-3101) [DOI: 10.11834/jig.210286]
- Han J M, Ding J, Xue N and Xia G S. 2021. ReDet: a rotation-equivariant detector for aerial object detection//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville, USA: IEEE: 2785-2794 [DOI: 10.1109/CVPR46437.2021.00281]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- Huang J C, Wang H and Lu H. 2022. One-stage detectors combining lightweight backbone and multi-scale fusion. *Journal of Image and Graphics*, 27(12): 3596-3607 (黄健宸, 王晗, 卢昊. 2022. 结合轻量化骨干与多尺度融合的单阶段检测器. *中国图象图形学报*, 27(12): 3596-3607) [DOI: 10.11834/jig.211028]
- Jiang Y Y, Zhu X Y, Wang X B, Yang S L, Li W, Wang H, Fu P and Luo Z B. 2017. R2CNN: rotational region CNN for orientation robust scene text detection [EB/OL]. [2023-09-21]. <https://arxiv.org/pdf/1706.09579.pdf>
- Kim B, Lee J, Lee S, Lee S, Kim D and Kim J. 2022. TricubeNet: 2D kernel-based object representation for weakly-occluded oriented object detection//*Proceedings of 2022 IEEE/CVF Winter Conference on Applications of Computer Vision*. Waikoloa, USA: IEEE: 3421-3430 [DOI: 10.1109/WACV51458.2022.00348]
- Kim G N, Kim S H, Joo I, Kim G B and Yoo K H. 2023. Center deviation measurement of color contact lenses based on a deep learning model and hough circle transform. *Sensors*, 23(14): #6533 [DOI: 10.3390/s23146533]
- Li C Y, Li L L, Jiang H L, Weng K H, Geng Y F, Li L, Ke Z D, Li Q Y, Cheng M, Nie W Q, Li Y D, Zhang B, Liang Y F, Zhou L Y, Xu X M, Chu X X, Wei X M and Wei X L. 2022. YOLOv6: a single-stage object detection framework for industrial applications [EB/OL]. [2023-09-21]. <https://doi.org/10.48550/arXiv.2209.02976>
- Lin T Y, Goyal P, Girshick R, He K M and Dollár P. 2017. Focal loss for dense object detection//*Proceedings of 2017 IEEE International Conference on Computer Vision*. Venice, Italy: IEEE: 2999-3007 [DOI: 10.1109/ICCV.2017.324]
- Ma J Q, Shao W Y, Ye H, Wang L, Wang H, Zheng Y B and Xue X Y. 2018. Arbitrary-oriented scene text detection via rotation proposals. *IEEE Transactions on Multimedia*, 20(11): 3111-3122 [DOI: 10.1109/TMM.2018.2818020]
- Ou Q F, Xiao J B, Chen G F, Li X M and Xiong B S. 2021. Full-scene measurement and analysis of helicopter blade flaps based on vision. *Chinese Journal of Scientific Instrument*, 42(1): 146-156 (欧巧凤, 肖佳兵, 陈焱锋, 李新民, 熊邦书. 2021. 直升机桨叶挥舞量的全场景视觉测量及分析. *仪器仪表学报*, 42(1): 146-156) [DOI: 10.19650/j.cnki.cjsi.J2006688]
- Qian W, Yang X, Peng S L, Zhang X J and Yan J C. 2022. RSDe-t++: point-based modulated loss for more accurate rotated object detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(11): 7869-7879 [DOI: 10.1109/TCSVT.2022.3186070]
- Ren S Q, He K M, Girshick R and Sun J. 2017. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6): 1137-1149 [DOI: 10.1109/TPAMI.2016.2577031]
- Tian Z, Shen C H, Chen H and He T. 2020. FCOS: fully convolutional one-stage object detection//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*. Seoul, Korea (South): IEEE: 9626-9635 [DOI: 10.1109/ICCV.2019.00972]
- Ultralytics. 2020. YOLOv5 [EB/OL]. [2023-09-21]. <https://github.com/ultralytics/yolov5>
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser Ł and Polosukhin L. 2017. Attention is all you need//*Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach, USA: Curran Associates Inc.: 6000-6010
- Wang C Y, Bochkovskiy A and Liao H Y M. 2023. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver, Canada: IEEE: 7464-7475 [DOI: 10.1109/CVPR52729.2023.00721]
- Wang C Y, Liao H Y M, Wu Y H, Chen P Y, Hsieh J W and Yeh I H. 2020. CSPNet: a new backbone that can enhance learning capability of CNN//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. Seattle, USA: IEEE: 1571-1580 [DOI: 10.1109/CVPRW50498.2020.00203]
- Wang J, Yang L and Li F. 2021. Predicting arbitrary-oriented objects as points in remote sensing images. *Remote Sensing*, 13(18): #3731 [DOI: 10.3390/rs13183731]
- Xia G S, Bai X, Ding J, Zhu Z, Belongie S, Luo J B, Datcu M, Pelillo M and Zhang L P. 2018. DOTA: a large-scale dataset for object detection in aerial images//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA: IEEE: 3974-3983 [DOI: 10.1109/CVPR.2018.00418]
- Xie X X, Cheng G, Wang J B, Yao X W and Han J W. 2021. Oriented R-CNN for object detection//*Proceedings of 2021 IEEE/CVF International Conference on Computer Vision*. Montreal, Canada: IEEE: 3500-3509 [DOI: 10.1109/ICCV48922.2021.00350]

- Yang X, Yan J C, Feng Z M and He T. 2021. R3Det: refined single-stage detector with feature refinement for rotating object//Proceedings of the 35th AAAI Conference on Artificial Intelligence. [s.l.]: AAAI: 3163-3171 [DOI: 10.1609/aaai.v35i4.16426]
- Yang X, Zhang G F, Li W T, Zhou Y, Wang X H and Yan J C. 2022. H2RBox: horizontal box annotation is all you need for oriented object detection//Proceedings of the 11th International Conference on Learning Representations. Kigali, Rwanda: ICLR: 1-14
- Yuan Y, Li Z G and Ma D D. 2022. Feature-aligned single-stage rotation object detection with continuous boundary. IEEE Transactions on Geoscience and Remote Sensing, 60: #5538011 [DOI: 10.1109/TGRS.2022.3203983]
- Zheng Z H, Wang P, Liu W, Li J Z, Ye R G and Ren D W. 2020. Distance-IoU loss: faster and better learning for bounding box regression//Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York, USA: AAAI: 12993-13000 [DOI: 10.1609/aaai.v34i07.6999]
- Zhou Q and Yu C H. 2022. Point RCNN: an angle-free framework for rotated object detection. Remote Sensing, 14(11): #2605 [DOI:

10.3390/rs14112605]

- Zuo C L, Wei C H, Ma J, Yue T R, Liu L and Shi Z Y. 2021. Full-field displacement measurements of helicopter rotor blades using stereo-photogrammetry. International Journal of Aerospace Engineering, 2021: #8811601 [DOI: 10.1155/2021/8811601]

作者简介

徐飞龙,男,硕士研究生,主要研究方向为计算机视觉。

E-mail: xufeilong99@126.com

熊邦书,通信作者,男,教授,主要研究方向为图像处理、计算机视觉和直升机故障诊断技术。E-mail: xiongbs@126.com

欧巧凤,女,副教授,主要研究方向为图像处理和计算机视觉。E-mail: ou.qiaofeng@nchu.edu.cn

余磊,男,教授,主要研究方向为图像处理和计算机视觉。

E-mail: yulei@nchu.edu.cn

饶智博,男,讲师,主要研究方向为计算机视觉、深度学习和三维重建。E-mail: raoxi36@foxmail.com