

DOI:10.19650/j.cnki.cjsi.J2413616

面向多无人机协同的多模态目标检测方法

孙晓永, 孙 备, 郭润泽, 党昭洋, 周沛达

(国防科技大学智能科学学院 长沙 410072)

摘要:为解决现阶段单无人机目标检测过程中存在的探测视场有限、目标易被遮挡、单一光源图像信息薄弱等问题,提升无人机可靠、高效的感知计算能力,提出一种面向多无人机协同的多模态目标检测方法。首先,研究可见光和红外融合的多模态目标检测算法,提出了由视觉任务驱动的基于卷积融合网络的双光融合模型,将融合图像经过语义分割网络的结果反馈给融合网络,通过对融合网络参数的迭代训练出损失较小的双光图像融合模型。然后对融合后的图像输入视觉感知增强模块进行图像增强,消除了不良光照条件对图像质量的影响,提升了目标细节特征保持性,并在 MSRS 数据集验证了算法的有效性。此外,面向多无人机协同检测提出了基于分布式生物感知处理的主动感知流程,通过无人机在被感测目标位置计算检测置信度并通过释放信息素来分配主机和从机的检测优先级,完成多无人机协同检测任务的引导策略,实现不同光照条件下非结构化地面场景的目标检测。实验结果表明,该算法在无人机载智能边缘计算平台 RK3588 上具有 56.55 ms 延迟和 45.84 fps 的推理速度,能准确检测地面场景布置的典型军事目标,平均检测精度达到 78.5%。

关键词: 多模态数据; 多无人机协同; 主动感知; 目标检测

中图分类号: TP391.4 TH701 **文献标识码:** A **国家标准学科分类代码:** 510.4050

Multimodal target detection method for multi-UAV coordination

Sun Xiaoyong, Sun Bei, Guo Runze, Dang Zhaoyang, Zhou Peida

(College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410072, China)

Abstract: To solve problems such as limited detection field of view, easy occlusion of target and weak image information of single light source in the current process of single unmanned aerial vehicle (UAV) target detection, and improve the reliable and efficient perception and computing capability of UAV, a multi-modal target detection method for multi-UAV cooperation is proposed in this article. Firstly, a multimodal object detection algorithm based on visible light and infrared fusion is studied, and a dual light fusion model based on a convolutional fusion network driven by visual tasks is proposed. The fused image is fed back to the fusion network through the semantic segmentation network, and a dual light image fusion model with a smaller loss is iteratively trained on the fusion network parameters. Then, the fused image is input into the visual perception enhancement module for image enhancement, eliminating the impact of poor lighting conditions on image quality and improving the preservation of target detail features. The effectiveness of the algorithm is verified on the MSRS dataset. In addition, an active perception process based on distributed biosensing processing is proposed for multi-drone collaborative detection. By calculating the detection confidence of the drone at the location of the sensed target and allocating the detection priority of the host and slave through the release of pheromones, a guidance strategy for multi-drone collaborative detection tasks is completed, achieving target detection in unstructured ground scenes under different lighting conditions. Experimental results show that the algorithm has a 56.55 ms delay and 45.84 fps reasoning speed on the unmanned aerial intelligent edge computing platform RK3588, and can accurately detect typical military targets deployed in ground scenes, with an average detection accuracy of 78.5%.

Keywords: multimodal data; multi-UAV coordination; active perception; object detection

0 引 言

当前,人工智能正逐渐成为战争形态质变和社会科技进步的第一推动力,无人机(unmanned aerial vehicle, UAV)由于机动、灵活的特点成为智能化无人系统的典型代表得到了空前的重视和发展,逐渐在军事和社会生活中发挥着重要的作用^[1]。无人机协同具有高效费比、强抗毁性、低可探测性、功能分布化等方面的显著优势,可以弥补固定摄像头不灵活、检测距离短等缺点,它同时还能大大的减少人工成本,能够支撑在战场侦察、安防、环境监测、森林巡查等军民应用场景广域分布、多点多向的任务。

凭借低成本、零伤亡、操作简单、部署灵活、载荷多变等特点,无人机可在狭小空间起降、长时间空中停留、超低空或贴地飞行,并在城市、山地、丛林、海滩等复杂地形地貌中穿梭,可以在极端恶劣气象条件和极度危险的区域执行任务,实现对目标的精确检测、识别,形成独特的非对称优势,从而大幅提高探测效能。随着环境态势的日趋复杂和多样,目标在无人机拍摄的视频中会出现各种各样的突发情况,例如遮挡、尺度变化、视野进出、光照变化、移动速度变化等,故目前单一平台和单一传感器已无法满足探测需求。而智能多无人机通过搭载多传感器协同执行任务可以避免单架无人机在观测角度、自身安全、探测范围等方面的不足,提供更广泛的目标信息,大幅提高目标搜索与检测的成功率和准确率^[2]。

传统目标检测方法使用的特征信息多数属于浅层特征,对目标内容表达能力较差,传统分类器难以在这些浅层特征的基础上做出有效分类,导致准确率、实时性与泛化能力较弱^[3-5]。随着深度学习理论与技术的飞速发展,基于深度学习的目标检测算法在准确率与实效性上取得了远超传统方法的性能,大致分成基于区域提出的目标检测算法和基于回归的目标检测算法两类。前者典型代表为区域卷积神经网络(region-based convolution neural networks, R-CNN)^[6-7],后者典型代表为YOLO(you only look once)^[8-9]和单发多框检测(single shot multibox detector, SSD)^[10-11]。Ren等^[12]在R-CNN和金字塔池化网络(spatial pyramid pooling-net, SPP-Net)的基础上提出了快速区域生成卷积神经网络(faster region-based convolution neural networks, Faster R-CNN),实现了端到端训练,加快了检测速度。YOLO系列作为一阶检测器的代表作采用基于网格直接回归目标信息,检测速度快,适合实时性任务需求。SSD模型借鉴了Faster R-CNN的锚框和YOLO的回归设计思想,对密集目标和小目标十分友好。

在多无人机协同探测方面,Cai等^[13]结合传统层次

强化学习架构,提出了一种多UAV协作搜索策略,有效地完成了对静态独立目标和聚类目标的搜索任务。郭绍兴^[14]首先将单无人机目标检测框位置和类别置信度通过图像配准进行坐标系转换,然后使用决策层融合和改进YOLOv5实现多无人机目标检测。李洪瑶等^[15]通过多无人机图像配准和拼接并结合动态切换策略实现决策融合的目标检测。关显赫^[16]通过使用基于改进粒子群的多无人机协同救援任务分配方法和强化学习的多无人机协同搜索算法提高了多无人机协同在线搜索动态目标的准确率。Hayat等^[17]将多无人机协同搜索问题构建为博弈模型,提出了一种改进的二进制对数线性学习算法和新的无人机行动选择策略来提高协同控制性能,解决多无人机的最大区域的覆盖问题。史立峰^[18]采用离散粒子群算法和卷积神经网络来实现多无人搜索路径的优化和近空目标的识别。

基于此,首先设计了基于任务驱动的双光图像融合目标检测算法,通过设计的卷积融合网络、语义分割网络和视觉增强模块实现特征增强的双光图像融合,进而提高目标检测精度。然后通过构建分布式仿生集群的多无人机主动感知方法,利用仿生信息素原理设计多无人机协同任务分配方式。最后,将算法部署在多无人机搭载智能边缘计算平台实现了真实场景的地面目标检测,检测精度和推理速度均取得较好效果。

1 基于双光融合的多模态目标检测

随着环境和任务复杂性的不断提高,在远距离的感知过程中,仅通过单一传感器无法感知完整的目标信息,并且复杂环境背景会扰乱有效的视觉信息,从而出现低光照或遮挡导致图像丢失目标的情况。对目标的可见光和红外图像进行融合,可以显著提高整个检测任务的精度和鲁棒性。重点研究针对上述问题的双光图像融合增强,实现多模态目标检测。

双光融合作为一种增强方法,对各种尺寸目标检测均有促进作用。基于双光融合的多模态目标检测方法,总体思路为:利用卷积融合网络和语义分割网络对网络参数进行迭代更新,训练由视觉任务驱动的融合模型,对输入的双光图像进行融合;对输出的融合图像进行视觉感知增强;最后,将得到的增强图像输入目标检测网络得到检测结果。其中视觉感知增强包括照度信息弥补和拉普拉斯锐化。其总体基本结构如图1所示。

1.1 卷积融合网络

卷积融合网络中红外和可见光图像并行输入特征提取层,每个特征提取分支包含一个共同的卷积层和梯度残差密集块(gradient residual dense block, GRDB)。对输入图像执行卷积核大小为 3×3 的卷积操作($3 \text{ conv} \times$

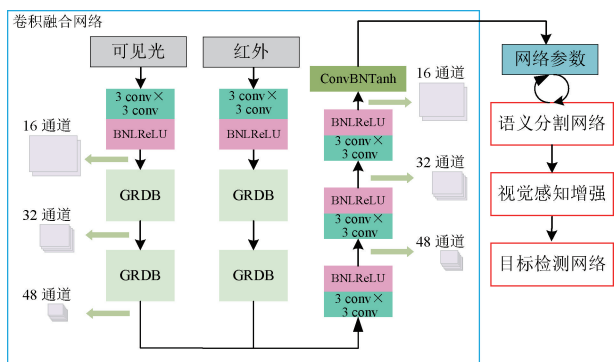


图1 基于双光融合的多模态目标检测流程

Fig. 1 Flowchart of multi-modal object detection based on dual-sensor fusion

3 conv)、批量归一化操作(batch normalization, BN)和采用激活函数(rectified linear unit, ReLU)实现更好地提取浅层特征。使用梯度残差密集块 RDG module 提取深层特征。融合后的特征进入解码器结构,进行特征聚合和图像重建。

GRDB 的具体结构如图 2 所示,通过跳跃连接卷积层提取充分的深层特征,残差流集成梯度操作,使梯度流有效传递,提高特征重用能力,梯度流从浅层特征中提取细粒度特征。最后将多尺度特征融合,兼顾深层语义信息与浅层细粒度特征。

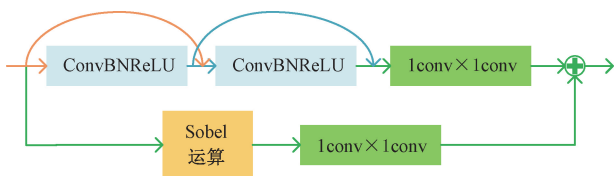


图2 梯度残差密集块结构

Fig. 2 Residual dense gradient module architecture

从图 2 中可以看出,梯度残差密集块的主流采用了两个 3×3 卷积核的卷积层和一个核大小为 1×1 的公共卷积层。其中,在主数据流中引入了跳跃密集的连接,以充分利用由各种卷积层所提取的特征。残差流采用梯度运算来计算特征的梯度幅度,并采用 1×1 的卷积层来消除通道的尺寸差异。然后,通过元素级叠加主要密集流和残差梯度流的输出,集成深度特征和细粒度细节特征。

在提取浅层特征时,使用 Sobel 梯度算子^[19]进行边缘检测,找出图像中亮度变化明显的转折点,区分边缘和噪声的高频带,计算梯度信息和方向,最后根据梯度信息进行检测边界。

1.2 语义分割网络

语义分割网络可采用双边注意机制来构建语义损失,该网络的输入为融合图像,将此融合图像的语义分割结果

反馈给融合网络,指导融合网络参数的迭代更新,最终训练出损失较小的双光图像融合模型,结构如图 3 所示。

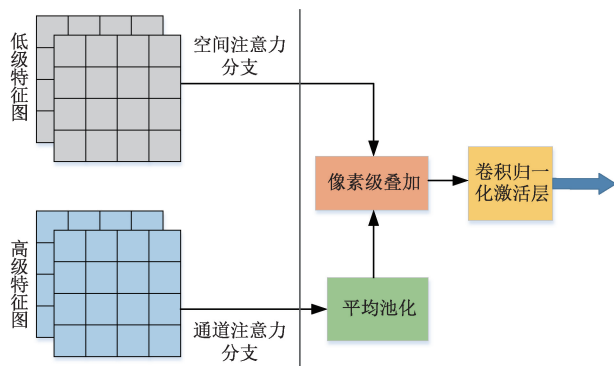


图3 语义分割网络结构

Fig. 3 Semantic segmentation network architecture

语义分割网络包括空间注意力分支、通道注意力分支以及池化融合单元(虚线右侧部分)。空间注意力分支从融合图像提取的低级特征图可以捕获更准确的空间信息,通道注意力分支从融合图像提取的高级特征图可以捕获更精确的语义信息,池化融合单元首先使用 3×3 的平均池化层来平滑上采样的通道注意力特征,再与空间注意力特征进行像素级相加,最后通过卷积归一化激活层(Conv+BN+ReLU)得到融合图像的语义分割结果,实现融合提取的高级和低级特征映射。

空间注意力分支将融合图像首先通过卷积归一化激活层(Conv+BN+ReLU,用于减少特征通道数)后再分别通过最大池化、平均池化操作,然后将最大池化和平均池化操作得到的特征求和后(使图像的空间细节特征得到充分的提取)通过 Conv+BN+Sigmoid 层获得的特征与原始的融合图像相加后作为空间注意力分支的输出特征,如图 4 所示。

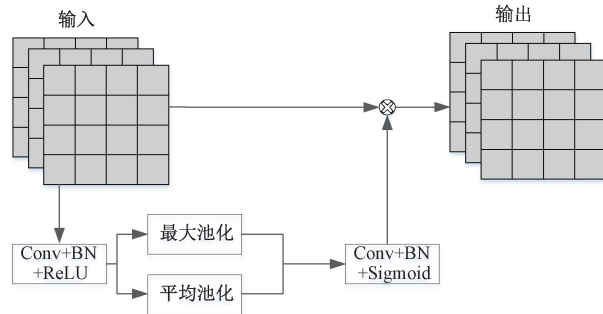


图4 空间注意力分支结构

Fig. 4 Spatial attention branch architecture

通道注意力分支首先使用全局平均池化将输入特征图降维,再连入 Conv+BN+Sigmoid 层将获得的特征与原始的融合图像相加后作为通道注意力分支的输出特征,如图 5 所示。

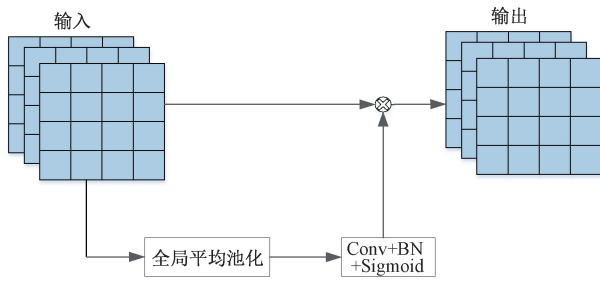


图5 通道注意力分支结构

Fig. 5 Channel attention branch architecture

双光融合图像经过语义分割后的图像如图6所示。从图中可以看出语义分割的准确性很高,对于图像中的小目标也可以精准分割,证明该语义分割方法对小目标分割的有效性。

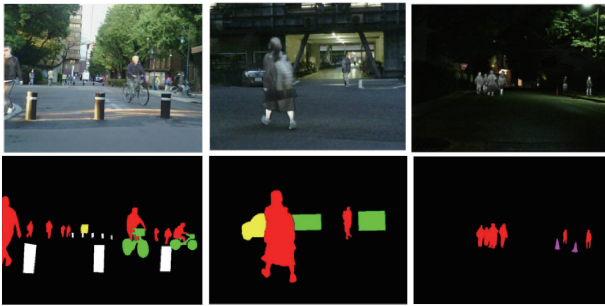


图6 融合图像语义分割结果

Fig. 6 Semantic segmentation results of fused images

1.3 损失函数

假设配准的一对可见光与红外图像,分别为 I_k, I_i , 融合图像为 I_f 。融合图像的质量很大一部分取决于损失函数。本融合网络设计了基于内容损失和语义损失的联合损失函数,利用特征提取模块 E_F , 从红外和可见光图像中提取具有丰富细粒度细节信息的深度特征,可以表示为:

$$\{F_{ki}, F_{hi}\} = \{E_F(I_{ki}), E_F(I_{hi})\} \quad (1)$$

经过梯度残差密集模块的特征表示为式(2),其中 F^i 为输入的特征。

$$F^{i+1} = \text{Conv}^n(F^i) \oplus \text{Conv}(\nabla F^i) \quad (2)$$

为了较好地融合红外图像中的突出目标和可见光图像中的纹理细节,设计内容损失确保融合图像的视觉保真度。设内容损失为 L_{con} , 包括了融合图像的整体强度损失 L_{int} 和纹理细节损失 L_{text} 。强度损失在像素级别衡量融合图像与原图像间的差异,纹理损失表示在细粒度特征方面的差异,内容损失函数为:

$$L_{con} = L_{int} + \alpha L_{text} \quad (3)$$

定义内容损失分量的强度损失为:

$$L_{int} = \frac{1}{HW} \|I_f - \max(I_{ki}, I_{hi})\|_1 \quad (4)$$

式中: H 和 W 为融合图像的高和宽; $\|\cdot\|_1$ 代表 l_1 范数; $\max(\cdot)$ 表示元素级的最大选择。

定义纹理损失如式(5)所示,它计算了梯度值之差的比较融合图像与原图像之间的纹理变化。

$$L_{text} = \frac{1}{HW} \| |\nabla I_f| - \max(|\nabla I_{ki}|, |\nabla I_{hi}|) \|_1 \quad (5)$$

α 参数起到调节两种损失的平衡,使强度损失和纹理损失限制在一定范围内的作用。基于梯度残差密集块的融合网络可以在内容损失的指导下,实现较好的强度分布,并保持丰富的细节信息。内容损失可以有效地提高融合图像的视觉质量和统计评价指标。

语义损失来自于语义分割网络,表达式为:

$$L_{sem} = \frac{-1}{HW} \sum_{h=1}^H \sum_{w=1}^W \sum_{c=1}^C L_{so}^{(h,w,c)} \log(I_s^{(h,w,c)}) \quad (6)$$

式中: H 和 W 为融合图像的高和宽; C 为融合图像的通道; L_{so} 表示分割标签; I_s 表示分割网络输出的语义分割结果。

最后得到的联合损失函数为:

$$L_{joint} = L_{con} + \beta L_{sem} \quad (7)$$

语义损失用来约束融合网络的方向,同时还可以指导分割网络训练。 β 用来调节语义损失的占比情况,根据融合网络训练进行调整。

1.4 视觉感知增强模块

对于图像中的小目标,视觉感知增强模块将图像恢复属性颜色,大大消除了不良光照条件对图像质量的影响,使难以检测到的小目标显示出来。同时进行拉普拉斯锐化,对图像中的小目标边缘细节进行增强,从而使检测网络提取到更丰富的浅层特征。

视觉感知增强模块首先从图像金字塔下采样得到的最小图像开始计算,比较每个像素与周围其余8个方向相邻的像素,采用高斯中心函数估算图像反射分量 R ;依次往下进行,在上一层计算完成后,对估计的图像反射分量进行插值,使得插值后的反射率估算图像与下一层图像大小,再对下一层进行相同的比较操作;最后完成原图像的相邻像素比较后得到最终结果,得到了去照度的增强图像。在该计算过程中假设 $S_1, S_2, \dots, S_m$ 是路径上的点,从远到近排列,其中 r_c 表示中心位置的最终反射率估计。

计算步骤为:

1) 将原始图像变换到对数域 $S(x, y)$, 获取金字塔层数;

2) 初始化常值图像矩阵 $R_0(x, y)$, 作为迭代计算的图像初始值;

3) 从最顶层到最后一层进行八邻域像素比较操作,根据 Retinex 理论^[20-21] 路径上的像素按照式(8)和(9)计

算反射率 $\mathbf{R}$ 。在第 n 层的操作完成后,将第 n 层的操作结果 $\mathbf{R}$ 插值为原来的两倍,这与第 $n+1$ 层的大小相同。当计算最底层时,得到的 $\mathbf{R}$ 是最终的去照度增强图像。

$$\mathbf{S}_m = r_m + l_m \quad (8)$$

$$\mathbf{R} = \mathbf{R}_0 + \frac{r_c - r_m}{2} + \frac{r_c - r_{m-1}}{4} + \dots + \frac{r_c - r_1}{2^m} \quad (9)$$

其中, r_m 为图像原始反射分量对数值, l_m 表示图像入射照度分量对数值。

针对去照度后的融合图像,进行锐化以补偿图像的轮廓、增强图像的边缘及灰度跳变的部分,提高图像清晰度和对比度。通过比较邻域内中心像素灰度与其他像素的平均灰度,当中心像素灰度低于其他像素的平均灰度时,将增强后的融合图像的模板卷积结果加在原像素上以进一步降低中心像素灰度;当中心像素灰度高于其他像素的平均灰度时,将增强后的融合图像的模板卷积结果加在原像素上以进一步提高中心像素灰度。

1.5 评价指标

为了验证提出算法的有效性,使用熵(entropy, EN)、平均梯度(average gradient, AG)、视觉信息保真度(visual information fidelity, VIF)、平均检测精度(mean average precision, mAP)、图像属性基础评价指标均值(Q)这5个指标对算法的性能进行评估。

$VIF = \frac{I(C; F|s)}{I(C; E|s)}$ 代表图像所包含的总体平均信息量,通过统计学定义计算图像中信息含量,定义为:

$$EN = - \sum_{i=0}^{L-1} P_i \log_2 P_i \quad (10)$$

其中, L 为灰度级数, P_i 为融合图像中对应灰度级的归一化直方图。

AG 能反映图像中细节纹理,能够反映了图像的清晰程度,定义为:

$$AG = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \sqrt{\frac{\nabla F_x^2(i,j) + \nabla F_y^2(i,j)}{2}} \quad (11)$$

其中, $\mathbf{F}$ 为待评价的图像, $M \times N$ 为尺寸大小, $\mathbf{F}(i,j)$ 为第 i 行、第 j 列的像素。 $\nabla F_x(i,j) = \mathbf{F}(i,j) - \mathbf{F}(i+1,j)$, $\nabla F_y(i,j) = \mathbf{F}(i,j) - \mathbf{F}(i,j+1)$ 。

VIF 为量化融合图像 $\mathbf{F}$ 与源图像之间共享的信息量的指标,定义为:

$$VIF = \frac{I(C; F|s)}{I(C; E|s)} \quad (12)$$

其中, $\mathbf{C}$ 为原始图像信息, $\mathbf{E}$ 为人眼获得的原始图像信息, $\mathbf{F}$ 为人眼获得的测试图像信息, $\mathbf{I}$ 为互信息操作。

mAP 则是将所有类别检测的平均正确率进行综合加权平均而得到的,表示全类的平均正确率,该值由 YOLOv5 算法计算得出。

图像属性基础评价指标均值定义为:

$$Q = \frac{1}{3}(EN + AG + VIF) \quad (13)$$

1.6 检测效果

本节首先在 MSRS 数据集(具备白天和夜晚不同光照可见光和红外数据)上进行训练和测试来检验算法性能,结果如表 1 所示。

表 1 不同算法的评价指标对比

Table 1 Comparison of evaluation indicators for different algorithms

算法	EN	AG	VIF	mAP@0.5	Q
NSCT-SR	5.684	0.080	0.935	0.655	2.233
Densefuse	5.022	0.042	0.705	0.706	0.923
IFCNN	5.450	0.081	0.931	0.635	2.154
本算法	5.698	0.079	0.940	0.750	2.239

算法基本参数如下:输入融合模型的双光图像的大小为 640×480 ,卷积融合网络训练 2 700 代,每次迭代处理 4 张图像,初始学习率设为 0.001,采用 Adam 优化器;视觉增强模块使用 Retinex 去照度算法,采用尺度因子大小为 $[15, 101, 301]$,拉普拉斯锐化的模板为 $([0, -1, 0], [-1, 5, -1], [0, -1, 0])$ 。实验的硬件配置环境为配备 Intel i7-7800X @ 3.5 GHz CPU 和 NVIDIA GeForce GTX 1080 Ti GPU 的计算机用于训练和测试。

分别采用了 NSCT-SR 方法^[22]、Densefuse 方法^[23]、IFCNN 方法^[24]与提出的融合方法进行对比。从表格可以看出,该算法在 5 种指标中有 4 项指标均高于其他算法,尤其是在检测精度方面具有较大的优势。但是该算法在 AG 指标落后 IFCNN 算法高于其他 3 个算法,也仅落后 0.002。结果表明提出的算法可以更好地克服低光条件下的不利影响,提高融合图像的信息丰富度,提高目标检测能力。

通过消融实验评价视觉感知增强模块的有效性,从量化指标角度评价视觉感知增强模块的作用。通过将原始可见光图像和经过融合增强后的图像进行对比,设置未添加和添加视觉感知增强模块进行消融实验,并使用目标检测网络 YOLOv5 对融合图像进行检测,比较两者检测精度的区别,结果如表 2 所示(P 为精度, R 为召回率)。从表中可以看出,添加双光融合增强模块和视觉感知增强模块的模型在目标检测的精度明显大幅提升。

2 基于生物感知处理的多无人机协同方法

由于检测任务中目标预先信息的不确定性、目标潜在的机动性,以及无人机性能约束,仅依靠被动感知,容易造成目标难以检测或目标丢失,导致任务的失败。

表2 不同增强算法的检测评价指标对比
Table 2 Comparison of detection and evaluation indicators for different enhancement algorithms

图像类型	增强方法	<i>P</i>	<i>R</i>	<i>mAP@0.5</i>
原始可见光图像	无	0.66	0.174	0.209
融合增强图像	双光融合增强	0.786	0.795	0.76
	双光融合增强+视觉感知增强	0.955	0.896	0.97

基于此,设计了多无人机协同方法,旨在利用多个无人机探测单元实现对未知环境未知目标的协同检测。其中,负责广域搜索、引导探测、覆盖扫描的定义为主机,并记为主机 UAV1,其飞行高度高,拍摄范围更大,目标更容易出现在其视野中;负责低空侦察、辅助辨识、精细检测的定义为从机,并记为从机 UAV2 和从机 UAV3,其飞行高度略低,可以拍摄到更为清晰的地面图像,便于进行更准确的检测。

基于此,考虑到环境和目标运动特性对无人机群体系统感知性能的影响,提出基于分布式仿生集群算法建立无人机集群分布式模型,流程如图7所示。

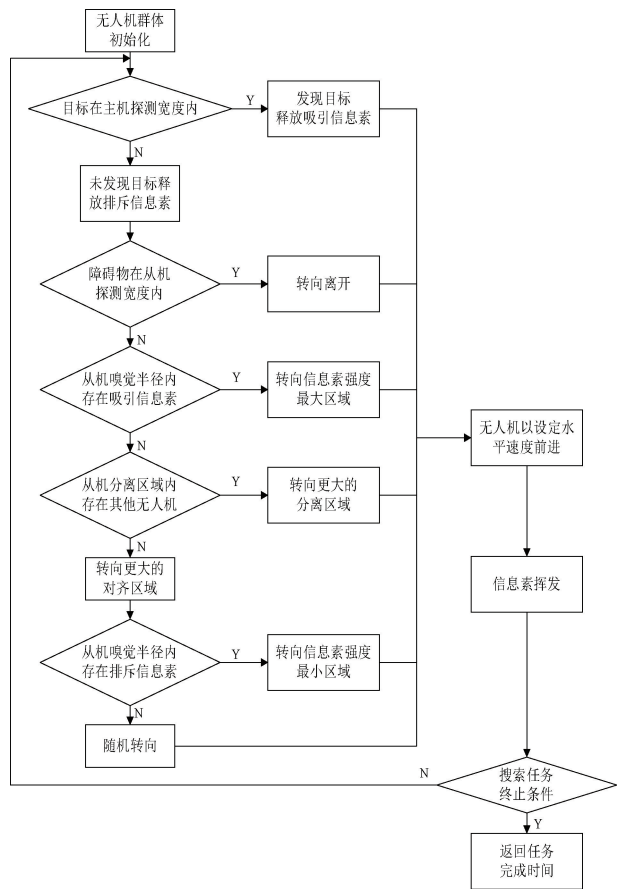


图7 基于分布式仿生集群算法的多无人机协同方法
Fig.7 Multi-UAV collaborative method based on distributed biomimetic cluster algorithm

具体来说,当主机的物理传感器对目标检测时,主机会在被感测目标的位置上释放一种特别数量有吸引力的信息素,信息素进行扩散并对邻近的从机具有吸引力。当被检测目标的置信度较高时,信息素的数量趋近于0,此时仅需主机即实现对目标的检测;当被测目标的置信度较低时,传递信息素的数量增加,此时需要引导从机来对目标进行进一步检测。一段时间后,从机的识别置信度高于阈值,主机的信息素强度无法增强就会因挥发而消失。此外,主机在没有感觉到目标的地方释放出排斥的信息素,从机也不会在该区域进行摸排,当在信息素饱和的地方移动时,无人机无法感觉到信息素。任务开始后,UAV1 起飞,执行广域搜索任务,当 UAV1 发现并检测目标时(置信度>0.5),则认为主机感知有效,无需释放信息素;当 UAV1 无法准确感知目标时(置信度位于0.1~0.5之间),UAV1 在疑似目标区域(如目标1所在位置)释放信息素,主动将目标位置结果通过分发给 UAV2 和 UAV3,随后引导 UAV2 和 UAV3 对目标进行辨识,并将检测结果回传,从而实现多无人机的主动感知和精细检测。

3 实验结果与分析

为证明所提出的算法用于典型无人机对地视角目标检测的性能,采用自己构建的地面场景不同目标对算法进行训练和测试,验证所提算法具有良好的检测性能。

3.1 实验环境

选取的实验场地位于湖南省常德市清水湖实验外场,场地内具有草地、丛林、水泥路、建筑物、湖泊等环境设施。在草丛中随机部署多个目标(包括坦克模型、海马斯模型、狙击手模型、迫击炮模型等,平视场景几乎无法观测到),如图8所示。

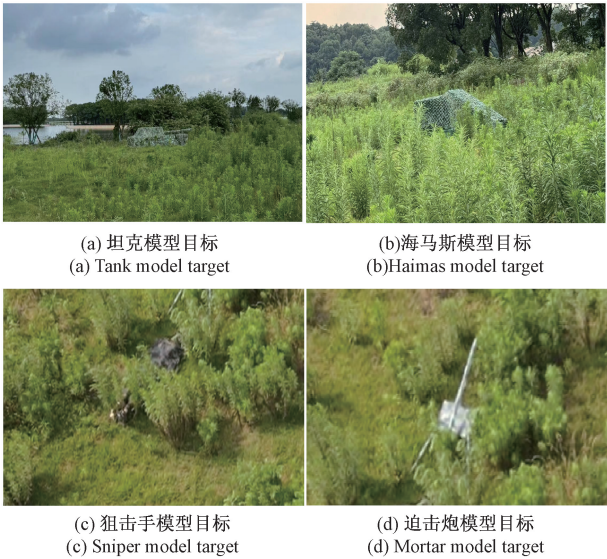


图8 实验场地典型目标布设
Fig.8 Typical target arrangement at the experimental area

多无人机通过预先设定位置从 3 个视角对实验场地中目标进行探测,示意图如图 9 所示。

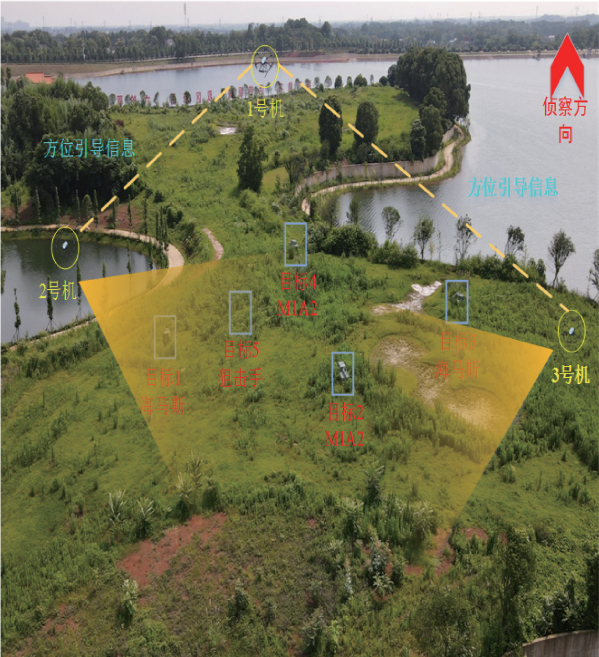


图 9 无人机布放及目标真实位置

Fig.9 Schematic of drone deployment and actual target location

3.2 实验平台

实验中搭建了如图 10 所示的无人机平台,用于典型地面目标数据的采集及算法的部署验证。无人机平台搭载了光电吊舱(具有可见光和红外传感器)用于获取地

面光电图像的实时采集以实现地面环境的感知。此外,无人机还搭载了惯导模块及全球定位系统(global positioning system,GPS)获取无人机自身的姿态和位置信息,然后通过边缘计算平台在机上进行在线的目标检测并使用通信电台模块建立与地面端之间的通信,实现数据的传输和信息的交互。



图 10 无人机平台
Fig.10 UAV platform

3.3 多无人机协同探测实验

1)大视场可见光引导红外精细辨识

光照较好条件下,引导算法自适应选择可见光引导红外的方式进行主动感知。图 11 所示流程主要用来测试无人机多模协同对某一广域范围区域内目标检测能力,UAV1 利用可见光传感器大视场侦察区域目标,发现目标后,通过方位引导 UAV2 和 UAV3 利用红外传感器小视场进行目标检测。

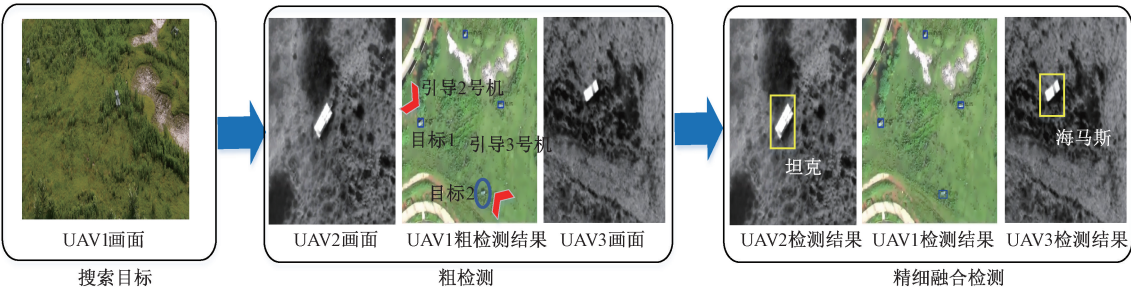


图 11 大视场可见光引导红外目标检测流程

Fig.11 Target detection flowchart of large field visible light guided infrared

详细步骤为:

- (1)任务开始后,3 架无人机巡飞至目标区域,其中 UAV1 在目标区域选择可见光传感器进行大视场搜索。此时,由于主机的飞行高度较高且变焦倍数较小,目标尺寸通常较小,特征不明显;
- (2)当 UAV1 发现目标后,执行单机检测算法,首先进行目标的粗检测,识别结果如图 11 中第 2 列“粗检测”

- 所示,图中显示两个目标已经被检测到,目标 1 和 2 未检测到。随后,UAV1 主动选择低置信度目标:目标 1(置信度为 35%),目标 2(置信度低于 30%),将空间位置发送给 UAV2 和 UAV3,引导从机的红外传感器进行精细识别;
- (3)UAV2 和 UAV3 接受目标位置信息,转动吊舱,使低置信度目标位于画面中央,得到目标精细识别结果。

目标 1 被识别为坦克,置信度为 83%,目标 2 被识别为海马斯,置信度为 67%。

通过与实际布置的目标类别、位置进行对比,UAV1 传输给 UAV2 和 UAV3 的方位信息准确,且 3 个无人机节点通过协同的主动感知能够给出场地范围内的所有设定目标的检测结果。

2)大视场红外引导可见光精细辨识

在恶劣天气或弱光照条件时,红外图像仍然能够保

持一定的清晰度和对比度,显示出目标和背景的辐射差异,但是无法提供目标详细的外观信息。可见光图像能提供一定的轮廓信息,有效弥补红外图像分辨率低的不足。因此,弱光照条件下,图 12 为自适应选择红外引导可见光的方式进行主动感知流程,UAV1 利用红外传感器大视场侦察区域目标,发现目标后,通过方位引导 UAV2 和 UAV3 利用可见光传感器小视场进行目标检测。

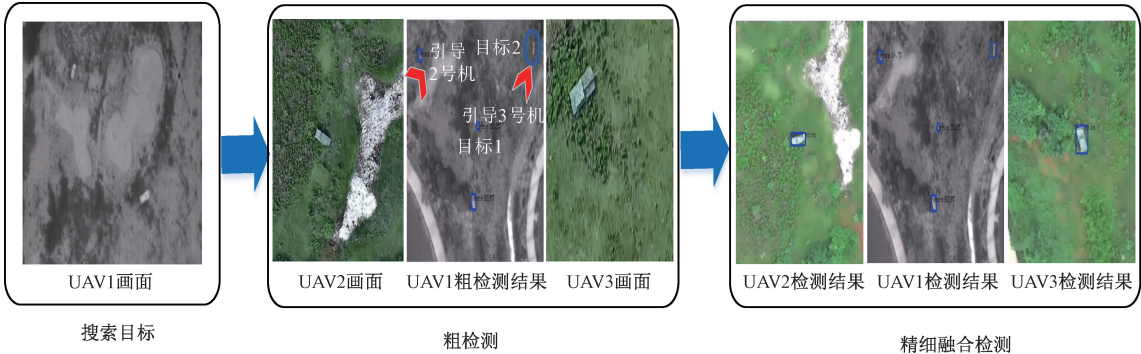


图 12 大视场红外引导可见光目标检测流程

Fig. 12 Target detection flowchart of large field infrared guided visible light

详细步骤为:

(1)任务开始后,3 架无人机巡飞至目标区域,其中 UAV1 在目标区域选择红外传感器进行大视场搜索。此时,由于主机的飞行高度较高且红外传感器变焦倍数较小,目标尺寸通常较小,特征不明显;

(2)当 UAV1 发现目标后,执行单机检测算法,首先进行目标的粗识别,识别结果如图 12 中第 2 列“粗检测”所示,图中显示两个目标已经被检测到,目标 1 和 2 未检测到。随后,UAV1 主动选择低置信度目标:目标 1(置信度为 39%),目标 2(置信度低于 30%),解算其空间位置,并发送给 UAV2 和 UAV3,引导从机的可见光传感器进行精细识别;

(3)UAV2 和 UAV3 接受目标位置信息,转动吊舱,使低置信度目标位于画面中央,得到目标精细检测结果。目标 1 被识别海马斯,目标 2 被识别为坦克,置信度均在 80% 以上。

3)静态目标检测效果

表 3 记录了 4 组不同时间、不同目标类型在静态条件下的试验数据,统计了主动感知算法应用前后静态目标检测效果的变化和检测置信度。在“是否使用主动感知算法”一列,“√”表示使用主动感知算法,“×”表示未使用主动感知算法。在“目标检测效果”一列,“√”表示检测到目标,“×”表示未检测到目标。由表 3 可知,在使用多无人机协同主动感知算法后,目标检测效果,检测置信度有了明显提升。

表 3 多无人机协同主动感知算法使用前后静态目标检测效果对比

Table 3 Comparison of detection performance before and after the use of multi UAV collaborative active perception algorithm

序号	目标类型	时间	是否使用主动感知算法	目标检测效果	检测置信度/%
1	海马斯	上午	×	√	39
	海马斯	上午	√	√	87
2	坦克	下午	×	×	
	坦克	下午	√	√	83
3	狙击手	上午	×	√	32
	狙击手	上午	√	√	85
4	迫击炮	下午	×	×	
	迫击炮	下午	√	√	80

图 13 为检测算法在不同光照条件下的融合效果,并将融合时可见光模态的权重打印在图像下方,从左至右分别为早晨、正午、傍晚 3 种不同场景。由图 13 可知,正午的光照条件最少,图像中所含的热辐射信息最少,因此,图像整体的背景色与可见光图像的背景最为接近。相比于中午,早晨的光照条件稍弱,因此,红外权重占比稍大,目标具备一定的红外热辐射信息。傍晚的光照条件最弱,因此,融合后的图像中红外权重占比最大,整体

图像也更接近灰度图像,目标与背景也被明显区分。经过计算,在早晨,红外权重的占比约在 50%~60% 左右。在正午,红外权重的占比约在 10%~30% 左右,在傍晚,红外权重的占比约在 80%~90% 左右。这说明本算法在不同光照条件下有着良好的自适应能力。

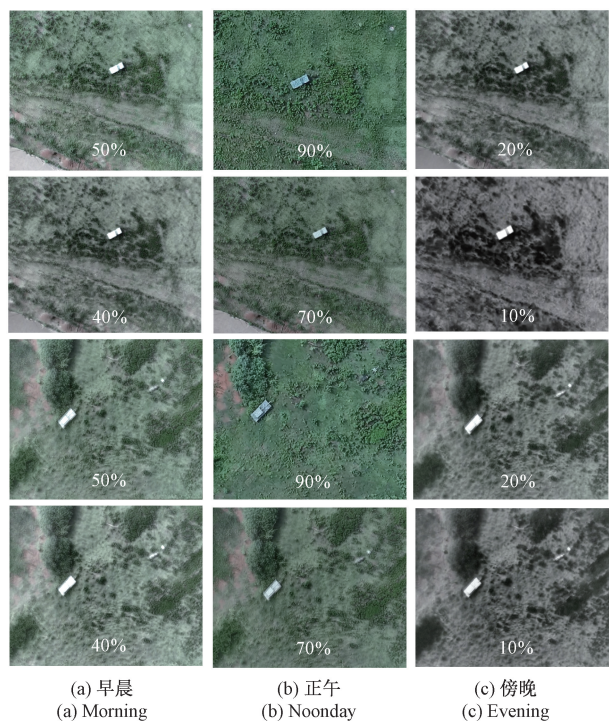


图 13 不同光照条件下目标融合效果

Fig. 13 Target fusion effect under different lighting conditions

4) 动态目标检测效果

无人机飞行高度较高,虽然增大了信息获取能力,但也会将更多建筑物、树木等背景元素纳入成像平面中,使得周围环境更加复杂。在这种复杂动态环境中,目标与目标之间存在相互遮挡,背景元素同样会遮挡目标,导致检测失败。图 14 记录了在遮挡等特殊情况下的动态目标检测效果。



图 14 动态目标检测效果

Fig. 14 The dynamic object detection performance of the algorithm in this article

由无人机 1 画面中圆框可知,所提出的目标检测算

法在遮挡等动态场景下都可以实现正确的目标检测和匹配,能够有效说明算法对于复杂场景下的目标检测的有效性和鲁棒性。此外,计算了经纬度方向上动态目标的定位误差来衡量目标检测的误差,得到如图 15 所示的目标变速运动情况下以及目标快速转弯情况下移动目标位置估值误差,其中上面线条为单探测单元计算得到的位置估值与真值的距离,下面线条为多探测单元计算得到的位置估值与真值的距离。由图 15 可知,所提算法大大减少了动态条件下的目标检测误差。

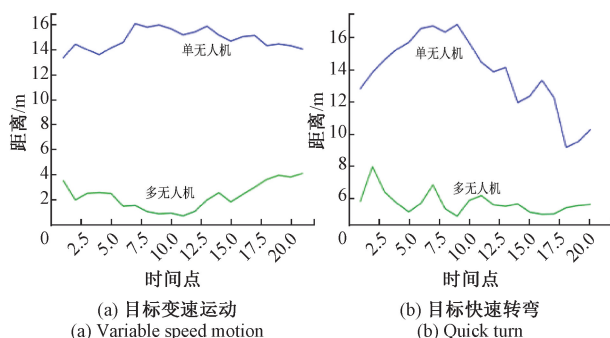


图 15 动态目标位置估计误差对比

Fig. 15 Comparison of dynamic target position estimation errors

5) 无人机多模目标检测

准备 3 个无人机节点,分别搭载边缘计算平台、任务调度板卡模块、多模光电吊舱、自组网电台和无人机姿态及定位设备,地面端布置地面端电台,连接显控电脑,用于显示无人机状态数据和融合感知处理结果,测试机载端与地面端的控制与通信链路系统是否正常。

如图 16 所示,为实验外场单无人机(第 1 列)和多无人机(后 3 列)协同目标检测结果。可以看出,单无人机检测结果出现漏检(图中椭圆位置),而通过多无人机主动感知可以检测到漏检目标,平均准确率可以达到 75.8%,证明了本方法的有效性。

6) 推理性能分析

推理性能是多无人机协同探测系统实际部署和应用的关键,提出的算法在 RK3588 边缘计算平台上进行网络模型的训练和测试,平台实物图如图 17 所示。在计算推理性能时,忽略数据准备的时间,仅考虑算法处理图像的时间。表 4 给出了输入为 3 路无人机视频流情况下本算法和其他框架在 RK3588 上的性能对比。本算法在使用同一 CUDA Context 下多 CUDA Stream 并行模式时性能最好,吞吐率为 45.84 fps,延迟为 56.55 ms,相比串行的 Darknet 吞吐率提高 4.48 倍,相比串行的 TensorRT 吞吐率提高 1.14 倍。

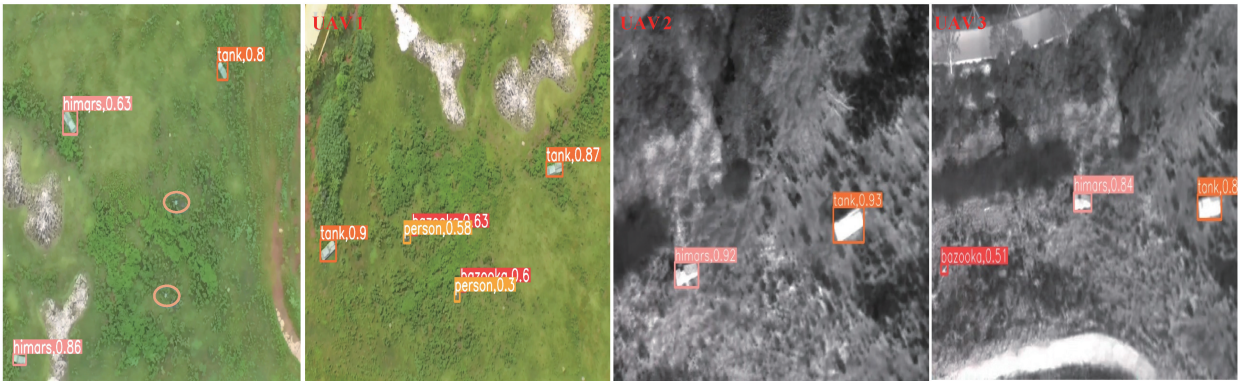


图 16 单无人机与多无人机协同目标检测结果

Fig. 16 Single UAV and multi UAV collaborative target detection results

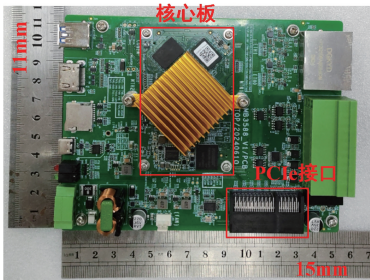


图 17 RK3588 图像处理平台实物

Fig. 17 Picture of the RK3588 image processing platform

表 4 所提方法和其他框架在 RK3588 上的性能对比

Table 4 Performance comparison between the method proposed in this article and other frameworks on RK3588

框架	策略	延迟/ms	吞吐量/fps
Darknet	串行	106. 16	10. 23
TensorRT	串行	37. 89	40. 26
本算法	3 线程并行	56. 55	45. 84

4 结 论

本研究对面向多无人机协同的多模态目标检测方法进行了研究。首先,研究了利用可见光和红外融合的基于卷积融合网络和视觉感知增强的任务驱动型目标检测方法,提升目标细节特征保持性,并在 MSRS 数据集进行了验证。此外,设计了基于分布式仿生集群的多无人机协同主动感知流程,充分利用多个无人机探测单元实现对未知环境未知目标的协同检测。实验结果表明,提出的多无人机对地目标检测算法平均精度达到 75. 8%,在智能边缘计算平台上具有 56. 55 ms 延迟和 45. 84 fps 的推理速度,具有一定的鲁棒性和实时性。提出的算法在

典型地面场景下能有效的对地面布设目标进行探测,未来的研究应该更加聚焦在复杂背景、恶劣天气和任务多样化下的目标检测、更加关注提升边缘计算平台的计算效率和多机协同检测精度、提升模型网络的鲁棒性和泛化性等。

参考文献

[1] 马捷,李雅. 人工智能在无人机领域的应用[J]. 无线电工程,2024,54(3):759-764.
MA J, LI Y. Application of artificial intelligence in the field of UAV[J]. Radio Engineering, 2024,54(3):759-764.
[2] 贾永楠,田似营,李擎. 无人机集群研究进展综述[J]. 航空学报,2020,41(S1):4-14.
JIA Y N, TIAN SH Y, LI Q. A review of research progress on drone swarms [J]. Acta Aeronautica et Astronautica Sinica, 2020,41(S1):4-14.
[3] VIOLA P, JONES M J. Robust real-time face detection[J]. International Journal of Computer Vision, 2004(57):137-154.
[4] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[J]. Computer Vision and Pattern Recognition, 2005:886-893.
[5] FELZENSZWALB P F, GIRSHICK P B, MCALLESTER D, et al. Object detection with discriminatively trained part-based models [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(9):1627-1645.
[6] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [J]. 2014 IEEE Conference on

- Computer Vision and Pattern Recognition, 2014; 580-587.
- [7] 伍锡如,邱涛涛,王耀南. 改进 Mask R-CNN 的交通场景多目标快速检测与分割[J]. 仪器仪表学报, 2021, 7(42): 242-249.
- WU X R, QIU T T, WANG Y N. Multi-object detection and segmentation for traffic scene based on improved Mask R-CNN[J]. Chinese Journal of Scientific Instrument, 2021, 7(42): 242-249.
- [8] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [9] 范云生,张凯,牛龙辉,等. 海上无人机对运动船舶的长期检测跟踪算法[J]. 仪器仪表学报, 2024, 45(3): 326-335.
- FAN Y SH, ZHANG K, NIU L H, et al. Long-term detection and tracking algorithm for moving vessels by maritime UAVs [J]. Chinese Journal of Scientific Instrument, 2024, 45(3): 326-335.
- [10] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. Computer Vision-ECCV 2016, 2016: 21-37.
- [11] 孙备,左震,吴鹏,等. 面向无人艇环境感知的改进型 SSD 目标检测方法[J]. 仪器仪表学报, 2021, 9(42): 52-61.
- SUN B, ZUO ZH, WU P, et al. Object detection for environment perception of unmanned surface vehicles based on the improved SSD [J]. Chinese Journal of Scientific Instrument, 2021, 9(42): 52-61.
- [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 & Machine Intelligence, 2017, 39(6): 1137-1149.
- [13] CAI Y F, YANG S X, XU X. A combined hierarchical reinforcement learning based approach for multi-robot cooperative target searching in complex unknown environments [J]. Control and Intelligent Systems, 2013, 41(4): 218-230.
- [14] 郭绍兴. 基于改进 YOLOv5 算法的多无人机协同目标检测[D]. 哈尔滨: 哈尔滨工业大学, 2023.
- GUO SH X. Collaborative target detection of multiple unmanned aerial vehicles based on improved YOLOv5 algorithm[D]. Harbin: Harbin Institute of Technology, 2023.
- [15] 李洪瑶,李小强,韩心中,等. 基于决策融合的多无人机协同目标检测识别算法[J]. 系统工程与电子技术, 2022, 44(3): 746-754.
- LI H Y, LI X Q, HAN X ZH, et al. Cooperative object detection and recognition algorithm for multiple UAVs based on decision fusion [J]. Systems Engineering and Electronic, 2022, 44(3): 746-754.
- [16] 关显赫. 基于强化学习的多无人机协同区域搜救技术研究[D]. 大连: 大连海事大学, 2021.
- GUAN X H. The research of cooperative area search and rescue technology for multi-UAVs based on reinforcement learning [D]. Dalian: Dalian Maritime University, 2021.
- [17] HAYAT S, YANMAZ E, BETTSTETTER C, et al. Multi-objective drone path planning for search and rescue with quality-of-service requirements [J]. Autonomous Robots, 2020, 44(7): 1183-1198.
- [18] 史立峰. 基于多无人机协同的近空目标识别与跟踪技术研究[D]. 上海: 东华大学, 2019.
- SHI L F. Research of near-space target recognition and tracking technology based on multi-UAV cooperation [D]. Shanghai: Donghua University, 2019.
- [19] YAO G X. Design of edge detection algorithm for image sobel based on FPGA [C]. 2015 4th International Conference on Computer Science and Network Technology, 2015: 851-853.
- [20] TANG L F, YUAN J T, MA J Y. Image fusion in the loop of high-level vision tasks: A semantic-aware real-time infrared and visible image fusion network [J]. Information Fusion, 2022, 82: 28-42.
- [21] 尚砚娜,石晶欣,赵岩,等. 大型结构体裂缝检测中的定位方法[J]. 仪器仪表学报, 2017, 38(3): 681-688.
- SHANG Y N, SHI J X, ZHAO Y, et al. Novel positioning method for detecting large structure surface cracks [J]. Chinese Journal of Scientific Instrument, 2017, 38(3): 681-688.
- [22] LIU Y, LIU SH P, WANG Z F. A general framework for image fusion based on multi-scale transform and sparse

representation[J]. Information Fusion, 2015(24):147-164.

[23] LI H, WU X J. DenseFuse: A fusion approach to infrared and visible images[J]. IEEE Transactions on Image Processing, 2019, 28(5):2614-2623.

[24] ZHANG Y, LIU Y, SUN P, et al. IFCNN: A general image fusion framework based on convolutional neural network[J]. Information Fusion, 2020(54):99-118.

作者简介



孙晓永,2020 年获得国防科技大学博士学位,现为国防科技大学助理研究员,主要研究方向为无人平台智能侦测。
E-mail:sunxiaoyong14@nudt.edu.cn

Sun Xiaoyong received his Ph. D. degree

from National University of Defense Technology in 2020. He is currently an assistant researcher at National University of Defense Technology. His main research interest is the unmanned platform intelligent detection.



孙备(通信作者),2018 年获得国防科技大学博士学位,现为国防科技大学副研究员,硕士生导师,主要研究方向为无人平台智能侦测。
E-mail:sunbei08@nudt.edu.cn

Sun Bei (Corresponding author) received his Ph. D. degree from National University of Defense Technology in 2018. He is currently an associate research fellow at National University of Defense Technology. His main research interest is the unmanned platform intelligent detection.