

DOI: 10.19650/j.cnki.cjsi.J2311334

多尺度互交叉注意力改进的单无人机 对地伪装目标检测定位方法*

孙 备, 党昭洋, 吴 鹏, 袁书东, 郭润泽
(国防科技大学智能科学学院 长沙 410072)

摘要:为了提升无人机对地伪装目标探测能力,本文提出了多尺度互交叉注意力改进的单机对地目标检测定位方法。首先,设计了一种多尺度互交叉注意力模块,在原始多尺度金字塔基础上,进行互交叉注意力增强,提升对伪装目标的边界区分能力;其次,搭建了开源无人机目标检测定位系统,通过融合无人机载定位模块、惯导传感器和光电吊舱等数据,在获取目标图像位置后对其空间位置进行解算;最后,自行构建了丛林伪装数据集进行了相关实验验证。实验结果表明,该方法在典型伪装场景下对地目标平均检测精度(mAP)为70.2%,相较于改进前提升5.7%,且能有效输出目标与无人机(UAV)的方位距离,算法平均运行效率可达29.4 fps,满足UAV对地目标检测定位的实时性需求。

关键词:伪装目标探测;互交叉注意力;多尺度;预测定位;无人机

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

Multi scale cross attention improved method of single unmanned aerial vehicle for ground camouflage target detection and localization

Sun Bei, Dang Zhaoyang, Wu Peng, Yuan Shudong, Guo Runze

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

Abstract: To enhance the detection ability of unmanned aerial vehicles (UAV) in ground camouflage targets, this article proposes a multi-scale cross attention improved single machine ground camouflage target detection and localization method. Firstly, a multi-scale cross attention module is designed to enhance cross attention based on the original multi-scale pyramid. The ability to distinguish the boundaries of camouflaged targets is enhanced. Secondly, an open-source drone target detection and positioning system is established, which integrates data such as drone carrier positioning modules, inertial navigation sensors, and optoelectronic pods to calculate the spatial position of the target image after obtaining its position. Finally, a jungle camouflage dataset is constructed and validated through relevant experiments. The results show that the method has a ground target detection accuracy mAP of 70.2% in typical camouflage scenarios, which is 5.7% higher than before improvement. It can effectively output the azimuth distance between the target and the UAV, and the average operating efficiency of the algorithm can reach 29.4 fps, which can meet the real-time requirement of UAV ground target detection and positioning.

Keywords: camouflage object detection; cross attention; multi scale; predictive location; UAV

0 引 言

小型无人机具有操作灵活、部署方便、成本低廉等优势,在区域搜索、信息勘探、物流运输、战场侦察等领域具有重要的应用价值^[1]。尤其是随着人工智能技术的发

展,无人机自主能力得到极大拓展,复杂环境下的无人机精准感知能力得到广泛研究关注。

为了完成精准环境感知,无人机通常会携带可见光、红外、激光测距等载荷对环境目标进行数据采集、检测定位等任务。对于无人机目标检测,学术上主要关心街景为主的人、车目标^[2],而在一些环境光照度弱、环境/目标

纹理接近、目标信噪比低的场景,目标往往具有隐蔽、伪装、难分辨的特点,对无人机目标检测算法提出了更高要求^[3]。当前,无人机目标检测方法大多直接来源于学术上通用场景目标检测识别方法,在复杂环境下目标检测效果还非常有限。对于无人机目标定位,主要采取视觉跟踪与激光测距联合的对地目标定位或基于纯视觉的单机对地目标定位^[4],其中方式1精度高但流程复杂且对无人平台载体、同步控制等有较高要求,方式2操作简单但精度有限,基于纯视觉的单无人机对地目标定位还需实践验证。

本文针对单无人机对地伪装目标的检测定位问题开展研究,主要工作包括:首先,设计了一种多尺度互交叉注意力增强模块,将其嵌入到传统特征金字塔网络(feature pyramid network, FPN)结构,提高复杂场景下伪装目标检测精度;其次,构建单机对地目标定位系统,通过融合无人机载定位数据、惯导数据以及目标检测数据,对目标空间位置进行解算;最后,通过在丛林背景下布放伪装网、迷彩车模型等目标,构建丛林伪装数据集,进行单机对地目标视觉检测定位的全流程验证。实验结果表明:本文方法对典型丛林伪装目标在交并比(intersection over union, IoU)取0.5时的mAP达到70.2%,相较于原生方法提升5.7%,在无人机飞行高度 ≥ 12 m时对地目标定位误差 < 2 m,平均运行效率29.4 fps,初步验证了复杂场景下单机对地伪装目标检测定位的可行性。

1 无人机视觉检测定位相关工作

近年来,小型无人机的应用越来越广泛。如图1所示,尤其是人工智能技术的发展,具有自主感知决策能力的智能无人机系统得到越来越多的研究关注^[5]。



图1 无人机典型应用场景

Fig. 1 Typical application scenarios of UAV

1.1 无人机目标检测方面

视觉目标检测是无人机环境感知的基础。随着深度学习技术的发展,自然场景下的目标检测技术已取得一定突破。如图2所示,为不同场景下目标检测的实验数据对比。鉴于复杂场景下的无人机目标与自然场景目标的异质性,如何将通用目标检测算法改进应用到无人机目标检测场景成为研究关注的重点^[6]。

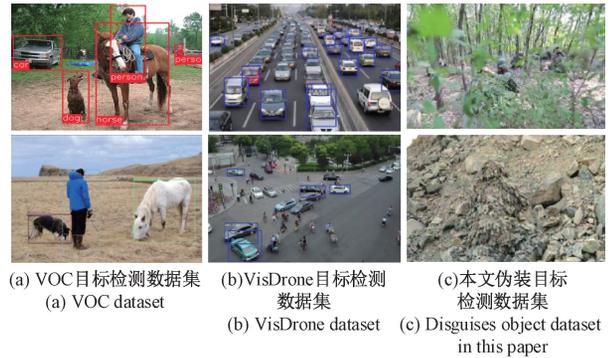


图2 不同场景下目标检测数据

Fig. 2 Target detection data in different scenarios

近年来,针对无人机航拍目标的检测问题,国内外研究者提出了一系列改进方法,总结起来主要包括两个方面:1)通过多尺度方式提升目标特征感受野,常见方式包括多尺度分支级联、空洞卷积等,例如Sun等^[7]提出语义分割引导区域推荐网络(semantic segmentation guided region proposal network, SRPN)模块,利用空洞空间金字塔池化^[8](atrous spatial pyramid pooling, ASPP)和卷积运算提取多层金字塔特征,抑制航拍图像的背景杂波,以获得更精确的分割结果;2)引入视觉注意力机制,使网络关注视野敏感区域,以提升敏感区域特征显著性,能在网络后向传输中保留目标信息。Yang等^[9]提出了一种面向旋转、稠密、小尺寸的目标检测网络(small cluttered and rotated objects detection, SCRDet),使用有监督多维注意力模块(multi dimensional attention learner, MDANET)突出目标特征,弱化背景特征。

目前,以一阶段目标检测算法(you only look once, YOLO^[10])系列为代表的目标检测方法和以位置引导的分割方法(segmenting objects by locations, SOLO^[11])为代表的实例分割方法,都可以获取目标的图像坐标和类别。其中,检测方法以矩形框形式表征目标位置;实例分割以轮廓掩码形式表征目标位置,而像素级分割方法能更加准确的描述目标位置。

因此,本文以SOLO实例分割为基础网络,结合多尺度和注意力的优势,提出了一种多尺度互交叉注意力模块,对原始网络的FPN^[12]多尺度特征进行互交叉注意力

加权融合,以提升对弱纹理目标的特征保持和语义表征能力,从而实现伪装目标的实时检测。

1.2 无人机目标定位方面

目标定位是在检测基础上对目标位置进行估计解算,随着机载光电吊舱的广泛应用,不仅可以实施目标搜索跟踪,还可以结合各类机载传感数据,对地面目标进行位置解算,在精准引导、搜救、投放等方面具有重要的应用价值。

按照采用的传感器不同可将单无人机对地目标定位分为以下两种:1)多传感器融合方式^[13],例如光电实现对地目标检测跟踪及方位估计,激光测距再进行目标距离估计,两者结合实现目标位置解算,这类方法精度较高,单成本高、操作复杂,且光电跟踪与激光测距对平台运动状态要求较高;2)基于单机纯视觉定位^[14],由光电进行目标检测跟踪及方位估计,然后结合平台位置高度及惯导信息实现方位解算,这种方式成本低,操作简单,估测精度相对较低,但满足大部分场景的应用需求,本文主要探索单无人机对地目标定位方法。

1.3 无人机视觉探测数据集

无人机航拍图像自身有着背景复杂、小目标、大视场等特点,相比于自然场景目标检测(common objects in context, COCO^[15])数据集,无人机目标检测数据集出现时间较晚,在类别数量以及人工标注数据量方面相差较多。

如图3所示,目前使用较多的航拍图像数据集例如Okutama数据集^[16]、Stanford数据集^[17]以及VisDrone数据集^[18]等,这类数据集大多拍摄于街道场景,以街道的行人、行车目标为主,其主要难点是目标分布较为密集,而对于复杂环境下目标背景纹理缺乏描述。随着无人机作业场景逐渐丰富,复杂环境下的目标数据集构建越显重要,本文针对典型丛林背景下的伪装目标检测,进行了部分数据集构建工作。

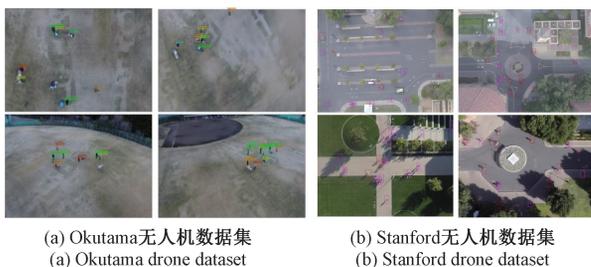


图3 公开的无人机目标检测数据集

Fig. 3 Public drone target detection datasets

2 无人机对地目标视觉检测平台设计

无人机对地目标视觉检测包括对地目标检测和对地目标定位两个部分,其中检测是为了获得目标的图像坐标,定位实际上是将图像坐标转换到相机坐标。

如图4所示,为无人机对地目标检测平台总体结构。本文搭建了一套开源的四旋翼无人机对地目标检测定位系统^[19],无人机搭载光电吊舱(包括视觉、陀螺仪加速器、编码器、磁方位传感器等)、GPS模块、测高传感器、加速度计等以及开源飞控和图数传一体的传输模块。机载端主要负责光电吊舱、定位模块、惯导传感器等数据采集,然后基于图数传一体模块传输到地面处理端进行目标检测与位置解算。为了简化成本,采取在地面端实时接收机载数据,进行在线的目标检测定位,后续随着无人机电载重和功耗能力的提升,也可基于机载平台进行实时处理。

如图5所示,为无人机对地目标检测定位数据流程。其中,G2光电吊舱的网口数据和串口控制直接连接到Homer模块的网口和串口1,其他传感器数据通过机载飞控连接到Homer模块的串口2。在地面端接收到机载数据后,进行实时目标检测识别与位置解算。

3 典型数据集构建

数据集构建包括两个部分,目标检测数据和目标定位数据。选择与丛林背景高度相似的迷彩车、伪装网等目标随机摆放,然后起飞无人机进行不同高度、不同角度的拍摄。

为了构建目标定位数据,在目标端搭载带有无线传输功能的定位模块,无人机在采集数据的同时,目标的位置数据会由搭载的模块实时传输到地面端进行记录,最后通过时间戳将无人机记录的数据和目标的真实数据进行对齐。如图6所示,为构建的伪装目标数据集示例,其中迷彩车又细分为迷彩拖车、迷彩导弹车和迷彩坦克。如图7所示,为每个类别的样本数量,拖车、导弹车和坦克都是丛林迷彩涂装,荒漠网、丛林网和军绿网选用了3种较为常见的伪装网。

4 多尺度互交叉注意力检测方法设计

4.1 多尺度互交叉注意力的提出

考虑到伪装目标具有与环境高度相似的纹理性,为了能有效分割目标,本文引入注意力模块^[20],对不同尺度的卷积特征进行互交叉增强,以提升对弱纹理目标的边界提取能力。

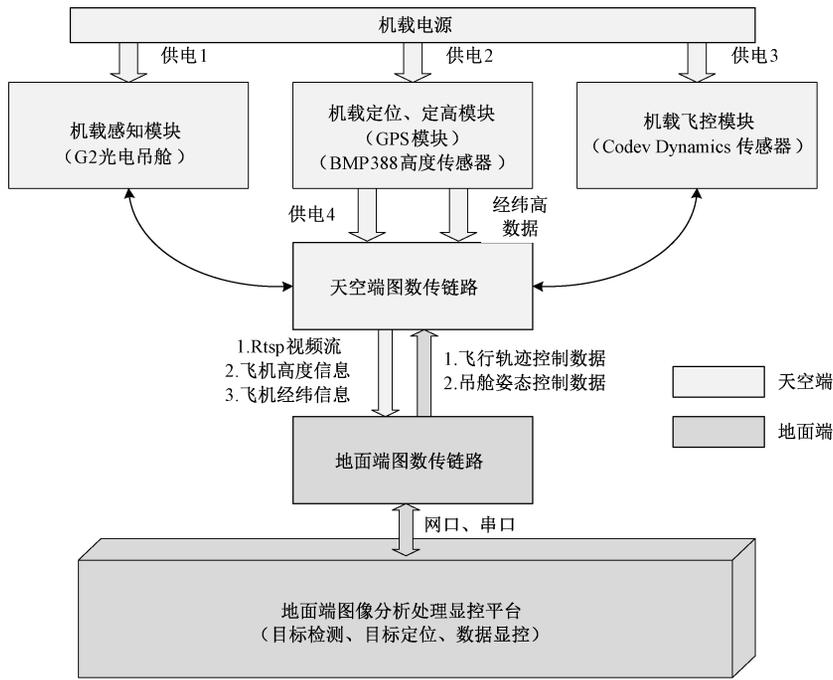


图 4 无人机对地伪装目标探测平台

Fig. 4 UAV ground camouflage target detection platform

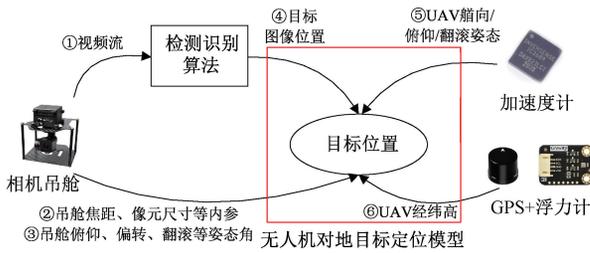


图 5 无人机对地目标定位流程

Fig. 5 UAV ground target positioning process

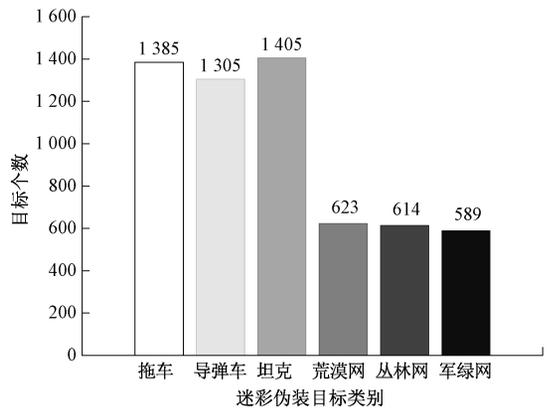


图 7 数据集各类目标数量统计

Fig. 7 Statistics on the number of various targets in the datasets



图 6 地面伪装目标数据集

Fig. 6 The ground camouflages the target datasets

多尺度特征增强是深度学习的通用模块,现有方法大多基于 FPN 结构实现。如图 8 所示,FPN 中的多尺度融合模块结构示意图。如图 8(a)所示,为经典 FPN 结构。其主要思想是将骨干网络两相邻 Block 的卷积特征进行融合,其中深层 Block 的卷积特征 C_{i+1} 经过 2 倍插值将空间维度扩大到与 C_i 一致, C_i 经过 1×1 的卷积将通道维度调整为与 C_{i+1} 一致,依次将多尺度融合模块作用到 18 层的深度残差网络 (deep residual network, ResNet) 的输出 $\{C_2, C_3, C_4, C_5\}$, 得到 3 个经过融合后的特征 $\{P_2, P_3, P_4\}$, 最终将 $\{P_2, P_3, P_4, C_5\}$ 作为 FPN 结构输出。

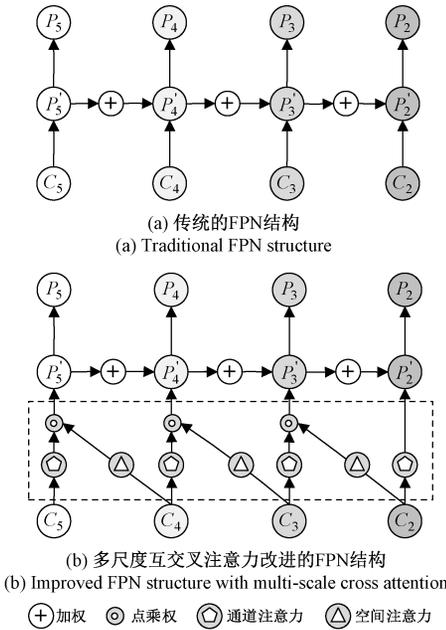


图8 FPN中的多尺度融合模块

Fig. 8 Multi scale fusion module in FPN

本文在经典 FPN 结构的基础上进行改进,如图 8(b)所示,为本文提出的多尺度互交叉注意力的结构。其主要思想是考虑用互交叉注意力模块替换经典 FPN 的加权运算。本文借鉴注意力机制和多尺度融合思想,对于两个相邻分支,首先对 C_{i+1} 进行通道注意力增强,然后计算 C_i 的空间注意力,对 C_{i+1} 分支的数据进行互交叉增强,得到新的特征 P'_{i+1} ,依次对 $\{C_2, C_3, C_4, C_5\}$ 进行计算,输出 $\{P'_2, P'_3, P'_4, P'_5\}$,再按照 FPN 结构输出新的 $\{P_2, P_3, P_4, P_5\}$ 特征。

4.2 多尺度互交叉注意力的设计与实现

如算法 1 所示,为多尺度互交叉注意力的实现流程。给定骨干网络的两个 Block 层输出的不同尺度特征,对深层卷积得到的空间注意力权值乘权到通道注意力增强后的浅层卷积特征,最终得到输出特征。

算法 1 多尺度互交叉注意力

已知:

第 l 层 Block 输出: C_i

第 $l + 1$ 层 Block 输出: C_{i+1}

步骤:

S1:对 C_i 进行通道注意力增强,得到增强后的卷积特征 C'_i ;

S2:计算 C_{i+1} 的空间注意力权值 C''_{i+1} ,将其向上 2 倍插值,空间维度变换为与 C'_i 一致,通道维度保持不变,得到变换后的空间权值 C'_{i+1} ;

S3:将 C'_{i+1} 与 C'_i 进行对应位点乘,得到互交叉注意力增强后的卷积特征 P'_i ;

S4:将 P'_i 与 C_i 进行对应位加权,输出最终卷积特征 P_i 。

如图 9 所示,为多尺度互交叉注意力模块的结构图。设 C_i 卷积特征维度为 $[b, c, w, h]$,计算通道注意力权值,维度为 $[b, c, 1, 1]$,并对 C_i 进行逐通道增强,得到增强后特征,维数为 $[b, c, w, h]$;记 C_{i+1} 卷积特征维度为 $[b_1, c_1, w_1, h_1]$,计算空间注意力权值,维数为 $[b_1, 1, w_1, h_1]$,然后进行维度变换,维数为 $[b, 1, w, h]$,然后将 C_{i+1} 的空间注意力对 C_i 分支的增强特征进行互交叉逐像素空间增强,维数为 $[b, c, w, h]$,将互交叉增强后的特征与 C_i 分支增强特征进行残差加权,得到互交叉注意力模块的最终输出。

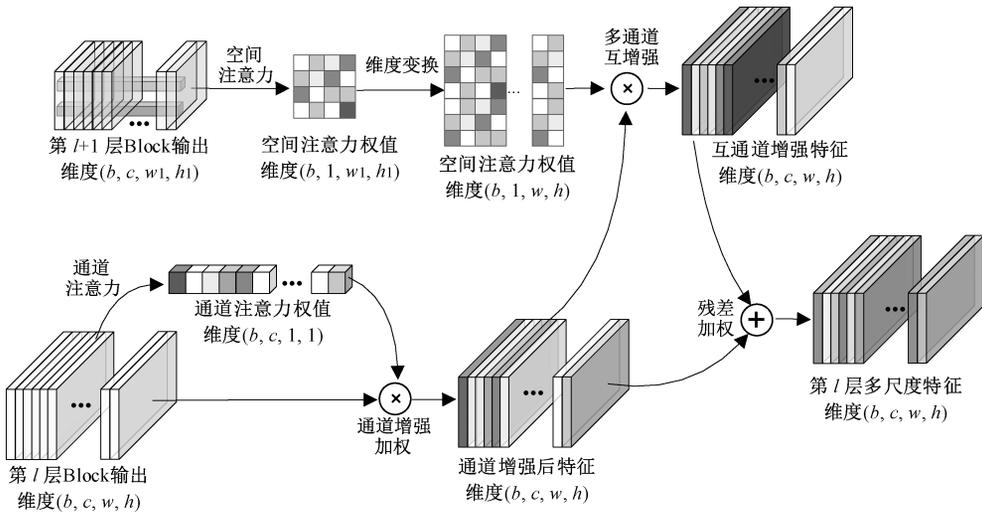


图9 多通道互交叉注意力模块

Fig. 9 Multi channel cross attention module

互交叉注意力模块实际上结合了注意力、多尺度和残差模块等深度学习近几年的主流改进方法,目的是为了提升弱对比度条件下目标特征显著性。

如图 10 所示,为互交叉注意力模块中用到的通道注意力模块和空间注意力模块。其中通道注意力模块基于通道维度平均池化,将特征维数由 $[b, c, w, h]$ 降采样到 $[b, c, 1, 1]$,然后通过降采样、上采样和激活函数等进行通道维度的变换,空间注意力模块基于空间维度的最大池化和平均池化,将特征维数由 $[b, c, w, h]$ 降采样到 2 个独立的维数为 $[b, 1, w, h]$ 的卷积特征,然后将 2 个独立的卷积特征进行通道维度的拼接,生成一个 $[b, 2, w, h]$ 的卷积特征,再通过二维卷积、激活函数等进行维度变换,最终生成 $[b, 1, w, h]$ 的卷积特征。 b 是网络的批处理大小, c, w 和 h 分别是通道维数、卷积特征的宽高。

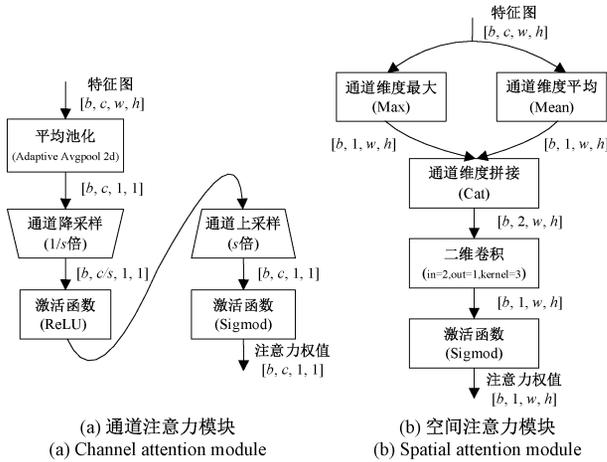


图 10 通道注意力模块与空间注意力模块

Fig. 10 Channel attention module and spatial attention module

4.3 多尺度互交叉注意力的嵌入

本文选择 SOLO 实例分割算法作为目标检测基础网络。SOLO 网络包括骨干网络、FPN 融合网络和头结构,其中骨干网络采用 ResNet18 模块,主要提取图像特征;FPN 结构主要对骨干网络的多个不同尺度输出进行融合;头结构主要是输出与任务相关的掩码和类别预测。

参考原生 SOLO 网络,对骨干网络第 2~5 层卷积块输出进行相邻通道的互交叉增强。如图 11 所示,为多通道互交叉注意力模块改进后的 SOLO 网络结构。互交叉注意力嵌入位置分为为:骨干网络第 2 与 3 个卷积块,第 3 与 4 个卷积块,第 4 与 5 个卷积块。由于互交叉注意力模块不改变数据维数,本文沿用 SOLO 的头结构对融合后特征进行处理。由于 SOLO 的其他模块已经较为成熟,本文不做过多介绍。

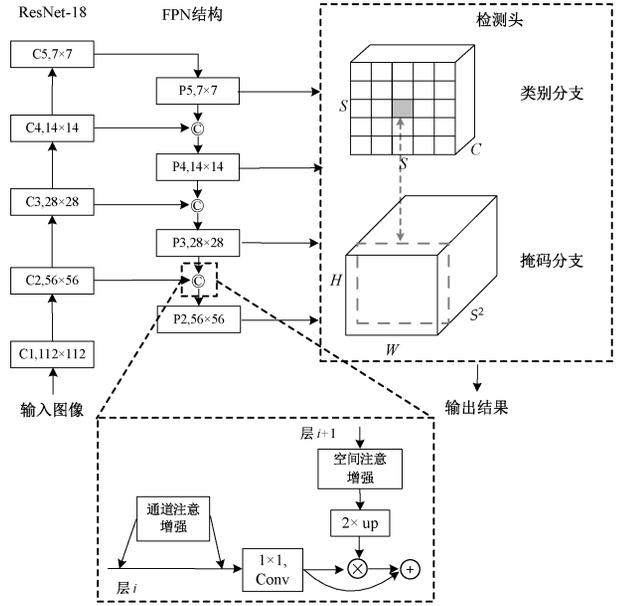


图 11 多通道互交叉注意增强的 SOLO 网络

Fig. 11 The improved SOLO with multi-channel cross attention

4.4 损失函数

损失是评估算法预测值与真实值偏离程度的标准。本文实例分割算法的损失函数由分类损失和掩码损失两个部分组成,其计算方法如式(1)所示。

$$L = L_{conf} + \lambda L_{mask} \quad (1)$$

其中, L_{conf} 是分类分支的损失函数, L_{mask} 掩码分支的损失函数, λ 为常数,用于调节两部分的权重,本文设为 3。 L_{conf} 采用 Focal 损失函数计算,定义如式(2)所示。

$$L_{conf} = -\alpha(1 - p_{cls})^\tau \log(p_{cls}) \quad (2)$$

其中, p_{cls} 是类别概率, α 和 τ 取值 0.25 和 2,分别解决正、负样本比例和难、易样本比例不平衡的问题。 L_{mask} 的计算如式(3)所示。

$$L_{mask} = \frac{1}{N_{pos}} \sum_k l_{|p_{cls} > 0|} d(m_o^k, m_t^k) \quad (3)$$

其中, N_{pos} 是正样本数量, $l_{p_{cls}}$ 是指标函数,当 $p_{cls} > 0$, 函数 $l_{p_{cls}}$ 为 1,反之为 0。 $d(\cdot, \cdot)$ 选用 Dice 损失函数,其中 m_o^k 和 m_t^k 分别是第 k 块上的预测掩码和真实掩码。

5 基于视觉检测的单 UAV 对地目标定位

对地目标定位实际上是由图像坐标到世界坐标的转换,单无人机对地目标定位可以借助两者的高程差作为先验知识^[21],对目标位置进行初步估算。

如图 12 所示,为无人机对地目标定位流程。在基于深度学习方法获得目标的图像位置后,融合无人机经纬高等自身姿态位置信息、光电吊舱方位角、俯仰角以及无

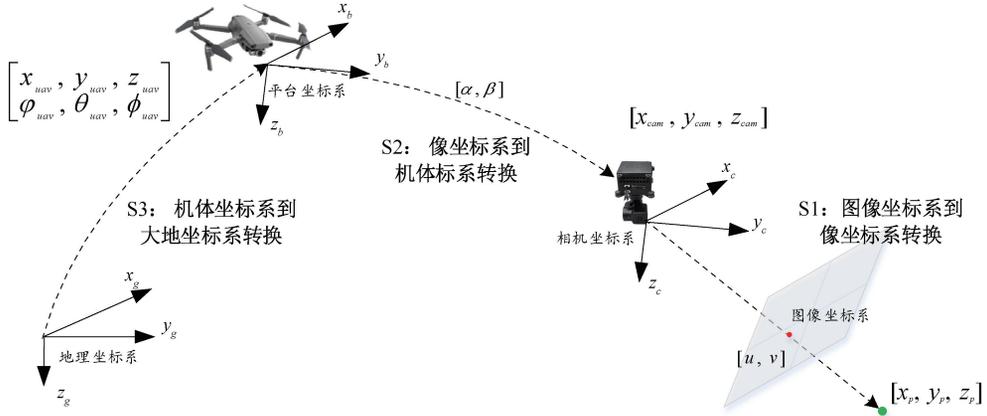


图12 无人机对地目标定位流程

Fig. 12 UAV ground target positioning process

人机与地面目标斜距等信息,即可精确解算出地面伪装目标的绝对位置,实现无人机对地面目标的定位。

具体流程包括:首先地面端处理平台从机载光电吊舱获取的图像中识别到目标并获得图像位置,然后按照图像坐标到相机坐标的转换方式解算目标的相机坐标 $[x_{cam}, y_{cam}, f]^T$ 。在此基础上,按照下式求解目标的世界坐标。

$$\begin{bmatrix} x_p \\ y_p \end{bmatrix} = \begin{bmatrix} x_{uav} \\ y_{uav} \end{bmatrix} + \frac{\Delta z_{p-uav}}{[0, 0, 1] \mathbf{C}_c^n \mathbf{C}_b^h \mathbf{C}_c^h [x_{cam}, y_{cam}, f]^T} \times \begin{bmatrix} 1, 0, 0 \\ 0, 1, 0 \end{bmatrix} \mathbf{C}_b^h \mathbf{C}_c^h [x_{cam}, y_{cam}, f]^T \quad (4)$$

其中, Δz_{p-uav} 是无人机到地面的飞行高度, f 是无人机吊舱焦距, \mathbf{C}_c^h 是图像坐标系到机体坐标系的转换矩阵, \mathbf{C}_b^h 是从机体坐标系到地理坐标系的转换矩阵。光电吊舱翻滚角设为 0° , 记 (α, β) 为光电吊舱的方位角和俯仰角, 则转换矩阵 \mathbf{C}_c^h 为:

$$\mathbf{C}_c^h = \begin{bmatrix} \cos \alpha & \sin \alpha & 0 \\ -\sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} \cos \beta & 0 & \sin \beta \\ 0 & 1 & 0 \\ -\sin \beta & 0 & \cos \beta \end{bmatrix} \quad (5)$$

记 (φ, θ, ϕ) 为无人机姿态位置信息, 分别表示滚转角、俯仰角和航向角, 则转换矩阵 \mathbf{C}_b^h 可表示为:

$$\mathbf{C}_b^h = \begin{bmatrix} \cos \varphi & 0 & \sin \varphi \\ 0 & 1 & 0 \\ -\sin \varphi & 0 & \cos \varphi \end{bmatrix} \times \begin{bmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \phi & -\sin \phi \\ 0 & \sin \phi & \cos \phi \end{bmatrix} \quad (6)$$

以无人机经纬位置为原点, 即 $(x_p, y_p) = (0, 0)$, 在求解目标相对于无人机的坐标 (x_p, y_p) 的基础上, 根

据 (x_p, y_p) 坐标转经纬度坐标公式, 计算目标点经纬位置。

$$\begin{aligned} x_p &= (lat_p - lat_{uav}) \times \frac{\pi}{180} \times R_{earth} \\ y_p &= (lon_p - lon_{uav}) \times \frac{\pi}{180} \times R_{earth} \times \left(lat_p \times \frac{\pi}{180} \right) \end{aligned} \quad (7)$$

其中, R_{earth} 是地球半径, 取值 6 378 245。在具体实施前, 还需对吊舱内参进行标定, 以校正像元畸变等误差。标定工作可选择常见的黑白格板, 基于 MATLAB、OpenCV 等标定工具箱实现, 标定参数最终用于图像坐标到相机坐标的解算, 本文不作过多阐述。

6 实验与分析

6.1 实验设置

为了保持对比实验的一致性, 将所有的网络都设置为不输入预训练模型。网络训练参数迭代次数设置为 200, 批处理大小设置为 10, 初始学习率设置为 0.01, PC 环境为 Intel Core i7-11800H 型 CPU, TITAN RTX 显卡, Ubuntu 20.04 系统, Pytorch 1.8.0 软件环境。

为了证明互交叉注意力模块的有效性, 设计了如下不同的模块嵌入方式。如图 13 所示, 按照互交叉注意力结构嵌入到 FPN 结构不同层, 共形成了 3 种不同嵌入结构。为了便于区分, 将只对第 2、3 层的互交叉注意力增强称为改进 1, 对第 2、3、4 层的互交叉注意力增强称为改进 2, 对第 2、3、4、5 层的互交叉注意力增强称为改进 3。

6.2 评价指标

评价指标包括精度评价和效率评价两个方面。其中, 分割精度采用 COCO 的 mAP 标准指标体系, 分别选

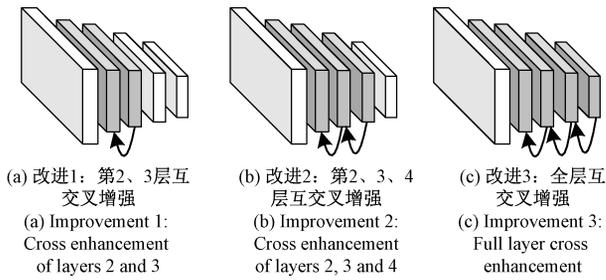


图 13 互交叉注意力不同嵌入位置

Fig. 13 Cross attention with different embedding positions

用 $mAP@0.5$ 、 $mAP@0.75$ 和 $mAP@0.5:0.95$ 等指标, $mAP@t$ 中的 t 表示判定为正确预测的 IoU 阈值;分割效率采用 fps 来进行衡量。IoU 是预测掩码与真实掩码的交并比,具体计算方式为:

$$IoU = \frac{R_{pre} \cap R_{true}}{R_{pre} \cup R_{true}} \quad (9)$$

式中: R_{pre} 表 R_{true} 分别表示预测掩码与真实掩码区域, IoU 大于阈值 t ($t \in [0, 1]$), 则认为预测结果正确。

6.3 消融实验

1) 训练损失对比

对图 13 所示的 3 种改进方式进行消融对比实验。如图 14 所示,列举了 200 个迭代次数下,各个网络的模型收敛情况,其中训练初始状态设置相同。

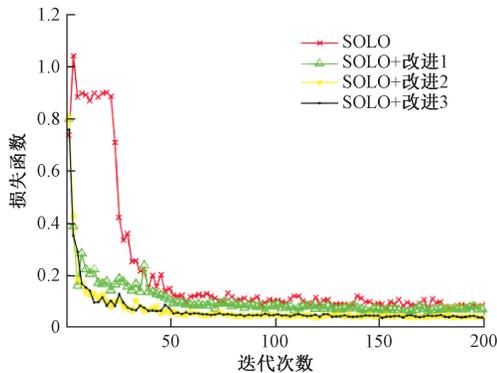


图 14 不同网络损失函数对比

Fig. 14 Comparison of different network loss function

由图 14 可知,改进后的方法具有更高训练效率:一方面互交叉注意力改进模型的损失收敛速度更快,大约在 10 个迭代次数时的损失降为 0.2 以下,而原始 SOLO 需要在 25 个迭代次数后才开始下降;另一方面改进后模型稳定收敛的损失更小。此外,发现改进 2 和改进 3 的损失收敛较为一致,这说明对深层卷积的多尺度互交叉注意力增强的提高更明显,后续将结合实际检测结果进行分析。

2) AP 指标对比

如表 1 所示,列举了不同改进方法的指标评测结果,选取 $AP.5$ 、 $AP.75$ 和 $AP.5:0.95$ 和 $AR.5:0.95$ 值。

表 1 添加到不同位置的对比实验

Table 1 Comparison experiments added to different locations

模型	AP.5	AP.75	AP.5:0.95	AR.5:0.95
原生网络	0.645	0.374	0.362	0.427
改进 1	0.687	0.476	0.426	0.467
改进 2	0.691	0.543	0.457	0.490
改进 3	0.702	0.561	0.460	0.491

由表 1 可知,改进后方法的模型检测精度相较于原生 SOLO 有一定提升, $AP.5$ 最大提高 5.7%, $AP.75$ 最大提高 18.7%, $AP.5:0.95$ 最大提高 9.8%, 结果充分证明了互交叉注意力对原生网络的检测精度具有明显作用。 $AP.75$ 和 $AP.5:0.95$ 的提升幅度明显大于 $AP.5$, 说明互交叉注意力改进方法具有较高的目标检测置信度。

如图 15 所示,列举了 3 种改进策略的分割结果。由图可知,只有改进 3 有效分割了图中迷彩导弹车模型,而其他两种方法都漏检了该类目标。

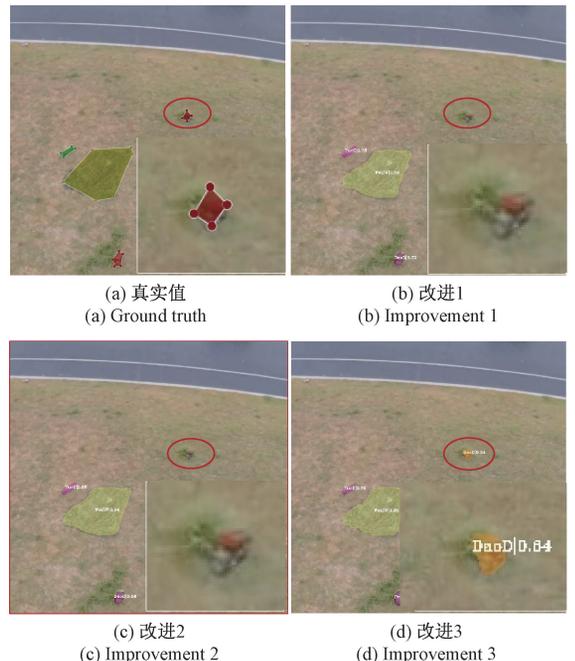


图 15 不同迭代次数阶段的 AP.5 输出

Fig. 15 AP.5 output at different epoch stages

对比互交叉注意力嵌入到网络不同位置的性能,发现改进 3(第 2~5 层嵌入)的性能最优,改进 2(第 2~4 层嵌入)的性能次优,改进 2(只对第 2、3 层嵌入)的提升幅

度最小,实验结果符合预期。即更充分的利用深层卷积对算法检测性能的提升作用更加显著,通过在原始 FPN 结构上进行简答的互交叉加权,可有效提升不同尺度感受野的相互作用,即对网络结构有更显著的提升作用。

3) 不同尺寸检测对比

进一步,按照 COCO 评价标准对多尺度目标检测结果进行了对比。如表 2 所示,改进方法在小、中和大尺度上的检测性能都有一定提升,其中中尺度的提升效果最佳(提升 10.6%),大尺度次之(提升 9.2%),小尺度提升幅度最小(提升 7.1%)。这一方面说明了互交叉注意力对多尺度目标特征的增强作用,另一方面也是由于本文的互交叉注意力结构嵌入在骨干网络的后几层,对中、大尺度目标的特征描述能力更强。此外,小目标检测精度受限主要是由于 SOLO 骨干网络本身表征小尺度目标的浅层卷积网络特征描述能力较弱的限制影响。

表 2 对不同尺度检测结果的对比实验

Table 2 Comparison of detection results at different scales

模型	AP			AR		
	小尺度	中尺度	大尺度	小尺度	中尺度	大尺度
原生网络	0.221	0.335	0.465	0.270	0.399	0.514
改进 1	0.282	0.409	0.523	0.305	0.442	0.558
改进 2	0.291	0.436	0.557	0.312	0.469	0.581
改进 3	0.292	0.441	0.557	0.310	0.469	0.582

6.4 与其他注意力对比实验

1) 注意力嵌入形式

互交叉注意力实际上使用了通道注意力和空间注意力模块,为了进一步分析互交叉注意力的性能,分别对只加通道注意力和只加空间注意力的结构进行对比。

如图 16 所示,为两种结构嵌入方式,挤压-激励网络(squeeze and excitation networks, SENet)和空间注意力模块(spatial pyramid networks, SPNet)分别作为通道、空间注意力。其中图 16(a)只对某个分支进行了通道注意力,未进行多尺度分支间的互交叉注意力;图 16(b)用浅层分支的空间注意力对深层分支进行了互交叉增强。

2) 不同注意力检测结果

如表 3 所示,列举了 3 种注意力结构检测性能的量化结果。由表 3 可知,图 16(a)只对某个分支进行通道注意力增强的检测结果相较于原生 SOLO 网络的检测性能未有明显改进,甚至略低;图 16(b)用浅层分支的空间注意力对深层分支进行互交叉增强后,相较于原生 SOLO 网络的提升较为显著,其中 AP@0.5 提高 4.6%, AP@0.75 提高 16.3%, AP@0.5:0.95 提高 8.9%;而本文方法的提升作用最为显著。

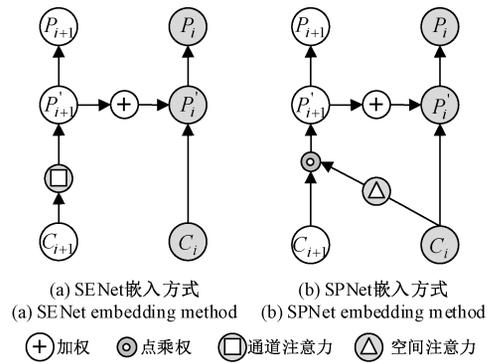


图 16 不同注意力模块嵌入方式

Fig. 16 Different attention module embedding methods

表 3 不同注意力模块对比

Table 3 Comparison of different attention modules

模型	AP@0.5	AP@0.75	AP@0.5:0.95	AR@0.5:0.95
原生网络	0.645	0.374	0.362	0.427
SOLO+SE	0.642	0.369	0.365	0.424
SOLO+SP	0.691	0.537	0.451	0.491
本文方法	0.702	0.561	0.460	0.491

由表 3 可知,有无多尺度之间的互交叉模块区别最大,引入互交叉注意力模块后,模型性能的提升幅度最为显著,这是因为浅层 C_i 相较于深层 C_{i+1} 的特征图分辨率更高,对目标特征的细节描述更好,用 C_i 层增强 C_{i+1} 层相当于在低分辨率特征图上赋予了更高分辨率的细节指导,这一定程度上证明了对不同尺度卷积特征进行相互作用,可获得较好的实验性能。

如图 17 所示,列举了不同方法在几种场景的检测结果。由图可知,第 1 行中,SOLO+SE 方法对迷彩坦克的边界分割出现较大失真;第 2 行和第 3 行中,SOLO+SE 方法未检测出迷彩拖车。结果表明,SOLO+SP 和本文方法都引入互交叉注意力,网络检测效果得到较明显提升。

6.5 单机对地目标视觉检测定位应用

在视觉检测的基础上,对单机定位进行了初步验证。如图 18 所示,为构建的开源无人机对地目标检测定位原型系统。

如图 18(a)所示,为开源无人机,按照图 12 的定位流程构建;如图 18(b)所示,为目标携带的“定位+数传”模块,用于实时将目标位置传输到 PC 端进行记录保存;如图 18(c)所示,为用于目标伪装的迷彩网与迷彩服。实验主要模拟 UAV 在飞行高度 ≥ 10 m 时,对地面伪装的人进行检测定位。

如表 4 所示,随机列举了几组定位数据,可知在 UAV ≥ 10 m 下,对地面伪装的人具有一定的定位效果,定位

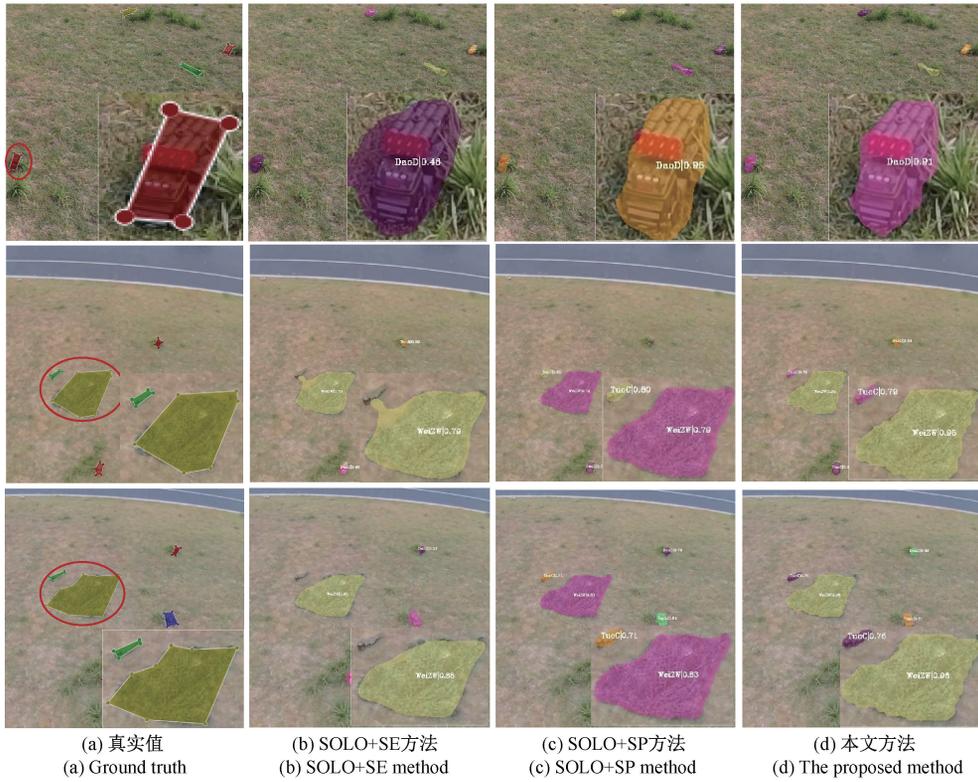


图 17 不同方法的明细对比指标

Fig. 17 Detailed comparison indicators for different methods



图 18 实验无人机及伪装目标设置

Fig. 18 Experimental drone and camouflage target setting

表 4 UAV 定位结果分析

Table 4 Analysis of UAV positioning results m

UAV 高度	实际距离	预测距离	距离误差
13.34	17.02	18.58	1.56
13.37	11.40	13.10	1.30
13.65	9.84	8.51	1.33
13.58	9.65	7.46	1.19

误差 <2 m。这在单 UAV 视觉定位中已经具有一定效果。

如图 19 所示,为连续检测定位结果的实时显示。在傍晚时分丛林背景下对伪装的人具有较好的检测效果,并能输出实时定位估测。此外,实际检测结果表明:平均运行帧率为 29.4 fps,满足实时性需求。

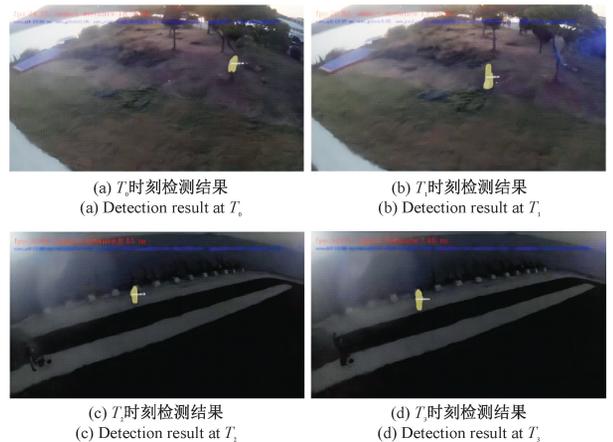


图 19 单无人机对地目标实时定位结果

Fig. 19 Real time location results of single drone to ground targets

7 结 论

无人机搭载光电吊舱实现对地目标检测定位,在区域搜索、信息勘探、物流运输、战场侦察等领域具有重要的应用价值。本文针对复杂环境下的无人机视觉检测定位问题,首先提出了一种多尺度互交叉注意力增强的深度学习目标检测方法,以提升背景/目标纹理高度相似场景下的特征分辨能力;其次构建了单机对地目标定位系统,结合检测结果、无人机经纬、惯导姿态以及吊舱状态等进行目标位置解算;最后通过在丛林背景放置迷彩类的伪装网、迷彩车等目标,构建伪装目标数据集,并基于该数据对所提出方法进行了检测定位实验。结果表明,本文方法能明显提升无人机伪装目标检测性能,初步实现目标定位,并具有较高的实时效率。未来的研究一方面重点考虑 UAV 平台和目标存在较大相对运动时的检测定位问题;另一方面将考虑实时机载检测应用;此外,在进一步完善的基础上考虑无人机伪装目标数据集的开源工作。

参考文献

- [1] 江波, 屈若锟, 李彦冬, 等. 基于深度学习的无人机航拍目标检测研究综述[J]. 航空学报, 2021, 42(4):137-151.
JIANG B, QU R K, LI Y D, et al. Review of UAV aerial photography target detection based on deep learning[J]. Journal of Aeronautics, 2021, 42(4): 137-151.
- [2] 程旭, 宋晨, 史金钢, 等. 基于深度学习的通用目标检测研究综述[J]. 电子学报, 2021, 49(7): 1428-1438.
CHENG X, SONG CH, SHI J G, et al. Review of general object detection based on deep learning[J]. Chinese Journal of Electronics, 2021, 49(7): 1428-1438.
- [3] 彭继慎, 孙礼鑫, 王凯, 等. 基于模型压缩的 ED-YOLO 电力巡检无人机避障目标检测算法[J]. 仪器仪表学报, 2021, 42(10): 161-170.
PENG J SH, SUN L X, WANG K, et al. ED-YOLO electric patrol UAV obstacle avoidance target detection algorithm based on model compression [J]. Chinese Journal of Scientific Instrument, 2021, 42(10): 161-170.
- [4] 石风, 朱有德. 一种无人机载光电吊舱目标定位方法[J]. 计算机与数字工程, 2014, 42(12): 2259-2262.
SHI F, ZHU Y D. A target positioning method of UAV-borne photoelectric pod [J]. Computer and Digital Engineering, 2014, 42(12): 2259-2262.

- [5] 宋谱怡, 陈红, 苟浩波. 改进 YOLOv5s 的无人机目标检测算法[J]. 计算机工程与应用, 2023, 59(1): 108-116.
SONG P Y, CHEN H, GOU H B. Improved YOLOv5s UAV target detection [J]. Computer Engineering and Applications, 2023, 59(1): 108-116.
- [6] ZOU ZH X, SHI ZH W, GUO Y H, et al. Object detection in 20 years: A survey [C]. IEEE Computer Vision and Pattern Recognition, 2019.
- [7] SUN Y B, HUANG X R, ZHOU H Y, et al. SRPN: Similarity-based region proposal networks for nuclei and cells detection in histology images [J]. Medical Image Analysis, 2021, 72: 102-142.
- [8] 张瑞倩, 邵振峰, ALEKSEI P, 等. 多尺度空洞卷积的无人机影像目标检测方法[J]. 武汉大学学报(信息科学版), 2020, 45(6): 895-903.
ZHANG R Q, SHAO ZH F, ALEKSEI P, et al. UAV image target detection method based on multi-scale void convolution [J]. Geomatics and Information Science of Wuhan University, 2020, 45(6): 895-903.
- [9] YANG X, YANG J, YAN J, et al. SCRDet: Towards more robust detection for small, cluttered and rotated objects [C]. ICCV 2018: Proceedings of the 2018 IEEE International Conference on Computer Vision, Piscataway, IEEE Press, 2019: 8232-8241.
- [10] 徐晓光, 李海. 多尺度特征在 YOLO 算法中的应用研究[J]. 电子测量与仪器学报, 2021, 35(6): 96-101.
XU X G, LI H. Application research of multi-scale features in YOLO algorithm [J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(6): 96-101.
- [11] WANG X L, ZHANG R F, KONG T, et al. SOLOv2: Dynamic and fast instance segmentation [C]. Proceedings of Advances in Natural Information Processing Systems, ArXiv E-prints, 2020: 17721-17732.
- [12] KRISHNA H, JAWAHAR C. Improving small object detection [J]. In Proceedings of the 4th IAPR Asian Conference on Pattern Recognition, 2017, 10: 340-345.
- [13] 王贺, 卜智勇, 谭冲. 基于核相关滤波视觉检测的多旋翼无人机对地目标跟踪与逼近[J]. 中国科学院大学学报, 2022, 39(2): 217-223.
WANG H, BU ZH Y, TAN CH. Multi rotor unmanned aerial vehicle ground target tracking and approximation based on kernel correlation filtering visual detection [J]. Journal of the University of the Chinese Academy of Sciences, 2022, 39(2): 217-223.
- [14] 孙聪, 贾萌萌, 于起峰. 基于仿射近似投影模型的无人机对地目标定位方法[J]. 中国惯性技术学报, 2022,

- 30(1): 104-112, 120.
- SUN C, JIA M N, YU Q F. A method for ground target localization of unmanned aerial vehicles based on affine approximate projection model [J]. Journal of Chinese Inertial Technology, 2022, 30(1): 104-112, 120.
- [15] 李云红, 张轩, 李传真, 等. 融合 DBSCAN 的改进 YOLOv3 目标检测算法[J]. 计算机工程与应用, 2022, 58(10): 208-215.
- LI Y H, ZHANG X, LI CH ZH, et al. Improved YOLOv3 target detection algorithm combined with DBSCAN [J]. Computer Engineering and Applications, 2022, 58(10): 208-215.
- [16] AHMAD T, CAVAZZA M, MATSUO Y, et al. Detecting human actions in drone images using YOLOv5 and stochastic gradient boosting [J]. Sensors, 2022, 22(18): 7020.
- [17] 张垚, 鲜斌, 于琰平, 等. 四旋翼无人机可视化半实物仿真平台研究 [J]. 仪器仪表学报, 2012(11): 2572-2578.
- ZHANG Y, XIAN B, YU Y P, et al. Research on visual hardware in the loop simulation platform of quadrotor UAV [J]. Chinese Journal of Scientific Instrument, 2012(11): 2572-2578.
- [18] CAO Y, HE Z, WANG L, et al. VisDrone-DET2021: The vision meets drone object detection challenge results[C]. IEEE International Conference on Computer Vision, 2021.
- [19] 白冠冰, 宋悦铭, 左羽佳, 等. 机载光电平台的对地多目标定位 [J]. 光学精密工程, 2020, 28(10): 2323-2336.
- BAI G B, SONG Y M, ZUO Y J, et al. Ground multiple target localization of airborne opto electronic platform[J]. Optics and Precision Engineering, 2020, 28(10): 2323-2336.
- [20] JIE H, LI SH, GANG S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 9(42): 2011-2023.
- [21] 鲁亚飞, 吴岸平, 陈清阳. 无人机对地目标多帧融合定位与误差收敛特性分析 [J]. 国防科技大学学报, 2021, 43(2): 66-73.
- LU Y F, WU AN P, CHEN Q Y. Analysis of multiframe fusion localization and error convergence characteristics for ground target of unmanned aerial vehicles[J]. Journal of National University of Defense Technology, 2021, 43(2): 66-73.

作者简介



孙备, 分别于 2012 年、2014 年和 2018 年于国防科技大学获得学士、硕士和博士学位, 现为国防科技大学副研究员, 主要研究方向为无人平台智能侦测与对抗。

E-mail: sunbei08@nudt.edu.cn

Sun Bei received his B. Sc., M. Sc. and Ph. D. degrees all from National University of Defense Technology in 2012, 2014 and 2018. He is currently an associate researcher at National University of Defense Technology. His main research interest includes intelligent detection and countermeasure of unmanned platforms.



党昭洋 (通信作者), 2022 年获得成都信息工程大学学士学位, 正在国防科技大学攻读硕士学位, 主要研究方向为无人平台智能侦测与对抗。

E-mail: dzy0329@nudt.edu.cn

Dang Zhaoyang (Corresponding author) received his B. Sc. degree from Chengdu University of Information Technology in 2022. He is currently a master degree at National University of Defense Technology. His main research interests include intelligent detection and countermeasure of unmanned platforms.