

DOI:10.19650/j.cnki.cjsi.J2413498

无人机弱光条件下多模态融合目标检测方法^{*}

郭润泽, 孙 备, 孙晓永, 卜德森, 苏绍璟
(国防科技大学智能科学学院 长沙 410072)

摘 要:弱光条件下, 图像亮度低、对比度弱、成像质量差, 且由于机载算力的制约, 算法部署应用难以落地, 极大地影响了无人机对目标的识别定位精度。因此, 无人机弱光照条件下目标检测方法具有重要的理论意义和应用价值。针对此问题, 提出了耦合光照条件 and 对比度的多尺度差分注意力融合检测方法。首先, 设计了信息感知引导的多尺度差分注意力融合检测网络, 通过信息感知模块计算图像的光照信息和目标的局部对比度, 来引导多尺度差分注意力模块对可见光和红外图像的模态内和模态间特征进行深度交叉融合, 以提升弱光条件下无人机对地目标的检测识别精度; 其次, 基于多模吊舱、边缘计算模块和自组网电台构建了一套旋翼无人机多模目标检测系统, 针对可见光和红外数据, 在通信交互上具有规范的传输协议和统一的任务管理机制, 可实现同步解码; 随后, 设计了对比和消融实验, 实验结果显示该方法在典型暗光照数据集 LLVIP 上 mAP 达到 69.2%, 较改进前提升 3.9%, 并优于典型的双流网络 LRAF-Net。最后, 在机载端对本文算法进行了轻量化部署和验证, 结果表明在真实弱光场景下该算法能显著提升无人机对目标的检测能力, 且平均运行效率可达 21.2 FPS, 满足机载端应用需求。

关键词: 弱光条件; 无人机; 多尺度差分注意力; 多模; 目标检测

中图分类号: TP391.4 TH865

文献标识码: A

国家标准学科分类代码: 510.4050

Multimodal fusion object detection method for UAVs under low light conditions

Guo Runze, Sun Bei, Sun Xiaoyong, Bu Desen, Su Shaojing

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

Abstract: Under low light conditions, factors such as low image brightness, weak contrast, poor imaging quality and the constraints of on-board arithmetic greatly affect the detection accuracy from the UAV's point of view. Therefore, researches based on object detection under low light conditions in UAVs is of great significance. Aiming at this problem, this paper proposes a multiscale differential attention fusion detection method with coupled illumination conditions and contrast. First, an information-aware module is designed to guide the multiscale differential attention module. This module deeply fuses the intra- and inter-modal features of visible and infrared images through calculating the light information and local contrast, thereby enhancing the recognition ability under low light conditions. Second, a rotary-wing UAV multimodal target detection system is constructed based on multimodal pods, edge computing modules and self-organizing network radios. This system has a standardized transmission protocol and a unified task management mechanism for communication interaction and realizes synchronous decoding. Subsequently, comparison and ablation experiments are designed, and the results show that the mAP of this method on the LLVIP is 69.2%, which is 3.9% better than before the improvement, and outperforms LRAF-Net. Finally, the proposed algorithm is validated at the airborne end of USVs, demonstrating that it can significantly improve the detection capability of UAVs on targets under low light conditions. The average operation efficiency can reach 21.2 FPS, which meets the requirements of airborne applications.

Keywords: low light conditions; UAVs; multiscale differential attention; multimodal; object detection

0 引言

当前,人工智能正逐渐成为社会形态质变的第一推动力,以无人机为代表的低空经济产业得到了空前的重视和发展。凭借成本低廉、操作便捷等优势,无人机可部署于人类难以生存的区域^[1],在城市、山地、丛林等复杂地形地貌中穿梭,在广域搜索、测量测绘、跟踪检测、引导定位等领域发挥着不可替代的作用^[2]。然而,在恶劣的气象环境特别是弱光条件下,图像亮度低、对比度弱,成像质量差,极大地影响了无人机对目标的识别定位精度,为无人机端自主探测和环境感知带来挑战。因此,研究无人机弱光照条件下目标检测方法具有重要的理论意义和应用价值。

部分工作通过目标增强方法提升弱光条件下的图像质量,进而提高感知精度^[3-4]。陈清江等^[5]结合交叉残差结构和注意力模块提出多通道融合注意力网络,有效提高了弱光照条件下图像的亮度。闫晓阳等^[6]提出基于颜色空间的图像增强网络,抑制了弱光照条件下的噪声,解决了弱光照条件下目标和背景耦合紧密,难以分辨的问题。然而,基于单一模态的图像增强技术往往会导致颜色失真,影响图像的细节信息呈现,缺乏鲁棒性。

为缓解单一模态在弱光条件下的局限性,越来越多的研究人员开始关注多模态融合检测,通过融合可见光、红外等不同传感器的视觉信息实现全天候、多时段的探测。由于其强大的特征提取能力,双分支卷积神经网络被广泛用于多模态特征的融合。Fang 等^[7]提出一种多模态特征融合方法,通过上下文信息的指导学习可见光和红外模态之间的潜在联系。郝博等^[8]以 YOLO (you look only once, YOLO) 网络为基础,在双分支网络中加入注意力模块,提升了对辨别性区域和小尺寸目标的检测能力。然而,现有双分支方法大多采用简单的融合策略^[9-11],如元素相加或相乘,虽然比单模态检测具有更高的性能,但是它们并没有充分考虑和利用模态间的互补信息。更糟糕的是,简单的融合机制缺乏长期依赖性,随着训练的进行,可能会加剧网络的不平衡,从而导致检测结果不理想。此外,双分支方法存在架构冗余、网络参数量大等问题,由于机载算力的制约,部署应用难以落地。因此,本文的目标是设计高效且轻量化的多模态特征提取网络和融合机制,以增强弱光条件下无人机的精确感知能力。

对无人机弱光条件下多模态融合检测问题开展研究,主要工作包括:首先,提出了一种耦合光照信息和对比度的多尺度差分注意力融合模块,通过信息感知模块的计算,引导可见光和红外模态深度交叉融合,提高弱光条件下的目标检测精度;其次,构建了一套旋翼无人机多模目标检测系统,通过搭载三光吊舱和机载边缘计算板

卡,实现了机载端的数据采集和处理;最后,在公开数据集低光可见光-红外配对数据集(low light visible-infrared paired, LLVIP)^[12]和机载端,对本文算法进行了全流程验证。结果表明,本文方法在 LLVIP 数据集上平均精度均值(mean average precision, mAP)达到 69.2%,优于典型双流网络长距离注意力融合网络(long-range attention fusion network, LRAF-Net)^[13]。在机载端,本文算方法能显著提升弱光条件下对目标检测性能,且平均运行效率的每秒帧数(frame per second, FPS)为 21.2 fps,满足无人机实时的探测应用需求。

1 无人机弱光条件下目标检测相关工作

如图 1 所示,无人机以其独特的优势广泛应用在军用和民用领域,特别是随着高分辨率成像技术的发展,基于无人机的自主检测和识别技术成为各国机构的热门研究课题。



图1 无人机典型应用案例

Fig. 1 Typical application cases of UAVs

1.1 弱光条件下航拍数据特性分析

无人机弱光条件下的视觉检测是指无人机在光照不足环境下,如雾霾、雨雾、夜晚等天气,捕捉实时图像,对目标进行检测识别的过程^[14]。图2所示为典型弱光条件下无人机所拍摄的图像,矩形框为目标位置。从图中可以看出,弱光照条件下图像的整体亮度降低,噪声增大,目标边缘和纹理特征丢失,目标的可见性变差,可能导致漏检、误检等问题的出现。

受拍摄角度和飞行高度的影响,与自然图像数据相比,无人机航拍数据具有目标尺寸差异大、视角多样等特点,从而导致算法在处理无人机端数据时性能不均衡,难以准确分类。图3展示了典型航拍数据集 Visdrone^[15]的

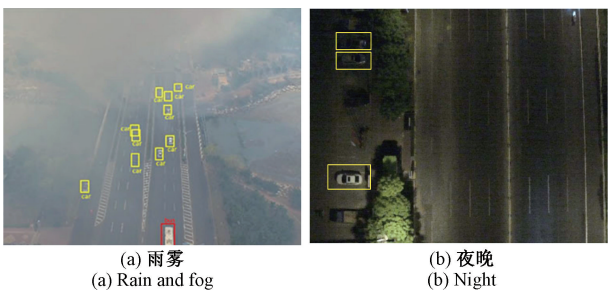


图 2 无人机弱光条件下航拍数据实例
Fig. 2 Examples of aerial data from UAVs in low light conditions

数据分布特点。该数据集中小于 $16\text{ pixels}\times 16\text{ pixels}$ 和介于 $16\text{ pixels}\times 16\text{ pixels}$ 和 $32\text{ pixels}\times 32\text{ pixels}$ 的目标在数据集中的占比高达 12.05% 和 32.65%，均高于自然数据集^[16]中的 7.35% 和 14.38%，这表明在无人机航拍图像里中小目标的占比较高，且不同目标尺寸差异也较大。因此，本文算法采用了多尺度、多梯次的特征提取方式，以提升无人机视角下航拍数据的检测性能。

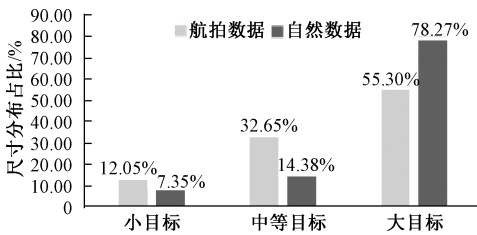


图 3 典型航拍数据的尺寸分布
Fig. 3 The distribution of sizes in typical aerial data

1.2 弱光条件下目标多模特性分析

受限与物理参数和成像方式，单一传感器在弱光照条件下的感知能力有限，多模载荷集成、融合成为有效的感知方式。图 4 展示了典型弱光条件下可见光和红外图像的视觉差异以及前景、背景的降维聚类结果。

从视觉效果分析，如图 4(a) 所示，可见光图像属于被动成像，分辨率较高，与人眼所获取的信息一致。然而，弱光照条件下图像亮度和对比度降低，纹理特征弱化，难以区分前景和背景。如图 4(b) 所示，红外成像属于主动成像，虽然分辨率较低，且缺乏纹理和颜色信息，但是红外图像在夜晚仍然可以获得目标辐射信息，突破了光照条件的限制，可有效区分前景和背景。从降维聚类结果分析，样本被分为前景和背景 2 个类簇，图 4(c) 中样本类内差异较大。图 4(d) 中样本类间差异大，类内差异小。这说明弱光照条件下可见光图像中目标分类难度较大，红外图像中前景和背景更容易被区分。因此，该方法采取双分支 YOLOv8 作为基础网络，耦合光照条件和对比度，知道可见光和红外特征进

行加权融合，提升弱光照条件下目标的表征和融合检测能力。

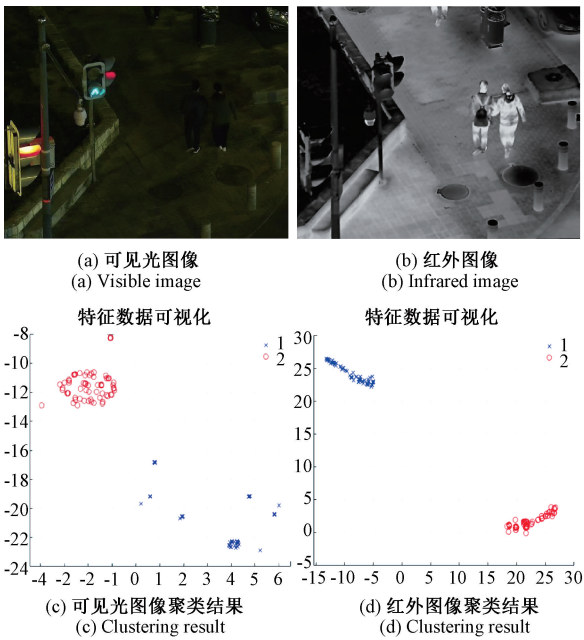


图 4 多模数据及其降维结果
Fig. 4 Multimodal data and its downscaling results

2 无人机多模目标检测系统设计

针对弱光照条件，本文基于旋翼无人机和视觉传感器，构建了无人机多模视觉检测系统。系统由无人机平台、多模吊舱、边缘计算模块和自组网电台构成^[17]。具体模块和数据流如图 5 所示。

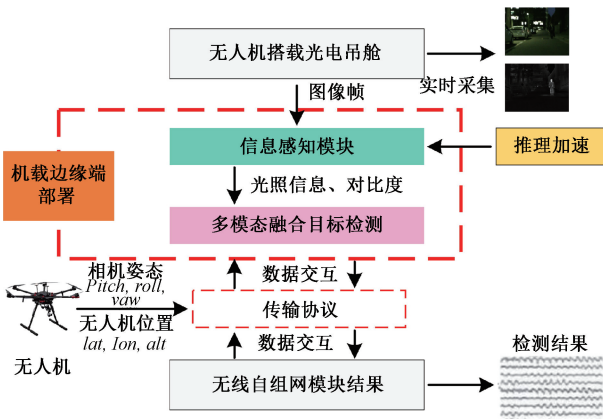


图 5 多模目标检测系统设计
Fig. 5 The design of multimodal object detection system

其中，目标检测算法可在机载边缘端部署，利用无人机采集的实时多模图像，得到目标的位置和类别。无人机机载光学相机为品灵 q30tirm plus 高精度三轴增稳吊舱，配有 213 万像素的 30 倍光学变焦索尼相机，25 mm

镜头 640 pixels×48 pixels 分辨率热成像仪以及 2 000 m 测距仪。为提升目标成像质量,本系统中的视觉伺服云台可利用目标的图像位置,控制云台旋转,将待检测目标保持在图像中心。同时,针对可见光和红外等多模数据,在通信交互上具有规范的传输协议和统一的任务管理机制。

3 数据集选取

可见光、红外融合检测通常应用于自动驾驶、遥感探测等领域,相比于常规检测,具有视觉差异显著、图像成对出现等特点。目前学术界常见的融合检测数据集有前视红外(forward looking infrared, FLIR)^[18]、遥感目标检测数据集(vehicle detection in aerial imagery, VEDAI)^[19]和 LLVIP 等,其中 FLIR 数据集以街道和公路为背景,VEDAI 数据集以田野、森林为背景,仅 LLVIP 数据集用于解决低照度场景的行人检测问题。经统计,LLVIP 数据集共包含 15 488 个图像对,其中 9 292 对图像用于训练,3 098 对图像分别用于验证和测试。与其他数据集相比,LLVIP 数据集中可见光和红外模态的图像在时间和空间上严格对齐,可以使网络更专注于融合检测性能的提升,而不用考虑校准带来的误差。因此,为更好地验证算法在弱光照条件下的检测性能,选择 LLVIP 作为基准数据集进行测试和验证。图 6 所示为 LLVIP 数据集的实例和宽、高等变量的分布情况,说明所用的数据集涵盖不同尺度的实例,具有较好的代表性。

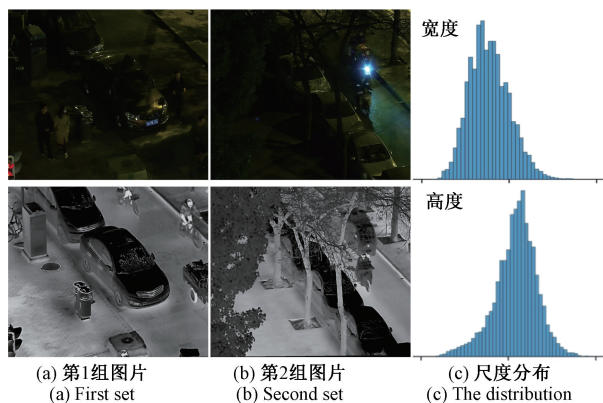


图 6 LLVIP 数据集及尺度分布

Fig. 6 The LLVIP dataset and scale distribution

4 耦合光照和对比度的多尺度差分注意力检测方法设计

4.1 信息感知模块

为缓解复杂融合结构带来的模型参数量过大、无法在边端部署等问题,现有方法主要通过并行计算光照条

件等先验信息来引导可见光和红外特征融合^[20-21]。尽管通过先验信息引导多模态特征深度融合的方式是有效的,但是在弱光照条件下,使用光照信息有以下局限性:1)基于照明信息引导的方法在白天等光照充足的场景下通常表现更优异,但在黄昏、黎明等弱光照条件下,照明信息无法量化,难以辅助网络进行判别;2)照明信息被用来指导不同模态特征的融合,但是对于模态内特征的增强是不足的。因此,使用耦合光照和对比度的方法感知来替代单一的光照信息,进而更深入地挖掘不同模态间的关联性。如图 7 所示,信息感知模块计算光照和对比度作为加权信息,从而更加高效地指导模态内和模态间特征的深度交叉融合。

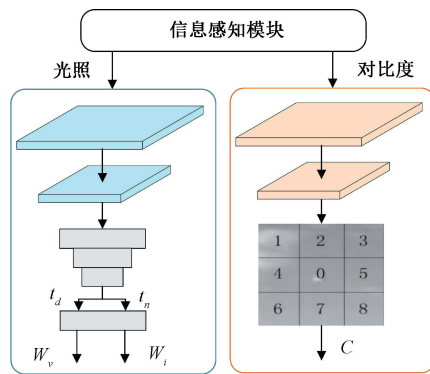


图 7 信息感知模块结构

Fig. 7 The architecture of information awareness module

由于谱段的的不同,可见光图像在白天质量更高,而红外图像则恰恰相反,在夜晚更加清晰,目标轮廓较为显著。如图 8 所示,记录了不同时刻的光照强度和该时刻主动融合权重的变化情况。随着光照强度的不断变化,可见光和红外模态的加权系数也在不断变化,且光照强度越高,可见光权重的占比越大。

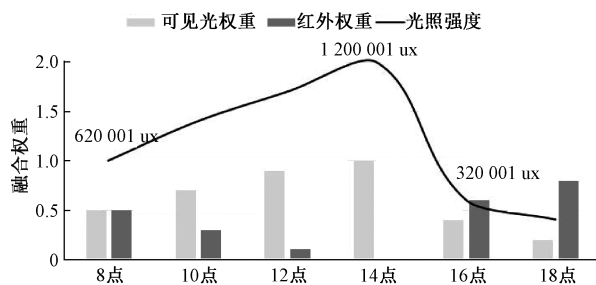


图 8 不同时刻融合权重占比情况

Fig. 8 Fusion weights at different moments

因此,在弱光条件下红外图像所占融合权重更高。为准确衡量不同模态图像融合时所占比的权重,避免因可见光图像占比较高影响弱光条件下的检测效果,设计了一个分类网络,网络的输出为 l_d 和 l_n ,分别代表图像的

亮光信息和暗光信息, $l_d + l_n = 1$ 。图像亮度越大则 l_d 越大, 融合时可见光模态占比的权重也越高。但是弱光条件下光源较弱, 计算所得的值大多会接近 0, 如果直接用此权重引导融合, 那么可见光模态的特征将极大概率被抑制。为优化权重, 设计了一个门函数来重新调整两种模态的权重, 门函数为:

$$W_v = \frac{l_d - l_n + 1}{2} \quad (1)$$

$$W_i = \frac{l_n - l_d + 1}{2} \quad (2)$$

其中, W_v 和 W_i 分别为可见光和红外模态深度融合时占比的权重, 两者之和为 1。弱光条件时, 则 $W_i > 1/2$ 。且由于 W_v 不再趋近于 0, 可以较充分地促进两种模态的融合, 反之亦然。具体来说, 光照信息预测网络由卷积层和全连接层组成。在卷积层之后, 添加激活层和 2×2 自适应池化。随后输入全连接层进行亮光和暗光信息, 将模型的输出转换为所需权重。

弱光条件下仅依靠光照信息是无法辅助网络分辨前景和背景。例如, 可见光图像中地物和背景在颜色上的相似性使目标难以分辨, 而红外图像的温度分布表征可以缓解因视觉特征弱化带来的检测性能的下降。因此, 信息感知模块在照明信息的基础上引入对比度这一先验知识对高相似度目标进行增强, 指导红外特征的融合。首先, 设置滑动窗口, 在图像上以固定的步长移动。如图 7 中右侧矩形框内容所示, 将每个窗口划分成 3×3 的网格, 并编号为 0~8。定义变量 m , 代表 1~8 号区域内所有点的像素均值, m 的定义如式(3)所示。

$$m_i = \frac{1}{N_c} \sum_{j=1}^{N_j} I_j^i \quad (3)$$

其中, i 代表区域的序号, 取值范围为 1~8, N_c 为除中心区域外的区域数量, N_j 代表每个区域内像素点的序号, I 代表像素点的灰度值。引入变量 L , L 代表 0 号区域中最大的像素点灰度值。用式(4)来定义对比度 C 。

$$C = \frac{1}{N_c} \sum_{i=1}^{N_c} \frac{L}{m_i} \quad (4)$$

其中, i 代表区域的序号, N_c 为除中心区域外的区域数量, 此处为 8。通过计算对比度, 网络可以在融合过程中进一步增强目标区域, 抑制地物背景, 弥补因光照条件无法量化对加权融合带来的影响。

4.2 多尺度差分注意力融合模块的设计

由于成像方式和通道数的不同, 可见光和红外模态特征存在本质上的差异。目前, 学界上主要通过双流卷积网络实现多模特征的提取和融合。然而, 大部分网络采用简单的融合策略, 无法充分发挥融合的优势。粗糙的连接也会增加模型收敛和拟合的难度, 从而导致性能

的下降。此外, 目标的多尺度特性也会影响检测的精度。因此, 如何稳定和高效地融合两种模态特征是该研究的关键。受差分放大器中差模信号被增强和共模信号被抑制这一特性的启发, 设计多尺度差分注意力融合模块, 该模块由差模注意力和共模注意力两部分组成。设可见光和红外模态输入为 M^R 和 M^T , 则差模特征 M^D 和共模特征 M^C 可以表示为:

$$M^D = \frac{M^C + M^D}{2} - \frac{M^C - M^D}{2} = M^R - M^T \quad (5)$$

$$M^C = \frac{M^C + M^D}{2} + \frac{M^C - M^D}{2} = M^R + M^T \quad (6)$$

由式(5)~(6)可知, 差模特征即模态间的差值, 旨在得到模态间的互补性特征; 共模特征即模态的和, 旨在得到模态间的一致性特征。由此, 多尺度差分注意力融合模块通过模态间的加减法定义了两个全新的特征, 进而提供更加丰富的表示。

1) 差模注意力

差模注意力的灵感来自于差分放大器中的差模信号, 通过计算可见光和红外模态的差值和双分支的通道加权来增强模态间的互补性特征。如图 9 所示, M^R 和 M^T 作为输入特征, 首先通过直接相加得到 M^D ; 其次, 如式(7)所示, M^D 通过全局平均池化得到向量 s_1 , 如式(8)所示, 通过全局最大池化得到向量 s_2 , 由此将差模特征编码到全局向量中。其次, 通过共享卷积将两分支相加得到通道注意力图 V_{DA} ; 随后, 将注意力图与两模态的输入特征相乘, 通过跳跃连接得到与输入模态维度相匹配的特征图。最后, 将增强后的特征按照信息感知模块得到的权重 W_v 和 W_i 按比例相加得到差模注意力输出 M_{DA} 。

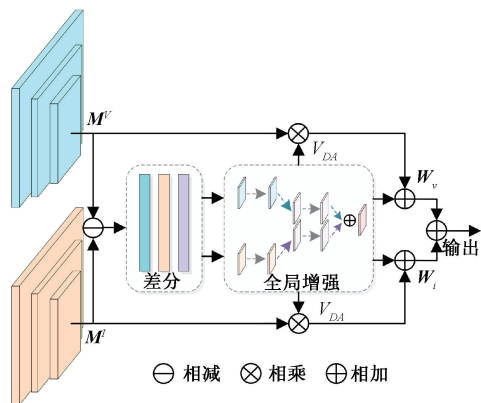


图9 差模注意力结构

Fig. 9 The structure of differential attention

$$s_1 = GAP(M^D) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W M^D \quad (7)$$

$$s_2 = GMP(M^D) = \max(M^D) \quad (8)$$

$$M_{DA} = W_v \cdot M^V \otimes (1 + V_{DA}) + W_i \cdot M^I \otimes (1 + V_{DA}) \quad (9)$$

其中, DA 表示差模注意力。为确保网络的轻量化,在池化后共享了两分支的网络卷积,减少了参数的冗余。通过模态相减、全局编码、共享卷积和加权融合,网络学习到可见光和红外模态不同通道间的关联性和重要性,增强了模态间的互补性特征。此外,为缓解网络在传递时导致的原始特征的衰减,在增强过程中采用了跳跃连接的方式,在得到模态间互补的辨别性特征的同时,保留了输入模态的原始特征。

2) 共模注意力

此外,模态内部噪声也会影响模型的检测性能,例如相似的外观和不确定的纹理,仅靠差模注意力是不够的。因此,借鉴差分放大器中的共模信号,设计了共模注意力,利用信息感知模块得到对比度权重,关注模态内的固有特征,抑制背景带来的噪声,引导网络学习模态内的特征。如图 10 所示,共模注意力的输入为可见光和红外特征。首先,通过直接相加获得共模特征 M^C ;其次,利用全局平均值池化,并与全连接层相连得到注意力图 V_{CA}^V 和 V_{CA}^I ;最后,分别将注意力图与输入特征相乘,其中,红外分支利用对比度进行加权,得到共模注意力输出 M_{CA} 。

因此,共模注意力的输出如式 (10) 所示。

$$M_{CA} = M^V \otimes V_{CA}^V + C \cdot M^I \otimes V_{CA}^I \quad (10)$$

通过求和、全局压缩和对比度加权,本模块实现了自

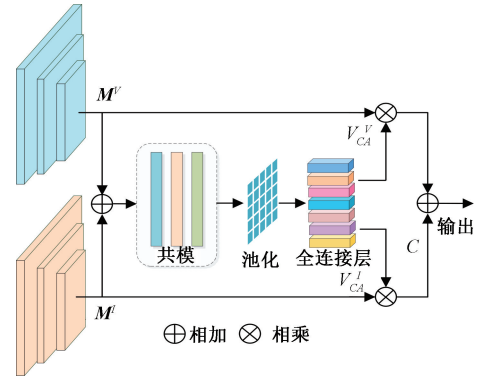


图 10 共模注意力结构

Fig. 10 The structure of common attention

适应的通道筛选。为避免参数冗余导致的推理速度下降,共模注意力通过信道筛选和参数共享减少了特征维度,提升了计算效率。此外,本模块同样使用了跳跃连接,能够复用浅层特征,提升对复杂场景表征能力。

3) 多尺度联合策略

将差模注意力和共模注意力交叉融合是本模块的最后一步。为确保模型的轻量化,避免注意力串行连接导致的模型速度减缓,针对弱光条件下目标的多尺度特性,设计了多尺度模块交叉并行的联合策略。如图 11 所示,以 YOLOv8 为骨干网络,通过大、中、小 3 种不同尺寸特征的交叉并行连接实现了可见光和红外模态的融合。

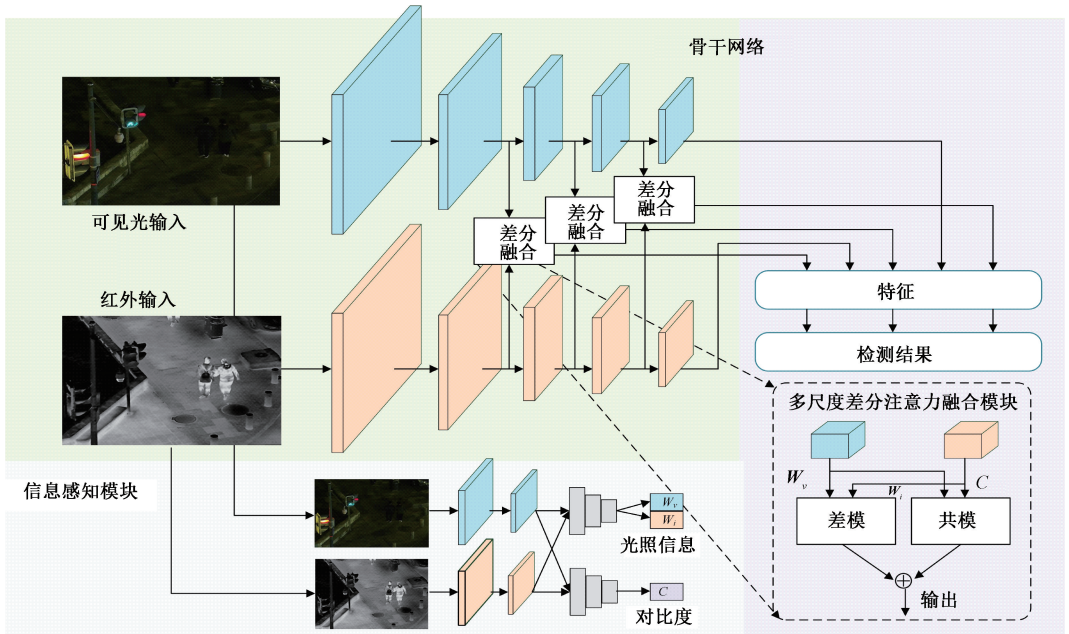


图 11 多尺度差分注意力融合框架

Fig. 11 The framework of multiscale differential attention fusion

图 11 联合步骤如式 (11) 所示。

$$F_{FUSE} = \sum_{i=1}^3 (M_{CA}^i + M_{DA}^i) \quad (11)$$

其中, i 的取值范围为 1, 2, 3, 分别代表大、中、小 3 种不同尺度的特征图。为保持全局的权重, 在每个分支后使用卷积将维度减小至初始值的 1/2, 然后进行双线性插值, 恢复到输入的长宽上。最后, 将不同尺度的特征拼接联合起来作为多尺度的全局特征。

4.3 损失函数

为自适应地捕捉模态间和模态内特征, 引导网络进行学习, 创新性地提出信息感知损失。信息感知损失由检测损失和感知损失组成。检测损失用于衡量检测结果, 感知条件损失用于校准模态内和模态间的加权融合权重。信息感知损失的定义如式 (12) 所示。

$$L = L_D + L_I \quad (12)$$

检测损失由类别、置信度和边框损失组成, 类别损失采用二元交叉熵方法, 通过计算预测和真实值的交叉熵和关联性, 从而衡量模型预测类别能力的准确性, 其定义如式 (13) 所示。本文将置信度损失定义为式 (14), 主要解决不同尺度目标比例不均衡的问题, 有助于模型性能的提升。边框损失采用的是 GIOU, GIOU 在原始 IoU 的基础上增加了一个惩罚项, 可以缓解检测框未重叠时出现的梯度问题, 其定义如式 (15)。

$$L_{cls} = -[y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \quad (13)$$

$$L_{conf} = -\alpha_i (1 - p_i)^\tau \log(p_i) \quad (14)$$

$$L_{reg} = 1 - IoU(B, A) + \frac{|C - (B \cup A)|}{|C|} \quad (15)$$

其中, i 为样本序号, p 为类别预测概率, y 是取值为 0, 1 的二值变量。 A 为真实框大小, B 为预测框大小, C 为同时包含两者最小框的大小。式 (14) 中 α 用于解决样本的不均衡问题, τ 用于难识别样本的不均衡问题, 在本研究中分别被设置为 0.25, 2。

本研究中, 特征的融合依赖于信息感知模块的指导。由于对比度通过局部窗口滑动计算获得, 且用于模态内特征增强, 不需要训练和收敛。而光照条件的计算与二分类器 (白天或黑夜) 相似, 判断图像属于白天的概率。与上文相似, 同样使用交叉熵损失来进行训练, 具体定义为:

$$L_I = -z \log \sigma(x) - (1 - z) \log(1 - \sigma(x)) \quad (16)$$

其中, z 为图像的光照标签, x 为白天的概率, σ 为激活函数, 可将概率归一化为 [0, 1]。为丰富光照条件的取值空间, z 被设置为 0, 0.5, 1.0, 分别表示黑夜、弱光、白天场景。

5 实验与分析

5.1 实验条件及评价指标

为确保算法性能评价的有效性, 将所有算法在

ubuntu 20.04 下进行测试, 算法框架采用 pytorch 1.9。由于算法收敛较快, 训练参数中迭代次数设为 100, 批处理大小设为 16, 学习率为 0.01。此外, 采用精确度 (precision, P)、召回率 (recall, R) 和 mAP 等公开指标来评估网络性能。

$$P = \frac{TP}{TP + FP} \quad (17)$$

$$R = \frac{TP}{TP + FN} \quad (18)$$

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i = \frac{1}{n} \sum_{i=1}^n P_i(r) dr \quad (19)$$

其中, TP 表示被正确分类的正样本数。否则, 它们会被分为 FP 。 FN 表示被错误分类的负样本数。 mAP_{50} 是交并比为 0.5 时检测精度的平均值。 mAP 是交并比为 0.5 : 0.95 时检测精度的平均值, 其阈值更高, 通常更具挑战性。

5.2 消融实验

首先, 探讨了不同融合策略对检测性能的影响, 并验证了多尺度融合策略的有效性。随后, 测试了信息感知模块和多尺度交叉注意力融合模块对模型的增益。

1) 不同融合策略对比

本部分重点讨论不同阶段的融合策略对检测性能的影响。不同融合方法的示意图如图 12 所示。早期融合也叫像素级融合, 最简单的方法可通过像素值的叠加来实现。然而, 简单融合可能会干扰初始特征, 导致性能的下滑, 甚至不如单模态检测的性能。与早期融合不同, 中期融合旨在通过不同分支特征间的交互, 来捕捉模态间的相关性, 从而提高模型性能。后期融合也叫决策级融合, 更加侧重于检测结果的融合。

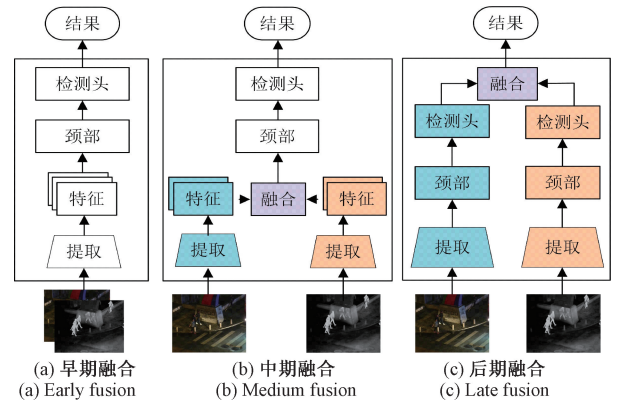


图 12 不同融合策略

Fig. 12 Different fusion strategies

因此, 以 3 种融合方式进行消融研究: 早期融合、中期融合和晚期融合。结果如表 1 所示。从加粗的字体中可以看出, 中期融合的检测性能最佳, 比可见光和红外单

模态检测的 mAP 值分别高出 19.2% 和 7.9%,后期融合方法的性能次之。原因在于,随着网络的前向传递,中期融合中的交叉连接可以保留两种不同层次的特征,而早期或后期融合无法保留保留浅层的细节信息和深层的语义信息,因此,不如中期融合的性能。此外,弱光条件下算法在红外图像上的性能明显优于在可见光图像上的性能,也反映出红外传感器在弱光条件下的优势。

表 1 不同融合策略的对比实验

Table 1 Comparison of different fusion strategies (%)

方法	骨干网络	mAP_{50}	mAP
可见光	Darknet53+CSP	90.8	50.0
红外	Darknet53+CSP	94.6	61.9
早期融合	Darknet53+CSP	95.2	67.9
中期融合(本文方法)	Darknet53+CSP	97.9	69.2
后期融合	Darknet53+CSP	96.3	68.7

图 13 展示了 120 次迭代下早期、中期、后期 3 种融合网络的收敛情况对比,所有网络的初始参数和设置均相同。由图可知,中期融合方法的收敛速度最快,且损失值最低。

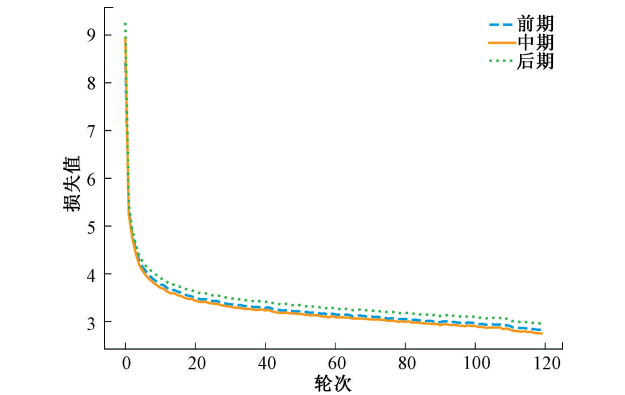


图 13 不同策略损失值对比

Fig. 13 The comparison of loss for different strategies

2) 不同模块性能对比

本文中,信息感知模块根据图像的背景环境,输出光照条件和对比度权重。多尺度差分注意力融合模块则包含差模注意力和共模注意力,在上述先验知识的引导下,实现 2 种模态的自适应融合。表 2 展示了使用信息感知模块(光照条件和对比度权重)前后和使用多尺度差分注意力模块(共模和差模注意力)前后网络的性能。“√”代表模型中使用了该方法。空白处代表模型中没有使用该方法。从粗体脚注中可以看出,使用信息感知模块后, mAP_{50} 和 mAP 分别提高了 3.3% 和 1.4%,超过了只使用光信息或只使用对比度时的网络性能。这说明

信息感知模块可以利用先验知识更好地指导网络进行融合,从而适应更复杂的背景。

表 2 不同方法的消融实验

Table 2 Comparison of different modules (%)

信息感知模块		多尺度差分注意力		mAP_{50}	mAP
光照条件	对比度	差模	共模		
				91.9	64.1
√				94.0	64.9
	√			92.7	64.4
√	√			95.2	65.5
		√		95.5	65.1
			√	94.9	64.7
		√	√	96.6	66.7
√	√	√	√	97.9	69.2

本研究的多尺度差分注意力模块由差模注意力和共模注意力组成,前者关注模态间的互补特征,后者关注模态内的固有特征。表 2 显示,整个模块将 mAP 提高了 2.6%,仅差模注意力将 mAP 提高了 1.0%,仅共模注意力将 mAP 提高了 0.6%。图 14 展示了特征的热力图可视化结果,图 14(b)、(c) 展示了单分支网络特征和融合特征的热力图。由图可知,单模态特征较为分散,而融合特征则聚焦在目标附近,说明通过本文方法后目标得到了显著增强。

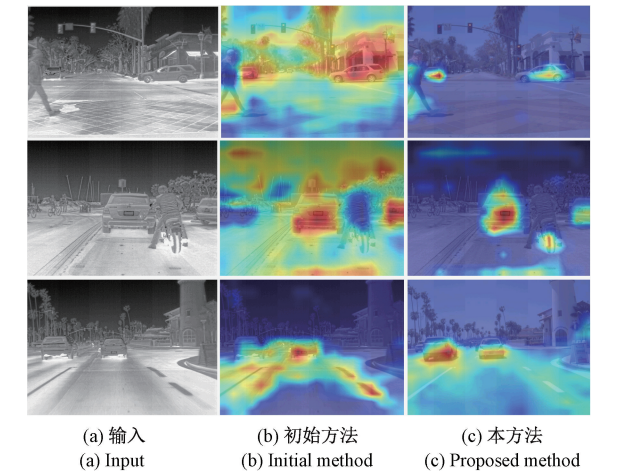


图 14 热力图分布

Fig. 14 The sidtribution of heat map

图 15 列出了信息感知模块在图像右上角预测的照明条件权重,白天场景下可见光模式的权重较高,夜间场景下红外模式的权重较高。在第 1 张图中,可见光权重可达 0.557。



图 15 光照条件计算结果

Fig. 15 Calculation results of light conditions

5.3 与其他方法对比实验

表 3 列举了本研究算法与其他算法在弱光场景下检测性能的对比结果。与其他算法相比,本研究算法在弱光条件数据集上具有最先进的性能。特别是与单分支网络 YOLOv9 和双分支网络 LRAF-Net 相比, mAP 还分别高出 7.9% 和 2.9%。此外,从前 4 行的结果可知,算法在红外图像中 mAP 明显高于可见光图像中的 mAP ,这表明弱光条件下红外图像通常表现更好。而本研究算法的检测性能(mAP ,69.2%)也显著优于单模态的检测性能。

表 3 检测性能对比结果

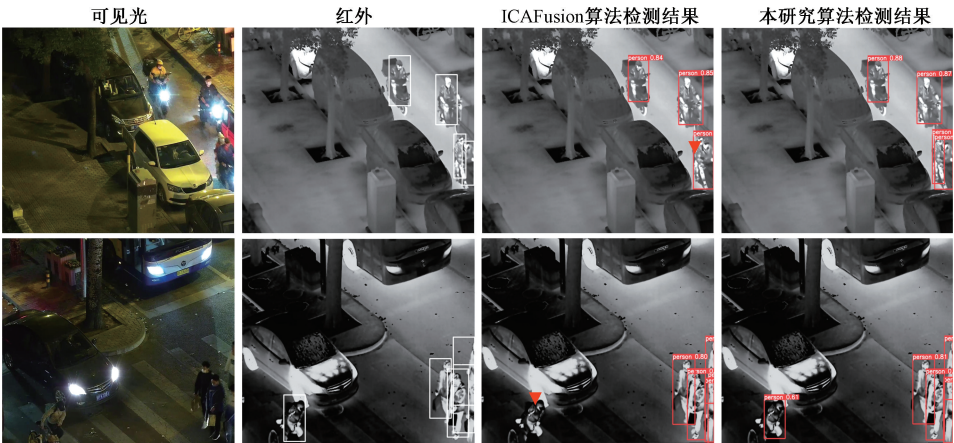
Table 3 Comparison of detection performance (%)				
模型	模态	基础网络	mAP_{50}	mAP
Faster R-CNN ^[22]	可见光	ResNet50	91.4	49.2
Faster R-CNN	红外	ResNet50	96.1	61.1
YOLOv9	可见光	CSPNet+ELAN	90.8	50.0
YOLOv9	红外	CSPNet+ELAN	94.6	61.9
CFT	双分支	CFB	97.5	63.6
LRAF-Net	双分支	DarkNet53	97.9	66.3
TFDet ^[23]	双分支	ResNet18	96.0	59.4
ICAFusion ^[24]	双分支	DarkNet53	97.8	66.1
Baseline	双分支	DarkNet53+CSP	95.9	65.3
本文	双分支	DarkNet53+CSP	97.9	69.2

为验证算法的实时性能,在分辨率大小为 512×512 的 LLVIP 数据集上对算法的推理时间进行了验证,结果如表 4 所示。所有实验设置相同,均在 NVIDIA RTX 3090 上完成。从表中可以看出,本研究算法运行速度略慢于典型的单分支网络 Faster R-CNN 和 YOLOv9,原因在于双分支网络的模型参数通常大于单分支网络。但是,本算法运行速度显著优于 CFT 和 TFDet,说明本研究设计的运行速度优于常见双分支算法。

表 4 推理时间对比结果

Table 4 Comparison of inference time			
方法	模态	基础网络	推理时间/fps
Faster R-CNN	单模态	ResNet50	21.7
YOLOv5	单模态	Darknet53	42.8
YOLOv9	单模态	CSPNet+ELAN	25.6
CFT	双分支	CFB	21.0
TFDet	双分支	ResNet18	7.7
本文方法	双分支	Darknet53+CSP	25.2

图 16 展示了典型的双流算法 ICAFusion 与本研究算法的检测结果。可以看出,弱光照条件下红外图像更能清晰地区分前景和背景。如图 16(a) 中第 2 行第 3 列图像左下角三角形处所示,当目标互相重叠或存在遮挡时,ICAFusion 算法会存在漏检。将图 16(a) 中第 1 行第 3 列图像右侧三角形与第 1 行第 1 列图像对照来看,弱光条件下其他光源的存在也会影响检测性能,导致目标的漏检。此外,算法还在其他弱光场景公开数据集上进行了测试。如图 16(b) 第 3 列所示,图中的车辆由于遮挡的原因被误检测到 2 次(三角形所在位置)。如图 16 第 4 列所示,为本研究算法的检测结果。由图可知,本研究提出的多光谱检测算法通过深度融合两种模态特征有效解决了上述漏检或误检问题,取得了更优的检测性能。



(a) 在LLVIP数据集上的检测结果
(a) Detection results on the LLVIP dataset



(b) 在其他弱光照数据上的检测结果
(b) Detection results on other weak light data

图 16 可视化检测结果

Fig. 16 Visualization of detection result

5.4 弱光条件下算法机载端部署验证

在上述基础上,在机载端对算法进行了轻量化部署,并对算法在真实弱光场景下的检测性能进行了初步验证。构建的旋翼无人机多模融合检测系统如图 17 所示。



图 17 本文搭建的无人机多模检测系统

Fig. 17 Multimodal detection system for UAVs in this paper

由图 17 可知,无人机平台上分别搭载了装有 RK3588 拓展板的边缘计算平台、多模光电吊舱、自组网电台和无人机姿态及定位设备,机载数据的采集、处理和交互则通过交换机实现^[25]。拓展板实际尺寸为 15 cm×11 cm,聚类小型无人机载能力。机载端部署验证时,有两点值得注意:1)多路传感器的数据在访问和缓存时可能因延迟、访问速度等原因无法保持同步;2)机载端两路数据采集、融合检测等任务在对同一内存读取访问时可能存在冲突,导致硬件计算资源的浪费。因此,在部署验证时,首先通过网络时间协议实现不同数据流之间的时间同步,而后设计无人机节点上的多线程计算任务,包括数据解码、数据读取、多模融合检测等,通过互锁结构实现多线程优化。设计的多线程资源互锁结构如图 18 所示。例如,解码线程在接收解码前访问互锁变量,当期空闲时进行数据的硬解码,在完成后释放互锁资源供读取线程使用。进一步列举了 CPU 资源占用率,如图 19 所示。随着线程数的增加,CPU 占用率增加较快,考虑到线程数为 3 时,占用率已达到 260%,本算法在实际部署时设置线程数量为 3。

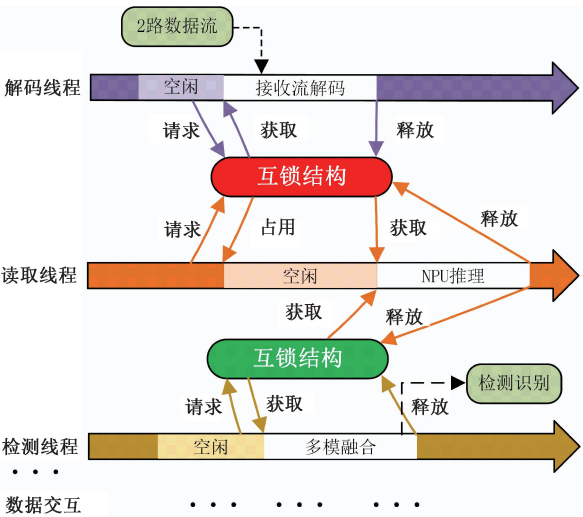


图 18 多线程互锁结构设计

Fig. 18 Multi-thread interlock structure

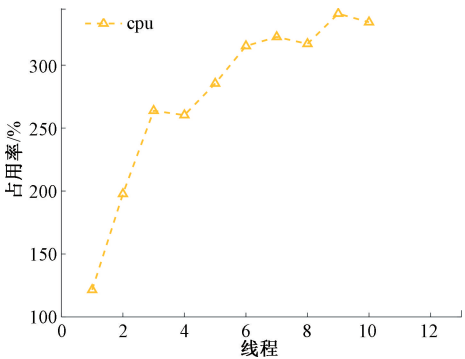


图 19 CPU 占用率分析

Fig. 19 The analysis of CPU usage

无人机多模融合检测结果如图 20 所示,图左侧为红外图像检测结果,右侧为可见光图像检测结果。由图可知,弱光照条件下可见光图像中目标模糊不清楚,难以分辨,而本研究算法目标具有较好的检测结果,并能够实时输出,结果表明,本文算法在机载端的平均运行效率达到 21.2 FPS,满足机载端应用需求。



图 20 多模融合检测结果

Fig. 20 Multimodal fusion detection results

6 结 论

弱光条件下,图像亮度低、对比度弱,成像质量差,极大地影响了无人机的探测精度。因此,无人机弱光照条件下目标检测方法具有重要的研究意义和应用价值。针对此问题,提出了一种耦合光照和对比度的多尺度差分注意力检测方法,通过信息感知模块计算得到的光照和对比度权重,来引导多尺度差分注意力模块对可见光和红外模态进行深度融合,以提升弱光照条件下网络的特征提取能力和目标识别能力。其次,构建了旋翼无人机多模目标检测系统,并使用 LLVIP 作为实验部分的基准数据集。最后,进行了大量的对比实验,在真实弱光照场景下对车辆等目标进行了验证。结果表明,本研究方法能够显著提升弱光照条件下无人机对目标的检测性能,并具备较好的实时性。未来的研究一方面考虑多视角或遮挡情况下目标检测性能的提升,另一方面考虑多机间的融合检测应用。此外,无人机弱光照条件下目标检测数据集的构建和开源也是本文下一步努力的方向。

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] 孙备,党昭洋,吴鹏,等. 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[J]. 仪器仪表学报, 2023,44(6): 54-65.
- SUN B, DANG ZH Y, WU P, et al. Multi-scale intersecting attention improved localization method for ground camouflage target detection by single UAV[J]. Chinese Journal of Scientific Instrument, 2023,44(6): 54-65.
- [3] 韩浩,何赞泽,杜旭,等. 基于偏振信息图像增强的多目标检测[J]. 电子测量与仪器学报,2023,37(3): 29-38.
- HAN H, HE Y Z, DU X, et al. Multi-target detection based on polarization-informed image enhancement[J]. Journal of Electronic Measurement and Instrumentation, 2023,37(3): 29-38.
- [4] 范馨月,张阔,张干,等. 细微特征增强的多级联合聚类跨模态行人重识别算法[J]. 电子测量与仪器学报, 2024,38(3): 94-103.
- FAN X Y, ZHANG K, ZHANG G, et al. A multilevel joint clustering cross-modal pedestrian re-identification algorithm with fine feature enhancement[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(3): 94-103.
- [5] 陈清江,顾媛. 多通道融合注意力网络的低照度图像增强[J]. 光学精密工程,2023,31(14): 2111-2122.
- CHEN Q J, GU Y. Multi-channel fusion attention network for low-light image enhancement[J]. Optics and Precision Engineering, 2023,31(14): 2111-2122.
- [6] 闫晓阳,王华珂,侯兴松,等. YCbCr 空间分治的双分支低照度图像增强网络[J]. 中国图象图形学报, 2023,28(11): 3415-3427.
- YAN X Y, WANG H K, HOU X S, et al. YCbCr spatial partitioning for two-branch low illumination image enhancement network[J]. Journal of Image and Graphics, 2023,28(11): 3415-3427.
- [7] FANG Q Y, HAN D P, WANG ZH K. Cross-modality fusion transformer for multispectral object detection[J]. ArXiv preprint arXiv:2111.00273, 2021.
- [8] 郝博,谷继明,刘力维. 基于 BF-YOLOv5 的红外及可见光图像融合的目标检测[J]. 光电与控制,2024, 31(5): 72-76.
- HAO B, GU J M, LIU L W. Target detection based on

- infrared and visible image fusion with BF-YOLOv5[J]. *Electronics Optics & Control*, 2024,31(5):72-76.
- [9] 王昱婷,刘志明,万亚平,等. 基于可见光与红外图像的弱光条件下目标检测[J]. *计算机工程*, 2024, 50(8):270-281.
- WANG Y T, LIU ZH M, WAN Y P, et al. Target detection in low light conditions based on visible and infrared images[J]. *Computer Engineering*, 2024, 50(8):270-281.
- [10] GUAN D Y, CAO Y P, YANG J X, et al. Fusion of multispectral data through illumination-aware deep neural networks for pedestrian detection [J]. *Information Fusion*, 2019, 50(1):148-157.
- [11] SONG K CH, BAO Y Q, WANG H, et al. A potential vision-based measurements technology: Information flow fusion detection method using RGB-thermal infrared images[J]. *IEEE Transactions on Instrumentation and Measurement*, 2023, 72:5004813.
- [12] 毕程程,黄妙华,刘若璿,等. 基于双模态图像关联式融合的行人实时检测[J]. *激光与光电子学进展*, 2024,61(8):463-470.
- BI CH CH, HUANG M H, LIU R Y, et al. Real-time pedestrian detection based on associative fusion of bimodal images[J]. *Advances in Laser and Optoelectronics*, 2024,61(8):463-470.
- [13] FU H L, WANG S X, DUAN P H, et al. LRAF-Net: Long-range attention fusion network for visible-infrared object detection [J]. *IEEE Transaction on Neural Network and Learning System*, 2023, 35(10):1-14.
- [14] 李淦,牛洛第,陈路,等. 融合视觉特征增强机制的机器人弱光环境抓取检测[J]. *计算机应用*, 2023, 43(8):2564-2571.
- LI G, NIU M D, CHEN L, et al. Fusion of visual feature enhancement mechanism for robot grasping detection in low light environment [J]. *Computer Applications*, 2023,43(8):2564-2571.
- [15] 苏雨蕾,黄丹丹,刘智,等. 无人机对地小目标检测方法研究[J]. *电子测量与仪器学报*, 2024,38(9):144-154.
- SU Y L, HUANG D D, LIU ZH, et al. Research on ground small target detection method for UAV [J]. *Journal of Electronic Measurement and Instrumentation*, 2024,38(9):144-154.
- [16] 罗会兰,陈鸿坤. 基于深度学习的目标检测研究综述[J]. *电子学报*, 2020, 48(6):1230-1239.
- LUO H L, CHEN H K. Survey of object detection based on deep learning [J]. *Acta Electronica Sinica*, 2020, 48(6):1230-1239.
- [17] 孙备,孙晓永,钱翰翔,等. 动态大视角场景融合帧间信息与模板匹配的低慢小无人机目标检测[J]. *仪器仪表学报*, 2024,45(7):64-74.
- SUN B, SUN X Y, QIAN H X, et al. Dynamic large-view scene fusion of inter-frame information and template matching for low-slow-small UAV target detection [J]. *Chinese Journal of Scientific Instrument*, 2024, 45(7):64-74.
- [18] ZHU Y H, SUN X Y, WANG M, et al. Multi-modal feature pyramid transformer for RGB-infrared object detection[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2023, 24(9):9984-9995.
- [19] MANISH S, MAYUR D, SRIVALLABHA K, et al. YOLOrs: Object detection in multimodal remote sensing imagery[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021, 14:1497-1508.
- [20] ZHANG Y, YU H, HE Y J, et al. Illumination-guided RGBT object detection with inter-and intra-modality fusion [J]. *IEEE Transactions on Instrumentation and Measurement*, 2023, 72:2508013.
- [21] TANG L F, YUAN J T, ZHANG H, et al. PIAFusion: A progressive infrared and visible image fusion network based on illumination aware [J]. *Information Fusion*, 2022, 83:79-92.
- [22] LI CH Y, SONG D, TONG R F, et al. Illumination-aware faster R-CNN for robust multispectral pedestrian detection[J]. *Pattern Recognition*, 2019, 85:161-171.
- [23] ZHANG X, ZHANG X H, WANG J T, et al. TFDet: Target-aware fusion for RGB-T pedestrian detection[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2024:3443455.
- [24] SHEN J F, CHEN Y F, LIU Y, et al. ICAFusion: Iterative cross-attention guided feature fusion for multispectral object detection[J]. *Pattern Recognition*, 2023, 145:109-113.
- [25] 石欣,卢灏,秦鹏杰,等. 一种远距离行人小目标检测方法[J]. *仪器仪表学报*, 2022,43(5):136-146.

SHI X, LU H, QIN P J, et al. A long-distance pedestrian small target detection method [J]. Chinese Journal of Scientific Instrument, 2022,43(5):136-146.

作者简介



郭润泽, 分别于 2018 年、2021 年于国防科技大学获得学士学位、硕士学位, 现为国防科技大学博士研究生, 主要研究方向为智能侦测与对抗。
E-mail: guorunze14@ nudt. edu. cn

Guo Runze received his B.Sc. degree and M.Sc. degree degrees from National University of Defense Technology in 2018 and 2021, respectively. He is currently a doctoral student at National University of Defense Technology, with a main research

focus on intelligent detection and countermeasures.



孙备(通信作者), 于 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与对抗。
E-mail: sunbei08@ nudt. edu. cn

Sun Bei (Corresponding author) received his Ph. D. degrees from National University of Defense Technology in 2018. He is currently an associate researcher and graduate advisor at National University of Defense Technology. His main research interest includes intelligent detection and countermeasure of unmanned platforms.