文章编号:1001-9014(XXXX)XX-0001-12

DOI:10.11972/j.issn.1001-9014.XXXX.XX.001

红外图像暗弱目标轻量级检测网络

李朝旭[1], 徐清宇[1], 安 玮[1*], 贺 旭[1], 郭高伟[1], 李 淼[1*], 凌 强[1], 王龙光[2], 肖 超[1], 林再平[1]

(1. 国防科技大学 电子科学学院,湖南 长沙 410073;
2. 空军航空大学,吉林 长春 130000)

摘要:弱小目标检测一直是红外图像处理领域的经典问题,通常所关注的弱小目标在亮度上高于所在的局部背景。然而在一些场景下,目标辐射能量会弱于背景,如在高空中巡航的民航飞机,由于机身蒙皮温度低于地表,在中等空间分辨率的热红外卫星图像上呈现为暗弱点目标。针对暗弱目标形态特征少、现有目标检测网络结构冗余的问题,提出了一种基于可形变注意力机制的极轻量级暗弱目标单帧检测网络AirFormer,参数量仅为37.1 K,在256×256尺寸的图像上浮点运算次数仅有46.2 M。此外,针对当前红外图像暗弱目标检测数据集缺乏的问题,作者通过对热红外卫星图像民航飞机的特性进行分析,提出了一种中等空间分辨率热红外卫星图像民航飞机的简易仿真方法,并以民航飞机为仿真对象构建了红外图像暗弱目标检测数据集——IRAir数据集。在IRAir数据集上进行验证,所提的AirFormer网络对暗弱点目标的召回率可达71.0%,检测准确率可达82.6%。此外,基于仿真数据训练,AirFormer可有效检出热红外卫星图像上真实的民航飞机。

关 键 词:人工智能;红外目标检测;轻量化网络;暗弱目标检测

中图分类号:TP753 **文献标识码**:A

A lightweight dark object detection network for infrared images

LI Zhao-Xu[1], XU Qing-Xu[1], AN Wei[1*], HE Xu[1], GUO Gao-Wei[1], LI Miao[1*], LING Qiang[1], WANG Long-Guang[2], XIAO Chao[1], LIN Zai-Ping[1]

(1. College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China;
2. Aviation University of Air Force, Changchun 130000, China)

Abstract: Small target detection has been a classic research topic in the field of infrared image processing, and the objects are usually brighter than the local background. However, in some scenarios, the target brightness may be lower than the background brightness. For example, the civil airplanes usually have low-temperature skin when cruising, appearing as dark points on medium spatial resolution thermal infrared satellite images. There are few features of these objects, so the current detection networks are redundant. Hence, we proposed a lightweight dark object detection network, AirFormer. It only has 37.1 K parameters and 46.2 M floating-point operations on a 256×256 image. Considering the lack of infrared dark object detection dataset, the authors analyzed the characteristics of airplanes on thermal infrared satellite images, and then developed a simulated flying aircraft detection dataset called IRAir. AirFormer achieves 71.0% at recall and 82.6% at detection precision on the IRAir dataset. In addition, after training on simulated data, AirFormer has achieved detection of real flying airplanes on the thermal infrared satellite images.

Key words: artificial intelligence, infrared target detection, lightweight network, dark object detection

收稿日期:2024-05-08,**修回日期**:2024-07-12 **Received date**:2024-05-08,**revised date**:2024-07-12

基金项目:湖南省研究生科研创新项目(QL20230012, CX20240120);国防科技大学自主创新科学基金项目(22-ZZCX-042);国家自然科学基金创新群体项目(61921001);国家自然科学基金项目(62401591, 62401589);中国博士后科学基金项目(GZB20230982, 2023M744321)

Foundation items: Supported by the Hunan Provincial Innovation Foundation For Postgraduate (QL20230012, CX20240120), the Science Technology Innovation Program of National Defense University (22-ZZCX-042), Innovative Research Groups of the National Natural Science Foundation of China (61921001), National Natural Science Foundation of China (62401591, 62401589), China Postdoctoral Science Foundation (GZB20230982, 2023M74432)

作者简介(Biography):李朝旭(1996-),男,山东青岛人,博士研究生,主要研究领域为空间信息处理、天基光学弱小目标检测. E-mail: lizhaoxu@nudt.edu.cn

[*] **通讯作者**(Corresponding authors): E-mail: anwei@nudt.edu.cn, lm8866@nudt.edu.cn

引言

随着全球经济规模不断扩大，民用航空作为最快捷的长途交通运输方式，将在未来几十年内持续保持稳定的增长态势。据国际航空运输协会预测，2024年全球航班将超4千万班次，客运量将突破47亿人次[1]。随着民用航空业持续繁荣，民航机队规模不断增大，全球民航活动实时监测重要性日益凸显。一方面，获得民航飞机实时位置，有助于及时应对空中突发情况，缩小航空事故发生后的搜救范围，提高救援效率；另一方面，航空业碳排放量占全球碳排放总量的2%~3%，掌握民航活动的时空分布情况，有助于优化航线设计，节约航空燃料，可推动民航业绿色、低碳、循环发展，助力于我国2030年前碳达峰目标。

当前民航交通管控主要依赖地基雷达（工作距离通常为几百公里）和自动相关监视系统（ADS-B、ADS-C），对于海洋区域上空非合作状态下的飞机缺乏有效实时监控手段。而卫星遥感图像可以提供全球范围的地理影像，通过图像检测算法可以获取民航位置信息，可以为民航交通管制提供补充信息。中等空间分辨率卫星影像（通常为空间分辨率在5~100 m的卫星图像）可以对地表实现较宽的幅宽覆盖，同时空中民航目标在图像上仍然能呈现一定的信息。近年来，一些学者在中等空间分辨率卫星影像上进行了空中民航目标检测的探索。Zhao[2]等人在Landsat8卫星OLI传感器拍摄的1.3 μm水汽强吸收通道图像（空间分辨率为30 m）上利用幅值差异特性实现了空中民航飞机的检测。Liu[3]等人利用Sentinel-2卫星推扫式传感器采集的多光谱卫星图像异常成像的视差特性进行空中民航目标检测。然而，上述工作研究对象为可见光至短波红外波段的卫星影像，不能在夜间进行空中民航的监控。长波红外载荷主要捕获地表物体自身所发出的电磁辐射能量，可在无外界光源（如太阳）的情况下进行成像。当前民航业所运营民航均为亚音速飞机，在亚音速巡航状态下，飞机表面蒙皮温度受周围大气温度影响，通常在-25 ~ 30 ℃[4]，地表温度会高于亚音速民航飞机蒙皮温度，因此亚音速巡航中的民航被长波红外载荷观测为暗目标。“可持续发展科学卫星1号”（SDG卫星1号）[5]热红外载荷采集的三个长波波段影像的空间分辨率为30 m，与民航飞机尺寸接近，因此空中民航目标在其上呈现为暗点状（如图1所示）。

当前红外小目标检测技术通常关注于自身能量高于背景的亮目标，对长波红外卫星图像在飞民航等暗弱目标的研究较少。现有技术可分为基于模型驱动的方法和基于数据驱动的方法。基于模型驱动的方法主要集中在基于图像滤波和基于稀疏低秩重建两类方法。基于图像滤波方法利用目标的亮度异常性以及背景区域灰度缓变性来实现对背景区域的抑制[6]。这类算法所需的计算资源较少，被广泛应用于实际工程中。基于稀疏低秩重建方法假设缓变背景具有低秩性，小目标具有稀疏性，利用目标稀疏性和背景低秩性对图像重建进而实现目标与背景的分离[7,8]。上述工作可在简单背景的红外图像上实现较好的检测效果，但是依赖于对目标的先验认知对关键参数进行人工设置，难以适配复杂场景，容易产生虚警和漏警。以深度神经网络为核心的基于数据驱动的方法成为近年来红外小目标检测领域新的研究热点，可以从大量数据中自动学习有效特征，避免了特征模板人工设置，对复杂场景具备一定的适应性和泛化性。然而，当前这类红外小目标检测方法普遍认为像素级目标标签有助于神经网络学习小目标的形状特征，因此将红外小目标检测任务建模为语义分割问题，使用分割网络来在红外图像上实现亮目标与背景的分离[9~10]。然而，对于SDG卫星1号长波红外图像民航目标而言，一些目标像素与背景像素灰度值接近，难以获得精细的逐像素分割标签。Li[5]等人在SDG卫星1号三波段长波红外图像利用了尾迹云特征检测出了部分民航目标，构建了带有框标签的三波段长波红外卫星图像民航目标数据集，并验证了现有RGB图像目标检测网络用于检测红外图像暗弱目标的可行性。然而现有RGB图像目标检测网络针对尺寸较大目标设计，由视觉基础模型和检测头模块两部分组成。视觉基础模型需要使用大量的卷积计算或者注意力计算等操作将RGB图像转化为反映目标纹理和形状特性的特征图，检测头模块依据视觉基础模型输出的特征图实现目标的定位和分类。然而，民航飞机在长波红外卫星图像上的像素数量极少，形状和纹理特征较少，现有RGB图像目标检测网络的网络结构和参数规模冗余性较大，不利于在成像卫星等计算资源有限的边缘端侧进行实时检测。因此，如何针对长波红外卫星图像在飞民航等暗弱目标的特点实现低计算资源消耗的目标检测需要进一步深入研究。

此外,基于数据驱动的目标检测方法需要大量有标注样本进行学习,但是红外图像暗弱目标尺寸极小、灰度与背景接近,部分目标不能占据一个完整像素,人眼难以辨识,人工标注工作量较大且易产生错标、漏标。Li[5]等人在SDG卫星1号红外图像上利用民航飞机自身产生的尾迹云实现了对民航目标的初步自动定位并构建了首个长波红外卫星图像民航目标检测数据集。然而,民航飞机尾迹云的产生需要满足一定的气象条件,利用尾迹云特征仅能实现部分民航目标的定位。此外,文献[5]所创建的数据集在尾迹云初定位的基础上依然需要人眼判读,因此所提供的目标其尾迹明显、与背景灰度差异较大。标注精确的数据集是基于数据驱动的检测技术发展的基础,而长波红外卫星图像上的部分民航飞机尺寸较小、可辨识度低,增大了精确数据集的构建难度,进而阻碍了检测技术的发展。

针对上述红外图像暗弱目标检测方法和数据两方面问题,本文分别进行了相应的探索。在检测方法上,本文提出了一种非卷积结构的极轻量级单帧图像暗弱目标检测网络,模型参数量仅为37.1 K,在256×256尺寸的图像上浮点运算次数仅为46.2 M。针对真实暗弱目标数据集缺乏、弱目标人工标注难的问题,本文提出了一种简单易行的暗弱目标仿真方法,并在卫星热红外图像上进行验证,可生成与空中民航飞机高度相似的仿真暗弱目标。以实测的长波红外(Infrared)卫星图像为背景图像,以巡航态民航飞机(Airplanes)为仿真目标,本文构建了红外图像暗弱目标检测仿真数据集,将其命名为IRAir数据集。考虑到单帧单通道图像所提供信息难以去除与民航飞机具有相似形态的虚警,IRAir数据集仿真了2 000段长波红外图像序列,可用于运动暗弱目标检测方法的研究与开发。本数据集是领域内第二个热红外卫星图像民航目标检测数据,也是首个热红外卫星视频民航目标检测数据集。此外,本文利用IRAir数据集提供的仿真目标训练集对所提的单帧暗弱目标检测网络进行训练,进而在包含真实民航目标的SDG卫星1号长波红外图像进行推理,发现基于仿真数据集训练的检测模型对真实民航目标仍具有较好的检测性能,验证了本文所提方法和所仿数据集的有效性。

1　极轻量级暗弱目标检测网络

如图1所示,在SDG卫星1号热红外图像上,民航飞机尺寸极小,有显著异常性的暗像素集中在2×2窗口范围内,而灰度相关的像素基本在7×7窗口范围内。而现有通用目标检测网络是针对人类视觉观测的有形状目标进行设计。如在计算机视觉领域中常用的COCO数据集[11]上,一些目标覆盖整个

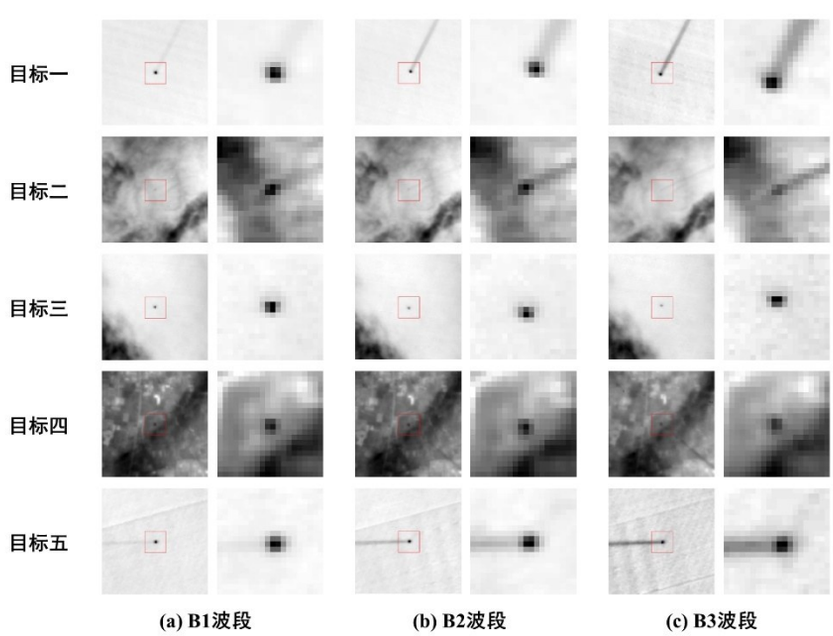


图1　SDG卫星1号实测空中民航飞机的热红外图像:(a)8~10.5 μm;(b)10.3~11.3 μm;(c)11.5~12.5 μm

Fig. 1　The thermal infrared images of real civial airplanes capured by SDGSAT-1: (a) 8~10.5 μm; (b) 10.3~11.3 μm; (c) 11.5~12.5 μm

图像范围，而另一些目标则只占据较小的空间范围。因此，通用目标检测网络通常需要使用基于卷积结构或者视觉Transformer结构的视觉基础模型来从图像中提取深层特征，并且需要利用多尺度特征图来兼顾不同尺寸目标的检测性能。而对于暗弱目标检测任务来说，只需要关注较小空间范围内的信息即可，不需要进行跨较大图像空间尺度的特征提取。

近年来，基于Transformer[12-17]架构的深度学习模型在自然语言处理领域和图像处理领域取得了一系列突破性进展。其中，Detection Transformer（DETR）[13]是首个基于Transformer架构的目标检测网络，将目标检测转换为集合预测问题，为目标检测提供了一种全新的范式。Deformable DETR[14]将可形变卷积和自注意力机制向结合实现了可形变注意力（Deformable Attention）机制，显著降低了注意力计算量，进而实现多尺寸特征的聚合。DAB-DETR[15]将解码器阶段的目标查询中位置信息部分显式建模为锚框，在DETR结构中重新引入了锚框机制。SpecDETR[16]在高光谱图像弱小目标检测任务中移除了视觉基础模型，仅使用Transformer编码器-解码器结构实现了较佳的检测效果。遵循Transformer在自然语言处理任务中数据处理方式，这些DETR类检测网络对骨干网络输入的特征图上各空间位置上的特征向量进行注意力计算处理。本文认为这种特征提取机制比卷积操作在像素尺度上有更精细的特征提取能力。然而，在视觉基础模型之外，DETR类检测网络通常还包含多层Transformer编码器层和多层Transformer解码器层，参数规模庞大，计算资源需求大。而经过上文分析，暗弱目标有相关性的像素较少，针对通用检测任务设计的网络结构有较大冗余性。受当前DTER发展的启发，本文提出了一种基于可形变注意力机制的极轻量级暗弱目标检测网络，将其命名为AirFormer。

图2为所提的AirFormer网络结构示意图。AirFormer网络结构简单，仅由可形变注意力模块、位置预测模块和置信度预测模块构成。AirFormer的输入图像为单帧单波段图像$\boldsymbol{I} \in \mathbb{R}^{H\times W\times 1}$，其中$H$和$W$分别为图像长和宽。为了使网络能有充足的特征表征空间，需对图像通道数进行扩充：

$$\boldsymbol{F} = \mathrm{linear}(\boldsymbol{I}) \quad , \quad (1)$$

其中linear表示线性层，$\boldsymbol{F} \in \mathbb{R}^{H\times W\times C}$为扩维后的图像。由于单帧图像上暗弱目标形状特征极少，再使用骨干网络提取图像深层特征意义不大，因此本文将扩维图像$\boldsymbol{F}$直接视为网络特征图。在自然语言处理领域，Transformer网络处理的基本单元为表征单词信息的一维特征向量，被称为token，译作“词元”或者“令牌”。在当前DETR类检测网络中，token被视为特征图上各像素上的一维特征向量。借鉴于遥感领域中的“像元”概念，本文将特征图$\boldsymbol{F}$上各特征向量称为特征像元。

相比于通用检测任务所处理的目标，暗弱目标

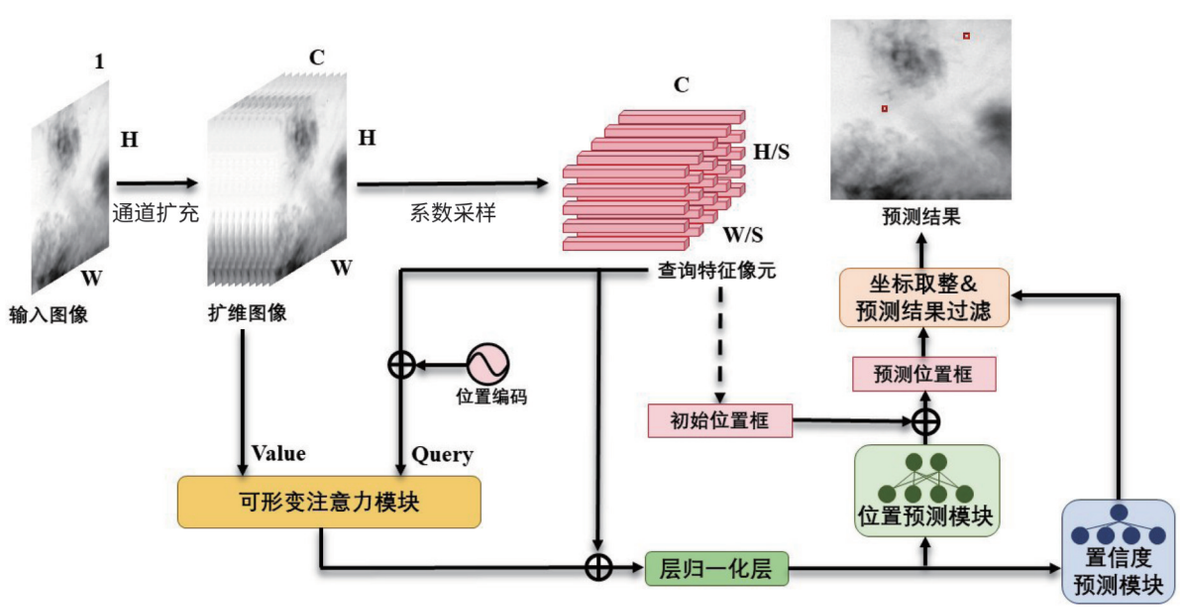


图2 AirFormer网络结构示意图

Fig. 2 Schematic diagram of AirFormer network structure

形态较为简单,网络所需要关注空间范围较小,因此使用一层线性层得到的特征图$\boldsymbol{F}$可以支撑后续目标定位和识别。Deformable DETR采用了一种局部空间稀疏采样的注意力机制,对每一个像素位置只关注邻域内有限数量的采样点。本文对Deformable DETR单层编码器结构进行简化,直接在特征图$\boldsymbol{F}$上利用简化后的可形变注意力模块提取各特征像元的邻域空间特征,再利用所提取的特征来预测该特征像元存在目标的可能性以及目标的位置框。具体而言,对选定的特征像元$\boldsymbol{f}_{i,j}\in\mathbb{R}^C$,将其视为查询特征像元,然后使用正弦编码计算出对应的位置编码$\boldsymbol{p}_{i,j}\in\mathbb{R}^C$,进而得到嵌入特征像元位置信息的query向量$\boldsymbol{q}_{i,j}\in\mathbb{R}^c$:

$$\boldsymbol{q}_{ij}=\mathrm{linear}\left(\boldsymbol{f}_{ij}+\boldsymbol{p}_{ij}\right)\quad,\quad(2)$$

其中,i表示垂直位置,j表示水平位置,$c<C$。此外,线性层对query向量进行降维,来降低后续采样点位置映射的计算量。利用简化后的可形变注意力模块,计算特征像元$\boldsymbol{f}_{i,j}$对应的注意力特征向量:

$$\boldsymbol{a}_{ij}=\mathrm{linear}\left[\sum_{m=1}^{M}\sum_{k=1}^{K}A_{m,k}g\left(\boldsymbol{F},\Delta i_{m,k},\Delta j_{m,k}\right)\right]\quad,\quad(3)$$

其中,$\boldsymbol{a}_{i,j}\in\mathbb{R}^C$,$M$为注意力模块的多头数目,$K$为各注意力头的采样点数目,$m$和$k$为注意力头和采样点的索引。$A_{m,k}$表示各采样点的注意力权重,为可学习变量,范围为$[0,1]$,满足$\sum_{k=1}^{K}A_{m,k}=1$。$\Delta i_{m,k}$和$\Delta j_{m,k}$表示采样点相对与$\boldsymbol{f}_{i,j}$的位置偏移量,而$g\left(\boldsymbol{F},\Delta i_{m,k},\Delta j_{m,k}\right)$表示使用双线性插值从特征图$\boldsymbol{F}$采样得到的特征像元。$\Delta i_{m,k}$和$\Delta j_{m,k}$均是由$\boldsymbol{q}_{i,j}$经过线性映射得到的采样点位置:

$$\Delta i_{m,k}=\frac{i}{H}+\mathrm{linear}\left(\boldsymbol{q}_{ij}\right)\quad,\quad(4)$$

$$\Delta j_{m,k}=\frac{j}{W}+\mathrm{linear}\left(\boldsymbol{q}_{ij}\right)\quad.\quad(5)$$

各采样点的位置和注意力权重可以随网络训练动态调整。实际上,AirFormer的注意力模块可以视为自适应的目标特征提取模板。在计算出注意力特征向量$\boldsymbol{a}_{i,j}$后,再对特征像元$\boldsymbol{f}_{i,j}$进行更新:

$$\boldsymbol{f}'_{ij}=\mathrm{norm}\left(\boldsymbol{f}_{ij}+\boldsymbol{a}_{ij}\right)\quad,\quad(6)$$

其中,norm表示层归一化(layer normalization)层。

得到融合局部空间特征和像素自身特征的向量$\boldsymbol{f}'_{i,j}$后,再将其分别送入置信度预测模块和位置预测模块,来分别得到像素位置附近存在目标的置信度和目标矩形预测框。置信度预测模块和位置预测模块均为有均为由4层线性层构成的多层感知机(multi layer perceptron,MLP),两层相邻的线性层由ReLU激活函数相连,最后一层线性层的输出特征量送入Sigmoid激活函数。置信度预测模块和位置预测模块前三层线性层的输入维度和输出维度以及最后一层线性层的输入维度均为C。置信度预测模块最后一层线性层的输出特征维度为1,输出特征用于表征目标存在的置信度。置信度预测可用下式来描述:

$$\alpha=\mathrm{Sigmoid}\left[\mathrm{MLP}\left(\boldsymbol{f}'_{ij}\right)\right]\quad,\quad(7)$$

其中Sigmoid表示Sigmoid激活函数,MLP表示多层感知机,α为目标预测框的置信度,经过Sigmoid激活函数后α的数值区间为$(0,1)$,α越接近1表示预测框属于目标的可能性越大。位置预测模块最后一层线性层的输出特征维度为4,输出特征量分别对应目标中心点横轴坐标、中心点纵轴坐标、目标框宽度和目标框高度。为了便于网络训练,在$\boldsymbol{f}_{i,j}$的像素位置上预设了一个初始锚框$\boldsymbol{b}_0=\left[\frac{x_0}{W},\frac{y_0}{H},\frac{w_0}{W},\frac{h_0}{H}\right]$,其中,$x_0$和$y_0$为锚框的中心,设为$\boldsymbol{f}_{i,j}$对应的像素中心,$w_0$和$h_0$为锚框的宽和高,分别设为图像宽和高的0.01。将$\boldsymbol{b}_0$和$\boldsymbol{f}'_{i,j}$送入位置预测模块,来得到目标的预测矩形框$\boldsymbol{b}$:

$$\boldsymbol{b}=\mathrm{Sigmoid}\left[\mathrm{MLP}\left(\boldsymbol{f}'_{ij}\right)-\ln\left(\frac{1}{\boldsymbol{b}_0}-1\right)\right]\quad.\quad(8)$$

考虑到经过归一化后的目标矩形框的中心坐标、宽和长范围在$(0,1)$内,因此先对$\boldsymbol{b}_0$进行了逆Sigmoid操作,加上由MLP输出的位置偏移预测量后再用Sigmoid激活函数映射到$(0,1)$。

将特征图$\boldsymbol{F}$每一个特征像元都当做查询特征像元来处理存在较大冗余。本文对特征图$\boldsymbol{F}$进行稀疏采样来选择查询特征像元。将特征图$\boldsymbol{F}$划分成若干个S×S大小的不重叠的图像块,S为正整数,每个图像块上选择相对位置固定的特征像元作为该图像块的查询像元。设选择出的查询像元数目为Q,在注意力模块中,查询向量降维操作的计算复杂度为$O\left(QCc\right)$,注意力权重和采样点生成的计算复杂度为$O\left(2QcMK+MK\right)$,而双线性插值采样以及注意力加权求和的计算复杂度为$O\left(5QKC\right)$,而最后线性层的运算复杂度为$O\left(QC^2\right)$。此外,注意力模块后续的模块总的计算复杂度也为$O\left(Q\right)$。因此,查询像元数目对AirFormer计算量有较大的影响。本文将S设为

2，即每2×2邻域产生一个查询像元，这样既不会使查询像元分布过于稀疏，又可以实现较大程度的计算量降幅。

在训练阶段，AirFormer使用DETR的基于匈牙利算法的标签分配策略，损失函数也采用DETR所使用的由预测框GIoU损失函数、预测框L1损失函数以及Focal类别预测损失函数[18]组成的联合损失函数。

在推理阶段，AirFormer对预测的目标框依次进行了坐标取整以及预测结果过滤等后处理操作。本文所关注的目标其尺寸与图像最小单元（单个像素）接近，而经过网络输出的坐标预测结果均为带小数，当预测坐标与真值坐标相差为小于0.5的纯小数时，即可认为与预测坐标与真值坐标完全贴合。但若结果保留小数，预测坐标框与真值坐标框之间的交并比可能偏小。因此对目标的坐标预测值进行四舍五入取值处理，若取整后的长或宽为0，取未取整时长或宽中心点所在像素的长或宽替换。在坐标取整后，使用目标检测任务中常用的非极大值抑制方法（non-maximum suppression，NMS）来处理预测框目标重叠情况，只保留重叠预测框中置信度最大的预测框。此外，可根据实际检测任务对漏检和虚警的不同需求来设置置信度阈值，移除置信度α低于阈值的预测框。如需低漏检率，置信度阈值需设置较低数值；如需低虚警率，置信度阈值需设置较高数值。

2 红外图像暗弱目标检测数据集

2.1 实测暗弱目标特性分析

本节对图1所示的五个空中民航飞机进行分析，其成像地点以及成像时间由表1给出。五个目标观测条件所有不同，目标一、目标三和目标四在白天成像，而目标二和目标五在夜间成像，可看出热红外载荷在夜间仍有较好的探测能力。此外，目标一、目标三和目标五所在的局部背景为海背景，目标二和目标四所在局部背景为陆地和云层的混合背景。从图1可以看出，空中民航目标的灰度值低于海背景和陆地背景，而一些云背景像素的灰度值则低于空中民航目标。

表2给出了目标和其局部背景的灰度值统计信息。将目标上的最小灰度值视为目标灰度值V_t，所在像素视为目标中心。所选的五个民航目标在三个波段图像上均成点扩散状，目标形态边界扩散到3×3到7×7之间的空间范围内。计算以目标为中心的11×11窗口到15×15窗口之间的背景像素的平均灰度值V_b，再通过下式计算目标灰度值和背景灰度值之间的差值比例r：

$$r = \frac{V_b - V_t}{V_b} \quad . \quad (9)$$

表1 实测民航飞机成像信息

Table1 The imaging information of real civial airplanes

序号	经度	纬度	日期	当地时间	局部背景
目标一	125.22° E	30.86°N	23.03.16	09:37	海
目标二	115.16°E	39.65°N	23.03.22	20:30	陆地、云
目标三	123.71°E	37.32°N	23.08.03	09:42	海
目标四	118.42°E	34.28°N	23.10.01	09:01	陆地、云
目标五	126.34°E	37.21°N	23.10.17	20:41	海

表2 实测民航飞机与背景灰度值信息

Table 2 Grayscale value information of real civial airplanes and background

序号	波段	目标灰度值	局部背景平均灰度值	差值比例
目标一	B1	1419	1562	9.17%
	B2	1644	1817	9.53%
	B3	1177	1260	6.59%
目标二	B1	1014	1044	2.92%
	B2	1198	1239	3.34%
	B3	858	875	1.98%
目标三	B1	1804	1917	5.91%
	B2	2036	2164	5.96%
	B3	1398	1461	4.37%
目标四	B1	1652	1724	4.21%
	B2	1880	1967	4.44%
	B3	1318	1359	3.08%
目标五	B1	1668	1773	5.95%
	B2	1928	2041	5.58%
	B3	1370	1425	3.88%

从表2可以看出，在相近的成像时刻和同一背景类型下，目标一和目标三的灰度值及其局部背景灰度值仍有较大差异。这种差异可能是由成像时大气状况差异造成的。尽管目标的灰度值在不同波段和不同观测场景下差异较大，但目标灰度值与其局部背景平均灰度值之间的Pearson相关系数为0.9965，p值为2.18×10^{-15}，线性相关性显著。对所选的五个目标进行统计，目标的灰度值和其局部背景平均灰度值差值比例在1.98%~9.53%范围内。

对同一个目标而言,B1 和 B2 波段的差值比例接近,而 B3 波段的差值比例明显小于 B1 和 B2 波段。根据文献[5]给出的三个波段不同高度大气透射率曲线,相比于 B1 和 B2 波段,B3 波段地表处的大气透过率远低于高空处的大气透过率,造成卫星入瞳处 B3 波段地表与空中民航目标的电磁辐射能量差值要低于 B1 和 B2 波段。

2.2　目标仿真

本文基于上一节分析的热红外卫星图像空中民航飞机特性提出了一个简易的暗弱目标仿真方法,可为基于深度学习的暗弱目标检测技术发展提供大规模目标样本支撑。该仿真方法的主要输入为热红外卫星图像、目标长度 l、目标中心点位置(x, y)、目标航向角 θ(航向与水平线的夹角)、灰度值差值比例 r 和高斯模糊核标准差 σ,其中目标中心点位置取值范围为图像尺寸范围内的任意实数。然后通过目标全像元灰度计算、目标丰度矩阵计算、目标灰度注入三步在图像上生成仿真目标。

1)目标全像元灰度计算

由上一节分析可知,热红外卫星图像空中民航飞机灰度值与背景灰度值存在正线性相关性。考虑到图像可能存在云背景,计算图像中前 30% 高的灰度值求平均值 V_I,则目标全像元灰度值设为 $(1\sim r)V_I$。

2)目标丰度矩阵计算

如图 3(a)所示,当前民航飞机机身长度与翼展接近,因此本文将民航飞机星下观测形状简化为长宽一致的十字形,十字最外侧边长设为其长度的 0.3。根据目标中心点位置和目标航向角,将简化后的目标形状模型贴到图像上。计算目标形状模型在图像每个像素上所占比例(即目标丰度),构建与图像同尺寸的目标丰度矩阵 $\boldsymbol{A}$。考虑到光学成像系统存在的像差和衍射效应,点光源在传感器平面上扩散成弥散斑。为模拟该特性,再使用标准差为 σ 的高斯模糊核对目标丰度矩阵进行高斯模糊处理。

3)目标灰度注入

依据线性混合模型,将目标灰度注入到图像上:

$$V'_{i,j} = \left(1 - \boldsymbol{A}_{i,j}\right)V_{i,j} + \boldsymbol{A}_{i,j}(1 - r)V_I \quad , \quad (10)$$

其中,$V_{i,j}$ 是原始热红外图像上 (i,j) 像素处的灰度值,$V'_{i,j}$ 是添加仿真目标后的像素灰度值。

如图 4 和图 5 所示,本文分别裁剪出目标一和目标五周围 30×30 大小的图像,依据表 3 仿真参数分别对两个目标进行仿真复现。真实目标用蓝色圈出,仿真目标用红色圈出,子图(b)和子图(c)给出其局部放大图。可以看出,除了尾迹云外,仿真目标基本复刻出了真实暗弱目标的灰度特征和形态特征,验证了本文所提仿真方法的有效性。

2.3　数据集介绍

基于 SDG 卫星 1 号的实测热红外图像和上一节所提的暗目标仿真方法,本文构建了红外序列图像暗弱动目标数据集,命名为 IRAir。数据集包含 2 000 段序列图像,训练集和测试集各 1 000 段序列,每段序列包含 50 张相同背景基底的单波段仿真图像,图像尺寸为 256×256,保存为 TIFF 图像格式。为贴合人眼样本标注,本文将目标丰度大于 0.1 的像素集合的最大外接矩形作为目标的标注框。考虑到动目标检测方法通常需要利用多帧序列图像信息,将每段序列后 30 帧图像作为训练和测试图像,前 20 帧图像留作动目标检测方法开发的备用前序图像。下文将进一步介绍目标仿真设置和仿真场

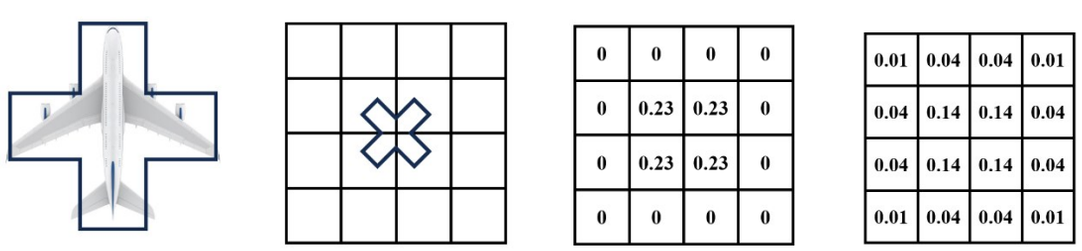
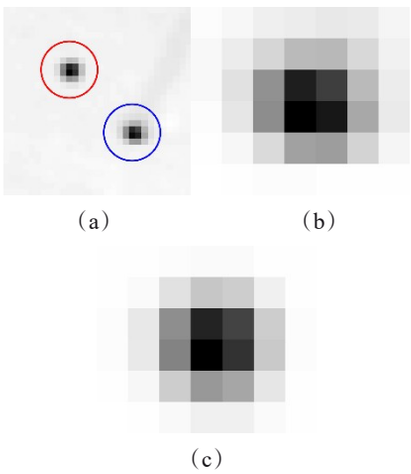


图 3　目标丰度矩阵计算流程示意图:(a)目标形状建模;(b)形状模型嵌入图像;(c)目标丰度矩阵计算;(d)丰度矩阵高斯模糊

Fig. 3　Schematic diagram of the calculation for object abundance matrix: (a) object shape modeling; (b) shape model embedding in image; (c) object abundance matrix calculation; (d) Gaussian blurring of the abundance matrix

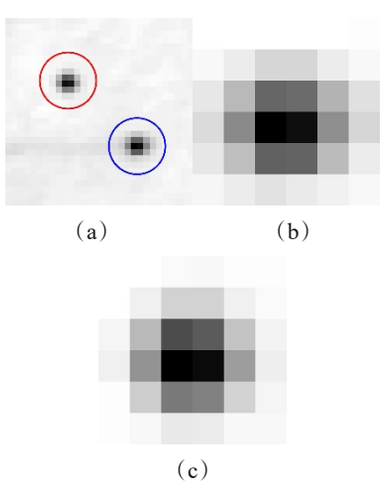


（a） （b）

（c）

图4 实测目标一及其仿真目标：(a)仿真图像；(b)真实目标；(c)仿真目标

Fig. 4 The 1st real civial aircraft and its simulation: (a) simulated image; (b) real object; (c) simulated object



（a） （b）

（c）

图5 实测目标五及其仿真目标：(a)仿真图像；(b)真实目标；(c)仿真目标

Fig. 5 The 5th real civial aircraft and its simulation: (a) simulated image; (b) real object; (c) simulated object

表3 真实目标的仿真参数设置

Table 3 Simulation parameter settings for real objects

参数	目标1	目标2
坐标	(10.2,10.7)	(10.3,9.95)
目标长度	80m	80m
航向角	45°	0°
差值比例	0.18	0.11
高斯模糊标准差	0.7	0.7

景设置。

(1)目标仿真参数设置

目标长度：不同类型民航飞机尺寸差异较大，如波音747-8洲际客机机身长76.3 m，翼展68.4 m，而空客A320中短程客机机身长37.6 m，翼展34.1 m。因此，本文将目标长度l取值设为40 m、50 m、60 m、70 m和80 m，各取值的目标数目比例设为3:3:1:1:1。

目标速度：当前民航飞机基本为亚音速飞机，本文将目标速度范围设为800~900 km/h，在30 m空间分辨率的图像上对应的目标速度范围为7.4~8.3 pixel/s。

目标灰度：目标灰度值差值比例r取值范围设0.1~0.2。

高斯模糊：高斯模糊核的标准差σ固定为0.7。

目标轨迹：本文设置了匀速直线运动和匀速圆周运动两类民航目标运动轨迹，目标航向取值范围为-180^{o}~180^{o}。

目标起始状态：50%的目标在起始帧出现，剩余50%的目标在第5帧到第35帧之间出现。起始帧出现的目标初始位置取值范围为整张图像，非起始帧出现的目标初始位置位于图像边界。

目标长度、目标速度、目标灰度、目标轨迹和目标初始状态都在取值范围内随机组合，以保证目标样本的多样性和丰富性。

(2)仿真场景设置

为进一步丰富场景多样性，除了目标相关参数设置之外，本文还考虑了图像帧频、成像波段、帧间位移、图像噪声等

帧频：取值范围为1~10帧每秒(FPS)。

波段：每段序列从B1、B2和B3三个波段中任选一个。

帧间位移：由于观测平台震颤以及运动等因素，视频图像帧间存在位移。在仿真中，本文引入帧间位移量e。起始帧不做位移处理，后续的每一帧图像都相对起始帧偏移e像素。对于第k帧图像，其沿水平轴和垂直轴的偏移量分别为：

$$\Delta x = e\cos\theta_k \quad , \quad (11)$$

$$\Delta y = e\sin\theta_k \quad , \quad (12)$$

其中，θ_k为第k帧图像的偏移方向角，取值范围为$[-\pi,\pi]$。e取值为0、1和2像素。在计算出偏移量后，使用双线性插值法获得偏移后的图像。

图像噪声强度：在添加仿真目标后的图像上，通过下式向其添加噪声：

$$V''_{ij} = V'_{ij}(1+\eta) \quad , \quad (13)$$

其中，$V'_{i,j}$是仿真图像上(i,j)像素处的灰度值，$V''_{i,j}$是添加噪声后的像素灰度值，η为服从均值为0标准差为n的随机变量。n取值为0.002或0.005，本文称为噪声强度。

目标数量：每段序列目标数目设为3~10之间

的随机整数。

目标遮掩场景:当以目标位置为中心的5×5窗口的背景均值小于仿真目标全像元灰度,认为目标被遮掩,不添加该目标。

表4为各场景类型的序列数和目标数统计信息。图6给出了4段仿真序列的第50帧图像。方框为第50帧的目标位置框,圆点为前49帧目标位置框的中心。表5为4段仿真序列的参数设置。由图6可以看出,不同帧频和帧间位移设置下目标轨迹差异较大,当帧频为1FPS且无帧间位移时目标轨迹较为规则,而当帧频为10FPS且帧间位移为2像素时目标轨迹点呈现随机波动的状态。

表4　IRAir数据集不同载荷参数下样本数

Table4　Sample numbers under different load parameters on the IRAir dataset

	场景类型	训练集序列数	训练集目标数	测试集序列数	测试集目标数
波段	B1	301	1970	338	2260
	B2	363	2271	345	2332
	B3	336	2131	317	1977
帧频	1~5 FPS	480	3059	499	3229
	6~10 FPS	520	3313	501	3340
帧间位移	0像素	339	2158	315	2063
	1像素	339	2157	361	2374
	2像素	322	2057	324	2132
噪声强度	0. 002	497	3200	467	3079
	0. 005	503	3172	533	3490
总计		1000	6372	1000	6569

3　实验结果与分析

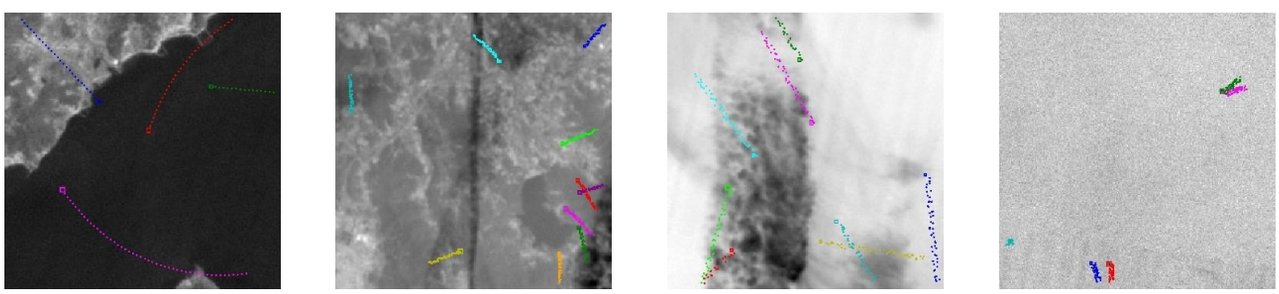
3. 1　实施细节

本文在IRAir数据集上对所提出的AirFormer网络进行了评测。AirFormer的特征维度C设为64,c设为8,注意力模块的多头数目M设为32,每个注意力头的采样点数设为4。网络输入图像尺寸为256×256,训练迭代轮数设为50轮,批次大小设为8,使用Adam优化器,学习率初始化为0. 0001,在40轮时学习率减小为原始的0. 1。此外,本文还对RGB图像通用目标检测网络CornerNet[19]、YOLOv3[20]、Deformable DETR[14]、RTMDET-tiny[21]和YOLOX-tiny[22],以及可见光卫星视频运动目标检测算法DSFNet[23]在IRAir数据集进行了评测。为公平对比,CornerNet、YOLOv3、Deformable DETR、RTMDET-tiny和YOLOX-tiny均在单帧256×256图像上训练50轮,只采用图像翻转作为训练时的数据增强手段。YOLOv3、RTMDET和YOLOX算法为单阶段检测算法,较双阶段检测算法整体结构简洁,推理实时性较好。RTMDET-tiny和YOLOX-tiny分别为RTMDET和YOLOX的最小参数版本,是近年目标检测领域在轻量化上取得的代表性成果。RTMDET-tiny和YOLOX-tiny是面向RGB图像常规目标检测任务的轻量级网络,在IRAir数据集评测中发现直接采用原始设置RTMDET-tiny和YOLOX-tiny均不能有效预测目标,因此在本文实验中对RTMDET-tiny和YOLOX-tiny的结构进行了针对性调整。RTMDET-

表5　示例仿真序列参数设置

Table 5　The parameter settings of example simulated sequences

参数	序列0041	序列0077	序列0266	序列0393
波段	B2	B2	B3	B3
帧频	1 FPS	6 FPS	2 FPS	10 FPS
帧间位移	0像素	1像素	2像素	2像素
噪声强度	0. 002	0. 002	0. 005	0. 002
目标数量	4	10	8	5



(a)　(b)　(c)　(d)

图6　仿真序列示例:(a)序列0041;(b)序列0077;(c)序列0266;(d)序列0393

Fig. 6　Simulated sequence examples: (a) sequence 0041; (b) sequence 0077; (c) sequence 0266; (d) sequence 0393

tiny和YOLOX-tiny均默认使用骨干网络输出的P3、P4和P5特征图进行检测，其尺寸为原始输入图像尺寸的1/8、1/16和1/32，IRAir数据集中的目标本身尺寸极小，在缩小后的P3至P5特征图上目标特征更加微弱，因此，在本章实验中，RTMDET-tiny和YOLOX-tiny改为使用P1特征图（其尺寸为原始输入图像尺寸的1/2）进行检测。DSFNet则是针对于可见光卫星视频小尺寸运动车辆检测任务[24]所提的检测算法，联合利用空间域形态信息以及多帧时间域上的运动信息进行检测。

3.2 评价指标

本文采用COCO数据集[11]中的评测指标平均精度（average precision，AP），每张图取置信度在前300的预测框，计算各检测算法的AP和AP_{20}，其中AP在0.5到0.95共10个交并比阈值下平均精度的平均值。考虑暗弱目标尺寸极小，与真值框交并比较低的预测框也有一定的目标定位意义，故也评估了交并比阈值0.2时的平均精度AP_{20}。此外，本文还使用遥感小目标检测任务常用的召回率（Recall，Re，又称检出率）、准确率（Precision，Pr）以及F1分数等三个指标。召回率是正确检出的目标数目占真实目标总数的比例，准确率为正确检出的目标数目占所有预测出的目标数目的比例。F1分数是召回率和准确率的调和平均数，计算公式为：

$$F1 = \frac{2 \times Re \times Pr}{Re + Pr} \quad , (14)$$

F1分数越高，表示检测方法性能越好。在召回率Re、准确率Pr和F1分数的计算中，需要考虑预测框与真值框的交并比阈值以及预测框置信度阈值来判断预测框是否为虚警。本文将预测框与真值框的交并比阈值设为0.2。各类算法的预测框置信度阈值以0.1的步长从0.1遍历到0.9，取F1分数最大时的召回率、准确率和F1分数作为被测算法性能的评测结果。所有预测框均值坐标取整处理。考虑到同一序列各帧图像较为相似，本文只评测各测试序列第30帧、第40帧和第50帧图像。

本文还评测了各算法的网络模型参数量、输入图像尺寸为256×256时的浮点运算次数（floating point operations，FLOPs）用于比较算法的复杂度。此外，本文测试了输入单张256×256图像各单帧目标检测算法的推理时间。由于DSFNet为视频动目标检测算法，因此测试了DSFNet同时输入5张256×256序列图像进行推理所用时间。推理耗时只计算图像在网络中前传所用时间，不包含图像加载时间和后处理时间。

3.3 结果分析

如表6所示，在没有使用网络剪枝和网络量化等模型压缩操作的情况下，AirFormer实现了极小规模的模型参数量和计算量，参数量仅为37.1 K，浮点运算次数仅为46.2 M，单图推理耗时仅为5.7 ms，均显著优于对比方法。在参数量和浮点运算次数方面，AirFormer相较当前常规目标检测网络可下降3到4个数量级，较本章实验中优化后的RTMDET-tiny和YOLOX-tiny两个轻量级目标检测网络下降2个数量级。

在网络规模显著降低的情况下，AirFormer在IRAir数据集上实现了对暗弱目标较佳的检测效果，AP性能可达0.349，召回率可达0.710，准确率可达0.826。AirFormer以及仅使用一层低层特征图的RTMDET-tiny和YOLOX-tiny在AP指标均高于CornerNet、YOLOv3、Deformable DETR，但是在召回率、准确率和F1分数上有较大劣势。在不同评价指标下算法检测性能对比结果并不一致，AP指标过多关注预测框对真值框的贴合能力，因此使用原始尺寸或者仅一次下采样的特征图的检测网络在AP指标上取得较好的评估结果。但实际上，对于IRAir数

表6 检测方法性能比较

Table 6 Performance comparison of detection methods

方法	CornerNet	YOLOv3	Deformable DETR	RTMDET-tiny	YOLOX-tiny	DSFNet	AirFormer
AP	0.336	0.270	0.274	0.350	0.398	0.233	0.349
AP_{20}	0.770	0.752	0.709	0.812	0.765	0.528	0.737
召回率	0.738	0.766	0.688	0.544	0.716	0.504	0.710
准确率	0.904	0.902	0.897	0.675	0.843	0.932	0.826
F1	0.812	0.828	0.779	0.603	0.774	0.653	0.764
参数量	201.0M	61.5M	41.1M	2.7M	2.7M	17.0M	37.1K
FLOPs	112.8G	12.4G	15.0G	5.9G	5.5G	12.2G	46.2M
推理耗时	29.4 ms	11.7 ms	32.3 ms	10.6 ms	9.1 ms	50.1 ms	5.7 ms

据集中的极小尺寸目标而言,目标与背景的区分边界并不明显,此类目标的检测方法设计不必过于追求预测框与真值框的高重合度。此外,基于多帧图像的运动目标检测网络DSFNet并没有显示出优于当前单帧目标检测网络的性能,可能是因为各序列的帧间位移和帧频类型多样、时域特征不稳定导致的。利用低质量序列图像数据实现运动目标检测仍是日后需要进一步研究的难点。

表7评估了目标尺寸对检测性能的影响。尽管在仿真中40 m目标和50 m目标占据更多样本比例,但是这两类目标的召回率仍显著低于其他三类更大尺寸的目标。在3 000张测试图片上,40 m长度的目标真值数为6 222,但是只正确检出了3 016个目标,召回率为48.5%。随着目标尺寸增大,检测网络对目标的召回率也在提升,对80 m长度的目标的召回率可达92.6%。尽管仿真目标的长度参数不能与真实目标尺寸精确对应,但是仍能反应出暗弱目标尺寸对检测性能的影响。

表7　AirFormer对不同尺寸目标检测性能对比

Table7　Detection performance comparison of AirFormer for objects with different sizes

目标长度/m	真值数	检出数	召回率/%
40	6 222	3 016	48.5
50	6 154	4 626	75.2
60	2 109	1 838	87.2
70	2 016	1 798	89.2
80	2 062	1 909	92.6

利用IRAir数据集训练出来的AirFormer网络对2.1节中所分析的五个实测空中民航飞机共15张单通道红外图像进行检测,检测结果可视化情况如图7所示。绿框表示正确检测的目标,黄框表示漏检的目标,红框表示虚警。在15张测试图像上,AirFormer共报出33个预测框,其中14个预测框为真实民航飞机,剩余19个预测框为虚警,漏检真实目标1个。实测目标一、三和五所在场景为海背景,AirFormer检出全部目标,仅在图像拼接错位处产生1个虚警。实测目标二和三所在场景场景为陆地背景和云背景,场景复杂,AirFormer报出多个虚警。实测目标三B3波段图像有两条由暗点组成的噪声条带,AirFormer在这两条条带上报出了9个虚警。此外,实测目标二在B3波段与背景在灰度上接近,灰度差值比例仅为1.98%,AirFormer未能检出。

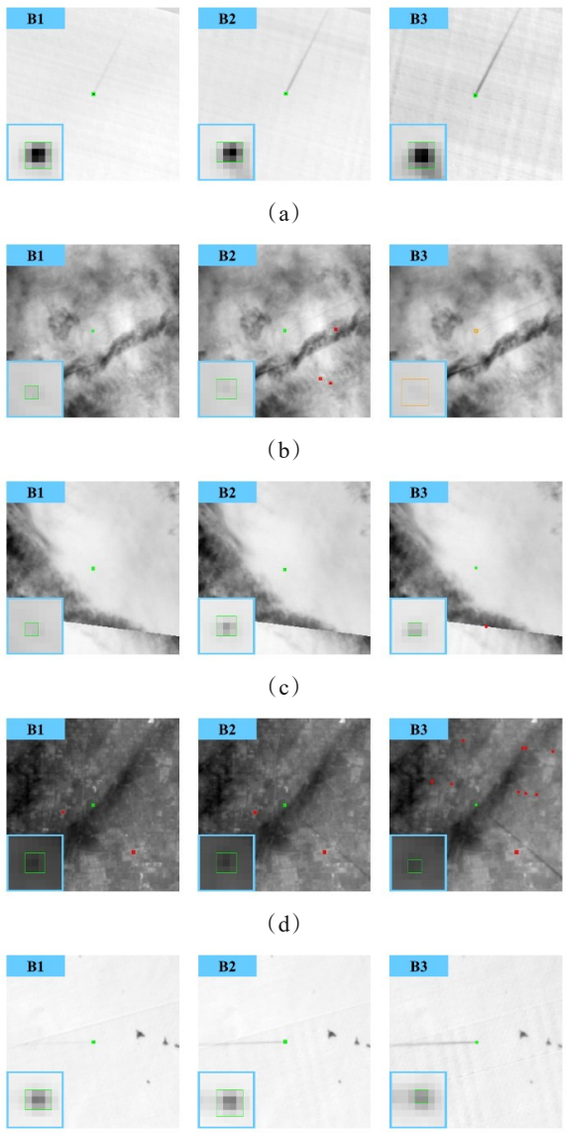


图7　AirFormer对实测暗弱目标检测结果图:(a)实测目标一;(b)实测目标二;(c)实测目标三;(d)实测目标四;(e)实测目标五

Fig. 7　The detection results of AirFormer for real civil airports: (a) the 1st real airport; (b) the 2nd real airport; (c) the 3rd real airport; (d) the 4th real airport; (e) the 5th real airport

4　结论

针对于红外图像暗弱目标检测这一新挑战,本文进行了数据集和检测方法两方面上的工作。在检测方法方面,本文提出了一种基于可形变注意力机制的极轻量级暗弱目标检测网络,参数量仅为37.1 K,在256×256尺寸的图像上浮点运算次数仅为46.2 M。在数据集方面,针对真实暗弱目标数据集缺乏、弱目标人工标注难的问题,本文构建了红

外图像暗弱目标检测仿真数据集IRAir。所提的网络利用单帧图像信息在IRAir数据集上实现了暗弱目标71.0%召回率和82.6%检测准确率。在实测红外图像上进行验证,利用仿真数据训练出的网络在海背景上对真实暗弱目标实现了较佳的检测结果。同时也发现,复杂背景以及极小尺寸的目标仍是检测的难点和挑战,需要后续工作中进一步探索。

References

[1] IATA. Airlines Set to Earn 2.7% Net Profit Margin on Record Revenues in 2024 [EB/OL]. (2023-12-06) [2024-04-28]. https://www.iata.org/en/pressroom/2023-releases/2023-12-06-01/.

[2] Zhao F, Xia L, Kylling A, et al. Detection flying aircraft from Landsat 8 OLI data [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2018, 141: 176-184.

[3] Liu Y, Xu B, Zhi W, et al. Space eye on flying aircraft: From Sentinel-2 MSI parallax to hybrid computing [J]. Remote Sensing of Environment, 2020, 246: 111867.

[4] Fehrm B. Bjorn's Corner: Supersonic transport revival, Part 6 [EB/OL]. (2018-09-14) [2024-04-29]. https://leehamnews.com/2018/09/14/bjorns-corner-supersonic-transport-revival-part-6/.

[5] Li L, Zhou X, Hu Z, et al. On-orbit monitoring flying aircraft day and night based on SDGSAT-1 thermal infrared dataset [J]. Remote Sensing of Environment, 2023, 298: 113840.

[6] Zhu H, Zhang X, Chen X, et al. Dim small targets detection based on horizontal-vertical multi-scale grayscale difference weighted bilateral filtering [J]. Journal of Infrared and Millimeter Waves, 2020, 39(4): 513-522.

[7] Gao C, Zhang T, Li Q. Small infrared target detection using sparse ring representation [J]. IEEE Aerospace and Electronic Systems Magazine, 2012, 27(3): 21-30.

[8] Liu T, Yang J, Li B, et al. Infrared small target detection via nonconvex tensor tucker decomposition with factor prior [J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-17.

[9] Lin Z P, Li B Y, Li M, et al. Light-weight infrared small target detection combining cross-scale feature fusion with bottleneck attention module [J]. Journal of Infrared and Millimeter Waves, 2022, 41(6): 1102-1112.
林再平, 李博扬, 李淼, 等. 结合跨尺度特征融合与瓶颈注意力模块的轻量型红外小目标检测网络[J]. 红外与毫米波学报, 2022, 41(6): 1102-1112.

[10] Lin Z P, Luo Y H, Li B Y, et al. Gradient-aware channel attention network for infrared small target image denoising before detection [J]. Journal of Infrared and Millimeter Waves, 2024, 43(2): 254-260.

[11] Lin T Y, Maire M, Belongie S, et al. Microsoft coco: Common objects in context [C]. Proceedings of the European Conference on Computer Vision. Cham: Springer International Publishing, 2014: 740-755.

[12] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [C]. Proceedings of the 31st International Conference on Neural Information. New York: ACM, 2017, 6000-6010.

[13] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers [C]. Proceedings of the European Conference on Computer Vision. Cham: Springer International Publishing, 2020: 213-229.

[14] Zhu X, Su W, Lu L, et al. Deformable DETR: Deformable transformers for end-to-end object detection [J]. arxiv preprint arxiv:2010.04159, 2020.

[15] Liu S, Li F, Zhang H, et al. DAD-DETR: Dynamic anchor boxes are better queries for DETR [J]. arxiv preprint arxiv:2201.12329, 2022.

[16] Li Z, An W, Guo G, et al. SpecDETR: A transformer-based hyperspectral point object detection network [J]. arXiv preprint arXiv:2405.10148, 2024.

[17] Xu Q, Wang L, Sheng W, et al. Heterogeneous graph transformer for multiple tiny object tracking in RGB-T videos [J]. IEEE Transactions on Multimedia, 2024, 26: 9383-9397.

[18] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection [C]. Proceedings of the IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2017: 2980-298.

[19] Law H, Deng J. Cornernet: Detecting objects as paired keypoints [C]. Proceedings of the European Conference on Computer Vision. Cham: Springer International Publishing, 2018: 734-750.

[20] Redmon J, Farhadi A. YOLOv3: An incremental improvement [J]. arxiv preprint arxiv:1804.02767, 2018.

[21] Lyu C, Zhang W, Huang H, et al. RTMDET: An empirical study of designing real-time object detectors [J]. arXiv preprint arXiv:2212.07784, 2022.

[22] Ge Z, Liu S, Wang F, et al. Yolox: Exceeding yolo series in 2021 [J]. arXiv preprint arXiv:2107.08430, 2021.

[23] Xiao C, Yin Q, Ying X, et al. DSFNet: Dynamic and static fusion network for moving object detection in satellite videos [J]. IEEE Geoscience and Remote Sensing Letters, 2021, 19: 1-5.

[24] Yin Q, Hu Q, Liu H, et al. Detecting and tracking small and dense moving objects in satellite videos: A benchmark [J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1-18.