文章编号:1001-9014(2024)06-0859-13
DOI:10.11972/j.issn.1001-9014.2024.06.017

基于渐进时空特征融合的红外弱小目标检测

曾 丹[1], 卫建铭[1], 张俊杰[1], 常 亮[2], 黄 微[1*]
(1. 上海大学 通信与信息工程学院,上海 200444;
2. 中国科学院 微小卫星创新研究院,上海 201203)

摘要:为避免现有多帧红外弱小目标检测算法在显式对齐多帧特征时产生的估计误差累积,并缓解网络降采样导致的目标特征丢失,提出了一种渐进时空特征融合网络,采用渐进时序特征累积模块隐式地聚合多帧信息,并利用多尺度空间特征融合模块增强浅层细节特征与深层语义特征之间的交互。针对多帧红外弱小目标数据集稀缺的现状,构建了一个高度真实的半仿真数据集。与主流算法相比,提出的算法在所提出数据集和公开数据集上的检测概率分别提升了4.69%与4.22%。

关 键 词:红外弱小目标检测;时空特征融合;渐进时序特征累积;多尺度空间特征融合;多帧数据集
中图分类号:TP753 **文献标识码**:A

Progressive spatio-temporal feature fusion network for infrared small-dim target detection

ZENG Dan[1], WEI Jian-Ming[1], ZHANG Jun-Jie[1], CHANG Liang[2], HUANG Wei[1*]
(1. School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China;
2. Innovation Academy for Microsatellites, Chinese Academy of Sciences, Shanghai 201203, China)

Abstract: To avoid the accumulation of estimation errors from explicitly aligning multi-frame features in current infrared small-dim target detection algorithms, and to alleviate the loss of target features due to network downsampling, a progressive spatio-temporal feature fusion network is proposed. The network utilizes a progressive temporal feature accumulation module to implicitly aggregate multi-frame information and utilizes a multi-scale spatial feature fusion module to enhance the interaction between shallow detail features and deep semantic features. Due to the scarcity of multi-frame infrared dim target datasets, a highly realistic semi-synthetic dataset is constructed. Compared to the mainstream algorithms, the proposed algorithm improves the probability of detection by 4.69% and 4.22% on the proposed dataset and the public dataset, respectively.

Key words: infrared small-dim target detection, spatio-temporal feature fusion, progressive temporal feature accumulation, multi-scale spatial feature fusion, multi-frame dataset

引言

由于红外传感器具有可全天候工作、抗电磁干扰性能强以及弹载方便的特点,红外弱小目标检测技术作为红外探测系统的关键技术,被广泛应用于早期预警系统、精确制导等领域[1]。但在实际应用场景中,由于背景复杂,信杂比(Signal to Clutter Ratio, SCR)低,目标容易淹没在背景中,呈现出“弱”的特点;同时,由于成像距离长,目标在整个图像中仅占据很少的像素,呈现出“小”的特点,因此红外弱小目标检测是一个具有较大挑战性的课题[2]。

为了检测红外弱小目标,研究人员提出了许多传统的检测方法,这些方法主要通过图像处理技术或手工设计的特征抑制图像中背景及噪声,实现目标检测。传统的检测方法包括基于滤波的方法,如最大中值滤波(Max-median filter)[3]方法、形态学顶

收稿日期:2024-03-24,**修回日期**:2024-04-17 **Received date**:2024-03-24,**Revised date**:2024-04-17
基金项目:国家科学自然基金(62372284)
Foundation items:Supported by the National Natural Science Foundation of China (62372284)
作者简介(Biography):曾 丹(1982-),女,湖南邵东人,教授,博士,主要研究领域为计算机视觉、模式识别. E-mail: dzeng@shu.edu.cn
*通讯作者(Corresponding author): E-mail: lyxhw@shu.edu.cn

帽变换(Top-Hat)[4]方法等;基于局部对比度衡量(local contrast measure, LCM)[5]的方法,如加权局部对比度衡量(weighted strengthened local contrast measure, WSLCM)[6]方法等;基于低秩的方法,如红外补丁图像(infrared patch-image, IPI)[7]、部分和张量核范数(partial sum of the tensor nuclear norm, PSTNN)[8]方法等。然而,这些传统方法过度依赖于手工设计的特征及超参数的调整,当目标的大小、形状、信杂比剧烈变化时,容易出现虚警和漏检的情况,算法鲁棒性较差。

与传统方法相比,基于深度学习的方法由于其强大的建模能力,可以从覆盖复杂场景的大量训练数据中自动提取特征,算法的检测性能得到了显著提升,近年来吸引了越来越多的研究兴趣[9]。Dai等人[10]提出了使用非对称上下文调制(asymmetric contextual modulation, ACM)的单帧红外弱小目标检测方法,该方法通过对语义信息和空间细节进行更丰富地交互,实现了良好的检测性能。Li等人[11]提出了一种密集嵌套注意网络(dense nested attention network, DNANet)来增强这种交互作用。Wu等人[12]将一个较小的UNet嵌入到一个较大的UNet主干(UNet In UNet, UIUNet)中,以实现红外弱小目标的多层次和多尺度表示学习。Zhang等人[13]提出了一种红外形状网络(infrared shape network, ISNet),建立了一个受泰勒有限差分启发的边缘块来关注形状问题,检测具有边缘特征的红外弱小目标。Li等人[14]提出了一种使用点监督实现红外弱小目标检测的方法,通过在聚类过程中引入随机性,对多次聚类结果求平均得到可靠的伪掩码进行网络训练。近年来,部分工作通过引入瓶颈结构[15]或使用混合精度量化[16]提高了网络检测效率,便于网络模型进行部署应用。

然而,红外弱小目标往往淹没在大量的杂波和复杂的背景中,一旦目标在单帧图像中视觉特征不明显,基于单帧的检测方法性能会急剧下降。基于多帧的检测算法能同时利用目标的视觉信息和运动信息,具有更好的检测性能。Liu等人[17]开发了一种非凸张量低秩近似(non-convex tensor low-rank approximation, NTLA)方法来实现背景估计,抑制杂波。Sun等人[18]在此基础上引入了加权Schatten范数,获得了更精准的背景估计。Liu等人[19]提出了一种具有因子先验的非凸张量Tucker分解模型,采用群稀疏正则化的总变分抑制背景杂波。Yan等人[20]提出了一种时空差异多尺度注意网络(spatio-temporal differential multiscale attention network, STDMANet),该网络以差分的形式从多帧和多尺度特征注意模块中捕获时空特征。Chen等人[21]提出了一种切片时空网络(sliced spatial-temporal network, SSTNet),设计了一种跨切片的卷积长短期记忆(convolutional long short-term memory, ConvLSTM)节点捕获切片内部及切片间的时空运动特征。Li等人[22]提出了方向编码时间U型模块(direction-coded temporal U-Shape module, DTUM),通过构建运动到数据的映射来区分目标和杂波的运动,并设计了一种方向编码的卷积块来提取目标的运动信息。

上述的多帧方法中,虽然传统方法获得了更精准的背景估计,抑制了大部分的杂波,但也会在一定程度上增强如建筑物角点或非目标运动对象等干扰信号,造成较高的虚警率。基于深度学习的方法通常使用光流法、仿射变换等方法进行帧间配准,显式地对齐不同帧中的目标特征,但在进行多帧对齐时会产生误差累积,难以应对目标复杂的运动情况;其次,由于网络模型中存在大量降采样操作,红外弱小目标的视觉特征和位置信息会不可避免地逐渐丢失,深层网络特征难以捕捉到目标的信息;最后,目前公开的多帧红外弱小目标数据集较为稀缺,无法满足多帧深度学习算法对于数据集的需求。Hui等人[23]提出了真实拍摄的多帧红外图像弱小目标数据集,该数据集只提供了中心点标注,没有提供检测框和掩码标注,限制了使用场景。Sun等人[24]提出了仿真的多帧红外弱小运动目标检测数据集,该数据集简单地使用符合高斯分布的亮斑作为目标,且只提供了中心点标注。Sun等人[25]开发了一个红外弱小目标数据集(infrared dim small target dataset, IRDST),带有三种类型的标签(掩码、边界框和中心像素),但其目标运动轨迹较为简单,镜头抖动明显,且手工标注较为粗糙。Li等人[22]提出了多帧红外弱小目标仿真数据集,该数据集背景仅有轻微的抖动,且目标与背景融合较为突兀。

为了解决上述问题,本文提出了一种新的深度学习框架,称为渐进时空特征融合网络(progressive spatio-temporal feature fusion network, PSTFNet),该网络利用连续帧图像隐式地增强检测帧中的目标特征,并融合不同层次的特征,兼顾了细节信息丰富的浅层特征和语义信息丰富的深层特征,增强了

目标在深层网络中的特征表示。一方面,本文设计了一个渐进时序特征累积模块(progressive temporal accumulation module, PTAM),该模块使用感受野渐进增大的2D卷积将连续帧中的特征累加到检测帧上,同时使用不同尺度的3D卷积提取时间维度特征,以增强检测帧中目标的特征。另一方面,本文设计了一个多尺度空间特征融合模块(multi-scale spatial feature fusion module, MSFM),该模块使用具有不同感受野的卷积层融合不同层次的特征,强化了深层网络对于弱小目标的关注能力,在获得弱小目标语义特征表达的同时兼顾了目标定位的准确性,提高了检测效果。此外,本文构建了一个包含100段序列,共10 000帧图像的多帧红外弱小目标数据集SHU-MIRST,该数据集以半仿真的方式,在真实拍摄的红外背景图像中嵌入红外运动目标,涵盖了多种目标运动方式、背景运动类型及典型场景,且目标的运动轨迹同时考虑了目标自身的移动及背景的位移。同时本文设计了一种基于区域重采样的图像融合算法,使得目标插入背景后符合视觉真实性和物理特性合理性。

1　网络结构及数据集制作

本节首先介绍PSTFNet的整体结构,接着介绍网络的特征融合模块,包括PTAM模块和MSFM模块,最后介绍红外弱小目标数据集SHU-MIRST的制作方法。

1.1　整体结构

为了更好地利用序列图像中的时序特征信息和空间特征信息,本文提出了渐进时空特征融合网络(PSTFNet)。如图1(a)所示,PSTFNet的主干网络是ResUNet[26],其中编码器部分共有4个阶段,每个阶段会通过ResUNet卷积块对特征进行下采样及维度扩充,以提取更深层次的特征。本文使用ResUNet对连续5帧图像中的每一帧图像单独提取特征F_i^j,其中$i\in[t-4,\ t]$,代表连续5帧中的某一帧,$j\in[1,\ 4]$,代表特征的阶段。

以图1(a)中编码器中第j阶段提取的5帧特征为例,使用PTAM模块对5帧的特征进行处理,最终增强第5帧中红外弱小目标的特征。本文一共使用了4个PTAM模块对4个阶段的特征进行了增强,得到了$F_e^j, j\in[1,\ 4]$,增强后的特征与增强前单帧特征的维度相同。本文认为,经过PTAM模块,前4帧中的时序信息已经被提取并累积到了第5帧中,后续部分将仅使用增强后的特征F_e^j进行处理及检测,以减少网络的参数量。

由于本文的主干网络ResUNet中包含4次降采

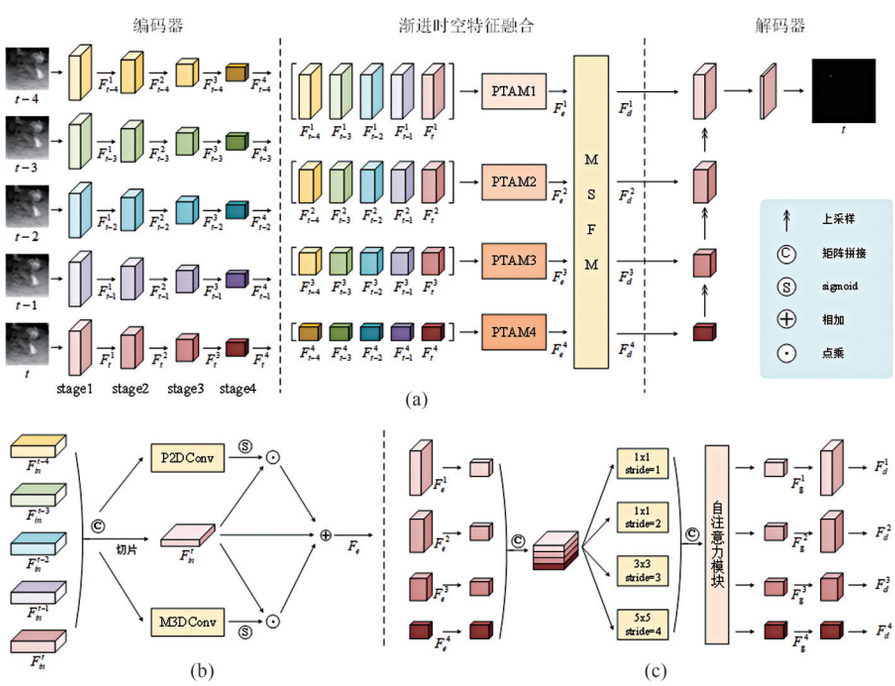

图1　渐进时空特征融合网络结构:(a)PSTFNet网络整体架构;(b)PTAM模块架构;(c)MSFM模块架构

Fig. 1　Progressive spatio-temporal feature fusion network structure: (a) overall architecture of PSTFNet; (b) progressive temporal accumalation module; (c) multi-scale spatial feature fusion module

样操作，提取到的最小特征图尺寸仅有原图大小的1/16，目标很容易淹没在降采样操作中，且经过增强后的不同阶段的特征尺寸不同，因此本文使用MSFM模块对F_e^j进行特征融合，增强深层网络中目标的特征。MSFM模块将输出$F_d^j, j \in \{1, 4\}$，将F_d^j输入到ResUNet的解码器中，使用ResUNet卷积块和上采样来解码多层特征，并最终通过sigmoid函数以及阈值分割得到检测结果。

1.2 渐进时序特征累积模块

在渐进时序特征累积模块（PTAM）中，本文使用了两种方式分别提取时序特征。如图1(b)所示，本文使用感受野渐进增大的2D卷积操作（progressive increasing 2D convolution，P2DConv）将之前帧的特征累加到检测帧上，同时使用多尺度3D卷积操作（multi-scale 3D convolution，M3DConv）提取时序特征。

本文认为，红外弱小目标是连续运动的，任意两帧中目标特征的位移应随着时间隔增大而增加，例如第1帧与第3帧之间目标的位移大于第1帧和第2帧之间的位移。本文根据选定两帧之间的帧间间隔选择合适大小的卷积核提取特征，获取相应的感受野，更好地提取时序特征。具体而言，对于相邻的两帧特征，本文使用3 × 3的卷积核提取前一帧特征累加至后一帧的特征上，对于间隔一帧的特征（例如第$t-2$帧和第t帧），使用5 × 5的卷积核提取前帧的特征累加至后帧特征上。依次类推，最终，本文使用9 × 9的卷积核提取第$t-4$帧的特征累加至第t帧的特征上。特别的，对于每帧的原始特征，本文将其通过一个1 × 1的卷积核，以平衡不同帧中特征的大小。

如图2(a)所示，F_{in}^i代表输入P2DConv模块的第i帧特征，P_{out}为P2DConv的输出，计算过程如下所示：

$$P_{t-4} = C_{1\times1}(F_{\text{in}}^{t-4}) \quad , \quad (1)$$

$$P_{t-3} = C_{1\times1}(F_{\text{in}}^{t-3}) + C_{3\times3}(P_{t-4}) \quad , \quad (2)$$

$$P_{t-2} = C_{1\times1}(F_{\text{in}}^{t-2}) + C_{3\times3}(P_{t-3}) + C_{5\times5}(P_{t-4}) \quad , \quad (3)$$

$$P_{t-1} = C_{1\times1}(F_{\text{in}}^{t-1}) + C_{3\times3}(P_{t-2}) + C_{5\times5}(P_{t-3}) + C_{7\times7}(P_{t-4}) \quad , \quad (4)$$

$$P_t = C_{1\times1}(F_{\text{in}}^{t}) + C_{3\times3}(P_{t-1}) + C_{5\times5}(P_{t-2}) + C_{7\times7}(P_{t-3}) + C_{9\times9}(P_{t-4}) \quad , \quad (5)$$

$$P_{\text{out}} = \sigma(C_{3\times3}(\text{cat}([P_{t-4}, P_{t-3}, P_{t-2}, P_{t-1}, P_t]))[-1]) \quad , \quad (6)$$

其中，$C_{a\times a}$代表卷积核大小为$a \times a$的2D卷积操作，$a = \{1, 3, 5, 7, 9\}$，$\sigma(\cdot)$为sigmoid函数，$\text{cat}(\cdot)$为矩阵拼接操作，[-1]表示对特征的时间维度进行切片操作，取该维度最后一个切片的特征进行后续处理。

此外，3D卷积被广泛应用于视频目标分割领域[27]，3D卷积具有$T \times W \times H$的三维卷积核，可以对连续帧特征张量中的时间维度进行卷积，从而提取时序信息。基于以上想法，本文在PTAM模块中增加了多个不同尺度的3D卷积操作，卷积核大小分别为1 × 3 × 3、3 × 3 × 3、5 × 3 × 3，通过使用T维度大小不同的卷积核，提取不同时间跨度内的红外弱小目标的时序特征，三个3D卷积操作是并行的，

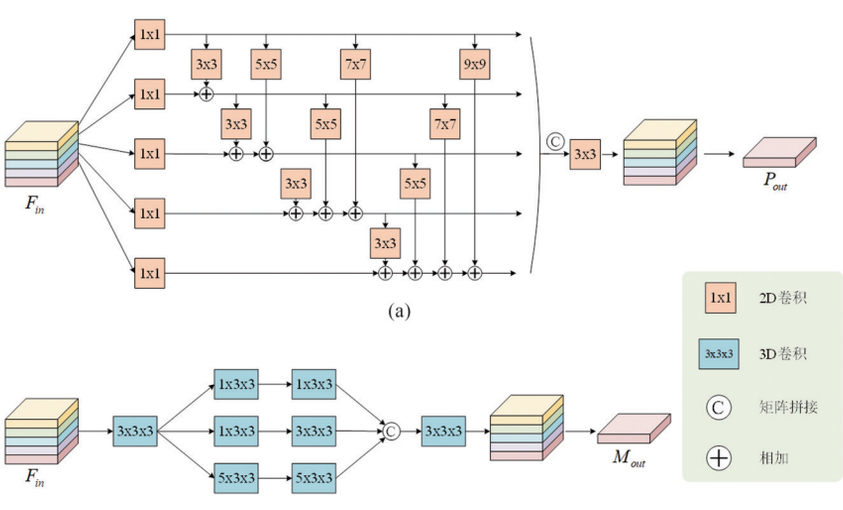

图2 渐进时序特征累积模块：(a)P2DConv模块架构；(b)M3DConv模块架构

Fig. 2 Progressive temporal accumulation module: (a)architecture of the P2DConv module; (b)architecture of the M3DConv module

将卷积结果的通道维度拼接到一起,并使用一个3 × 3 × 3的3D卷积融合不同时间跨度的时序特征,最终通过切片操作得到第t帧中的目标特征。需要注意的是,为了减少网络参数量,本文采用(2 + 1)D卷积代替3D卷积操作[28]。如图2(b)所示,F_{in}^{i}代表输入M3DConv模块的第i帧特征,M_{out}为M3DConv的输出,计算过程如下所示:

$$F_{\text{s}} = \text{cat}(\left[F_{\text{in}}^{t-4},F_{\text{in}}^{t-3},F_{\text{in}}^{t-2},F_{\text{in}}^{t-1},F_{\text{in}}^{t}\right]) \quad , \quad (7)$$

$$M_1 = C_{1\times1\times1}(C_{1\times3\times3}(F_{\text{s}})) \quad , \quad (8)$$

$$M_2 = C_{3\times1\times1}(C_{1\times3\times3}(F_{\text{s}})) \quad , \quad (9)$$

$$M_3 = C_{5\times1\times1}(C_{1\times3\times3}(F_{\text{s}})) \quad , \quad (10)$$

$$M_{\text{out}} = \sigma(C_{3\times3\times3}(\text{cat}(\left[M_1,M_2,M_3\right]))[-1]) \quad , (11)$$

其中,$C_{a\times b\times c}$代表卷积核大小为$a \times b \times c$的3D卷积操作。

将通过两种方式提取到的时序特征通过一个sigmoid层,并使用短连接汇聚到检测帧的特征上,从而实现时序特征的提取。最终,经过PTAM模块的输出F_e如式(12)所示:

$$F_e = F_{\text{in}} + F_{\text{in}} \odot P_{\text{out}} + F_{\text{in}} \odot M_{\text{out}} \quad , \quad (12)$$

1.3　多尺度空间特征融合模块

考虑到红外弱小目标大小在1 × 1到6 × 6之间[25],容易淹没在降采样操作中。本文设计了多尺度空间特征融合模块(MSFM)用于融合不同深度的红外弱小目标特征,使目标特征在深层网络中得以更好地留存,兼顾深层语义信息和浅层细节信息,并使用自注意力模块[29]获取全局感受野,提高弱小目标检测效果。

具体而言,MSFM模块的输入为经PTAM模块增强后的不同阶段特征,即前文提到的F_e^j,$j \in [1,4]$,对4个阶段增强后的特征执行交互操作,如图1(c)所示。对于输入的4个阶段的特征,它们的深度与尺寸各不相同,为了提取空间特征信息,本文首先将这些特征通过若干个3 × 3卷积降采样到最小特征层大小,再根据其通道维度进行拼接,拼接后的特征为F_m,计算过程如下所示:

$$F_m^j = C_{3\times3}(F_e^j) \quad , \quad (13)$$

$$F_m = \text{cat}(\left[F_m^1,F_m^2,F_m^3,F_m^4\right]) \quad . \quad (14)$$

接着,为了实现多尺度空间特征融合,本文设计了三个具有不同感受野的卷积层进行并行处理,卷积核的大小分别为1 × 1、3 × 3、5 × 5,并且设置合适的填充和步幅参数对F_m特征分别进行2倍、3倍、4倍的下采样,从而兼顾到不同大小的目标。将卷积得到的结果与原始特征按照通道维度进行拼接得到F_n,计算过程如下:

$$F_n = \text{cat}([C_{1\times1}(F_m),C_{1\times1}(F_m),C_{3\times3}(F_m),C_{5\times5}(F_m)]) \quad . \quad (15)$$

为了获取全局感受野,对于多尺度卷积后的特征F_n,将其H、W维度展开,形成一个二维特征F_a,作为自注意力模块的键向量k和值向量v,F_m作为查询向量q,通过自注意力模块获得输出F_g:

$$F_g = \text{softmax}(\frac{F_m W_q \cdot F_a W_k}{\sqrt{\text{C}}}) \cdot F_a W_v \quad , \quad (16)$$

其中,W_q、W_k和W_v为卷积核,C为F_m的通道维度大小,softmax(·)为归一化指数函数,$F_m \in \mathbb{R}^{C\times W\times H}$,$F_a \in \mathbb{R}^{C\times(\sum_1^4 W\times H)}$。

对于自注意力模块的输出F_g,为了更好地保存局部细节,首先将其通过两个3 × 3的卷积层进行特征提取,接着按照特征拼接前的维度重新将F_g划分为四个阶段的特征,并加到融合前的特征上,此时认为经过多尺度卷积、自注意力模块的特征F_g具有全图的感受野,融合前的特征F_e为局部特征。为了融合每个阶段的局部特征F_e和全局特征F_g,将全局特征上采样到原始大小,使用两个1 × 1卷积层分别对局部特征和全局特征进行归一化处理,并将其乘积通过sigmoid函数加权到原始特征上得到最终结果F_d^i,计算过程如下所示:

$$F_g = C_{3\times3}(C_{3\times3}(F_g)) \quad , \quad (17)$$

$$F_d^i = \text{sigmoid}\left(C_{1\times1}(F_e^i)\cdot C_{1\times1}(\text{upsample}(F_g))\right)\cdot F_e^i + F_e^i \quad , \quad (18)$$

其中,$i \in [1,4]$,代表4个阶段的特征,upsample(·)为上采样操作。MSFM模块输出的特征将作为后续解码器的输入,以得到最终的检测结果。

1.4　数据集制作

基于深度学习的算法对数据集的质量、数量及场景多样性有较高的要求。目前,公开可用的多帧红外弱小目标数据集较为稀缺,并且现有数据集没有考虑红外弱小目标的大小、方向、运动轨迹等特性,阻碍了多帧红外弱小目标检测的发展。本文构建了一个半仿真的多帧红外弱小目标数据集SHU-MIRST,该数据集图像分辨率为384 × 384,由100段序列组成,每段序列包含100帧,共10 000张图像,提供了中心点、边界框以及掩码标注信息,包括城市、天空、河流、植被等场景,该数据集具有多种目标模板、目标运动方式、背景运动类型及典型场景。

SHU-MIRST数据集的仿真流程如图3所示,包

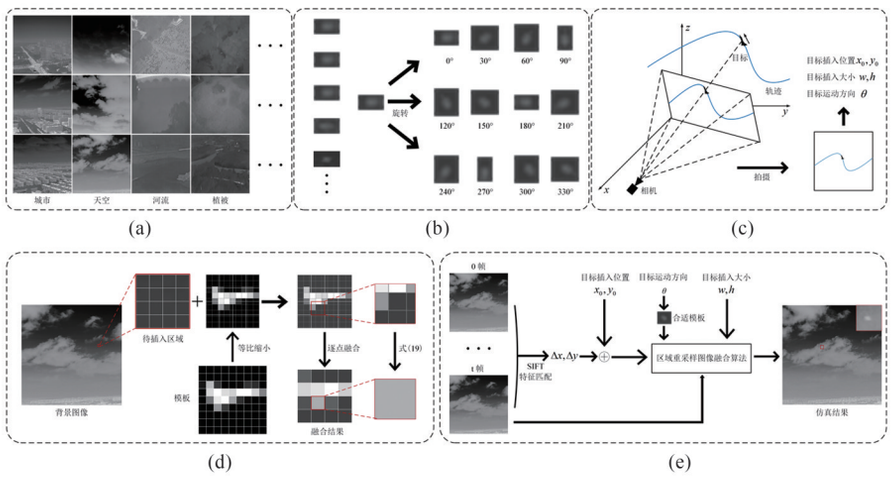

图3 SHU-MIRST数据集仿真流程:(a)背景拍摄;(b)目标模板制作;(c)目标三维建模;(d)区域重采样图像融合算法;(e)目标模板插入

Fig. 3 SHU-MIRST dataset simulation flowchart: (a)background shooting; (b)target template production; (c)target 3D modeling; (d)image fusion algorithm for region resampling; (e)target template embedding

括背景拍摄、目标模板收集、目标三维软件建模、目标模板插入四个步骤。

红外背景图像由波长范围为8~14μm、搭载于无人机上的非制冷氧化钒型红外摄像头拍摄,为了保证背景的多样性,本文针对无人机运动方式、摄像头朝向、相机云台运动方式、拍摄时天气、拍摄地温度等多种情况,多次拍摄得到了134段背景序列图像,并筛选掉高度相似、背景模糊的序列,保留了100段不同场景的序列。考虑到背景运动速度的多样性,根据无人机的运动状态,本文以2~5帧的帧间间隔对拍摄数据进行抽帧,拍摄的背景如图3(a)所示。

对于红外弱小目标模板,如图3(b)所示,考虑到真实的红外目标特性,本文从自主拍摄的数据及公开的红外弱小目标数据集中收集了30种红外弱小目标模板,并以30°为间隔对原始目标模板进行旋转,得到不同朝向的红外弱小目标模板。

本文使用三维建模软件仿真红外弱小目标的运动轨迹,如图3(c)所示,通过三维软件中的摄像头拍摄仿真目标的运动情况。具体而言,本文使用三维的贝塞尔曲线模拟目标运动轨迹,通过设置贝塞尔曲线控制点调整轨迹,获得每一帧中目标需要插入的粗略位置、插入大小及目标运动方向。

对于目标插入阶段,如图3(e)所示,考虑到随着拍摄平台的移动,图像中背景也会产生偏移,目标的运动轨迹应是目标本身运动与背景偏移的结合,因此本文使用尺度不变特征变换(scale invariant feature transform, SIFT)[30]算子对背景图像进行特征点匹配,求出每帧图像与第一帧之间的位置偏移,并将该偏移加到目标的粗略位置上,得到目标需要插入的具体位置。目标的朝向应与目标运动方向相同,选择对应朝向的目标模板插入背景中。

高质量的仿真数据集需要保证目标插入背景后符合物理特性上的合理以及视觉特征上的真实。为了保证目标物理特性上合理,本文在目标上施加了一个与目标外接矩形相同大小的高斯模糊函数,保证目标图像融合的平滑度[25],同时在整个图像上施加一个符合高斯分布的噪声,该噪声的均值为0,标准差为5。为了保证目标视觉特征上的真实,如图3(d)所示,本文设计了一种基于区域重采样的图像融合算法,根据模板图像和待插入区域的映射关系,一个目标模板像元可能会跨越多个背景像元,多个目标像元也可能会落入同一个背景像元内,因此,融合后像元的灰度值应同时取决于对应背景像元的灰度值以及落入该像元的多个目标模板像元的灰度值,上述过程可用式(19)表示:

$$I_{(x,y)} = \frac{\sum_{i,j} T_{(i,j)} \cdot S^{\mathrm{T}}_{(i,j)} + B_{(x,y)} \cdot S^{\mathrm{B}}_{(x,y)}}{\sum_{i,j} S^{\mathrm{T}}_{(i,j)} + S^{\mathrm{B}}_{(x,y)}} \quad , \quad (19)$$

其中,$I_{(x,y)}$为融合后图像(x,y)处的灰度值,$T_{(i,j)}$为目标模板中(i,j)处的灰度值,$S^{\mathrm{T}}_{(i,j)}$表示目标模板中(i,j)像元占融合后图像(x,y)像元的面积,$B_{(x,y)}$为插入前背景图像(x,y)处的灰度值,$S^{\mathrm{B}}_{(x,y)}$为背景图像中(x,y)像元占融合后图像(x,y)像元的面积,由此可

得$\sum_{i,j} S^{\mathrm{T}}_{(i,j)} + S^{\mathrm{B}}_{(x,y)} = 1$。

通过以上的步骤,数据仿真过程考虑了目标运动的方向、轨迹及透视关系等情况,实现了红外弱小目标数据集的高质量仿真。本文使用平均信杂比(mean signal to clutter ratio, mSCR)评价所提出数据集不同序列的检测难度,SCR为目标灰度值与周围背景区域灰度值差的归一化值,mSCR为一个序列中所有图像SCR的均值,SCR计算公式如式(20)所示:

$$\mathrm{SCR} = \left| \frac{\mu_{\mathrm{T}} - \mu_{\mathrm{B}}}{\sigma_{\mathrm{B}}} \right| \quad , \quad (20)$$

其中,μ_{T}为目标的均值,μ_{B}和σ_{B}为背景区域的均值与方差,本文取目标外接框向四周分别扩充20像素作为背景区域。

SHU-MIRST数据集的目标大小分布及序列mSCR分布如图4(a)和图4(b)所示,可以看出,本文仿真的红外弱小目标大小在1~36像素之间,且序列mSCR的分布合理,涵盖了1~10之间不同的mSCR,并有约一半序列的mSCR在3以下,说明SHU-MIRST数据集目标的大小及亮度设置合理,满足红外弱小目标检测算法对于数据集的要求。图5展示了SHU-MIRST数据集中部分序列的目标运动轨迹,图中目标运动轨迹为目标自身运动与背景位移的复合,轨迹点的大小指示了目标在图像中的大小,体现了目标距离摄像头的远近,且目标大小的变化与目标的运动轨迹是相符的。

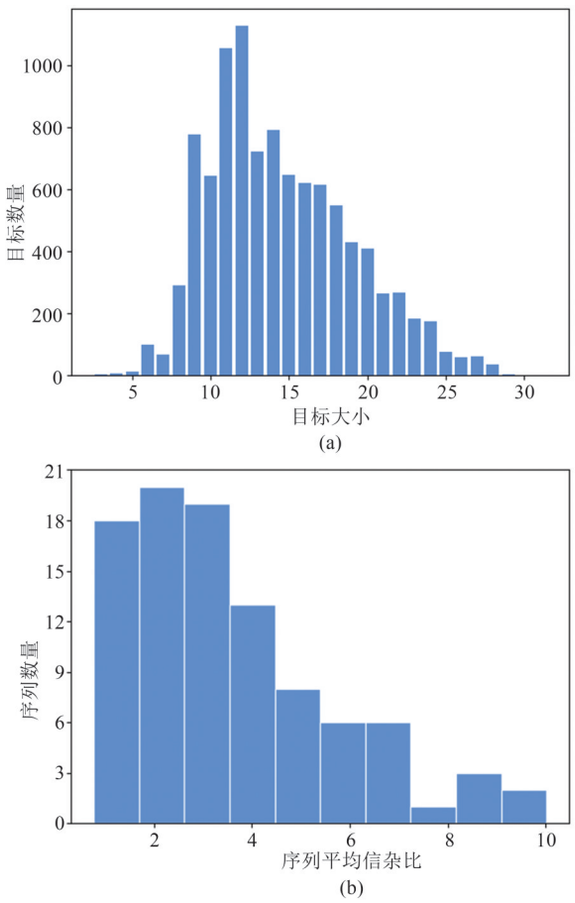

图4　SHU-MIRST数据集统计信息:(a)目标大小分布图;(b)序列平均信杂比分布图

Fig. 4　SHU-MIRST dataset statistical information: (a) distribution of target sizes; (b) distribution of mean SCR

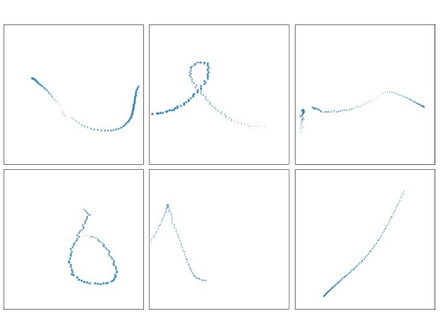

图5　SHU-MIRST数据集目标运动轨迹样例图

Fig. 5　Examples of target motion trajectory in the SHU-MIRST dataset

2　实验分析

2.1　评价指标

本文采用红外弱小目标检测领域常见的交并比(intersection over union, IoU)、检测概率(probability of detection, Pd)和虚警率(false-alarm rate, Fa)作为评价指标[11]。此外,本文中使用受试者工作特征(receiver operating characteristic, ROC)曲线评估不同检测算法的性能,各个评价指标的定义与公式如下所示:

(1)交并比:交并比是图像分割领域常见的指标,用于评估检测到目标形状的精度,越高表示算法检测目标形状的能力越好。交并比是预测结果与标签之间交集区域面积和两者并集区域面积的比值,计算公式如下:

$$\mathrm{IoU} = \frac{A_{\mathrm{inter}}}{A_{\mathrm{union}}} \quad , \quad (21)$$

其中,A_{inter}为交集区域面积,A_{union}为并集区域面积。

(2)检测概率:检测概率是一个目标级别的评估指标,它评估了算法准确寻找目标及定位目标的性能,越高表示算法定位目标的能力越好。检测概率的定义如下:

$$\mathrm{P_d} = \frac{T_{\mathrm{TP}}}{T_{\mathrm{ALL}}} \quad , \quad (22)$$

其中,T_{TP}代表正确检测的目标数,T_{ALL}代表标签中

所有目标的数量。

(3)虚警率:虚警率用于评估算法抑制虚假警报的能力,越低表示算法抑制虚警的能力越好。虚警率是错误检测目标的数量与所有像素个数的比值,定义如下:

$$F_{\mathrm{a}}=\frac{T_{\mathrm{FP}}}{\sum_{i=1}^{N}H_i\times W_i}\quad,\quad(23)$$

其中,T_{FP}代表错误检测的目标数,$H_i\times W_i$是第i张输入图像的像素数,N是测试集图像的数量。

(4)ROC曲线:ROC曲线用于评估假阳性率(false positive rate, FPR)与真阳性率(true positive rate, TPR)之间的变化趋势,它展示了不同阈值下检测器的整体性能,ROC曲线下的面积(area under curve, AUC)越大表示算法的整体性能越好。TPR和FPR的定义如下:

$$\mathrm{TPR}=\frac{\mathrm{TP}}{P},\ \mathrm{FPR}=\frac{\mathrm{FP}}{P}\quad,\quad(24)$$

其中,TP表示真正例像素数,FP为假正例像素数,P表示预测结果中正例像素数。

在评估过程中,本文认为如果预测的目标的质心与标签中目标的质心偏差小于预定义的偏差阈值D_{thresh},则认为这些该目标是正确检测的目标;如果质心偏差大于D_{thresh},则认为这些目标是错误检测的。本文中,D_{thresh}被设置为3。

2.2 实验设置

本文在所提出的SHU-MIRST数据集和IRDST-Real数据集[25]评估了本文的方法。IRDST-Real数据集为IRDST数据集中真实拍摄的部分,包含85段序列,共40 650帧图像,图像分辨率为720×480及934×696。两个数据集都被分为训练集和测试集,比例约为8:2。其中IRDST-Real数据集原有的数据划分方法仅适用于单帧算法,因此本文使用前65段序列作为训练集,后17段序列作为测试集。在图像输入网络前,图像分辨率被调整为512×512,并进行旋转、反转、裁剪等数据增强操作。

本文中所有基于深度学习的方法都是基于Pytorch实现的,计算设备采用Intel Xeon E5-2683 CPU @ 2.10 GHz以及两块Nvidia Titan Xp GPU。在训练过程中,本文使用Adam优化器对网络进行36轮迭代训练,初始学习率设置为0.001,并且每三轮迭代学习率会衰减一半,本文使用Kaiming初始化为网络中所有的卷积层进行初始化。为了缓解目标与背景之间的不平衡,本文采用Soft-IoU损失来训练本文的模型。

2.3 对比实验及分析

为了验证提出模型的效果,本文将PSTFNet与多种基于模型驱动的传统算法进行对比,包括单帧传统算法如Top-Hat[4]、PSTNN[8]、WSLCM[6]以及IPI[7]算法,以及多帧传统算法如WSNM-STIPI[18]、IMNN-LWEC[31]以及ASTTV-NTLA[17]算法。同时与目前较为先进的深度学习算法进行对比,包括单帧的深度学习算法如DNANet[11]、ISNet[13]、UIUNet[12]、RDIAN[25]算法,以及多帧深度学习算法SSTNet[21]、DNANet-DTUM[22]及ResUNet-DTUM[22]算法。所有的传统方法都采用其默认参数实现,所有基于深度学习的方法都采用0.5作为阈值。其中,SSTNet算法为目标检测算法,为了公平比较,本文使用检测框的中心点作为预测目标的质心,计算Pd及Fa指标,不计算IoU指标。

2.3.1 定量结果分析

对于所提出的数据集SHU-MIRST,为了考察所提出模型对于目标不同强度时的检测效果,本文计算了测试集中每段序列的mSCR,并根据mSCR将测试集分为mSCR≤3和mSCR>3两部分,分别评估所提出方法及对比方法的IoU、P_{d}以及F_{a}指标,结果如表1所示,其中最好的指标加粗表示,次好的指标加下划线表示,7种基于深度学习方法的ROC曲线如图6所示。可以看出,与其他方法相比,本文提出的方法在全部序列上具有最好的效果。对于使用单帧及多帧的传统方法,本文提出的PSTFNet在P_{d}、F_{a}以及IoU指标上具有显著的提升,这是由于传统的检测方法过于依赖手工设计的特征,并且此类方法中存在大量需要人工选择的超参,通常只能针对于特定场景检测,对于场景更有挑战、目标信杂比变化剧烈的SHU-MIRST数据集,此类算法的检测性能被极大约束。对于基于数据驱动的深度学习方法,使用多帧的深度学习算法P_{d}和IoU优于使用单帧的深度学习算法5%以上,这是由于使用多帧检测的算法可以使用多帧图像抑制背景噪声,从而获得更好的检测效果。对于使用多帧的深度学习算法,本文提出的PSTFNet算法相较于DNANet-DTUM算法IoU和P_{d}分别高3.60%和4.69%,这是由于DNANet-DTUM算法仅对主干网络提取到的特征进行位置编码,并没有将提取到的目标运动信息应用于主干网络解码操作中,对于本文提出的算法,使用PTAM模块提取到目标的时序特征

表 1　不同算法在 SHU-MIRST 数据集上的定量比较

Table 1　Quantitative comparison of different algorithms on the SHU-MIRST dataset

方法	SHU-MIRST(mSCR≤3)			SHU-MIRST(mSCR>3)			SHU-MIRST(all)		
	IoU/(%)	P_d/(%)	F_a(10^{-6})	IoU/(%)	P_d/(%)	F_a(10^{-6})	IoU/(%)	P_d/(%)	F_a(10^{-6})
Top-Hat	0.00	0.83	856.81	2.67	11.17	185.81	0.93	4.45	621.96
IPI	0.19	2.75	80.23	2.72	14.75	57.34	1.08	6.95	72.22
PSTNN	0.00	0.14	122.94	2.41	10.31	129.36	0.84	3.70	125.19
WSLCM	0.45	45.80	4 623.48	5.61	80.22	3 562.33	2.26	57.85	4 252.08
WSNM-STIPI	9.61	53.61	35.95	13.67	66.01	36.35	11.03	57.95	36.09
IMNN-LWEC	0.00	0.00	32.24	0.12	3.96	139.76	0.04	1.38	69.87
ASTTV-NTLA	0.00	0.30	80.29	0.40	5.02	34.67	0.14	1.95	64.34
RDIAN	36.40	52.07	36.46	67.36	84.84	15.40	47.23	63.54	29.09
DNANet	38.74	61.82	39.75	74.19	85.56	10.60	51.14	70.13	29.55
ISNet	36.17	49.01	13.15	65.33	82.46	13.23	46.38	60.72	13.18
UIUNet	43.54	55.93	11.88	74.29	90.61	3.28	54.30	68.07	8.87
SSTNet	–	64.09	18.55	–	93.56	8.92	–	74.40	15.17
ResUNet-DTUM	51.78	68.51	13.32	75.53	93.83	6.60	60.09	77.37	10.97
DNANet-DTUM	51.91	69.19	21.63	76.71	93.98	2.67	60.59	77.86	15.00
Ours	57.68	75.80	10.80	76.28	95.08	2.69	64.19	82.55	7.97

后,会将时序特征输入解码器中,从而获得更好的检测效果。

对于背景复杂、目标更加暗弱的 mSCR≤3 的部分序列上,本文提出算法 IoU 比使用多帧的深度学习如 DNANet-DTUM 以及 ResUNet-DTUM 算法高 5% 以上,比使用单帧的深度学习算法高 10% 以上,而对于背景信息简单、目标较为显著的 mSCR>3 的部分序列上,本文提出的 PSTFNet 与 DNANet-DTUM 算法各项指标接近,说明本文提出的 PSTFNet 对低信杂比目标检测时具有更大的优势。

对于真实拍摄的公开数据集 IRDST-Real,对比实验结果如表 2 所示,从表中可以看出,本文提出的 PSTFNet 相较于 DNANet-DTUM 算法 IoU 和 P_d 分别高 2.95% 和 4.22%,同时具有最低的 F_a,说明本文提出的方法在不同的数据集上都具有更好的性能,本文提出的 PSTFNet 具有较强的鲁棒性。

2.3.2　定性结果分析

如图 7 所示,本文从 SHU-MIRST 数据集中选取了 6 段序列,展示不同检测算法的输出图像,图 7 中(a)、(b)、(c)为 mSCR>3 序列图像的检测结果,(d)、(e)、(f)为 mSCR≤3 序列图像的检测结果,为了更好地展示结果,本文使用蓝色框标识出目标位置,并在图像右上角放大显示,使用黄色虚线圈出检测到的虚警目标。

如图 7 所示,除了图 7(a)所示序列为天空背景,检测较为简单,其余序列传统算法都有大量虚警目标,对于 mSCR>3 的部分序列图像,目标强度较强,所有深度学习算法均可正确检测出目标,仅 RDIAN 算法具有少量虚警点。对于 mSCR≤3 的部分序列,背景较为复杂、目标暗弱,对比的深度学习算法出现漏检,并具有较高的虚警率,本文提出的 PSTFNet 能在保持较高检测概率的同时,具有更低的虚警率,PSTFNet 可以使用连续帧中的时序信息增强检测帧中目标的特征强度,在平均信杂比低的场景更具有优势。

2.4　消融实验

为了验证本文提出的 PTAM 模块以及 MSFM 模块的效果,找到 PSTFNet 网络中模块的最佳配置,本文在 SHU-MIRST 数据集上进行了消融实验,结果如表 3 所示,与主干网络相比,单独增加 PTAM 模块后,IoU 和 P_d 分别提高 20.93% 和 25.82%,单独增加 MSFM 模块后,IoU 和 P_d 分别提高 3.41% 和 6.37%,同时使用 PTAM 与 MSFM 模块,IoU 和 P_d 分别提高 27.02% 和 32.14%,实验表明,引入 PTAM 模块和 MSFM 模块能有效提高网络性能。

本文接下来将对 PTAM 模块的具体配置与组成进行更详细的消融实验,本文设计了两组实验验证 PTAM 模块的有效性。首先验证 PTAM 模块的数量对网络性能的影响,本文逐层移除 PSTFNet 中每一阶段的 PTAM 模块,实验结果如表 4 所示,随着 PTAM 模块的移除,网络的各项性能开始下降,PSTFNet 相较于不添加任何 PTAM 模块的网络结构,在

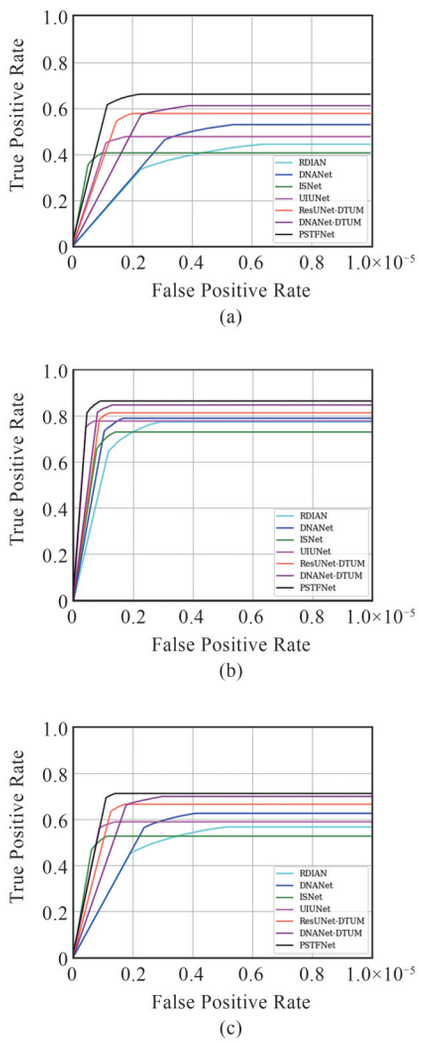

图6 PSTFNet在不同信杂比条件下的ROC曲线：(a)mSCR≤3；(b)mSCR>3；(c)所有序列

Fig. 6 ROC curves of PSTFNet under different mSCR: (a) mSCR≤3; (b) mSCR>3; (c) all sequences

IoU和P_d上分别取得了23.61%和25.77%的增益，同时从表4中可以看出，第一层的PTAM模块对检测性能的影响最大，这是由于随着层级的加深，特征会进行下采样，较深层次中的特征维度较小，PTAM模块提取到的时间特征较少。该实验表明了所提出的PTAM模块可以有效地利用连续多帧的特征增强检测帧中目标的特征，获得更好的检测性能。

此外，PTAM模块中包括P2Dconv模块和M3Dconv模块两部分，本文分别对两部分进行了消融实验，结果如表5所示，仅使用P2Dconv模块提取时序信息使网络分别获得18.16%和18.21%的IoU及P_d提升，仅使用M3Dconv模块获得的IoU和P_d提升分别为13.25%和9.73%，而同时使用两个模块获得的提升为23.61%和25.77%，说明单独使用P2Dconv模块或M3Dconv模块都无法提取全部的时序特征，两者共同作用能增强时序特征的提取效果。

表2 不同算法在IRDST-Real数据集上的定量比较

Table 2 Quantitative comparison of different algorithms on the IRDST-Real dataset

方法	IoU/(%)	P_d/(%)	F_a(10^{-6})
Top-Hat	5.39	24.66	489.28
IPI	9.38	36.55	37.11
PSTNN	5.79	17.58	57.05
WSLCM	4.92	37.44	1 389.62
WSNM-STIPI	17.79	59.66	38.92
IMNN-LWEC	3.10	7.99	641.05
ASTTV-NTLA	0.27	1.82	395.59
RDIAN	47.69	86.04	3.95
DNANet	50.34	82.57	5.15
ISNet	50.35	82.38	3.86
UIUNet	48.73	81.54	<u>2.70</u>
SSTNet	-	85.11	4.83
ResUNet-DTUM	50.31	86.19	2.87
DNANet-DTUM	<u>50.98</u>	<u>87.03</u>	3.62
Ours	53.93	91.25	2.26

MSFM模块使用不同的卷积核对F_m特征进行不同尺度的下采样，并通过自注意力模块进行多尺度空间特征融合，以增强深层网络中弱小目标的特征表征，同时获取全局感受野。为了验证多尺度卷积核及自注意力模块的作用，本文针对MSFM模块设计了两种变体：(1)将MSFM模块中多尺度卷积核(Multi-scale Convolution, MC)全部替换为大小为3×3的卷积核，并设置下采样倍率为2，其余部分不变，该变体记作PSTFNet w/o MC；(2)去除MSFM模块中的自注意力模块(Self-Attention, SA)，将不同尺度的特征上采样到F_m大小并按元素相加，作为具有全局感受野的特征F_g，其余部分不变，该变体记作PSTFNet w/o SA。本文分别对两种变体进行了消融实验，结果如表6所示，单独移除MSFM模块中的多尺度卷积核，IoU和P_d分别下降了3.67%和3.39%，单独移除MSFM模块中的自注意力模块，IoU和P_d分别下降了1.72%和1.17%，移除整个MSFM模块后，网络下降了6.09%和6.32%的IoU和P_d指标，证明了MSFM模块中多尺度卷积核及自注意

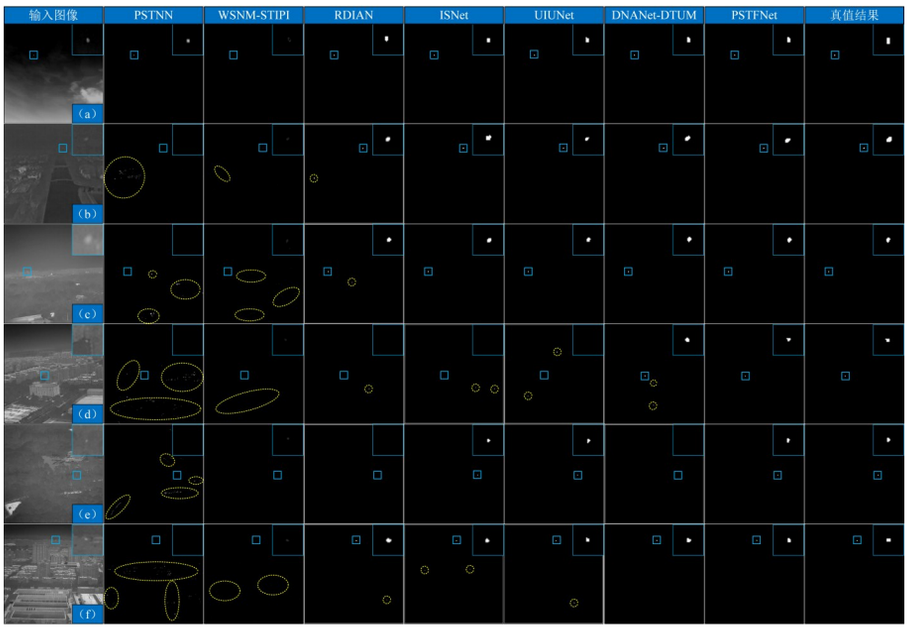

图 7　PSTFNet 与 6 种对比算法在 SHU-MIRST 数据集上的定性实验结果对比图

Fig. 7　Qualitative comparison results of PSTFNet and 6 benchmark algorithms on the SHU-MIRST Dataset

表 3　PSTFNet 组成模块消融实验结果

Table 3　Results of the ablation experiment for the PSTFNet component modules

方法	IoU/(%)	P_d/(%)	$F_a(10^{-6})$
Backbone	37. 17	50. 41	24. 89
Backbone + PTAM	58. 10	76. 23	12. 13
Backbone + MSFM	40. 58	56. 78	11. 08
PSTFNet	64. 19	82. 55	7. 97

表 4　PTAM 模块层数消融实验结果

Table 4　Results of the PTAM layer ablation experiment

方法	IoU/(%)	P_d/(%)	$F_a(10^{-6})$
PSTFNet w/o PTAM	40. 58	56. 78	11. 08
PSTFNet w/o PTAM L123	43. 26	58. 95	13. 14
PSTFNet w/o PTAM L12	49. 04	64. 02	9. 95
PSTFNet w/o PTAM L1	55. 33	72. 60	10. 29
PSTFNet	64. 19	82. 55	7. 97

表 5　PTAM 模块组成消融实验结果

Table 5　Results of the PTAM composition ablation experiment

方法	IoU/(%)	P_d/(%)	$F_a(10^{-6})$
PSTFNet w/o PTAM	40. 58	56. 78	11. 08
PSTFNet w/o M3DConv	58. 74	74. 99	7. 17
PSTFNet w/o P2Dconv	53. 83	66. 51	13. 23
PSTFNet	64. 19	82. 55	7. 97

力模块的作用。

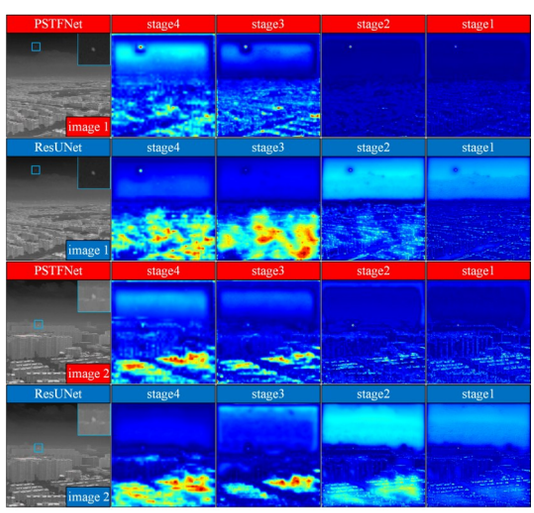

图 8　PSTFNet 与主干网络 ResUNet 在解码器不同阶段的特征响应图

Fig. 8　Visualization map of PSTFNet and the backbone network ResUNet at different stage of decoder

最后,本文提出的 PSTFNet 算法通过 PTAM 模块利用多帧图像对检测帧目标特征进行增强,并使用 MSFM 模块融合不同层次的特征。从图 8 中可以看出,对于深层网络(如第 3、4 阶段),红外弱小目标在 PSTFNet 中的特征与 ResUNet 相比在响应图中响应幅值更高,并且每个阶段的特征响应幅值都高于 ResUNet,说明了本文提出的 PSTFNet 增强了深层网

络对于红外弱小目标的关注能力，缓解了由于网络降采样导致的目标特征丢失。

表6 MSFM模块组成消融实验结果

Table 6 Results of the MSFM composition ablation experiment

方法	IoU/(%)	P_d/(%)	F_a(10^{-6})
PSTFNet w/o MSFM	58.10	76.23	12.13
PSTFNet w/o MC	60.52	79.16	9.25
PSTFNet w/o SA	62.47	81.38	17.18
PSTFNet	64.19	82.55	7.97

3 结论

本文提出了一种基于渐进时空特征融合的多帧红外弱小目标检测网络，该网络提取连续多帧图像中目标的时序特征以增强检测帧中的目标，同时融合了不同深度的目标特征，避免了深层网络中目标的丢失。设计了一种渐进时序特征累积模块，使用感受野渐进增大的2D卷积以及多尺度的3D卷积提取目标的时序特征，同时设计了一个多尺度空间特征融合模块，使用不同大小的卷积核融合红外弱小目标的空间特征，获得同时具有目标细节信息以及高级语义信息的目标特征，实现红外弱小目标的鲁棒性检测。此外，本文构建了一个半仿真的多帧红外弱小目标数据集SHU-MIRST，该数据集考虑了目标的自身运动及背景位移的复合，涵盖了多种目标模板、运动轨迹及场景类型，同时设计了一种基于区域重采样的图像融合算法，实现了高质量的红外目标嵌入，具有较好的仿真效果。在SHU-MIRST数据集和公开的IRSTD-Real数据集上的测试表明，与多个主流算法相比，所提出网络的交并比、检测概率及虚警率指标均有更好的表现，尤其是在目标强度较弱、背景较为复杂的序列上，验证了本文提出的算法在红外弱小目标检测方面的有效性和鲁棒性。本文所提出的数据集及相关代码会在不久后公开在https://github.com/danZengSHU/PSTFNet。

References

[1] Zhang M, Zhang R, Zhang J, *et al*. Dim2Clear network for infrared small target detection [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, **61**: 1-14.

[2] Luo K. Space-based infrared sensor scheduling with high uncertainty: issues and challenges [J]. *Systems Engineering*, 2014, **18**(1): 102 - 113. DOI: 10.1002/sys.21295.

[3] Deshpande S D, Er M H, Ronda V, *et al*. Max-mean and max-median filters for detection of small targets[C]. Signal and Data Processing of Small Targets 1999. SPIE, 1999, **3809**: 74-83.

[4] Zeng M, Li J, Peng Z. The design of Top-Hat morphological filter and application to infrared target detection [J]. *Infrared Physics & Technology*, 2006, **48**(1): 67-76. DOI: 10.1016/j.infrared.2005.04.006.

[5] Chen C L P, Li H, Wei Y, *et al*. A local contrast method for small infrared target detection [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2013. DOI: 10.1109/TGRS.2013.2242477.

[6] Han J, Moradi S, Faramarzi I, *et al*. Infrared small target detection based on the weighted strengthened local contrast measure [J]. *IEEE Geoscience and Remote Sensing Letters*, 2020, PP(99): 1-5. DOI: 10.1109/LGRS.2020.3004978.

[7] Gao C, Meng D, Yang Y, *et al*. Infrared patch-image model for small target detection in a single image [J]. *IEEE Transactions on Image Processing*, 2013, **22**(12): 4996-5009. DOI: 10.1109/TIP.2013.2281420.

[8] Zhang L, Peng Z. Infrared small target detection based on partial sum of the tensor nuclear norm [J]. *Remote Sensing*, 2019, **11**(4): 382. DOI: 10.3390/rs11040382.

[9] Li H, Wang N, Ding X, *et al*. Adaptively learning facial expression representation via cf labels and distillation [J]. *IEEE Transactions on Image Processing*, 2021, **30**: 2016-2028.

[10] Dai Y, Wu Y, Zhou F, *et al*. Asymmetric contextual modulation for infrared small target detection[C]. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2021: 950-959.

[11] Li B, Xiao C, Wang L, *et al*. Dense nested attention network for infrared small target detection [J]. *IEEE Transactions on Image Processing*, 2022, **32**: 1745-1758.

[12] Wu X, Hong D, Chanussot J. UIU-Net: U-Net in U-Net for infrared small object detection [J]. *IEEE Transactions on Image Processing*, 2022, **32**: 364-376.

[13] Zhang M, Zhang R, Yang Y, *et al*. ISNet: Shape matters for infrared small target detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 877-886.

[14] Li B, Wang Y, Wang L, *et al*. Monte Carlo linear clustering with single-point supervision is enough for infrared small target detection[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 1009-1019.

[15] Lin Z, Li B, 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* (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。**红外与毫米波学报**), 2022, **41**(6): 1102-1112.

[16] Li B, Wang L, Wang Y, *et al*. Mixed-precision network quantization for infrared small target segmentation [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[17] Liu T, Yang J, Li B, *et al*. Nonconvex tensor low-rank approximation for infrared small target detection [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, **60**: 1-18.

[18] Sun M A W. Infrared small target detection via spatial-

temporal infrared patch-tensor model and weighted Schatten p-norm minimization [J]. *Infrared Physics and Technology*, 2019, 102.

[19] 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.

[20] Yan P, Hou R, Duan X, *et al*. STDMANet: Spatio-temporal differential multiscale attention network for small moving infrared target detection [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, **61**: 1-16.

[21] Chen S, Ji L, Zhu J, *et al*. SSTNet: Sliced spatio-temporal network with cross-slice ConvLSTM for moving infrared dim-small target detection [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[22] Li R, An W, Xiao C, *et al*. Direction-coded temporal U-shape module for multiframe infrared small target detection [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[23] Hui B W, Song Z Y, Fan H Q, *et al*. A dataset for infrared image dim-small aircraft target detection and tracking under ground / air background[J/OL]. Science Data Bank (回丙伟, 宋志勇, 范红旗, 等. 地/空背景下红外图像弱小飞机目标检测跟踪数据集. **中国科学数据**), 2020,5(3).DOI:10.11922//csdata.2019.0074.2h.

[24] Sun X L, Guo L C, Zhang W L, *et al*. A dataset for small infrared moving target detection under clutter background [J/OL]. Science Data Bank(孙晓亮, 郭良超, 张文龙, 等. 复杂背景下红外弱小运动目标检测半仿真数据集. **中国科学数据**), 2022[2024-03-18].

[25] Sun H, Bai J, Yang F, *et al*. Receptive-field and direction induced attention network for infrared dim small target detection with a large-scale dataset IRDST [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, **61**: 1-13.

[26] Xiao X, Lian S, Luo Z, *et al*. Weighted Res-UNet for high-quality retina vessel segmentation[C]. 2018 9th International Conference on Information Technology in Medicine and Education (ITME). IEEE Computer Society, 2018, 327-331. DOI: 10.1109/ITME.2018.00080.

[27] Zhang L, Zhu G, Shen P, *et al*. Learning spatiotemporal features using 3dcnn and convolutional lstm for gesture recognition [C]. Proceedings of the IEEE International Conference on Computer Vision Workshops. 2017: 3120-3128.

[28] Li Y, Ji B, Shi X, *et al*. Tea: Temporal excitation and aggregation for action recognition [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 909-918.

[29] Vaswani A, Shazeer N, Parmar N, *et al*. Attention is all you need [J]. *Advances in Neural Information Processing Systems*, 2017, 30.

[30] Lowe D G. Object recognition from local scale-invariant features[C]. Proceedings of the Seventh IEEE International Conference on Computer Vision. IEEE, 1999, **2**: 1150-1157.

[31] Luo Y, Li X, Chen S, *et al*. IMNN-LWEC: A novel infrared small target detection based on spatial - temporal tensor model [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, **60**: 1-22.