文章编号:1001-5078(2023)11-1702-10　　·红外技术及应用·

远距离探测条件下红外序列中弱特征目标检测

李　淼[1],胡铭原[1],樊建鹏[1],林再平[1],高金艳[2],安　玮[1]
(1. 国防科技大学电子科学学院,湖南 长沙 410073;2. 中国空间技术研究院北京空间飞行器总体设计部,北京 100094)

摘　要:远距离探测情况下红外目标无纹理特征、易受噪声和杂波干扰,属于弱特征目标,检测难度大。将红外序列中的多尺度目标检测视为图像中的逐像素分割问题,更好适应复杂环境下的小尺度目标;提出了基于注意力全卷积的弱特征目标分割方法,通过构建时域显著图,将源自不同相机的图像在同一空间进行表征,提高模型的泛化能力和收敛速度;通过注意力机制自适应学习双模态多层次信息,强化对目标区域的特征学习能力。利用实测数据和仿真数据对检测方法在不同条件下的性能进行了对比分析,实验结果表明,所提方法能够有效提升多尺度弱特征目标检测性能。

关键词:红外弱特征目标;目标检测;全卷积神经网络;注意力机制
中图分类号:TP391　　**文献标识码:**A　　**DOI:**10.3969/j.issn.1001-5078.2023.11.013

Weak feature target detection in infrared sequence under remote imaging

LI Miao[1], HU Ming-yuan[1], FAN Jian-peng[1], LIN Zai-ping[1], GAO Jin-yan[2], AN Wei[1]
(1. College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China;
2. Institute of Spacecraft System Engineering, China Academy of Space Technology, Beijing 100094, China)

Abstract: Infrared targets have no texture featurein the case of remote imaging, and are susceptible to noise and clutter. It is difficult to detect such weak feature targets. An attention based full convolutional neural network is proposed, and the multiscale-target detection is regarded as a pixel-wise image segmentation problem to adapt small target extraction in complex backgrounds. The time domain saliency map is constructed, and the images generated by different cameras can be projected into the same representational space. It can improve the generalization ability of the network model. The attentional mechanism is used to learn bimodal and multi-level information, and to adaptively enhance the learning ability around targets. The performance comparison under different conditions is implemented based on real data and simulation data. The experimental results show that the proposed method can effectively improve detection performance for multiscale and weak feaure targets. Meanwhile, the neural network has quicker convergent rate.

Keywords: weak feature target; target detection; full convolutional neural network; attention mechanism

1　引　言

红外探测技术被广泛应用于军事和民用领域,例如空中交通管制、小型无人机探测、战场监视、机载红外搜索与跟踪、天基光学监视系统等[1-3],具有覆盖范围广、受天气影响小、可全天候工作、不主动辐射信号、抗干扰性高等优点。在远距离探测情况

基金项目:国家自然科学基金资助项目(No. 62001478);国防科技大学自主创新科学基金项目(No. 22-ZZCX-042)资助。
作者简介:李　淼,男,博士,副研究员,主要研究方向为智能光电感知。E-mail:lm8866@nudt.edu.cn
通讯作者:安　玮,女,博士,教授,博士生导师,主要研究方向为空间信息获取与处理。E-mail:anwei_nudt@126.com
收稿日期:2023-05-02

下,目标检测存在如下挑战[4-6]:

(1)目标在像平面上往往无纹理、几何、颜色等特征,可用信息少。

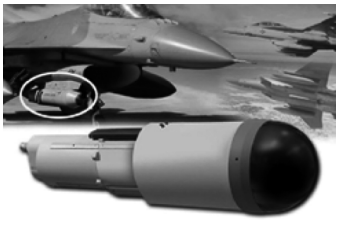
(2)远距离条件下成像分辨率低,主要呈现为斑点状,例如5像素,当观测距离较近时可达300像素,甚至更大,目标尺度变化范围大,形状不规则。

(3)红外探测机制主要感知热辐射,难以直观反映目标本体形态,且不同相机的分辨率、灵敏度等各不相同,目标和背景的差异性建模难度大。

(4)热辐射信号受大气衰减作用影响,目标被大量相机噪声、背景杂波等淹没,特别是容易被云层、树木、建筑物等干扰。

具备上述特点的目标可称之为弱特征目标。弱特征目标检测是远距离红外感知系统(图1)的首要环节,直接影响系统级性能,在传感器性能难以提升条件下,研究自适应不同场景的弱特征目标检测技术对增强远距离探测系统的整体效能十分重要。

(a)天基广域监视系统



(b)机载红外搜索跟踪系统

图1 远距离红外探测应用

Fig. 1 Remote infrared detection system

目前,目标检测一般包括两类方法:基于经验驱动的传统检测方法、基于数据驱动的智能检测方法。

基于经验驱动的传统检测方法,一般包括背景抑制、阈值分割等步骤,对人的经验依赖程度高,以不同滤波器的形式设置多种规则,例如顶帽(Top-hat)滤波[7]、LCM[8]、RLCM[9]等,滤波器的尺度和权重是预先设定的。此类方法的优点是可解释性好,数学模型清晰;缺点是主要适用于线性特征建模,对复杂非线性特征的建模能力弱,当目标形态与滤波器假设不同时,检测性能退化明显,难以在不同场景中自适应调整。

基于数据驱动的智能检测方法,模仿生物神经网络,建立深层人工神经网络,从海量数据中自动挖掘目标和背景的差异特征,目前主要以深度学习方法为代表。此类方法的优点是自动学习能力强,在学习过程中对人工干预需求少,通过多层神经网络级联可更好建模深层模糊特征;缺点是可解释性弱,针对性的神经网络架构设计是核心难点。

目前人工智能技术,在计算机视觉领域发展迅速,但所针对的目标主要是近景成像中的有形态物体(人脸、车辆、动物等),以输出合理的目标框为目的,例如FasterRCNN[10]、YOLO[11]、SSD[12]。由于红外监视的特殊用途,对弱特征目标(图2)检测的研究较少,上述方法难以直接应用于尺度较小的弱特征目标检测。

(a)弱特征目标 (b)有形态物体

图2 弱特征目标和有形态物体示例

Fig. 2 Example for weak feature target and shaped object

针对远距离探测条件下红外序列中弱特征目标检测问题,本文将目标检测视为逐像素分割问题,提出基于注意力全卷积的弱特征目标分割方法,采用自注意力学习等方式强化不同层级特征、不同相关性特征的自主融合学习,提升复杂条件下多尺度目标检测性能。

2 基于注意力全卷积的弱特征目标分割

2.1 整体算法架构

在计算机视觉中,目标检测指的是用边界框(Boundingbox)将目标大概位置框出,不追求精确到每个像素;语义分割指的是判决每个像素的属性,属于逐像素处理。考虑到弱特征目标的尺度相对较小,边界框过大则丧失检测意义,边界框过小则难以保证性能。因此,本文将弱特征目标检测问题直接建模为二分类图像分割问题,每个像素的属性包括两类(目标或背景),更好实现远距离探测下动态多尺度目标提取。

考虑到对弱特征目标分割的任务需求,受UNet启发[13],建立多尺度特征学习网络,更好适应动态多

尺度目标。主要思想是采用多卷积层和池化层进行多次下采样,每进行一次下采样感知野就扩大一次。通过多次感知野扩大,可学习不同尺度中的图像特征,有利于背景中较大范围内的结构特征学习(例如较大的云层、建筑物等),更好抑制连续性背景成分。

但与有形态物体不同,在弱特征目标检测中,每个像素的判决结果都十分有意义,因此需要通过连续反卷积实现上采样,更好的保留连续像素细节信息,如图 3。

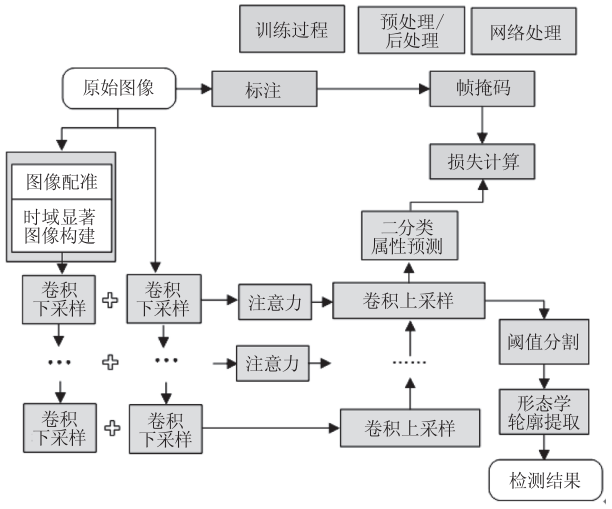


图 3 本文算法整体架构

Fig. 3 The overall architecture of the proposed algorithm

在网络框架中,主要步骤包括以下几点:

(1)图像配准:利用 SIFT 算法[14],从当前帧和参考图像中提取特征点,生成特征向量并进行匹配,根据匹配结果计算帧间几何变化,从而降低相机运动的影响。

(2)时域显著图构建:通过多帧差分预处理剔除明显背景成分,使网络集中学习较难区分的目标和杂波。实验表明,通过多帧差分大部分背景得到抑制,且多帧差分法相较于深度学习方法处理算法简单,能够提高整体算法的处理时效,一定程度上也强化了时域信息利用。

(3)双模数据特征学习:将原始图像和时域显著图分别进行卷积和下采样,基于时域显著图的学习突出时域变化成分之间的相关性,基于原始图像的学习突出空域相关性,二者结合更好学习背景和目标之间的时空差异。

(4)通过跳层连接进行特征融合:在多次下采样过程形成了多尺度浅层特征,在多次上采样过程中形成了多尺度深层特征。浅层特征更接近原始信息,细节丰富,但感知野小,非线性化表征程度低;深层特征感知野大,且非线性化表征程度高,但容易缺失细节特征。在多个网络层级设计跳层连接,将浅层特征与深层特征拼接,进行一体学习,实现更丰富维度特征融合。

(5)通过注意力机制进行自适应权重学习:为加快网络收敛速度,在融合原始图像和时域显著图学习结果的基础上,重点突出目标局部区域,设计自注意力机制,对于目标相关的疑似区域加强学习关注度,以有限参数聚焦目标区域。

(6)目标分割后处理:对网络输出预测结果进行二值化分割,并对疑似目标像素进行形态学滤波,获得完成目标像素。

2.2 时域显著图构建

为将不同相机生成的图像进行解耦,将多模态数据解耦至相近属性空间。本文通过预处理构建时域显著图,本质上是通过当前帧与序列中历史多帧分别进行差分预处理,突出所关注的目标运动或目标辐射强度变化的像素。红外图像中背景杂波起伏大,弱特征目标尺度较小且与背景界限不明显,传统单帧差分方法可能会出现目标重叠区域被抵消,背景未对齐被增强的情况。为了利用序列信息来降低上述误差,本算法进行多帧差分并对相应像素按位与叠加。

将与当前帧的前后间隔为 τ 的多帧做配准和差分,将叠加帧差图灰度化作为当前帧的背景抑制,流程图如图 4 所示。具体流程如下:

(1)对于当前帧图像 I_c,寻找在同序列中有步进 τ 的多帧图像 I_i。对于所用红外小目标序列数据集,如果取较远帧做差分,时间间隔过长,背景内部相对运动,造成无法抑制的背景杂波,引发虚警,降低检测准确率;如果取相邻帧做帧差,由于无人机等目标距离航拍设备远,常态运动速度较慢,目标会产生重叠,帧间相减会导致目标重叠部分被抵消,检测率降低。选取多帧 I_i,使其与 I_c 有步进间隔 τ,能减少重复特征并减少目标重叠情况,且多帧图叠加能增加目标的能量。因此,为了减少检测误差,提升检测准确率,本文采用多帧帧差进行背景抑制。

(2)对选取的 I_i,以当前帧为参考配准。由于采集序列图像时相机晃动等因素影响,实测数据的

相邻帧间的背景可能会有旋转、位移、拉伸。直接帧差背景抑制的背景难以匹配。对帧差对象 I_i 以当前帧做参考配准，再进行差分，降低相机运动的影响。

(3)对配准后的多 I_i 与当前帧分别做截 0 处理。与可见光图像相比，红外图像测量的是物体的热辐射信号，在图像中目标和背景等属于正叠加效应。因此，相减为负的被减方为背景，将这部分的灰度值截断为0。

(4)将帧差后的序列叠加得到当前帧差分图像。灰度图的像素数据范围是 0 -255 间的整数，多个像素叠加的结果可能会大于最大范围，为了防止叠加后像素值溢出，本文没有直接将图像像素的相应位置相加，而是对图像像素进行按位与操作。

对序列中的每一帧重复步骤(1) ~ (4)实现背景抑制。

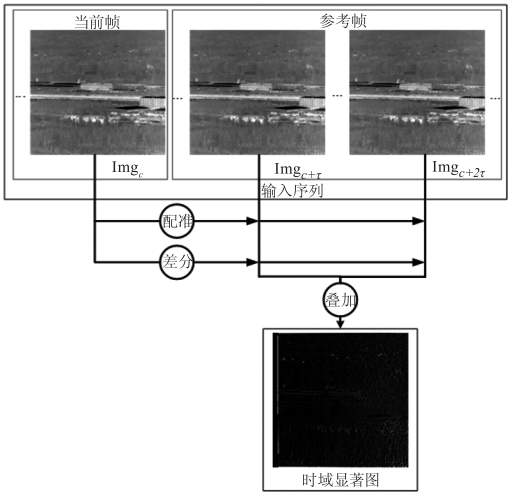
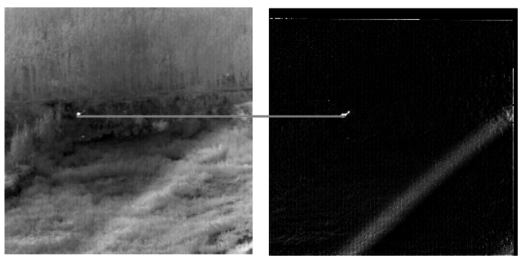


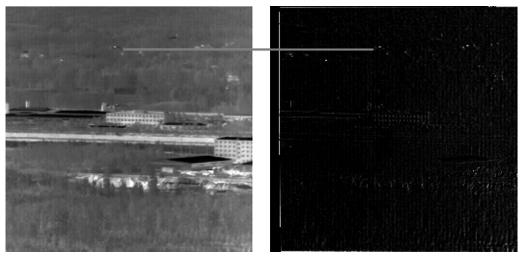
图 4 时域显著图构建流程图

Fig. 4 The flow chart of time domain saliency image

不同场景下构建的时域显著图如图 5 所示，可见针对不同的地物背景，目标并未丢失，且大部分背景得到抑制。且预处理相较于神经网络方法时延消耗非常小，能够提高算法整体处理时效。此外，通过帧间差分，构建的时域显著图，抑制了不同相机、不同场景的数据差异，提高网络鲁棒性。



(a) 场景1的显著图



(b) 场景2的显著图

图 5 时域显著成分示意图

Fig. 5 The time domain saliency composition

2.3 基于注意力机制的双模态特征学习

神经网络中的注意力机制与神经网络均为仿生技术。注意力机制以人体视觉选择注意力为蓝本，借鉴人体通过扫视聚焦关键区域信息的能力。由于注意力机制形式简便，通常不会提高模型的复杂度，被广泛应用于深度学习领域[15]。

注意力的计算公式如下：

$$I_iA(Q,S) = \sum_{i=1}^{L_s} R(Q,K_i) * V_i \tag{1}$$

Q 是目标中的某个元素；S 是输入序列；L_s 是 S 的长度；K 是 S 的一个元素；V 是对应元素的值；$R(Q,K)$ 代表输入元素对输出元素的相关性或相似性，作为输入的权重影响注意力的聚焦程度；$A(Q,S)$ 代表输入序列对输出元素的注意力多少。

通过注意力门，网络将注意力聚焦在与目标相关性高的输入上。相关性高的输入，其 R 大，对 Q 的权重高。如果 R 的取值范围为[0,1]，被认为是 soft attention；如果 R 的取值为 0 或 1，被认为是 hard attention。soft attention 被用于分类、分割、检测、生成模型、视频处理等，本文采用该方式。

通常情况下，目标检测仅需要考虑与目标相关性强的区域，聚焦目标区域能够节省计算资源，且提升目标特征提取准确性。在原始 UNet 等网络中需要在反卷积层重新对目标区域定位，浪费网络训练资源，收敛变慢。

针对小目标检测需求，本文在网络的跳层连接处加装注意力门模块，如图 6 所示，对不同层次的空域特征进行自适应权重调整。

X^l 是跳层连接的输入特征图，是反卷积层同层的池化层的输出。g 是网络在前一层反卷积层的输出。$\hat{X}^l$ 是本文经注意力门处理后的跳层连接输入特征图，与跳层连接恒等变换相比，由于 X^l 参考了与 g 中目标特征的相关性，在目标区域的特征权重会加

强,在非目标区域会受到抑制,网络对小目标特征提取的准确性会加强。

为了更好实现时空信息融合,本文对双模态数据同步进行学习,基于原始图像的学习和基于时域显著图的学习,分别利用了时域运动信息和空域相关信息。本文中,注意力机制的输入是双模态学习的综合结果。

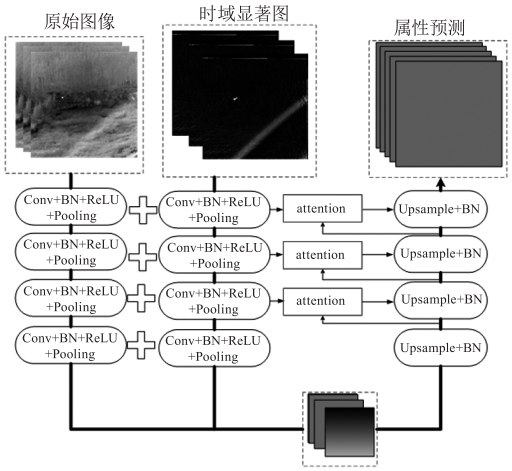


图 6 注意力全卷积神经网络结构

Fig. 6 The architecture of full convolutional neural network with attention mechanism

注意力门的具体作用流程如下:

(1)对 X^l 下采样使 X^l 和 g 尺寸相同。

(2) X^l 和 g 分别进行核为 1×1×1 的卷积得到 $W_x^T X^l$ 与 $W_g^T g$ 。

(3)对 $W_x^T X^l$ 与 $W_g^T g$ 做线性相加再用 ReLU 函数激活。

(4)对步骤 3 的输出做 1×1×1 卷积,再经过 sigmoid 函数激活。

(5)对步骤 4 的输出进行上采样得到 X^l 的注意力权重 α , α 乘以 X^l 得到注意力加权输入 $\hat{X}^l$ 。

3 实验设置及评估

3.1 红外弱特征目标数据集特点

红外弱特征目标序列数据集较少,检测物体一般为飞机、舰船等,红外弱特征目标数据集与物体检测领域的可见光数据集相比存在多个差异。下面针对红外弱特征目标数据集的特点进行介绍。

3.1.1 红外目标数据集为灰度图

红外小目标数据集用红外相机拍摄,一般呈现灰度图。灰度图中的像素只有单通道,用灰度表示,位深一般为 8。比较而言,可见光拍摄的图片有 RGB 三原色,要用三通道 24 位深表示,信息更为丰富。

3.1.2 红外目标数据集的目标特征少

红外图像的灰度值由温度决定,目标核心部分的温度高,外围温度低,呈现灰度梯度同心圆。相对而言,可见光目标往往有颜色差异,具有纹理特征。如图 7,红外无人机检测图所示,发动机部分往往温度高,灰度值大,在目标中所占像素有限,但机翼部分的温度较低,与背景对比度低,难以准确提取目标。只有机身 1/6 的部分灰度值大于周围背景,且该部分灰度值大于 200 的像素有 6 个,大于 180 的也仅有 22 个,周围的平均灰度为 150。且红外图像往往成像距离长,目标所占像素少,无法提供足够的纹理及形状信息。

3.1.3 红外目标数据集的分辨率低

红外手段测量的是物体向外辐射的热量,热量辐射后杂波干扰严重,相对而言,分辨率、信噪比均比可见光图像低,且目标与背景边界模糊,检测难度更高。

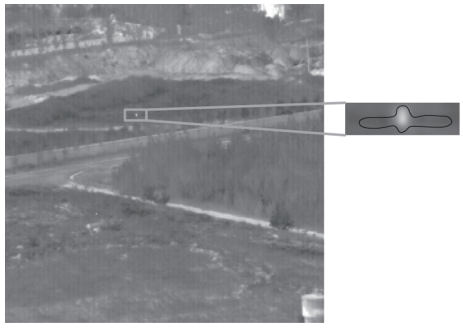


图 7 无人机探测红外图像

Fig. 7 The unmanned aerial vehicle in infrared image

3.2 真实数据集设计及分析

本文的真实数据集采用空天杯的红外序列数据集[16]。拍摄由无人机航拍完成,拍摄目标为无人机。数据集所用背景包括天空、山、森林、地物等。

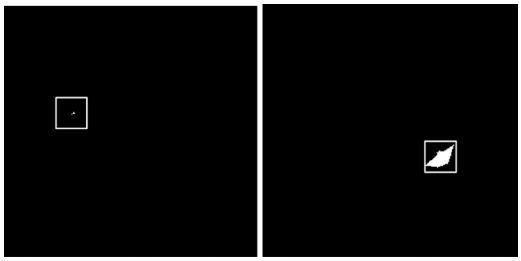
其中真实数据集从原数据集中剔除了模糊图片,最终选定了 22 段共 352 帧航拍红外无人机序列作为本文数据集,每段序列截取 16 帧,每帧图像大小 256×256 像素,在 Labelme 上对数据集进行了标注。

真实数据集中的小目标在全图中的像素及占比如表 1 所示。在真实数据集中大部分目标尺寸小于 15×15 像素,与原图尺寸的相对占比小于 5.86 %。如图 8 所示,目标最小像素数约 5 个,最大像素数约 295 个,相差 50 倍以上,属于大动态多尺度目标。

表1 真实数据集中不同尺度目标占比

Tab.1 Proportion of targets with different sizes in real data

目标像素数	目标数量	占比/%
(0,25]	117	33
(25,80]	90	26
(80,100]	97	28
(100,225]	32	9
(225,300]	16	5
(0,300]	352	100



(a)最小尺度目标 (b)最大尺度目标

图8 最小和最大尺度目标

Fig.8 The smallest target and the biggest target

3.3 仿真数据集设计及分析

真实数据集的数据量少,且目标运动、尺度等种类少,为了充分训练网络,本文制作了红外弱特征目标序列仿真数据集,在数据集中仿真了不同目标尺寸、灰度、运动及相机运动情况。

仿真数据集的背景图像源自真实红外数据集,其图像从数百个不同场景中选择,背景具有代表性和真实性。选取330个不同图像背景,设计成13200段仿真红外小目标序列。仿真数据集用每个真实图像制作出40段随机背景,并在仿真背景上各模拟一段红外小目标序列的运动,每段序列有5帧图像,每帧图像大小均为128×128。

3.3.1 仿真数据集背景设计

仿真数据集的背景由真实背景剪切、旋转、翻转、拼接而成。首先在单帧真实红外图像中随机剪切多个大小为128×128的不含目标的背景作为仿真数据集的背景源,然后将这些背景源旋转、翻转,从这里面选取9个来自同一帧图像的背景源紧凑拼接为大小为384×384的仿真背景。

每张真实红外图像的背景用随机算法组合得到40段仿真背景,背景源相同的仿真背景会被一起分配,以保证对算法泛化能力的准确评估。序列按照7:2:2的比例分配到训练集、测试集、验证集内。训练集共有210个背景,测试集共有60个背景,验证集共有60个背景。

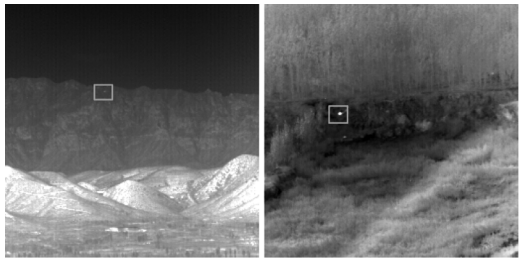
3.3.2 仿真数据集目标设计

目标尺度动态变化,例如5×5、10×10、15×15,代表不同大小的目标。目标灰度为动态变化,代表不同强度的目标。目标设定为随机向上下左右四个方向做匀速直线运动,目标运动速度动态变化,以模拟目标在短时间内的运动。目标各种参数的出现权重均相同,以设置对照。

3.3.3 仿真数据集运动设计

仿真数据集在已有目标和背景的基础上,仿真了多段连续运动目标。为更好的模拟动态背景和静态背景,相机也设置了运动和抖动,每次选取一定区域的背景,模拟相机保持原地或在上下左右四个方向做匀速直线运动。相机运动速度动态变化,并且相机在每帧会有50%的概率向随机方向2个像素内抖动。最后在图像上加标准差为20的高斯噪声,模拟各种各样干扰和噪声。

制作仿真数据集有两个目的:其一为扩充数据,增加训练样本,避免欠拟合;其二为制作各类数据集,对比算法在不同条件下的检测效果。



(a)山 (b)森林

图9 真实数据集图片

Fig.9 Images of real data

3.4 评价指标

本文用准确率、召回率、F1值和交并比(IoU,Intersection over Union)评价模型的优劣。首先对中间变量进行介绍,虚警(False Positives,FP)代表将背景误判为目标,漏警(False Negatives,FN)代表将目标误判为背景。真阳率(True Positives,TP)代表将目标正确预测目标。

准确率(Precision)与检测率同义,表示在预测为目标的情况中,有多少是正确的真正目标。准确率低代表虚警多,理想状况下的准确率为1,这时所有预测出的目标均为实际目标。准确率计算公式如下。

$$P = \frac{TP}{TP + FP} \tag{2}$$

召回率(Recall)表示在实际样本中,有多少是正确预测的。召回率与漏警相关。理想状况下的召回率为 1,这时所有目标均被正确预测。召回率低代表漏检多。召回率计算公式如下。

$$R = \frac{TP}{TP + FN} \tag{3}$$

F1 值为准确率和召回率的调和平均数,用于测量不均衡数据的精度。准确率和召回率越高,F1 值越高,检测的总体性能越好。F1 值计算公式如下。

$$F_1 = \frac{2}{\frac{1}{P} + \frac{1}{R}} = 2 \times \frac{P \times R}{P + R} \tag{4}$$

IoU 交并比是实际样本与预测样本的交集面积/并集面积,同样是准确率和召回率的变形,IoU 比 F1 指数更小,但可以与图形直观对应,理想状况下交并比为 1,这时预测样本与实际目标重合。交并比计算公式如下。

$$IoU = \frac{TP}{TP + FP + FN} \tag{5}$$

4 实验结果及分析

本节用仿真实验分析所提检测算法的性能,进行了不同算法性能对比分析、多类型场景性能分析、注意力门影响对比分析、时域显著图预处理影响对比分析。

运行算法的 GPU 为 NVIDIA GEFORCE RTX1650,显存 4G,网络采用 4 层卷积,初始学习率 0.01,采用 Adagrad 优化器以及 DICE 损失函数。下面分别对上述对比实验进行详细介绍。

4.1 不同算法性能对比

对传统序列小目标检测方法(帧差法)[17]、UNet 网络方法[13]和本文设计方法在 4 段航拍真实红外序列数据集上进行性能对比分析,测试结果分别如表 2 所示。

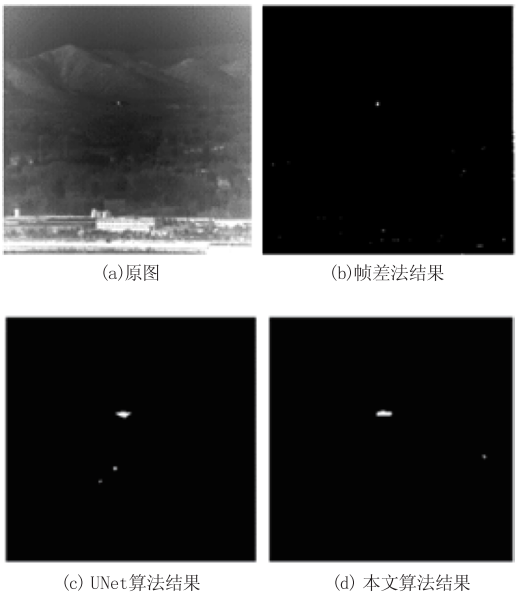
从表 2 可知,检测性能上本文算法 > UNet 算法 > 传统帧差法。下面选取典型目标序列进行分析,结果如图 10 所示。

从准确率情况分析,当目标与背景的灰度相近时,传统帧差法要权衡阈值控制虚警率和漏警率平衡,在二值化分割时保留了较多的背景高灰度点,且对帧差法来说,无法将目标无人机外围低温低灰度部分正确分类,正确检测率和提取目标像素完整性偏低。而深度学习算法能学习到目标的空间分布特征,且所提算法由于添加注意力机制,对目标空间特征学习更强,能够准确提取目标的边缘部分,分割更准确,正确检测率更高。

表 2 不同方法测试结果

Tab. 2 The test result of different methods

算法	准确率/ %	召回率/ %	F1/ %	IoU/ %
传统帧差法	50.71	45.31	42.03	28.45
UNet 算法	68.40	72.26	69.96	53.95
本文算法	93.95	80.66	86.79	76.78



(a)原图 (b)帧差法结果

(c) UNet算法结果 (d) 本文算法结果

图 10 不同算法的结果比较

Fig. 10 The resultof different methods

从虚警情况分析,传统帧差法的虚警背景中存在较多空间高频杂波,特别是建筑物边缘,深度学习方法能够感知整体背景分布,虚警较少,其中所提方法的虚警最少,对背景的抑制最强。

实验证明,所提算法运用目标提取更准确,运用空间注意力,对空间特征提取更精准。比较而言,实现了对检测交并比、F1 值、准确率和召回率的综合提高。

4.2 多类型场景性能分析

对不同类型场景的仿真数据集进行对照实验,采用的为同一个训练集、测试集、训练模型,并在不同条件下验证检测效果。为不同的目标强度、目标大小、目标运动速度、相机运动速度、相机抖动序列数据分别设置验证集,评估算法的泛化能力和鲁棒性。

验证集包含60个共同背景,每一个验证集在一个背景下有4段序列,共计240段序列。动态背景实验组的灰度为223,目标大小5×5像元,目标运动速度3像素/帧,相机运动速度1像素/帧,相机无抖动。静态背景为第4组,既为实验组又为对照组,对照组除一项不同外与实验组无其他差异。网络在不同场景的交并比、准确率、召回率、F1值如表3所示。其中,第1-1组的目标灰度为191,第1-2组的目标灰度为255,第2-1组目标运动速度为5像素/帧,第2-2组目标运动速度为7像素/帧,第3-1组目标大小为15×15像素,第3-2组目标大小为10×10像素,第4组相机静止拍摄,第5组相机带有3像元抖动。

表3 所提算法在不同场景的测试结果

Tab.3 The result of the proposed method in different scenes

组别	准确率/%	召回率/%	F1/%	IoU/%
总体	93.95	80.66	86.79	76.78
实验组	90.32	79.14	84.36	72.91
1-1组	85.01	71.99	77.96	63.64
1-2组	95.09	85.25	89.90	81.64
2-1组	93.96	79.52	86.14	75.55
2-2组	98.41	81.07	88.90	80.00
3-1组	98.37	84.62	90.98	83.43
3-2组	98.63	84.45	90.99	83.45
4组	92.56	80.77	86.26	75.70
5组	93.22	79.15	85.61	74.71

下面对比实验组与各情形下检测性能,分析虚警和漏警原因:

(1)从实验组与第1组的对比中可见,随目标灰度的升高,交并比、准确率、召回率、F1值均有提高,且提高幅度较均匀。召回率的提升比准确率提升更大,灰度上升32时,召回率平均上升5%,准确率平均上升7%。

(2)从实验组与第2组的对比中可见,随目标运动速度提高,检测性能均有提高。这时准确率的提升比召回率提升更大,且速度越快,提升越快,准确率的提升比召回率更明显。

初步分析,背景运动速度慢或静止,而目标运动明显,通过目标位移规律可以将目标和噪声区分。目标运动速度变快,提高了目标的检测率,同时降低噪声的虚警率,召回率也有所上升。

(3)从实验组与第3组的对比中可见,目标尺寸从5到10,检测性能均有提高,尺寸从10到15各项检测指标的变化不大。

初步分析,三种尺寸的图片数量一致,不存在数据集中某种尺寸出现多,而导致精度更高的问题。因此,可能有两方面的原因导致5×5的小目标检测性能下降:一是小目标与杂波尺寸相近,小目标被分类为杂波或杂波被分类为小目标;二是小目标在特征提取的过程中能被提取的特征更少,这些都不利于小目标检测。

(4)从实验组与第4组的对比中可见,相机静止的检测性能均比实验组高,其中准确率升高2%,召回率升高1%。

对序列背景运动变化的影响进行分析,根据仿真数据集的设计,目标和相机会随机在上下左右四个方向运动,有1/4的可能性运动方向相同,1/4的可能性运动方向相反。运动目标检测的背景按相机是否运动可分为静态背景和动态背景,相机为了保持稳定避免运动模糊,往往运动速度不会太快,仿真数据集中背景的最大运动速度为2像素/帧,小于目标的最小运动速度3像素/帧,在动态背景下,区分背景与目标的难度增大。

(5)在第4组的静止拍摄基础上加入随机抖动得到第5组,从第4、5组的对比中可看出,相机抖动会降低检测召回率,导致交并比与F1值下降。相机抖动的序列中,部分目标的运动速度降低,可能会被误判为背景杂波,造成漏警,召回率下降。

上述分析了目标强度、目标大小、目标运动速度、相机运动速度、相机抖动因素对检测性能的影响,总体而言本文算法对不同场景下的弱特征目标的鲁棒性较好。

4.3 注意力机制影响分析

为了分析注意力机制的影响,在训练序列仿真数据的基础上,对跳层连接输入的浅层特征加入空间注意力门。二者损失函数和F1值随训练迭代次数变化如图11所示。

与原始UNet网络相比,加入注意力门后网络的损失更早收敛,且收敛后的损失相差不大。从F1的变化来看,加入注意力后的F1值更大,检测效果比未加注意力的时候更优。

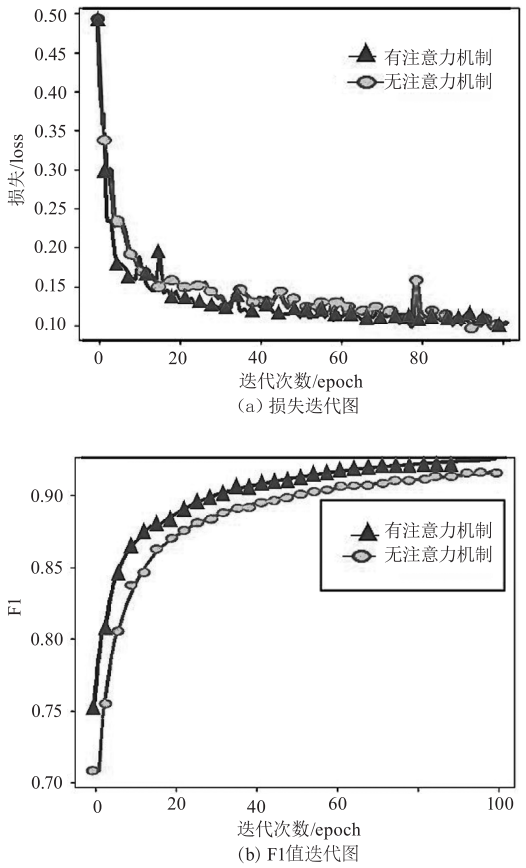


图11 对比注意力门的F1值和损失迭代图

Fig. 11 The comparison of loss and F1

结合损失和F1值的变化图可知,加入注意力机制后能节省计算资源,收敛更快,加快网络训练和收敛速度,加注意力门的训练时间为原网络的70 %左右。且网络在目标区域的注意力权重增加,在背景区域的注意力权重减少,对目标分割更精准。

4.4 时域显著图预处理影响分析

本实验首先对原图进行时域显著图预处理,再将预处理的输出送入神经网络训练。为了分析预处理的影响,将该算法的在训练过程中的F1值和损失变化情况与未做预处理时对比,评估二者各方面检测性能的差异。

在图12中,圆形点线为不做时域显著图预处理的训练F1值和损失,三角形点线代表预处理后的训练F1值和损失。从F1值的变化对比可知,加入时域显著图预处理后收敛加快,训练所需时长是不做预处理的60 %。从损失的变化情况可见,没有预处理的损失曲线的损失较大,在20次迭代后就一直在75 %到85 %区间波动。而预处理后的损失曲线的中值一直在下降,在100次迭代时的中值与不预处理的500次左右的损失相当,在270次迭代附近收敛。

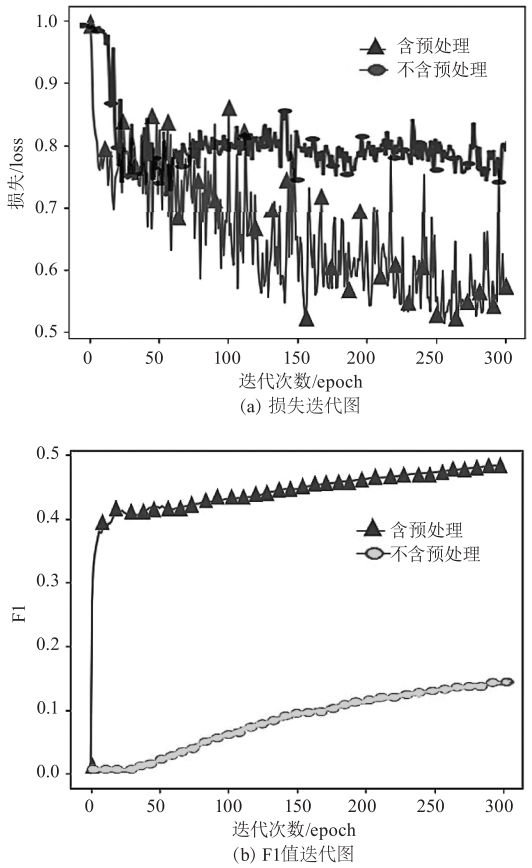


图12 时域显著图对F1值和损失迭代影响图

Fig. 12 Influenceof saliency map for loss and F1

仿真结果证实,采用时域显著图预处理,一方面可以实现背景抑制并增强目标,对提升检测率与召回率均有帮助;另一方面强化对运动特征的关注,特征更加稀疏,减少了网络需要学习的特征量,使损失下降更快,模型更早达到收敛。

5 结 论

远距离探测条件下红外序列中目标属于弱特征目标,为检测此类目标,提出了基于注意力全卷积的弱特征目标分割方法,首先采用多帧差分叠加构建时域显著图,将不同相机中的变化成分在同一域中表征,强化对运动特征的学习,继而构建深度学习网络从时域显著图和原始图像中同步学习时空特征,并通过注意力机制从综合多维特征中自适应学习目标相关信息。通过真实数据和仿真数据测试表明,在真实多类型场景中,本文方法对大小相差50倍以上的目标平均检测准确率优于93 %。

参考文献:

[1] Liu Z, Yang D Z, Li J Y, et al. A review of infrared single-framedim small target detection algorithms[J]. Laser & Infrared, 2022, 52(3): 154 – 162. (in Chinese)
刘征, 杨德振, 李江勇, 等. 红外单帧弱小目标检测算法研究综述[J]. 激光与红外, 2022, 52(3): 154 – 162.

[2] Wang H H, Cao D, Zhao Y, et al. Survey of infrared dim small target detection algotithm based on deep learning [J]. Laser & Infrared, 2016, 52(1): 60 – 72. (in Chinese)
王恒慧, 曹东, 赵杨, 等. 基于深度学习的红外弱小目标检测算法研究综述[J]. 激光与红外, 2022, 52(9): 1274 – 1279.

[3] Liu H K, Zhang L, Huang H. Small target detection in infrared videos based on spatio-temporal tensor model[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(12): 8689 – 8700.

[4] Luo K P. Space-based infrared sensor scheduliing weith high uncertainty: issues and challenges[J]. Systems Engineering, 2015, 18(1): 102 – 113.

[5] Chen Y Y, Han J H, Zhang H H, et al. Infrared small dim target detection using local contrast measure weighted by reversed local diversity[J]. Infrared Laser Ingineering, 2021, 50(8): 151 – 167.

[6] Barnett J. Statistical analysis of median subtraction filtering with application to point target detection in infrared backgrounds[C]//Proc. SPIE 1050 on Infrared System and Components III, 2016: 779 – 788.

[7] Tom V T, Peli T, Leung M, et al. Morphology-based algorithm for point target detection in infrared backgrounds [C]//Proc. SPIE 1954 on Signal and Data Processing of Small Targets, 1993: 2 – 11.

[8] Chen C L P, Li H, Wei Y T, et al. A local contrast method for small infrared target detection[J]. IEEE Transcations on Geoscience and Remote Sensing, 2014, 52(1): 574 – 581.

[9] Han J H, Liang K, Zhou B, et al. Infrared small target detection utilizing the multiscale relative local contrast measure[J]. IEEE Geoscience and Remote Sensing Letters, 2018, 15(4): 612 – 616.

[10] Ren S Q, He K M, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137 – 1149.

[11] Jredmon, Divvaia S, Girshick R, et al. You only look once: unified, real-time object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 27 – 30.

[12] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector[C]//Springer International Publishing, 2016: 21 – 37.

[13] Olaf R, Philipp F, Thomas B. U-Net: convolutional networks for biomedical image segmentation. [C]//Medical Image Computing and Computer-Assisted Intervention, 2015: 234 – 241.

[14] Rafael C G, Richard E W. Digital image procesing[M]. 4th Edition. Beijing: Electronic Industry Press, 2020. (in Chinese)
拉斐尔, 查理德. 数字图像处理[M]. 第四版. 北京: 电子工业出版社, 2020.

[15] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Adcances in Neural Information Processing Systems, 2017, 30: 1 – 11.

[16] Hui B W, Song Z Y, Fan H Q, et al. A dataset for infrared detection and tracking of dim-small aircraft targets under ground/air background[J]. China Scientific Data, 2020, 5(3): 291 – 302. (in Chinese)
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[J]. 中国科学数据, 2020, 5(3): 291 – 302.

[17] Du W H, Li D X, Wang Q N, et al. Moving target detection based on improved frame difference and edge extraction algorithm[J]. Science Technology and Engineering, 2022, 26(5): 1944 – 1945. (in Chinese)
杜文汉, 李东兴, 王倩楠, 等. 融合改进帧差和边缘提取算法的运动目标检测[J]. 科学技术与工程, 2022, 26(5): 1944 – 1949.