激光与光电子学进展 先进成像

基于位置感知的热红外目标跟踪方法

杨静，马龙*

西安工业大学兵器科学与技术学院，陕西 西安 710021

摘要 针对热红外图像中目标缺少细节信息而导致跟踪精度不高的问题，提出一种基于位置感知的目标跟踪方法。首先使用深度空洞残差网络(D-ResNet)提取语义特征，鲁棒表征热红外目标；然后设计位置感知模块，有效感知目标在特征图上的空间位置，提高算法的定位精度；并引入通道注意力模块，在通道域上筛选特征图信息，抑制干扰信息；接着引入区域提取网络，完成目标分类和边框回归；最后使用RGBT234热红外序列对网络进行微调，确保网络能有效学习热红外目标信息。所提方法在VOT-TIR2019、GTOT数据集上分别获得75.3%和91.4%的准确率，速度为30 frame/s。实验结果表明：所提方法在热红外场景下能获得较高的跟踪精度，并能有效应对遮挡、相似物干扰、尺度变化等目标跟踪过程中的常见挑战。

关键词 热红外目标跟踪；位置感知；通道注意力；深度空洞残差网络

中图分类号 TP391.9 **文献标志码** A **DOI**：10.3788/LOP220929

Thermal Infrared Object Tracking Method Based on Positional Perception

Yang Jing, Ma Long*

School of Ordnance Science and Technology, Xi'an Technological University, Xi'an 710021, *Shaanxi, China*

Abstract A target tracking method based on positional perception is proposed to address the issue of low tracking accuracy caused by the absence of target detail information in thermal infrared images. First, semantic characteristics were extracted and thermal infrared objects were robustly characterized using the deep dilated residual network (D-ResNet). Second, a positional perception module was designed to efficiently detect the object position on the feature map and enhance the positioning accuracy of the algorithm. Third, the channel attention module was introduced to suppress interference information and filter feature map data in the channel domain. Then, the region proposal network was implemented to complete border regression and target categorization. Finally, RGBT234 thermal infrared sequences were used to adjust the network to successfully learn the thermal infrared object information. The proposed method is tested on VOT-TIR2019 and GTOT datasets and achieves accuracy of 75.3% and 91.4%, respectively, and a speed of 30 frame/s. Experimental results also demonstrate that the proposed method can realize high tracking accuracy in dealing with common difficulties, such as occlusion, analog interference, and scale change, effectively in the thermal infrared scene.

Key words thermal infrared object tracking; positional perception; channel attention; deep dilated residual network

1 引 言

随着热红外技术的发展，基于热红外图像的目标跟踪问题成为计算机视觉领域研究的热点。热红外图像分辨率较低，缺乏对目标纹理信息的描述；伴随光照、遮挡、旋转、尺度等问题引起的物体表观变化[1]，热红外目标跟踪的难度极大增强。因此，研究适用于复杂场景的高效热红外目标跟踪算法是至关重要的。

目标跟踪算法的主要流程包含目标初始化、目标表观建模、运动预测和目标定位。其中，目标表观建模是关键。深度学习由于强大的表征能力，广泛应用于机器人导航、人机交互等民用领域[2]。孪生网络作为基于深度学习的目标跟踪算法的重要组成部分，最早被用于银行系统的签名验证中[3]，后来Siamese instance search for tracking(SINT)[4]将孪生网络看作相似度度量问题，为目标跟踪任务提供了新思路。Bertinetto等[5]基于相似度计算提出一种基于全卷积孪生神经网络(SiamFC)的跟踪算法，证明了孪生网络在速度和精度上具有巨大的潜力[6]。Siamese region proposal network(SiamRPN)在SiamFC的基础上引入

收稿日期：2022-03-08；**修回日期**：2022-03-28；**录用日期**：2022-06-13；**网络首发日期**：2022-06-23

通信作者：*malong@xatu.edu.cn

区域提取网络(RPN)[7],实现了孪生网络在速度和精度方面的良好平衡。Li等[8]提出基于超深网络的孪生区域提取网络(SiamRPN++),该网络将深度残差网络应用于孪生网络中,在多个基准数据集上表现优异。

近年伴随热红外技术的蓬勃发展,基于热红外图像的目标跟踪算法得到国内外学者的广泛关注。杨福才等[9]将稀疏编码直方图应用在红外目标跟踪方法中,有效提升跟踪算法的稳定性。文献[10]提出基于全局感知孪生网络的红外目标跟踪算法。钱琨[11]将引导滤波和卷积神经网络(CNN)应用于红外目标跟踪问题。以上研究尽管取得了一定程度的进展,但仍无法解决跟踪网络对目标位置感知能力不足的问题。

基于此,本文结合孪生网络和注意力机制,提出了一种基于位置感知的热红外目标跟踪方法。首先,使用深度空洞残差网络(D-ResNet)获取红外目标深层语义特征和背景细节信息;然后,利用空间位置和通道信息的语义相互依赖性建模,使提取到的特征得到充分利用,进一步突出特征表示;接着,利用区域提取网络对目标进行分类和回归;最后,使用热红外数据对网络进行微调,增强算法的位置感知能力。实验结果表明,所提方法可以有效应对红外场景下光照、遮挡、尺度问题引起的物体表观变化,速度满足实时性要求。

2　基本原理

2.1　模型构建

所使用的神经网络框架如图1所示,可分为4部分:第一,使用D-ResNet提取热红外目标深层语义信息;第二,在残差模块后设计位置感知模块,利用特征图的上下文信息来增强对跟踪目标的特征表达;第三,通道注意力模块通过建模不同特征通道的重要程度,针对不同的目标类别来增强或者抑制不同的通道,增强特定通道下的语义响应能力;第四,利用区域提取网络获取目标的准确定位。所使用的孪生神经网络跟踪算法可简单看作目标模板与搜索图像之间的相似度度量问题,公式为

$$f(\boldsymbol{z},\boldsymbol{x})=f\left[\varphi(\boldsymbol{z}),\varphi(\boldsymbol{x})\right]=\varphi(\boldsymbol{z})*\varphi(\boldsymbol{x})+b\cdot\boldsymbol{I},\tag{1}$$

式中:$*$代表互相关运算;$b\cdot\boldsymbol{I}$为响应图中每一个位置的偏差。将目标模板$\boldsymbol{z}$和搜索区域$\boldsymbol{x}$输入到权值共享的特征提取网络φ中,得到$\varphi(\boldsymbol{z})$和$\varphi(\boldsymbol{x})$,经过相似度量函数f得到响应图,响应得分越高,二者相似度越高。

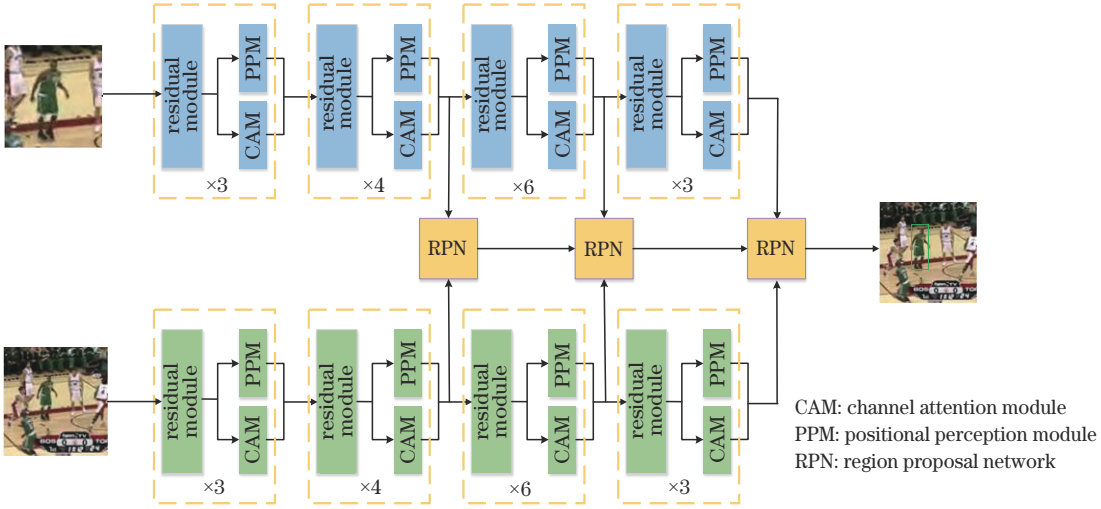

图1　所提网络结构

Fig. 1　Structure of the proposed network

2.2　D-ResNet

经典的孪生目标跟踪算法SiamFC和SiamRPN使用层数较少的VGG或AlexNet作为主干网络来提取目标特征,因热红外图像信息描述少,浅层语义信息对特征的描述力不足,导致跟踪失败。ResNet-50因特征提取能力强被广泛应用在计算机视觉领域,本文的主干网络是基于ResNet-50改进而来的。

传统ResNet使用32 pixel的大步幅,不利于孪生网络对热红外目标进行有效定位。文献[12]指出,8倍上采样比32倍上采样得到的效果更好,保留的图像细节更多。因此,考虑将原ResNet-50中Conv4和Conv5的大步幅改为8 pixel小步幅,提高网络提取红外目标特征信息的能力。同时为保证特征提取网络对目标尺度的感受能力,采用空洞卷积(dilated convolution)改进ResNet-50,将其命名为D-ResNet。

空洞卷积在卷积核之间加入空间进行“膨胀”,扩大模型的感受野(receptive field),如图2所示。空洞率(l)表示将卷积核放宽,当l=1、感受野为3×3时,此时的卷积和普通卷积一样;当l=2、感受野为7×7时,实际的卷积核大小还是3×3,如图2(b)所示,9个圆点的权重不为0,其余都为0;当l=4时,跟在图2(a)和图2(b)后面,感受野为15×15。对比传统的Conv操作,3层3×3的卷积加起来,stride为1的话,只能达到$(k_{\text{kernel}}-1)\times l_{\text{layer}}+1=7$的感受野,空洞卷积的感受野

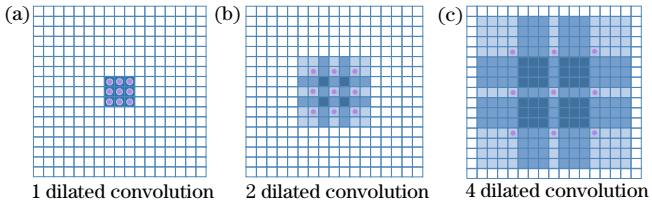

图 2　空洞卷积示意图

Fig. 2　Schematic of dilated convolution

呈指数级增长，不需要增加参数运算成本就能观察大的图像范围，充分利用背景信息，提高算法对目标的辨别能力，避免出现跟踪漂移现象。

深度残差神经网络中残差模块的映射是信息传递的主要途径，根据热红外图像的特性，该算法首先使用 ResNet-50 提取图像深层语义信息，其次通过减小步幅保留更多图像背景信息，最后使用空洞卷积获取大范围的感受野从而获得更加丰富的数据信息，帮助算法更准确地定位红外目标。

2.3　位置感知模块（PPM）

在跟踪任务中，相比于可见光图像，红外图像纹理信息少，特征提取困难，易出现对目标定位错误即位置感知能力差的情况。受计算机视觉中经典的非局部均值方法启发[13]，设计一个位置感知模块，通过对所有空间位置上的特征信息进行加权，有选择性地聚合每个位置的特征，完成对跟踪目标的重新建模。通过位置感知模块，可以充分利用热红外图像背景信息，筛选出更重要的目标特征并抑制背景中的干扰信息，使后续深度互相关得出的特征响应更加有效，提高算法对热红外目标的位置感知能力，实现目标准确定位。

图 3 为位置感知模块结构，首先将局部特征 $\boldsymbol{A}$ 分别送入到具有正则化和 ReLU 的卷积层，得到两个特征映射 $\boldsymbol{B}$ 和 $\boldsymbol{C}$。将 $\boldsymbol{B}$ 转置后与矩阵 $\boldsymbol{C}$ 相乘，经 Softmax 函数得到特征响应图 $\boldsymbol{S}$。$\boldsymbol{S}$ 中每一行计算的是所有像素与某个像素之间的依赖关系，表达式为

$$s_{ji}=\frac{\exp(\boldsymbol{B}_i\cdot\boldsymbol{C}_j)}{\sum_{i=1}^{N}\exp(\boldsymbol{B}_i\cdot\boldsymbol{C}_j)},\tag{2}$$

式中：$\boldsymbol{B}_i$ 表示第 i 个位置的特征映射；$\boldsymbol{C}_j$ 表示第 j 个位置的特征映射；s_{ji} 表示第 i 个位置对第 j 个位置的影响。Softmax 函数的值越大，说明置信度越高，空间位置相

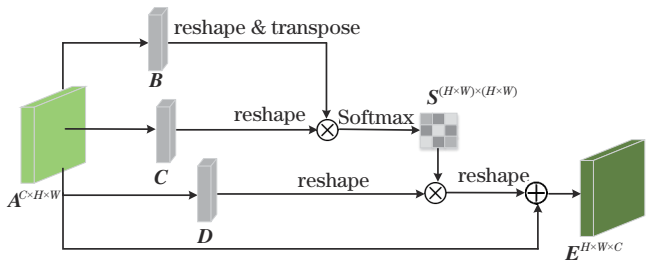

图 3　位置感知模块的结构

Fig. 3　Structure of positional perception module

对依赖性也越强。同时，将特征 $\boldsymbol{A}$ 送入带有正则化和 ReLU 的卷积层，得到新的特征 $\boldsymbol{D}$，将 $\boldsymbol{D}$ 和 $\boldsymbol{S}$ 相乘结果与局部特征 $\boldsymbol{A}$ 相加，得到最终特征图 $\boldsymbol{E}$，表达式为

$$\boldsymbol{E}_j=\alpha\sum_{i=1}^{N}(s_{ji}\boldsymbol{D}_i)+\boldsymbol{A}_j,\tag{3}$$

式中：$\boldsymbol{D}_i$ 表示第 i 个位置的特征映射；$\boldsymbol{A}_j$ 表示第 j 个位置的原始特征图；$\boldsymbol{E}_j$ 为第 j 个位置的特征；α 为尺度因子。

所提方法将残差模块得到的特征图输入到位置感知模块，该模块根据空间注意力图谱选择性地聚合上下文信息，对目标实现准确定位。图 4 为热红外目标线索可视化示意图，通过相互关联的特征找到相关线索来支持红外目标预测。这样，根据不同位置的权值大小，可以有效抑制无用信息，突出有效信息。为进一步描述目标特征，所提方法利用空间维度和通道维度的语义相互依赖性[14]，将位置感知的输出结果与通道注意力模块结果相加，使目标在复杂场景中仍然具有突出特性。

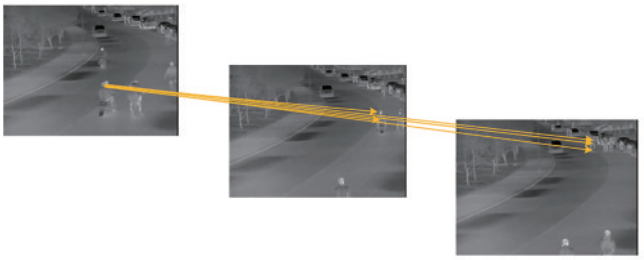

图 4　热红外目标线索可视化

Fig. 4　Thermal infrared object visualization

2.4　通道注意力模块（CAM）

通道注意力模块可看作不同通道特征图语义属性重新分配权重的过程[15]。在不同的跟踪环境下，通道表示的特征信息是不同的：在目标定位时需要浅层信息，如颜色、形状等特征；当受到相似目标干扰时，则需要更深层次的语义信息来描述红外目标。使用的通道注意力模块将残差模块提取到的特征通道重新分配权重，通道包含的信息与模板特征越相似，权重系数越大。另外，通道注意力模块有即插即用的特点，使用方便，模块结构如图 5 所示。

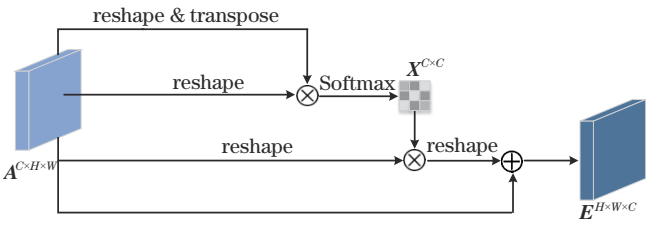

图 5　通道注意力模块的结构

Fig. 5　Structure of channel attention module

通道注意力模块从原始的特征 $\boldsymbol{A}$ 直接计算通道映射 $\boldsymbol{X}$，x_{ji} 表示第 i 个通道对第 j 个通道的影响，表达

式为

$$x_{ji}=\frac{\exp(\boldsymbol{A}_i\cdot\boldsymbol{A}_j)}{\sum_{i=1}^{C}\exp(\boldsymbol{A}_i\cdot\boldsymbol{A}_j)},\tag{4}$$

式中：$\boldsymbol{A}_i$表示第i个通道的原始特征；$\boldsymbol{A}_j$表示第j个通道的原始特征。对矩阵$\boldsymbol{A}$和$\boldsymbol{A}$的转置矩阵执行相乘操作，最后应用Softmax层得到通道响应映射$\boldsymbol{X}$。除此之外，对$\boldsymbol{X}$的转置和$\boldsymbol{A}$执行矩阵相乘操作，将得到的结果乘以尺度因子后与$\boldsymbol{A}$执行逐元素相加，得到最终特征图$\boldsymbol{E}$，表达式为

$$\boldsymbol{E}_j=\beta\sum_{i=1}^{C}(x_{ji}\boldsymbol{A}_i)+\boldsymbol{A}_j,\tag{5}$$

式中：$\boldsymbol{E}_j$为第j个通道的特征图；β为尺度因子。

所提方法中通道注意力模块和位置感知模块并行，通过捕获任何两个通道映射之间的通道相关性，用所有的通道映射加权来更新每个通道。其中，高层特征的每一个通道映射可以看作一个类别明确的响应与不同的语义响应之间的互相联系。通道注意力模块通过不同通道映射之间的相互依赖性可以有效增强特征图对特定语义的表征能力，充分挖掘红外图像的特征信息。

2.5　区域提取网络(RPN)

SiamRPN[8]引入RPN模块代替了尺度搜索策略来进行多尺度的目标估计。RPN主要由分类、回归、上通道互相关3部分组成。在RPN模块中，分类分支用来区分前景和背景；回归分支用来定位候选框，生成边界框的准确位置，预先定义k个锚框完成对目标的多尺度估计，RPN会输出$2k$个分类通道和$4k$个回归通道；上通道互相关操作把模板分支的特征图$\varphi(\boldsymbol{z})$当作卷积核，$\varphi(\boldsymbol{z})$与搜索分支的特征图$\varphi(\boldsymbol{x})$进行卷积运算，得到响应图。然后分类分支与回归分支根据响应图得到目标的分类得分和边界框的位置。

由于RPN模块的上通道互相关存在特征提取模块和RPN模块参数严重不平衡的问题，SiamRPN++在此问题上提出了深度可分离互相关，极大地简化参数量，平衡两支的参数，同时让训练更加稳定，也能更好收敛。采用SiamRPN++中的RPN模块，如图6所示。其中深度可分离互相关运算得到的多通道响应图具有正交特性，结合位置感知模块和通道注意力模块，能更好地提高跟踪精度。

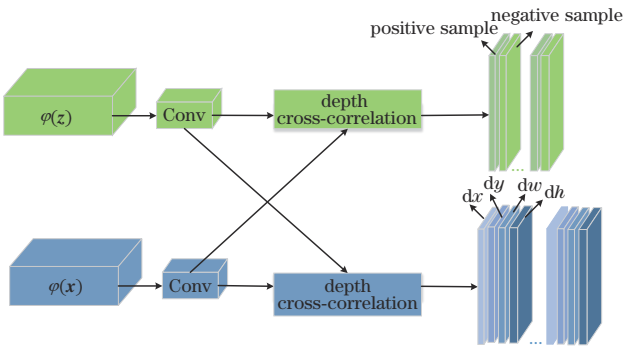

图6　RPN模块

Fig. 6　RPN module

3　分析与讨论

3.1　实验配置

本实验平台使用的操作系统是Ubuntu18.04。软件配置：PyTorch1.7.1、cuda11.2。硬件配置：CPU为Intel(R) Core(TM) i9-10900X CPU@3.70 GHz、显卡为NVIDIA TITAN RTX，内存大小为24 GB。

3.2　数据集介绍

训练集：所提方法的主干网络D-ResNet是在Imagenet2012数据集上完成预训练的。在YouTube-BoundingBoxes[16]、ImageNet VID[17]、ImageNet DET[17]、COCO[18]四个数据集上端到端训练整个网络。其中YouTube-BoundingBoxes数据集包含38万个视频片段，23个类别的物体，560万个人工标注框；ImageNet VID和ImageNet DET数据集分别有230个类别和200个子集；COCO数据集包含91个对象类型和328千张图片，250万个标注框。最后将热红外数据集RGBT234[19]划分为训练集和测试集，训练集包含173个序列，使用训练集对网络进行微调，以增强网络在红外场景下的鉴别能力。RGBT234数据集包含红外视频序列和目标跟踪中常见的难点，如尺度变化、相似物干扰等复杂场景。

测试集：为验证所提方法的有效性，选取热红外数据集VOT-TIR2019和GTOT[20]对包括所提方法在内的6种方法进行了评估。其中VOT-TIR2019数据集共包含60个热红外视频序列，该数据集中每组序列图像的分辨率为320×240～1920×480，包含多种目标跟踪难点，如动态摄像头、遮挡、动态变化等。GTOT数据集由50个视频序列组成，包含红外目标跟踪中目标尺度变化、光照变化等因素。以上视频序列对评估不同的跟踪器具有足够的挑战。

3.3　网络参数设置

采用SGD优化器训练了25个周期(epoch)，批大小为64。在第1～10 epoch，以0.1的热身学习速率训练SiamRPN++；从第11个epoch开始到第20个epoch结束，对整个网络进行端对端训练；在后5个epoch，利用红外数据对D-ResNet的layer 2、layer 3、layer 4三部分进行训练。学习率从0.005指数衰减到0.0005，动量为0.9。整个损失函数是分类的损失和回归的标准平滑L_1损失(standard smooth L_1)之和。

3.4　结果可视化

为了更加直观地展示位置感知模块和通道注意力模块对目标定位的作用，引入类激活热力图对输入图像进行计算，来确定每个位置对特定类别的重要程度，对网络中D-ResNet的最后一个卷积层完成可视化。图7分别为未加入注意力模块和加入注意力模块的类

激活热力图，网络认为某个区域越接近目标类别，类激活热力值就越大，温度也就越高。由图 7 可知，相比未加入注意力模块的网络，加入注意力模块的网络对感兴趣目标的辨别能力更强，对特定目标的判别能力得到有效提升，应用在跟踪任务的特征提取部分时可以更准确地把网络限制在跟踪目标上。

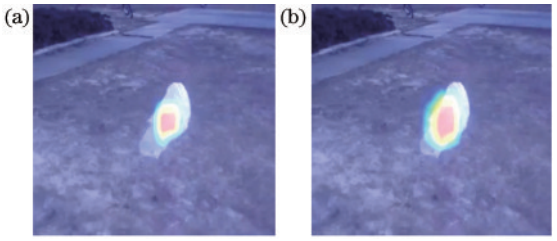

图 7　类激活热力图。(a)无注意力模块；(b)有注意力模块

Fig. 7　Class activated thermal maps. (a) Without attention module; (b) with attention module

3.5　实验结果

对所提方法与目前主流的目标跟踪方法 SiamFC、SiamRPN、ECO[21]、C-COT[22]、DSST[23]进行实验对比。ECO、C-COT 和 DSST 为基于相关滤波的跟踪方法，SiamFC 和 SiamRPN 为基于孪生网络的跟踪方法。C-COT 将学习检测过程推广到连续空间域，可以获得亚像素精度的位置。ECO 在 C-COT 的基础上将速度提升到 60 frame/s，并且将样本分组，解决过拟合问题。SiamFC 使用全卷积孪生网络进行相似性学习，解决跟踪任意对象的问题。SiamRPN 省去多尺度测试耗费时间，借鉴了目标检测的 RPN 结构，让跟踪框更加准确。此外，实验使用公开的代码，以保证实验的公平性。

3.5.1　定性分析

表 1 是来自热红外测试集 VOT-TIR2019 的 4 个视频序列，包含遮挡(OCC)、尺度变化(SV)、相似物干扰(AI)、光照变化(IV)、运动变化(MC)、相机移动(CM)等热红外目标跟踪中的常见挑战。在不同复杂场景下对所提方法与其他 4 种效果较好的主流方法进行定性评估，评估结果如图 8 所示。

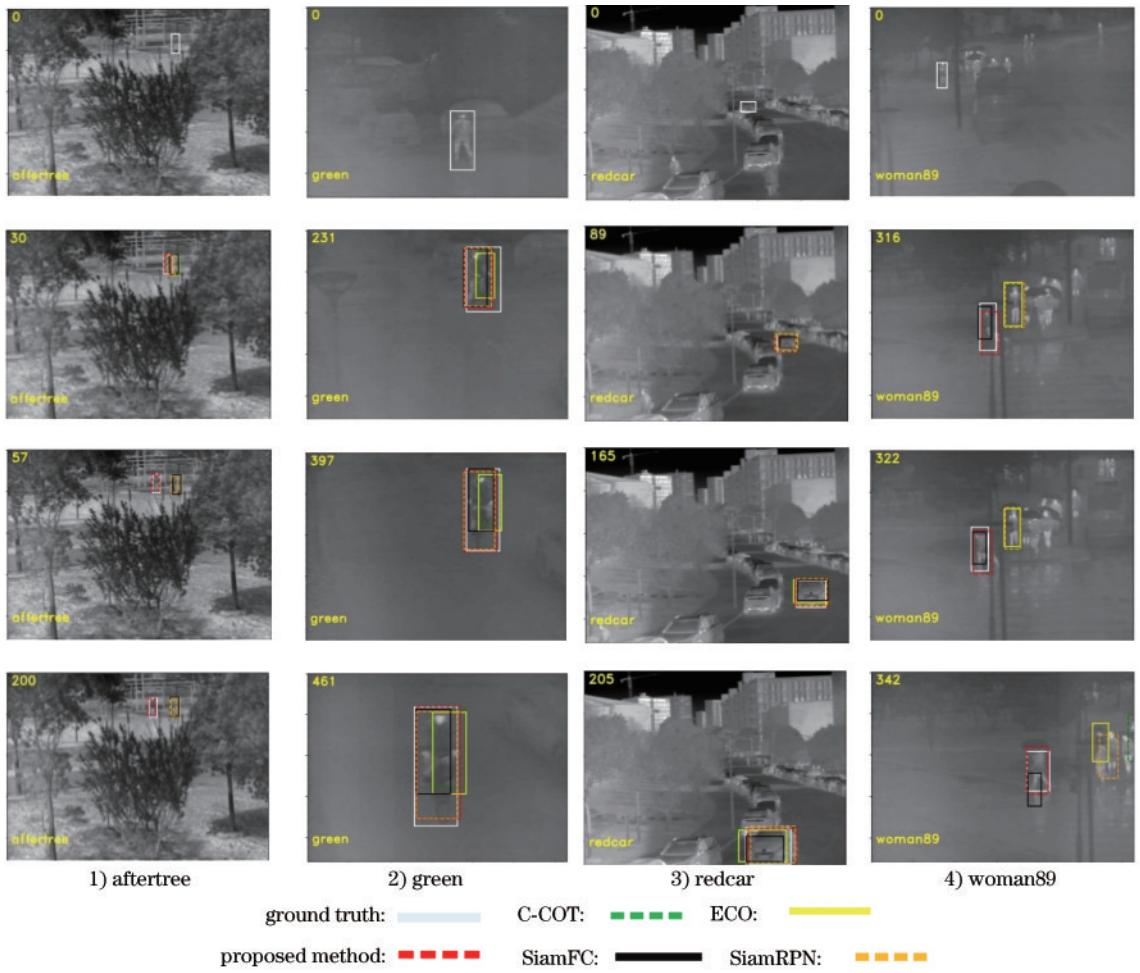

图 8　部分跟踪结果

Fig. 8　Some tracking results

1) aftertree 序列：本序列中的目标存在严重的相似物干扰且产生了一定程度的遮挡。在第 30 帧时，目标物被相似物遮挡，特征信息减少，极易产生跟踪错误，但是所提方法利用 D-ResNet 提取深度语义信息，实现目标准确定位。在第 57 帧，相似物对目标物造成严重的干扰，由于它们具有相同的目标特征，其他方法

表 1　评估序列

Table 1　Evaluation sequence

Sequence	Total frames	Challenge
aftertree	313	MC, OCC, SV, AI
green	491	MC, SV, CM
redcar	218	MC, SV
woman89	434	CM, IV, OCC, AI, MC

均出现跟踪错误甚至找不到目标的问题，所提方法利用位置感知模块和通道注意力模块突出关键信息，有效抵制相似物干扰，提高了辨别能力，实现了对红外目标的准确定位。

2) green 序列：在本序列中，目标在由远及近的运动过程中，相机发生移动。在第 231 帧镜头视角发生转变的情况下，所提方法仍能迅速准确捕捉目标，与真实框保持一致的运动变化，应对跟踪过程中的突发状况。

3) redcar 序列：随着目标的移动，本序列的目标尺度发生了变化。目标尺寸变大后，两种相关滤波算法由于使用多尺度目标估计，不能很好地适应目标变化。所提方法和 SiamRPN 由于有多个 RPN 模块融合，可自适应目标尺度变换，实现准确定位。

4) woman89 序列：本序列面临光照变化、相似物干扰的挑战。在第 316 帧，背景光变亮，ECO 等其他三种算法跟踪失败，所提方法得到的结果和真实框仍具有较高的重叠率。在第 342 帧，存在背景信息对正样本严重干扰的情况，所提方法利用位置感知模块建立与背景信息的联系，提取到更为丰富的语义信息，提高了在复杂场景中的稳定性。

所提方法对 4 种具有挑战性的热红外序列进行跟踪，在遮挡、尺度变化、相似物干扰、光照变化、运动变化等情况下仍然可以准确定位目标，可见在实际应用中所提方法有着较好的辨认能力和抗干扰能力，可有效应对目标跟踪过程中的难题。

3.5.2　定量分析

为进一步验证所提方法的能力，本实验在 VOT-TIR2019 和 GTOT 测试集上对 6 种方法采用一次通过评估模式（OPE）、成功率（success rate）、准确率（precision）进行评估[24]。成功率是指跟踪算法得到的 bounding box（a）与 ground truth（b）重叠区域的像素数目，表达式为

$$o_s = |a \cap b| / |a \cup b|。 \tag{6}$$

当某一帧的交并比（IoU）大于设定的阈值时，则该帧跟踪成功。准确率指追踪的目标位置中心点与人工标注目标位置的中心点两者的距离小于给定阈值的视频帧的百分比，不同的阈值具有不同的百分比

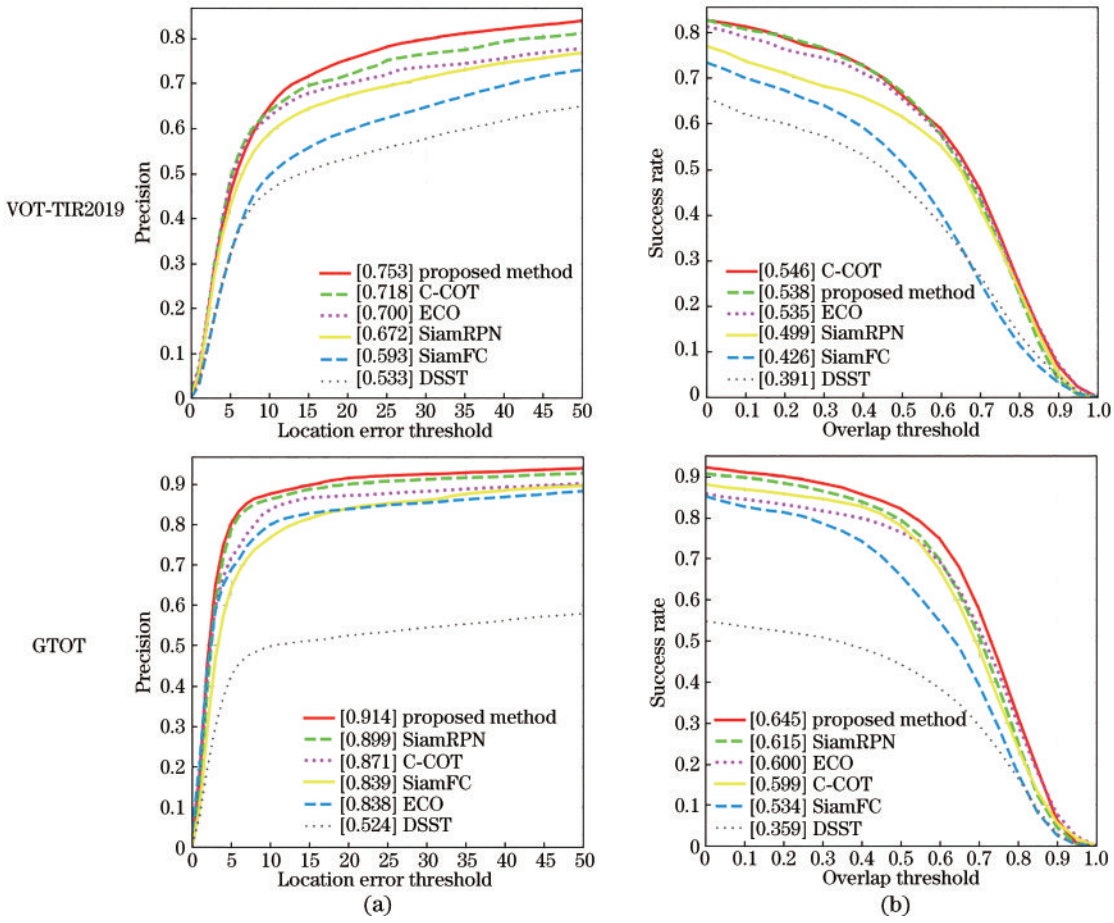

图 9　评估结果。(a)准确率；(b)成功率

Fig. 9　Evaluation results. (a) Precision; (b) success rate

(一般为20个像素)。如图9所示,所提方法在VOT-TIR2019数据集上的成功率和准确率分别为53.8%和75.3%,相比于ECO提升了0.3个百分点和5.3个百分点;在GTOT数据集上准确率高达91.4%,比C-COT高出4.3个百分点,超过了SiamFC和SiamRPN。表2为在VOT-TIR2019、GTOT数据集上6种跟踪器的实验结果对比,其中AUC为ROC曲线下的面积,ROC为受试者工作特征。在运行速率方面,所提方法的帧率高达30 frame/s,超过C-COT、ECO和DSST,满足实时性要求。

SiamFC由于网络结构简单,特征提取能力有限,不足以应对目标被遮挡、尺度发生变化等问题;ECO相关滤波算法采用传统特征提取方式建立目标的表观模型,在训练时引入背景像素,影响模型性能;C-COT算法精度尚可,但速度堪忧。相比于使用浅层特征提取网络的跟踪器,所提方法使用深度空洞残差网络提取到了更好的特征,加入位置感知模块和通道注意力模块,优化了提取到的目标特征信息,在目标不具有颜色特性的情况下仍然具有较强的鉴别性,更适应于对热红外目标的跟踪。

表2 在VOT-TIR2019、GTOT数据集上6种跟踪器的实验结果对比

Table 2 Comparison of experimental results of 6 trackers on VOT-TIR2019 and GTOT datasets

Method	VOT-TIR2019		GTOT		Speed /(frame·s^{-1})
	AUC	Precision	AUC	Precision	
Proposed method	0.538	0.753	0.645	0.914	30
C-COT	0.546	0.718	0.599	0.871	0.3
ECO	0.535	0.700	0.600	0.838	6
SiamRPN	0.499	0.672	0.615	0.899	160
SiamFC	0.426	0.594	0.534	0.839	58
DSST	0.391	0.533	0.359	0.524	24

4 结 论

跟踪方法对复杂场景位置感知能力不足会降低对热红外目标跟踪的准确率。针对这一问题,提出了基于位置感知的热红外目标跟踪方法。首先,使用深度空洞残差网络D-ResNet提取语义特征;其次,设计位置感知模块对空间位置信息进行建模;然后结合通道注意力模块进一步提取关键特征,获取红外目标的表观位置信息;最后,利用RPN模块实现对目标的准确定位。所提方法在VOT-TIR2019和GTOT数据集上与其他主流方法进行了对比实验,分别取得75.3%和91.4%的准确率。结果表明:在复杂环境中,所提方法可应对运动变化、相似物干扰等问题,跟踪速度虽满足实时性要求,但仍有提升空间。未来的研究重点是结合热红外目标特性对特征提取网络进行剪枝优化,在保证精度的情况下,提高跟踪速率,实现性能的全面提升。

参 考 文 献

[1] 王鹏, 孙梦宇, 王海燕, 等. 一种目标响应自适应的通道可靠性跟踪算法[J]. 电子与信息学报, 2020, 42(8): 1950-1958.
Wang P, Sun M Y, Wang H Y, et al. An object tracking algorithm with channel reliability and target response adaptation[J]. Journal of Electronics & Information Technology, 2020, 42(8): 1950-1958.

[2] 孟琭, 杨旭. 目标跟踪算法综述[J]. 自动化学报, 2019, 45(7): 1244-1260.
Meng L, Yang X. A survey of object tracking algorithms [J]. Acta Automatica Sinica, 2019, 45(7): 1244-1260.

[3] Bromley J, Bentz J W, Bottou L, et al. Signature verification using a "Siamese" time delay neural network [J]. International Journal of Pattern Recognition and Artificial Intelligence, 1993, 7(4): 669-688.

[4] Tao R, Gavves E, Smeulders A W M. Siamese instance search for tracking[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 1420-1429.

[5] Bertinetto L, Valmadre J, Henriques J F, et al. Fully-convolutional Siamese networks for object tracking[M]// Hua G, Jégou H. Computer vision-ECCV 2016 workshops. Lecture notes in computer science. Cham: Springer, 2016, 9914: 850-865.

[6] 金国栋, 薛远亮, 谭力宁, 等. 基于孪生神经网络的目标跟踪算法进展研究[J]. 系统工程与电子技术, 2022, 44(6): 1805-1822.
Jin G D, Xue Y L, Tan L N, et al. Advances in object tracking algorithm based on Siamese network[J]. Systems Engineering and Electronics, 2022, 44(6): 1805-1822.

[7] Ren S Q, He K M, Girshick R, et al. 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.

[8] Li B, Wu W, Wang Q, et al. SiamRPN: evolution of Siamese visual tracking with very deep networks[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 4277-4286.

[9] 杨福才, 杨德东, 毛宁, 等. 基于稀疏编码直方图的稳健红外目标跟踪[J]. 光学学报, 2017, 37(11): 1115002.

Yang F C, Yang D D, Mao N, et al. Robust infrared target tracking based on histograms of sparse coding[J]. Acta Optica Sinica, 2017, 37(11): 1115002.

[10] 李畅, 杨德东, 宋鹏, 等. 基于全局感知孪生网络的红外目标跟踪[J]. 光学学报, 2021, 41(6): 0615002.
Li C, Yang D D, Song P, et al. Global-aware Siamese network for thermal infrared object tracking[J]. Acta Optica Sinica, 2021, 41(6): 0615002.

[11] 钱琨. 基于机器学习理论的红外目标跟踪技术研究[D]. 西安: 西安电子科技大学, 2018.
Qian K. Study on infrared target tracking technology based on machine learning theory[D]. Xi'an: Xidian University, 2018.

[12] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition, June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 3431-3440.

[13] Wang X L, Girshick R, Gupta A, et al. Non-local neural networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 7794-7803.

[14] Fu J, Liu J, Tian H J, et al. Dual attention network for scene segmentation[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 3141-3149.

[15] Wang Q, Teng Z, Xing J L, et al. Learning attentions: residual attentional Siamese network for high performance online visual tracking[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 4854-4863.

[16] Real E, Shlens J, Mazzocchi S, et al. YouTube-BoundingBoxes: a large high-precision human-annotated data set for object detection in video[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 7464-7473.

[17] Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211-252.

[18] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: common objects in context[M]//Fleet D, Pajdla T, Schiele B, et al. Computer vision-ECCV 2014. Lecture notes in computer science. Cham: Springer, 2014, 8693: 740-755.

[19] Li C L, Liang X Y, Lu Y J, et al. RGB-T object tracking: benchmark and baseline[J]. Pattern Recognition, 2019, 96: 106977.

[20] Li C L, Cheng H, Hu S Y, et al. Learning collaborative sparse representation for grayscale-thermal tracking[J]. IEEE Transactions on Image Processing, 2016, 25(12): 5743-5756.

[21] Danelljan M, Bhat G, Khan F S, et al. ECO: efficient convolution operators for tracking[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6931-6939.

[22] Danelljan M, Robinson A, Shahbaz Khan F, et al. Beyond correlation filters: learning continuous convolution operators for visual tracking[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9909: 472-488.

[23] Danelljan M, Häger G, Shahbaz Khan F, et al. Accurate scale estimation for robust visual tracking[C]//Proceedings of the British Machine Vision Conference 2014, September 1-5, 2014, Nottingham, UK. London: British Machine Vision Association, 2014: 1-11.

[24] Danelljan M, Häger G, Khan F S, et al. Learning spatially regularized correlation filters for visual tracking [C]//2015 IEEE International Conference on Computer Vision, December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 4310-4318.