第 60 卷 第 4 期/2023 年 2 月/激光与光电子学进展 研究论文

激光与光电子学进展 先进成像

基于非局部感知网络的运动目标跟踪方法

张立国，马子荐*，金梅，李义辉

燕山大学电气工程学院，河北 秦皇岛 066000

摘要 针对跟踪运动目标过程中网络对目标被遮挡或目标周围存在干扰物敏感，从而导致不可靠的响应位置和错误跟踪框的问题，提出一种基于深度学习的免锚框孪生卷积网络跟踪方法。首先，通过非局部感知网络来学习目标引导的特征权重，该权重用于细化目标模板分支和搜索分支的深度特征，以监督的方式利用两个分支特征的远程依赖性，从而有效抑制噪声干扰。其次，进一步开发一个包围框感知块将多维回归特征与跟踪质量相关联，这个模块加强目标模板分支和搜索分支之间的相互作用，提高网络定位准确性。在标准数据集上的实验结果表明，所提方法能实时跟踪目标，并在准确度上获得提升。

关键词 机器视觉；目标跟踪；深度学习；非局部感知网络；孪生卷积网络

中图分类号 TP391 **文献标志码** A **DOI**：10.3788/LOP212946

Nonlocal Neural Network-Based Moving Target Tracking Method

Zhang Liguo, Ma Zijian*, Jin Mei, Li Yihui

Institute of Electrical Engineering, Yanshan University, Qinhuangdao 066000, *Hebei, China*

Abstract Typically, networks are sensitive to targets being blocked or interference around the target when tracking moving targets, resulting in unreliable response positions and incorrect tracking frame. Thus, an anchor-free Siamese network-tracking approach based on deep learning is proposed. First, the feature weight of the target guidance is derived through the nonlocal perceptual network, which is then applied to refine the depth features of the target template branch and search branch, and to improve the remote dependence of the two branch features in a supervised manner to effectively suppress noise interference. Second, to correlate the multidimensional regression features with the tracking quality, a bounding box perception block is developed. This module strengthens the interaction between the target template branch and the search branch and enhances the accuracy of network positioning. Furthermore, the proposed method can track the target in real time and enhance accuracy, according to the experimental findings on standard data sets.

Key words machine vision; target tracking; deep learning; nonlocal network; Siamese network

1 引 言

运动目标跟踪任务无论在智能安防还是在智能机器人领域一直是国内外科研人员研究的重点。对于该任务，跟踪器或者跟踪系统根据给定视频第1帧中的已知目标位置来定位后续视频帧中的给定目标。随着研究的不断深入，近年来目标跟踪领域取得很大进展，但由于跟踪目标的图像信息会随着目标移动或摄像机的移动而变化，且跟踪过程中目标还会受到遮挡和背景杂波等因素的影响。因而，如何更加快速、准确地确定跟踪目标位置和大小仍然是视觉跟踪领域中具有挑战性的问题[1]。

传统的目标跟踪主要有卡尔曼滤波法和粒子滤波法等，这些方法无法处理和适应跟踪中复杂的变化情况，鲁棒性和准确度较差，渐渐变为一种辅助手段[2]。当前运动目标跟踪算法集中在特征提取、模板更新、目标定位和边界框回归等不同方面进行研究。早期的特征提取主要使用手工制作的特征如灰度特征或颜色特征等，如Danelljan等[3]使用颜色特征提高算法性能。随着卷积神经网络的深度卷积特性被开发出来，深度特征开始被广泛采用。2016年，SiamFC[4]跟踪方法被提出，它构建一个完全卷积的孪

收稿日期：2021-11-12；**修回日期**：2021-12-08；**录用日期**：2021-12-21；**网络首发日期**：2021-12-30

基金项目：河北省科学技术研究与发展计划科技支撑计划(20310302D)、河北省中央引导地方专项(199477141G)

通信作者：*1239038456@qq.com

生(Siamese)网络用于特征提取,在保证跟踪精确度的同时,跟踪速度可以达到实时。在跟踪过程中,模板更新可以提高模型的适应性,但这种实时更新模板的在线跟踪方法会遇到因干扰导致的跟踪漂移问题,并且运行速度往往很慢。目标定位和边界框回归的改进往往从目标检测领域获得启发。将Fast-RCNN[5]中基于锚点的区域建议网络应用到跟踪器设计中,Li等[6]重新设计SiamRPN跟踪器的分类和边框回归分支,减小目标尺度变化对跟踪器的影响,使跟踪器专注于目标前景背景分类和边界框回归。随着免锚框检测网络的发展,2020年,Xu等[7]设计SiamFC++跟踪网络,它舍弃了那些预定义的锚框从而让网络能够直接得到被跟踪目标的边框,这极大提高了目标跟踪精确度和效率。然而,当前孪生网络跟踪的瓶颈在于很难消除目标周围有背景干扰物或目标被遮挡等情况下跟踪器对特征表示的副作用,即背景中的语义干扰总是阻碍网络学习给定目标的区别性表示,使得网络在跟踪场景中很难将目标与其他背景准确地区分开来。

针对上述问题,本文设计一个在线训练和离线跟踪的免锚框孪生卷积网络跟踪器。在孪生卷积网络提取模板特征和搜索特征后,通过非局部感知网络来提高跟踪器的特征感知能力,进而提高算法的抗干扰能力。在跟踪器末端应用基于回归特征增强的分类结构来改善跟踪过程中在目标受遮挡等情况下跟踪器的定位效果,在不影响算法速度的前提下提高目标跟踪准确性。

2　基于孪生卷积网络的目标跟踪方法

2.1　免锚框孪生卷积网络目标跟踪器

直接估计目标状态的免锚框机制因其强大的跟踪性能和快速的推理速度而在视觉跟踪任务中受到广泛关注。免锚框孪生卷积网络跟踪器通过在视频某一帧中选择跟踪对象,在接下来的连续帧中将目标框选出来达到目标跟踪的目的,其中跟踪对象最初的图像信息为目标模板图像,待定位跟踪目标的图像为目标搜索图像。孪生卷积网络包括目标模板分支和搜索分支,目标模板分支将目标模板图像$\boldsymbol{Z}$转换为深度特征,目标搜索分支将目标搜索图像$\boldsymbol{X}$转换为深度特征,利用提取的两个分支的特征,通过互相关运算进一步计算相关性图。接着跟踪器用一个目标分类和目标边框回归的区域估计网络组成分类分支和回归分支,从分类分支相关性图中得到跟踪目标在搜索图像中的目标位置,从回归分支相关性图中得到跟踪目标包围框的相关信息,最后将信息映射回视频帧获得跟踪结果。在本研究的实现中,如图1所示,首先通过AlexNet[8]共享主干网络(backbone)构成孪生卷积网络的目标模板分支和搜索分支。目标模板分支和搜索分支共享相同的网络参数,以确保模板图像和搜索图像变换相同,这对于相关信息学习至关重要。然后,通过互相关层来计算相关性图。互相关层中的计算可描述为

$$\boldsymbol{A}^{i}=[\sigma(\boldsymbol{X})]*[\sigma(\boldsymbol{Z})],i\in\{\mathrm{cls},\mathrm{reg}\},\tag{1}$$

式中:$*$表示互相关运算;σ表示卷积特征嵌入;$\boldsymbol{X}$表示目标模板图像;$\boldsymbol{Z}$表示目标搜索图像;$\boldsymbol{A}^{\mathrm{cls}}$是分类分支得到的相关性图;$\boldsymbol{A}^{\mathrm{reg}}$是回归分支得到的相关性图。

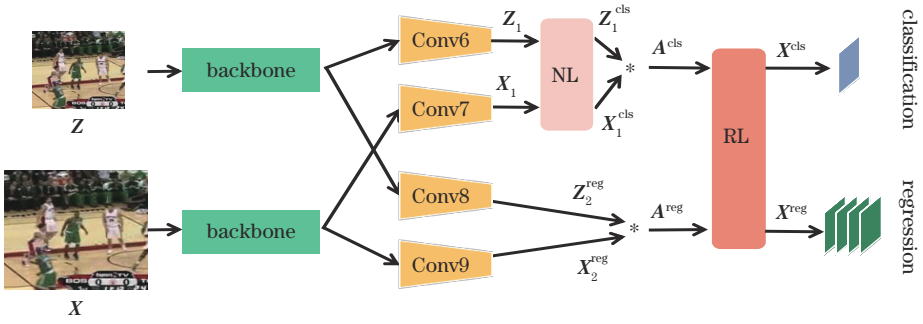

图1　跟踪方法示意图

Fig. 1　Schematic diagram of tracking method

在共享主干网络提取两个分支特征后,通过参数不同的卷积操作和批归一化来进一步对特征进行调整,使特征满足计算分类相关性图和回归相关性图的差异性表达。具体地,模板分支特征和搜索分支特征分别通过由3×3卷积和批归一化组成的Conv6和Conv7得到模板分类特征$\boldsymbol{Z}_1$和搜索分类特征$\boldsymbol{X}_1$,通过Conv8和Conv9得到回归任务的特征。在分类特征上应用非局部网络结构(NL)得到特征增强的分类特征图$\boldsymbol{X}_1^{\mathrm{cls}}$和$\boldsymbol{Z}_1^{\mathrm{cls}}$,以此来提高两分支的特征整合能力,抑制目标搜索分支可能存在的特征干扰。使用模板特征作为卷积核与搜索特征进行卷积相关运算,将相关特征调整到特定于分类和回归任务的特征空间中,得到$\boldsymbol{A}^{\mathrm{cls}}$和$\boldsymbol{A}^{\mathrm{reg}}$。接着将分类特征和回归特征通过包围框感知块(RL)进行整合使回归任务参与分类任务的质量估计,提高跟踪器面对干扰或遮挡等情况下目标跟踪精确度,最终得到相应的分类图$\boldsymbol{X}^{\mathrm{cls}}$和回归图$\boldsymbol{X}^{\mathrm{reg}}$。分类图中具有最大值的位置表示目标的位置,回归图中与目标位置对应的特征值为当前跟踪帧中目标的包围框大小。主干网络结构如图2所示。

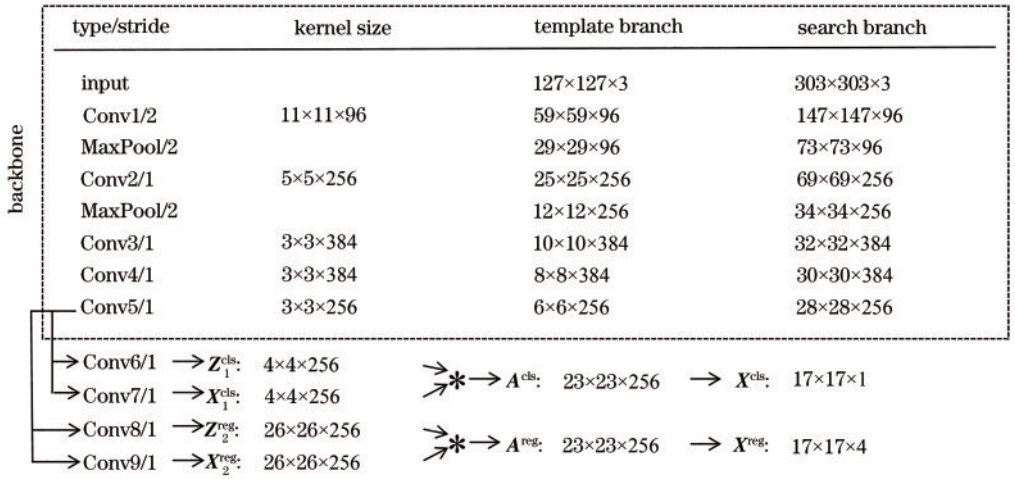

图 2　主干网络结构

Fig. 2　Backbone network structure

2.2　非局部感知网络结构

原始的孪生网络跟踪器通过大量图像对的训练来学习目标的跟踪特征。然而，这些特征的辨别力较弱，当类似的干扰物体出现时，跟踪器很容易被误导。此外，运动目标的搜索分支往往包含杂乱的背景，且目标可能不在搜索区域的中心。这些问题容易导致跟踪过程中跟踪器偏离目标位置[9]。因此，跟踪器设计过程中需要考虑减轻干扰物对特征的负面影响，增强搜索分支的语义识别能力和定位能力。从特征通道的角度来看，不同通道的视觉语义是不同的。在孪生网络跟踪器中，由于搜索分支与模板分支共享相同的主干网络，因此两个分支的相同语义在某些通道中具有高度重叠的性质。基于这一认识，在非局部感知网络中将模板分支的通道特征信息引入搜索分支中，通过两个分支特征中的依赖属性来获得有利于目标辨别的语义关系，进而提高目标跟踪过程中网络对搜索分支中目标的识别能力[10]。

具体地，非局部特征网络的结构如图 3 所示。给定来自运动目标的模板特征 $\boldsymbol{Z}_1$ 和搜索特征 $\boldsymbol{X}_1$。非局部特征关系感知块利用 3 种位置线索，即每个通道特征的平均值信息、最大值信息及不同通道的相关性信息来构成非局部特征网络块。该模块通过整合 3 种位置线索得到非局部感知网络的特征权重信息。每个通道特征的平均值信息和最大值信息由全局平均池化和全局最大池化来获得，通道间相关性信息由不同通道间特征经过卷积细化后获得。

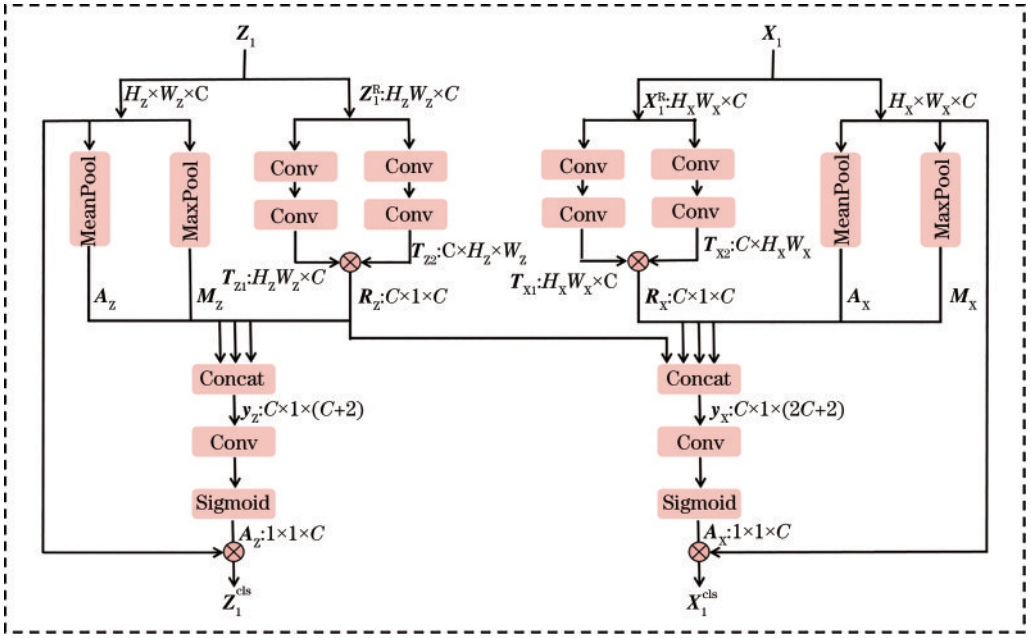

图 3　NL 模块结构

Fig. 3　NL module structure

非局部感知网络模块对目标模板分支和搜索分支采用不同的操作。对于目标模板分支的$\boldsymbol{Z}_1$特征，将平均池化全局信息$\boldsymbol{A}_Z$、最大池化全局信息$\boldsymbol{M}_Z$与通道间相关信息$\boldsymbol{R}_Z$按照通道维度拼接得到响应信息$\boldsymbol{y}_Z$。在通道间相关信息$\boldsymbol{R}_Z$生成过程中，将$\boldsymbol{Z}_1$特征每个通道作为特征节点，调整其维度使其变为$\boldsymbol{Z}_1^{\mathrm{R}}$。$\boldsymbol{Z}_1^{\mathrm{R}}$特征分为两个分支，通过卷积操作将特征通道数增加后继续通过卷积操作将该特征调整回原通道数，以此来整合特征信息，形成了两个节点图$\boldsymbol{T}_{Z1}$和$\boldsymbol{T}_{Z2}$。两个节点图中的元素相乘可以定义为嵌入空间中$\boldsymbol{T}_{Z1}$和$\boldsymbol{T}_{Z2}$的点积亲和度$\boldsymbol{R}_Z$[11]。相关信息$\boldsymbol{R}_Z$的值表示每个通道和其他通道的成对关系，即对于通道i，获得它与所有通道的关系。接着将R_Z通过拼接操作与最大池化和平均池化信息整合得到关系向量$\boldsymbol{y}_Z$来表示整合的响应信息。最后通过卷积来细化通道关系并由Sigmoid函数得到特征权重图$\boldsymbol{A}_Z$，并将特征权重图与原$\boldsymbol{Z}_1$特征相乘得到增强特征$\boldsymbol{Z}_1^{\mathrm{cls}}$。

$$\boldsymbol{y}_Z=\mathrm{Cat}\left[\mathrm{pool}(\boldsymbol{Z}_1),\max(\boldsymbol{Z}_1),\boldsymbol{R}_Z\right],\tag{2}$$

式中：Cat为按照通道维度对特征进行拼接操作；pool为全局平均池化操作；max为全局最大池化操作；$\boldsymbol{Z}_1$为目标模板特征；$\boldsymbol{R}_Z$为目标模板特征的通道间相关信息。

针对$\boldsymbol{X}_1$特征图，同样获得平均池化信息$\boldsymbol{A}_X$、最大池化信息$\boldsymbol{M}_X$和表示$\boldsymbol{X}_1$中所有通道间关系的相关信息$\boldsymbol{R}_X$。为了应用搜索分支和模板分支中的依赖属性进一步消除目标周围干扰物影响，将模板分支通道间相关信息$\boldsymbol{R}_Z$通过按维度拼接操作与搜索分支特征整合到一起，这利用了孪生卷积网络中两个分支的相同语义在某些通道中具有高度重叠的性质。通过拼接操作获得的向量$\boldsymbol{y}_X$表示搜索分支整合的响应信息。最后通过卷积和Sigmoid操作得到特征权重图$\boldsymbol{A}_X$并将特征权重图与原$\boldsymbol{X}_1$特征相乘得到增强特征$\boldsymbol{X}_1^{\mathrm{cls}}$。

$$\boldsymbol{y}_X=\mathrm{Cat}\left[\mathrm{pool}(\boldsymbol{X}_1),\max(\boldsymbol{X}_1),\boldsymbol{R}_X,\boldsymbol{R}_Z\right]。\tag{3}$$

原始的跟踪器没有目标相关信息的监督，搜索分支并不能保证那些与目标相关的区域不受干扰物的影响得到最大的关注。非局部感知网络模块的主要作用是引入全局信息和局部关联信息。该模块的这种关联信息相互作用可以减少背景干扰物对搜索分支的负面影响，从而有助于在搜索区域中定位目标。因此，采用非局部目标感知网络来学习特征权重的跟踪器可以通过对特征通道重要性的再分配来增强网络对目标的关注效果，进而提高跟踪器性能。图4展示了非局部感知网络模块改进后的模型效果。在使用非局部感知网络模块的图4(b)中，跟踪器的目标热值信息更加集中于跟踪目标中心，说明非局部感知网络模块提高了跟踪器的抗干扰能力。

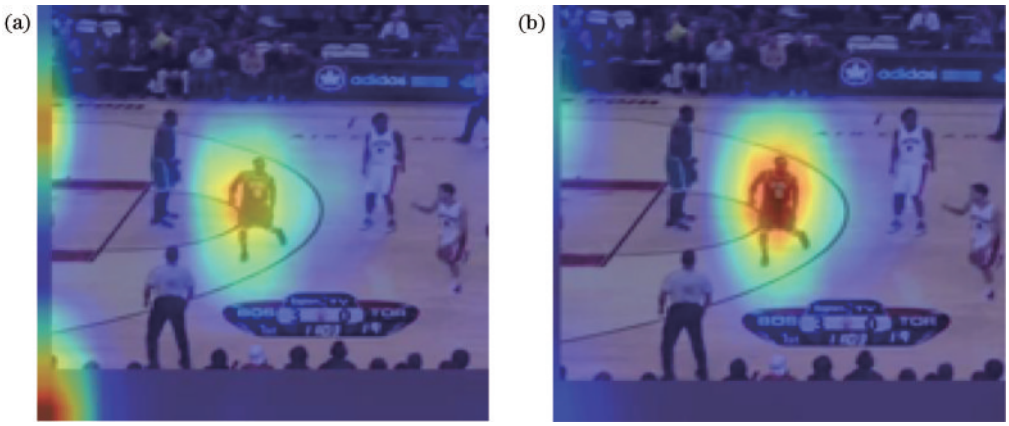

图4　分类分支特征可视化图。(a)未改进结果；(b)改进后结果

Fig. 4　Visualization diagrams of classified branch characteristics. (a) Original result; (b) improved result

2.3　基于回归特征增强的分类结构

原始的目标跟踪网络中分类分支负责将目标与背景区分开，回归分支负责回归目标边界框。利用同一骨干网络提取特征的孪生网络跟踪器将用于区分跟踪帧间目标前景和背景的目标分类任务与目标边框回归任务分开计算，这限制了跟踪器的性能[12]。

根据跟踪器分类任务相似度分数(score)、目标偏移距离(offset)和目标边框重叠率(overlap)对运动目标跟踪结果进行分析。分类相似度分数最大值所在位置为跟踪目标位置；目标偏移距离指的是当前帧中跟踪器预测目标位置中心和实际目标中心的欧氏距离，可以衡量目标跟踪过程中跟踪位置的偏离程度；边框重叠率指的是跟踪框和目标实际边框的交并比，可以衡量目标跟踪是否成功。图5(a)将不同目标偏移距离的跟踪结果所占跟踪成功结果总数的比例展示出来，图5(b)为不同边框重叠率所占跟踪成功总数的比例。

由图5(a)和5(b)可见，当分类分数大于0.5时，跟踪结果中更低目标偏移距离和更高边框重叠率的跟踪结果所占比例变高，这说明分类分数与跟踪器定位质量和边框回归质量具有一定联系。这是由于在目标周围存在干扰物情况下，目标周围的模糊或者遮挡会使

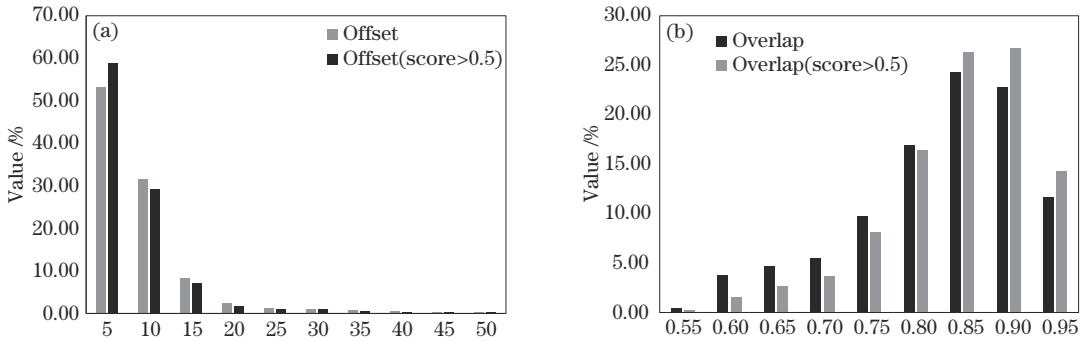

图 5　分类分数数据分布。(a)偏移距离;(b)边框重叠率

Fig. 5　Classification score data distribution. (a) Offset; (b) overlap

分类网络难以将目标位置准确估计出来,进而使网络对目标的区分能力降低,降低了分类分数。针对上述问题,本研究提出一种通过特征整合使边框回归特征指导分类相似度计算的包围框感知块来增强目标边框回归和目标分类之间的关系,进而提高网络对目标的定位能力。

基于回归特征增强分类效果的结构区别于传统跟踪器使分类特征和回归特征单独参与分类任务和回归任务,它的主要作用在于增加细化边框回归的特征信息,将整合后的回归信息用于指导分类任务。包围框感知模块结构如图 6 所示,通过卷积操作将分类相关性图 $\boldsymbol{A}^{\mathrm{cls}}$ 和回归相关性图 $\boldsymbol{A}^{\mathrm{reg}}$ 分别调整为用于分类任务和回归任务的分类图 $\boldsymbol{X}^{\mathrm{cls}}$ 和回归图 $\boldsymbol{X}^{\mathrm{reg}}$。不同的相关性图通过 3×3 卷积操作得到 $\boldsymbol{A}_1$ 和 $\boldsymbol{A}_2$,$\boldsymbol{A}_2$ 在 1×1 卷积作用下可以得到回归图。在通过回归特征信息对分类特征进行选择方面,回归特征信息 $\boldsymbol{A}_2$ 经过模块中的 3×3 卷积操作细化特征,通过 1×1 卷积操作将通道数降低,便于提高模型计算能力,接着通过 Sigmoid 激励层获得分类特征增强信息 $\boldsymbol{A}_4$。将分类特征增强信息 $\boldsymbol{A}_4$ 与分类特征 $\boldsymbol{A}_3$ 相乘后与 $\boldsymbol{A}_3$ 相加,得到更加具有区分力的分类图。在包围框感知块应用后,网络对目标定位能力增强,如图 7 所示,相比图 7(a)中目标实际边框与跟踪

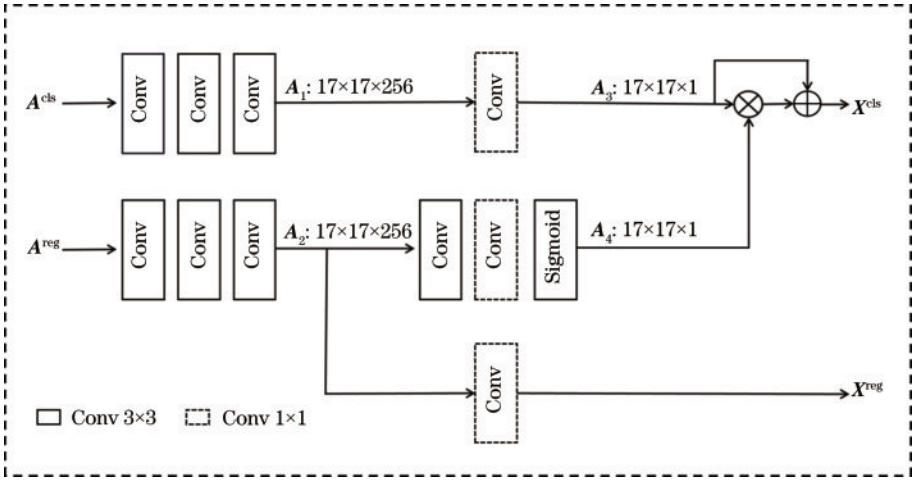

图 6　RL 模块结构

Fig. 6　RL module structure

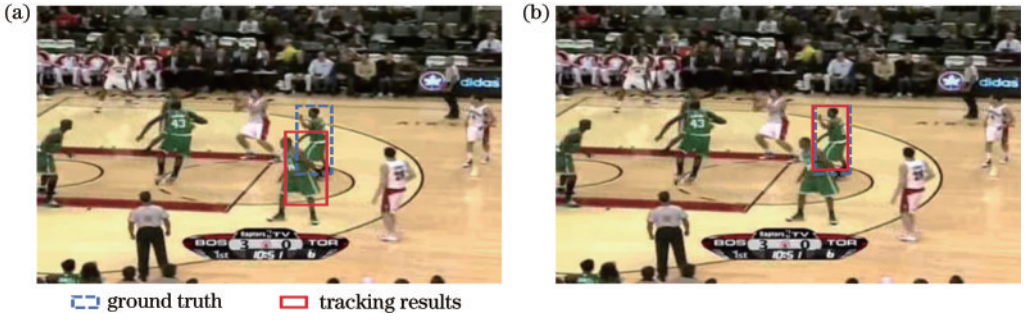

图 7　跟踪结果可视化。(a)未改进结果;(b)改进后结果

Fig. 7　Visualization of tracking results. (a) Original result; (b) improved result

边框偏移的情况，应用包围框感知模块后的图 7(b)中，两个边框几乎重叠，跟踪结果更好。这说明应用包围框感知模块的跟踪器对目标的定位更加准确，同时回归得到的边界框误差降低，跟踪器不易发生漂移，模型的跟踪精确度获得提高。

2.4　损失函数

在对孪生网络提取的不同分支特征进行互相关操作后，设计了分类网络和回归网络，接着将跟踪器的训练损失分为分类损失和回归损失。对于分类特征图中的每个像素，特征图上的每个位置落入目标真实边界框的值为正样本，否则为负样本。将与目标位置对应的像素分类为正样本和负样本后进行训练得到分类损失值[13]：

$$l_{\mathrm{cls}}=\begin{cases}-\partial_{\mathrm{t}}(1-p_{\mathrm{t}})^{\beta}\log p_{\mathrm{t}}, & y=1\\ -(1-\partial_{\mathrm{t}})^{\beta}\log(1-p_{\mathrm{t}}), & y=0\end{cases},\tag{4}$$

式中：p_{t}为网络的估计值；y为模板值(正样本时y为1，负样本时y为0)；∂取0.05；β为0.01。

对于回归分支来说，与分类图上最大值相同的位置对应着回归分支对目标边框的估计值。免锚框孪生卷积网络跟踪器意味着回归图上每个位置对应的回归头输出的4个偏移值可以不需要锚框来预测目标边界框位置[14]。将网络预测的目标边框4个边距目标真实边框的距离表示为向量$\boldsymbol{t}=(l,t,r,b)$。因此，位置(x,y)的回归目标可以表述为

$$\begin{cases}l=\left(\left|\dfrac{s}{2}\right|+xs\right)-x_0\\ t=\left(\left|\dfrac{s}{2}\right|+ys\right)-y_0\\ r=x_1-\left(\left|\dfrac{s}{2}\right|+xs\right)\\ b=y_1-\left(\left|\dfrac{s}{2}\right|+ys\right)\end{cases},\tag{5}$$

式中：(x_0,y_0)和(x_1,y_1)代表与(x,y)关联的目标真实框的左上角和右下角；s是主干网络的步长，取值为8。

通过计算目标真实框和跟踪器估计的目标边框的交并比(IOU)，可以得到网络的回归损失值。最终的损失函数包括分类损失和回归损失。

$$L_{\mathrm{reg}}=1-R_{\mathrm{IOU}},\tag{6}$$

$$L=L_{\mathrm{cls}}+\lambda L_{\mathrm{reg}},\tag{7}$$

式中：超参数λ表示损失计算参数，经实验发现，将λ设置为3时损失函数对模型优化效果最好；L_{cls}为分类端损失；L_{reg}为回归端损失。

3　实验与分析

3.1　神经网络训练

本研究主要采用GOT10K数据集[15]，一共筛选了74 GB图像数据，近9335个视频作为训练集。网络实际训练时，使用在ImageNet数据预训练的AlexNet骨干网络。训练使用动量为0.9的随机梯度下降算法进行优化。训练批量大小设为64，通过将训练集中帧差小于100的一对图像组成训练样本对神经网络进行训练，同时对训练图像进行随机移位和缩放以达到数据增强的目的。实验以瑞龙R3600 CPU和英伟达RTX 3060 GPU为硬件平台。软件平台为Ubuntu 16.04、CUDA 11.1、Python 3.8和Pytorch框架。

3.2　实验结果比较分析

使用精确度、成功率、速度等3个指标评价跟踪器在一个视频上的跟踪效果。跟踪精确度(DP)为跟踪过程中目标偏移距离大于某一阈值的帧数与跟踪总帧数的比例，跟踪精确度曲线将不同阈值下的跟踪精确度记录下来。使用成功率曲线下面积(AUC)来评价跟踪器在所有跟踪视频中的跟踪效果好坏。成功率为边框重叠率大于阈值的跟踪帧所占视频总帧数的比例，分别记录不同阈值下跟踪成功率的变化情况得到成功率曲线。为了充分评估所提算法的有效性，在标准视觉跟踪基准数据集OTB100[16]中存在遮挡、形变、运动模糊和背景干扰注释属性的视频数据上对改进策略进行消融分析，比较结果如表1所示。消融实验对NL模块形成$\boldsymbol{Z}_1$与$\boldsymbol{X}_1$的响应信息时在特

表1　跟踪结果消融分析

Table 1　Ablation analysis of tracking results

Test	NL							RL	DP	AUC
	$\boldsymbol{Z}_1$			$\boldsymbol{X}_1$						
	$\boldsymbol{A}_{\mathrm{Z}}$	$\boldsymbol{M}_{\mathrm{Z}}$	$\boldsymbol{R}_{\mathrm{Z}}$	$\boldsymbol{A}_{\mathrm{X}}$	$\boldsymbol{M}_{\mathrm{X}}$	$\boldsymbol{R}_{\mathrm{X}}$	$\boldsymbol{R}_{\mathrm{Z}}$			
1	√			√				√	0.822	0.629
2	√			√					0.802	0.611
3	√	√		√	√				0.803	0.615
4			√			√			0.808	0.618
5			√			√	√		0.813	0.627
6	√	√	√	√	√	√	√		0.829	0.626
7	√	√	√	√	√	√	√	√	0.836	0.633

征拼接阶段应用的3种感知线索(平均池化全局信息、最大池化全局信息与通道间相关信息)和RL模块的应用对跟踪器的影响进行了研究。由实验2、实验3和实验4可以看出,NL模块中3种感知线索对跟踪器性能均有影响,其中实验4情况下不同通道的相关信息$\boldsymbol{R}_{\mathrm{Z}}$和$\boldsymbol{R}_{\mathrm{X}}$的引入对跟踪精确度提高影响最大,跟踪精确度达到80%。实验6的跟踪精确度高于实验5,说明NL模块中搜索分支中引入模板分支相关信息$\boldsymbol{R}_{\mathrm{Z}}$的重要性。由实验6和实验7可以看出,跟踪器应用RL模块后跟踪结果获得了提高,说明所提改进方法能够提高跟踪器对运动目标跟踪的精确度和成功率。

为了进一步评估所提算法的有效性,将所提算法在OTB100数据集[16]和UAV123数据集[17]上与当前流行的SiamFC[4]、SiamRPN[6]和SiamFC++[7]算法进行对比评估。同时为了评估跟踪器在应对遮挡或干扰等能力方面的改善,将OTB100数据集中存在遮挡和形变注释属性的视频数据及运动模糊和背景干扰注释属性的视频数据分成两组作为测试数据。通过跟踪准确率(阈值20)对不同算法进行比较排序。表2、表3和表4为所提算法与其他跟踪算法在跟踪成功率和精确度上的评估结果。所提算法在3个评估实验上的评估精确度分别达到了84%、82%和77%,超过了其他具有竞争力的算法,并且跟踪速度可以达到198 frame·s^{-1},远远超过实时跟踪(25 frame·s^{-1})的要求。图8为跟踪结果分析图。从图8可以看出,相比于其他算法,所提算法在两组OTB评估视频中的精确度曲线和成功率曲线均取得了更好的结果,说明其在跟踪准确度和跟踪成功率上更优,跟踪器可以较好地应对运动目标跟踪过程中遇到的遮挡或干扰噪声等造成的跟踪效果下降问题。对比实验表明,通过非局部块增强特征信息,同时引入回归特征增强的分类结构可以使跟踪器更加关注目标本身的位置和边框信息,实现运动目标的准确跟踪。

图9为所提算法与其他3种算法的跟踪结果可视化对比。在典型的视频序列尤其是存在遮挡和背景干扰条件下,进行各算法的比较分析。从图9可以看出,在Girl2视频中存在大量遮挡和相似目标干扰情况,随着目标被周围相似目标遮挡情况的发生,另外3种跟踪算法均发生了不同程度的跟踪漂移,甚至出现跟踪目标丢失情况。所提算法由于更加关注目标在复杂环境中特征的变化情况,可以有效增强跟踪器对目标的判别能力,实现准确跟踪。Board视频序列的主要难度是存在背景干扰,尤其是在249帧中只有所提算法可以准确跟踪目标。这表明利用回归特征和分类特征的相干性可以使跟踪器更加关注目标的变化位置,进而可以准确区分目标。

表2 不同跟踪算法在OTB100遮挡和形变数据上对比结果

Table 2 Comparison results of different algorithms on OTB100 occlusion and deformation dataset

Algorithm	DP	AUC	Detection speed /(frame·s^{-1})
SiamFC[4]	0.707	0.520	85
SiamRPN[6]	0.717	0.535	180
SiamFC++[7]	0.801	0.593	228
Proposed algorithm	0.846	0.626	217

表3 不同跟踪算法在OTB100运动模糊和背景干扰数据上对比结果

Table 3 Comparison results of different algorithms on OTB100 motion blur and background interference dataset

Algorithm	DP	AUC	Detection speed /(frame·s^{-1})
SiamFC[4]	0.713	0.545	82
SiamRPN[6]	0.731	0.548	178
SiamFC++[7]	0.794	0.612	227
Proposed algorithm	0.820	0.623	215

表4 不同跟踪算法在UAV123上对比结果

Table 4 Comparison results of different algorithms on UAV123 dataset

Algorithm	DP	AUC	Detection speed /(frame·s^{-1})
SiamFC[4]	0.691	0.497	78
SiamRPN[6]	0.703	0.525	177
SiamFC++[7]	0.744	0.580	205
Proposed algorithm	0.775	0.595	198

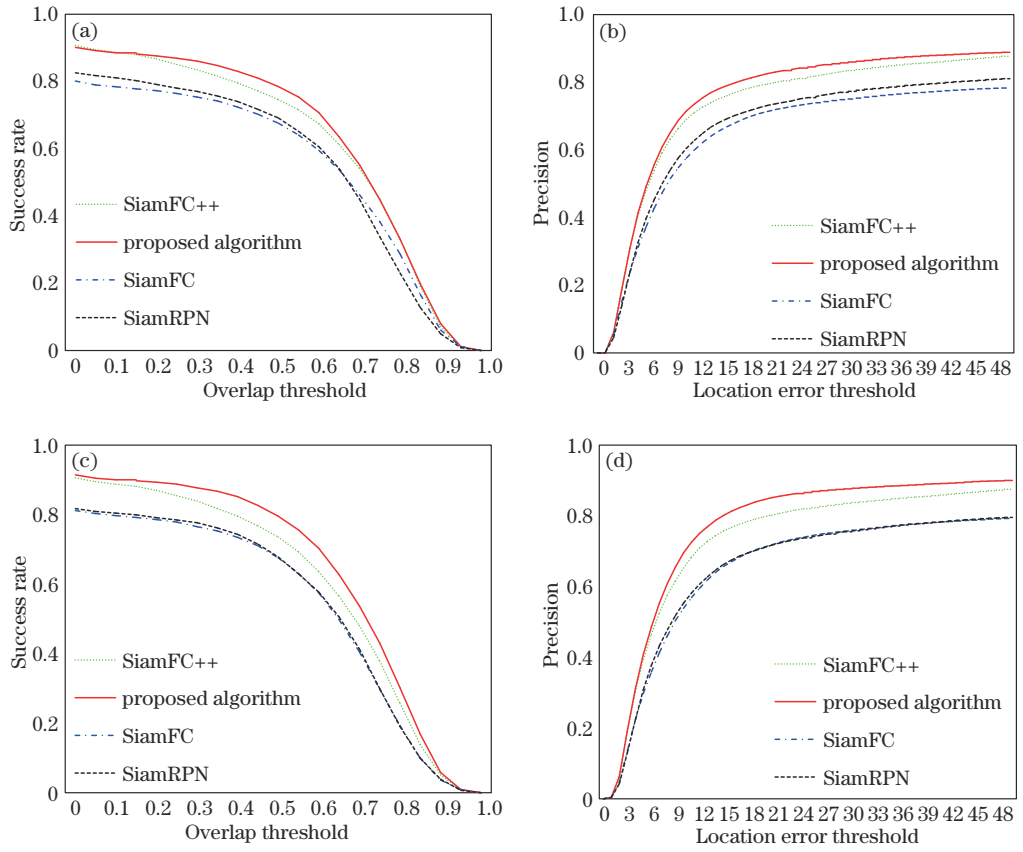

图 8　跟踪结果分析。(a)遮挡和形变-成功率;(b)遮挡和形变-精确度;(c)运动模糊和背景干扰-成功率;(d)运动模糊和背景干扰-精确度;

Fig. 8　Analysis of tracking results. (a) Occlusion and deformation-success rate; (b) occlusion and deformation -precision; (c) motion blur and background clutters-success rate; (d) motion blur and background clutters-precision

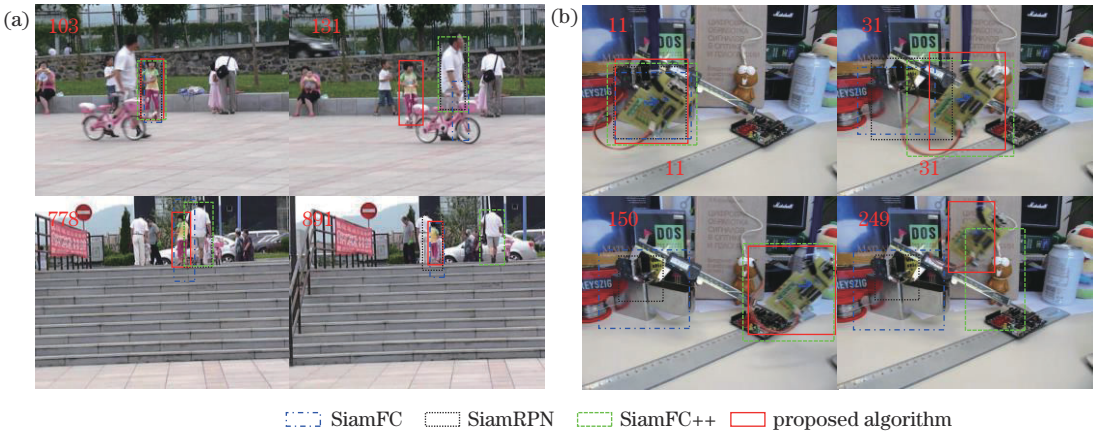

图 9　跟踪结果可视化 (a) Girl2; (b) Board

Fig. 9　Visualization of tracking results. (a) Girl2; (b) Board

4　结　　论

针对当前运动目标跟踪领域存在的跟踪受遮挡或背景干扰影响导致跟踪精确度下降甚至目标丢失的问题,以孪生卷积网络为基础,构建结合非局部感知网络和回归特征增强分类网络的免锚框孪生卷积网络跟踪器。非局部感知网络对跟踪特征进行的增强可以提高特征图的有效性,回归特征增强分类网络可以提高跟踪器的定位能力。在目标跟踪标准数据集上进行的实验结果表明,所提算法能够实现较好的跟踪效果,在形变、背景杂波、遮挡等影响因素下的跟踪准确度和成功率均获得提高。下一步工作将对实现超高的跟踪精确度进行深入研究。

参　考　文　献

[1]　许克应, 束平, 鲍华. 结合注意力与特征融合网络调制

的跟踪算法[J]. 激光与光电子学进展, 2022, 59(12): 1210013.
Xu K Y, Shu P, Bao H. Visual tracking combining attention and feature fusion network modulation[J]. Laser & Optoelectronics Progress, 2022, 59(12): 1210013.

[2] Xie G R, Qu Y, Jiang R Q. Tracking algorithms based on antiocclusion object models[J]. Laser & Optoelectronics Progress, 2022, 59(8): 0815001.
谢郭蓉, 曲毅, 蒋镕圻. 基于抗遮挡目标模型的跟踪算法综述[J]. 激光与光电子学进展, 2022, 59(8): 0815001.

[3] Danelljan M, Khan F S, Felsberg M, et al. Adaptive color attributes for real-time visual tracking[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 1090-1097.

[4] 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.

[5] Girshick R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision, December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1440-1448.

[6] Li B, Yan J J, Wu W, et al. High performance visual tracking with Siamese region proposal network[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: 8971-8980.

[7] Xu Y D, Wang Z Y, Li Z X, et al. SiamFC++: towards robust and accurate visual tracking with target estimation guidelines[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12549-12556.

[8] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[9] 刘宗达, 董立泉, 赵跃进, 等. 视频中快速运动目标的自适应模型跟踪算法[J]. 光学学报, 2021, 41(18): 1815001.
Liu Z D, Dong L Q, Zhao Y J, et al. Adaptive model tracking algorithm for fast-moving targets in video[J]. Acta Optica Sinica, 2021, 41(18): 1815001.

[10] 陈志旺, 张忠新, 宋娟, 等. 基于目标感知特征筛选的孪生网络跟踪算法[J]. 光学学报, 2020, 40(9): 0915003.
Chen Z W, Zhang Z X, Song J, et al. Tracking algorithm for Siamese network based on target-aware feature selection[J]. Acta Optica Sinica, 2020, 40(9): 0915003.

[11] Zhang Z Z, Lan C L, Zeng W J, et al. Relation-aware global attention for person re-identification[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 3183-3192.

[12] Li X, Wang W H, Hu X L, et al. Generalized focal loss V2: learning reliable localization quality estimation for dense object detection[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 11627-11636.

[13] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2999-3007.

[14] Tian Z, Shen C H, Chen H, et al. FCOS: fully convolutional one-stage object detection[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 9626-9635.

[15] Huang L H, Zhao X, Huang K Q. GOT-10k: a large high-diversity benchmark for generic object tracking in the wild[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(5): 1562-1577.

[16] Müller M, Bibi A, Giancola S, et al. TrackingNet: a large-scale dataset and benchmark for object tracking in the wild[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11205: 310-327.

[17] Mueller M, Smith N, Ghanem B. A benchmark and simulator for UAV tracking[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 445-461.