计算机科学与探索
Journal of Frontiers of Computer Science and Technology
1673-9418/2022/16(07)-1504-12
doi: 10.3778/j.issn.1673-9418.2111105

视频目标跟踪算法综述

刘　艺[1+],李蒙蒙[1],郑奇斌[2],秦　伟[1],任小广[1]
1. 国防科技创新研究院,北京 100071
2. 军事科学院,北京 100091
+ 通信作者 E-mail: albertliu20th@163.com

摘　要:视频目标跟踪是计算机视觉领域重要的研究内容,主要研究在视频流或者图像序列中定位其中感兴趣的物体。视频目标跟踪在视频监控、无人驾驶、精确制导等领域中具有广泛的应用,因此,全面地综述视频目标跟踪算法具有重要的意义。首先根据挑战来源不同,将视频目标跟踪技术面临的挑战分为目标自身因素和背景因素两方面,并分别进行总结;其次将近些年典型的视频目标跟踪算法分为基于相关滤波的视频目标跟踪算法和基于深度学习的视频目标跟踪算法,并进一步将基于相关滤波的视频目标跟踪算法分为核相关滤波算法、尺度自适应相关滤波算法和多特征融合相关滤波算法三类,将基于深度学习的视频目标跟踪算法分为基于孪生网络的视频目标跟踪算法和基于卷积神经网络的视频目标跟踪算法两类,并对各类算法从研究动机、算法思想、优缺点等方面进行分析;然后介绍了视频目标跟踪算法中常用的数据集和评价指标;最后总结了全文,并指出视频目标跟踪领域未来的发展趋势。

关键词:计算机视觉;视频目标跟踪;相关滤波;深度学习
文献标志码:A　　**中图分类号:**TP391.4

Survey on Video Object Tracking Algorithms

LIU Yi[1+], LI Mengmeng[1], ZHENG Qibin[2], QIN Wei[1], REN Xiaoguang[1]
1. Defense Innovation Institute, Beijing 100071, China
2. Academy of Military Science, Beijing 100091, China

Abstract: Video object tracking is an important research content in the field of computer vision, mainly studying the tracking of objects with interest in video streams or image sequences. Video object tracking has been widely used in cameras and surveillance, driverless, precision guidance and other fields. Therefore, a comprehensive review on video object tracking algorithms is of great significance. Firstly, according to different sources of challenges, the challenges faced by video object tracking are classified into two aspects, the objects' factors and the backgrounds' factors, and summed up respectively. Secondly, the typical video object tracking algorithms in recent years are classified into correlation filtering video object tracking algorithms and deep learning video object tracking algorithms. And further the correlation filtering video object tracking algorithms are classified into three categories: kernel correlation filtering algorithms, scale adaptive correlation filtering algorithms and multi-feature fusion correlation filtering algorithms. The deep learning video object tracking algorithms are classified into two categories: video object tracking algorithms based on siamese network and based on convolutional neural network. This paper analyzes various algorithms from the aspects of research motivation, algorithm ideas, advantages and disadvantages.

基金项目:国家自然科学基金青年基金项目(61802426)。
This work was supported by the National Natural Science Foundation for Young Scientists of China (61802426).
收稿日期:2021-11-22　　**修回日期:**2022-01-20

Then, the widely used datasets and evaluation indicators are introduced. Finally, this paper sums up the research and looks forward to the development trends of video object tracking in the future.

Key words: computer vision; video object tracking; correlation filtering; deep learning

视频目标跟踪是计算机视觉领域的重要问题，指利用视频或图像序列的上下文信息，对目标的外观和运动信息进行建模，从而对目标运动状态进行预测并标定位置的技术[1]。视频目标跟踪在视频监控[2-3]、无人驾驶[4]等实际环境中有着广泛的应用。尽管近年来关于视频目标跟踪算法的研究取得了很大的进展，但是由于跟踪目标的外观变化、尺寸变化、物体遮挡、运动模糊、跟踪背景干扰等因素的影响，现有方法的效果仍未达到理想状态。根据是否涉及背景环境，可以将视频目标跟踪面临的挑战分为目标自身因素和背景因素两方面。目标自身变化带来的挑战主要有外形变化、尺度变化、运动模糊和目标旋转等；除了目标自身变化带来的挑战，背景因素的影响也较为显著，主要包括遮挡与消失、光照变化和相似背景干扰等。具体分类如图1所示。

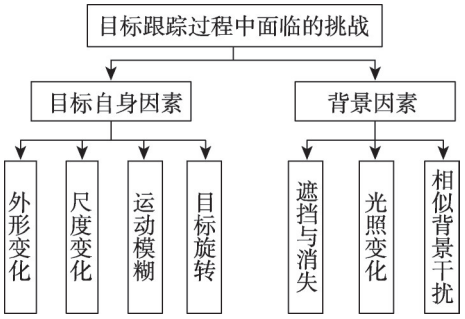


图1　视频目标跟踪面临的挑战

Fig.1　Challenges of video object tracking

按照跟踪方法的不同，本文将视频目标跟踪算法分为基于相关滤波的视频目标跟踪算法和基于深度学习的视频目标跟踪算法。基于相关滤波的视频目标跟踪算法在跟踪的过程中主要利用相关滤波器计算模板图像和预测图像之间的相似度来确定目标位置；而基于深度学习的视频目标跟踪算法主要是通过训练深层网络来学习目标特征，完成视频目标跟踪。相关滤波方法最早源自信号领域，旨在通过卷积操作判断两个信号的相似程度。文献[5]首次将相关滤波引入到视频目标跟踪问题中，提出了误差平方和最小滤波器（minimum output sum of squared error，MOSSE）算法，之后基于相关滤波的算法逐渐成为了视频目标跟踪领域的主流方法。但是，随着AlexNet网络[6]的提出，基于深度学习的视频目标跟踪算法逐渐兴起，近几年受到了广泛关注，已经产生了很多性能优异的算法模型。

本文对视频目标跟踪算法的研究做详细的总结，为从事视频目标跟踪方面研究的学者了解相关领域的进展提供参考。首先从核相关滤波算法、尺度自适应相关滤波算法和多特征融合相关滤波算法三方面描述了基于相关滤波的视频目标跟踪算法，从基于孪生网络的视频目标跟踪算法和基于卷积神经网络的视频目标跟踪算法两个角度总结了近几年基于深度学习的视频目标跟踪算法，然后概述了视频目标跟踪领域常用的数据集和评价指标，最后总结了全文并探讨了该领域未来的发展趋势。

1　基于相关滤波的视频目标跟踪算法

1.1　算法原理

MOSSE算法[5]是最典型的基于相关滤波的视频目标跟踪算法，其主要思想是将视频目标跟踪问题抽象为模板图像与候选区域的相似度匹配问题。该类算法在跟踪过程中首先训练一个滤波器，然后利用该滤波器对候选区域的特征做卷积操作，输出响应值，响应最大值在候选区域中对应的位置即为跟踪目标下一帧所在位置。相关滤波算法的流程如下所示。

（1）用边界框在视频第一帧中标定跟踪目标，生成模板图像 z；

（2）提取模板图像 z 的特征图 F_i；

（3）利用高斯函数生成特征图 F_i 的输出响应图 G_i；

（4）根据公式 $H_i^* = \frac{G_i}{F_i}$ 得到初始化的滤波器；

（5）后续图像特征图经过傅里叶变换之后与相关滤波器相乘，结果进行逆傅里叶变换，生成输出响应图，获得当前帧跟踪目标的位置；

（6）利用当前帧的目标位置训练更新相关滤波器，用于之后的预测。

1.2　核相关滤波算法

MOSSE算法虽然具有669 frame/s的实时速度，且针对亮度、尺寸和形状等不严格的变形具有很好的鲁棒性，但是也具有准确度不高等缺陷。针对此问题，研究学者提出了一系列基于MOSSE算法的优

化策略与改进算法，其中一类重要的改进算法是基于核函数的相关滤波算法。

文献[7]针对MOSSE算法中训练样本数量较少，容易产生过拟合的问题提出了CSK（circulant structure with kernels）算法，该算法在MOSSE算法的基础上增加了正则化项，采用循环矩阵进行稠密采样，有效解决了滤波器过拟合的问题；此外，该算法还引入了核技巧，提高了在高维空间中的分类速度。文献[8]在CSK算法的基础上提出了KCF/DCF（kernelized/dual correlation filter）算法，该算法把视频目标跟踪问题抽象为目标检测问题。首先利用岭回归方法训练了一个目标检测器，然后利用训练的目标检测器预测候选位置是否为下一帧目标的位置。此外，该算法利用循环矩阵在傅里叶空间对角化的性质将矩阵运算转化为向量的Hadamad运算（即元素的点乘），提高了算法的运算速度；同时引入了高斯核函数，将低维空间中的线性不可分问题转化为高维空间中的线性可分问题。KCF/DCF算法进一步引进了基于多通道的方向梯度直方图（histogram of oriented gradient，HOG）[9]进行特征提取，进一步提升了算法的跟踪精度。KCF/DCF算法虽然在跟踪速度和跟踪精度两方面都有了很大改进，但是其对尺度变化较大的视频目标跟踪效果不太理想，主要是由于其仅采用单一尺度的候选图像。针对此问题，相关学者提出多尺度缩放策略用于解决尺度变化较大的视频目标跟踪问题[10-12]。

近几年，一些核相关滤波视频目标跟踪算法也相继被提出。文献[13]针对KCF/DCF算法在目标遮挡和尺度变化问题中的局限性提出了一个基于核相关滤波的鲁棒跟踪算法。该算法针对目标遮挡问题提出了损失辨别和重定位策略，根据当前帧和第一帧的相似度判定目标是否被遮挡，当目标被遮挡时，算法在一定范围内重新定位目标；此外，该算法针对尺度变化问题引入了多尺度滤波器以缓解目标漂移问题。文献[14]针对卫星数据中目标较小且目标与背景相似的问题提出了一个混合核相关滤波算法（hybrid kernel correlation filter，HKCF）。该算法利用光流和方向梯度直方图两个互补的特征进行自适应融合以检测目标变化。文献[15]为了提高视频目标跟踪算法在遇到背景杂波、遮挡等问题时的鲁棒性，提出了一种基于自适应更新策略和再检测技术的关联跟踪算法。该算法的自适应更新策略根据跟踪结果的置信度自适应调整模板更新系数，当目标遭受遮挡时，利用再检测策略对目标进行重新检测，不仅降低了目标漂移概率，还提高了算法的纠错能力。

1.3 尺度自适应相关滤波算法

在跟踪过程中由于目标与相机的距离经常发生变化导致跟踪目标尺度不一。为适应目标尺度缩放的问题，文献[10]提出的SAMF（scale adaptive with multiple features tracker）算法提出了尺度池策略，其主要思想是对候选区域的目标做七个尺度的缩放，再与上一帧样本进行匹配，选择相似度最高的候选区域作为最终的跟踪目标。尺度池策略的引入使得算法能够在小范围内实现尺度自适应，提高了跟踪精度。文献[11]提出的DSST（discriminative scale space tracker）算法将视频目标跟踪看成平移跟踪和尺度跟踪两个问题，算法除了训练平移滤波器之外，还训练了尺度滤波器以解决目标尺度变化的问题。尺度滤波器以目标位置为中心进行空间位置采样，距离原始目标越近抽样越精细，获得33个不同尺度的样本。文献[12]在DSST算法的基础上提出了一种鲁棒的旋转估计算法。该算法基于等角度间隔策略在目标中心区域进行采样，并训练角度滤波器，结合DSST算法中的平移滤波器和尺度滤波器形成了一个由三层滤波器组成的跟踪器，实现了对旋转目标的精确跟踪。

结合卷积神经网络深层特征的相关滤波跟踪算法具有较好的跟踪性能，但是无残差的卷积神经网络深层特征缺乏目标局部信息，容易受到相似物和背景噪声的影响。针对此问题，文献[16]提出尺度自适应的视频目标跟踪算法。该算法从ResNet网络的不同层提取特征生成响应图，然后基于AdaBoost算法进行融合，再利用尺度滤波器估计目标尺寸，实现准确跟踪。文献[17]提出了一种可变尺度因子学习方法，该方法克服了常用的多尺度搜索方法中固定尺度因子的局限性，其次使用多尺度纵横比方法替换固定尺度纵横比方法进一步缓解目标尺度变化问题。

尺度池策略和尺度自适应方法的提出使得基于相关滤波的视频目标跟踪算法在目标尺度缩放、目标外观变化等挑战下的跟踪精度得到较大提升，但是如何得到一个合适的尺度滤波器对候选图像进行采样仍是该领域面临的难题。

1.4 多特征融合相关滤波算法

文献[18]认为提取合适的特征能够显著提升模型的跟踪效果。因此，通过多特征融合的方法来提高视频目标跟踪精度成为了当下研究的热点。

在早期的视频目标跟踪算法中[19-20]，主要采用颜色直方图或者单通道的灰度特征来辨别目标。该类方法简单高效，但是学习到的目标信息较少，跟踪精度较低。为了提高算法跟踪精度，文献[7-8,11]采用了HOG特征，该特征是在图像的局部方格单元上进行操作，对图像几何变化和光照变化都具有较好的鲁棒性。文献[21]将RGB三通道细化为11种颜色，在跟踪过程中将11维颜色特征降为两维，自适应选择颜色特征。文献[22-24]则根据不同的颜色特征方法进行视频目标跟踪。

自深度学习快速发展以来，基于深度特征的相关滤波跟踪算法得到了广泛的研究和发展。文献[25]将SRDCF（spatially regularized discriminative correlation filters）算法[26]中的传统手工特征替换为基于卷积神经网络的深度特征，提出了deepSRDCF算法，取得了较好的跟踪效果。C-COT（continuous convolution operator tracker）算法[27]结合深度特征和传统的手工特征共同进行跟踪。首先采用深度网络VGGNet[28]进行特征提取，然后将提取的深度特征与HOG和颜色直方图等手工特征进行融合实现视频目标跟踪，深浅层特征的融合显著提升了算法的跟踪精度。

采用深层特征的视频目标跟踪算法虽然在性能上得到了显著的提升，但是跟踪速度却明显地下降。针对此问题，ECO（efficient convolution operators）算法[29]深入分析了影响算法速度的三个主要原因：模型复杂度、训练集尺寸和模型更新策略。并针对不同的原因提出了相应的解决方案：（1）跟踪过程中仅选择贡献较大的滤波器进行线性组合，减少模型参数，实现快速跟踪；（2）去除冗余样本，简化训练集；（3）提出间隔N_s帧更新一次模型，提升算法的实时性。针对当前的跟踪算法仅使用深度网络中浅层特征的问题，UPDT（unveiling the power of deep tracking）算法[30]系统地阐述了深层和浅层特征对视频目标跟踪的影响，并指出深层特征能提升网络的鲁棒性，浅层特征能获得更好的定位精度，提出了一种深浅层特征自适应融合的跟踪算法。深层和浅层特征的优缺点如表1所示。文献[31]针对基于深度互相关操作的视频目标跟踪算法容易被相似物干扰且对目标边界的辨别能力较弱等问题提出了一种可学习模块，称为不对称卷积模型（asymmetric convolution module，ACM）。ACM可以在大规模数据的离线训练中学习如何更好地捕捉语义相关信息，有效地融合目标和搜索区域中不同尺寸的特征图，结合先验信息和视觉特征，可以很容易地集成到现有跟踪器中，具有较好的泛化性能。

表1　深层特征与浅层特征的对比

Table 1　Comparison of deep and shallow features

特征	优点	缺点
深层特征	包含高层语义信息，对目标外观变化具有不变性，鲁棒性较强	空间分辨率较低，无法精确定位，容易导致目标漂移，准确性较弱
浅层特征	空间分辨率高，适合高精度定位，准确性较高	目标跟踪的鲁棒性较弱

多特征融合算法的提出显著提升了基于相关滤波视频目标跟踪算法的跟踪精度和鲁棒性，尤其是传统手工特征和深层特征的融合，使得在运动模糊、目标旋转等复杂情况下视频目标跟踪算法的鲁棒性也能得到较大提升。

1.5　算法对比

基于相关滤波的算法是视觉目标跟踪领域中发展较为成熟的一类算法，具有速度快、精度高等优点，但是该类算法通常采用手工浅层特征，因此鲁棒性较差。现对典型的基于相关滤波的视频目标跟踪算法进行简单对比，如表2所示。

2　基于深度学习的视频目标跟踪算法

2.1　基于孪生网络的视频目标跟踪算法

基于孪生网络的视频目标跟踪算法自提出以来得到了研究学者们的广泛关注。孪生网络架构如图2所示，输入1和输入2分别代表模板图像和搜索区域图像，经过两个结构相同、参数共享的子网络之后生成相应的特征图，然后通过计算生成两个图像的相似度。由于孪生网络可以进行离线训练，可以使用大规模的图像数据集进行预训练，很好地缓解了视频目标跟踪领域中训练样本数量较少的问题。

2.1.1　算法原理

SiamFC（fully-convolutional siamese networks）算法[32]是由Bertinetto等人提出的全卷积孪生网络。它首次将孪生网络引入到视频目标跟踪领域，把视频目标跟踪问题转化为图像匹配问题，通过选择与模板图像最相似的候选图像实现对目标的跟踪。

SiamFC网络的两个输入分别为模板图像z和搜索区域x。其中模板图像通常是视频第一帧选定的跟踪目标，跟踪期间模板图像不进行更新；搜索区域一般以上一帧目标所在位置为中心选出固定尺寸大小的区域。在跟踪过程时，算法对目标图像进行多

表2 基于相关滤波的视频目标跟踪算法

Table 2 Video object tracking algorithms based on correlation filter

类型	文献	算法名称	特点	优点	缺点
相关滤波	[5]	MOSSE	将相关滤波引入到视频目标跟踪领域,用滤波器与候选区域的特征图做卷积操作,响应最大值所在位置即为当前帧跟踪目标所在位置	速度快,可达669 frame/s	精度较低,单通道灰度特征
核相关	[7]	CSK	增加了正则化项,有效地防止了滤波器的过拟合;采用循环矩阵的方法进行稠密采样;引入了核技巧,提高了算法在高维空间中的速度	速度快,计算量有所减少	单一尺度,单通道灰度特征
	[8]	KCF/DCF	训练了一个目标检测器,判断预测位置是否为目标位置;引进了基于多通道的HOG特征	速度快,可达172 frame/s;多通道HOG特征,精度显著提升	单一尺度
	[13]	鲁棒跟踪算法	将灰度特征、HOG特征、LAB颜色特征进行融合;提出损失辨别和重定位策略缓解目标遮挡问题;采用多尺度滤波器缓解目标漂移的问题	中心位置误差较低	仅采用手工特征,未结合深度特征
	[14]	HKCF	针对卫星数据进行研究,有效缓解了目标较小且与背景相似的问题	特征融合,速度快,可达100 frame/s	仅采用手工特征,未结合深度特征
多尺度跟踪	[11]	DSST	将视频目标跟踪看作目标中心平移和目标尺度变化两个独立的问题,训练了两个滤波器:平移滤波器和尺度滤波器	33个尺度,多尺度跟踪,精度较高	速度较慢25.4 frame/s,边界效应
	[10]	SAMF	HOG特征、颜色特征和灰度特征融合;提出尺度池策略,小范围内实现了尺度自适应跟踪	HOG、颜色、灰度特征融合,7个尺度跟踪,提高精度	仅在尺度池内效果较好,没有做到真正意义的自适应
	[16]	尺度自适应算法	从ResNet网络的不同层提取特征生成响应图,然后基于AdaBoost算法进行融合,再利用尺度滤波器估计目标尺寸,实现准确跟踪	多特征融合,尺度滤波器	速度较慢;未采用手工特征,鲁棒性较差
	[17]	可变尺度学习跟踪算法	尺度因子可学习,不断调整;多尺度跟踪框纵横比方法共同缓解目标尺度变化问题	针对尺度变化问题效果较好	未进行特征融合
多特征融合	[27]	C-COT	将深度特征和手工特征(HOG特征和颜色特征)进行融合	13个滤波器,跟踪精度较高	速度较慢1.5 frame/s,算法参数较多
	[30]	UPDT	系统地分析了深层和浅层特征在视频目标跟踪中的影响,提出一种深层和浅层特征自适应融合的跟踪算法	精度较高	虽然速度有所提升,但仍较慢
	[31]	ACM	融合目标和搜索区域中不同尺寸的特征图,结合先验信息和视觉特征,可以容易地集成到现有跟踪器中	泛化性能较好,可直接集成到其他跟踪器中	跟踪效果与选用的跟踪器关系较大

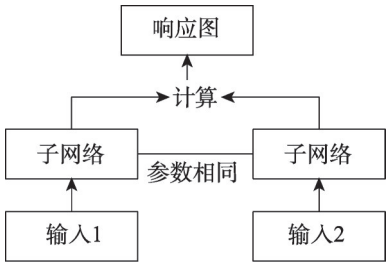


图2 孪生网络架构

Fig.2 Architecture of siamese network

种尺度缩放,并以不同尺寸的滑动窗口在整个搜索区域进行滑动匹配。两个分支骨干网 φ 的结构相同,参数共享,骨干网对两个输入进行相同的变换后,将提取的特征图送入到相似性度量函数 g 中,利用式(1)得到相似度。

$$f(z,x)=g(\varphi(z),\varphi(x)) \tag{1}$$

其中,g 一般为卷积操作,$\varphi(z)$ 为卷积核。

2.1.2 孪生网络的改进

SiamFC算法虽然具有实时的跟踪速度,但是跟踪精度并不理想,主要原因在于SiamFC算法并不能较好地处理目标尺度变化问题。因此,为了更精确地跟踪目标,文献[33]在SiamFC算法的基础上提出了SiamRPN(siamese region proposal network)算法。SiamRPN算法引入了候选区域生成网络(region proposal network,RPN)模块。该模块取代了传统的多尺度检测方法,实现了高精度跟踪。RPN网络架构有两个分支:分类分支和回归分支。分类分支用于区分目标和背景,实现对目标的检测分类;回归分支用于对目标边界框回归预测,实现对目标的精确定

位。SiamMask算法[34]把视频目标跟踪与实例分割结合起来，在SiamRPN算法的基础上对RPN模块进行了扩展，增加了目标二值掩码分支，该分支通过一个两层神经网络得到目标的实时像素级标注信息，进一步完成目标的精确定位。C-RPN（siamese cascaded region proposal networks）算法[35]利用特征转换模块融合多层特征，并将融合后的特征图输入到不同的RPN模块中，采用多RPN模块级联的方式进行候选区域选择，该方法不仅可以充分利用深层特征和浅层特征，还可以精确地计算目标边界框、定位目标。SPM-Tracker（series-parallel matching tracker）算法[36]将视频目标跟踪分为两个阶段：粗略匹配阶段和精细匹配阶段。粗略匹配阶段采用SiamRPN网络，分离目标和相似干扰物；精细匹配阶段通过两层全连接网络分类相似物体，选出真正的目标。文献[37]在SiamFC的基础上提出了一种融合注意力机制的孪生网络视频目标跟踪算法。该算法通过融合注意力机制，由神经网络学习模板图像的通道相关性和空间相关性，增大前景贡献，抑制背景特征，提升网络对目标特征的辨别力。文献[38]针对SiamFC在严重遮挡、旋转、光照变化和尺度变化等情况下容易造成跟踪失败的问题，提出了一种融合扰动感知模型的孪生神经网络视频目标跟踪算法。该算法将孪生网络提取的浅层特征和深层语义特征进行有效融合，提高了特征表征能力；此外，该算法引入了颜色直方图特征的扰动感知模型，通过加权融合的方式获得目标响应图，以此来实现目标跟踪。

为了进一步提升算法的跟踪速度，文献[39]提出了一种目标感知模块，并将其与SiamFC框架结合。目标感知模块选择当前跟踪目标所需通道，同时去除非必要的通道，提高了跟踪速度。为了降低目标漂移概率，DaSiamRPN（distractor-aware siamese region proposal network）算法[40]提出了干扰物感知模型，降低了目标漂移到相似干扰物的概率；同时提出了局部-全局搜索策略，当目标跟踪失败时，以目标消失位置为中心重新检测目标，实现长时跟踪。DSiam（dynamic siamese network）算法[41-42]提出了动态孪生网络，在SiamFC架构上增加了形变学习层，利用视频前几帧图像学习目标外观变化，抑制背景噪声干扰，提升长时跟踪精度。DCFNet（discriminant correlation filters network）算法[43]提出了一种轻量级的端到端网络架构。首先利用预训练的卷积网络进行特征提取，然后利用相关滤波器层进行目标匹配和跟踪。为了降低跟踪成本，该架构将卷积层设置成轻量级的网络，将相关滤波器层的计算转换到傅里叶频域中进行。在测试阶段，DCFNet的跟踪速度可达60 frame/s，实现了实时跟踪。CFNet（correlation filter network）算法[44]提出一个非对称的网络架构，首先“训练图像”和“测试图像”经过相同的卷积层进行特征变换，然后“训练图像”通过相关滤波操作学习线性模板，通过互相关操作对“测试图像”进行搜索，最后通过实验证明两层的CFNet可以较好地平衡跟踪精度和跟踪速度，在跟踪精度较理想的情况下，跟踪速度可以达到75 frame/s。文献[45]针对跟踪目标在平面内旋转的问题展开研究，提出了旋转等变孪生网络（rotation-equivariant siamese networks，RE-SiamNets），以无监督的方式估计目标旋转方向变化，促进算法跟踪性能。文献[46]针对现有视频目标跟踪算法目标边界框精度不高，且算法耦合严重、各阶段难以移植的问题提出了一种精确通用的跟踪模块，称为AR（alpha-refine）。该模块以孪生网络架构为基础，两个分支分别对模板图像和测试图像进行特征提取，然后利用相关模块进行特征融合，在跟踪过程中将边界框设定为目标尺寸的2倍。比普通跟踪器更小的边界框可以让跟踪器更关注目标空间信息，有利于精确定位。该模块轻量级的设计降低了跟踪成本，但具体效能仍与完整的跟踪框架相关。

2.1.3　孪生子网络的改进

跟踪算法中使用的孪生子网络相对较浅，目标信息利用率不高，若直接将子网络替换为深度网络，算法性能也并不能得到显著提升，这主要是由于深度网络一般都具有填充操作，但是在跟踪过程中填充操作将引入目标位置偏差，影响跟踪效果。针对此问题，文献[47]提出了内部裁剪残差单元来增强SiamRPN的性能，该方法删除了受填充操作影响的特征，把深度网络有效地应用到了视频目标跟踪领域。文献[48]提出了具有空间感知采样策略的SiamRPN++算法，该策略较好地解决了填充操作带来的负面影响，同时采用了多SiamRPN级联策略，增强了算法的判别能力。

2.1.4　双重孪生网络

为了更加充分地利用孪生网络对称性的优势，SA-Siam（semantic features and appearance features siamese network）算法[49]提出了基于双重孪生网络的视频目标跟踪算法。该算法由语义分支对和外观分支对组成，语义分支对生成的语义特征用于图像分类，外观

分支对生成的外观特征用于相似度匹配。SiamBM(better match in siamese network)算法[50]在SA-Siam算法的基础上添加了旋转角度模块来预测目标的旋转。SA-Siam++算法[51]提出了一种基于语义和外观双分支孪生网络的跟踪方法,双分支网络分别是通过沙漏-通道注意力机制提取语义信息的语义分支网络和采用SiamFC算法提取外观特征的外观分支网路,提高了算法的跟踪性能。

近几年,基于孪生网络的算法由于其思想简单,架构可伸缩,在视频目标跟踪领域取得了快速的发展。该类算法在跟踪速度较为可观的情况下,有效提升了跟踪器对背景和相似干扰物的辨别能力,是视频目标跟踪领域未来研究的重点。

2.2 基于卷积神经网络的视频目标跟踪算法

2.2.1 针对卷积神经网络的改进

文献[52]提出了一种基于深度学习的支持向量机算法(deep learning support vector machines,DLSVM),该算法利用图像识别领域预训练的卷积神经网络(convolutional neural network,CNN)对目标进行特征提取,利用支持向量机分类器进行跟踪。该算法由于不需要大量训练样本学习网络模型,因此在一定程度上提高了算法的执行效率,但是直接将图像识别领域的神经网络应用到视频目标跟踪领域中并不能达到预期效果,主要原因在于图像分类关注类间差异,忽视类内区别,而视频目标跟踪关注目标实例与背景间的差异,同时需要消除同类别其他实例物体的干扰。针对此问题,文献[53]提出了针对视频目标跟踪的多域卷积神经网络(multi-domain network,MDNet)。MDNet网络最终生成一个二维向量分别表示该边界框中物体为目标或者背景的概率,该网络架构较小,参数较少,具有很好的实时性。文献[54]提出了一种基于树状结构的CNN跟踪算法。该算法在树状结构的不同分支中维护多个CNN网络,并对CNN网络生成的结果进行加权平均来估计目标外观的变化。文献[55]针对MDNet算法采样密集的问题提出了行为驱动策略,通过捕获目标的运动信息,搜索高质量的候选样本,提高算法泛化性能。

2.2.2 卷积神经网络与其他网络的结合

SANet(structure-aware network)算法[56]结合了CNN和循环神经网络(recurrent neural network,RNN),其中CNN负责类间判别,区分目标类与背景,RNN负责类内选择,区分目标实例与相似干扰物。Siam R-CNN(siamese R-CNN)算法[57]结合了孪生网络和Faster R-CNN,孪生网络用于特征提取,Faster R-CNN用于候选区域生成。ATOM(accurate tracking by overlap maximization)算法[58]结合了两层深度回归网络和IoUNet网络(intersection-over-union network)[59],前者用于目标粗略定位,生成候选区域,后者用于目标精细定位和尺度估计。文献[60]针对现有跟踪器网络架构越来越庞大、跟踪代价越来越高昂,在资源有限的应用中部署越来越受限的问题,提出一种轻量级的神经网络跟踪架构(LightTrack)。该架构使用神经网络架构搜索(neural architecture search)方法自动设计轻量级模型,首先将所有可能的架构编码为骨干超网络和头部超网络,骨干超网络在ImageNet数据集上进行训练,然后利用测试数据进行微调,而头部超网络则直接使用测试数据进行训练。所有超网络只训练一次,然后每个候选架构直接从超网络中继承权重。此外,该算法构建了新的搜索空间,促使算法搜索更紧凑的神经架构。该算法在跟踪性能和计算成本之间取得了较好的平衡。

基于卷积神经网络的视频目标跟踪算法虽然可以利用不同的网络架构提取不同深度的目标特征,但是由于网络架构庞大,模型参数较多,该类算法的跟踪代价通常较大,因此轻量级的跟踪模型具有十分重要的研究意义。

3 视频目标跟踪常用数据集和评价指标

3.1 视频目标跟踪常用数据集

随着对视频目标跟踪算法的深入研究,传统的数据集已经不能有效地评估算法的综合性能,为此,研究人员提出了更多高质量的数据集。这些数据集除了能够有效评估算法的性能,也进一步推动了视频目标跟踪领域的发展。下面总结该领域常用的数据集及其特点。表3按照时间线给出了常用视频目标跟踪数据集的信息,其中数据集包括:OTB-2013[61]、OTB-2015[62]、VOT2013[63]、VOT2014[64]、VOT2015[65]、V0T2016[66]、VOT2017[67]、VOT2018[68]、VOT2019[69]、UAV123[70]、UAV20L[70]、TrackingNet[71]、GOT-10K[72]、LaSOT[73]。

3.2 视频目标跟踪评价指标

随着数据集的不断更新,更加准确高效的评价指标也在不断完善,优异的评价指标可以更加公平客观地反映算法的优劣。在视频目标跟踪算法中最常用的评价指标为精确度、交并比、成功率和跟踪速度等。

精确度(precision plot)主要评估的是目标中心位置误差,指跟踪目标中心位置与目标真值中心位置

表3　视频目标跟踪领域常用数据集

Table 3　Datasets widely used in field of video object tracking

数据集	年份	视频数	帧数	平均长度/帧	类别	特点
OTB-2013	2013	51	29 000	578	10	包含25%的灰度序列；11种常见的视频属性标注：光照变化、尺度变化、遮挡、形变、运动模糊、快速移动、平面内旋转、平面外旋转、消失、相似背景干扰、低分辨率；随机帧开始
OTB-2015	2015	98	59 000	598	16	在OTB-2013的基础上增加了视频序列
VOT	2013	16	—	—	—	为彩色序列，平均时长较短，分辨率较高；第一帧初始化开始；VOT2018和VOT2019均在VOT2017的基础上加入了长时跟踪视频序列
	2014	25	10 000	409	11	
	2015	60	22 000	358	24	
	2016	60	22 000	358	24	
	2017	60	22 000	356	24	
	2018	60	22 000	356	24	
	2019	60	22 000	356	24	
UAV123	2016	123	113 000	915	9	特殊场景数据集，均由低空无人机捕获；视频序列背景干净，视角变化丰富
UAV20L	2016	20	59 000	2 934	5	视频序列平均时长较长，常应用于长时跟踪
TrackingNet	2018	30 643	14 432 000	467	27	规模较大，主要针对野外目标的短时跟踪；训练集和测试集互不相交
GOT-10K	2019	10 000	1 500 000	150	563	数据集种类较多，时长较短，常应用于短时跟踪；训练集和测试集互不相交
LaSOT	2019	1 400	3 520 000	2 506	70	大规模的长时跟踪数据集；提供了可视化的边界框注释，当目标消失时，出现“目标不存在”的注释

之间的平均欧氏距离小于给定阈值的视频帧占整个视频序列帧数的百分比，公式如式(2)所示。

$$p=\frac{f_{l<L}}{f} \tag{2}$$

其中，f 是视频序列长度，l 为跟踪目标与目标真值之间的距离误差，L 为设定的阈值。跟踪精度虽然能直观地反映算法的优劣，但是不能很好地处理目标尺寸发生变化的情况。因此，在OTB数据集中同时采用了成功率指标。

成功率(success plot)主要依据的是交并比，指当某一帧图像的交并比大于规定阈值时，则认为该帧跟踪成功，跟踪成功的帧数占整个视频序列的百分比设置为成功率，公式如式(3)所示。

$$s=\frac{f_{a>A}}{f} \tag{3}$$

其中，a 为某一帧的交并比，A 为设定的阈值。

交并比(intersection over union，IoU)是指“预测图像”与“目标真值图像”之间面积的交集与并集的比值，如式(4)和图3所示。

$$IoU=\frac{|A\bigcap B|}{|A\bigcup B|} \tag{4}$$

其中，A 和 B 分别代表“预测图像”和“目标真值图像”。

成功率指标可以很好地评估目标尺寸变化情况，但是并不能很好地体现跟踪目标与目标真值未

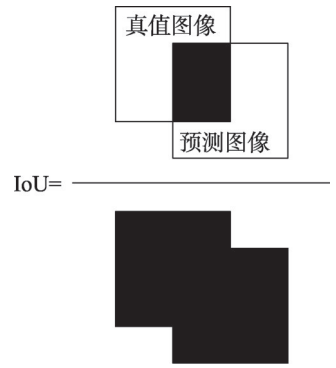


图3　交并比计算图

Fig.3　Calculation graph of IoU

重叠和目标旋转等问题。当跟踪目标与目标真值未重叠时，简单地认为跟踪失败并不能提供有用信息，导致算法无法在下一帧作出有效改进；当目标发生旋转时，某一帧预测图像与目标真值重叠率很高，但是角度相差较大，此时回归效果很差，却被误判为跟踪成功，导致成功率不可信。针对以上问题，文献[74]提出了GIoU(generalized intersection over union)指标，如式(5)所示。

$$GIoU=IoU-\frac{|C\backslash(A\bigcup B)|}{|C|} \tag{5}$$

其中，C 代表包围 A 和 B 的最小面积框。GIoU是指用传统的IoU减去“‘从 C 中去除 A 和 B 的面积框’

与‘C’之间的比值”。该指标通过引入同时覆盖预测图像和目标真值图像的最小封闭面积框，使得两者即使不重叠，预测图像也会逐渐移向目标真值。

视频目标跟踪领域中算法的实时性很重要，通常用跟踪速率（单位为frame/s）来评估算法实时性能。

4 总结与展望

尽管近些年视频目标跟踪技术取得了快速发展，但是在复杂的应用场景中，现有的视频目标跟踪算法仍未取得理想效果。现对视频目标跟踪领域存在的问题进行简单总结与展望，希望可以为相关学者的研究提供思路。

（1）基于相关滤波的视频目标跟踪算法把视频目标跟踪问题抽象为模板图像与候选区域的相似度匹配问题，因此模板图像的选择更新策略对算法的跟踪精度具有显著影响，加大对模板图像的关注和研究具有重要意义。

（2）基于深度学习的视频目标跟踪算法主要依托深度网络架构实现视频目标跟踪，因此网络架构的设计和构建非常重要。网络架构较深会限制算法长时跟踪性能，参数较多会降低算法跟踪效率，耦合性较高会加大算法改进优化难度，因此，设计轻量级的、耦合性较低的视频目标跟踪模块或者算法具有较大现实意义和应用价值。

（3）现有的视频目标跟踪数据集涵盖的类别较多，视频序列场景丰富，虽然可以综合评估算法性能，但是不能针对性地评估应用在某一具体领域的跟踪算法。因此，为了更有效地评估实用可靠的跟踪算法，根据具体应用领域制作相应的数据集和评价指标具有一定的必要性。

参考文献：

[1] 李玺, 查宇飞, 张天柱, 等. 深度学习的目标跟踪算法综述[J]. 中国图象图形学报, 2019, 24(12): 2057-2080.
LI X, ZHA Y F, ZHANG T Z, et al. Survey of visual object tracking algorithms based on deep learning[J]. Journal of Image and Graphics, 2019, 24(12): 2057-2080.

[2] LIANG J W, JIANG L, NIEBLES J C, et al. Peeking into the future: predicting future person activities and locations in videos[C]//Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, Jun 16-20, 2019. Piscataway: IEEE, 2019: 5725-5734.

[3] JANI D, MANKODIA A. Comprehensive analysis of object detection and tracking methodologies from surveillance videos[C]//Proceedings of the 2021 International Conference on Computing Methodologies and Communication, Erode, Apr 8-10, 2021. Piscataway: IEEE, 2021: 963-970.

[4] LI P, CHEN X, SHEN S. Stereo R-CNN based 3D object detection for autonomous driving[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, Jun 16-20, 2019. Piscataway: IEEE, 2019: 7644-7652.

[5] BOLME D S, BEVERIDGE J R, DRAPER B A, et al. Visual object tracking using adaptive correlation filters[C]//Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, Jun 13-18, 2010. Washington: IEEE Computer Society, 2010: 2544-2550.

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

[7] HENRIQUES J F, CASEIRO R, MARTINS P, et al. Exploiting the circulant structure of tracking-by-detection with kernels[C]//LNCS 7575: Proceedings of the 12th European Conference on Computer Vision, Florence, Oct 7-13, 2012. Berlin, Heidelberg: Springer, 2012: 702-715.

[8] HENRIQUES J F, CASEIRO R, MARTINS P, et al. High-speed tracking with kernelized correlation filters[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3): 583-596.

[9] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, Sep 21-23, 2005. Washington: IEEE Computer Society, 2005: 886-893.

[10] BAO F, CAO Y, ZHANG S, et al. Using segmentation with multi-scale selective kernel for visual object tracking[J]. IEEE Signal Processing Letters, 2022, 29: 553-557.

[11] LI H, PU L. Correlation filtering tracking algorithm with joint scale estimation and occlusion processing[C]//Proceedings of the 2021 International Conference on Intelligent Transportation, Big Data & Smart City, Xi’an, Mar 27-28, 2021. Piscataway: IEEE, 2021: 663-667.

[12] FANG Y, JO G S, LEE C H. RSINet: rotation-scale invariant network for online visual tracking[C]//Proceedings of the 2021 International Conference on Pattern Recognition, Milan, Jan 10-15, 2021. Piscataway: IEEE, 2021: 4153-4160.

[13] SHAO J, DU B, WU C, et al. Can we track targets from space? A hybrid kernel correlation filter tracker for satellite video[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(11): 8719-8731.

[14] 马珺, 王昱皓. 结合自适应更新策略和再检测技术的跟踪算法[J]. 计算机工程与应用, 2021, 57(9): 217-224.
MA J, WANG Y H. Object tracking algorithm based on adaptive update strategy and re-detection technology[J]. Computer Engineering and Applications, 2021, 57(9): 217-224.

[15] YUAN Y, CHU J, LENG L, et al. A scale-adaptive object-

tracking algorithm with occlusion detection[J]. Eurasip Journal on Image and Video Processing, 2020(1): 7.

[16] HE X, ZHAO L, CHEN Y C. Variable scale learning for visual object tracking[J]. Journal of Ambient Intelligence and Humanized Computing, 2021. DOI: 10.1007/s12652-021-03469-2.

[17] HAN W, DONG X, KHAN F S, et al. Learning to fuse asymmetric feature maps in siamese trackers[C]//Proceedings of the 2021 IEEE Conference on Computer Vision and Pattern Recognition, Jun 19-25, 2021. Washington: IEEE Computer Society, 2021: 16570-16580.

[18] LI C, LIU X, ZHANG X, et al. Design of UAV single object tracking algorithm based on feature fusion[C]//Proceedings of the 2021 Chinese Control Conference, Shanghai, Jul 26-28, 2021. Piscataway: IEEE, 2021: 3088-3092.

[19] ZHANG K H, ZHANG L, LIU Q S, et al. Fast visual tracking via dense spatio-temporal context learning[C]//LNCS 8693: Proceedings of the 13th European Conference on Computer Vision, Zurich, Sep 6-12, 2014. Cham: Springer, 2014: 127-141.

[20] COMANICIU D, RAMESH V, MEER P. Kernel-based object tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(5): 564-575.

[21] DANELLJAN M, KHAN F S, FELSBERG M, et al. Adaptive color attributes for real-time visual tracking[C]//Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, Jun 23-28, 2014. Washington: IEEE Computer Society, 2014: 1090-1097.

[22] YIN Y K, DU X P, CHU W B, et al. A color histogram based large motion trend fusion algorithm for vehicle tracking[J]. IEEE Access, 2021, 9: 83394-83401.

[23] MORRIS R, MIRZAEI S. Efficient FPGA implementation of parameterized real time color based object tracking[C]//Proceedings of the 2021 IEEE Annual Information Technology, Electronics and Mobile Communication Conference, Vancouver, Oct 27-30, 2021. Piscataway: IEEE, 2021: 102-105.

[24] ZHANG P, ZHAO J, BO C, et al. Jointly modeling motion and appearance cues for robust RGB-T tracking[J]. IEEE Transactions on Image Processing, 2021, 30: 3335-3347.

[25] DANELLJAN M, HAGER G, KHAN F S, et al. Convolutional features for correlation filter based visual tracking[C]//Proceedings of the 2015 International Conference on Computer Vision, Santiago, Dec 7-13, 2015. Washington: IEEE Computer Society, 2015: 621-629.

[26] DANELLJAN M, HAGER G, KHAN F S, et al. Learning spatially regularized correlation filters for visual tracking[C]//Proceedings of the 2015 International Conference on Computer Vision, Santiago, Dec 7-13, 2015. Washington: IEEE Computer Society, 2015: 4310-4318.

[27] DANELLJAN M, ROBINSON A, KHAN F S, et al. Beyond correlation filters: learning continuous convolution operators for visual tracking[C]//LNCS 9909: Proceedings of the 14th European Conference on Computer Vision, Amsterdam, Oct 8-16, 2016. Cham: Springer, 2016: 472-488.

[28] WANG L, GUO S, HUANG W, et al. Places205-VGGNet models for scene recognition[J]. arXiv:1508.01667v1, 2015.

[29] DANELLJAN M, BHAT G, KHAN F S, et al. ECO: efficient convolution operators for tracking[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Jul 21-26, 2017. Washington: IEEE Computer Society, 2017: 6931-6939.

[30] BHAT G, JOHNANDER J, DANELLJAN M, et al. Unveiling the power of deep tracking[C]//LNCS 11206: Proceedings of the 15th European Conference on Computer Vision, Munich, Sep 8-14, 2018. Cham: Springer, 2018: 493-509.

[31] LI D L, LU R T, YANG X G. Object tracking based on kernel correlation filter and multi-feature fusion[C]//Proceedings of the 2019 Chinese Automation Congress, Hangzhou, Nov 23-24, 2019. Piscataway: IEEE, 2019: 4192-4196.

[32] BERTINETTO L, VALMADRE J, HENRIQUES J F, et al. Fully-convolutional siamese networks for object tracking[C]//LNCS 9914: Proceedings of the 14th European Conference on Computer Vision, Amsterdam, Oct 8-16, 2016. Cham: Springer, 2016: 850-865.

[33] LI B, YAN J, WU W, et al. High performance visual tracking with siamese region proposal network[C]//Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Jun 18-22, 2018. Washington: IEEE Computer Society, 2018: 8971-8980.

[34] WANG Q, ZHANG L, BERTINETTO L, et al. Fast online object tracking and segmentation: a unifying approach[C]//Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, Jun 16-20, 2019. Piscataway: IEEE, 2019: 1328-1338.

[35] FAN H, LING H. Siamese cascaded region proposal networks for real-time visual tracking[C]//Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, Jun 16-20, 2019. Piscataway: IEEE, 2019: 7952-7961.

[36] WANG G, LUO C, XIONG Z, et al. SPM-Tracker: series-parallel matching for real-time visual object tracking[C]//Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, Jun 16-20, 2019. Piscataway: IEEE, 2019: 3643-3652.

[37] 王玲, 王家沛, 王鹏, 等. 融合注意力机制的孪生网络目标跟踪算法研究[J]. 计算机工程与应用, 2021, 57(8): 169-174.
WANG L, WANG J P, WANG P, et al. Siamese network tracking algorithms for hierarchical fusion of attention mechanism[J]. Computer Engineering and Applications, 2021, 57(8): 169-174.

[38] 李勇, 杨德东, 韩亚君, 等. 融合扰动感知模型的孪生神经网络目标跟踪[J]. 光学学报, 2020, 40(4): 114-125.
LI Y, YANG D D, HAN Y J, et al. Siamese neural network object tracking with distractor-aware model[J]. Acta Optica

Sinica, 2020, 40(4): 114-125.

[39] LI X, MA C, WU B Y, et al. Target-aware deep tracking[C]// Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, Jun 16-20, 2019. Piscataway: IEEE, 2019: 1369-1378.

[40] ZHU Z, WANG Q, LI B, et al. Distractor-aware siamese networks for visual object tracking[C]//LNCS 11213: Proceedings of the 15th European Conference on Computer Vision, Munich, Sep 8-14, 2018. Cham: Springer, 2018: 103-119.

[41] GUO Q, FENG W, ZHOU C, et al. Learning dynamic siamese network for visual object tracking[C]//Proceedings of the 2017 International Conference on Computer Vision, Venice, Oct 22-29, 2017. Washington: IEEE Computer Society, 2017: 1781-1789.

[42] ABDELPAKEY M H, SHEHATA M S. DP-Siam: dynamic policy siamese network for robust object tracking[J]. IEEE Transactions on Image Processing, 2020, 29: 1479-1492.

[43] WANG Q, GAO J, XING J, et al. DCFNet: discriminant correlation filters network for visual tracking[J]. arXiv: 1704.04057v1, 2017.

[44] VALMADRE J, BERTINETTO L, HENRIQUES J, et al. End-to-end representation learning for correlation filter based tracking[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Jul 21-26, 2017. Washington: IEEE Computer Society, 2017: 5000-5008.

[45] GUPTA D K, ARYA D, GAVVES E. Rotation equivariant siamese networks for tracking[C]//Proceedings of the 2021 IEEE Conference on Computer Vision and Pattern Recognition, Jun 19-25, 2021. Washington: IEEE Computer Society, 2021: 12362-12371.

[46] YAN B, ZHANG X, WANG D, et al. Alpha-refine: boosting tracking performance by precise bounding box estimation[C]//Proceedings of the 2021 IEEE Conference on Computer Vision and Pattern Recognition, Jun 19-25, 2021. Washington: IEEE Computer Society, 2021: 5289-5298.

[47] ZHANG Z, PENG H. Deeper and wider siamese networks for real-time visual tracking[C]//Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, Jun 16-20, 2019. Piscataway: IEEE, 2019: 4591-4600.

[48] LI B, WU W, WANG Q, et al. SiamRPN++: evolution of siamese visual tracking with very deep networks[C]//Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, Jun 16-20, 2019. Piscataway: IEEE, 2019: 4282-4291.

[49] HE A, LUO C, TIAN X, et al. A twofold siamese network for real-time object tracking[C]//Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Jun 18-22, 2018. Washington: IEEE Computer Society, 2018: 4834-4843.

[50] HE A F, LUO C, TIAN X M, et al. Towards a better match in siamese network based visual object tracker[C]//LNCS 11129: Proceedings of the 15th European Conference on Computer Vision, Munich, Sep 8-14, 2018. Cham: Springer, 2018: 132-147.

[51] 田朗, 黄平牧, 吕铁军. SA-Siam++: 基于双分支孪生网络的目标跟踪算法[J]. 北京邮电大学学报, 2019, 42(6): 105-110. TIAN L, HUANG P M, LV T J. SA-Siam++: two-branch siamese network-based object rracking algorithm[J]. Journal of Beijing University of Posts and Telecommunications, 2019, 42(6): 105-110.

[52] TANG Y. Deep learning using linear support vector machines [J]. arXiv:1306.0239v4, 2013.

[53] NAM H, HAN B. Learning multi-domain convolutional neural networks for visual tracking[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, Jun 27-30, 2016. Washington: IEEE Computer Society, 2016: 4293-4302.

[54] NAM H, BAEK M, HAN B. Modeling and propagating CNNs in a tree structure for visual tracking[J]. arXiv:1608.07242v1, 2016.

[55] YUN S, CHOI J, YOO Y, et al. Action-decision networks for visual tracking with deep reinforcement learning[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition，Honolulu, Jul 21-26, 2017. Washington: IEEE Computer Society, 2017: 1349-1358.

[56] FAN H, LING H. SANet: structure-aware network for visual tracking[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Jul 21-26, 2017. Washington: IEEE Computer Society, 2017: 2217-2224.

[57] VOIGTLAENDER P, LUITEN J, TORR P H S, et al. Siam R-CNN: visual tracking by re-detection[C]//Proceedings of the 2020 IEEE Conference on Computer Vision and Pattern Recognition, Seattle, Jun 14-19, 2020. Washington: IEEE Computer Society, 2020: 6577-6587.

[58] DANELLJAN M, BHAT G, KHAN F S, et al. ATOM: accurate tracking by overlap maximization[C]//Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, Jun 16-20, 2019. Piscataway: IEEE, 2019: 4655-4664.

[59] JIANG B, LUO R, MAO J, et al. Acquisition of localization confidence for accurate object detection[C]//LNCS 11218: Proceedings of the 15th European Conference on Computer Vision, Munich, Sep 8-14, 2018. Cham: Springer, 2018: 816-832.

[60] YAN B, PENG H, WU K, et al. LightTrack: finding lightweight neural networks for object tracking via one-shot architecture search[C]//Proceedings of the 2021 IEEE Conference on Computer Vision and Pattern Recognition, Jun 19-25, 2021. Washington: IEEE Computer Society, 2021: 15180-15189.

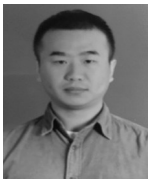
[61] WU Y, LIM J, YANG M H. Online object tracking: a benchmark[C]//Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, Jun 23-28, 2013. Washington: IEEE Computer Society, 2013: 2411-2418.
[62] WU Y, LIM J, YANG M H. Object tracking benchmark[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1834-1848.
[63] KRISTAN M, PFLUGFELDER R, LEONARDIS A, et al. The visual object tracking VOT2013 challenge results[C]// Proceedings of the 2013 International Conference on Computer Vision, Berlin, Oct 1-8, 2013. Washington: IEEE Computer Society, 2013: 98-111.
[64] KRISTAN M, PFLUGFELDER R P, LEONARDIS A, et al. The visual object tracking VOT2014 challenge results[C]// LNCS 8926: Proceedings of the 13th European Conference on Computer Vision, Zurich, Sep 6-12, 2014. Cham: Springer, 2014: 191-217.
[65] KRISTAN M, MATAS J, LEONARDIS A, et al. The visual object tracking VOT2015 challenge results[C]//Proceedings of the 2015 International Conference on Computer Vision, Santiago, Dec 7-13, 2015. Washington: IEEE Computer Society, 2015: 564-586.
[66] KRISTAN M, LEONARDIS A, MATAS J, et al. The visual object tracking VOT2016 challenge results[C]//LNCS 9914: Proceedings of the 14th European Conference on Computer Vision, Amsterdam, Oct 8-16, 2016. Cham: Springer, 2016: 777-823.
[67] KRISTAN M, LEONARDIS A, MATAS J, et al. The visual object tracking VOT2017 challenge results[C]//Proceedings of the 2017 International Conference on Computer Vision, Venice, Oct 22-29, 2017. Washington: IEEE Computer Society, 2017: 1949-1972.
[68] KRISTAN M, LEONARDIS A, MATAS J, et al. The sixth visual object tracking VOT2018 challenge results[C]//LNCS 11129: Proceedings of the 15th European Conference on Computer Vision, Munich, Sep 8-14, 2018. Cham: Springer, 2018: 3-53.
[69] KRISTAN M, BERG A, ZHENG L, et al. The seventh visual object tracking VOT2019 challenge results[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision, Seoul, Oct 27-28, 2019. Piscataway: IEEE, 2019: 2206-2241.
[70] MUELLER M, SMITH N, GHANEM B. A benchmark and simulator for UAV tracking[C]//LNCS 9905: Proceedings of the 14th European Conference on Computer Vision, Amsterdam, Oct 8-16, 2016. Cham: Springer, 2016: 445-461.
[71] MÜLLER M, BIBI A, GIANCOLA S, et al. TrackingNet: a large-scale dataset and benchmark for object tracking in the wild[C]//LNCS 11205: Proceedings of the 15th European Conference on Computer Vision, Munich, Sep 8-14, 2018. Cham: Springer, 2018: 310-327.
[72] HUANG L, ZHAO X, HUANG K. 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.
[73] FAN H, LIN L, YANG F, et al. LaSOT: a high-quality benchmark for large-scale single object tracking [C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, Jun 16-20, 2019. Piscataway: IEEE, 2019: 5374-5383.
[74] REZATOFIGHI H, TSOI N, GWAK J, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, Jun 16-20, 2019. Piscataway: IEEE, 2019: 658-666.

刘艺（1990—），男，安徽蚌埠人，博士，助理研究员，主要研究方向为机器人操作系统、数据质量、演化算法。

LIU Yi, born in 1990, Ph.D., assistant researcher. His research interests include robot operating system, data quality and evolutionary algorithms.

李蒙蒙（1992—），女，河北邯郸人，硕士研究生，主要研究方向为演化算法、数据质量、目标跟踪等。

LI Mengmeng, born in 1992, M.S. candidate. Her research interests include evolutionary algorithms, data quality, object tracking, etc.



郑奇斌（1990—），男，甘肃兰州人，博士，助理研究员，主要研究方向为数据工程、数据挖掘、机器学习等。

ZHENG Qibin, born in 1990, Ph.D., assistant researcher. His research interests include data engineering, data mining, machine learning, etc.

秦伟（1983—），男，安徽阜阳人，硕士，助理研究员，主要研究方向为智能信息系统管理。

QIN Wei, born in 1983, M.S., assistant researcher. His research interest is intelligent information system management.

任小广（1986—），男，湖北随州人，博士，副研究员，主要研究方向为高性能计算、数值计算和模拟、机器人操作系统等。

REN Xiaoguang, born in 1986, Ph.D., associate research fellow. His research interests include high performance computing, numerical computation and simulation, robot operation systems, etc.