光电技术及应用

DOI：10.16818/j.issn1001－5868.2023020401

结合全局光流的孪生区域提名网络目标跟踪算法

吴 非[1]，张建林[2*]

（1. 中国科学院光电技术研究所，成都 610200；2. 中国科学院大学 电子电气与通信工程学院，北京 101408）

摘 要： 基于孪生网络的跟踪器受限于孪生网络跟踪框架固有的跟踪机制和搜索区域选择机制，当目标处在被遮挡、快速运动和出视野等困难场景下时，如何稳定、鲁棒地进行目标跟踪始终是孪生网络跟踪器亟需解决的问题。为此，文章提出一种结合光流的孪生区域提名网络目标跟踪算法（GOF-SiamRPN）。通过全局光流对目标的运动趋势信息进行补充，该方法可以有效地解决在这些困难场景下的跟踪问题。在VOT2019和UAV123上的实验结果表明，相比基准方法，该算法分别取得了2.0%和1.8%的性能提升。与其他先进的跟踪器相比，该算法也取得了有竞争力的跟踪效果。

关键词： 目标跟踪；光流；孪生区域提名网络

中图分类号： TP391 **文章编号：** 1001－5868(2023)03－0422－07

Siamese Region Proposal Network Object Tracking Algorithm with Global Optical Flow

WU Fei[1], ZHANG Jianlin[2]

（1. Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu 610200, CHN; 2. School of Electrical, Electronics and Communication Engineering, University of Chinese Academy of Sciences, Beijing 101408, CHN）

Abstract: The Siamese network-based trackers are limited by the inherent tracking mechanism and search area selection mechanism of the Siamese network tracking framework. When the object is under challenging scenarios such as occlusion, fast motion, and out-of-view, how to perform stable and obust object tracking is always an urgent problem for Siamese trackers. To this end, in this paper, an object-tracking algorithm that combines the Siamese region proposal network with the global optical flow (GOF-SiamRPN) is proposed. By assisting the motion trend information of the object with global optical flow, the proposed method could effectively solve the potential tracking issues in these challenging scenarios. Extensive experimental results on VOT2019 and UAV123 show that the proposed method achieves a performance gain of 2.0% and 1.8% compared with the baseline method. It also achieves a competing performance compared to other state-of-the-art trackers.

Key words: object tracking; optical flow; Siamese region proposal network

0 引言

目标跟踪作为基本的计算机视觉任务，在人机交互、虚拟现实、交通监控、军事打击、自动驾驶等方面有广泛的应用[1-2]。目标跟踪任务要求在一个连续的视频序列中估计每帧目标的位置和状态信息，其中目标在视频首帧以矩形框或掩膜的形式指定。由于目标跟踪在现实世界应用上的广泛性和重要性，众多学者针对目标跟踪任务开展了集中研究[2-3]，这一领域也有其清晰的发展脉络：从最初的生成式方法聚焦在描述和建模目标本身信息，通过

收稿日期：2023－02－04.

* 通信作者：张建林 E-mail：jlin@ioe.ac.cn

最小化重构误差的方式搜寻目标[4]。如代表性的子空间学习方法[5]和稀疏表示方法[6-7]，到之后聚焦在训练一个分类器，对前景（目标）和背景进行判断进而搜索目标的判别式模型。如代表性的基于相关滤波的跟踪算法和基于深度学习的跟踪算法。作为基于深度学习目标跟踪的主流分支，基于孪生网络的目标跟踪算法由于其均衡的跟踪精度和速度而受到研究人员的青睐。孪生网络将目标跟踪过程视为相似性学习，一般包括三个部分：分别以目标模板和搜索区域作为输入；一个共享参数的神经网络作为特征提取器；相应的嵌入特征通过互相关操作得到目标模板在搜索区域上的响应图，进而通过该响应图确定目标的位置和状态。大数据量的离线训练使得孪生网络具备较强的判别能力。此外，大多数孪生网络无需在线更新，显著提升了目标跟踪的速度，这对于目标跟踪算法在实际工程项目中的落地十分重要[2]。近年来，尽管产生了许多基于孪生网络的跟踪算法，但孪生网络目标跟踪的机制从未改变，即将跟踪过程视为一个局部的检测任务，并以当前帧检测目标的位置作为下一帧搜索区域裁剪的基础。因此，当视频中目标遭遇困难场景，如遮挡、快速运动和出视野等情况时，当前帧的目标跟踪失败，更重要的是，失败的跟踪位置将影响后续搜索区域的选择，产生积累误差，最终导致目标丢失。为此，本文提出结合全局光流的孪生区域提名网络目标跟踪算法（GOF-SiamRPN）。通过光流信息辅助孪生区域提名网络搜索并选择真正目标位置，尤其是在困难场景跟踪器跟踪不稳定的情况下。此外，本算法在公开、挑战性强的 VOT2019[8] 和 UAV123[9] 上的实验结果表明，相比基准方法，本算法在性能上分别提升了 2.0%和 1.8%。与其他先进的跟踪器相比，本算法也取得有竞争力的跟踪效果，充分说明本算法的有效性。

1　相关工作

1.1　深度学习在目标跟踪中的应用

在 2012 年 AlexNet[10] 取得 ImageNet 大规模视觉识别挑战赛冠军之后，卷积神经网络（CNN）开始吸引研究人员的注意。当时，基于相关滤波器（DCF）的算法由于其高效的频域计算和模型更新机制成为目标跟踪领域的研究热点。因此深度学习与相关滤波器的结合成为了一大研究趋势，DeepSRDCF[11] 将深度特征引入相关滤波器。C-COT[12] 引入连续域插值获取多分辨率的特征，赢得 2016 年 VOT 视觉目标跟踪比赛的冠军。之后 ECO[13] 通过减少滤波器数量和降低更新频率的方式优化了 C-COT[12]，实现了优秀的跟踪精度。另一种结合深度特征实现目标跟踪的方式是训练一个分类器[14-15]，在线微调模型[16-17]。此外，孪生网络在目标跟踪中的成功应用使研究重点迅速偏向孪生网络，将在下一节详细介绍。

1.2　基于孪生网络的目标跟踪算法

孪生网络跟踪算法开始于 SINT[18] 和 SiamFC[19]，通过将跟踪过程视为搜索帧中候选框与目标模板的相似性匹配过程，孪生网络跟踪器在速度和精度上同时取得了良好的效果。GOTURN[20] 使用一个深度回归网络来学习物体外观和运动之间的通用关系。CFNet[21] 将相关滤波器和孪生网络相结合，实现了端到端的相似性表征学习。SiamRPN[22] 将目标检测中的区域提名网络引入孪生跟踪器，通过在预设的锚框上进行分类和回归，网络在提升跟踪精度的同时，大幅度提升了跟踪速度。C-RPN[23] 在孪生网络中级联了从高层到低层的一系列区域提名网络。SiamMask[24] 设计了一个用于视觉目标跟踪和视频对象分割的统一框架。SiamRPN＋＋[25] 和 SiamDW[26] 首次将深度网络作为孪生网络的骨干网络，取得了优秀的跟踪性能。之后，Transformer[27] 被引入目标跟踪[28-29]，在取得更先进跟踪效果的同时依然遵循着孪生网络跟踪框架。

1.3　光流算法在计算机视觉中的应用

光流通过计算相邻帧图像序列目标运动的二维矢量来反映目标的运动趋势及周围环境结构变化。在计算机视觉领域有广泛的应用。例如，姿态估计、目标检测、异常行为检测和医学图像处理等。文献[30]提出一种利用多帧图像结合光流用于姿态估计的卷积神经网络。文献[31]提出应用光流进行下一帧的运动估计。文献[32]通过使用光流信息辅助视频识别任务。文献[33]提出使用光流信息增强对目标的判别能力进而进行目标检测，取得同时期先进的检测效果。文献[34]在相关滤波器上引入光流信息辅助目标跟踪。文献[35]提出一个在相关滤波器上利用光流和时空注意力的端到端目标跟踪算法。文献[36]通过预训练一个光流网络[37]，在 SiamFC[19] 上通过密集连接实现更优的跟踪效果。

2 结合全局光流的孪生区域提名网络目标跟踪算法

2.1 算法框架

本文提出的结合全局光流的孪生区域提名网络目标跟踪算法(GOF-SiamRPN)如图 1 所示。它由两大分支和一个连接器构成,分别为孪生网络跟踪器,全局光流网络和难例判断模块。其中,在孪生网络跟踪器的设计上本文遵循了深度孪生区域提名网络 SiamRPN++[25]的框架,并针对该框架的不足提出了引入全局光流提取运动信息,以辅助跟踪器实现在困难场景上对目标的稳定跟踪。本算法的整体框架和细节模块将在后续章节详细描述,现对深度孪生区域提名网络的网络结构细节和其在设计上的两个特点进行介绍。

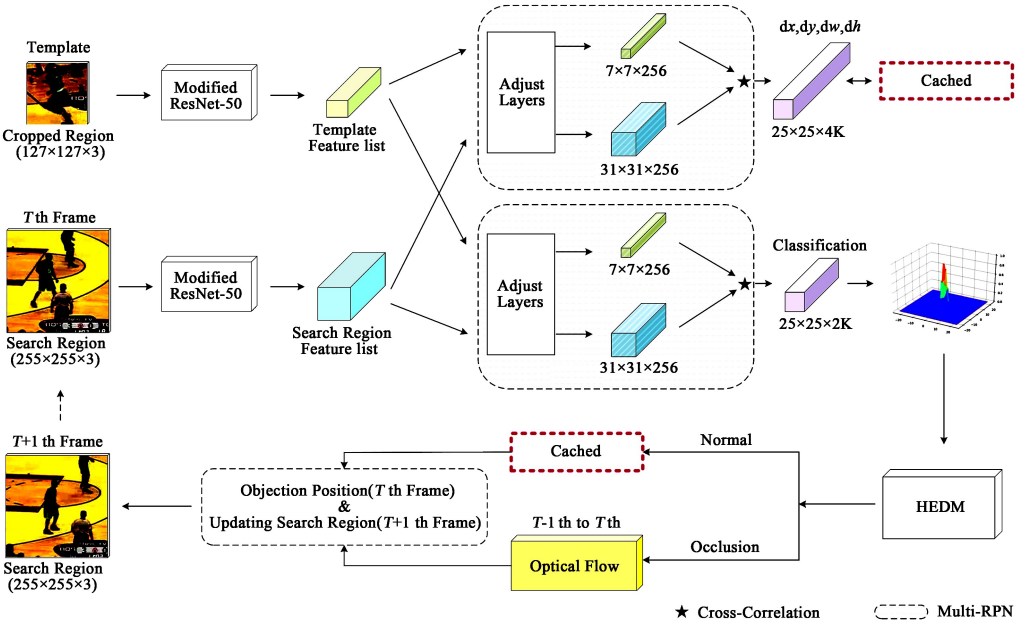


图 1 GOF-SiamRPN 网络结构

深度孪生区域提名网络有两个输入:视频首帧初始目标模板(z)和第 T 帧搜索区域(X_t)。通过深度网络 ResNet-50[38]($\varphi(*)$)提取特征,其中 z 和 X_t 在 $\varphi(*)$中共享参数,以实现相似性学习。其过程可以表达为如下形式:

$$S(z,X_t)=\mathrm{Corr}(\varphi(z),\varphi(X_t)) \tag{1}$$

式中,$Corr(*)$表示互相关操作,$S(z,X_t)$表示 z 与 X_t 特征嵌入的相似性度量,可以通过该度量值确定目标位置。具体而言,在深度孪生区域提名网络中,相似性度量被进一步分为分类分支和回归分支用于目标的识别与定位,通过引入锚框设计,响应图上的每个点对应 5 个不同尺度的锚框。分类分支和回归分支的响应可以表达为如下形式:

$$A^{\mathrm{CLS}}_{W\times H\times 2K}=Corr([\varphi(X_t)]_{\mathrm{CLS}},[\varphi(z)]_{\mathrm{CLS}}) \tag{2}$$

$$A^{\mathrm{REG}}_{W\times H\times 4K}=Corr([\varphi(X_t)]_{\mathrm{REG}},[\varphi(z)]_{\mathrm{REG}}) \tag{3}$$

式中,CLS 和 REG 分别表示分类分支和回归分支;$A^{\mathrm{CLS}}_{W\times H\times 2K}$ 表示区域提名网络在 $W\times H$ 的尺度上,每个响应点对应 $2K$ 个分类置信度,K 为锚框数量;同理,$A^{\mathrm{REG}}_{W\times H\times 4K}$ 表示其对应的 $4K$ 个回归偏移。通过学习分类置信度和回归偏移,网络得以实现对目标的识别与定位。此外,在网络结构设计上,深度孪生区域提名网络具备以下两个特点:

(1)通过聚合 ResNet-50[38]的最后三个代表不同语义和细节信息的卷积层特征并依次通过三个区域提名网络(Multi-RPN),且在分类分支和回归分支上分别进行线性加权得以显著提升网络的判别能力和回归能力。

(2)使用深度可分离卷积进行目标模板和搜索区域的互相关操作相比之前方法如 SiamFC[19],SiamRPN[22]的直接卷积可以显著降低参数数量,尤其是在 2 倍锚框数量乘以通道数量的前提下,这有利于网络的优化。

仅使用深度孪生区域提名网络进行目标跟踪无法避免孪生网络跟踪框架下的一个根本问题,即目标在当前帧中以局部区域裁剪搜索区域,且此搜索区域与上帧网络预测的目标位置直接相关。因此,

当网络对当前帧预测不可靠时，不仅会错误输出当前帧的目标位置，并且会影响下一帧搜索区域的选择，导致连续时序目标跟踪任务失败。为此，我们提出引入全局光流网络对目标的跟踪信息进行补偿，以克服在困难场景下目标跟踪容易失败的问题。

2.2　在引入光流信息

在引入光流信息之前，首先需要准确判断目标是否处于困难场景中导致当前深度孪生区域提名网络的，本文采用本人所提出的 HEDM[39] 对此进行检测。HEDM 与峰值旁瓣比（PSR）和分类分数的置信度（Score）相比可以更加有效地判断目标处于困难场景的情况[39]，进而对目标位置和运动信息进行补偿。HEDM 分别由遮挡系数 TOI 和分类分数 Score 确定。如下所示：

$$TOI=\frac{R_{\text{peak}}-R_{\text{low}}}{R_{\text{std}}} \tag{4}$$

$$HEDM=\theta_1\times Score_{\text{Max}}^{-1}+\theta_2\times TOI \tag{5}$$

式中，R_{peak} 是当前帧响应图上的最大值，R_{low} 是当前帧响应图上的最小值，R_{std} 是当前帧响应图的标准差，θ_1，θ_2 分别是用以平衡该项分布的超参数。

当 HEDM 大于本文设置的阈值时，当前帧被确定为困难帧，此时，引入光流网络对上一帧到当前帧的全局输入进行光流估计，以保证光流输出的稳定性，如图 2 所示。

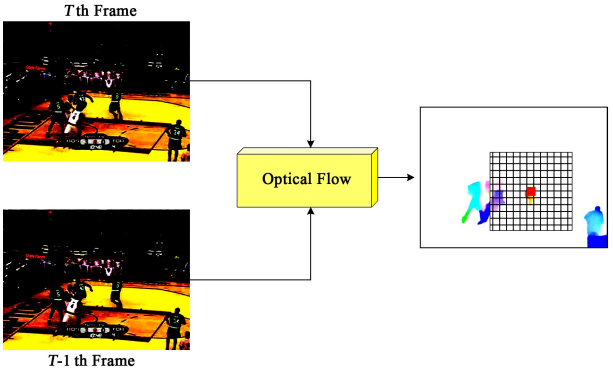


图 2　目标在光流信息上的检索

本文使用 RAFT[40] 作为光流基准网络，RAFT[40] 具有模型小、精度高和推理速度快的优点，非常适合跟踪任务。在光流信息的辅助下，本文使用一个 MaxPooling 层检索目标在光流层中运动信息，如下所示：

$$Motion=MaxPooling(Flow_Feature_{5\times5}) \tag{6}$$

$$ObjectPosition_T=ObjectPosition_{T-1}+Motion \tag{7}$$

以第 $T-1$ 帧预测的目标位置为中心在光流图上裁剪 5×5 的区域，通过 MaxPooling 采集响应最大的点输出运动信息，将其引入第 $T-1$ 帧预测的目标位置得到第 T 帧的目标位置，成功避免困难场景下深度孪生区域提名网络对目标的错误预测，并提升了搜索区域的选择稳定性。

3　实验结果与分析

3.1　实验设备与平台

实验平台为 Ubuntu 20.04LTS，Pytorch1.2，实验设备为 Intel i7-7820X，Nvidia RTX6000 24 GB。为验证本算法对孪生网络跟踪框架的高效性和即插即用性，本文使用在 ResNet-50[38] 上预训练的孪生区域提名网络作为特征提取器和位置回归器，预训练的 RAFT[40] 作为光流基准网络，公式（5）中的 θ_1，θ_2 分别被设置为 0.05 和 1.00。

3.2　在 VOT2019 上的实验

VOT[38] 视觉目标跟踪竞赛是目标跟踪领域在国际上最权威的竞赛之一，自 2013 年以来每年举办一次。总共包含 60 个测试数据集，数据集涵盖各种困难场景，如遮挡、光照变化、运动变化、尺度变化、出视野等，每年更新部分测试集。VOT 对跟踪器的评测包括三个指标，其中精度（Accuracy）指的是预测结果与目标标注信息的平均重叠率。鲁棒性（Robustness）指的是算法跟踪失败的次数，失败是指跟踪结果与目标真实位置无重叠。由于 VOT 有跟踪失败重新初始化的机制，平均期望重叠（EAO）综合反映了跟踪器的性能。

本算法 GOF-SiamRPN 在 VOT2019 上的实验结果如图 3 所示。本算法的性能在众多算法中处于领先地位。具体与多个孪生跟踪器包括 SiamMargin[38]，SiamFCOT[38]，SiamCRF[38]，SiamDW[26]，SiamMask[24]，SA-Siam-R[41] 和 SiamMsST[38] 等相比的三个指标性能如表 1 所示。

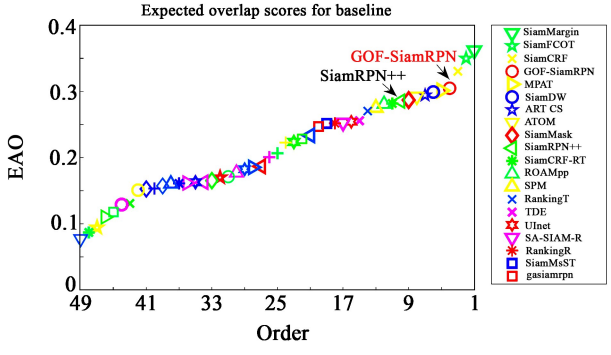


图 3　GOF-SiamRPN 在 VOT2019 上的评测图

与基准方法：SiamRPN＋＋[25]相比，本算法在EAO，Accuracy 和 Robustness 上分别提升了2.0%，0.6%和1.0%。值得注意的是，本算法旨在缓解孪生网络跟踪框架中的搜索机制选择问题，因此可以集成在这些孪生网络跟踪器中并带来性能提升。

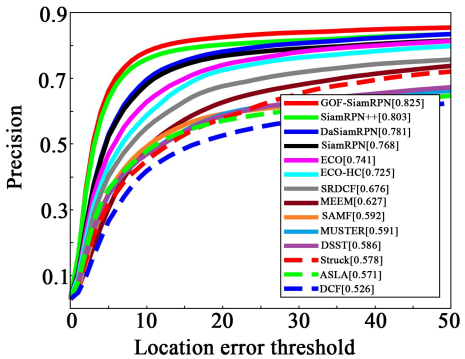
表 1 在 VOT2019 上的性能比较

跟踪器	EAO	Accuracy	Robustness
SiamMargin	0.362	0.578	0.326
SiamFCOT	0.350	0.601	0.386
SiamCRF	0.330	0.625	0.296
GOF-SiamRPN	0.305	0.605	0.472
MPAT	0.301	0.632	0.414
SiamDW	0.299	0.600	0.467
ARTCS	0.294	0.602	0.456
ATOM	0.292	0.603	0.411
SiamMask	0.287	0.594	0.461
SiamRPN＋＋	0.285	0.599	0.482
SiamCRF-RT	0.282	0.550	0.301

3.3 在 UAV123 上的实验

UAV123[39]包含了无人机在空中拍摄的 123 个视频，共 9 种类别。相比通用的视频拍摄场景如 VOT，UAV123 中标注的目标往往尺度更小、变化更大，挑战性较强。UAV123 采用成功率（Success Rate）和精度（Precision）两个指标评估跟踪器的性能。成功率是指预测目标与真实位置的重叠率大于所设定的阈值。精度是指预测的目标中心位置与真实目标中心位置的偏移小于所设定的阈值。

本算法 GOF-SiamRPN 在 UAV123 上的评估结果如图 4 所示，其中包括多个先进的跟踪器如 SiamRPN＋＋[25]，DaSiamRPN[42]，SiamRPN[22]，ECO[13]和 SRDCF[43]等。本算法的性能在众多算法中处于领先地位。与基准方法：SiamRPN＋＋[25]相比，本算法在成功率和精度上分别提升了 1.8% 和 2.2%，充分说明了本算法的有效性。



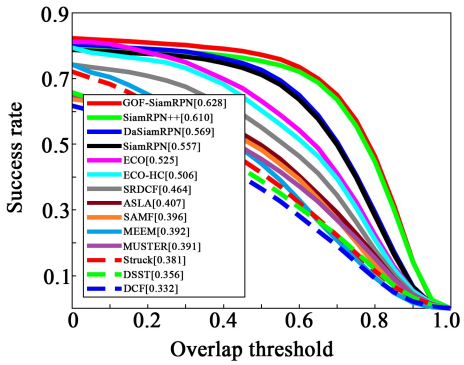


图 4 本算法在 UAV123 上的评测图

4 结论

本文针对孪生网络框架搜索机制选择的固有约束，提出了结合全局光流的孪生区域提名网络目标跟踪算法（GOF-SiamRPN）。通过结合全局光流对目标在困难场景的跟踪信息进行补充，以应对目标被遮挡、快速运动和出视野等情况。在 VOT2019 和 UAV123 上展开的实现及结果表明，本算法有效提升了基准跟踪器的目标跟踪精度，与其他优秀的跟踪器相比本算法也占据优势地位。更重要的是，本算法作为针对孪生网络跟踪器固有缺陷的方法，可以即插即用的整合在任意孪生网络跟踪框架中。

参考文献：

[1] Su K，Li J，Fu H. Smart city and the applications[C]// 2011 Inter. Conf. on Electronics，Communications and Control (ICECC)，2011：1028-1031.

[2] Marvasti-Zadeh S M，Cheng L，Ghanei-Yakhdan H，et al. Deep learning for visual tracking：A comprehensive survey[J]. IEEE Trans. on Intelligent Transportation Systems，2022，23(5)：3943-3968.

[3] Smeulders A W M，Chu D M，Cucchiara R，et al. Visual tracking：An experimental survey[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence，2014，36(7)：1442-1468.

[4] Yu Q，Dinh T B，Medioni G. Online tracking and reacquisition using co-trained generative and discriminative trackers[C]// Computer Vision-ECCV 2008，2008：678-691.

[5] Hager G，Belhumeur P. Efficient region tracking with parametric models of geometry and illumination[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence，1998，20(10)：1025-1039.

[6] Zhang T，Bibi A，Ghanem B. In defense of sparse tracking：Circulant sparse tracker[C]// 2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)，2016：3880-3888.

[7] Zhong W，Lu H，Yang M H. Robust object tracking via sparsity-based collaborative model[C]// 2012 IEEE Conf. on

Computer Vision and Pattern Recognition，2012：1838-1845.

[8] Kristan M，Matas J，Leonardis A，et al. The seventh visual object tracking VOT2019 challenge results[C]// 2019 IEEE/CVF Inter. Conf. on Computer Vision Workshop（ICCVW），2019：2206-2241.

[9] Mueller M，Smith N，Ghanem B. A benchmark and simulator for UAV tracking[C]// Computer Vision-ECCV 2016，2016：445-461.

[10] Krizhevsky A，Sutskever I，Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Commun. ACM，2017，60(6)：84-90.

[11] Danelljan M，Häger G，Khan F S，et al. Convolutional features for correlation filter based visual tracking[C]// 2015 IEEE Inter. Conf. on Computer Vision Workshop（ICCVW），2015：621-629.

[12] Danelljan M，Robinson A，Shahbaz Khan F，et al. Beyond correlation filters：Learning continuous convolution operators for visual tracking[C]// Computer Vision-ECCV 2016，2016：472-488.

[13] Danelljan M，Bhat G，Khan F S，et al. Eco：Efficient convolution operators for tracking[C]// 2017 IEEE Conf. on Computer Vision and Pattern Recognition（CVPR），2017：6931-6939.

[14] Wang L，Ouyang W，Wang X，et al. Visual tracking with fully convolutional networks[C]// 2015 IEEE Inter. Conf. on Computer Vision（ICCV），2015：3119-3127.

[15] Wang L，Ouyang W，Wang X，et al. Stct：Sequentially training convolutional networks for visual tracking[C]// 2016 IEEE Conf. on Computer Vision and Pattern Recognition（CVPR），2016：1373-1381.

[16] Jung I，Son J，Baek M，et al. Real-time mdnet[C]// Computer Vision-ECCV 2018，2018：89-104.

[17] Park E，Berg A C. Meta-tracker：Fast and robust online adaptation for visual object trackers[C]// Computer Vision-ECCV 2018，2018：587-604.

[18] Wang G，Luo C，Sun X，et al. Tracking by instance detection：A meta-learning approach[C]// 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition（CVPR），2020：6287-6296.

[19] Bertinetto L，Valmadre J，Henriques J F，et al. Fullyconvolutional siamese networks for object tracking[C]// Computer Vision-ECCV 2016，2016：850-865.

[20] Held D，Thrun S，Savarese S. Learning to track at 100 fps with deep regression networks[C]// Computer Vision-ECCV 2016，2016：749-765.

[21] Valmadre J，Bertinetto L，Henriques J，et al. End-to-end representation learning for correlation filter based tracking[C]// 2017 IEEE Conf. on Computer Vision and Pattern Recognition（CVPR），2017：5000-5008.

[22] Li B，Yan J，Wu W，et al. High performance visual tracking with siamese region proposal network[C]// 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition，2018：8971-8980.

[23] Fan H，Ling H. Siamese cascaded region proposal networks for real-time visual tracking[C]// 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition（CVPR），2019：7944-7953.

[24] Wang Q，Zhang L，Bertinetto L，et al. Fast online object tracking and segmentation：A unifying approach[C]// 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition（CVPR），2019：1328-1338.

[25] Li B，Wu W，Wang Q，et al. Siamrpn＋＋：Evolution of siamese visual tracking with very deep networks[C]// 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition（CVPR），2019：4277-4286.

[26] Zhang Z，Peng H. Deeper and wider siamese networks for real-time visual tracking[C]// 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition（CVPR），2019：4586-4595.

[27] Vaswani A，Shazeer N，Parmar N，et al. Attention is all you need[EB/OL]. arXiv，2017. http://arxiv.org/abs/1706.03762.

[28] Lin L，Fan H，Xu Y，et al. Swintrack：A simple and strong baseline for transformer tracking[EB/OL]. arXiv，2021. http://arXiv.org/abs/2112.00995.

[29] Wang N，Zhou W，Wang J，et al. Transformer meets tracker：Exploiting temporal context for robust visual tracking[C]// 2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition（CVPR），2021：1571-1580.

[30] Pfister T，Charles J，Zisserman A. Flowing convnets for human pose estimation in videos[C]// 2015 IEEE Inter. Conf. on Computer Vision（ICCV），2015：1913-1921.

[31] Patraucean V，Handa A，Cipolla R. Spatio-temporal video autoencoder with differentiable memory[EB/OL]. arXiv，2015. http://arxiv.org/abs/1511.06309.

[32] Zhu X，Xiong Y，Dai J，et al. Deep feature flow for video recognition[C]// 2017 IEEE Conf. on Computer Vision and Pattern Recognition（CVPR），2017：4141-4150.

[33] Zhu X，Wang Y，Dai J，et al. Flow-guided feature aggregation for video object detection[C]// 2017 IEEE Inter. Conf. on Computer Vision（ICCV），2017：408-417.

[34] Gladh S，Danelljan M，Khan F S，et al. Deep motion features for visual tracking[C]// 2016 23rd Inter. Conf. on Pattern Recognition（ICPR），2016：1243-1248.

[35] Zhu Z，Wu W，Zou W，et al. End-to-end flow correlation tracking with spatial-temporal attention[C]// 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition，2018：548-557.

[36] Zhou L，Yao X，Zhang J. Accurate positioning siamese network for real-time object tracking[J]. IEEE Access，2019，7：84209-84216.

[37] Ilg E，Mayer N，Saikia T，et al. Flownet 2.0：Evolution of

optical flow estimation with deep networks[C]// 2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2017: 1647-1655.

[38] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]// 2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2016: 770-778.

[39] Wu F, Zhang J, Xu Z. Stably adaptive anti-occlusion siamese region proposal network for real-time object tracking[J]. IEEE Access, 2020, 8: 161349-161360.

[40] Teed Z, Deng J. Raft: Recurrent all-pairs field transforms for optical flow[C]// Computer Vision-ECCV 2020, 2020: 402-419.

[41] He A, Luo C, Tian X, et al. A twofold siamese network for real-time object tracking[C]// 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition, 2018: 4834-4843.

[42] Zhu Z, Wang Q, Li B, et al. Distractor-aware siamese networks for visual object tracking[C]// Computer Vision-ECCV 2018, 2018: 103-119.

[43] Danelljan M, Häger G, Khan F S, et al. Learning spatially regularized correlation filters for visual tracking[C]// 2015 IEEE Inter. Conf. on Computer Vision (ICCV), 2015: 4310-4318.

作者简介：

吴　非(1995—)，男，博士生，研究方向为基于深度学习的目标跟踪与分割；

张建林(1976—)，男，四川省乐山市人，博士，研究员，博士生导师，主要研究方向为计算机视觉、深度学习、图像分析与理解技术。