



一种低参数的孪生卷积网络实时目标跟踪算法

罗 朔,侯 进,谭光鸿,韩雁鹏

(西南交通大学 信息科学与技术学院,成都 611756)

摘 要:针对基于深度学习的目标跟踪算法模型参数多、难以部署于嵌入式设备上的问题,提出一种改进的孪生卷积网络实时目标跟踪算法。设计一个非对称卷积模块来构建整个网络框架,通过非对称卷积模块的压缩层减少模型参数量,利用非对称层进行特征融合,以在保证精度的同时压缩模型大小。使用三元组损失函数代替逻辑损失函数进行模型训练,在输入不变的情况下提取表达性更强的深度特征,从而完成目标跟踪任务并提高模型的跟踪精度。在 GOT-10K、OTB100 和 VOT2016 基准上对算法性能进行测试,结果表明,该算法能够将模型大小降为 3.8×10^6 ,且速度与精度均优于 SiamFC、KCF 和 DAT 等跟踪算法。

关键词:目标跟踪;低参数模型;孪生卷积网络;实时性;非对称卷积;三元组损失

开放科学(资源服务)标志码(OSID):



中文引用格式:罗朔,侯进,谭光鸿,等.一种低参数的孪生卷积网络实时目标跟踪算法[J].计算机工程,2021,47(2):84-89.

英文引用格式:LUO Shuo, HOU Jin, TAN Guanghong, et al. A low parameter real-time target tracking algorithm based on siamese convolutional network[J]. Computer Engineering, 2021, 47(2): 84-89.

A Low Parameter Real-Time Target Tracking Algorithm Based on Siamese Convolutional Network

LUO Shuo, HOU Jin, TAN Guanghong, HAN Yanpeng

(The School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, China)

[Abstract] The existing target tracking algorithms based on deep learning face deployment problems on embedded devices due to their large number of parameters. To address the problems, this paper proposes a real-time target tracking algorithm based on low parameter siamese convolutional network. By designing an asymmetric convolution module to build the entire network framework, the compression layer of the asymmetric convolution module is used to reduce the number of parameters, and the asymmetric layer is used for feature fusion to reduce the size of the model while maintaining accuracy. At the same time, the triplet loss function is used to replace the logical loss function for model training, and the more expressive depth features are extracted for target tracking with the same input, which improves the tracking accuracy of the model. The performance of the algorithm is tested on the GOT-10K, OTB100 and VOT2016 benchmarks. Experimental results show that the algorithm reduces the model size to 3.8×10^6 , and outperforms SiamFC, KCF, DAT and other tracking algorithms in terms of speed and accuracy.

[Key words] target tracking; low parameter model; siamese convolutional network; real-time performance; asymmetric convolution; triplet loss

DOI: 10.19678/j.issn.1000-3428.0057076

0 概述

目标跟踪是计算机视觉领域的重要研究课题之一,广泛应用于人机交互、智能视频监控和医学诊断等场景^[1]。目标跟踪定义为根据给定视频中第一帧的目标位置来预估随后帧中的目标位置。在一般情况下,目标跟踪技术主要在连续的视频序列中建立

所要跟踪物体的位置关系,然后得到物体完整的运动轨迹,从而为分析视频内容提供数据依据并对目标行为进行分析。

传统的目标跟踪算法在面对目标快速移动、目标模糊、物体形变和光照变化等复杂场景时,难以准确跟踪到目标^[2-4],存在较大的局限性。随着大数据时代的到来,深度学习技术凭借其强大的特征提取

基金项目:浙江大学CAD&CG国家重点实验室开放课题“基于深度学习的调制识别技术的研究”(A1823)。

作者简介:罗 朔(1994—),男,硕士研究生,主研方向为计算机视觉、深度学习;侯 进(通信作者),副教授;谭光鸿、韩雁鹏,硕士研究生。

收稿日期:2019-12-31 修回日期:2020-02-05 E-mail:1907128121@qq.com

能力,在图像识别和目标分割等计算机视觉领域引起广泛关注,同时也推动了目标跟踪技术的发展。2013年,WANG等人提出的DLT^[5]将深度学习与单目标跟踪任务相结合,虽然效果比某些传统算法差,但是其为深度学习应用于目标跟踪提供了新思路。2016年,BERTINETTO等人提出了一种新的全卷积孪生网络跟踪算法SiamFC^[6],其用孪生网络进行相似度比较,将跟踪问题转换为相似度学习问题,在ILSVRC15的目标跟踪视频数据集上进行端到端的训练,达到了实时跟踪的效果。2017年,SONG等人提出CREST算法^[7],其建立一种端到端的跟踪模型,将特征提取和响应生成融合在深度学习框架中,仅采用单层卷积的端到端结构就达到了利用深度特征的传统协同滤波器的效果。2018年,WANG等人提出RASNet算法^[8],其在SiamFC的基础上加入一般注意力、残差注意力、通道注意力3种机制,将SiamFC网络结构改为端到端的网络,有效缓解了过拟合问题,提升了网络的判别能力和适应能力。

近年来,越来越多的研究人员利用深度学习强大的特征提取能力来提高跟踪精度,其中性能较优的深度学习目标跟踪算法包括MDNet、SiamRPN和SiamMask^[9-11]等。但是,基于深度学习的跟踪算法模型大,参数量多,很难部署于其他嵌入式设备上。SiamFC算法虽然能够实现实时跟踪,但是其模型依然很大,且训练时没有充分利用样本之间的关系。

本文提出一种低参数的孪生卷积网络实时目标跟踪算法,设计基于非对称卷积模块的孪生网络框架,使用非对称卷积模块来降低模型的参数量,同时利用三元组损失函数进行训练,提取出表达性较强的深度特征完成目标跟踪,从而提高算法的目标跟踪精度。

1 SiamFC跟踪算法

全卷积孪生网络SiamFC的核心思想是将跟踪一个任意目标当作一种相似度学习,通过比较上一帧图像和下一帧图像来返回一个相似度值,该相似度值越高,说明2个图像越相似。SiamFC采用逻辑损失函数进行训练,其表达式为:

$$L_l(y, v) = \sum_{x_i \in x} w_i \ln(1 + e^{-y_i v_i}) \quad (1)$$

其中, y, v, x 分别为标签集、相似性评分集、实例输入集, v_i 是单个样本中跟踪器跟踪框的真实得分, $y_i \in \{1, -1\}$ 是单个样本的参考标签分数, w_i 为实例 x_i 的权重,在SiamFC中,根据正负实例的数量将平衡权重应用于损失函数中。平衡权重的定义如下:

$$w_i = \begin{cases} \frac{1}{2M}, & y_i = 1 \\ \frac{1}{2N}, & y_i = -1 \end{cases} \quad (2)$$

其中, M 为正实例输入集 x_p 的数量, N 为负实例输入集 x_n 的数量, $M = |x_p|, N = |x_n|$ 。在SiamFC中, $M = 13, N = 212$ 。

虽然SiamFC利用深度网络来提取特征,但其只利用了样本之间的两两关系,忽略了正样本和负样本之间的潜在关系,同时,SiamFC模型也存在一定冗余。因此,设计一种参数量低且特征提取效率高的网络模型十分有必要。

2 孪生卷积网络实时目标跟踪算法

2.1 网络结构设计

SiamFC所用的网络结构为AlexNet,虽然其层数较少,但是模型参数量依然很大。参数量小的卷积神经网络模型可以进行更高效的训练,能更方便地部署在嵌入式设备上。为了压缩模型大小,文献[12]提出将 $d \times d$ 的卷积分解为 $1 \times d$ 和 $d \times 1$ 卷积的方法,以减少参数量。文献[13]提出的ENet也采用上述方法来设计高效的语义分割网络,该网络虽然精度有所下降,但是其降低了33%的参数量。本文提出一种基于非对称卷积模块的网络结构,如图1所示。

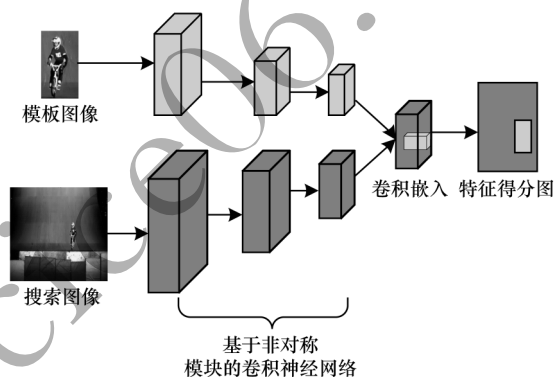


图1 基于非对称卷积模块的网络结构

Fig.1 Network structure based on asymmetric convolution module

图1所示网络结构通过训练孪生网络来学习一个函数 $f(z, x)$,将模板图像 z 与相同大小的搜索图像 x 进行比较,如果2个图像描述相同的对象,则返回高分;否则,返回低分。基于非对称模块的卷积神经网络中有一个特征提取器 ϕ ,其对2个输入图像进行相同的特征提取,再通过一个卷积嵌入函数比较上述特征,卷积嵌入函数可以定义为:

$$f(z, x) = g(\phi(z), \phi(x)) \quad (3)$$

其中, g 是一个距离度量或相似度度量。本文采用的神经网络为全卷积神经网络,将卷积神经网络最后的全连接层换成卷积层。全卷积神经网络应用于目标跟踪的优势在于待搜索图像不需要与模板图像具有相同的尺寸,可以为网络提供更大的搜索图像作为输入,然后在密集网格中计算所有平移窗口的相

似度。全卷积神经网络定义为:

$$f(z, x) = \varphi(z) * \varphi(x) + b \quad (4)$$

其中, b 为偏置项。式(4)的输出是一个标量值的分数映射, 大小为搜索区域和模板图像经过一系列卷积后的尺度, 得分最高的位置对应搜索区域中需要跟踪的目标位置。

一维卷积核通常被用于逼近正方形卷积核的特征提取效果, 以此进行模型压缩和加速。如果几个大小互为转置的一维卷积核在相同的输入上以相同的步幅进行卷积, 产生相同分辨率的特征输出, 可以将这些特征与 3×3 卷积核提取后的特征在通道上进行融合, 从而得到一个等效的特征输出, 并且不会增加额外的计算负担。非对称卷积模块包括压缩层和非对称层 2 个卷积层, 其中, 压缩层只包含 1×1 的卷积核, 非对称层则包含 1×3 、 3×1 和 3×3 3 种卷积核。与 AlexNet 相比, 非对称卷积模块使用大量 1×1 的卷积核来替换 3×3 的卷积核, 可以将参数降低为原先的 $1/9$ 。在压缩层中使用 1×1 的卷积核降低输入到 3×3 卷积核中的通道数, 也可以降低模型的参数量。在非对称层中应用 1×3 、 3×1 和 3×3 的卷积核进行特征再融合, 非对称卷积模块结构如图 2 所示, 整个网络结构参数如表 1 所示, 其中, S1 为压缩层中 1×1 的卷积核个数, 即输入到压缩层中的通道数。

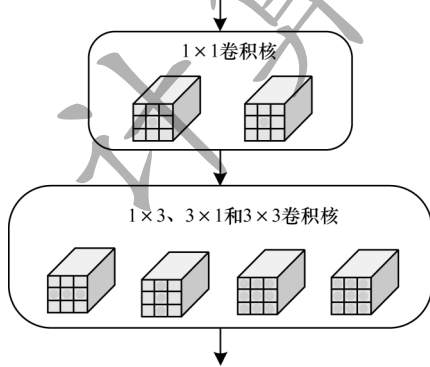


图 2 非对称卷积模块结构

Fig.2 Structure of asymmetric convolution module

表 1 网络结构参数

Table 1 Network structure parameters

网络层	搜索图像大小	模板图像大小	S1	输出通道数
Conv1	255×255	127×127		96
Maxpool1	123×123	63×63		
Asymmetry module1	61×61	31×31	64	128
Asymmetry module2	61×61	31×31	64	128
Asymmetry module3	61×61	31×31	128	256
Maxpool2	61×61	31×31		
Asymmetry module4	30×30	15×15	128	256
Asymmetry module5	30×30	15×15	128	256
Asymmetry module6	30×30	15×15	256	512

整个网络的输入通道数为 3, 所有卷积层都有 ReLU 非线性激活函数, 都加入了批处理规范化层来进行数据的归一化处理, 使得在进行 ReLU 之前不会因为数据过大而导致网络性能不稳定。

2.2 三元组损失

三元组损失广泛应用于计算机视觉领域中的人脸识别、图像检索和行人再识别^[14-16]等任务。本文提出一种新的三元组损失, 将其加入到孪生网络结构中, 以充分挖掘输入之间的内在联系。如同分割实例集 x , 将相似度的评分集 v 也分割为正评分集 v_p 和负评分集 v_n , 然后直接在这些得分对上定义三元组损失。为了测量每个得分对, 本文应用匹配概率, 即使用 softmax 函数将正实例分配给示例的概率。匹配概率的定义如下:

$$\text{prob}(v_p, v_n) = \frac{e^{v_p}}{e^{v_p} + e^{v_n}} \quad (5)$$

本文目标是使所有得分对之间的联合概率最大, 即所有概率的乘积最大。通过使用负对数可以得出损失公式如下:

$$L_t = -\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \ln \text{prob}(v_p, v_n) \quad (6)$$

将式(5)代入式(6)可得:

$$L_t = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \ln(1 + e^{v_n - v_p}) \quad (7)$$

将式(2)代入式(1)可以得出逻辑损失函数如下:

$$L_l = \sum_{i=1}^M \frac{1}{2M} \ln(1 + e^{-v_p}) + \sum_{j=1}^N \frac{1}{2N} \ln(1 + e^{v_n}) \quad (8)$$

进一步可得:

$$L_l = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \frac{1}{2} (\ln(1 + e^{-v_p}) + \ln(1 + e^{v_n})) \quad (9)$$

从式(8)、式(9)可以看出, 2 种损失函数的区别在于求和项不同, 可设:

$$T_l = \frac{1}{2} (\ln(1 + e^{-v_p}) + \ln(1 + e^{v_n})) \quad (10)$$

$$T_t = \ln(1 + e^{v_n - v_p}) \quad (11)$$

梯度在深度学习的训练中起重要作用, 因为涉及反向传播阶段, 所以可以通过梯度来分析 2 个损失函数的特点。对于逻辑损失函数项, 梯度为:

$$\frac{\partial T_l}{\partial v_p} = -\frac{1}{2(1 + e^{v_p})}, \quad \frac{\partial T_l}{\partial v_n} = \frac{1}{2(1 + e^{v_n})} \quad (12)$$

对于三元组损失函数项, 其梯度为:

$$\frac{\partial T_t}{\partial v_p} = -\frac{1}{1 + e^{v_p - v_n}}, \quad \frac{\partial T_t}{\partial v_n} = \frac{1}{1 + e^{v_p - v_n}} \quad (13)$$

由式(12)、式(13)可以看出, 逻辑损失的梯度 $\partial T_l / \partial v_p$ 和 $\partial T_l / \partial v_n$ 分别只依赖 v_p 和 v_n , 这意味着逻辑损失函数不能充分利用 v_p 和 v_n 的信息, 而三元组损失的梯度 $\partial T_t / \partial v_p$ 、 $\partial T_t / \partial v_n$ 与 v_p 、 v_n 都有联系, 与逻辑损失相比, 三元组损失可以同时利用 v_p 和 v_n 的信息, 从而实现

更强大的表示。同时,因为本文的三元组损失是在原始分数的基础上定义的,使用正分数和负分数的组合,所以可以使用相同的输入来满足网络,在训练过程中不需要额外的计算也可进行深度网络的特征提取。

3 实验结果与分析

3.1 训练数据集

GOT-10K数据集^[17]是中科院在2018年发布的一个目标追踪数据集,其包含了超过10 000条视频,分成560多个类别,87种运动模式,人工标注边界框超过150万个。数据集又分为训练集、验证集和测试集。GOT-10K数据集与其他数据集的区别在于其子集之间不存在交集,可以使得训练出的模型有更强的泛化能力。

3.2 结果分析

对整个数据集训练50轮,每个阶段包括9 335对样本,训练的batchsize大小为8,学习率在 10^{-2} ~ 10^{-5} 之间进行衰减。计算机CPU为Intel I7-6800K处理器,主频为3.4 GHz,内存为32 GB,显卡为NVIDIA GeForce GTX1080Ti,实验环境为ubuntu16.04,Pytorch框架。为了充分验证本文算法的性能,选择在当前比较流行的跟踪基准GOT-10K、OTB100^[18]和VOT2016^[19]上进行测试。

3.2.1 GOT-10K基准

在GOT-10K测试集上验证算法性能,GOT-10K包含180段视频,一共有84个目标类别和32个动作类别。采用平均重叠率(AO)、成功率(SR)和FPS 3个评价指标。AO为所有帧跟踪结果与事实之间重叠率的平均值,SR为重叠率超过一定阈值的成功跟踪帧所占的百分比,本文选择0.50和0.75这2个阈值,FPS为每秒传输帧数。GOT-10K基准下的实验结果如表2所示。

表2 GOT-10K基准下的评估结果

Table 2 Evaluation results under GOT-10K benchmark

算法	AO/%	SR(0.50) /%	SR(0.75) /%	FPS	模型大小 /10 ⁶
SiamFC	35.6	39.5	11.4	88	9.4
Asymmetry module	35.2	39.3	11.2	97	3.8
Triplet loss	37.9	41.2	12.6	88	9.4
本文算法	37.4	40.6	12.3	97	3.8

在表2中,本文算法同时使用非卷积模块和改进的三元组损失函数。从表2可以看出,在仅使用非对称卷积模块时,模型大小只有 3.8×10^6 ,算法精度略低于SiamFC算法,但是速度提升了9FPS。在仅使用三元组损失函数时,算法精度有所提升,模型大小保持不变。当同时使用非对称卷积模块和三元组损失函数时,模型精度和速度均优于SiamFC算法,其中,AO提升了1.8个百分点,SR(0.50)与SR(0.75)分别提升了1.1和0.9个百分点,速度也提升了9FPS,模型大小为 3.8×10^6 ,

只有SiamFC算法的40%。本文算法在精度、速度都提升的情况下降低了模型大小,为跟踪算法部署于嵌入式设备提供了可能。

3.2.2 OTB基准

OTB也是视觉跟踪领域广泛使用的基准库^[20],其包括OTB50和OTB100 2个数据集,本文选择OTB100数据集,该数据集包含100个人工标注的视频帧,每个序列包括11个不同的属性,如尺度变换、运动模糊、光照变化和遮挡等。本文采用准确率和成功率2个评价指标,准确率表示中心点距离小于给定阈值的视频帧所占的百分比,成功率表示重合率得分超过某个阈值的帧所占的百分比。各算法准确率和成功率对比结果如图3所示。图3(a)中的横坐标为中心点位置误差的阈值,其为一个像素值,当中心点位置误差的阈值越大时,算法的准确率越高。图3(b)中的横坐标为重叠率阈值,重叠率阈值越高,算法的成功率越低。从图3可以看出,在仅使用三元组损失函数时,模型的准确率和成功率分别达到79.8%和59.5%,本文算法同时使用非卷积模块和三元组损失函数,准确率和成功率也达到78.9%和59.2%,相比SiamFC算法都有一定提升。

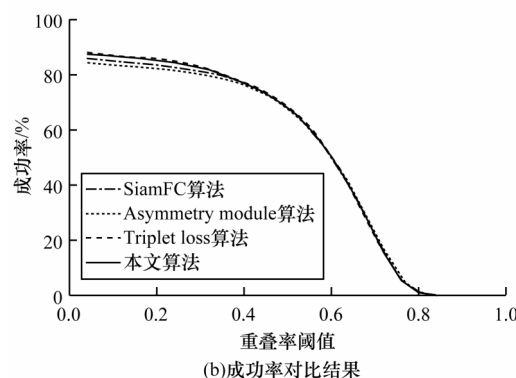
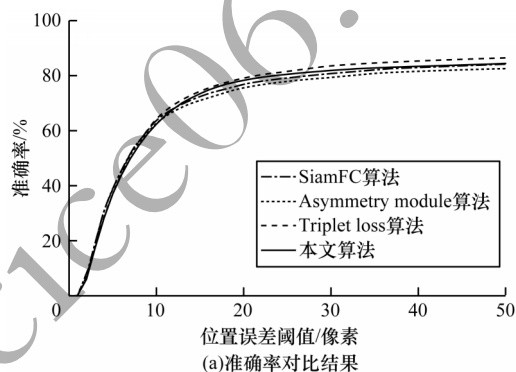


图3 OTB100基准下算法性能对比结果

Fig.3 Comparison results of algorithms performance under OTB100 benchmark

3.2.3 VOT2016基准

VOT是一个针对单目标跟踪的测试平台,本文选择在VOT2016数据集上进行测试,评价标准为预期平均重叠率(Expected Average Overlap rate, EAO)、准确率

(Accuracy)、EFO(Equivalent Filter Operations)和鲁棒性(Robustness)。其中,EFO为等效滤波,为了减小不同网络的编程语言和硬件配置对跟踪速度的影响,本文首先在600像素×600像素的图像上进行30×30的滤波运算,然后将跟踪算法处理每帧图像的时间除以滤波运算的时间,得到一个归一化的参数,即EFO,其可以比较客观地评价跟踪器的性能。鲁棒性数值为跟踪过程中的失败总次数。一个性能较优的跟踪器应该有较高的EAO、准确率和EFO,但鲁棒性分数应该较低。

在VOT2016中,将本文算法与SiamFC^[6]、KCF^[21]、SAMF^[22]和DAT^[23]4个主流跟踪算法进行对比,结果如表3所示。从表3可以看出,在VOT2016基准下,本文算法的EAO高出SiamFC算法0.8个百分点,准确率也提高了1.27个百分点,并且与其他3个主流算法KCF、SAMF、DAT相比,本文算法都有很大的性能提升。虽然在速度评价指标EFO上本文算法低于KCF算法和DAT算法,但也高于SiamFC算法和SAMF算法,达到实时跟踪的效果。在EAO指标上,本文算法取得最高值24.38%,高于KCF的19.35%和DAT的21.67%。在

鲁棒性方面,本文算法取得最小的鲁棒性值0.447,跟踪失败次数与其他算法相比最少。综上,本文算法EAO和准确率最高,鲁棒性能最好,在5种算法中具有较好的性能表现。

表3 VOT2016基准下的评估结果

Table 3 Evaluation results under VOT2016 benchmark

算法	EAO/%	Accuracy/%	EFO	Robustness
KCF ^[21]	19.35	50.84	15.74	0.569
SAMF ^[22]	18.51	49.36	3.65	0.587
DAT ^[23]	21.67	45.13	9.68	0.480
SiamFC ^[6]	23.58	53.20	9.21	0.461
本文算法	24.38	54.47	9.36	0.447

图4所示为5种跟踪算法的实验效果对比,左上角标号为测试视频序列中的图片帧数编号。从图4可以看出,在整个视频序列中本文算法一直保持平滑的跟踪效果,KCF算法和DAT算法的跟踪效果越来越差,SAMF算法最后完全丢失了跟踪目标。

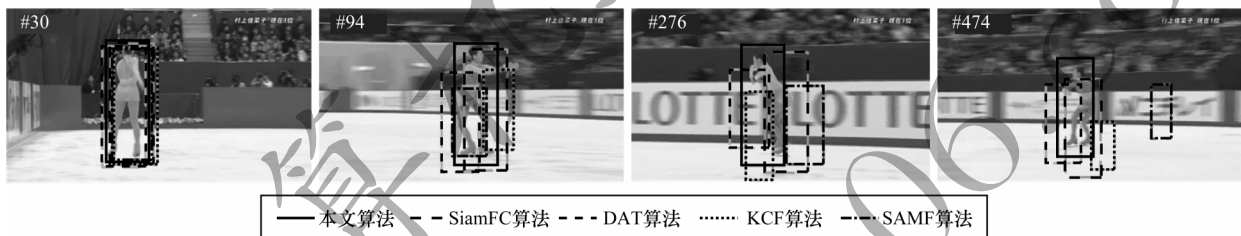


图4 5种算法的跟踪效果比较

Fig.4 Comparison of tracking effects of five algorithms

4 结束语

本文提出一种孪生卷积网络实时目标跟踪算法。构建基于非对称卷积模块的网络结构,通过非对称模块减少模型的参数量,使用三元组损失函数进行模型训练以提高算法精度。实验结果表明,该算法能够大幅降低模型大小并实现实时跟踪,且跟踪精度优于KCF、DAT、SAMF和SiamFC 4种算法。后续将引入再检测机制,结合目标检测与目标跟踪进一步提升算法的跟踪性能。

参考文献

- [1] GUAN Hao, CUI Xiangyang, AN Zhiyong. Advances on application of deep learning for video object tracking[J]. Acta Automatica Sinica, 2016, 42(6): 834-847. (in Chinese) 管皓, 薛向阳, 安志勇. 深度学习在视频目标跟踪中的应用进展与展望[J]. 自动化学报, 2016, 42(6): 834-847.
- [2] BOLME D, BEVERIDGE J R, DRAPER B A, et al. Visual object tracking using adaptive correlation filters [C]// Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2010: 2544-2550.
- [3] LIU Tianjian, QIU Lida, ZHANG Ning. Improved object

tracking algorithm based on mean shift [J]. Computer Engineering, 2015, 41(9): 281-285. (in Chinese)

刘天健, 邱立达, 张宁. 基于均值漂移的改进目标跟踪算法[J]. 计算机工程, 2015, 41(9): 281-285.

- [4] LEICHTER I. Mean shift trackers with cross-Bin metrics[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(4): 695-706.
- [5] WANG N, YEUNG D Y. Learning a deep compact image representation for visual tracking [EB/OL]. [2019-11-10]. <http://winsty.net/papers/dlt.pdf>.
- [6] BERTINETTO L, VALMADRE J, HENRIQUES J F, et al. Fully-convolutional Siamese networks for object tracking [M]//ELGAMMAL A M, HARWOOD D, DAVIS L S. Lecture notes in computer science. Berlin, Germany: Springer, 2016: 850-865.
- [7] SONG Y B, MA C, GONG L J, et al. CREST: convolutional residual learning for visual tracking [C]// Proceedings of 2017 IEEE International Conference on Computer Vision. Washington D. C. , USA: IEEE Press, 2017: 2574-2583.
- [8] WANG Qiang, TENG Zhu, XING Junliang, et al. Learning attentions: residual attentional Siamese network for high performance online visual tracking [C]// Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2018: 4854-4863.

- [9] NAM H, HAN B. Learning multi-domain convolutional neural networks for visual tracking[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA; IEEE Press, 2016; 4293-4302.
- [10] LI Bo, YAN Junjie, WU Wei, et al. High performance visual tracking with Siamese region proposal network[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA; IEEE Press, 2018; 8971-8980.
- [11] WANG Q, ZHANG L, BERTINETTO L C, et al. Fast online object tracking and segmentation: a unifying approach[EB/OL]. [2019-11-10]. <https://arxiv.org/abs/1812.05050>.
- [12] DENTON E, ZAREMBA W, BRUNA J, et al. Exploiting linear structure within convolutional networks for efficient evaluation[EB/OL]. [2019-11-10]. <https://cs.nyu.edu/~denton/papers/compressconv.pdf>.
- [13] PASZKE A, CHAURASIA A, KIM S, et al. ENet: a deep neural network architecture for real-time semantic segmentation[EB/OL]. [2019-11-10]. <https://arxiv.org/abs/1606.02147>.
- [14] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: a unified embedding for face recognition and clustering[C]//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA; IEEE Press, 2015; 815-823.
- [15] SONG H O, XIANG Y, JEGELKA S, et al. Deep metric learning via lifted structured feature embedding [C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA; IEEE Press, 2016; 4004-4012.
- [16] CHEN Qiaoyuan, CHEN Ying. TriHard loss based multi-task person re-identification[J]. Journal of Computer-Aided Design and Computer Graphics, 2019, 31(7): 1156-1165. (in Chinese)
- 陈巧媛, 陈莹. 基于困难样本三元组损失的多任务行人再识别[J]. 计算机辅助设计与图形学学报, 2019, 31(7): 1156-1165.
- [17] HUANG Lianghua, ZHAO Xin, HUANG Kaiqi. GOT-10K: a large high-diversity benchmark for generic object tracking in the wild [EB/OL]. [2019-11-10]. <https://arxiv.org/pdf/1810.11981.pdf>.
- [18] WU Y, LIM J, YANG M H. Online object tracking: a benchmark[C]//Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA; IEEE Press, 2013; 2411-2418.
- [19] KRISTAN M, LEONARDIS A, MATAS J, et al. The visual object tracking VOT2016 challenge results [C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2016; 236-256.
- [20] WANG Renhua, SHEN Jianyu, JIANG Min. Target tracking algorithm based on adaptive multi-model joint[J]. Computer Engineering, 2019, 45(8): 266-274. (in Chinese)
- 王任华, 沈剑宇, 蒋敏. 基于自适应多模型联合的目标跟踪算法[J]. 计算机工程, 2019, 45(8): 266-274.
- [21] 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.
- [22] LI Yang, ZHU Jianke. A scale adaptive kernel correlation filter tracker with feature integration [EB/OL]. [2019-11-10]. http://vigir.missouri.edu/~gdesouza/Research/Conference_CDs/ECCV_2014/workshops/w09/W9-07.pdf.
- [23] CEHOVIN I, LEONARDIS A, KRISTAN M. Robust visual tracking using template anchors [C]//Proceedings of 2016 IEEE Winter Conference on Applications of Computer Vision. Washington D. C. , USA; IEEE Press, 2016; 1-8.

编辑 吴云芳