



文章编号:1001-9081(2019)06-1652-05

DOI:10.11772/j.issn.1001-9081.2018112419

基于双重注意力孪生网络的实时视觉跟踪

杨 康¹, 宋慧慧^{2*}, 张开华¹

(1. 江苏省大数据分析技术重点实验室(南京信息工程大学), 南京 211800;
2. 大气环境与装备技术协同创新中心(南京信息工程大学), 南京 211800)
(* 通信作者电子邮箱 songhuihui@nuist.edu.cn)

摘要:为了解决全卷积孪生网络(SiamFC)跟踪算法在跟踪目标经历剧烈的外观变化时容易发生模型漂移从而导致跟踪失败的问题,提出了一种双重注意力机制孪生网络(DASiam)去调整网络模型并且不需要在线更新。首先,主干网络使用修改后表达能力更强的并适用于目标跟踪任务的VGG网络;然后,在网络的中间层加入一个新的双重注意力机制去动态地提取特征,这种机制由通道注意机制和空间注意机制组成,分别对特征图的通道维度和空间维度进行变换得到双重注意特征图;最后,通过融合两个注意机制的特征图进一步提升模型的表征能力。在三个具有挑战性的跟踪基准库即OTB2013、OTB100和2017年视觉目标跟踪库(VOT2017)实时挑战上进行实验,实验结果表明,以40 frame/s的速度运行时,所提算法在OTB2013和OTB100上的成功率指标比基准SiamFC分别高出3.5个百分点和3个百分点,并且在VOT2017实时挑战上面超过了2017年的冠军SiamFC,验证了所提出算法的有效性。

关键词:卷积神经网络;视觉跟踪;注意力机制;孪生网络

中图分类号: TP391.4 **文献标志码:**A

Real-time visual tracking based on dual attention siamese network

YANG Kang¹, SONG Huihui^{2*}, ZHANG Kaihua¹

(1. Jiangsu Key Laboratory of Big Data Analysis Technology
(Nanjing University of Information Science and Technology), Nanjing Jiangsu 211800, China;
2. Collaborative Innovation Center of Atmospheric Environment and Equipment Technology
(Nanjing University of Information Science and Technology), Nanjing Jiangsu 211800, China)

Abstract: In order to solve the problem that Fully-Convolutional Siamese network (SiamFC) tracking algorithm is prone to model drift and results in tracking failure when the tracking target suffers from dramatic appearance changes, a new Dual Attention Siamese network (DASiam) was proposed to adapt the network model without online updating. Firstly, a modified Visual Geometry Group (VGG) network which was more expressive and suitable for the target tracking task was used as the backbone network. Then, a novel dual attention mechanism was added to the middle layer of the network to dynamically extract features. This mechanism was consisted of a channel attention mechanism and a spatial attention mechanism. The channel dimension and the spatial dimension of the feature maps were transformed to obtain the double attention feature maps. Finally, the feature representation of the model was further improved by fusing the feature maps of the two attention mechanisms. The experiments were conducted on three challenging tracking benchmarks: OTB2013, OTB100 and 2017 Visual-Object-Tracking challenge (VOT2017) real-time challenges. The experimental results show that, running at the speed of 40 frame/s, the proposed algorithm has higher success rates on OTB2013 and OTB100 than the baseline SiamFC by the margin of 3.5 percentage points and 3 percentage points respectively, and surpass the 2017 champion SiamFC in the VOT2017 real-time challenge, verifying the effectiveness of the proposed algorithm.

Key words: convolutional neural network; visual tracking; attention mechanism; siamese network

0 引言

视觉目标跟踪在计算机视觉领域是一个基础性但充满挑战的研究方向,被应用于各种视觉领域,比如无人驾驶、人机交互和视频监控等。由于存在目标发生剧烈的外观变化、目标遮挡、光照变换等干扰因素,除此之外,还要考虑实时的因素,所以尽管最近几年目标跟踪算法研究取得了显著性的提升,但到目前为止仍然是一个极具挑战性的任务。

基于相关滤波的跟踪器可以通过一个循环矩阵在傅里叶域快速求解来实现快速目标跟踪,出现了很多速度快且简单的跟踪器^[1-5]。最近几年,深度卷积神经网络在计算机视觉领域取得了显著的成功,比如分类任务、目标检测等任务。所以也有很多研究者将深度学习应用到目标跟踪任务上去,其中取得突破性的且能够达到实时要求的算法就是全卷积孪生网络(Fully-Convolutional Siamese network, SiamFC)^[5], SiamFC把目标跟踪任务当作相似性匹配任务,即利用外部训练数据

收稿日期:2018-12-07;修回日期:2019-01-10;录用日期:2019-01-10。 基金项目:国家自然科学基金资助项目(61872189, 61876088);江苏省自然科学基金资助项目(BK20170040);江苏省研究生科研与实践创新计划项目(SJCX19_0311)。

作者简介:杨康(1993—),男,江苏徐州人,硕士研究生,主要研究方向:目标跟踪; 宋慧慧(1986—),女,山东聊城人,教授,博士,主要研究方向:遥感图像处理; 张开华(1983—),男,山东日照人,教授,博士,CCF会员,主要研究方向:图像分割、目标跟踪。



训练一个修改后的 AlexNet^[6]卷积网络作为通用的匹配函数,再把匹配函数作为目标跟踪的图像特征提取器,如果匹配函数能够学习更好的特征表达能力,那么对于提升跟踪器的性能是有帮助的。孪生实例搜索跟踪(Siamese Instance Search Tracking, SINT)^[7]将跟踪任务看作是一个验证任务并利用光流进一步提升性能表现,但是速度只有4 frame/s,很难应用到现实场景中;提前停止跟踪(Early-Stopping Tracker, EAST)^[8]主要判断低级的特征,如果能够跟踪到目标时就停止特征提取进行加速;相关滤波网络(Correlation Filter Network, CFNet)跟踪^[9]将相关滤波作为一个可微的层加入低层的网络特征中去学习目标变换,大大降低了网络参数量的同时仍然保持很好的跟踪性能。动态孪生网跟踪(Dynamic Siamese Network, DSiam)^[10]尝试在线学习目标的外观变化去进一步提升孪生网络的表征能力。

尽管基于孪生网络的跟踪算法取得了显著的进步,但是这种孪生网络框架仍然有一些问题没有解决。首先,用于孪生网络的框架一般都是比较浅层的 AlexNet 网络,在深度学习任务中,已经证明了更深的网络具有更强的信息表征能力^[11];其次,在目标发生剧烈的变化时,由于孪生网络缺少动态的调节模型机制,只能等价地对待每一个特征图和特征空间,没有重点关注的目标区域,这样限制了模型丰富的表征能力。

针对基于孪生网络的跟踪器出现的上述问题,本文在 SiamFC 的跟踪算法框架之下,把特征提取网络换成了修改过的且适用于目标跟踪任务的 VGG(Visual Geometry Group)^[12]网络,在此基础之上,为了进一步增强网络模型的判别能力,提出了一种新的双重注意力机制去调节模型。最后为了验证该算法的有效性,在三个具有挑战性的视频库上进行详尽的实验,并与几个经典的跟踪算法进行比较,实验结果表明所提方法得到了很有竞争力的结果。

1 双重注意力孪生网络算法

为了实现高效的视觉跟踪任务,本文提出了一种新的基于双重注意孪生(Dual Attention Siamese network, DASiam)网络的视觉跟踪算法,如图1所示。该算法由一个修改后的深度卷积神经网络VGG和一个双重注意模块组成,其中双重注意模块包括通道注意模块和空间注意模块,最后将提取到的模板图像和搜索图像的高维语义信息特征进行相关操作得到最终的目标位置。

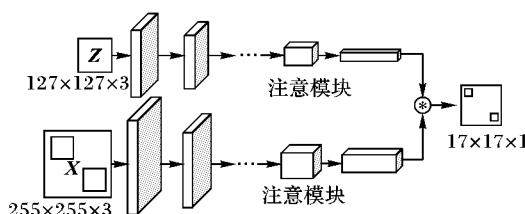


图1 DASiam 算法原理图

Fig. 1 Schematic diagram of DASiam algorithm

1.1 基于孪生网络的跟踪算法

最近几年在目标跟踪领域的开创性工作是全卷积孪生网络(SiamFC)目标跟踪算法,如图1所示,孪生网络的输入是

从视频第一帧(目标)和后续帧进行裁剪的一堆图像,分别用 Z 和 X 表示,其中 $Z \in \mathbf{R}^{W_t \times H_t \times 3}$ 且 $X \in \mathbf{R}^{W_s \times H_s \times 3}$,然后通过一个离线训练的匹配函数 $F(Z, X)$ 在模板图像 Z 和搜索图像 X 进行相关运算得到一个相似性响应得分图,响应得分最大的位置就是新的目标位置,其中用于特征提取的卷积网络关于搜索图像 X 是全卷积的,这样就可以输入不同尺度大小的搜索图像以便选择合适的尺度作为新的预测框。相似性响应得分图可以由式(1)得到:

$$F(Z, X; \theta) = \varphi(Z; \theta) * \varphi(X; \theta) + b \cdot 1 \quad (1)$$

其中:“*”表示相关运算; θ 是网络的参数且 b 是一个偏置项。最后的输出是一个定义在有限的网格区域中的具有空间结构的得分图而不是一个链式向量或者标量,其中网络中最优的参数是采用随机梯度下降从头开始训练而得到的,具体而言,从视频目标检测数据集(ImageNet Large Scale Visual Recognition Challenge, ILSVRC)^[13]获得大量的图像对 (Z_i, X_i) ,给定相应的标签响应图 $Y_i \in \{+1, -1\}$,然后最小化如下的逻辑回归损失函数 $L(\cdot)$:

$$\arg \min_{\theta} \frac{1}{N} \sum_{i=1}^N \{L(Y_i, F(Z_i, X_i; \theta))\} \quad (2)$$

其中, N 表示训练样本数。

虽然SiamFC取得了很好的结果,但是与现有的相关滤波跟踪器的结果有着很大的差距,这是因为SiamFC用于特征提取的全卷积网络是使用修改后的AlexNet,由于AlexNet层数较浅,学到的特征表征能力有限,当跟踪目标发生剧烈形变时模型容易发生漂移,导致跟踪失败。本文采用更深的修改后的适用于目标跟踪任务的VGG网络模型作为新的特征提取网络,并在网络中加入双重注意力机制调整模型的特征提取过程,进而选择性地强调有用的信息而抑制不太有用的信息,而不是等价地对待所有的特征信息。

1.2 双重注意力孪生网络框架

图1展示了本文算法的基础框架,由修改后的VGG网络作为主干网络,除了最后一个卷积(Convolutional, Conv)层,每一层卷积之后立即加入批归一化(Batch Normalization, BN)层,然后再经过非线性激活函数(Rectified Linear Unit, ReLU)层,没有填充,并且在网络的第10层后面加入一个注意力调节机制,具体的网络参数如表1所示。由于深度卷积网络中高语义信息对于目标的外观变化具有很强的鲁棒性,但是当出现相似性目标时,由于高级语义信息缺少判别性,就容易导致模型出现漂移。所以为了增强网络的判别能力,在网络的中间层加入一个动态的特征调节机制,这个机制由双重注意力机制实现,包括通道注意机制和空间注意机制,在后面将详细介绍双重注意力机制算法,所有的网络参数在训练完成后都是固定的,不需要在线微调从而满足实时性的要求。

1.3 双重注意力机制算法

注意力机制在图像领域取得了很大的成功,因为它参考了人类的一个习惯:当我们看到一张图片的时候并不是一次性能看到所有的信息,而是仅仅关注某个被选定的位置,然后再向四周蔓延。神经网络在处理图像的时候,每次网络的关注点可能只是图像中的某个小部分,因此如果能在网络模型关注图像某个部分时都能够强调这个部分的话,这样对于模



型的特征表达能力是有提升的。为此,本文设计了一种适用于目标跟踪任务的双重注意力机制,当目标发生剧烈形变的时候,网络能够通过注意力机制关注目标的主要部分,从而提升模型的鲁棒性。

表1 双重注意力孪生网络参数
Tab. 1 Dual attention siamese network parameters

卷积层	卷积核大小 (w, h, in, out)	步长	模板图像 127×127	搜索图像 255×255
Conv1	$3 \times 3 \times 3 \times 96$	1	125×125	253×253
Conv2	$3 \times 3 \times 96 \times 96$	1	123×123	251×251
Pool1	3×3	2	61×61	125×25
Conv3	$3 \times 3 \times 96 \times 128$	1	59×59	123×123
Conv4	$3 \times 3 \times 128 \times 128$	1	57×57	121×121
Pool2	3×3	2	28×28	60×60
Conv5	$3 \times 3 \times 128 \times 256$	1	26×26	58×58
Conv6	$3 \times 3 \times 256 \times 256$	1	24×24	56×56
Conv7	$3 \times 3 \times 256 \times 256$	1	22×22	54×54
Conv8	$3 \times 3 \times 256 \times 256$	1	20×20	52×52
Attention	—	—	—	—
Pool3	2×2	2	10×10	26×26
Conv9	$3 \times 3 \times 256 \times 256$	1	8×8	24×24
Conv10	$3 \times 3 \times 256 \times 512$	1	6×6	22×22

本文所提出的双重注意力机制分别对通道维度和空间维度上的语义特征进行建模,如图2所示。通道依赖性由通道注意模块得到,由于高级特征图的每个通道可以被视为对特定类的响应,并且不同的语义特征响应彼此之间相互关联,利用通道特征图之间的相互依赖关系,如可以通过增强特征图之间的相互依赖关系去提高特征的表达能力;对于空间注意模块,通过引入自我注意机制来建立特征图中任意两个位置之间的联系,对于某个位置的特征可以通过加权求和所有位置的特征信息来更新,最后把通道注意特征与空间位置特征进行元素相加来进一步加强网络的特征表征能力。

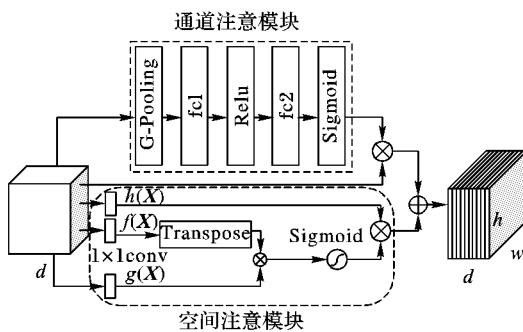


图2 双重注意力模块

Fig. 2 Dual attention module

具体而言,通道注意模块是以特征图为单位,对每一个通道都配一个权值,如图2中通道注意模块所示。它由一个多层次感知器实现,输入特征 $M \in \mathbb{R}^{w \times h \times d}$ 首先经过一个全局平均池化层得到一个特征向量 $\mathbf{m} = (m_1, m_2, \dots, m_d)$ 作为全连接层的输入,其中 $m_i \in \mathbb{R}$, 经过一个隐藏层,再经过一个非线性激活函数 Sigmoid 层得到输出向量 $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_d)$, 其中 $\beta_i \in \mathbb{R}$, 然后将激活向量 $\boldsymbol{\beta}$ 与输入特征 M 进行元素相乘,最终生成通道注意特征图 $U \in \mathbb{R}^{w \times h \times d}$ 。

对于空间注意模块来说,它是以特征图中的每个像素点

为单位的,对特征图中的每个像素点都配一个权重,以便建立空间信息的结构依赖关系去增强模型的特征表达能力。如图2中空间注意模块所示,给定一个输入特征图 $A \in \mathbb{R}^{d \times w \times H}$,首先经过三个变换函数 h, f, g 得到变换后的特征图 B, C, D ,其中 $\{B, C, D\} \in \mathbb{R}^{d \times w \times H}$, 变换函数包括 1×1 的卷积层, BN 层和 ReLU 层,然后把 C, D 变换为 $\mathbb{R}^{d \times WH}$,用 C 的转置乘 D ,再经过一个 Sigmoid 激活函数计算得到空间注意图,计算式为:

$$s_{i,j} = \exp(\mathbf{C}_i^T \cdot \mathbf{D}_j) / \sum_{j=1}^{WH} \exp(\mathbf{C}_i^T \cdot \mathbf{D}_j) \quad (3)$$

其中 $s_{i,j}$ 表示第 i 个区域与第 j 个区域之间的权重,与此同时,特征图 B 也变换为 $\mathbb{R}^{d \times WH}$, 然后再将 B 与 S 的转置进行矩阵相乘并且将得到的结果重新变换为 $\mathbb{R}^{d \times w \times H}$, 由式(4)计算得到最终的空间注意特征输出:

$$\mathbf{V}_i = \lambda \sum_{j=1}^{WH} s_{i,j} \mathbf{B}_j + \mathbf{A}_i \quad (4)$$

其中, λ 是一个可学习的变量因子, 初始化为 0, 然后渐渐赋予更大的权重,这样可以允许网络首先学习简单的任务然后再慢慢增加学习任务的复杂度。

最终双重注意力机制的输出是将通道注意特征和空间注意特征进行元素相加,以便获得更好的特征表征信息。

$$\mathbf{O}_i = \mathbf{U}_i + \mathbf{V}_i; \quad i = 1, 2, \dots, d \quad (5)$$

1.4 数据集和网络训练细节

本文的网络是在视频目标检测数据集 ILSVRC 上使用彩色图像离线训练的,其中包含了 4500 个视频序列且有大约有 130 万个人工标注的边界框,最近被广泛应用在跟踪领域。采用动量为 0.9 的随机梯度下降最优化网络并设置权重衰减为 0.0005, 学习率以指数衰减方式从 10^{-2} 到 10^{-5} , 训练周期大约为 65 个周期且每次小批量训练样本数为 16。最后为了解决尺度变换问题,在搜索图像上采用三个不同的尺度缩放因子 $\{q^s | q = 1, 0.25, -1, 0, 1\}$ 去搜索图像,通过一个因子为 0.35 的线性插值去更新当前目标的尺度。

本文所提出的网络模型是在 TensorFlow 1.4.1 框架^[14]上训练的,且实验评估是在一台配置为英特尔 i7-8700K CPU 和显卡 GTX1080Ti 电脑上进行的,平均帧率是 40 frame/s。

2 实验结果及分析

为了评估本文所提算法的有效性,在三个具有挑战性并且被广泛使用的视频基准库上进行实验,分别是: OTB2013^[15]、OTB100^[16]、2017 年视觉目标跟踪库(2017 Visual-Object-Tracking challenge, VOT2017)^[17] 实时挑战,并且与基准算法 SiamFC 和几个经典的算法进行对比实验。

在本文实验中,选择了三个具有代表性的跟踪器进行对比,包括本文算法基准 SiamFC 和经典的相关滤波算法判别尺度空间跟踪器(Discriminative Scale Space Tracker, DSST)^[18]、核化相关滤波跟踪(Kernelized Correlation Filter, KCF)^[1]、空间正则判别相关滤波跟踪(Spatially Regularized Discriminative Correlation Filter, SRDCF)^[19]。

2.1 在 OTB2013 和 OTB100 上的评估

OTB2013 和 OTB100 是视觉跟踪领域广泛使用的基准库, 分别包含了 51 个和 100 个人工标注的视频帧, 并且包含



了11个不同的属性,例如尺度变换、光照变化、平面内旋转、快速运动等。算法的性能由两个性能指标衡量:成功率和精确率。成功率表明重合率得分超过某个阈值的帧的个数占视频总帧数的百分比,精确率表明了中心位置误差在一个特定阈值内的视频帧数占总帧数的百分比。重合率计算如下:

$$os = |\mathbf{G}_{rec} \cap \mathbf{P}_{rec}| / |\mathbf{G}_{rec} \cup \mathbf{P}_{rec}| \quad (6)$$

其中, \mathbf{G}_{rec} 、 \mathbf{P}_{rec} 分别表示人工标定的边界框和跟踪器预测的边界框。

OTB2013 和 OTB100 基准库上不同算法的成功率对比结果如图3所示。从图3(a)可以看出,在视频数据集 OTB2013 上,本文算法 DASiam 能够排在第一并且比基准算法 SiamFC 提高了3.5个百分点,显著地提高了跟踪性能;由图3(b)可以看出,在更具有挑战性的100个视频数据集 OTB100 上,DASiam 也比 SiamFC 高出了3个百分点,很好地验证了本文跟踪算法的有效性。

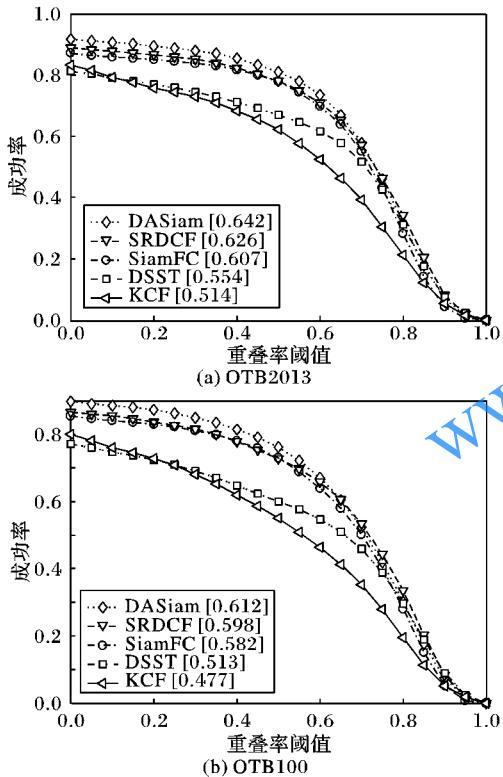


图3 不同视频基准库上成功率对比

Fig. 3 Comparison of success rates on different video benchmarks

2.2 基于 OTB100 属性的分析

本文在 OTB100 上对所提出的跟踪器进行了11种不同属性的对比分析实验。图4(a)、4(b)分别展示了当目标经历了运动模糊和平面内旋转两种属性的成功率,这两种属性表明了跟踪的目标经历了比较大的外观变化,与给定的第一帧的目标外观变化差别较大。由图4可以看出,在运动模糊的属性下本文算法取得了62.4%的得分,比基准算法 SiamFC 高出7.4个百分点;同时,本文算法在平面内旋转的属性下也取得了较好的表现。在目标经历了运动模糊或者旋转导致目标外观发生变化的时候,SiamFC 的跟踪成功率得分比较低,表明该算法的鲁棒性较低;而本文的 DASiam 加入了双重注意力机制能够很好地建立通道和空间的联系,充分利用目标

的有用信息而抑制周围的干扰因素,从而提升了算法的鲁棒性,并且充分利用深度网络的优势进一步提取表达能力更强的特征。

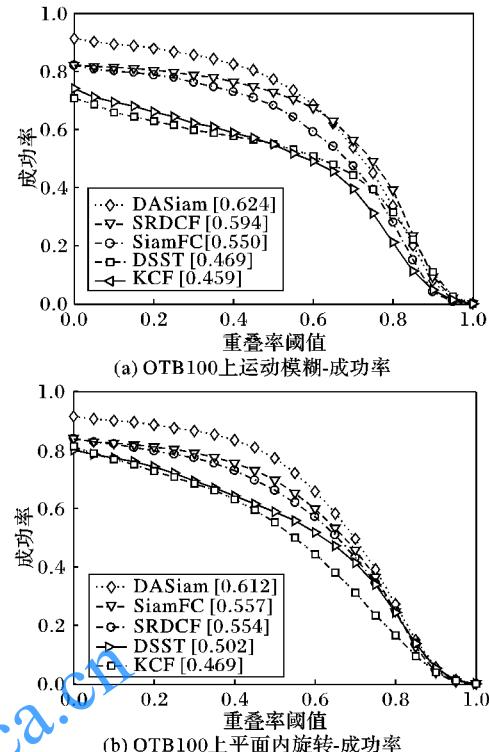


Fig. 4 Comparison of success rates of different attributes on OTB100

2.3 在 VOT2017 实时挑战上的结果

在 VOT2017 数据库中包含了60个更精细的人工标注的视频序列并且更具有挑战性,最近几年在跟踪领域中也被广泛采用,除此之外,VOT2017 还包含了一项新的实时实验,要求所有的跟踪器必须以超过实时的25 frame/s的速度处理视频流,这就意味着跟踪器如果达不到实时,评估器将以上一帧的预测结果作为当前帧的跟踪结果,这就很容易导致跟踪器跟踪失败。图5给出了本文算法 DASiam 和其他5个实时的跟踪器在 VOT2017 实时实验上的排名,其中基准 SiamFC 是2017年实时挑战赛上的冠军。

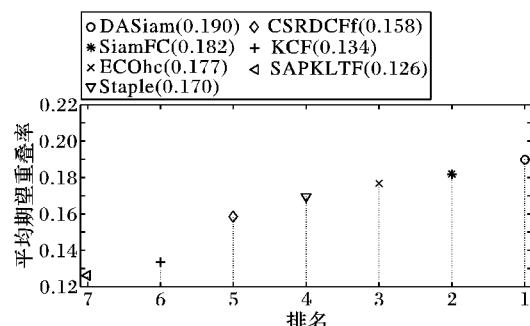


图5 VOT2017 实时平均期望重叠率排名

Fig. 5 Average expected overlapping ranking for VOT2017

由图5可以看出,本文算法的性能已经超过了 SiamFC 排到了第一,具体而言超过基准 SiamFC 大约 0.8 个百分点,更好地验证了本文的双重注意力机制孪生网络能够很好地适用



于基于孪生网络的跟踪器。

3 结语

本文在全卷积孪生网络(SiamFC)跟踪的基础上改进了用于特征提取的卷积神经网络,提出了双重注意力机制孪生网络跟踪器(DASiam),通过在修改后的VGG网络中嵌入了通道注意模块和空间注意模块提升网络模型的判别能力,去解决目标外观变化等问题。本文方法能够在跟踪标准测试集OTB2013和OTB100上取得很有竞争力的实验结果,在VOT2017实时挑战上的性能表现甚至超过了2017年实时的冠军SiamFC,表明本文方法能够在实际场景中,如无人驾驶、智能安防等,可以实现更好的跟踪效果以满足实际要求。但是,本文方法对于强烈光照变化、尺度变化较大等其他干扰因素出现时,跟踪结果不太理想,接下来将针对强烈光照变化、尺度变化较大等问题进行进一步研究改进。

参考文献 (References)

- [1] 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.
- [2] 熊昌镇, 车满强, 王润玲. 基于稀疏卷积特征和相关滤波的实时视觉跟踪算法[J]. *计算机应用*, 2018, 38(8): 2175 – 2179, 2223. (XIONG C Z, CHE M Q, WANG R L. Real-time visual tracking algorithm based on correlation filters and sparse convolutional features [J]. *Journal of Computer Applications*, 2018, 38(8): 2175 – 2179, 2223.)
- [3] 樊佳庆, 宋慧慧, 张开华. 通道稳定性加权补充学习的实时视觉跟踪算法[J]. *计算机应用*, 2018, 38(6): 1751 – 1754. (FAN J Q, SONG H H, ZHANG K H. Real-time visual tracking algorithm via channel stability weighted complementary learning [J]. *Journal of Computer Applications*, 2018, 38(6): 1751 – 1754.)
- [4] 朱明敏, 胡茂海. 基于相关滤波器的长时视觉目标跟踪方法[J]. *计算机应用*, 2017, 37(5): 1466 – 1470. (ZHU M M, HU M H. Long-term visual object tracking algorithm based on correlation filter [J]. *Journal of Computer Applications*, 2017, 37(5): 1466 – 1470.)
- [5] BERTINETTO L, VALMADRE J, HENRIQUES J F, et al. Fully-convolutional Siamese networks for object tracking [C]// ECCV 2016: Proceedings of the 2016 European Conference on Computer Vision, LNCS 9914. Cham: Springer, 2016: 850 – 865.
- [6] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C]// NIPS 2012: Proceedings of the 25th International Conference on Neural Information Processing Systems. North Miami Beach, FL: Curran Associates Inc., 2012: 1097 – 1105.
- [7] TAO R, GAVVES E, SMEULDERS A W M. Siamese instance search for tracking [C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 1420 – 1429.
- [8] HUANG C, LUCEY S, RAMANAN D. Learning policies for adaptive tracking with deep feature cascades [C]// Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2017: 105 – 114.
- [9] 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. Piscataway, NJ: IEEE, 2017: 5000 – 5008.
- [10] GUO Q, FENG W, ZHOU C, et al. Learning dynamic Siamese network for visual object tracking [C]// Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2017: 1781 – 1789.
- [11] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 770 – 778.
- [12] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. [2019-10-16]. <http://www.cs.cmu.edu/~jeanoh/16-785/papers/simonyan-iclr2015-vgg.pdf>.
- [13] 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.
- [14] ABADI M, BARHAM P, CHEN J M, et al. TensorFlow: a system for large-scale machine learning [C]// OSDI 2016: Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation. Berkeley, CA: USENIX Association, 2016: 265 – 283.
- [15] WU Y, LIM J, YANG M H. Online object tracking: a benchmark [C]// CVPR 2013: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2013: 2411 – 2418.
- [16] WU Y, LIM J, YANG M H. Object tracking benchmark [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1834 – 1848.
- [17] KRISTAN M, LEONARDIS A, MATAS J, et al. The visual object tracking VOT2017 challenge results [C]// ICCVW 2017: Proceedings of the 2017 IEEE International Conference on Computer Vision Workshop. Piscataway, NJ: IEEE, 2017: 1949 – 1972.
- [18] DANELLJAN M, HÄGER G, KHAN F, et al. Accurate scale estimation for robust visual tracking [C]// Proceedings of the 2014 British Machine Vision Conference. Durham, UK: BMVA Press, 2014: 65.1 – 65.11.
- [19] DANELLJAN M, HÄGER G, KHAN F S, et al. Learning spatially regularized correlation filters for visual tracking [C]// ICCV 2015: Proceedings of the 2015 IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2015: 4310 – 4318.

This work is partially supported by the National Natural Science Foundation of China (61872189, 61876088), the Natural Science Foundation of Jiangsu Province (BK20170040), the Postgraduate Research & Practice Innovation Program of Jiangsu Province (SJCX19_0311).

YANG Kang, born in 1993, M. S. candidate. His research interests include object tracking.

SONG Huihui, born in 1986, Ph. D., professor. Her research interests include remote sensing image processing.

ZHANG Kaihua, born in 1983, Ph. D., professor. His research interests include image segmentation, object tracking.