

DOI: 10.19650/j.cnki.cjsi.2006906

基于联合注意力孪生网络目标跟踪算法*

杨梅¹, 贾旭¹, 殷浩东¹, 孙福明²

(1. 辽宁工业大学电子与信息工程学院 锦州 121001; 2. 大连民族大学信息与通信工程学院 大连 116600)

摘要:为改进在发生形变、尺度变化及相似目标等多种干扰因素时视频中运动目标跟踪精度,提出了一种联合注意力的孪生网络模型。首先,采用一种轻量级网络 MobileNetV3 作为主干网络对目标进行特征提取;然后,为提高模型对于目标关键特征的关注度,提出了通道联合空间注意力与孪生网络结合的模型结构;最后,对基于注意力模块与非注意力模块的特征向量互相关结果进行加权融合获得响应图,并利用该响应图获得目标跟踪结果。实验结果表明,所提算法在 OTB50 与 OTB100 数据集上能够获得较好的跟踪效果,两个数据集平均精确率和成功率达到 78.5% 和 58.3%。此外,当存在形变、尺度变化及相似目标等不合作因素时,所提算法仍能取得较好的跟踪效果,从而表明该算法具有良好的鲁棒性。

关键词:目标跟踪;孪生网络;联合注意力;MobileNetV3

中图分类号: TP391.4 TH865 **文献标识码:** A **国家标准学科分类代码:** 520.2040

Object tracking algorithm based on siamese network with combined attention

Yang Mei¹, Jia Xu¹, Yin Haodong¹, Sun Fuming²

(1. School of Electronics and Information Engineering, Liaoning University of Technology, Jinzhou 121001, China;

2. School of Information and Communication Engineering, Dalian Minzu University, Dalian 116600, China)

Abstract: In order to improve the tracking accuracy of moving targets in video when various interference factors such as deformation, scale variation and similar targets occur, a siamese network model with combined attention is proposed. Firstly, a lightweight network, i. e., MobileNetV3, is adopted as the backbone network to extract object feature. Then, in order to improve the attention of the model to the key features of the target, a model structure combining channel combined spatial attention and siamese network is proposed. Finally, through weighting and fusing the cross-correlation results of the feature vectors of attention module and non-attention module, the response map can be obtained, which can be used to obtain the tracking result. Experiment results show that the proposed algorithm can achieve good tracking effect on the OTB50 and OTB100 datasets, the average accuracy and success rate for the two datasets reach 78.5% and 58.3%, respectively. In addition, when multiple uncooperative factors, such as deformation, scale variation and similar targets exist, the proposed algorithm can still achieve good tracking effect, which shows that the proposed algorithm has good robustness.

Keywords: object tracking; siamese network; combined attention; MobileNetV3

0 引言

目标跟踪技术在智能视频监控、无人驾驶和人机交互^[1-2]等领域得到了广泛应用,在许多现实场景中,所需跟踪的目标主体通常具有形变、被遮挡、或相似物体干扰的特点,一定程度上影响跟踪算法的性能。因此,提出一

种准确且具有较强鲁棒性的目标跟踪算法将具有重要意义。

目标跟踪技术主要包括以下两类:1)是传统的滤波方法,这类算法均以上一帧的目标位置作为初始位置,进行迭代搜索目标,如 mean shift^[3]、粒子滤波^[4]等算法。随后相关滤波类的方法^[5-7]也被广泛应用到目标跟踪中,相关滤波类方法结合深度特征的目标跟踪算法^[8-10]提升了跟

收稿日期:2020-09-14 Received Date: 2020-09-14

* 基金项目:国家自然科学基金(61572244)、辽宁省自然科学基金计划指导计划项目(2019-ZD-0700)资助

踪性能。2) 是基于深度学习的方法, 基于 AlexNet^[11], VGGNet^[12], 残差神经网络^[13] (residual network, ResNet) 等经典卷积网络的目标跟踪算法在现有的测试数据集上均有较好的表现, 如多域卷积神经网络^[14] 和 SANet^[15]。近年来, 以相似性学习为基础的孪生网络被引入到目标跟踪领域, 如 Siam-FC^[16]、CGACD^[17] 算法, 以及将区域推荐网络 (region proposal network, RPN)^[18] 引入到孪生网络的目标跟踪算法^[19-20]。以上方法在不同程度下均取得一定的效果, 但由于传统的滤波方法大部分都是取一定面积的搜索区域进行操作, 这种限定区域的搜索方法具有一定的局限性, 跟踪性能受到一定影响。此外, 基于深度学习的方法不断被改进, 但是在复杂场景下无法适应目标和背景信息的变化, 当目标发生形变或者出现严重的干扰时, 对算法的鲁棒性有一定的影响。

针对以上问题, 本文提出一种基于联合注意力的孪生网络目标跟踪模型, 实现对目标连续的跟踪。该方法的创新性主要体现为以下 3 个方面: 1) 将改进后的轻量级网络作为特征提取骨干网络, 消除了填充 (padding) 操作对于跟踪模型带来的影响, 以及调整其网络结构适应

于跟踪算法; 2) 为获得更有利于模型的特征, 设计全局通道联合空间注意力模块 (global channel combined spatial attention, GCCSA), 对提取的特征进一步操作, 增强网络模型的判别能力; 3) 相似度得分响应图的生成, 使用不同部分的特征向量做互相关结果做加权平均, 提高模型的性能。

1 本文算法

为了提高算法对于形状变化、尺度变化及相似目标等多种不合作因素的鲁棒性, 这里利用了轻量级网络和注意力模块在处理该问题上的优势, 提出了一种基于联合注意力孪生网络的跟踪模型, 该模型框架如图 1 所示。图 1 中, Z 与 X 分别表示模板分支图像与检测分支图像, 经过 MobileNetV3^[21] 处理后得到特征图; 然后, 将特征图输入到 GCCSA 模块中, 进行特征修饰, 以增强模型的判别能力; 最后, 为了提高跟踪的准确性, 在融合响应分数图过程中, 对各分支的注意力模块和非注意力模块互相关结果进行权值为 μ 的加权融合, 得到响应分数图, 从而得到跟踪结果。

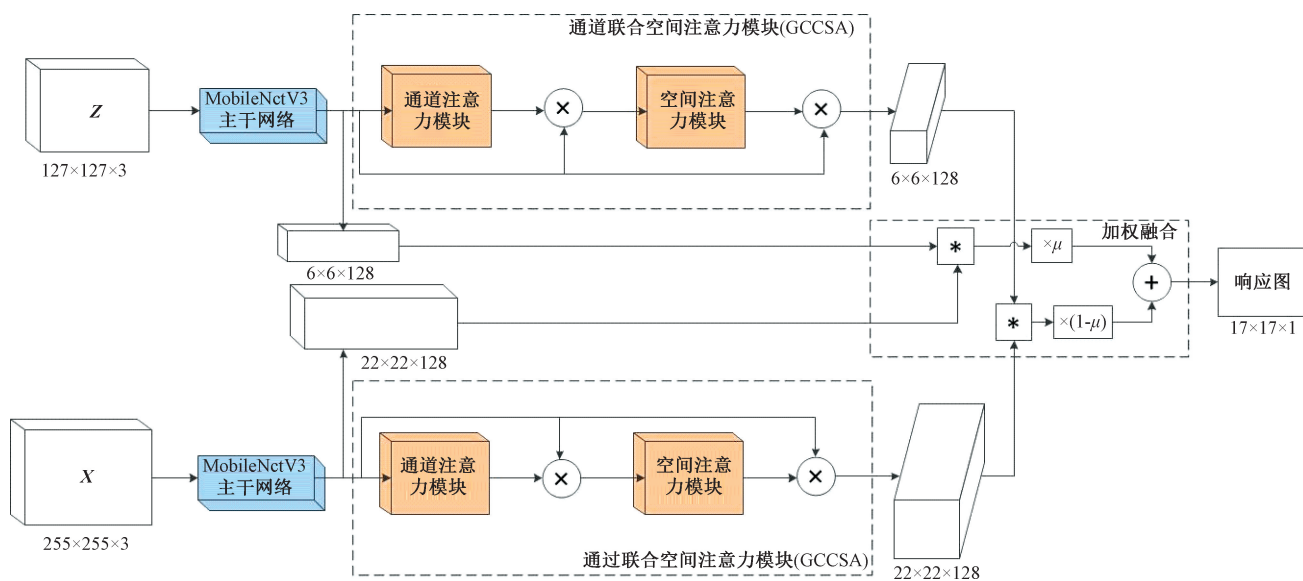


图 1 模型框架

Fig. 1 Framework of the model

1.1 全卷积孪生网络架构

所提的跟踪模型采用了孪生网络思想, 通过相似性学习的方法来解决目标追踪的问题, 即对比视频的前一帧模板图像与后一帧搜索图像, 通过相似度度量函数得到相似度得分 E_w , 分数越高说明两个图像越相似, 如图 2 所示。

整个过程可以看做是模板图像 Z 和搜索图像 X 的函

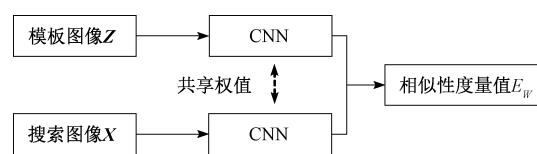


图 2 孪生网络结构

Fig. 2 Siamese network structure

数输出进行卷积运算,可表示为:

$$f(\mathbf{Z}, \mathbf{X}) = \Phi(\mathbf{Z}) * \Phi(\mathbf{X}) + b \quad (1)$$

式中: $b \in \mathbf{R}$ 表示偏置值; $\Phi(\mathbf{Z})$ 和 $\Phi(\mathbf{X})$ 为 $\mathbf{Z}$ 和 $\mathbf{X}$ 经过特征提取网络得到的特征图; $*$ 为卷积运算, $f(\mathbf{Z}, \mathbf{X})$ 为相似性得分图。

训练阶段采用判别方法,逻辑损失函数为:

$$l(y, v) = \log(1 + \exp(-yv)) \quad (2)$$

式中: v 是分数图中的真实得分; $y \in \{+1, -1\}$ 是单个样本的标签,该标签的生成表达式为:

$$y(u) = \begin{cases} +1, & k \| u - c \| \leq R \\ -1, & \text{其他} \end{cases} \quad (3)$$

式中: k 为网络的步长; c 为目标中心; u 为分数图中每个点的位置; R 是 c 以为为中心的圆的半径。

为了解决训练样本不均衡的问题,全卷积孪生网络采用的是分数图中每个点损失的均值。

1.2 主干网络 MobileNetV3

MobileNetV3 是一种参数少、复杂度较小的轻量级网络,在提出的目标跟踪模型中作为孪生网络的特征提取主干网络。为使 MobileNetV3 能够有效地解决目标跟踪问题,对该网络从以下 3 个方面做出改进。

1) MobileNetV3 原网络是针对分类和检测任务而提出,而对于跟踪任务,由于孪生网络的输入图像尺寸较小,适当减少网络步长,可以较有效地保留深层特征图中足够多的特征信息,使得网络适应于跟踪任务。因此删除 2 个步长为 2 的卷积层,将网络的总步长由 32 改为 8。

2) padding 会使特征图中的最大响应点会发生偏移,在做相似性度量时,得到的最大响应值也会发生很大的偏差。为了减少 padding 对跟踪效果的影响,用 3×3 卷积核替换原网络的 5×5 卷积核,然后通过剪裁操作除去特征图最外层的特征。

3) 原网络使用步长为 2 的卷积层将特征图分辨率降低为原来的 $1/2$,容易丢失图像边缘信息,不利于跟踪。因此本文网络将所有 bneck 的 3×3 卷积的步长改为 1,并使用最大池化层来降低特征图分辨率。为了增加网络的特征维度,得到更多可利用的特征信息,使用 1×1 卷积核调节网络通道数,同时,使之与 SiamFC 中原来的 Alexnet 通道数保持一致。

具体网络结构如表 1 所示。

1.3 通道联合空间注意力模块

人类视觉注意力机制被应用到目标跟踪领域后^[22],提高了计算机视觉处理目标信息的效率和准确率,该机制使网络关注具有重要信息的局部区域,抑制背景中无用的信息。

为了充分利用特征图中通道和空间的重要信息,卷积块注意力模块^[23](convolutional block attention module,

表 1 MobileNetV3 网络结构

Table 1 MobileNetV3 network structure

输入	卷积层	膨胀参数	输出维度	SE	最大值池化	非线性激活函数	步长
255×255×3	conv2d, 3×3	-	16	-	√	HS	1
112×112×16	bneck, 3×3	16	16	√	-	RE	1
110×110×16	bneck, 3×3	72	24	-	-	RE	1
108×108×24	bneck, 3×3	96	40	-	√	RE	1
54×54×40	bneck, 3×3	240	40	√	-	RE	1
52×52×40	bneck, 3×3	120	48	√	-	HS	1
50×50×48	bneck, 3×3	144	48	√	-	HS	1
48×48×48	bneck, 3×3	288	96	√	√	HS	1
24×24×96	bneck, 3×3	576	96	√	-	HS	1
22×22×96	bneck, 3×3	576	96	√	-	HS	1
22×22×128	conv2d, 1×1	-	256	-	-	HS	1

CBAM) 注意力模块从这两个维度进行独立特征的计算。在目标跟踪过程中,目标形变较大或者被环境中相似物干扰时,会导致目标跟踪效果不佳,这里通过非局部注意(non-local attention)对视觉场景进行全局理解,捕获长距离依赖关系,从而对于目标形态的变化更为敏感。CBAM 模型中的通道注意力,使用全局上下文对不同通道分配权重,调整通道的依赖程度,这种注意力方式并没有充分考虑到全局上下文信息,此外,目标跟踪是一种实时的任务,non-local attention 带来的计算量是不可忽视的。因此,本文结合孪生网络的结构特点,参考轻量级通用模块 CBAM,提出 GCCSA 模型,如图 3 所示。

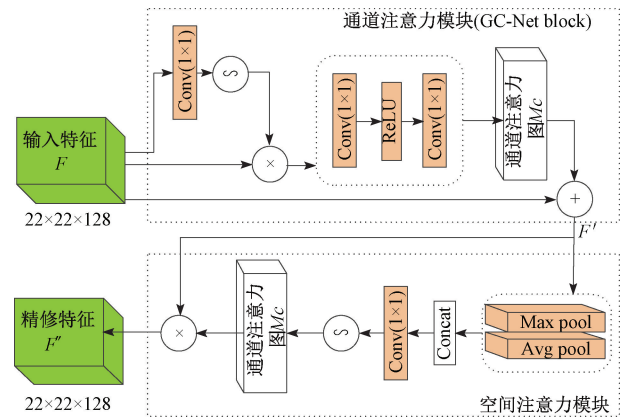


图 3 全局通道联合空间注意力模块 (GCCSA 模块)

Fig. 3 Global channel combined spatial attention module (GCCSA module)

GCCSA 模型的描述如下:

1) GCCSA 模型的输入为 MobileNetV3 的输出,即模

板图像特征与搜索图像特征;

2)在通道注意力模块中引入 GC-net block 作为通道注意力模块,集成到 MobileNetV3 网络中,使网络充分考虑到全局上下文信息;

3)GC-net block:首先,中间特征图 F 作为输入,做 1×1 卷积和 softmax 激活函数得到的通道注意力权重与输入做乘积得到全局上下文特征;然后,通过 1×1 卷积做依赖关系转换,得到注意力图 $M_c \in \mathbf{R}^{C \times 1 \times 1}$;最后对特征图每一个位置加上全局上下文信息,实现特征融合,得到 F' ;

4)空间注意力模块:首先, F' 作为输入,通过平均池化 (avg pool) 和最大池化 (max pool) 对特征图 F' 对通道维度进行压缩,其结果做合并操作 (concat);然后,通过 1×1 卷积操作和归一化指数函数 (softmax) 生成注意力图 $M_s \in \mathbf{R}^{1 \times H \times W}$;最后,再与 F' 做乘积运算得到最终的输出精修特征 F'' 。

整体计算过程描述如下:

$$F' = M_c(F) \otimes F \quad (4)$$

$$F'' = M_s(F') \otimes F' \quad (5)$$

式中: F 是中间特征图,作为输入; $M_c(F)$ 和 F' 分别为通道注意力模块注意力图和输出; $M_s(F)$ 和 F'' 分别为空间注意力模块注意力图和输出。

通道注意力模块描述如下:

$$M_c(F) = f^{1 \times 1}(\sigma_r(f^{1 \times 1}(\sigma_s(f^{1 \times 1}(F)) \otimes F))) + F \quad (6)$$

式中: σ_s 和 σ_r 分别为激活函数 sigmoid 和 ReLU; $f^{1 \times 1}$ 是 1×1 卷积层。

空间注意力模块描述如下:

$$M_s(F') = \sigma(f^{1 \times 1}([AvgPool(F'); MaxPool(F')])) \quad (7)$$

式中: $f^{1 \times 1}$ 是 1×1 的卷积层。

1.4 分数图加权融合

本文为进一步提高算法的鲁棒性,基于 SiamFC 做出改进。将 Z 和 X 基于注意力和非注意力生成的特征图做互相关的结果做加权平均:

$$f(Z, X) = \mu f_m(Z, X) + (1 - \mu) f_a(Z, X) \quad (8)$$

式中: f_m 是经过主干网络 MobileNetV3 的互相关结果; f_a 是加入注意力的互相关结果, μ 为权重,本文 $\mu = 0.5$ 。

2 实验结果与分析

2.1 数据集

本实验训练数据集采用了 ILSVRC2015_VID 数据集^[24],该数据集包含 5 000 多个被逐帧标注的视频序列和约 130 万个人工标注数据,目标类别数为 30。此外,为进一步验证算法的有效性,在训练后选用了数据集

OTB50^[25] 和数据集 OTB100^[26] 做进一步的测试,两个数据集分别包含 50 和 100 个人工标注的视频帧。

OTB 数据集的跟踪效果评价标准是一次通过评估 (one-pass evaluation, OPE),评价指标为精确度图 (precision plot) 和成功率图 (success plot)。精确度值表示的是算法预测的跟踪框的中心点与真实框中心点误差在阈值为 20 帧的帧数占总帧数的百分比。成功率用重合率 (overlap score, OS) 进行评估,OS 表达式为:

$$OS = \frac{|A \cap B|}{|A \cup B|} \quad (9)$$

式中: A 为算法跟踪得到的边界框; B 为人工标记的真实边界框; $|\cdot|$ 为区域像素数目。当 OS 超过设定的阈值,被视为成功帧,总的成功帧占总帧数的百分比即为成功率。本实验的成功率是范围为 0~1 内的不同阈值的成功率的平均值。

2.2 实验环境与参数设定

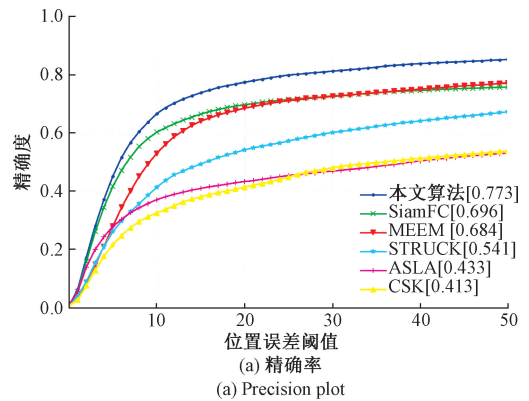
本实验是在 ubuntu16.04 系统下进行的,采用了 Python 语言与 tensorflow1.8.0 框架, CPU 处理器为 Intel(R) Xeon(R) Platinum 8176, GPU 选用的是两块 NVIDIA GeForce 2080ti。

本实验采用的离线训练方式,忽略目标类别,在 ILSVRC2015_VID 数据集上选取 3 862 个视频序列进行训练。从同一个视频序列中随机选取两帧,作为模板图像和搜索图像,两帧间的距离不超过 100 帧。整个训练过程包括 50 个阶段,每个阶段包括 5 000 对样本,学习率的变化为 $10^{-2} \sim 10^{-6}$,训练过程中,采用动量梯度下降法进行网络的优化。

2.3 定量分析

本文算法分别在 OTB50 数据集和 OTB100 数据集上与算法 SiamFC、MEEM^[27]、ASLA^[28]、CSK^[29]、STRUCK^[30] 做对比实验。

对于 OTB50 数据集,本文算法精确度为 77.3%、成功率达到 56.6%,相对于 SiamFC 算法,提升了 7.5% 和 5.3%。6 个算法的成功率和精确度对比如图 4 所示。



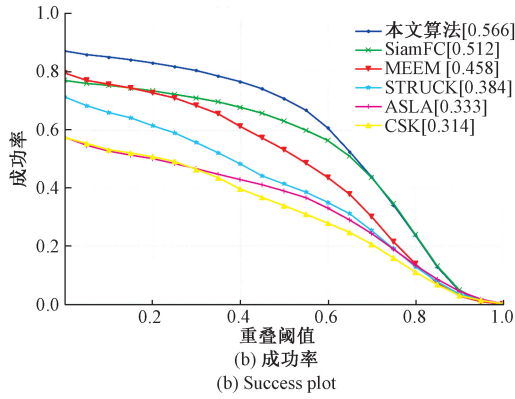


图 4 6 种算法在 OTB50 数据库中的对比

Fig. 4 Comparison of 6 algorithms in the OTB50 database

在 OTB100 数据集中, 本文算法精确度达到 79.8%、成功率达到 59.9%, 相对于 SiamFC 算法, 分别提升了 3.1% 和 2.2%。6 个算法的成功率和精确度对比如图 5 所示。

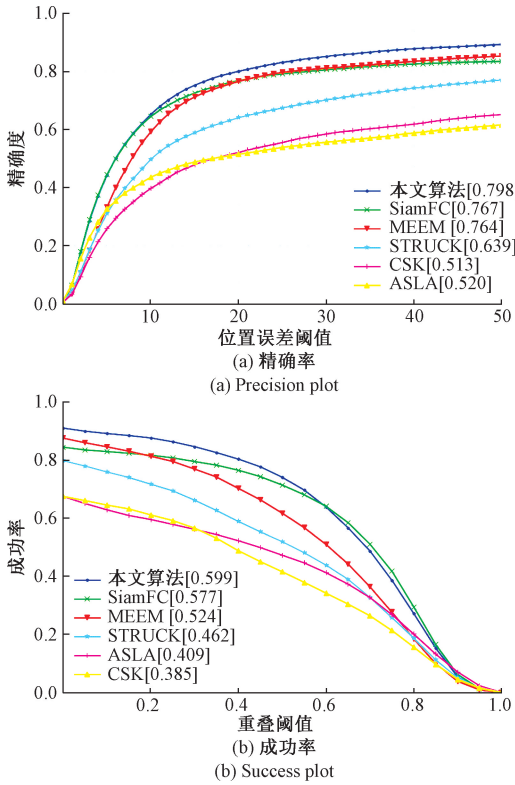


图 5 6 个算法在 OTB100 数据库中的对比

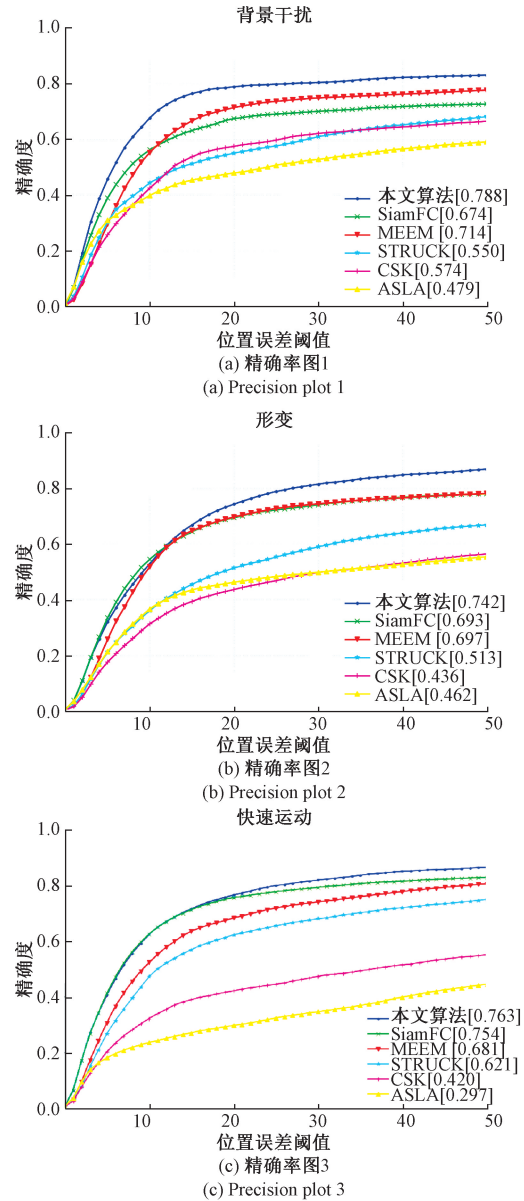
Fig. 5 Comparison of 6 algorithms in the OTB100 database

算法的实时性方面, 在一块 GPU 上 SiamFC 的 FPS 为 92, 本文算法由于加入注意力模块, 导致速度有所下降, FPS 为 67, 但精确度和成功率有所提高。

2.4 定性分析

OTB 数据集的标注可分为 11 种不同的属性: 尺度变化 (scale variation, SV)、光照变化 (illumination variation, IV)、形变 (deformation, DEF)、遮挡 (occlusion, OCC)、快速运动 (fast motion, FM)、运动模糊 (motion blur, MB)、平面内旋转 (in-plane rotation, IPR)、平面外旋转 (out-of-plane rotation, OPR)、超出视野 (out-of-view, OV)、低分辨率 (lower-resolution, LR)、背景干扰 (background clutters, BC)。本文在 OTB100 数据集中将包括本算法在内的 6 个算法在各个难点属性下的视频序列中做对比, 比较算法在不同场景下的表现。

图 6 所示为 6 个算法在不同属性上的精确率图, 图 7 所示为 6 个算法在不同属性上的成功率图。



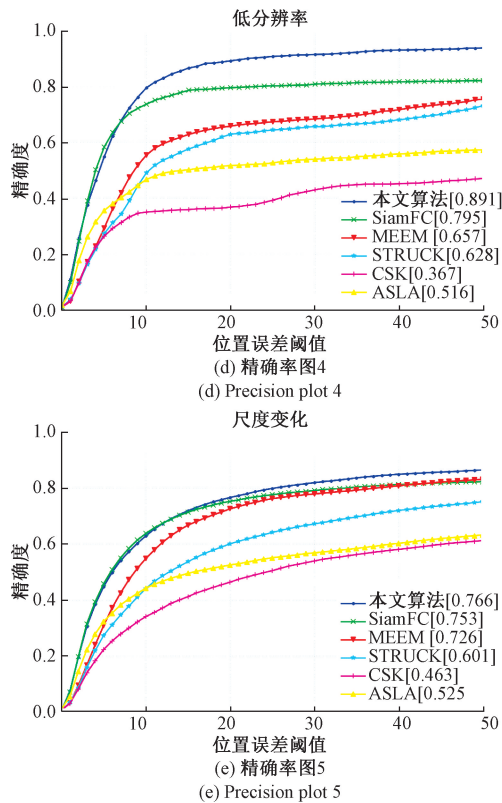


图6 OTB100不同属性下各算法跟踪结果的精确率

Fig. 6 The precision plots of the tracking results for various algorithms on OTB100 dataset with different attributes

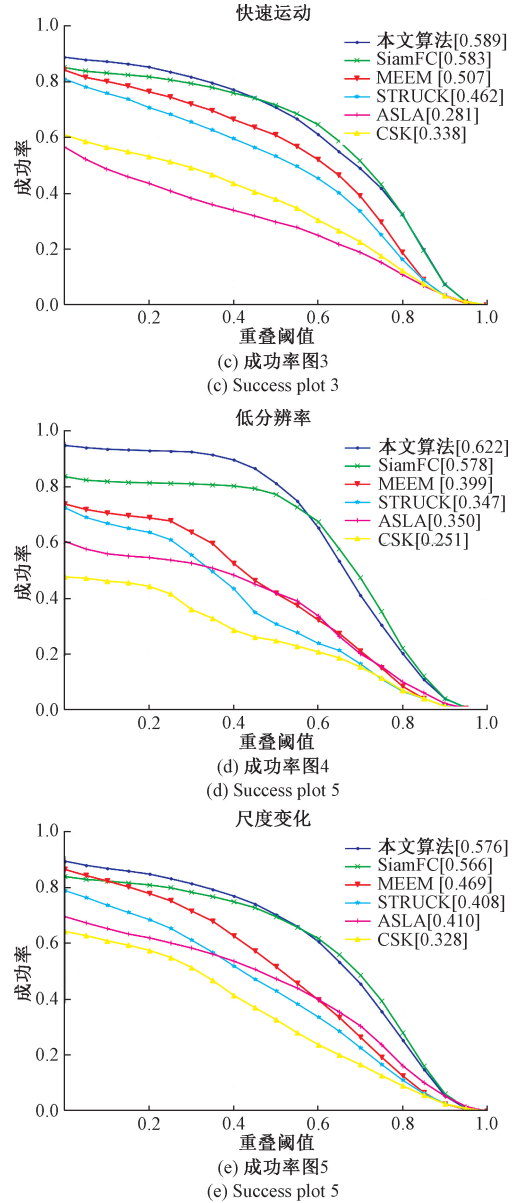
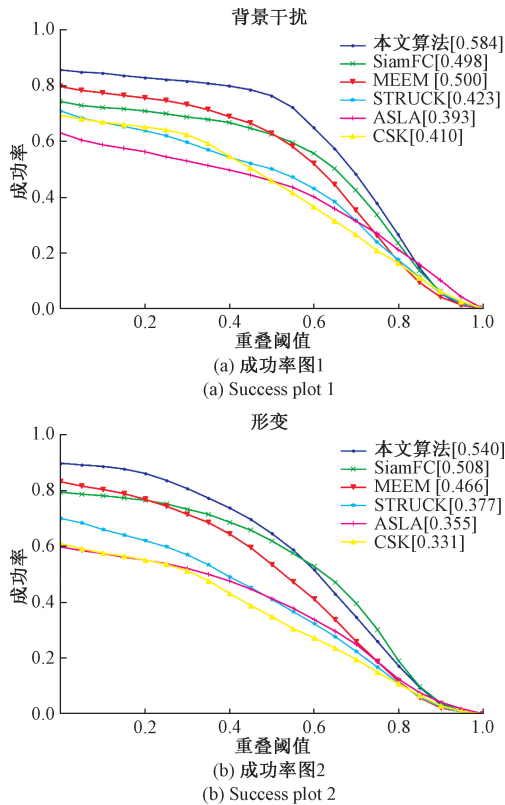


图7 OTB100不同属性下各算法跟踪结果的成功率

Fig. 7 The success plots of the tracking results for various algorithms on OTB100 dataset with different attributes

本文算法和 SiamFC 算法在具有背景干扰、形变、快速运动、低分辨率和尺度变化 5 种属性的视频序列中做对比, 依次为 Couple、FleetFace、Human3、Jump 和 MotorRolling。图 8 所示为可视化结果, 本文算法能较好地应对这些场景。

针对属性分析实验结果, 进行以下分析:

1) 背景干扰 (BC): 相似干扰问题是跟踪领域的难点之一, 对算法的特征提取能力要求高。由数据可知, 未提取到深度特征的算法表现略差, 如图 6 和 7(a) 可知, 本文算法精确率和成功率分别达到 78.8% 和 58.4%。

2) 形变 (DEF): 当目标发生形变时, 跟踪算法需要较



(a) SiamFC在Couple序列上测试结果
(a) Test results for SiamFC on Couple sequence



(b) 本文算法在Couple序列上测试结果
(b) Test results for the proposed algorithm on Couple sequence



(c) SiamFC在FleetFace序列上测试结果
(c) Test results for SiamFC on FleetFace sequence



(d) 本文算法在FleetFace序列上测试结果
(d) Test results for the proposed algorithm on FleetFace sequence



(e) SiamFC在Human3序列上测试结果
(e) Test results for SiamFC on Human3 sequence



(f) 本文算法在Human3序列上测试结果
(f) Test results for the proposed algorithm on Human3 sequence



(g) SiamFC在Jump序列上测试结果
(g) Test results for SiamFC on Jump sequence



(h) 本文算法在Jump序列上测试结果
(h) Test results for the propose algorithm on Jump sequence



(i) SiamFC在MotorRolling序列上测试结果
(i) Test results for SiamFC on MotorRolling sequence



(j) 本文算法在MotorRolling序列上测试结果
(j) Test results for the proposed algorithm on MotorRolling sequence

图8 不同属性视频序列中各算法的实际跟踪效果

Fig. 8 The actual tracking effects of various algorithms in the video sequences with different attributes

高的鲁棒性。由图6和7(b)可知,本文算法性能表现良好,精确率和成功率分别达到74.2%和54.0%。

3)快速运动(FM):目标的快速运动对跟踪算法的稳健性带来极大的挑战。本文算法和SiamFC均有良好的表现,如图6和7(c)可知,相比SiamFC,本文算法由于加入注意力模块精确率和成功率分别达到76.3%和58.9%。

4)低分辨率(LR):图像分辨率较低时,会加大特征提取的难度,容易导致跟踪漂移。由图6和7(d)可知,本文算法相对于其他算法,有较强的特征表达能力,成功率和精确率均分别达到89.1%和62.2%。

5)尺度变化(SV):目标尺度变化时,跟踪算法需对变化的目标进行正确的尺度估计。本文算法均比其他算法表现较好,如图6和7(e)所示,精确率达和成功率分别达到76.6%和57.6%。

3 结 论

为实现复杂场景中对目标实时的跟踪,同时保证跟踪准确率和成功率,提出通道联合空间的孪生网络跟踪方法。该算法在SiamFC网络结构中将改进后的MobileNetV3作为特征提取主干网络,结合GCCSA注意力模块,然后融合得分响应图,将得到的网络模型在ILSVRC2015_VID数据集中进行端到端离线训练,并且在OTB系列上对算法进行验证,算法取得较好的效果,并且在目标快速移动和相似目标干扰时具有良好的鲁棒性。

而当目标在较为复杂的背景下,出现长期遮挡时会影响算法的适应性,下一步工作针对目标跟踪的特点对注意力网络进行改进,并且关注目标在时空序列上的连续性,进一步提高算法的跟踪性能。

参考文献

- [1] 余铎,王耀南,毛建旭,等. 基于视觉的移动机器人目标跟踪方法[J]. 仪器仪表学报, 2019, 40(1): 227-235.
YU D, WANG Y N, MAO J X, et al. VISION-BASED object tracking method of mobile robot [J]. Chinese Journal of Scientific Instrument, 2019, 40(1): 227-235.
- [2] 王耀南,罗琼华,毛建旭,等. 旋翼飞行机器人的行人视觉跟踪方法及系统[J]. 电子测量与仪器学报, 2018, 32(11): 97-107.
WANG Y N, LUO Q H, MAO J X, et al. Pedestrian visual tracking method and system of rotorcraft [J]. Chinese Journal of Scientific Instrument, 2018, 32(11): 97-107.
- [3] COMANICIU D, MEER P. Mean shift: A robust approach toward feature space analysis [J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2002, 24(5): 603-619.
- [4] NUMMIAROA K, KOLLER M E, VAN G L. An adaptive color based particle filter [J]. Image & Vision Computing, 2003, 21(1): 99-110.
- [5] WANG M, LIU Y, HUANG Z. Large margin object tracking with circulant feature maps [J]. Conference on

- Computer Vision and Pattern Recognition (CVPR), 2017: 4021-4029.
- [6] VALMADRE J, BERTINETTO L, HENRIQUES J F, et al. End-to-end representation learning for correlation filter based tracking [C]. Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 2805-2813.
- [7] 江南. 基于动态自适应相关滤波的复杂场景目标跟踪算法[J]. 电子测量与仪器学报, 2019, 33(2): 140-147.
JIANG N. Visual target tracking algorithm based on dynamic adaptive filtering [J]. Journal of Electronic Measurement and Instrument, 2019, 33(2): 140-147.
- [8] 黄树成, 张瑜, 张天柱, 等. 基于条件随机场的深度相关滤波目标跟踪算法[J]. 软件学报, 2019, 30(4): 927-940.
HUANG SH CH, ZHANG Y, ZHANG T ZH, et al. Improved deep correlation filters via conditional random field[J]. Journal of Software, 2019, 30(4): 927-940.
- [9] DANELLJAN M, HAGER G, KHAN F S, et al. Discriminative scale space tracking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(8): 1561-1575.
- [10] MARTIN D, GOUTAM B, FAHAD S K, et al. ECO: Efficient convolution operators for tracking [C]. Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 6638-6646.
- [11] KRIZHEVSKY A, SUTSKEVER I, HINTON G. Imagenet classification with deep convolutional neural networks [C]. Conference on Neural Information Processing Systems, 2012: 1106-1114.
- [12] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. ArXiv Preprint, 2014, ArXiv:1409.1556.
- [13] HE K M, ZHANG X Y, RE S Q, et al. Deep residual learning for image recognition [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [14] NAM H, HAN B. Learning multi-domain convolutional neural networks for visual tracking [C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 4293-4302.
- [15] NAM H, BAEK M, HAN B. Modeling and propagating CNNs in a tree structure for visual tracking [C]. European Conference on Computer Vision, 2016: 1-10.
- [16] BERTINETTO L, VALMADRE J, HENRIQUES J F, et al. Fully-convolutional siamese networks for object tracking [C]. European Conference on Computer Vision, 2016: 850-865.
- [17] DU F, LIU P, ZHAO W, et al. Correlation-guided attention for corner detection based visual tracking [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 6836-6845.
- [18] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.
- [19] BO L, JUNJIE Y, WEI W, et al. High performance visual tracking with siamese region proposal network [C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 8971-8980.
- [20] ZHU Z, WANG Q, LI B, et al. Distractor-aware siamese networks for visual object tracking [C]. European Conference on Computer Vision, 2018: 103-119.
- [21] HOWARD A, SANDLER M, CHU G, et al. Searching for MobileNetV3 [C]. IEEE/CVF International Conference on Computer Vision, 2019: 1314-1324.
- [22] ZENG Y, WANG H, LU T. Learning spatial-channel attention for visual tracking [C]. 2019 IEEE/CIC International Conference on Communications in China (ICCC), 2019: 277-282.
- [23] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module [J]. European Conference on Computer Vision, 2018: 11211.
- [24] 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.
- [25] WU Y, LIM J, YANG M H. Online object tracking: A benchmark [C]. Computer Vision and Pattern Recognition, 2013: 2411-2418.
- [26] WU Y, LIM J, YANG M H. Object tracking benchmark [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 37(9): 1834-1848.
- [27] ZHANG J, MA S, SCLAROFF S. MEEM: Robust tracking via multiple experts using entropy minimization [C]. European Conference on Computer Vision, 2014: 188-123.
- [28] JIA X, LU H, YANG M H. Visual tracking via adaptive structural local sparse appearance model [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2012: 1822-1829.
- [29] HENRIQUES J F, CASEIRO R, MARTINS P, et al. Exploiting the circulant structure of tracking-by-detection with kernels [C]. Proceedings of the 12th European Conference on Computer Vision, 2012: 702-715.
- [30] HARE S, SAFFARI A, TORR P H S. Struck:

Structured output tracking with kernels [C]. IEEE International Conference on Computer Vision, 2011: 6-13.

作者简介



杨梅, 2018 年于辽宁工业大学获得学士学位, 现为辽宁工业大学硕士研究生, 主要研究方向为机器学习, 计算机视觉。

E-mail: glyangmei1994@163.com

Yang Mei received her B. Sc. degree in 2018 from Liaoning University of Technology.

Now, she is an M. Eng. candidate in Liaoning University of Technology. Her main research interest includes machine learning and computer vision.



贾旭 (通信作者), 2005 年于沈阳航空工业学院获得学士学位, 2009 年于东北大学获得硕士学位, 2012 年于东北大学获得博士学位, 现为辽宁工业大学副教授, 主要研究方向为机器学习, 计算机视觉。

E-mail: gbjdjiayu@163.com

Jia Xu (Corresponding author) received his B. Sc. degree in 2005 from Shenyang Aerospace University, received his M. Sc. degree in 2009 and Ph. D. degree in 2012 both from Northeastern University. Now, he is an associate professor in Liaoning

University of Technology. His main research interest includes machine learning and computer vision.



殷浩东, 2018 年于辽宁工业大学获得学士学位, 现为辽宁工业大学硕士研究生, 主要研究方向为机器学习, 计算机视觉。

E-mail: haodongyin23@163.com

Yin Haodong received his B. Sc. degree in 2018 from Liaoning University of Technology.

Now, he is an M. Eng. candidate in Liaoning University of Technology. His main research interest includes machine learning and computer vision.



孙福明, 1996 年于华北工学院获得学士学位, 2001 年于华北工学院获得硕士学位, 2007 年于中国科学技术大学获得博士学位, 现为大连民族大学教授, 主要研究方向为机器学习, 多媒体处理。

E-mail: sunfuming@dlmu.edu.cn

Sun Fuming received his B. Sc. degree in 1996 and M. Sc. degree in 2001 both from North University of China, received his Ph. D. degree in 2007 from University of Science and Technology of China. Now, he is a professor in Dalian Minzu University. His main research interest includes machine learning and multimedia processing.