引入全局上下文特征模块的DenseNet孪生网络目标跟踪

谭建豪 殷 旺* 刘力铭 王耀南
(湖南大学电气与信息工程学院 长沙 410082)
(机器人视觉感知与控制技术国家工程实验室 长沙 410082)

摘 要：近年来，采用孪生网络提取深度特征的方法由于其较好的跟踪精度和速度，成为目标跟踪领域的研究热点之一，但传统的孪生网络并未提取目标较深层特征来保持泛化性能，并且大多数孪生网络只提取局部领域特征，这使得模型对于外观变化是非鲁棒和局部的。针对此，该文提出一种引入全局上下文特征模块的DenseNet孪生网络目标跟踪算法。该文创新性地将DenseNet网络作为孪生网络骨干，采用一种新的密集型特征重用连接网络设计方案，在构建更深层网络的同时减少了层之间的参数量，提高了算法的性能，此外，为应对目标跟踪过程中的外观变化，该文将全局上下文特征模块(GC-Model)嵌入孪生网络分支，提升算法跟踪精度。在VOT2017和OTB50数据集上的实验结果表明，与当前较为主流的算法相比，该文算法在跟踪精度和鲁棒性上有明显优势，在尺度变化、低分辨率、遮挡等情况下具有良好的跟踪效果，且达到实时跟踪要求。

关键词：目标跟踪；孪生网络；全局上下文特征；DenseNet网络
中图分类号：TN911.73; TP391.41 **文献标识码**：A **文章编号**：1009-5896(2021)01-0179-08
DOI: 10.11999/JEIT190788

DenseNet-siamese Network with Global Context Feature Module for Object Tracking

TAN Jianhao YIN Wang LIU Liming WANG Yaonan
(*College of Electrical and Information Engineering*, *Hunan University*, *Changsha* 410082, *China*)
(*National Engineering Laboratory for Robot Visual Perception and Control Technology*, *Hunan University*, *Changsha* 410082, *China*)

Abstract: In recent years, the method of extracting depth features from siamese networks has become one of the hotspots in visual tracking because of its balanced in accuracy and speed. However, the traditional siamese network does not extract the deeper features of the target to maintain generalization performance, and most siamese architecture networks usually process one local neighborhood at a time, which makes the appearance model local and non-robust to appearance changes. In view of this problem, a densenet-siamese network with global context feature module for object tracking algorithm is proposed. This paper innovatively takes densenet network as the backbone of siamese network, adopts a new design scheme of dense feature reuse connection network, which reduces the parameters between layers while constructing deeper network, and enhances the generalization performance of the algorithm. In addition, in order to cope with the appearance changes in the process of object tracking, the Global Context feature Module (GC-Model) is embedded in the siamese network branches to improve the tracking accuracy. The experimental results on the VOT2017 and OTB50 datasets show that comparing with the current mainstream tracking algorithms, the Tracker has obvious advantages in tracking accuracy and robustness, and has good tracking effect in scale change, low resolution, occlusion and so on.

Key words: Object tracking; Siamese network; Global context feature; DenseNet network

1 引言

目标跟踪是计算机视觉领域最基本也是最有挑战的热点研究问题之一，基于视觉的运动目标跟踪已经广泛应用在监控系统、无人机视觉系统、军事侦查、人机交互以及无人驾驶等领域[1]。

近年来，目标跟踪主要分为两类，基于相关滤波的方法和基于深度网络方法。相关滤波方法如

收稿日期：2019-10-16；改回日期：2020-11-13；网络出版：2020-11-19
*通信作者： 殷旺 yinwang@hnu.edu.cn

Henrique等人[2]提出的核相关滤波器(Kernelized Correlation Filter, KCF)算法、Danelljan等人[3]提出的空间正则化判别相关滤波器(Spatially Regularized Discriminative Correlation Filters, SRDCF)算法，该类方法引入了核技巧，提高了跟踪器效率，但相关滤波方法仅考虑相邻帧间的相关特征信息，当目标出现漂移或遮挡时容易出现跟丢。随着深度卷积神经网络的发展，以孪生卷积网络来提取深度特征，进行相似度衡量的方法具有较好的跟踪性能。孪生全卷积(Siamese Fully-Convolutional, SiamFC)算法[4]采用两个网络分支，模板分支和目标分支，通过相关层计算相似性，在速度和精度上获得较好的性能。基于相关滤波器的跟踪(Correlation Filter based tracking, CFNet)算法[5]在目标分支中引入相关滤波层对文献[4]进行改进，在线调整目标模型。动态孪生网络(Dynamic Siamese network, DSiam)[6]通过设计在线动态调整模型，提高了性能。上述Siamese系列算法虽然取得了一定的跟踪精度和速度，但仍存在一些问题。首先，大多数孪生网络算法是基于AlexNet骨干网络，其提取的特征都是浅层的外观特征,缺乏深度特征，双分支孪生神经网络(twofold Siamese network, SA-Siam)算法[7]使用两个Siamese网络，一个用于提取语义信息的网络，另一个用于构建外观模型，将语义信息合并到响应图中，弥补深度信息的不足，但它们都是直接从卷积神经网络(Convolutional Neural Networks, CNN)中获取的局部特征，并没有获取全局上下文特征。

针对以往骨干网络难以提取深层特征，且外观模型不具有全局上下文特征两个问题，本文在Siamese网络思想的基础上重新进行网络设计与搭建，提出一种引入全局上下文信息模块的DenseNet孪生网络目标跟踪算法。其创新有：(1)采用密集网络DenseNet作为骨干网络，提出一种全新端到端深度密集连体结构网络，它在减少网络参数的同时，将层与层之间的特征在channel上进行拼接从而达到特征重用，提高了泛化能力；(2)在网络中加入全局上下文模块(Global Context feature Module, GC-Model)，通过全局池化、1×1特征变化、特征融合等步骤将全局上下文信息进行聚合，用以提升该算法的跟踪性能。

2 骨干网络结构

2.1 残差网络结构

网络的深度对于模型的性能是至关重要的，He等人[8]在实验中发现，网络层数增加到一定程度时，网络准确度会出现饱和，甚至出现下降，并且不是过拟合所导致的问题，因此，残差网络由此产生，对于一个堆基层结构，当输入为x时其所学习到的特征记作$H(x)$，我们期望真实地学习到残差

$$F(x)=H(x)-x \tag{1}$$

因此，原始的学习特征为$F(x)+x$，残差学习相比原始特征学习容易，且实际残差不会为0，这让堆基层在输入特征的基础上能够学习到新特征，从而具有更好的性能。

ResNet网络是在VGG19网络的基础上进行修改，并引入残差模块，该网络为后续密集型网络DenseNet提供了理论与经验基础。

2.2 DenseNet网络结构

有研究表明，如果卷积网络包含有输入层和输出层之间的较短连接，则卷积网络可以更加深入，更加精确有效地进行训练，因此，本文选择密集连接卷积网络DenseNet网络结构。

ResNet模型的核心是通过建立前面层与后面层之间的一种“短路连接”，这有利于训练过程中梯度的反向传播，DenseNet提出了一种更激进的密集连接机制。

DenseNet网络是一种密集连接方式，每个层都会与前面所有层在通道维度上进行拼接，假设网络层数为N，则DenseNet共包含有 $N(N+1)/2$个连接，实现特征重用。考虑通过卷积神经网络传递的单幅图像x_0，第n层的输入为前面所有层的特征映射，如式(2)所示

$$x_n=H_l([x_0,x_1,\cdots,x_{n-1}]) \tag{2}$$

其中，$[x_0,x_1,\cdots,x_{n-1}]$指$0,1,\cdots,n-1$中产生的特征映射的连接；$H_l(\sim)$表示归一化(Batch Normalization, BN)修正线性单元，(REctified Linear Unit, RELU)、池化(pool)、卷积(convolution)等复合函数变换。

所有DenseBlock中各个卷积之后均输出K个通道的特征图，特征重用会使得后面层的输入很大，因此在DenseBlock内部结构中加入1×1卷积[9,10]，减少计算量，提高特征效率，如图1所示，DenseBlock的结构为BN+RELU+1×1 Conv+BN+RELU+3×3 Conv。

3 全局上下文特征模型(GC-Model)

3.1 长距离依赖捕获方法

非局部神经网络(Non Local neural Network, NLNet)[11]采用自注意力机制来建模像素对关系，但是其对于每一个位置学习不受限制依赖的注意力图(attention map), 造成了很大的计算资源浪费。

NLNet旨在从其他位置聚集信息来增强当前位

置的特征，x和z定义为该网络结构的输入与输出，则NLNet可以表示为

$$z_i = x_i + W_z \sum_{j=1}^{N_p} \frac{f(x_i, x_j)}{C(x)} (W_v, x_j) \tag{3}$$

其中，$C(x)$为归一化因子，W_z, W_v表示类似于1×1卷积等线性转换矩阵，i为位置的索引，j为枚举的所有可能的位置，其网络结构如图2(a)所示。

NLNet[11]将每个查询位置进行全局上下文聚合，提供了一种非局部特征捕获的开创性方法，该类方法旨在提取视觉场景的全局理解，广泛应用于识别、物体检测、分割等领域。为了模拟全局上下文特征，SENet[12],GENet[13]对不同通道执行重新加权操作，以重新校准具有全局上下文的通道依赖性。

如图2(b)所示为SENet网络结构，其可以大致理解成3个过程：网络中全局平均池化用于上下文建模，增强位置的特征；通道权值计算，即1×1卷积、RELU和Sigmoid等计算，使用特征转换来获取通道间的依赖；通道特征重标定。

3.2 GC-Model网络模型

GC-Model是结合了SENet计算量小以及NLNet全局上下文能力等优点提出了非局部操作网络[14]，其计算量相对较小，又能够很好地融合全局信息，在目标检测中取得了重要的提升。

GC-Model中非局部操作可分为3个过程[14]：(1)用于上下文建模的全局注意力集中机制，采用1×1卷积W_k和Softmax函数来得到自注意权重，然后进行注意力集中获取全局背景特征；(2)特征转换获取通道依赖性；(3)特征融合，全局上下文模型的详细结构如图3所示，可表示为式(4)

$$z_i = x_i + W_{v2}\mathrm{RELU}\left(\mathrm{LN}\left(W_{v1}\sum_{j=1}^{N_p}\frac{\mathrm{e}^{W_k x_j}}{\sum_{m=1}^{N_p}\mathrm{e}^{W_k x_m}} x_j\right)\right) \tag{4}$$

其中，$\dfrac{\mathrm{e}^{W_k x_j}}{\sum_{m=1}^{N_p}\mathrm{e}^{W_k x_m}}$表示全局关注池的权重，$W_{v2}\mathrm{RELU}(\mathrm{LN}(W_{v1}(\sim)))$表示特征转换。

GC-Model是轻量级的模型，能够获取远程非

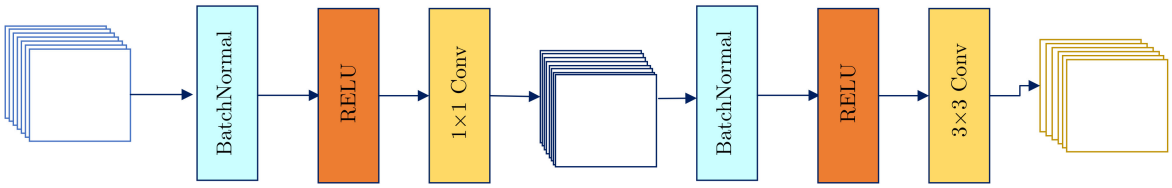
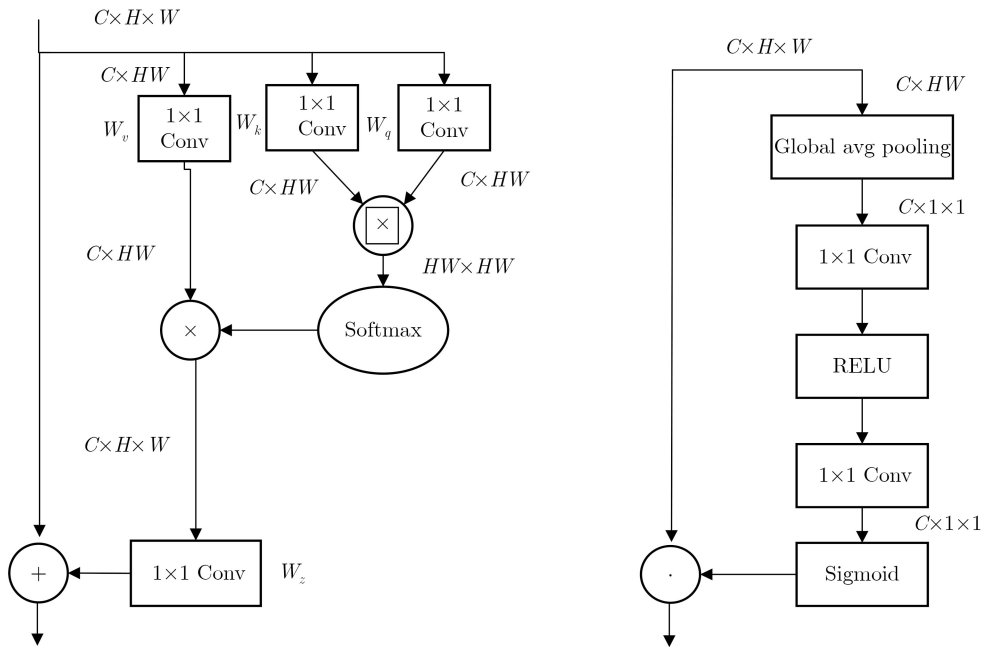


图 1 DenseNet的网络结构



(a) NLNet结构　　(b) SENet结构

图 2 两种长距离依赖模型图

局部特征，且能灵活地插入各个视觉问题的网络架构中，本文将GC-Model放入骨干DenseNet网络架构中，用以提升网络训练的泛化性能。

4 本文SD-GCNet算法

4.1 孪生网络目标跟踪框架

近年来，SiamFC开启了深度学习方法在目标跟踪领域的大门，通过端到端网络学习，使用相似度学习的方法来实现目标跟踪。其网络框架如图4所示。

孪生网络通过建立两个分支进行训练，两分支所使用的骨干网络完全一致，在SiamFC中，采用互相关函数$f(z,x)$作为相似度函数，计算经过φ之后的特征提取后的特征图相似性[15]

$$f(z,x)=\varphi(z)*\varphi(x)+kI \tag{5}$$

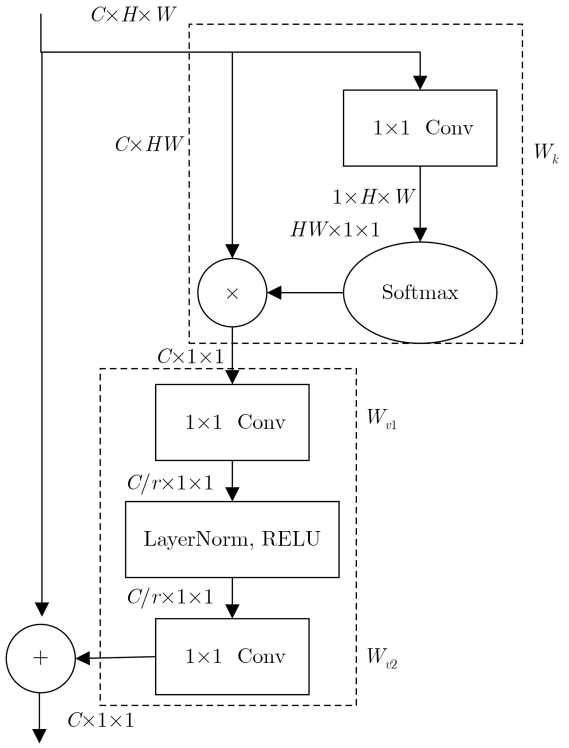


图 3 全局上下文GC-Model模块

其中，*表示卷积，kI表示响应图在每个位置的取值。

4.2 本文目标跟踪算法框架

上述第2节，第3节详细地介绍了骨干网络架构和孪生网络架构的基本信息。(1)DenseNet网络是一种密集连接型网络，在构建更深层网络的同时减少了层之间的参数量，能够增强算法的泛化性能，并且能够解决训练过程中的梯度消失问题；(2)GCNet综合了SENet计算量小以及NLNet全局上下文能力等优点，其计算量相对较小，又能够很好地融合全局信息，可融入任何骨干网络当中用以提升性能；(3)孪生网络的网络架构方式已经在目标跟踪上取得了较好的跟踪性能，且实时性较好。据此以上述3个研究成果为出发点，整理思路，本文通过假设、组合、实验验证等一系列步骤，最终得出本文SD-GCNet目标跟踪网络框架。其网络框图如图5所示。

本文提出一种引入全局上下文信息模块的DenseNet孪生网络目标跟踪算法SiamDenseNet+GC-Model，简称SD-GCNet，其核心思想是以密集型网络DenseNet作为孪生网络的骨干[16]，在骨干网络中引入GC-Model，搭建SD-GCNet网络框架。

为了更加明显地表示网络结构，特以表1形式进行展示。

4.3 损失函数

SD-GCNet网络实际上是一种判别的二分类方法，在正负样本对上采用极大似然估计进行训练，本文采用Logistic损失函数，如式(6)

$$l(y,v)=\lg(1+\exp(-yv)) \tag{6}$$

其中，v为目标候选模板得分，$y\in\{-1,+1\}$表示类别真实标签，采用一个目标候选图像和一个较大搜索区域图像来训练SD-GCNet网络，定义响应图的损失为整张图的损失平均值，如式(7)

$$L(y,v)=\frac{1}{M}\sum_{o\in M}l(y[o],v[o]) \tag{7}$$

对每一个像素位置o，都有对应的标签$\boldsymbol{y}$。当位

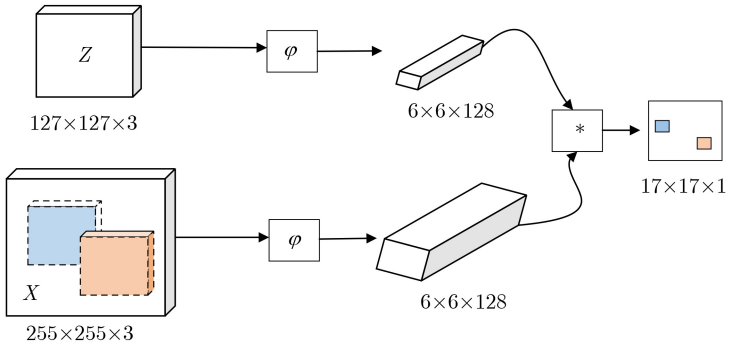


图 4 孪生网络目标跟踪框架图

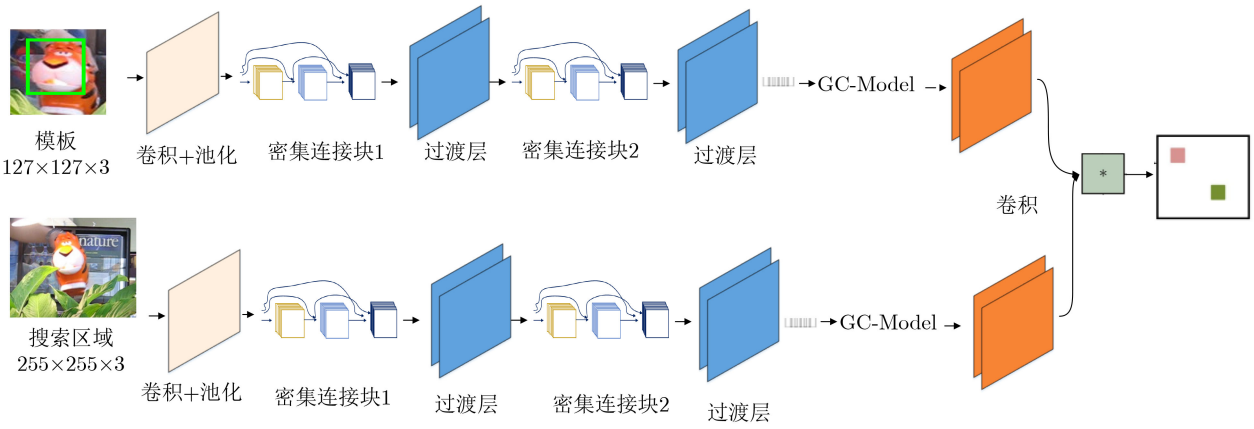


图 5 SD-GCNet算法框架

表 1　网络结构

层名称	模板分支	搜索分支	输出
卷积层	7×7Conv, stride 2	7×7Conv, stride 2	61×61×72
密集连接1	1×1Conv ×2 +3×3Conv×2	1×1Conv ×2+3×3Conv×2	61×61×144
过渡层1	1×1Conv+average pool	1×1Conv+average pool	30×30×36
密集连接2	1×1Conv ×4+3×3Conv×4	1×1Conv ×4+3×3Conv×4	30×30×180
过渡层2	1×1Conv+average pool	1×1Conv+average pool	15×15×36
密集连接3	1×1Conv ×6+3×3Conv×6	1×1Conv ×6+3×3Conv×6	15×15×252
密集连接3	3×3Conv×3	3×3Conv×3	9×9×128
GC-Model	图3	图3	9×9×128

置o与图像中心位置的距离在一个阈值内时，将其视为正样本，否则视为负样本。

5　实验与结果

5.1　算法实现细节

本文算法实现与调试在ubuntu16.04操作系统下，计算机硬件配置为Intel Core i7-8700k 主频3.7 GHz, GeForce RTX2080TI显卡。

SD-GCNet算法在ILSVRC2015和GOT-10K数据集共6000个视频序列上进行训练，该数据集具有各种各样的视频目标对象，具备一定的普遍性。本文采用随机梯度下降(Stochastic Gradient Descent, SGD)优化算法以动量参数为0.9训练网络，学习率从$10^{-8}\sim10^{-3}$在训练过程中逐渐递减，用高斯函数初始化参数，batchsize设置为16。通过5种尺寸1.0327$\{-2,-1,0,1,2\}$上的搜索对象来调整尺寸变化，输入候选图像尺寸为127×127，搜索图像尺寸为255×255，使用线性插值来更新尺寸。

5.2　实验结果分析

为验证本文提出的SD-GCNet算法可靠性，特在VOT2017数据集上对算法进行定量评估，在OTB50数据集上对算法进行定性分析，从多个数据集多种角度验证算法的有效性和优越性。

5.2.1　定量分析

如表2所示，为本文算法在VOT2017[17]数据集上与目前较为主流的6种目标跟踪算法SiamFC, SiamVGG[18], DCFNet[19], SRDCF[3], DeepCSRDC, Staple[20]在精确度、鲁棒性等指标上的性能对比，其中表中鲁棒性用跟丢次数来衡量，SiamFC, SiamVGG, DCFNet为深度学习算法，SRDCF, DeepCSRDCF, Staple为相关滤波算法。

由表2可知，本文算法在目标跟踪的精确度上均高于其余算法，与用VGG-16作为骨干网络SiamVGG算法相比，在VOT2017数据集上，其精确度提升了1.9%，平均重叠期望提升了1.0%，与以AlexNet作为主干网络的SiamFC算法相比，则性能提升更多，这更进一步验证了本文DenseNet作为主干网络的优越性。

为了进一步分析该算法的优缺点，本文提供了其在VOT2017数据集上的具体属性对比，包括相机移动、目标丢失、光照变化、运动变化、目标遮挡、尺度变化共6种属性。

表3和表4分别列出了上述6种属性下算法的跟踪精度和跟踪鲁棒性，其中，加粗数字表示排名第1，蓝色数字表示排名第2，从表中可以看出，本文算法除光照变化外，其跟踪精度均处于最优位置，

在跟踪鲁棒性上，相机移动和尺度变化也处于领先位置，其余均排在前列。由此表明，本文算法具有较好的跟踪精度，在较多复杂的条件下也能有较好的跟踪鲁棒性。

5.2.2 定性分析

图6给出了本文算法与另外4种算法SiamFC, SRDCF, Staple, Struck[21]在OTB50[22]上的跟踪结果，表5表示了测试序列的影响因素。

根据图6的跟踪结果和表5的影响因素对算法进行如下定性分析：

(1) 快速运动：以测试序列Bolt和Ironman为例，目标快速移动，目标外观和背景都发生快速变化，对匹配性算法和更新模板类算法都会产生较大的影响。SRDCF和Struck算法在序列Bolt上第10帧就完全丢失了目标，并且基于模板更新，后续不能恢复跟踪，在序列Ironman第38帧，SRDCF, Staple, Struck已经完全丢失目标，只有本文算法在两种干扰因素下保持良好跟踪。

(2) 背景干扰、杂波，光照变化：以测试序列carDark为例，在背景干扰严重，光照变化明显的条件下，对于前景特征提取的准确性显得尤为重要。在carDark序列第295帧，匹配类算法SiamFC已经出现目标丢失，这进一步说明DenseNet骨干网络优于AlexNet在背景干扰上的特征提取能力。

(3) 遮挡：以测试序列Jogging-2为例，在该序列第53帧时出现跟踪目标完全被遮挡情况，当遮挡消失，Staple算法和Struck算法全部跟丢，本文算法，SiamFC, SRDCF能够重新恢复跟踪，本文算法和SiamFC采用第1帧目标匹配方法，能够在目标消失遮挡时恢复跟踪。

本文所提算法有上述优点，在快速运动、背景干扰、遮挡等方面具备一定的性能，但其涉及较深的深度网络，在运行时对计算机性能要求颇高，如果将其运用在机器人或旋翼无人机等实际设备上，对小型机载计算机性能有要求，才能确保达到实时跟踪状态，且本文算法并没有设定自适应目标跟踪框，也没有使用动态孪生网络方法进行参数更新，后续可以考虑在这几个方面进行进一步的研究，以便达到更好的跟踪性能。

表 2 在VOT2017数据集上与主流算法的基础模型结果对比

跟踪算法	精确度	鲁棒性	平均重叠期望
本文算法	**0.544**	20.090	**0.297**
SiamFC	0.500	34.031	0.188
SiamVGG	0.525	20.453	0.287
DCFNet	0.465	35.202	0.183
SRDCF	0.480	64.114	0.119
DeepCSRDCF	0.483	**19.007**	0.293
Staple	0.524	44.019	0.169

6 结论

本文提出了一种引入全局上下文特征模块的DenseNet孪生网络目标跟踪算法。使用较深层的密集型DenseNet网络，获取更深层的前景外观特征和语义背景，增强了算法的泛化性能；将全局上下文特征模块嵌入孪生网络分支，提高算法跟踪精度。在两个流行的数据集VOT2017, OTB50上评估，实验结果表明了该算法具备良好的跟踪精度与

表 3 不同属性下算法的跟踪精度对比

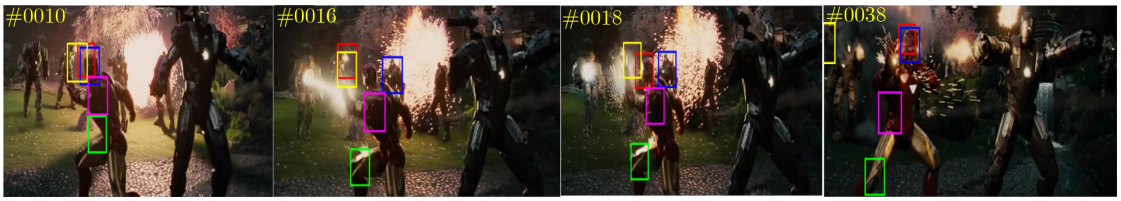
跟踪算法	相机移动	目标丢失	光照变化	运动变化	目标遮挡	尺度变化
本文算法	**0.561**	**0.562**	0.543	**0.554**	**0.461**	**0.543**
SiamFC	0.513	0.513	0.556	0.514	0.416	0.474
SiamVGG	0.542	0.531	0.538	0.540	0.442	0.514
DCFNet	0.485	0.472	0.532	0.464	0.377	0.450
SRDCF	0.484	0.511	**0.588**	0.453	0.419	0.447
Staple	0.554	0.528	0.5371	0.523	0.459	0.492

表 4 不同属性下算法的跟踪鲁棒性对比(数字表示失败次数)

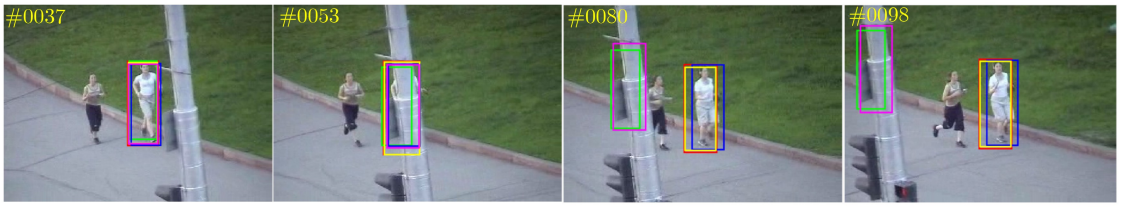
跟踪算法	相机移动	目标丢失	光照变化	运动变化	目标遮挡	尺度变化
本文算法	**29.0**	18.0	3.0	16.0	22.0	**11.0**
SiamFC	40.0	31.0	5.0	42.0	32.0	25.0
SiamVGG	35.0	**15.0**	**2.0**	**15.0**	**19.0**	**11.0**
DCFNet	50.0	34.0	8.0	31.0	24.0	21.0
SRDCF	76.0	86.0	9.0	49.0	32.0	29.0
Staple	62.0	53.0	5.0	27.0	27.0	17.0

(a) Bolt

(b) carDark



(c) Ironman



(d) Jogging-2

本文算法　SiamFC　SRDCF　Staple　Struck

图 6 本文算法与4种算法的跟踪结果对比

表 5　OTB50中测试序列与其影响因素

测试序列	帧数	影响因素
Bolt	18	快速移动、相机移动、尺度变化等
carDark	244～363	运动模糊、低分辨率、背景杂波等
Ironman	38	平面内旋转、快速运动、光照变化等
Shaking	55	光照变化、背景模糊等
Jogging-2	53	遮挡

鲁棒性，在尺度变化、低分辨率、遮挡等情况下具有良好的跟踪效果。

参 考 文 献

[1] 孙彦景, 石韫开, 云霄, 等. 基于多层卷积特征的自适应决策融合目标跟踪算法[J]. 电子与信息学报, 2019, 41(10): 2464–2470. doi: 10.11999/JEIT180971.
SUN Yanjing, SHI Yunkai, YUN Xiao, *et al*. Adaptive strategy fusion target tracking based on multi-layer convolutional features[J]. *Journal of Electronics & Information Technology*, 2019, 41(10): 2464–2470. doi: 10.11999/JEIT180971.

[2] HENRIQUE 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. doi: 10.1109/tpami.2014.2345390.

[3] DANELLJAN M, HÄGER G, KHAN F S, *et al*. Learning spatially regularized correlation filters for visual tracking[C]. 2015 IEEE International Conference on Computer Vision, Santiago, Chile, 2015: 4310–4318.

[4] BERTINETTO L, VALMADRE J, HENRIQUES J F, *et al*. Fully-convolutional Siamese networks for object tracking[C]. European Conference on Computer Vision, Amsterdam, The Netherlands, 2016: 850–865. doi: 10.1007/978-3-319-48881-3_56.

[5] VALMADRE J, BERTINETTO L, HENRIQUES J, *et al*. End-to-end representation learning for correlation filter based tracking[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 5000–5008. doi: 10.1109/CVPR.2017.531.

[6] GUO Qing, WEI Feng, ZHOU Ce, *et al*. Learning dynamic Siamese network for visual object tracking[C]. 2017 IEEE

International Conference on Computer Vision, Venice, Italy, 2017: 1781–1789. doi: 10.1109/ICCV.2017.196.

[7] HE Anfeng, LUO Chong, TIAN Xinmei, *et al*. A twofold siamese network for real-time object tracking[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 4834–4843. doi: 10.1109/CVPR.2018.00508.

[8] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, *et al*. Deep residual learning for image recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 770–778. doi: 10.1109/CVPR.2016.90.

[9] SZEGEDY C, VANHOUCKE V, IOFFE S, *et al*. Rethinking the inception architecture for computer vision[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 2818–2826. doi: 10.1109/CVPR.2016.308.

[10] 侯志强, 陈立琳, 余旺盛, 等. 基于双模板Siamese网络的鲁棒视觉跟踪算法[J]. 电子与信息学报, 2019, 41(9): 2247–2255. doi: 10.11999/JEIT181018.
HOU Zhiqiang, CHEN Lilin, YU Wangsheng, *et al*. Robust visual tracking algorithm based on Siamese network with dual templates[J]. *Journal of Electronics & Information Technology*, 2019, 41(9): 2247–2255. doi: 10.11999/JEIT181018.

[11] WANG Xiaolong, GIRSHICK R, GUPTA A, *et al*. Non-local neural networks[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 7794–7803. doi: 10.1109/CVPR.2018.00813.

[12] HU Jie, SHEN Li, and SUN Gang. Squeeze-and-excitation networks[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 7132–7141. doi: 10.1109/CVPR.2018.00745.

[13] HU Jie, SHEN Li, ALBANIE S, *et al*. Gather-excite: Exploiting feature context in convolutional neural networks[C]. The 32nd International Conference on Neural Information Processing Systems, Montréal, Canada, 2018: 9423–9433.

[14] CAO Yue, XU Jiarui, LIN S, *et al*. GCNet: Non-local networks meet squeeze-excitation networks and beyond[C]. 2019 IEEE/CVF International Conference on Computer Vision Workshop, Seoul, Korea (South), 2019: 1971–1980. doi: 10.1109/ICCVW.2019.00246.

[15] 刘畅, 赵巍, 刘鹏, 等. 目标跟踪中辅助目标的选择、跟踪与更新[J]. 自动化学报, 2018, 44(7): 1195–1211.
LIU Chang, ZHAO Wei, LIU Peng, *et al*. Auxiliary objects selecting, tracking and updating in target tracking[J]. *Acta Automatica Sinica*, 2018, 44(7): 1195–1211.

[16] ABDELPAKEY M H, SHEHATA M S, and MOHAMED M M. DensSiam: End-to-end densely-Siamese network with self-attention model for object tracking[C]. The 13th International Symposium on Visual Computing, Las Vegas, USA, 2018: 463–473.

[17] KRISTAN M, LEONARDIS A, MATAS J, *et al*. The visual object tracking VOT2017 challenge results[C]. 2017 IEEE International Conference on Computer Vision, Venice, Italy, 2017: 1949–1972. doi: 10.1109/ICCVW.2017.230.

[18] LI Yuhong and ZHANG Xiaofan. SiamVGG: Visual tracking using deeper Siamese networks[J]. arXiv: 2019, 1902.02804.

[19] WANG Qiang, GAO Jin, XING Junliang, *et al*. Dcfnet: Discriminant correlation filters network for visual tracking[J]. arXiv: 2017, 1704.04057.

[20] BERTINETTO L, VALMADRE J, GOLODETZ S, *et al*. Staple: Complementary learners for real-time tracking[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 1401–1409. doi: 10.1109/CVPR.2016.156.

[21] HARE S, GOLODETZ S, SAFFARI A, *et al*. Struck: Structured output tracking with kernels[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 38(10): 2096–2109. doi: 10.1109/TPAMI.2015.2509974.

[22] WU Yi, LIM J, and YANG M H. Online object tracking: A benchmark[C]. 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, USA, 2013: 2411–2418. doi: 10.1109/CVPR.2013.312.

谭建豪：男，1962年生，教授，硕士生导师，研究方向为计算机视觉、飞行机器人、模式识别.

殷 旺：男，1995年生，硕士生，研究方向为计算机视觉、目标跟踪.

刘力铭：男，1996年生，硕士生，研究方向为计算机视觉、目标跟踪、图像分割.

王耀南：男，1957年生，教授，博士生导师，研究方向为智能控制、模式识别技术等.

责任编辑：余 蓉