计算机科学与探索
Journal of Frontiers of Computer Science and Technology
ISSN 1673-9418,CN 11-5602/TP

《计算机科学与探索》网络首发论文

题目：　改进的 Siamese 自适应网络和多特征融合跟踪算法
作者：　李睿，连继荣
网络首发日期：　2021-06-16
引用格式：　李睿，连继荣．改进的 Siamese 自适应网络和多特征融合跟踪算法．计算机科学与探索. https://kns.cnki.net/kcms/detail/11.5602.TP.20210615.1641.008.html

计算机科学与探索
Journal of Frontiers of Computer Science and Technology doi: 10.3778/j.issn.1673-9418.2103044

改进的 Siamese 自适应网络和多特征融合跟踪算法

李睿，连继荣+

兰州理工大学 计算机与通信学院，兰州 730050
+ 通信作者 E-mail: 649094574@qq.com

摘 要: 针对当前目标跟踪领域中跟踪精确度和跟踪速度难以平衡的问题。例如基于相关滤波实现的跟踪器能够以很高的速度运行，但跟踪准确性极低；基于深度学习实现的跟踪器能够实现较高的跟踪准确性,但跟踪速度较低。在此基础上，提出一种改进的 Siamese 自适应网络和多特征融合目标跟踪算法。首先在 Siamese 网络每个分支上同时构建 AlexNet 网络和改进的 ResNet 网络，用于特征提取。其次通过端到端的方式同时进行训练，将跟踪问题分解为分类每个位置标签和回归边界框子问题。最后对浅层特征和深层特征进行自适应选择以及基于多特征融合进行目标识别和定位。将提出的算法与现有的一些跟踪器在目标跟踪标准数据集上进行测试。实验结果表明，提出的算法能够在确保跟踪速度的同时实现较高的跟踪精确度和成功率。同时，在光照变化、形变、背景杂波等复杂情况下，算法具有较强的鲁棒性。

关键词: 目标跟踪；Siamese 网络；特征融合；尺度自适应；ResNet 网络

文献标志码：A **中图分类号：**TP391

Improved Siamese Adaptive Network and Multi-feature Fusion Tracking Algorithm

LI Rui, LIAN Jirong+

College of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China

Abstract: Aiming at the problem that tracking accuracy and tracking speed are difficult to balance in the current target tracking field. For example, a tracker based on correlation filtering can run at a very high speed, but the tracking accuracy is extremely low; a tracker based on deep learning can achieve high tracking accuracy, but the tracking speed is extremely low. On this basis, an improved Siamese adaptive network and multi-feature fusion target tracking algorithm are proposed. First, the AlexNet network and the improved ResNet network are constructed on each branch of the Siamese network at the same time for feature extraction. Secondly, through end-to-end training at the same time, the tracking problem is decomposed into sub-problems of classifying each position label and returning to the bounding box. Finally, adaptive selection of shallow features and deep features, and target recognition and location based on multi-feature fusion. The proposed algorithm and some existing trackers are tested on the target tracking standard data set. Experimental results show that the proposed algorithm can achieve high target tracking accuracy and success rate while ensuring tracking speed. At the same time, the algorithm has strong robustness in complex situations such as illumination changes, deformations, and background clutter.

Key words: object tracking; Siamese network; feature fusion; scale adaptation; ResNet network

基金项目：国家自然科学基金项目（61761028）；甘肃省重点研发计划-工业类（18YF1GA060）。

This work was supported by the National Natural Science Foundation of China (No. 61761028) and the Key R&D Program of Gansu Province-Industrial (No. 18YF1GA060).

目标跟踪是计算机视觉中一项基本但具有挑战性的任务[1-2]。给定视频序列初始帧中的目标状态，跟踪器需要预测每个后续帧中目标的位置和大小。尽管近年来取得了很大进展，但由于遮挡、尺度变化、背景杂波、快速运动、光照变化和形变等因素的影响[3]。视觉跟踪仍然面临巨大挑战。

在现实生活中，目标的大小和宽高比例随着目标的移动、摄像机的移动和目标外观的变化而变化。在目标跟踪任务中能够快速、准确的确定目标的位置和大小是视觉跟踪领域一个难以解决的问题。近几年视觉跟踪方法都是基于 Siamese 网络的架构来实现[4-7]。许多研究者对此提出了大量改进方法以实现准确的目标跟踪。

Siamese 网络将目标跟踪看作是目标匹配问题来处理，核心思想是学习目标模板和搜索区域的相似图，一个常见的策略是在搜索区域的多个尺度上进行匹配，以确定目标尺度的变化，这就是这些跟踪器耗费时间、耗费空间的原因。其中文献[5]引入区域建议网络以获取更加准确的目标边界框，通过联合一个分类分支和一个回归分支进行视觉跟踪，避免了由于目标尺度不变性而费时提取多特征的步骤，在许多基准上取得了较好的结果。但是为了处理不同的尺度大小和高宽比，他们基于启发式知识设计锚框，如此做将会引入大量的超参数以及计算复杂度很高。紧接着 DaSiam[7]，和 SiamRPN++[8]针对以上问题对 SiamRPN 做了改进。然而，由于为区域建议引入了锚点，所以这些跟踪器对锚盒的数量、大小和长宽比都很敏感，超参数调优技术对于成功地使用这些跟踪器进行跟踪至关重要[9]。

在这篇文章中，我们按照 Siamese 网络的特点，将跟踪问题分解为两个子问题：一个分类问题和一个回归问题。其中分类任务是将每个位置预测为一个标签，而回归任务将每个位置回归为一个相对的边界框。通过这种分解，可以将跟踪任务进行按模块求解，我们设计一个简单有效的 Siamese 自适应网络用于特征提取，同时进一步进行分类和回归，以端到端的方式同时进行学习。

1 相关工作

近年来，随着大数据、机器学习等的快速发展，凭借计算机强大的计算能力极大地推动人工智能地快速发展。目标跟踪成为计算机视觉领域最活跃的研究主题之一[10-13]。深度学习算法相比传统的相关滤波算法，在目标跟踪精确度和成功率方面得到巨大的改善和提高。本节我们主要回顾基于 Siamese 网络设计的一系列跟踪器。因为近几年这些跟踪器在性能方面遥遥领先。

目标跟踪领域的研究者主要从特征提取[14-15]、模板更新[16-17]、分类器设计[18]、边界框回归[19]等不同方面，致力于设计更快、更准确的跟踪器。早期的特征提取主要使用颜色特征、纹理特征或其他手工制作的特征[20]。得益于深度学习的发展，卷积神经网络（Convolutional Neural Network, CNN）的深度卷积特性被广泛采用。模板更新可以提高模型的适应性，但在线跟踪效率很低。

此外，模板更新的跟踪漂移问题还有待解决。相关滤波方法[14]的引入使得跟踪的效率和准确率都达到了前所未有的高度[21-22]。目前的研究表明，基于 Siamese 的在线训练和带有深度神经网络的离线跟踪方法在准确率和效率之间取得了最好的平衡。

作为开创性的工作之一，SiamFC[6]构建了一个完全卷积的 Siamese 网络用于特征提取，由于 SiamFC 的结构简单。因此，跟踪速度可以达到 86FPS。受其成功的鼓舞，许多研究者认可了这项工作并基于 SiamFC 提出了一些改进方法。

CFNet[23]在 SiamFC 框架中引入相关滤波层，进行在线跟踪,提高精度。DSiam 学习了一个特征变换，用于解决目标外观变化以及背景干扰。通过动态的 Siamese 网络，在可接受的速度损失的情况下，提高了跟踪精度[7]。SAsiam 构建了一个双重 Siamese 网络，包括语义分支和外观分支，两个分支分开训练以保证输出特征的异质性，提高跟踪精度。为了解决目标尺度变化问题，这些跟踪器需要进行多尺度搜索,这会造成大量的时间消耗和空间

浪费。

SiamRPN[5]通过联合训练一个分类分支和一个回归分支进行区域建议，避免了由于目标尺度不变性而费时提取多尺度特征图的步骤，取得了非常高效的结果。然而，它很难处理与物体外观相似的干扰物。至今，已对 SiamFC 做了很多修改和改进，但是使用 AlexNet[24]作为主干网络，跟踪器的性能无法进一步提高。 针对这个问题，SiamRPN ++ 通过使用 ResNet[25]作为主干网络，优化了网络架构。为了消除中心位置偏差，在训练期间随机移动目标在搜索图像区域的位置。经过以上改进，可以使用非常深的网络结构实现更高精度的目标跟踪。

本文的主要创新点：

1、我们设计了一个 Siamese 自适应网络，即在 Siamese 网络的每个分支同时构建 AlexNet 浅层网络和改进的 ResNet 深层网络，用于特征提取。

2、提出一种全新的跟踪策略，对浅层特征和深层特征进行自适应选择以及基于多特征融合进行识别和定位，增强网络判别力，提高目标跟踪精度。同时采用由局部到全局的搜索策略，减小计算复杂度，降低时间资源和空间资源的浪费。

3、经实验比较，提出的算法能够达到较好的效果。与一些跟踪器比较，具有较好的性能改善。

2 研究方法

本节主要详细介绍提出的网络结构和实现方法。首先分析视觉目标跟踪的特点，需要说明的是，本文提出的方法对于目标的快速运动不稳定，在此基础上提出一个假设：在视频序列中，物体在相邻帧之间的位移不大。

事实上，这个假设对于大多数数据集来说是成立的。因为对于一个视频序列而言，相邻帧之间的时间间隔极小。因此，在极小的时间间隔里常规运动导致的位移很小。基于此假设，本文提出一种全新的目标跟踪策略。

2.1 网络结构

随着 Siamese 网络的提出，研究学者将该网络模型应用于视觉跟踪领域，得到很好的效果。首先是基于全卷积的 Siamese 网络，只有简单的几层就能够达到很好的效果。随后研究学者对其进行改进，将 AlexNet 加入 Siamese 网络中，得到一定的改善，但也遇到瓶颈。之后又将更深层次的 ResNet 替换浅层的 AlexNet。

如图 1 所示，SiamFC 网络结构由两个分支构成，一个是目标分支，输入数据为模板图像块（$z:127\times127\times3$）。另一个是搜索分支，输入数据为搜索图像块（$x:255\times255\times3$）。

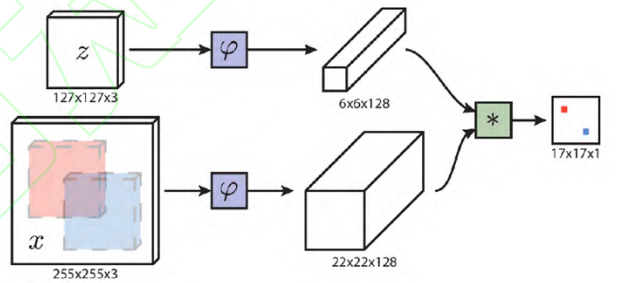


Fig.1 SiamFC network structure

图 1 SiamFC 网络结构

两个分支的卷积神经网络共享参数，确保相同的变换应用于不同的两个图像块。分别输出两个特征图 φ_Z 和 φ_X，为了结合两个分支的信息，对 ϕ_Z 和 φ_X 执行互相关操作，得到响应图 R，为了后续获得目标的位置信息和比例信息，需要 R 包含大量的特征信息。因此，响应图 R 为：

$$R=\varphi_X * \varphi_Z \tag{1}$$

根据 Siamese 网络的结构特点，将目标跟踪问题分为两个分支：分类分支、回归分支。

如图 2 所示，我们在 Siamese 网络的每个分支同时构建 AlexNet 浅层网络和 ResNet 深层网络。低层次特征如边缘、角、颜色、形状等代表较好的视觉属性，是定位不可或缺的特征，而高层次特征对语义属性具有较好的表征，对识别更为关键。

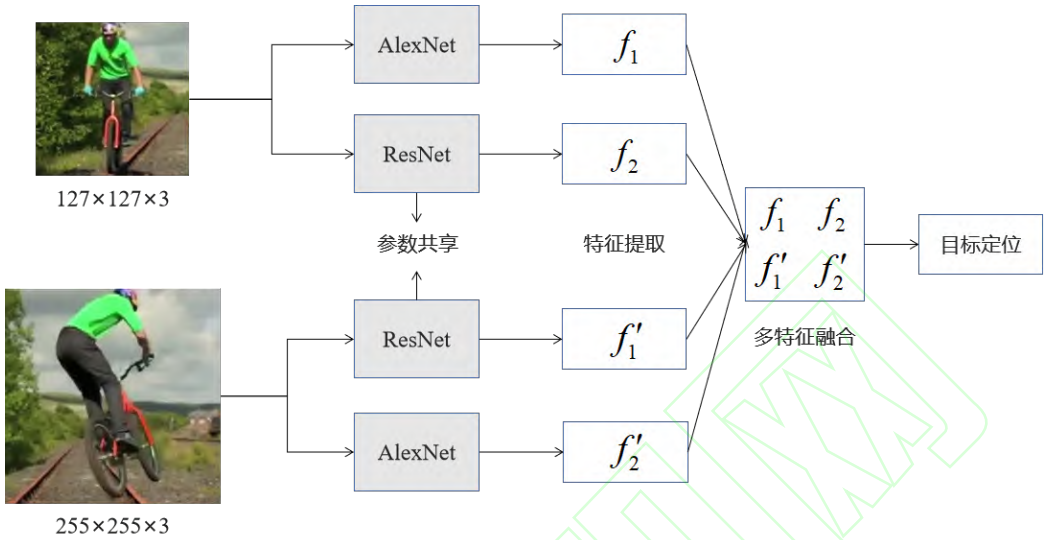


Fig.2 Improved Siamese adaptive network structure

图 2 改进的 Siamese 自适应网络结构

在 Siamese 自适应网络学习过程中，根据图像帧背景的复杂程度，对网络设置不同的权重。对于浅层的 AlexNet 网络设置权重为 α，对于深层的 ResNet 网络设置权重为 β。

当正样本数量大于负样本数量时，赋予 α 较大值。当图像帧背景复杂，负样本数量大于正样本数量时，赋予 β 较大值。则选择用于特征提取的神经网络为：

$$C_k=\frac{\alpha C_A}{\alpha C_A+\beta C_R} \tag{2}$$

$$C_l=\frac{\beta C_R}{\alpha C_A+\beta C_R} \tag{3}$$

$$C=\max(C_k,C_l) \tag{4}$$

其中，C_A 表示选择 AlexNet 网络，C_k 为该网络得分。C_R 表示选择 ResNet 网络，C_l 为该网络得分。根据网络得分可以得知该帧图像背景复杂程度，进而选择两个网络得分较大值用于该图像帧特征提取的网络。

在本文算法中，为了适应提出的网络结构，需要对 ResNet-50 作为主干网络并进行修改。基本的残差单元如图 3 所示。

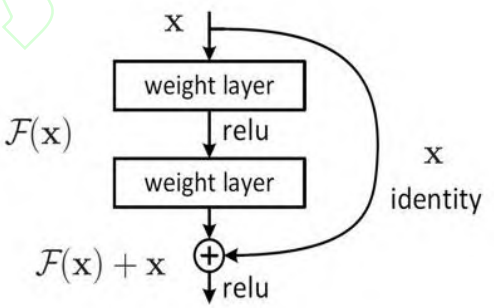


Fig.3 Residual unit structure

图 3 残差单元结构

由图 3 可以看出，X 为上一层特征图的输出。跳转连接，被称为 Identity Function。$G(X)=F(X)+X$ 为深层输出。

原始的 ResNet-50 的总步长为 32，与本文构建的网络结构不匹配。因此将 cov4 和 cov5 的步长改为 1，使得总步长减少为 8。并且对每个块添加步长为 1 的卷积层。将 cov3-3、cov4-6、cov5-3 的特征图输出，用于计算分类和回归。

改进的残差网络结构如图 4 所示，以 cov5 为例。

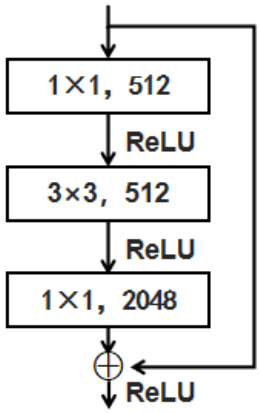


Fig.4 Improved residual unit structure

图 4 改进的残差单元结构

2.2 多特征融合

在神经网络中，一般浅层特征如边缘、颜色、形状等，包含更多的位置信息，是用于定位不可或缺的特征。而深层特征鲁棒性好，包含更多的语义信息，对识别更为关键。

在本文算法中，提出两个多特征融合的方法，其一：两个分支含有完全相同的两个网络。对于不同的输入，提取更加完善、多样性的特征。最后对两个分支得到的特征进行相加融合。通过将浅层特征和深层特征共同使用，能够更好的进行识别。其二：对于不同的图像帧，由于其背景复杂程度不同，对于简单背景来说，浅层特征可以轻松识别、定位目标。但是如果选择使用深层特征则会造成大量的时间消耗，增加计算复杂度，降低目标跟踪速度。因此，进行自适应的特征选择和多特征融合，根据图像帧的复杂程度，自动选择使用浅层特征、深层特征还是浅层、深层混合使用。

给定图像 I_j 、使用哪些特征进行组合的决策 $P(\cdot)$。因此，每个特征上的响应图为：

$$R^{f_j} = P(I_j) \tag{5}$$

其中，$f_j \in \{L_j, H_j, M_j\}$。$L$ 、H 、M 分别表示为浅层特征、深层特征、混合特征。

对于改进的 ResNet 网络提取的深层特征进行加权总和，最终融合得到的自适应特征图 ψ 为：

$$\psi = \sum_{i=3}^{5} \alpha_i C_i^k \oplus \sum_{i=3}^{5} \beta_i C_i^l \tag{6}$$

其中，α_i 、β_i 为每个图对应的权重，与网络一起参与训练，$\oplus$ 表示特征融合操作。进一步为了确保网络自主学习每个特征图的重要性，运用 SoftMax 函数规范化权重，表示每个特征图的重要性：

$$S_i = \sum_i \frac{e^{w_i}}{\sum_j e^{w_j}} \cdot C_i \tag{7}$$

其中，w_i 和 w_j 表示学习的权值。C_i 表示第 i 层的特征。

通过以上方法，能够得到更加精确、更加精细的特征，用于特定的图像帧进行鲁棒、快速的识别和定位。

2.3 边界框回归

本文通过端到端的完全卷积来训练网络，直接对每个目标位置进行分类和回归，避免了人工干预和多余的参数调整。我们用交叉熵损失用于分类，用具有标准化坐标的 Smooth L1 损失用于回归。对于跟踪数据集来说，每个图像帧都有已标注的真实边界框。因此，用 T_w 、T_h 、(x_1, y_1) 、(x_0, y_0) 、(x_2, y_2) 分别表示真实边界框的宽度、高度、左上角坐标、中心点坐标、右下角坐标。则以 (x_0, y_0) 为中心，$\frac{T_w}{2}$ 、$\frac{T_h}{2}$ 为轴长，可以得到椭圆 Q_1：

$$\frac{(p_i - x_0)^2}{(T_w/2)^2} + \frac{(p_j - y_0)^2}{(T_h/2)^2} = 1 \tag{8}$$

同理，以(x_0,y_0)为中心，$\frac{T_w}{4}$、$\frac{T_h}{4}$为轴长，可以得到椭圆Q_2：

$$\frac{(p_i-x_0)^2}{(T_w/4)^2}+\frac{(p_j-y_0)^2}{(T_h/4)^2}=1 \tag{9}$$

此时，如果目标位置(p_i,p_j)在椭圆Q_2内，则将其标记为正。如果在椭圆Q_1之外，则标记为负。如果位于椭圆Q_2和Q_1之间，则忽略不计。然后将标记为正的位置(p_i,p_j)用于边界框回归，回归目标可以公式化为：

$$D=\begin{cases}d_1=p_i-x_1\\ d_2=p_j-y_1\\ d_3=x_2-p_i\\ d_4=y_2-p_j\end{cases} \tag{10}$$

其中，d_1、d_2、d_3、d_4分别表示目标位置(p_i,p_j)到边界框四条边的距离。为此，我们定义多任务损失函数：

$$L=\lambda_1 L_c+\lambda_2 L_r \tag{11}$$

其中，L_c为交叉熵损失，L_r表示 Smooth L1 损失。在训练期间，我们根据多次实验设定$\lambda_1=1$，$\lambda_2=2$。

Smooth L1 损失函数如公式 12 所示：

$$smooth_{L_1}=\begin{cases}0.5x^2 & if\ |x|<1\\ |x|-0.5 & otherwise\end{cases} \tag{12}$$

3 实验与分析

3.1 数据集与评价指标

本文实验使用的数据集为目标跟踪标准数据集 VOT(Visual Object Tracking)[26]和 OTB(Object Tracking Benchmark)[27]，视频序列均经过精心标注，更具权威性。OTB 数据集包括 OTB50 和 OTB100。其中 50 和 100 代表该数据集中视频序列的数目。VOT 是官方竞赛的数据集，有 VOT2015、VOT2016 等，且每年均会更新。OTB 和 VOT 数据集存在一定的差别，其中 OTB 数据集含有 25%的灰度图像，VOT 中均为彩色图像。

本文实验主要使用以下 4 种评价指标对提出的算法进行分析。

（1）中心位置误差

中心位置误差（Center Location Error, CLE）是计算预测目标位置中心点和真实目标中心点之间的欧氏距离。假设真实目标的中心位置坐标为(x_g,y_g)，预测的目标中心位置坐标为(x_p,y_p)。因此，中心位置误差计算如下：

$$CLE=\sqrt{(x_p-x_g)^2+(y_p-y_g)^2} \tag{13}$$

一般来说，计算视频序列中所有图像帧的平均中心位置误差，在一定程度上能够近似看作是目标跟踪准确度。但是，跟踪算法在某些图像帧中难免会丢失目标，造成跟踪目标中心位置的预测具有随机性。因此，此时的平均中心位置误差值难以评价跟踪器的准确性。为此，在中心位置误差的基础上，通常采用准确率拟合曲线来反映跟踪器的准确度，统计不同阈值下，成功跟踪目标的中心位置误差的比例，使用误差阈值为 20 个像素点时所对应的数值，作为跟踪算法在各个测试视频序列中的准确率。

（2）精确度

精确性表示目标跟踪算法预测的目标框与真实目标框的重叠程度，数值越大，表示该算法的精确性更好，如公式 14 所示。

$$\phi_t(i)=\frac{1}{N}\sum_{k=1}^{N}\phi_t(i,k) \tag{14}$$

其中，$\phi_t(i,k)$表示经过k次重复后，第t帧图像的精确性，N表示重复的次数。则平均准确率为：

$$\rho_A(i)=\frac{1}{M}\sum_{t=1}^{M}\phi_t(i) \tag{15}$$

其中，M 表示有效跟踪图像帧的数量。

（3）成功率

成功率用预测框和真实框之间的交并比表示。通过重叠率（Overlap Ratio，OR）表示预测目标区域和真实目标区域的重叠比率。即两个边界框的交并比，如公式 16 所示。

$$OR=\frac{R\cap G}{R\cup G} \tag{16}$$

其中，OR 表示区域重叠比率，R 表示预测目标区域，G 表示真实目标区域。

（4）速度

在目标跟踪领域，跟踪速度通常指算法所用时间与视频序列帧数的比值，即平均每秒跟踪的视频帧数，值越大表示跟踪的速度越快。

3.2 实验

该方法基于 PyTorch 框架在 Python 中实现。实验设备硬件为一台装备 NVDIA Titan X 显示处理核心并配备 i7-7700k 处理器的计算机。

构建的网络在 ImageNet[28] 上进行了预训练，然后使用该参数作为初始化来重新训练我们构建的网络模型。在 OTB-50 数据集上实现的一些具有代表性的跟踪效果如图 5 所示。

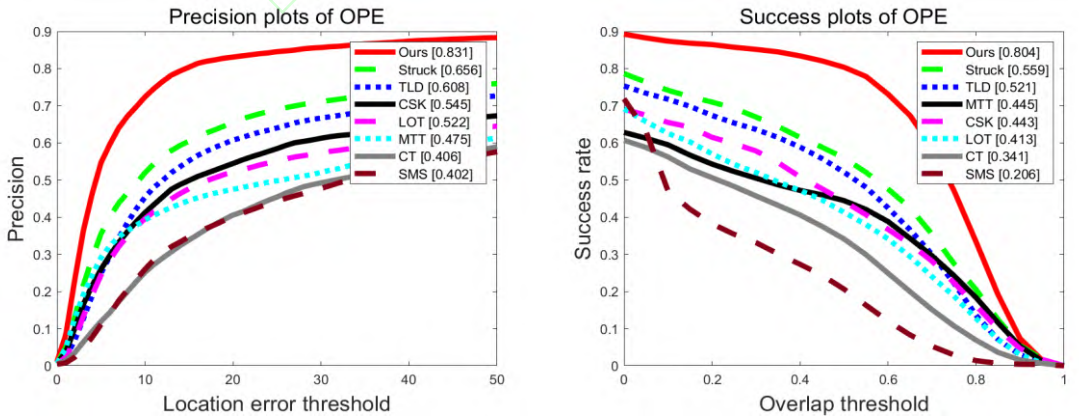
Fig.5 Tracking performance comparison

图 5 跟踪效果比较

由图 5 可以看出，在各种影响因素下，本文算法能够稳定地跟踪目标。

我们将提出的算法与现有的跟踪器在标准数据集上进行公平比较。用一次性通过评估（One-Pass Evaluation，OPE）绘制精度曲线图和成功曲线图。由图 6 可以看出，我们提出的方法比一些现有跟踪器的效果好，能够在一些影响因素下进行鲁棒跟踪。

我们在 OTB 数据集上对不同算法在形变、背景杂波、遮挡等影响因素下进行测试，绘制精度图和成功图。如图 7、图 8、图 9 所示。

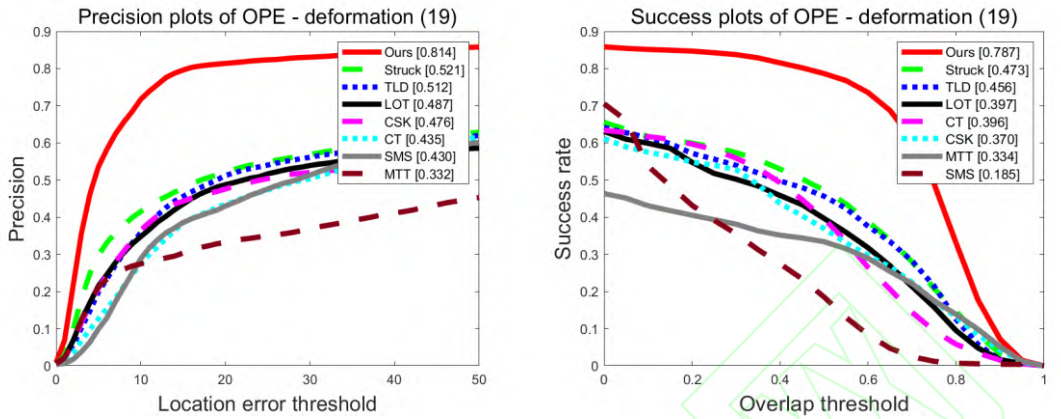


(a) Precision (b) Success rate

(a) 精确度 (b) 成功率

Fig.6 Accuracy graph and success graph

图 6 精度曲线图和成功曲线图

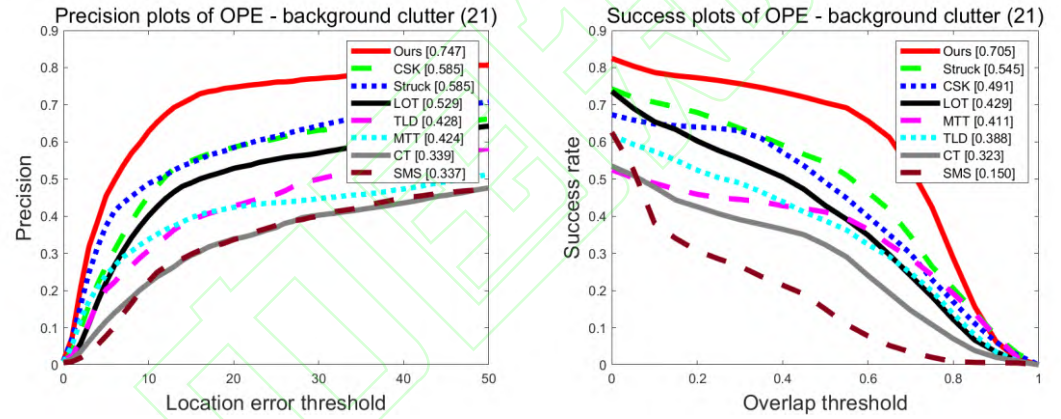


(a) Precision (b) Success rate

(a) 精确度 (b) 成功率

Fig.7 Accuracy graph and success graph drawn under the influence of deformation

图 7 形变影响下的精度曲线图和成功曲线图

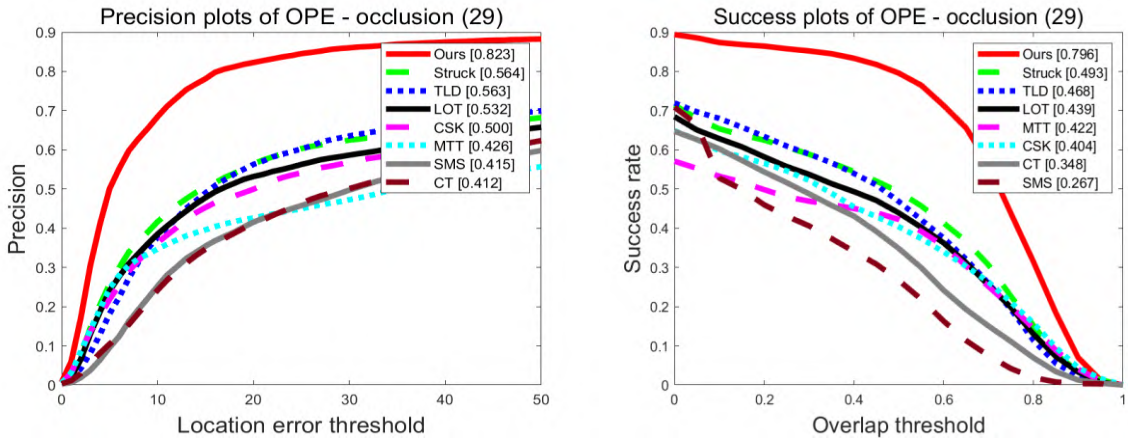


(a) Precision (b) Success rate

(a) 精确度 (b) 成功率

Fig.8 Accuracy graph and success graph drawn under the influence of background clutter

图 8 背景杂波影响下的精度曲线图和成功曲线图



(a) Precision (b) Success rate

(a) 精确度 (b) 成功率

Fig.9 Accuracy graph and success graph drawn under the influence of occlusion

图 9 遮挡影响下的精度曲线图和成功曲线图

将本文提出的算法与已有跟踪器：Struck (Struck: Structured Output Tracking with Kernels)[29]、LOT(Locally Orderless Tracking)[30]、TLD(P-N learning: Bootstrapping binary classifiers by structural constraints)[31]、CT(Real-Time Compressive Tracking)[32]、SMS(Mean-shift blob tracking through scale space)[33]、MTT(Robust visual tracking via multi-task sparse learning)[34]、CSK(Exploiting the Circulant Structure of Tracking-by-Detection with Kernels)[35]在精确度、成功率、速度三项指标在 OTB 数据集上进行详细评估，如表 1 所示。为了公平评估，均使用各指标的平均值。

Table 1 Performance comparison between the proposed algorithm and existing trackers

表 1 提出的算法与已有跟踪器性能对比

	成功率	精度	速度/fps
Ours	0.804	0.831	63.82
Struck	0.559	0.656	14.45
LOT	0.413	0.522	35.62
TLD	0.521	0.608	45.53
CT	0.348	0.406	20.59
SMS	0.185	0.337	89.57
MTT	0.445	0.475	49.83
CSK	0.443	0.545	94.66

由表 1 显示，本文提出的算法在保证速度的前提下，能够实现较好的跟踪准确性和成功率。且对光照变化、形变、背景杂波、遮挡等影响较鲁棒。

3.3 消融研究

通过对提出的算法和已有的跟踪器进行实验比较，发现本文提出的算法实现效果较好，提出的跟踪方法可行性较高。为此，本节从网络结构、特征图选择、跟踪方法等方面对提出的方法进行内部比较。用 N 表示网络结构，N_1 表示 Siamese 网络每个分支仅使用 AlexNet 网络。N_2 表示每个分支同时使用 AlexNet 和 ResNet。N_3 表示每个分支同时使用 AlexNet 和改进的 ResNet。F 表示用于识别提取到的特征，其中 F_1 表示仅用浅层特征、F_2 表示仅用深层特征、F_3 表示本文提出的多特征融合。M 表示跟踪方法，M_1 表示全局搜索、M_2 表示本文提出的由局部到全局的搜索方法，实验结果如表 2 所示。

Table 2 Algorithm internal comparison

表 2 算法内部比较

	成功率	精度	速度/fps
N1	0.586	0.642	80.35
N2	0.713	0.686	68.61
N3	0.786	0.813	62.54
F1	0.467	0.516	98.96
F2	0.424	0.498	25.31
F3	0.658	0.579	56.32
M1	0.864	0.882	48.69
M2	0.753	0.806	92.73

由表 2 数据可得，改进的网络结构和多特征融合的方法能够极大提高目标跟踪的精确度和成功率，但是跟踪速度会有所下降。提出的由局部到全局的搜索策略导致在准确度和成功率上效果不明显，但是在跟踪速度上有明显提升。

4 结束语

本文针对目标跟踪领域存在的跟踪精度和跟踪速度不平衡问题，以 Siamese 网络为基础，构建结合 AlexNet 网络和改进的 ResNet 网络的 Siamese 自适应网络。通过对提取到的特征进行多特征融合和自适应选择提高特征图的高效性，提高网络的识别和定位能力。进一步，通过加入一种由局部到全局的搜索策略，极大地降低网络计算复杂度，能够节约时间资源和空间资源。在目标跟踪标准数据集上进行实验对比，结果表明，本文提出的算法能够实现较好的效果，同时在形变、背景杂波、遮挡等影响因素下具有较强的鲁棒性。下一步工作将对实现超高的跟踪精确度进行深入研究。

参考文献：

[1] CHEN C, DENG Z H, GAO Y L, et al. Single target tracking algorithm based on multi- fuzzy kernel fusion[J]. Journal of Frontiers of Computer Science and Technology, 2020, 14(5): 848-860.
陈晨，邓赵红，高艳丽，等. 多模糊核融合的单目标跟踪算法[J]. 计算机科学与探索, 2020, 14(5): 848-860.

[2] LU H C, LI P X, WANG D. Visual object tracking: a survey[J]. Pattern Recognition and Artificial Intelligence, 2018, 31(1): 61-76.
卢湖川，李佩霞，王栋. 目标跟踪算法综述[J]. 模式识别与人工智能, 2018, 31(1): 61-76.

[3] Yang H, Shao L, Zheng F, et al. Recent advances and

trends in visual tracking: a review[J]. Neurocomputing, 2011, 74(18): 3823-3831.
[4] Kalal Z, Mikolajczyk K, Matas J. Tracking learning-detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(7): 1409-1422.
[5] Li B, Yan J, Wu W, et al. High performance visual tracking with siamese region proposal network[C]. IEEE Conference on Computer Vision and Pattern Recognition. 2018,26(3):8971-8980.
[6] Bertinetto L,Valmadre J,Henriques J F,et al. Fully-Convolutional Siamese Networks for Object Tracking[C].European Conference on Computer Vision,2016,32(2):850-865.
[7] Zhu Z, Wang Q, Li B, et al. Distractor-aware Siamese Networks for Visual Object Tracking[C].European Conference on Computer Vision,2018,43(1):103-119.
[8] Li B,Wu W,Wang Q,et al. SiamRPN++:evolution of siamese visual tracking with very deep networks[C].IEEE Conference on Computer Vision and Pattern Recognition, 2019, 23(3):4282-4291.
[9] Guo D, Wang J, Cui Y, et al. SiamCAR: Siamese Fully Convolutional Classification and Regression for Visual Tracking[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).IEEE, 2020,26(1):1-13.
[10] Smeulders A W M, Chu D M, Cucchiara R, et al. Visual tracking: an experimental survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 36(7):1442-1468.
[11] Wang H Q, Cai Y N, Sun F C, et al. Adaptive sequence learning and application of multiscale kernel method[J]. Pattern Recognition and Artificial Intelligence, 2011, 24(1):72-81.
汪洪桥, 蔡艳宁, 孙富春, 等. 多尺度核方法的自适应序列学习及应用[J]. 模式识别与人工智能, 2011, 24(1): 72-81.
[12] Wang L, Wang J P, Wang P, et al. Research on target tracking algorithm of twin network fusion attention mechanism[J]. Computer Engineering and Applications, 2021, 57(08):169-174.
王玲,王家沛,王鹏,孙爽滋.融合注意力机制的孪生网络目标跟踪算法研究[J].计算机工程与应用, 2021, 57(08): 169-174.
[13] Shan Y G, Hu W G. A Survey of Adaptive Vision Target Tracking Methods in Scale and Direction[J]. Computer Engineering and Applications, 2020, 56(9): 13-23.
单玉刚, 胡卫国. 尺度方向自适应视觉目标跟踪方法综述[J]. 计算机工程与应用, 2020, 56(9): 13-23
[14] João F. Henriques, Caseiro R, Martins P, et al. High-Speed Tracking with Kernelized Correlation Filters[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014:583-596.
[15] P Horst, M Thomas, and B Horst. In defense of color-based model-free tracking[J]. In CVPR, 2015: 2113-2120.
[16] Valmadre J, Bertinetto L, Henriques, João F, et al. End-to-End Representation Learning for Correlation Filter Based Tracking[C]. 2017,32(2):5000-5008.
[17] J Y Gao, T Z Zhang, and C S Xu. Graph convolutional tracking[C]. In CVPR, 2019,2(3):4644-4654.
[18] L Zhang, V Jagannadan, N S Ponnuthurai, et al. Robust visual tracking using oblique random forests[C]. In CVPR, 2017,21(4):5825-5834.
[19] Danelljan M, Bhat G, Khan F S, et al. ATOM: Accurate Tracking by Overlap Maximization[C].2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2019,3(4):4655-4664.
[20] MAC, HUANG J B, YANG X K, et al. Hierarchical convolutional features for visual tracking[C]//Proceedings of the 2015 IEEE International Conference on Computer Vision, Santiago, Dec 7-13, 2015. Washington: IEEE Computer Society,2015: 3074-3082.
[21] Bolme D S, Beveridge J R, Draper B A, et al. Visual object tracking using adaptive correlation filters[C]// The Twenty-Third IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2010, San Francisco, CA, USA, 13-18 June 2010. IEEE, 2010:2544-2550.
[22] Li F, Yao Y, Li P, et al. Integrating Boundary and Center Correlation Filters for Visual Tracking with Aspect Ratio Variation[C].IEEE International Conference on Computer Vision Workshop.2017,4(2):2001-2009.
[23] Zhang J, Ma S, Sclaroff S. MEEM: Robust Tracking via Multiple Experts Using Entropy Minimization[C]. European Conference on Computer Vision, 2014, 35(3): 188-203.
[24] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6):84-90.
[25] K. He, X. Zhang, S. Ren and J. Sun, Deep Residual Learning for Image Recognition[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).2016,45(3):770-778.
[26] Kristan M, Leonardis A, Matas J, et al. The visual object tracking vot2017 challenge results[C] // Proceedings of the IEEE International Conference on Computer Vision Workshops. 2017,52(5): 1949-1972.
[27] Wu Y, Lim J, Yang M-H. Object tracking benchmark[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1834-1848.
[28] 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.
[29] Hare S, Golodetz S, Saffari A, et al. Struck: Structured Output Tracking with Kernels[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 65(5): 2096-2109.
[30] Oron S, Bar-Hillel A, Dan L, et al. Locally Orderless Tracking[C]. International Journal of Computer Vision, 2015, 111(2):213-228.
[31] Kalal Z, Matas J, Mikolajczyk K. P-N learning: Boot-

strapping binary classifiers by structural constraints[C]// Computer Vision & Pattern Recognition. IEEE, 2010, 32(5): 49-56.

[32] Zhang K H, Zhang L. Real-Time Compressive Tracking[C]//European Conference on Computer Vision, 2012, 65(1): 866-879.

[33] Collins R T. Mean-shift blob tracking through scale space[C]// IEEE Computer Society Conference on Computer Vision & Pattern Recognition. IEEE, 2003, 102(2): 229-234.

[34] T Zhang, B Ghanem, S Liu ,et al. Robust visual tracking via multi-task sparse learning[C].2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012,43(6):2042-2049.

[35] Henriques J F, Rui C, Martins P, et al. Exploiting the Circulant Structure of Tracking-by-Detection with Kernels[C]//Proceedings of the 12th European conference on Computer Vision - Volume Part IV. Springer, Berlin, Heidelberg, 2012,53(4):702-715.

李睿（1971-），女，甘肃秦安，硕士，教授，主要研究方向为智能信息处理、模式识别与人工智能。

LI Rui, born in 1971, M.S., professor. Her research interests include intelligent information processing, Pattern recognition and artificial intelligence.

连继荣（1995-），男，甘肃定西，硕士研究生，主要研究方向为模式识别与人工智能。

LIAN Jirong, born in 1995, M.S. candidate, His research interests include Pattern recognition and artificial intelligence.