# 基于自适应深度稀疏网络的在线跟踪算法

侯志强 王 鑫 余旺盛 戴 铂 金泽芬芬 (空军工程大学信息与导航学院 西安 710077)

摘 要:视觉跟踪中,高效鲁棒的特征表达是解决复杂环境下跟踪漂移问题的关键。该文针对深层网络预训练复杂 费时及单网络跟踪易漂移的问题,在粒子滤波框架下,提出一种基于自适应深度稀疏网络的在线跟踪算法。该算法 利用 ReLU 激活函数,针对不同类型目标构建了一种具有自适应选择性的深度稀疏网络结构,仅通过有限标签样 本的在线训练,就可得到鲁棒的跟踪网络。实验数据表明:与当前主流的跟踪算法相比,该算法的平均跟踪成功率 和精度均为最好,且与同样基于深度学习的 DLT 算法相比分别提高了 20.64%和 17.72%。在光照变化、相似背景 等复杂环境下,该算法表现出了良好的鲁棒性,能够有效地解决跟踪漂移问题。 关键词:视觉跟踪;在线训练;深度学习;自适应深度稀疏网络 中图分类号:TP391 文献标识码:A 文章编号:1009-5896(2017)05-1079-09 DOI: 10.11999/JEIT160762

# **Online Visual Tracking via Adaptive Deep Sparse Neural Network**

HOU Zhiqiang WANG Xin YU Wangsheng DAI Bo JIN Zefenfen (Institute of Information and Navigation, Air Force Engineering University, Xi'an 710077, China)

Abstract: In visual tracking, the efficient and robust feature representation is the key factor to solve the problem of tracking drift in complex environments. Therefore, to solve the problems of the complex and time-consuming of the pre-training process of deep neural network and the drift of the single network tracking, an online tracking method based on an adaptive deep sparse network is proposed under the tracking structure of particle filter. A deep sparse neural network architecture, which can be adaptively selected according to different types of targets, is constructed with the implementation of the Rectifier Linear Unit (ReLU) activation function. The robustness of deep tracking network can be easily achieved only through the online training of limited labeled samples. The results of experiments show that, compared with the state-of-the-art tracking algorithm, the average success ratio and precision of the proposed algorithm are both the highest, and they are raised by 20.64% and 17.72% respectively contrasted with the Deep Learning Tracker (DLT) algorithm based on deep learning. The proposed method can solve the problems of tracking drift efficiently, and shows better robustness, especially for the complex environment such as illumination changes, background clutter and so on.

Key words: Visual tracking; Online training; Deep learning; Adaptive deep sparse network

1 引言

视觉跟踪技术是当前计算机视觉及其相关领域 的研究热点之一<sup>[1,2]</sup>,广泛应用于视频监控、人机交 互、医学成像等领域。其目的是给定视频中目标的 初始状态(如位置、尺寸等),通过估计后续帧中目 标的状态来实现目标的跟踪。

近年来,视觉跟踪技术取得了很大的进步,

IVT<sup>[3]</sup>, MIL<sup>[4]</sup>, TLD<sup>[5]</sup>, CT<sup>[6]</sup>等多种跟踪算法的出现, 在简单可控环境下取得了比较满意的跟踪结果。但 在实际复杂环境中,目标会受到光照变化、形变、 遮挡、运动模糊等<sup>[7,8]</sup>多种因素的影响,这些算法往 往会出现跟踪漂移或者目标丢失现象。如何在复杂 条件下对目标进行更为鲁棒、快速的跟踪,仍是当 前视觉跟踪领域极具挑战性的问题。

深度学习<sup>[9]</sup>的出现为解决上述问题提出了一种 新的思路。与浅层网络相比,深度学习利用其深层 架构自动地从样本中学习更加抽象和本质的特 征<sup>[10]</sup>,成为了视觉跟踪领域的一个新趋势。Jin 等 人<sup>[11]</sup>通过训练一个两层的神经网络,提取目标的尺 度不变特征,在长时间持续稳定跟踪中取得了较好 的效果;文献[12]基于消噪自编码器原理,提出了一

收稿日期 :2016-07-20 ;改回日期 :2016-12-16 ;网络出版 :2017-02-28 \*通信作者:王鑫 wangxiin@foxmail.com

基金项目:国家自然科学基金(61473309),陕西省自然科学基础研究 计划项目(2015JM6269,2016JM6050)

Foundation Items: The National Natural Science Foundation of China (61473309), The Project Supported by Natural Science Basic Research Plan in Shaanxi Province (2015JM6269, 2016JM6050)

种新的跟踪算法—DLT(Deep Learning Tracker), 其采用离线训练与在线跟踪相结合的方法,得到了 更加鲁棒的跟踪结果。

然而,基于深度学习架构的跟踪方法还存在一 些难点问题。首先,深层网络的训练需要大量的样 本数据<sup>[13,14]</sup>,但实际问题中标签样本十分有限,容 易产生"过拟合"。尽管非监督预训练方法<sup>[9]</sup>可以较 好地解决这个问题,但其对实验硬件要求比较高, 且训练过程相当费时和复杂。其次,利用梯度下降 法训练深层网络时,传统的 BP 算法<sup>[15]</sup>容易出现"梯 度消失"问题,导致难以完成深层网络的训练。另 外,传统的基于单一网络进行跟踪的方法,容易丢 失原有的结构信息,一定程度上更易造成跟踪漂移。

针对以上问题,本文提出一种基于自适应深度 稀疏网络的在线跟踪方法。该方法利用 ReLU 激活 函数<sup>[16,17]</sup>构造深度稀疏网络,网络本身具有高度稀 疏性,仅通过有限样本的在线训练就可以得到鲁棒 的深层网络;同时由于 ReLU 梯度恒定,有效解决 了"梯度消失"问题;根据跟踪目标的不同类型, 构造自适应的多网络跟踪模型,较好地保留了目标 的结构信息,具有更好的鲁棒性和自适应性。实验 结果表明:本文算法不仅能较好地适应复杂背景和 目标外观、尺度变化情况,而且具有较高的跟踪速 度。

#### 2 深度稀疏网络模型

所谓深度稀疏网络<sup>[16]</sup>(deep sparse network),就 是指在构建之初就只有极少的神经元被激活的深层 网络结构,与传统神经网络相比,其具有更好的稀 疏性和鲁棒性,在深层架构中具有更好的表现性。 本节首先介绍了堆栈式消噪自编码器(Stacked Denoising Auto-Encoder, SDAE)的基本原理,然后 通过引入 ReLU(Rectifier Linear Unit)激活函数, 构建深度稀疏神经网络模型。

### 2.1 堆栈式消噪自编码器

SDAE <sup>[18]</sup>是一个由多层消噪自编码器(DAE)逐 层贪婪训练而成的深度神经网络结构。DAE<sup>[12,19]</sup>采 用无监督学习方法,通过学习叠加了噪声的腐蚀数 据 $\tilde{x}$ ,使得输出值 $\hat{x}$ 接近于原始数据x,从而得到 更加鲁棒的深层特征。DAE 的基本架构如图1所示。

对于训练样本  $x = \{x_1, x_2, \dots, x_m\}$ ,用 $x_i$ , $\tilde{x}_i$ 分 别表示第i个样本的原始数据和其加噪腐蚀数据, $\hat{x}_i$ 表示 $x_i$ 的重构输出,W, W', b, b'分别表示编码 器、解码器的权重矩阵和偏置向量。其编码和解码 过程可表示为

$$\begin{array}{c} h_i = f\left(\boldsymbol{W}\tilde{x}_i + \boldsymbol{b}\right) \\ \hat{x}_i = g\left(\boldsymbol{W}'h_i + \boldsymbol{b}'\right) \end{array}$$
(1)



其中,  $f(\cdot)$ 和  $g(\cdot)$ 分别为编码器和解码器的激活函数,通常为非线性函数。训练 DAE 的过程就是在样本集 x上寻找参数集 $\theta = \{W, W', b, b'\}$ ,使其代价函数最小化的过程<sup>[12]</sup>。本文的代价函数定义为

$$J(\boldsymbol{x}, \hat{\boldsymbol{x}}) = \sum_{i=1}^{m} \|\boldsymbol{x}_{i} - \hat{\boldsymbol{x}}_{i}\|_{2}^{2} + \lambda \left(\|\boldsymbol{W}\|_{\mathrm{F}}^{2} + \|\boldsymbol{W}'\|_{\mathrm{F}}^{2}\right) \qquad (2)$$

其中, m为样本的个数;  $\lambda$ 为权重惩罚因子, 用来 平衡重构损失和权重的关系;  $\|\|_{F}$ 为 F 范数。

堆叠多层 DAE,并利用逐层贪婪算法进行训练,得到堆栈式消噪自编码器(SDAE)。之后将所有 层作为一个整体,采用 BP 算法对 SDAE 进行微调, 即可得到更适应任务需要的深层网络。

2.2 ReLU 激活函数

激活函数是神经网络的核心所在,它通过模拟 生物神经元的激活特征,可以实现对任意非线性函 数的逼近。传统神经网络中,最常用的激活函数有 sigmoid 函数和 tanh 函数,其在信号的特征空间映 射上有着很好的效果。

随着神经生物学研究的不断推进,2001年,神 经科学家 Dayan 和 Abott 从生物学的角度提出了更 加精确的脑神经元激活模型。在此基础上,Glorot 等人<sup>[16]</sup>对传统激活函数进行了修正,提出了一种与 脑神经元模型更加相似的激活函数——ReLU函数。 其函数表达式为

$$\operatorname{ReLU}(z) = \max(0, z) \tag{3}$$

由图 2 可以看出,与 sigmoid 和 tanh 函数相比, ReLU 函数是一个单侧抑制函数,这种单侧抑制性 使其具有稀疏表达性<sup>[16]</sup>,有助于降低原始高维信息 的相关性,使数据具有更大可能线性可分,降低了 网络复杂度,有助于得到更加鲁棒的特征。其次,



由于 ReLU 的单侧响应端为非饱和的线性函数,其 梯度恒为 1,不用进行复杂的指数运算就可将梯度 很好地传递到较前面的网络,有效解决了传统神经 网络训练过程中的梯度衰减甚至消失问题<sup>[15]</sup>,同时 提高了网络的训练速度。

目前, ReLU 激活函数在图像检测<sup>[20]</sup>、语音识 别<sup>[17]</sup>以及图像去噪<sup>[21]</sup>等领域都取得了显著的效果。 在图像识别领域中, Glorot 通过实验证明,采用不 进行非监督预训练 ReLU 网络可以得到与采用非监 督预训练的 sigmoid 或 tanh 网络相近的结果<sup>[16]</sup>。这 是因为 ReLU 激活函数的应用,使得网络在构建之 初就具有了稀疏性,从而缩小了深层网络做与不做 非监督预训练之间的差距,更容易训练深层且稀疏 的网络结构。

#### 2.3 深度稀疏神经网络模型

深度稀疏网络构建之初就具有高度稀疏性,这种固有的稀疏性,可以等效于传统神经网络进行预训练所获得的稀疏性,从而缩小了模型做与不做预训练的差距。基于 ReLU 激活函数的稀疏表达性,在 SDAE 深层框架下,构建深度稀疏的神经网络模型,其结构如图 3 所示。

提高网络的稀疏性可以得到更加鲁棒的特征<sup>[16]</sup>。利用 ReLU 函数搭建的深层网络,在随机初始化后,约有 50%的隐含层节点被抑制。为进一步提高网络的稀疏性,我们在式(2)中添加稀疏惩罚项<sup>[17]</sup>:

$$L = \sum_{j=1}^{m_k} \rho\left(a_j\right) \tag{4}$$

$$\rho\left(a_{j}\right) = \lg\left(1 + a_{j}^{2}\right) \tag{5}$$

其中, $m_k$ 表示第k个隐含层节点数, $a_j$ 表示第k个 隐含层j个节点的激活值, $\rho(a_j)$ 表示 $a_j$ 的稀疏惩 罚。此时,通过求解式(6)的最优化问题,可以训练 出更加稀疏的深层网络。



$$\min_{\boldsymbol{W}, \boldsymbol{W}', \boldsymbol{b}, \boldsymbol{b}'} \left( \sum_{i=1}^{m} \| x_i - \hat{x}_i \|_2^2 + \lambda \left( \| \boldsymbol{W} \|_{\mathrm{F}}^2 + \| \boldsymbol{W}' \|_{\mathrm{F}}^2 \right) + \mu \cdot L \right)$$
(6)

其中, $\mu$ 为稀疏惩罚因子。

### 3 基于自适应深度稀疏网络的在线跟踪

本文将 ReLU 激活函数与粒子滤波跟踪框架相 结合,针对单一跟踪网络易出现跟踪漂移问题,提 出一种自适应的深度稀疏网络在线跟踪方法。所谓 自适应深度稀疏网络,是通过构造多个不同结构的 跟踪网络,根据目标的不同类型而自适应地选择跟 踪网络的网络结构。其更好地利用了目标的原始结 构信息,使得跟踪结果更加稳定。

3.1 构建自适应深度稀疏跟踪网络

视觉跟踪问题可以看作是一个跟踪目标与背景的二值分类问题,跟踪的目的是将图像中的目标与 背景区分开<sup>[22]</sup>。跟踪过程中,利用图 3 中的前向编 码网络可提取采样样本的鲁棒特征,然后添加 sigmoid 分类层,对样本进行分类,从而达到跟踪的 目的。跟踪模型的基本结构如图 4 所示。

利用这种单一网络进行跟踪时,为减少网络输入节点数量,降低复杂度,通常会在网络输入端对 初始目标及采样样本进行统一的归一化处理,如将 其归一化为 32 × 32 的低分辨率图像。但在实验中发现,这种对不同目标做统一归一化处理的方法,会 造成目标在水平或者垂直方向上产生不同程度的拉 伸,丢失了其原有的结构特征,使得跟踪过程中很 难得到足够和合适的正样本,容易产生跟踪漂移。 在测试视频"Subway"中,由图 5 的跟踪结果和中 心误差曲线可以看出,当采用 32 × 32 的归一化输入 进行跟踪时,在相似背景情况下,80 帧时出现了跟 踪漂移;但当采用 32 × 16 归一化输入时,目标始终 被正确跟踪。

通过实验可以看出,在跟踪过程中更好地利用目标的原有结构信息可以得到更加鲁棒的特征,从而提高跟踪的可靠性。同时由表1可以看出,当采用32×16归一化输入时,由于其输入节点的减少,使得网络结构变窄,网络训练速度和更新速度加快,平均跟踪速度由原来的10.70帧/s提高到了14.93帧/s。





图 5 Subway 测试视频跟踪结果对比 (红色、绿色矩形框分别表示 32×16, 32×32 归一化输入的跟踪结果)

表1 两种不同归一化输入方法跟踪速度对比(帧/s)

归一化输入方式	Basketball	Coke	Jogging1	Walking2	Subway	Woman	平均
32×32 归一化	10.39	9.80	11.53	11.55	9.36	11.58	10.70
32×16 归一化	14.05	13.64	17.42	16.66	11.70	16.12	14.93

基于此,本文提出了一种自适应的深度稀疏跟 踪网络。首先我们将跟踪目标分为 3 类,分类依据 为目标初始的宽高比r = w/h,其中w, h分别表示 目标的宽和高。当 $2/3 \le r \le 3/2$ 时为 类目标,当 r < 2/3时为 类目标,当r > 3/2时为 类目标。 针对这 3 类跟踪目标,在传统单一跟踪网络的基础 上,搭建 3 个不同结构的跟踪网络,网络的基本结 构如图 6 所示。

跟踪开始时,根据r值大小确定目标类别,若 为类目标,则选择跟踪网络进行跟踪;若类,

类目标,则分别选择跟踪网络,进行跟踪。 在跟踪网络输入端,分别对 类、 类、 类目标 进行 32 × 32, 32 × 16, 16 × 32 的归一化处理,以减 少目标的形变,更好地利用其结构特征训练和更新 网络。

对于层间节点的连接权重,我们对其进行随机

初始化,使得初始化后的权重为[-0.5,0.5]的微量权 重,且不同层间的权重不同。

#### 3.2 在线训练网络

通过网络选择算法确定跟踪网络后,随机初始 化的网络并不能满足特定的跟踪任务。因此,为了 使跟踪网络能够更好地发挥性能,要使用特定的标 签样本对跟踪网络进行在线训练,调整更新网络参 数。

在特定的视觉跟踪任务中,仅知道目标初始状态<sub>s0</sub>,为了提高跟踪网络的鲁棒性,在首帧图像中,除了采集目标样本(正样本),还需要采集背景样本 (负样本)进行训练。我们在跟踪目标附近±2个像素 点的位置采集 10 个正样本,在距离目标稍远位置随 机采样得到 100 个负样本,分别构建正、负样本库, 样本尺寸与目标初始尺寸一致。



图 6 自适应深度稀疏跟踪网络

利用采集的正负样本作为标签数据在线训练网 络参数,就可以得到更适应跟踪任务的跟踪网络。 3.3 粒子滤波算法

粒子滤波算法<sup>[23]</sup>是视觉跟踪中常用的算法。假 设 $s_t$ ,  $z_t$ 分别表示目标在t时刻的目标状态和观测 值,目标跟踪过程可以看成是根据观测序列 $\{z_{1:t}\}$ 寻 找t时刻目标最大可能状态的过程:

$$s_t = \arg\max p\left(s_t \,|\, z_{1:t}\right) \tag{7}$$

其中,  $p(s_t|z_{1:t})$ 表示目标在t时刻的后验概率分布, 根据贝叶斯准则,可以得到

$$p(s_t|z_{1:t}) = \frac{p(z_t|s_t) p(s_t|z_{1:t-1})}{p(z_t|z_{1:t-1})}$$
(8)

而粒子滤波通过一组带有权值的随机粒子集  $\left\{s_{t}^{i}, w_{t}^{i}\right\}_{i=1}^{N}$ 来逼近目标状态的后验概率分布。其中, 粒子初始权值为 1/N,由于粒子滤波容易出现权值 退化问题,因此在跟踪过程中通过式(9)更新粒子权 值,且使其和始终保持为 1。

$$w_t^i = w_{t-1}^i \cdot \left( z_t \big| s_t \right) \tag{9}$$

$$\sum_{i=1}^{N} w_t^i = 1$$
 (10)

在本文算法中,利用粒子滤波算法进行随机采 样,将N个采样样本送入跟踪网络,得到每个采样 粒子的置信度 $s_i = p(s_t | z_{1:t})$ ,即其后验概率。利用 式(7),我们选择 $s_i$ 最大的粒子作为当前帧目标的估 计位置,即得到了目标的跟踪结果。

3.4 网络更新策略

跟踪过程受到光照变化、物体遮挡、目标形变 等因素影响时,容易出现跟踪漂移,此时必须对网 络参数进行更新。网络更新条件为

$$\max(\varsigma_i) < \boldsymbol{\tau} \mid\mid fn \ge \eta \tag{11}$$

其中, $\tau$ 为更新阈值,fn为上一次更新后的累计跟踪帧数, $\eta$ 为最大累计帧数。

更新策略为:将当前跟踪结果加入正样本库, 重新随机采样构建负样本库,此时共有 11 个正和 99 个负样本;利用更新后的正负样本库重新训练跟 踪网络,实现网络参数的更新。

3.5 算法流程

基于上述构建的自适应深度稀疏网络,本文跟 踪算法的主要步骤如表2所示,算法流程如图7。

## 4 仿真实验

为充分验证本文算法的有效性,我们在 Windows 操作系统下,采用 MATLAB 2014a 编程 实现本文算法。在 Intel Xeon 2.4 GHz 的处理器上 对本文算法进行测试,并利用了 GPU(TITANX)进

#### 表 2 基于自适应深度稀疏网络的在线跟踪

输入:图像序列 $I_1, I_2, \cdots, I_n$ ,目标初始状态 $s_0$ 。
输出:每帧图像中目标的跟踪结果,即目标的估计状态 $\hat{s}_i$ 。
$步骤 1$ 根据目标初始状态 $s_0$ 选择跟踪网络类型;
步骤 2 采集正负样本,在线训练跟踪网络;
步骤 3 对于 <i>i</i> = 1,2,…, <i>n</i> , do:
$(\mathbf{a})$ 在 $(x_{i-1},y_{i-1})$ 附近进行粒子采样,得到 $N$ 个采样样本;
(b)将 N个采样样本送入相应跟踪网络,得到每个粒子置信
度 <sub>si</sub> ;
(c)利用式(7),选择置信度最大的粒子所在位置作为跟踪结
果;
(d) <b>根据式</b> (11) <b>对网络进行更新;</b>
the second se

步骤4 结束。



图 7 基于自适应深度稀疏网络的跟踪算法

行加速。测试参数设置如下:粒子采样中设定粒子 数为 N=1000,置信度阈值  $\tau=0.9$ ,学习率  $\xi=0.01$ , 惩罚因子  $\lambda=0.002$ ,稀疏惩罚因子  $\mu=0.2$ ,最大累计 帧数  $\eta=50$ 。

在文献[7]中给出的 51 组测试视频中,将本文算 法与其他 7 种主流跟踪算法进行对比分析。对比算 法包括:DLT<sup>[12]</sup>, MIL<sup>[4]</sup>, CT<sup>[6]</sup>, OAB<sup>[24]</sup>, Frag<sup>[25]</sup>, TLD<sup>[5]</sup>, IVT<sup>[3]</sup>。这些算法均使用默认参数。

4.1 定性分析

在 51 组实验结果中,我们选取 10 组具有代表 性的视频用于定性分析展示。图 8 给出了实验结果 的部分截图,左上角数字为图像帧数,不同算法的 跟踪结果用不同颜色的矩形框表示,红色为本文算 法。针对图 8 中的 10 组视频,下面从 7 种情况对算 法进行定性分析。

(1)目标形变:非刚体目标的形变,以视频 "Basketball"为例。跟踪目标始终处于运动状态, 且身体发生了不同程度的非刚体形变,在第 300 帧 以后,仅有本文算法和 DLT 能够正确跟踪目标,基 于分块思想的 Frag 算法只跟踪上了目标的一部分 区域,其余算法均跟踪失败。

(2)复杂背景:目标附近的背景与目标有很大的 相似性,以视频"CarDark","Walking2"为例。 在视频"CarDark"第 355 帧时,只有本文算法还 能正确跟踪目标,其余算法都出现了跟踪漂移;在 视频"Walking2"中,第 340 帧出现了与跟踪目标 相似的背景,CT,Frag,MIL,TLD 失效,493 帧相 似背景消失后,只有本文算法,DLT 和 OAB 能正 确跟踪目标,但 OAB 不能适应目标的尺度变化。

(3)运动模糊:目标运动过快导致目标区域变模 糊。以视频"Deer"为例,从第5帧开始,由于目 标的快速运动,除本文算法外,其余算法都出现了 不同程度的目标漂移;第29帧时,只有本文算法, MIL, OAB 能够正确跟踪目标,且 MIL, OAB 的跟 踪结果出现了偏离。

(4)光照变化:目标区域中的光照剧烈变化,以 视频"CarDark","Singer2"为例。背景光照的剧 烈变化,要求跟踪算法具有较高的鲁棒性,在2组



视频中,只有本文算法准确地完成了整个跟踪过程。

(5)尺度变化:目标在跟踪过程中的尺度大小发 生变化,以视频"Walking2"为例。由于跟踪目标 逐渐远离摄像机,目标尺度变得越来越小, "Walking2"中只有本文算法和DLT能在对目标准 确跟踪的同时,较好地适应目标尺度的变化。

(6)遮挡:目标被部分或者全部遮挡,以视频 "Jogging-2","FaceOcc2","Woman","Suv" 为例。当目标被全部遮挡或部分遮挡时,在目标重 新出现后,仅本文算法能够及时准确地跟踪目标。

(7)旋转:目标在图像平面内发生旋转。以视频 "MountainBike"为例,目标的旋转使得 CT 和 Frag 算法失效,但本文算法仍能始终准确地跟踪目标。 4.2 定量分析

本文分别对算法在单个视频下的跟踪性能和 51 组视频的综合性能进行定量分析。

(1)算法对单个测试视频的定量分析:针对上述 10 组测试视频,采用跟踪成功率和平均中心位置误 差 2 个评价指标对算法进行对比分析。成功率是指算 法在整个视频中成功跟踪的帧数占总帧数的百分比, 其大小由跟踪结果的覆盖率  $(|S_T \cap S_G|)/(|S_T \cup S_G|)$ 决定,其中 $S_G$ 和 $S_T$ 分别表示目标的真实区域和跟踪 器的跟踪区域, $\cap$ 和 $\cup$ 分别表示两个区域的交集和 并集, $\bowtie$ 表示区域的面积。当其覆盖率大于给定阈 值 $t_0$ 时,则认为此帧跟踪成功。平均中心位置误差 是指跟踪结果的中心位置与手工标定的准确位置之 间的平均欧氏距离,其值越小表示算法跟踪准确率 越高。

表3给出了8种跟踪算法对于10组视频的跟踪 对比结果,括号前的数字表示覆盖率为0.5时的成 功率(%),括号内的数字表示平均中心位置误差(像 素)。红色为最好的结果,绿色为次好结果可以看出, 在大多数视频中,本文算法保持了较高的跟踪成功 率和较低的平均中心位置误差,表现出较好的跟踪 性能。

(2)算法综合性能的定量分析:针对文献[7]中给 出的 51 组视频进行对比测试,我们采用跟踪精度和 成功率 2 个通用的评价指标来对算法的综合性能进 行对比分析。其中,精度指的是跟踪算法在给定的 平均中心位置误差之内的帧数占总帧数的百分比。 实验过程中,根据第1帧中的准确位置进行初始化, 然后在整个测试序列中运行跟踪算法,得到精度曲 线和成功率曲线<sup>[7]</sup>,根据曲线下的 AUC(Area Under Curve)值对算法性能进行排序,如图 9。

可以看出,在 51 组视频测试结果下,本文算法 的精度曲线和成功率曲线的 AUC 值分别为 0.691 和 0.526,均优于其它 7 种对比算法,且比同样基于深 度学习的 DLT 算法分别提高了 17.72%和 20.64%。 同时,在跟踪速度上,本文算法的平均速度约为 13.6 帧/s,可以取得实时的跟踪效果。

## 5 结束语

本文提出了一种基于自适应深度稀疏网络的在 线跟踪算法。该算法针对传统深层网络预训练过程 复杂耗时的问题,构建了一种自然稀疏的深度神经 网络结构,其仅通过有限标签样本的在线训练,就 可以得到针对特定跟踪任务的鲁棒的网络模型。同 时,将这种深度稀疏的网络结构与粒子滤波框架相 结合,针对单一网络不能充分利用目标原始结构信 息的问题,提出了一种自适应多网络选择的在线跟 踪算法。实验结果表明,本文算法在复杂环境下具

表 3 部分视频跟踪结果对比

	算法	DLT	MIL	CT	OAB	Frag	TLD	IVT
CarDark	100(1.6)	67.9(18.7)	8.9(45.7)	0.3(121.1)	90.1(4.5)	7.6(79.6)	17.8(34.1)	54.7(21.5)
Deer	100(5.2)	38.0(49.2)	45.1(57.6)	4.2(243.6)	94.4(15.1)	5.6(98.9)	78.9(-)	4.2(188.2)
Basketball	79.6(11.6)	49.7(12.1)	28.7(106.3)	25.9(89.1)	1.7(205.5)	69.5(13.0)	2.5(-)	9.2(114.9)
FaceOcc-2	69.1(10.8)	62.8(11.3)	72.7(16.9)	62.4(22.8)	67.2(27.4)	56.0(39.8)	65.0(18.0)	61.2(15.7)
Jogging2	96.4(3.4)	16.3(161.2)	12.1(135.1)	1.95(138.3)	39.4(32.9)	11.7(76.7)	94.1(13.4)	14.0(128.0)
MountainBike	100(12.0)	84.2(14.1)	100(7.8)	25.9(189.8)	91.2(13.8)	14.0(206.0)	31.6(-)	89.9(12.7)
Singer2	95.4(10.0)	3.5(172.5)	3.3(169.1)	2.7(182.3)	3.0(188.4)	18.0(97.3)	3.0(-)	3.6(181.7)
Suv	98.4(4.2)	82.5(22.4)	14.2(73.8)	15.1(85.0)	78.8(35.4)	73.1(41.0)	89.7(-)	45.4(59.9)
Walking2	98.2(2.7)	98.6(2.7)	37.0(43.5)	7.8(64.7)	84.4(12.4)	38.4(64.0)	5.3(41.4)	46.6(6.0)
Woman	68.8(9.0)	88.8(5.3)	18.8(124.2)	12.1(115.9)	60.5(33.8)	15.9(103.8)	62.0(-)	18.9(187.9)
平均	90.6(7.1)	59.2(46.9)	34.1(78.0)	15.8(125.3)	61.1(56.9)	31.0(82.0)	45.0(-)	34.8(91.7)



图 9 51 组测试结果的精度曲线和成功率曲线

有良好的鲁棒性,有效地解决了目标丢失和跟踪漂 移问题。

在实验中发现,当目标出现较大角度的旋转和 严重的运动模糊现象时,如在"MotorRolling"视 频中,本文算法的跟踪效果不是很好。如何提高算 法在目标旋转和运动模糊下的鲁棒性将是下一步研 究的重点方向。

## 参考文献

- 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, 2014, 36(7): 1442-1468. doi: 10.1109/TPAMI.2013.230.
- WANG Naiyan, SHI Jianping, YEUNG Dityan, et al. Understanding and diagnosing visual tracking systems[C].
   IEEE International Conference on Computer Vision, Santiago, Chile, 2015: 3101-3109. doi: 1109/ICCV.2015.355.
- [3] ROSS D A, LIM J, LIN R S, et al. Incremental learning for robust visual tracking[J]. International Journal of Computer Vision, 2008, 77(1-3): 125-141. doi: 10.1007/s11263-007-0075-7.
- [4] BABENKO B, YANG M, and BELONGIE S. Robust object tracking with online multiple instance learning [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(8): 1619-1632. doi: 10.1109/TPAMI.2010.226.
- [5] KALAL Z, MIKOLAJCZYK K, and MATAS J. Trackinglearning-detection[J]. *IEEE Transactions on Pattern Analysis* and Machine Intelligence, 2012, 34(7): 1409-1422. doi: 10.1109/TPAMI.2011.239.
- [6] ZHANG Kaihua, ZHANG Lei, and YANG Minghsuan. Real-time compressive tracking[C]. European Conference on Computer Vision, Florence, Italy, 2012: 864-877.
- [7] WU Yi, LIM Jongwoo, and YANG Minghsuan. Online object tracking: A benchmark[C]. IEEE Conference on Computer Vision and Pattern Recognition, Portland, USA, 2013, 9(4): 2411-2418. doi: 10.1109/CVPR.2013.312.

- [8] MA Chao, HUANG Jiabin, YANG Xiaokang, et al. Hierarchical convolutional features for visual tracking[C].
   IEEE International Conference on Computer Vision, Santiago, Chile, 2015: 3074-3082.
- [9] NAJAFABADI M M, VILLANUSTRE F, KHOSHGOFTAAR T M, et al. Deep learning applications and challenges in big data analytics[J]. Journal of Big Data, 2015, 2(1): 1-21. doi: 10.1186/s40537-014-0007-7.
- [10] 李寰宇,毕笃彦,杨源,等.基于深度特征表达与学习的视觉
  跟踪算法研究[J].电子与信息学报,2015,37(9):2033-2039.
  doi: 10.11999/JEIT150031.

LI Huanyu, BI Duyan, YANG Yuan, et al. Research on visual tracking algorithm based on deep feature expression and learning[J]. Journal of Electronics & Information Technology, 2015, 37(9): 2033-2039. doi: 10.11999/JEIT150031.

- [11] JIN J, DUNDAR A, BATES J, et al. Tracking with deep neural networks[C]. Annual Conference on Information Sciences and Systems, Baltimore, MD, USA, 2013: 213-217.
- [12] WANG Naiyan and YEUNG Dityan. Learning a deep compact image representation for visual tracking[C]. Advances in Neural Information Processing Systems, South Lake Tahoe, Nevada, USA, 2013: 809-817.
- [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. doi: 10.1007/ s11263-015-0816-y.
- [14] 侯志强,戴铂,胡丹,等. 基于感知深度神经网络的视觉跟踪
  [J]. 电子与信息学报, 2016, 38(7): 1616-1623. doi: 10.11999/ JEIT151449.
   HOU Zhiqiang, DAI Bo, HU Dan, et al. Robust visual tracking via perceptive deep neural network[J]. Journal of Electronics & Information Technology, 2016, 38(7): 1616-1623. doi: 10.11999/JEIT151449.
- [15] GLOROT X and BENGIO Y. Understanding the difficulty of training deep feedforward neural networks[C]. International Conference on Artificial Intelligence and Statistics, Sardinia,

Italy, 2010: 249-256.

- [16] GLOROT X, BORDES A, and BENGIO Y. Deep sparse rectifier neural networks[C]. International Conference on Artificial Intelligence and Statistics, Fort Lauderdale, FL, USA, 2011: 315-323.
- [17] TOTH L. Phone recognition with deep sparse rectifier neural networks[C]. IEEE International Conference on Acoustics, Speech and Signal Processing, Vancounver, BC, Canada, 2013: 6985-6989. doi: 10.1109/ICASSP.2013.6639016.
- [18] VINCENT P, LAROCHELLE H, LAJOIE I, et al. Stacked denoising autoencoders: Iearning useful representations in a deep network with a local denoising criterion[J]. Journal of Machine Learning Research, 2010, 11(6): 3371-3408.
- [19] KALMANOVICH A and CHECHIK G. Gradual training method for denoising auto encoders[OL]. http://arxiv.org/ abs/1504.02902v1, 2015.
- [20] HE K, ZHANG X, REN S, et al. Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification[C]. IEEE International Conference on Computer Vision, Santiago, Chile, 2015: 1026-1034.
- [21] WU Y W, ZHAO H H, and ZAHNG L Q. Image denoising with retified linear units[J]. Lecture Notes in Computer Science, 2014, 8836: 142-149. doi: 10.1007/978-3-319-12643-2\_18.
- [22] LI X, HU W, SHEN C, et al. A survey of appearance models

in visual object tracking[J]. ACM Transactions on Intelligent Systems and Technology, 2013, 4(4): 1-48. doi: 10.1145/2508037.2508039.

- [23] WANG F S. Particle filters for visual tracking[C]. International Conference on Advanced Research on Computer Science and Information Engineering, Zhengzhou, China, 2011: 107-112.
- [24] GRABNER H, GRABNER M, and BISCHOF H. Real-time tracking via on-line boosting[C]. British Machine Vision Conference, Edinburgh, Scotland, 2006: 47-56.
- [25] ADAM A, RIVLIN E, and SHIMSHONI I. Robust fragments-based tracking using the integral histogram[C]. IEEE Conference on Computer Vision and Pattern Recognition, New York, USA, 2006: 798-805.
- 侯志强: 男,1973年生,教授,主要研究方向为图像处理、计算 机视觉和信息融合.
- 王 鑫: 男,1992年生,硕士生,研究方向为计算机视觉、机器
  学习.
- 余旺盛: 男,1985年生,讲师,研究方向为图像处理、模式识别.
- 戴 铂: 男,1992 年生,硕士生,研究方向为计算机视觉、机器 学习.
- 金泽芬芬:女,1993年生,硕士生,研究方向为计算机视觉、模式 识别.