

# 基于预判式学习更新策略孪生全卷积网络的目标跟踪

卢盼成 丁 勇 黄鑫城

(南京航空航天大学自动化学院 江苏 南京 211106)

**摘 要** 针对孪生全卷积网络缺乏有效的模型更新策略,跟踪目标周围存在相似目标干扰容易出现跟踪丢失的问题,提出一种基于预判式学习更新策略孪生全卷积网络的目标跟踪算法。确定目标模板和搜索区域的直方图置信度估计;模拟学习率的自重启机制,由给定正确标注初始化学习模块;根据置信度估计决定预判式学习模块更新,实现跟踪目标和相似目标的有效区分。实验结果表明,该算法具有良好的跟踪效果,在满足实时性跟踪的基础上,具有很好的跟踪精度和成功率。

**关键词** 目标跟踪 深度学习 孪生全卷积网络 置信度估计 预判式学习

中图分类号 TP394.41

文献标志码 A

DOI:10.3969/j.issn.1000-386x.2020.12.027

## TARGET TRACKING BASED ON SIAMESE FULL-CONVOLUTION NETWORK WITH PREJUDGMENT LEARNING UPDATE STRATEGY

Lu Pancheng Ding Yong Huang Xincheng

(College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, Jiangsu, China)

**Abstract** Aiming at the lack of effective model updating strategy for the Siamese full-convolutional network and the problem of tracking loss due to the similar target interference around the target, a target tracking algorithm based on the prejudgment learning update strategy for the Siamese full-convolutional network is proposed. The histogram confidence estimate for the target template and the search region was determined; the self-restart mechanism of the simulation learning rate was initialized by the given correct annotation; the prejudgment learning module was updated according to the confidence estimation to achieve effective differentiation between the tracking target and the similar target. The experimental results show that the algorithm has good tracking effect and good tracking accuracy and success rate on the basis of real-time tracking.

**Keywords** Target tracking Deep learning Siamese full-convolution network Confidence estimation Prejudgment learning

## 0 引 言

目标跟踪是计算机视觉领域的核心问题,在智能交通<sup>[1]</sup>、医学辅助诊断<sup>[2]</sup>、军事制导、航空视觉导航等各方面都发挥着重要的作用。近年来随着人工智能技术的飞速发展,基于卷积神经网络的深度学习方法在目标跟踪领域展现出巨大潜力。

近 20 年来,涌现出了大量的目标跟踪算法,根据其工作原理主要分为生成式模型和判别式模型两类。早期的目标跟踪算法研究主要集中在生成式模型上,

如光流法<sup>[3]</sup>、粒子滤波<sup>[4]</sup>、Mean-shift 算法、Cam-shift 算法等。此类方法首先建立目标模型或者提取目标特征,然后对后续帧进行相似特征搜索实现对目标的定位。但此类方法也存在明显的缺点,如图像的背景信息没有得到全面的利用且目标本身的外观变化有随机性和多样性的特点。判别式模型则是同时考虑目标模型和背景信息,通过比较两者的差异,将目标模型提取出来,从而得到当前帧的目标位置。由于判别式模型将背景信息引入跟踪模型,可以很好地实现背景区分,因此具有很大的优势。2000 年以来,人们逐渐尝试采用经典的机器学习方法训练分类器,例如 MIL (Matrox

Imaging Library)、TLD(Tracking Learning Detection)、支持向量机(Support Vector Machine, SVM)<sup>[5]</sup>等。尽管这些分类器种类繁多并具有一定的学习能力,但分类效果存在不确定性。2010年,文献[6]首次将通信领域的相关滤波(Correlation Filtering)方法引入到目标跟踪中,提高了目标跟踪的速度和精度,但对于多尺度的目标跟踪效果并不理想。2015年以后,随着卷积神经网络在图像领域的广泛应用,人们开始将深度学习的方法应用于目标跟踪<sup>[7]</sup>。由于该方法缺乏正样本训练数据,同时,随着神经网络规模的增加跟踪实时性很低。因此深度学习的方法难以训练出性能优良的分类器,不能满足快速运动目标的实时跟踪。

针对深度学习方法在目标跟踪中的问题,孪生全卷积网络的目标跟踪方法由于其端到端的训练能力和跟踪的实时性备受关注<sup>[8-11]</sup>。虽然孪生全卷积网络通过相似性学习的方法解决了深度学习在目标跟踪领域的实时性问题,但它的一个关键限制是缺少有效的模型更新策略,无法将来自背景区域或先前跟踪帧的信息融合到模型预测中。近年来相关文献的研究旨在解决这个问题。Valmadre等<sup>[9]</sup>将相关滤波器整合到深层网络中,但由于滤波器判别能力有限,难以有效区分跟踪目标与相似目标。Yao等<sup>[12]</sup>在训练阶段,通过特征提取器去学习跟踪参数,但参数本身不能应用于多个样本,需要特殊的线性滤波器组合来适应模型。Park等<sup>[13]</sup>从初始帧出发,通过独立的学习框架优化梯度下降过程和学习步长,然而这种策略只适用于模型的初始自适应,并不能改善迭代时的模型本身。此外,Guo等<sup>[14]</sup>通过学习变化特征来处理目标外观变化带来的影响,从而抑制周围相似目标的干扰;Zhu等<sup>[15]</sup>在目标跟踪期间,从目标模板中减去相似目标图像特征达到排除相似目标干扰的目的。以上两种方法为在线更新,定位精度高但模型学习效果较差。综上所述,孪生全卷积网络在进行目标跟踪时,仍然无法有效解决跟踪目标易受相似目标干扰的问题。

针对上述问题,本文提出一种基于预判式学习更新策略孪生全卷积网络(Pre-judgment Learning Update Strategy Full-Convolutional Siamese Networks, Prl-Siam-FC)的目标跟踪算法。由目标模板和搜索区域模型建立置信度估计,将模型更新策略加入到端到端的跟踪架构中,解决了孪生全卷积网络在目标跟踪时相似目标干扰的问题。

## 1 总体框架

传统孪生全卷积网络(Full-Convolutional Siamese

Networks, SiamFC)的结构如图1所示。在进行目标跟踪时,首先构建目标模板和搜索区域,然后采用两个相同的卷积神经网络,分别获取目标模板和搜索区域的卷积特征。最后对特征再进行卷积操作得到预测分数矩阵,从而定位到原始图片的目标跟踪区域。

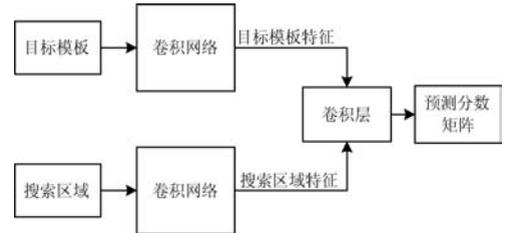


图1 传统孪生全卷积网络结构

本文所提出的预判式学习更新策略的孪生全卷积网络在孪生全卷积网络基础上加入了置信度估计模块和预判式学习模块,如图2所示。从输入端到输出端,总体框架包括卷积网络模块、置信度估计模块和预判式学习模块三个部分。卷积网络模块的网络构架采用孪生全卷积网络的 AlexNet;置信度估计模块对目标模板和搜索区域建立颜色直方图密度估计;预判式学习模块包括初始化模块和优化器模块两个部分。

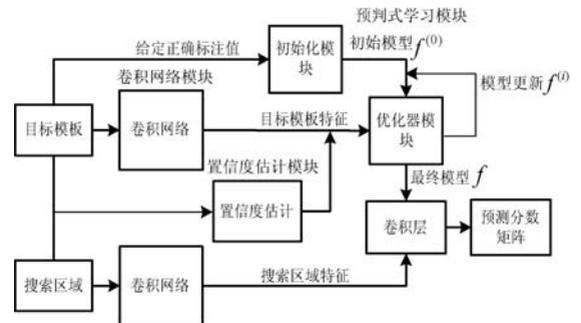


图2 预判式学习更新策略孪生全卷积网络总体框架

图2中,视频序列输入后,根据给定序列坐标中心建立目标模板及搜索区域;卷积网络模块用来提取目标模板和搜索区域的卷积特征;置信度估计模块用来对目标模板与搜索区域进行置信度估计,作为卷积权重滤波器 $f$ 更新的参考标准;在预判式学习模块中,将视频序列的初始正确标注数据(Ground Truth, GT)平均分布在数据集 $s$ 上,获得初始模型 $f^{(0)}$ ,同时根据置信度估计判别跟踪序列中是否存在相似目标干扰,从而决定是否更新 $f^{(i)}$ 。

## 2 目标模板与搜索区域置信度估计

预判式学习更新策略孪生全卷积网络的置信度估计模块,对跟踪算法的精度、成功率和速度具有至关重要的影响。目标模板与搜索区域的卷积特征具有良好的鲁棒性,但不能从中有效地得到相似目标干扰的相关特征。若要实现对跟踪目标和相似目标的有效区

分,则需要原卷积网络模块中附加用于相似目标提取的卷积模块,这种方法在提高跟踪成功率的同时会降低跟踪速度。

本文选取具有较高判别精度且计算量小的颜色直方图密度估计<sup>[16]</sup>的方法,在保证目标跟踪速度的基础上,建立目标模板与搜索区域的置信度估计,作为模型更新的参考标准,可以最大限度在不降低目标跟踪速度的基础上提高模型跟踪的成功率。

## 2.1 目标模板与搜索区域颜色直方图描述

假设目标模板的中心坐标为原点,其彩色空间被划分为  $m$  个区间,  $u$  表示每个区间的颜色簇索引,取值范围为  $\{1, 2, \dots, m\}$ , 目标模板的各像素点位置由集合  $\{x_i^*\} (i = 1, 2, \dots, n)$  表示,则构成目标的每簇颜色的概率密度函数  $\hat{q}_u$  为:

$$\hat{q}_u = C \sum_{i=1}^n k(\|x_i^*\|^2) \delta[b(x_i^*) - u] \quad (1)$$

式中:  $C$  是归一化系数函数;  $k(\cdot)$  是核密度估计函数  $K(\cdot)$  的轮廓函数,即  $K(x_i^*) = k(\|x_i^*\|^2)$ ;  $\delta(\cdot)$  是 Kronecker Delta 函数;  $b(\cdot): R^2 \rightarrow \{1, 2, \dots, m\}$  是位于  $x_i^*$  的像素向颜色索引的映射。

搜索区域描述与目标模板描述方法类似,设  $y$  为搜索区域中心坐标,  $\{x_i\} (i = 1, 2, \dots, n_h)$  表示搜索区域像素位置的集合,搜索区域的概率密度函数  $\hat{p}_u(y)$  可描述为:

$$\hat{p}_u(y) = C_h \sum_{i=1}^{n_h} k\left(\left\|\frac{y - x_i}{h}\right\|^2\right) \delta[b(x_i) - u] \quad (2)$$

式中:  $h$  是核函数的带宽;  $C_h$  是基于  $h$  的归一化系数函数。

## 2.2 置信度估计

目标模板与搜索区域的置信度可以简化为  $\hat{q}_u$  与  $\hat{p}_u(y)$  的相似度,由式(1)和式(2)可得,目标模板和搜索区域在整个颜色空间上的概率分布可以分别表示为  $\hat{q} = (\hat{q}_1, \hat{q}_2, \dots, \hat{q}_m)$ 、 $\hat{p}(y) = (\hat{p}_1(y), \hat{p}_2(y), \dots, \hat{p}_m(y))$ ,采用 Bhattacharyya 系数  $\rho(y)$  度量,定义该相似度值  $C_{\text{siam}}$  即为目标模板与搜索区域置信度值:

$$C_{\text{siam}} = \hat{\rho}[\hat{p}(y), \hat{q}] \approx \frac{1}{2} \sum_{u=1}^m \sqrt{\hat{p}_u(y_0) \hat{q}_u} + \frac{C_h}{2} \sum_{i=1}^{n_h} w_i k\left(\left\|\frac{y - x_i}{h}\right\|^2\right) \quad (3)$$

式中:  $y_0$  为搜索区域初始中心坐标。 $w_i$  计算如下:

$$w_i = \sum_{u=1}^m \delta[b(x_i) - u] \sqrt{\frac{\hat{q}_u}{\hat{p}_u(y_0)}} \quad (4)$$

由式(3)可知,置信度值  $C_{\text{siam}} \in [0, 1]$ 。 $C_{\text{siam}}$  越接近 1,表示目标模板与搜索区域相似度越高;  $C_{\text{siam}}$  越接

近 0,表示搜索区域中越可能存在相似目标,则需要进行预判式学习模块更新。为保证跟踪实时性,间隔 20 帧进行置信度估计及预判式学习模块更新。

## 3 预判式学习模块

在目标跟踪过程中,影响跟踪的三个重要因素是特征、分类器和模型更新策略<sup>[17]</sup>。孪生全卷积网络中特征采用的是卷积神经网络 CNN 提取到的图像特征,分类器根据目标模板与搜索区域的相似性进行学习,而整个网络缺乏有效的模型更新策略。本文给出预判式学习模块  $D$ ,在整个模型更新过程中,  $S$  为数据集,  $f$  是具有卷积层权重的滤波器,预判式学习模块的作用是给出预测函数  $f = D(S)$ ,用于在特征空间中区分跟踪目标和背景外观。

### 3.1 损失函数和自重启学习率的设计

为保证预判式学习更新策略孪生全卷积网络的稳健性,定义预判式学习的损失函数  $L(f)$  为:

$$L(f) = \frac{1}{|S|} \sum_{(x,c) \in S} \|r(s,c)\|^2 + \|\lambda f\|^2 \quad (5)$$

式中:  $S = \{(x_j, c_j)\}_{j=1}^n, x_j \in \mathcal{X}, \mathcal{X}$  为卷积网络模块生成的深度特征图,  $c_j$  为每个样本目标中心坐标  $(x_0, y_0)$ ; 函数  $r(s, c)$  根据目标置信度得分  $s = x \times f$  和给定的真值目标中心坐标  $c$  计算每个空间位置的残差;  $\lambda$  是正则化因子。

在实际的样本训练中,如果单纯以图片的方式进行训练,传统的岭回归问题只会让网络过分注意于优化负样本的影响,而忽略正样本特征本身的判别能力。为此引入空间权重系数  $v_c$ ,通过权重系数使得前景和背景之间具有一定的距离,改善分类结果。定义残差函数  $r(s, c)$  为:

$$r(s, c) = v_c(m_c s + (1 - m_c) \max(0, s) - y_c) \quad (6)$$

式中:  $m_c$  为目标分割系数,具体参数选择通过网络学习得到<sup>[18]</sup>,若目标中心在跟踪目标区域,则  $m_c = 1$ ;若目标中心在跟踪背景区域,则  $m_c = 0$ ;  $s$  为目标置信度得分;  $y_c$  是以  $c = (x_0, y_0)$  为中心的二次高斯函数,表示每个位置的目标期望得分。 $y_c$  具体计算如下:

$$y_c = y(x_0, y_0) = \frac{1}{2\pi\sigma^2} e^{-\frac{x_0^2 + y_0^2}{2\sigma^2}} \quad (7)$$

式中:  $\sigma$  为高斯函数分布方差。

在用预判式学习更新策略孪生全卷积网络进行目标跟踪的过程中,将固定学习率  $\eta$  设为自重启式学习率  $\eta_i$ ,对  $T_i$  次已经完成数据集训练的学习率采用余弦退火,即:

$$\eta_t^i = \eta_{\min}^i + \frac{1}{2}(\eta_{\max}^i - \eta_{\min}^i) \left(1 + \cos\left(\frac{T_{\text{cur}}}{T_i} \pi\right)\right) \quad (8)$$

式中:  $i$  是数据集训练的运行索引;  $\eta_{\min}^i$  和  $\eta_{\max}^i$  表示学习率  $\eta$  变化的最小值和最大值;  $T_{\text{cur}}$  用来记录自上次重启以来训练的次数。

上述的自重启随机梯度下降算法通过给定  $T_i$  和  $T_{\text{cur}}$ , 使得学习率  $\eta_t$  出现衰减变化, 当学习率  $\eta_t$  下降到最小值后, 再次变为初始值, 模拟出现自重启现象。这种优化策略避免了神经网络使用固定学习率易于陷入到鞍点中, 收敛于局部最优, 有效解决了孪生全卷积网络训练中的损失误差和病态函数问题。

### 3.2 卷积层权值滤波器的预判式更新

传统的目标跟踪算法通常在确定跟踪目标后, 再来衡量模型的可靠性, 进而采取相应的更新策略。然而这种更新策略存在滞后性, 并不能进行动态调整实现提前更新, 以应对复杂的目标跟踪。针对此问题, 本文对卷积层权值滤波器  $f$  进行了预判式更新。

在预判式学习模块中, 我们引入初始化模块, 该模块由一个卷积层和一个精确的池化层组成, 用来生成初始模型  $f^{(0)}$ 。初始模型  $f^{(0)}$  只提供一个合理的初始估计, 用来进一步减少预判式学习模块的优化递归次数, 然后由优化器模块进行处理, 以提供最终的识别模型。初始化模块完成初始化后, 对式(5) 预判式学习损失函数  $L(f)$  求偏导得:

$$\nabla L(f) = \frac{2}{|S|} \sum_{(s,c) \in S} \left( \frac{\partial r(s,c)}{\partial f} \right)^T r(s,c) + 2\lambda^2 f \quad (9)$$

式中:  $\frac{\partial r(s,c)}{\partial f}$  为式(5) 的雅可比矩阵, 根据式(6) 残差函数  $r(s,c)$  的定义对其求偏导得:

$$\frac{\partial r(s,c)}{\partial f} = \text{diag}(v_c m_c) \frac{\partial s}{\partial f} + \text{diag}((1 - m_c) \cdot l_s) \frac{\partial s}{\partial f} \quad (10)$$

令  $q_c = v_c m_c + (1 - m_c) \cdot l_s$ , 对于  $q_c$  仅使用点操作, 对于正样本,  $l_s > 0$ , 对于负样本,  $l_s = 0$ 。由此得到:

$$\frac{\partial r(s,c)}{\partial f} = \text{diag}(q_c) \frac{\partial s}{\partial f} \quad (11)$$

最终, 由式(9) 和式(11) 推导出卷积权重滤波器  $f$  的梯度下降表达式为:

$$\nabla L(f) = \frac{2}{|S|} \sum_{(s,c) \in S} \left( \frac{\partial s}{\partial f} \right)^T (q_c \cdot r(s,c)) + 2\lambda^2 f \quad (12)$$

在得到上述梯度下降值的基础上, 采用 3.1 节中的自重启学习率  $\eta_t$ , 则滤波器  $f$  的更新表达式为:

$$f^{(i+1)} = f^{(i)} - \eta_t^i \nabla L(f^{(i)}) \quad (13)$$

综上, 本文算法在卷积网络获取目标模板和搜索区域的卷积特征时, 每隔 20 帧同时对目标模板和搜索区域进行了置信度估计, 由置信度值  $C_{\text{siam}}$  提前预判序列帧中是否存在相似目标。当置信度值  $C_{\text{siam}}$  小于给定阈值时, 则跟踪目标周围可能存在相似目标, 进而考虑卷积层权值滤波器  $f$  的更新。这种更新策略在确定跟踪目标区域前进行预判确定是否更新, 实现了卷积层权值滤波器  $f$  的实时更新, 在保证跟踪的准确性和稳定性的同时, 有效解决了相似目标干扰的问题。

## 4 算法流程

本文算法流程描述如下:

**步骤 1** 输入视频数据集  $S = \{(x_j, c_j)\}_{j=1}^n$ , 构建目标模板和搜索区域。

**步骤 2** 由数据集的给定 GT 进行模块初始化, 即  $f^{(0)} \leftarrow S$ , 得到初始模型  $f^{(0)}$ 。

**步骤 3** 卷积网络模块提取目标模板和搜索区域卷积特征。

**步骤 4** 间隔 20 帧计算目标模板及搜索区域置信度值  $C_{\text{siam}}$  为:

$$C_{\text{siam}} \approx \frac{1}{2} \sum_{u=1}^m \sqrt{\hat{p}_u(y_0) \hat{q}_u} + \frac{C_h}{2} \sum_{i=1}^{n_h} w_i k \left( \left\| \frac{y - x_i}{h} \right\|^2 \right)$$

**步骤 5** 判断置信度值  $C_{\text{siam}}$  是否大于阈值, 若大于阈值则执行步骤 9, 否则执行步骤 6 - 步骤 8, 对滤波器  $f$  进行更新。

**步骤 6** 由给定学习率极值  $\eta_{\min}^i$  和  $\eta_{\max}^i$ , 对学习率  $\eta_t$  采用自重启机制:

$$\eta_t^i = \eta_{\min}^i + \frac{1}{2}(\eta_{\max}^i - \eta_{\min}^i) \left(1 + \cos\left(\frac{T_{\text{cur}}}{T_i} \pi\right)\right)$$

**步骤 7** 采用步骤 6 中的学习率, 计算卷积权重滤波器  $f$  的梯度下降值:

$$\nabla L(f^{(i)}) \leftarrow (f^{(i)}, S)$$

**步骤 8** 卷积层权值滤波器  $f$  进行预判式更新:

$$f^{(i+1)} \leftarrow f^{(i)} - \eta_t^i \nabla L(f^{(i)})$$

**步骤 9** 对搜索区域和经过滤波器  $f$  过滤掉相似目标的目标模板进行卷积操作。

**步骤 10** 卷积操作输出预测分数矩阵, 矩阵经线性插值定位到原始图片的目标位置, 实现目标跟踪。

## 5 实验

### 5.1 实验设计

本文实验所使用的测试基准是 OTB 2013, 包含 50

个长短时序列,视频序列影响因素包括相似目标干扰、平面内旋转、光照变化、快速运动等 11 种类型。预判式学习更新策略孪生全卷积网络构架采用 AlexNet,如表 1 所示,其中前两层卷积层后均接有池化层,除了最后一层,每一层均使用 ReLU 激活函数。

表 1 预判式学习更新策略孪生全卷积网络构架

网络层	卷积核	通道参数	步长	输出
卷积层 1	11 × 11	96 × 3	2	× 96
池化层 1	3 × 3	—	2	× 96
卷积层 2	5 × 5	256 × 48	1	× 256
池化层 2	3 × 3	—	2	× 256
卷积层 3	3 × 3	384 × 256	1	× 192
卷积层 4	3 × 3	384 × 192	1	× 192
卷积层 5	3 × 3	256 × 192	1	× 128

在实验中,预判式学习更新策略孪生全卷积网络初始化参数设置如表 2 所示。

表 2 预判式学习更新策略孪生全卷积网络初始化参数

参数名称	参数大小
正则化因子 $\lambda$	0.01
空间权重系数 $v_c$	1.20
高斯函数分布方差 $\sigma$	2.00
学习率最小值 $\eta_{\min}^i$	$10^{-8}$
学习率最大值 $\eta_{\max}^i$	$10^{-2}$
目标模板与搜索区域置信度 $C_{\text{siam}}$	0.70

为了有效地评估本文所提算法的性能,实验中的对比算法均来自相应作者的公开代码。所有实验都在 Inter(R) Core(TM) i5-2450M CPU@ 2.50 GHz、4 GB 内存、NVIDIA GeForce GTX 750 Ti 台式机上进行,算法通过 MATLAB 2018a、Visual Studio 2015、CUDA 9.0、CUDNN 7.0 实现。

实验从预判式学习更新策略孪生全卷积网络不同模块消融对比、数据集训练识别精度、对比孪生全卷积网络和跟踪算法整体性能四个方面进行验证。

## 5.2 不同模块消融对比

本文采用一次通过测试 (One-pass Evaluation, OPE) 对预判式学习更新策略孪生全卷积网络 Prl-SiamFC 的置信度估计模块和预判式学习模块分别进行了测试实验,结果如表 3 所示。可以看出,与孪生全卷积网络 SiamFC 相对比,预判式学习模块提高了 4.05% 的跟踪成功率。与无置信度估计模块的 Prl-SiamFC 相比,置信度估计模块提高了 7.41% 的跟踪精度,置信度估计模块和预判式学习对跟踪精度和成功

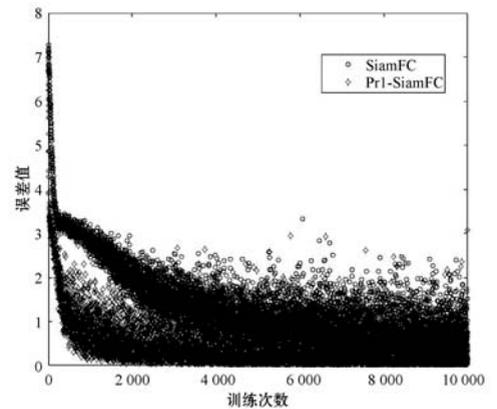
率均有较大提升效果;与孪生全卷积网络 SiamFC 的跟踪速度相比,采用置信度模块和预判式学习模块基本上对跟踪速度没有太多影响。

表 3 预判式学习更新策略孪生全卷积网络不同模块消融对比

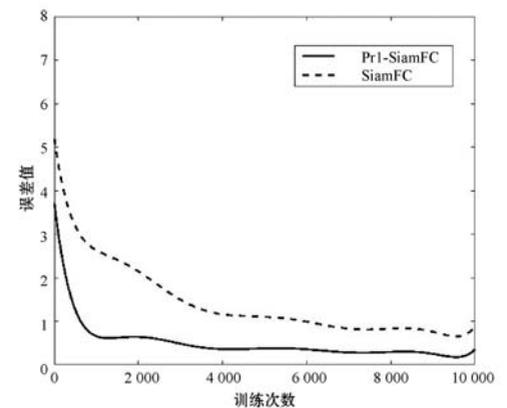
算法名称	置信度估计模块	预判式学习模块	精度	成功率	平均 FPS
SiamFC	无	无	0.815	0.791	25.87
Prl-SiamFC	无	有	0.837	0.823	24.64
	有	有	0.899	0.832	23.34

## 5.3 数据集训练识别精度

采用 Minist 手写数字数据集进行 10 000 次迭代, Prl-SiamFC 和 SiamFC 的数据集训练识别精度效果如图 3 所示,(a) 为原始识别误差散点分布情况,(b) 为识别误差拟合曲线。



(a) 识别误差散点分布



(b) 识别误差拟合曲线

图 3 Prl-SiamFC 与 SiamFC 数据集训练识别精度比较

可以看出,与 SiamFC 相比,采用自重启式学习率的 Prl-SiamFC 初始收敛速度更快,收敛过程更加平稳,数据集训练识别误差精度平均减少 93% 且不存在较大波动。因此, Prl-SiamFC 比 SiamFC 学习效果更好。

## 5.4 相似目标干扰下的不同网络跟踪性能比较

为了验证本文所提的 Prl-SiamFC 在存在相似目标

干扰情况下跟踪性能的优越性, 本文将其与 SiamFC 和三种尺度孪生全卷积网络(简称: SiamFC\_3S)对具体帧的跟踪情况进行比较。如图 4 所示, 选取 4 组视频序列, 视频图中左上角数字为当前图像帧数, 黑框为 Prl-SiamFC, 白框为 SiamFC 和 SiamFC\_3S。

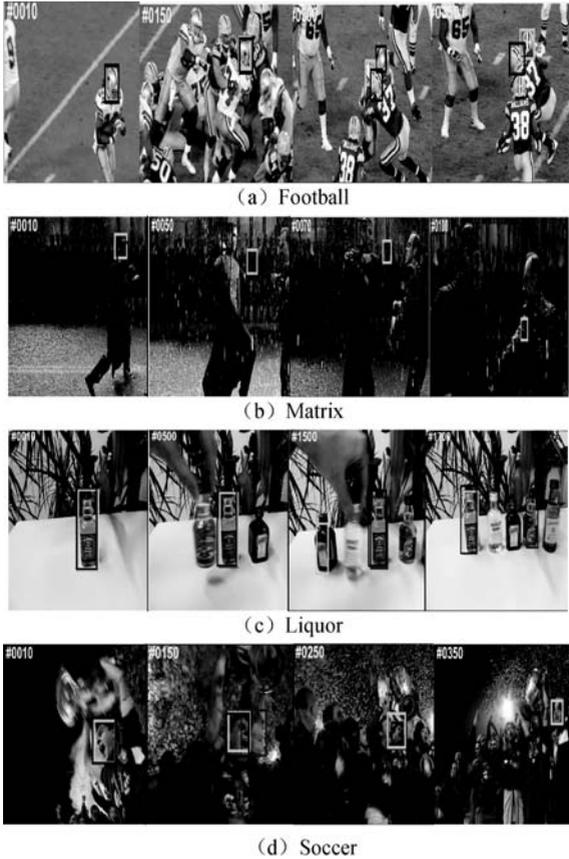
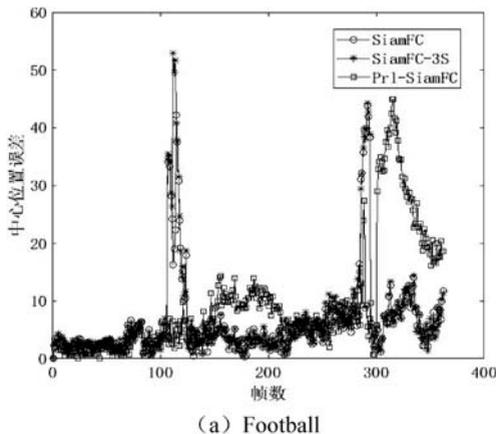


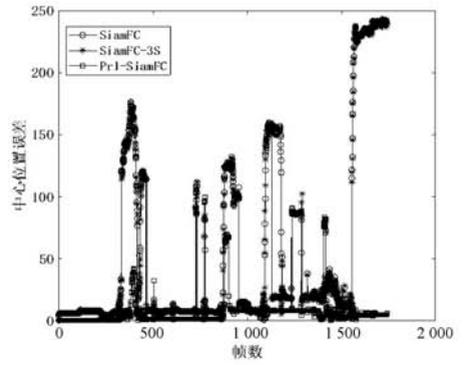
图 4 相似目标干扰下跟踪性能比较

可以看出, 4 组视频的第 10 帧均能确定跟踪目标且跟踪精度较高, 但序列 Matrix 第 50 帧、序列 Liquor 第 1 500 帧和序列 Soccer 第 350 帧以后, SiamFC 和 SiamFC\_3S 均出现跟踪丢失的情况, 而 Prl-SiamFC 始终跟踪目标, 具有较好的跟踪效果。

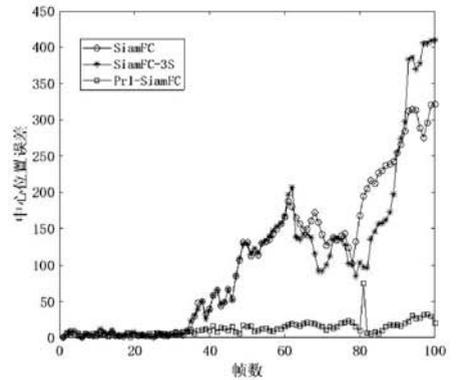
对上述 4 组视频序列, 如图 5 所示, 采用中心位置误差进行对比分析, 这里的中心位置误差是指视频跟踪结果中心点与真值中心点之间的欧氏距离。



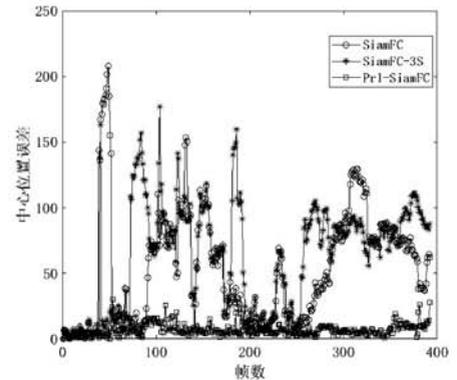
(a) Football



(b) Matrix



(c) Liquor



(d) Soccer

图 5 不同算法在相似目标干扰下中心位置误差比较

可以看到, Prl-SiamFC 的中心位置误差除了在视频序列 Football 中最后帧数出现波动外, 序列 Matrix、Liquor 和 Soccer 的中心误差始终保持在较低水平, 序列 Matrix 和 Soccer 的中心位置误差平均小于 20, 在跟踪过程中有效解决了跟踪目标周围存在相似目标干扰问题。

### 5.5 算法跟踪性能比较

在 OTB 2013 测试下, 采用 OPE、空间鲁棒性测试 (Spatial Robustness Evaluation, SRE)、时间鲁棒性测试 (Temporal Robustness Evaluation, TRE) 三种标准, 分别测试 Prl-SiamFC 与 SiamFC 的跟踪效果, 如图 6 所示。各标准下的整体跟踪效果提升等于跟踪精度和成功率提升的平均值。在 OPE、SRE、TRE 三种标准测试下, Prl-SiamFC 与 SiamFC 相比, 整体跟踪效果分别提升了

7.75%、10.14%、8.98%。

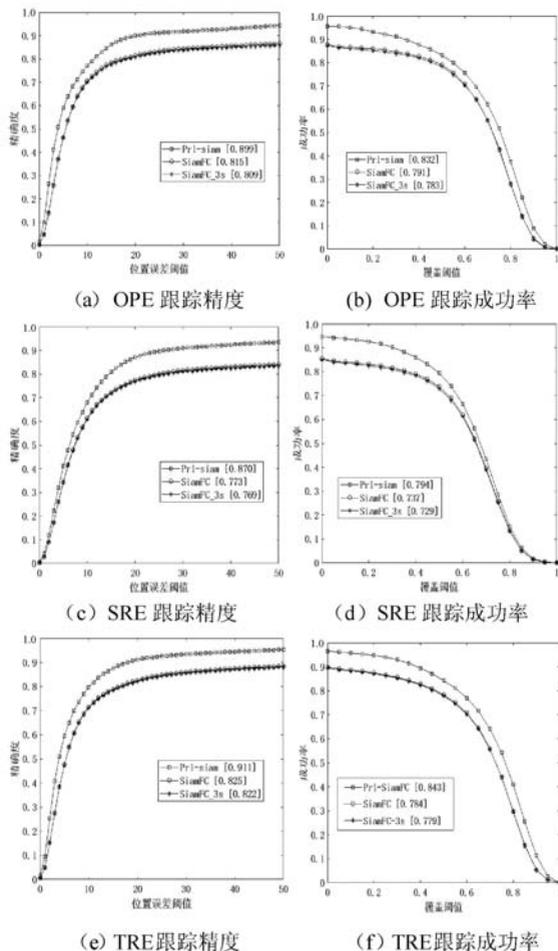


图 6 Prl-SiamFC 改善 SiamFC 跟踪效果曲线

除了整体跟踪精度和成功率的提升,与 SiamFC、SiamFC\_3S 相比,Prl-SiamFC 对于跟踪目标的变形、遮挡、尺度变化、平面内外旋转等问题,均有跟踪效果的明显提升,如表 4 所示。特别是对跟踪目标出现运动模糊和视野外的问题,Prl-SiamFC 相比较于 SiamFC、SiamFC\_3S,跟踪效果的平均提升效果在 20% 以上。

表 4 Prl-SiamFC 在其他方面跟踪效果的提升 %

跟踪分类	对比 SiamFC	对比 SiamFC_3S
变形	18.64	19.59
遮挡	16.63	17.39
视野外	21.33	22.79
相似目标	10.20	11.72
快速运动	14.12	15.80
照明变化	16.28	14.90
低分辨率	6.04	14.54
运动模糊	20.58	22.86
尺度变化	6.52	6.87
平面内旋转	6.55	7.45
平面外旋转	10.02	10.29

除了与 SiamFC 对比外,本文在 OTB 2013 上还测试了 18 种经典的跟踪算法,其中,前 10 名算法的跟踪精度和成功率曲线,如图 7 所示。图中括号里数字表示中心位置误差取值为 20 时对应的跟踪精度和覆盖阈值取值为 0.5 时对应的跟踪成功率。

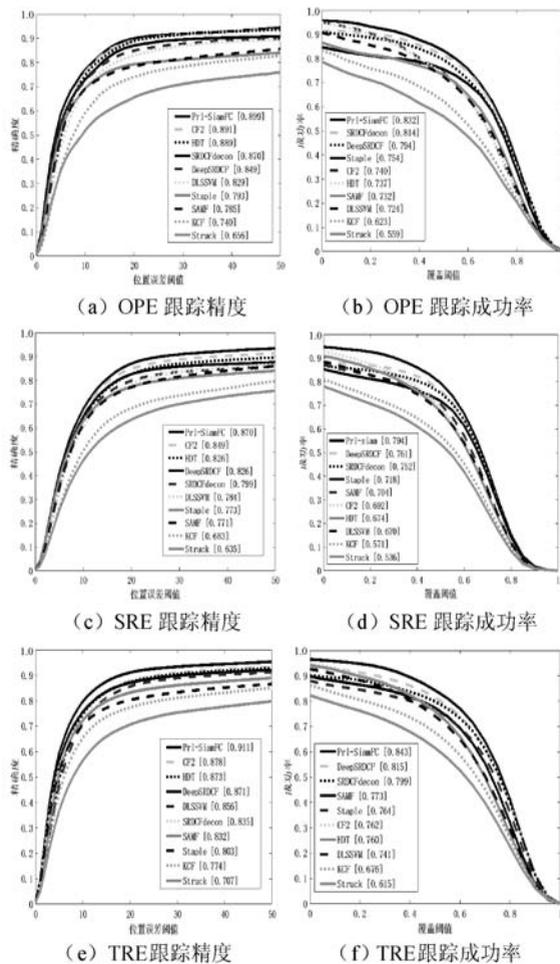


图 7 不同算法跟踪性能比较

可以看出,本文的跟踪算法精度和成功率明显高于其他算法。在图 7(a)、(c)和(e)中,跟踪精确度在中心位置误差阈值为 20 像素时,在 OPE、SRE、TRE 下,本文跟踪精确度高于 CF2 0.90%、2.47%、3.76%;当中心误差阈值小于 20 像素时,本文算法的精确度值更明显高于其他算法,这说明在高精度约束下,本文算法的跟踪效果更好。在图 7(b)、(d)和(f)中,本文算法的跟踪成功率比深度学习跟踪算法 SRDCFdecon 提高了 2.21%、4.34%、3.44%;当覆盖阈值处于 0.2 ~ 0.8 位置时,本文算法的成功率明显高于其他对比算法;当覆盖阈值大于 0.8 时,本文算法也能保证最优性能。

综上所述,本文所提的预判式学习更新策略孪生全卷积网络满足了目标跟踪的各项性能要求,具有训练识别误差小、跟踪精度高和成功率高的特点,算法整体具有良好的跟踪性能。

## 6 结 语

本文提出预判式学习更新策略孪生全卷积网络目标跟踪算法,将模型更新策略加入到孪生全卷积网络的端到端跟踪架构中,由目标模板和搜索区域模型确定置信度并判别是否进行预判式学习模块更新,解决了孪生全卷积网络在目标跟踪过程中相似目标干扰的问题。预判式学习更新策略孪生全卷积网络与孪生全卷积网络相比,数据集识别训练精度更高,采用 Minist 数书写数据集测试,数据集训练识别精度误差平均减少 93%。在 OTB 2013 的 OPE、SRE、TRE 标准测试下,目标跟踪精度提高了 10.31%、12.55%、10.42%,成功率提高了 5.18%、7.73%、7.53%,整体跟踪精度和成功率与其他经典算法相比更优,性能表现优异。

## 参 考 文 献

- [ 1 ] 黄铎,应娜,蔡哲栋. 基于强化学习的多人姿态检测算法优化[J]. 计算机应用与软件,2019,36(4):186-191.
  - [ 2 ] 罗建豪,吴建鑫. 基于深度卷积特征的细粒度图像分类研究综述[J]. 自动化学报,2017,43(8):1306-1318.
  - [ 3 ] Tao D P, Cheng J, Song M L, et al. Manifold ranking-based matrix factorization for saliency detection[J]. IEEE Transactions on Neural Networks and Learning Systems, 2016, 27(6):1122-1134.
  - [ 4 ] Cheng M M, Mitra N J, Huang X L, et al. Global contrast based salient region detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3):569-582.
  - [ 5 ] Li S Q, Zeng C, Fu Y, et al. Optimizing multi-graph learning based salient object detection[J]. Signal Processing: Image Communication, 2017, 55:93-105.
  - [ 6 ] Hou X D, Zhang L Q. Saliency detection: A spectral residual approach[C]//2007 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2007:1-8.
  - [ 7 ] Zhang Q R, Zhang M, Chen T H, et al. Recent advances in convolutional neural network acceleration[J]. Neurocomputing, 2019, 323:37-51.
  - [ 8 ] Bertinetto L, Valmadre J, Henriques J F, et al. Fully-convolutional siamese networks for object tracking[C]//2016 European Conference on Computer Vision. Springer, 2016: 850-865.
  - [ 9 ] Valmadre J, Bertinetto L, Henriques J F, et al. End-to-end representation learning for Correlation Filter based tracking [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017:5000-5008.
  - [ 10 ] Gundogdu E, Alatan A A. Good features to correlate for visual tracking[J]. IEEE Transactions on Image Processing, 2018, 27(5):2526-2540.
  - [ 11 ] Liu F, Yang A Z, Wu Z W. Adaptive Siamese network based UAV target tracking algorithm[J]. ACTA Aeronauticae Astronautica Sinica, 2020, 41(1):248-260.
  - [ 12 ] Yao Y J, Wu X H, Zhang L, et al. Joint representation and truncated inference learning for correlation filter based tracking[C]//2018 European Conference on Computer Vision. Springer, 2018:560-575.
  - [ 13 ] Park E, Berg A C. Meta-tracker: Fast and robust online adaptation for visual object trackers[EB]. arXiv: 1801. 03049, 2018.
  - [ 14 ] Guo Q, Feng W, Zhou C, et al. Learning dynamic siamese network for visual object tracking[C]//2017 IEEE International Conference on Computer Vision (ICCV). IEEE, 2017: 1781-1789.
  - [ 15 ] Zhu Z, Wang Q, Li B, et al. Distractor-aware siamese networks for visual object tracking[C]//2018 European Conference on Computer Vision. Springer, 2018:103-119.
  - [ 16 ] Ning J, Zhang L, Zhang D, et al. Scale and orientation adaptive mean shift tracking[J]. IET Computer Vision, 2012, 6(1):52-61.
  - [ 17 ] Wang N Y, Shi J P, Yeung D Y, et al. Understanding and diagnosing visual tracking systems [C]//IEEE International Conference on Computer Vision. IEEE, 2015:3101-3109.
  - [ 18 ] Bhat G, Danelljan M, Gool L V, et al. Learning discriminative model prediction for tracking [EB]. arXiv: 1904. 07220, 2019.
- 
- (上接第 112 页)
- [ 16 ] Fayek H M, Lech M, Cavedon L. Evaluating deep learning architectures for Speech Emotion Recognition [J]. Neural Networks, 2017, 92:60-68
  - [ 17 ] 田熙燕,徐君鹏,杜留锋. 基于语谱图和卷积神经网络的语音情感识别[J]. 河南科技学院学报(自然科学版), 2017, 45(2):62-68.
  - [ 18 ] 刘戈锋. 基于浅层学习引导深度学习的行人检测[D]. 武汉:武汉大学, 2016.
  - [ 19 ] Chan T H, Jia K, Gao S, et al. PCANet: A simple deep learning baseline for image classification[J]. IEEE Transactions on Image Processing, 2015, 24(12):5017-5032.
  - [ 20 ] 顾凌云,吕文志,杨勇,等. 基于 PCANet 和 SVM 的谎言测试研究[J]. 电子学报, 2016, 44(8):1969-1973.
  - [ 21 ] 赵小蕾,毛启容,詹永照. 融合功能性副语言的语音情感识别新方法[J]. 计算机科学与探索, 2014, 8(2):186-199.