

DOI:10.19650/j.cnki.cjsi.J2412935

改进转换模型预测网络的外观变化下目标跟踪方法*

陈仁祥¹,何家乐¹,杨黎霞²,余腾伟¹,梁 栋¹

(1. 重庆交通大学 交通工程应用机器人重庆市工程实验室 重庆 400074; 2. 重庆科技学院 经济与金融学院 重庆 401331)

摘要:针对跟踪过程中因目标发生外观变化导致跟踪效果不佳的问题,提出改进转换模型预测(ToMP)网络的外观变化下目标跟踪方法。首先,在ToMP网络基础上添加目标外观状态判别模块,利用级联长短时记忆网络(LSTM)输出目标外观状态判别信息;其次,优化网络在线样本储存判据,在置信度分值判别基础上添加正常外观判别信息,利用可靠样本优化模型权重,增强网络对目标的分类能力;然后,改进网络外观变化时在线样本利用机制,结合最新样本优化模型权重,提高网络对新外观目标的分类效果;最后,利用中心点轨迹预测位置对网络生成目标响应分数进行加权,增强目标特征映射效果的同时减少相似物体干扰以稳定跟踪目标。在公共数据集上准确率分别达到93.9%和68.9%,优于其他方法,并利用机器人进一步证实了可行性。

关键词:目标跟踪;改进转换模型;连续状态判别;预测距离加权

中图分类号:TH39 文献标识码:A 国家标准学科分类代码:510.40

Improved ToMP network appearance changes under target tracking method

Chen Renxiang¹, He Jiale¹, Yang Lixia², Yu Tengwei¹, Liang Dong¹

(1. Chongqing Engineering Laboratory for Transportation Engineering Application Robot, Chongqing Jiaotong University, Chongqing 400074, China; 2. Chongqing University of Science & Technology, Economics and Finance College, Chongqing 401331, China)

Abstract: To address the issue of poor tracking performance caused by target appearance changes during tracking, a target tracking method under appearance change based on improved ToMP network is proposed. Firstly, a target appearance state discrimination module was added to the ToMP network, using a cascaded Long Short-Term Memory (LSTM) network to output the target appearance state discrimination information. Secondly, the online sample storage criterion of the network was improved by adding the normal appearance discrimination information to the confidence score-based evaluation. This optimizes the model weights using reliable samples, enhancing the network's classification ability for the target. Then, the mechanism for utilizing online samples during appearance changes was improved, updating the model weights with the latest samples to enhance classification performance for newly appeared targets. Finally, a center point trajectory prediction was used to weight the target response score generated by the network, improving the target feature mapping while reducing interference from similar objects, thereby stabilizing target tracking. The accuracy on public datasets reached 93.9% and 68.9%, respectively, outperforming other methods and feasibility is further confirmed through robotics experiments.

Keywords: target tracking; transforming model; continuous state discrimination; predictive distance weighting

0 引言

目标跟踪是计算机视觉领域的研究热点之一,其流

程为:给定视频序列初始帧边界框,跟踪后续帧中该目标的位置和形状^[1]。目前主流的目标跟踪方法有基于相关滤波的跟踪算法和基于深度学习的跟踪算法。文献[2]对目标跟踪相关滤波器进行实验评价,文献[3]提出了

收稿日期:2024-05-14 Received Date: 2024-05-14

* 基金项目:国家自然科学基金项目(52475548)、重庆市教委科学技术研究项目(KJZD-M202200701)、重庆市自然科学基金创新发展联合基金项目(CSTB2023NSCQ-LZX0127)、重庆市研究生联合培养基地项目(JDLHPYJD2021007)、重庆市专业学位研究生教学案例库项目(JDALK2022007)、重庆市自然科学基金项目(CSTB2023NSCQ-MSX0177)资助

全天候的无人机目标跟踪方法。实际工程中,目标外观可能会发生不同程度的变化,导致跟踪模型无法做出及时调整,是亟需解决的问题之一^[4]。

近年来,国内外学者在目标跟踪领域展开了大量研究。文献[5]提出了基于加权在线样本更新的目标跟踪方法,解决了长时跟踪过程中因丢失视野导致目标跟踪失败的问题。文献[6]利用多层特征位置和语义信息,获得了更准确的目标位置估计。文献[7]引入目标分割方法,在跟踪过程中更准确获取目标前景区域,有效应对目标变化问题。上述方法为解决目标变化问题,将目标作为前景区域在图像中区分,以更好获取目标特征表示,但这些方法大多采用离线方式,无法在目标发生外观变化时稳定跟踪目标。

针对该问题,文献[8]为适应目标运动时产生的外观变化,利用外观动作自适应跟踪方法对跟踪过程目标特征模板进行更新。文献[9]利用高斯混合模型对跟踪结果建模并建立目标模板库,再选出最可靠模板进行匹配自适应目标外观变化。文献[10-11]采用在线学习方式,使用目标训练样本快速优化模型权重,有效应对目标外观变化问题。文献[12]构建转换模型预测(transforming model prediction, ToMP)网络,利用其编码器与解码器在线更新模型权重,对外观变化目标具有较好跟踪效果。这些方法均利用目标可靠模板在线匹配跟踪目标,在目标发生规律变化时能稳定跟踪,但目标快速形变时,无法在连续帧信息中获益进而及时调整。文献[13]利用连续时间的目标跟踪信息优化在线学习样本储存机制,使模型有效利用目标新外观样本。文献[14-15]将每

一帧目标及背景信息进行关联,在跟踪目标的同时对相似物体进行建模。文献[16]基于 Transformer 网络充分利用连续帧中的时间信息,提高模型质量,进而有效跟踪目标。

这些方法为目标外观变化时稳定跟踪问题提供了思路,但仍存在问题:1)未考虑目标连续外观状态信息,无法判别目标外观变化情况;2)利用置信度分值判别样本可靠性,无法有效利用外观变化下目标样本;3)网络专注于获取目标特征表示,无法处理目标外观变化时相似物体的干扰。

综上,充分发挥 ToMP 的优势并对其改进,提出改进 ToMP 网络的外观变化下目标跟踪方法。即在 ToMP 网络基础上添加目标外观状态判别模块,增强网络对目标外观变化情况的适应能力,并优化网络在线样本储存机制,结合可靠样本增强网络对目标的分类能力。同时,改进网络样本利用机制,提高网络对新外观目标的分类效果。最后利用距离加权网络生成目标响应分数,增强目标特征映射效果并缓解相似物体对跟踪的干扰。

1 改进转换模型预测网络原理

1.1 网络结构

改进 ToMP 网络充分利用 ToMP 编码器与解码器权重更新机制,结合级联长短时记忆网络(long short-term memory, LSTM)对判别当前帧,并使用时间序列信息预测目标位置。网络结构如图 1 所示,主要由特征提取、权重更新、状态判别和距离加权等构成。

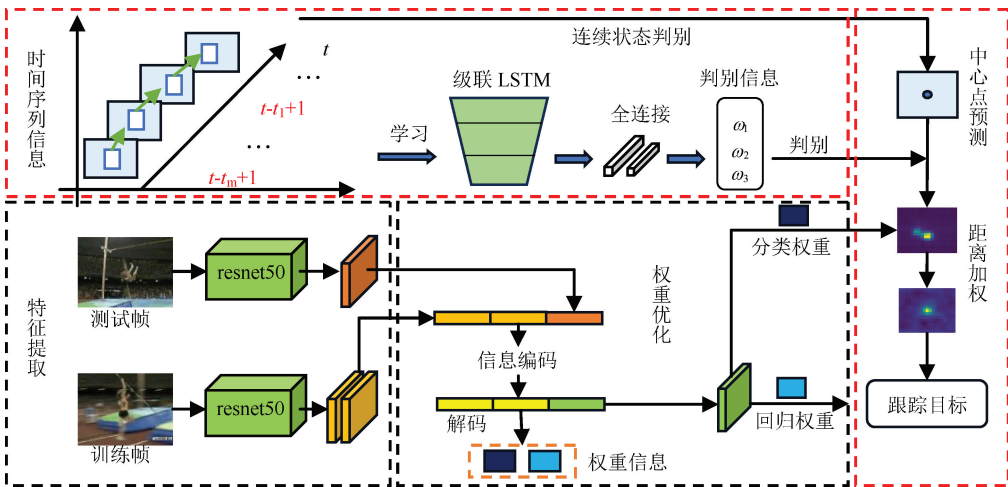


图 1 网络结构

Fig. 1 Network structure

1) 特征提取。使用 ResNet50 作为主干网络提取目标训练与测试特征信息,并将目标标签信息编码融合到训练特征中。

2) 权重优化。将处理后的训练和测试特征进行拼接,使用 Transformer 模型预测器进行联合处理,优化目标分类权重和边界框回归权重。

3) 状态判别。通过收集目标连续帧跟踪信息,利用级联 LSTM 网络判别当前帧目标外观状态,在目标发生外观变化时及时优化模型权重。

4) 距离加权。学习目标连续帧中心点轨迹并估计当前目标位置信息,同时结合状态判别方法,利用距离加权目标响应分数,减少相似物体干扰。

1.2 连续状态判别

为充分运用在线跟踪网络,通过收集跟踪过程目标序列信息,进一步挖掘目标状态,为模型在线更新权重提供可靠判别信息。在跟踪过程中,目标边界框直接反应当前帧目标几何状态,来自连续帧的边界框序列包含了目标的运动信息。同样,目标中心点反映目标在图像中的坐标位置,连续中心点坐标即可描述目标在图像中的运动轨迹。此外,通过收集目标前后帧中心点距离以反映目标运动信息,结合连续帧边界框 $b_t = (x_t, y_t, w_t, h_t)$, 目标中心点 $p_t = (x_{0t}, y_{0t})$ 为判别器提供数据支持。其中 (x, y) 为边界框左上角坐标, (w, h) 为目标的宽度和高度。

在线跟踪利用置信度分数判别当前帧目标状态,当主干网络提取目标图像特征后,利用模型权重获取目标在图像中的响应映射 R_t , 其响应映射最大值即为当前目标在图像中的位置。

$$S_t = \max(R_t) \quad (1)$$

其中, S_t 为目标置信度分值,以连续帧目标置信度分值判断目标状态变化。进一步,将响应映射 R_t 作为数据输入卷积神经网络以挖掘响应映射中的目标信息,获得连续的响应向量 $\mathbf{v}_t^R$ 。

$$\mathbf{v}_t^R = f^R(R_t; \mathbf{W}^R) \quad (2)$$

在获取到足够数量序列信息后,将前后帧距离、边界框 b_t 、中心点 p_t 、置信度分值 S_t 、响应向量 $\mathbf{v}_t^R$ 整合到一个序列矩阵中,

$$\mathbf{X}_t = [\mathbf{x}_{t-t_m+1}; \cdots; \mathbf{x}_{t-1}; \mathbf{x}_t] \quad (3)$$

其中, $\mathbf{X}_t \in \mathbf{R}^{d \times t_m}$, d 为序列信息维度, t_m 为历史序列到当前帧的时间步长。利用级联 LSTM 网络学习整合的序列信息,从而构建出稳定的目标状态判别器。

LSTM 可解决长序列训练过程中的梯度消失和梯度爆炸问题,其基本概念如下:

$$\begin{cases} f_t = \sigma(\mathbf{W}_f \mathbf{x}_t + \mathbf{U}_f \mathbf{h}_{t-1} + \mathbf{b}_f) \\ i_t = \sigma(\mathbf{W}_i \mathbf{x}_t + \mathbf{U}_i \mathbf{h}_{t-1} + \mathbf{b}_i) \\ o_t = \sigma(\mathbf{W}_o \mathbf{x}_t + \mathbf{U}_o \mathbf{h}_{t-1} + \mathbf{b}_o) \\ c_t = f_t \times c_{t-1} + i_t \times \tanh(\mathbf{W}_c \mathbf{x}_t + \mathbf{U}_c \mathbf{h}_{t-1} + \mathbf{b}_c) \\ h_t = o_t \times \tanh(c_t) \end{cases} \quad (4)$$

其中, f, i, o, c 分别表示遗忘门、输入门、输出门和储存单元, $\mathbf{W}, \mathbf{U}, \mathbf{b}$ 分别表示需要学习的权重矩阵和偏置向量。

使用三级级联 LSTM 模型提取序列信息并关注最新

帧状态信息。使用两个全连接层处理 LSTM 输出生成一个三元分类分数,最终判别目标状态。

为获取足够的时间序列训练样本,将 20 个样本作为一个样本集,间隔 4 个样本进行重叠采样,并为每个样本集进行标注。

$$l(Y_t) = \begin{cases} [0, 0, 1] & \text{IOU}(\mathbf{b}_t, \mathbf{g}_t) > 0.5 \\ [0, 1, 0] & \text{IOU}(\mathbf{b}_t, \mathbf{g}_t) = 0 \text{ and } S_t < 0.25 \\ [1, 0, 0] & \text{IOU}(\mathbf{b}_t, \mathbf{g}_t) < 0.25 \\ & \text{and IOU}(\mathbf{b}_0, \mathbf{g}_0) > 0.5 \end{cases} \quad (5)$$

其中, $Y_t = \{y_{t-t_m+1}, y_{t-t_m+2}, \cdots, y_{t-1}, y_t\}$, IOU 为模型跟踪框与真实框交并比, S_t 为目标置信度分值。

目标跟踪过程中,将储存正常跟踪样本用于在线学习,如目标发生外观变化,模型将舍去初始帧样本,并使用最新帧样本更新模型权重,以适应目标外观变化。同时,为避免模型崩坏,将初始帧信息间隔数帧和稳定跟踪后进行重新利用。

1.3 预测距离加权

跟踪目标所在区域背景信息复杂多变,特别在外观发生变化时目标在图像中的响应映射分数将降低,相似物体也易导致跟踪失败。因此,如能预测目标位置,就能有效判别目标外观变化和避免相似物体的干扰。由 1.2 节可知,连续中心点坐标可描述目标在图像中的运动轨迹,使用 LSTM 网络学习目标运动轨迹从而预测目标位置。

在目标初始定位后,后续帧通过不断收集目标可靠中心点坐标信息,同时预测下一帧目标位置信息,实现目标在图像中的轨迹预测。因此,每一帧跟踪结果将获得预测目标中心点位置 $p_p = (x_p, y_p)$, 当判别器输出当前帧出现外观变化时,将预测中心点作为目标中心位置重新剪裁图像,并基于距离对目标响应映射分数进行加权。

$$\mathbf{R}'_t = \omega_b e^{-\omega_d \|p_t - p_p\|_2} \times \mathbf{R}_t \quad (6)$$

其中, ω_b, ω_d 分别为映射分数和距离参数, p_t, p_p 分别为图像和预测中心点的响应映射位置坐标。

在修正目标响应映射后,通过边界框回归权重对目标准确定位,获取准确目标边界框,并及时将当前帧图像样本信息更新,让模型适应目标新外观。

2 跟踪方法流程

改进 ToMP 网络模型主要包括特征提取、权重更新、状态判别和距离加权构成,以 ResNet50 提取目标特征后使用 Transformer 优化目标分类权重和边界框回归权重,再利用三级级联 LSTM 判别目标外观状态,最后利用 LSTM 预测目标位置获取目标边界框,完成跟踪。跟踪流程如图 2 所示,主要步骤如下:

- 1) 提取视频序列图像特征信息,将可靠目标样本特征融入训练特征信息中优化模型权重。
- 2) 将最终优化模型权重对测试图像进行分类。
- 3) 利用连续目标中心点轨迹信息,预测当前帧目标位置。
- 4) 将连续目标边界框、中心点、置信度分值等信息输入级联 LSTM 网络进行学习,输出目标外观状态判别信息。
- 4) 目标正常,跟踪定位。返回 1) 或结束跟踪。
- 5) 目标发生外观变化,添加新目标外观特征优化模型权重对测试图像进行分类,利用目标中心点预测结果,使用距离加权目标响应映射,跟踪新外观目标。返回 1) 或结束跟踪。

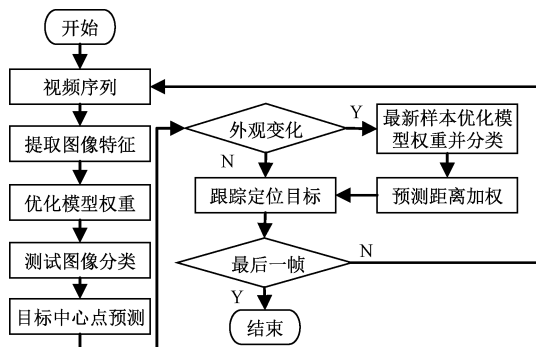


图2 跟踪方法流程

Fig. 2 Flow of the tracking method

3 实验结果分析

3.1 实验环境

运行环境为 Ubuntu18.04 系统,硬件为 Nvidia 1080Ti。映射加权参数 $\omega_b = 1.25$,距离加权参数 $\omega_d = 0.75$ 。训练模型实验设置训练轮数为 1 000,批处理大小为 1,训练期间的初始学习率为 0.001,并在后续使用自适应学习率调整。判别模型实验设置训练轮数为 10 000,批处理大小为 32,训练期间的学习率为 0.000 1。

实验使用 OTB100 和 LaSOT 公共数据集。其中 OTB100 数据集包括 100 个视频序列,每个视频序列都逐帧用边界框标注,并分类了 11 种挑战性属性。包括的属性有光照变化、尺寸改变、遮挡、形态变化、运动模糊、快速移动、内旋转、外旋转、视野外、背景杂乱和分辨率低。在分析目标外观变化情况时,选择其中 44 个标注目标形态变化属性视频序列进行实验。

LaSOT 数据集是一个大型目标跟踪数据集,包括了 1 400 个视频序列,分布在 70 个不同的对象类别中。为测试模型在处理目标外观变化时的跟踪性能,选择了其中具有目标形态变化属性的 142 个序列进行实验。

3.2 评价指标

使用准确率和成功率作为目标跟踪有效性指标。准确率定义为跟踪目标中心点与真实中心点间的平均欧式距离,通过计算序列中所有帧平均定位误差来评估整个序列的跟踪效果。准确率通过估计模型输出目标位置与真实位置在特定阈值距离内的帧数占比来计算每个模型的代表性准确率得分,以 20 pixels 为阈值用作衡量准确率的标准。

成功率为跟踪目标边界框和真实边界框的交并比值,如式(7)所示。

$$S = \frac{|r_t \cap r_a|}{|r_t \cup r_a|} \quad (7)$$

其中, r_t 为模型输出边界框, r_a 为目标真实边界框。通过计算交并比值 S 大于给定阈值的成功帧数量,成功率曲线显示了在阈值处成功帧的比例,并使用每个成功率曲线下面积来衡量模型性能。

3.3 公共数据集对比实验

为验证所提方法跟踪过程的优势以及有效性,将所提方法与现有跟踪方法 PrDiMP50^[11]、LTMU^[13]、SiamRCNN^[14]、KeepTrack^[15]、TrDiMP^[16] 等方法进行比较,结果如图 3 和 4 所示。

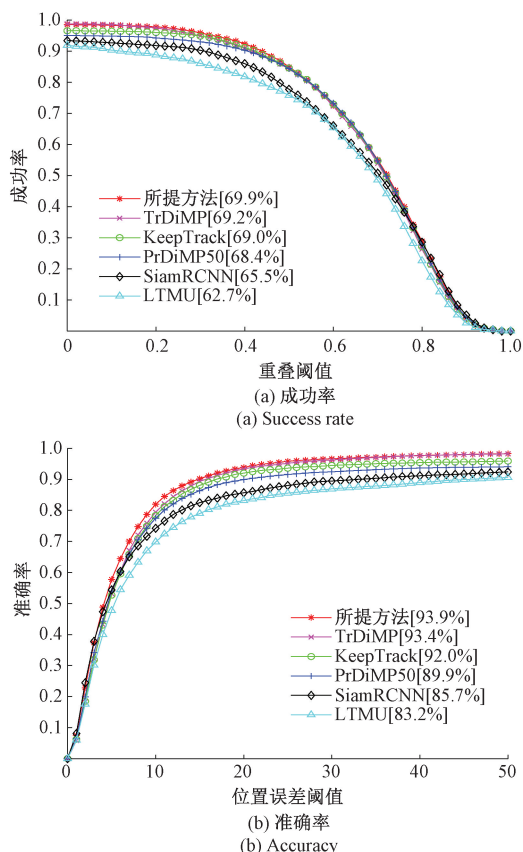


图3 OTB100 数据集对比

Fig. 3 Comparison of OTB100 data sets

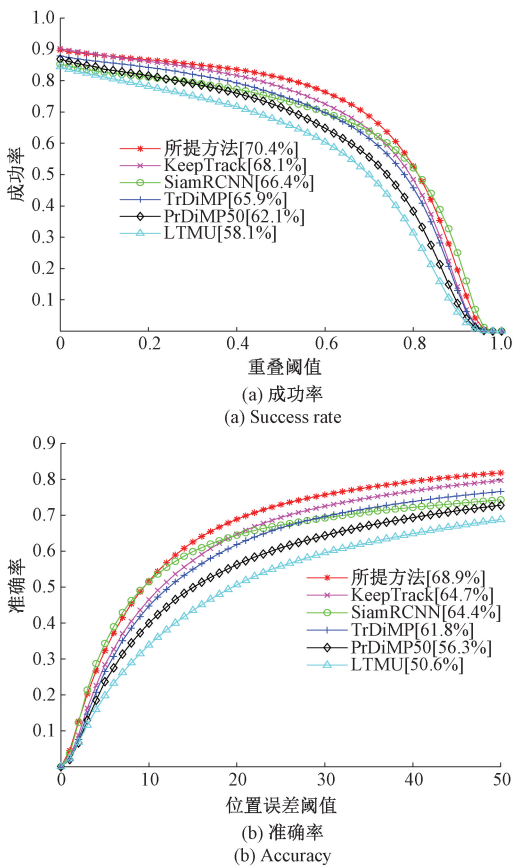


图4 LaSOT数据集对比

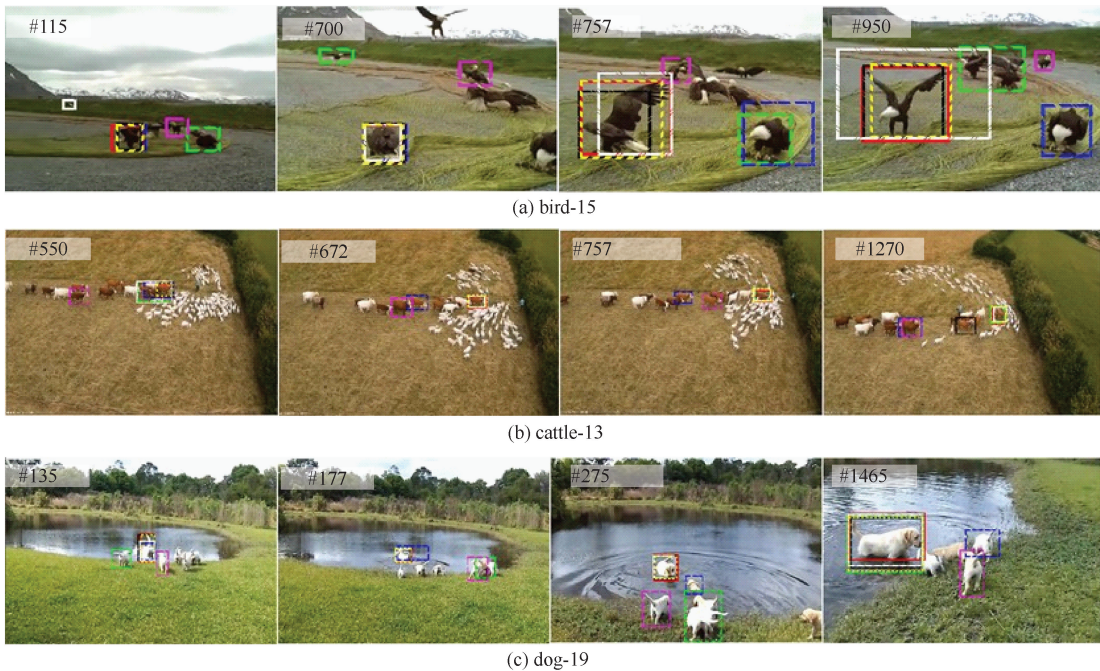
Fig. 4 Comparison of LaSOT data sets

根据比较结果,所提方法在 OTB100 数据集中准确率和成功率分别达到 93.9% 和 69.9%,均优于其他方法。在 LaSOT 数据集上,准确率和成功率分别为 68.9% 和

70.4%,明显优于其他方法。如图 3 所示,TrDiMP、KeepTrack 和 PrDiMP50 均具有较好准确率和成功率,明显优于其他 2 种方法,主要因为这 3 种方法均使用在线学习方式,利用目标跟踪过程中可靠样本进行模型权重优化,能较好适应目标外观变化情况。TrDiMP 方法利用 Transformer 编码器和解码器结构,充分利用连续帧之间的图像特征信息,具有较好的跟踪效果。SiamRCNN 方法仅利用目标初始模板作为样本跟踪目标,短时跟踪效果较差。LTMU 方法同样采用在线学习方式,但该方法的基础网络特征提取能力不强,在复杂环境下无法有效跟踪目标,跟踪效果不佳。

如图 4 所示,KeepTrack 和 SiamRCNN 在 LaSOT 数据集上具有较高的准确率和成功率。在跟踪过程中,目标外观变化的同时可能存在其他相似物体干扰,而这两种方法在跟踪过程中充分利用图像背景信息,降低了相似物体被错误跟踪的问题。PrDiMP50 和 LTMU 虽采用在线学习方式适应目标外观变化,但无法抵抗复杂环境下相似物体干扰问题,在跟踪目标发生外观变化时,将错误跟踪到相似物体,因此跟踪效果较差。所提方法使用在线学习方式适应目标外观变化,且利用判别模型区分目标外观状态,将跟踪过程的可靠样本合理利用,通过对目标中心点轨迹进行预测,充分利用目标时间序列位置信息,并进一步使用距离加权目标响应分数,有效抑制相似物体的干扰,具有更高的跟踪准确率和成功率。

为更直观分析所提方法优势及有效性,选取 5 组视频序列 (bird-15、cattle-13、dog-19、rabbit-10、sheep-5) 进行比较。视频序列中目标外观变化时不同方法跟踪结果如图 5 所示,准确率和成功率如表 1、2 所示。



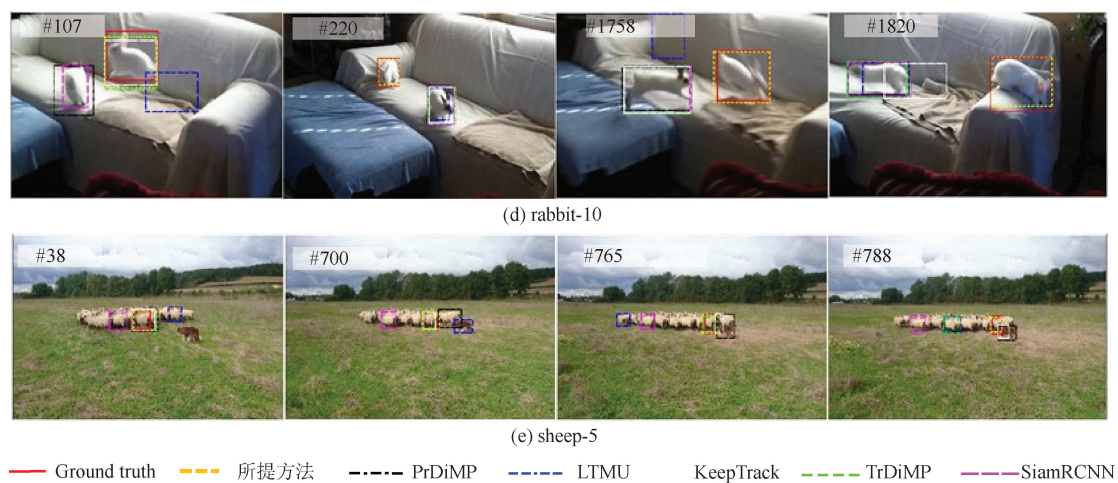


图 5 不同方法跟踪结果

Fig. 5 Tracking results by different methods

表 1 跟踪准确率对比

Table 1 Comparison of tracking accuracy (%)

方法	bird-15	cattle-13	dog-19	rabbit-10	sheep-5
TrDiMP	77.8	67.6	51.9	58.0	74.6
KeepTrack	94.0	69.5	64.9	61.0	82.6
PrDiMP	98.4	54.9	55.2	44.7	64.8
SiamRCNN	26.5	18.9	43.3	58.7	50.8
LTMU	61.7	36.2	16.8	45.9	68.8
本文方法	98.5	71.6	59.3	70.6	81.6

表 2 跟踪成功率 (AUC) 对比

Table 2 Comparison of tracking success rate (AUC) (%)

方法	bird-15	cattle-13	dog-19	rabbit-10	sheep-5
TrDiMP	67.9	54.3	65.6	61.0	58.3
KeepTrack	83.8	56.5	77.2	59.6	64.4
PrDiMP	82.0	44.6	65.1	50.5	53.7
SiamRCNN	23.7	15.9	47.3	58.7	44.7
LTMU	52.5	33.0	33.4	49.2	50.0
本文方法	87.8	58.8	75.6	65.0	68.3

根据图 5 和表 1、2,所提方法在 5 组视频序列中准确率和成功率明显优于其他方法。在 bird-15 序列中,所提方法准确率达到 98.5%,相比 SiamRCNN 方法提升 72%。该序列中,SiamRCNN 方法在目标发生外观变化时无法及时修正,导致长时间跟踪失败;在 cattle-13 序列中,所提方法和 KeepTrack 方法具有较高的跟踪准确率和成功率,在第 550 帧中,目标周围出现大量相似物体干扰,其他方法均无法准确跟踪目标,而所提方法得益于判别模型提供的有效目标样本,且通过距离加权了响应分数,成

功持续跟踪到目标;在 dog-19 序列中,所提方法略低于 KeepTrack 方法,该方法通过跟踪相似物体可有效缓解相似干扰问题,在第 275 帧中,目标发生外观变化,大部分方法将相似物体作为跟踪目标,而所提方法利用预测中心点加权方式,同样有效减少了远处相似物体干扰问题,从而持续跟踪目标;在 rabbit-10 序列中,所提方法明显优于其他方法,在第 1 758 帧中,目标发生快速移动且产生一定外观变化,其他方法无法有效跟踪到目标,而所提方法充分利用目标最新样本信息进行学习,并预测到目标新位置,最终成功跟踪目标;在 sheep-5 序列中,大部分方法在大量相似物体干扰下无法准确跟踪到目标,而所提方法跟踪准确率达到 80% 以上,在相似物体干扰环境下仍持续跟踪目标。

3.4 效果分析实验

为分析所提连续状态判别、预测距离加权方法的有效性,进行消融实验,结果如表 3 所示。

表 3 消融实验对比

Table 3 Comparison of ablation experiments (%)

连续状态判别	预测距离加权	准确率	提升
×	×	90.8	
✓	×	91.5	0.7 ↑
✓	✓	93.9	3.1 ↑

观察该表,使用连续状态判别方法后,在线学习方式得到进一步优化,跟踪准确率提升 0.7%。在目标外观变化时使用预测中心点距离加权方法,跟踪准确率再次提升 2.4%,最终跟踪准确率提升 3.1%,说明所提方法提高了目标外观变化时的跟踪能力。

为说明轨迹预测方法在跟踪过程中的有效性,使用

Bird1、Crossing、Diving 等 6 个视频序列轨迹预测中心点与真实目标中心点的平均欧氏距离作为参考,结果如表 4 所示。根据表 4,轨迹预测方法的平均欧式距离均小于 20 个 pixels,Diving 序列的平均欧氏距离仅为 7.1 个 pixels,说明了轨迹预测方法的有效性。将轨迹预测中心点与目标真实中心点在图像中绘制,如图 6 所示,轨迹预测曲线基本与真实轨迹线吻合,进一步证明了轨迹预测方法的有效性。

表 4 平均欧氏距离误差/像素点

Table 4 Average Euclidean distance error/pixels					
Bird1	Crossing	Diving	Human8	Jump	Panda
12.35	16.51	7.10	9.42	17.65	9.87



图 6 预测与真实轨迹对比

Fig. 6 Comparison of predicted and real trajectories

为分析所提方法应对不同外观变化程度时的跟踪效果,使用多种目标外观变化程度情况进行实验验证,不同情况下跟踪结果分数和跟踪效果图如表 5 和图 7 所示。观察图 7,所提方法在目标发生外观变化时具有较好的跟踪效果,在变化过程中可对不同程度目标外观进行稳定跟踪。观察表 5,所提方法在多种目标外观变化情况下成功率均达到 87% 以上,进一步证明了所提方法可有效应对目标外观变化并稳定跟踪。

表 5 外观变化情况跟踪结果

Table 5 Appearance change tracking results (%)			
外观变化情况	形态变化	特征变化	视角变化
成功率(AUC)	90.7	88.0	87.9

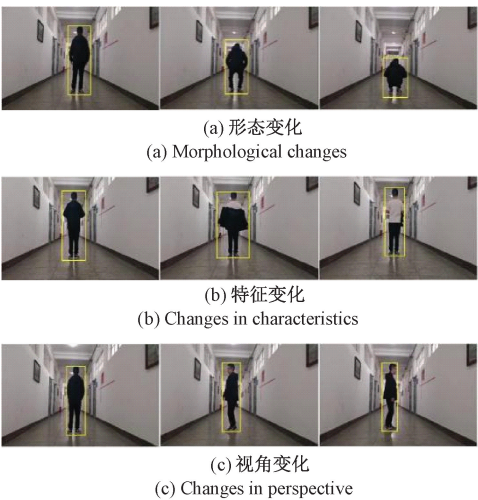


图 7 不同外观变化程度跟踪效果

Fig. 7 Tracking effect of different appearance change degree

3.5 移动机器人实验

为验证所提方法在真实场景中跟踪有效性,使用四足移动机器人进行跟踪实验,如图 8 所示。四足移动机器人配置了 Intel RealSense D435i 摄像头,摄像头分辨率为 1 920×1 080,帧速率为 30。对真实场景中目标发生外观变化情况进行实验分析,给出目标发生外观变化时的跟踪效果,如图 9 所示。



图 8 四足移动机器人平台

Fig. 8 Quadruped mobile robot platform



图 9 真实场景跟踪效果

Fig. 9 Real scene tracking effect

由图 9,所提方法在目标发生外观变化时具有较好的跟踪效果。在第 168 帧中,目标出现外观变化,所提方法仍持续跟踪新外观状态下的目标;在第 603 帧中,目标

快速跳起且发生一定外观变化,所提方法仍持续跟踪快速运动目标,说明了所提方法可有效处理目标外观变化问题。

4 结 论

为解决跟踪过程中目标出现外观变化导致跟踪效果不佳的问题,提出了改进 ToMP 网络的外观变化下目标跟踪方法。利用连续目标边界框、中心点、置信度分值等信息判别目标外观状态,改进了在线学习样本利用机制,让网络更快适应目标最新外观,并基于预测位置加权了目标响应分数,增强目标特征映射效果的同时有效缓解了相似物体的干扰。所提方法实现了外观变化下目标稳定跟踪,弥补了当前跟踪方法在目标外观变化时易丢失的问题,进一步丰富和发展了目标跟踪理论和方法,具有良好的工程应用潜力。

参考文献

- [1] 张津浦,王岳环. 融合检测技术的孪生网络跟踪算法综述[J]. 红外与激光工程,2022,51(10):318-331.
- [2] FU CH H, LI B W, DING F Q, et al. Correlation filters for unmanned aerial vehicle-based aerial tracking: A review and experimental evaluation[J]. Geoscience and Remote Sensing, 2022, 10(1):125-160.
- [3] LI B W, FU CH H, DING F Q, et al. All-day object tracking for unmanned aerial vehicle [J]. IEEE Transactions on Mobile Computing, 2023, 22(8):4515-4529.
- [4] 韩瑞泽,冯伟,郭青,等. 视频单目标跟踪研究进展综述[J]. 计算机学报,2022,45(9):1877-1907.
- [5] 陈仁祥,何家乐,杨黎霞,等. 基于加权在线样本更新的目标长时跟踪方法[J]. 仪器仪表学报,2023,44(6):66-73.
- [6] CHAN S X, TAO J, ZHOU X L, et al. Siamese implicit region proposal network with compound attention for visual tracking [J]. IEEE Transactions on Image Processing, 2022, 31:1882-1894.
- [7] HU W M, WANG Q, ZHANG L, et al. Siammask: A framework for fast online object tracking and segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(3):3072-3089.
- [8] 熊珺瑶,王蓉,孙义博. 外观动作自适应目标跟踪方法[J]. 北京航空航天大学学报, 2022,48(8):1525-1533.
- [9] 刘芳,杨安喆,吴志威. 基于自适应 Siamese 网络的无人机目标跟踪算法[J]. 航空学报, 2020,41(1):248-260.
- [10] LIU F, YANG AN ZH, WU ZH W. Uav target tracking algorithm based on adaptive Siamese network[J]. Acta Aeronautica Sinica, 2020,41(1):248-260.
- [11] BHAT G, DANELLJAN M, GOOL L V, et al. Learning discriminative model prediction for tracking[C]. 2019 IEEE/CVF International Conference on Computer Vision, 2019: 6181-6190.
- [12] DANELLJAN M, GOOL L V, TIMOFTE R. Probabilistic regression for visual tracking [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020:7181-7190.
- [13] MAYER C, DANELLJAN M, BHAT G, et al. Transforming model prediction for tracking[C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022:8721-8730.
- [14] DAI K, ZHANG Y H, WANG D, et al. High-performance long-term tracking with meta-updater[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020:6297-6306.
- [15] VOIGTLAENDER P, LUITEN J, TORR P H S, et al. Siam R-CNN: Visual tracking by re-detection[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020:6577-6587.
- [16] MAYER C, DANELLJAN M, PAUDEL D P, et al. Learning target candidate association to keep track of what not to track [C]. 2021 IEEE/CVF International Conference on Computer Vision. 2021:13424-13434.
- [17] WANG N, ZHOU W G, WANG J, et al. Transformer meets tracker: Exploiting temporal context for robust

visual tracking [C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 1571-1580.

作者简介



陈仁祥, 分别于 2007 年和 2012 年在重庆大学获得学士学位和博士学位, 现为重庆交通大学教授、博士生导师, 主要研究方向为智能测试技术与信号处理。
E-mail: manlou.yue@ 126. com

Chen Renxiang received his B. Sc. degree and Ph. D. degree both from Chongqing University in 2007 and 2012, respectively. He is currently a professor and a master advisor at Chongqing Jiaotong University. His main research

interests include intelligent testing technology and signal processing.



杨黎霞(通信作者), 分别于 2012 年和 2018 年于重庆交通大学获得硕士和博士学位, 现为重庆科技大学讲师, 主要研究方向为信号处理与大数据分析。
E-mail: lixiayang1207@ 126. com

Yang Lixia (Corresponding author) received her M. Sc. degree in 2012 and Ph. D. degree in 2018 from Chongqing Jiaotong University, Chongqing, China. Now she is a lecturer at Chongqing University of Science and Technology. Her main research interest is signal processing and big data analysis.