

多信息融合的卫星视频单目标跟踪^①



王丽黎, 张 慧

(西安理工大学 自动化与信息工程学院, 西安 710048)

通信作者: 张 慧, E-mail: 2359498987@qq.com

摘 要: 针对卫星视频中存在目标特征信息少、前景背景对比性低等问题, 在 SiamCAR 的基础上提出了一种融合运动信息和注意力机制的目标跟踪方法; 首先引入运动激励模块和通道注意力模块以增强目标特征提取信息; 然后将相邻帧作为新模板添加到网络里形成三重网络补充模板信息; 最后加入卡尔曼滤波算法进行目标轨迹预测, 将预测模板添加到网络中形成四重网络增加目标的运动信息; 选取 SatSOT 卫星视频数据集中的 10 组数据进行测试, 实验结果表明与 SiamCAR 网络相比, 改进算法的跟踪准确率和成功率分别提升了 6% 和 6.2%。

关键词: 卫星视频; 目标跟踪; SiamCAR; 卡尔曼滤波; 注意力机制; 信息融合

引用格式: 王丽黎, 张慧. 多信息融合的卫星视频单目标跟踪. 计算机系统应用, 2023, 32(2): 266-273. <http://www.c-s-a.org.cn/1003-3254/8995.html>

Single Target Tracking of Satellite Video Based on Multi-information Fusion

WANG Li-Li, ZHANG Hui

(School of Automation and Information Engineering, Xi'an University of Technology, Xi'an 710048, China)

Abstract: Aiming at the problems of less target feature information and low contrast between foreground and background in satellite video, this study proposes a target tracking method integrating motion information and attention mechanism based on SiamCAR. First, the motion excitation and channel attention modules are introduced to enhance the target feature extraction information. Then, adjacent frames are regarded as new templates and added to the network to form a triple network and supplement template information. Finally, the Kalman filter algorithm is added to predict the target's trajectory, and a prediction template is introduced to the network to construct a quadruple network and increase the motion information of the target. In addition, 10 sets of data in the SatSOT satellite video data set are selected for testing. The experimental results show that compared with those of the SiamCAR network, the tracking accuracy and success rate of the improved algorithm are increased by 6% and 6.2%, respectively.

Key words: satellite video; target tracking; SiamCAR; Kalman filter; attention mechanism; information fusion

随着卫星技术的发展, 卫星视频中的目标跟踪成为新的研究领域. 而与普通视频相比, 卫星视频中的目标相对于背景只占少量像素, 这会使卫星视频中的目标跟踪更容易产生漂移现象.

目标跟踪分为基于相关滤波和基于深度学习的方法. 基于相关滤波的方法依赖手工特征通过相关计算进行目标的位置预测, 常见的相关滤波算法有 MOSSE^[1]、CSK^[2]、KCF^[3]、CN^[4] 等. 基于深度学习的孪生网络跟

踪方法通过提取深度特征在精度和速度之间达到了平衡, 这种方法简单高效, 受到广泛学者的关注. Bertinetto 等^[5] 提出 SiamFC 算法, 该算法使用相同的网络结构提取模板和搜索区域的特征, 然后利用相关计算得到目标位置. Li 等^[6] 在 SiamFC 的基础上加入区域候选网络, 提出了 SiamRPN 算法, 有效解决了目标尺度变化问题. Li 等^[7] 在 SiamRPN 的基础上引入残差网络和通道互相关策略, 提出了 SiamRPN++ 算法. Guo 等^[8] 在

① 收稿时间: 2022-07-12; 修改时间: 2022-09-07, 2022-09-15; 采用时间: 2022-09-27; csa 在线出版时间: 2022-12-09

CNKI 网络首发时间: 2022-12-13

SiamRPN++的基础上将分类和回归并行计算,提出了 SiamCAR 算法,实现无先验框跟踪,避免了很多先验框超参数的设置. Shi 等^[9]在 SiamFC 的基础上增添一个新分支提出了 TripFC 算法.

这些方法主要针对传统视频的目标跟踪进行的改进,但在卫星视频跟踪中还需要根据目标特征不明显的特点进行优化. Shao 等^[10]在孪生网络中加入目标的像素级运动特征保证目标更精确的定位. Zhang 等^[11]利用全卷积网络预测像素位置提高跟踪精度. Wu 等^[12]引入慢特征和运动特征提高目标跟踪性能. Ruan 等^[13]通过拟合目标运动轨迹缓解模型漂移问题. Song 等^[14]在孪生网络中加入通道、空间和交叉注意力机制,实现更精确的目标定位.

与上述算法不同,本文以无锚框 SiamCAR 算法为基本框架,将通道注意力模块和运动激励模块结合,更多的关注得分较高的通道和运动位置,并且添加相邻帧分支和卡尔曼预测分支形成四重网络,提高算法在卫星视频中的目标跟踪能力. 我们的算法仅需额外训练注意力机制模块,增添的相邻帧和卡尔曼预测模块

不需要训练,可以直接使用.

1 相关理论

SiamCAR 将跟踪问题分解成分类和回归两个子问题,以像素的方式进行求解. 该框架主要包括:一个用于特征提取的骨干网络和一个用于边界框预测的分类回归子网.

1.1 特征提取

特征提取子网采用改进的 ResNet-50 网络^[15],如图 1, ResNet-50 网络包含 49 个卷积层、1 个全连接层,第 1 部分结构主要对输入进行卷积、正则化、激活函数和最大池化计算,第 2-5 部分结构中都增添了残差块,解决了原 CNN 网络过深时准确率下降的问题.

改进的 ResNet-50 将后 3 层的特征进行拼接分别得到 template patch 的特征图 $\varphi(Z)$ 和 search region 的特征图 $\varphi(X)$, 之后执行深度互相关计算获得响应图 R :

$$R = \varphi(Z) \star \varphi(X) \quad (1)$$

其中, \star 表示深度相关. 然后将响应图 R 通过 1×1 的卷积层进行降维得到 R^* , 以减少模型参数量, 加快后续的计算.

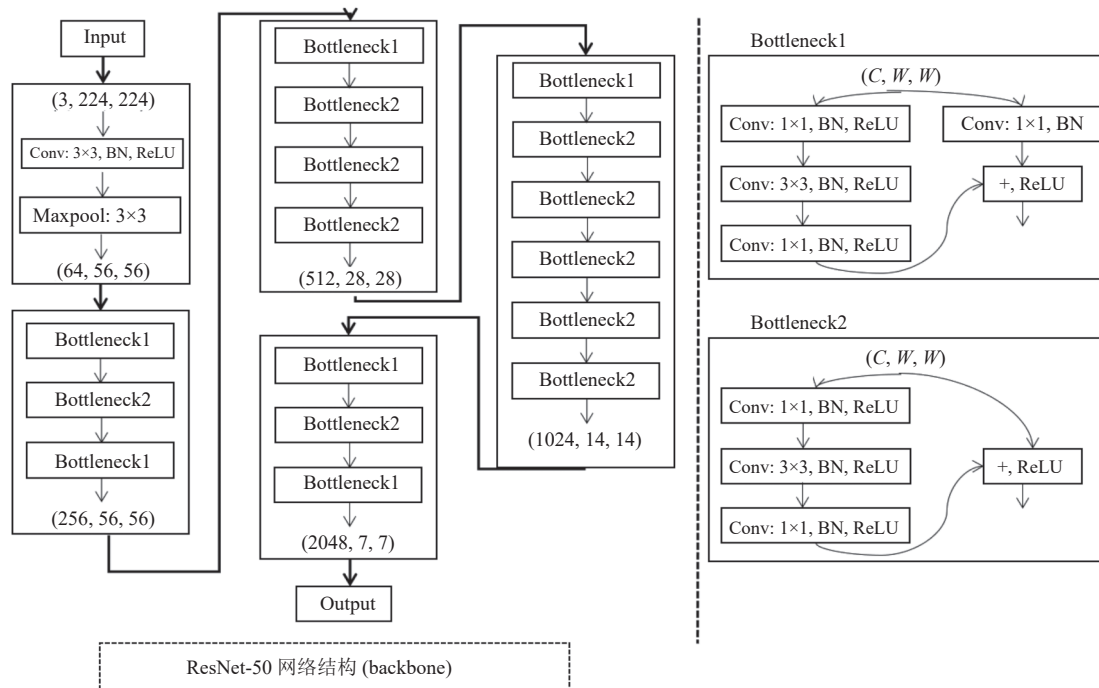


图 1 ResNet-50 网络结构

1.2 边界框预测

将特征提取阶段得到的 $R^*_{w \times h \times m}$ 作为输入, 分类回归子网通过分类分支和回归分支输出, 其中分类分支

输出分类特征图 $A_{w \times h \times 2}^{cls}$ 和中心度特征图 $A_{w \times h \times 1}^{cen}$, 回归分支输出回归特征图 $A_{w \times h \times 4}^{reg}$. $A_{w \times h \times 2}^{cls}$ 中每个点都包含一个二维向量, 该向量表示输入搜索区域中相应位置的前

景和背景分数; $A_{w \times h \times 4}^{res}$ 中每个点都包含一个四维向量 $t(i, j) = (l, t, r, b)$, 该向量表示从相应位置到输入搜索区域中边界框四边的距离, 设 (x_0, y_0) 和 (x_1, y_1) 表示实际边界框的左上角和右下角, 并设 (x, y) 表示点 (i, j) 的对应位置, 则回归目标的公式为:

$$\tilde{l} = x - x_0 \quad (2)$$

$$\tilde{t} = y - y_0 \quad (3)$$

$$\tilde{r} = x_1 - x \quad (4)$$

$$\tilde{b} = y_1 - y \quad (5)$$

$A_{w \times h \times 1}^{cen}$ 中每个点给出相应位置的中心度分数, 公式定义为:

$$C(i, j) = I(\tilde{t}_{i,j}) * \sqrt{\frac{\min(\tilde{l}, \tilde{r})}{\max(\tilde{l}, \tilde{r})} \times \frac{\min(\tilde{t}, \tilde{b})}{\max(\tilde{t}, \tilde{b})}} \quad (6)$$

1.3 跟踪阶段

对于每个位置都会生成对应的 6 维向量 $T_{i,j} = (cls, cen, l, t, r, b)$, 其中 cls 表示分类的前景分数, cen 表示

中心度分数, $l+r$ 和 $t+b$ 表示当前帧中目标的预测宽度和高度. 然后加入尺度惩罚 $p_{i,j}$, 将跟踪阶段的公式定义为:

$$q = \arg \max_{i,j} \{(1 - \lambda_d) cls_{ij} \times p_{ij} + \lambda_d H_{ij}\} \quad (7)$$

其中, H 是余弦窗, λ_d 是平衡权重, q 是最大分数像素对应的位置.

2 改进的卫星视频目标跟踪算法

将 SiamCAR 算法作为基本框架, 首先在特征提取模块中加入运动激励模块 (ME) 和改进的高效通道注意力 (ECA) 模块; 然后加入相邻帧模板和卡尔曼预测模板补充对象模板以形成四重网络结构.

2.1 特征提取模块

特征提取模块以改进的 ResNet-50 为基础, 在骨干网络的残差单元加入了运动激励模块和改进的高效通道注意力模块, 改进的残差单元结构如图 2 所示.

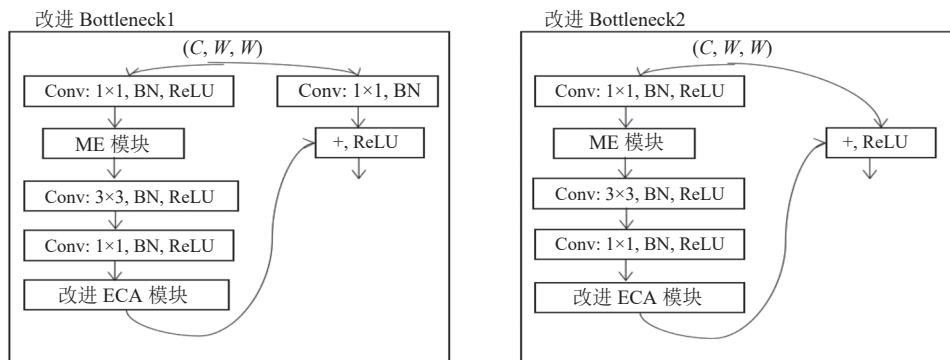


图 2 改进的残差单元结构

2.1.1 运动激励模块

运动激励模块^[16] 将短距离运动建模与整个时空特征学习方法相结合, 激发特征的运动敏感通道, 以解决卫星视频中前景背景对比性低的问题. 首先输入形状为 $[N, T, C, H, W]$ 的时空特征 X , 其中 N 是批量大小, T 和 C 分别表示时间维度和特征通道, H 和 W 表示形状. 将输入特征 X 经过 1×1 的 2D 卷积得到信道压缩特征 X' . 接着对 X' 进行通道变换, 计算 t 时刻的运动特征为:

$$M(t) = conv_{trans} * X'(t+1) - X'(t), 1 \leq t \leq T-1 \quad (8)$$

其中, $conv_{trans}$ 是一个 3×3 的 2D 通道卷积层.

接着, 连接所有运动特征得到最终运动矩阵 M , 使用全局平均池层来总结空间信息得到 M^s , 将得到的空

间信息进行 1×1 卷积, 并使用 Sigmoid 函数获得运动注意力权重 A :

$$A = 2Sigmoid(conv_{exp} * M^s) - 1, A \in R^{N \times T \times C \times 1 \times 1} \quad (9)$$

最后, 将输入特征 X 与注意力权重 A 进行通道乘法, 得到了激励运动敏感通道和抑制静态背景信息的特征, 计算公式如下:

$$X^o = X + X \otimes A, X^o \in R^{N \times T \times C \times H \times W} \quad (10)$$

其中, X^o 是运动模式被激发的输出, \otimes 表示通道乘法.

2.1.2 改进的高效通道注意力模块

高效通道注意力模块^[17] 是在 SE-Net 通道注意力基础上采用无降维的局部跨通道交互策略的一种注意

力机制,该模块涉及少量参数,通过增强有用信息的通道,抑制信息较少的通道,带来明显的性能增益.卫星视频中目标相对于背景只占少量像素,原 ECA 模块利用通道的全局平均值计算,会带来很多与目标无关的背景信息,因此本文加入全局最大值计算,将全局平均值和全局最大值得到两个空间信息进行叠加优化成更好的通道注意力模块,如图 3 所示.

首先将大小为 $H \times W \times C$ 的输入特征分别进行全局最大池化和全局平均池化,将两个结果进行叠加得到

$1 \times 1 \times C$ 的压缩特征, 然后进行自适应 k 相邻 1 维卷积获得局部跨信道交互信息, 自适应 k 值的计算公式如下:

$$k = \left\lfloor \frac{\log_2 C}{\gamma} + \frac{b}{\gamma} \right\rfloor_{\text{odd}} \quad (11)$$

其中, C 为通道数, γ 和 b 为调节参数, $\lfloor f(x) \rfloor_{\text{odd}}$ 为获取值的最近奇数.

最后,使用 *Sigmoid* 函数将得到的结果转换成 0-1 之间的数值,作为权值与输入的特征对应相乘,得到带有通道注意力的特征图.

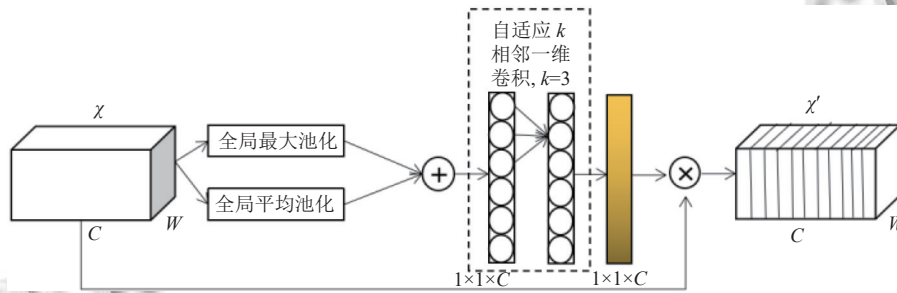


图 3 改进的 ECA 模块

2.2 四重网络模型构建

SiamCAR 使用初始帧作为对象模板,在跟踪过程中,通常会面临着遮挡、光照等影响,此时初始帧模板

中包含的信息不足以支持后续帧的跟踪.为了提高跟踪器的性能,我们加入了相邻帧模板和卡尔曼预测模板,用来补充对象模板以形成四重网络,如图 4 所示.

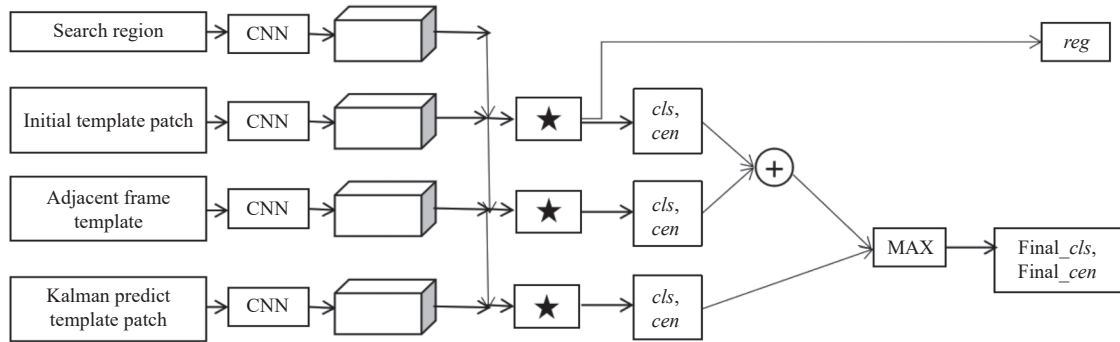


图 4 改进算法的网络框架

2.2.1 加入相邻帧模板

SiamCAR 算法仅利用首帧模板进行目标跟踪,当目标在后续跟踪中受到干扰时会导致模板信息不准确,于是在 SiamCAR 的基础上添加一个相邻帧模板,将初始帧和相邻帧的信息结合以增强模板特征的表达能.分别将初始帧模板和相邻帧模板与搜索区域进行深度互相关计算,计算公式如下:

$$R_1 = \varphi(Z_1) \star \varphi(X) \quad (12)$$

$$R_2 = \varphi(Z_2) \star \varphi(X) \quad (13)$$

其中, R_1 和 R_2 的大小相同,将两者进行叠加得到最终响应图,计算公式如下:

$$A_{cls} = \varphi_{cls}(R_1) + \varphi_{cls}(R_2) \quad (14)$$

$$A_{cen} = \varphi_{cen}(R_1) + \varphi_{cen}(R_2) \quad (15)$$

$$A_{reg} = \varphi_{reg}(R_1) \quad (16)$$

$$P_{score1} = A_{cls} \times A_{cen} \quad (17)$$

2.2.2 加入卡尔曼预测模块

由于卫星视频中目标相对于背景较小,特征信息

不明显,于是在 SiamCAR 的基础上增添一个卡尔曼^[18]预测的模板分支,对目标的运动轨迹进行预测,将预测的目标位置与跟踪算法得到的目标位置进行最优值计算,减少目标漂移问题,提高目标的跟踪性能.卡尔曼滤波器系统的状态方程和观测方程分别为:

$$x_t = A_{t-1}x_{t-1} + w_{t-1} \quad (18)$$

$$z_t = H_t x_t + v_t \quad (19)$$

其中, A_{t-1} 是系统在 $t-1$ 帧的状态转移矩阵, H_t 是系统在 t 帧的观测矩阵, x_t 和 x_{t-1} 分别是系统 t 帧和 $t-1$ 帧的系统状态的最佳值, z_t 是状态矩阵的观测量, w_{t-1} 和 v_t 分别是均值为 0、协方差矩阵为 Q 和 R 的高斯白噪声.

然后利用上一时刻运动状态预测此时此刻的状态变量值,如式(20)所示:

$$\hat{x}_t^- = A_t \hat{x}_{t-1} \quad (20)$$

利用上一时刻误差协方差值预测此时此刻的误差协方差值,如式(21)所示:

$$P_t^- = A_{t-1} P_{t-1} A_{t-1}^T + Q_{t-1} \quad (21)$$

使用卡尔曼增益系数作为权重,如式(22)所示:

$$K_t = P_t^- H^T (H P_t^- H^T + R_t)^{-1} \quad (22)$$

然后取状态变量预测值与观测值的加权和作为最终状态变量,如式(23)所示:

$$\hat{x}_t = \hat{x}_t^- + K_t (z_t - H_k \hat{x}_t^-) \quad (23)$$

误差协方差值更新,如式(24)所示:

$$P_t = P_t^- - K_t H_t P_t^- \quad (24)$$

利用卡尔曼滤波预测的模板与搜索区域进行深度互相关计算,计算公式如下:

$$R_3 = \varphi(Z_3) \star \varphi(X) \quad (25)$$

$$A_{cls}^* = \varphi_{cls}(R_3) \quad (26)$$

$$A_{cen}^* = \varphi_{cen}(R_3) \quad (27)$$

$$P_{score2} = A_{cls}^* \times A_{cen}^* \quad (28)$$

2.2.3 跟踪阶段

根据前两个阶段分别得到的响应图 P_{score1} 和 P_{score2} 进行最大值比较,选取响应值最大位置作为目标跟踪的最终位置,计算公式如下:

$$P_{score} = \max(P_{score1}, P_{score2}) \quad (29)$$

$$q = \arg \max(P_{score}) \quad (30)$$

3 实验结果分析

3.1 实验环境及配置

本文实验所使用的环境配置为 Windows 11、CUDA 11.6. 硬件配置为 Intel-CPU-i5-11260H 处理器,显卡为 Nvidia GeForce RTX 3050Ti. 本文选取 SatSOT 卫星视频数据集^[19]里的 10 组数据作为本文实验的测试序列, SatSOT 是第 1 个密集注释的卫星视频目标跟踪数据集,覆盖了卫星视频中特定的移动目标类别,包括汽车、火车、飞机和船舶.为了验证本文算法的鲁棒性和有效性,采取成功率和准确率作为跟踪评价标准,将 SiamCAR 算法与本文算法进行对比.

3.2 算法结果比较分析

3.2.1 定量分析

为了验证算法的跟踪性能,在选取的 10 组数据集上测试了 SiamCAR 算法和本文算法.表 1 为实验所得数据.本文算法在 10 组数据集的准确率和成功率相比 SiamCAR 算法分别提高了 6% 和 6.2%;本文算法在 plane_02 序列中的准确率和成功率相比 SiamCAR 算法分别提高了 0 和 45.1%;本文算法在 ship_01 序列中的准确率和成功率相比 SiamCAR 算法分别提高了 57.2% 和 5.1%;本文算法在 train_03 序列中的准确率和成功率相比 SiamCAR 算法分别提高了 3.3% 和 7.5%. 分别如图 5-图 8 所示.

表 1 在 SatSOT 数据集上两种算法对比

数据集	指标	OURS	SiamCAR
SatSOT中10组数据集	Precision (%)	83.6	77.4
	Success rate (%)	69.1	57.1
	Speed (FPS)	8.1	21.0
plane_02数据集	Precision (%)	100	100
	Success rate (%)	85.5	80.4
	Speed (FPS)	8.6	17.2
ship_01数据集	Precision (%)	100	42.8
	Success rate (%)	35.3	16.8
	Speed (FPS)	8.6	20.9
train_03数据集	Precision (%)	25.8	22.5
	Success rate (%)	41.8	34.3
	Speed (FPS)	7.2	22.2

3.2.2 定性分析

为了更直观评估算法的跟踪性能,将本文算法与 SiamCAR 算法分别在 plane_04、ship_01 和 train_03 视频序列上进行了对比实验,如图 9-图 11 所示.其中红色框代表本文算法跟踪结果、蓝色框代表 SiamCAR 算法跟踪结果.图 9-图 11 表明,在卫星视频目标跟踪中,本文算法相比于 SiamCAR 算法可以更贴合目标的尺寸,更好的跟踪目标.

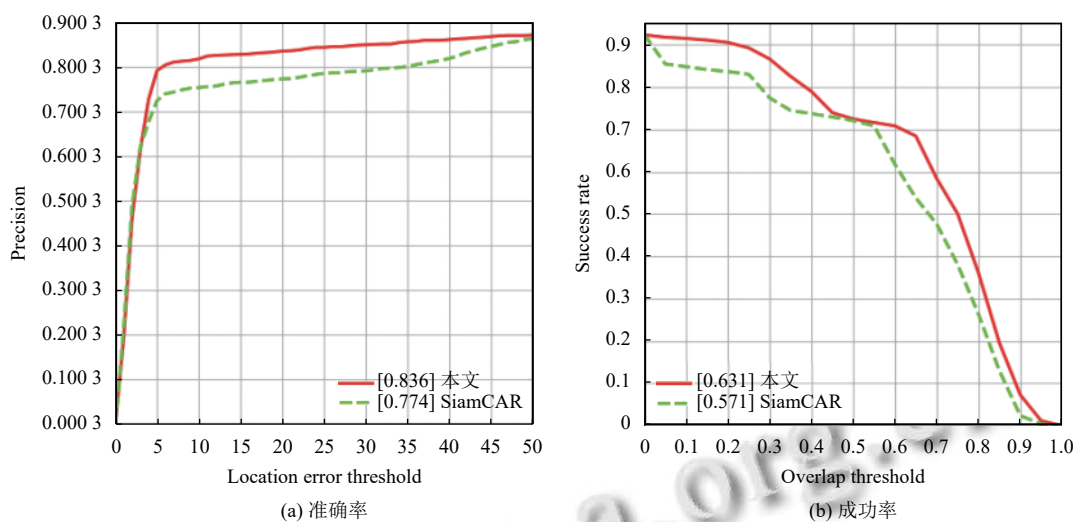


图5 SatSOT 中 10 组数据集算法准确率和成功率对比图

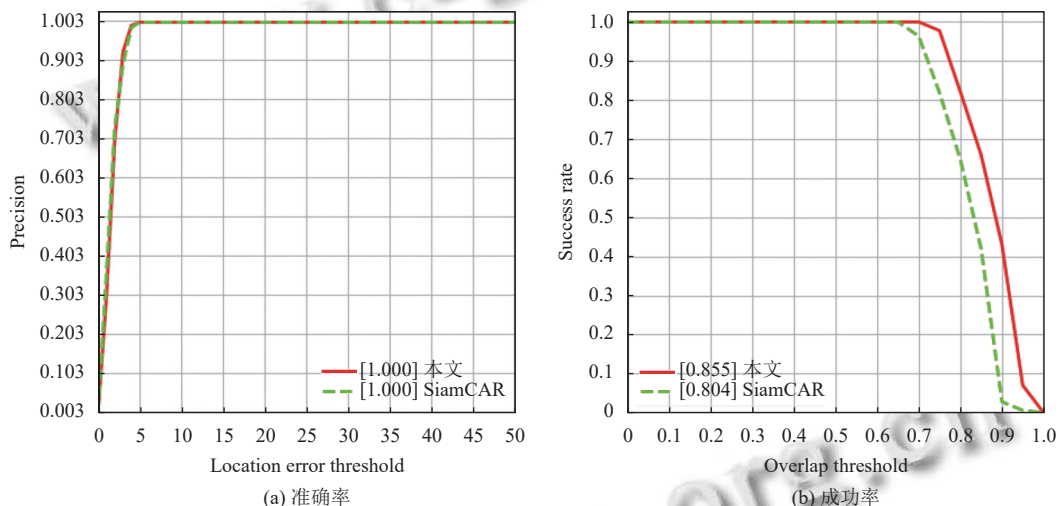


图6 plane_02 序列中算法准确率和成功率对比图

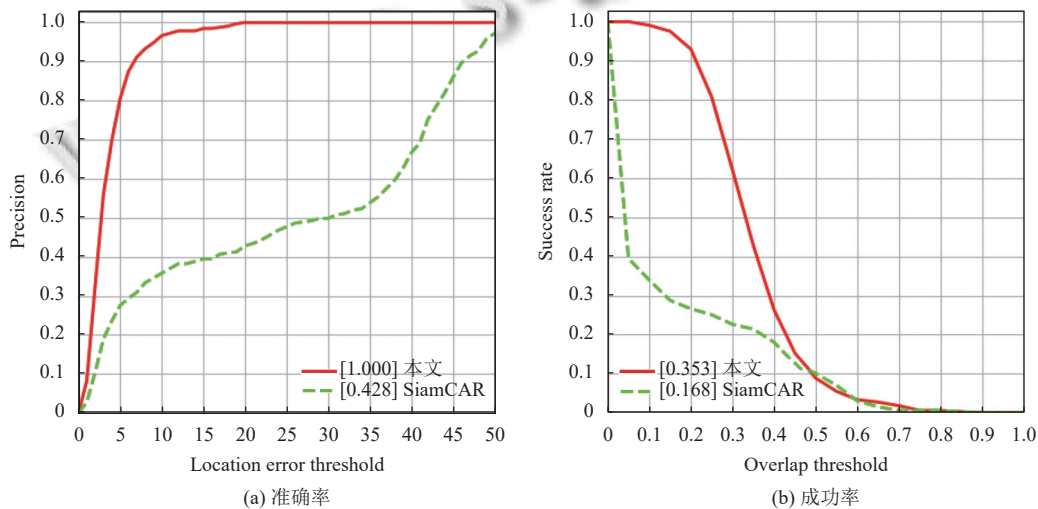


图7 ship_01 序列中算法准确率和成功率对比图

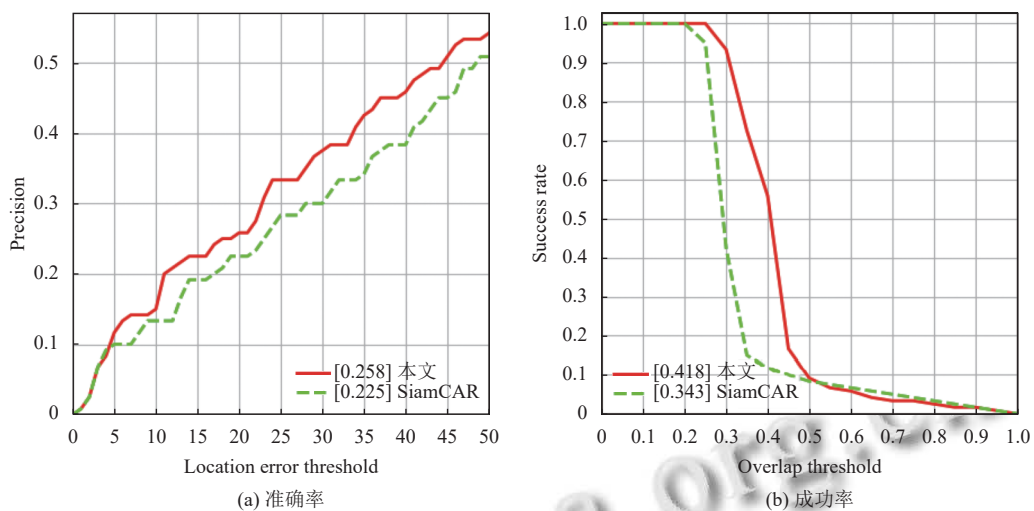


图8 train_03序列中算法准确率和成功率对比图



图9 两种算法在 plane_04 序列中的跟踪结果



图10 两种算法在 ship_01 序列中的跟踪结果



图11 两种算法在 train_03 序列中的跟踪结果

4 结论

本文主要介绍了改进后的卫星视频目标跟踪算法网络,针对卫星视频中存在的目标特征信息不明显、目标与背景对比性低等问题,将运动激励模块和改进的高效通道注意力模块加入到骨干网络的残差单元,激励目标的运动敏感通道和有用信息通道,增强目标的特征提取信息;并增加了相邻帧模板和卡尔曼预测帧模板与首帧模板和搜索区域形成四重网络结构,补充目标的运动信息和外观信息.实验表明,改进后的算法网络有效地提高了目标跟踪的准确率和成功率.

参考文献

- 1 Bolme DS, Beveridge JR, Draper BA, *et al.* Visual object tracking using adaptive correlation filters. Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco: IEEE, 2010. 2544–2550. [doi: [10.1109/CVPR.2010.5539960](https://doi.org/10.1109/CVPR.2010.5539960)]
- 2 Henriques JF, Caseiro R, Martins P, *et al.* Exploiting the circulant structure of tracking-by-detection with kernels. Proceedings of the 12th European Conference on Computer Vision. Florence: Springer, 2012. 702–715.
- 3 Henriques JF, Caseiro R, Martins P, *et al.* High-speed tracking with kernelized correlation filters. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3): 583–596. [doi: [10.1109/TPAMI.2014.2345390](https://doi.org/10.1109/TPAMI.2014.2345390)]
- 4 Danelljan M, Khan FS, Felsberg M, *et al.* Adaptive color attributes for real-time visual tracking. Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014. 1090–1097. [doi: [10.1109/CVPR.2014.143](https://doi.org/10.1109/CVPR.2014.143)]
- 5 Bertinetto L, Valmadre J, Henriques JF, *et al.* Fully-convolutional Siamese networks for object tracking. Proceedings of the European Conference on Computer Vision. Amsterdam: Springer, 2016. 850–865.
- 6 Li B, Yan JJ, Wu W, *et al.* High performance visual tracking with Siamese region proposal network. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 8971–8980.
- 7 Li B, Wu W, Wang Q, *et al.* SiamRPN++: Evolution of Siamese visual tracking with very deep networks. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 16–20. [doi: [10.1109/CVPR.2019.00441](https://doi.org/10.1109/CVPR.2019.00441)]
- 8 Guo DY, Wang J, Cui Y, *et al.* SiamCAR: Siamese fully convolutional classification and regression for visual tracking. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 6269–6277. [doi: [10.1109/CVPR42600.2020.00630](https://doi.org/10.1109/CVPR42600.2020.00630)]
- 9 Shi T, Wang DH, Ren HG. Triplet network template for Siamese trackers. IEEE Access, 2021, 9: 44426–44435. [doi: [10.1109/ACCESS.2021.3066294](https://doi.org/10.1109/ACCESS.2021.3066294)]
- 10 Shao J, Du B, Wu C, *et al.* HRSiam: High-resolution Siamese network, towards space-borne satellite video tracking. IEEE Transactions on Image Processing, 2021, 30: 3056–3068. [doi: [10.1109/TIP.2020.3045634](https://doi.org/10.1109/TIP.2020.3045634)]
- 11 Zhang WH, Jiao LC, Liu F, *et al.* MBLT: Learning motion and background for vehicle tracking in satellite videos. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 4703315. [doi: [10.1109/TGRS.2021.3109028](https://doi.org/10.1109/TGRS.2021.3109028)]
- 12 Wu JL, Su X, Yuan QQ, *et al.* Multivehicle object tracking in satellite video enhanced by slow features and motion features. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 5616426. [doi: [10.1109/TGRS.2021.3139121](https://doi.org/10.1109/TGRS.2021.3139121)]
- 13 Ruan L, Guo YJ, Yang DQ, *et al.* Deep Siamese network with motion fitting for object tracking in satellite videos. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 6508005. [doi: [10.1109/LGRS.2022.3158652](https://doi.org/10.1109/LGRS.2022.3158652)]
- 14 Song W, Jiao LC, Liu F, *et al.* A joint Siamese attention-aware network for vehicle object tracking in satellite videos. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5625617. [doi: [10.1109/TGRS.2022.3184755](https://doi.org/10.1109/TGRS.2022.3184755)]
- 15 He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 770–778. [doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90)]
- 16 Li Y, Ji B, Shi XT, *et al.* TEA: Temporal excitation and aggregation for action recognition. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 909–918. [doi: [10.1109/CVPR42600.2020.00099](https://doi.org/10.1109/CVPR42600.2020.00099)]
- 17 Wang QL, Wu BG, Zhu PF, *et al.* ECA-Net: Efficient channel attention for deep convolutional neural networks. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 11531–11539. [doi: [10.1109/CVPR42600.2020.01155](https://doi.org/10.1109/CVPR42600.2020.01155)]
- 18 Kalman RE. A new approach to linear filtering and prediction problems. Journal of Fluids Engineering, 1960, 82(1): 35–45.
- 19 Zhao MQ, Li SY, Xuan SY, *et al.* SatSOT: A benchmark dataset for satellite video single object tracking. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5617611. [doi: [10.1109/TGRS.2022.3140809](https://doi.org/10.1109/TGRS.2022.3140809)]

(校对责编:孙君艳)