

结合掩膜与孪生网络的目标跟踪方法研究

石胜斌,王曙光,刘 桢,杨传栋
(陆军炮兵防空兵学院,安徽合肥 230031)

摘要:针对在舰船目标跟踪过程中由于视场角变化导致的跟踪精度下降的问题,基于全连接孪生神经网络的目标跟踪方法,融合了目标分割策略,提出了基于 SiamMask 的实时目标跟踪算法。该算法将二元掩膜运算作为孪生网络的一个分支,以实现对目标的分割,在获取目标位置信息的同时,获取目标的外观信息,使得该网络的损失函数得到显著增强。由传统的轴对称目标跟踪框,改进为可根据目标形状、外观而自适应调整的可旋转矩形框。将该算法与传统基于孪生网络的跟踪算法 SiameseNet,以及基于相关滤波的 KCF 跟踪算法进行实验仿真测评对比,通过计算各算法的中心误差,结果表明该算法的跟踪精度较两者分别提升了 19.5% 和 24.5%,且运行速度可达 30 fps,满足了舰船目标跟踪对于准确性和实时性的要求。

关键词:舰船目标;目标跟踪;孪生网络;目标分割;中心误差

中图分类号:TP391.4

文献标识码:A

文章编号:1673-629X(2020)05-0061-05

doi:10.3969/j.issn.1673-629X.2020.05.012

Research on Target Tracking Method Based on Mask and Siamese Network

SHI Sheng-bin, WANG Shu-guang, LIU Zhen, YANG Chuan-dong
(Army Artillery and Air Defense Officer Academy of PLA, Hefei 230031, China)

Abstract: Aiming at the problem that the tracking accuracy decreases due to the change of the field of view during the ship target tracking, a real-time target tracking algorithm based on SiamMask is proposed, which combines the target segmentation strategy with the method of full-connected Siamese neural network. In this algorithm, binary mask operation is taken as a branch of Siamese network to realize the segmentation of the target. While acquiring the location information of the target, the appearance information of the target is obtained, so that the loss function of the network is significantly enhanced. The traditional axis-symmetrical target tracking frame is improved to a rotatable rectangular frame which can be adaptively adjusted according to the shape and appearance of the target. Compared with traditional Siamese-network-based tracking algorithm SiameseNet and KCF-based tracking algorithm based on correlation filtering, the simulation results show that the tracking accuracy of the proposed algorithm is improved by 19.5% and 24.5% respectively, and the running speed can reach 30 fps, which meets the requirement of accuracy and real-time for ship target tracking.

Key words: ship target; target tracking; Siamese network; target segmentation; center error

0 引言

中国的领海面积十分广阔,海洋资源丰富;与此同时,海洋周边环境复杂,存在一定的安全隐患,为了维护海洋安全稳定、准确锁定及跟踪可疑海上目标,舰船目标跟踪系统的搭载势在必行,因此对于海上目标跟踪技术的研究具有一定的现实意义^[1]。

目标跟踪在计算机视觉领域仍然是极富挑战性的一个子领域。很多国家都投入了大量的人力、物力和财力进行该项研究。当前,基于计算机视觉的跟踪具

有非常广泛的应用价值,国内外学者已提出了很多运动目标跟踪的方法,并在不同的应用领域取得了一定的效果。在交通应用领域,主要有交通流量控制,车辆、船舶的异常检测,利用智能交通监控系统提高交通的管控系统效率^[2-3]。在精确制导领域,随着目标跟踪技术的不断发展,制导精度也大幅提升。以往的跟踪算法通常只能用简单的轴对齐的边界框来表示目标,这种方法固然存在一定的弊端,如目标发生外观及形态变化时,传统轴对齐的跟踪框中势必会引入大量

收稿日期:2019-07-14

修回日期:2019-11-19

网络出版时间:2020-01-10

基金项目:军队“十二五”预研基金项目(301070103)

作者简介:石胜斌(1994-),男,研究生,通信作者,研究方向为目标检测与跟踪技术;王曙光,教授,博导,研究方向为智能弹药技术。

网络出版地址:<http://kns.cnki.net/kcms/detail/61.1450.TP.20200110.1123.054.html>

背景信息,例如在导弹打击目标时极有可能造成任务失败。文中基于全卷积孪生网络 (Siamese) 框架^[4]并结合目标分割策略,使得跟踪器能够在目标发生变化时,自适应地调整跟踪框的形状,该方法能够有效减小跟踪器的跟踪误差。

1 基于 SiamMask 的目标跟踪方法

1.1 孪生网络简介

孪生神经网络 (SiameseNet: Siamese neural network)最早是由 Bromley 和 LeCun^[5]提出用来处理签名验证的问题,该网络主要包含两个或多个具有相同结构且共享网络参数的网络模型,并经过不断的发展,逐渐被应用到图像匹配、分类等领域。孪生神经网络

的目标是寻找两个可比较对象的相似程度(例如,签名验证、人脸和指纹识别等)。

在目标跟踪中,该网络有两个相同的子网络,并且这两个子网络有相同的参数和权重^[6]。孪生网络中间隐含层为传统的卷积神经网络,不同的是一个孪生网络通常包含两个甚至多个输入,每个网络分支的参数共享,通过对输出相似度对比函数的学习,进而比较各分支输入样本的相似度,最终得到的是一个得分图,进而确定目标位置。其跟踪框架如图 1 所示。相比于传统卷积神经网络的单输入特性,多输入的孪生网络能代入一定的先验信息,即能够将更多的特征信息代入跟踪器,使得训练模型的小型化得以实现。

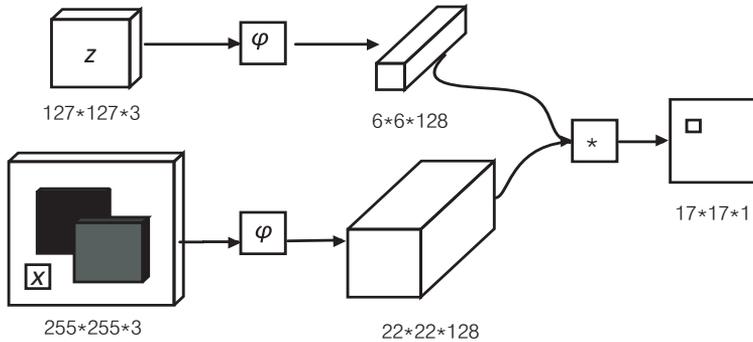


图 1 基于孪生网络目标跟踪框架

图中 * 代表相关性计算, φ 代表简化的卷积神经网络,结构与 AlexNet 类似^[7],不过只有卷积层与池化

层,没有最后的全连接层,其结构如表 1 所示。

表 1 网络结构

层级	卷积核	个数 * 层级	步长	模板图像	搜索图像
				127x127	255x255
conv1	11x11	96x3	2	59x59	123x123
pool1	3x3		2	29x29	61x61
conv2	5x5	256x48	1	25x25	57x57
pool2	3x3		2	12x12	28x28
conv3	3x3	384x256	1	10x10	26x26
conv4	3x3	384x192	1	8x8	24x24
conv5	3x3	256x192	1	6x6	22x22

其中池化层采用 max-pooling 的方法,除第五层外,每个卷积层后都连有非线性激活层,同时在训练时对每个激活层进行批标准化,以降低训练过程中的过拟合^[8-9]。该算法虽能较为准确地确定目标位置,但由于采用的是轴对齐跟踪框,在目标视角发生变化则会引入一定的背景信息。

1.2 基于 SiamMask 的跟踪算法框架

文中采用实时在线目标分割与目标跟踪统一架构 (SiamMask)^[10]来解决因目标旋转变换所带来的跟踪

精度下降的问题。SiamMask 算法基于全卷积孪生网络架构,通过在原有损失函数的基础上增加对目标分割损失的计算,即对传统孪生网络的离线训练过程进行改进。该算法的基本框架^[11]如图 2 所示。

与传统孪生网络不同的是,该算法中 * d 是对图像神经网络得到的特征进行逐通道的相关性计算,所以保证了响应的通道数不变^[12],得到的响应值称为 RoW (response of candidate window),然后在该响应的基础上得到了分割 mask 和分类 score 两个分支。

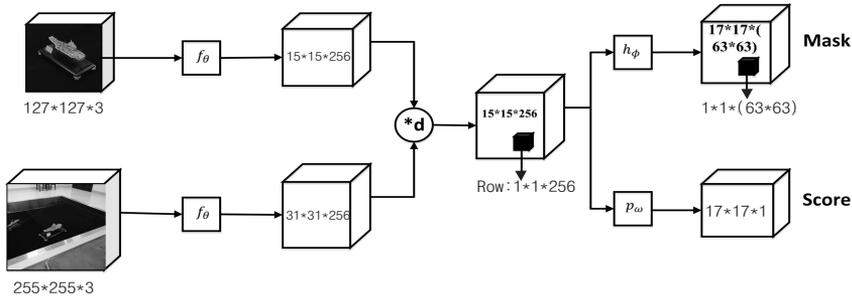


图2 SiamMask 算法目标跟踪框架

由于每个 Row 最终生成的 mask 是一个 63×63 的向量,因此最终得到的分割结果的图像较原图的尺寸会小很多,因此该算法在输出结果的基础上加入了

如图3所示的上采样过程,最终得到了 127×127 的 mask 图像,进而取得了更加精细的分割结果。

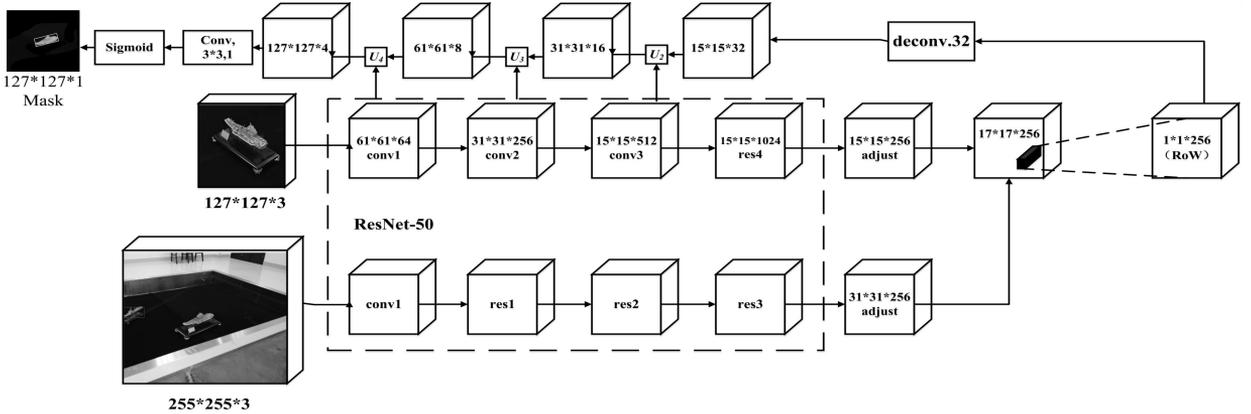


图3 mask 图像上采样

该方法使用具有可学习参数 φ 的简单的两层神经网络 h_φ 对每个 RoW 预测 $w \times h$ 二元掩模 (Mask)^[13]。令 m_n 表示对应于第 n 个 RoW 的预测 mask,则可表示为:

$$m_n = h_\varphi(g_\theta^n(z, x)) \quad (1)$$

$$g_\theta(z, x) = f_\theta(z) * f_\theta(x) \quad (2)$$

其中, x 和 z 分别表示待搜索区域以及目标模板, $g_\theta(z, x)$ 为响应值,经计算响应值最大的部分即为预测的目标位置。

由上式可知,最后生成的 mask 是待分割图像 x 和目标图像 z 的一个函数,因此可以看成是用 z 去指导完成 x 上的分割。而最终跟踪结果的边界框则根据分割所得结果的最小外接矩形绘制,故可以根据目标的外观做出最佳预测。

1.3 模型训练

对于模型的训练采取端到端的方式,两个分支同时进行训练,即根据每一对训练样本都应给出每个分支的标签,分割对应的损失函数^[14]为:

$$L_{\text{mask}}(\theta, \varphi) = \sum_n \left[\frac{1 + y_n}{2wh} \sum_{i,j} \log(1 + e^{-c_{ij}^{m_n}}) \right] \quad (3)$$

式3表明模型对于分割损失函数的计算采用二元逻辑回归损失的方式,每个 RoW 利用标准值的二进制标签 $y_n \in \{ \pm 1 \}$ 进行标记,即当 $y_n = 1$ 时表示该样本

为正样本;同时结果还与尺寸为 $w \times h$ 的像素级 Mask 相关, $c_{ij}^n \in \{ \pm 1 \}$ 为第 n 个候选区域 (i, j) 像素对应物体的 Mask。

分类分支对应损失函数为:

$$L_{\text{score}} = \log(1 + \exp(-y_n v)) \quad (4)$$

即采用逻辑损失函数的方式, v 为图1中最终输出得分图中每点的真实值, $y_n = \pm 1$ 为该点对应标签。而得分图中整体损失函数为所有点损失函数的均值:

$$l_{\text{score}} = \frac{1}{|D|} \sum_{u \in D} l(y[u], v[u]) \quad (5)$$

其中, $u \in D$ 表示该点在 score map 中的位置,最终可以得到该模型整体的损失函数:

$$L_{2B} = \lambda_1 \cdot L_{\text{mask}} + \lambda_2 \cdot L_{\text{score}} \quad (6)$$

其中, L_{2B} 表示模型整体的损失函数, L_{score} 为分类的损失函数。该算法令 $\lambda_1 = 32$, $\lambda_2 = 1$,如果预测与真实的回归框有至少 0.6IOU (intersection-over-union, 交并比) 则是正例, 否则是负例^[15]。

网络训练所用数据集来自 ILSVRC (ImageNet large scale visual recognition challenge) 竞赛中用于视频目标检测的 ImageNet 视频集^[16], 这个数据集共有 4 500 个视频,其中也包含大量的舰船目标的视频,通过对这些视频的训练能够有效提升跟踪器中孪生网络的判别能力。

2 实验与分析

2.1 实验环境

实验环境如表 2 所示。

表 2 实验环境

名称	配置
CPU	Core i7-8700
GPU	NVIDIA GTX1070Ti
操作系统	Windows 10
内存	16 G
图像处理	Matlab2015b、OpenCV3.2.1

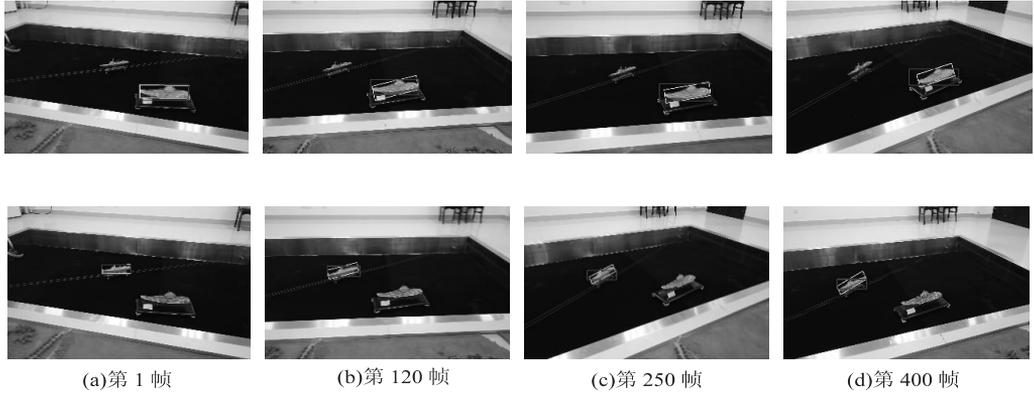


图 4 舰船模型跟踪效果对比

为定量分析算法的实验效果,传统评估跟踪效果方法计算预测框与标准框的重叠率,但由于文中算法与传统算法的标注方式不同,即采用非轴对称框的标注方式,使得算法的重叠率指标不具有可比性。因此通过计算对比 SiamMask 算法、SiameseNet 算法以及经典检测跟踪算法 KCF^[18]的中心误差,并绘制中心误差曲线^[19]来反映算法的跟踪精度。

所谓中心误差即算法对于目标中心像素的预测值与标准值的偏差。对采集的十组包含舰船模型的视频进行逐帧标注,将各算法在这十个序列的测试所得中心误差的平均值绘制成曲线,如图 5 所示,曲线上的点代表跟踪预测结果小于误差阈值的帧数占视频总帧数的百分比。

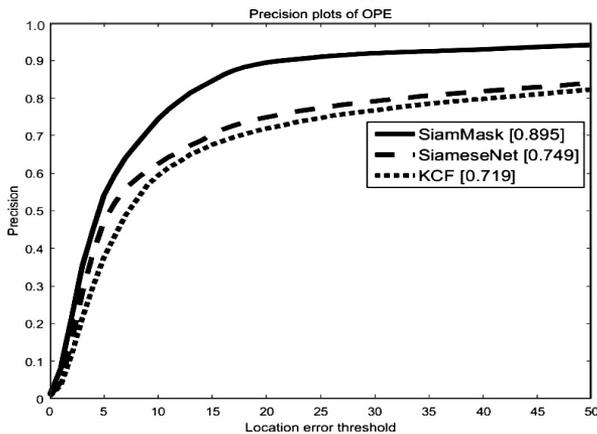


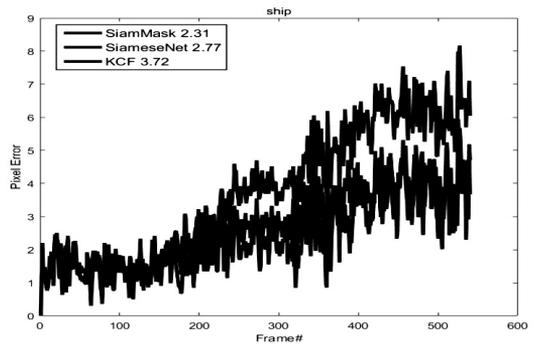
图 5 平均中心误差

2.2 对比实验

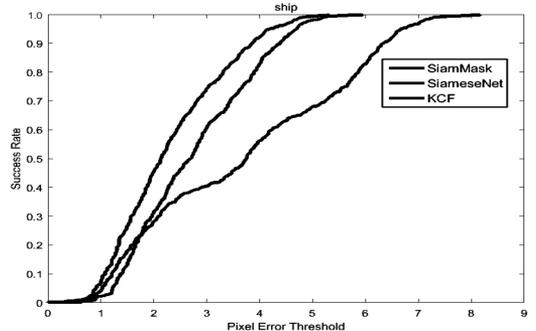
分别采用传统基于孪生网络 (SiameseNet) 跟踪算法^[17]以及文中的结合分割策略的 (SiamMask) 跟踪算法在同一图像测试集上进行实验,以舰船模型为跟踪目标进行跟踪效果对比,结果如图 4 所示。

可以看出传统跟踪算法的跟踪框 (图中正矩形) 在包含目标的同时不可避免地引入了大量背景信息,而文中算法能够根据分割结果自适应地调整跟踪框 (图中斜矩形) 的角度与位置,充分减少了背景信息的干扰。

由实验结果可知,SiamMask 算法的跟踪效果更好,其平均跟踪成功率较 SiameseNet 以及 KCF 算法分别提升了 19.5% 和 24.5%。选择了其中舰船目标旋转变化较明显的一个视频序列,对每一帧的中心误差分析,如图 6 所示。



(a) 中心误差



(b) 成功率

图 6 单一视频中心误差曲线

可以看出,当目标视角发生变化时,文中算法较其他算法的中心误差增量更小,跟踪成功率更高。与此同时,算法在 GPU 加速的情况下其运行速度可达 30 fps,可以较好地满足实时性要求。

3 结束语

传统跟踪方法在目标发生外观变化,尤其是旋转变化时,其跟踪精度明显下降。而 SiamMask 算法从目标的 Mask 获得旋转边界框的策略比简单地提供轴对齐边界框有显著的优势。通过对比 SiamMask 算法、SiameseNet 以及 KCF 算法在标注好的包含舰船目标的图像序列上的测试结果,进一步验证了该算法的突出性能,能够较好地满足舰船目标跟踪的准确性和实时性,具有较高的应用价值。

参考文献:

- [1] 蔡丹. 水面运动目标检测与跟踪算法研究[D]. 南京:江苏科技大学,2017.
- [2] 尹宏鹏,陈波,柴毅,等. 基于视觉的目标检测与跟踪综述[J]. 自动化学报,2016,42(10):1466-1489.
- [3] 张微,康宝生. 相关滤波目标跟踪进展综述[J]. 中国图象图形学报,2017,22(8):1017-1033.
- [4] BERTINETTO L, VALMADRE J, HENRIQUES J F, et al. Fully-convolutional Siamese networks for object tracking [C]//European conference on computer vision. Netherlands, Amsterdam: Springer, 2016:850-865.
- [5] BROMLEY J, BENTZ J W, BOTTOU L, et al. Signature verification using a "Siamese" time delay neural network [J]. International Journal of Pattern Recognition and Artificial Intelligence, 1993, 7(4):669-688.
- [6] 史璐璐,张索非,吴晓富. 基于 Tiny Darknet 全卷积孪生网络的目标跟踪[J]. 南京邮电大学学报:自然科学版,2018, 38(4):89-95.
- [7] 闫若怡,熊丹,于清华,等. 基于并行跟踪检测框架与深度学习的目标跟踪算法[J]. 计算机应用,2019,39(2):343-347.
- [8] 闵召阳,赵文杰. 基于深度学习的目标抗干扰跟踪算法[J]. 红外技术,2018,40(2):176-182.
- [9] 贾静平,覃亦华. 基于深度学习的视觉跟踪算法研究综述[J]. 计算机科学,2017,44(S1):19-23.
- [10] MA C, HUANG J B, YANG X, et al. Adaptive correlation filters with long-term and short-term memory for object tracking[J]. International Journal of Computer Vision, 2018, 126(8):771-796.
- [11] HE A, LUO C, TIAN X, et al. A twofold Siamese network for real-time object tracking [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. [s. l.]: IEEE, 2018:4834-4843.
- [12] 曹诗雨,刘跃虎,李辛昭. 基于 Fast R-CNN 的车辆目标检测[J]. 中国图象图形学报,2017,22(5):671-677.
- [13] 王琳,刘强. 基于局部特征的多目标图像分割算法[J]. 激光与光电子学进展,2018,55(6):103-110.
- [14] FANG Y, KO S, JO G S. Multi-scale region proposal network trained by multi-domain learning for visual object tracking [C]//International conference on neural information processing. Guangzhou: Springer, 2017:330-339.
- [15] RAHMAN M A, YANG W. Optimizing intersection-over-union in deep neural networks for image segmentation [C]//International symposium on visual computing. Las Vegas, NV, USA: Springer, 2016:234-244.
- [16] 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.
- [17] TAO R, GAVVES E, SMEULDERS A W, et al. Siamese instance search for tracking [C]//Computer vision and pattern recognition. Las Vegas: [s. n.], 2016:1420-1429.
- [18] HENRIQUES J F, CASEIRO R, MARTINS P, et al. High-speed tracking with kernelized correlation filters [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3):583-596.
- [19] WU Y, LIM J, YANG M H. Online object tracking: a benchmark [C]//IEEE conference on computer vision and pattern recognition. Portland, OR, USA: IEEE, 2013:2411-2418.