

〈图像处理与仿真〉

## 机载平台下基于深度检测网络的目标跟踪重捕算法

沈旭<sup>1</sup>, 孟巍<sup>2</sup>, 程小辉<sup>3</sup>, 王新政<sup>3</sup>

(1. 岭南师范学院 信息工程学院, 广东 湛江 524048; 2. 山东电力科学研究院, 山东 济南 250012;  
3. 桂林理工大学 信息科学与工程学院, 广西 桂林 541004)

**摘要:** 目标检测与跟踪是机载光电设备至关重要的功能模块, 其检测跟踪的性能直接关系到目标感知的精度。近年来基于 Siamese 网络的改进跟踪算法在各种挑战性的数据集上取得了优异的效果, 但大多数改进算法采用局部搜索策略, 无法更新模板, 且模板会引入背景干扰, 最终因跟踪点漂移导致跟踪失败。为了解决这些问题, 本文提出了一种结合目标边缘检测的改进全连接 Siamese 跟踪算法, 该算法利用目标的轮廓模板代替边界框模板, 减少了背景杂波的干扰; 同时, 在 Siamese 网络的基础上增加了一路改进 tiny-YOLOv3 目标检测网络, 利用 K 均值聚类找到最合适的锚框 (anchor box), 引入了扩张模块层来扩展感受野, 增加了系统的抗遮挡能力, 提高机载光电设备的目标捕获概率。在基准测试数据集以及挂飞数据集基础上的仿真测试性能表明本文提出的改进模型特别适合机载光电设备在跟踪与重捕复杂环境下的运动目标, 在长期跟踪中能够更好地适应目标的变形和遮挡, 提升系统响应时间与适应性。

**关键词:** 目标跟踪; 深度学习; Siamese 网络; 轮廓模板; 目标检测

中图分类号: TN219, TP181, TP391 文献标识码: A 文章编号: 1001-8891(2020)07-0624-08

## Object Tracking and Recapture Model Based on Deep Detection Network Under Airborne Platform

SHEN Xu<sup>1</sup>, MENG Wei<sup>2</sup>, CHENG Xiaohui<sup>3</sup>, WANG Xinzheng<sup>3</sup>

(1. School of Information Engineering, Lingnan Normal University, Zhanjiang 524048, China;

2. Shandong Electric Power Research Institute, Jinan 250012, China;

3. College of Information Science and Engineering, Guilin University of Technology, Guilin 541004, China)

**Abstract:** Object detection and tracking is an essential module in airborne optoelectronic equipment, and its performance is directly related to the accuracy of object perception. Improved Siamese network tracking algorithms have produced excellent results for various challenging datasets recently, but most of the improved algorithms use local fixed search strategies, which cannot update the template. In addition, the template will introduce background interference, which will result in tracking drift and eventually cause tracking failure. To solve these problems, this paper proposes an improved fully connected Siamese tracking algorithm combined with object contour extraction and object detection; the algorithm uses the contour template of the target instead of the bounding box template to reduce the background clutter interference. A branch is added to the Siamese network to improve the tiny-YOLOv3 object detection network, where K-means clustering is used to find the most suitable anchor box. An expansion module layer is introduced to expand the receptive field. Therefore, our proposed model increases the anti-occlusion ability of the system and improves the object recapture probability of airborne optoelectronic equipment. The results of a simulation of benchmark test data set and a flight dataset show that the improved model is especially suitable for tracking and recapture of moving objects in complex environments; in addition, it can better adapt to

收稿日期: 2019-08-27; 修订日期: 2020-07-09.

作者简介: 沈旭 (1979-), 男, 硕士, 讲师, 主要研究方向图像处理、智能控制应用、模式识别等。

基金项目: 国家自然科学基金 (61662017, 61402399); 湛江市科技发展专项资金竞争性分配项目 (2019A01042); 岭南师范学院教育教学改革项目 (LSJGMS1811)。

deformed or occluded objects in long-term tracking, which improves the system response time and adaptability.

**Key words:** object tracking, deep learning, siamese network, contour extraction network, object detection

## 0 引言

无人机具有高度的灵活性和可操作性,广泛应用于视频监控、态势感知、应急响应等领域<sup>[1]</sup>。目标检测与跟踪是机载光电设备至关重要的功能模块,其检测跟踪的性能直接关系到目标感知的精度<sup>[2]</sup>。

目标跟踪是对视频序列的分析,其过程是在第一帧中选择目标,然后通过跟踪算法逐帧从序列中找到与所选目标相同的区域作为跟踪目标<sup>[3]</sup>。尽管机载目标跟踪算法已经取得了优异的跟踪效果,尤其是彩虹系列无人机可以实现地面大尺寸目标的稳定跟踪,但由于机载平台的抖动、翻滚和视角突变、背景杂波、目标旋转、快速运动和变形,这使得机载目标跟踪仍然是一项艰巨而富有挑战性的任务。现有的目标跟踪算法可以分为基于人工低层次特征的目标跟踪与基于深度高层次特征的目标跟踪,前者是通过手工设计的算子提取低层次特征;后者则是用深度网络去学习目标的高层次特征,但最终都是采用生成(Generative)模型或判别(Discriminative)模型实现跟踪<sup>[4]</sup>。虽然深度模型获取的深度特征提高了目标的表征能力,但同时也增加了算法的复杂度。

为了实现复杂场景下机载跟踪算法的精度与速度的平衡,国内外专家提出了许多有效的算法。基于判别模型的核相关滤波器<sup>[5]</sup>是目前速度与精度都非常高的算法,该算法利用初始样本循环移位构建循环矩阵,并采用岭回归获取傅里叶域最优鲁棒滤波器,实现目标跟踪。然而,循环移位获取的样本存在边界效应,产生无效负样本,影响跟踪精度。

随着深度学习在识别、分类等信号处理领域的不断渗透,一些学者开始尝试将深度学习应用于目标跟踪领域,DLT(Deep Learning Tracker)<sup>[6]</sup>是第一个将深度模型应用于单目标跟踪的跟踪算法,该方法是在粒子滤波框架下引入稀疏自编码模型作为测量分类器,需要海量的数据进行离线训练与在线微调。FCNT(Fully Convolutional Networks for Tracking)<sup>[7]</sup>和HCFT(Hierarchical Convolutional Features for Tracking)<sup>[8]</sup>则是利用VGG-Net在大规模分类数据集上训练模型,获取目标的特征表示,并将观测模型用于分类以获得跟踪结果,虽然这些方法避免了在线训练海量样本的不足,充分利用了深度特征的强大表征能力。但由于分类任务更加关注目标与背景之间的差

异,因此这种经过预训练的网络会忽略类内的差异,造成跟踪精度不如KCF(Kernelized Correlation Filter)、TLD(Tracking-Learning-Detection)等跟踪算法<sup>[9]</sup>,其跟踪速度也不高。Siamese跟踪网络<sup>[10]</sup>是目前深度模型中实时性较高的跟踪性算法,该网络提取搜索区域与模板的特征,输入到相似度量函数里计算相似度,实现目标跟踪。

虽然近年来基于Siamese网络的跟踪算法在各种挑战性的数据集上取得了优异的效果,但大多数Siamese跟踪算法采用局部搜索策略,无法更新模板,且模板会引入背景干扰,最终因跟踪点漂移而导致跟踪失败。为了解决这些问题,本文提出了一种结合目标边缘检测的改进全连接Siamese跟踪算法,该算法利用目标的轮廓模板代替边界框模板,减少了背景杂波的干扰;同时,在Siamese网络的基础上增加了一路目标检测分支,不仅能够获取目标的边界,还能增加抗遮挡能力,提高机载光电设备的目标捕获概率。

## 1 算法基础

### 1.1 HED 边缘检测网络

边缘检测是红外图像处理中一个重要的领域,尤其是桥梁、道路、跑道等典型目标大多是以目标的边缘作为先验信息进行感知。传统的边缘检测算法是利用微分算子对局部区域的灰度信息进行处理,通过邻域灰度的波动情况实现边缘的检测,但该方法容易受背景干扰。若目标内部纹理信息丰富则很难获取目标的边缘轮廓。随着深度学习为代表的机器学习算法的发展,已经可以提取到封闭的边缘轮廓<sup>[11]</sup>。HED(Holistically-Nested Edge Detection)算法<sup>[12]</sup>是目前边缘检测领域的最优算法,该方法是以VGG16网络作为基础网络,VGG16含有5个block,3个全连接层和1个softmax输出层,前两个block分别含有两个卷积层和一个池化层,后3个block分别含有3个卷积层和一个池化层,HED边缘检测网络去掉了VGG16网络的softmax层、全连接层和最后一个block的池化层,并将从每一个block的最后一个卷积层引出一个侧边输出层,最后将侧输出层融合得到最终的候选区域边缘图。HED网络在训练过程中采用了6个类均衡交叉熵损失函数,如式(1)所示:

$$l_{\text{side}}^{(m)}(W, w^{(m)}) = -\beta \sum_{j \in Y_+} \log P_r(y_j = 1 | X; W, w^{(m)}) + \\ (\beta - 1) \sum_{j \in Y_-} \log P_r(y_j = 0 | X; W, w^{(m)}) \quad (1)$$

式中： $Y_+$ 与 $Y_-$ 分别表示边缘与非边缘的标签集合； $\beta$ 是该集合中边缘标签样本的比例因子； $m$ 是网络分支数； $P_r$ 可以通过 Sigmoid 函数的激活值确定，即：

$$\log P_r(y_j = 1 | X; W, w^{(m)}) = \sigma(\alpha_j^{(m)}) \in [0, 1]$$

在每一个侧边输出层的边缘预测响应值 $Y_{\text{side}}^{(m)} = \sigma(A_{\text{side}}^{(m)})$ 可以通过逻辑回归得到，其中 $A_{\text{side}}^{(m)} = \{\alpha_j^{(m)}\}$ ；然后对所有侧边的损失函数进行加权融合，最终通过梯度下降法进行迭代优化，实现多尺度和多层次的端到端边缘轮廓特征的学习，该方法提升了网络的尺度适应性，解决了因网络层过深导致的梯度消失信息。

HED 网络使用经典的分类网络架构 VGG-16<sup>[13]</sup>作为基础网络，得到的边缘检测效果很大程度上已经超过了传统边缘检测算法。若能准确地提取目标的封闭轮廓，就可以获得目标的尺度特征与内部纹理信息。在红外目标跟踪过程中，由于红外图像对比度过低，目标特性难以捕捉。本文提出的跟踪算法采用 HED 边缘检测网络检测目标的边缘，获取目标的封闭外轮廓，从而确定目标的尺度，也能够对目标的位置进行进一步的修正，提高目标跟踪精度。

## 1.2 Siamese 全卷积跟踪网络

Siamese 网络也称孪生网络，是用于衡量输入样本相似性度量的网络架构，该网络由两个共享权值参数的相同结构网络组成。图 1 是 Siamese 全卷积跟踪网络的模型结构。Siamese 全卷积跟踪网络采用 Alex Net 作为基本网络，该网络可以对输入的模板 $z$ 与搜索区域 $x$ 进行深度特征编码<sup>[14]</sup>，然后对获得的深度特征进行交叉相关，最终获得响应矩阵，如式(2)所示：

$$f(x, z) = g(\phi(x), \phi(z)) \quad (2)$$

式中： $\phi$ 是去掉全连接层的 AlexNet 网络<sup>[14]</sup>； $g(\cdot)$ 是卷积操作； $f$ 为最终获取的相似性响应矩阵。为了获得准确的目标定位，对响应图 $f$ 进行非线性插值，得到与原图尺寸一样的响应图。最大的响应值为预测到的最终目标位置，最终实现目标跟踪。

Siam FC 网络结构包括两个相同的特征提取网络，每个特征提取网络为包含 7 个卷积层，2 个全连接层的深度卷积网络，该网络对特征提取网络中的卷积层和全连接层的数量并不做限制，但是要包含卷积层和全连接层。Siam FC 仅仅利用孪生网络获取了深度特征，其跟踪过程仍然采用的相关匹配选最优的方法。跟踪性能的提升主要是以下两方面：深度特征提

升了目标的表征能力，进一步增强了跟踪的精度；跟踪过程中不进行在线更新，加快了跟踪速度。

然而，跟踪过程不更新模板，意味着长时间跟踪过程中模板不会随着目标状态而自适应刷新。另一方面，Siam FC 获得的相似性响应图是由相似性矩阵上采样插值得到，存在较大的定位误差，且目标的尺寸很难确定。

## 2 改进 Siamese 跟踪网络模型

### 2.1 总体模型框架

由于 HED 边缘检测网络可以获取目标准确的封闭边界，只要选定目标就可以知道目标的轮廓，不仅可以提升系统反应时间，还能通过封闭轮廓提升目标的跟踪能力。当目标灰度突变、部分遮挡、姿态变化时，采用轮廓模板可以增强系统的抗干扰能力。因此，本文提出了一种结合目标边缘检测的改进全连接 Siamese 跟踪算法，该算法利用目标的轮廓模板代替边界框模板，减少了背景杂波的干扰；同时，在 Siamese 网络的基础上增加了一路目标检测分支，不仅能够获取目标的边界，还能增加抗遮挡能力，提高机载光电设备的目标捕获概率。

### 2.2 尺度自适应轮廓模板更新

模板是现有跟踪算法的基准，模板的准确性直接关系跟踪性能，错误的模板会降低场景匹配适应性，导致跟踪失败。模板更新可以分为固定帧更新、逐帧更新及自适应更新。机载设备跟踪过程目标存在姿态突变、遮挡、光照、运动模糊等干扰影像，其模板只有随着环境变化而刷新，才能获取稳定的跟踪性能。图 2 是车辆在不同环境下的跟踪结果，白色是逐帧刷新的相关跟踪算法<sup>[15-16]</sup>，黑色是模板不刷新的结果。可以看出，目标在机动运动过程中，由于姿态与运动模糊，目标的形状发生了较大变化，尤其是 46 帧时目标与路边的背景灰度类似，且目标的尺度逐渐变小。只有刷新模板才能更适应状态的变化。

理想的模板应该完整的包含目标，并尽可能降低引入背景信息。然而现有的模板都是以边界框的形式表征目标的模板，由于目标的不规则性，不可避免会引入背景信息。相比之下，轮廓特征清楚地表征了目标的尺度、形状与位置信息。若能用目标的轮廓模板代替传统模板，不仅能获取准确地目标尺度信息，还能有助于提高目标匹配跟踪精度。

### 2.3 轮廓提取网络

受 HED 侧边输出层结构的启发，本文提出采用一种具有特征重提取功能的边缘检测网络，如图 3 所

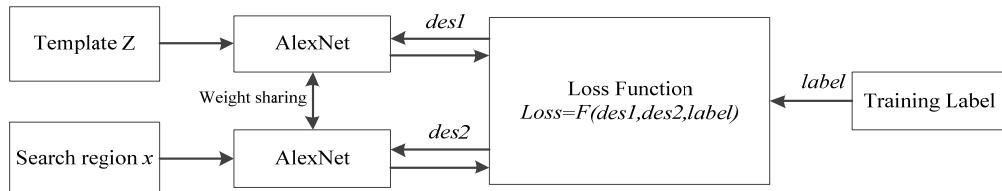


图1 Siam FC 全卷积跟踪网络

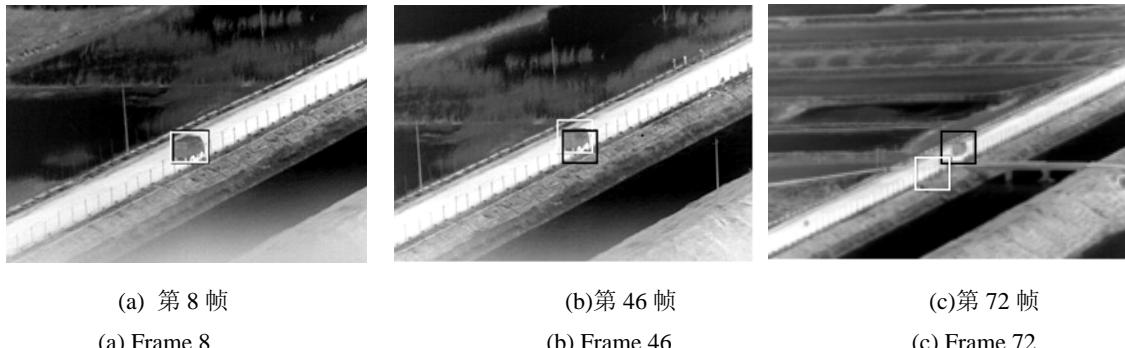
示, 该网络包含基本网络、侧边输出模块以及融合模块。

基本网络模块主要是提取深度特征, 常用的网络架构有 VGG16, VGG19<sup>[17]</sup>, ResNet-50<sup>[18]</sup>, ResNet-101<sup>[19]</sup>等。本文选用 VGG16 用作边缘提取基本网络。虽然 VGG19 和 ResNet 网络获得的深度特征表征能力更强, 但其网络深度与参数数量比 VGG16 大得多, 需要花费大量的时间进行网络训练与微调, 影响边缘检测的实时性。原始的 VGG16 包含 5 个卷积层, 每一层连接一个步长为 2 的池化层, 获得不同感受野的特征图。表 1 展示了每个卷积层的感受野的大小。在 VGG 网络中, 不同的卷基层包含不同的层次的深度特征信息。特别是在较深的层次上, HED 网络中的边缘输出更容易受到密集边问题的影响。例如, Stage 3

Fig.1 Siam FC tracking network

表达图像纹理特征, Stage 5 包含图像类别信息等。因此, 从每个层次提取的特征信息不能直接作为目标边缘轮廓, 需要对输出边界进行细化, 以生成清晰和准确的目标轮廓。为了生成边缘图像, 需要将各卷积层提取的不同层次信息映射到边缘像素空间。

侧边输出模块由特征重新提取模块和上采样层组成。在轮廓检测网络中, 特征重提取模块连接在基本网的 conv2\_2, conv3\_3, conv4\_3 和 conv5\_3 层之后, 可以将特征映射到边缘像素空间中。考虑到 GPU 显存的限制, conv1\_2 仅连接  $1 \times 1 \times 1$  卷积层以减少特征维度并实现特征融合。特征重提取模块由 3 个卷积层组成, 分别为  $1 \times 1 \times 32$ 、 $3 \times 3 \times 32$  和  $1 \times 1 \times 128$ 。在网络的末端, 特征融合模块使用卷积  $1 \times 1$  卷积核生成最终的边缘检测图像。



(a) 第 8 帧

(a) Frame 8

(b) 第 46 帧

(b) Frame 46

(c) 第 72 帧

(c) Frame 72

图2 目标尺度逐渐变小导致跟踪失败, 其中(a), (b)和(c)为不同帧下的跟踪结果

Fig.2 Tracking failure with scale change of object appearance, where (a), (b)and(c)are the results of different frames

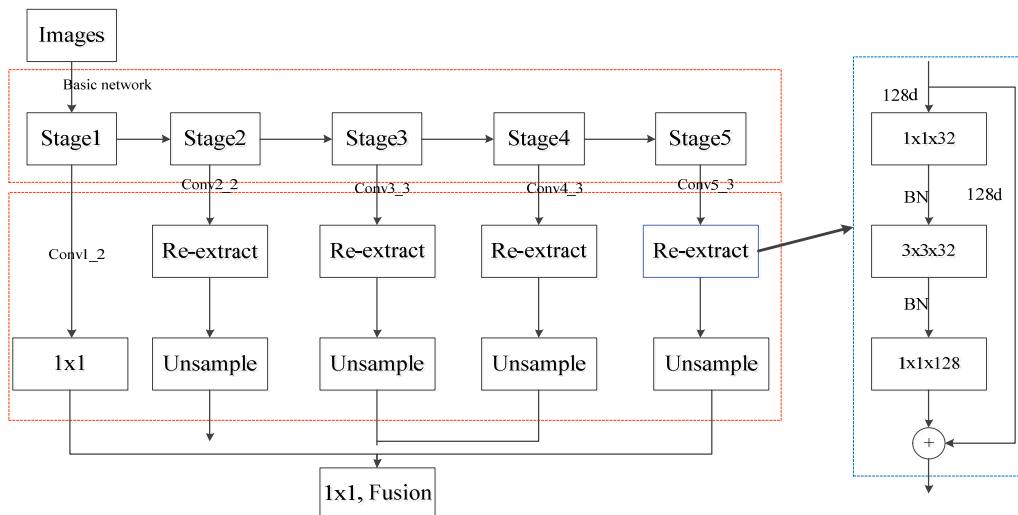


图3 轮廓提取网络

Fig.3 Contour-extraction networks

表1 轮廓检测网络中每层参数设置

Table 1 The parameter setting of each layer in contour extraction network.

Layer	S1-1	S2-2	S3-3	S4-3	S5-3
RF size	5	14	40	92	196
Stride	1	2	4	6	15

## 2.4 目标检测网络

机载光电设备在实时搜跟目标的过程中，阴影变化、目标重叠、遮挡、形变、旋转等干扰会直接影响跟踪的精度。尤其是地面运动目标在穿越树林，建筑物的过程中，目标会出现短暂丢失的现象。虽然大多数算法具备抗干扰能力，但如何重捕目标是光电跟踪系统研究的热点问题之一。目前大多数策略采用的全视场搜索检测，如相关配准、R-CNN、SSD 和 YOLO 等算法。YOLO 是近年来精度与效率最优的目标检测算法，该算法将目标检测视为回归问题，一次 CNN 计算就可以实现目标在端到端网络上的位置和分类，但对小目标的检测精度较差，不适应机载光电设备远距离跟踪弱小目标。为了提升光电平台的目标重捕能力与小目标的检测能力，本文在 tiny-YOLOv3 网络<sup>[20]</sup>的基础上提出了一种改进的目标检测网络，利用 K 均值聚类找到最合适的锚框（anchor box），在空洞卷积（dilated convolution）启发下引入了扩张模块层来扩展感受野，并增加了基本网络的深度，以便可以提取更多的语义信息，从而满足了机载空对地复杂环境下弱小目标检测与目标重捕的需求，其基本网络模型如图 4 所示。

传统的  $K$  均值聚类方法使用欧氏距离函数，这意味着较大的锚框比较小的锚框具有更多的误差，并且聚类结果可能会发生偏差。为此，本文采用 IOU 得分来评估聚类结果，从而避免了由框的大小引起的误差。距离函数可以通过以下公式计算：

$$d(\text{box}, \text{centriod}) = 1 - \text{IOU}(\text{box}, \text{centriod}) \quad (3)$$

式中：box 与 centriod 分别表示候选样本框与聚类中心，IOU 是候选框与原始标记框的重叠率。 $K$  均值聚类方法用于比较具有不同  $K$  值的 IOU 分数，考虑到模型的复杂性，本文最终选择  $K$  值为 6。

## 2.5 基于自适应轮廓模板的 Siamese 目标跟踪网络

机载平台逼近目标以及目标高机动不规则运行时，目标的尺度是时刻变化的，传统的 Siamese 跟踪网络很难进行长期稳定的目标跟踪。尤其是视场中存在多个类似目标的情形，跟踪过程中目标易被相似背景干扰。对于大多数改进的 Siamese 跟踪网络，如 GOTURN (Generic Object Tracking Using Regression Networks) 和 Siam FC，通过金字塔策略进行多尺度目标融合，可以提升跟踪精度，但也会损失小目标特征信息，并导致目标尺度降低，引入更多背景信息。本文提出了一种基于自适应轮廓模板 Siamese 跟踪网络。

现有的改进的 Siamese 跟踪网络通常选择以上一帧的跟踪结果作为搜索区域的中心，并从当前帧图像中提取一块固定大小的图像块作为搜索区域，若目标位于视场边缘则可能无法跟踪到最优区域，因此采用轮廓模板将最大限度降低干扰的影响，利用相关层作为融合张量，提高了融合张量的准确性。

针对滑窗搜索的不足，本文提出采用稀疏区域搜索，减少计算冗余；同时，根据卷积神经网络在不同的特征层的感受野不同，可以将不同层次的特征进行融合，实现多层次多尺度目标特征表示；为了获得更精确的位置估计，在网络末端预测置信度的基础上增加了位置回归模块，建立端对端的网络结构。

## 3 实验结果及其性能分析

### 3.1 实验设置及其性能评价指标

本文提出的改进 Siamese 跟踪网络采用 Python3.5 作为开发语言，Tensor Flow 作为深度学习框架。实验平台选用多核 Intel Xeon，最高频率达 2.4 GHz，以及多张 GTX 1080Ti GPU 显卡的服务器。离线训练数据是用于检测任务的 ILSVRC2015，测试视频来自当前目标跟踪中常用的 DARPA VIVID<sup>[15]</sup>，OTB 2015，VOT-TIR 2016，以及挂飞收集的自建数据库。为了便于性能分析，所有数据都进行逐帧尺度与位置标注，并按干扰状态进行分类。

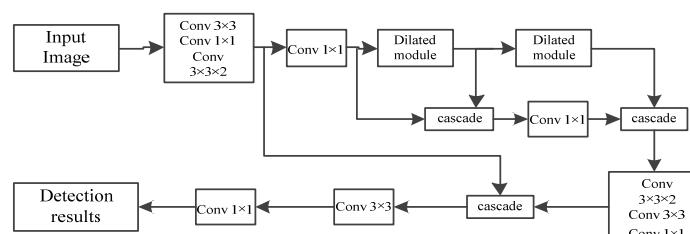


图 4 目标检测网络框架 Fig.4 Main framework for object detection model

网络的基础骨干网包括 5 个从 Alexnet 网络派生的卷积层。值得注意的是, Alexnet 网络包括 5 个卷积层, 其中前两个卷积层后面是一个最大池化层。除第五卷积层之外, 每个卷积层后面都有一个 ReLU 层。在训练期间, 在每个 ReLU 层之前使用批归一化以减少过度拟合。

中心误差 (Center Error, CE) 与重叠率 (Overlap Rate, OR) 是衡量跟踪性能的主要定量指标, 前者是表征跟踪点与基准点之间的差异, 误差越小, 表明跟踪点越接近理想状态, 但不能衡量跟踪目标的尺度变化; 后者则是衡量跟踪区域与基准区域之间的重叠率。由于不同阈值下, 定量分析存在较大差异。因此, 本文采用精确度图和成功率图定量分析算法性能, 其中前者表示平均中心误差在测试序列上的整体跟踪性能, 后者表征重叠率大于给定的阈值下的帧数比例<sup>[21]</sup>。本文选用位置误差为 20, 重叠率为 0.5 作为跟踪成功阈值。

### 3.2 参数设置

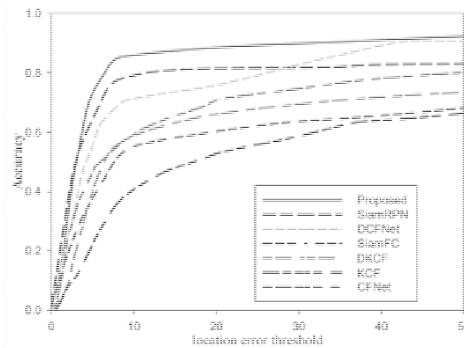
本文提出的改进跟踪网络分为 3 部分: 轮廓检测网络, 目标检测网络与 Siamese 跟踪网络, 各网络模型的基本网络参数与原模型保持一致, 其中学习率  $\eta$  设定为 0.001。Siamese 网络在离线训练期间, 所有卷积层都将更新。进行在线更新后, 浅层卷积层参数保持固定, 最后两个卷积层将根据测试数据进行微调。采用峰值旁瓣比选用最优跟踪点, 归一化响应的阈值设置为 0.65。若回归响应分数小于 0.65, 则认为跟踪失败, 则利用改进的 tiny-Yolov3 实时执行目标检测网络, 重捕最优目标。

### 3.3 定性定量分析

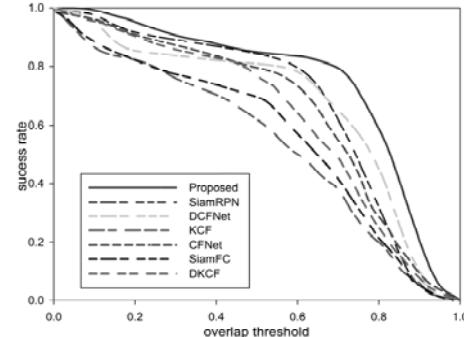
为了定性定量分析本文提出的改进 Siamese 深度模型的目标跟踪算法的应用性能, Siam FC, DCF

Net<sup>[22]</sup>, CF Net<sup>[23]</sup>, Siam RPN<sup>[24]</sup>, KCF 和 DKCF<sup>[25]</sup>, 其中前 4 种对比算法是基于 Siamese 的改进算法, KCF 是基于判别模型的核相关算法; DKCF 是将核相关滤波嵌入到深度网络中, 以端到端的方式学习目标的深度特征, 并进行目标跟踪。由于选用的基准数据集已经标注了目标大小与位置, 为了便于对比分析, 本文提出的算法跟踪点由第一帧的基准点给定, 但其目标大小由轮廓检测模型自适应计算。图 5 展示了不同算法在基准数据集上的平均准确度图和成功率图。表 2 展示了位置误差为 20, 重叠率为 0.5 下不同算法的定量指标。

从图 5(a)可以看出, 本文算法可以大部分情况下超过其他基准算法, 其精度为 0.8421, AUC 为 0.8631, 与 Siam FC 相比分别提高了 29.4% 和 11.9%。由于 Siam FC 是在第一帧中标记目标, 获取目标, 在跟踪期间不再更新, 因此不能适应目标变化, 抗干扰能力较弱。本文提出方法的性能仅次于 Siam RPN, 在目标快速运动, 背景干扰, 旋转, 遮挡, 光照变化等挑战序列中表现出色。Siam RPN 在 Siamese 网络之后添加了额外的区域提取网络来细化边界框, 而本文算法在 Siamese 网络基础上增加了一路轮廓检测网络, 用于获取目标的轮廓信息, 并利用轮廓模板降低干扰因素的影响。值得注意的是, 本文提出的算法中有一路目标检测网络, 用于跟踪丢失后目标重捕。实验结果可以看出, 一旦目标从穿出遮挡物, 本文的算法很容易检测到目标, 并再次启动跟踪模型进行逐帧跟踪。但若目标附近存在多个类似目标时, 存在误跟踪, 主要是本文模型没有考虑轨迹预测, 而其他算法一旦丢失目标则完全不能重捕。因此, 基准序列定量分析结果表明本文算法能够准确地定位模板, 减少了跟踪点漂移现象。



(a)精确度图 (a) Accuracy



(b)成功率图 (b) Success rate

图 5 不同跟踪算法的跟踪性能对比

Fig.5 Comparison of tracking performance for different tracking algorithms

表2 跟踪结果定量对比, 其中位置误差为20, 重叠率为0.5

Table 2 Quantitative comparison of tracking results, where the location error is 20 and the overlap rate is 0.5

Model	Siam RPN	DCF Net	Siam FC	CF Net	KCF	DKCF	Proposed
Accuracy	79.92%	75.75%	56.28%	66.01%	54.31%	60.31%	84.21%
Success rate	86.30%	83.61%	75.92%	83.42%	63.13%	82.97%	86.31%

本文提出的算法通过离线训练来学习类似目标的共性特征, 采用在线训练来捕捉当前目标的个性特征, 提高了复杂背景下目标特征的辨别能力。为了验证这一点, 本文从基准数据与挂飞数据中选择了6个具有多种复杂干扰场景的序列进行跟踪测试, 并且与所选的对比算法进行了比较。从图6可以看出, 所提出的算法对于大多数复杂场景都是鲁棒的, 尤其是对于跟踪丢失后的目标重捕, 具有较高的重捕概率。为了可视化定性分析本文算法的跟踪性能, 本文选择几段序列的跟踪结果进行定性分析。图6(a)场景是对地面运动车辆的跟踪过程, 主要挑战是背景中多个类似目标, 目标本身的快速移动导致跟踪模糊, 且出现多次全遮挡现象。一旦目标被完全遮挡, 本文算法将在视场中心显示一个黑色边界框, 表明跟踪目标丢失, 进入目标重捕状态。从跟踪结果可以看出, KCF与DKCF的跟踪波门已经偏离目标中心。Siam RPN跟踪效果与本文提出的算法类似, 但本文算法具有重捕功能。Siam FC则偏离了目标中心, 在目标模糊时突然跳转跟踪到道路边缘的树冠。

图6(b)图像序列的跟踪目标是地面运动的黑色车辆, 由于视角差异导致目标外形变化较大。除CFNet, SiamRPN和本文方法外, 其他跟踪算法都丢失了目标。尽管SiamFC在跟踪丢失后又在67帧重新跟踪上目标, 主要是由于67帧目标的状态与建立模板的状

态类似, 其目标正好进入SiamFC的搜索区域。本文提出的一种结合目标边缘检测的改进全连接Siamese跟踪算法能够在长期跟踪中更好地适应目标的变形和遮挡, 特别适合机载光电设备在捕获复杂环境下的运动目标, 增强机载装备态势感知、对地观瞄与跟踪的应用效果, 提升系统响应时间与适应性。

## 4 结论

本文结合了目标检测网络、轮廓提取网络和全卷积Siamese跟踪网络的特点, 提出一种有效的机载光电系统目标跟踪算法, 该算法利用目标的轮廓模板代替边界框模板, 减少了背景杂波的干扰; 同时, 在Siamese网络的基础上增加了一路目标检测分支, 不仅能够获取目标的边界, 还能增加抗遮挡能力, 提高机载光电设备的目标捕获概率。在tiny-YOLOv3网络的基础上提出了一种改进的目标检测网络, 利用K均值聚类找到最合适的锚框, 引入了扩张模块层来扩展感受野, 并增加了基本网络的深度, 实现目标重捕的需求。仿真试验结果表明本文提出的改进模型特别适合机载光电设备在捕获与跟踪复杂环境下的运动目标, 具有较高的工程应用价值。

## 致谢

本项研究得到了桂林中电科三十四所提供的帮助, 在此表示感谢。

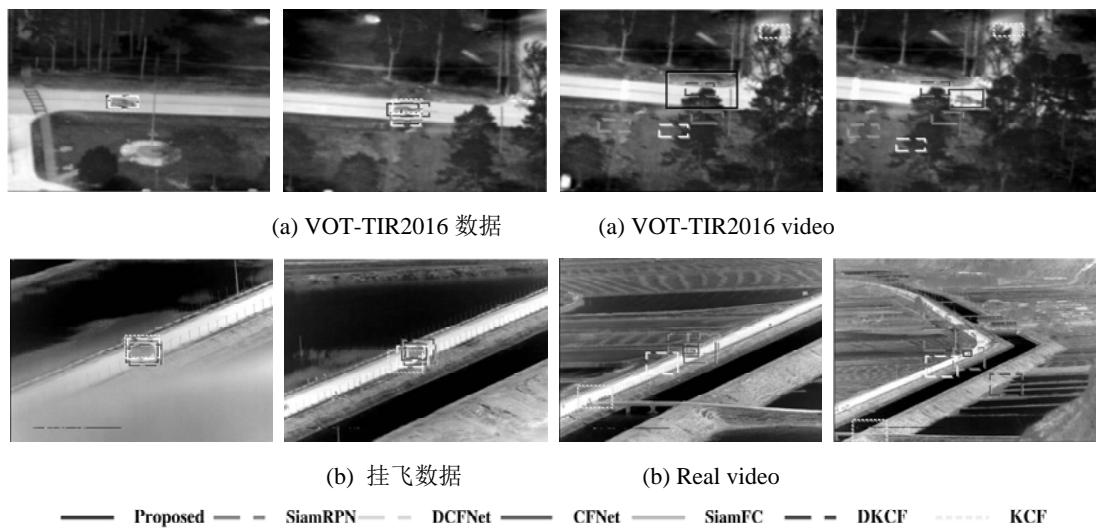


图6 不同对比算法对视频序列的跟踪结果

Fig.6 Tracking results for different models

## 参考文献:

- [1] 王爽, 邵艳明, 宋敏敏, 等. 一种面向目标跟踪性能的红外图像自适应增强方法[J]. 红外技术, 2019, 41(7): 646-653.
- WANG S, SHAO Y M, SONG M M, et al. An adaptive enhancement method for infrared image based on target tracking performance[J]. *Infrared Technology*, 2019, 41(7) : 646-653.
- [2] KONG L, HUANG D, QIN J, et al. A Joint Framework for Athlete Tracking and Action Recognition in Sports Videos[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2019, 23(5): 1-21.
- [3] YANG H, WEN J, WU X J, et al. An Efficient Edge Artificial Intelligence Multi-pedestrian Tracking Method with Rank Constraint[J]. *IEEE Transactions on Industrial Informatics*, 2019, 12(34): 1-1.
- [4] MA Chao, YANG M, ZHAN C M, et al. Hierarchical features for visual tracking[C]//*Proceedings of the IEEE International Conference on Computer Vision*, 2017, 48(9): 21-31.
- [5] 齐永锋, 王梦媛. 基于 LBP 与核相关滤波器的运动目标跟踪算法[J]. 红外技术, 2019, 41(6): 572-576.  
QI Yongfeng, WANG Mengyuan. Moving Target Tracking Algorithm Based on LBP and Kernel Correlation Filter[J]. *Infrared Technology*, 2019, 41(6): 572-576.
- [6] Poppenk J, Norman K A . Multiple-object Tracking as a Tool for Parametrically Modulating Memory Reactivation[J]. *Journal of Cognitive Neuroscience*, 2017: 1-15.
- [7] Galteri L, Seidenari L, Bertini M, et al. Spatio-Temporal Closed-Loop Object Detection[J]. *IEEE Transactions on Image Processing*, 2017, 26(3): 1253-1263.
- [8] WANG N, YEUNG D Y. Learning a Deep Compact Image Representation for Visual Tracking[C]//*Advances in Neural Information Processing Systems*, 2013, 38(18): 5045-5068.
- [9] WANG L, OUYANG Wanli, WANG Xiaogang, et al. Visual tracking with fully convolutional networks[C]//*Proceedings of the IEEE International Conference on Computer Vision*, 2015, 125(1-3): 3-18.
- [10] Bertinetto L, Valmadre J, Henriques J F, et al. Fully-Convolutional Siamese Networks for Object Tracking[C]//*European Conference on Computer Vision*. Springer, 2016: 850-865.
- [11] LUO H, XU L, HUI B, et al. Status and prospect of target tracking based on deep learning[J]. *Infrared & Laser Engineering*, 2017, 46(5): 502002.
- [12] 焦安波, 何淼, 罗海波. 一种改进的 HED 网络及其在边缘检测中的应用[J]. 红外技术, 2019, 41(1): 072-77.  
JIAO Anbo, HE Miao, LUO Haibo. Research on Significant Edge Detection of Infrared Image Based on Deep Learning[J]. *Infrared Technology*, 2019, 41(1): 072-77.
- [13] WU Y, LIM J, YANG M H. Online Object Tracking: A Benchmark[C]//*Computer Vision and Pattern Recognition (CVPR)*, 2013: 019-27.
- [14] WANG P, CHEN P, YE Y, et al. Understanding Convolution for Semantic Segmentation[C]//*IEEE Winter Conference on Applications of Computer Vision*, 2018: 1287-1299.
- [15] Held D, Thrun S, Savarese S. Learning to Track at 100 FPS with Deep Regression Networks[C]// *European Conference on Computer Vision*, 2016: 29-37.
- [16] Erickson K J, Hanna P M, Westerkamp L A, et al. Evaluation of the VIVID confirmatory identification module[C]//*Proc. SPIE*, 2007, 6566: 65660B-65660B-12.
- [17] Carvalho T, De Rezende E R S, Alves M T P, et al. Exposing computer generated images by eye's region classification via transfer learning of VGG19 CNN[C]//*16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2017: 866-870.
- [18] Akiba T, Suzuki S, Fukuda K. Extremely large minibatch SGD: training resnet-50 on imagenet in 15 minutes[J]. *arXiv preprint arXiv*, 2017, 1711: 04325.
- [19] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks BT[J]. *Comput. Vis ECCV*, 2014: 818-833.
- [20] TAO R, Gavves E, Smeulders A W M . Siamese Instance Search for Tracking[C]//*Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016: 1063-6919.
- [21] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement[J]. *Computer Vision and Pattern Recognition*, 2018, 12(7): 64-78.
- [22] WANG Q, GAO J, XING J, et al. DCFnet: Discriminant correlation filters network for visual tracking[J]. *Computer Vision and Pattern Recognition*, 2017, 1704: 04057.
- [23] Valmadre J, Bertinetto L, Henriques J F, et al. End-to-end representation learning for correlation filter based tracking[C]//*Proc. IEEE Conf Comput. Vis. Pattern Recognit.*, 2017: 5000-5008.
- [24] LI B, YAN J, WU W, et al. High performance tracking with Siamese region proposal network[C]//*Proc. IEEE Comput. Vis. pattern Reconnit. (CVPR)*, 2018: 8971-8980.
- [25] Uzkent B, Seo Y W. EnKCF: Ensemble of Kernelized Correlation Filters for High-Speed Object Tracking[C]//*IEEE Winter Conference on Applications of Computer Vision*, 2018: 77-89.