doi: 10.7690/bgzdh.2025.08.016

基于机器学习的涉密终端违规外联自动检测系统

钱　锦，徐李冰，李　昂，徐汉麟
（国网杭州供电公司，杭州 310000）

摘要：为保证电力系统中涉密终端的数据传输安全，设计基于机器学习的涉密终端违规外联自动检测系统。采用机器学习技术构建检测系统框架，按照多个阶层实现违规外联的涉密信息过滤；划分违规外联主要方式，通过模块化松耦合模式统一违规外联监测接口；以聚类算法获取涉密信息聚类中心，按照非线性映射关系划分终端违规外联数据类型；基于机器学习计算信息增益率，以特征信息熵为基础自动检测涉密终端违规外联。以 3 组涉密终端作为测试对象，在多种违规外联方式下验证检测系统的应用效果。结果表明：该系统可实现快速阻断，且具有较高的查全率和查准率，保证涉密信息的安全性。

关键词：机器学习；涉密终端；违规外联；自动检测系统
中图分类号：TP309.1　　**文献标志码：**A

Automatic Detection System for Violation of Secret Terminal Based on Machine Learning

Qian Jin, Xu Libing, Li Ang, Xu Hanlin
(*State Grid Hangzhou Power Supply Comapny*, *Hangzhou* 310000, *China*)

Abstract: In order to ensure the security of data transmission of classified terminals in power system, an automatic detection system for illegal external connection of classified terminals based on machine learning is designed. Constructing a detection system framework by adopting a machine learning technology, and filtering secret-related information of violation outreach according to a plurality of levels; dividing a main violation outreach mode, and unifying a violation outreach monitoring interface through a modularized loose coupling mode; acquiring a secret-related information clustering center by a clustering algorithm, and dividing a terminal violation outreach data type according to a nonlinear mapping relationship; The information gain rate is calculated based on machine learning, and the illegal outreach of the secret-related terminal is automatically detected based on the feature information entropy. Taking three groups of classified terminals as the test objects, the application effect of the detection system is verified in a variety of illegal outreach modes. The results show that the system can achieve rapid blocking, and has a high recall rate and precision rate to ensure the security of classified information.

Keywords: machine learning; classified terminal; illegal external connection; automatic detection system

0　引言

在大数据技术的蓬勃发展下，网络用户的数量不断增加，网络终端的泄密违规问题越来越严重。为实现对网络和终端设备的长期监督，需要对可疑数据进行及时判断，建立一个有效的监督和检测系统，为网络的信息安全提供保障。文献[1]通过边缘样本技术设计了检测系统，主要针对入侵信息进行数据拦截，以此保证网络的运行安全。该方法主要通过 GAN 算法和模糊对抗网络进行拟合模型设计，对数据样本进行有效清洗，在分析经典入侵数据的前提下对信息进行过滤，构建了一个有效的污染数据检测模型。通过边缘样本技术能够对网络中的污染数据进行精准检测，适用于多种常见的网络入侵，具有一定的应用效果。但由于对抗网络识别的时间较久，在整体污染数据处理中会耗费大量时间，导致检测系统的过滤时效不佳。为解决检测效率问题，文献[2]设计了一个新的跟踪系统，该系统通过人机交互技术进行实时跟踪，通过点云特征建立特殊数据的提取框架，促使检测目标被圈定在一定区域内，实现有目的和有顺序的交互跟踪。通过人机交互技术能够确定问题数据的流转过程，在系统检测到该数据时可以实现场景标定，对需要被保护的数据进行隔离，实现了有效的检测和跟踪；但在输出过程中容易受到环境因素影响，不能满足复杂场景的需求。为有效保证信息数据的安全，需要设计一个快速且有效的检测系统。机器学习技术是人工智能的核心技术，可以对大量的数据信息进行快速标记。笔者以此为研究前提，设计一个新的涉密终端违规

收稿日期：2024-09-13；**修回日期：**2024-10-14
第一作者：钱　锦(1988—)，男，江苏人，硕士。

外联检测系统，为网络安全提供理论支持。

1 涉密终端违规外联自动检测系统硬件设计

1.1 基于机器学习设计监督检测系统框架

随着电力系统中用户的不断增加，整个电力系统中产生的数据信息急速增长，需要设计一个有效的监督检测系统。为保证电力行业的信息运行安全，需要对传输过程中的设备以及终端设备进行实时监督，笔者以机器学习技术中的半监督聚类方式构建检测系统框架，对电力系统内的数据进行有效分析。由于泄漏违规信息与常规数据存在差异性，在框架内对泄漏的途径和样式进行划分，建立一个多层的过滤阶级，使其能够在流量数据外输时进行有效检测，保证系统的检测性能。具体框架如图 1 所示。

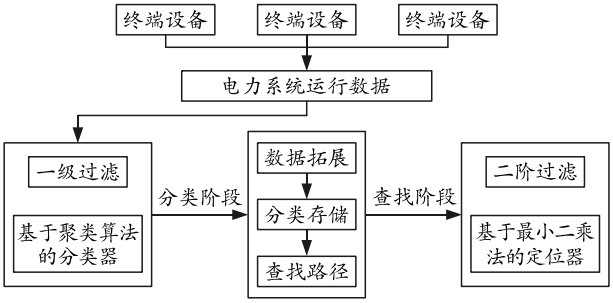


图 1 基于机器学习的自动检测系统框架

如图 1 所示，此次设计的框架主要是为降低数据分类的复杂性，将整体检测过程划分为 3 部分，分别为一阶过滤、数据分类和二阶过滤[3]。从机器学习技术中选择聚类算法和最小二乘法进行框架内部的检测设备构建，通过聚类算法设计数据分类器，主要是对传输的数据进行分类，通过最小二乘法设计定位器，主要是对泄漏设备进行标定。在 2 个检测设备的作用下，能够实现电力系统的数据聚类和监督，完成泄漏检测的初步检验。

在框架内以遗传算法作为一级过滤层次，最小二乘法作为二阶过滤，2 个部分相结合实现数据分类和标定，主要过程为：将电力系统运行的数据划分为正常数据和外泄数据 2 种类型，对正常的数据进行传输和保存，而外泄的数据需要对其路径进行判断，以此标定出具体的泄漏终端设备。在路径标记过程中采用机器学习技术进行数据扩展，定义泄漏数据的所属类型，验证其对电力系统的威胁程度，以此排序不同泄漏数据的检测顺序。

1.2 模块化松耦合模式统一违规外联监测接口

涉密终端要实行物理隔离，按照国家现行要求，在物理环境中确定涉密终端在安全场所内，但违规外联行为是一种内部的违规行为，内部的涉密信息通过违规外联作为跳板，从而将内部信息传输至外部环境中[4]。对常见的违规外联形式进行说明，如表 1 所示。

表 1 违规外联主要方式

具体方式	详细说明
多网卡	以多网卡或交叉网线等方式接入电力系统
拨号上网	在调制器设备中进行拨号连接
无线	WIFI 形式或者 5G 联网
外设	串口
跳板	设备互通

通过表 1 中内容所示，在电力系统中存在多种违规外联方式，为保证涉密终端的信息安全，选择模块化的松耦合形式建立统一接口，使涉密终端设备按照统一的形式进行信息传导，防止信息被窃取和入侵。主要结构如图 2 所示。

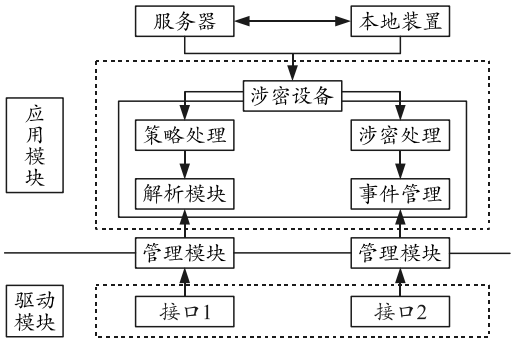


图 2 涉密终端统一接口结构

如图 2 所示，在涉密终端设备中同时接入 2 组接口，分别执行本地电网信息接收和传输，在服务器接收到外部的信息后，通过通信模块进行数据分析和事件处理，以此实现多个设备的有效管 理[5]。对违规外联的监控主要包括代理监控和组件模块，其中代理监控实现数据的存储和转发功能，组件模块为设备提供统一接口，在检测框架和监测结构的双重布置下，实现系统的硬件设计。

2 基于机器学习的自动检测系统软件设计

2.1 聚类算法下划分终端违规外联数据类型

通过硬件框架设计，对自动检测系统的软件进行构建，按照框架内的分类器以及定位器进行匹配，采用聚类算法划分违规外联数据类型。在聚类算法的作用下随机选择聚类中心，依次计算外联数据与初始中心的距离，公式为：

$$m^{(n)} = \arg\min_x \left\| v^{(n)} - c_b \right\|^2 。 \tag{1}$$

式中：n 为外泄数据；$m^{(n)}$为 n 的表现形式；$v^{(n)}$为被分类的聚类中心；b 为选择的初始聚类中心；x

为总计聚类中心；c_b 为 x 个中心的第 b 个聚类中心。采用欧氏距离进行检测：

$$z(l,k)=\sqrt{\sum_{h=1}^{j}(l_h-k_h)^2}\ 。\tag{2}$$

式中：$z(l, k)$为欧氏距离；l_h 为输入数据的第 n 个特征；j 为特征总量，且 $h\in j$；k_h 为该特征在聚类中心的对应值。在对每个外泄数据分配后，每个类型的数据均可以移动至所属的聚类中心，在重复操作下保证每个类型不再移动，实现泄密终端违规外联数据的聚类。

为保证后续检测的准确性和定位精度，对聚类后的数据进行精准划分，按照均值采样和最大值采样 2 种形式，对数据进行非线性映射处理，使其在较快时间内归属到所属类型中：

$$g_f^d=s\left(\sum_{a\in i_f}g_a^{d-1}\times o_{af}^d+p_f^d\right)。\tag{3}$$

式中：$s(\bullet)$为激活函数；在非线性映射处理中存在多个层级，d 为层级数量[6]；g_f^d 为在该层的输出；g_a^{d-1} 为在 d 层的输入；a、f 为泄漏数据量，且 $a\in i_f$；o 为泄漏数据的聚类中心；p 为定位偏差；i_f 为数据的特征集合。以此最大值采样和均值采样划分模式为：

$$g_f^{d''}=s\left(u_f^d\times\max_{a\in i_f}\left(g_a^{d-1}\right)+p_f^d\right);\tag{4}$$

$$g_f^{d'}=s\left(u_f^d\times\frac{1}{t\times y}\sum_{a\in i_f}g_a^{d-1}\times o_{af}^d+p_f^d\right)。\tag{5}$$

式中：$g_f^{d''}$ 为最大值采样分类模式；$g_f^{d'}$ 为均值采样分类模式；u 为分类偏差[7]；t、y 为分类维数。将划分后的数据进行统计，按照所属类型对应传输路径，并在机器学习方式下对泄密终端进行定位，以此实现对违规操作设备的自动检测。

2.2　基于机器学习自动检测涉密终端违规外联

上述实现了对外联传输数据的分类，为保证每种分类具有查全指标，在聚类算法的基础上增加机器学习技术的决策算法，针对每种涉密终端违规外联情况进行处理。

假定存在训练数据集(Q_W, E_W)，其中 W=1, 2, …, R；Q_W 表示第 W 个涉密终端中的输入数据；E_W 为其类型标签，共计有 T 类[8]。每个设备中的涉密数据均存在多种特征，以此对其信息熵进行计算：

$$\mathrm{Info}(Y)=-\sum_{W=1}^{T}U_W\log_2(U_W)。\tag{6}$$

式中：U_W(W=1, 2, …, T)为输入数据 Q 属于第 W 个类别的概率；Y 为数据集合[9]；Info(Y)为 Y 的信息熵。当特征的信息熵越大说明形成的集合就越混乱，假定数据集 Y 被特征 P 划分，最终形成 I 个数据集，则计算方式如下：

$$\mathrm{Info}_P(Y)=\sum_{A=1}^{I}\left(\frac{|Y_A|}{|Y|}\times\mathrm{Info}(Y_A)\right)。\tag{7}$$

式中 Info_P 为特征 P 的信息熵，表示按照特征 P 划分的第 A 个数据集的信息熵。以此作为违规外联的检测基础，将特征 P 划分的信息增益结果作为原始数据集合进行分类，具体为：

$$\mathrm{Infogain}(F,P)=\mathrm{Info}(Y)-\mathrm{Info}_P(Y)。\tag{8}$$

式中：F 为增益率；Infogain(•)为度量函数。在分类过程中对涉密信息进行度量，以此实现该信息是否被传入至外联通路中，而被外联的信息成为危险数据，其特征值向量为：

$$\mathrm{splitInfo}_P(Y)=-\sum_{A=1}^{I}\left(\frac{|Y_A|}{|Y|}\times\log_2\frac{|Y_A|}{|Y|}\right)。\tag{9}$$

式中 $\mathrm{splitInfo}_P(Y)$为被外联的涉密数据特征向量。其本质上即为含有特征 P 的信息熵，可以衡量信息包含特征 P 的信息量，则涉密数据被外联的增益率计算方式如下：

$$\mathrm{Infogainration}(F,P)=\frac{\mathrm{Infogain}(F,P)}{\mathrm{splitInfo}_P(Y)}。\tag{10}$$

式中 Infogainration(F, P)为涉密信息外联过程中含有特征 P 的增益率。当信息增益率增加时表示其含有的内在信息减少，可能已经被外联通道所连接，以此影响涉密终端信息的安全性[10]。在此基础上对不同设备的涉密信息进行计算，实现违规外联的自动检测。笔者通过机器学习中的多个算法，分别设计检测系统的硬件结构和软件结构，实现自动检测系统设计。

3　实验测试分析

上述采用机器学习技术设计了自动检测系统，为验证该系统能够实现涉密终端的违规外联检测，采用对比测试的方法进行论证。分别选择基于边缘样本的检测系统和基于人机交互的检测系统作为对照组，与本文中系统进行比较，验证不同系统的应用效果。

3.1　搭建测试环境

涉密终端违规外联检测系统主要是对涉密终端进监控，以此在测试过程中需要对涉密终端安装代理程序，并设定违规外联策略，而选择的监控系统

直接连接在服务器中，直接对测试终端的违规行为进行检测。在 Matlab 平台中搭建测试环境，配置要求如下：

1) 硬件环境：设备类型为电力系统中的数据接收计算机，硬盘内存超过 40 G，主频在 16.2 GHz 以上。

2) 软件环境：操作系统为 Windows 7，并分配了多个固定用户 IP，数据存储库的版本在 5.2 以上。

根据搭建的网络环境模拟电力系统的数据接收和传输过程，选择 3 组计算机终端作为局域网，根据电力系统的行业要求，选择的设备为物理隔离状态，测试环境结构如图 3 所示。

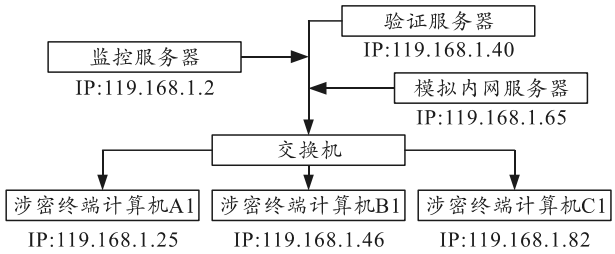
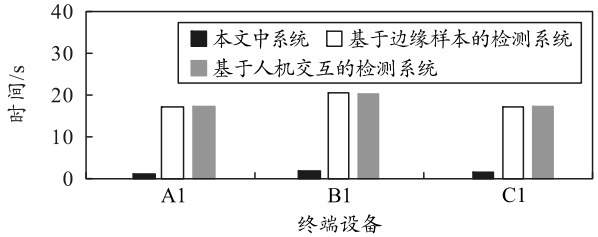


图 3 测试环境

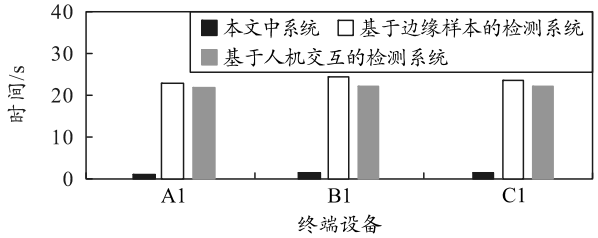
如图 3 所示，此次共选择 3 组计算机作为涉密终端进数据传输，并通过内网服务器模拟违规外联形式，通过不同的违规外联检验各组系统的检测效果。

3.2 不同违规外联方式下系统检测时间

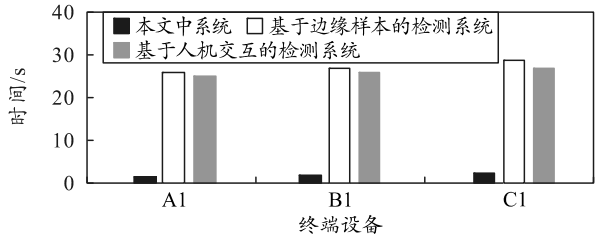
为验证系统的有效性，保证测试的真实度，将测试分为 2 部分，第 1 阶段验证各系统的检测时间，选择 3 种违规外联方式，分别为非授权接入、串口接入和多网卡接入。设定各组检测系统在监测到违规行为后会发出报警信号并进行阻断，具体结果如图 4 所示。



(a) 非授权接入违规外联方式



(b) 串口接入违规外联方式



(c) 网卡接入违规外联方式

图 4 不同系统的阻断时间测试结果

图 4 中，在本文中系统应用下能够在 3 s 内完成不同违规外联的阻断，具有较强的检测效果，而传统系统的阻断时间超过 15 s，尤其在多网卡违规外联方式下，阻断时间超过 25 s，综合说明本文中系统更加有效。

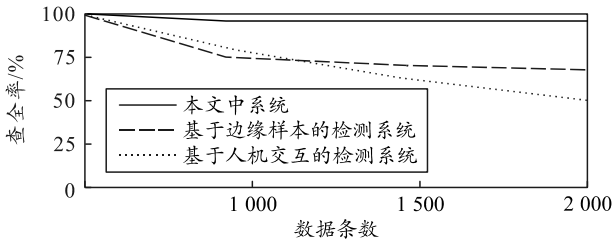
3.3 涉密信息违规外联查全率与查准率测试结果

上述阶段验证了本文中系统的有效性，为进一步分析本文中系统的应用价值，对涉密信息违规外联后的查全率和查准率进行检测。选择多网卡接入违规外联方式进行检验，3 组设备数据传输量分别为 1 000、1 500 和 2 000 条。根据设定情况可知，A1 设备传输数据量较少，而 C1 设备的传输量较多，分别对各系统的查全率和查准率情况进计算，公式为：

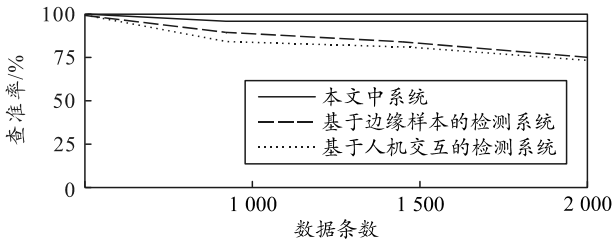
$$N=(M/B)\times 100\%; \tag{11}$$

$$V=(C/X)\times 100\%。 \tag{12}$$

式中：N 为查全率；M 为检测到的数据量；B 为全部外泄数据；V 为查准率；C 为正确检测到的外泄数据；X 为实际的外泄数据。将各系统检测到的数据代入，结果如图 5 所示。



(a) 查全率



(b) 查准率

图 5 不同检测系统的测试结果

如图 5 所示，在本文中系统应用下能够保证较高的查全率和查准率，基本可以控制在 98%以上，

而 2 组传统系统的查全率与查准率，会随着数量的增加逐渐降低，在数据量为 2 000 条时，传统系统的查全率不足 60%，查准率不足 80%，说明本文中系统更具有应用价值。

4　结束语

机器学习的研究领域十分广泛，包括数据预处理和人工神经网络以及支持向量机等多种技术。笔者以其为研究前提设计了新的检测系统，并在实验论证的基础上证明了自动检测系统的有效性，能够对泄密终端的违规现象进行精准识别，较传统系统具有更强的应用价值。研究过程中仍存在不足之处，如只选择了一种芯片技术进行分析，受限于硬件多方设计。后续研究中会针对问题，提出更多适用于自动检测系统的硬件结构，为终端信息的安全运行提供理论支持。

参考文献：

[1] 刘广睿, 张伟哲, 李欣洁. 基于边缘样本的智能网络入侵检测系统数据污染防御方法[J]. 计算机研究与发展, 2022, 59(10): 2348-2361.

[2] 吴超, 吴绍斌, 李子睿, 等. 基于人机交互的免锚检测和跟踪系统设计[J]. 兵工学报, 2022, 43(10): 2565-2575.

[3] 周璇, 王馨瑶, 闫军威, 等. 基于机器学习的建筑复杂用能系统运行状态异常检测[J]. 华南理工大学学报(自然科学版), 2022, 50(7): 144-154.

[4] 朱海亮, 潘巨龙, 刘鹏达. 基于 PCA-ANN 的跌倒检测系统设计与实现[J]. 微电子学与计算机, 2022, 39(6): 108-114.

[5] 张洪亮. 基于大数据的煤矿违规行为分析识别系统研究[J]. 煤矿安全, 2022, 53(5): 133-137.

[6] 邵晋梁, 石磊, 李彤, 等. 合作竞争网络下的多智能体系统链路故障检测[J]. 中国科学:信息科学, 2022, 52(8): 1500-1512.

[7] 徐钢, 黎敏, 吕志民, 等. 基于机器学习的产品质量在线智能监控方法[J]. 工程科学学报, 2022, 44(4): 730-743.

[8] 马海洲, 丁爱萍. 基于机器学习的舰船信息系统入侵检测技术[J]. 舰船科学技术, 2021, 43(22): 163-165.

[9] 孟令雯, 张锐锋, 李鑫卓, 等. 基于机器学习的变电站设备异常状态数据清洗[J]. 电力系统及其自动化学报, 2021, 33(12): 79-86.

[10] 苟升异, 宿翀, 王磊, 等. 一种基于阵列式 PVDF 触觉传感器和机器学习的针刺手法识别系统[J]. 针刺研究, 2021, 46(6): 474-479.

****

（上接第 67 页）

参考文献：

[1] 王翔. 一种复杂海空背景下的红外小目标检测跟踪算法[J]. 光学与光电技术, 2022, 20(2): 113-119.

[2] 徐亚杰, 王海星. 基于新特征算子的复杂背景红外弱目标检测算法[J]. 国外电子测量技术, 2021, 40(12): 7-11.

[3] 戴亚峰, 陶青川, 杨波. 联合 YOLOv3 和核相关滤波算法的红外视频目标检测及跟踪[J]. 计算机应用与软件, 2022, 39(1): 200-205, 230.

[4] 王笛, 沈涛. 复杂天空背景下的红外弱小目标检测算法研究[J]. 光学学报, 2020, 40(5): 103-110.

[5] 马永杰, 马芸婷, 程时升, 等. 基于改进 YOLOv3 模型与 Deep-SORT 算法的道路车辆检测方法[J]. 交通运输工程学报, 2021, 21(2): 222-231.

[6] 张宏鸣, 汪润, 董佩杰, 等. 基于 DeepSORT 算法的肉牛多目标跟踪方法[J]. 农业机械学报, 2021, 52(4): 248-256.

[7] 王思奎, 刘云鹏, 亓琳, 等. 基于背景约束与卷积特征的目标跟踪方法[J]. 计算机工程与应用, 2020, 56(8): 205-214.

[8] 王博, 董登峰, 周维虎, 等. 面向激光跟踪仪跟踪恢复的合作目标视觉检测[J]. 光学精密工程, 2020, 28(2): 271-282.

[9] 侯春萍, 张倩文, 王晓燕, 等. 轮廓匹配的复杂背景中目标检测算法[J]. 哈尔滨工业大学学报, 2020, 52(5): 121-128.

[10] 王鑫瑞, 陈明. 基于循环相关滤波算法的复杂背景目标跟踪方法研究[J]. 国外电子测量技术, 2020, 39(4): 22-25.

[11] 寇志强, 艾斯卡尔·艾木都拉. 局部最大熵的红外小目标快速检测方法[J]. 激光杂志, 2020, 41(7): 18-22.

[12] 李大维, 莫波, 高可, 等. 针对弹载平台红外小目标的检测与跟踪算法[J]. 宇航学报, 2020, 41(11): 1440-1448.

[13] 蔡伟, 徐佩伟, 杨志勇, 等. 复杂背景下红外图像弱小目标检测[J]. 应用光学, 2021, 42(4): 643-650.

[14] 李继泉, 时勤功, 胡春松. 一种复杂背景下红外目标稳定跟踪算法[J]. 红外技术, 2020, 42(5): 434-439.

[15] 杨博, 蔺素珍, 禄晓飞, 等. 基于多特征融合与分层数据关联的空中红外多目标跟踪方法[J]. 计算机应用, 2020, 40(10): 3075-3080.

[16]向涛. 基于显著区域提取的红外图像舰船目标检测[J]. 电讯技术, 2020, 60(7): 785-790.

[17] 王芳, 李传强, 伍博, 等. 基于多尺度特征融合的红外小目标检测方法[J]. 红外技术, 2021, 43(7): 688-695.