

引用格式: 邓佳坤, 尹益卓, 张彦博, 等. 结合图表示学习和多特征融合的红外小目标检测 [J]. 电子科技大学学报, 2026, 55(1): 100-108.
DENG J K, YIN Y Z, ZHANG Y B, et al. Infrared small target detection based on multi-feature fusion combined with graph representation learning[J].
Journal of University of Electronic Science and Technology of China, 2026, 55(1): 100-108.



结合图表示学习和多特征融合的 红外小目标检测

邓佳坤, 尹益卓, 张彦博, 龙 畅, 李科萱, 崔兴晔, 彭真明*

(电子科技大学信息与通信工程学院, 成都 611731)

摘要: 红外目标检测是红外搜索与跟踪系统的核心技术之一。在复杂背景下红外目标信号微弱, 且存在大量不规则干扰源, 容易引发虚警。针对这一问题, 提出了一种结合图表示学习和多特征融合的红外小目标检测算法。首先采用形态学方法提取目标候选区域; 由于不规则虚警源和目标在视觉上难以协同表征, 将目标候选区从图像领域转换到图领域, 并分别提取基于图像的手工特征和基于图表示学习的深度特征; 最后使用全连接网络进行特征融合和分类, 筛除虚警区域, 得到目标区域。在公开的红外小目标数据集上进行了性能对比实验, 结果表明该算法在复杂场景下具有较好的检测性能。

关键词: 红外小目标; 目标检测; 图表示学习; 特征融合; 深度特征

中图分类号: TP39

文献标志码: A

DOI: 10.12178/1001-0548.2025060

Infrared small target detection based on multi-feature fusion combined with graph representation learning

DENG Jiakun, YIN Yizhuo, ZHANG Yanbo, LONG Chang, LI Kexuan, CUI Xingye, and PENG Zhenming*

(School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China)

Abstract: Infrared target detection is one of the core technologies in infrared search and tracking systems. In complex backgrounds, infrared target signals are weak and there are numerous irregular sources of interference, which can easily lead to false alarms. To address this issue, this paper proposes an infrared small target detection algorithm that combines graph representation learning and multi-feature fusion. Initially, morphological methods are used to extract candidate target regions. Then, considering irregular false alarm sources and targets are difficult to represent visually in a coordinated manner, the candidate target regions are transform from the image domain to the graph domain to extract both handcrafted features based on images and deep features based on graph representation learning. Finally, a fully connected network is used for feature fusion and classification, thereby filtering out the false alarm regions and obtaining the target regions. The performance comparison experiments are conducted on a public infrared small target dataset, and the results show that the proposed algorithm has good detection performance in complex scenarios.

Key words: infrared small target; object detection; graph representation learning; feature fusion; deep features

红外目标检测是红外搜索与跟踪系统的核心技术, 目前已广泛应用于军事侦察、安防监控、遥感分析等领域。在实际红外成像过程中, 由于成像设备与目标距离较远, 目标往往成像模糊、面积较小、对比度低且缺乏纹理细节。此外, 背景中存在多种干扰源, 如树木、高亮建筑物、水面反光以及

空中的高亮卷云边缘等, 这些因素对检测算法的有效性和稳定性提出了巨大挑战。

当前红外弱小目标检测方法主要分为多帧检测和单帧检测两类。多帧检测方法利用序列图像中目标的连续性和相关性实现目标检测, 而单帧检测方法则依靠目标与背景的成像差异进行检测。多帧检

收稿日期: 2025-03-20

基金项目: 国家自然科学基金(61571096); 四川省自然科学基金(2025ZNSFSC0522)

作者简介: 邓佳坤, 博士, 主要从事遥感目标检测、识别及跟踪等方面的研究。

*通信作者 E-mail: zmpeng@uestc.edu.cn

测需要依赖一定数量的帧数才能实现检测, 且容易受到背景运动的影响。相比之下, 单帧检测计算简单、适用性更广。目前的单帧检测方法通常分为基于人类视觉显著性^[1-4]、低秩稀疏分解^[5-9]、基于深度学习^[10-12]以及基于特征提取和分类^[13]的方法。基于人类视觉显著性的方法以局部对比度系列算法为主, 如局部对比度 (local contrast measure, LCM)^[1]、改进的局部对比度测量 (ILCM)^[2]、基于块的多尺度局部对比度 (MPCM)^[3]以及比率-差分联合局部对比度 (RDLCM)^[4]等。这些方法假设目标区域具有高对比度, 通过计算显著性图来检测目标, 但当目标对比度较低或干扰较多时, 其效果较差。低秩稀疏分解方法基于目标稀疏和背景低秩的假设, 将检测问题转化为矩阵 (如 IPI^[5]、NRAM^[6]) 或张量 (如 RIPT^[7]、PSTNN^[8]、SRSTT^[9]) 的恢复问题。然而, 这类方法通常计算复杂度较高、参数鲁棒性较低, 且在背景不满足低秩条件时效果不佳。基于深度学习的方法 (如 ACM^[10]、ISTDU-Net^[11]、UIUNet^[12]) 利用深度神经网络自动学习目标特征并用于检测和分类, 但其性能依赖于大量红外图像数据。由于当前真实红外数据集较少, 且目标在红外图像中仅占据少量像素, 存在严重的类别失衡问题, 导致端到端的检测模型缺乏场景鲁棒性和泛化能力。基于候选区提取加分类的方法 LCI^[13]结合了模型驱动和数据驱动的各自优势, 首先采用模型驱动方法提取目标候选区, 然后通过特征提取和分类器去除其中虚警。这种框架比较适合当前的小样本红外图像环境, 且模型具有更强的场景鲁

棒性。

基于此, 本文采用了特征提取加分类的框架思路。为了解决不规则虚警源难以建模的问题, 并充分利用深度特征在分类中的有效性, 提出了一种结合图表示学习和多特征融合的红外小目标检测方法。该方法包括 3 个部分: 候选区域提取、特征提取和分类。首先, 使用自适应阈值的 Top-Hat^[14]对可能含有红外弱小目标的区域进行初步检测, 以剔除大部分无关背景, 从而缓解类别失衡问题; 再将目标候选区从图像领域转换为图领域, 并分别提取图像上的手工特征和基于图表示学习的深度特征; 最后使用一个 4 层全连接网络作为特征融合和分类网络筛选虚警区域, 最终得到目标区域。

1 方法

本文提出的红外弱小目标检测算法综合了传统模型驱动方法与数据驱动方法的优势, 同时将计算机视觉算法与图表示学习相结合, 总体框架如图 1 所示。该方法分为 3 个主要部分: 1) 采用 Top-Hat 快速提取红外弱小目标的候选区域 (candidate region, CR) 二值图像, 初步确定目标位置, 并根据候选区域的位置, 依次从原图、Top-Hat 响应图和候选区域图像中提取前景图像片、响应图像片和模板图像片; 2) 基于各图像片提取视觉层次的手工特征并构建候选图结构, 然后采用图注意力网络 (graph attention networks, GAT)^[15]挖掘其深度特征; 3) 对提取的手工特征和深度特征进行融合, 并训练分类器完成目标分类。

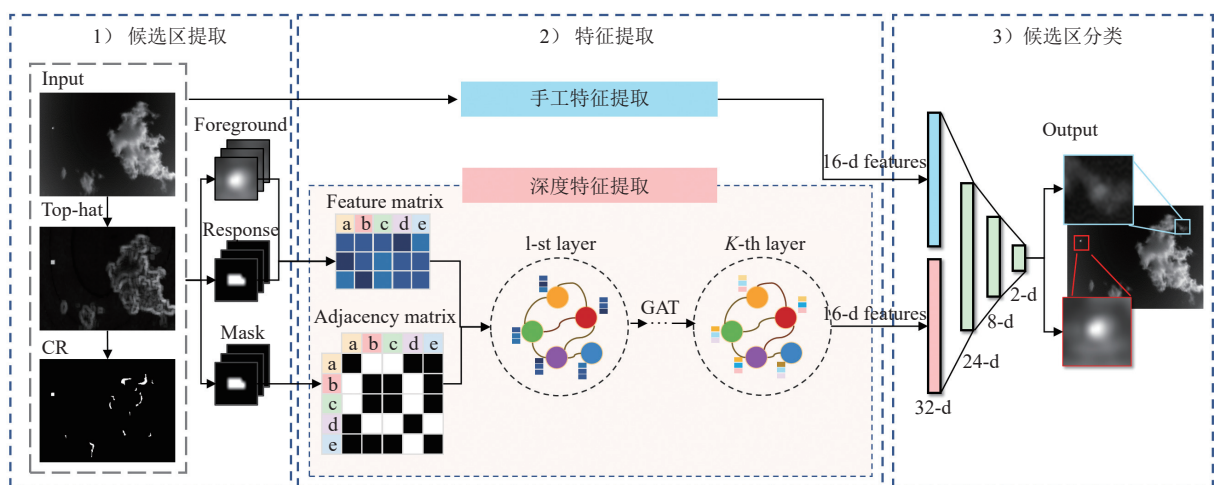


图 1 算法总体框架

1.1 候选区提取

红外弱小目标通常指像素面积小于 11×11 且图

像占比小于 0.1% 的目标。这类目标通常对比度较低、纹理细节较少, 且灰度分布接近高斯分布。针

对此类目标, 可以首先采用形态学中的 Top-Hat 方法来初步检测弱小目标。Top-Hat 方法通过对输入图像进行形态学开操作去除小的亮物体, 从而估算背景图像, 再通过输入图像减去背景图像得到响应图像, 如式 (1) 所示:

$$I_{\text{res}} = I_{\text{open}}(\varphi) - I \quad (1)$$

式中, I 是输入图像; I_{open} 是输入图像执行以 φ 为结构元的开操作后的图像; I_{res} 是获得的响应图像; φ 是以 r_φ 为半径的圆形结构元, 如式 (2) 所示:

$$r_\varphi = \min\left(\lceil \sqrt{0.001w_I h_I} \rceil, w_0\right) \quad (2)$$

式中, w_I 和 h_I 分别是输入图像的宽和高; $\lceil \cdot \rceil$ 表示向下取整; w_0 表示最大目标的尺寸宽度, 这里取 11。

对 I_{res} 进行以 θ 为阈值的阈值分割得到弱小目标的前景二值图 I_{BW} 。 θ 值通过二分搜索来确定, 通过测试, 0.3 是一个较好的起始值, 终止条件是从 I_{BW} 中获取的除去图像边缘部分的连通域数在 $1 \sim K$ 之间, 以防止丢失目标区域或检出太多非目标区域导致算法整体性能降低。如果无法满足终止条件, 取检出连通域数量最少的结果。

本文采用 4 连通准则对前景二值图 I_{BW} 进行连通域分析。连通域可能为单个像素的噪声和长边缘, 为了提升检测效率, 将像素面积小于 2 和大于 121 的连通域剔除, 得到候选区域二值图像 I_{CR} 。对于 n 个连通域 $I_{\text{CR}}^{(i)}$, 为了关注更多的目标周围信息, 以每个连通域的中心 $(c_x^{(i)}, c_y^{(i)})$ 为中心, 在原图 I 中取长宽为 r_c 的矩形区域作为候选区域 $I_{\text{cand}}^{(i)}$, r_c 为:

$$r_c^{(i)} = \max(2 \max(w_c^{(i)}, h_c^{(i)}) + 1, w_0) \quad (3)$$

式中, $w_c^{(i)}$ 和 $h_c^{(i)}$ 分别是该连通域最小外接矩形的宽和高。得到至少 11×11 的候选区域 $I_{\text{cand}}^{(i)}$ 。对于长或宽大于 5 个像素的目标, $2 \max(w_c^{(i)}, h_c^{(i)}) + 1$ 的候选区域包含更多局部背景信息。

1.2 手工特征提取

通过候选区域提取, 红外小目标检测问题转换为判断候选区域 $I_{\text{cand}}^{(i)}$ 是否为真实目标的二分类问题。该类方法基于目标候选区域挖掘大量特征, 并通过机器学习模型训练分类器。本文参考文献 [13], 需要获取必要的图像特征作为分类器的基础特征。首先, 候选目标的形状是区分目标跟虚警的重要因素之一, 因此基于 $I_{\text{CR}}^{(i)}$ 提取了 4 维形状特

征, 包括面积 (Area)、周长 (Perimeter)、离心率 (Eccentricity)、圆度 (Circularity)。考虑到红外目标的局部显著性^[1], 目标与局部背景的灰度信息也是区分真实目标和虚警的关键。候选区域 $I_{\text{cand}}^{(i)}$ 包含了目标区域 $I_{\text{TG}}^{(i)}$ 和局部背景区域 $I_{\text{BG}}^{(i)}$, 基于此本文设计了 12 维的灰度特征, 包括了目标区域和背景区域各自的灰度标准差 (StdTgtInt、StdBgInt)、灰度平均值 (AvgTgtInt、AvgBgInt) 和灰度熵 (EntTgt、EntBg), 以及描述整个区域的灰度平均值 (AvgFullInt)、灰度对比度 (Contrast1、Contrast2)、灰度熵 (EntFull)、熵差异 (EntDiff1、EntDiff2)。注意灰度熵主要用于描述像素灰度的纹理复杂度, 红外场景中目标通常具有结构性纹理 (中等熵值), 而背景可能是低熵 (天空) 或者高熵 (杂波)^[5]。最终的手工特征共计 16 维, 具体描述如表 1 所示。

表 1 手工特征提取

特征名称	特征描述	维度
Area	Pixel area of $I_{\text{CR}}^{(i)}$	1
Perimeter	Pixel perimeter of $I_{\text{CR}}^{(i)}$	1
Eccentricity	Ratio between foci and major axis length of target region	1
Circularity	$4\pi \text{Area} / \text{Perimeter}^2$	1
StdTgtInt	Standard deviation of intensities in $I_{\text{TG}}^{(i)}$	1
StdBgInt	Standard deviation of intensities in $I_{\text{BG}}^{(i)}$	1
AvgTgtInt	Average intensity in $I_{\text{TG}}^{(i)}$	1
AvgBgInt	Average intensity in $I_{\text{BG}}^{(i)}$	1
AvgFullInt	Average intensity in $I_{\text{cand}}^{(i)}$	1
Contrast1	$\text{AvgTgtInt} / \text{AvgBgInt}$	1
Contrast2	$\text{AvgTgtInt} - \text{AvgBgInt}$	1
EntTgt	Entropy of $I_{\text{TG}}^{(i)}$	1
EntBg	Entropy of $I_{\text{BG}}^{(i)}$	1
EntFull	Entropy of $I_{\text{cand}}^{(i)}$	1
EntDiff1	$\text{EntTgt} / \text{EntFull}$	1
EntDiff2	$\text{EntBg} / \text{EntFull}$	1

1.3 图表示学习

表示学习是指从原始数据中自动提取到有用的特征, 无需人工设计。这些特征能够揭示数据的内在结构和模式。卷积神经网络 (convolutional neural networks, CNN)^[16] 是表示学习最常见的手段, 目前已广泛用在各类计算机视觉任务中^[17-19]。它通过训练一个浅层的二分类卷积神经网络, 将分

类层舍弃, 并将某个全连接层的特征向量作为输出。基于 CNN 的表示学习要求输入尺度统一, 需要将候选区域 $I_{\text{cand}}^{(i)}$ 经过插值缩放为统一尺度才能得到自动的特征提取器。因此, 基于 CNN 的表示学习方法无法面向复杂场景中的不规则虚警源建模, 强行统一候选区尺度会改变它们的原始特性, 影响最终的特征提取效果。在机器学习任务中, 图数据结构具有更高的灵活性和表达能力, 能够准确地模拟现实世界中各种复杂的关系。因此, 本文将候选区域 $I_{\text{cand}}^{(i)}$ 转换为图结构, 并提出了一个更加灵活的基于图表示学习的特征自动提取器。

1.3.1 候选图构建

一个图描述为 $G = (\mathbf{V}, \mathbf{A})$, 其中 \mathbf{V} 表示 N 个节点构成的特征矩阵 (Feature matrix), $\mathbf{V} \in \mathbb{R}^{N \times F}$, F 是每个节点特征 \vec{h} 的维度。 $\mathbf{A} \in \{0, 1\}^{N \times N}$ 是一个稀疏的邻接矩阵 (Adjacency matrix), 表示节点之间的边连接关系, 其中 $A_{i,j} = 1$ 表示节点 V_i 与节点 V_j 通过一条边相连接, $A_{i,j} = 0$ 表示不存在边关系。根据边是否存在方向, 一个图又被分为无向图和有向图。

将候选区域 $I_{\text{cand}}^{(i)}$ 转换为候选图 $G^{(i)}$ 时, 首先需要定义图的拓扑结构, 包括节点和边的定义, 将连通域 $I_{\text{CR}}^{(i)}$ 中每个像素定义为一个节点。由于采用 4 连通的连通域定义方法, 上下相邻或左右相邻的 2 个像素点 i, j 对应的 2 个节点 V_i, V_j 被视为存在一条边。同时, 由于图像中像素之间没有物理意义上的方向关系, 本文采用无向图来表示每一个候选对象, 在邻接矩阵 \mathbf{A} 中, $A_{i,j} = A_{j,i} = 1$, 其他将表示为 0。对于每个节点的特征, 最简单的方式是将 $I_{\text{cand}}^{(i)}$ 对应位置的像素值作为节点的特征, 以保证原始图像的灰度信息不丢失。同时, 将候选区提取算法结果图 I_{res} 对应像素位置的响应值也作为节点特征, 最终每个节点的特征维数为 2 维, 包括 1 维的灰度特征和 1 维的响应特征。为了聚合并编码所有的节点特征, 定义每个图一个虚拟的中心节点, 该节点与所有节点存在一条有方向的边, 方向为其他节点指向该中心节点。最终的图可描述为 $G^{(i)} = (\mathbf{V}, \mathbf{A})$, 其中 $\mathbf{V} \in \mathbb{R}^{(N+1) \times 2}$, $\mathbf{A} \in \{0, 1\}^{(N+1) \times (N+1)}$, N 表示连通域 $I_{\text{CR}}^{(i)}$ 的像素个数。

中心节点提供了一种有效的机制, 能够在图内编码候选区的拓扑特性, 并帮助图神经网络在图分类过程中处理来自不同像素的灰度信息。本文用全零向量初始化中心节点的特征。在整个图中,

仅有该节点存在单个方向。通过这种设计, 中心节点在更新节点特征时可以考虑所有其他节点, 而其他节点只考虑它们的相邻节点, 不受中心节点的影响。

1.3.2 基于图注意力网络的深度特征提取

本文采用了图注意力机制网络进行深度特征表示。图注意力机制网络由多层图注意力层组成。考虑一个已构建的图 $G = (\mathbf{V}, \mathbf{A})$, 它有一组节点特征 $\vec{h} = \{\vec{h}_1, \vec{h}_2, \dots, \vec{h}_{N+1}\}$, $\vec{h}_i \in \mathbb{R}^F$ 作为输入, 一层注意力层能产生一组新的节点特征, 如 $\vec{h}' = \{\vec{h}'_1, \vec{h}'_2, \dots, \vec{h}'_{N+1}\}$, $\vec{h}'_i \in \mathbb{R}^{F'}$, F' 为每个节点新特征的维度。这种学习过程的详细描述如下。

首先, 一个由权重矩阵 $\mathbf{W} \in \mathbb{R}^{F \times F'}$ 参数化的共享线性变换应用于每个节点。然后利用自注意力机制计算注意力系数 e_{ij} , 该系数表示节点 j 的特征对节点 i 的重要性。注意力系数 e_{ij} 的计算公式如下:

$$e_{ij} = \text{LeakyReLU}\left(\vec{a}^T \left[\mathbf{W} \vec{h}_i \parallel \mathbf{W} \vec{h}_j \right]\right) \quad (4)$$

式中, \parallel 表示连接操作; $\vec{a}^T \in \mathbb{R}^{2F'}$ 是一个参数化的权重向量, 由一个具有 LeakyReLU^[20] 非线性的全连接层实现 (负输入斜率为 0.2); $\mathbf{E} \in \mathbb{R}^{(N+1) \times (N+1)}$ 是注意力系数矩阵, 由 e_{ij} 构成。之后, 利用 softmax^[21] 函数对系数 e_{ij} 进行归一化, 使其在不同节点间具有可比性:

$$\alpha_{ij} = \text{softmax}_j(e'_{ij}) = \frac{\exp(e'_{ij})}{\sum_{k \in N_i} \exp(e'_{ik})} \quad (5)$$

式中, N_i 表示图中节点 i 的邻居节点, 即对于相邻节点 $j \in N_i$ 只考虑 e'_{ij} , 以避免引入任何不相关的节点。最后, 使用归一化关注系数 α_{ij} 计算相关特征的加权和, 作为每个节点的最终输出特征:

$$\vec{h}'_i = \text{ELU}\left(\sum_{j \in N_i} \alpha_{ij} \mathbf{W} \vec{h}_j\right) \quad (6)$$

式中, ELU (exponential linear unit) 表示指数线性单元非线性操作^[22]。根据 Velickovic P (2018)^[23], 本文采用了多头 (head) 注意来稳定自注意力的学习过程。

为避免过度平滑问题导致的性能下降, GAT 由 3 层图注意层组成 (层 1 输入维度为 2, 输出维度为 16, head=3; 层 2 输入维度为 48, 输出维度为 16, head=1; 层 3 输入维度为 16, 输出维度为 8, head=1)。在最后的特征汇聚阶段, 未对最后一层的所有节点特征进行平均, 而是将中心节

点的特征作为最终的集成特征。集成特征作为整个图 G^0 的深度特征表示,描述了图的全局拓扑结构和主题语义。本文最终的集成特征维数为 16 维。随后,通过接上 softmax 激活函数,并采用二分类交叉熵损失函数对整个图识别模型进行训练。训练完成后,去除 GAT 最后的分类层,将中心节点的 16 维特征输出,得到自动的深度图特征表示器。

1.4 候选区分类

为了结合深度表示学习和传统手工设计的各自优势,将 16 维深度特征和 16 维手工特征通过一个 4 层全连接网络进行特征融合并分类,如图 1 所示。网络的第 1 层接收前后连接的共 32 维深度特征与手工特征,第 2 层将特征编码融合为 24 维特征,第 3 层是一个包含 8 个单元的全连接层,用于增强网络的非线性能力,第 4 层是带有 softmax 激活函数的二分类输出层,通过激活可以得到候选区域属于红外小目标的概率。在使用深度特征时,将训练集上训练完成的 GAT 视为一个固定的特征提取器,每次输入数据时直接提取中心节点的 16 维特征作为输出,而无需在后续分类任务中重新训练。

2 实验

2.1 数据集

算法测试使用 3 个公开数据集,分别为 NUDT-SIRST^[24]、IRSTD-1k^[25]和 NUAA-SIRST^[25]。NUDT-SIRST 包含 1 327 幅图像(训练集 663 幅/测试集 664 幅),图像尺寸为 256×256 像素;IRSTD-1K 包含 1 001 幅图像(训练集 800 幅/测试集 201 幅),图像尺寸为 512×512 像素;NUAA-SIRST 则包含 427 幅图像(训练集 213 幅/测试集 214 幅)。这 3 个数据集覆盖城市、乡村、海域、密云等场景环境,具有斑驳云层、建筑边缘、高亮地面等干扰源。训练集主要用于离线构建特征库和训练识别器;在测试集上,首先采用候选区提取算法提取目标候选区,然后使用训练好的识别器进行识别,最后整合得到最终的检测结果。

2.2 评价指标

本文采用受试者工作特征曲线(receiver operating characteristics, ROC)^[26]来评估特征的重要性及候选区识别的性能。ROC 曲线以真阳率(true positive rate, TPR)为纵轴,假阳率(false positive rate, FPR)为横轴绘制而成。TPR 表示真实目标被

正确识别的概率,FPR 表示假目标被错误识别的概率。ROC 曲线下方的面积(area under roc curve, AUC)越大,表明模型在较低误报率的情况下具有较高的灵敏度,对目标和虚警的区分能力更强。一般而言,AUC 值大于 0.9 才能表示模型的性能较好。

采用目标检出率^[27](probability of detection, PD)和虚警率(false-alarm rate, FA)对最终检测结果进行评价。PD 是一个目标级别的评价指标,表示正确检测的目标数量占有所有目标数量的比例。FA 是一个像素级别的指标,表示误检测的像素数量占图像总像素数量的比例。目标检出率 PD 和虚警率 FA 的计算分别为:

$$PD = \frac{N_{\text{detected}}}{N_{\text{target}}}, \quad FA = \frac{P_{\text{false}}}{P_{\text{all}}} \quad (7)$$

式中, N_{detected} 和 N_{target} 分别表示正确检测的目标数量和图像中的目标总数; P_{false} 和 P_{all} 分别表示假正类的像素个数和图像像素总数。PD 越高表示目标检出的性能越好,FA 越低表示算法的虚警越少。

2.3 实验配置

基于 GAT 深度特征提取阶段,输入为候选区域转化成的图结构数据集,包括 6 218 个图结构,900 个正样本个数,5 461 个负样本。学习策略采用 adam 优化器,初始学习率为 0.01,训练 2 000 epoch, batch size 为 128,为归一化,每 200 epoch 学习率降低为上一阶段的 0.1。完成网络训练后,将候选区域提取出的样本正向通过网络,并将中心节点的 16 维特征向量作为深度特征,与 16 维手工特征结合,采用多层感知器构建识别网络并训练,训练参数与 GAT 训练阶段相同。

2.4 实验结果

2.4.1 候选区提取算法

不同场景下候选区域提取结果如图 2 所示。从左往右为测试数据集中的 5 张典型图像,真实弱小目标被红色框框出,且在左下角放大。从上到下分别为原始红外图像、Tophat 结果、候选区域二值图像。Tophat 结果表示该算法对背景杂波也会产生一定的响应,但真实目标均被检出,通过连通域分析可筛除一部分大面积虚警,得到最终的候选区域。该算法在测试集上的检出率达到 1.0,平均每张图像的负样本个数为 11.8,正负样本比率约为 1:10。尽管最终结果依然存在大量误检,但高检出率和运算较快的特点非常有利于后续的候选区

分类。

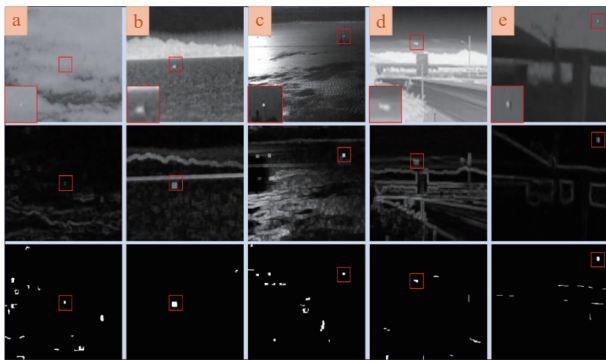


图 2 候选区域提取结果示意图

2.4.2 特征提取与分类算法

为了全面评估不同特征表示的性能差异, 本文设计了一套系统的消融实验方案, 重点对比分析了 3 类特征表示: 手工特征 (hand-crafted feature, HCF)、基于 GAT 的深度特征 (deep feature, DPF) 以及二者的联合表示 (HCF+DPF)。实验采用五折交叉验证策略, 每组特征均获得 5 组独立实验结果, 通过 ROC 曲线及其置信区间展示模型性能的稳定性, 并以平均 AUC 值作为最终评价指标。实验结果如图 3 所示, 3 类特征表示的平均 AUC 值分别为: HCF (0.876 8)、DPF (0.933 8) 和联合特征 (0.971 5)。值得注意的是, DPF 和联合特征的性能优于传统手工特征。从 ROC 曲线的波动范围来看, DPF 表现出更好的稳定性 (AUC 波动范围 ± 0.015), 而 HCF 的 AUC 值波动较大 (0.45~0.89), 这充分证明了基于图表示学习方法的特征具有更强的鲁棒性。如表 2 所示, 深度特征和联合特征在准确率、F1-score 等各项指标均展现出明显优势。

在计算效率方面, 由于候选区提取以及后续基

于全连接层识别的时间消耗很少, 本文主要记录了单个候选区的特征提取时间。虽然单个候选区的图特征提取时间与传统特征提取相比增加了 1~2 倍 (主要来自图结构的表示), 但带来的虚警抑制能力和检测稳定性明显增强。考虑到实际应用中候选区数量有限, 这种时间代价完全可接受。

为了深入理解不同特征表示的可分性, 本文采用 t-分布随机邻域嵌入 (t-distributed stochastic neighbor embedding, t-SNE)^[28] 技术将高维特征降至二维空间进行可视化分析, 如图 4 所示。其中, 蓝色点为虚警源, 红色点为目标。从图中可以发现 HCF 特征空间中存在大量重叠区域, DPF 特征空间中目标与虚警的分离度显著提高, 仅剩少量重叠, 而联合特征空间展现出最佳的可分性, 两类样本几乎完全分离。

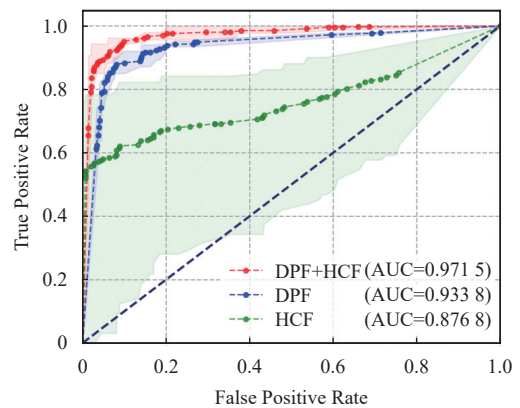


图 3 ROC 曲线

表 2 基于不同特征的候选区分类算法性能比较

特征	AUC	准确率	F1-score	Time/ms
HCF	0.876 8	0.935 0	0.913 9	2.13
DPF	0.933 8	0.945 0	0.929 8	3.18
HCF+DPF	0.971 5	0.970 0	0.944 4	5.39

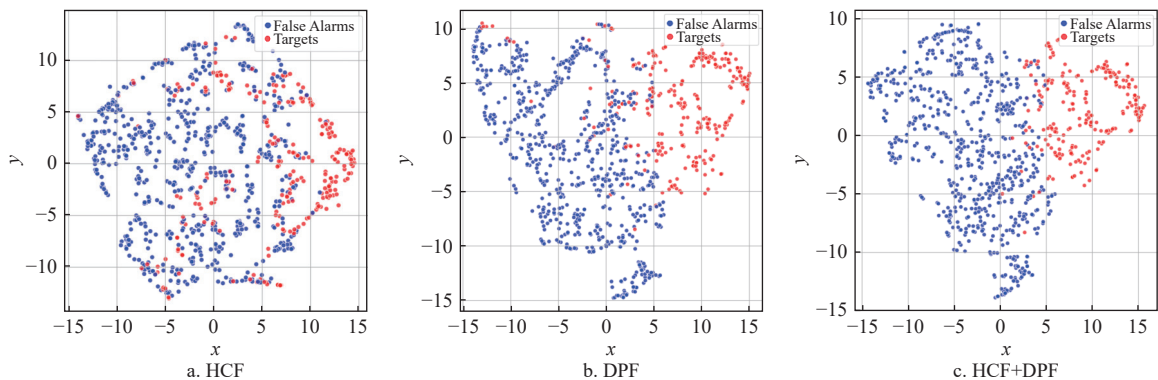


图 4 各类特征的 t-SNE 可视化

进一步地, GAT 各层特征的 t-SNE 可视化结果如图 5 所示, 随着网络深度的增加, 特征的判别能力逐步增强: 第 1 层 GAT 仅能区分少量目标,

而经过 3 层 GAT 后, 大多数目标都能与虚警源有效分离。这一现象验证了本文设计的图表示学习方法的有效性。

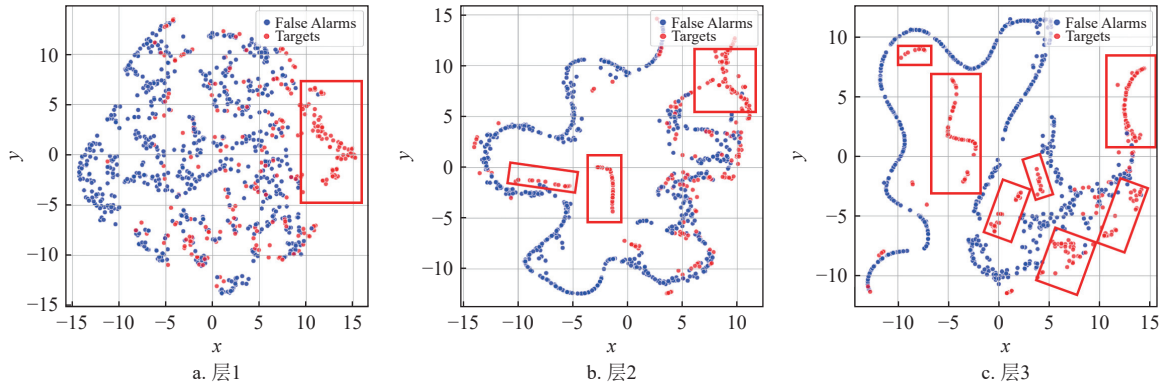


图 5 各层 GAT 特征的 t-SNE 可视化

2.4.3 最终检测算法

表 3 展示了本文算法与现有方法在 NUDT-SIRST、IRSTD-1K 和 NUAA-SIRST 这 3 个公开数据集上的对比实验结果。参与对比的算法包括基于局部对比度的方法 (LCM^[1]、MPCM^[3])、基于低

秩稀疏分解的方法 (IPI^[5]、NRAM^[6]、RIPT^[7]、PSTNN^[8]、MSLSTIPT^[25])、基于 CNN 的方法 (ACM^[10]、UIUNet^[12]、MDIGCNet^[29]、RDIAN^[30]、ALCNet^[31]、RPCANet^[32]) 以及基于候选区分类的方法 (LCI^[13])。

表 3 不同方法在目标检出率与虚警率上的性能比较

方法	NUDT-SIRST		IRSTD-1K		NUAA-SIRST	
	目标检出率 ($\times 10^{-2}$)	虚警率 ($\times 10^{-6}$)	目标检出率 ($\times 10^{-2}$)	虚警率 ($\times 10^{-6}$)	目标检出率 ($\times 10^{-2}$)	虚警率 ($\times 10^{-6}$)
LCM ^[1]	56.82	1 309	67.12	1 237	77.89	123.8
MPCM ^[3]	66.14	116.0	68.73	65.12	80.06	51.43
IPI ^[5]	74.49	41.23	72.28	37.28	83.27	38.92
NRAM ^[6]	56.4	19.27	71.44	20.12	74.52	<u>13.85</u>
RIPT ^[7]	91.85	344.3	91.15	161.3	79.08	22.61
PSTNN ^[8]	66.13	44.17	89.14	23.79	77.95	29.11
ACM ^[10]	95.97	10.18	90.57	92.64	91.89	28.51
ALCNet ^[31]	<u>96.51</u>	9.261	92.58	69.45	96.57	30.47
RPCANet ^[32]	96.14	28.73	89.35	43.90	95.62	47.21
MDIGCNet ^[29]	96.67	37.28	91.25	40.38	93.65	56.84
LCI ^[13]	90.64	9.833	91.35	<u>12.26</u>	92.38	14.27
RDIAN ^[30]	94.51	22.81	91.57	36.44	94.42	58.58
UIUNet ^[12]	95.12	<u>7.91</u>	90.91	20.21	92.40	18.99
本文方法	95.28	6.95	<u>92.16</u>	9.12	<u>95.78</u>	10.03

表 3 中, 最优检测性能被加粗显示, 次优性能用下划线表示。结果表明, 基于局部对比度的方法在目标检出率和虚警率方面表现欠佳, 其中 LCM 的虚警率尤为突出; 基于低秩稀疏分解的方法性能优于局部对比度方法, 其中 RIPT 在目标检出方面表现较优但虚警率较高, 而 NRAM 在 NUAA-SIRST 数据集上取得了次优的虚警率结果; 基于深度学习的方法显著优于传统模型驱动算

法, 在 NUDT-SIRST 数据集上的目标检出率均超过 94%, 其中 ALCNet 在 3 个数据集上的检出率分别达到 96.51%、92.58%、96.57%, 取得了次优和最优的检出性能, 但可能由于全黑盒特征建模的方式导致其虚警率表现不稳定 (9.261×10^{-6} 、 69.45×10^{-6} 、 30.47×10^{-6}); 基于候选区分类的 LCI 方法通过手工特征建模实现了稳定的虚警率控制 (均稳低于 15×10^{-6}), 但依然存在优化空间; 本文方

法同样属于候选区分类方法, 但创新性地融合了手工设计特征与图表示学习的优势, 在保持较高检出率(95.28%、92.16%、95.78%)的同时, 取得了最优的虚警率表现(6.95×10^{-6} 、 9.12×10^{-6} 、 10.03×10^{-6})。综合来看, 本文算法在目标检出率和虚警率控制方面实现了最佳平衡。

为了直观评估不同算法在复杂场景下的目标检测性能, 本文对5种典型背景(天空、地面、建筑等)的示例图像进行了多算法对比实验, 并将MPCM、IPI、PSTNN、ACM、RDIAN、LCI以及UIUNet等算法的检测结果进行可视化分析, 如图6所示。真实目标采用红框标注并辅以左下角局部放大显示, 虚警和漏检目标分别用黄色框和蓝色框标记。实验结果表明: 在天空背景的示例图6a和图6b中, 传统算法MPCM、IPI和PSTNN均存在背景杂波残留问题, 其中IPI和PSTNN还表现出明显的暗区噪声干扰; 对于高亮地面背景的示例图6c,

上述3种传统算法均未能有效检测目标。在深度学习算法中, ACM虽然在所有测试场景中都能有效抑制背景噪声, 但对示例图6a中微弱目标的检测敏感性不足; RDIAN和UIUNet则出现明显的漏检现象, 这可能与模型过拟合有关。在建筑背景的示例图6d中, 由于存在高亮干扰源, 大多数算法(MPCM、IPI、PSTNN、ACM、LCI)均产生了不同程度的虚警。值得注意的是, 基于候选区分类的LCI方法通过目标知识建模能够检测到所有示例中的真实目标, 但因虚警特征表示不足仍存在少量杂波干扰。本文算法在所有测试图像中均表现优异: 不仅实现了背景杂波的完全抑制, 同时有效增强了目标显著性, 展现出最优的综合检测性能。这一结果验证了手工设计与表示学习联合特征建模策略在复杂场景目标检测中的有效性。

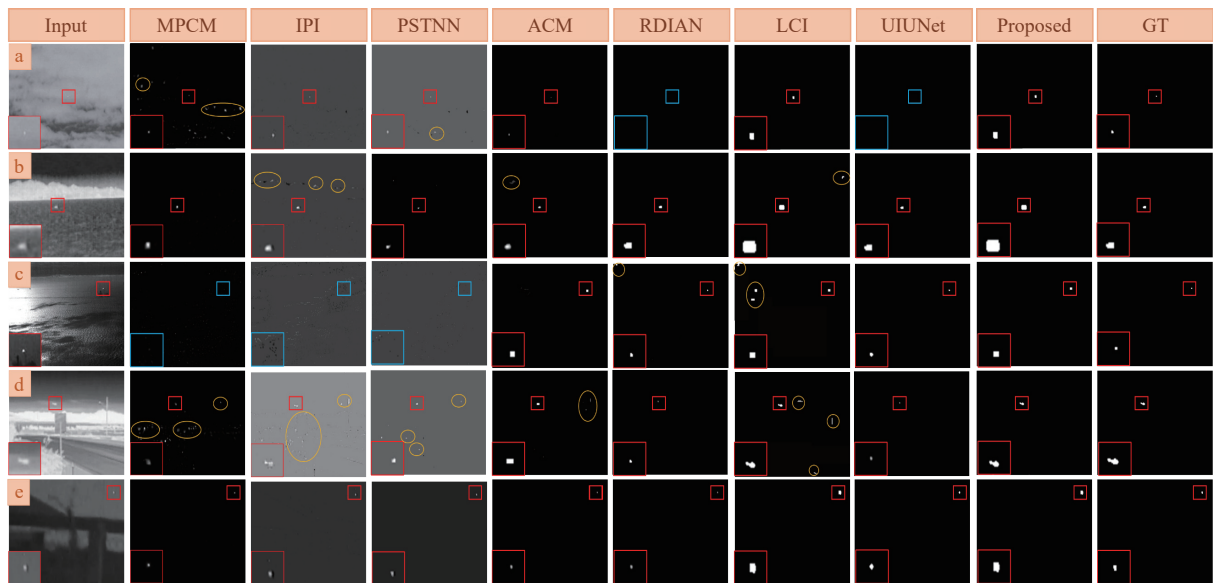


图6 检测结果可视化

3 结束语

本文提出了一种结合图表示学习和多特征融合的红外弱小目标检测算法, 以提高复杂背景中红外弱小目标的检测精度和鲁棒性。算法采用图表示学习方法实现目标和不规则虚警源的协同建模和深度特征表示, 最后联合传统特征实现目标的精确检测。在未来的工作中, 将进一步优化异构候选区的图结构构建和图学习网络结构, 提升算法的检测率和鲁棒性。

参考文献

- [1] CHEN C, LI H, WEI Y, et al. A local contrast method for small infrared target detection[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2014, 52(1): 574-581.
- [2] HAN J, MA Y, ZHOU B, et al. An improved local contrast method for small infrared target detection[J]. *Infrared Physics & Technology*, 2017, 81: 202-209.
- [3] QIN Y, LI B. Effective infrared small target detection utilizing a novel local contrast method[J]. *IEEE Geoscience and Remote Sensing Letters*, 2015, 13(12): 1890-1894.
- [4] WEI Y, YOU X, LI H. Multiscale patch-based contrast measure for small infrared target detection[J]. *Pattern Recognition*, 2016, 58: 216-226.
- [5] GAO C, LI L, LI W, et al. Infrared patch-image model for

- small target detection in a single image[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2013, 51(2): 891-903.
- [6] ZHANG L, PENG L, ZHANG T, et al. Infrared small target detection via non-convex rank approximation minimization joint l_2, l_1 norm[J]. *Remote Sensing*, 2018, 10: 1821.
- [7] DAI Y, WU Y. Reweighted infrared patch-tensor model with both non-local and local priors for single-frame small target detection[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2017, 10(8): 3752-3767.
- [8] ZHANG L, PENG Z. Infrared small target detection based on partial sum of the tensor nuclear norm[J]. *Remote Sensing*, 2019, 11(4): 382.
- [9] LI J, ZHANG P, ZHANG L Y, et al. Sparse regularization-based spatial-temporal twist tensor model for infrared small target detection[EB/OL]. [2025-01-10]. <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10007842&tag=1>.
- [10] DAI Y, WU Y, ZHOU F, et al. Asymmetric contextual modulation for infrared small target detection[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Pasadena: IEEE, 2021: 950-959.
- [11] HOU Q, ZHANG L, TAN F, et al. ISTDU-Net: Infrared small-target detection U-Net[J]. *IEEE Geoscience and Remote Sensing Letters*, 2022, 19: 7506205.
- [12] WU X, HONG D. UIU-Net: U-Net in U-Net for infrared small object detection[J]. *IEEE Trans Image Process*, 2023, 20: 1-5.
- [13] CAO S, DENG J, LUO J, et al. Local convergence index-based infrared small target detection against complex scenes[J]. *Remote Sensing*, 2023, 15(5): 1464.
- [14] LI Y, LI Z, LI J, et al. Robust small infrared target detection using weighted adaptive ring top-hat transformation[J]. *Signal Processing*, 2024, 217: 14.
- [15] YANG F, PENG D. Spatio-temporal contrastive heterogeneous graph attention networks for session-based recommendation[J]. *Mathematics*, 2024, 12(8): 16.
- [16] PENG Y, GONG X, LU H, et al. Human visual pathways for action recognition versus deep convolutional neural networks: Representation correspondence in late but not early layers[J]. *Journal of Cognitive Neuroscience*, 2024, 36(11): 23.
- [17] LIU H. Research on the construction method of curriculum teaching knowledge graph based on Bi-LSTM and CNN algorithm[J]. *International Journal of High Speed Electronics and Systems*, 2025, 34(3): 2540029.
- [18] ZHONG S S, ZHOU H B, XU C, et al. A local-patch and global attention network with multiscale asymmetry fusion for infrared small-target detection[J]. *Measurement Science and Technology*, 2025, 36(6): 065412.
- [19] WANG P, WANG J, CHEN Y, et al. Paying more attention to local contrast: Improving infrared small target detection performance via prior knowledge[J]. *Engineering Applications of Artificial Intelligence*, 2025, 146: 110244.
- [20] NAIR V, HINTON G E. Rectified linear units improve restricted boltzmann machines[C]//International Conference on Machine Learning. [S.l.]: ACM, 2010: 807-814.
- [21] WANG Y Y, GU G, LIU J L, et al. Double softmax attention backbone for visual tracking by unifying feature extraction and fusion[J]. *Signal, Image and Video Processing*, 2025, 19(10): 865.
- [22] CLEVERT D A, UNTERTHINER T, HOCHREITER S. Fast and accurate deep network learning by exponential linear units (ELUs)[C]//International Conference on Learning Representations. San Juan: [s.n.], 2016: 1-13.
- [23] VELICKOVIC P, CUCURULL G, CASANOVA A, et al. Graph attention networks[C]//International Conference on Learning Representations (ICLR). Vancouver: [s.n.], 2018: 1-13.
- [24] LI B Y, XIAO C, WANG L G, et al. Dense nested attention network for infrared small target detection[J]. *IEEE Transactions on Image Processing*, 2023, 32: 1745-1758.
- [25] ZHANG M J, ZHANG R, YANG Y X, et al. ISNet: Shape matters for infrared small target detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2022: 867-876.
- [26] FAWCETT T. An introduction to ROC analysis[J]. *Pattern Recognition Letters*, 2006, 27(8): 861-874.
- [27] PAN L H, LIU T, CHENG J H, et al. AIMED-net: An enhancing infrared small target detection net in UAVs with multi-layer feature enhancement for edge computing[J]. *Remote Sensing*, 2024, 16(10): 1776.
- [28] SHI K, WANG W Z, LIU Q C, et al. Topological information aggregation network for few-shot cross-domain hyperspectral image classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2025, 63: 1-15.
- [29] ZHANG L P, LUO J H, HUANG Y A, et al. MDIGCNet: Multidirectional information-guided contextual network for infrared small target detection[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2025, 18: 2063-2076.
- [30] SUN H, BAI J X, YANG F, et al. Receptive-field and direction induced attention network for infrared dim small target detection with a large-scale dataset IRDST[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 1-13.
- [31] DAI Y, WU Y, ZHOU F, et al. Attentional local contrast networks for infrared small target detection[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 99: 1-12.
- [32] WU F Y, ZHANG T F, LI L, et al. RPCANet: Deep unfolding RPCA based infrared small target detection[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. [S.l.]: IEEE, 2024: 4809-4818.