

难点注意力感知红外小目标检测网络

王伯霄 宋延嵩 董小娜

Indistinguishable points attention-aware network for infrared small object detection

WANG Bo-xiao, SONG Yan-song, DONG Xiao-na

引用本文:

王伯霄, 宋延嵩, 董小娜. 难点注意力感知红外小目标检测网络[J]. *中国光学*, 2024, 17(3): 538–547. doi: 10.37188/CO.2023–0178

WANG Bo-xiao, SONG Yan-song, DONG Xiao-na. Indistinguishable points attention-aware network for infrared small object detection[J]. *Chinese Optics*, 2024, 17(3): 538–547. doi: 10.37188/CO.2023–0178

在线阅读 View online: <https://doi.org/10.37188/CO.2023–0178>

您可能感兴趣的其他文章

Articles you may be interested in

轻量型YOLOv5s车载红外图像目标检测

Lightweight YOLOv5s vehicle infrared image target detection

中国光学 (中英文). 2023, 16(5): 1045 <https://doi.org/10.37188/CO.2022–0254>

基于双注意力机制的车道线检测

Lane detection based on dual attention mechanism

中国光学 (中英文). 2023, 16(3): 645 <https://doi.org/10.37188/CO.2022–0033>

融合视觉显著性和局部熵的红外弱小目标检测

Infrared dim small target detection based on visual saliency and local entropy

中国光学 (中英文). 2022, 15(2): 267 <https://doi.org/10.37188/CO.2021–0170>

基于 L_{12} 时空域总变分正则项的红外弱小目标检测算法

Infrared small target detection via L_{12} spatial–temporal total variation regularization

中国光学 (中英文). 2023, 16(5): 1066 <https://doi.org/10.37188/CO.2022–0229>

基于深度学习的空间脉冲位置调制多分类检测器

Spatial pulse position modulation multi–classification detector based on deep learning

中国光学 (中英文). 2023, 16(2): 415 <https://doi.org/10.37188/CO.2022–0106>

氨气泄漏混洗自注意力轻量化红外检测

Lightweight infrared detection of ammonia leakage using shuffle and self–attention

中国光学 (中英文). 2023, 16(3): 607 <https://doi.org/10.37188/CO.2022–0127>

文章编号 2097-1842(2024)03-0538-10

难点注意力感知红外小目标检测网络

王伯霄¹, 宋延嵩^{1,2*}, 董小娜^{1,2}

(1. 长春理工大学 光电工程学院, 吉林 长春 130000;
2. 长春理工大学 空间光电技术研究所, 吉林 长春 130000)

摘要:随着飞行器机动性能的提升,多帧红外小目标检测方法不足以满足检测要求。近年来,基于深度学习的单帧红外小目标检测方法取得了巨大成功。然而,红外小目标通常缺少形状特征,而且边界与背景模糊不清,给准确分割带来了一定的挑战。针对上述问题,本文提出难点注意力感知红外小目标检测网络。通过基于点的区域建议模块获取目标潜在区域,同时滤除多余背景。为实现高质量分割、细化掩码边界模块、判断粗掩码中无序、非局部难以分辨点,融合这些难点的多尺度特征,进行逐像素注意力建模。最后,由点检测头对难点注意力感知特征重新预测,生成精细分割掩码。在公开数据集 NUDT-SIRST 和 IRDST 上进行测试,平均精度均值 mAP 达到 87.4 和 63.4, F 值达到 0.8935 和 0.7056。本文提出的难点注意力感知红外小目标检测网络可在多检测场景、多目标形态下实现准确分割,抑制误报信息,同时控制计算开销。

关键词:目标检测;深度学习;红外成像;红外小目标检测;注意力机制

中图分类号:V19;E928.9;TN911.73

文献标志码:A

doi:10.37188/CO.2023-0178

Indistinguishable points attention-aware network for infrared small object detection

WANG Bo-xiao¹, SONG Yan-song^{1,2*}, DONG Xiao-na^{1,2}

(1. *The School of Electro-Optical Engineering, Changchun University of Science and Technology, Changchun 130000, China;*
2. *Institute of Space Photoelectronic Technology, Changchun University of Science and Technology, Changchun 130000, China*)

* *Corresponding author, E-mail: songyansong2006@126.com*

Abstract: As aircraft maneuverability increases, multi-frame infrared small target detection methods are becoming insufficient to meet detection requirements. In recent years, significant progress has been achieved in single-frame infrared small-target detection method based on deep learning. However, infrared small targets often lack shape features and have blurred boundaries and backgrounds, obstructing accurate segmentation.

收稿日期:2023-10-11; 修订日期:2023-10-30

基金项目:国家重点研发计划(No. 2022YFB3902505); 国家自然科学基金重点项目(No. U2141231); 国家自然科学基金(No. 62305032)

Supported by National Key Research and Development Program (No. 2022YFB3902505); Key Project of National Natural Science Foundation of China (No. U2141231); National Natural Science Foundation of China (No. 62305032)

According to the problems, an indistinguishable points attention-aware network for infrared small object detection was proposed. First, potential target areas were acquired through a point-based region proposal module while filtering out redundant backgrounds. Then, to achieve high-quality segmentation, the mask boundary refinement module was utilized to identify disordered, non-local indistinguishable points in the coarse mask. Multi-scale features of these difficult points were then fused to perform pixel-wise attention modeling. Finally, A fine segmentation mask was generated through re-predicting the indistinguishable points attention-aware features by point detection head. The mAP of the proposed method reached 87.4 and 63.4 on the publicly available datasets NUDT-SIRST and IRDST, and the F-measure reached 0.893 5 and 0.705 6, respectively. It can achieve accurate segmentation in multi-detection scenarios and multi-target morphology, suppressing false alarm information while controlling the computational overhead.

Key words: object detection; deep learning; infrared imaging; infrared small object detection; attention mechanism

1 引言

红外传感器在低照明环境或恶劣大气条件下能够捕获目标的辐射热量^[1], 故成为快预警系统^[2]、精确制导^[3]、地面监控等领域不可或缺的重要组成部分。然而, 远距离目标在红外图像中仅占几个像素大小、缺乏明显的形状特征、易埋在复杂低空背景中^[4], 故探测红外小目标是项具有挑战的任务。

随着飞行器机动性增强, 目标在图像序列中的轨迹变化较大, 这严重影响基于连续帧方法检测效果^[5]。基于单帧检测方法继而受到广泛关注。传统单帧检测方法大多数依赖目标与邻域的局部对比度信息, 即寻找缓慢过渡背景中灰度不连续的异常值。LCM(Local Contrast Method)^[6]将红外目标所处位置理解为邻域灰度值对比度最大的中心点。LEF(Local Energy Factor)^[7]在LCM基础上添加局部相异性, 丰富局部差异描述。TLLCM(Tri-layer Local Contrast Measure)^[8]提出核心层、保留层和周围层过滤窗口, 分别有目的地增强核心与背景的对比度, 提取多尺度目标。然而, 这些传统方法通常基于人为的先验假设来预测红外小目标, 无法准确分辨具有相似预设特征的背景杂波, 易产生误报。此外, 这类算法对预设窗口大小、分割阈值等超参数十分敏感, 需在特定场景下进行调整。当面对结构复杂的非点目标时, 难以精确预测分割边界, 泛化性有限。

受监督的深度学习方法由已标注数据驱动,

可以根据设定的损失函数修正特征参数, 在目标检测任务中具有极佳的鲁棒性^[9]。在面对深度学习检测常规小目标时产生的像素少、特征少以及目标背景实例不平衡等问题, 研究人员提出4种解决方法: (1) 多特征图融合^[10-11]; (2) 添加上下文信息^[12-14]; (3) 平衡类示例^[13]; (4) 超分辨率^[15]。基于深度学习的红外小目标检测方法也相继被提出。ALCNet(Attentional Local Contrast Networks)^[16]和DNANet(Dense Nested Attention Network)^[17]着重关注小目标上下文信息与高级特征融合。ALCNet运用自下而上的注意力调节来突出和保留小目标特征, 并运用循环移位加速方案进行长距离信息交互。DNANet提出了密集嵌套交互模块和通道空间注意模块来实现渐进特征融合和自适应特征增强。IAANet(Interior Attention-Aware Network)^[18]提出一种Two-stage分割方法, 利用RPN(Region Proposal Network)^[19]结构获取目标边框提取特征图, 最后由Transformer结构^[20]对特征图编码以获取感知特征。

受到上述深度学习检测方法的启发, 本文提出了一种基于难点注意力感知的红外小目标检测网络, 旨在解决红外弱小目标的边界与背景模糊不清, 从而导致网络分割边界不够准确的问题。首先采用简单且高效的基于点的区域建议模块来稀疏提取目标潜在区域, 并滤除冗余背景。随后, 使用细化掩码模块对建议区域进行进一步精细化处理。对深层特征预测的粗糙掩码进行逐级上采样, 融合上采样过程中难以分辨点对应的多尺度特征, 并嵌入位置编码获取难点位置信息。接下

来, 细化模块对逐个难点进行内部注意力建模, 使用点检测头对注意力感知特征单独预测, 从而生成高分辨率精细预测掩码。本文的主要贡献如下:

(1) 利用基于点的区域建议模块, 对红外小目标的潜在区域进行裁剪, 实现了高效粗检测。

(2) 提出了细化掩码边界模块, 对分割难点注意力进行建模, 显著提升了检测结果。

在 NUDT-SIRST^[17] 和 IRDST^[21] 数据集进行了实验分析。实验结果表明, 本文方法相较于现有红外小目标检测算法性能优越。此外, 还进行了消融实验, 以证明基于点的区域建议模块与细化掩码边界模块对网络整体性能的提升效果。

2 难点注意力感知网络

本文提出一种高质量红外小目标分割方法, 其整体算法结构如图 1(彩图见期刊电子版)所示。采用常规目标检测方法中的 FPN(Feature Pyramid Network)结构作为骨干网络。通过基于点的区域建议模块预测中心点热图, 以截取红外小目标所在区域, 同时利用该区域低分辨率特征预测初始粗糙掩码。利用细化掩码边界模块选择粗糙掩码上采样时的无序、非局部难点, 融合难点细粒度特征后嵌入位置信息。然后, 将其展平为特征序列, 进行注意力建模, 以捕获各难点内部关系。最后重新解码得到高分辨率的精细化分割图。

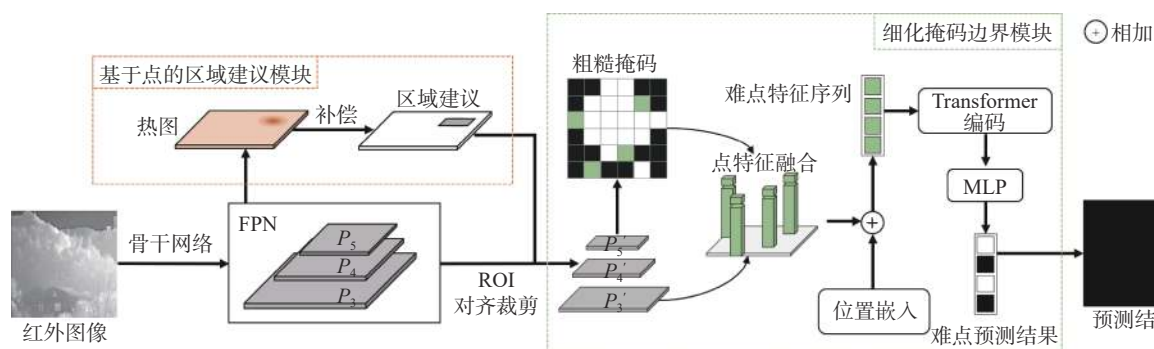


图 1 难点注意力感知网络结构图

Fig. 1 Indistinguishable points attention-aware network architecture

2.1 基于点的区域建议模块

为了准确高效地提取红外小目标潜在区域, 并滤除冗余背景。本文通过目标中心点实施区域建议。此模块分为中心点预测分支和补偿分支, 具体结构如图 2 所示。

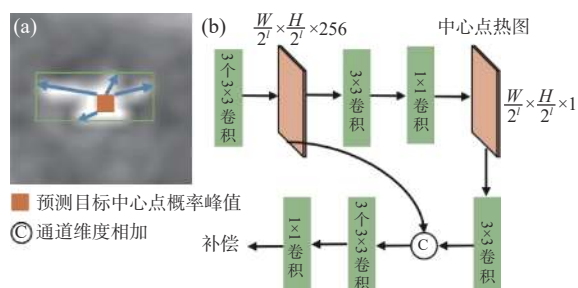


图 2 (a) 中心点补偿边界; (b) 基于点的区域建议模块
Fig. 2 (a) Centre point offset to boundary; (b) point-based region proposal module

在中心点预测分支中, 将预测中心点坐标视为关键点预测的一种形式^[22]。设一张宽为 W 、高

为 H 、颜色通道数为 1 的红外图像 $I \in \mathbf{R}^{W \times H \times 1}$ 。本模块以 FPN 层级 P_l 为输入, 预测热图 $\hat{V}_l \in [0, 1]^{(W/2^l) \times (H/2^l) \times 1}$, 其中 2^l 为骨干网络输出特征图步长, 1 为预测类别: 红外小目标。 $\hat{V}_l^{i,j}$ 表示各像素坐标 (i, j) 为红外小目标中心的概率, 将其视为检测置信度, 取前 128 个 3×3 邻域峰值视为目标中心位置, 并在后文分析了中心点数量给检测性能带来的影响。为了缓解由于小目标数目过少而导致的类别不平衡问题, 使用焦点损失 L_{heat} 监督热图的逐像素生成^[23]。

在补偿分支中, 直接基于中心点回归补偿位移量, 表示为 $\{(\Delta x_k, \Delta y_k)\}_{k=1}^N$, 其中 N 为目标顶点, 表示所需点数。考虑到红外小目标通常相对较小, 无需设置过多顶点来表示目标轮廓, 在本实验中, 默认将 N 设为 4。将中心点坐标加补偿位移量作为目标顶点, 表示为 $\{(x_k, y_k)\}_{k=1}^N$ 。在 RepPoints^[24] 方法中, 预测 9 个无序采样点集偏移量。但在本

文实验中,一些目标仅占用 2×2 个像素,无法提供足够的采样点。因此,本文实验直接基于中心点回归轮廓表示,可以视为RepPoints的一种特殊设置。通过回归得到的顶点坐标用于确定区域建议的边界:中心坐标以及宽度和高度:

$$\mathcal{B} = \left(\frac{x^+ + x^-}{2}, \frac{y^+ + y^-}{2}, x^+ - x^-, y^+ - y^- \right), \quad (1)$$

其中 x^+ 、 x^- 分别为顶点横坐标最大值、最小值,纵坐标同理,如图2(a)所示。通过计算 \mathcal{B} 和真实边界左上角和右下角点之间的平滑 L_1 距离^[19]表示补偿分支损失 L_{offset} 。

训练时,目标中心点的真值由张量 $\mathbf{V}_l \in [0, 1]^{W \times H \times 1}$ 表示,位于真值中心点 3×3 邻域内的像素都视为正样本:目标,即 $\mathbf{V}_l^{i,j} = 1$;否则为负样本:背景,即 $\mathbf{V}_l^{i,j} = 0$ 。只有预测坐标位于正样本像素内时,才会对位移补偿进行预测。

两个分支各自通过4个 3×3 卷积层和检测头进行预测。每个卷积层都连接ReLU(Rectified Linear Unit)激活函数,各自的检测头为 1×1 卷积^[25]。补偿分支经过第一个卷积后,和中心点分支中间特征进行通道维度相加,丰富感知信息,具体结构如图2(b)所示。

传统4维边界框表示方法仅考虑目标在空间中的矩形坐标,而没有考虑局部空间内形状和语义上的相对位置信息,本文方法可以更好地为后续掩码分割提供对象特征信息。

基于点的区域建议模块损失函数为:

$$L_r = L_{\text{heat}} + L_{\text{offset}}. \quad (2)$$

2.2 细化掩码边界模块

细化掩码边界模块仅在建议区域内进行掩码预测。首先,通过ROI对齐剪裁后的低分辨率特征图 \mathbf{P}_5 生成等分辨率粗糙掩码。为了获得高分辨率预测掩码,需要逐级进行双线性插值上采样。然而,这个过程难免会产生较多的误报像素。为了在推理时不消耗太多计算资源,本模块只对上采样过程中少量难以分辨的点进行重新预测。选择 M 个二值化掩码预测概率最接近0.5的点作为难点。由于小目标占用较少像素,本实验中仅选择 $M = 8^2$ 个难点。设粗糙掩码分辨率为 $W_0 \times H_0$,若想得到图像原始分辨率 $W \times H$ 的分割掩码,只需计算 $M \log_2 \sqrt{W/W_0 \times H/H_0}$ 个点,这远

小于对整张图像进行预测的数量。训练时,采用过度采样策略,以形成足够多的正负参考。具体做法是,过度生成 kM 个均匀分布的点($k > 1$),对这些点估计不确定性,然后从过度生成的点中选择最不确定的 γM 个点($\gamma \in [0, 1]$),而其余 $(1 - \gamma)M$ 个点从均匀分布中采样,故 γ 为调节选点策略的超参数。本文中设置 $k = 3$, $\gamma = 0.75$,并在后文分析了不同选点策略对检测结果的影响。预测和损失函数都作用在采样的难点上,使整个过程简单高效。

为了丰富难以分辨点的特征,在选点位置提取以下3种特征进行融合:

(1) 初始粗糙掩码 $\mathbf{P}_{\text{coarse}}$ 丰富上下文和高级语义信息:

$$\mathbf{P}_{\text{coarse}} = \sigma(\delta(\text{MLP}(\delta(\text{CONV}_{1 \times 1}(\mathbf{P}_5))))), \quad (3)$$

其中 $\text{CONV}_{1 \times 1}$ 、 δ 、 MLP 、 σ 分别表示 1×1 卷积、ReLU激活函数、多层感知器和Sigmoid函数。 MLP 逐像素预测粗糙掩码。

(2) 细粒度特征 \mathbf{P}_{fine} 可提供分割细节信息:从难点所对应的FPN层级提取特征向量,可以是单层特征 \mathbf{P}_3 层,也可是多层特征 $\mathbf{P}_3 \sim \mathbf{P}_5$ 层。

(3) 相对位置编码 \mathbf{P}_{pos} 使距离相近的像素具有更相似的位置嵌入,可获取难点之间的局部依赖关系和相关信息。

难点对应位置的3种特征信息融合公式如下:

$$\mathbf{P}_{\text{point}} = \text{FC}(\text{Concat}[\mathbf{P}_{\text{coarse}}, \mathbf{P}_{\text{fine}}]) + \mathbf{P}_{\text{pos}}, \quad (4)$$

其中Concat表示通道维度相加,FC表示全连接层,用作维持原始特征通道维度 C 。

Transformer结构在目标检测领域已经取得了不错的成绩^[26]。本文利用Transformer对难点特征序列编码,利用注意力机制整合所有被选点的特征信息,以建立无序点之间的内部关系,这对重新预测非顺序排列的难点至关重要。设映射矩阵为 $\mathbf{U}^{\text{QKV}} \in \mathbf{R}^{C \times C_{\text{in}}}$,将 $\mathbf{P}_{\text{point}}^{s-1}$ 映射于3个可学习矩阵: $\mathbf{Q}_t^{s-1} \in \mathbf{R}^{C_{\text{in}} \times M}$; $\mathbf{K}_t^{s-1} \in \mathbf{R}^{C_{\text{in}} \times M}$; $\mathbf{V}_t^{s-1} \in \mathbf{R}^{C_{\text{in}} \times M}$,其中 $s = 1, \dots, S$ 表示第 s 个编码阶段, $t = 1, \dots, T$, T 为注意力头数量。则注意力矩阵计算公式为:

$$\mathbf{A}_t^{s-1} = \text{softmax} \left(\frac{\mathbf{Q}_t^{s-1} (\mathbf{K}_t^{s-1})^T}{\sqrt{C_{\text{in}}}} \right) \mathbf{V}_t^{s-1}, \quad (5)$$

其中softmax()为softmax函数, $\sqrt{C_{\text{in}}}$ 为避免

$Q_i^{s-1} (K_i^{s-1})^T$ 运算结果过大或过小而造成经过 softmax 函数运算产生梯度消失现象的平衡因子, 多头注意力计算公式为:

$$TA^s = \text{Concat}[A_1^{s-1}, \dots, A_T^{s-1}]U^O + P_{\text{point}}^{s-1}, \quad (6)$$

其中映射矩阵 $U^O \in \mathbf{R}^{T C_m \times C}$ 。由于嵌入了位置编码, 使编码结构可集中感知当前位置的信息。采用多头注意力并行处理可提高模型对注意力权重的合理分配。 T 和 $\sqrt{C_{\text{in}}}$ 均采用 Transformer 中的默认设置。第 s 层注意力编码可写为:

$$P_{\text{point}}^s = \xi(TA^s) + TA^s, \quad (7)$$

其中 ξ 为前馈层^[20]。编码过程由 3 个标准的编码阶段串联组成。同时添加粗糙掩码特征, 为点特征向量提供足够的正负参考。最后, 运用 1×1 卷积解码, 预测难点最终掩码标签, 结构如图 1 右侧所示。计算机视觉通用注意力模块 CBAM(Convolutional Block Attention Module) 通过神经网络学习到的重要性表示注意力, 解释性较差。本文注意力机制通过计算点编码序列内积获得相似度, 相似度越高, 权重越大。同时, 这种方式可以让无序难点进行跨距离交互, 这是卷积神经网络难以实现的。

训练时, 粗糙掩码预测损失 L_{coarse} 和点检测头损失 L_{point} 均使用平均交叉熵损失函数监督:

$$L_{\text{coarse}} = L_{\text{point}} = \frac{-1}{G} \sum_{m=0}^G [y_m \ln \hat{y}_m + (1 - y_m) \ln(1 - \hat{y}_m)], \quad (8)$$

其中 y_m 为掩码真值, \hat{y}_m 为预测掩码, G 为所需预测像素数, y_m 与 \hat{y}_m 预测同为目标或背景时损失函数降低, 可有效对掩码及难点进行预测监督。在粗糙掩码预测时 $G = W_0 \times H_0$, 而在点检测时 $G = M$ 。

原则上, 本文方法属于 Two-stage 实例分割的一种形式。与基于锚点的 Two-stage 检测方法大量生成区域建议不同, 本文方法允许生成更少的区域建议。而与 One-stage 实例分割方法相比, 避免了对小目标的全局计算, 通过建议区域截取, 可以保留更多的细节信息。此外, 由于小目标布满建议区域, 这也有助于均衡正负样本。整体网络损失函数为:

$$L = L_r + L_{\text{coarse}} + L_{\text{point}}. \quad (9)$$

3 实验结果与分析

3.1 实验细节

难点注意力感知的红外小目标检测网络在 Detectron2 框架实现^[27], 同时参照 Detectron2 中的训练策略。本实验使用 SGD(Stochastic Gradient Descent) 优化器训练迭代 5 万次。学习率经由 4 千次热身迭代后达到基础学习率 0.02, 并在 3 万次和 4 万次迭代后分别下降十倍。设置训练图像多尺度从 256 至 608, 以达到数据增强的目的。骨干网络使用携有 4 层 BiFPN(Bidirectional Feature Pyramid Network)^[28] 的 DLA-34(Deep Layer Aggregation)^[29], 其中 BiFPN 层级通道为 256 层。FPN 输出层数为 $P_3 - P_5$, 步长分别为 [8, 16, 32]。所有实验均在一台带有 Nvidia A4000 GPU 和 Intel i7-8550U CPU 的计算机上实现。

在 NUDT-SIRST 和 IRDST 数据集上对本文算法进行评估。NUDT-SIRST 数据集由 1327 张图片组成, 大概有 37% 的图片至少拥有 2 个目标, 96% 的目标符合国际光学工程学会对小目标的定义: 目标应小于整幅图像面积的 0.15%。IRDST 数据集作为补充, 丰富更多红外目标场景。训练图像和测试图像比例设为 7:3。数据集改为 COCO 数据集^[30] 标注格式以便运用 Detectron2 框架。

3.2 评价指标

本文选用常规红外小目标检测性能评价指标: ROC 曲线(Receiver Operating Characteristic curve)、F 值^[31]。除此之外, 还选用经典实例分割指标平均精度均值(mean Average Precision, mAP)。ROC 曲线可直观体现目标检测真阳率 TPR(True Positive Rate)随着假阳率 FPR(False Positive Rate)变化关系。F 值是准确率 Precious 和召回率 Recall 调和平均值。TPR、FPR、Recall 及 Precious 的计算公式如下:

$$\text{TPR} = \text{Recall} = \frac{T_p}{T_p + F_N}, \quad (10)$$

$$\text{FPR} = \frac{F_p}{W \times H}, \quad (11)$$

$$\text{Precious} = \frac{T_p}{T_p + F_p}, \quad (12)$$

其中 T_p 、 F_N 、 F_p 分别表示真阳、假阴、假阳像素数量。F值计算公式如下:

$$F = \frac{(1 + \beta^2) \text{Precious} \times \text{Recall}}{\beta^2 \text{Precious} + \text{Recall}}, \quad (13)$$

参照文献 [31] 设置 F 值的超参数 β 为 1。

3.3 对比实验

与现阶段主流红外小目标检测方法做对比实验,其中包括传统算法: Top-hat^[32]、LEF^[7]、AAD-CDD^[33]、TLLCM^[8]。基于深度学习算法: ALCNet^[16]、DNANet^[17]、RDIAN^[21]。其中传统方法超参数设置如表 1 所示。为确保实验公平性,所有算法将在相同配置环境下运行。

表 1 传统算法超参数设置

Tab. 1 Hyperparameter settings of traditional algorithms

传统算法	超参数设置
Top-hat	Nhood=ones(5)
LEF	$h=0.2, \alpha=0.5, P=9$
AADCDD	内窗口尺寸={3, 5, 7, 9}, 外窗口尺寸=19
TLLCM	窗口尺寸={3, 5, 7, 9}, $k=9$

在表 2 中,分别列出了所选方法在两个数据

集上的定量结果。每项数据的最佳结果已加粗标识,次佳数据用下划线标识。可见传统算法表现相对较差。两个数据集包含了多种复杂背景及目标类型,导致传统算法针对特定场景设计的先验性特征遭到限制,同时超参数的选择也无法达到较好的泛化性。当面对背景与目标具有相似的预设特征或者背景与目标边界对比度不高时,传统算法难以产生精确的分割掩码。深度学习算法性能远超传统算法,表明数据驱动的方法有助于提高模型的泛化能力。本文算法相较于其他深度学习算法有较好的改进。在 NUDT-SIRST 数据集上, mAP 达到了 87.4, 在 IRDST 数据集上, mAP 达到了 63.4, 同时 F 值评价指标也很优异。这得益于基于点的区域建议模块对目标所在区域准确提取和细化掩码边界模块对目标边缘精细分割。

ROC 曲线如图 3(彩图见期刊电子版)所示,传统算法在低假阳率时无法保证真阳率,不能有效滤除背景信息。深度学习算法由图像信息驱动生成分割掩码,因此可以实现非常低的假阳率。本文算法通过基于点的区域建议模块滤除冗余背景,从而保证了低假阳率,同时细化掩码边界模块提高了真阳率。

表 2 各方法在 NUDT-SIRST 及 IRDST 数据集定量结果对比

Tab. 2 Comparison of quantitative results of different methods on NUDT-SIRST and IRDST datasets

检测算法	NUDT-SIRST		IRDST	
	mAP	F值(Pre, Rec)	mAP	F值(Pre, Rec)
Top-hat	1.5	0.3599(0.2850, 0.4884)	0.7	0.0088(0.0045, 0.4107)
LEF	6.4	0.1151(0.0748, 0.2498)	2.5	0.1219(0.0686, 0.5470)
AADCDD	1.6	0.1490(0.3838, 0.0924)	1.4	0.0705(0.0521, 0.1090)
TLLCM	16.5	0.0724(0.0479, 0.1476)	6.1	0.1881(0.1254, 0.3759)
ALCNet	69.3	0.7595(0.7035, 0.8251)	46.5	0.5929(0.5461, 0.6486)
DNANet	<u>86.9</u>	0.8645(0.9070 , 0.8259)	<u>62.1</u>	0.6697(<u>0.7124</u> , 0.6319)
RDIAN	82.4	<u>0.8900</u> (<u>0.8990</u> , <u>0.8811</u>)	60.0	0.7102 (0.7092, 0.7113)
本文方法	87.4	0.8935 (0.8923, 0.8948)	63.4	0.7056(0.7183 , 0.6935)

深度学习算法在两个数据集中单张图片推理时间如表 3 所示,数值越低表示性能越好。NUDT-SIRST 数据集中图像尺寸为 256×256, 本文算法单张图像平均推理速度为 0.099 s。在图像尺寸

更大(512×512)的 IRDST 数据集中,由于细化模块只需要处理建议区域内的像素,受图片尺寸变化影响较小,推理单张图像平均时间只增加了 0.022 s,控制计算开销效果可观。

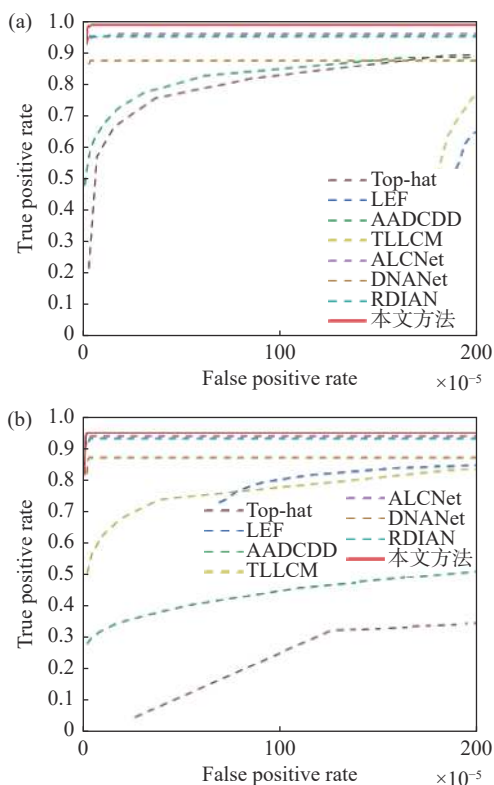


图 3 各算法在 (a) NUDT-SIRST 数据集和 (b) IRDST 数据集 ROC 曲线

Fig. 3 ROC curves for methods on (a) the NUDT-SIRST dataset and (b) the IRDST dataset

表 3 深度学习单张图片平均推理时间
Tab. 3 Average inference times of a single image for deep learning methods (s)

检测算法	NUDT-SIRST	IRDST
ALCNet	0.104	0.166
DNANet	0.089	0.259
RDIAN	0.065	0.114
本文算法	0.099	0.121

图 4(彩图见期刊电子版)展示了所选方法在不同代表性场景下的红外小目标检测结果。可以看出,本文方法在不同场景下都能够对目标进行精确定位,同时产生精细分割边缘。传统算法在点目标场景中表现较好,但在复杂背景下易对背景边缘高频区域产生误报。相比之下,深度学习算法能够更准确地定位目标。本文方法经过细化掩码边界模块对粗糙边界重新预测,可以得到相较于同类算法更精细的分割边缘。图 5(彩图见期刊电子版)为各方法不同场景的 3D 可视化结果,可以看出本文方法可以明显抑制复杂背景。

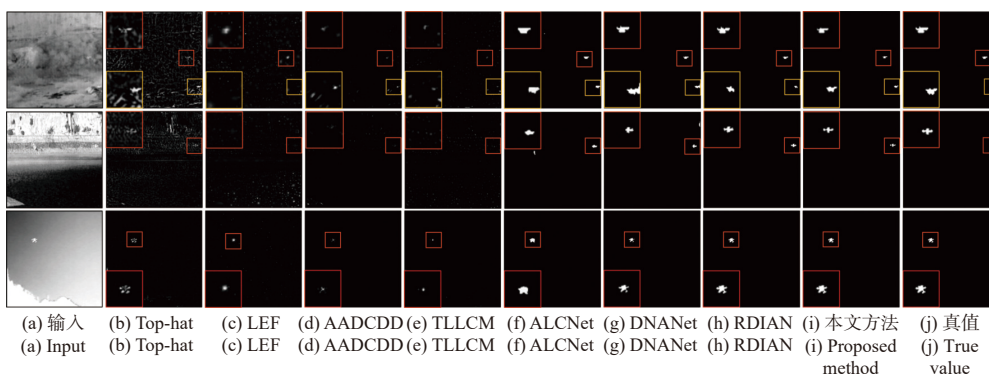


图 4 各方法不同场景检测结果,相同目标放大显示于同颜色框内

Fig. 4 Different scene detection results of different methods, with the same target zoomed in the same color box

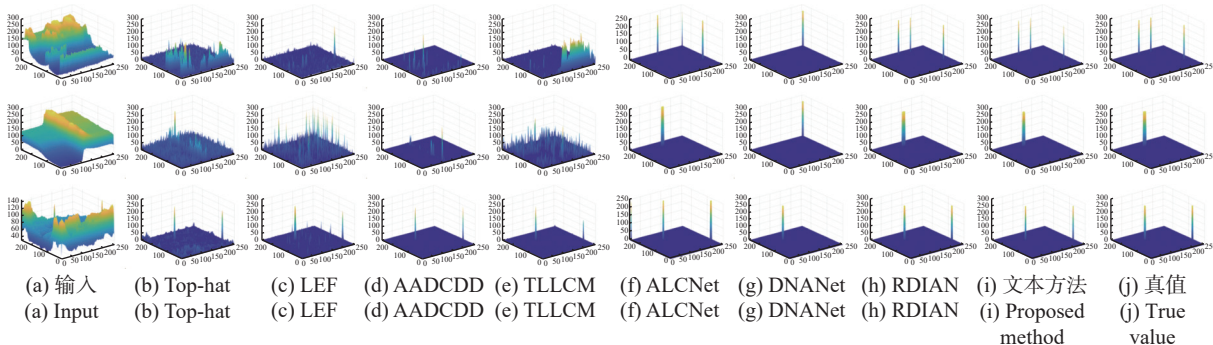


图 5 各方法检测结果 3D 可视化

Fig. 5 3D visualization of detection results for each method

3.4 消融实验

本文在 NUDT-SIRST 数据集上进行详细的消融实验, 分析基于点的区域建议模块及细化掩码边界模块对检测性能的影响。

模型中目标中心点热图及区域建议边界可视化结果如图 6(彩图见期刊电子版)所示。小目标中心位置获得更高的权重加持, 从而可以准确提供目标区域建议。

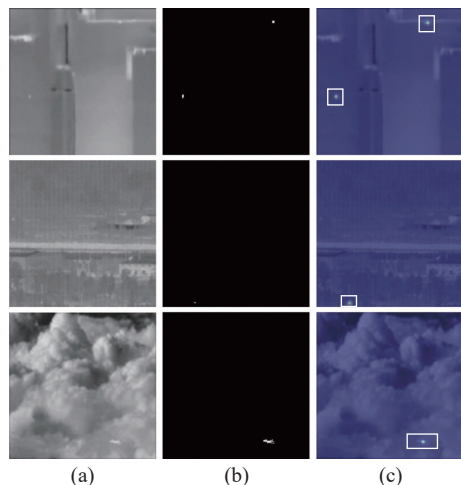


图 6 区域建议可视化结果。(a) 原图; (b) 真值; (c) 中心点热图及区域建议边界

Fig. 6 Visualization of regional proposals. (a) Original maps; (b) groundtruths; (c) centroid heatmap and regional proposal boundaries

表 4 对比了基于点的区域建议模块与常规目标检测区域建议模块 RPN 在不同区域建议数量下的性能。本文方法在相同建议数量下 mAP 高于 RPN。随着区域建议数量的减少, RPN 的性能受到更明显的影响。本文方法在建议区域数量为 64 时的 F 值高于建议区域数量为 1000 的 RPN。这表明基于点的区域建议模块在提供更精确的区域建议方面具有明显优势。

表 4 不同区域建议模块对比表

Tab. 4 Comparison of different region proposal modules

建议数量	基于点的区域建议		RPN	
	mAP	F值	mAP	F值
1000	87.9	0.8927	86.2	0.8425
256	87.5	0.8962	85.8	0.8412
128	87.4	0.8935	85.2	0.8406
64	86.0	0.8901	84.5	0.8397

不同选点策略下细化掩码边界模块的性能如表 5 所示。针对难点采样 mAP 高于均匀取点, 但是过于倾向于难点反而导致性能下降。在图 7(彩图见期刊电子版)中, 将所选的难点像素用粉色表示, 可以看出所选难点分布于目标边缘, 因此对产生精细化分割提升较大。难点特征表示信息对网络表现的影响如表 6 所示。细粒度特征引入粗糙掩码后 mAP 值提高了 0.3。由于分割任务对位置的敏感性, 而且本文选点均是无序点, 位置掩码带来的性能提升更为显著, mAP 提高了 1.6, 可见难点的位置信息至关重要。将细化掩码边界模块中的点检测头与其余细化方案结构进行比较, 其中 CNN 结构来自 Mask-rcnn^[34], MLP 结构来自 Pointrend^[35]。同时, 在表 7 中还分析了 Transformer 结构深度对网络性能的影响。本文方法优于其余两种方法, 这得益于 Transformer 对非局部分布的难点进行注意力建模。而随着 Transformer 模块深度的增加, 性能提升有限。

表 5 不同选点策略检测结果

Tab. 5 Detection results of different point selection strategies

选点策略	mAP
均匀选点	86.7
$k=1, \gamma=0.00$	86.9
$k=3, \gamma=0.75$	87.4
$k=10, \gamma=1.00$	85.8

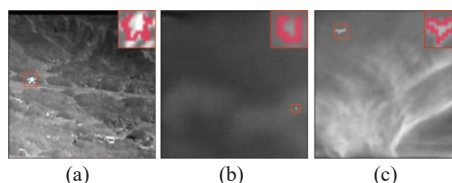


图 7 难点分布可视化。(a) 无人机目标; (b) 点目标; (c) 飞机目标

Fig. 7 Visualization of indistinguishable points distribution. (a) UAV target; (b) point target; (c) aircraft target

表 6 难点不同特征融合结果

Tab. 6 Fusion results of different features at indistinguishable points

细粒度特征	粗糙掩码	位置嵌入	mAP
√			85.5
√	√		85.8
√	√	√	87.4

表 7 不同细化方案检测结果

Tab. 7 Results of different refinement strategies

细化方案	mAP
CNN(16×16)	85.5
MLP(16×16)	86.2
细化掩码边界模块(S=3)	87.4
细化掩码边界模块(S=6)	87.6

4 结 论

本文提出一种基于难点注意力感知的红外小

目标检测网络。基于点的区域建议模块准确定位目标,提供区域建议,在相同建议数量(128)下使 mAP 提升 2.2。细化掩码边界模块,选择上采样预测难点,融合难点特征,建模逐点注意力,点检测头重新预测生成精细掩码,使得 mAP 提升 1.2。

实验结果表明,本文方法在诸多具有挑战性背景下均能够实现精确分割,于 NUDT-SIRST 和 IRDST 数据集上 mAP 分别达到 87.4 和 63.4。同时通过消融实验证明各模块的有效性。

参考文献:

- [1] 单秋莎, 谢梅林, 刘朝晖, 等. 制冷型长波红外光学系统设计[J]. *中国光学*, 2022, 15(1): 72-78.
SHAN Q SH, XIE M L, LIU ZH H, *et al.*. Design of cooled long-wavelength infrared imaging optical system[J]. *Chinese Optics*, 2022, 15(1): 72-78. (in Chinese).
- [2] MA T L, YANG ZH, WANG J Q, *et al.*. Infrared small target detection network with generate label and feature mapping[J]. *IEEE Geoscience and Remote Sensing Letters*, 2022, 19: 6505405.
- [3] SUN Y, YANG J G, AN W. Infrared dim and small target detection via multiple subspace learning and spatial-temporal patch-tensor model[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 59(5): 3737-3752.
- [4] 赵鹏鹏, 李庶中, 李迅, 等. 融合视觉显著性和局部熵的红外弱小目标检测[J]. *中国光学*, 2022, 15(2): 267-275.
ZHAO P P, LI SH ZH, LI X, *et al.*. Infrared dim small target detection based on visual saliency and local entropy[J]. *Chinese Optics*, 2022, 15(2): 267-275. (in Chinese).
- [5] GAO C Q, MENG D Y, YANG Y, *et al.*. Infrared patch-image model for small target detection in a single image[J]. *IEEE Transactions on Image Processing*, 2013, 22(12): 4996-5009.
- [6] CHEN C L P, LI H, WEI Y T, *et al.*. A local contrast method for small infrared target detection[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2014, 52(1): 574-581.
- [7] XIA CH Q, LI X R, ZHAO L Y, *et al.*. Infrared small target detection based on multiscale local contrast measure using local energy factor[J]. *IEEE Geoscience and Remote Sensing Letters*, 2020, 17(1): 157-161.
- [8] HAN J H, MORADI S, FARAMARZI I, *et al.*. A local contrast method for infrared small-target detection utilizing a tri-layer window[J]. *IEEE Geoscience and Remote Sensing Letters*, 2020, 17(10): 1822-1826.
- [9] 刘彦磊, 李孟喆, 王宣宣. 轻量型 YOLOv5s 车载红外图像目标检测[J]. *中国光学(中英文)*, 2023, 16(5): 1045-1055.
LIU Y L, LI M ZH, WANG X X. Lightweight YOLOv5s vehicle infrared image target detection[J]. *Chinese Optics*, 2023, 16(5): 1045-1055. (in Chinese).
- [10] PANG Y W, WANG T C, ANWER R M, *et al.*. Efficient featurized image pyramid network for single shot detector[C]. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2019: 7328-7336.
- [11] YANG X, YAN J CH, FENG Z M, *et al.*. R3Det: Refined single-stage detector with feature refinement for rotating object[C]. *Thirty-Seventh AAAI Conference on Artificial Intelligence*, AAAI Press, 2020.
- [12] LIU Y ZH, CAO S, LASANG P, *et al.*. Modular lightweight network for road object detection using a feature fusion approach[J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2021, 51(8): 4716-4728.
- [13] ZHANG SH F, WEN L Y, BIAN X, *et al.*. Single-shot refinement neural network for object detection [C]. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, 2018: 4203-4212.
- [14] YUAN Y, XIONG ZH T, WANG Q. VSSA-NET: Vertical spatial sequence attention network for traffic sign detection[J]. *IEEE Transactions on Image Processing*, 2019, 28(7): 3423-3434.
- [15] PANG Y W, CAO J L, WANG J, *et al.*. JCS-Net: Joint classification and super-resolution network for small-scale pedestrian detection in surveillance images[J]. *IEEE Transactions on Information Forensics and Security*, 2019, 14(12): 3322-3331.

- [16] DAI Y M, WU Y Q, ZHOU F, *et al.*. Attentional local contrast networks for infrared small target detection[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 59(11): 9813-9824.
- [17] LI B Y, XIAO CH, WANG L G, *et al.*. Dense nested attention network for infrared small target detection[J]. *IEEE Transactions on Image Processing*, 2023, 32: 1745-1758.
- [18] WANG K W, DU SH Y, LIU CH X, *et al.*. Interior attention-aware network for infrared small target detection[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 5002013.
- [19] GIRSHICK R. Fast R-CNN[C]. *2015 IEEE International Conference on Computer Vision (ICCV)*, IEEE, 2015: 1440-1448.
- [20] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, *et al.*. An Image is worth 16x16 words: Transformers for image recognition at scale[C]. *9th International Conference on Learning Representations*, OpenReview. net, 2021.
- [21] 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.
- [22] ZHOU X Y, KARPUR A, LUO L J, *et al.*. StarMap for category-agnostic keypoint and viewpoint estimation[C]. *Proceedings of the European Conference on Computer Vision*, Springer, 2018: 328-345.
- [23] LIN T Y, GOYAL P, GIRSHICK R, *et al.*. Focal loss for dense object detection[C]. *2017 IEEE International Conference on Computer Vision (ICCV)*, IEEE, 2017: 2999-3007.
- [24] YANG Z, LIU SH H, HU H, *et al.*. RepPoints: Point set representation for object detection[C]. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, 2019: 9656-9665.
- [25] XU B, WANG N Y, CHEN T Q, *et al.*. Empirical evaluation of rectified activations in convolutional network[Z]. *Computerence*, 2015. DOI: 10.48550/arXiv.1505.00853.
- [26] ZHU X ZH, SU W J, LU L W, *et al.*. Deformable DETR: deformable transformers for end-to-end object detection[C]. *9th International Conference on Learning Representations*, OpenReview. net, 2020.
- [27] WU Y, KIRILLOV A, Massa F, *et al.*. Detectron2[CP/OL]. (2019)[2023-8-24]. <https://github.com/facebookresearch/detectron2>.
- [28] TAN M X, PANG R M, LE Q V. EfficientDet: Scalable and efficient object detection[C]. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2020: 10778-10787.
- [29] YU F, WANG D Q, SHELHAMER E, *et al.*. Deep layer aggregation[C]. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, 2018: 2403-2412.
- [30] LIN T Y, MAIRE M, BELONGIE S, *et al.*. *Microsoft COCO: Common Objects in Context*[M]. FLEET D, PAJDLA T, SCHIELE B, *et al.*. *Computer Vision – ECCV 2014*. Cham: Springer, 2014: 740-755.
- [31] WANG H, ZHOU L P, WANG L. Miss detection vs. false alarm: adversarial learning for small object segmentation in infrared images[C]. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, 2019: 8508-8517.
- [32] RIVEST J F, FORTIN R. Detection of dim targets in digital infrared imagery by morphological image processing[J]. *Optical Engineering*, 1996, 35(7): 1886-1893.
- [33] AGHAZIYARATI S, MORADI S, TALEBI H. Small infrared target detection using absolute average difference weighted by cumulative directional derivatives[J]. *Infrared Physics & Technology*, 2019, 101: 78-87.
- [34] HE K M, GKIOXARI G, DOLLÁR P, *et al.*. Mask R-CNN[C]. *2017 IEEE International Conference on Computer Vision (ICCV)*, IEEE, 2017: 2980-2988.
- [35] KIRILLOV A, WU Y X, HE K M, *et al.*. PointRend: Image segmentation as rendering[C]. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2020: 9796-9805.

作者简介:



宋延嵩(1983—),男,吉林长春人,长春理工大学空间光电技术研究所研究员,博士生导师,主要从事空间激光通信等方面的研究。E-mail: songyansong2006@126.com