基于级联嵌套U-Net的红外小目标检测

薛雅丽[1],俞潼安[1],崔 闪[2],周李尊[1]

(1. 南京航空航天大学 自动化学院,南京 210000; 2. 上海机电工程研究所,上海 201109)

摘 要:针对复杂场景下红外小目标尺寸差异大、检测效果欠佳的问题,提出了一种基于级联嵌套U-Net的红外小目标检测方法。首先,为解决不同场景下小目标尺寸差异大的问题,搭建了3种深度的U-Net网络,并将3个U-Net网络以级联嵌套的方式组成检测模型;其次,引入对比度信息抽取模块,以进一步丰富特征信息,抑制密集背景噪声的干扰;最后,将所提出的算法与5种主流的算法进行比较。实验结果表明:本文算法性能优于其他算法,并且平均交并比、精确率和召回率分别达到了78.61%、93.36%和81.78%。

关键词:信息处理技术;红外小目标;对比度信息;注意力机制;检测

中图分类号:TP391 **文献标志码**:A **文章编号**:1671-5497(2025)05-1714-08

DOI:10.13229/j.cnki.jdxbgxb.20230785

Infrared small target detection based on cascaded nested U-Net

XUE Ya-li[1], YU Tong-an[1], CUI Shan[2], ZHOU Li-zun[1]

(1. *College of Automation Engineering*, *Nanjing Univercity of Aeronautics and Astronautics*, *Nanjing* 210000, *China*; 2. *Shanghai Electro-Mechanical Engineering Institute*, *Shanghai* 201109, *China*)

Abstract: Aiming at the problem of large size differences in infrared small targets and poor detection results in complex scenes, an infrared small target detection method based on cascaded nested U-Net is proposed. First, in order to solve the problem of large size differences between small targets in different scenarios, three depths of U-Net networks were built, and the three U-Net networks were cascaded and nested to form a detection model; secondly, contrast was used information extraction module to further enrich feature information and suppress the interference of dense background noise; finally, the proposed algorithm is compared with five mainstream algorithms. The experimental results show that the performance of this algorithm is better than other algorithms, and the average intersection and union ratio, the precision rate and recall rate reached 78.61%, 93.36% and 81.78% respectively.

Key words: information processing technology; infrared small target; contrast information; attention mechanism; detection

收稿日期:2023-07-26.

基金项目:国家自然科学基金项目(62073164);航天集成基金项目(U22B6001);上海市航天科技创新基金项目(SAST2022-013).

作者简介:薛雅丽(1974-),女,副教授,博士.研究方向:目标检测及识别.E-mail:xueyali@nuaa.edu.cn

0　引　言

20世纪初，红外成像技术因具有较强穿透能力、可全天时工作等特点被广泛应用于军事领域[1]。随着红外成像技术的不断发展，红外传感器在战场环境中扮演着关键角色，尤其是在探测和识别小目标方面。然而，在红外成像中，小目标往往会因为海杂波或云杂波而难以被识别，而且红外小目标的大小和形状在不同场景下差异显著。因此，红外弱小目标的识别一直是一个具有挑战性的任务。

在该领域早期研究中，学者们提出了一些传统的方法。Top-hat[2]是一种经典的形态学滤波方法，它通过将原图像与经过膨胀腐蚀等处理后的图像进行差分运算来抑制背景，但该算法过度依赖目标与背景的尺寸差异，因此使用范围受限。随后学者们又关注到小目标与领域之间的像素差异，从而提出了基于视觉机制的检测方法。Chen等[3]提出了局部对比度测量（Local contrast measure, LCM），该算法利用目标与领域的局部对比度信息进行检测。之后陆续有学者提出了许多基于对比度的改进方法，例如张文杰等[4]将空间关系与区域关系联系起来，得到了较精确的显著图。但上述方法最终都需要进行阈值分割来区分目标与背景，因此对具有高亮噪声点的图像检测效果较差。

随着深度学习的兴起，神经网络强大的特征学习能力为红外小目标检测提供了新的方法。Faster R-CNN[5]网络的提出，有效的提升了基于候选框算法的检测效率；YOLO[6]算法解决了候选框系列算法步骤复杂、训练速度慢的问题。但这些方法无法直接应用于红外小目标的检测工作，因为随着网络的深入，小目标极可能因池化或下采样操作而丢失。Wang等[7]提出了MDvsFA（Miss detection vs. false alarm），使用两个模型分别进行训练，以实现虚警率和漏检率的平衡，但该算法没有考虑敏感噪声对模型的影响；为了消除敏感噪声对小目标的影响，Deng等[8]以生成对抗网络[9]为基础，丰富了小目标的特征信息；Wu等[10]提出的UIU-Net将微小的U-Net嵌入更大的U-Net[11]骨干网络中，实现了全局和局部对比信息的增强。但是，这些方法仍没有考虑到不同场景下目标尺寸差异大的问题。针对上述问题，本文提出了一种基于级联嵌套U-Net（Cascaded nested U-Net, CNU-Net）的网络。

1　本文方法

1.1　总体框架

不同尺寸目标的最优感受野差异较大，因此不同深度的网络适用于不同尺寸的目标。而红外小目标在不同场景中的像素尺寸大小不一且形状变化较大。基于上述思想，本文提出的CNU-Net框架如图1所示，它由3个不同深度的U-Net级联嵌套组成。首先由3层U-Net网络对输入图像进行浅层特征挖掘，提取小尺寸目标的特征信息，并保留大尺寸目标的边缘特征。其次，将特征信息通过对比度抽取模块提取对比度信息，同时抑制密集背景噪声，并将原图像、特征信息及对比度信息通过注意力机制加权融合作为4层U-Net的输入。4层U-Net以较深的网络结构来扩大感受野，从而提取大尺寸目标的特征信息，同时在网络内部嵌套2层U-Net来防止极小目标的特征丢失。最后，4层U-Net输出的特征图通过特征金字塔融合结构[12]进行多尺度特征融合，并通过由卷积层组成的8连通域聚类模块对特征图进行预测，得到最终的图像分割结果。因此本文提出的网络可以用于识别复杂背景下不同尺寸的红外小目标。

1.2　特征提取模块

本文的特征提取模块如图2所示，由卷积层（Conv）、批归一化层（BN）、ReLU激活函数组成，

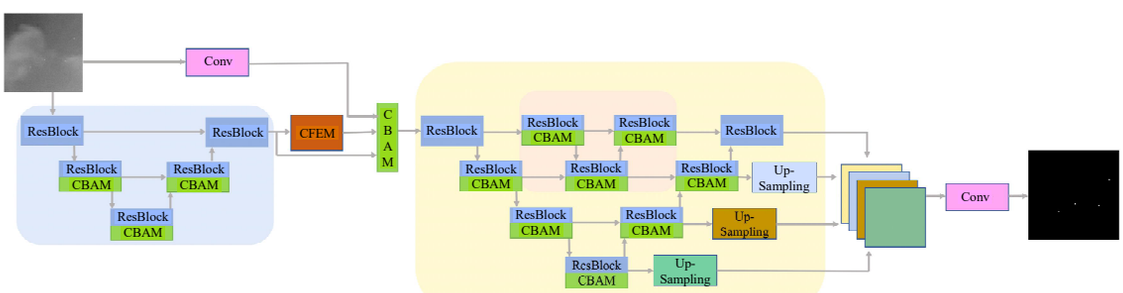

图1　基于CNU-Net的红外小目标检测框图

Fig. 1　Infrared small target detection block diagram based on CNU-Net

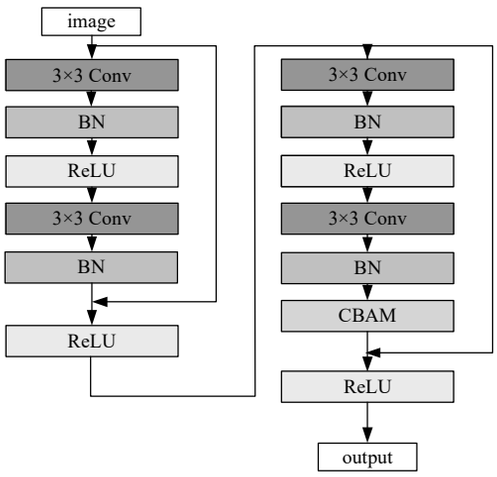

图2 特征提取模块

Fig. 2 Feature extraction module

并通过ResNet残差结构连接,最后利用通道空间注意力机制模块[13](Convolutional block attention module,CBAM)引导网络学习重要特征。考虑到3个U-Net组合的情况下参数量较大且易出现梯度消失的情况,为降低模型复杂度,并且在加速网络模型的同时防止梯度消失,本文引入了残差结构进行特征提取。

在U-Net的特征提取过程中,特征图的通道数随着编码过程增加,而随着解码过程减少。为提升特征提取模块接收信息的质量,本文在输入输出层以外的所有特征提取模块中都加入了CBAM(如图3所示),实现特征的自适应增强。CBAM由2个注意力单元串联而成,输入特征图先通过通道注意力单元得到通道注意力权重并与输入特征图相乘,再通过空间注意力单元得到空间注意力权重,最后加权得到自适应增强的特征图。

通道注意力单元用于关注图像中哪些通道的信息更重要,主要由池化操作与多层感知机共同组成,可以表示为:

$$\boldsymbol{W}_{\mathrm{c}}(\mathrm{Input})=\sigma\Big[\mathrm{MLP}\big(\mathrm{Avgpool}(\mathrm{Input})\big)+\mathrm{MLP}\big(\mathrm{Maxpool}(\mathrm{Input})\big)\Big] \tag{1}$$

式中:$\boldsymbol{W}_{\mathrm{c}}$为一维通道注意力权重;MLP为多层感知机;Avgpool、Maxpool分别为平均池化和最大池化层;σ为sigmoid概率函数。

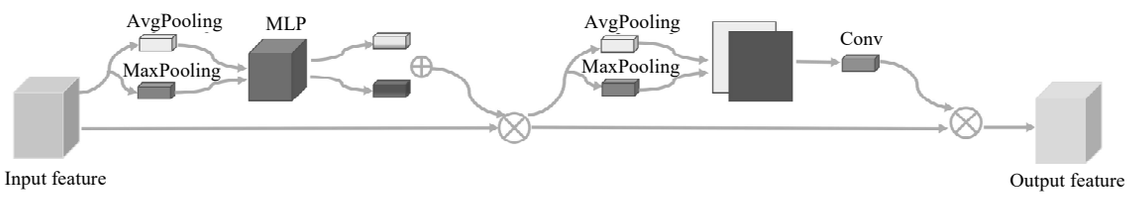

图3 通道空间注意力机制模块

Fig. 3 Convolutional block attention module

空间注意力与通道注意力不同,它用于关注输入图像哪些部分信息更加重要。空间注意力机制单元主要由池化操作与卷积操作组成,可以表示为:

$$\boldsymbol{W}_{\mathrm{s}}(\mathrm{Input})=\sigma\Big[\mathrm{Conv}\big(\mathrm{Avgpool}(\mathrm{Input}),\mathrm{Maxpool}(\mathrm{Input})\big)\Big] \tag{2}$$

式中:$\boldsymbol{W}_{\mathrm{s}}$为二维空间注意力权重;Conv为卷积层。

1.3 对比度特征抽取模块

由于3层U-Net在提取浅层特征时会被大面积的密集背景噪声干扰,为在抑制噪声的同时增强对比度信息,本文采用了一种基于局部对比度测量的对比度特征抽取模块(Contrast feature extraction module,CFEM)。该算法原本用于检测,本文将该算法迁移至特征的增强上,以有效提取对比度信息。

如图4所示,假设图中每一小格为一个像素点,图像大小为$a\times b$,滑动窗口大小为6×6,将其分成9个区域,那么每个区域所占大小为2×2,C为中心区域,$B_i(i=1,2,\cdots,8)$为背景区域,滑动窗口从左到右、从上到下遍历整幅图像。

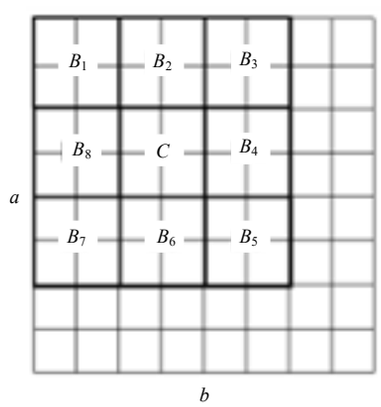

图4 滑动窗口

Fig. 4 Sliding window

当滑动窗口停留在某位置时，该位置中心区域与背景区域的相异度为：

$$d(C,B_i)=|M_c-M_{B_i}|(i=1,2,\cdots,8) \quad (3)$$

式中：M_c为中心区域像素均值；M_{B_i}为背景区域的像素均值，由这8个值来描述当前位置中心区域与背景区域的差异。

在特征图中，提取到的目标特征往往与背景特征的强度不同，本文用以下公式来描述该性质，即相似性：

$$d_n=d(C,B_i)\times d(C,B_{n+4})(n=1,2,3,4) \quad (4)$$

为了最大限度地提取对比度信息，采用相似性最小值来代替中心区域C的像素值，最终得到了对比度信息图，即：

$$c(x,y)=\min_{n=1,2,3,4}\sum_{t=1}^{N}\tilde{d}_n(x,t) \quad (5)$$

$$x=1,2,\cdots,a,y=1,2,\cdots,b,t=1,2,\cdots,b$$

本文的模型是一个端到端的网络，因此将该算法嵌入网络时，需要将原本的数值运算转化成卷积运算，并设计成一个模块。如图5所示，首先对特征图进行边缘扩充，然后通过步长为1的平均池化操作计算各区域的像素均值，接着定义一个不参与反向传播的卷积层来计算相异度，最后通过相似性计算取最小值得到对比度特征图。

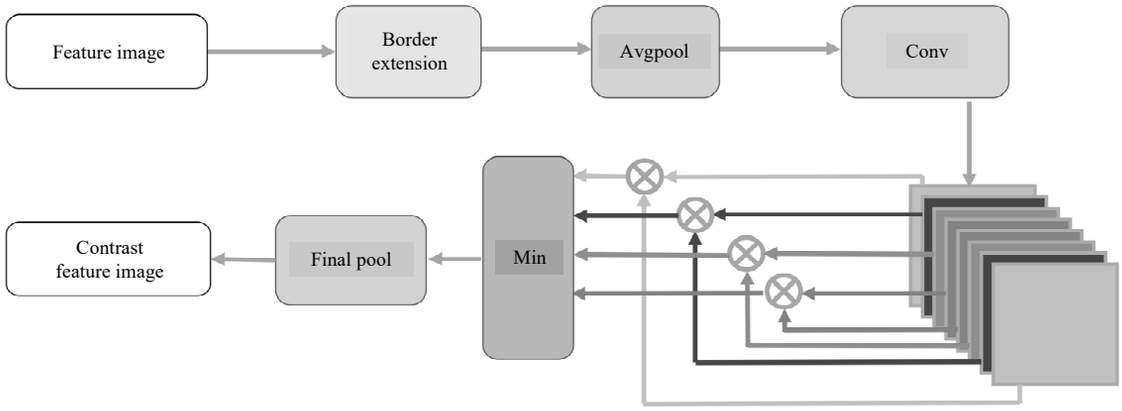

图5　对比度特征抽取模块

Fig. 5　Contrast feature extraction module

2　实验方法与结果分析

2.1　数据集介绍

本文从NUAA-SIRST[14]、NUDT-SIRST[15]以及NUST-SIRST 3个不同的数据集中整合了红外弱小目标的图像并构成了新的数据集。根据国际光学工程学会(Society of photo-optical instrumentation engineers，SPIE)定义的小目标(即目标应小于整幅图像的0.15%)，本数据集有99.58%的目标符合标准，其中27.53%的目标只占整幅图像的0.02%，视为极小目标。整个数据集共1 384张图片，其中单目标图像占比65.97%，双目标图像占比30.71%，多目标图像占比3.32%。综上，该数据集的检测工作有一定难度，需要做到多目标检测以及极小目标的检测。在训练模型时，随机对数据集中的样本以50%概率进行水平翻转、亮度增强的操作，以实现数据增强。

2.2　训练环境及实验设置

本文基于PyTorch1.13.1平台对所提算法进行验证，图形处理器(Graphics processing unit，GPU)为NVIDIA GeForce RTX 3070，操作系统为Windows 11。本文采用AdaGrad作为优化器，初始学习率为0.05；采用余弦退火的学习率衰减策略，训练轮次为1 000，batch大小为8。

2.3　评价指标

本文采用平均交并比(Mean intersection over union，mIoU)、精确率(Precision)、召回率(Recall)、受试者操作特征(Receiver operation characteristics，ROC)曲线这几个度量指标对网络性能进行评价。

交并比(IoU)是一个像素级的评价指标，用于评估模型对目标的轮廓描述能力。IoU定义为预测图像与掩码图像的交集面积和并集面积的比值，即：

$$IoU=\frac{A_{\text{inter}}}{A_{\text{outer}}} \quad (6)$$

式中:A_{inter}和A_{outer}分别为交集区域和并集区域的大小。

mIoU可通过计算所有样本IoU的平均值得到,即:

$$mIoU=\frac{1}{N}\sum_{i=1}^{N}\mathrm{IoU}(i) \tag{7}$$

式中:N为样本总数;$IoU(i)$为第i个样本的IoU。

精确率(Precision)是一个像素级的评价指标,定义为正类被预测为正类的像素点数与所有被预测为正类的像素点数的比值,即:

$$Precision=\frac{TP}{TP+FP} \tag{8}$$

式中:TP为正类被预测为正类的像素点数;FP为负类被预测为正类的像素点数。

召回率(Recall)是一个像素级的评价指标,用于评估模型判别背景的能力,定义为正类被预测为正类的像素点数与实际为正类的像素点数的比值,即:

$$Recall=\frac{TP}{TP+FN} \tag{9}$$

式中:FN为图像中正类被预测为负类的像素点数。

ROC曲线是指在不同虚警率下的精准率变化情况。

2.4 损失函数

本文采用的损失函数为softIoU Loss,该损失函数常用于图像分割中小目标的分割任务,定义为:

$$\mathrm{Loss}=1-\frac{1}{N}\sum_{N}\frac{\sum_{\mathrm{pixels}}y_{\mathrm{true}}y_{\mathrm{pred}}}{\sum_{\mathrm{pixels}}y_{\mathrm{true}}+y_{\mathrm{pred}}-y_{\mathrm{true}}y_{\mathrm{pred}}} \tag{10}$$

式中:y_{true}为标签图像;y_{pred}为预测图像。

2.5 实验结果

2.5.1 网络深度实验结果

针对红外小目标检测,首先需要确定网络的深度,即明确U-Net对小目标的最佳感受野大小。为此,本文分别对2~5层的U-Net进行了实验,结果如图6所示。

从实验结果可以得出,2层的U-Net只对小尺寸目标敏感,因此检测效果最差;3、4层的U-Net检测性能明显优于其他层的U-Net;5层的U-Net由于网络感受野过大而丢失目标较多,检测效果开始下降。综上,本文选择了2~4层的U-Net作为识别网络的基础架构。

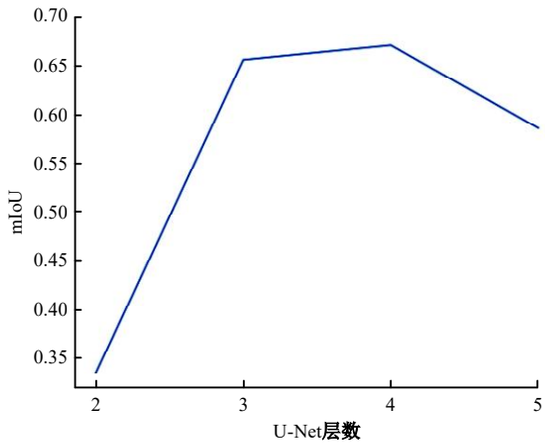

图6 不同层数U-Net的检测性能

Fig. 6 Detection performance of different layers of U-Net

2.5.2 消融实验结果

为探究不同模块对网络性能的影响,本文设置了消融实验,具体包括:是否采用级联结构(Nested structure, NS)、嵌套结构(Cascading structure,CS)进行多尺度小目标检测;是否采用CFEM提取对比度信息来增强特征。实验结果如表1所示,同时本文还对模型识别过程中的特征图进行了可视化分析,如图7所示。

分析上述实验结果,实验1在去除CFEM结构的情况下,Precision、Recall、mIoU指标均比实验5有所下降,说明对比度信息对于本文模型的识别性能具有正向作用;在实验2中,CNU-Net

表1 消融实验结果

Table 1 Ablation experimental results %

实验序号	NS	CS	CFEM	Precision	Recall	mIoU
1	√	√	×	90.11	78.66	74.68
2	√	×	√	91.14	79.64	75.19
3	√	×	×	88.04	76.83	71.16
4	×	√	×	86.47	74.89	68.94
5	√	√	√	93.36	81.78	78.61

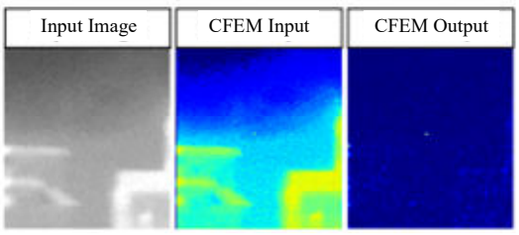

图7 CFEM输入输出的可视化特征图

Fig. 7 Visual feature map of CFEM input and output

没有嵌套2层的U-Net，其Precision、Recall、mIoU指标比实验5略微下降，这是因为有一部分极小目标识别有误，造成性能下滑，也证明了嵌套结构对极小目标特征的捕捉能力出色；实验3、4考虑到CFEM结构与NS的关联性，因此在去除CFEM结构的同时分别取消NS和CS，各项指标比实验1有一定程度的下降，说明了这两种结构在CNU-Net中的必要性。

由图7可以明显看出，3层U-Net的输出特征图存在密集的高亮噪声，识别模型的注意力分散，对后续的目标识别造成一定影响；而经过CFEM后，其对比度信息中目标特征显著增强，背景噪声被大幅度抑制，为网络提供了高质量的特征图。如图8所示，原始图像中的小目标像素点占比极小，属于极小目标，无嵌套结构的CNU-Net容易将其识别成背景，造成漏检，而带有嵌套结构的CNU-Net可以保留住极小目标的特征并实现精准识别，验证了本文采用的CFEM以及嵌套结构的有效性。

2.5.3　算法对比实验结果

根据上述参数设置，本文对目前主流的6种红外小目标检测算法进行对比，包括CNU-Net（本文算法）、ALC-Net[16]、MDvsFA、PSTNN[17]、Top-hat、LCM。实验结果见表2。本文提出的CNU-Net模型在各指标上都取得了最优的性能，

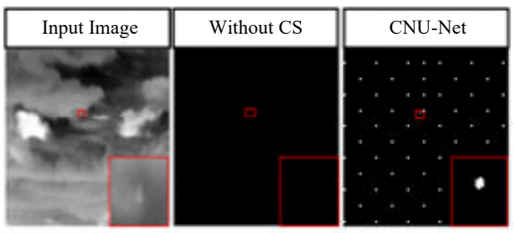

图8　有无嵌套结构的CNU-Net识别目标效果对比图

Fig. 8　CNU-Net recognition target effect comparison diagram with or without nested structure

表2　对比实验结果

Table 2　Comparative experimental results

方法	mIoU/%	Precision/%	Recall/%	Inference time
Top-hat	9.05	78.03	0.39	—
PSTNN	32.99	73.48	35.27	—
LCM	4.53	55.22	0.45	—
MDvsFA	63.83	86.96	68.33	0.021
ALC-Net	74.69	89.94	80.13	0.032
CNU-Net	78.61	93.36	81.78	0.028

Precision、Recall、mIoU分别达到了93.36%、81.78%、78.61%，与传统方法中基于滤波的算法Top-hat、基于对比度的算法LCM相比，各项指标显著提高；与基于局部先验信息的算法PSTNN相比，mIoU提高了45.62百分点，Precision提高了19.88百分点，Recall提高了46.51百分点；与基于神经网络的算法相比，本文提出的CNU-Net依然表现优异；相较于MDvsFA算法，mIoU提高了14.78百分点，Precision提高了6.4百分点，Recall提高了13.45百分点；而与在许多数据集上都表现出色的ALC-Net相比，mIoU提高了3.92百分点，Precision提高了3.42百分点，Recall提高了1.65百分点。另外，本文选择了识别性能较好的3种算法比较了它们的推理速度，结果见表2。从表中可知，CNU-Net的推理速度比ALC-Net稍快，且与MDvsFA相近；在推理速度相差不大的情况下，CNU-Net在各项识别的指标上都有相比于其他算法更优越的表现，证明了其具备出色的识别效率。

图9为不同算法的ROC曲线，从图中可以发现，随着虚警率的增加，本文模型的真阳性率始终保持最大，且ROC曲线下面积（Area under the ROC curve，AUC）值最大，表明了本文模型在不同阈值下，综合性能最佳。

本文对各算法的红外小目标检测结果进行了可视化，如图10所示，可以看出，Top-hat等传统算法容易受到密集背景噪声的干扰而出现大面积虚警区域甚至漏检，性能较差；而MDvsFA等基于深度学习的算法也在局部高亮区域仍会识别出

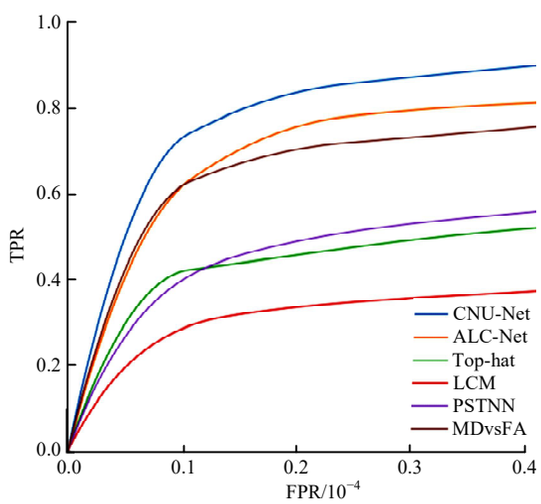

图9　不同算法的ROC曲线

Fig. 9　ROC curves of different algorithms

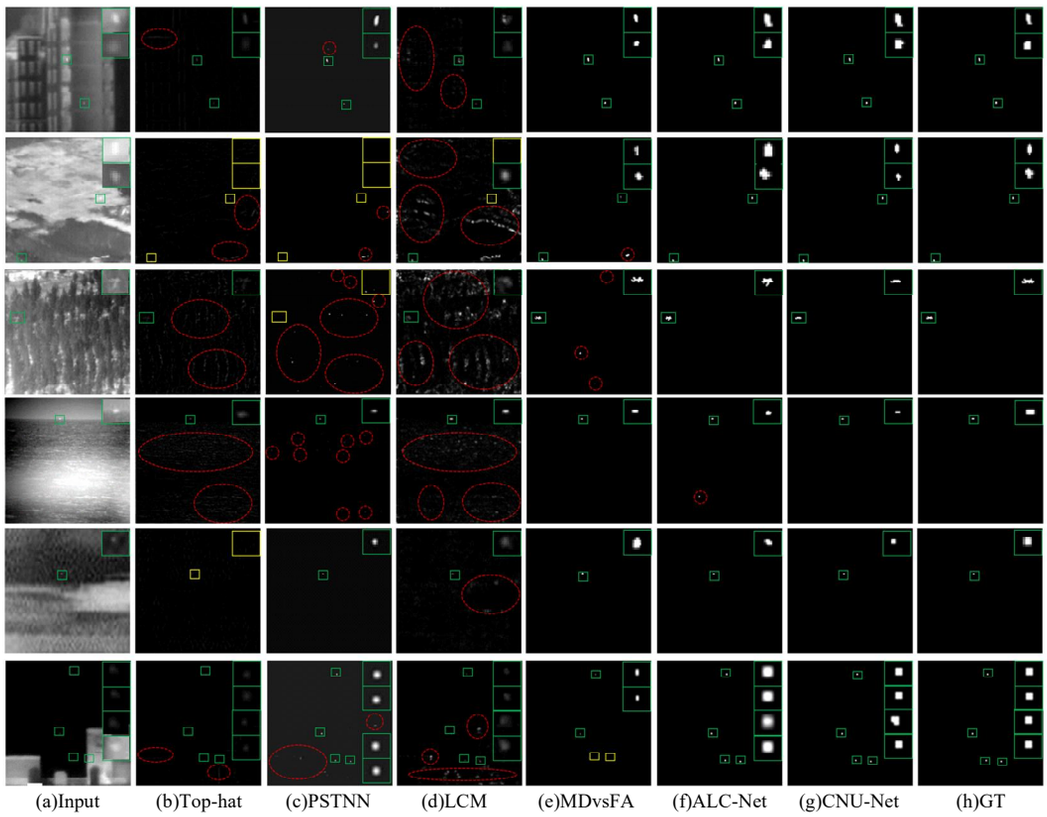

图10 不同算法的红外小目标图像检测结果

Fig. 10 Infrared small target image detection results by different algorithms

小面积虚警区域,无法完成高质量的多尺度小目标检测;本文模型CNU-Net在信噪比低的环境下仍可对不同尺寸的目标实现精准检测,同时有效避免了噪声干扰,降低了虚警率。

3 结束语

本文提出了一种具有级联嵌套结构的检测模型,针对小目标在不同场景下的尺寸形状变化大以及密集背景噪声干扰性强的问题给出了解决方案。具体来说,该模型利用U-Net保留浅层特征的优势来减少小目标特征的丢失;通过级联嵌套结构来完成多尺度小目标检测;采用局部对比度测量来减少密集背景噪声干扰,增强目标特征。从检测结果来看,本文算法可以有效避免漏检、误检等情况;对比实验结果表明:本文算法具有优越的性能,能完成不同的复杂背景下红外小目标的检测任务。

参考文献:

[1] 查月, 汤晔. 红外海面小目标检测算法综述[J]. 宇航计测技术, 2022, 42(6): 57-65.
Zha Yue, Tang Ye. Review of infrared sea surface small target detection algorithm [J]. Journal of Astronautic Metrology and Measurement, 2022, 42(6): 57-65.

[2] 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.

[3] Chen C L P, Li H, Wei Y, et al. A local contrast method for small infrared target detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2013, 52(1): 574-581.

[4] 张文杰, 熊庆宇. 基于对比度与空间位置特征的显著性区域检测[J]. 吉林大学学报: 工学版, 2015, 45(5): 1709-1716.
Zhang Wen-jie, Xiong Qing-yu. Joint segmentation of optic cup and disc based on high resolution network [J]. Journal of Jilin University (Engineering and Technology Edition), 2015, 45(5): 1709-1716.

[5] Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. Advances in Neural Information Processing Systems, 2015, 28: 7485869.

[6] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]// Proceedings of the IEEE Conference on Computer Vi-

sion and Pattern Recognition, Las Vegas, USA, 2016: 779-788.

[7] Wang H, Zhou L, Wang L. Miss detection vs false alarm: adversarial learning for small object segmentation in infrared images[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea (South), 2019: 8509-8518.

[8] Deng C, Wang M, Liu L, et al. Extended feature pyramid network for small object detection[J]. IEEE Transactions on Multimedia, 2021, 24: 1968-1979.

[9] Goodfellow I, Pouget A J, Mirza M, et al. Generative adversarial nets[J]. Advances in Neural Information Processing Systems, 2014, 2: 2672-2680.

[10] Wu X, Hong D, Chanussot J. UIU-Net: U-Net in U-Net for infrared small object detection[J]. IEEE Transactions on Image Processing, 2022, 32:364-376.

[11] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation [C]// Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015: 18th International Conference, Munich, Germany, 2015: 234-241.

[12] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 2117-2125.

[13] Woo S, Park J, Lee J Y, et al. Cbam: convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision, Munich, Germany, 2018: 3-19.

[14] 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, Waikoloa, USA, 2021: 950-959.

[15] Li B, Xiao C, Wang L, et al. Dense nested attention network for infrared small target detection[J]. IEEE Transactions on Image Processing, 2022, 32: 1745-1758.

[16] 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, 59(11): 9813-9824.

[17] Sun Y, Yang J, 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, 2020, 59(5): 3737-3752.