基于扩散模型的红外小目标检测

屠晨浩[1]，叶文亚[1]，杜妮妮[2]，郑彬淏[1]，徐 生[2]
（1. 宁波工程学院 建筑与交通工程学院，浙江 宁波 315211；2. 浙江工商职业技术学院 建筑与艺术学院，浙江 宁波 315100）

摘要：红外小目标检测作为一项复杂且关键的计算机视觉任务，面临着目标尺寸微小、对比度低、背景噪声干扰强烈及数据稀缺等多重挑战，这些问题极大地制约了检测精度与实时性。现有基于深度学习的算法大多基于分割范式，通过设计结构较深的编码器-解码器网络实现分割掩码的生成，由于缺乏足够的特征表示和学习能力，在应对各种复杂场景时检测精度较低。鉴于此，受启发于人工智能领域扩散模型技术所取得的巨大成功，本文提供了一种新的解决思路，将红外小目标检测问题描述为生成式任务，并提出了一个条件去噪网络 diff-ISTD。该网络利用逐步去噪与重建优势，挖掘图像内在深层次统计特性，从而能够更精确地区分并捕获微弱且易于混淆的小目标特征。具体来说，该网络包含条件分支网络以及去噪分支网络，分别用于充分提取红外图像的先验知识和细化含有噪声的掩码。此外，本文还设计了一种并行双维自注意力计算（PDSA）模块，融合空间与通道维度分析，极大增强了模型对全局结构和局部细节的把握力，克服了由分辨率和环境多样性引起的目标模糊难题。综合实验结果显示，diff-ISTD 在面对极端检测条件时，相比目前先进的分割方法，展现出卓越的性能与更高的检测效率，为克服小目标检测领域的长期挑战开辟了新路径。

关键词：红外图像；弱小目标检测；并行双维自注意力机制；扩散模型
中图分类号：TP753 **文献标识码：**A **文章编号：**1001-8891(2025)06-0757-08

Diffusion Model for Infrared Small Target Detection

TU Chenhao[1]，YE Wenya[1]，DU Nini[2]，ZHENG Binhao[1]，XU Sheng[2]
(1. *School of Architecture and Transportation Engineering*, *Ningbo Institute of Technology*, *Ningbo* 315211, *China*;
2. *School of Architecture and Art*, *Zhejiang Business Technology Institute*, *Ningbo* 315100, *China*)

Abstract: Infrared small-target detection, a complex and critical task in computer vision, faces numerous challenges—including tiny target sizes, low contrast, severe background noise, and limited data availability. These factors significantly impair detection accuracy and real-time performance. Existing deep learning–based algorithms, which predominantly adopt segmentation paradigms via deep encoder–decoder architectures for generating segmentation masks, often exhibit limited precision in complex scenarios due to inadequate feature representation and learning capabilities. Inspired by the notable success of diffusion models in artificial intelligence, this paper introduces a novel approach by reframing infrared small-target detection as a generative task and proposes a conditional denoising network, termed diff-ISTD. By leveraging the strengths of progressive denoising and image reconstruction, diff-ISTD captures the deep statistical properties of infrared images, enabling more precise identification of weak and ambiguous small-target features. The proposed network consists of conditional branching modules for extracting prior knowledge from infrared inputs and denoising branches for refining noisy segmentation masks. In addition, a parallel dual-dimensional self-attention (PDSA) block is introduced to integrate spatial and channel information, significantly enhancing the model's sensitivity to global structures and local details. This design effectively addresses the challenges of target blurring caused by resolution limitations and environmental variability. Comprehensive experiments demonstrate that, under rigorous detection conditions, diff-ISTD

收稿日期：2024-05-24；**修订日期**：2024-06-15.
作者简介：屠晨浩（2002-），男，本科，主要研究方向为道路与桥梁工程。E-mail：956250283@qq.com。
通信作者：叶文亚（1974-），女，高级工程师，主要研究方向为图像检测、道路智能化。E-mail: 763425011@qq.com。
基金项目：宁波市交通运输科技计划项目（202216）；宁波市科技计划项目（2024S076）。

outperforms current state-of-the-art segmentation methods in terms of performance and detection efficiency, offering a promising direction for advancing infrared small-target detection technologies.

Key words: infrared images, small target detection, parallel dual-dimension self-attention, diffusion model

0 引言

红外探测技术相较于常规可见光探测，在极端环境如低光照、云雾条件下展现出卓越的抗干扰性能，因而在民用安全监控、应急响应及军事侦察与精确制导等领域发挥着关键作用[1-3]。然而，随着红外成像技术的发展，如何从复杂红外图像中精准辨认与定位微小目标成为研究前沿。由于成像距离较长及硬件条件限制，这些目标占比不到图像的0.15%，且受环境热效应影响，其轮廓与纹理常被背景噪声湮没，造成图像模糊，检测难度极大。

早期传统算法多基于目标与背景的局部差异性设计滤波策略，因此通过设计了一系列滤波器来将目标从背景中分离出来[4]。然而，这类方法在增强微弱目标信号的同时也放大了背景杂讯，导致虚警率偏高。而基于人眼视觉系统感知的方法，例如IPI算法[5]和LCM算法[6-7]，也仅在目标与背景对比鲜明时表现出较好的检测精度。

近年来，得益于红外小目标检测领域数据集[8-9]的发布和深度学习技术的飞速发展，许多基于数据驱动的深度学习方法应运而生。Dai 等人设计了一种专门用于检测红外小目标的非对称上下文调制模块ACM[8]，采用反向自下而上的上下文调制路径，将较小尺度的视觉细节编码到更深的层次，有效增强了网络的检测能力。接着，Dai 等人还在 ACM 的基础上引入了局部对比度设置并提出了 ALCNet[9]，进一步提升了红外小目标检测的准确率。Zhang 等人依据泰勒有限差分理论设计了 ISNet[10]，对不同层次的边缘结构信息进行聚合和增强，以提高目标与背景的对比度，为用数学理论解释提取目标边界信息奠定基础。Wu 等人提出了一种简单却十分有效的框架UIU-Net[11]，通过在 U-Net 网络中嵌套 U-Net 结构，同时在全局和局部尺度上改进上下文表示和多尺度特征的提取。Li 等人提出了一种密集嵌套的注意力网络 DNA-Net[12]，通过重复的特征融合和增强，可以很好地整合和充分利用小目标的上下文信息。Wang 等人基于 U-Net 进行改进，提出了一种注意力引导特征增强网络 AFE-Net[13]，能够有效地从高噪声和低对比度的红外图像中提取和增强小目标的特征。Zhang 等人构建了由精细细节引导的多层次特征补偿（F-MFC）模块以及跨层次特征相关（CFC）模块组成的 FC3-Net[14]。最近一段时间，Transformer 结构在各类计算机视觉任务中取得广泛应用，Liu 等人首先将自注意力机制应用于该任务中[15]，并在检测精度层面取得突破。然而，传统的 Transformer 结构可能会引入大量的计算开销，给实际应用和部署带来挑战。在此基础上，杜等人进行了相关改进工作[16-17]，降低计算量的同时，还提升了网络检测的准确性。

近期，扩散模型在包括 Sora、图像生成[18-19]以及图像编辑[20]等众多计算机视觉任务中取得广泛应用。受此启发，与此前依赖于分割范式的算法不同的是，本文将红外小目标检测任务定义为一个生成范式，并设计了一个条件去噪模型 diff-ISTD。在训练阶段，将高斯噪声添加到干净的目标掩码中，并依据相应的红外图像作为条件先验训练模型来反转这一过程。在推理阶段，该模型通过学习到的去噪模型逐步细化初始噪声掩码，直到它们符合目标图像的分布。具体来说，diff-ISTD 包含一个条件网络，用于从红外图像中提取先验知识，以及一个去噪网络，专注于细化受噪声污染的目标掩码。此外，diff-ISTD 并未采用常规的拼接或者相加操作实现两分支网络信息的交互，而是采用了一种交叉注意力机制[21]，高效地将条件网络中所提取到的相关先验信息融入到去噪网络中。最后，尽管针对红外特征的空间和通道维度上已有广泛研究，但同时结合空间和通道维度的 Transformer 结构来有效模拟深层次的非局部空间及通道相关性的探索仍然有限。本文还提出了一种并行双维 Transformer（parallel dual-dimension self-attention，PDSA）模块作为 diff-ISTD 的基本构建单元。

1 本文方法

1.1 扩散模型介绍

在分布 $p(x,y)$未知情况下，给定由红外图像以及目标掩码样本构成的数据集$(x_i,y_i)_{i=1}^{N}$。鉴于条件分布 $p(y|\ x)$ 的不确定性，本文的目标是通过迭代细化过程学习该分布的一个参数化近似。在本文中，我们将红外小目标检测任务视为一个条件目标掩码生成的过程，并引入一个基于扩散模型的新框架 diff-ISTD。该框架使用纯噪声图像 $Y_T\sim N(0,I)$进行初始化，接着按照学习到的条件分布 $p_\theta(y_{t-1}|\ y_t,x)$ 迭代地优化输出掩码，直到 y_0 服从分布 $p(y|\ x)$。

如图 1 所示，diff-ISTD 包含前向扩散 q 和逆向去噪 p 两个过程。其中，在前向过程中，掩码 y_0 通过马

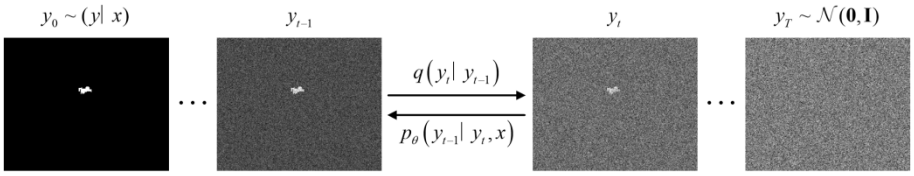

图 1 前向过程与逆向去噪过程示意图

Fig. 1 Illustration of the forward diffusion and reverse denoising processes

尔可夫链逐步加入噪声；逆向去噪过程中，带噪掩码依据条件先验 x 通过逆马尔可夫过程逐步去除噪声。为学习逆向去噪过程，本文设计了去噪模型 f_θ，更多细节如下：

1.1.1 前向扩散过程

受到 Ho 等人工作[22]的启发，给定干净掩码 y_0，其噪声版本$(y_t)_{t=1}^T$ 可以通过迭代添加高斯噪声来获得：

$$q(y_t|\ y_{t-1}) = \mathcal{N}\left(y_t;\sqrt{1-\beta_t}y_{t-1},\beta_t I\right) \tag{1}$$

式中：时间步长 $t\in(1,T)$，$\beta\in(0,1)$为超参数，决定高斯噪声的方差。通过舍弃中间步骤，y_t 的分布可描述为：

$$q(y_t|\ y_0) = \mathcal{N}\left(y_t;\sqrt{\bar{\alpha}_t}y_0,(1-\bar{\alpha}_t)I\right) \tag{2}$$

式中：$\bar{\alpha}_t = \prod_{i=1}^{t}\alpha_t$，$\alpha_t = 1-\beta_t$。此外，为方便后续对 y_0 的预测，已知 y_0，y_t 条件下 y_{t-1} 的后验概率分布为：

$$q(y_{t-1}|\ y_0,y_t) = \mathcal{N}\left(y_{t-1};\mu,\sigma^2 I\right) \tag{3}$$

式中：

$$\mu = \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{1-\bar{\alpha}_t}y_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}y_t$$

$$\sigma^2 = \frac{(1-\bar{\alpha}_{t-1})(1-\alpha_t)}{1-\bar{\alpha}_t}$$

1.1.2 逆向去噪过程

该过程基于纯高斯噪声 y_T 作为输入，沿着与前向扩散过程相反的方向进行，利用去噪模型 f_θ逐步执行去噪操作，最终生成干净的目标掩码 y_0。表示为：

$$p_\theta(y_{t-1}|\ y_t,x) = \mathcal{N}\left(y_{t-1};\mu_\theta(y_t,\bar{\alpha}_t),\sigma_t^2 I\right) \tag{4}$$

式中：分布 $p_\theta(y_{t-1}|\ y_t,x)$ 通过参数θ参数化。

在 diff-ISTD 框架中，我们通过训练网络 $f_\theta(y_t,x,t)$ 来预测最终的干净掩码。依据上述推导，每次迭代表示为：

$$y_{t-1} = \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{1-\bar{\alpha}_t}f_\theta(y_t,x,\bar{\alpha}_t) + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}y_t + \sqrt{1-\alpha_t}\epsilon, \tag{5}$$

式中：ϵ 服从标准正态分布 $N(0,I)$。

1.2 网络结构

如图 2 所示为本文所提出的 diff-ISTD 框架，采用了一种双分支架构，包括一个用于从红外图像中提取先验知识的条件网络，以及一个用于从噪声掩码中去除噪声的去噪网络。此外，作为 diff-ISTD 的核心计算单元，本文将空间维度以及通道维度自注意力机制进行整合，开发出了并行双维自注意力（parallel dual-dimension self-attention, PDSA）模块，能够有效地模拟深层次、长距离的空间和通道相关性。

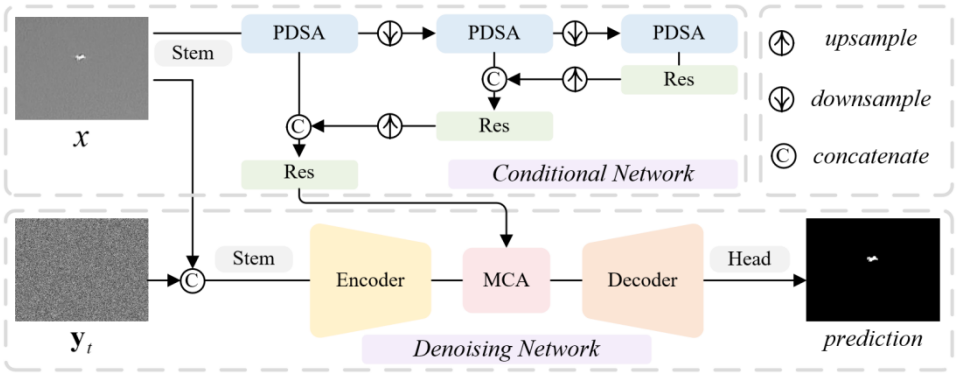

图 2 diff-ISTD 的整体框架：PDSA 表示并行双维自注意力模块，Res 表示残差块，MCA 表示多头交叉注意模块。

Fig. 2 Overall framework of diff-ISTD, where PDSA represents the parallel dual-dimension self-attention block, Res represents the residual block, and MCA represents the multi-head cross-attention module

此外，为了使网络有效感知噪声水平，根据 Ho 等人[22]的做法，通过利用正弦位置编码，将时间步长 t 转换为编码形式，并将其作为输入馈送到每一个 DDT 块中。上述融合过程参照了 Saharia 等人在文献[23]中的做法，出于简化考虑，这部分并未在框架图中展示。

1.2.1 条件网络分支（Conditional network）

图 2 上半区域详细阐述该网络分支的具体结构，本文首先使用基础的 stem 操作[9]从输入的红外图像 x 中提取浅层特征。随后，为了深度挖掘特征的空间与通道维度关联性，本文级联多组不同层级的 PDSA 模块，并通过残差块（Res）[9]配合上采样操作实现多尺度特征的融合。这一系列操作不仅捕获了丰富的多层次特征信息，而且通过综合这些信息，显著增强了模型解析复杂图像构成和细腻表达的能力，为后续的去噪任务奠定了坚实的基础。

1.2.2 去噪网络分支（Denoising network）

如图 2 下半区域所示，去噪网络主体采用了一种编码器-解码器结构。本文首先将噪声掩码 y_t 以及红外图像 x 进行融合并作为该分支的输入，同时为了增强对噪声掩码与红外特征之间空间和通道相关性的建模能力，编码器和解码器均配备了 3 组串联 PDSA 模块。此外，为了有效整合来自条件网络的先验信息，本文引入了多头交叉注意力（MCA）模块[21]，通过交叉注意力的形式实现两种特征的有效融合，从而提升该分支去噪性能。最终，通过 head 操作[9]，获得干净目标掩码的生成。

1.2.3 并行双维自注意力模块（PDSA）

在去噪任务中，特征图中的非局部相似性被作为一种先验知识广泛用于提升图像的去噪效果。然而大多基于 CNN 的算法往往受限于局部感受野导致建模能力有限，难以充分利用这一先验信息。鉴于多头自注意力（MSA）机制在捕捉长距离依赖关系方面的有效性，本文开创性地拓展了其应用范畴，不仅局限于空间维度，同时深入到通道维度，这与以往大多仅侧重空间维度而忽略通道信息处理的常规做法[15-16,24]形成鲜明对比。具体而言，如图 3(a)所示，PDSA 模块中集成了空间维自注意力机制、通道维度自注意力机制以及前馈网络组件[25]。给定该模块的输入为 Z_{in}，可以分别得到：

$$Z_{\text{cat}}=f_{\text{LP}}[f_{\text{spa-MSA}}(f_{\text{LN}}(Z_{\text{in}})),f_{\text{cha-MSA}}(f_{\text{LN}}(Z_{\text{in}}))]+Z_{\text{in}},$$

$$Z_{\text{out}}=f_{\text{FFN}}(f_{\text{LN}}(Z_{\text{cat}}))+Z_{\text{cat}} \tag{6}$$

式中：$[\cdot,\cdot]$表示特征拼接操作；$f_{\text{LP}}(\cdot)$表示线性映射操作；$f_{\text{LN}}(\cdot)$表示层归一化操作；$f_{\text{spa-MSA}}(\cdot)$和$f_{\text{cha-MSA}}(\cdot)$分别为空间维多头自注意力操作和通道维多头自注意力操作。这一设计通过全面利用 Transformer 的核心优势，不仅增强了对空间特征全局上下文的把握，还同时深入剖析了特征在通道层面的复杂交互，共同构建起一个强大且细腻的特征表示模型，并且由于采用了并行的操作，避免了两种特征提取方式的相互干扰，从而在红外小目标检测任务中取得更优异的表现。

空间维自注意力模块的结构如图 3(b)所示，给定输入特征的尺寸为 $H\times W\times C$，为了避免过大的计算开销，首先将其划分为 HW/M^2 个窗口，其中 M 表示窗口大小。令 d 表示 head 的个数，MSA 中每一个 head 会被分配 $C_{\text{h}}=\left\lfloor\frac{C}{d}\right\rfloor$ 通道个数，为简化，本文选择 $Z_{\text{in}}^{\text{spa}}\in R^{M\times M\times C}$ 作为该模块的输入，来展示每一 head 在某一个窗口内的计算过程。首先通过对输入进行线性映射分别获得 q、k 以及 v 分别表示 query、key 以

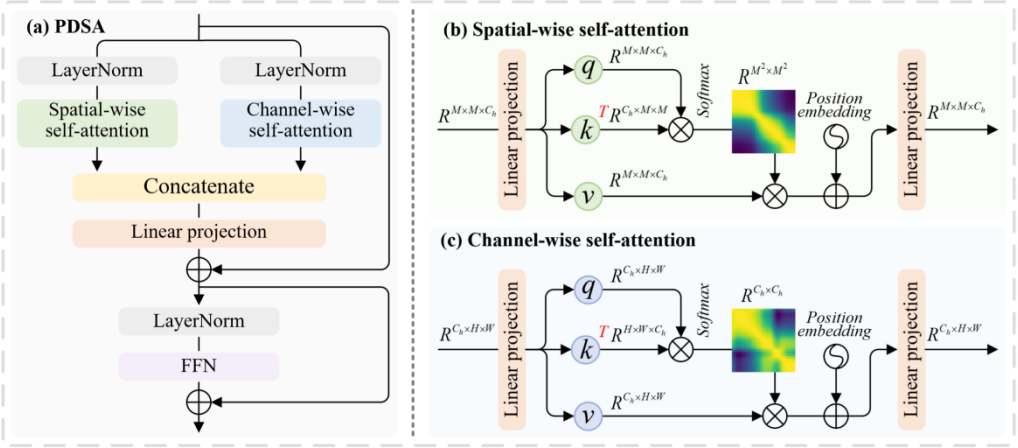

图 3 diff-ISTD 子模块结构示意图：(a)并行双维自注意力模块结构；(b) 空间维自注意力模块；(c) 通道维自注意力模块。

Fig. 3 Illustration of the diff-ISTD submodule structure: (a) The architecture of the PDSA module, (b) the architecture of the spatial-wise self-attention module, and (c) the architecture of the channel-wise self-attention module.

注：红色 T 表示矩阵转置操作。Note: The red T represents the matric transpose operation

及 value。接着，非局部空间自注意矩阵 $\boldsymbol{A}^{\text{spa}} \in R^{M^2 \times M^2}$ 可以表示为：

$$\boldsymbol{A}^{\text{spa}} = \text{Softmax}\left(q \otimes k^T / \sqrt{C_{\text{h}}}\right) \tag{7}$$

该 MSA 的输出 $Z_{\text{out}}^{\text{spa}}$ 可以表示为：

$$Z_{\text{out}}^{\text{spa}} = f_{\text{LP}}(\boldsymbol{A}^{\text{spa}} \otimes v + P^{\text{spa}}) \tag{8}$$

式中：$f_{\text{LP}}(\cdot)$表示线性映射；P^{spa} 表示空间维自注意力模块的位置编码信息，最终通过特征拼接操作将 d 个 head 的输出结果进行合并。

通道维自注意力模块的结构如图 3(c)所示，类似于空间维自注意力模块，首先沿着通道维划分 d 个 head，$C_{\text{h}} = \left\lfloor \frac{C}{d} \right\rfloor$。再根据输入 $Z_{\text{in}}^{\text{cha}} \in R^{C_{\text{h}} \times H \times W}$ 计算得到 q、k 以及 v，并得到每个 head 的自注意力矩阵 $\boldsymbol{A}^{\text{cha}} = \text{Softmax}\left(q \otimes k^T / \sqrt{C_{\text{h}}}\right)$，尺寸为 $R^{C_h \times C_h}$。进一步得到：

$$Z_{\text{out}}^{\text{cha}} = f_{\text{LP}}(\boldsymbol{A}^{\text{cha}} \otimes v + P^{\text{cha}}) \tag{9}$$

式中：P^{cha} 表示通道维自注意力模块的位置编码信息，同时通过特征拼接操作将 d 个 head 的输出结果合并。

2 实验结果与分析

2.1 数据集介绍

与主流的红外小目标检测算法一致，本研究在NUAA-SIRST[8]以及IRSTD-1k[9]两个数据集上进行一系列实验来验证本文所采用扩散模型框架在该任务中的检测性能。其中，NUAA-SIRST 数据集包含 427 张红外图像，覆盖了 480 个小目标实例；IRSTD-1k 数据集中包含有 1000 张红外图像，背景包含多种场景。在本研究中，对于这两个数据集，分别将其中 50%用于训练，20%用于验证，30%用于测试。

2.2 训练环境及实验设置

本研究使用 PyTorch 平台，在 GTX2080Ti GPU 计算设备上对 diff-ISTD 的各项性能进行验证。本文选择了 AdaGrad 作为网络训练时的优化器，初始学习率设定为 0.04，衰减率为 10^{-4}。此外，实验中设置了 batchsize 为 12，共进行了 3000 个 epoch 的训练。

为了有效验证 diff-ISTD 的检测性能，本文选取了一些该领域较为先进的检测算法进行对比，包括：AGPCNet[26]，ALCNet[10]，ACMNet[8]，MDvsFA[27]，WSLCM[7]，TLLCM[6]，IPI[5]，NRAM[28]，PSTNN[29]以及 MSLSTIPT[30]。

2.3 评价指标

为了量化验证本文提出的 diff-ISTD 的性能，我们采用了与该领域主流相同的评价指标，包括 IoU、nIoU、P_{d} 以及 F_{a} 这几个常用指标。其中，IoU 指交并比，表示为：

$$\text{IoU} = A_i / A_{\text{u}} \tag{10}$$

式中：A_i 和 A_{u} 分别表示相交区域和并集区域的大小。

nIoU 表示 IoU 指标的标准化，定义为：

$$\text{nIoU} = \frac{1}{N}\sum_{i=1}^{N}(\text{TP}[i] / (\text{T}[i] + \text{P}[i] - \text{TP}[i])) \tag{11}$$

式中：N 表示总样本数；TP[·]表示模型正确预测为正样本的像素数；T[·]和 P[·]分别表示图像中真实和预测为正样本的像素数目。

P_{d} 为检出率，$P_{\text{d}} = N_{\text{pred}}/N_{\text{all}}$，为正确检测出的目标 N_{pred} 与所有目标 N_{all} 的比值。

F_{a} 为虚警率，$F_{\text{a}} = N_{\text{false}}/N_{\text{all}}$ 表示错误预测目标像素数 N_{false} 与总像素数 N_{all} 的比值。

2.4 消融实验

为了评估 diff-ISTD 中所采用的各个模块的有效性以及对整体性能的影响，本文在 NUAA-SIRST 数据集上进行一系列消融实验，结果如表 1 所示。

1）条件扩散模型有效性验证：为了验证本文引入的条件扩散模型相较于传统回归模型在红外小目标检测任务中的优越性，我们在保留网络结构的基础上，将 diff-ISTD 修改为回归范式。实验结果如表 1 中实验 1 所示，与采用完整条件扩散模型的 diff-ISTD（实验 8）相比，基于回归范式的检测网络在 IoU 指标上下滑了 2.83，从而有效证明了条件扩散模型在红外小目标检测任务中的有效性。

2）并行双维自注意力模块有效性验证：作为 diff-ISTD 网络的核心结构，通过全面利用 Transformer 结构的核心优势，不仅强化了对空间维度全局上下文的理解，还深化了对通道间复杂特征互动的解析。为确保该模块的有效性，本文首先将 PDSA 替换为传统的基于 CNN 结构的残差块结构[9]，结果如表 1 中实验 3 所示，通过与完整网络（实验 8）的结果进行对比发现，替换后的结果在多个性能指标上均有所下降，其中 IoU 指标降低了 1.58；此外，本文还将 PDSA 替换为 Squeeze-and-Excitation (SE) attention[31]模块，结果如表 1 中实验 4 所示，各项指标相比于原始 diff-ISTD 结构仍存在一定差距。归功于自注意力计算模型相比于 CNN 结构在捕捉远距离依赖关系方面更为有效，从而有效提升了小目标的检测精度。除了上述将 PDSA 模块替换为一些基于 CNN 的结构以外，本文还验证了各类基于 Transformer 结构的自注意力

机制在本文所提出扩散模型框架下的测试结果，表 1 中实验 5，6，7 分别表示将 PDSA 模块替换为基于空间维 transformer 结构、通道维 transformer 结构以及串联形式的空间维和通道维 transformer 结构，可以看到，相比于本文所提出的 PDSA，以上结构均无法充分发挥基于扩散模型的红外小目标检测框架的性能，从而导致检测精度受限。而 PDSA 模块由于强大且细腻的特征表征能力，取得了更优的检测结果，证明其在该任务上的有效性。

综合以上消融实验结果，充分验证了本文采用的扩散模型架构以及 PDSA 模块在提升 diff-ISTD 红外小目标检测性能方面的重要性和必要性。

2.5 实验结果

为了评估本文提出的 diff-ISTD 在红外小目标检测任务中的表现，我们将其与几种经典算法进行了比较，我们在 NUAA-SIRST 以及 IRSTD-1k 这两个红外数据集上分别进行对比实验，比较结果如表 2 所示。传统算法往往过度依赖手工设置的先验信息，在处理具有挑战性的样本时性能受到挑战，并且与基于深度学习的方法相比存在明显差距。基于卷积神经网络（CNN）构建的算法由于表达能力有限且缺乏对全局信息的有效建模，因此难以准确定位和识别小目标，各项评价指标的数值较低。此外，这些算法在噪声背景下的学习判别能力也较弱，容易导致小目标的漏检和误检。相比之下，本文所提出的 diff-ISTD 算法由于基于扩散模型范式，利用其独特的逐步去噪与重建优势，挖掘图像内在深层次统计特性，从而能够更精确地区分并捕获微弱且易于混淆的小目标特征，在两个数据集上的所有评价指标上都表现出最佳性能。本文还通过如图 4 所示的可视化方法对比了不同算法的检测结果，即使在一些具有低信噪比和低对比度的红外图像输入情况下，diff-ISTD 仍能准确定位目标，并且检测到的目标形状基本完整且准确。最后，本文还通过绘制两个数据集的 ROC 曲线（如图 5 所示），表明 diff-ISTD 网络的性能同样优于其他算法。

表 1 消融实验结果

Table 1 Ablation study results

Experiment	Model	IoU	nIoU	P_d	F_a
1	w/o diffusion+PDSA	70.04	69.38	93.26	39.45
2	w/o diffusion	71.62	70.83	95.42	32.02
3	w/o PDSA	72.87	70.65	96.72	28.75
4	diffusion+SE attention	72.94	70.41	96.88	32.42
5	diffusion+spatial-wise transformer	73.16	71.06	97.45	28.07
6	diffusion+channel-wise transformer	72.89	70.82	97.78	26.92
7	diffusion+spatial & channel-wise transformer	74.01	71.98	98.33	25.53
8	diff-ISTD	74.45	72.81	98.52	20.13

表 2 对比实验结果

Table 2 Experimental results on different algorithms

Method	NUAA-SIRST				IRSTD-1k				Params
	IoU	nIoU	P_d	F_a	IoU	nIoU	P_d	F_a	
WSLCM	4.41	33.82	91.74	22593	3.45	0.68	72.44	6619	-
TLLCM	3.51	21.75	92.66	26498	3.31	0.78	77.39	6738	-
IPI	2.62	4.16	84.40	203.07	27.92	20.46	81.37	16.18	-
NRAM	45.68	55.49	85.32	161.15	15.25	9.90	70.68	16.93	-
PSTNN	51.95	62.66	82.57	394.29	24.57	17.93	71.99	35.25	-
MSLSTIPT	20.21	24.74	82.57	259.75	11.43	5.93	79.03	1524	-
MDvsFA	45.28	48.16	76.15	166.07	42.45	44.31	83.21	78.54	3.77 M
ACM	67.96	71.05	97.25	72.92	58.64	56.94	90.42	23.57	0.39 M
ALCNet	73.43	71.44	97.84	25.68	61.02	57.98	91.24	26.53	0.38 M
AGPCNet	74.26	70.05	98.16	20.56	61.53	58.32	92.02	24.43	12.36 M
diff-ISTD	74.45	72.81	98.52	20.13	62.65	60.18	93.53	21.03	0.29 M

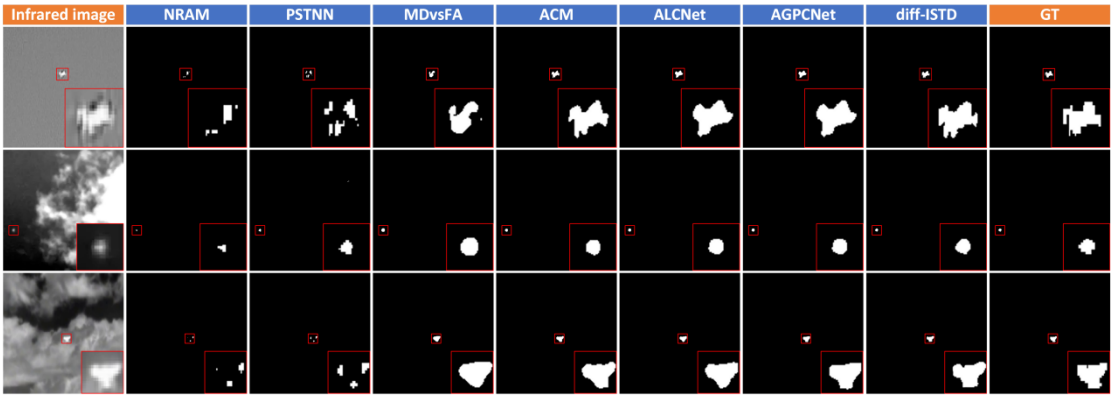

图 4　不同算法在 NUAA-SIRST 数据集上红外图像检测结果

Fig. 4　Infrared image detection results of different algorithms on NUAA-SIRST datasets

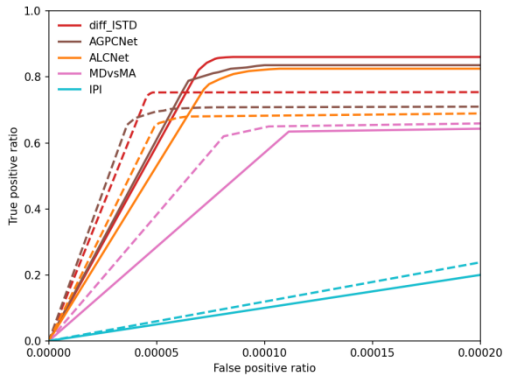

图 5　不同算法在 NUAA-SIRST（实线）以及 IRSTD-1k（虚线）数据集上的 ROC 曲线

Fig.5　ROC curves of different algorithms on NUAA-SIRST (solid line) and IRSTD-1k datasets (dotted line)

3　结论

本文针对现有红外小目标检测算法难以克服红外图像中目标尺寸微小、对比度低以及背景噪声干扰强烈等挑战，不同于以往基于分割范式的红外小目标检测算法在应对复杂场景存在着特征表示和学习能力不足的问题，本文基于扩散模型，提出了一种新型的条件去噪网络 diff-ISTD。该算法利用逐步去噪与重建优势，能够挖掘图像内在深层次统计特性，从而能够更精确地区分并捕获微弱且易于混淆的小目标特征。同时，为了提升模型对于红外特征的全局结构和局部细节的敏感性，本文还提出一种融合了空间与通道维度的并行双维自注意力计算（PDSA）模块作为本模型的骨干结构。通过在公开数据集上进行广泛试验，与其他现有先进的红外小目标检测算法相比，本文所提出的 diff-ISTD 实现了较高的检测精度，证明了该算法的有效性，为红外小目标检测任务在算法设计层面提供了新思路。

参考文献：

[1] ZHAO M, LI W, LI L, et al. Three-order tensor creation and tucker decomposition for infrared small-target detection[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, **60**: 1-16.

[2] ZHAO M, LI L, LI W, et al. Infrared small-target detection based on multiple morphological profiles[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, **59**(7): 6077-6091.

[3] ZHANG J, TAO D. Empowering things with intelligence: a survey of the progress, challenges, and opportunities in artificial intelligence of things[J]. *IEEE Internet of Things Journal*, 2020, **8**(10): 7789-7817.

[4] Deshpande S D, Er M H, Venkateswarlu R, et al. Max-mean and max-median filters for detection of small targets[C]//*Signal and Data Processing of Small Targets*, *Proc. of SPIE*, 1999, **3809**: 74-83.

[5] GAO C, MENG D, 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] HAN J, 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*, 2019, **17**(10): 1822-1826.

[7] HAN J, Moradi S, Faramarzi I, et al. Infrared small target detection based on the weighted strengthened local contrast measure[J]. *IEEE Geoscience and Remote Sensing Letters*, 2020, **18**(9): 1670-1674.

[8] 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*, 2021: 950-959.

[9] 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.

[10] ZHANG M, ZHANG R, YANG Y, et al. ISNet: Shape matters for infrared small target detection[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022: 877-886.

[11] 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.

[12] 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.

[13] WANG K, WU X, ZHOU P, et al. AFE-Net: Attention-guided feature enhancement network for infrared small target detection[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024, **17**: 4208-4221.

[14] ZHANG M, YUE K, ZHANG J, et al. Exploring feature compensation and cross-level correlation for infrared small target detection[C]// *Proceedings of the* 30t*h ACM International Conference on Multimedia*, 2022: 1857-1865.

[15] LIU F, GAO C, CHEN F, et al. Infrared small and dim target detection with transformer under complex backgrounds[J]. *IEEE Transactions on Image Processing*, 2023, **32**: 5921-5932.

[16] 杜妮妮，单凯东，卫莎莎. LPformer: 基于拉普拉斯金字塔多级Transformer的红外小目标检测[J]. **红外技术**, 2023, **45**(6): 630-638.
DU Nini, SHAN Kaidong, WEI Shasha. LPformer: Laplacian pyramid multi-level transformer for infrared small target detection[J]. *Infrared Technology*, 2023, **45**(6): 630-638.

[17] 杜妮妮，单凯东，王建超. HRformer：基于多级回归Transformer网络的红外小目标检测[J]. **红外技术**, 2024, **46**(2): 199-207.
DU Nini, SHAN Kaidong, WANG Jianchao. HRformer: hierarchical regression transformer for infrared small-target detection[J]. *Infrared Technology*, 2024, **46**(2): 199-207.

[18] Samuel D, Ben-Ari R, Raviv S, et al. Generating images of rare concepts using pre-trained diffusion models[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*, 2024, **38**(5): 4695-4703.

[19] LI Y, WANG H, JIN Q, et al. Snapfusion: text-to-image diffusion model on mobile devices within two seconds[J]. *Advances in Neural Information Processing Systems*, 2023, **36**: 20662-20678.

[20] ZHANG Z, LI B, NIE X, et al. Towards consistent video editing with text-to-image diffusion models[J]. *Advances in Neural Information Processing Systems*, 2023, **36**: 58508-58519.

[21] CHEN C F R, FAN Q, Panda R. Crossvit: cross-attention multi-scale vision transformer for image classification[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021: 357-366.

[22] HO J, JAIN A, Abbeel P. Denoising diffusion probabilistic models[J]. *Advances in Neural Information Processing Systems*, 2020, **33**: 6840-6851.

[23] Saharia C, Ho J, Chan W, et al. Image super-resolution via iterative refinement[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 2022, **45**(4): 4713-4726.

[24] XI Y, ZHANG J, LIU K. Nanetformer: nested attention network with auxiliary transformer enhancement for infrared small target detection[C]//*IGARSS* 2023-2023 *IEEE International Geoscience and Remote Sensing Symposium*, 2023: 6596-6599.

[25] Zamir S W, Arora A, Khan S, et al. Restormer: efficient transformer for high-resolution image restoration[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022: 5728-5739.

[26] ZHANG T, LI L, CAO S, et al. Attention-guided pyramid context networks for detecting infrared small target under complex background[C]//*IEEE Transactions on Aerospace and Electronic Systems*, 2023, **59**(4): 4250-4261. doi: 10.1109/TAES.2023.3238703

[27] WANG Huan, ZHOU Luping, WANG Lei. Miss detection vs. false alarm: adversarial learning for small object segmentation in infrared images[C]//*Proceedings of* 2019 *IEEE/CVF International Conference on Computer Vision* (ICCV), 2019: 8508-8517. doi: 10.1109/ICCV.2019.00860.

[28] ZHANG L, PENG L, ZHANG T, et al. Infrared small target detection via non-convex rank approximation minimization joint l 2, 1 norm[J]. *Remote Sensing*, 2018, **10**(11): 1821.

[29] ZHANG L, PENG Z. Infrared small target detection based on partial sum of the tensor nuclear norm[J]. *Remote Sensing*, 2019, **11**(4): 382.

[30] 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.

[31] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 7132-7141.