

利用扩散先验的高斯无知表示学习以增强红外小目标检测

Junyao Li^a, Yahao Lu^a, Xingyuan Guo^a, Xiaoyu Xian^b, Tiantian Wang^c and Yukai Shi^{a,*}

^aSchool of Information Engineering, Guangdong University of Technology, Guangzhou, 510006, China

^bCRRC Institution, Beijing 100000, China

^cGuangzhou National Laboratory, Guangzhou 510006, China

ARTICLE INFO

Keywords:

Self-supervised learning

Infrared

Data-centric Representation Learning

ABSTRACT

红外小目标检测 (ISTD) 在众多实际应用中发挥着至关重要的作用。为了确定性能边界, 研究人员使用大量且昂贵的手动标记数据进行表示学习。然而, 这种方法使得最先进的 ISTD 方法在实际挑战中显得非常脆弱。在本文中, 我们首先研究了在各种稀缺性条件下——即高质量红外数据的缺乏——几种主流方法的检测性能变化, 这对当前关于实际 ISTD 的理论构成挑战。为了解决这一问题, 我们引入了高斯无知表征学习。具体来说, 我们提出了高斯组挤压器, 利用高斯采样和压缩进行非均匀量化。通过利用多样化的训练样本, 我们提高了 ISTD 模型面临各种挑战时的弹性。然后, 我们引入了两阶段扩散模型用于现实世界重建。通过将量化信号与现实世界分布密切对齐, 我们显著提升了合成样本的质量和保真度。在各种稀缺性情境下, 与最先进的检测方法的比较评估证明了所提方法的有效性。

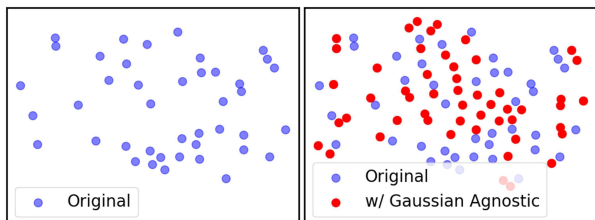


Fig. 1: 红外图像特征的 t-SNE (van der Maaten and Hinton, 2008) 可视化。蓝色点代表来自原始 SIRST (Dai et al., 2021a) 数据集的样本, 而红色点代表通过高斯无关表示学习增强的样本。显然, 增强后的红色点有助于生成比原始蓝色样本更丰富、紧凑和多样的特征点。

1. 介绍

红外小目标检测 (ISTD) (Zhang et al., 2025; Shen et al., 2025; Goodall et al., 2016) 在众多实际应用中发挥着至关重要的作用, 特别是在视频监控 (Xiao et al., 2023; Hu et al., 2023; Xiao et al.)、预警系统 (Deng et al., 2016) 和环境监测 (Rawat et al., 2020) 领域。为了有效检测红外图像中的小目标, 研究人员提出了多种基于滤波器和神经网络的方法来处理真实世界的数据 (Lu et al. (2024); Shi et al. (2024))。与传统方法相比, 基于神经网络的 (Hong et al., 2021) 方法从昂贵且繁多的手动标注中学习红外小目标的特征表示, 从而能够更好地定位这些目标。

深度学习方法通常依赖于如此广泛的大规模训练数据集。然而, 获取足够数量的高质量标记数据仍然是一个主要挑战。在海洋监测 (Teutsch and Krüger, 2010; Ying et al., 2022) 的背景下, 小目标如远程舰艇或无人机, 往往在复杂背景环境中表现出低对比度。在复杂环境中——如海浪、植被和云层形成——噪声可以很容易地掩盖小目标 (Wang et al., 2022), 这进一步使得它们的检测和准确标记变得复杂。此外, 小目标的检测常常受到其距离远以及成像传感器分辨率有限的阻碍。总体而言, 这些物理挑战显著阻碍了数据获取的可行性。

因此, ISTD 技术在数据稀缺的情况下尤为关键。然而, 大多数现有的 ISTD 模型在极端条件下表现不佳。在现实场景中, 获得足够且高精度的数据难以实现。数据不足使得这些模型无法有效学习基本目标特征, 导致检测精度显著下降。此外, 由于复杂和动态的背景结合非常低的目标对比度, 从有限样本中提取稳定且区分度高的特征变得异常困难。因此, 在少样本学习场景下, 现有方法的稳健性和准确性受到显著限制。为了定量评估数据的影响, 我们利用 SIRST (Dai et al., 2021a) 数据集 30% 和 10% 的训练数据模拟训练场景。?? 展示了在全尺度和少样本场景下几个主流方法检测性能的变化。这些结果定量表明, 模型可以在有足够数据的情况下有效学习特征并表现出优越性能。相比之下, 现有方法在少样本场景下表现出较差的泛化能力和不稳定的性能。如 ?? 所示, 有时, 最先进的模型在学习足够的特征表示时遇到困难, 导致检测精度急剧下降。

为了使 ISTD 模型在数据稀缺的情况下保持鲁棒性, 我们提出了高斯不敏感表示学习。这个方法旨在增加用于红外小目标的可用数据, 并提升检测模型在各种挑战中的表现。具体而言, 我们提出了基于高斯采样和压缩的高斯组压缩器。这个模块执行非均匀量化操作, 以为扩散模型生成大量训练样本。扩散模型包括两个阶段: 粗重建阶段和扩散阶段。同时, 粗重建阶段学习从量化图像到重建图像的映射表示, 便利初始像素复原。此外, 我们训练扩散阶段重新采样重建的数据, 从而进一步将其与真实世界分布对齐。我们的目标是通过使用训练过的生成模型生成高质量的合成数据来扩大 IR 数据集。如 Fig. 1 所示, 相比于代表原始 SIRST (Dai et al., 2021a) 样本的蓝点, 这些合成样本展示了一个更为紧凑的特征分布。生成的数据可以被整合到任意模型中, 并带来增强的 ISTD 方法。?? 展示了所提出的增强策略在改进表示学习和检测性能方面的有效性。面对极端挑战, 我们的方法显著提高了从小样本到全规模场景中的检测性能。主要贡献如下:

- 我们引入了高斯不可知表示学习, 专为解决红外小目标检测 (ISTD) 模型中的数据稀缺问题而定制。该高级技术采用高斯组压缩器, 利用高斯采样和压缩进行非均匀量化。通过生成多样化的训

*Corresponding author

✉ <ykshi@gdut.edu.cn> (Y. Shi)

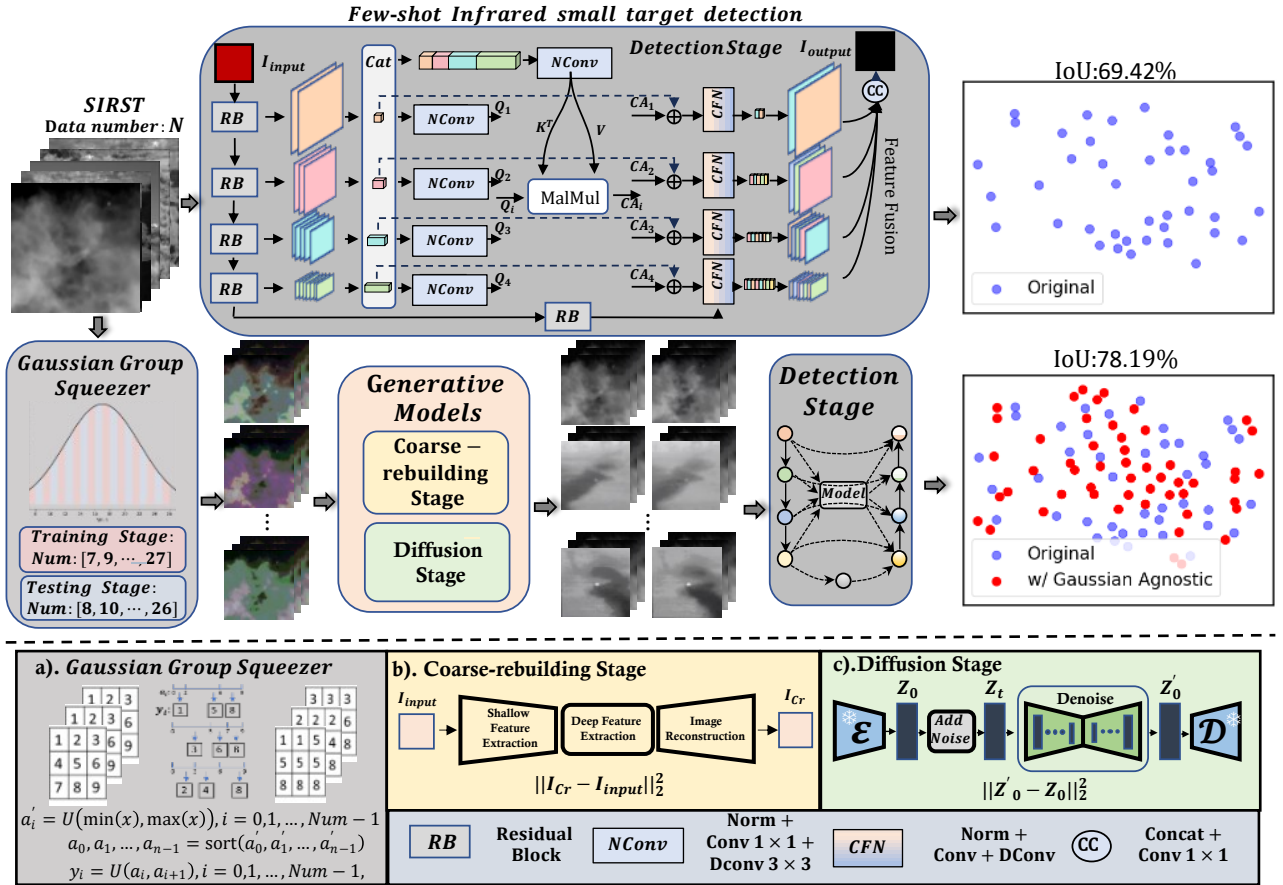


Fig. 2: 高斯无关表示学习。我们的方法使用高斯无关策略生成大量训练数据。通过将高斯无关表示学习与生成模型结合，ISTD 的性能显著提高，在各种极端挑战中达到更高的 IoU 分数。在训练阶段，框架包含三个阶段：a) 高斯组压缩器：从高斯分布中采样参数以量化图像。b) 粗重建阶段：融合浅层和深层特征以初步重建量化内容。c) 扩散阶段：利用 VQ-VAE (Van Den Oord et al., 2017) 编解码隐藏变量，通过扩散先验生成逼真的内容。在推理阶段，我们打乱高斯参数化的区间以强制扩散模型生成多样但统一的数据。生成样本（红点）的特征分布比原始样本（蓝点）更紧凑，有助于实现通用的表示学习。

训练样本，我们增强了 ISTD 模型在各种挑战下的弹性，并通过高质量的合成数据扩展了红外数据集。

- Ö 用于真实世界重建的两阶段生成模型。生成模型学习将量化图像与真实世界分布紧密对齐，从而促进精确的像素重采样。该集成过程显著提高了合成样本的质量和真实性。
- Ö 在全尺度和少样本情境中，与最先进的检测方法相比的对比评估展示了所提方法的有效性。

2. 相关工作

2.1. 单帧红外小目标检测

单帧红外小目标检测与常规物体检测相比，红外小目标检测具有几个独特的特征：1) 变化的背景杂波：云层、建筑物边缘和树线会产生高频噪声。来自道路、机器或海浪的动态热信号。2) 低目标-背景对比度：在长距离上有散射和吸收。小目标辐射的热量有限 (Li et al., 2022)。3) 小目标尺寸：由于成像距离较远，红外目标通常较小，在图像中范围从一个像素到几十个像素 (Li et al., 2022)。

在复杂背景中的精确分割 ISTD Peng et al. (2025); Yang et al. (2025) 仍然是一个具有挑战性的任务。红外成像非常容易受到噪声干扰、小目标尺寸以及较远成像距离的影响，这显著增加了检测的难度。传统方法，如 Top-hat (Rivest and Fortin, 1996)、LCM (Chen et al., 2013) 和 IPI 模型 (Gao et al., 2013)，由于缺乏语义信息的利用，在从复杂背景中有效分离小目标方面存在困难。

相比之下，基于 CNN 的模型 Kumar and Singh (2025); Zhong et al. (2024); Zhang et al. (2025) 具有提取高级语义特征的能力，从而能够更有效地识别红外小目标。MDvsFA-GAN (Wang et al., 2019) 使用生成对抗网络执行多模态数据转换。ISTDU-Net (Xi et al., 2023) 通过特征图分组来增强小目标特征。ACM (Dai et al., 2021b) 结合全局上下文与局部细节以提高检测精度。DNA-Net (Li et al., 2022) 利用密集区域嵌套和注意机制来捕捉多尺度特征。UIU-Net (Wu et al., 2022) 整合了局部和全局信息。ISNet (Zhang et al., 2022) 强调目标形状特征以提高检测精度。MSHNet (Liu et al., 2024) 适用于目标尺寸变化较大的场景。SpirDet (Mao et al., 2024) 通过高效且轻量化的设计优化了红外小目标检测。SCAFNet (Zhang et al., 2024) 通过语义引导的融合提高了目标识别精度。SCTransNet (Yuan et al.,

Table 1

在 NUDT-SIRST 数据集上的 30 % 少样本设置下，扩散基础的 ISTD 方法比较。我们的方法采用扩散进行数据增强，并表现出卓越的性能。

Method	Role of Diffusion	Architecture	IoU	P_d	F_a
ISTD-Diff	Detection	Attn + Diffusion	-	-	-
Diff-Mosaic	Data Augmentation	Mosaic + Diffusion	89.82	98.41	12.70
Ours	Data Augmentation	GGs + 2-Stage Diff	93.39	98.94	2.06

Table 2

SIRST (Dai et al., 2021a) 和 NUDT-SIRST (Li et al., 2022) 数据集的主要特征

Datasets	Target size percent of image area		
	0 $\leq$ 0.03 %	0.03 $\leq$ 0.15 %	>0.15 %
SIRST	58 % (percent of images)	35 %	7 %
NUDT-SIRST	58 %	38 %	4 %

2024) 利用空间和通道注意机制来增强检测性能。尽管这些基于 CNN 的方法取得了显著进步，但获取红外图像数据的高成本、繁琐的标注过程以及有限的数据可用性仍然是阻碍性能进一步提升的主要瓶颈。

近年来，去噪扩散概率模型 (DDPM) 在图像生成领域取得了显著突破。DDPM 的成功显著推动了高质量、逼真图像生成技术的发展。为了进一步提升性能，研究人员提出了潜在扩散模型 (LDM)。LDM 不仅能生成高质量的图像，还结合了真实的先验知识，展示了其在多个应用场景中的强大能力。通过自编码器将数据压缩到低维潜在空间，LDM 显著提高了训练效率和生成质量。ControlStyle 将文本和图像数据结合起来，通过扩散先验实现了文本驱动的高质量风格化图像生成。这些应用表明，基于扩散的生成技术在图像处理和创意内容生成中具有巨大的潜力和广泛的适用性。

最近，一些研究人员将强大的扩散模型引入了 ISTD 任务，以提升模型性能。ISTD-diff (Du et al., 2024) 通过从噪声中迭代生成目标掩码，将 ISTD 任务框定为生成任务。然而，这种方法仅带来了有限的检测性能提升，并增加了计算复杂度，导致推理时间减少。Diff-mosaic (Shi et al., 2024) 方法旨在通过利用扩散先验来有效解决数据增强方法在多样性和保真度方面的挑战。然而，在小样本条件下，缺乏足够的数据和有效的数据增强技术来微调扩散模型，可能导致模型过拟合和随后的性能下降。

为了验证我们方法在少样本场景中的有效性，我们在相同的训练流程下，使用仅 30 % 的 NUDT-SIRST 训练数据来评估所有竞争方法。如 Table 1 所示，我们的方法实现了最高的交并比 (IoU) 和检测概率 (P_d)，同时与 Diff-Mosaic 相比，显著降低了虚警率 (F_a)。

3. 方法论

3.1. 问题定义



Fig. 3: 现实世界中的红外场景表现出严重的不均匀背景。

红外小目标检测 (ISTD) 面临三大基本挑战：异质背景杂波、极低的目标与背景对比度，以及亚像素目标尺度。如在 Table 2 中所证明，超过 90 % 的目标在 SIRST (Dai et al., 2021a) 和 NUDT-SIRST 数据集中占据的图像面积不到 0.15 %。关键的信杂比 (SCR) 度量进一步量化了对比度的退化：SIRST (Dai et al., 2021a) 的 SCR 平均值为 6.24，而 NUDT-SIRST 下降到 1.83 (接近检测边界)。Fig. 3 以视觉形式印证了这些问题，其中复杂背景 (例如，云团、海面反射) 显示出高强度变化，使亚像素目标难以辨识。

ISTD 方法通常依赖于大规模、高质量的训练数据集。然而，它们在现实世界场景中的性能往往会显著下降。为了定量评估这种效果，我们在 SIRST (Dai et al., 2021a) 数据集上进行了实验，使用训练数据的 30 % 和 10 % 来模拟每种极端情况。使用 30 % 训练数据时，SCTansNet (Yuan et al., 2024) 的 IoU 从 76.00 % 下降到 69.42 %，DNANet (Li et al., 2022) 从 76.48 % 下降到 74.56 %，ACM (Dai et al., 2021b) 从 66.14 % 下降到 62.91 %。在使用 10 % 训练数据时，SCTansNet (Yuan et al., 2024) 进一步减少到 59.74 %，DNANet (Li et al., 2022) 减少到 56.49 %，ACM (Dai et al., 2021b) 减少到 20.84 %。结果表明，高质量的训练数据能够使模型更有效地学习特征。相反，样本较少时，模型的泛化能力容易受到侵犯，导致在 ISTD 中表现脆弱。

3.2. 高斯不可知表示学习

我们提出了一种基于生成模型的高斯无关表示学习方法，用于解决在各种极端挑战中 ISTD 的性能瓶颈。我们的框架由三个模块组成：高斯组压缩器、粗重建阶段和扩散阶段。

3.3. 高斯群挤压器

高斯群压缩器通过从高斯分布中采样参数 Num 并将其与扩散模型相结合来以不同程度压缩图像，以生成未知的增强样本。压缩器离散化连续像素值，从而减少数据复杂性。不同的区间数量对应不同的量化水平。Fig. 4 的上半部分以三个区间为例说明图像的量化。具体来说，非均匀压缩器具有随机分配的区间大小。它包括一组不重叠的区间 $S_i = [a_i, a_{i+1})$ ，一个对应的复制值集 y_i ，以及代表区间数量的参数 Num 。

$$a_0, a_1, \dots, a_{Num*1} = \text{sort}(a_0, a_1, \dots, a_{Num*1}) \quad (1)$$

$$a_i = U(\min(x), \max(x)), i = 0, 1, \dots, Num * 1, \quad (2)$$

其中 U 表示在区间内的随机采样， $[\min(x), \max(x)]$ 表示图像 x 每个通道的最小/最大像素值。

$$y_i = U(a_i, a_{i+1}), i = 0, 1, \dots, Num * 1, \quad (3)$$

从 $U(a_i, a_{i+1})$ 中随机抽取 y_i ，替换区间内的像素 x 以生成量化图像。如 Fig. 7 所示，较大的量化参数保留了更多的图像信息，可能导致生成模型训练时的过拟合。较小的量化参数导致更大的图像细节损失，阻碍模型的收敛和泛化。为解决此问题，我们从高斯分布中采样量化参数，以捕获零均值的中等量化图像，从而减少极端量化实例并确保平衡的样本大小。通过采样不同的量化参数，生成不同压缩水平的图像。这些量化图像被用作生成模型训练和测试的输入，以生成大量目标域数据。

由于对红外小目标图像的直接量化常常导致小目标的丢失，从而阻碍生成模型对这些小目标周围像素的重

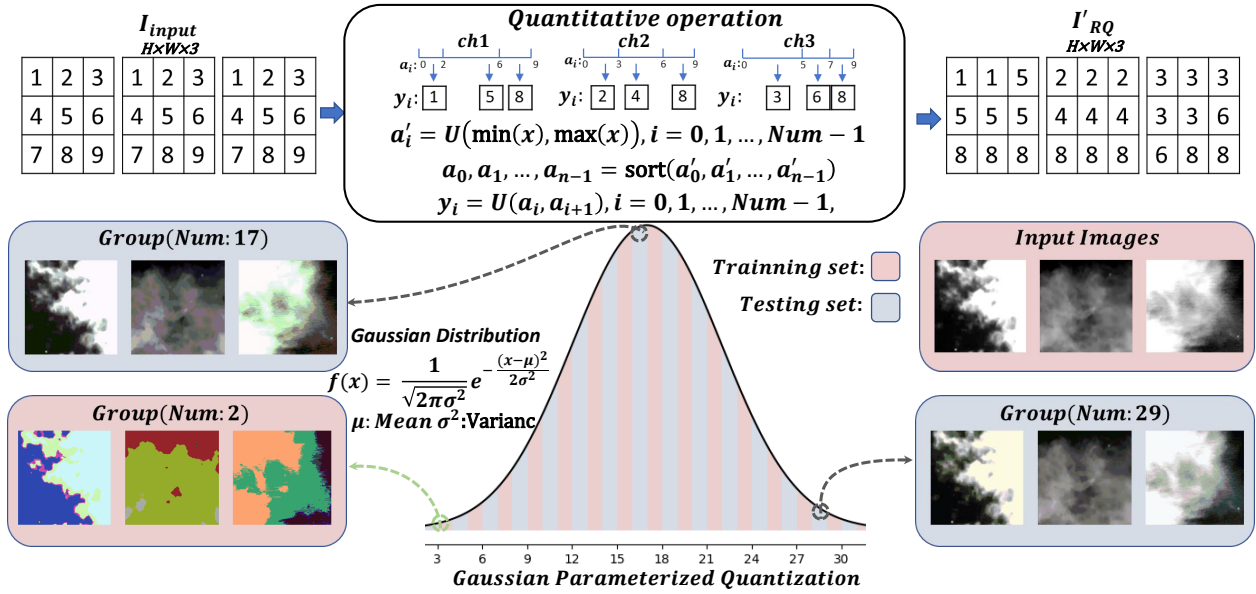


Fig. 4: 使用高斯群压缩器对红外数据进行多样化。我们设计了一种非均匀量化器，随机化高斯区间大小，以实现交叉峰值采样策略。输入图像的像素值首先被采样并排序。接下来，从每个相邻区间 $U(a_i, a_{i+1})$ 中选择随机值，并用作复制值 y_i ，替换其对应的像素 y_i 以生成量化后的图像。Num 从高斯分布中采样。我们在训练和测试阶段应用不同的量化参数作为生成模型的输入。

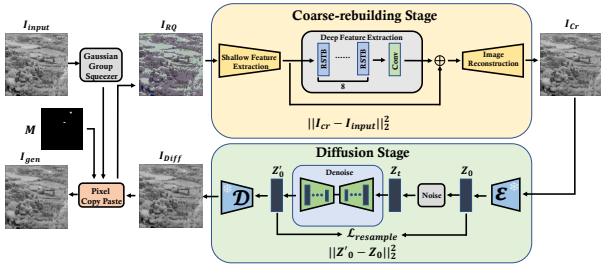


Fig. 5: 用于不可知表示学习的双阶段生成模型。量化图像 I_{RQ} 首先通过粗重建阶段以修复量化损坏的像素，生成初步恢复的图像 I_{Cr} 。随后， I_{Cr} 进入扩散阶段，在此阶段通过扩散过程学习高质量图像分布，进而生成综合图像 I_{Diff} 。最后，使用像素复制粘贴模块将小目标像素复制到生成的图像中，以获得最终的综合图像 I_{gen} 。

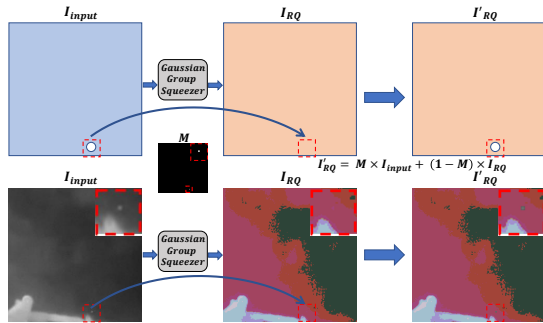


Fig. 6: 像素复制粘贴。量化操作仅应用于红外图像的背景，而小目标的像素保持未量化，例如 I'_{RQ} 。

建。在量化过程中，图像的背景会进行量化处理，而小目标像素则保持不变。

$$I_{RQ} = \begin{cases} y_i & \text{if } M_{j,k} \neq 1 \text{ and } I_{j,k} \in [a_i, a_{i+1}) \\ I_{j,k} & \text{otherwise} \end{cases} \quad (4)$$

在这里， $I_{j,k}$ 表示位置 (j,k) 处的原始像素值， $M_{j,k} \in \{0,1\}$ 是一个二值掩码，指示像素是属于小目标 ($M_{j,k} = 1$) 还是背景 ($M_{j,k} = 0$)。 y_i 是区间 $[a_i, a_{i+1})$ 的代表量化值。量化仅应用于背景像素 ($M_{j,k} \neq 1$)，而小目标像素则被保留以维持细节的完整性。在 Fig. 6 中，对整个图像进行量化会导致小目标的丢失，而仅量化背景同时保留小目标像素则允许模型更有效地编码周围的像素。由于小目标像素的数量有限，且常在图像重建时丢失，我们将小目标像素从原图复制到新生成的图像中。

3.4. 两阶段生成模型

粗重建阶段旨在学习从量化压缩图像到真实世界表示的映射。通过为训练模型提供不同量化水平的图像，可以生成具有不同细节水平的目标图像。如在 Fig. 5 中所示，其中量化间隔和复制值通过高斯分组挤压器为每个图像随机选择，以增强训练数据的多样性。量化图像经过使用 3×3 卷积的浅特征提取，然后通过残差 Swin Transformer 块 (RSTB) (Liu et al., 2021) 提取更深的特征。浅层和深层特征被融合并通过三个插值阶段上采样到原始分辨率。每个插值阶段后接一个卷积层和一个 Leaky ReLU 激活层。最后，通过使用以下损失函数最小化原图 I_{input} 和生成图像 I_{Cr} 之间的像素损失来优化粗重建过程：

$$I_{Cr} = Cr(I_{RQ}), \quad \mathcal{L}_2 = ||I_{Cr} * I_{input}||_2^2 \quad (5)$$

$Cr(\cdot)$ 表示应用于量化输入 I_{RQ} 的粗重建模型。像素级的 L2 损失 $\mathcal{L}_2$ 用于将生成的图像 I_{Cr} 与真实情况 I_{input} 对齐。这个损失促进了重建的红外图像与原始图像之间的结构一致性。尽管粗重建处理后的图像 I_{Cr} 在量化像素质量和分布对齐方面有所改善，但仍然无法达到真实世界的表现。我们将 I_{Cr} 输入扩散阶段，使其能够从大量原始红外小目标数据中学习。

在扩散阶段，我们采用潜在扩散模型 (LDM) (Rombach et al., 2022)，通过在潜在表示空间中逐

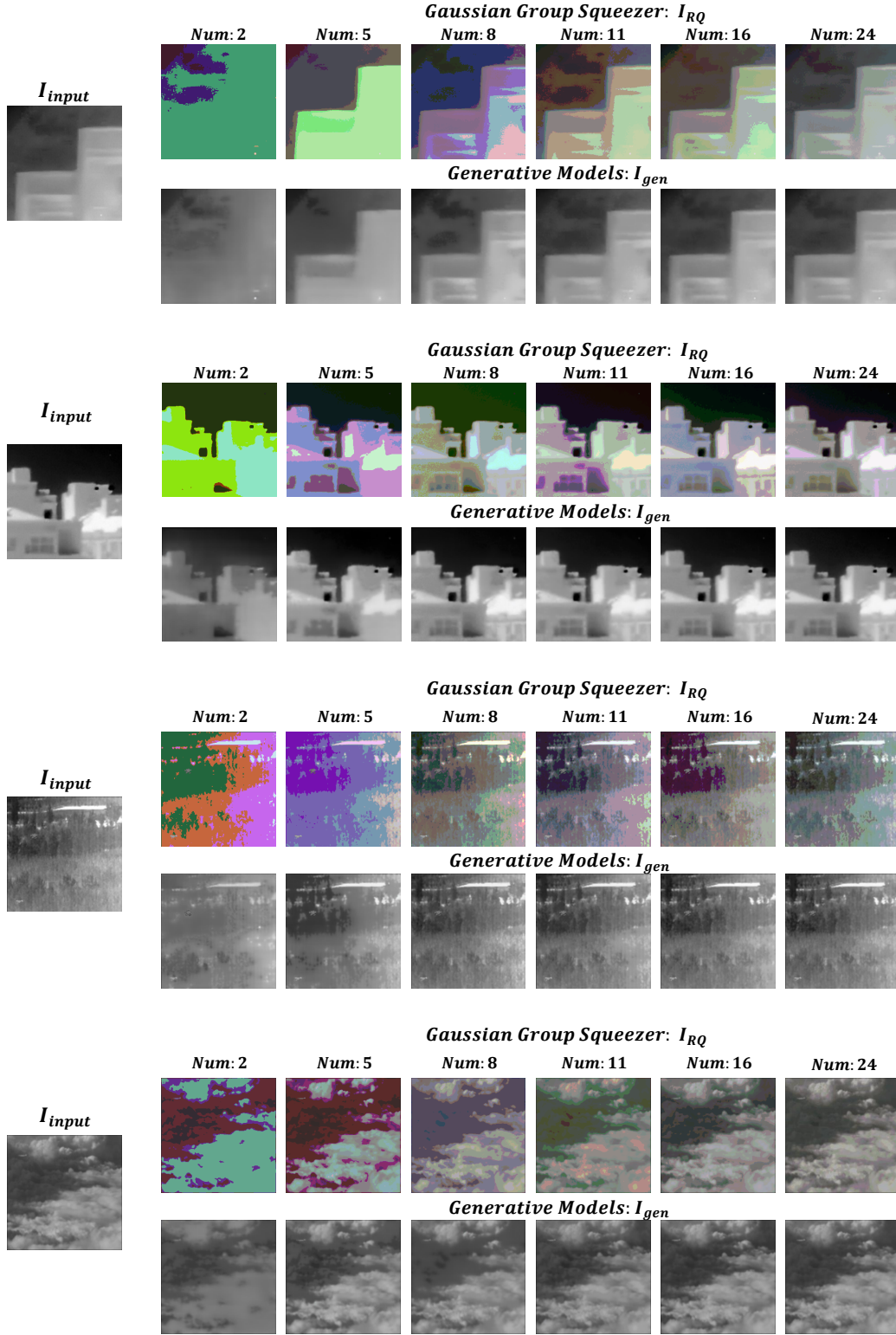


Fig. 7: 红外图像 I_{gen} 是使用基于扩散的生成模型生成的。在训练和推理阶段，输入图像通过不同的高斯参数化间隔进行量化。为此，我们生成了不具偏见但一致的红外数据。

步去噪数据来生成先验信息，并随后将其解码成完整图像。LDM 基于自动编码器架构，由编码器 $\mathcal{E}$ 和解码器 $\mathcal{D}$ 组成。输入图像 I 首先通过编码器 $\mathcal{E}$ 被压缩为潜在表示 z 。在每个时间步 t ，加入方差为 $\beta_t \tilde{\epsilon}(0, 1)$ 的高斯噪声以生成噪声潜在图像 z_t 。

当 t 足够大时， z_t 接近标准高斯分布。噪声预测网络 ϵ_θ 被训练为去噪模型，并使用损失函数 $\mathcal{L}_{ldm}$ 进行优化：

$$\mathcal{L}_{ldm} = \mathbb{E}_{z, t, c, \epsilon} \left(\left\| \epsilon^* \epsilon_\theta(z_t, c, t) \right\|_2^2 \right) \quad (6)$$

Table 3

各方法在 NUDT-SIRST (Li et al., 2022) 数据集上的全规模和少样本场景中, 对 $IoU(10^2)$ 、 $P_d(10^2)$ 、 $F_a(10^6)$ 、 $Precision(10^2)$ 和 $F1(10^2)$ 的比较。训练集由 100 个%、30 个% 和 10 个% 的图像组成。较大的 IoU 、 P_d 、 $Precision$ 和 $F1$ 值及较小的 F_a 值表示更好的性能。最佳和次佳的结果分别以红色和蓝色突出显示。显然, 我们的方法在表现上达到了最高水平, 特别是在少样本场景中展示了其有效性。

Method	Full-scale Data (100 %)					Few-shot (30 %)					Few-shot (10 %)				
	$IoU \uparrow$	$P_d \uparrow$	$F_a \downarrow$	$Prec \uparrow$	$F1 \uparrow$	$IoU \uparrow$	$P_d \uparrow$	$F_a \downarrow$	$Prec \uparrow$	$F1 \uparrow$	$IoU \uparrow$	$P_d \uparrow$	$F_a \downarrow$	$Prec \uparrow$	$F1 \uparrow$
ACM	67.76	96.4	13.65	76.83	80.78	64.49	95.76	15.60	74.26	78.42	53.93	91.00	64.94	62.71	70.08
ALCNet	66.48	95.97	16.36	78.67	79.86	62.56	94.6	11.53	75.19	76.95	22.61	95.55	514.13	22.88	36.91
UIUNet	88.39	98.62	4.87	94.26	93.84	83.7	97.67	9.65	89.56	91.13	76.68	96.61	25.94	85.82	88.70
MSHNet	77.68	95.13	10.43	90.53	87.44	75.83	95.02	21.44	87.47	86.26	64.08	91.32	68.64	80.13	78.10
DNANet	85.21	98.51	4.06	97.26	96.85	83.17	98.09	11.42	89.11	90.82	74.07	95.66	22.45	86.16	85.11
SCTrans	94.09	99.04	3.95	97.67	96.96	90.63	98.83	6.94	95.54	95.09	71.64	95.76	27.02	84.70	83.47
Ours	95.37	99.04	0.80	98.06	97.63	93.39	98.94	2.04	97.17	96.59	86.04	97.67	14.13	93.94	92.49

Table 4

使用不同的方法, 在 SIRST (Dai et al., 2021a) 数据集的全尺寸和少样本场景中比较 $IoU(10^2)$ 、 $P_d(10^2)$ 、 $F_a(10^6)$ 、 $Precision(10^2)$ 和 $F1(10^2)$ 。训练集包括 100 %、30 % 和 10 % 的图像。较大的 IoU 、 P_d 、 $Precision$ 和 $F1$ 值, 以及较小的 F_a 值, 表明性能更好。最好和第二好的结果以红色和蓝色突出显示。显然, 我们的方法达到了最高的性能, 证明了其有效性, 尤其是在少样本场景中。

Method	Full-scale Data (100 %)					Few-shot (30 %)					Few-shot (10 %)				
	$IoU \uparrow$	$P_d \uparrow$	$F_a \downarrow$	$Prec \uparrow$	$F1 \uparrow$	$IoU \uparrow$	$P_d \uparrow$	$F_a \downarrow$	$Prec \uparrow$	$F1 \uparrow$	$IoU \uparrow$	$P_d \uparrow$	$F_a \downarrow$	$Prec \uparrow$	$F1 \uparrow$
ACM	66.14	91.63	24.83	82.68	80.38	62.91	91.25	33.20	76.64	77.96	20.84	85.17	657.5	21.79	34.80
ALCNet	46.58	93.91	121.35	22.26	36.30	33.76	93.15	263.29	18.29	30.80	12.72	91.63	2192	6.93	12.95
UIUNet	76.08	92.77	13.44	92.53	87.24	73.61	92.01	15.50	93.25	85.62	62.52	91.63	54.81	83.20	77.73
MSHNet	73.02	95.43	19.89	79.69	79.08	69.66	94.67	46.13	75.44	76.26	59.43	93.91	116.36	67.55	73.27
DNANet	76.48	96.57	26.47	88.50	87.44	74.56	93.15	18.65	89.58	86.04	56.49	88.97	143.58	73.96	72.90
SCTrans	76.00	95.81	19.34	88.43	87.17	69.42	95.43	63.73	89.26	88.38	59.74	90.87	94.87	83.06	79.72
Ours	78.88	96.19	12.69	89.29	88.98	78.19	96.19	16.67	88.44	88.55	67.96	92.77	80.74	81.83	81.71

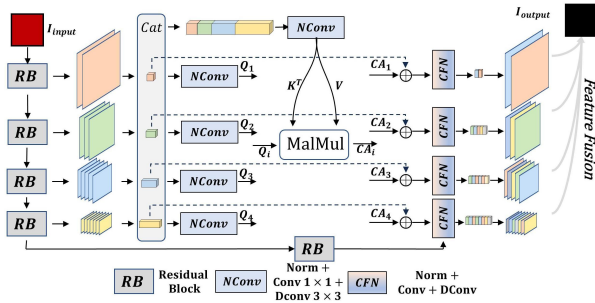


Fig. 8: 检测框架。我们将生成模块生成的图像 I_{gen} 作为增广样本输入到检测网络中进行训练。

其中 c 表示一个条件变量, ϵ 是从标准高斯噪声中采样的。

为了使扩散模型能够生成目标域的高质量图像, 我们使用一个红外小目标数据集对其进行了微调。如图 5 所示, 在训练阶段, 图像 I_{Cr} 通过编码器 $\mathcal{E}$ 被编码为潜在空间变量 $z_0 = \mathcal{E}(I_{Cr})$ 。然后逐渐加入高斯噪声以获得 z_t , 这近似为高斯分布。然后, 通过 t 重采样步骤生成一个新的潜在表示 z_0^* , 并由 $\mathcal{D}$ 解码以生成重采样图像 I_{gen} 。损失函数 $\mathcal{L}_{resample}$ 用于最小化生成图像与真实图像之间的差异。

$$\mathcal{L}_{resample} = ||z_0^* - z_0||_2^2 \quad (7)$$

通过最小化真实噪声和预测噪声之间的偏差, 该模型学习了红外图像的数据分布, 理解了背景语义和小目标特征, 并将真实世界和多样化的知识融入到图像中。

在推理阶段, 我们采用了交叉峰值采样策略。具体来说, 在训练和推理过程中, 不同的高斯组压缩参数集从同一个高斯分布中采样, 以增强视觉多样性。与传统的数据增强方法不同, 它在再生成阶段保留了结构特征。借助扩散先验, 我们的方法显著增强了生成数据的真实性, 为判别学习提供了更具代表性的样本。如图 7 所示, I_{gen} 随后被用于检测网络的训练。由于在训练和推理过程中高斯组压缩器的参数集不同, 生成的图像展示了更丰富的纹理和模式。

3.5. 检测

为了验证生成数据集的有效性, 我们将生成的数据 I_{gen} 与 I_{input} 结合用于训练。如图 8 所示, 利用 ResNet 模块提取特征, 然后通过空间注意力模块进行增强。最后, 将不同尺度下的特征融合生成稳健的特征图以预测 I_{output} 。地面实况 I_{label} 和预测输出 I_{output} 都是分割图, 并通过如下定义的损失函数 $\mathcal{L}_{IOU}$ 最小化它们的差异:

$$\mathcal{L}_{IOU} = 1 * \frac{I_{output} \cdot I_{label} + a}{I_{output} + I_{label} * I_{output} \cdot I_{label} + a} \quad (8)$$

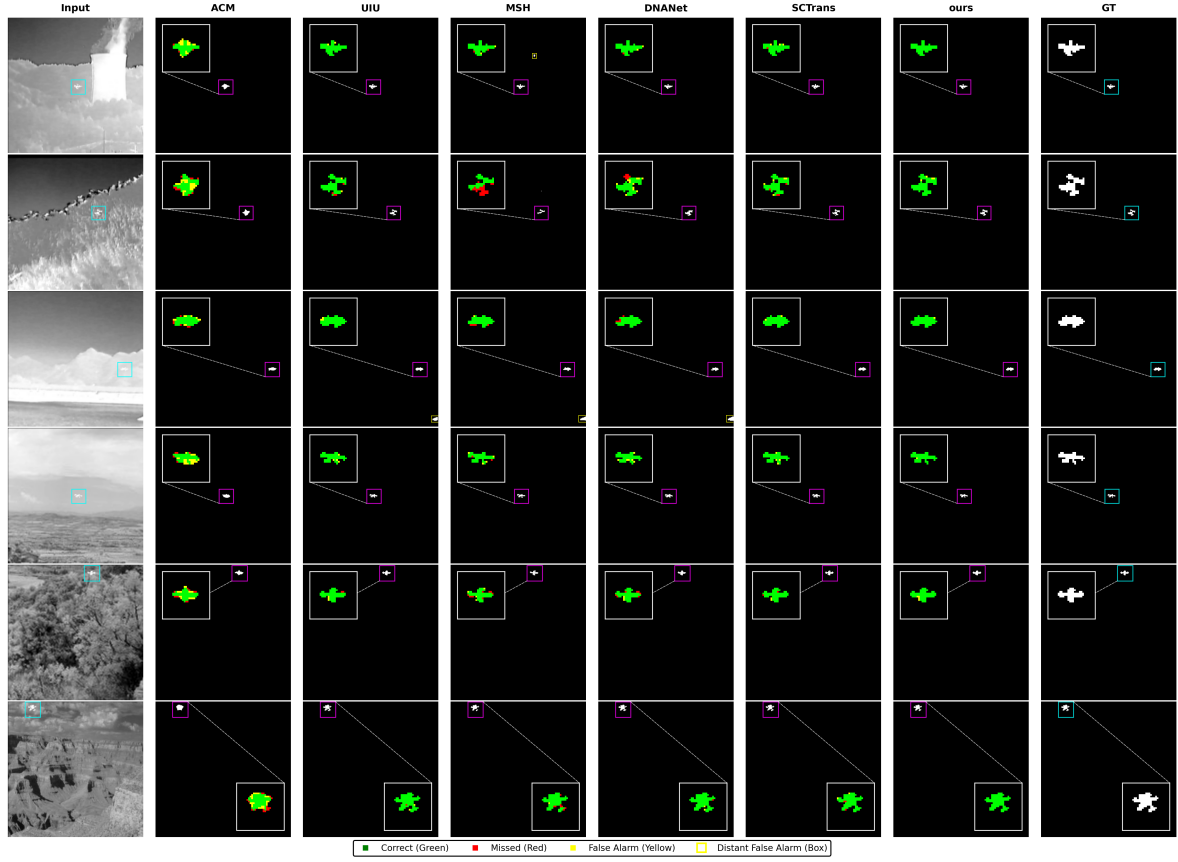


Fig. 9: NUDT-SIRST (Li et al., 2022) 和 SIRST (Dai et al., 2021a) 数据集上的可视化比较。我们的方法生成的输出与真实标签更加一致。在放大的区域中，我们使用颜色编码的高亮：绿色表示正确预测的目标像素，红色标记为漏检的目标（假阴性），黄色表示不是在真实标签中存在的假阳性。

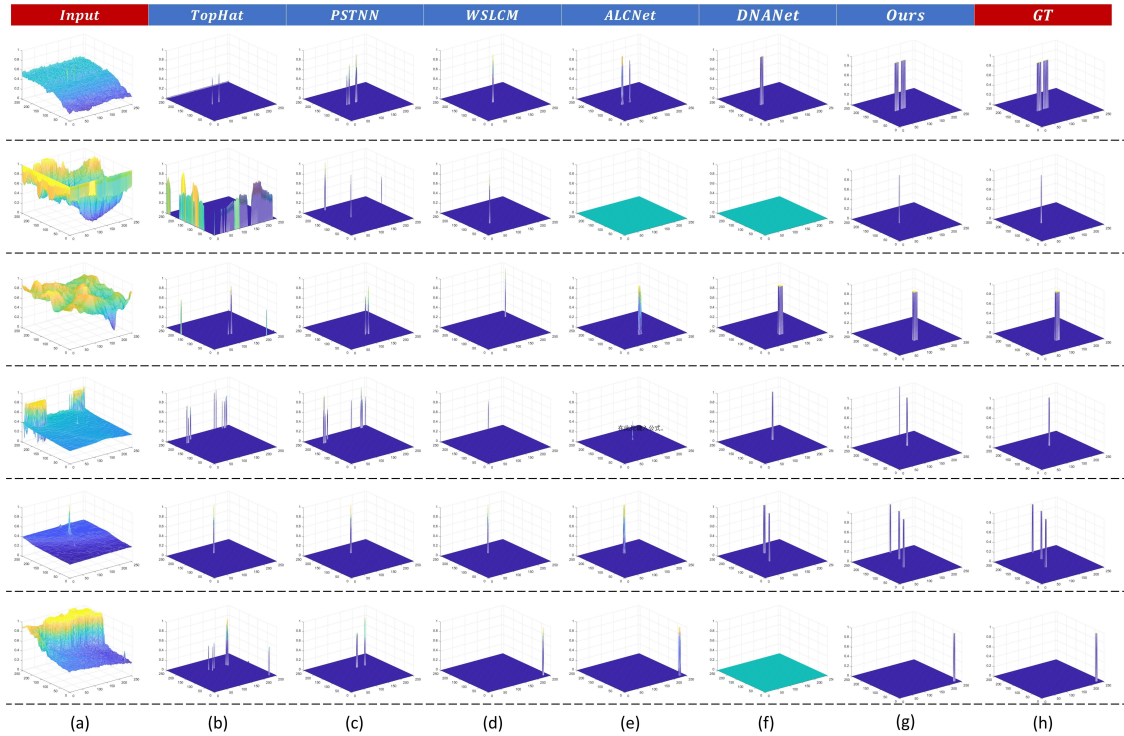


Fig. 10: 不同方法的 3D 可视化结果。

4. 实验

4.1. 评估指标

CNN 相关工作采用的传统像素级指标 (IoU、精度、召回率) (Dai et al., 2021b) (Dai et al., 2021c) (Wang et al., 2019) 对于红外小目标表现出两个关键限制。首先, 它们的极端边界敏感性导致了不成比例的惩罚: 一个 3×3 目标中仅有一个像素误分类就会使精度降低 11.1 %。其次, 它们与以目标检测信心为优先而非精确形状描绘的军事需求不符。为了解决这些问题, 我们的评估框架结合了战术和算法指标: 目标级别标准: P_d 用于检测率 (在预警系统中至关重要), F_a 用于量化每 1000 个像素的误报。增强的像素指标: IoU 用于形状保真度; F1-score 平衡召回率 (检测完整性) 和精度 (误报抵抗)。

为了证明我们方法的有效性, 我们选择了 SIRST (Dai et al., 2021a) 和 NUDT-SIRST (Li et al., 2022) 数据集进行实验。数据集被分割, 其中 50 % 用于训练, 50 % 用于测试。完整场景中利用了训练数据的 100 %, 而数据中的 30 % 和 10 % 被用来代表不同尺度的小样本场景, 以验证该方法在极端挑战下的有效性。

在训练阶段, 高斯组挤压器通过从均值为 17, 方差为 4 的高斯分布中采样, 生成量化参数。粗重建阶段的批量大小设为 3, 训练步数为 25000, 学习率为 0.001。我们借用了 Diffbir (Lin et al., 2023) 的训练方法, 将预训练于 ImageNet (Deng et al., 2009) 的扩散模块加载作为扩散阶段的模型。量化恢复的图像用于微调, 批量大小、周期数和学习率分别设置为 1、20 和 0.01。在推理阶段, 采样了一组不同的量化区间。来自 NUDT-SIRST (Li et al., 2022) 和 SIRST (Dai et al., 2021a) 数据集的样本被量化, 并且扩散阶段使用了时间步长为 50 的 DDPM 跳跃采样。我们生成了与原始训练数据量相同的数据用于基线训练。检测网络的批量大小设为 4, 学习率为 0.05, 周期数为 1000。

4.2. 比较

我们选择了几种具有代表性的方法进行比较, 包括 ACMNet (Dai et al., 2021b)、ALCNet (Dai et al., 2021c)、UIU-Net (Wu et al., 2022)、DNANet (Li et al., 2022)、MSHNet (Liu et al., 2024) 和 SCTransNet (Yuan et al., 2024)。我们按照各自论文中指定的配置, 在 NUDT-SIRST (Li et al., 2022) 和 SIRST (Dai et al., 2021a) 数据集上对所有方法进行了重新训练。

如 Table 3 和 Table 4 所示, 我们使用生成的数据集进行基线网络训练, 并将测试结果与当前的先进检测网络 (SOTA) 进行比较。与所有 SOTA 方法相比, 我们的模型在联合交集 (IoU)、检测概率 (P_d) 和虚警率 (F_a) 指标上, 对 NUDT-SIRST (Li et al., 2022) 和 SIRST (Dai et al., 2021a) 数据集均表现出色。生成的样本显著提高了性能, 证明了我们方法的可靠性。此外, 在少样本场景中, 即使在完整和有限样本数量下, 我们的方法也表现出极小的性能降级。值得注意的是, 在 30 % 的少样本条件下, SIRST (Dai et al., 2021a) 的表现优于 SOTA 模型在全尺度场景下的表现。

我们将最先进的方法与我们的方法在 SIRST (Dai et al., 2021a) 和 NUDT-SIRST (Li et al., 2022) 上进行了比较。评估使用了三个指标: IoU、 P_d 和 F_a 。如图 Fig. 9 和 Fig. 10 所示, 由于图像中目标像素数量有限,

Table 5

我们的模型与各种数据增强技术在 SIRST (Dai et al., 2021a) 数据集上的定量比较。

Method	Metrics		
	$IoU \uparrow$	$P_d \uparrow$	$F_a \downarrow$
Mixup (Zhang et al., 2017)	76.89	94.67	24.97
CutOut (DeVries and Taylor, 2017)	77.57	96.19	28.40
CutMix (Yun et al., 2019)	71.69	92.77	25.31
Mosaic (Bochkovskiy et al., 2020)	76.10	95.81	30.59
Ours	78.88	96.19	12.69

Table 6

在标准基准上的实时检测性能比较。我们的方法在保持模型大小和计算成本之间的良好平衡的同时, 实现了最佳的 IoU。

Method	IoU (%) $\uparrow$	Params (M) $\downarrow$	FLOPs (G) $\downarrow$
ACM	67.76	0.39	0.40
ALCNet	66.48	0.42	0.37
UIU-Net	88.39	50.54	54.42
DNANet	85.21	4.69	14.26
Ours	95.37	11.19	10.11

Table 7

在 NUDT-SIRST (Li et al., 2022) 上的消融结果表明, GGS 和扩散模块都对性能有显著贡献, 尤其是在减少误报方面。

Method	$IoU (10^2) \uparrow$	$P_d (10^2) \uparrow$	$F_a (10^6) \downarrow$
Baseline	94.09	99.04	3.95
w/o GGS	93.14	98.51	3.81
w/o Diff	93.64	98.51	4.18
Ours	95.37	99.04	0.80

提供了放大视图以进行详细观察。DNA-Net (Li et al., 2022) 存在泄漏和错误检测, 而 UIU-Net (Wu et al., 2022) 和 SCTansNet (Yuan et al., 2024) 虽然减少了错误预测, 但在一些困难样本中未能准确检测出目标轮廓。利用生成模型生成的样本进行训练的检测模型在处理困难样本的预测方面表现良好。这表明本文提出的方法生成的数据可以有效地帮助模型学习数据分布和特征, 并提供更具挑战性的样本。与其他 SOTA 方法相比, 我们的方法在 P_d 、 F_a 和 IoU 指标上表现更好, 显著提高了 ISTD 的准确性和鲁棒性。

我们将常用的数据增强方法与我们的方法进行比较, 以帮助提高检测模型的性能。Table 5 展示了各种数据增强方法对 SIRST (Dai et al., 2021a) 数据集上的模型性能的影响。我们的方法在 IoU 、 P_d 和 F_a 指标方面展示了卓越的性能。这些结果表明, 传统的数据增强技术可能会对小目标的检测产生负面影响, 使其更难以识别。相比之下, 我们提出的增强方法显著提高了检测模型的能力, 并降低了误报率。

在 Table 6 中, 我们的检测模型效率竞争力 (10.11 GFLOPs)。基于扩散的增强仅在训练期间严格离线运行, 因此推理阶段没有额外开销。这种计算解耦符合数据预处理优化原则 (Wang et al., 2021)。

4.3. 消融实验

我们使用基础检测网络在 SIRST (Dai et al., 2021a) 和 NUDT-SIRST (Li et al., 2022) 数据集上进行了多组实验, 以评估我们提出的方法中各个模块的有效性。

Table 8

高斯参数敏感性分析。测试不同均值 (μ) 和标准差 (σ) 的组合, 以评估它们对检测性能的影响。

μ	σ	IoU (%) $\uparrow$	P_d (%) $\uparrow$	F_a (10^{-3}) $\downarrow$
8	4	93.26	98.94	2.71
28	4	93.33	98.83	6.31
17	2	92.99	98.73	5.53
17	8	93.19	98.94	4.48
17	4	95.37	99.04	0.80

Table 9

在 RealScene-ISTD 和 IRSTD-1K 数据集上的性能比较。评估包括交并比 (IoU)、检测概率 (P_d) 和误报率 (F_a)。我们的方法在真实世界和大规模基准测试中取得了更加一致的结果。

Method	RealScene-ISTD			IRSTD-1K		
	IoU	P_d	F_a	IoU	P_d	F_a
ACM	36.86	88.03	948.4	61.79	87.98	21.82
UIUNet	37.06	86.32	680.3	64.11	91.18	29.24
DNANet	36.25	83.76	529.1	65.47	91.45	19.11
SCTrans	46.67	88.88	193.12	64.22	91.18	23.24
Ours	50.93	90.59	146.16	64.23	93.05	23.46

Table 10

比较使用 RSamp 和 GSamp 对 NUDT-SIRST (Li et al., 2022) 数据集上的检测性能指标的影响。

	IoU (10^2) $\uparrow$	P_d (10^2) $\uparrow$	F_a (10^6) $\downarrow$
baseline	94.09	99.04	3.95
w/ RSamp	93.50	98.51	1.88
w/ GSamp	95.37 ($\uparrow 1.87\sim$)	99.04 ($\uparrow 0.53\sim$)	0.80 ($\uparrow 1.08\sim$)

Table 11

比较 Cr 阶段和 Diff 阶段对 SIRST (Dai et al., 2021a) 数据集检测性能指标的影响。

	IoU (10^2) $\uparrow$	P_d (10^2) $\uparrow$	F_a (10^6) $\downarrow$
baseline	76.00	95.81	19.34
w/ Cr Stage	78.50	96.19	13.24
w/ Diff Stage	78.88 ($\uparrow 0.38\sim$)	96.19	12.69 ($\uparrow 0.55\sim$)

高斯组压缩器 (GGs) 和扩散模块的有效性。正如 Table 7 所示, 扩散模块和 GGS 都对性能有显著贡献。移除扩散阶段会导致 IoU 明显下降和误报显著增加, 表明其在维护特征保真度方面的重要性。同样地, 禁用 GGS 会降低检测精度, 并导致误报的频率增加, 突出了其在增强特征多样性和泛化方面的作用。Table 8 表明, 我们的高斯参数设置 ($\mu = 17$, $\sigma = 4$) 在准确性和稳定性之间实现了最佳平衡。其他配置始终表现较差, 证实了我们所选参数的稳健性。

高斯采样 (GSamp) 和随机采样 (RSamp) 的效果。我们比较了使用 GSamp 和随机采样策略选择量化参数时模型的性能。正如 Table 10 所示, 结果表明在 IoU、 P_d 和 F_a 指标上, 采用 GSamp 的模型显著优于使用随机采样的模型。GSamp 确保了量化参数的合理分布, 从而生成更具代表性的样本并增强了泛化能力。

粗重建阶段 (Cr 阶段) 与扩散阶段 (Diff 阶段) 的效果评估。为了评估 Diff 阶段的有效性, 我们将仅用 Cr 阶段增强的模型与结合 Cr 阶段和 Diff 阶段重采样的模型进行了比较。实验结果表明, 加入 Diff 阶段显著提高了检测性能, 尤其是在细节恢复和特征重建方面。

Table 12

在 SIRST (Dai et al., 2021a) 上, DNANet 在有和没有高斯未知数据时的定量结果。

Method	Metrics		
	IoU $\uparrow$	P_d $\uparrow$	F_a $\downarrow$
w/o Gauss	76.48	96.57	26.47
w/ Gauss	77.11 ($\uparrow 0.63\sim$)	94.67 ($\downarrow 1.90\sim$)	14.95 ($\uparrow 11.52\sim$)

Table 13

UIUNet (Wu et al., 2022) 在 NUDT-SIRST (10~) 上使用和不使用高斯无关数据的定量结果。可以看出, 'w/ Gauss' 的结果相比于 'w/o Gauss' 有显著提高。

Method	Metrics		
	IoU $\uparrow$	P_d $\uparrow$	F_a $\downarrow$
w/o Gauss	76.68	96.61	25.94
w/ Gauss	79.17 ($\uparrow 2.49\sim$)	96.61 ($\uparrow 0\sim$)	8.29 ($\uparrow 17.65\sim$)

通过重采样过程, Diff 阶段引入了额外的真实信息, 从而更有效地增强了小红外目标的检测。经过扩散阶段增强的数据更真实, 并引入了现实世界的知识。我们将增强后的结果输入到基线模型中进行实验。Table 11 中的定量结果显示, 扩散阶段在检测小目标时提高了性能。这表明扩散阶段产生的增强样本是多样且真实的, 有效地提升了模型的泛化能力。

对更多骨干网络的有效性。为了评估使用高斯无关表示学习的有效性, 我们利用这些样本来训练各种检测模型。在本节中, 我们应用 UIU-Net (Wu et al., 2022) 和 DNA-Net (Li et al., 2022) 作为检测模型, 并比较它们在有无高斯无关数据情况下的性能 ('w/o Gauss' vs. 'w/ Gauss')。如 Table 12 所示, 我们比较了 DNA-Net (Li et al., 2022) 模型在 SIRST (Dai et al., 2021a) 数据集上是否使用高斯无关数据的性能。为了突出性能的提升或下降, 我们使用上箭头 $\uparrow$ 来表示性能的提高 (IoU、 P_d 和 F_a) 和下箭头 $\downarrow$ 来表示性能的下降。结果表明, 使用高斯无关扩展进行训练显著提高了 DNA-Net (Li et al., 2022) 在 IoU 和 F_a 方面的性能。

如 Table 13 所示, 我们在 NUDT-SIRST (Li et al., 2022) (10 %) 数据集上评估了 UIU-Net (Wu et al., 2022) 模型的检测性能。实验结果表明, UIU-Net (Wu et al., 2022) 在使用生成的样本进行训练后, 检测性能显著提高, 从而充分验证了这些样本的有效性。

表 9 的结果进一步验证了我们方法的跨领域泛化能力。尽管 RealScene-ISTD (Lu et al., 2025) 和 IRSTD-1K (Zhang et al., 2023) 之间传感器类型、场景结构和目标特性存在差异, 我们的方法仍然保持高精度和低误报率, 表明其对真实世界红外环境的强适应性。

5. 结论

在本文中, 我们提出了一种新颖的高斯不敏感表示学习框架, 以解决红外小目标检测的极端挑战。所提出的高斯组挤压器通过非均匀量化生成多样化的训练数据, 而粗重建和扩散阶段通过重建目标结构和结合扩散模型的先验知识来增强数据质量。实验结果表明, 与最先进的方法相比, 我们的方法在极端条件下显著提高了

检测精度。我们的方法的有效性突出了其在克服红外目标检测中的数据规模限制方面的潜力，为未来的研究提供了一个有前途的方向。

References

- Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M., 2020. Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934 .
- Chen, C.P., Li, H., Wei, Y., Xia, T., Tang, Y.Y., 2013. A local contrast method for small infrared target detection. IEEE Transactions on Geoscience and Remote Sensing 52, 574–581.
- Dai, Y., Wu, Y., Zhou, F., Barnard, K., 2021a. Asymmetric contextual modulation for infrared small target detection, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 950–959.
- Dai, Y., Wu, Y., Zhou, F., Barnard, K., 2021b. Asymmetric contextual modulation for infrared small target detection, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 950–959.
- Dai, Y., Wu, Y., Zhou, F., Barnard, K., 2021c. Attentional local contrast networks for infrared small target detection. IEEE Transactions on Geoscience and Remote Sensing 59, 9813–9824.
- Deng, H., Sun, X., Liu, M., Ye, C., Zhou, X., 2016. Small infrared target detection based on weighted local difference measure. IEEE Transactions on Geoscience and Remote Sensing 54, 4204–4214.
- Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE. pp. 248–255.
- DeVries, T., Taylor, G.W., 2017. Improved regularization of convolutional neural networks with cutout. arXiv preprint arXiv:1708.04552 .
- Du, N., Gong, X., Liu, Y., 2024. Istd-diff: Infrared small target detection via conditional diffusion models. IEEE Geoscience and Remote Sensing Letters 21. doi:10.1109/LGRS.2024.3401838.
- Gao, C., Meng, D., Yang, Y., Wang, Y., Zhou, X., Hauptmann, A.G., 2013. Infrared patch-image model for small target detection in a single image. IEEE Transactions on Image Processing 22, 4996–5009.
- Goodall, T.R., Bovik, A.C., Paulter, N.G., 2016. Tasking on natural statistics of infrared images. IEEE Transactions on Image Processing 25, 65–79. doi:10.1109/TIP.2015.2496289.
- Hong, D., Gao, L., Yao, J., Zhang, B., Plaza, A., Chanussot, J., 2021. Graph convolutional networks for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing 59, 5966–5978. doi:10.1109/TGRS.2020.3015157.
- Hu, M., Jiang, K., Wang, Z., Bai, X., Hu, R., 2023. Cycmunet+: Cycle-projected mutual learning for spatial-temporal video super-resolution. IEEE Transactions on Pattern Analysis and Machine Intelligence .
- Kumar, N., Singh, P., 2025. Small and dim target detection in infrared imagery: A review, current techniques and future directions. Neurocomputing , 129640.
- Li, B., Xiao, C., Wang, L., Wang, Y., Lin, Z., Li, M., An, W., Guo, Y., 2022. Dense nested attention network for infrared small target detection. IEEE Transactions on Image Processing 32, 1745–1758.
- Lin, X., He, J., Chen, Z., Lyu, Z., Fei, B., Dai, B., Ouyang, W., Qiao, Y., Dong, C., 2023. Diffbir: Towards blind image restoration with generative diffusion prior. arXiv preprint arXiv:2308.15070 .
- Liu, Q., Liu, R., Zheng, B., Wang, H., Fu, Y., 2024. Infrared small target detection with scale and location sensitivity, in: Proceedings of the IEEE/CVF Computer Vision and Pattern Recognition, pp. 17490–17499.
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021. Swin transformer: Hierarchical vision transformer using shifted windows, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10012–10022.
- Lu, Y., Li, Y., Guo, X., Yuan, S., Shi, Y., Lin, L., 2025. Rethinking generalizable infrared small target detection: A real-scene benchmark and cross-view representation learning. arXiv preprint arXiv:2504.16487 URL: <https://arxiv.org/abs/2504.16487>.
- Lu, Y., Lin, Y., Wu, H., Xian, X., Shi, Y., Lin, L., 2024. Sirst-5k: Exploring massive negatives synthesis with self-supervised learning for robust infrared small target detection. IEEE Transactions on Geoscience and Remote Sensing .
- van der Maaten, L., Hinton, G., 2008. Visualizing data using t-sne. Journal of Machine Learning Research 9, 2579–2605.
- Mao, Q., Li, Q., Wang, B., Zhang, Y., Dai, T., Chen, C.L.P., 2024. Spirdet: Toward efficient, accurate, and lightweight infrared small-target detector. IEEE Transactions on Geoscience and Remote Sensing 62, 5006912.
- Peng, S., Ji, L., Chen, S., Duan, W., Zhu, S., 2025. Moving infrared dim and small target detection by mixed spatio-temporal encoding. Engineering Applications of Artificial Intelligence 144, 110100.
- Rawat, S.S., Verma, S.K., Kumar, Y., 2020. Review on recent development in infrared small target detection algorithms. Procedia Computer Science 167, 2496–2505.
- Rivest, J.F., Fortin, R., 1996. Detection of dim targets in digital infrared imagery by morphological image processing. Optical Engineering 35, 1886–1893.
- Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B., 2022. High-resolution image synthesis with latent diffusion models, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10684–10695.
- Shen, Y., Li, Q., Xu, C., Chang, C., Yin, Q., 2025. Graph-based context learning network for infrared small target detection. Neurocomputing 616, 128949.
- Shi, Y., Lin, Y., Wei, P., Xian, X., Chen, T., Lin, L., 2024. Diff-mosaic: augmenting realistic representations in infrared small target detection via diffusion prior. IEEE Transactions on Geoscience and Remote Sensing .
- Teutsch, M., Krüger, W., 2010. Classification of small boats in infrared images for maritime surveillance, in: 2010 International WaterSide Security Conference, IEEE. pp. 1–7.
- Van Den Oord, A., Vinyals, O., et al., 2017. Neural discrete representation learning. Advances in neural information processing systems 30.
- Wang, A., Li, W., Wu, X., Huang, Z., Tao, R., 2022. Mpanet: Multi-patch attention for infrared small target object detection, in: IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium, IEEE. pp. 3095–3098.
- Wang, H., Zhou, L., Wang, L., 2019. Miss detection vs. false alarm: Adversarial learning for small object segmentation in infrared images, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 8509–8518.
- Wang, S., Celebi, M.E., Zhang, Y.D., Yu, X., Lu, S., Yao, X., Zhou, Q., Miguel, M.G., Tian, Y., Gorriz, J.M., et al., 2021. Advances in data preprocessing for biomedical data fusion: An overview of the methods, challenges, and prospects. Information Fusion 76, 376–421.
- Wu, X., Hong, D., Chanussot, J., 2022. Uiu-net: U-net in u-net for infrared small object detection. IEEE Transactions on Image Processing 32, 364–376.
- Xi, Y., Zhou, Z., Jiang, Y., Zhang, L., Li, Y., Wang, Z., Tan, F., Hou, Q., 2023. Infrared moving small target detection based on spatial-temporal local contrast under slow-moving cloud background. Infrared Physics & Technology 134, 104877.

- Xiao, Y., Yuan, Q., Jiang, K., He, J., Jin, X., Zhang, L., 2023. Ediffr: An efficient diffusion probabilistic model for remote sensing image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*.
- Xiao, Y., Yuan, Q., Jiang, K., Jin, X., He, J., Zhang, L., Lin, C., . Local-global temporal difference learning for satellite video super-resolution. *arXiv preprint arXiv:2304.04421*.
- Yang, H., Wang, J., Bo, Y., Wang, J., 2025. Istd-detr: A deep learning algorithm based on detr and super-resolution for infrared small target detection. *Neurocomputing* 621, 129289.
- Ying, X., Wang, Y., Wang, L., Sheng, W., Liu, L., Lin, Z., Zho, S., 2022. Mocopnet: Exploring local motion and contrast priors for infrared small target super-resolution. *arXiv preprint arXiv:2201.01014*.
- Yuan, S., Qin, H., Yan, X., Akhtar, N., Mian, A., 2024. Sc-transnet: Spatial-channel cross transformer network for infrared small target detection. *IEEE Transactions on Geoscience and Remote Sensing* 62, 1–15. doi:10.1109/TGRS.2024.3383649.
- Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J., Yoo, Y., 2019. Cutmix: Regularization strategy to train strong classifiers with localizable features, in: *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 6023–6032.
- Zhang, F., Hu, H., Zou, B., Luo, M., 2025. M4net: Multi-level multi-patch multi-receptive multi-dimensional attention network for infrared small target detection. *Neural Networks* 183, 107026.
- Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D., 2017. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*.
- Zhang, M., Zhang, R., Yang, Y., Bai, H., Zhang, J., Guo, J., 2022. Isnet: Shape matters for infrared small target detection, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE/CVF*. pp. 877–886.
- Zhang, M., Zhang, R., Yang, Y., Bai, H., Zhang, J., Guo, J., 2023. Isnet: Shape matters for infrared small target detection. *Information Fusion* 97, 1–13. doi:10.1016/j.inffus.2023.05.005.
- Zhang, S., Wang, Z., Xing, Y., Lin, L., Su, X., Zhang, Y., 2024. Scafnet: Semantic-guided cascade adaptive fusion network for infrared small targets detection. *IEEE Transactions on Geoscience and Remote Sensing* 62, 1.
- Zhong, S., Zhou, H., Zheng, Z., Ma, Z., Zhang, F., et al., 2024. Hierarchical attention-guided multiscale aggregation network for infrared small target detection. *Neural Networks* 171, 485–496.