

LAFI-Diffusion: 基于可控扩散模型的卫星遥感舰船目标检测数据集构建与生成方法

何达¹, 李泽裕¹, 刘浩然¹, 胡锡坤², 钟平², 石茜¹

1. 中山大学 地理科学与规划学院, 广州 510006;
2. 国防科技大学 自动目标识别全国重点实验室, 长沙 410073

摘要: 卫星遥感图像舰船目标检测与识别在海洋监视、航运管理和军事侦察等领域具有重要应用价值。然而, 现有数据集在规模、多样性和细粒度方面的不足严重制约了该领域的发展。为突破这一瓶颈, 本文构建了大规模细粒度舰船实例检测数据集 LAFI (LArge-scale FIne-grained Ship Instance Detection), 并基于稳定扩散模型进一步构建了一个百万级舰船目标检测数据集 LAFI-Diffusion。该研究具有以下创新性贡献: (1) LAFI 是目前规模最大、类别最丰富的公开舰船数据集, 包含 8000 张高分辨率遥感图像、49 种舰船类型以及 48717 个精细标注实例; (2) 采用一个可控的扩散模型, 由船舶-海洋文本提示引导, 在不同的海况、天气条件和时间变化下生成数百万幅合成图像, 可以支撑目标检测模型的预训练, 显著提升了模型在复杂海洋场景下的泛化能力, 同时减少环境干扰。最后, 基于 LAFI 数据集, 本文系统评估了当前主流的 7 种有向框检测算法, 为后续研究提供了重要基准。

关键词: 遥感图像, 舰船目标检测, 数据集, 扩散模型, 合成数据生成, 细粒度识别, 有向边界框, 数据增强
中图分类号: P2

引用格式: 何达, 李泽裕, 刘浩然, 胡锡坤, 钟平, 石茜. XXXX. LAFI-Diffusion: 基于可控扩散模型的卫星遥感舰船目标检测数据集构建与生成方法. 遥感学报, XX(XX): 1-16

HE Da, LI Zeyu, LIU Haoran, HU Xikun, ZHONG Ping, SHI Qian. XXXX. LAFI-diffusion: Construction and Generation Method of Satellite Remote Sensing Ship Target Detection Dataset Based on Controllable Diffusion Model. National Remote Sensing Bulletin, DOI:10.11834/jrs.20265303]

1 引言

随着经济全球化的深入发展, 船运业已成为推动国际贸易发展和促进全球经济一体化的重要支柱。舰船作为海上运输的核心载体和关键军事装备, 基于卫星遥感图像对其进行目标检测与识别意义重大 (Li K 等, 2020; Nie G 和 Huang H, 2023; Yang X 等, 2023; Lu Y 等, 2025; Zheng Z 等, 2023; 禹文奇等, 2023), 在民用和军事领域均具有广阔的应用前景和巨大的发展潜力。在民用领域, 自动检测与识别技术可以用于监控特定港口或海域的海上交通, 辅助遇难船只的救援行动, 并协助安全管理部门打击非法捕鱼、非法倾倒油污、走私等违法行为以及海盗活动; 在军事

领域, 该技术能够用于监视重点港口和海域的舰船部署与动态, 分析海上作战实力, 支撑作战情报生成 (姚艳清等, 2021; 周培诚等, 2021)。

随着遥感技术的快速发展, 基于卫星遥感数据的舰船目标智能化检测与识别技术在海洋监测、航运安全和环境保护等领域展现出广阔的应用前景 (黄洁等, 2017)。舰船目标检测的核心任务是判断图像中是否存在舰船目标, 并对舰船进行准确的检测、识别和定位。 (谢星星等, 2022) 提出了一种动态特征融合网络, 根据输入目标尺度动态学习融合权重。基于合成孔径雷达 (SAR) 影像的方法因其在全天候条件下卓越的监测性能而受到学界广泛关注 (Xu X 等, 2025)。然而, SAR 成像技术的局限性导致 SAR 图像缺乏结构细节和视

收稿日期: 2025-08-23; 预印本: XXXX-XX-XX

基金项目: 重点实验室稳定支持科研项目 (WDZC20245250206), 湖南省科技创新计划资助 (2024RC3119)

第一作者简介: 何达, 研究方向为遥感大数据智能解译。E-mail: heda@mail.sysu.edu.cn

通信作者简介: 胡锡坤, 研究方向为智能目标识别遥感野火监测。E-mail: xikun@nudt.edu.cn

觉特征的表达,从而给高精度模型的检测和识别带来困难。随着卫星传感器成像技术的发展,光学卫星遥感图像的分辨率显著提高,其直观易懂、语义信息丰富的特征为舰船目标的准确识别提供了有力支撑;也为细粒度船舶检测(fine-grained ship detection)提供了坚实的数据基础(王彦情等,2011)。

现有的研究大多关注船舶检测任务,多用于区分船舶目标和非船舶目标。一些粗糙的舰船识别方法只将舰船粗略地划分为2~5类,如军用舰船和民用舰船,而对其细粒度类别的研究较少。在细粒度船舶的识别中,需要将舰船目标进一步划

分为更加具体的船型类别,如拖船、邮轮和货船等。目前已公开的船舶数据集中只有类似于HRSC2016、OFSD、MCSD等专注于细粒度船舶检测的数据集。从表1中可以发现,主流的细粒度舰船检测数据集的图像数量多在3000张以下,实例数量不超过10000个,类别数量大多数在20类附近。考虑到现实中舰船的丰富性,目前的数据集在图像数量,实例规模和类别数量方面仍显不足。因此,迫切需要构建实例数量更多、图像数量更多且舰船细粒度更丰富的船舶识别数据集,以提高细粒度船舶检测的精度和实用性。

表1 现有数据集的对比

Table 1 Comparison of existing datasets

数据集	分辨率/米	图像数	实例数	类别数	边界框	年份
DOTA v1.0(Xia G S等,2018)	<0.5	2806	52516	1	HBB, OBB	2018
HRSC2016(Liu Z等,2017)	0.4-2	1061	2886	25	HBB, OBB	2016
FGSD(Chen K等,2020)	0.12-1.93	2612	5634	43	OBB	2020
HRSD(Liu L等,2021)	0.4-4	2499	9269	1	OBB	2022
OFSD(Qur'ang L等,2022)	0.5-6	9541	41937	13	OBB	2022
MCSD(Guo B等,2023)	0.25-2	4000	17850	16	OBB	2023
ShipRSImageNet(Zhang Z等,2021)	0.12-6	3435	17573	49	OBB	2021
LAFI-Diffusion(Ours)	0.12-6	8000	48717	49	OBB	2025

当前,主流研究使用的舰船数据集普遍存在的样本规模局限、场景多样性匮乏以及细粒度不足的问题,这种数据瓶颈严重制约了深度学习模型的性能上限(张寅等,2022)。尤其在军事应用场景中,传统依赖有限样本训练的检测模型在复杂多变的海洋环境下识别性能易下降(徐志京和柏雪,2022),而大参数量深度学习模型的训练需求又使得构建超大规模细粒度舰船目标数据集愈发迫切。对航母、驱逐舰等关键舰型的细粒度识别需求,更凸显了构建百万级细粒度船舶检测数据集的战略价值。

为了适应复杂多样化场景下的细粒度船舶识别需求,同时提高模型在真实场景中的适应性,本研究创新性地提出“真实数据+生成式扩展”的数据集构建方法。首先,我们构建了一个新的大规模细粒度船舶实例检测(LAFI, Large-scale Fine-grained Ship instance detection)基准数据集,该数据集采集自全球36个主要港口,包含8000张高分辨率图像和48717个标注实例,涵盖49类不

同尺寸、分辨率和成像质量的舰船,其规模和多样性显著超越现有公开数据集,适用于遥感图像中的细粒度船舶识别任务。

为了进一步扩展数据集规模并增强环境多样性,扩散生成模型(Diffusion Models)作为一种新兴的深度生成技术为数据增强提供了新的解决方案(Rombach R等,2022)。该模型通过模拟数据的扩散过程,能够学习复杂的数据分布并生成高质量的合成样本(Ye J等,2025)。特别是在文本条件引导下,扩散模型能够根据语义描述生成符合特定要求的图像,为基于LAFI数据集的进一步数据扩充提供了强有力的技术支撑。

基于扩散生成模型的优势,本研究使用船舶知识引导的可控数据生成方法(Controllable Ship-based Knowledge-guided Data Generation, CSDG),通过大语言模型将不同层次的环境因素转换为船舶知识向量,结合稳定扩散生成模型实现具有船舶位置约束的图像扩散生成。该方法以LAFI数据集为基础,不仅能够生成包含多样化海洋场景、

天气条件和时间变化的高质量船舶数据, 还能通过精准的文本提示控制生成内容, 将数据集规模扩展至百万级别 (LAFI-Diffusion), 有效扩展了舰船数据集的规模和多样性, 为细粒度船舶识别技术的发展提供了重要的数据支撑。

本文的主要贡献如下:

1) 提出了一个结合真实数据与可控扩散模型的数据集构建方法, 并据此发布了一个大规模、细粒度的舰船检测数据集及其百万级扩展集, 有效解决了船舶目标检测领域高质量数据稀缺的问题;

2) 通过充分的实验验证了合成数据对于提升目标检测模型在复杂海洋场景下泛化能力的显著效果, 并提供了充分的实验分析, 同时也对不同

检测方法的性能进行了分析与讨论。

2 LAFI数据集

2.1 数据收集

本研究共收集了8000张光学遥感图像, 图像尺寸大多数为930×930像素。首先, 我们从ShipRSImageNet数据集的3435幅光学遥感影像中, 优先选取2748张公开标注的训练集与验证集数据, 并对其中错误标注的实例进行了人工校正, 显著提升了标注质量。另外, 从Google Earth平台获取了全球36个主要港口的0.5米分辨率影像, 经裁剪生成5252张标准化的930×930像素图像, 通过人工精细标注, 获得了31144个舰船实例, 如图1所示。

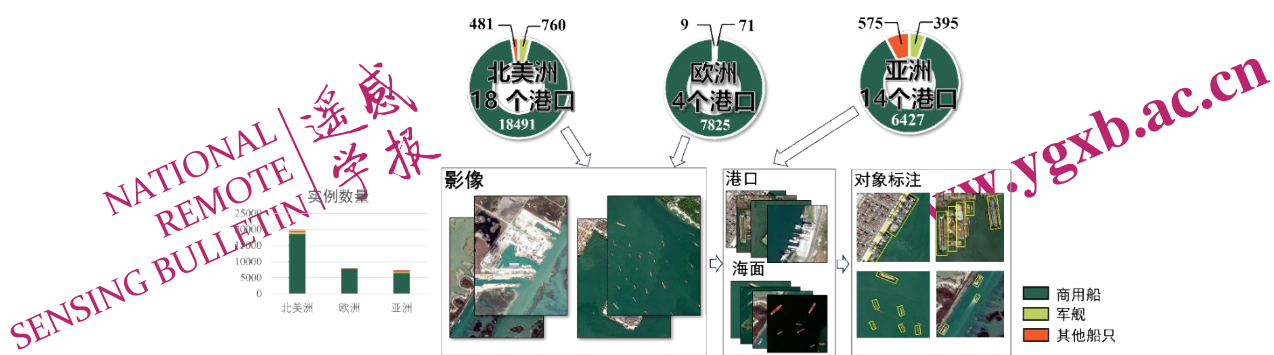


图1 数据收集、地理分布和注释过程概述。Google Earth 图片来自北美、欧洲和亚洲, 分别覆盖 18、4 和 14 个港口, 将这些图像分类为港口和海洋场景以进行详细注释。

Fig. 1 Overview of data collection, geographical distribution, and annotation process. The Google Earth images are sourced from North America, Europe, and Asia, covering 18, 4, and 14 ports, respectively. These images are then categorized into harbor and sea scenes for detailed annotation.

相较于现有数据集 (如 HRSC2016 仅覆盖 6 个港口, FGSD 局限于 17 个港口), LAFI 数据集扩展至全球 36 个港口, 除中国、日本、美国、西班牙外, 还新增英国、韩国、印度等地区。这种全球化的地理分布大幅增强了数据集的场景多样性和环境适应性, 使得基于 LAFI 训练的模型具备更强的泛化能力, 能够适用于不同海域、不同气候条件下的船舶检测任务。

2.2 数据分类

构建科学合理的船舶分类体系是实现细粒度目标识别的基础, 其设计必须兼顾实用性、军事价值和技术可行性。例如, 将舰船检测为“阿利·伯克驱逐舰”比将其检测为“驱逐舰”能提供更多有价值的信息。在实际应用中, 过于粗糙

的分类无法满足精细化识别需求, 而过于细化的分类则可能导致类别间特征差异不明显, 影响识别准确性。

基于上述考虑, 本研究参考 ShipRSImageNet 数据集的分类框架, 并结合实际应用需求进行了优化调整。在第一级分类中, 根据舰船的主要用途和应用场景, 将目标划分为军舰 (Warship)、商船 (Merchant) 和其他船舶 (Other Ship) 三个主要类别。在第二级分类中, 商船划分为 13 个子类, 主要基于货物类型和运输功能进行区分; 军舰划分为 35 个子类, 按照具体舰型和作战功能进行精细分类, 以满足军事探测的精确需求。

2.3 数据标注

在领域专家的指导和帮助下, 我们对 Google

Earth 图像进行了仔细的标记工作。标注时，专家知识体系 (Tupper E C, 2013) 能有效辅助特定舰船的识别，例如对于“尼米兹级” (Nimitz) 与“中途岛级” (Midway) 航空母舰，可根据其飞行甲板的形状进行区分 (前者采用梯形飞行甲板设计，而后者为不规则多边形甲板)。对于难以确定的特殊船型，由专业标注团队利用船舶所在港口和附近可识别的船舶作为辅助信息，进行进一步的判定。在注释时，我们使用开源图像标注工具 LabelMe (Russell B C 等, 2008)，在 Google Earth 图像上手动勾画出每艘船的边界框。对于难以确定的特殊船型，将船舶所在港口和附近可识别的船舶作为可靠确定的辅助信息。最后，由另一个工作小组对标注进行复核，纠正质量较差的标注、剔除错误标注，共获得 48717 个目标实例。最终构建的 LAFI 数据集可视化结果如图 2 所示。

标注时采用定向边界框 (OBB, oriented bounding boxes) 标注方式，在 LabelMe 工具中，对于每一个舰船目标，首先选择目标的左顶点作为起始点 (x_1, y_1) ，而后按顺时针方向依次标注其余三个顶点 $(x_2, y_2, x_3, y_3, x_4, y_4)$ 构成每艘船的边界框，并与该舰船的类型字段形成一组数据，记录于本幅图像对应标注文件的多边形字段中。

本研究采用了与 ShipRSImageNet 数据集相同的类别缩写：DD (Destroyer) 代表驱逐舰，FF (Frigate) 代表护卫舰，LL (Landing) 代表登陆舰，AS (Auxiliary Ship) 代表辅助舰，LSD (Landing Ship Dock) 代表登陆舰码头，LHA (Landing Helicopter Assault Ship) 代表登陆直升机突击舰，AOE (Fast Combat Support Ship) 代表快速战斗支援舰，EPF (Expeditionary Fast Transport Ship) 代表远征快速运输舰，RoRo (Roll-on Roll-off Ship) 代表滚装船。

2.4 数据集特点

LAFI 数据集具有以下三个特点：(1) 大规模：LAFI 数据集目前规模最大的细粒度船舶目标检测数据集，包含 8000 张光学遥感图像、48717 个船舶实例和 49 种细粒度船型类别。原始图像的尺寸介于 512×512 至 930×930 像素之间，并以 $0.64:0.2:0.16$ 的比例随机划分为训练集 (5120 张)、验证集 (1600 张) 和测试集 (1280 张)。(2) 细粒度

多样性：LAFI 数据集涵盖 49 个细粒度类别，包括商船和军舰两大类型。如图 3 (a) 所示，各船型实例数量分布均衡，平均每类包含 994 个实例。其中，“摩托艇” (Motorboat) 实例最多 (26232 个)，而“中途岛级航空母舰” (Midway) 最少 (仅 18 个)。此外，图 3 (b) 展示了每个类别的图像覆盖情况，平均每类出现在 282 张图像中，且大多数类别覆盖超过 100 张图像，有效保证了数据集的类别平衡性。这种细粒度分类设计不仅满足了军事应用中对精确目标识别的需求，也为民用船舶监测提供了详细的类型区分能力。(3) 多源数据融合：LAFI 数据集在数据来源多样化方面实现了重要突破，融合了 WorldView-3、GanFen-2、QuickBird 等八种不同卫星平台的影像数据。这些卫星平台的空间分辨率覆盖范围从 0.12 米到 6 米，形成了多尺度、多光谱的图像采集体系，显著提升了数据集的图像质量多样性和技术适应性。

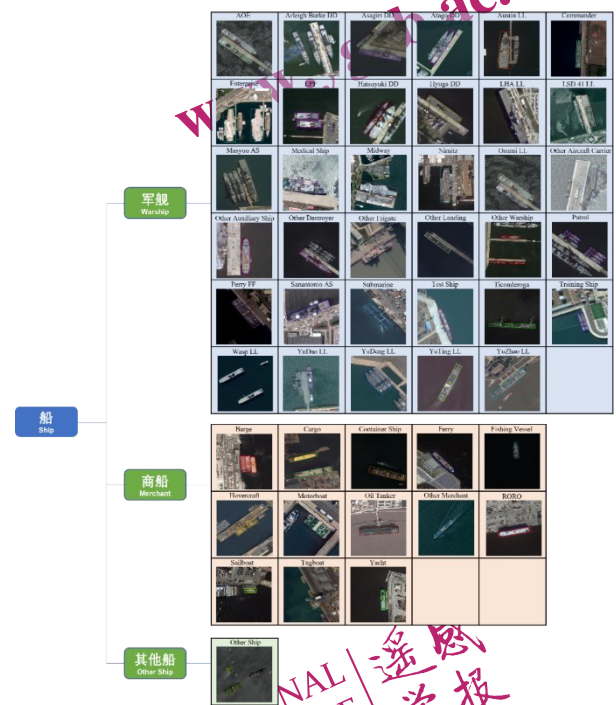


图 2 在 LAFI 中标注的图像样本，显示了军舰、商船和其他船舶三个一级类别中所包含的特殊类别。

Fig. 2 Samples of annotated images in LAFI, showing the special categories which included in three first-level category of warship, merchant and other ship.

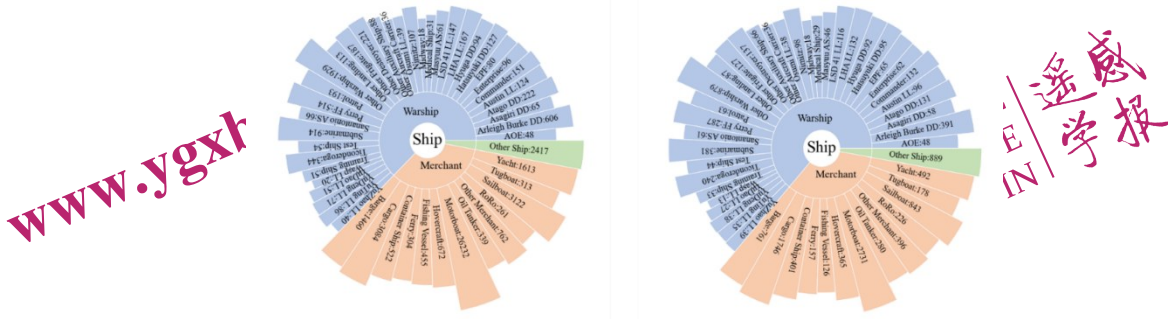


图3 LAFI数据集分布特征

Fig. 3 LAFI dataset distribution characteristics.

(a) 包含每个类别的图像数 (b) 每个类别的实例数

(a) the number of images containing each category. (b) the number of ship instances per category.

3 LAFI-Diffusion: 百万数据集

LAFI-Diffusion 是基于 LAFI 构建的百万级细粒度舰船检测数据集, 通过融合真实采集与智能生成技术, 在数据规模和质量上实现了重大突破。该数据集不仅包含 LAFI 中经过严格标注的 8000 张光学遥感影像, 并完整保留了其原有的 49 类舰船类别体系, 更创新性地引入船舶知识引导的可控数据生成方法, 充分考虑了真实海洋环境的复杂性和多样性, 利用有限的真实船舶数据自动合成大规模、高质量的船舶图像, 将总体规模提升至百万级别。通过构建分层环境描述体系和文本条件引导机制, LAFI-Diffusion 涵盖了不同海洋场景、天气条件和时间状态的合成船舶图像, 在保持类别一致性的基础上, 解决了现有数据集在图像数量、实例规模和类别多样性方面的不足, 为细粒度船舶检测模型的训练提供了充足且多样化的数据支撑。

3.1 扩散模型理论基础

3.1.1 稳定扩散模型架构

稳定扩散模型是一种基于潜在空间的生成模型, 其核心机制包含正向扩散过程和反向扩散过程两个阶段。正向扩散过程通过马尔可夫链将原始船舶图像映射至潜在空间, 并采用逐步注入高斯噪声的策略实现数据变换。反向扩散过程则利用船舶场景提示中的文本语义信息, 在每个去噪步骤中对神经网络的噪声预测进行约束, 从而实现合成数据样本分布的有效调节, 并将海洋环

境信息融入生成过程 (如图 4 所示)。

3.1.2 正向扩散过程

正向扩散过程定义为一个马尔可夫链, 通过逐步添加高斯噪声将图像转换为完全高斯图像。具体实现中, 首先使用 LAFI 数据集中的标记图像生成船舶掩模图像作为输入, 然后通过逐步增加噪声强度获得完全高斯图像, 作为反向扩散过程的输入。每一时间步的条件概率 $q(z_t|z_{t-1})$ 符合高斯分布 $\mathcal{N}(\mu, \sigma^2 I)$, 则每一步的扩散过程定义为:

$$q(z_t|z_{t-1}) = \mathcal{N}(z_t; \sqrt{1 - \beta_t} z_{t-1}, \beta_t I) \quad (1)$$

其中, z_{t-1} 表示前一时间步 $t-1$ 的高斯图像, z_t 表示当前时间步 t 的高斯图像, β_t 是预定义的噪声调度参数, I 为单位矩阵。每一步的噪声添加量计算如下:

$$z_t = \sqrt{\alpha_t} z_{t-1} + \sqrt{1 - \alpha_t} \epsilon_{t-1} \quad (2)$$

其中 $\alpha_t = 1 - \beta_t$, 表示信号保留系数, ϵ_{t-1} 为时间步 t 添加的标准高斯噪声, 用于维持生成过程的随机性, 经过 N 次加噪, 得到完全高斯图像 z_N 。

3.1.3 反向扩散过程

反向扩散过程通过学习条件概率分布 $p_\theta(z_{t-1}|z_t, c)$ 来逐步去噪, 其中 c 表示条件信息, 以船舶场景提示为条件约束。神经网络被训练来预测添加的噪声:

$$p_\theta(z_{t-1}|z_t, c) = \mathcal{N}(z_{t-1}; \mu_\theta(z_t, t, c), \sigma_t^2 I) \quad (3)$$

其中, $\mu_\theta(z_t, t, c)$ 和 σ_t^2 为神经网络预测的均值和预定义的方差, ϵ_θ 为神经网络预测的噪声, 预测均值计算为:

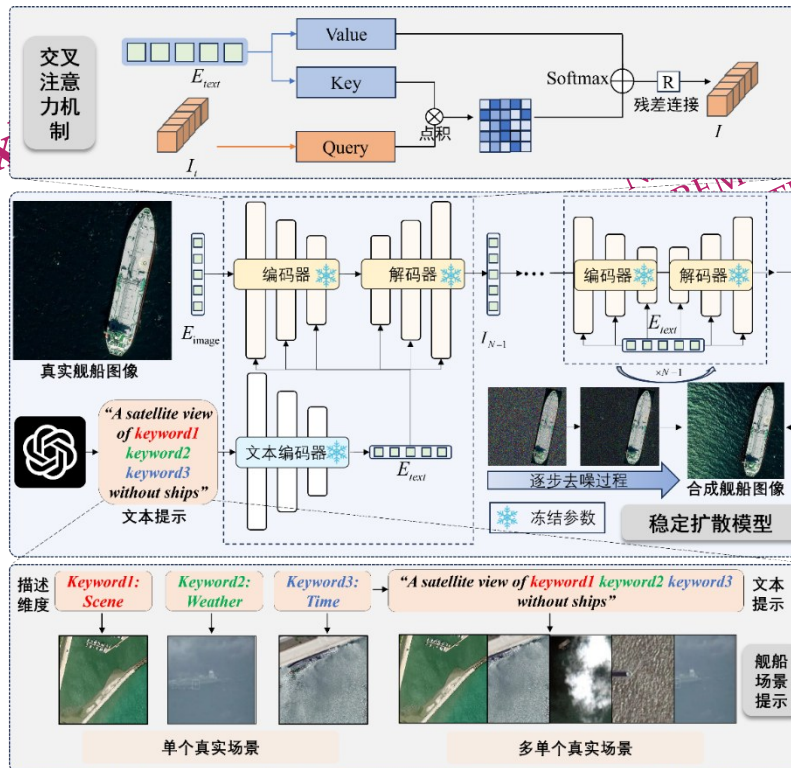


图4 从扩散模型生成合成图像的具体过程

Fig. 4 Specific process for generating synthetic images from diffusion models

$$\mathcal{K}_\theta(z_i, t, c) = \frac{1}{\sqrt{\alpha_i}} \left(z_i - \frac{\beta_i}{\sqrt{1 - \alpha_i}} \epsilon_\theta(z_i, t, c) \right) \quad (4)$$

基于完全高斯图像 z_N 和船舶场景提示，去噪过程通过逐步预测并移除各时间步的噪声分量，以逆向方式重建合成图像。具体而言，首先利用预训练编码器将完全高斯图像 z_N 转换为初始图像嵌入 I_n ，该编码器继承稳定扩散模型的冻结编码器权重。同时，通过 CLIP 文本编码器将船舶场景提示编码为文本特征 E_{text} 。随后，将 I_n 和 E_{text} 输入编码器-解码器网络，从标准高斯噪声 $\epsilon_t \sim \mathcal{N}(0, I)$ 开始，通过 DDPM 采样器逐步去噪，从而获得时间步 $t-1$ 的去噪图像 z_{t-1} 。

$$z_{t-1} = \frac{1}{\sqrt{\alpha_i}} \left(z_i - \frac{\beta_i}{\sqrt{1 - \alpha_i}} \epsilon_\theta(z_i, t, c) \right) + \sigma_t z \quad (5)$$

其中， α_t 和 β_t 为时间步 t 的高斯噪声参数，这些超参数可在训练过程中进行优化； z 为 DDPM 采样中的随机噪声项。

3.1.4 迭代去噪机制

在反向扩散过程中，需要经过多次的去噪才能得到最终的合成图像，这个过程叫做迭代去噪。

在迭代去噪过程中，基于时间步 t 的高斯图像 z_t ，采用相同的预训练编码器提取潜在表示 I_t ，即时间步 t 的图像嵌入。结合船舶场景提示的文本嵌入 E_{text} ，计算时间步 t 的高斯噪声 ϵ_t ：

$$\epsilon_t = UNet(I_t, E_{text}) \quad (6)$$

在计算 ϵ_t 时，为了融入文本条件信息，模型在 U-Net 的每个注意力层中采用交叉注意力机制，通过注意力计算将海洋环境的语义信息融入图像特征。交叉注意力机制的本质就是让图像生成过程（U-Net 处理的图像特征）随时“询问”或“参考”文本描述（海洋环境的语义信息），以确保生成的每一个像素都与文本提示保持一致。

具体而言，在 U-Net 的任意一个注意力层中，交叉注意力模块接收三个输入：从 U-Net 当前层的图像特征图获取来自图像的 Query 和来自文本的 Key 和 Value，计算每个图像 Query 与所有文本 Key 的点积，然后进行缩放和 Softmax 操作，生成一个注意力图：

$$Attention(Q, K, V) = softmax \left(\frac{QK^T}{\sqrt{d}} \right) V \quad (7)$$

其中， $Q = W_Q \cdot \phi_i(z_i)$ ， $K = W_K \cdot \tau_\theta(c)$ ， $V = W_V \cdot \tau_\theta(c)$ ， $\phi_i(z_i)$ 是第 i 层的图像特征， $\tau_\theta(c)$ 是文本编

码器 CLIP 的输出。而后, 使用上一步得到的注意力图对文本的 Value 向量进行加权求和, 输出一个新的特征向量, 通过一个残差连接与原始的图像特征相加, 从而将文本条件信息注入到图像特征中。经过 N 个迭代去噪步骤后, 最终从 z_N 生成包含多样化海洋环境的合成船舶图像。

这种架构设计使得稳定扩散模型能够在保持生成质量的同时显著降低计算成本, 并通过条件机制实现精确的文本到图像生成控制。因而, 该模型能够将船舶场景提示中蕴含的复杂海洋环境信息有效融入生成图像, 从而实现基于 LAFI 数据集的海洋条件多样性扩展。

3.2 数据生成

本研究利用 CSDG 方法实现船舶位置约束的图像扩散生成。该方法通过大语言模型将分层环境因素 (如海洋场景、天气、时间等) 的文本描述转换为船舶知识向量, 然后基于稳定扩散模型, 利用 LAFI 数据集中的真实船舶图像与船舶知识向量合成各种海事条件下的大规模船舶数据。为了确保合成图像与真实图像的一致性, 方法中设计了文本提示过滤策略来排除低质量的文本描述。

首先, 利用大语言模型生成多样化的船舶场景文本描述。基于对真实海洋环境的观察, 船舶可能出现在各种场景中 (如海岸线和开阔海域), 在不同的天气条件下 (如阴天和海浪), 以及在不同的时间 (如傍晚和日出)。因此, 本文将海洋条件分层概括为三个描述维度: 场景、天气和时间 <scene, weather, and time>, 这些维度可以全面表征大多数包含船舶的遥感影像。将这些维度组织成标准的句子格式 “A satellite view of {keyword1: scene} {keyword2: weather} {keyword3: time}”, 用作船舶场景提示。

为了确保每个描述维度的多样性并避免信息冗余, 本文使用 ChatGPT (Biswas S S, 2023) 为每个维度生成 100 个提示词, 每个提示词代表一种唯一的真实场景特征。为了防止背景中出现模糊或不真实的船舶物体, 在提示中添加了 “without ships” 限制词, 以避免在背景中生成船舶物体。通过这三个描述维度的组合, 船舶场景提示相比单一维度具有更高的语义丰富度, 能够更好地捕捉现实世界多样化的海洋环境。

最后, 以船舶场景提示为文本引导, 通过稳

定扩散模型生成大规模合成图像。需要注意的是, 为了保证合成图像中船舶物体的真实性, 方法在 LAFI 数据集中固定了船舶位置, 并仅通过船舶场景提示调整船舶物体周围的环境。

3.3 prompt 筛选

在合成图像生成过程中, 语义模糊的文本提示会导致低质量样本的产生。为保证合成数据集质量, 本文设计了基于 CLIP 的文本提示过滤策略。该策略的核心思想是通过计算合成图像与船舶类别标签文本基于 CLIP 模型的语义相似度来评估每个提示的质量 (图 5)。相似度得分越低, 表明海洋环境对船舶类别识别的干扰越大。

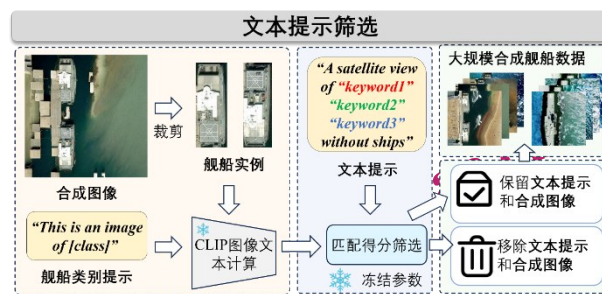


图5 文本提示筛选的具体机制,用于丢弃低质量的文本提示和合成图像

Fig. 5 Specific Mechanism of Textual Prompt Filtering, which are used to discard the low-quality textual prompts and synthetic images.

具体而言, 提示质量评估基于其生成的合成图像质量进行。如果某个提示生成的图像中超过 70% 的 CLIP 相似度得分低于预定阈值, 则该提示被视为低质量并被丢弃。考虑到一张合成图像可能包含多种类型的船舶, 本文将每个船舶实例及其周围的局部区域进行裁剪, 并将其与对应的船舶类别文本 (例如, “This is an image of [class label]”) 进行相似度评估。高相似度得分表示海洋环境对船舶类别识别的干扰较小, 低相似度得分则表示存在较大干扰。筛选后描述符的详细信息如表 2 所示。

最后, 我们使用筛选后的文本提示及其对应的合成图像, 构建了一个百万级数据集, 称为 LAFI-Diffusion, 数据集中各类别的数量信息如表 3 所示。

图 6 展示了基于 LAFI 生成的具有多种海洋条件的合成船舶图像。例如, 在提示 “A satellite

view of a vast ocean with turbulent waves at sunset without ships”下，合成图像显示海面波涛汹涌，复杂海况中的真实状态得到了逼真的呈现，如图6(a)所示。在提示“A satellite view of a coastal with

clam waters at midday without ships”下，合成图像显示岸边的海面平静，船只轮廓清晰，波浪平滑，准确反映了海洋状况，如图6(b)所示。

表2 文本提示关键词

Table 2 Keywords in textual prompt

Keyword	Description1	Description2	Description3	Description4	Description5
Keyword1:Scene	a deep sea	a vast ocean	a coastal	a shallow sea	a rocky coastline
Keyword2:Weather	with turbulent waves	with gentle waves	with mist covering	with calm waters	with stormy waves
Keyword3:Time	at sunset	at dawn	at twilight	at midday	at sunrise

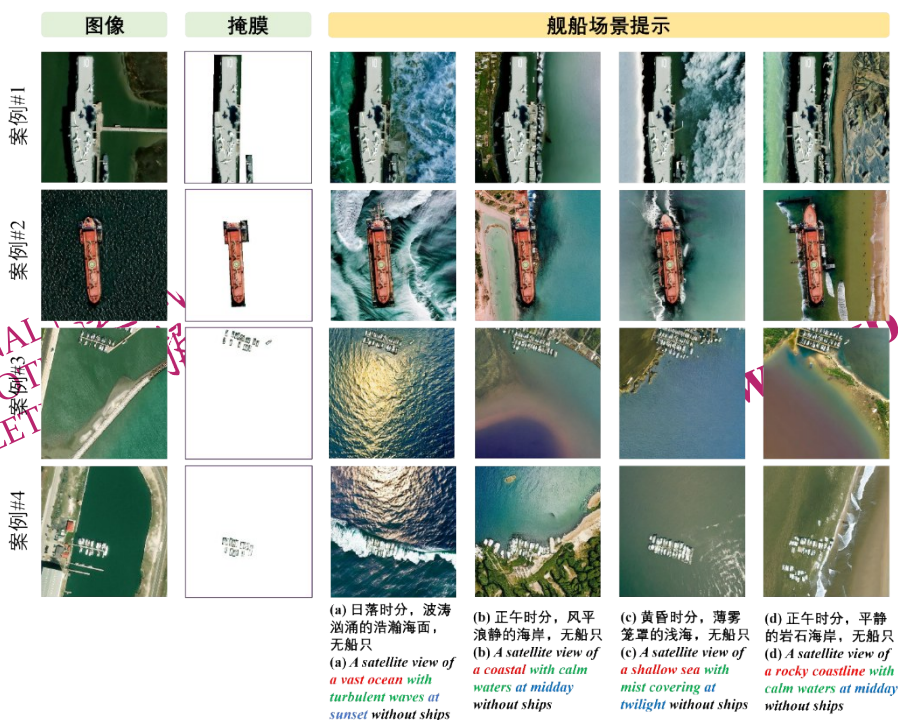


图6 通过提供相应的视觉-文本提示,显示生成的合成遥感图像

Fig. 6 Display of synthetic remote sensing images generated from CSDG by providing the corresponding visual - textual prompt.

上述视觉分析表明，LAFI-Diffusion 是 CSDG 方法生成的包含多样化海洋场景、天气条件和时间的高质量船舶数据，同时能通过精准的文本提示控制生成内容，实现船舶与背景之间的自然融合，从而验证了 LAFI-Diffusion 基于 LAFI 数据集扩展多样化海洋环境的有效性。

表3 LAFI-Diffusion 数据集中舰船的分类及其数量

Table 3 Categories and quantities of ships in the LAFI-Diffusion dataset.

class	LAFI	diffuse@	diffuse @	diffuse @
		1:10	1:50	1:125
Other_Ship	2417	24170	120850	302125
Other_Warship	1929	19290	96450	241125

续表

class	LAFI	diffuse@	diffuse @	diffuse @
		1:10	1:50	1:125
Submarine	914	9140	45700	114250
Other_Aircraft_Carrier	36	360	1800	4500
Enterprise	96	960	4800	12000
Nimitz	107	1070	5350	13375
Midway	18	180	900	2250
Ticonderoga	344	3440	17200	43000
Other_Destroyer	221	2210	11050	27625
Atago_DD	222	2220	11100	27750
Arleigh_Burke_DD	606	6060	30300	75750
Hatsuyuki_DD	127	1270	6350	15875

续表

class	LAFI	diffuse @		
		1:0	1:50	1:125
Hyuga_DD	94	940	4700	11750
Asagiri_DD	65	650	3250	8125
Other_Frigate	187	1870	9350	23375
Perry_FF	514	5140	25700	64250
Patrol	193	1930	9650	24125
Other_Landing	113	1130	5650	14125
YuTing_LL	86	860	4300	10750
YuDeng_LL	71	710	3550	8875
YuDao_LL	51	510	2550	6375
YuZhao_LL	40	400	2000	5000
Austin_LL	124	1240	6200	15500
Osumi_LL	39	390	1950	4875
Wasp_LL	20	200	1000	2500
LSD_41_LL	147	1470	7350	18375
LHA_LL	167	1670	8350	20875
Commander	154	1540	7550	18875
Other_Auxiliary_Ship	88	880	4400	11000
Medical_Ship	31	310	1550	3875
Test_Ship	54	540	2700	6750
Training_Ship	51	510	2550	6375
AOE	48	480	2400	6000
Masyuu_AS	61	610	3050	7625
Sanantonio_AS	66	660	3300	8250
EPF	80	800	4000	10000
Other_Merchant	762	7620	38100	95250
Container_Ship	522	5220	26100	65250
RoRo	261	2610	13050	32625
Cargo	3084	30840	154200	385500
Barge	1460	14600	73000	182500
Tugboat	313	3130	15650	39125
Ferry	304	3040	15200	38000
Yacht	1613	16130	80650	201625
Sailboat	3122	31220	156100	390250
Fishing_Vessel	455	4550	22750	56875
Oil_Tanker	339	3390	16950	42375
Hovercraft	672	6720	33600	84000
Motorboat	26232	262320	1311600	3279000
all	48717	487170	2435850	6089625

4 实验

4.1 实验设置

本文的所有实验均在四块 NVIDIA RTX 4090 GPU 上进行, 每块 GPU 配备 24GB 显存。模型在 LAFI-Diffusion 中训练 12 个周期 (epoch), 采用随机梯度下降 (SGD, Stochastic Gradient Descent) 优化器, 动量因子和衰减系数分别设置为 0.9 和 0.0001。实验初始学习率为 0.01, 学习率经过每个衰减步长下降为原值的十分之一。在本次研究的所有实验中, 我们均采用在 ImageNet-1K 数据集 (Deng J 等, 2009) 上预训练过的权重来初始化检测模型的主干网络。实验默认使用 4 块 NVIDIA RTX 4090 GPU 进行训练 (batch size 为 8), 并使用一块 NVIDIA RTX 4090 GPU 进行推理。所有方法均基于 mmrotate 框架 (Zhou Y 等, 2022) 实现。

4.2 对比方法

对于目标检测性能的比较, 本文选择了 7 种基于 OBB 的目标检测方法, 包括 RoI Transformer (Ding J 等, 2019)、R3Det (Yang X 等, 2021)、Gliding Vertex (Xu Y 等, 2020)、Oriented R-CNN (Xie X 等, 2021)、Rotated Faster R-CNN (Yang S 等, 2020)、Rotated FCOS (Li Z 等, 2023) 和 S2A-Net (Han J 等, 2021)。

在评价指标方面, 本文遵循 PASCAL VOC 2007 (Everingham M 等, 2010) 基准, 并使用平均精度 (mAP) 来评估模型性能。同时, 本文还对每个模型的参数规模和计算效率进行了统计。

4.3 实验结果与讨论

本研究通过不同的真实-合成图像混合比例和不同检测方法的对比, 评估了 LAFI-Diffusion 数据集对舰船检测性能的提升效果。首先建立仅使用真实数据的基线模型进行训练, 真实-合成图像混合比例为 1: 0; 随后采用 1: 10、1: 50 和 1: 125 三种不同的真实-合成图像混合比例进行对比实验; 最后在真实影像的测试集上评估 7 种基于有向边界框 (OBB) 的先进目标检测算法的性能。

各方法在 LAFI-Diffusion 数据集上的性能表现如表 4 所示。实验结果显示当真实图像与生成图像的比例为 1: 50 时, 各模型达到了最优的检测精度。这一混合比例既充分发挥了合成数据增强模

型泛化能力的优势，又有效避免了因过度依赖生成数据而导致的性能下降，实现了检测精度与模型鲁棒性的最佳平衡。不同方法之间的对比表明，在 LAFI-Diffusion 数据集上，S2A-Net 方法展现出卓越的检测性能，其 mAP 值在所有混合比例下均保持领先。Oriented R-CNN 同样表现优异，进一步验证了数据增强策略的有效性。以上实验结果证实了 LAFI-Diffusion 数据集的有效性。

表 4 LAFI-Diffusion 数据集上的性能比较 (AP)

Table 4 Comparison of performance (AP) on LAFI dataset

方法	backbone	mAP			
		mAP @1:0	mAP @1:10	mAP @1:50	mAP @1:125
RoI Transformer	ResNet50	0.509	0.523	0.533	0.514
R3Det	ResNet50	0.278	0.280	0.281	0.280
Gliding Vertex	ResNet50	0.489	0.496	0.501	0.494
Oriented R-CNN	ResNet50	0.568	0.568	0.574	0.568
Rotated Faster R-CNN	ResNet50	0.476	0.485	0.490	0.477
Rotated FCOS	ResNet50	0.421	0.422	0.424	0.423
S2A-Net	ResNet50	0.585	0.605	0.620	0.609

注：粗体表示最优精度

图 7 的定性分析进一步验证了本研究所用方法的优势。与其他基于 OBB 的目标检测器相比，S2A-Net 方法有效避免了对大型船舶的重复预测问题，并能够准确识别油轮和滚装船等细粒度类别。在船舶分散的场景下，该方法也能正确识别小型目标及其细粒度类别，如摩托艇，这得益于其对旋转和分布特征的协同优化，从而显著提升了模型对不同形态船舶的适应能力。

然而，合成数据集与真实数据集在多个维度上存在一定差异，这些差异可能直接影响模型的训练效果与实际应用中的泛化能力。合成图像虽然能生成逼真图像，但在材质反射、光学畸变、阴影过渡等物理特性方面与真实成像存在细微差距。缺少自然的光照变化、大气扰动等真实噪声模式；通过参数控制可以生成不同海况、气象条件，但环境组合可能不够自然；有时合成数据倾向于生成“典型”场景，难以完全复现真实世界中罕见的异常情况，因而可能会产生分布偏移，如过度强调某些显性的上下文特征。此外，我们的数据集整合了多源遥感数据，合成数据在跨传感器迁移时可能出现性能下降，因不同卫星的成

像特性产生差异。

为了检验生成图像的质量，本文引入了生成模型中广泛使用的 FID 以及 LPIPS 指标对生成图像的质量和差异性进行定量评估。FID (Fréchet Inception Distance) 是 (Heusel M 等, 2018) 提出的一种生成图像质量的衡量标准，通过计算两个分布之间的 Fréchet 距离来衡量生成模型和真实数据分布之间的差异，FID 值越小代表数据生成质量越好。LPIPS (Learned Perceptual Image Patch Similarity) (Zhang R 等, 2018) 通过深度学习模型来评估两个图像之间的感知差异，LPIPS 值越小代表合成数据和真实数据越相似，数据的生成质量越好。表 5 展示了我们的方法和其他遥感图像生成的精度对比。可以发现，我们的遥感图像生成质量相较于其他文章的遥感生成图像质量而言，在 FID 和 LPIPS 指标上均有明显的优势。这表明我们的方法在生成图像质量上更接近真实图像，且能够有效保留图像的细节和结构信息，体现出更高的生成质量。

表 5 图像生成质量的比较

Table 5 Comparison of image generation quality

方法	Stable-diffusion+CLIP	SwinIR(Liang 等, 2021)	TTST(Xiao Y 等, 2024)
	数据集	LAFI-diffusion	DIV2K
FID	98.1136	390	327.5
LPIPS	0.7708	0.903	0.875

为了更全面地评估细粒度船型识别性能，本文评估了各方法在所有类型船只上的检测精度。实验结果如表 6 与表 7 所示，S2A-Net 和 Oriented R-CNN 方法在绝大多数船型类别上达到最优检测精度，与定量分析结论一致。其中，S2A-Net 方法在几乎所有商船和部分军舰中达到了最高精度，表明该方法在舰船的识别中更具优势。这一优势主要源于对局部舰船部件特征和海洋环境特征的高效建模，同时有效抑制了复杂背景的干扰，从而提升了细粒度分类的鲁棒性。

然而，S2A-Net 作为单阶段检测方法，其检测精度却整体优于双阶段 Oriented R-CNN 方法，我们认为这主要得益于它更优异的特征表示能力以及与数据集的适配性。

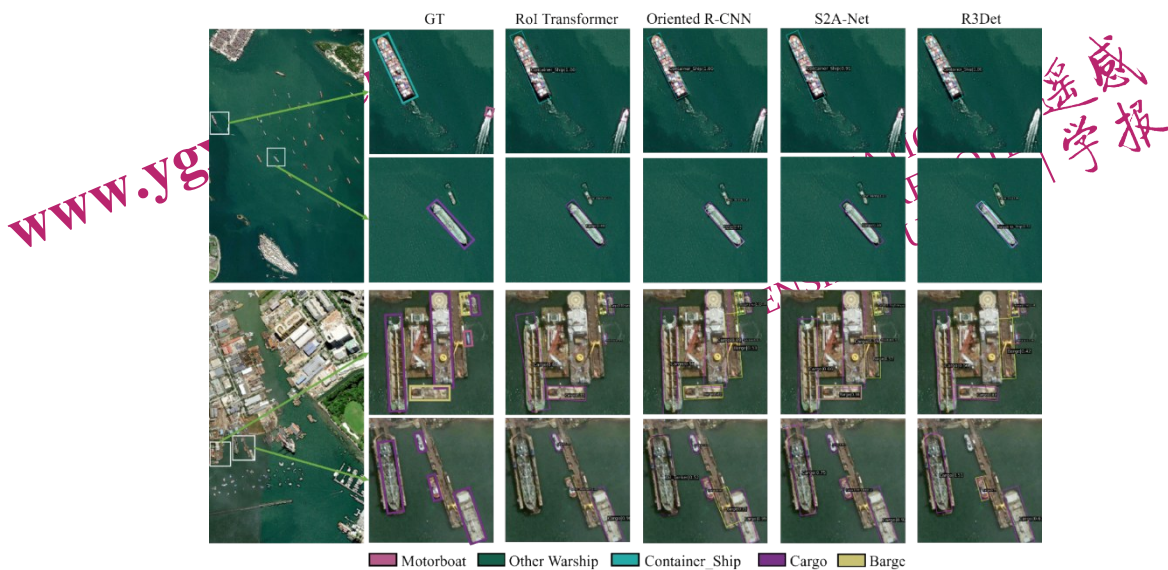


图7 LAFI-Diffusion数据集上不同旋转目标检测方法的检测结果,包括港口和开阔海域。

Fig. 7 Detection results from different rotate object detection methods on LAFI datasets, including the scenario of port and open sea.

具体而言, Oriented R-CNN 首先通过区域提议网络生成旋转的候选框,然后通过 RoI Align 或类似的池化操作从特征图中提取固定大小的特征,这在处理大长宽比、任意方向的舰船时,可能引入未对齐的特征,导致后续的分类和回归基于有噪声的特征进行预测。相比之下, S2A-Net 虽然是单阶段,但其内置的隐式特征对齐机制实现了特征的重新对齐和对齐后特征的细化。该特征对齐模块通过卷积层隐式地学习每个候选位置与目标方向最匹配的特征表示,这种端到端的学习方式,能够更灵活、更精准地适应各种旋转角度和形状,为后续的分类和回归提供了质量更高的特征。这是其精度更高的核心原因之一。

此外, S2A-Net 的方法对数据集特性更加匹配也是一个重要原因。港口中的舰船常密集排列且方向各异, S2A-Net 优秀的特征对齐能力和稳定的旋转框回归方法,能更好地处理舰船间的紧邻和重叠问题。对于大长宽比的舰船目标, S2A-Net 的八参数回归对这种不规则形状的拟合能力也优于 Oriented R-CNN 使用的传统五参数法。

尽管各类方法在整体指标上取得了良好的性能(见表 8),但通过分析各类别的检测结果(见表 5 和表 6),我们发现部分舰船类别的检测精度(AP)显著低于平均水平,尤其是“Training Ship”、“Fishing_Vessel”等类别。我们分析了导致这些类别精度较低的主要原因,归纳如下:

首先,样本数量不足是导致小样本类别精度

偏低的核心原因。有限的训练样本使得模型难以学习到这些类别鲁棒的特征表示,如“Training Ship”仅有 51 个实例。尽管 LAFI-Diffusion 数据集通过扩散生成技术缓解了此问题,但在极端小样本类别上仍存在不足。

其次,类间视觉特征相似性高导致一些类别难以检测。某些精度低的类别与一些高频类别在视觉上具有高度的相似性,导致模型难以区分。如“Fishing_Vessel”,该类别在尺寸、空间形态上与“Sailboat”“Motorboat”较为接近,尤其在大尺度遥感图像中区分度很低,模型容易出现误判。

此外,目标尺度与成像质量的限制也是原因之一。部分低精度类别本身成像条件较差,例如“Fishing_Vessel”等小型船舶成像模糊,属于小目标检测难题。其特征信息难以识别,更容易被复杂海况背景(如波浪、泡沫)所淹没。

基于以上分析,后续工作中可以采取以下针对性措施:首先,设计针对性的生成策略,在扩充数据时,为重点小样本类别设置更高的生成权重,并强化其与易混淆类别的差异化特征。另外,引入动态重加权等先进的损失函数,以减轻长尾分布带来的模型偏差。

表 6 各种方法在 LAFI-Diffusion 的军舰与其他舰船类别上的性能比较

Table 6 Comparison of performance on LAFI-Diffusion of S2A-Net and our methods in warship and other ship.

Class	RotTransformer	R3det	Gliding Vertex	Oriented R-CNN	Rotated Faster R-CNN	Rotated FCOS	S2A-Net
Other_Ship	0.23	0.192	0.221	0.215	0.212	0.211	0.463
Other_Warship	0.415	0.291	0.354	0.348	0.382	0.385	0.581
Submarine	0.662	0.59	0.64	0.79	0.622	0.739	0.643
Other_Aircraft_Carrier	0.795	0.132	0.599	0.545	0.69	0.563	0.727
Enterprise	0.83	0.743	0.758	0.768	0.739	0.732	0.931
Nimitz	0.711	0.418	0.574	0.636	0.582	0.495	0.728
Midway	0.625	0.292	0.697	0.737	0.818	0.351	0.673
Ticonderoga	0.748	0.626	0.829	0.871	0.719	0.447	0.798
Other_Destroyer	0.328	0.221	0.294	0.488	0.271	0.344	0.626
Atago_DD	0.614	0.465	0.494	0.702	0.431	0.473	0.73
Arleigh_Burke_DD	0.857	0.583	0.752	0.878	0.784	0.598	0.854
Hatsuyuki_DD	0.197	0.209	0.389	0.375	0.275	0.116	0.646
Hyuga_DD	0.875	0.947	0.676	0.873	0.921	0.82	0.975
Asagiri_DD	0.14	0.062	0.275	0.466	0.391	0.167	0.214
Other_Frigate	0.572	0.036	0.246	0.341	0.218	0.272	0.407
Perry_FF	0.739	0.712	0.663	0.834	0.673	0.667	0.836
Patrol	0.103	0.086	0.129	0.132	0.121	0.105	0.152
Other_Landing	0.129	0.016	0.186	0.207	0.041	0.091	0.212
YuTing_LL	0.545	0.009	0.545	0.545	0.545	0.341	0.467
YuDeng_LL	0.244	0.039	0.432	0.778	0.198	0.709	0.759
YuDao_LL	0.561	0.089	0.608	0.761	0.514	0.41	0.656
YuZhao_LL	0.818	0.311	0.948	0.851	0.818	0.818	0.818
Austin_LL	0.737	0.209	0.762	0.773	0.582	0.506	0.727
Osumi_LL	1	0.128	0.818	0.877	0.924	0.724	1
Wasp_LL	1	0.035	1	1	1	0.273	0.545
LSD_41_LL	0.744	0.215	0.586	0.761	0.656	0.24	0.633
LHA_LL	0.775	0.691	0.698	0.714	0.658	0.653	0.761
Commander	0.838	0.222	0.652	0.811	0.646	0.587	0.615
Other_Auxiliary_Ship	0.439	0.022	0.388	0.367	0.276	0.255	0.53
Medical_Ship	0	0.003	0.061	0.508	0.209	0.001	0.008
Test_Ship	0.492	0.022	0.583	0.692	0.523	0.395	0.733
Training_Ship	0.455	0.005	0.455	0.459	0.462	0.459	0.33
AOE	0.916	0.105	0.667	0.8	0.224	0.311	0.796
Masyuu_AS	0.289	0.065	0.104	0.244	0.255	0.209	0.463
Sanantonio_AS	0.689	0.114	0.687	0.729	0.62	0.154	0.673
EPF	0.495	0.164	0.442	0.556	0.485	0.464	0.631

注：粗体表示最优精度

表 7 各种方法在 LAFI-Diffusion 的商船类别上的性能比较

Table 7 Comparison of performance on LAFI-Diffusion of S2A-Net and other methods in merchant ship.

Class	RoI Transformer	R3det	Gliding Vertex	Oriented R-CNN	Rotated Faster R-CNN	Rotated FCOS	S2A-Net
Other_Merchant	0.196	0.146	0.138	0.164	0.144	0.189	0.486
Container_Ship	0.447	0.413	0.366	0.431	0.427	0.371	0.632
RoRo	0.536	0.393	0.52	0.541	0.527	0.539	0.684
Cargo	0.501	0.414	0.392	0.52	0.432	0.501	0.664
Barge	0.376	0.32	0.343	0.37	0.354	0.369	0.629
Tugboat	0.424	0.392	0.457	0.433	0.49	0.446	0.758
Ferry	0.346	0.143	0.336	0.322	0.375	0.242	0.602
Yacht	0.548	0.392	0.495	0.488	0.472	0.523	0.611
Sailboat	0.333	0.272	0.306	0.32	0.251	0.375	0.498
Fishing_Vessel	0.291	0.136	0.247	0.237	0.259	0.203	0.411
Oil_Tanker	0.434	0.329	0.338	0.387	0.421	0.431	0.549
Hovercraft	0.735	0.618	0.659	0.734	0.704	0.734	0.744
Motorboat	0.744	0.716	0.723	0.734	0.723	0.789	0.77

注: 粗体表示最优精度

表 8 各种方法在 LAFI-Diffusion 上的综合性能比较

Table 8 Comparison of performance on LAFI-Diffusion of S2A-Net and other methods

	RoI Transformer	R3det	Gliding Vertex	Oriented R-CNN	Rotated Faster R-CNN	Rotated FCOS	S2A-Net
mAP@1:50	0.533	0.281	0.501	0.574	0.490	0.424	0.620
mAR@1:50	0.701	0.831	0.689	0.774	0.694	0.858	0.888
F1 score	0.563	0.373	0.568	0.648	0.561	0.541	0.713

注: 粗体表示最优精度

5 结论

本文构建的 LAFI-Diffusion 数据集在舰船目标检测领域具有重要的研究价值和应用意义。针对现有数据集在规模、多样性和细粒度方面的不足, 我们通过整合全球 36 个主要港口的 8000 张高分辨率遥感图像, 构建了包含 49 类舰船、48717 个实例的大规模基准数据集, 并创新性地采用稳定扩散模型将数据规模扩展至百万级别, 有效解决了细粒度舰船检测领域的数据瓶颈问题。

在 LAFI-Diffusion 数据集上, 我们系统评估了 7 种主流目标检测算法的性能, 其中 S2A-Net 表现最优, 在不同真实-合成图像混合比例下均保持领先。实验结果表明, 当真实图像与生成图像比例为 1:50 时, 模型能够获得最佳的检测性能, 这一发现为后续研究提供了重要的数据使用参考。我们期待该数据集能够推动遥感图像舰船检测技术的进一步发展, 并为海洋监视、航运管理等实际应用提供有力支持。

参考文献 (References)

- Huang J, Jiang Z G, Zhang H P and Yao Y. 2017. Ship target detection in remote sensing images based on convolutional neural networks. Journal of Beijing University of Aeronautics and Astronautics. (黄洁, 姜志国, 张浩鹏, 姚远. 2017. 基于卷积神经网络的遥感图像舰船目标检测. 北京航空航天大学学报[DOI: 10.13700/j.bh.1001-5965.2016.0755])
- Wang Y Q, Ma L and Tian Y. 2011. Review of ship detection and recognition in optical remote sensing images. Acta Automatica Sinica, 37(9): 1029 - 1039. (王彦清, 马雷, 田原. 2011. 光学遥感图像舰船目标检测与识别综述. 自动化学报, 37(9): 1029-1039 [DOI: 10.3724/SP.J.1004.2011.01029])
- Xie X X, Cheng C, Yao Y Q, Yao X W and Han J W. 2022. Dynamic feature fusion for object detection in remote sensing images. Chinese Journal of Computers, 45(4): 735 - 747. (谢星星, 程琳, 姚艳清, 姚西文, 韩军伟. 2022. 动态特征融合的遥感图像目标检测. 计算机学报, 45(4): 735-747[DOI: 10.11897/SP. J. 1016.2022. 00735])
- Xu Z J and Bai X. 2022. Small ship detection in remote sensing images based on dual feature enhancement. Acta Optica Sinica, 42

- (18): 1828002. (徐志京, 柏雪. 2022. 基于双重特征增强的遥感舰船小目标检测. *Acta Optica Sinica*, 42(18): 1828002[DOI: 10.3788/AOS202242.1828002])
- Yao Y Q, Cheng G, Xie X X and Han J W. 2021. Multi-resolution feature fusion for object detection in optical remote sensing images. *National Remote Sensing Bulletin*, 25(5): 1124 - 1137. (姚艳清, 程塔, 谢星星, 韩军伟. 2021. 多分辨率特征融合的光学遥感图像目标检测. *遥感学报*, 25(5): 1124-1137[DOI: 10.11834/jrs.20210505])
- Yu W Q, Cheng G, Wang M J, Yao Y Q, Xie X X, Yao X W and Han J W. 2023. MAR20: A military aircraft recognition dataset for remote sensing images. *National Remote Sensing Bulletin*, 27(12): 2688 - 2696. (禹文奇, 程塔, 王美君, 姚艳清, 谢星星, 姚西文, 韩军伟. 2023. MAR20: 遥感图像军用飞机目标识别数据集. *遥感学报*, 27(12): 2688-2696[DOI: 10.11834/jrs.20222139])
- Zhang Y, Zhu G Y, Shi T J, Zhang K and Yan J H. 2022. Small object detection in remote sensing images based on feature fusion and attention mechanism. *Acta Optica Sinica*, 42(24): 2415001. (张寅, 朱桂熠, 施天俊, 张琨, 闫钧华. 2022. 基于特征融合与注意力的遥感图像小目标检测. *Acta Optica Sinica*, 42(24): 2415001 [DOI: 10.3788/AOS202242.2415001])
- Zhou P C, Cheng G, Yao X W and Han J W. 2021. Machine learning paradigms in high-resolution remote sensing image interpretation. *National Remote Sensing Bulletin*, 25(1): 182 - 197. (周培诚, 程塔, 姚西文, 韩军伟. 2021. 高分辨率遥感影像解译中的机器学习范式. *遥感学报*, 25(1): 182-197[DOI: 10.11834/jrs.20210164])
- Biswas S. 2023. Potential use of chat gpt in global warming. *Annals of biomedical engineering*, 51(6): 1126-1127
- Chen K, Wu M, Liu J, Zhang C. 2020. FGSD: A dataset for fine-grained ship detection in high resolution satellite images. arXiv preprint arXiv:2003.06832
- Deng J, Dong W, Socher R, Li L J, Li K, Fei-Fei L. 2009. Imagenet: A large-scale hierarchical image database. 2009 IEEE conference on computer vision and pattern recognition. Ieee: 248-255
- Ding J, Xue N, Long Y, Xia G S, Lu Q. 2019. Learning RoI transformer for oriented object detection in aerial images. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2849-2858
- Everingham M, Van Gool L, Williams C K, Winn J, Zisserman A. 2010. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88: 303-338
- Guo B, Zhang R, Guo H, Yang W, Yu H, Zhang P, Zou T. 2023. Fine-Grained Ship Detection in High-Resolution Satellite Images With Shape-Aware Feature Learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16: 1914-1926
- Han J, Ding J, Li J, Xia G S. 2021. Align deep features for oriented object detection. *IEEE transactions on geoscience and remote sensing*, 60: 1-11
- Heusel M, Ramsauer H, Unterthiner T, Nessler B, Hochreiter S. 2018. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. arXiv
- Li K, Wan G, Cheng G, Meng L, Han J. 2020. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS journal of photogrammetry and remote sensing*, 159: 296-307
- Li Z, Hou B, Wu Z, Ren B, Yang C. 2023. FCGSR: A simple anchor-free rotated detector for aerial object detection. *Remote Sensing*, 15(23): 5499
- Liu L, Bai Y, Li Y. 2021. Locality-aware rotated ship detection in high-resolution remote sensing imagery based on multiscale convolutional network. *IEEE Geoscience and Remote Sensing Letters*, 19: 1-5
- Liu Z, Yuan L, Weng L, Yang Y. 2017. A high resolution optical satellite image dataset for ship recognition and some new baselines. *International conference on pattern recognition applications and methods: 卷 2*. SciTePress: 324-331
- Nie G, Huang H. 2023. Multi-Oriented Object Detection in Aerial Images With Double Horizontal Rectangles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4): 4932-4944
- Ouyang L, Fang L, Ji X. 2022. Multigranularity self-attention network for fine-grained ship detection in remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15: 9722-9732
- Rombach R, Blattmann A, Lorenz D, Esser P, Ommer B. 2022. High-resolution image synthesis with latent diffusion models. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10684-10695
- Russell B C, Torralba A, Murphy K P, Freeman W T. 2008. LabelMe: a database and web-based tool for image annotation. *International journal of computer vision*, 77: 157-173
- Tupper E C. 2013. *Introduction to naval architecture*. Butterworth-Heinemann
- Xia G S, Bai X, Ding J, Zhu Z, Belongie S, Luo J, Datcu M, Pelillo M, Zhang L. 2018. DOTA: A large-scale dataset for object detection in aerial images. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3974-3983
- Xie X, Cheng G, Wang J, Yao X, Han J. 2021. Oriented R-CNN for object detection. *Proceedings of the IEEE/CVF international conference on computer vision*. 3520-3529
- Xu X, Zhang X, Wei S, Shi J, Zhang W, Zhang T, Zhan X, Xu Y, Zeng T. 2025. DiffSARShipInst: Diffusion model for ship instance segmentation from synthetic aperture radar imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 223: 440-455
- Xu Y, Fu M, Wang Q, Wang Y, Chen K, Xia G S, Bai X. 2020. Gliding vertex on the horizontal bounding box for multi-oriented object detection. *IEEE transactions on pattern analysis and machine intelligence*, 43(9): 1452-1459
- Yang S, Pei Z, Zhou F, Wang G. 2020. Rotated faster R-CNN for oriented object detection in aerial images. *Proceedings of the 2020 3rd International Conference on Robot Systems and Applications*. 35-39
- Yang X, Yan J, Feng Z, He T. 2021. R3det: Refined single-stage detector with feature refinement for rotating object. *Proceedings of the AAAI conference on artificial intelligence: 卷 35*. 3163-3171

- Yang X, Yan J, Liao W, Yang X, Tang J, He T. 2023. SCRDet++: Detecting Small, Cluttered and Rotated Objects via Instance-Level Feature Denoising and Rotation Loss Smoothing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2): 2384-2399
- Ye J, He L, Zhang X, Lin Y, Lin H, He C, Li W. 2025. Satellite Image Synthesis From Street View With Fine-Grained Spatial Textual Guidance: A novel framework. *IEEE Geoscience and Remote Sensing Magazine*, 13(3): 395-414
- Yu Y, Yang X, Li Y, Han Z, Da F, Yan J. 2025. Wholly-WOOD: Wholly Leveraging Diversified-Quality Labels for Weakly-Supervised Oriented Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(6): 4438-4454
- Zhang R, Isola P, Efros A A, Shechtman E, Wang O. 2018. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. arXiv
- Zhang Z, Zhang L, Wang Y, Feng P, He R. 2021. ShipRSImageNet: A large-scale fine-grained dataset for ship detection in high-resolution optical remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14: 8458-8472
- Zheng Z, Ye R, Hou G, Ren D, Wang P, Zuo W, Cheng M M. 2023. Localization Distillation for Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8): 10070-10083
- Zhou Y, Yang X, Zhang G, Wang J, Liu Y, Hou L, Jiang X, Liu X, Yan J, Lyu C, others. 2022. Mmrotate: A rotated object detection benchmark using pytorch. *Proceedings of the 30th ACM International Conference on Multimedia*. 7331-7334

LAFI-diffusion: Construction and Generation Method of Satellite Remote Sensing Ship Target Detection Dataset Based on Controllable Diffusion Model

HE Da¹, LI Zeyu¹, LIU Haoran¹, HU Xikun², ZHONG Ping², SHI Qian

¹School of Geography and Planning, Sun Yat-sen University, Guangzhou 510006, China;

²National Key Laboratory of Science and Technology on Automatic Target Recognition, National University of Defense Technology, Changsha 410073, China

Abstract: Objective Fine-grained ship detection in optical remote sensing imagery is an important research topic in maritime monitoring and ocean observation. Compared with coarse-grained ship detection, fine-grained ship detection aims to distinguish specific ship types with highly similar visual appearances, which places higher demands on dataset scale, category diversity, and annotation accuracy. However, existing ship detection datasets are generally limited in the number of fine-grained categories, instance scale, and scene diversity, which restricts the performance and generalization ability of deep learning - based oriented object detection models. The objective of this study is to construct a large-scale fine-grained ship detection dataset and to further expand data diversity through controllable synthetic data generation, thereby providing a reliable data foundation and benchmark for fine-grained ship detection in optical remote sensing imagery. Method First, a large-scale fine-grained ship detection dataset, named LAFI, is constructed using high-resolution optical remote sensing images collected from 36 representative ports worldwide. The dataset contains 8,000 images acquired under diverse imaging conditions and complex maritime environments. A total of 49 fine-grained ship categories are defined, and 48,717 ship instances are manually annotated using oriented bounding boxes to accurately describe ship orientation and geometric characteristics. Second, to alleviate the limitations of real data in terms of scale and scene coverage, a controllable diffusion-based data generation framework is designed to extend LAFI into a million-scale synthetic dataset, referred to as LAFI-Diffusion. By incorporating structured textual prompts that describe scene types, weather conditions, and temporal information, the diffusion model is guided to generate realistic ship images under diverse maritime scenarios. The generated synthetic samples are further filtered and combined with real data to form large-scale training sets suitable for fine-grained ship detection. Finally, several representative oriented object detection methods are selected and evaluated on the constructed datasets to analyze the effectiveness of synthetic data augmentation and to establish benchmark results. Result Experimental results show that the proposed LAFI dataset provides improved category richness, instance scale, and scene diversity compared with existing fine-grained ship detection datasets. Moreover, incorporating synthetic data from LAFI-Diffusion into the training process consistently improves detection performance and generalization ability across different oriented object detection models. Performance gains are particularly evident in complex maritime environments, such as crowded ports and scenes with varying sea states and illumination conditions. The benchmark evaluations also indicate that the contribution of synthetic data varies across detection methods, suggesting that appropriate integration strategies are important for fully exploiting synthetic data. Conclusion This study presents a large-scale fine-grained ship detection dataset that integrates real optical remote sensing imagery with controllable diffusion-based synthetic data generation. By substantially expanding data scale and enhancing scene diversity, the proposed LAFI-Diffusion dataset effectively addresses the limitations of existing fine-grained ship detection benchmarks. The experimental results confirm that synthetic data can serve as an effective complement to real-world samples, improving detection accuracy and robustness for oriented ship detection models. The released datasets and benchmark results provide valuable support

for future research on fine-grained ship detection and related remote sensing applications.

Key words: Remote sensing imagery, ship detection, dataset, diffusion model, synthetic data generation, fine-grained recognition, oriented bounding box, data augmentation

Supported by Supported by the Key Laboratory Stable Support Research Program (WDZC20245250206) and the Hunan Provincial Science and Technology Innovation Program (2024RC3119)

www.ygxb.ac.cn

NATIONAL
REMOTE
SENSING BULLETIN | 遥感学报

NATIONAL
REMOTE
SENSING BULLETIN | 遥感学报

www.ygxb.ac.cn

www.ygxb.ac.cn

NATIONAL
REMOTE
SENSING BULLETIN | 遥感学报