

本文网址: <http://www.ship-research.com/cn/article/doi/10.19693/j.issn.1673-3185.04536>

期刊网址: www.ship-research.com

引用格式: 李海鑫, 于林, 李祥弘, 等. 避障场景下基于成像声呐的轻量化实时感知方法 [J]. 中国舰船研究, 2026, 21(2): 184-191.

LI H X, YU L, LI X H, et al. Lightweight real-time perception method based on imaging sonar for obstacle avoidance scenarios[J]. Chinese Journal of Ship Research, 2026, 21(2): 184-191 (in Chinese).

避障场景下基于成像声呐的 轻量化实时感知方法



扫码阅读全文

李海鑫, 于林, 李祥弘, 乔磊*

上海交通大学 船舶海洋与建筑工程学院, 上海 200240

摘要: [目的] 针对水下声呐成像过程所面临的高强度噪声和大目标障碍物结构特性, 以及实时水下避障任务对感知算法轻量化部署和高推理效率的严苛要求, 提出一种低计算成本和短时推理特性的声呐图像语义分割算法, 以应对避障需求下感知算法中计算复杂度与实时响应效率之间的矛盾。 [方法] 基于编码器-解码器网络结构, 通过引入轻量化卷积操作以显著降低计算复杂度, 并针对避障场景将大核可分离注意力引入到跳跃连接中。通过对真实采集并标注的 6936 张声呐图像进行训练与对比, 以及在 Gazebo 仿真平台中对基于感知算法的避障策略进行验证。 [结果] 改进后的算法针对性地提高了大目标分割精度, 相较基础模型, 计算量和参数量分别降低 69.2% 和 83%, 同时推理时间减少 22.6%, 感知精度提升 10.8%。仿真实验验证了感知算法在避障过程中的有效性, 可充分满足基于前视声呐水下避障场景下的实时感知任务需求。 [结论] 所提基于声呐图像的感知算法可有效满足水下机器人机载场景下的避障需求, 具有良好的工程应用前景。

关键词: 水下潜航器; 避障; 避碰; 导航; 成像声呐; 传感感知; 图像识别; 语义分割; 轻量化实时感知

中图分类号: U674.941; U666.7

文献标志码: A

DOI: 10.19693/j.issn.1673-3185.04536

0 引言

随着海洋资源勘探、水下基础设施维护及军事安防等领域的快速发展, 水下机器人自主作业已成为深海探测与作业任务的核心技术手段^[1-2]。作为实现复杂水下作业的基础, 自主避障系统直接决定了机器人能否在动态未知环境中安全、高效地执行目标导向型任务。然而, 相较于陆地和空中避障, 水下环境的特殊性对避障问题提出了双重挑战: 其一, 由于电磁波信号在水下存在严重衰减, 光学及无线电传感设备难以稳定工作, 迫使水下机器人依赖单一声呐传感器实现环境感知, 而声呐系统易受水下多径效应、混响噪声及悬浮颗粒散射干扰, 且感知数据低信噪比衍生出的目标漏检与虚警问题, 使得高精度环境感知成为保障水下避障安全性亟待突破的技术瓶颈; 其二, 受限于水下机器人机载平台的有限计算资源与能源供给, 如何在满足避障实时性要求的前提下提升感知方法的精度, 成为平衡感知系统鲁棒

性与效率之间的关键难题。

在声呐图像处理领域, 噪声干扰与特定场景数据的鲁棒性分割始终是技术难点。针对声呐图像噪声和阴影问题, 何义才等^[3]采用 NSCT (nonsub-sampled contourlet transform) 图像分解、目标边缘提取与高低频分割融合的思路, 设计增强模型和多重分形奇异值检测机制, 实验结果表明, 所提方法在抗噪性方面和边缘保持性能方面具有显著优势。针对水下声呐图像质量差和特征模糊问题, 张家敏等^[4]通过降低传统方法的计算量, 采用基于 Hough 变换的快速椭圆检测和三帧差分法实现目标提取, 实验结果表明, 该算法可较好地分离目标周围噪声。Chen 等^[5]基于深度学习框架实现语义分割, 采用未标注/弱标注数据与精细标注样本开展半监督训练, 有效缓解了合成孔径声呐图像标注数据稀缺的瓶颈, 实验结果表明, 所提网络对目标和背景的分割效果较好。

然而, 现有研究在面向水下避障任务时仍存在技术挑战。首先, 成像过程中引入的高强度环

收稿日期: 2025-05-29 修回日期: 2025-09-11 网络首发时间: 2026-03-30 18:37

基金项目: 国家自然科学基金资助项目 (52571358); 上海市自然科学基金资助项目 (25ZR1401190)

作者简介: 李海鑫, 男, 2001 年生, 硕士。研究方向: 海洋机器人的感知与避障。E-mail: lihaixin@sjtu.edu.cn

乔磊, 男, 1989 年生, 博士, 副教授, 博士生导师。研究方向: 海洋智能机器人和无人系统。E-mail: qiaolei@sjtu.edu.cn

*通信作者: 乔磊

境噪声与复杂散射效应导致声呐图像信噪比显著下降^[6], 现有基于人工特征的图像处理方法在噪声抑制与特征保留的平衡方面面临固有局限; 其次, 实时避障任务对算法的轻量化部署与推理效率提出了严苛要求, 尤其在移动设备资源受限的条件下, 现有基于神经网络的分割模型在计算复杂度与动态避障需求下的实时响应效率之间存在矛盾性冲突; 最后, 现有常用算法大多采用专用框架, 其设计往往过于依赖其他特定场景的假设和固定的数据特征, 尚未充分考虑避障场景中水下复杂结构特性及成像设备的固有参数, 导致避障场景下泛化效果不佳。

因此, 本文研究的核心目标在于针对水下避障场景的复杂性和特殊性, 基于水下结构体特征和成像声呐物理特性, 开发专用适配的感知算法方案, 在满足任务实时性需求的同时, 提高感知模型鲁棒性和可靠性。为此, 本文将提出一种具有轻量化和短时推理特性的声呐图像语义分割算法, 通过引入轻量化深度卷积操作, 显著降低计算复杂度, 同时针对大目标场景的特征改进有效感受野 (effective perceptive field) 机制, 实现基于前视声呐水下实时避障场景的高精度轻量化感知。实验验证部分采用分割效果和避障验证的双重实验结构, 目的在于从感知性能和任务完成效果两个层面系统评估所提出方法的有效性与实用性。通过适应机载场景的感知性能实验支持感知鲁棒性论证, 同时结合 MPC 避障策略的 Gazebo 平台仿真实验验证任务的可行性。

1 基于语义分割网络的感知方法

本文采用语义分割神经网络实现水下避障任务的感知功能。作为常见的像素级分类方法, 语义分割的核心任务是为输入图像每个像素分配对应语义标签 (“障碍物”或“背景”)。语义分割无需严格区分同类实例且能够有效提取可通行区域与障碍物分布特征, 比较适用于水下避障场景。整个感知模块的处理流程是: 基于 U-Net 解码-编码器结构网络对声呐图像进行目标分割, 之后基于推理的障碍物轮廓实施最小外接圆建模计算, 结合声呐成像参数, 经过比例换算和坐标变换最终输出障碍物的空间和尺寸信息。

1.1 U-Net 介绍

U-Net^[7] 凭借其紧凑的拓扑结构与特征复用特性, 已成为医学影像、遥感检测等小样本密集预测任务的基准模型, 许多研究在其框架上进行了改进。网络核心架构采用对称的编码器-解码

器结构, 通过跳跃连接 (skip connection) 实现多尺度特征融合。通过编码器的空间压缩-语义抽象与解码器的分辨率恢复-细节重构, 构建了端到端的像素级预测框架。然而, 受限于推理时间和固定大小的卷积核对感受野的限制, U-Net 对大目标或全局场景的建模能力较差。此外, 跳跃连接直接将未经充分处理的编码器特征传递至解码器, 可能导致噪声对解码结果产生干扰, 从而降低模型的泛化能力以及对水下复杂噪声的鲁棒性。同时, 模型参数和计算复杂度冗余, 限制了其在资源受限移动设备上的实际应用。

1.2 改进后网络结构

为应对实时避障任务下资源受限移动设备算力限制, 本文基于 UNet 语义分割网络结构, 针对避障场景进行针对性改进。通过深度可分离和通道分组循环模块, 替代基准网络 U-Net 的两次常规卷积, 在显著降低计算复杂度和参数量的同时进一步加深网络深度, 平衡参数规模和语义提取能力。同时, 针对水下感知特殊场景, 在解码器与编码器之间的跳跃连接处引入大核可分离注意力模块, 增强对整体目标的感知能力, 确保全局语义一致性, 避免局部特征带来的干扰。

1) 轻量化卷积模块设计。

MobileNet V1^[8] 中首次系统性地将深度可分离卷积作为核心算子, 并证明了其在移动端设备上的高效性。该结果虽会造成轻微精度损失, 但可通过增加网络深度或宽度补偿性能。本文通过将 U-Net 中标准卷积分解为深度卷积 (depthwise convolution) 和逐点卷积 (pointwise convolution) 两个独立操作 (图 1), 显著降低了计算复杂度。

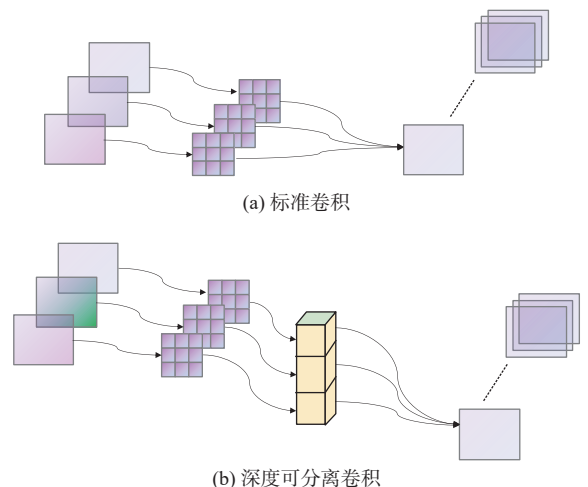


图 1 深度可分离卷积操作

Fig. 1 Depthwise separable convolution operations

通道分组递归卷积 (CSR-conv)^[9] 首次将递归深度模型作为通用卷积层以实现轻量级卷积神经

网络功能。该卷积的核心思想是保持模型性能的同时,通过通道分组和循环结构的权重共享来减少模型参数量,避免了标准卷积因同时建模空间维度与平面通道关系导致的参数量及计算量增加。具体的 CSR-conv 操作如下:首先,将输入通道均匀划分为 T 个组,形成长度为 T 的序列数据。每个时间步处理的通道数为 d/T (d 为输入通道数);之后,将序列数据馈送至递归神经网络中,并随时间输出隐藏状态;最后,将每个时间步的隐藏状态拼接为最终输出。

然而,CSR-conv 原论文^[9]所采用的循环神经网络(vanilla RNN)存在梯度消失/爆炸的局限性,建模长序列数据时对早期时间步信息的记忆能力较弱,难以学习长期依赖特征,同时无法灵活地选择保留和遗忘信息,难以处理更复杂的序列表示。为弥补上述不足,本文通过门控循环单元(GRU)^[10]作为替代,并将跳跃连接^[11]作为可选项进一步增强模型推理能力。通过引入门控机制,提升模型对长期依赖的建模能力。

本文采用的跳跃连接结构计算过程如式(1)所示。其中,concat为拼接操作,所设计的进阶网络整体结构增加了3个附带跳跃连接的门循环单元,将第1层和第2层GRU的输出 $h_1^{(t)}$ 和 $h_2^{(t)}$ 拼接输入到第3层GRU中。通过实验表明:相较于基准模型,在仅增加4.8%推理时间下,分割精度提升了16.96%。

$$\begin{cases} h_1^{(t)} = \text{GRU}_1(x_t, h_1^{(t-1)}) \\ h_2^{(t)} = \text{GRU}_2(h_1^{(t)}, h_2^{(t-1)}) \\ h_3^{(t)} = \text{GRU}_3(\text{concat}(h_2^{(t)}, h_1^{(t)}), h_3^{(t-1)}) \end{cases} \quad (1)$$

组合深度可分离和通道分组递归卷积,其初衷是实现二者性能的针对性互补。其中,深度可分离卷积在轻量化方面优势明显^[8],但会削减跨通道交互能力并可能造成欠拟合风险。分离计算带来的空间维度和通道维度混合特征较差,因此选用通道分组递归卷积中的循环神经网络结构实现多维度融合。同时,通过加深网络应对结构简化造成的性能下降,以平衡计算开销与精度。

2) 大感受野跳跃连接模块。

Attention U-Net^[12]在跳跃连接部分进行改进,首次将基于全局特征向量的注意力门控机制集成到U-Net中,以一定计算开销为代价,有效提升模型的灵敏度和预测精度,适用于图像小且不均匀分布的目标分割任务。然而,该机制其过度强调局部区域特征,不适用于本文的大目标复杂场景。针对本文任务需求,需要重点提高模型的有效感受野,确保全局语义一致性,避免局部特征

干扰。

现有策略通常采用大核卷积增强全局信息感知能力,但其计算量和参数量显著增加,在资源受限的水下环境中难以实现。大核注意力(LKA)^[13]结合了CNN和ViT的优点,可在较小计算开销下获得更大感受野。本文采用基于改进的LSKA模块^[14](结构如图1所示),通过将立体卷积分解成单维度深度卷积、深度膨胀卷积和逐点卷积,在更低计算复杂度和内存占用下,实现与标准大核卷积相当的性能。

1.3 基于语义分割结果的障碍物建模

综上所述,本研究针对水下避障场景,基于前视声呐图像采用轻量化语义分割网络,实现水下障碍物高精度探测和定位。该方法能够准确提取机体前视方向的有效障碍物区域信息,并进一步通过几何特征提取算法获取障碍物在图像坐标系中的几何表征;随后,通过系列坐标变换操作,精确建立图像坐标系与世界坐标系之间的映射关系,最终获取障碍物在世界坐标系下的准确表达。坐标系定义如图2所示。

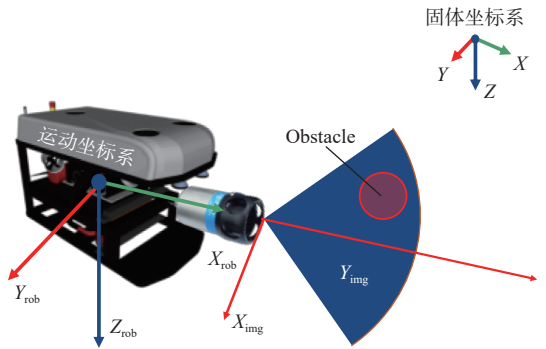


图2 坐标变换

Fig. 2 Coordinate transformations

首先,实现图像坐标系和声呐坐标系之间的尺度变换:

$$\begin{bmatrix} x_{\text{sonar}} \\ y_{\text{sonar}} \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & -\frac{1}{2}W_{\text{img}} \\ 0 & \frac{1}{dy} & -H_{\text{img}} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{\text{img}} \\ y_{\text{img}} \\ 1 \end{bmatrix} \quad (2)$$

式中: $(x_{\text{sonar}}, y_{\text{sonar}})$ 表示障碍物在声呐坐标系中的位置; $(x_{\text{img}}, y_{\text{img}})$ 表示图像坐标系的像素位置; H_{img} 和 W_{img} 分别表示图像的高度和宽度; dx 和 dy 分别表示 X_{img} 和 Y_{img} 方向的每像素实际映射。

障碍物的建模半径通过如下公式转换:

$$r_{\text{obs}} = \frac{R_{\text{max}}}{H_{\text{img}}} \cdot r_{\text{pixel}} \quad (3)$$

式中: r_{obs} 为障碍物的实际空间半径; r_{pixel} 为障碍物在图像中的像素半径; R_{max} 为声纳的最大探测范围。

然后, 实现声纳坐标系和世界坐标系的转换:

$$\begin{bmatrix} x_{\text{obs}} \\ y_{\text{obs}} \\ z_{\text{obs}} \end{bmatrix} = \begin{bmatrix} \cos(\psi) & -\sin(\psi) & 0 \\ \sin(\psi) & \cos(\psi) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{\text{sonar}} + d_{\text{offset}} \\ y_{\text{sonar}} \\ 0 \end{bmatrix} + \begin{bmatrix} x_{\text{rob}} \\ y_{\text{rob}} \\ z_{\text{rob}} \end{bmatrix} \quad (4)$$

式中: $(x_{\text{obs}}, y_{\text{obs}}, z_{\text{obs}})$ 和 $(x_{\text{rob}}, y_{\text{rob}}, z_{\text{rob}})$ 分别表示障碍物和机器人在世界坐标系中的位置; d_{offset} 为声纳中心相对于机器人中心的前向偏移量。

2 基于真实数据的训练对比实验

2.1 数据集和训练细节

本文基于声纳 M750D 采集并标注了一批真实图像数据, 用于模型训练与测试。为增强数据的多样性和鲁棒性, 对原始图像进行了随机水平翻转、旋转、缩放和裁剪等数据增强处理。原始标注数据集包含 1156 张图像, 通过数据增强得到总数据集数量为 6936 张。以 9:1 的比例随机划分数据集。

实验均在单个 NVIDIA GeForce RTX4070 GPU 上完成。训练时图像尺寸均被调整为 512×512 , 采用式(5)所示 Dice 损失函数, 使用 Adam 作为优化器, 初始学习率设置为 0.0001, 并结合余弦退火学习率调度器, 动态调整训练过程。最大迭代次数为 3121, 最小学习速率设定为 0.000001, 总训练轮数 epoch 为 300, 批量大小 (batch size) 为 2。

$$L_{\text{Dice}} = 1 - \frac{2|X \cap Y|}{|X| + |Y|} \quad (5)$$

式中: X, Y 分别表示真实值和预测值。

2.2 实验结果与分析

感知算法的 6 个指标(式(6))用于测量分割性能, 包括联合平均交集 ($mIoU$)、相似性得分 ($Dice$)、精确率 (Pre)、准确性 (Acc)、灵敏度 (Sen)、特异性 (Spe)。其中 $mIoU$ 是预测结果与真实标注间重叠率的统计结果, 是衡量感知算法噪声场景下鲁棒性的最关键指标。另外, 使用 $Params$ 表示参数个数, 单位为百万 (M)。计算复杂度根据浮点运算次数 ($FLOP$) 的数量计算, 模型参数和 $FLOP$ 采用 $1 \times 3 \times 512 \times 512$ 输入大小测量。

$$\left\{ \begin{array}{l} mIoU = \frac{TP}{TP + FP + FN} \\ Dice = \frac{2TP}{2TP + FP + FN} \\ Pre = \frac{TP}{TP + FP} \\ Acc = \frac{TP + TN}{TP + TN + FP + FN} \\ Sen = \frac{TP}{TP + FN} \\ Spe = \frac{TN}{TN + FP} \end{array} \right. \quad (6)$$

式中: TP 和 FP 分别为正类像素预测为正、负的情况; TN 和 FN 分别为负类像素预测为正、负的情况。

通过对真实采集并标注的图像数据集进行训练与测试, 对各算法性能指标进行定量评估, 结果如表 1 所示。通过对比基准 U-Net^[5] 网络以及在跳跃连接部分改进的 Attention U-Net^[12], 可以看出, 相较于基准模型, 所提算法的计算量和参数量分别降低了 69.2% 和 83%, 同时, 推理时间减少了 22.6%, 且在 $mIoU$ 精度上 (即感知精度) 提升了 10.8%。此外, 在引入跳跃连接仅增加 4.8% 推理时间的代价下, 分割精度相较于基准模型提升了 16.96%。这表明, 所提算法在保持高精度的同时, 显著优化了计算效率, 更加符合机载场景下资源约束和实时性需求。图 3 展示了网络训练过程中验证集评估的 $mIoU$ 曲线, 由图可见, 所提算法在收敛效果和性能提升方面均优于基准模型。

图 4 展示了在实际数据上测试的分割效果。其中, 图 4(a) 为采集图像的标注结果, 高亮区域 (蓝色) 为人工标注的障碍物 (墙壁), 图 4(b) 和图 4(c) 的高亮区域 (红色) 为模型推理出的障碍物标注。通过与真实标签的对比可以看出, U-Net 分割模型对散斑噪声、运动模糊的鲁棒性较差, 导致其输出呈现显著的语义区域不连续性与伪影敏感特性, 严重劣化了感知结果的拓扑一致性; 本文所提算法推理真实图像的抗噪声干扰能力更强, 分割目标更为完整且方便后续目标整合, 实现了更清晰准确的分割效果。这是因为在网络设计上主动加深了模型深度, 并引入多尺度上下文融合策略, 显著扩大了模型感受野。这一策略有助于模型更全面地理解图像全局信息, 从而减少对局部噪声的依赖。

3 基于 Gazebo 仿真的避障场景验证

为验证所提感知算法在实际避障场景中的应用性能, 本章将感知模块输出的障碍物信息整合

表 1 感知算法性能指标比较

Table 1 Comparison of perception algorithm performance metrics

性能指标/感知算法	<i>mIoU</i> %	<i>Dice</i> %	<i>Pre</i> %	<i>Acc</i> %	<i>Sen</i> %	<i>Spe</i> %	<i>FLOP</i> /G	<i>Params</i> /M	推理时间/s
U-Net	72.30	82.75	83.40	90.71	94.90	69.39	54.99	7.766	0.062
Attention U-Net	78.95	87.71	84.65	92.50	92.66	91.68	66.97	8.726	0.061
卷积轻量化设计 附带跳跃连接改进(a)	80.08	88.41	87.37	93.45	95.35	83.78	16.94	1.32	0.048
(a)改进的基础上, 部分GRU附带跳跃连接	84.56	91.33	90.17	95.10	96.31	88.94	20.26	1.40	0.065

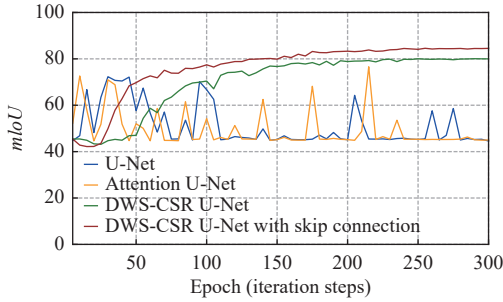
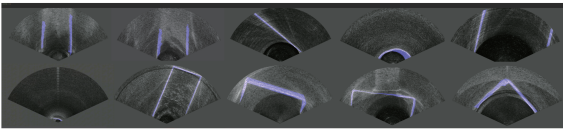
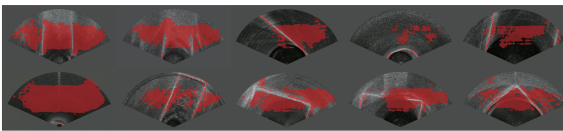


图 3 模型训练效果对比

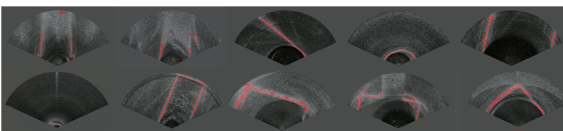
Fig. 3 Model training performance comparison



(a) 真实标签



(b) U-Net



(c) 改进后算法

图 4 实际数据上测试的分割效果

Fig. 4 Segmentation results on real-world data

到模型预测控制模块中作为避障约束,从而实现航行器的安全航行。从感知下游的任务需求出发,通过仿真实验部署感知算法,验证其可行性。

3.1 避障流程

基于前文感知模块所获取的障碍物位置和尺寸信息设计避障实验验证。在避障运行过程中,感知结果深度嵌入到 MPC(模型预测控制)求解器中,构成一种强耦合的协同机制,即将障碍物作为约束条件引入到 MPC 框架中,将避障控制问题转化为优化求解问题,进而实现精准跟踪与避障功能。二者协同机制如图 5 所示。

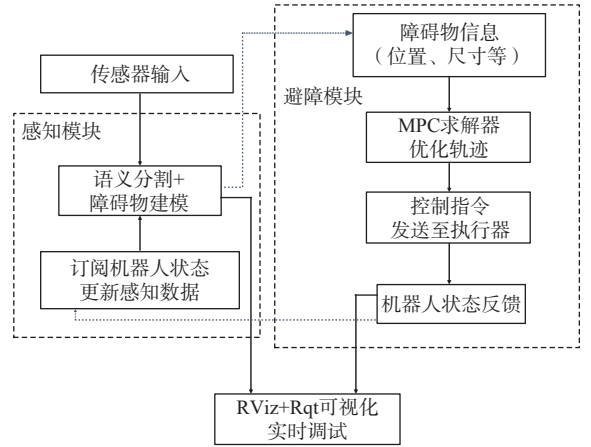


图 5 精准跟踪与避障控制协同机制

Fig. 5 Collaborative mechanism of precise tracking and obstacle avoidance

3.2 模型预测控制方法构建

本文采用非线性模型预测控制(NMPC)方法实现水下机器人避障。该方法将障碍物约束显式引入优化问题,通过求解在线优化问题实现对障碍物的实时规避。优化问题需分别满足系统状态、控制输入的动力学约束以及避障的安全距离约束。其中,实时传入的障碍物位置和边界信息作为模型预测控制的硬约束条件,确保规划轨迹与障碍物无碰撞。

构建如下有限时域优化问题:

$$\begin{aligned}
 \min_{\tau_{0:N-1}} & \left\{ \sum_{k=0}^{N-1} (\| \mathbf{x}_k - \mathbf{x}_{\text{ref},k} \|_{\mathbf{W}_x}^2 + \| \tau_k \|_{\mathbf{W}_u}^2) + \| \mathbf{x}_N - \mathbf{x}_{\text{ref},N} \|_{\mathbf{W}_x}^2 \right\} \\
 \text{s.t.} & \begin{cases} \mathbf{x}_{k+1} = f_d(\mathbf{x}_k, \tau_k) \\ \tau_{\min} \leq \tau_k \leq \tau_{\max} \\ \| \mathbf{p}_{\text{rob}}(\mathbf{x}_k) - \mathbf{p}_{\text{obs}} \|_2 \geq (r_{\text{obs}} + r_{\text{rob}}) + d_{\text{safe}} \end{cases} \quad (7)
 \end{aligned}$$

式中: \mathbf{W}_x 和 \mathbf{W}_u 为对称正定权重矩阵,分别用于调节状态跟踪精度与控制能耗之间的权衡; \mathbf{x}_k , $\mathbf{x}_{\text{ref},k}$ 和 τ_k 分别表示第 k 步 ($k = 1, 2, \dots, N-1$) 的状态向量、参考状态向量以及控制输入; \mathbf{x}_N 和 $\mathbf{x}_{\text{ref},N}$ 分别表示预测时域末端状态和参考状态。将前文感知算法获取的障碍物空间信息整合到模型预测控制框架中: d_{safe} 为预设安全裕度; \mathbf{p}_{rob} 为机器人中

心位置; p_{obs} 为障碍物中心位置; r_{obs} 和 r_{rob} 分别为障碍物建模和机器人机体的半径。其中, 状态约束 $x_{k+1} = f_d(x_k, \tau_k)$ 为状态方程式(8)的离散化, 本文采用的水下机器人运动学和动力学方程分别为

$$\begin{cases} \dot{\eta} = J(\eta)v \\ M\dot{v} + C(v)v + D(v)v + G = \tau \end{cases} \quad (8)$$

式中: $x = [x, y, z, \phi, \theta, \psi, u, v, w, p, q, r]^T$, 为状态向量; $\eta = [x, y, z, \phi, \theta, \psi]^T$, 为位置坐标以及横摇角、纵摇角和艏向角; $v = [u, v, w, p, q, r]^T$, 为对应的线速度和角速度; $\tau = [u_1, u_2, u_3, u_4]^T \in \mathbf{R}^4$, 为控制向量, 其中 $u_i (i = 1, 2, 3, 4)$ 分别表示 x, y, z 以及 ψ 自由度上的控制输入; 雅可比矩阵 $J(\eta)$ 实现从运动坐标系下速度到固定坐标系下位姿变化率的线性变换; M 为惯性矩阵(包含附加质量); $C(v)$ 为科氏力矩阵; $D(v)$ 为阻尼矩阵; G 为重力 and 浮力向量。机器人模型由 12 维状态向量和 4 维控制向量来描述。

3.3 感知算法避障效果验证实验

本文中算法验证在开源仿真平台 UUV Simulator^[15] 上实现, 其传感器仿真模块支持包括成像声纳、深度传感器、IMU(惯性测量单元)等多种水下感知定位设备, 能够生成与实际硬件相仿的感知数据。UUV Simulator 基于 Gazebo 的三维可视化环境(图 6), 可直观呈现水下机器人运动轨迹、姿态变化及环境交互, 易于显式结果分析与调试优化。此外, 在仿真运行时将文献 [16] 中收集的真实水池环境干扰引入控制器。图 7 所示为 Gazebo 水下场景仿真平台。其中, 机器人几何模型尺寸为 $0.46 \text{ m} \times 0.34 \text{ m} \times 0.26 \text{ m}$ 。在此平台中构建一个具有 512 个波束成像声纳传感器, 其带宽为 2.95 kHz, 探测距离为 25 m, 水平和垂直探测角分别为 45° 和 20° 。声源级为 220 dB re μPa , 图像分辨率为 997×512 像素。通过物理驱动的点散射模型和相干成像系统建模, 实现了相干斑点噪声以及水下混响等相关高保真设计^[17]。噪声模拟效果见图 6。

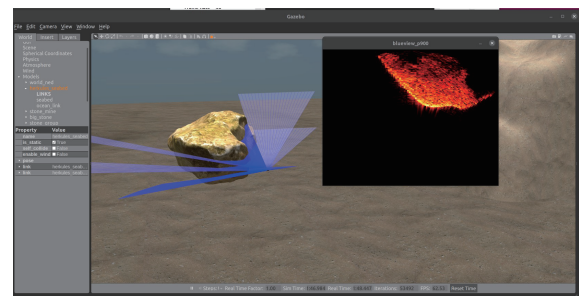
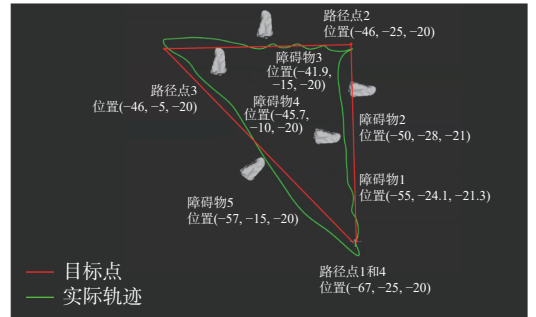


图 6 Gazebo 仿真环境

Fig. 6 Gazebo simulation environment

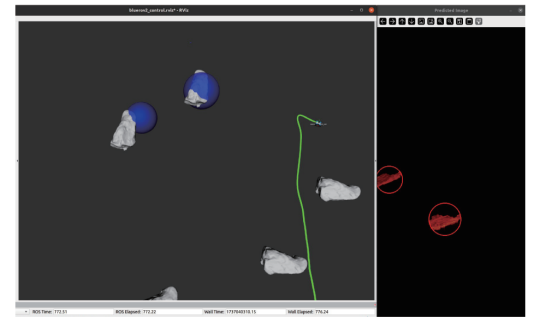
如图 7(a) 所示, 在 Gazebo 仿真环境中布置了静态避障测试场景。实验任务要求机器人依次到达 3 个预设路径点, 并在相邻路径点连线区域设置礁石障碍物(如图 6 所示)。实验中, 通过 ROS(机器人操作系统)实时发布 MarkerArray 类型消息主题(可视化效果见图 7(b)~图 7(d)), 在可视化工具 RViz 中同步呈现算法的感知结果(蓝色球体)。可见, 通过声纳生成的障碍物空间建模能



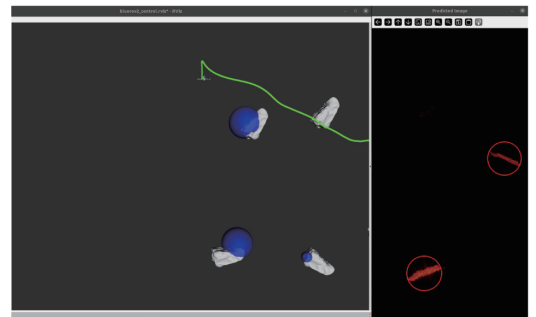
(a) 完整避障路径



(b) 第 1 个路径点



(c) 第 2 个路径点



(d) 第 3 个路径点

图 7 避障场景下障碍物建模可视化

Fig. 7 Visualization of obstacle modeling in obstacle avoidance scenarios

较精准地覆盖有限视域下障碍物空间分布,感知模块可为模型预测控制提供可靠的障碍物约束。

图7(a)展示了水下机器人在实际避障任务中的运行轨迹。仿真平台实验表明,基于感知算法的避障策略能够高质量完成基于成像声呐传感器的水下避障任务。同时,控制器的平均求解时间为0.0319 s,满足20 Hz控制周期需求,保证了避障系统的实时性。

4 结 语

本文面向水下机器人避障任务需求和特性,针对性地设计了改进U-Net解码器-编码器结构的轻量化、高精度实时语义分割网络。在设计过程中充分考虑了移动设备的计算性能限制与任务的实时性要求,通过轻量化卷积等效标准卷积,大幅降低了模型的计算量。此外,通过引入能够综合全局上下文信息的跳跃连接模块,有效提升了模型在水下大目标场景中的适应能力和感知精度。

基于真实采集标注的6936张水下声呐图像数据集训练对比实验结果表明,在水下避障场景下,相较于基准模型,本文所提感知算法的适应性和鲁棒性更优,计算量和参数量分别降低了69.2%和83%。通过适应避障场景的网络结构调整,感知精度提升了10.8%。此外,基于该感知算法的避障策略的高保真仿真实验,验证了算法的实时性和有效性。上述实验结果证明了本文方法的可行性及其实用价值,可为复杂环境下的自主避障问题提供理论支撑和工程实现方案。然而,当前研究实验部分仅通过ROS平台验证了各算法模块的通信和功能,未来计划将本文方法移植到具体航行器上,对水下机器人自主感知避障的机载资源和任务时效进行更全面的评估和研究。

参考文献:

- [1] 赵蕊,许建.自主式水下航行器水下回收融合引导技术方案及算法[J].中国舰船研究,2022,17(1):212-220.
ZHAO R, XU J. Fusion guiding technology solution and algorithm for underwater docking of autonomous underwater vehicles[J]. Chinese Journal of Ship Research, 2022, 17(1): 212-220 (in both Chinese and English).
- [2] 李祥弘,于林,苏婧,等.有限视域与多元约束下大型AUV三维高速避障方法[J].中国舰船研究,2025,20(1):181-190.
LI X H, YU L, SU J, et al. High-speed 3D obstacle avoidance method for large-scale AUVs under limited field of view and multiple constraints[J]. Chinese Journal of Ship Research, 2025, 20(1): 181-190 (in Chinese).
- [3] 何义才,赵建虎,张红梅,等.联合NSCT与多重分形的高噪声侧扫声呐图像分割[J].测绘学报,2020,49(2):162-170.
HE Y C, ZHAO J H, ZHANG H M, et al. Segmentation of polluted SSS image by combining NSCT and multifractal[J]. Acta Geodaetica et Cartographica Sinica, 2020, 49(2): 162-170 (in Chinese).
- [4] 张家敏,曾庆军,朱志宇,等.自主式水下机器人控制系统及声呐目标识别[J].中国舰船研究,2018,13(6):94-100.
ZHANG J M, ZENG Q J, ZHU Z Y, et al. AUV control system and sonar target identification[J]. Chinese Journal of Ship Research, 2018, 13(6): 94-100 (in Chinese).
- [5] CHEN J, SUMMERS J E. Deep convolutional neural networks for semi-supervised learning from synthetic aperture sonar (SAS) images[J]. Proceeding of Meetings on Acoustics, 2017, 30(1): 055018.
- [6] 张家发,周兴华,赵洪臣,等.基于弱纹理块的噪声估计方法在侧扫声呐图像去噪中的应用[J].海岸工程,2018,37(1):17-25.
ZHANG J F, ZHOU X H, ZHAO H C, et al. Application of noise level estimation algorithm based on weak textured patches in side-scan sonar image denoising[J]. Coastal Engineering, 2018, 37(1): 17-25 (in Chinese).
- [7] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation[C]//18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich: Springer, 2015: 234-241. DOI: 10.1007/978-3-319-24574-4_28.
- [8] HOWARD A G, ZHU M L, CHEN B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications [EB/OL]. (2017-04-17)[2025-4-23]. <https://arxiv.org/abs/1704.04861>.
- [9] WU G J, ZHANG X, ZHANG Z M, et al. Learning lightweight neural networks via channel-split recurrent convolution[C]//2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa: IEEE, 2023: 3847-3857. DOI: 10.1109/WACV56688.2023.00385.
- [10] BALLAKUR A A, ARYA A. Empirical evaluation of gated recurrent neural network architectures in aviation delay prediction[C]//2020 5th International Conference on Computing, Communication and Security (ICCCS). Patna: IEEE, 2020. DOI: 10.1109/ICCCS49678.2020.9276855.
- [11] ULLAH A, MUHAMMAD K, DING W P, et al. Efficient activity recognition using lightweight CNN and DS-GRU network for surveillance applications[J]. Applied Soft Computing, 2021, 103: 107102.
- [12] OKTAY O, SCHLEMPER J, LE FOLGOC L, et al. Attention U-Net: Learning where to look for the pancreas [EB/OL]. (2018-05-20)[2025-4-23]. <https://arxiv.org/abs/1804.03999>.
- [13] GUO M H, LU C Z, LIU Z N, et al. Visual attention net-

- work[J]. *Computational Visual Media*, 2023, 9(4): 733–752.
- [14] LAU K W, PO L M, REHMAN Y A U. Large separable kernel attention: rethinking the large kernel attention design in CNN[J]. *Expert Systems with Applications*, 2024, 236: 121352.
- [15] MANHÃES M M M, SCHERER S A, VOSS M, et al. UUV Simulator: a gazebo-based package for underwater intervention and multi-robot simulation[C]//OCEANS 2016 MTS/IEEE Monterey. Monterey: IEEE, 2016: 1–8. DOI: 10.1109/OCEANS.2016.7761080.
- [16] 陈盼召, 郭凯, 葛彤, 等. 外部扰动下无人水下航行器鲁棒快速轨迹跟踪控制 [J]. *中国舰船研究*, 2025, 20(6): 294–303.
- CHEN P Z, GUO K, GE T, et al. Robust and fast trajectory tracking control of UUVs with external disturbances [J]. *Chinese Journal of Ship Research*, 2025, 20(6): 294–303 (in Chinese).
- [17] CHOI W S, OLSON D R, DAVIS D, et al. Physics-based modelling and simulation of multibeam echosounder perception for autonomous underwater manipulation[J]. *Frontiers in Robotics and AI*, 2021, 8: 706646.

Lightweight real-time perception method based on imaging sonar for obstacle avoidance scenarios

LI Haixin, YU Lin, LI Xianghong, QIAO Lei^{*}

School of Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

Abstract: [**Objectives**] To address the challenges posed by high-intensity noise and the structural characteristics of large obstacle targets in underwater sonar imaging, as well as the stringent requirements for lightweight deployment and high inference efficiency of perception algorithms in real-time underwater obstacle avoidance tasks, a semantic segmentation algorithm for sonar images with low computational cost and short inference time is proposed. The method aims to resolve the trade-off between the computational complexity of perception algorithms and the real-time response requirements in obstacle avoidance applications.

[**Methods**] Based on an encoder-decoder network architecture, lightweight convolution operations were introduced to significantly reduce computational complexity. In addition, a large-kernel separable attention mechanism was incorporated into the skip connections to enhance feature fusion for obstacle avoidance scenarios. A dataset of 6936 sonar images collected and manually annotated from real environments was used for training and comparative experiments. Furthermore, the obstacle avoidance strategy based on the proposed perception algorithm was validated on the Gazebo simulation platform. [**Results**] The improved algorithm specifically enhances the segmentation accuracy of large targets. Compared with the baseline model, the FLOP and the number of parameters are reduced by 69.2% and 83%, respectively. At the same time, the inference time is shortened by 22.6%, while perception accuracy improves by 10.8%. In addition, simulation experiments verify the effectiveness of the perception algorithm during the obstacle avoidance process, demonstrating that it fully satisfies the requirements of real-time perception tasks in underwater obstacle avoidance scenarios based on forward-looking sonar. [**Conclusions**] The proposed sonar-image-based perception algorithm can effectively meet the obstacle avoidance requirements of unmanned underwater vehicles in onboard operating scenarios and shows promising potential for engineering applications.

Key words: unmanned underwater vehicle; obstacle avoidance; collision avoidance; navigation; imaging sonar; sensory perception; image recognition; semantic segmentation; lightweight real-time perception