

文章编号 1004-924X(2025)14-2303-15

联合双任务和双重注意力机制的 遥感影像舰船提取方法

邹毓杰, 林雨准, 王淑香, 左溪冰, 刘 潇, 芮 杰^{*}
(信息工程大学 地理空间信息学院, 河南 郑州 450001)

摘要:针对遥感影像像素级舰船提取任务中存在的基准不足、解译精度较差等问题,本文从数据集和网络结构两个层面展开研究,构建了一个数量充足、类型丰富、分布广泛的舰船语义分割数据集(Ship Semantic Segmentation Dataset, SSSD),并提出了一种联合双任务和双重注意力机制的舰船提取方法。在数据集层面,合理选取全球范围内的港口区域,在 HRSC2016 数据集的基础上完成 SSSD 的构建,最终形成的数据集包含训练影像 3 760 张,测试影像 923 张。在网络结构层面,通过融合多头自注意力和通道注意力,设计双重注意力机制,有效提升了分割精度。此外,引入了轮廓监督分支作为辅任务,联合语义分割形成双任务网络框架,增强了舰船与背景的切割效果。实验结果表明,本文方法在 IoU 上比 U-Net, ASPUNet, DeepLabV3+ 和 MANet 分别提升 11.44%, 17.84%, 12.25% 和 1.64%;双重注意力和辅任务的设计在 IoU 上分别提升了 2.02% 和 1.36%;同时,在泛化性实验中本文方法展现了更强的适应性和鲁棒性。本文构建的 SSSD 可为舰船像素级解译提供可靠的数据基准,本文提出的舰船提取方法相较于主流方法具有更优的性能以及更强的稳定性和泛化能力。

关键词:遥感影像;舰船提取;数据集;注意力机制;双任务

中图分类号: TP391.41 **文献标识码:** A

doi: 10.37188/OPE.20253314.2303

CSTR: 32169.14.OPE.20253314.2303

Joint dual-task and dual-attention mechanism for ship extraction from remote sensing images

ZOU Yujie, LIN Yuzhun, WANG Shuxiang, ZUO Xibing, LIU Xiao, RUI Jie^{*}

(information engineer university, Zhengzhou 450001, China)

^{*} Corresponding author, E-mail: zzwjllhtt@scut.edu.cn

Abstract: Aiming at the problems of insufficient data baseline and poor decoding accuracy in the pixel-level ship-extraction task of remote-sensing images, this paper conducted research from the two levels of dataset and network structure, constructed a Ship Semantic Segmentation Dataset (SSSD) that was sufficient in quantity, rich in type, and wide in distribution, and proposed a ship-extraction method with joint dual-task and dual-attention mechanisms. At the dataset level, port areas worldwide were reasonably selected, the construction of SSSD was completed on the basis of the HRSC2016 dataset, and the final dataset contained 3 760 training images and 923 test images. At the network-structure level, segmentation accuracy was effectively improved by fusing multi-head self-attention and channel attention and by designing the du-

收稿日期: 2025-02-16; 修订日期: 2025-04-18.

基金项目: 国家自然科学基金项目 (No. 42301464)

al-attention mechanism. In addition, contour-supervised branching was introduced as a secondary task, and the dual-task network framework was formed by jointly performing semantic segmentation, which enhanced the separation between the ship and the background. The experimental results show that this paper's method improves 11.44%, 17.84%, 12.25%, and 1.64% over U-Net, ASPPUNet, DeepLabV3+, and MANet, respectively, on IoU; the design of dual-attention and auxiliary tasks improves 2.02% and 1.36%, respectively, on IoU; meanwhile, this paper's method exhibits, in generalization experiments, a stronger adaptability and robustness. The SSSD constructed in this paper can provide a reliable data benchmark for ship pixel-level decoding, and the ship extraction method proposed in this paper has better performance as well as stronger stability and generalization capability compared with mainstream methods.

Key words: remote sensing image; ship extraction; data set; attention mechanism; dual task

1 引言

我国拥有广阔的海洋领土和丰富的海洋资源,舰船作为维护海洋权益的核心作战力量和海洋侦察的主要目标,历来备受关注。舰船检测技术为战略决策提供了关键信息,有效缩短了决策周期,是大规模运行环境下实现实时战场态势感知的重要基础^[1-2]。因此及时准确地掌握舰船数量、位置和类型等信息,对于获取海域空间信息和海上机动力量具有重要意义。

近年来,随着深度学习技术的快速发展,基于深度学习的遥感影像舰船提取成为了当下研究的热点。卷积神经网络(Convolutional Neural Networks, CNN^[3])可避免传统算法中人工提取特征的不足,模型能学习更深层次、更深广度、更本质的特征^[4]。全卷积网络(Fully Convolutional Networks, FCN^[5])作为首个将深度学习应用到图像分割中的方法,在CNN的基础上,将CNN的全连接层替换为卷积层,并利用反卷积进行特征上采样,得到端对端的特征输出。然而,FCN的网络结构相对简单,对深层特征的提取能力不足。

为实现深层特征的提取,基于编码器-解码器结构的分割方法开始涌现,代表性网络有U-Net^[6], ASPPUNet^[7], DeepLab 系列网络^[8-11]和MANet^[12]等。然而,这些方法在解译复杂背景的高分辨率舰船影像时,往往忽视了高分辨率影像信息量多、目标小、边界模糊等特点,导致分割效果不理想、计算效率低等问题。为此,后续研究者结合特定理论或算法提出了改进方法。DeepLabV3+^[9]引入编码器和解码器,能更好地获取

影像空间信息,细节特征更丰富,使得舰船的边界提取更清晰,颗粒度更细。Li^[13]提出GAM-EDNet,其全局关注的新型编码器-解码器框架能提高对小目标的语义分割性能。Liu^[14]等人提出PLFCN,该网络通过引入深度金字塔监督来探索多尺度空间上下文信息,Chen^[15]等人提出EFCNet,他们设计了一个混合模型基本卷积块,以获得更大的接受场。尽管上述算法在目标检测上取得了较好的效果,然而由于其仅采用单分支网络特征提取架构,一定程度上限制了算法检测性能的进一步提升。

注意力机制是深度学习算法中一种常用技术^[16-18]。Su^[19]等人通过使用通道注意力机制改进了U-Net架构。张吉友^[20]等人针对语义分割的编码器加入了注意力机制,取得了不错的效果。Li^[21]等人提出的MSCSANet结合自注意力和CNN机制的优点来提高分割质量,Fan^[22]等人融合了通道和空间注意力,提高了河流分割的准确性。自注意力以其全局捕捉的

能力,在影像分割中展现出强大的应用潜力。然而,研究人员只是简单地添加注意力结果,如何识别复杂背景的语义分割是一个需要不断解决的问题。因此,有必要设计有效的融合模型来满足更高精度要求。

数据集方面,当前用于舰船检测识别的数据集有HRSC2016^[23], FAIRIM^[24]等。其中,HRSC2016数据集包含616幅高分辨率舰船影像,集中分布在6个重要港口区域涵盖多种舰船类型,但场景分布较为单一。FAIRIM数据集涵盖了近岸港口和近海地区等场景,船舶类型多样,但其舰船样本数量偏少,不能满足当前遥感

影像舰船提取的数据集要求。

基于深度学习的目标检测技术具有高精度、高效率、广泛适用性的特点,并在遥感图像舰船检测任务中取得了一定成效。然而,当前研究在实际应用中仍面临两个方面的问题。第一是数据基准不足的问题。当前适用于遥感影像舰船提取的数据集极少,且已有数据集受限于样本数量偏少、场景类别不全、区域分布过密等特点,不足以代表实际场景。第二是解译精度较差的问题,在高分辨率遥感影像中,受云雾遮挡、地物背景复杂、舰船类型多样等现实条件干扰,舰船提取存在较为明显的漏检、误检、边缘模糊等现象,进而导致解译精度不理想。

2 舰船语义分割数据集

基于遥感影像的地物目标智能提取需要大量的样本对网络模型进行训练,样本质量的高低在一定程度上会直接影响整个网络模型的性能。因此,建立一个以高分辨率遥感影像为基础,且标注准确、舰船边界清晰度高、场景多样性强的数据集显得尤为重要。

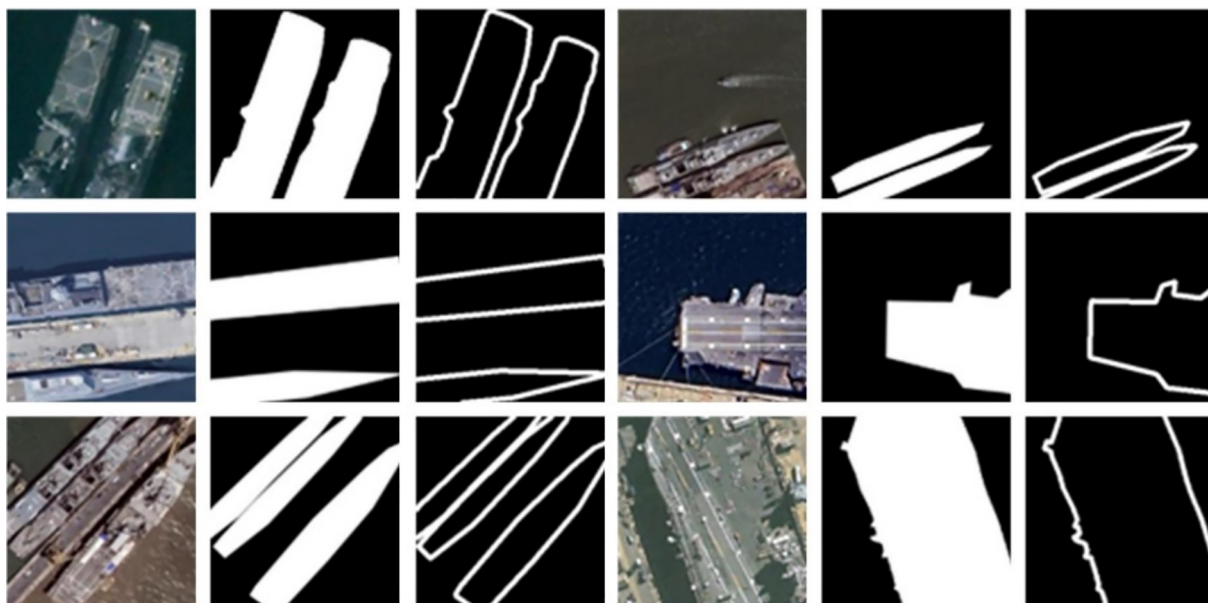
本文以具有像素级标注的 HRSC2016 数据集为基础,在全球范围(横须贺、佐世保、舟山、布雷默顿、圣迭戈、布雷斯特、土伦、诺福克、孟买、库塔巴尔)内选取典型港口区域,以人工目视判读的方式,根据舰船在影像上的色彩色调特征、

形状大小特征、纹理活动特征识别舰船,再结合人工标注最终形成数量充足、分布合理、场景多样的舰船语义分割数据集(Ship Semantic Segmentation Dataset, SSSD)。

2.1 数据预处理

针对 HRSC2016 数据集,本文将原有的多类别像素级标签后处理为二值标签。具体而言,通过对 HRSC2016 数据集中的标签影像进行遍历,将代表舰船标签的像素点灰度值统一设置为 255,同时将其余像素点灰度值设置为 0,从而得到符合要求的像素级二值标签数据。

针对自制数据集,本文首先依托水经注软件获取全球各个港口区域的遥感影像数据,然后进行人工标注,从而获取舰船的像素级二值标签数据。此外,舰船轮廓标签对于提高地物解译的准确性和效率至关重要。它不仅有助于精确识别和分类舰船,而且还能增强模型在学习过程中聚焦于舰船的关键特征,因此本文引入舰船的轮廓标签作为新的标签数据。先利用 Canny 算子从上述二值标签数据中提取边缘并查找轮廓线。同时考虑到网络预测值的聚集特性,过窄的轮廓线不足以提供足够的监督信息,导致损失值偏大,进而影响梯度的反向传播和模型的优化效果,导致模型训练效果较差。鉴于此,本文在原始轮廓标签的基础上绘制宽度为 5 pixels 的轮廓线,以生成舰船的二值轮廓标签。部分数据集如图 1 所示。



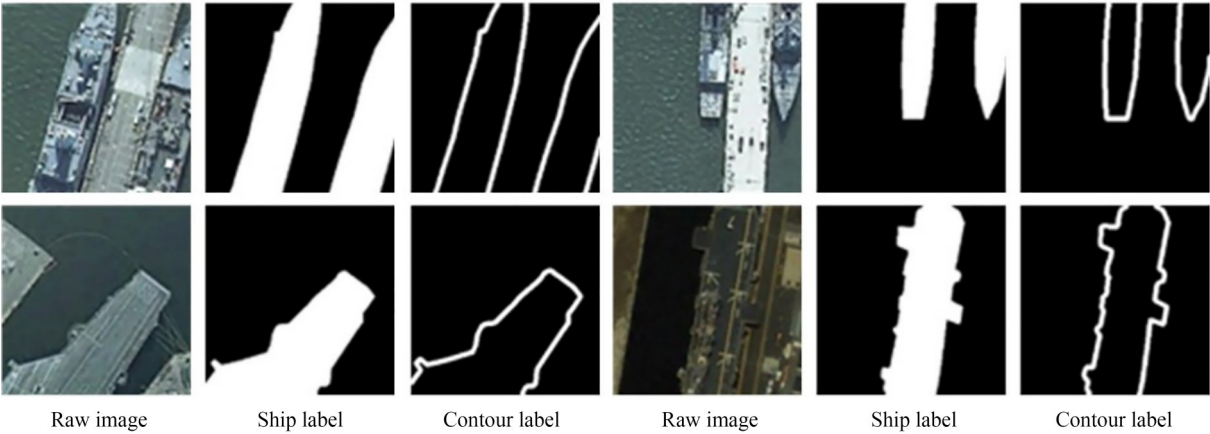


图1 数据集示例影像

Fig. 1 Sample dataset image

由于HRSC2016数据集和自制数据集其各自的舰船影像大小差异较大,不利于直接输入计算机进行训练。因此,本文选择对两个数据集的影像数据进行切片处理,将影像统一裁切成 $256\text{ pixels}\times 256\text{ pixels}$ 大小。同时,为增强数据在大场景中的应用效果,本文在切片过程中设计了10%的重叠度(上下、左右重叠度均为10%),对于影像右边缘和下边缘部分小于 $256\text{ pixels}\times 256\text{ pixels}$ 的影像,本文直接予以舍弃。针对舰船标签和轮廓标签,本文采用与舰船影像相同的切割方式,最终得到完整的数据集。

2.2 数据集分析

2.2.1 分布背景分析

舰船在空间上的分布情况对于海洋监测和安全保障具有重要意义。根据舰船所处的背景,将其分成离岸、靠岸和陆地三种类型。离岸舰船指周边无地面建筑物的舰船,靠岸舰船指停靠在各类码头的舰船,在舰船的某一方向毗邻码头建筑物,而陆地舰船指位于干船坞内,周边无水域环境,全是陆地建筑物的舰船。三种类型的舰船以靠岸舰船居多,具体而言,三种类型的舰船数量如表1所示,离岸舰船共有67艘,靠岸舰船共有1 734艘,陆地上舰船共有27艘。

不同背景其提取难易程度不同,离岸舰船背

景单一,对模型提取舰船特征影响较小,便于提取,而陆地舰船背景复杂,对舰船提取干扰性强。这一分布情况既符合舰船大量时间都在港口停泊的现实情况,又体现了数据集包含多样的应用场景。示例影像如图2所示。

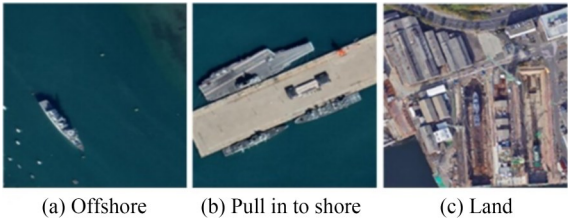


图2 不同舰船背景示例影像

Fig. 2 Sample images of different ship backgrounds

2.2.2 几何结构分析

通过对舰船空间特征的量化分析,可以更好地理解数据集中舰船的分布情况。能够确保模型在训练过程中接触到各种不同的场景和条件,从而使模型在面对未知数据时能够更好地进行预测和决策。长宽比揭示了舰船的几何形状和结构特点,图3为SSSD中部分不同长宽比的示例影像。可以看到舰船在长宽比上差异很大。此外,SSSD中存在大量长宽比较大的实例。其影像大小均为 $256\text{ pixels}\times 256\text{ pixels}$,分辨率均为 0.6 m/pixel 。通过图3可以清晰地看到不同类型舰船的长宽比往往存在差异,尤其是部分护卫舰的形状接近条状地物目标,这对卷积等操作中的关注空间提出了要求。

表1 舰船分布类型及数量

Tab. 1 Ship distribution types and numbers

Ship type	Offshore	Docked	Land-based
Number of ships	67	1 734	27

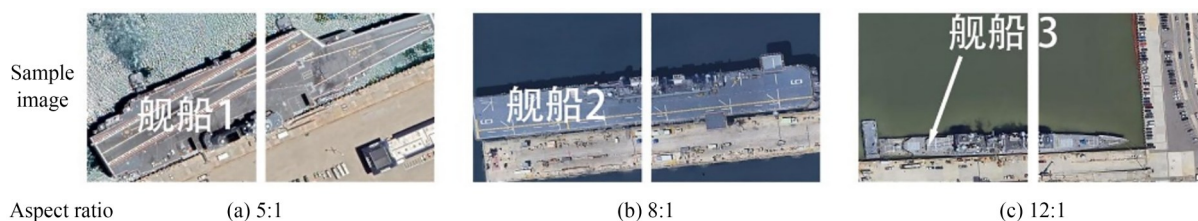


图3 不同长宽比舰船示例影像

Fig. 3 Sample images of ships with different aspect ratios

通过对数据集中舰船空间分布和长宽比特征的分析,SSSD在舰船数据组成上具有多样性和代表性,使得在SSSD上训练的模型更能适应各种复杂和极端的情况,从而在实际应用中表现出更强的鲁棒性和广泛的适应性。

3 本文方法

3.1 网络架构

本文提出的双任务和双重注意力的遥感影像舰船提取模型结构如图4所示,包含双任务网络模块和双重注意力模块。双任务网络模块包含主辅两个分支,主任务分支以舰船标签作为监督数据。舰船轮廓可以在空间上反映舰船的整体结构,对提升舰船提取的完整性有直接促进作用,故引入辅任务分支,用以提升舰船提取结果的完整性。辅任务分支输入数据来自编码层,输出数据为轮廓标签(非最终结果)。注意力模块能有效地提升网络模型对舰船特征的关注意度,从而提升舰船提取精度,而不同注意力模块作用效果不同,考虑到通道注意力模块能增强舰船目标的特征表示,同时抑制背景噪声;多头自注意

力模块能并行处理和聚焦舰船目标来增强网络模型对舰船目标的识别能力。两者结合使用可以在复杂场景下更有效地提取舰船目标。双重注意力模块包含多头自注意力机制和通道注意力机制,作用于主任务分支。

3.2 双任务网络模块

双任务网络模块详细架构如图5所示,主任务是舰船提取,基本单元由编码残差模块、双重注意力和解码卷积模块构成,输入数据是遥感影像,输出数据是舰船标签;辅任务是轮廓检测,输入数据来自主任务编码层,输出数据是轮廓标签。

具体而言,主任务中的编码层采用预先训练的ResNet34^[25]作为主干网络,共设置5个下采样层。

通过上述操作后将原始遥感影像的通道数从3提升到512,尺寸缩小至8 pixels×8 pixels,从而充分提取影像的语义信息。解码层部分使用跳跃相加的方式关联编码层(双重注意力处理后)和解码层,避免信息的过多丢失。最后阶段则是利用上采样、卷积和Sigmoid激活等操作获得与原始影像尺寸一致的舰船提取结果。

为提升舰船分割结果的精细化程度,在主任务网络的基础上引入轮廓监督分支(辅任务)。辅任务与主任务共享编码层,在解码层部分进行跳跃连接、转置卷积等操作,输出轮廓标签数据。

该分支进一步利用舰船的形状、边缘等辅助信息,以加强舰船与背景的区分程度。具体结构如图5所示,为强化其对主任务网络的监督作用,轮廓监督分支与主任务网络共享编码器。在解码器部分,以编码器下采样层(第2,3,4个下采样层)的输出特征为输入,顺序执行多层级的跳跃连接与解码操作,生成尺寸为64×64×64的解码特征图。在预测部分,通过转置卷积将特征图上

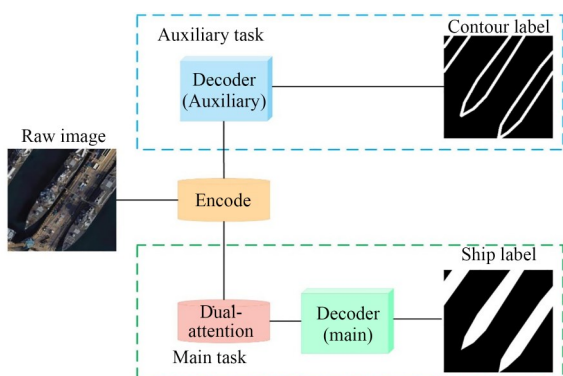


图4 模型框架流程图

Fig. 4 Model frame flow chart

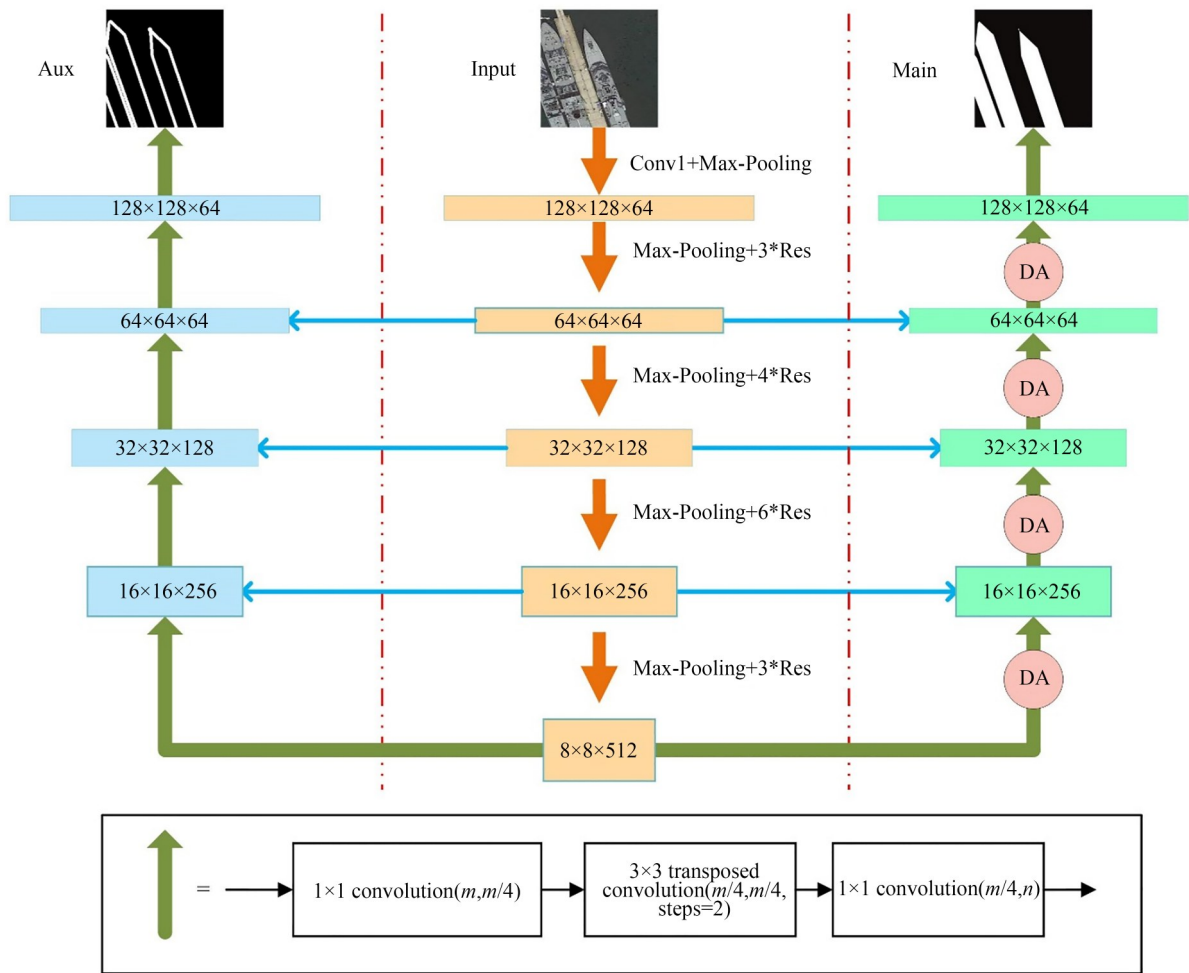


图5 模型框架详细流程图

Fig. 5 Model framework detailed flow chart

采样为 $256 \times 256 \times 32$ 的特征图,并利用 3×3 卷积和 Sigmoid 激活函数获取轮廓预测结果。轮廓预测结果与原始输入影像的尺寸保持一致,并以舰船轮廓的真实标签作为监督信号。

3.3 双重注意力模块

为了加强网络在深层信息提取中关注重点和多维度提取特征的能力,本文引入双重注意力机制,其不仅同时考虑序列内部的依赖关系以及不同通道之间的相关性,还解决了单个注意力机制模型语义信息提取不全的问题,对提高舰船分割精度有极大的促进作用。两种机制的有效结合是两种维度的有效融合,在深层信息提取中,可有效维持模型的关注重心,在本任务中即更加精准地提取关键特征,提高了分割任务的准确性和鲁棒性。

对输入数据分别进行多头自注意力机制和

通道注意力机制,得到两个输出数据,再对两个输出数据进行相加处理,输出最终的值。双重注意力机制结构如图 6 所示,以下是两个注意力机制的详细介绍。

3.3.1 多头自注意力块

光学影像蕴含了丰富的特征信息,传统的自注意力机制只能关注到不同位置的某一方面特征,但对光学舰船影像来讲,其影像数据的特征分布差异较大,不同特征之间的关联程度不同,单一的自注意力机制不足以充分利用光学影像的特征信息,甚至产生额外的冗余信息或者噪声,不利于最后的提取结果。综上所述,本文提出运用多头自注意力模块来适应光学影像的特征信息。

基于多组自注意力构件的组合,旨在帮助输入的特征更好地捕捉自身其他维度和层次的信

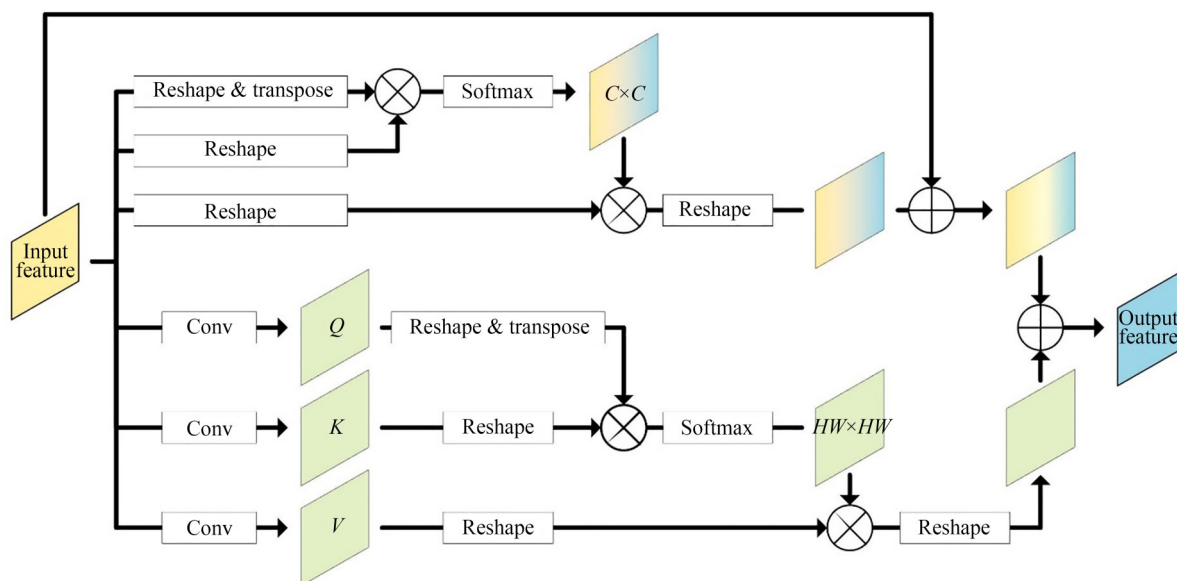


图6 双重注意力机制结构

Fig. 6 Structure of the dual attention mechanism

息,建立相适应的特征表达,从而增强模型对局部或全局信息的捕获能力。

多头自注意力模块具体实现的过程为:对每个注意力头,首先利用线性变换得到独立的查询向量 Q_i 、键向量 K_i 和值向量 V_i , W_{Qi} , W_{Ki} , W_{Vi} 分别表示三个向量的线性变换系数,如式(1)所示。然后,每个头先利用点积计算得到一个相似度 $Q_i K_i^T$,然后除以 $\sqrt{d_k}$ 后,利用softmax函数进行归一化得到权重矩阵,最后将得到的权重矩阵与 V_i 相乘得到自己的注意力输出 $\text{Attention}_i(Q_i, K_i, V_i)$, d_k 表示特征维度,如式(2)所示。最后拼接所有头的输出,即 $\text{Concat}(\text{Attention}_1, \text{Attention}_2, \dots, \text{Attention}_h)$,并进行线性变换得到最终的多头注意力输出 $\text{MultiHead}(Q, K, V)$, W_0 为线性变换系数,如式(3)所示:

$$Q_i = W_{Qi} X, \quad K_i = W_{Ki} X, \quad V_i = W_{Vi} X, \quad (1)$$

$$\text{Attention}_i(Q_i, K_i, V_i) = \text{softmax}\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right) V_i, \quad (2)$$

$$\begin{aligned} \text{MultiHead}(Q, K, V) = \\ \text{concat}(\text{Attention}_1, \text{Attention}_2, \dots, \text{Attention}_h) W_0. \end{aligned} \quad (3)$$

3.3.2 通道注意力块

通道注意力是一种针对深度学习模型中特征图通道的注意力机制。其核心思想是让模型能够在处理输入数据时,动态地关注输入的不同

部分,而不是一概而论地平均对待所有输入。根据各通道之间的相关性动态调整各个通道的权重,使得模型能够关注对当前任务更重要的通道,而降低对无关或冗余通道的关注度,提高网络模型在遥感影像中捕捉舰船特征的能力。

通道注意力权重矩阵 $S \in R^{C \times C}$ 由输入特征图 $X \in R^{C \times H \times W}$ 计算而来,其中 $H \times W$ 表示特征图的空间大小, C 代表通道个数。首先将 $X \in R^{C \times H \times W}$ 进行矩阵变换得到 $A \in R^{C \times N}$,然后对 $A \in R^{C \times N}$ 和其转置矩阵 A^T 进行相乘,最后将其通过softmax函数生成通道注意力权重矩阵 S ,公式如式(4)所示:

$$S_{ji} = \frac{\exp(A_i \times A_j^T)}{\sum_{i=1}^C \exp(A_i \times A_j^T)}, \quad (4)$$

其中: S 表示第 i 个通道对第 j 个通道的影响。然后将注意力权重矩阵 S 和 $X \in R^{C \times H \times W}$ 变换后的矩阵 $B \in R^{C \times N}$ 进行矩阵相乘,并变换为大小为 $R^{C \times H \times W}$ 的特征图使用比例系数 β 对其进行加权并与原始特征图进行叠加操作增强后的特征图 E_j ,公式如式(5)所示:

$$E_j = \beta \sum_{i=1}^C (S_{ji} B_j) + B_j. \quad (5)$$

3.4 损失函数

本文的网络模型设计了主任务和辅任务两个分支,分别利用舰船表面分割标签和舰船轮廓

标签进行损失函数计算,具体公式如式(6)所示:

$$L_{\text{total}} = L_S + \alpha L_B, \quad (6)$$

$$L = L_{\text{BCE}} + L_{\text{Dice}}, \quad (7)$$

$$L_{\text{BCE}} = -(y \log(p(x)) + (1 - y) \log(1 - p(x))), \quad (8)$$

$$\text{Dice} = \frac{2|X \cap Y|}{|X| + |Y|} \quad X \in (0, 1), Y \in (0, 1). \quad (9)$$

式(6)中 $L_{\text{total}}, L_S, L_B$ 分别表示总损失、主任务损失和辅任务损失, α 是平衡两项任务的超参数(具体取值在第四章中有针对性实验分析)。考虑到主任务与辅任务在数据集均存在类别不平衡的问题(即前景目标像素远少于背景像素),仅使用二进制交叉熵损失(BCE Loss),可能导致模型过于偏向背景预测。为了缓解这一问题并提高模型在处理稀有类别时的表现,联合Dice损失以优化模型在细节上的精确性和整体形状的恢复能力。因此,本实验的主任务损失函数和辅任务损失函数均采用结合了二进制交叉熵损失和Dice损失的复合损失函数,具体损失函数如式(7)所示,其中 L 表示主任务或辅任务损失, L_{Bce} 表示二进制交叉熵损失(具体公式如式(8)所示), y 为真实标签,取 0 或 1, $p(x)$ 为模型预测的值, L_{Dice} 表示Dice损失(具体公式如式(9)所示)其中的 $|x \cap y|$ 表示集合 x 和 y 的交集的元素个数, $|x|$ 和 $|y|$ 分别表示集合 x 和 y 的元素个数。

4 讨论与分析

4.1 实验配置

本文的所有实验配置一致,具体配置为:处理器 Intel(R) Xeon(R) Silver 4216@2.10 GHZ、显卡 Nvidia Tesla T4(内存 16 GB)、操作系统 Windows 10,并基于 python 编程语言和 PyTorch 深度学习框架实现。在模型训练过程中,采用 Adam 优化器动态调整网络参数。在训练过程中,将数据集的批量大小设置为 4,并进行 100 个 epoch 的训练。此外,为避免过拟合现象的出现,对输入数据采取随机水平翻转和垂直翻转等数据增强技术。

4.1.1 实验数据集情况

本文对 SSSD 中训练影像与测试影像按照 4:1 的比例进行随机划分,最终获得 3 760 张训练

影像和 923 张测试影像(均包含像素级标签),所有影像尺寸大小为 256 pixels $\times$ 256 pixels。具体影像数据分布情况如表 2 所示,SSSD 中公开数据集影像总数量为 2 643 张,自制数据集影像总数量为 2 040 张。

表 2 SSSD 影像数据分布情况

Tab. 2 Distribution of SSSD image data

—	Training images	Test image	All images
HRSC2016	2 291	352	2 643
Homemade	1 469	571	2 040
SSSD	3 760	923	4 683

4.1.2 实验精度评价指标

本文采用多个精度评价指标来全面评估方法的性能,包括精确度(Precision, P)、召回率(Recall, R)、F1 分数、总体分类精度(Overall Accuracy, OA)、交并比(Intersection over Union, IoU)和训练平均时间(Mean Time),计算公式如下:

$$P = \frac{TP}{TP + FP}, \quad (10)$$

$$R = \frac{TP}{TP + FN}, \quad (11)$$

$$F1 = \frac{2 \times P \times R}{P + R}, \quad (12)$$

$$OA = \frac{TP + TN}{TP + FP + TN + FN}, \quad (13)$$

$$IoU = \frac{TP}{TP + FP + FN}, \quad (14)$$

$$\text{Mean Time} = \frac{\text{time}}{\text{epoch}}, \quad (15)$$

其中: TP 表示正确分类的舰船像素数目, TN 表示正确分类的非舰船像素数目, FP 表示非舰船像素分类为舰船像素的数目, FN 表示舰船像素分类为非舰船像素的数目, time 代表训练总时间, epoch 代表训练的迭代次数。

4.2 超参数分析

为使模型性能达到最佳状态,明确辅任务与主任务所占的比重,对辅任务损失函数的超参数展开实验分析,共设置 0.5, 0.8, 1.0, 1.5 四个超参数比值,实验结果如表 3 所示。

实验结果表明,超参数比值取 1.0 时 R , $F1$, OA 和 IoU 指标精度最高, R 指标分别比超参数

表 3 损失函数超参数对比结果
Tab. 3 Results of hyperparameter comparison of loss function

α	P/%	R/%	F1/%	OA/%	IoU/%
0.5	90.53	85.11	87.74	97.26	78.15
0.8	92.63	82.07	87.03	97.18	77.04
1.0	91.10	87.55	89.29	97.58	80.66
1.5	90.35	87.34	88.82	97.47	79.89

比值为 0.5, 0.8, 1.5 时高 2.44%, 5.48%, 0.21%。F1 指标分别比超参数比值为 0.5, 0.8, 1.5 时高 1.55%, 2.26%, 0.47%。OA 指标分别比超参数比值为 0.5, 0.8, 1.5 时高 0.32%, 0.40%, 0.11%。IoU 指标分别比超参数比值为 0.5, 0.8, 1.5 时高 2.51%, 3.62%, 0.77%。P 指标比超参数比值为 0.8 时低 1.53%, 但其召回率比超参数比值为 0.8 时高 5.48%, 说明超参数为 0.8 时虽然能提高提取结果的正确率, 但同时降低了真实舰船被提取的比重, 而与超参数比值为 0.5 和 1.5 相比, 精确度和召回率都有提升, 证明了本网络架构下, 两项任务的损失约束采取等比例的关系将取得最佳结果。

4.3 对比试验

为了验证本文方法对舰船提取的有效性, 将其在 SSSD 上测试, 并与近些年典型和先进的网络模型进行对比实验, 实验的定性提取效果如图 7 所示、定量实验结果如表 4 所示。

图 7 展示了不同方法在测试数据集上的部分定性可视化结果。所选 5 张影像的背景包含不同水域环境、不同舰船分布位置 and 不同舰船类别, 基本覆盖了舰船提取的重难点区域。影像(a)为两艘舰船以不同方向靠码头停泊的状态, 背景仅涉及到海洋及陆地, 舰船位置靠近影像边缘且舰船不完整。从提取结果来看, U-Net, DeepLabV3+, MANet 出现大面积的漏检现象, ASPPUNet 在一艘舰船的船尾部分出现小部分错误漏检, 本文方法能实现舰船和海洋的准确提取。影像(b)为靠岸舰船, 其背景包含码头与海洋, 舰船与海洋背景的对比度较低, 与码头背景的对比度较高。从提取结果来看, 所有对比方法在船首位置均存在较为明显的误检现象, 而本文方法能获取更为准确的可视化结果。影像(c)涵盖部分航母影像, 目标在整幅影像中的占比较大, U-Net, ASPPUNet, Deep-

LabV3+ 均存在大面积的漏检现象, 而 MANet 和所提方法存在较少的误检像素。影像(d)涵盖两艘舰船, 在整个影像上舰船区域占比较大, ASPPUNet 出现大量的漏提取情况, U-Net 和 DeepLabV3+ 存在轻微的漏提取, MANet 和本文方法能完整地提取出影像中的两艘舰船。影像(e)的陆地背景较为复杂, 舰船仅分布在影像的左下角区域, 从提取结果来看, 对比方法均存在漏检现象, 其中 ASPPUNet 和 DeepLabV3+ 尤为严重, 本文方法能有效避免漏检, 但是仍存在舰船轮廓的过度提取现象。

表 4 展示了所提方法与对比方法(U-Net, ASPPUNet, DeepLabV3+ 和 MANet) 在 SSSD 上的定量实验结果。

由表 4 可知, 所提方法在 P, R, F1, OA, IoU 指标上均取得了领先。P 指标相比 U-Net, ASPPUNet, DeepLabV3+ 和 MANet 分别提升 11.10%, 11.78%, 6.65% 和 1.57%。R 指标相比 U-Net, ASPPUNet, DeepLabV3+ 和 MANet 分别提升 3.83%, 12.43%, 9.29% 和 0.49%。F1 指标相比 U-Net, ASPPUNet, DeepLabV3+ 和 MANet 分别提升 7.48%, 12.13%, 8.05% 和 1.74%。OA 指标相比 U-Net, ASPPUNet, DeepLabV3+ 和 MANet 分别提升 1.87%, 2.70%, 1.74% 和 0.24%。IoU 指标相比 U-Net, ASPPUNet, DeepLabV3+ 和 MANet 分别提升 11.44%, 17.84%, 12.25% 和 1.64%。说明本文方法与用于语义分割的典型与先进网络模型相比能够展现出更好的分割性能, 尽管所提方法比 ASPPUNet 和 U-Net 花费了更多的平均训练时间。然而, 相比于精度次优的 MANet, 所提方法在训练时间上取得了 46.18 s 的领先, 表明本文方法能同时兼顾模型性能与计算效率, 在基于光学影像的舰船提取任务中具有更强的实用性和竞争力。

表 4 不同网络模型提取舰船精度对比

Tab. 4 Comparison of ship accuracy extracted by different network models

Modle	P/%	R/%	F1/%	OA/%	IoU/%	Mean times/s
U-Net	80.00	83.72	81.81	95.71	69.22	65.97
ASPP	79.32	75.12	77.16	94.88	62.82	145.69
DeepLabV3+	84.45	78.26	81.24	95.84	68.41	269.68
MANet	89.53	87.06	88.28	97.34	79.02	248.64
Ours	91.10	87.55	89.29	97.58	80.66	202.46

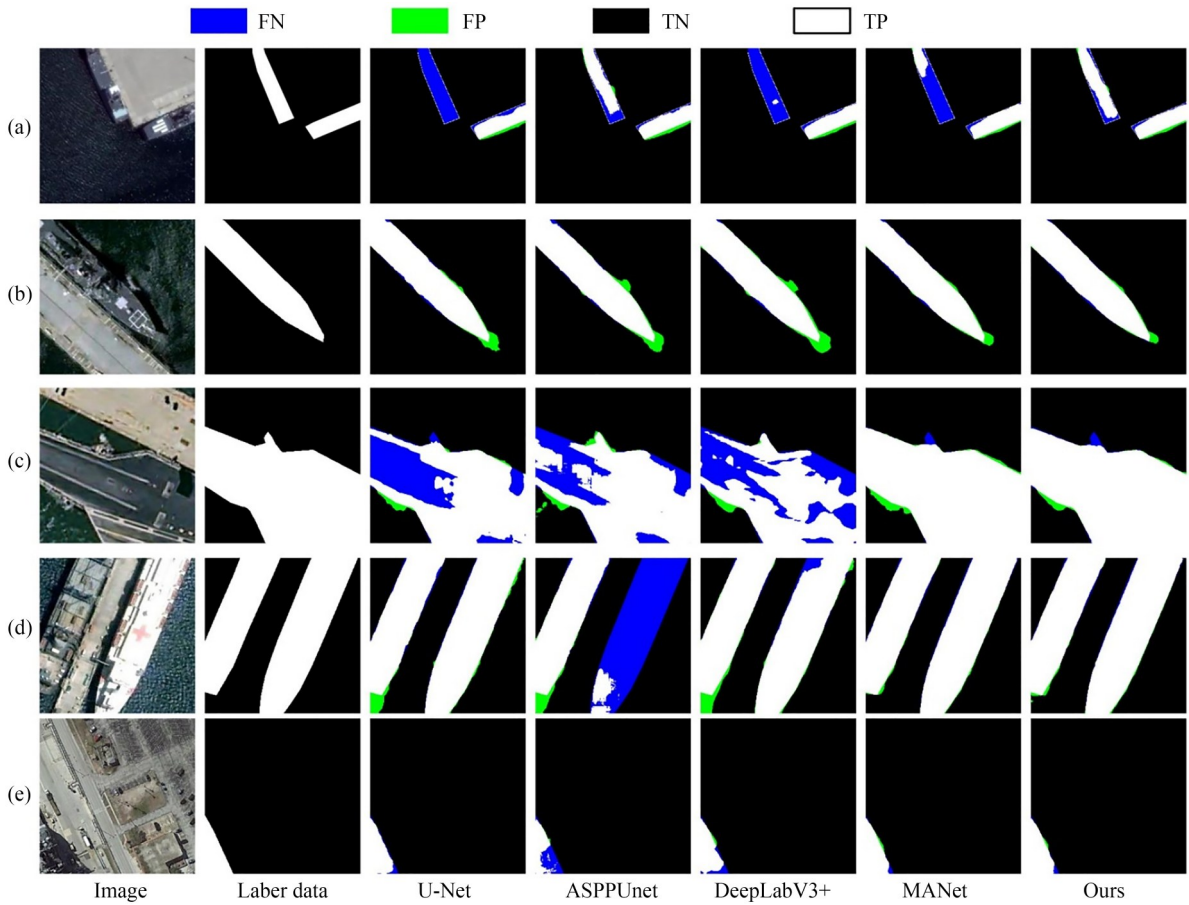


图 7 不同模型舰船分割结果示例

Fig. 7 Examples of ship segmentation results of different models

根据图 7 视觉层次的对比分析与表 4 具体数据的量化分析,证明了本文方法在定性的可视化结果和定量的精度评价指标上均优于对比方法,整体精度达到最佳效果,充分验证了所提方法在舰船提取任务上的优越性和鲁棒性。

4.4 消融实验

为证明本文所提模块的有效性,进行如下消融实验,具体情形主要包括以下 4 种,其中情形 1

为基础网络,情形 2 在情形 1 的基础上加入双重注意力,情形 3 在情形 1 的基础上加入轮廓分支,情形 4 是本文方法。消融实验精度统计情况如表 5 所示。

通过表 5 可以得出如下结论:(1)情形 1 与情形 2 相比,情形 3 与情形 4 相比,情形 2 和情形 4 在 R,F1,OA,IoU 四个指标上都有一定程度的提升。说明加入双重注意力机制对舰船提取精度

表 5 消融实验精度统计表
Tab. 5 Statistical Tab. of ablation experiment accuracy

Case	Case1	Case2	Case3	Case4
Double attention		✓		✓
Contour branching			✓	✓
P/%	90.32	90.16	92.03	91.10
R/%	85.23	87.80	85.33	87.55
F1/%	87.71	88.96	88.56	89.29
OA/%	97.25	97.49	97.46	97.58
IoU/%	78.10	80.12	79.46	80.66

有促进作用。

(2)情形 1 与情形 3 相比,情形 2 与情形 4 相比,情形 3 在 P,R,F1,OA,IoU 五个指标上均有提升,情形 4 在 P 和 R 指标上稍有落后,但在其他三个精度指标上占优,说明添加轮廓辅任务能更大地促进真实舰船被提取的比重,能满足实际应用的需要。(3)情形 4 与情形 1、情形 2、情形 3 相比,在 F1,OA,IoU 三个指标均有提升,两个模块的分别加入均有助于提升舰船提取精度,并且在协同加入后进一步提升舰船提取结果在量化评价上的精度,证明了本文所提两个模块对提升舰船提取效果均具有促进作用。

4.5 云雾实验

为了验证本文方法在复杂天气下的稳定性,本文在云雾条件下进行对比实验,考虑到没有足

够实际云雾数据,本文选择在 SSSD 中的影像上模拟云雾效果。模拟云雾效果通过 Photoshop 软件对原始影像进行批处理获得。与大场景实验一致,选取综合性能较好的两种网络(DeepLabV3+ 和 MANet)进行对比实验,在添加云雾处理后的 SSSD 上进行训练,然后用 SSSD 的测试集进行验证。

在测试影像有云雾条件下的实验结果精度统计如表 6 所示。可以看出,本文方法在 P,F1,OA 和 IoU 指标上分别比 DeepLabV3+ 高 7.61%,6.68%,1.49%,9.78%。比 MANet 高 0.76%,0.09%,0.03%,0.14%。只在 R 指标是比 MANet 略低 0.56%。本文方法在测试集有云雾条件下的提取精度优于对比方法,进一步证明了本文方法在复杂天气条件下的泛化和适应能力。

表 6 不同网络模型测试集有云雾条件提取精度对比
Tab. 6 Comparison of extraction accuracy of cloud conditions in different network model test sets

Modles	P/%	R/%	F1/%	OA/%	IoU/%
DeepLabV3+	83.06	82.00	82.23	96.00	70.25
MANet	89.91	87.77	88.82	97.46	79.89
Ours	90.67	87.21	88.91	97.49	80.03

图 8 展示了不同方法在云雾条件下的部分定性可视化结果。所选 5 张影像的背景包含不同水域环境、不同舰船分布位置和不同舰船类别,基本覆盖了舰船提取的重难点区域。影像(a)涵盖部分航母影像,目标在整幅影像中的占比较大,且背景几乎没有海洋。DeepLabV3+ 存在大量错误提取背景和错误提取舰船的情况,MANet 只在航母上方出现小部分错误提取舰船的情况,

只有本文方法较为完整和准确地提取出了影像中舰船部分。影像(b)为一艘沿着栈桥式码头分布的舰船。可以很明显地看到 DeepLabV3+ 把码头部分错误提取成了舰船,MANet 和本文方法能够较为完整地提取出舰船,但本文方法在舰船提取的完整度方面明显更占优势。影像(c)为舰船旁有小引渡船作为干扰的情况,DeepLabV3+ 错误地把小引渡船提取出来,MANet 和

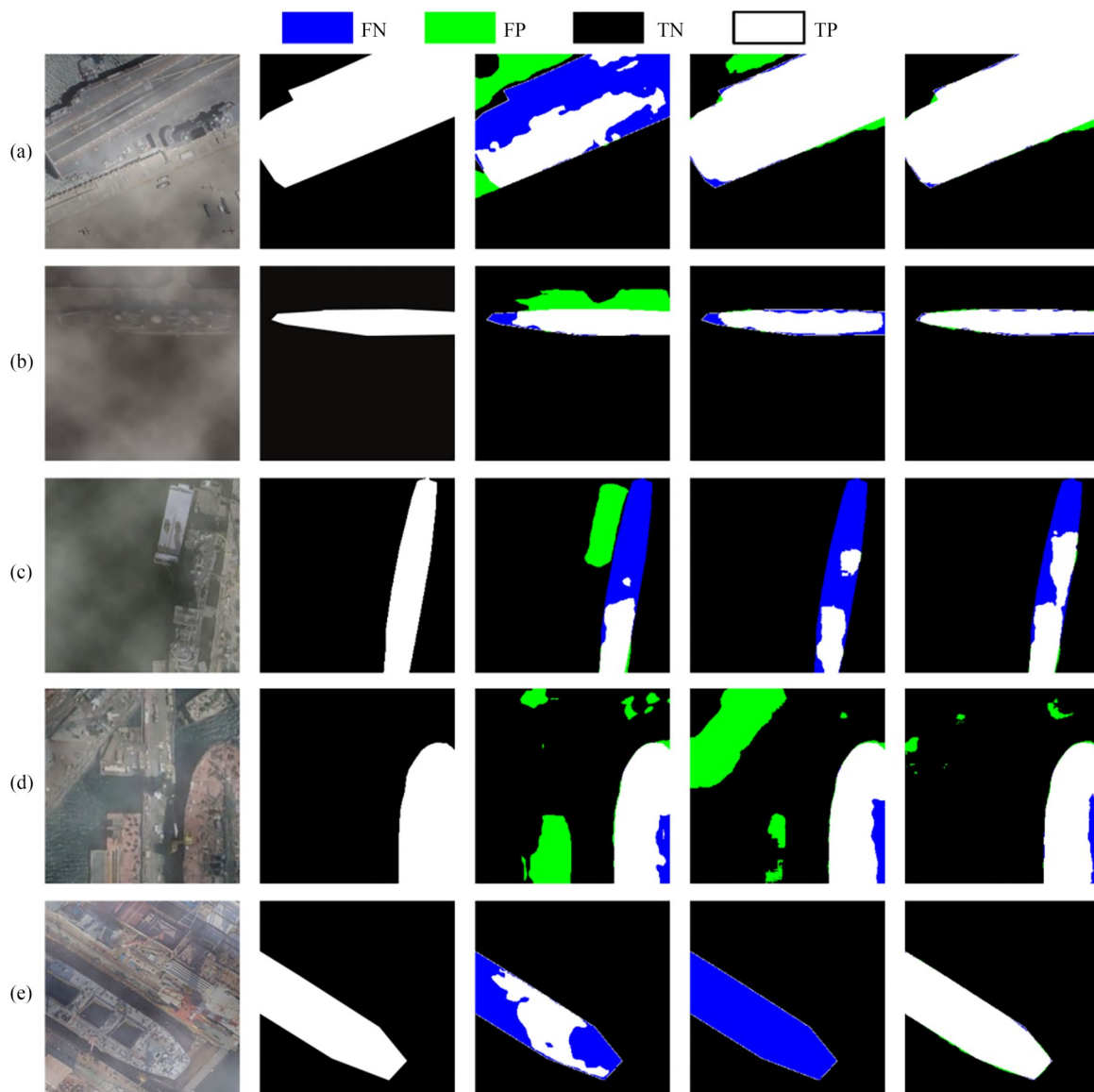


图 8 云雾条件实验效果图

Fig. 8 Experimental effect diagram of cloudy conditions

本文方法没有出现这一情况,但对于舰船的漏提取,本文方法明显要优于其他两种对比方法。影像(d)的主体为正在维修的舰船船头部分,DeepLabV3+和MANet均把码头错误提取成了舰船,而本文方法只在舰船边缘出现少量误提取。影像(e)为位于干船坞内正在建造的舰船,背景只在右下角有部分海水。DeepLabV3+对其存在一定的漏提取,MANet漏提取了整个舰船,而本文方法的结果在正确率和完整率两个方面均为理想。

根据图8视觉层次与表6具体数据的对比分析,证明了本文方法在复杂天气条件下相比对比

方法占据优势,整体精度达到最佳效果,充分验证了本文方法在舰船影像出现云雾下舰船提取任务的稳定性和普适性。

4.6 小样本实验

针对舰船所处位置多为敏感区域,数据集数量少、影像数据获取难度大、可用于训练的样本数量有限等现实情况,本文设计了小样本实验,以测试所提模型在样本数量较少时对舰船特征提取的效果,本文进行在样本数量为500,1 000,2 000条件下的对比实验,为了保证实验条件只与样本数量相关,数量大的训练样本均包含了数量少的训练样本。实验结果如图9所示。

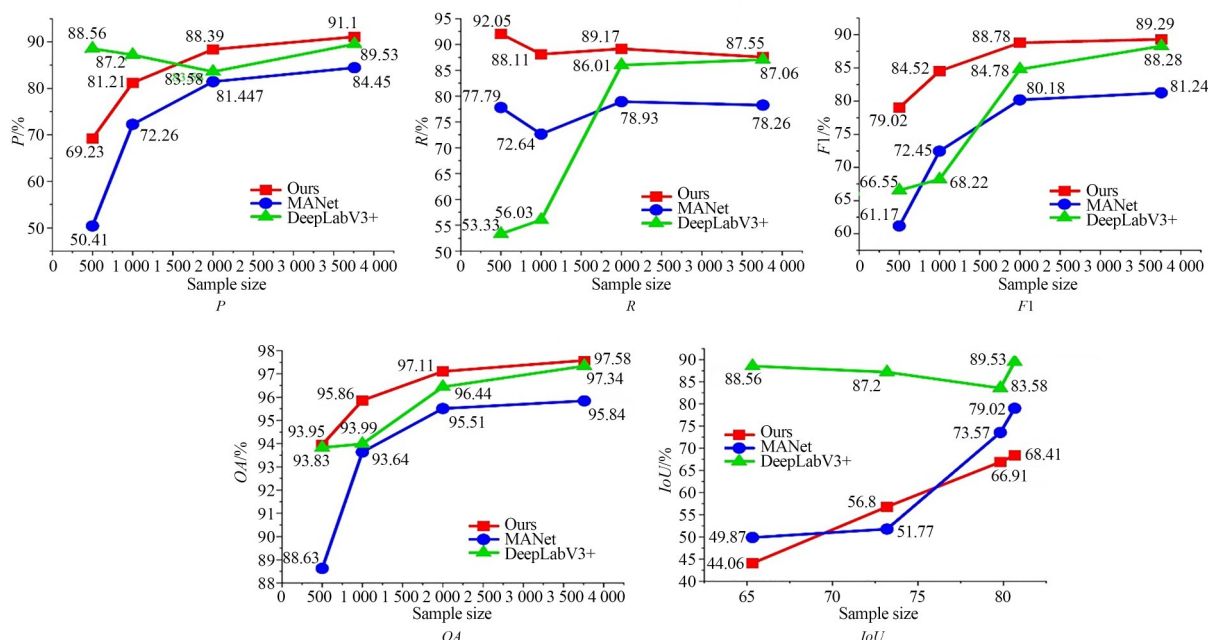


图9 小样本实验结果统计图

Fig. 9 Statistical chart of small-sample experimental results

分析折线统计数据我们可以得到以下结论:

(1)随着样本数量的增多,精度评价指标均呈现上涨趋势。(2)本文方法在R,F1,OA,IoU四个指标上不同样本数量下相比于DeepLabV3+和MANet都占据精度优势。(3)当样本数量超过2 000后,本文方法在F1和IoU两项综合性指标上提升缓慢,趋于稳定。

5 结 论

本文介绍了一种基于双任务和双重注意力的遥感影像舰船智能提取方法,旨在克服现有技术在舰船提取任务中的局限性,提升舰船提取的精度与效率,具体研究工作如下:

(1)数据集构建。整合HRSC2016数据集(公开数据集)和自制数据集,构建了舰船语义分割数据集SSSD,该数据集包含4683张舰船影像,整体场景复杂、舰船类型多样、空间分布丰富为模型训练提供了坚实的数据支撑。

(2)模型架构。提出了一种联合双任务和双重注意力的光学遥感影像舰船提取模型,该模型通过整合多头自注意力模块和通道注意力模块,以增强网络对舰船深层特征的提取能力,此外,

引入轮廓监督辅任务进一步提升了网络对深层特征的提取能力。

(3)实验验证。通过一系列对比实验、消融实验、云雾实验、小样本实验,验证了所提模型的有效性和稳定性;实验结果表明,在IoU上比U-Net, ASPPUNet, DeepLabV3+和MANet分别提升11.44%, 17.84%, 12.25%和1.64%;双重注意力和辅任务在IoU上分别提升2.02%, 1.36%;在泛化性实验中效果较好。尽管如此,本文所提模型在大场景应用中仍存在一定的局限性,与实际应用需求存在差距。受限于当前数据集的普遍问题和技术水平,仍存在一些亟须进一步解决,未来研究将聚焦于数据集的扩展、算法的优化、跨模态融合等方向,进一步提升模型的性能。

作者贡献声明:

邹毓杰:网络结构的提出,论文构思和撰写;
林雨准:实验的设计及数据整理;
王淑香:论文审核与编辑写作;
左溪冰:实验数据分析;
刘潇:论文文字校对与写作指导;
芮杰:论文总体指导与实验设计指导。

参考文献:

- [1] RIVEIRO M, FALKMAN G, ZIEMKE T. Improving maritime anomaly detection and situation awareness through interactive visualization [C]. 2008 11th International Conference on Information Fusion. June 30 - July 3, 2008, Cologne, Germany. IEEE, 2008: 1-8.
- [2] PEGLER , COLEMAN , ZHANG & Y, *et al.* The potential for using very high spatial resolution imagery for marine search and rescue surveillance [J]. *Geocarto International*, 2003, 18(3): 35-39.
- [3] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6): 84-90.
- [4] 孙国文, 罗小波, 张坤强. DeepLabV3_DHC: 城市无人机遥感图像语义分割[J]. *激光与光电子学进展*, 2024, 61(4): 394-403.
SUN G W, LUO X B, ZHANG K Q. DeepLabV3_DHC: semantic segmentation of urban unmanned aerial vehicle remote sensing image[J]. *Laser & Optoelectronics Progress*, 2024, 61(4): 394-403. (in Chinese)
- [5] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(4): 640-651.
- [6] RONNEBERGER O, FISCHER P, BROX T. *U-Net: Convolutional Networks for Biomedical Image Segmentation* [M]. Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. Cham: Springer International Publishing, 2015: 234-241.
- [7] HE H, YANG D F, WANG S C, *et al.* Road extraction by using atrous spatial pyramid pooling integrated encoder-decoder network and structural similarity loss [J]. *Remote Sensing*, 2019, 11 (9) : 1015.
- [8] CHEN L C, PAPANDREOU G, KOKKINOS I, *et al.* Semantic image segmentation with deep convolutional nets and fully connected CRFS[J]. *Computer Science*, 2014(4):357-361.
- [9] CHEN L C, PAPANDREOU G, KOKKINOS I, *et al.* DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40 (4): 834-848.
- [10] CHEN L C, PAPANDREOU G, SCHROFF F, *et al.* Rethinking atrous convolution for semantic image segmentation [J]. *arXiv*, 2017. DOI: 10.48550/arXiv.1706.05587.
- [11] CHEN L C, ZHU Y K, PAPANDREOU G, *et al.* *Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation* [M]. Computer Vision - ECCV 2018. Cham: Springer International Publishing, 2018: 833-851.
- [12] LI R, ZHENG S Y, DUAN C X, *et al.* Multiattention network for semantic segmentation of fine-resolution remote sensing images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 1-13.
- [13] LI J C, GOU S P, LI R M, *et al.* Ship segmentation via encoder - decoder network with global attention in high-resolution SAR images[J]. *IEEE Geoscience and Remote Sensing Letters*, 2021, 19: 4016605.
- [14] LIU S, DING W R, LI H G, *et al.* PLFCN: Pyramid loss reinforced fully convolutional network [C]. *Proceedings of the 13th International Conference on Distributed Smart Cameras. Trento Italy.* ACM, 2019.
- [15] CHEN L, DOU X, PENG J, *et al.* EFCNet: ensemble full convolutional network for semantic segmentation of high-resolution remote sensing images [J]. *IEEE Geoscience and Remote Sensing Letters*, 2021, 19: 8011705.
- [16] VASWANI A , SHAZEER N , PARMAR N , *et al.* Attention is all youneed [J]. *arXiv*, 2017. DOI:10.48550/arXiv.1706.03762.
- [17] FU J, LIU J, TIAN H J, *et al.* Dual attention network for scene segmentation [C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 15-20, 2019, Long Beach, CA, USA. IEEE, 2019: 3141-3149.
- [18] GUO M H, XU T X, LIU J J, *et al.* Attention mechanisms in computer vision: a survey [J]. *Computational Visual Media*, 2022, 8 (3) : 331-368.
- [19] SU Y C, LIU T J, LIUY K H. Multi-scale wavelet frequency channel attention for remote sensing image segmentation[C]. 2022 IEEE 14th Image, Video, and Multidimensional Signal Processing

- Workshop (IVMSP). June 26-29, 2022, Nafplio, Greece. IEEE, 2022: 1-5.
- [20] 张吉友, 张荣芬, 刘宇红, 等. 基于注意力机制的多模态图像语义分割[J]. 液晶与显示, 2023, 38(7): 975-984.
- ZHANG J Y, ZHANG R F, LIU Y H, *et al.* Multimodal image semantic segmentation based on attention mechanism [J]. *Chinese Journal of Liquid Crystals and Displays*, 2023, 38(7): 975-984. (in Chinese)
- [21] LI C H, LI X, XIA R L, *et al.* Hierarchical self-attention embedded neural network with dense connection for remote-sensing image semantic segmentation [J]. *IEEE Access*, 2021, 9: 126623-126634.
- [22] FAN Z Y, HOU J M, ZANG Q, *et al.* River segmentation of remote sensing images based on composite attention network [J]. *Complexity*, 2022, 2022(1): 7750281.
- [23] LIU Z K, *et al.* A high resolution optical satellite image dataset for ship recognition and some new baselines [C]. *International Conference on Pattern Recognition Applications and Methods*, 2017.
- [24] SUN X, WANG P J, YAN Z Y, *et al.* FAIR1M: a benchmark dataset for fine-grained object recognition in high-resolution remote sensing imagery [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2022, 184: 116-130.
- [25] HE K M, ZHANG X Y, REN S Q, *et al.* Deep residual learning for image recognition [C]. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 27-30, 2016, Las Vegas, NV, USA. IEEE, 2016: 770-778.

作者简介:



邹毓杰(2002—),男,湖南新化人,硕士研究生,2024年于网络空间部队信息工程大学获得学士学位,主要研究方向为影像智能解译和机器学习。E-mail:zouyujie_xd@163.com

通讯作者:



芮杰(1976—),男,安徽庐江人,博士,教授,硕士生导师,2017年于中国科学院大学获得博士学位,主要研究方向为多源遥感图像信息获取与应用研究。E-mail:ruijie@radi.ac.cn