

一种联合空间变换和置换注意力机制的 近岸水产养殖区信息提取方法

巫统仁^{1,2,3}, 张显³, 刘培^{1,2,4*}, 文婷婷^{1,2,3}, 邹振学^{1,2,3}

(1. 海南热带海洋学院 崖州湾创新研究院, 海南 三亚 572025; 2. 海南省海洋与渔业科学院, 海南 海口 571126; 3. 河南理工大学 测绘与国土信息工程学院, 河南 焦作 454003; 4. 北华航天工业学院 遥感信息工程学院, 河北 廊坊 065000)

摘要: 为解决水产养殖区遥感提取过程中近岸坑塘养殖目标和网箱养殖目标地物背景复杂, 受房屋、植被、海水和船只等干扰提取精度较低等问题, 提出了一种联合空间变换和置换注意力机制的近岸水产养殖区信息提取方法 SA-STN-Net, 选取海南省文昌市八门湾和万宁市坡头港为研究区域, 先利用光谱特征和纹理特征构建水产养殖目标先验知识, 然后在 U-Net 模型基础上联合空间变换网络 (spatial transformer network, STN) 和置换注意力机制 (shuffle attention, SA), 用于增强养殖目标空间特征、减少复杂地物的干扰并聚焦近岸水产养殖区域。结果表明: 与原始 U-Net 模型相比, SA-STN-Net 模型的总体提取精度和平均交并比提高了 3.3% 和 5.7%; 与当前较为先进的 A2fpn、Swin-Transformer 和 Dc-Swin 等深度学习分割算法相比, SA-STN-Net 模型具有更好的分割性能, F1 分数分别提高了 6.7%、4.2% 和 7.2%。研究表明, 本文提出的 SA-STN-Net 模型能适应近岸水产养殖目标地物背景复杂的情况, 可对近岸养殖目标进行有效提取, 本研究结果可为近岸规划与管理部门提供技术支持。

关键词: 空间变换网络 (STN); 置换注意力 (SA); 深度学习; 水产养殖区信息提取

中图分类号: S 967.3; TP 391

文献标志码: A

近岸养殖作为一种传统的海岸带经济活动, 在中国海洋经济中占有重要地位^[1]。坑塘养殖和网箱养殖是近岸养殖的重要方式, 随着养殖规模的逐步扩大, 其造成的海岸带环境压力日益增加^[2-3]。因此, 实现近岸水产养殖区域的精细提取与科学监管, 对水产养殖业的可持续发展具有重要意义^[4]。

卫星遥感技术具有实时性、大范围和长时序性等优点, 为近岸养殖目标的提取和监测提供了一种方便快捷的途径^[5]。传统的基于遥感影像的近岸养殖目标提取方法包括目视解译法^[6]、水体指数法^[7]、面向对象法^[8]和聚类法^[9]。其中, 目视解译法受人为主观因素影响较大, 效率低下且需要昂贵的人工成本。水体指数法简单有效且易于实现, 但提取精度较低, 对于水中养殖目标的提取效果不佳。除此之外, 随着近岸养殖目标类型的变化与养殖范围的扩大, 养殖目标环境背景也愈发复杂, 传

统算法的精度难以满足实际需求。

近年来, 卷积神经网络 (convolutional neural network, CNN) 凭借其强大的特征提取能力在包括近岸养殖目标分割在内的各个领域取得了令人瞩目的成果^[10-11]。利用深度学习方法进行影像分割不仅减少了人为因素的干扰, 还具有较强的灵活性和高效性, 如 A2fpn 网络^[12]、Ma-Net 网络^[13]、Swin-Transformer 网络^[14]和 Dc-Swin 网络^[15]等, 而 U-Net 网络^[16]和 Segnet 网络^[17]同为用于影像分割的全卷积神经网络, 二者因其具有简单的结构与良好的分割效果, 在水产养殖区域分割任务中被广泛应用^[18-19]。为进一步提高水产养殖区域的提取精度, 研究人员对模型进行了许多改进, 包括加入上采样压缩激励模块^[20]、引入坐标注意力机制^[21]等。然而, 这些研究多是针对模型的结构进行改进, 却未考虑水产养殖目标本身的特征。除此

收稿日期: 2023-10-23

基金项目: 海南热带海洋学院崖州湾创新研究院开放课题重点项目 (2022CXKYKFKT03); 海南热带海洋学院崖州湾创新研究院课题组学生创新重点项目 (2022CXYSXCXM07); 海南省自然科学基金 (423MS120); 海南省海洋与渔业科学院省本级课题项目 (KYL-2024-06); 河北省自然科学基金面上项目 (D2020409002)

作者简介: 巫统仁 (1999—), 男, 硕士研究生。E-mail: wtw2021yjs@163.com

通信作者: 刘培 (1985—), 男, 博士, 副研究员。E-mail: cumtlp@qq.com

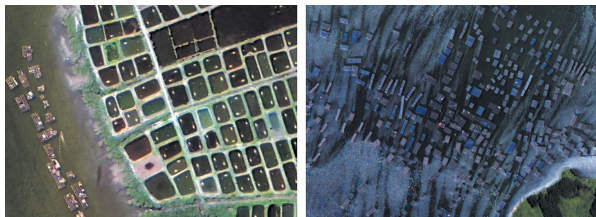
之外,已有的研究直接将遥感影像作为模型输入的对象,模型对养殖区域特征的学习完全依赖输入影像,缺少先验知识,而纹理、光谱指数等人工构建的特征对提高 CNN 的提取能力至关重要。

本研究中,以 U-Net 模型为基线方法,同时对陆基养殖区域与海面养殖区域(统称为“近岸养殖区域”)进行提取。针对以往研究缺少先验知识的问题,利用归一化水体指数与灰度共生矩阵法人工构建了近岸养殖的光谱与纹理特征,并将其作为先验知识与影像共同输入到模型中,在提高模型对近岸养殖区域分割精度的同时加速模型收敛;由于近岸养殖区域具有复杂的地物背景,受房屋、植被、海水和船只等因素的干扰,综合提取难度较大,为减少复杂地物背景对近岸养殖目标提取的干扰,使模型重点关注近岸养殖区域,在模型中添加置换注意力机制^[22];同时,单个养殖目标的形状多为不规则的矩形,由于不同养殖区域具有不同的空间位置、旋转角度与缩放尺度,为充分利用近岸养殖区域本身的空间特征,又将 U-Net 网络与空间变换网络^[23]相结合。最终提出了 SA-STN-Net 模型,并通过评价指标来评估所提出模型的提取效果,以期实现对近岸养殖区信息的有效提取。

1 研究区域及数据

1.1 研究区域

海南岛地处热带(18°10′~20°10′N, 108°37′~111°03′E),雨量充沛,沿岸海域水质优良,平均潮差小于 1.7 m,平均 pH 为 8.0,为近岸水产养殖提供了较为有利的条件。海南省文昌市八门湾和万宁市坡头港养殖类型以陆基养殖和网箱养殖为主(图 1),养殖动物以鱼虾为主,并已成为当地重要的经济来源,故以上述两个地区作为研究区域,对其进行水产养殖区信息提取具有代表意义。



(a) 文昌市 Wenchang City (b) 万宁市 Wanning City

图 1 研究区及养殖类型示意图

Fig. 1 Survey area and aquaculture type images

本文中的陆基养殖区是指位于近岸海域的陆地养殖塘区域,由池塘构成,池塘间通过垄地隔开。在遥感影像中几何特征较为规则,水体由海水或淡

水组成,水面面积小,边缘明显,其光谱特征受植被和房屋等因素影响。近岸网箱养殖一般指由框架系统、箱体系统和锚泊系统组合而成的养殖系统,其抗风浪能力较差,常设置在避风条件好的内湾、港湾等区域,大多靠近岸边且分布密集。在遥感影像中养殖形状规则,单个网箱为矩形,多个网箱连接为长条状矩形,光谱特征易受海水、船只等影响。

1.2 数据源

GF-2 遥感卫星是中国首颗自主研发的亚米级空间分辨率的民用光学卫星,搭载 1 m 全色和 4 m 多光谱两台相机,其包含红、绿、蓝、近红外和全色共 5 个波段,广泛应用于国土资源调查、土地监测和地质灾害监测等领域^[24]。本研究中,使用的数据为 GF-2 卫星遥感数据(表 1),坡头港的影像拍摄日期为 2020 年 8 月 10 日,八门湾为 2021 年 9 月 3 日,共 2 景影像。在养殖区信息提取之前,对原始数据进行辐射定标、大气校正等预处理,以消除大气和光照对地物反射所带来的影响。由于多光谱影像分辨率较低,利用 NNDiffuse Pan Sharpening 图像融合算法^[25]将多光谱影像空间分辨率提高至 1 m,再对其进行影像裁剪至研究范围。为获取相应的标签数据,对裁剪后的影像进行标注,分为陆地、海水、陆基养殖区和网箱养殖区共 4 类,其中,黄色代表陆地,蓝色代表海水,绿色代表陆基养殖区,红色代表网箱养殖区。将标注好的影像进行裁剪,大小为 256×256(图 2)。

表 1 数据来源

Tab. 1 Data source

卫星 satellite	覆盖区域 coverage area	位置 position		传感器类型 sensor type	拍摄日期 shooting date
		纬度 latitude	经度 longitude		
GF-2	八门湾	19.5N	110.8E	PMS	2021-09-03
	坡头港	18.6N	110.3E	PMS	2020-08-10

2 SA-STN-Net 网络结构

研究过程中,首先对原始高分遥感影像进行预处理,包括辐射定标、大气校正、几何校正、图像融合和图像裁剪等;然后利用归一化水体指数和灰度共生矩阵法提取养殖区的光谱信息和纹理信息,并完成数据集的制作,训练集、测试集和验证集比例为 7.5 : 1 : 1.5,利用联合空间变换和置换注意力机制的 SA-STN-Net 模型实现不同类型的水产养殖区信息提取;最后通过消融试验和对比试验对模型进行精度验证和分析,总体流程如图 3 所示。

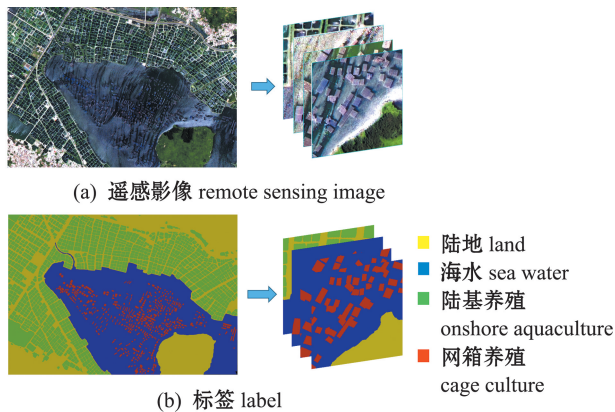


图 2 遥感影像与标签

Fig. 2 Remote sensing image and label

2.1 先验知识构建

纹理反映了遥感影像中地物的空间分布状况,光谱则描述了影像对象及对象中像元所有特征的集合,纹理信息和光谱信息是参与遥感图像目标识别

的重要特征。本研究中,首先,对预处理后的影像进行主成分变换,选取对比度、熵和均值这 3 种纹理特征参数,为避免数据冗余,利用灰度共生矩阵法对集中大多数信息的第一主分量进行纹理特征提取,作为水产养殖区纹理数据;归一化差异池塘指数(normalized difference pond index, NDPI)对识别池塘有较好的效果,由于高分影像缺少近红外通道,采用 NDWI 水体指数对遥感影像的特定波段进行差值处理,以此来突出遥感影像中的水体信息^[26],作为水产养殖区光谱数据。其表达式为

$$I_{NDWI} = (\rho_G - \rho_{NIR}) / (\rho_G + \rho_{NIR})。 \quad (1)$$

式中: ρ_G 为绿波段的反射率; ρ_{NIR} 为近红外波段的反射率。获取水产养殖区光谱特征和纹理特征后,将其作为先验知识与高分影像的波段数据一并输入到模型中参与训练,共 8 个通道数据,包括高分影像的 4 个波段数据、3 个纹理特征数据和 1 个光谱特征数据。

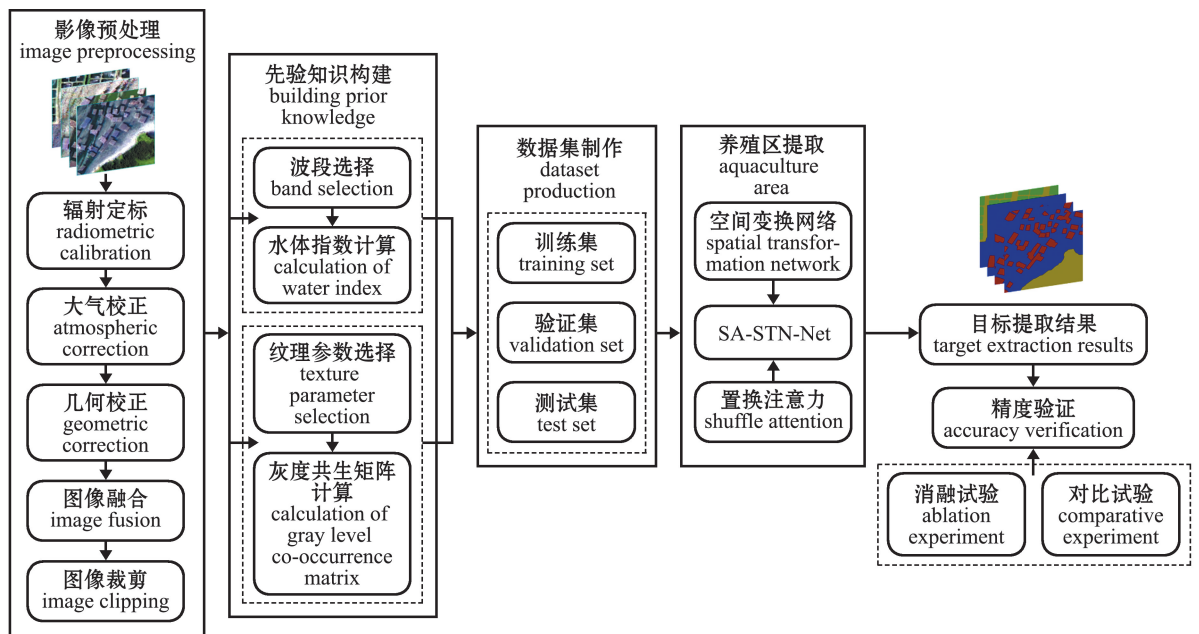


图 3 水产养殖区信息提取总体流程

Fig. 3 Overall process of information extraction in aquaculture areas

2.2 联合空间变换 (STN) 和置换注意力机制 (SA) 的 SA-STN-Net 模型

SA-STN-Net 模型是在 U-Net 模型基础上加以改进的。U-Net 模型由编码器和解码器两部分组成,编码器用于提取遥感影像的特征信息,解码器则将编码器获取的特征图逐渐恢复至原始图像大小。编码器输入尺寸为 256×256 的通道数据,首先经过通道数为 64 的卷积层,得到的特征图经下采样处理后,图像尺寸为原始通道数据的 $1/2$,即 $128 \times$

128 ;再经过通道数为 128 的卷积层,经下采样处理后图像尺寸为原始通道数据的 $1/4$,即 64×64 ;最后经过通道数为 256 的卷积层,经下采样处理后图像尺寸为原始通道数据的 $1/8$,即 32×32 。由于不同通道信息对于养殖目标提取的贡献不同,在面对多通道的遥感数据时有必要调整不同通道在特征提取中的权重,以达到突出重点信息,抑制或忽略无关信息的效果。SA-STN-Net 模型引入了置换注意力机制 SA 模块,在 SA-STN-Net 模型整个编码器和解码器的 4 层中,将经过 3×3 卷积层处理后的

特征图输入到 SA 模块, 对有利于近岸养殖目标提取的信息赋予更大的权重, 生成新的带有通道权重信息的特征图。同时为了最大程度地获取养殖目标自身的空间特征, 在每一层的编码-解码结构中, 通过残差连接引入空间变换网络 STN 模块, 将每一层带有通道信息的新特征图通过残差连接输入到 STN 模块, 对其进行仿射变换处理后再次通过残差连接输入到解码器中。该模块可以为每个输入的特征图进行仿射变换, 使神经网络能够准确学习到养殖目标的形状变换, 从而达到增强养殖目标空间特征的效果。SA-STN-Net 模型整体结构如图 4 所示。

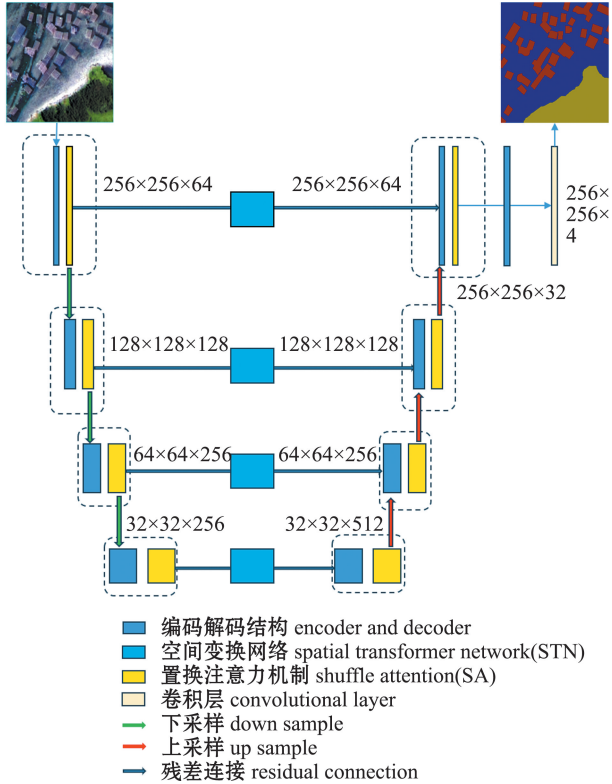


图 4 SA-STN-Net 网络结构

Fig. 4 Schematic of SA-STN-Net structure

SA 模块是 Zhang 等^[22] 提出的一种新的注意力机制, 该注意力机制是在空间注意力^[27] 与通道注意力机制^[28] 的基础上引入了特征分组和通道置换, 是一种超轻量型的注意力机制。其优势在于采用随机单元, 较好地实现了空间注意力和通道注意力机制的融合, 相比于压缩和激励模块^[29], 其在特征提取上具有更好的准确性和较低的模型复杂度。SA 模块包含 Feature Grouping、Channel Attention、Spatial Attention 和 Aggregation 4 个部分。第一部分 Feature Group 为特征分组, 用于对输入的特征数据 X ($X \in \mathbf{R}^{c \times h \times w}$) 进行分组, 沿着通道维度将特征图 X 拆分为 g 组, 此时目标特征 $X = [X_1, X_2, X_3, \dots, X_g]$ ($X_i \in \mathbf{R}^{c/g \times h \times w}$), 通过空间

注意力和通道注意力模块对分组后的每一组特征生成不同的重要性系数。具体来说, 分组后的目标特征 X_k 沿着通道维度拆分成两个分支 X_{k1} 和 X_{k2} ($X_k \in \mathbf{R}^{c/2g \times h \times w}$), X_{k1} 用于学习通道注意力特征, X_{k2} 用于学习空间注意力特征。第二部分 Channel Attention 为通道注意力模块, 为避免大量的参数, 采取了 GAP、Scale 及 Sigmoid 的组合, 数学表达式为

$$s = F_{gp}(X_{k1}) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X_{k1}(i, j), \quad (2)$$

$$X'_{k1} = \delta(F_c(s)) X_{k1} = \delta(W_{1s} + b_1) X_{k1}. \quad (3)$$

式中: $W_1 \in \mathbf{R}^{c/2g \times 1 \times 1}$, $b_1 \in \mathbf{R}^{c/2g \times 1 \times 1}$ 。第三部分 Spatial Attention 为空间注意力模块, 首先采用 Group Norm (GN) 对 X_{k2} 进行处理, 得到空域层面上的统计信息, 再利用 $F_c(\cdot)$ 对其进行增强处理, 该过程的数学表达式为

$$X'_{k2} = \delta(W_2 \times GN(X_{k2}) + b_2) X_{k2}. \quad (4)$$

式中: $W_2 \in \mathbf{R}^{c/2g \times 1 \times 1}$, $b_2 \in \mathbf{R}^{c/2g \times 1 \times 1}$ 。第四部分 Aggregation 用于对完成上述两种注意力机制计算后的特征进行集成, 该过程首先通过拼接实现特征融合, 得到 X'_k , 此时 $X'_k = [X'_{k1}, X'_{k2}]$ ($X'_k \in \mathbf{R}^{c/2g \times h \times w}$), 然后采用通道置换进行组间通信, 使得最终输出的特征尺寸与输入的特征尺寸一致。

STN 模块是基于可微分注意力机制的一种改进网络^[30], 其核心思想是利用网络内部学习的特征进行空间变换, 以自适应的方式对输入的样本进行空间变换和对齐, 主动对养殖目标特征进行旋转、缩放和平移的空间变换, 将感兴趣区域对准网络中的最优位置, 从而帮助模型充分利用近岸养殖目标自身的空间特征, 提高分割性能。STN 作为独立模块增强了深度卷积神经网络的几何不变性和学习能力, 使得网络的训练更加高效和稳定。STN 模块由本地网络、网格生成器及采样器 3 部分组成。将特征图输入后, 通过本地网络得到变换参数 θ :

$$\theta = \begin{bmatrix} a & c & e \\ b & d & f \end{bmatrix}. \quad (5)$$

其中, 参数 θ 包含 6 个参数, 利用变换参数 θ , 对输入的特征图进行旋转、缩放和平移等空间变换。网格生成器的作用是生成参数化的采样网格, 利用变换参数 θ 和空间坐标点 V , 反算出输入特征图相应的位置坐标点 U , 通过变换函数 $T_\theta(G)$ 对输入的特征图进行矩阵运算, 实现仿射变换。该过程的数学表达式为

$$\begin{bmatrix} x_i^u \\ y_i^u \\ 1 \end{bmatrix} = \theta \begin{bmatrix} x_i^v \\ y_i^v \\ 1 \end{bmatrix} = \begin{bmatrix} a & c & e \\ b & d & f \end{bmatrix} \begin{bmatrix} x_i^v \\ y_i^v \\ 1 \end{bmatrix} = \begin{bmatrix} ax_i^v & by_i^v & e \\ cx_i^v & dy_i^v & f \end{bmatrix}. \quad (6)$$

采样器的作用是通过填充的方式对输入的坐标点和原图 U 进行填充,从而输出 V ,将 U 和 V 若干个相应位置的特征信息进行融合,使得近岸养殖目标自身的空间特征信息表达得更加全面。

2.3 精度评价指标

试验选取 $F1$ 分数 ($F1$ score)、总体精度 (overall accuracy, OA)、平均交并比 (mean intersection over union, MIoU)、召回率 (recall rate, R)、准确率 (precision, P) 及戴斯系数 (Dice) 来评价不同模型对陆基养殖区和近岸网箱养殖区信息的提取能力。其中,召回率表示养殖区被模型正确分类的像元数占真实养殖区像元数量的比例;准确率表示养殖区被模型正确分类的像元数占模型分类结果为养殖区像元数量的比例; $F1$ 分数表示准确率和召回率的调和平均;总体精度表示模型正确分类的像元数占所有真实地物像元数的比例;平均交并比和 Dice 系数是语义分割常用的两个指标。

3 近岸水产养殖区信息提取试验

3.1 参数设置与模型性能分析

试验在 GPU 为 NVIDIA GeForce RTX 3060 的服务器上进行,模型采用 PyTorch 框架实现^[31]。在模型训练时利用数据增强手段来扩充数据集,试验使用 Adam 优化器对模型进行优化,超参数 β_1 为 0.9, β_2 为 0.99。初始学习率设置为 2×10^{-4} ,学习率衰减轮次设置为 150,衰减系数设置为 0.5,共训练 300 轮次。

为测试本文中所建模型 SA-STN-Net 的性能,将该模型与当前一些较为先进的语义分割模型进行了对比,包括 A2fpn、Ma-Net、Swin-Transformer、Dc-Swin 和 Segnet 模型。其中, A2fpn 是一种用于高分辨率遥感影像语义分割的改进特征金字塔网络,旨在有效聚合不同尺度的特征,以增强多尺度特征学习并提高影像分割的准确性; Ma-Net 是一种用于影像分割的多尺度注意力网络,该网络引入了自注意力机制,能对局部特征与全局特征进行自适应整合处理; Swin-Transformer 是一种基于 Transformer 结构的深度学习网络,通过引入滑动窗口机制和层次化结构,可实现跨窗口的信息交互; Dc-Swin 是一种基于 Swin-Transformer 的改进深度学习模型,通过结合 Swin-Transformer 网络和特征聚合模块,可实现对输入数据的高效特征提取和聚合。从图 5 可见,前 50 个轮次各个模型的性能迅

速提升,此时, SA-STN-Net 模型的性能表现最好并一路延续下去,在训练到 250 轮次的时候,各个模型的性能变化甚微,表明各个模型的性能达到饱和状态,其中,性能最好的是 SA-STN-Net 模型,而最差的是 Segnet 模型。

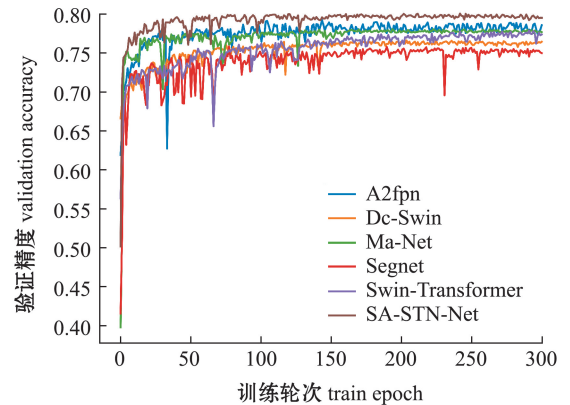


图 5 不同模型在训练过程中的性能变化

Fig. 5 Performance changes of different models during the training process

3.2 消融试验

为评价本文所提出的 SA-STN-Net 模型及先验知识的有效性,对其进行了消融试验,以验证先验知识、SA 模块和 STN 模块的有效性。U-Net 模型是指未改进的原始模型, U-Net-1 模型是指添加先验知识后的模型, SA-NET 是指添加 SA 模块后的模型, SA-STN-Net 模型是指添加 SA 模块和 STN 模块后的模型。所有的消融试验均在同一台服务器上进行,训练过程中的参数保持一致,消融试验结果如表 2 所示。根据试验结果可以看出: U-Net-1 模型的平均交并比和 Dice 系数比原始的 U-Net 模型分别提高了 3.1% 和 1.8%,这说明添加先验知识后的模型在获取近岸养殖目标的光谱特征和纹理特征后,模型通过学习这些特征,提高了对水产养殖目标的识别精度; SA-NET 模型的平均交并比和 Dice 系数比原始的 U-Net 模型提高了 4.3% 和 4.2%,说明 SA 模块的引入是有效的; SA-STN-Net 模型的性能最好,与原始的 U-Net 模型相比, $F1$ 分数提高 3.0%,总体精度提高了 3.3%,平均交并比提高了 5.7%,召回率提高了 4.3%, Dice 系数提高了 5.6%,模型的整体性能均有所提高,与 SA-NET 模型相比,总体精度提高了 1.1%,平均交并比提高了 1.4%, Dice 系数提高了 1.4%。这是因为 SA-STN-Net 模型的每个编码-解码部分引入了 STN 模块,将每一层带有通道信息的新特征图通

表 2 消融试验结果

Tab.2 Ablation results

模型	先验知识	置换注意力	空间变换网络	F1 分数/%	总体精度/%	平均交并比/%	召回率/%	Dice 系数/%
model	prior knowledge	SA	STN	F1 score	overall accuracy	MIoU	recall	Dice coefficient
U-Net				91.1	93.1	76.1	83.2	80.5
U-Net-1	✓			92.4	94.9	79.2	85.9	82.3
SA-Net	✓	✓		93.2	95.3	80.4	86.3	84.7
SA-STN-Net	✓	✓	✓	94.1	96.4	81.8	87.5	86.1

过残差连接输入到 STN 模块中, 该模块对输入的特征图进行仿射变换处理, 使网络能获取更多养殖目标自身的空间特征信息, 从而增强了模型对养殖区信息提取的平移、旋转与缩放不变形能力。

3.3 近岸养殖区信息提取结果与分析

为评价不同模型对近岸养殖目标的提取效果, 从分类结果中选择了 4 处陆基坑塘养殖区和 2 处海面网箱养殖区, 将各模型的预测结果与真实标签进行对比, 局部陆基养殖区和海面网箱养殖区预测结果如图 6 和图 7 所示, 其中, Image 代表原始影像, Label 代表标签数据。对于陆基坑塘养殖区, 从图 6 可以直观地看出, Segnet 模型和 Ma-Net 模型的效果不佳, 在提取 a 图中坑塘养殖区域的过程中, 受到复杂地物的干扰, 均出现大部分陆地被错分为海水的现象; Swin-Transformer 模型、A2fpn 模型及 Dc-Swin 模型提取效果较好, 这是因为这 3 种模型都引入了注意力机制, 在面对复杂地物干扰时, 能聚焦于近岸养殖目标, 分割精度较高, 未出现大片区域被错分的现象, 只有小部分陆地被错分

为陆基坑塘养殖区; SA-STN-Net 模型的效果最好, 基本上无粘连现象, 错分漏分现象最少, 大多数的陆基养殖区都能够被正确划分, 但也存在错分现象, 如 d 图中部分区域陆基养殖区被错分为海水。

对于近岸网箱养殖区, 从图 7 可直观地看出, Segnet 模型的提取效果最差, 大片的网箱养殖区域被错分为海水, 错分现象最为严重; Swin-Transformer 模型、Ma-Net 模型和 Dc-Swin 模型能提取出网箱养殖区域, 但存在养殖区边缘部分不整齐、网箱粘连等现象; Ma-Net 模型在提取陆基养殖区时表现不佳, 但其在提取网箱养殖区时表现较好; SA-STN-Net 模型提取网箱养殖区的效果最好, 网箱养殖区域均能较准确地被提取出来, 提取的网箱之间粘连现象较少且提取的网箱边界较为平整。总体来看, SA-STN-Net 模型在面对不同类型的养殖区时均表现出良好的提取效果, 这是因为 SA-STN-Net 模型联合了置换注意力机制和空间变换网络, 通过 SA 模块可使网络更加集中于近岸养殖区域, 对有利于近岸养殖目标提取的信息分配了更高的权重, 降低了复杂地物的干扰, 并通过 STN 模块实现

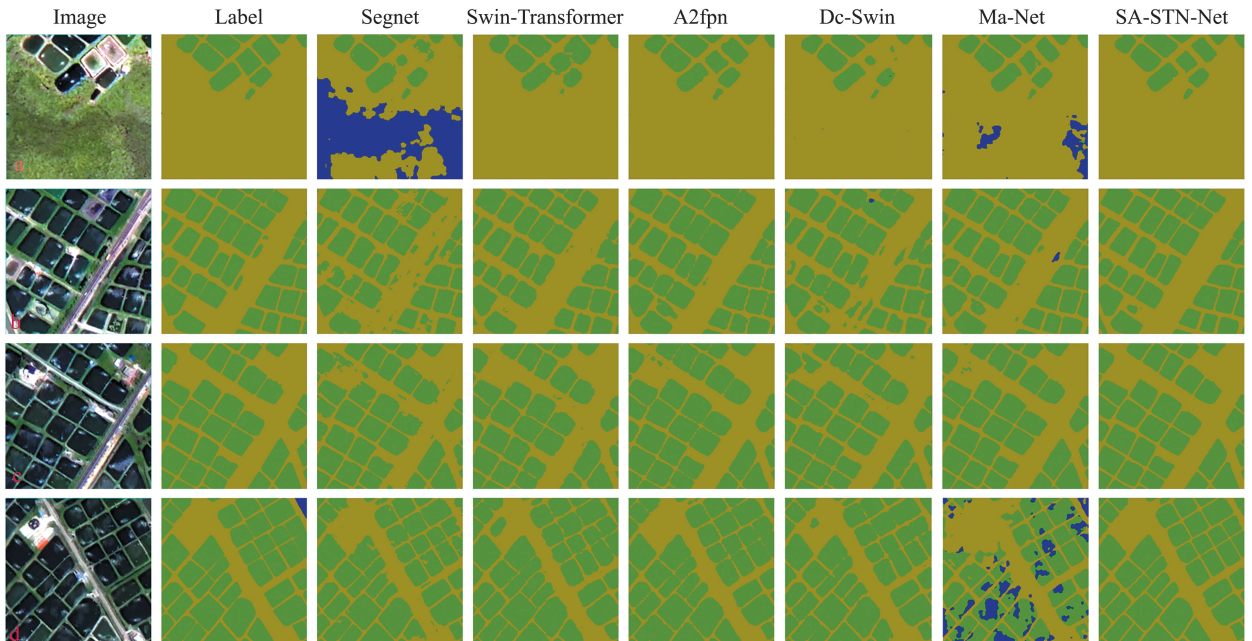


图 6 局部陆基养殖区信息提取结果

Fig. 6 Information extraction results of part onshore aquaculture area

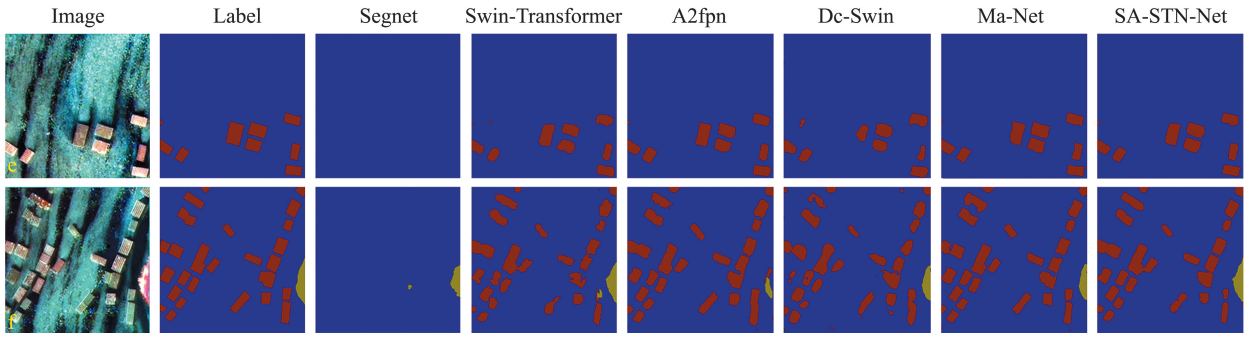


图 7 局部近岸网箱养殖区信息提取结果

Fig. 7 Information extraction results of part cage culture area

了网络对输入特征图的仿射变换,增加了近岸养殖目标自身的空间特征信息,从而提高了近岸养殖目标提取的准确性。

从表 3 可见: Segnet 模型的提取精度指标中除 $F1$ 分数值外,其他指标结果明显低于其他模型, $MIoU$ 仅为 54.6%,结合提取效果来看,该模型受复杂地物的干扰最为严重,不适用于近岸养殖区信息提取,特别是陆基坑塘养殖区域; Dc-Swin 和 Ma-Net 模型的提取精度指标结果相近, $MIoU$ 分别为 68.2%和 69.7%,均未突破 70%, $F1$ 分数分别为 86.9%和 86.1%,在提取近岸水产养殖区域的表现上相差不大,只能大致提取出养殖区域,精度

较低; A2fpm 和 Swin-Transformer 模型在提取近岸水产养殖区域的表现上可归为同一个档次,两者均引入了注意力机制,在一定程度上缓解了周边复杂地物的干扰,总体精度分别为 96.1%和 95.4%, $MIoU$ 分别为 76.6%和 73.6%,能较好地提取出近岸水产养殖区域;在所有指标中, SA-STN-Net 模型的 $F1$ 分数最高,相比于 Segnet 模型,总体精度从 84.7%提高至 96.4%,平均交并比从 54.6%提高到 81.8%, Dice 系数从 60.3%提高到 86.1%。

从养殖区域的细节提取上看, SA-STN-Net 模型的表现依然是最好的,相比于其他几种模型,其提取的养殖目标边缘更符合原始影像(图 8 和图 9)。

表 3 不同深度学习模型提取的近岸水产养殖区精度比较

Tab. 3 Comparison of information accuracy extracted by different depth learning models in inshore aquaculture area %

模型 model	$F1$ 分数 $F1$ score	总体精度 overall accuracy	平均交并比 $MIoU$	召回率 recall	Dice 系数 Dice coefficient
A2fpm ^[12]	87.4	96.1	76.7	81.6	81.5
Segnet ^[17]	89.2	84.7	54.6	69.9	60.3
Swin-Transformer ^[14]	89.9	95.4	73.6	85.7	78.4
Dc-Swin ^[15]	86.9	93.7	68.2	77.9	74.2
Ma-Net ^[13]	86.1	91.4	69.7	83.8	74.6
SA-STN-Net	94.1	96.4	81.8	87.5	86.1

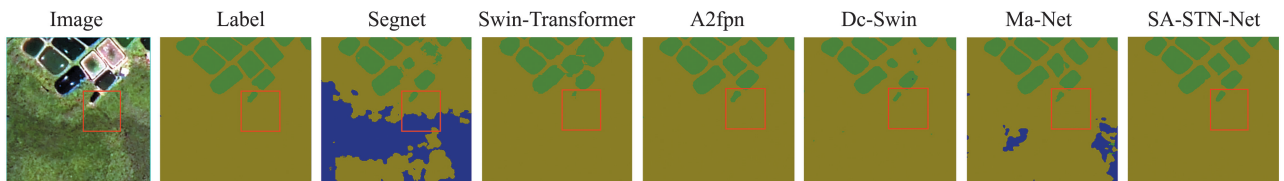


图 8 不同模型提取的局部陆基坑塘养殖区域细节对比

Fig. 8 Comparison of details of local onshore aquaculture area extracted by different models

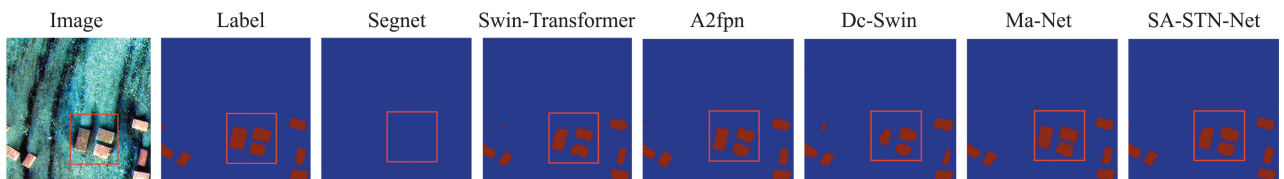


图 9 不同模型提取的局部近岸网箱养殖区域细节对比

Fig. 9 Comparison of details of local offshore cage culture area extracted by different models

综上所述,本文中提出的 SA-STN-Net 模型提取近岸养殖区域的整体效果最好, $F1$ 分数和 MIoU 得分最高也说明该模型具有更好的鲁棒性,这有助于其应用在近岸水产养殖区信息提取领域中。

4 讨论

4.1 先验知识的构建对近岸养殖目标提取的影响

大多数的研究者在利用深度学习提取近岸养殖目标时忽略了先验知识,使得模型对养殖区域特征的学习完全依赖输入的影像^[32]。如刘继鹏等^[33]基于高分辨率遥感影像,只利用了原始的遥感波段通道数据提取海水养殖区域,致使该方法的适用性较低。后来为解决模型缺少先验知识这一问题,李龙坤等^[34]构建了养殖目标的纹理特征来作为先验知识,该方法减少了海水对筏式养殖区的识别干扰,与不添加先验知识的方法相比,添加先验知识后的方法总体精度提高了 3.86%,但该方法构建的先验知识较少。本研究中,利用 NDWI 水体指数和灰度共生矩阵法分别提取了近岸养殖目标的光谱特征和纹理特征来作为先验知识,这与上述研究中的构建先验知识的思想保持一致,与未添加先验知识的方法相比,添加先验知识后的模型总体精度提高了 1.8%。然而,当添加的先验知识过多时会导致模型的计算量变大,如何平衡先验知识与模型训练的难度是未来有待研究的问题。

4.2 置换注意力机制对模型分割性能的影响

引入注意力机制能让模型对目标区域分配更高的权重,使其能更好地关注目标区域,从而提高提取精度。如舒睿^[35]为解决传统钢材表面缺陷检测方法精度低、速度慢等问题,在 YOLOv5s 模型的基础上,引入了 NAM 注意力机制,使得模型能更加关注钢材表面缺陷区域,相较于原始模型,该方法有着更好的检测精度。李龙坤等^[34]通过添加 SE 通道注意力机制,解决了筏式养殖区在弱信号区提取精度低的问题。由于近岸养殖区域有着复杂的地物背景,受房屋、植被、海水和船只等因素的干扰较大,本研究中提出的 SA-STN-Net 模型是在 U-Net 网络基础上引入了置换注意力机制 SA,利用置换注意力机制的强大特征提取能力使网络能更加聚焦于近岸养殖目标区域,降低了复杂地物背景对近岸养殖区信息提取的干扰, $F1$ 分数和总体精度分别达到了 94.1% 和 96.4%。这表明,引入置换注意力机制能有效提高近岸水产养殖区信息提取精

度,但所提取的边缘仍存在提升空间,养殖区之间的粘连、错分问题仍未完全解决。

4.3 空间变换网络对模型分割性能的影响

引入空间变换网络能增强卷积神经网络的几何不变性和学习能力,提高模型分割性能。如冀中等^[30]提出一种基于空间变换双线性网络的鱼类图像分类方法,通过 STN 模块对输入的鱼类样本图进行空间变换,与先进模型 DeepFish 相比,识别正确率提高了 0.56%。袁帆^[36]提出了基于空间变换网络的人员行为识别方法,在传统的卷积神经网络中引入空间变换网络,使得网络对输入样本的变化因素不敏感,从而让深度学习网络更加适应输入数据所造成的变化。空间变换网络最大的优势在于其仿射变换能力,本研究中提出的 SA-STN-Net 模型是在 U-Net 网络基础上引入了 STN 模块,使得模型能对近岸养殖目标进行空间变换和对齐,增加了近岸养殖目标自身的空间特征信息,进一步提升了模型的性能。

5 结论

针对近岸水产养殖目标地物背景复杂,受房屋、植被、海水和船只等干扰造成提取精度较低等问题,本研究中利用 GF-2 遥感影像,构建了近岸养殖目标的先验知识,在 U-Net 模型的基础上提出了一种联合空间变换和置换注意力机制的近岸水产养殖区信息提取方法,结论如下:

1) 联合空间变换网络和置换注意力机制能使模型在充分利用养殖目标自身空间特征的同时聚焦于养殖目标区域,降低了复杂地物背景对近岸养殖区信息提取的干扰,从而提升了模型的性能。

2) 本研究中提出的 SA-STN-Net 模型可以有效提取近岸水产养殖区域, $F1$ 分数、总体精度和召回率分别达到了 94.1%、96.4% 和 87.5%,说明 SA-STN-Net 模型能适应近岸水产养殖目标复杂地物背景的情况,水产养殖区信息提取结果可为近岸规划与海域管理部门提供技术支持。

致谢:感谢高分辨率对地观测系统海南数据与应用中心提供高分辨率遥感数据支撑!

参考文献:

- [1] WANG F. Comments on marine resources and environment conditions in Beibugulf and strategic conception for development[J]. Resource Industry, 2000(1):37-41.
- [2] YUAN J J, XIANG J, LIU D Y, et al. Rapid growth in greenhouse gas emissions from the adoption of industrial-scale aquaculture

- [J]. *Nature Climate Change*, 2019, 9: 318-322.
- [3] GU Y G, LIN Q, JIANG S J, et al. Metal pollution status in Zhelin Bay surface sediments inferred from a sequential extraction technique, South China Sea [J]. *Marine Pollution Bulletin*, 2014, 81 (1): 256-261.
- [4] STENTIFORD G D, BATEMAN I J, HINCHLIFFE S J, et al. Sustainable aquaculture through the One Health lens [J]. *Nature Food*, 2020, 1(8): 468-474.
- [5] 程田飞, 周为峰, 樊伟. 水产养殖区域的遥感识别方法进展 [J]. *国土资源遥感*, 2012, 24(3): 1-5.
CHENG T F, ZHOU W F, FAN W. Progress in the methods for extracting aquaculture areas from remote sensing data [J]. *Remote Sensing for Land & Resources*, 2012, 24(3): 1-5. (in Chinese)
- [6] 吴岩峻, 张京红, 田光辉, 等. 利用遥感技术进行海南省水产养殖调查 [J]. *热带作物学报*, 2006, 27(2): 108-111.
WU Y J, ZHANG J H, TIAN G H, et al. A survey to aquaculture with remote sensing technology in Hainan Province [J]. *Chinese Journal of Tropical Crops*, 2006, 27(2): 108-111. (in Chinese)
- [7] 李轶平, 吴英超, 尤广然, 等. 基于 Sentinel-2 影像的围海养殖信息提取 [J]. *海岸工程*, 2022, 41(2): 173-180.
LI Y P, WU Y C, YOU G R, et al. Information extraction of sea-enclosed aquaculture based on Sentinel-2 remote sensing images [J]. *Coastal Engineering*, 2022, 41(2): 173-180. (in Chinese)
- [8] KANG J M, SUI L C, YANG X M, et al. Sea surface-visible aquaculture spatial-temporal distribution remote sensing: a case study in Liaoning Province, China from 2000 to 2018 [J]. *Sustainability*, 2019, 11(24): 7186.
- [9] CELIK T. Unsupervised change detection in satellite images using principal component analysis and k-means clustering [J]. *IEEE Geoscience and Remote Sensing Letters*, 2009, 6(4): 772-776.
- [10] KUSSUL N, LAVRENIUK M, SKAKUN S, et al. Deep learning classification of land cover and crop types using remote sensing data [J]. *IEEE Geoscience and Remote Sensing Letters*, 2017, 14 (5): 778-782.
- [11] 张晓鹏, 许志远, 曲胜, 等. 基于改进 YOLOv5 深度学习的海上船舶识别算法 [J]. *大连海洋大学学报*, 2022, 37(5): 866-872.
ZHANG X P, XU Z Y, QU S, et al. Recognition algorithm of marine ship based on improved YOLOv5 deep learning [J]. *Journal of Dalian Ocean University*, 2022, 37(5): 866-872. (in Chinese)
- [12] HU M, LI Y L, FANG L, et al. A2-fpn: attention aggregation based feature pyramid network for instance segmentation [C]//2021 IEEE/CVF conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA; IEEE, 2021: 15338-15347.
- [13] LI R, ZHENG S Y, ZHANG C, et al. Multi-attention-network for semantic segmentation of fine-resolution remote sensing images [J/OL]. *IEEE Transactions on Geoscience and Remote Sensing*, 2020. DOI: 10.13140/RG.2.2.28977.81761.
- [14] LIU Z, LIN Y T, CAO Y, et al. Swin Transformer: hierarchical vision transformer using shifted windows [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, QC, Canada; IEEE, 2021: 9992-10002.
- [15] WANG L B, LI R, DUAN C X, et al. A novel transformer based semantic segmentation scheme for fine-resolution remote sensing images [J/OL]. *IEEE Geoscience and Remote Sensing Letters*, 2021. DOI: 10.48550/arXiv.2104.12137.
- [16] 周涛, 董雅丽, 霍兵强, 等. U-Net 网络医学图像分割应用综述 [J]. *中国图象图形学报*, 2021, 26(9): 2058-2077.
ZHOU T, DONG Y L, HUO B Q, et al. U-Net and its applications in medical image segmentation: a review [J]. *Journal of Image and Graphics*, 2021, 26(9): 2058-2077. (in Chinese)
- [17] AKHTAR N, MANDLOI M. Dense Res Segnet: a dense residual segnet for road detection using remote sensing images [C]//2023 international conference on Machine Intelligence for Geo Analytics and Remote Sensing (MIGARS). Hyderabad, India; IEEE, 2023: 1-4.
- [18] SUI B K, JIANG T, ZHANG Z, et al. A modeling method for automatic extraction of offshore aquaculture zones based on semantic segmentation [J]. *ISPRS International Journal of Geo-Information*, 2020, 9(3): 145.
- [19] 刘滨洪. 基于卷积神经网络的水产养殖区提取 [J]. *测绘与空间地理信息*, 2023, 46(11): 74-77.
LIU B H. Extraction of aquaculture area based on convolutional neural network [J]. *Geomatics & Spatial Information Technology*, 2023, 46(11): 74-77. (in Chinese)
- [20] CUI B G, FEI D, SHAO G H, et al. Extracting raft aquaculture areas from remote sensing images via an improved U-Net with a PSE structure [J]. *Remote Sensing*, 2019, 11(17): 2053.
- [21] 张欣, 戴佩玉, 李卫国, 等. 基于改进坐标注意力和 U-Net 神经网络的淡水养殖区提取 [J]. *农业工程学报*, 2023, 39(17): 153-162.
ZHANG X, DAI P Y, LI W G, et al. Extracting the images of freshwater aquaculture ponds using improved coordinate attention and U-Net neural network [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2023, 39(17): 153-162. (in Chinese)
- [22] ZHANG Q L, YANG Y B. SA-net: shuffle attention for deep convolutional neural networks [C]//ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Toronto, ON, Canada; IEEE, 2021: 2235-2239.
- [23] JADERBERG M, SIMONYAN K, ZISSERMAN A, et al. Spatial transformer networks [EB/OL]. 2015; arXiv: 1506.02025. <http://arxiv.org/abs/1506.02025>.
- [24] 潘腾. 高分二号卫星的技术特点 [J]. *中国航天*, 2015(1): 3-9.
PAN T. Technical characteristics of Gaofen-2 satellite [J]. *Aerospace China*, 2015(1): 3-9. (in Chinese)
- [25] SUN W H, CHEN B, MESSINGER D. Nearest-neighbor diffusion-based pan-sharpening algorithm for spectral images [J]. *Optical Engineering*, 2014, 53(1): 013107.
- [26] 巫统仁, 刘培, 于吉涛, 等. 平衡剖面模式下海岸线遥感提取中不同潮位校正方法对比研究 [J]. *热带海洋学报*, 2023, 42(6): 52-62.
WU T R, LIU P, YU J T, et al. Comparative study on different tide level correction methods in remote sensing extraction of coastline

- under balanced profile model[J]. *Journal of Tropical Oceanography*, 2023, 42(6): 52-62. (in Chinese)
- [27] DAI J F, QI H Z, XIONG Y W, et al. Deformable convolutional networks[C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017: 764-773.
- [28] WANG Q L, WU B G, ZHU P F, et al. ECA-Net: efficient channel attention for deep convolutional neural networks [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020: 11531-11539.
- [29] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 7132-7141.
- [30] 冀中, 赵可心, 张锁平, 等. 基于空间变换双线性网络的细粒度鱼类图像分类[J]. *天津大学学报(自然科学与工程技术版)*, 2019, 52(5): 475-482.
- Ji Z, ZHAO K X, ZHANG S P, et al. Fine-grained fish image classification based on a bilinear network with spatial transformation [J]. *Journal of Tianjin University (Science and Technology Edition)*, 2019, 52(5): 475-482. (in Chinese)
- [31] PASZKE A, GROSS S, MASSA F, et al. PyTorch: an imperative style, high-performance deep learning library [EB/OL]. 2019: arXiv:1912.01703. <http://arxiv.org/abs/1912.01703>. pdf.
- [32] LU Y M, SHAO W, SUN J. Extraction of offshore aquaculture areas from medium-resolution remote sensing images based on deep learning[J]. *Remote Sensing*, 2021, 13(19): 3854.
- [33] 刘继鹏, 王常颖, 初佳兰. 基于 U-Net 的国产高分卫星影像海水养殖区分类提取方法[J]. *海洋环境科学*, 2023, 42(3): 471-482.
- LIU J P, WANG C Y, CHU J L. Classification and extraction method of mariculture area from domestic high-resolution satellite images based on U-Net model [J]. *Marine Environmental Science*, 2023, 42(3): 471-482. (in Chinese)
- [34] 李龙坤, 蔡玉林, 徐慧宇, 等. 基于卷积神经网络和数据融合的筏式养殖区提取[J]. *海洋学报*, 2023, 45(8): 155-165.
- LI L K, CAI Y L, XU H Y, et al. Extraction of the raft aquaculture area based on convolutional neural networks and data fusion [J]. *Haiyang Xuebao*, 2023, 45(8): 155-165. (in Chinese)
- [35] 舒睿. 添加注意力机制的 YOLOv5s 算法对带钢表面缺陷检测[J]. *农业装备与车辆工程*, 2023, 61(12): 148-151, 160.
- SHU R. Steel surface defect detection based on improved YOLOv5s algorithm [J]. *Agricultural Equipment & Vehicle Engineering*, 2023, 61(12): 148-151, 160. (in Chinese)
- [36] 袁帆. 基于空间变换网络的人员行为识别方法[J]. *武汉大学学报(工学版)*, 2022, 55(7): 740-746.
- YUAN F. Spatial transformation network based human activity recognition method [J]. *Engineering Journal of Wuhan University*, 2022, 55(7): 740-746. (in Chinese)

An information extraction method from offshore aquaculture area based on spatial transformation and shuffle attention mechanism

WU Tongren^{1,2,3}, ZHANG Xian³, LIU Pei^{1,2,4*}, WEN Tingting^{1,2,3}, ZOU Zhenxue^{1,2,3}

(1. Yazhou Bay Innovation Institute, Hainan Tropical Ocean University, Sanya 572025, China; 2. Hainan Academy of Ocean and Fisheries Sciences, Haikou 571126, China; 3. School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo 454003, China; 4. School of Remote Sensing and Information Engineering, North China Institute of Aerospace Engineering, Langfang 065000, China)

Abstract: In order to solve the problem of low accuracy in extracting aquaculture areas using remote sensing technology due to complex background of onshore aquaculture and offshore cage culture areas and to easily disturbed by factors including houses, vegetation, seawater, and ships, a complex deep learning method that combines shuffle attention mechanisms and spatial transformation network was proposed, and tested in Bamen bay in Wenchang city and Potou Port in Wanning City. With the help of GF-2 high resolution remotely sensed data, the prior knowledge of aquaculture targets was constructed using spectral and texture features. Then, based on the U-Net model, the spatial transformation network (STN) and the shuffle attention (SA) mechanism are combined to enhance the spatial characteristics of the aquaculture area and to reduce the interference of complex backgrounds. The test results showed that the overall accuracy and mean intersection over union of SA-STN-Net model were enhanced by 3.3% and 5.7% compared with the preliminary U-Net model, respectively. Swin-Transformer, Dc-Swin, and F1 score of SA-STN-Net model were found to be increased by 6.7%, 4.2% and 7.2% in the score compared with the most state-of-art deep learn model such as A2fpn, respectively. The findings demonstrate that the proposed SA-STN-Net model is adapted to the complex environmental background of offshore aquaculture, effectively extracts offshore aquaculture targets, and can provide technical support for offshore planning and management departments.

Key words: spatial transformer network (STN); shuffle attention (SA); deep learning; extraction of aquaculture area