

王岚, 魏皓, 车亚辰, 等. 融合类别数量自适应深度数据增强和迁移学习的造礁珊瑚识别方法研究[J]. 海洋学报, 2024, 46(9): 120–130, doi:10.12284/hyxb2024096

Wang Lan, Wei Hao, Che Yachen, et al. Integration of category-quantity adaptive deep data augmentation and transfer learning for reef-building coral recognition[J]. Haiyang Xuebao, 2024, 46(9): 120–130, doi:10.12284/hyxb2024096

# 融合类别数量自适应深度数据增强和迁移学习的 造礁珊瑚识别方法研究

王岚<sup>1</sup>, 魏皓<sup>1</sup>, 车亚辰<sup>2</sup>, 张翠翠<sup>1,3\*</sup>

(1. 天津大学 海洋科学与技术学院 天津 300072; 2. 国家海洋技术中心 自然资源部海洋观测技术重点实验室 天津 300112; 3. 中国科学院 计算技术研究所 北京 100190)

**摘要:** 造礁珊瑚识别对于珊瑚礁生态系统的保护与监测具有重要意义。深度学习作为图像识别的前沿技术,在珊瑚识别领域逐渐得到应用。然而,其识别性能仍然面临挑战。其中,数据集中类别间样本数量不平衡和数据多样性欠缺是两个主要问题。前者使得深度学习模型在特征提取过程中更偏向于样本数较多的类,对少数类(尤其是濒危珊瑚)的学习能力不足进而影响其识别准确度。后者因为数据缺乏多样性使得模型无法充分学习各种珊瑚特征,进而限制了特征提取的能力。鉴于此,本文提出了一种融合类别数量自适应深度数据增强和迁移学习的造礁珊瑚类型识别方法。针对第一个问题,本文利用识别结果评价指标  $F_1$ -score 定义的数据生成量化公式对原始深度数据增强方法 DeepSMOTE 进行改进,提出了类别数量自适应的深度数据增强方法 DeepSMOTE- $F_1$ 。该方法根据每类珊瑚的识别结果自适应地增强其样本数量,确保模型充分学习各类珊瑚特征。针对第二个问题,利用迁移学习强化了模型的提取能力。实验结果表明,在 RSMAS、EILAT 和 EILAT2 这 3 个代表性珊瑚识别数据集上,相较于原始 DeepSMOTE,本文提出的 DeepSMOTE- $F_1$  识别准确率分别提升了 2.88%、0.39% 和 1.54%;与现有的珊瑚智能识别方法相比,准确率分别提升了 0.76%、1.40% 和 1.30%。

**关键词:** 珊瑚识别;深度学习;数据集不平衡;数据增强;迁移学习

中图分类号: P714<sup>+.5</sup>

文献标志码: A

文章编号: 0253-4193(2024)09-0120-11

## 1 引言

尽管珊瑚礁的面积不到海洋总面积的 0.2%,但却为超过 25% 的海洋生物提供了栖息地<sup>[1]</sup>。珊瑚礁多样的生物和复杂的环境构成了一个具有极高初级生产力的生态系统,对于维持海洋生态系统的健康、实现海洋资源的可持续开发、保护海岸线、提供资源和支持国防建设等起着重要作用<sup>[2]</sup>。造礁珊瑚作为珊瑚礁生态系统的主要构建者,其生存状况与珊瑚礁生态

系统密切相关<sup>[3]</sup>。然而,受全球气候变化、日益加剧的人类活动以及珊瑚病敌害等多重不利因素的共同影响,造礁珊瑚正面临着严重威胁<sup>[4]</sup>。根据国际自然与自然资源保护联盟(International Union for Conservation of Nature, IUCN)2022 年全球濒危物种红色目录报告显示,全球约 36% 的造礁珊瑚种类濒临灭绝<sup>[5]</sup>。为更好地监测珊瑚礁生态系统的健康状况,需要对各类造礁珊瑚进行识别。目前,这项工作主要依赖于人工识别与传统机器学习的方法。人工识别费时费力,

收稿日期: 2023-08-02; 修订日期: 2024-05-08。

基金项目: 国家重点研发计划项目(2022YFC3104600); 海南省重点研发计划项目(ZDYF2024SHFZ051); 国家自然科学基金面上项目(42076007); 自然资源部海洋观测技术重点实验室定向基金(klootB06)。

作者简介: 王岚(1997—),女,四川省绵阳市人,主要从事图像识别研究。E-mail: 17721910924@163.com

\* 通信作者: 张翠翠(1986—),女,山东省滨州市人,副教授,主要从事智能海洋计算、模式识别研究。E-mail: cuicui.zhang@tju.edu.cn

极大地限制了珊瑚研究的空间和时间范围。例如综合海洋观测系统(Integrated Marine Observing System, IMOS)每年都会收集澳大利亚周围数百万张珊瑚礁的图像,但只有不到5%的图像被分析<sup>[6]</sup>。传统机器学习的方法由于采用浅层学习框架,特征提取能力相对较弱。因此,在水下图像识别方面的性能较低<sup>[7-8]</sup>。

近年来,随着人工智能技术的发展,利用深度学习进行珊瑚识别已经成为可能。深度学习通过设定好的网络结构,能够从大量标注的图像数据中学习珊瑚的层级结构性特征,以提取更接近高级语义的抽象特征,从而推动了自动化珊瑚识别技术的进步。2015年,Elawady等<sup>[9]</sup>首次将卷积神经网络(Convolutional Neural Network, CNN)引入珊瑚识别研究。Mahmood等<sup>[10]</sup>提出了一种改进的残差神经网络ResFeats用于珊瑚识别。Modasshir等<sup>[11]</sup>采用DenseNet层间的密集连接以及多尺度的思想,以提取更具判别性的珊瑚特征。Gómez-Ríos等<sup>[12]</sup>对3种CNN模型(Inceptionv3、ResNet和DenseNet)在珊瑚数据集上的识别性能进行了对比分析。Lumini等<sup>[13]</sup>提出了多个CNN集成的方法,与单个模型相比多个CNN进行集成能显著提高识别精度。上述研究证明了深度学习在自动化珊瑚识别方面的可行性,但目前这项工作仍面临一定挑战。一方面,现有数据集中不同珊瑚样本数量间存在的失衡会影响模型训练。识别模型试图在训练过程中最小化总体识别误差,并且隐式地为所有类分配相同的错分代价,导致模型倾向于正确识别样本数较多的类<sup>[14]</sup>,这种情况下,样本数较少的类的特征无法被充分学习(尤其是濒危珊瑚),容易造成错分。另一方面,数据集小且缺乏多样性限制了深度学习方法在珊瑚识别方面的效果。受珊瑚类内多样性和类间相似性,拍摄角度以及水下环境因素的影响<sup>[6]</sup>,目前用于珊瑚自动化识别的公开数据集不足以完全覆盖现实世界中珊瑚的多样性特征且图像质量不高。此外,数据规模较小,无法充分训练一个包含成千上万参数的模型。这些因素共同导致了模型特征提取能力的不足,进而限制了其识别精度。

数据增强通常被用于解决数据集不平衡的问题。它在不改变样本类别标签的情况下,通过对原始样本进行一系列变换(如几何、颜色以及像素变换)来生成新的样本<sup>[12, 15]</sup>。由于图像间的差异非常复杂,传统方法容易导致数据冗余和过拟合,无法满足现代深度学习模型对多样化数据的需求。Dablain等<sup>[16]</sup>构建了一个基于深度学习的数据增强模型DeepSMOTE,它可以保留图像的原始属性并生成高质量、信息丰

富的新图像。然而,该方法未充分考虑每个类采样的具体数量,导致模型无法针对性地学习特征较为模糊或具有相似特征类别,使得这些类别的识别结果不理想。基于此,本文提出了一种融合类别数量自适应深度数据增强和迁移学习的造礁珊瑚类型识别方法。针对数据集中样本数量不平衡这一问题,本文利用识别结果评价指标 $F_1$ -score<sup>[17]</sup>定义的数据生成量化公式对原始的深度数据增强方法DeepSMOTE<sup>[18]</sup>进行改进,提出了类别数量自适应的深度数据增强方法DeepSMOTE- $F_1$ 。DeepSMOTE- $F_1$ 能够根据识别结果自适应地增强各珊瑚样本,使得模型更充分地学习每个造礁珊瑚类型的特征。同时,为进一步改善数据集小且缺乏多样性对模型识别精度的限制,本文引入了迁移学习<sup>[19]</sup>,将模型在大型数据集上学习到的特征提取能力应用于珊瑚识别上。本文利用该方法在3个不同的公开珊瑚识别数据集上进行了敏感性实验和对比试验,以证明该方法的有效性。

## 2 数据来源

本文使用3个公开的用于自动化珊瑚识别的数据集:RSMAS、EILAT和EILAT2<sup>[20]</sup>来验证所提出的方法的有效性。这些数据集由专业人员所拍摄,并经过统一标注和预处理。它们在珊瑚识别领域被研究人员所广泛使用<sup>[9-13]</sup>。这3个数据集均为彩色图像,具有数据量小和类别间样本数量不平衡的特点。这些数据集中的图像主要捕捉珊瑚不同部位的形状、颜色和纹理信息,不包含造礁珊瑚的完整结构信息。

### (1) RSMAS 数据集

RSMAS数据集是由迈阿密大学罗森斯蒂尔海洋和大气科学学院的潜水员拍摄的图像裁剪得到,共包含766个大小为 $256 \times 256$ 像素的图像。该数据集共有14个类别,各类别名称和样本数量如表1所示。在RSMAS数据集中,*Montipora* spp.(蔷薇珊瑚属)未细分至具体的种,而其他造礁珊瑚的类型划分已达到了种的层面<sup>[3, 21]</sup>。此外,Gorgonians(柳珊瑚)不属于造礁珊瑚,且作为一个概括性名称,未被细分至具体的种属。同时,*Diadema antillarum*、*Millepora alcicornis*、*Sponge fungus*和*Tunicates*分别属于长刺海胆、多孔螅、海绵真菌和被囊动物,均不在珊瑚类型的范畴之内<sup>[21-23]</sup>。

### (2) EILAT 数据集

EILAT数据集共1123张,大小为 $64 \times 64$ 像素。这些图像是在相同条件下,用相同的相机拍摄于红海埃拉特岛附近。EILAT共包含8个类,每个类别的名

表 1 RSMAS 数据集基本信息<sup>[21]</sup>

Table 1 The information of the RSMAS dataset

数据集	类	数量
RSMAS	<i>Acropora cervicornis</i>	109
	<i>Acropora palmata</i>	77
	<i>Colpophyllia natans</i>	57
	<i>Diadema antillarum</i>	63
	<i>Diploria strigosa</i>	24
	<i>Gorgonians</i>	60
	<i>Millepora alcicornis</i>	22
	<i>Montastraea cavernosa</i>	79
	<i>Meandrina meandrites</i>	54
	<i>Montipora</i> spp.	28
	<i>Palythoa palythoa</i>	32
	<i>Sponge fungus</i>	88
	<i>Siderastrea siderea</i>	37
	<i>Tunicates</i>	36
	总数	766

称和数量如表 2 所示。EILAT 对各珊瑚类型的划分较为粗糙, 未达到种属层面。此外该数据集中还包含非珊瑚: Sand(沙)、Urchin(海胆)。

表 2 EILAT 数据集基本信息<sup>[21]</sup>

Table 2 The basic information of the EILAT dataset

数据集	类	数量
EILAT	Sand	87
	Urchin	80
	Dead Coral	280
	Brain Coral	160
	Favid Coral	200
	Branches Type I	23
	Branches Type II	216
	Branches Type III	77
	总数	1 123

### (3) EILAT2 数据集

EILAT2 数据集共 304 张, 大小为  $128 \times 128$  像素, 和 EILAT 一样它也拍摄于红海埃拉特岛附近。EILAT2 只包含 5 个类, 与 EILAT 相比, 它没有死珊瑚类并且将树枝型珊瑚归为一个类。每个类别的名称和数量

如表 3 所示。与 EILAT 类似, EILAT2 对珊瑚类型的划分还未达到种属层面, 并包含两类非珊瑚: Sand(沙)、Urchin(海胆)。

表 3 EILAT2 数据集基本信息<sup>[24]</sup>

Table 3 The basic information of the EILAT2 dataset

数据集	类	数量
EILAT2	Sand	80
	Urchin	14
	Brain Coral	71
	Favid Coral	89
	Branches Type	49
总数	303	

这 3 个数据集均存在一定的不平衡性, 由公式 (1)<sup>[25]</sup> 计算得到 RSMAS、EILAT、EILAT2 数据集的不平衡率分别为 4.95、12.17、6.36。

$$\rho = \frac{\text{Max}\{c_i\}}{\text{Min}\{c_j\}}, \quad (1)$$

式中:  $c_i$  为数据集中珊瑚类别;  $\text{Max}\{c_i\}$  表示数量最多的类的图像数量;  $\text{Min}\{c_j\}$  表示数量最少的类的图像数量。

EILAT 数据集的不平衡率最高, 这是由于 Dead Coral(死珊瑚)的样本数量与其他类别相差过大, 死珊瑚的样本数占总数的 24.93%, 而最少的类 Branches Type I 仅占 2.05%; 其次是 EILAT2, 它的不平衡率源自于非珊瑚类, 而各个珊瑚类间的样本数量相差不大; RSMAS 数据集的不平衡率最低, 样本数量最多的类是 *Acropora cervicornis*(摩羯鹿角珊瑚), 占总数的 14.23%, 样本数量最少的类为 *Millepora alcicornis*(多孔螅), 占总数的 2.87%。

## 3 融合类别数量自适应深度数据增强和迁移学习的造礁珊瑚类型识别方法

### 3.1 造礁珊瑚识别骨干网络: ResNet-50

本研究采用 ResNet-50 作为造礁珊瑚识别的骨干网络。与传统神经网络直接学习期望的映射  $H(x)$  不同, ResNet-50 通过引入残差块的设计, 对网络的映射方式进行了改进。具体来说, ResNet 引入了残差映射  $F(x) = H(x) - x$ , 其中  $x$  表示输入。这种设计使得网络可以通过学习残差的方式, 将输入与期望输出之间的差异学习到残差块中, 而不是直接学习整个映射。这一思想有效地解决网络随着层数增加而出现的梯度消失和梯度爆炸问题。为更好地适应珊瑚识别任务,

本研究对其结构进行了修改,将原有的全连接层替换为两个不同结构的全连接层,并分别在其后面连接激活函数。具体结构如图1所示,第一个全连接层包含512个神经元,并连接一个ReLU激活函数。第二个全连接层的神经元个数与数据集的类别相同,也就是

说在进行RSMAS识别时,神经网络的全连接层有14个神经元;在进行EILAT识别时,有8个神经元;在进行EILAT2识别时,有5个神经元。最后,该层再连接一个Softmax激活函数。ReLU和Softmax激活函数的引入可以增加网络的非线性拟合能力<sup>[26]</sup>。

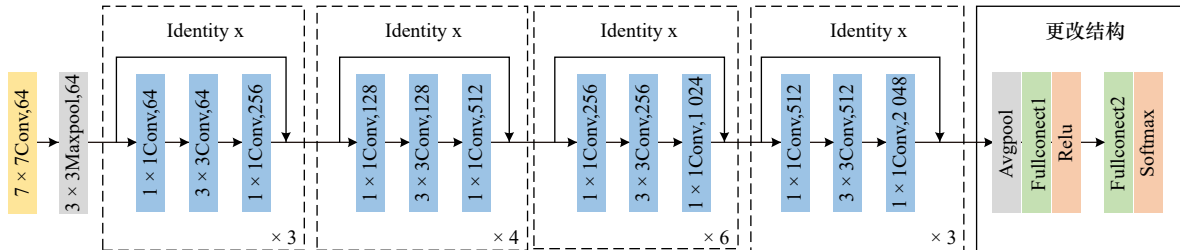


图1 更改后的 ResNet-50 网络结构

Fig. 1 The network structure of the modified ResNet-50

### 3.2 原始深度数据增强方法: DeepSMOTE

原始的深度数据增强方法 DeepSMOTE<sup>[16]</sup> 主要包含3个模块:编码器、SMOTE特征采样和解码器。编码器负责将图像映射为高维特征空间中的特征向量,这些向量捕捉了图像的重要信息。随后,SMOTE在特征空间中对图像的特征进行采样。最终,解码器将这些新的特征向量解码为新图像实现数据增强。DeepSMOTE的核心环节是SMOTE<sup>[27]</sup>,它通过在现有实例与最近邻实例间插入新的点来创建新实例。具体步骤如下:首先,对于某一类中的每一个样本 $x$ ,以欧氏距离为标准计算它到该类样本集中所有样本的距离,并得到其 $k$ 个近邻。然后,对于样本 $x$ ,从其 $k$ 近邻中随机选择1个样本作为近邻 $\tilde{x}$ 。最后,根据公式(2)构建新的样本 $x_{new}$ ,其中公式(2)的系数 $\text{rand}(0,1)$ 被称为比例因子,该值是随机的,介于0和1之间,用于决定新样本 $x_{new}$ 与 $x$ 和 $\tilde{x}$ 的相似程度。图2展示了具体过程。由于邻近点的特征是比较相似的,因此合成的新样本基本符合原始数据分布。

$$x_{new} = x + \text{rand}(0,1) \times (x - \tilde{x}). \quad (2)$$

DeepSMOTE考虑了样本之间的相似度,利用深

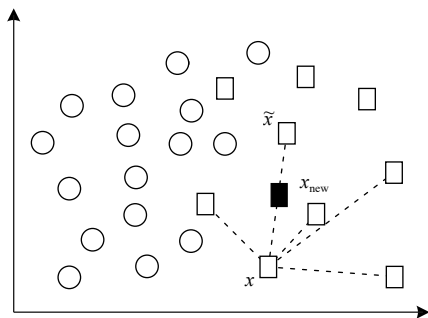


图2 SMOTE采样示意图

Fig. 2 The diagram of SMOTE sampling

度学习模型自动生成新样本,确保所生成的图像既能保留原始图像的属性,又具有高视觉质量和丰富的信息。图3展示了在RSMAS和EILAT数据集<sup>[20]</sup>上利用DeepSMOTE深度数据增强所生成的珊瑚图像。可以看出,SMOTE中的不同比例因子导致了不同效果的输出图像:一部分更类似于原始图像,另一部分更类似于最邻近图像,还有一部分同时具有两种图像的特征。

### 3.3 类别数量自适应深度数据增强方法: DeepSMOTE-F<sub>1</sub>

虽然DeepSMOTE对于深度学习模型是一种较为有效的数据增强方法,但它仍然面临一个重要难题,即如何确定每个类的采样数量。通常情况下,研究人员会选择一个阈值来决定该数量,以实现数据集的完全平衡。这个阈值通常选取最大类的样本数量或者介于最大类和最小类样本数量之间的某个人为设定值。这种方法虽然平衡了数据集,使得识别算法不会明显偏向某个类,但对于一些特征较为模糊或具有相似特征难以识别的类来说,数据平衡可能会造成负面影响。数据平衡后,模型对每个类的关注程度几乎相同,但并不意味着每个类特征学习的难易程度相同。因此,模型无法有针对性地学习难以识别的类的特征,导致这些类的识别结果仍然不理想。同时,如果阈值介于最大类别和最小类别数量之间,意味着会随机移除一些样本,这对于数据紧缺的珊瑚数据集来说非常不利,会导致信息的流失。

在识别模型训练过程中,识别算法以样本数量基本均衡作为前提,并以总体识别准确率作为目标,而这对不平衡数据集并不适用。比如,给定一个正样本只占1%的不平衡数据集,当识别器将所有示例标记为负样本,其识别准确率能高达99%,但是这个值并

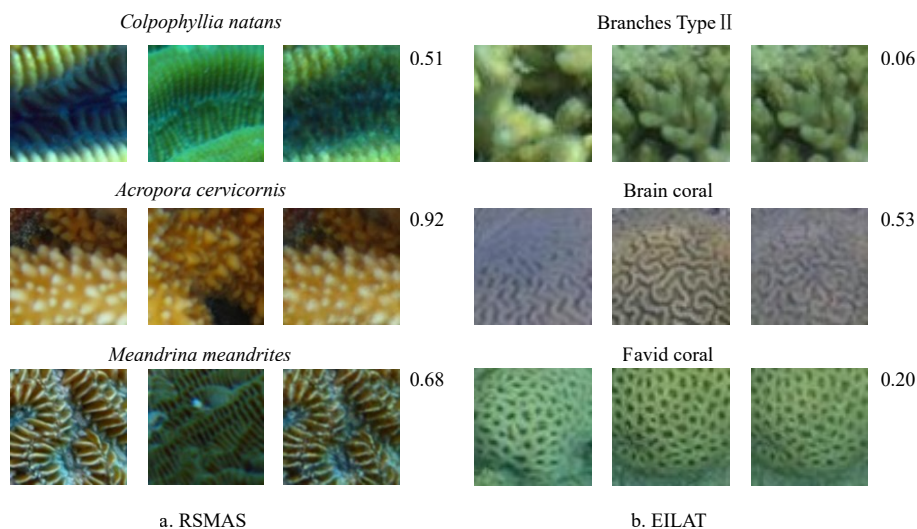


图3 DeepSMOTE 深度数据增强结果示例: a 中展示了 RSMAS 数据集<sup>[20]</sup> 中 3 个珊瑚类: *Colpophyllia natans*, *Acropora cervicornis* 和 *Meandrina meandrites*, b 中展示了 EILAT 数据集<sup>[20]</sup> 中 3 个珊瑚类: Branches Type II, Brain Coral 和 Favid Coral。每个类有 3 个示例, 从左到右依次为原始图像、近邻图像、新图像。右上角的数值是由 SMOTE 算法随机确定的比例因子, 介于 0~1, 表示新的图像与原始图像和近邻图像的接近程度。值越大则代表越接近原始图像, 值越小则代表越接近近邻图像

Fig. 3 The example results of deep data augmentation using DeepSMOTE: a. displays three classes of coral in the RSMAS dataset<sup>[20]</sup>: *Colpophyllia Natans*, *Acropora Cervicornis* and *Meandrina Meandrites*, b. displays three classes of coral in the EILAT dataset<sup>[20]</sup>: Branches Type II, Brain Coral 和 Favid Coral. There are three examples provided for each class. The 1st–3rd column shows the original image, the nearest neighbor image, and the generated image, respectively. The value in the top right corner is the scale factor randomly determined by the SMOTE, which ranges from 0 to 1. It indicates the degree of similarity between the generated image and the original image, as well as the nearest neighbor image. A higher value indicates greater similarity with the original image, while a lower value indicates greater similarity with the nearest neighbor image

不能真实反映识别器的价值。因此, 当数据集不平衡时, 在保证总体识别准确率的前提下提升每个类的识别效果比提升总体识别准确率更有意义。在本研究中, 我们重点关注模型对每个珊瑚类正确识别的能力。我们希望同时避免将某一类珊瑚错误地归类为其他类 (FN) 以及将其他类珊瑚错误地归类为该类 (FP)。F<sub>1</sub>-score 通过赋予召回率 (Recall) 和精确率 (Precision) 相等的权重, 综合考虑了 FP 和 FN 这两种情况, 见公式(3)~(5)。其中 Recall 表示被正确识别的正样本占实际正样本的比例, Precision 表示被正确识别的正样本占所预测的正样本的比例, 两者从不同维度描述了模型的性能。

我们根据 F<sub>1</sub>-score 来评估模型对每个类的识别情况, 以此来调整训练样本的数据分布, 从而间接影响模型训练过程。低 F<sub>1</sub>-score 表示模型在这个类别上的性能较差, 为了改善该类别的表现, 我们增加其训练样本数量, 使得模型有更多的机会学习其特征。相反, 如果 F<sub>1</sub>-score 较高, 说明模型能良好识别该类珊瑚, 因此在数据增强阶段无需对该类珊瑚过多采样, 以便模型更专注于其他识别结果较差的类。

$$F_1 - \text{score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (3)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (4)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \quad (5)$$

基于上述思想, 为解决 DeepSMOTE 采样数量难以确定的问题, 我们借助识别算法中的重要综合指标 F<sub>1</sub>-score, 提出了一种 F<sub>1</sub>-score 敏感的采样数量计算公式 [如公式(6)所示]。该公式根据初始识别准确度结果, 为每个类“量身定做”其需要增加的样本数量, 从而最大效能的兼顾每个类的识别效果。

$$\text{num} = \frac{1 - f_i}{\sum_{k \in n} 1 - f_k} \times N, \quad (6)$$

其中: num 表示每一类需采样的数量, *i* 代表类别, *n* 为数据集的类别数, *f<sub>i</sub>* 为每一类的 F<sub>1</sub>-score, 1 - *f<sub>i</sub>* 则表示类 *i* 的错误率, *N* 为原始数据集的总样本数量。公式(6)中, *N* 前面的系数表示类 *i* 的错误率与所有类错误率总和的比值。通过该比值, 我们可以确定类 *i* 需调整的样本数量的比例。将该比值乘以原始数据集中的样本总数 *N*, 即可得到类 *i* 需采样的数量。由于公式

中所有项都为正,因此不存在欠采样而导致信息丢失的情况发生。

我们将公式(6)与DeepSMOTE相结合,称为DeepSMOTE- $F_1$ 。首先,在训练集上对未经参数初始化的网络进行训练,并计算各个类的 $F_1$ -score。然后,通过

公式(6),计算出每个类别相应的采样数量。最后,对训练集进行深度数据增强。具体流程如图4所示。通过DeepSMOTE- $F_1$ ,我们对训练样本的分布进行了调整,使模型能够更加有针对性地学习各个类的特征,提高模型在珊瑚识别任务中的性能。

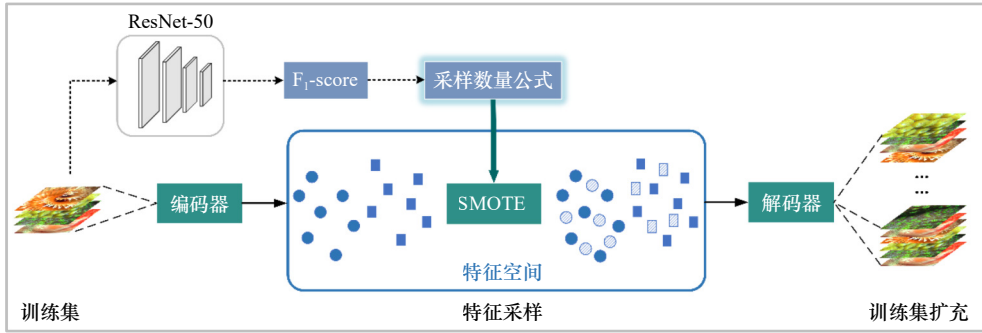


图4 DeepSMOTE- $F_1$  流程

Fig. 4 Flow chart of the DeepSMOTE- $F_1$

DeepSMOTE- $F_1$ 方法是一种基于评价指标的采样策略,其根据每一类珊瑚的原始识别结果 $F_1$ -score来计算采样数量。由于 $F_1$ -score可以直接从已知数据中计算得出,且原始样本数量也是已知的,因此DeepSMOTE- $F_1$ 不需要依赖于特定数据集的特性。这意味着无论数据集的分布如何,该方法都能自动地调整采样策略,使得模型能够充分学习每个珊瑚类型的特征,从而确保了模型的适用性和泛化能力。相比传统的采样策略,基于评价指标的采样方法DeepSMOTE- $F_1$ 更具普适性,为广泛存在数据集不平衡问题的珊瑚识别任务提供了一个有效的解决方案。

### 3.4 迁移学习

迁移学习是指将先前在某一领域中所学的知识与技能,有效地应用于与之不同但具有相关性的新领域中的一种学习方法<sup>[19]</sup>。本研究所采用的数据集RSMAS、EILAT和EILAT2均为小型数据集,图像数量分别为766张、1123张和303张。从深度学习在识别领域的效果来看,较深的网络能够提取语义更加丰富的高级特征,从而提高识别精度,但在小数据集上难以充分训练一个包含上万个参数的深度神经网络。另外,受数据多样性不足的影响,深度学习算法无法充分学习珊瑚的各种特征,导致识别准确度偏低。基于此,本文根据本研究数据集的特点,采用来自大型公共数据集ImageNet的迁移学习。ImageNet是一个包含超过100万张图像,涵盖了1000个类别的超大型数据集<sup>[28]</sup>。它虽与珊瑚图像数据集有一定差别,但是图像的边、线、颜色等低层特征提取的能力具有一定的通用性。因此,将官方在ImageN-

et上训练好的模型参数(除全连接层)迁移至更改结构后的ResNet-50(图1)中,实现网络参数的初始化,并根据珊瑚识别任务的需求对网络进行微调。

### 3.5 整体方法框架

本研究所提出的融合类别数量自适应深度数据增强和迁移学习的珊瑚识别方法的总体框架如图5所示。我们将数据集按8:2划分为训练集和测试集。首先,在训练集上对基础骨干网络ResNet-50进行训

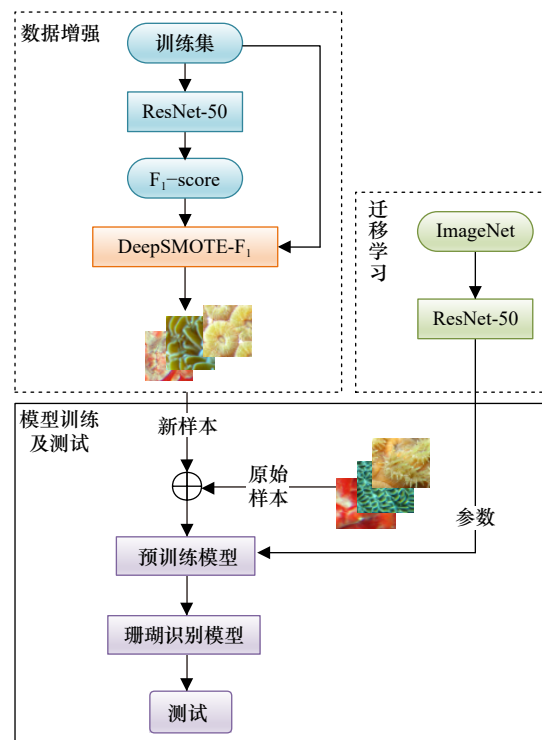


图5 整体方法框架

Fig. 5 The framework of the holistic method

练,得到各个类的  $F_1$ -score,并利用 DeepSMOTE- $F_1$  对训练集中各类别珊瑚样本进行深度数据增强。其次,使用来自 ImageNet 的迁移学习初始化基础骨干网络 ResNet-50 除最后两层全连接层外的所有网络参数,得到用于珊瑚识别的预训练模型。然后,在扩充后的训练集(新样本与原始样本的集合)上对预训练模型的参数进行优化,从而得到最终的珊瑚识别模型。最后,在测试集上评估模型的性能。在数据增强阶段,测试集仅被用于根据  $F_1$ -score 来计算采样数量,并未对测试集进行数据增强。这样做可以确保测试集保持与真实数据一致的分布,从而更准确地评估模型的泛化能力。

## 4 实验结果与分析

### 4.1 实验设置

为避免模型过大、内存占用较多以及训练时间增加的问题,本文将 RSMAS 数据集的图像大小调整为  $128 \times 128$  像素,而 EILAT 和 EILAT2 数据集的图像大小则保持不变,分别为  $64 \times 64$  像素和  $128 \times 128$  像素。

本实验采用 PyTorch 深度学习平台构建模型,基于 Python 3.9 版本。在训练过程中,我们分别设置了 3 组不同的参数,如表 4 所示。通过对比不同参数组合下的训练结果来评估各参数对模型稳定性、泛化能力以及训练效率的具体影响,进而找到最优的参数配置。为了优化训练过程,我们选择随机梯度下降(Stochastic Gradient Descent, SGD)作为优化算法,设置衰减率为  $10^{-6}$ ,同时启用 Nesterov 动量,并将其值设定为 0.9,以提升训练的稳定性和收敛速度。

表 4 参数设置

Table 4 The setting of parameters

模型	批次大小	学习率	迭代次数
ResNet-50	32	0.001/0.000 1/0.000 01	300/500/1 000
	64	0.001/0.000 1/0.000 01	300/500/1 000
	128	0.001/0.000 1/0.000 01	300/500/1 000

### 4.2 评价指标

本文所有的结果都是通过五折交叉验证获得,我们使用五次结果的平均值来比较和分析不同方法的性能。准确率[如公式(7)所示]是评估识别结果最常用的指标,表示模型识别正确的样本数与总样本数的比值。但对于不平衡数据集,准确率的说服力有限。因此,为更全面地评估模型性能,我们还采用了  $F_1$ -score[如公式(3)所示]。

$$\text{Accuracy} = \frac{\text{TP} + \text{FN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}. \quad (7)$$

### 4.3 敏感性实验

为验证本文所提方法的有效性和可靠性,我们分别在 3 个不同大小、珊瑚类型划分层次以及不平衡率的数据集 RSMAS、EILAT 和 EILAT2 上进行了敏感性实验。我们对表 4 的批次大小、学习率和迭代次数随机组合进行训练,选择最佳识别准确率,最终结果如表 5 所示。通过对比表 5 的第一行、第二行和第三行的结果,我们发现,相较于基线模型,采用 DeepSMOTE 和 DeepSMOTE- $F_1$  数据增强后,识别准确率均有一定提升,后者的提升幅度更大。这表明在处理不平衡数据时,与传统采样直至数据集达到平衡的方法相比,基于评价指标的采样策略能够取得更好的效果。进一步对比第三行和第四行的结果,可以看出 DeepSMOTE- $F_1$  和迁移学习都能够提升识别准确率,而迁移学习的效果更为显著。最后,相比于其他方法,本文提出的融合 DeepSMOTE- $F_1$  与迁移学习的识别方法在 3 个珊瑚数据集上均取得了最佳结果。具体而言,在具有较高不平衡率( $\rho_{\text{EILAT}} = 12.17$ ;  $\rho_{\text{RSMAS}} = 6.36$ )且类型划分相对粗糙的 EILAT 和 EILAT2 数据集上,我们的方法分别取得了 98.02% 和 99.01% 的准确率。在更为精细的珊瑚类型划分数据集 RSMAS 上,我们的方法也达到了 98.81% 的准确率。因此,本文所提方法不仅能够准确识别造礁珊瑚的类型,还能有效应对数据不平衡和多样性不足的问题,从而提升珊瑚识别的准确率。该方法提供的最佳识别准确率的参数设置如表 6 所示。

表 5 每个敏感性实验在 RSMAS、EILAT 和 EILAT2 数据集上的识别准确率

Table 5 The classification accuracy of each sensitivity experiment on RSMAS, EILAT, and EILAT2 datasets

方法	RSMAS	EILAT	EILAT2
基线	84.19%	74.28%	74.23%
DeepSMOTE	85.00%	79.22%	80.08%
DeepSMOTE- $F_1$	87.88%	79.61%	81.62%
迁移学习	97.54%	94.84%	97.06%
DeepSMOTE- $F_1$ + 迁移学习	98.81%	98.02%	99.01%

为进一步评估本文所提方法对存在数据集不平衡和缺乏多样性问题的珊瑚识别任务的有效性,我们分别在 3 个公开珊瑚识别数据集上对每一类的  $F_1$ -score 进行了分析。 $F_1$ -score 的值介于 0~1 之间,越高代表

表 6 本文方法提供最佳识别准确率的参数设置

Table 6 The setting of parameters of the proposed method on achieving the highest classification accuracy

数据集	批次大小	学习率	迭代次数
RSMAS	64	0.001	500
EILAT	32	0.001	300
EILAT2	32	0.001	300

识别结果越好。图 6 展示了 RSMAS、EILAT 和 EILAT2 数据集上各个实验的  $F_1$ -score 比较结果。图 7 展示了每一类珊瑚采用 DeepSMOTE- $F_1$  前后的图像数量变化情况。结合图 6 和图 7 的结果,我们观察到在原始数据集中,样本数量较少的类别往往表现出较低的基线识别结果,例如在 EILAT 数据集的 Branch I、Branch III 以及 RSMAS 数据集中的 *Montipora spp.*等。但是,采用 DeepSMOTE- $F_1$  后,针对这些基线中识别结果  $F_1$ -score 较低的类别,该方法能够自适应地增加

它们的样本数量,为模型提供更多学习该类特征的机会。上述提到的 Branch I、Branch III 和 *Montipora spp.* 在经过 DeepSMOTE- $F_1$  处理后,它们的采样数量明显高于其他类别。从识别结果来看,与基线相比,采用 DeepSMOTE- $F_1$  进行深度数据增强后,大多数类的  $F_1$ -score 有所提高,仅 RSMAS 数据集中 *Montastraea cavernosa* 的  $F_1$ -score 略微下降(从 0.808 下降至 0.796)。但采用 DeepSMOTE- $F_1$  与迁移学习相结合的方法仍取得了最佳识别结果。因为通过 DeepSMOTE- $F_1$  处理后,模型能够更加关注识别效果较差的类,但是数据增强后的数据集规模仍不足以充分训练包含上万参数的深度学习模型,影响了模型的特征提取能力。在这种情况下,迁移学习能够在大型数据集上学到的特征提取能力运用于小数据集的任务中,提供更可靠、丰富的特征表示,进而促进模型的学习与泛化能力。根据图 6 所示的实验结果,将 DeepSMOTE- $F_1$  与迁移学习相结合的方法在每个类别上均取得了最佳

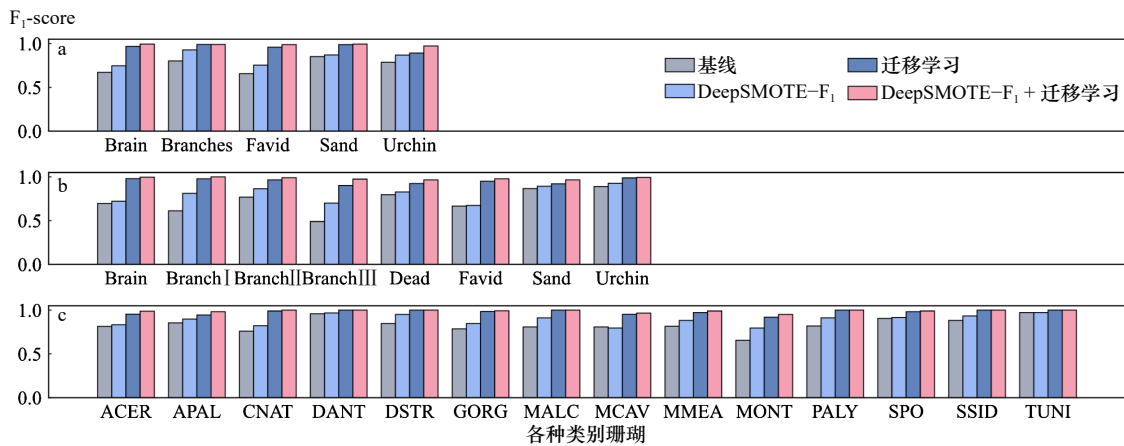


图 6 每个敏感性实验在 3 个数据集 (a. EILAT2; b. EILAT; c. RSMAS) 上的各类别  $F_1$ -score

Fig. 6 The  $F_1$ -score for each class on the three datasets (a. EILAT2; b. EILAT; c. RSMAS)

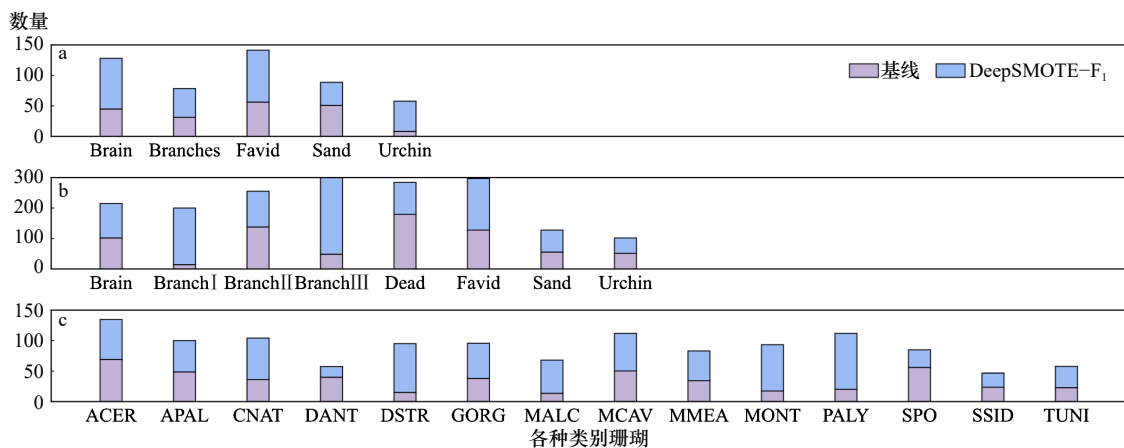


图 7 采用 DeepSMOTE- $F_1$  前后各类珊瑚图像数量对比 (a. EILAT2; b. EILAT; c. RSMAS)

Fig. 7 The comparison of the number of samples before and after data augmentation using DeepSMOTE- $F_1$

(a. EILAT2; b. EILAT; c. RSMAS)

表现。综上所述,本文所提出的方法能够有效地缓解数据集不平衡与多样性不足对珊瑚识别准确率的负面影响。

#### 4.4 与现有珊瑚识别方法对比分析

为进一步验证本文所提方法的优势,我们将融合类别数量自适应深度数据增强和迁移学习的珊瑚识别方法与现有的方法: ReasFeats<sup>[10]</sup>、MDNet<sup>[11]</sup>和 ResNet + Augmentation<sup>[12]</sup>进行了比较,结果如表 7 所示。结果表明,本文所提方法在 RSMAS、EILAT 和 EILAT2 数据集上均取得了更好的结果。

表 7 与现有珊瑚识别方法在 RSMAS、EILAT 和 EILAT2 数据集上的识别准确率对比

Table 7 The comparison of classification accuracy with existing coral classification methods on RSMAS, EILAT and EILAT2 datasets

方法	RSMAS	EILAT	EILAT2
ReasFeats <sup>[10]</sup>	97.42%	96.00%	96.83%
MDNet <sup>[11]</sup>	89.70%	94.70%	91.20%
ResNet+Augmentation <sup>[12]</sup>	98.05%	96.62%	97.71%
DeepSMOTE-F <sub>1</sub> +迁移学习	98.81%	98.02%	99.01%

针对数据集不平衡问题,这 3 种经典的珊瑚识别方法都存在一定不足。首先, ReasFeats 未考虑数据集不平衡的问题。其次,虽然 MDNet 为减轻数据集不平衡对识别效果的影响采用了代价敏感函数,并且代价敏感函数考虑了模型对各个类的识别情况,但它只是简单地复制样本,因此模型的泛化能力较低。最后 ResNet + Augmentation 方法仅采用了简单的数据增强操作,如平移、缩放、旋转和翻转等,容易导致过拟合问题。此外,该方法在每次迭代中对输入网络的图像进行相同概率的采样,导致模型无法有针对性地学习各个类的特征。在本研究中,我们采用 DeepSMOTE-F<sub>1</sub> 深度数据增强方法来解决数据集不平衡问题。DeepSMOTE-F<sub>1</sub> 能够根据每一类珊瑚的原始识别结果 F<sub>1</sub>-score 自动地调整采样策略,使得模型能够充分学习每个珊瑚类型的特征,从而保证了模型的适用

性和泛化能力。我们还利用迁移学习来进一步增强特征提取的能力。与现有方法最好的结果相比,我们的方法在 3 个不同大小、珊瑚类型划分和不平衡率的数据集 RSMAS、EILAT 和 EILAT2 上的准确率分别提升了 0.76%、1.40% 和 1.30%,在不平衡率较大的 EILAT ( $\rho = 12.17$ ) 和 EILAT2 ( $\rho = 6.36$ ) 数据集上的效果更加明显。

## 5 结论

本文针对目前深度学习在自动化珊瑚识别工作中面临的数据集不平衡和训练样本不足且缺乏多样性的问题,提出了一种融合类别数量自适应深度数据增强和迁移学习的珊瑚识别方法。具体而言,类别数量自适应的深度数据增强方法 DeepSMOTE-F<sub>1</sub> 能够根据珊瑚识别结果的综合评价指标 F<sub>1</sub>-score,自适应地调整每类珊瑚的样本增强数量,从而解决了样本间数量不平衡所导致的模型特征学习不充分的问题。同时, DeepSMOTE-F<sub>1</sub> 通过考虑样本之间的相似度,在高维特征空间中对图像数据进行复杂的数据增强操作,确保生成的图像既能保留原始图像的属性,又具有高视觉质量和丰富的信息。此外,该方法结合迁移学习技术进一步增强了模型对珊瑚特征的提取能力。我们在 3 个不同大小、珊瑚类型划分层次以及不平衡率的珊瑚数据集 RSMAS、EILAT 和 EILAT2 上对本文所提方法的有效性和可靠性进行了全面验证。实验结果显示,相较于原始 DeepSMOTE 方法,基于评价指标的采样策略 DeepSMOTE-F<sub>1</sub> 在识别准确率上分别实现了 2.88%、0.39% 和 1.54% 的提升。与现有珊瑚智能识别技术相比,本文方法同样展现出显著优势,准确率分别提高了 0.76%、1.40% 和 1.30%。

本文所提出的深度数据增强方法是一种数据自适应的方法,可以根据数据集内部每类珊瑚的识别结果自适应的调整其动态采样数量,不依赖于特定的珊瑚数据集,可以被广泛应用于存在样本数量不平衡的珊瑚数据集中,以提高其识别效果。在未来的研究中,我们计划将此方法推广应用于更多的珊瑚识别数据集中,以进一步验证其广泛适用性和实用性。

#### 参考文献:

- [1] 黄晖,陈竹,黄林韬. 中国珊瑚礁状况报告(2010–2019)[M]. 北京: 海洋出版社, 2021.  
Huang Hui, Chen Zhu, Huang Lintao. Status of Coral Reefs in China (2010–2019)[M]. Beijing: China Ocean Press, 2021.
- [2] Elliff C I, Silva I R. Coral reefs as the first line of defense: shoreline protection in face of climate change[J]. *Marine Environmental Research*, 2017, 127: 148–154.
- [3] 黄林韬,黄晖,江雷. 中国造礁石珊瑚分类厘定[J]. *生物多样性*, 2020, 28(4): 515–523.  
Huang Lintao, Huang Hui, Jiang Lei. A revised taxonomy for Chinese hermatypic corals[J]. *Biodiversity Science*, 2020, 28(4): 515–523.

- [4] 夏荣林, 宁志铭, 余克服, 等. 长棘海星暴发对珊瑚礁区沉积物营养盐动力学的影响研究 [J]. *海洋学报*, 2022, 44(8): 23–30.  
Xia Ronglin, Ning Zhiming, Yu Kefu, et al. Study on the impacts of crown-of-thorns starfish on nutrient dynamics in the coral reef sediments[J]. *Haiyang Xuebao*, 2022, 44(8): 23–30.
- [5] International Union for Conservation of Nature. The IUCN red list of threatened species, version 2022-2[R/OL]. [2023-07-06]. <https://www.iucnredlist.org>.
- [6] Mahmood A, Bennamoun M, An S, et al. Deep learning for coral classification[M]//Samui P, Sekhar Roy S, Balas V E. Handbook of Neural Computation. London: Academic Press, 2017: 383–401.
- [7] Marcos M S A C, Soriano M N, Saloma C A. Classification of coral reef images from underwater video using neural networks[J]. *Optics Express*, 2005, 13(22): 8766–8771.
- [8] Pizarro O, Rigby P, Johnson-Roberson M, et al. Towards image-based marine habitat classification[C]//Proceedings of the OCEANS 2008. Quebec City, Canada: IEEE, 2008: 1–7.
- [9] Elawady M. Sparse coral classification using deep convolutional neural networks[J]. arXiv: 1511.09067, 2015.
- [10] Mahmood A, Bennamoun M, An S, et al. Resfeats: residual network based features for image classification[C]//Proceedings of 2017 IEEE International Conference on Image Processing. Beijing, China: IEEE, 2017: 1597–1601.
- [11] Modasshir M, Li A Q, Rekleitis I. MDNet: multi-patch dense network for coral classification[C]//Proceedings of the OCEANS 2018 MTS/IEEE Charleston. Charleston, USA: IEEE, 2018: 1–6.
- [12] Gómez-Ríos A, Tabik S, Luengo J, et al. Towards highly accurate coral texture images classification using deep convolutional neural networks and data augmentation[J]. *Expert Systems with Applications*, 2019, 118: 315–328.
- [13] Lumini A, Nanni L, Maguolo G. Deep learning for plankton and coral classification[J]. *Applied Computing and Informatics*, 2023, 19(3/4): 265–283.
- [14] Khan S H, Hayat M, Bennamoun M, et al. Cost-sensitive learning of deep feature representations from imbalanced data[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, 29(8): 3573–3587.
- [15] Lee H, Park M, Kim J. Plankton classification on imbalanced large scale database via convolutional neural networks with transfer learning[C]//Proceedings of 2016 IEEE International Conference on Image Processing. Phoenix, USA: IEEE, 2016: 3713–3717.
- [16] Dablain D, Krawczyk B, Chawla N V. DeepSMOTE: fusing deep learning and SMOTE for imbalanced data[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, 34(9): 6390–6404.
- [17] Chinchor N. MUC-4 evaluation metrics[C]//Proceedings of the 4th Conference on Message Understanding. McLean, USA: Association for Computational Linguistics, 1992: 22–29.
- [18] Alzubaidi L, Bai J, Al-Sabaawi A, et al. A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications[J]. *Journal of Big Data*, 2023, 10(1): 1–82.
- [19] Pan S J, Yang Qiang. A survey on transfer learning[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2010, 22(10): 1345–1359.
- [20] Shihavuddin A S M. Coral reef dataset[EB/OL]. [2024-01-22]. <https://data.mendeley.com/datasets/86y667257h/2>.
- [21] WoRMS. An authoritative classification and catalogue of marine names[EB/OL]. [2024-04-05]. <https://www.marinespecies.org>.
- [22] Nanglu K, Lerosey-Aubril R, Weaver J C, et al. A mid-Cambrian tunicate and the deep origin of the ascidiacean body plan[J]. *Nature Communications*, 2023, 14(1): 3832.
- [23] Rodríguez L, López C, Casado-Amezua P, et al. Genetic relationships of the hydrocoral *Millepora alcicornis* and its symbionts within and between locations across the Atlantic[J]. *Coral Reefs*, 2019, 38(2): 255–268.
- [24] Shihavuddin A S M, Gracias N, Garcia R, et al. Image-based coral reef classification and thematic mapping[J]. *Remote Sensing*, 2013, 5(4): 1809–1841.
- [25] Buda M, Maki A, Mazurowski M A. A systematic study of the class imbalance problem in convolutional neural networks[J]. *Neural Networks*, 2018, 106: 249–259.
- [26] Nair V, Hinton G E. Rectified linear units improve restricted boltzmann machines[C]//Proceedings of the 27th International Conference on International Conference on Machine Learning (ICML-10). Haifa, Israel: Omnipress, 2010: 807–814.
- [27] Chawla N V, Bowyer K W, Hall L O, et al. SMOTE: synthetic minority over-sampling technique[J]. *Journal of Artificial Intelligence Research*, 2002, 16(1): 321–357.
- [28] Deng Jia, Dong Wei, Socher R, et al. ImageNet: a large-scale hierarchical image database[C]//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE, 2009: 248–255.

## Integration of category-quantity adaptive deep data augmentation and transfer learning for reef-building coral recognition

Wang Lan<sup>1</sup>, Wei Hao<sup>1</sup>, Che Yachen<sup>2</sup>, Zhang Cuicui<sup>1,3</sup>

(1. School of Marine Science and Technology, Tianjin University, Tianjin 300072, China; 2. Key Laboratory of Ocean Observation Technology of Ministry of National Resources, National Ocean Technology Center, Tianjin 300112, China; 3. Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China)

**Abstract:** Recognition of reef-building corals is important for protecting and monitoring coral reef ecosystems. Deep learning, as an advanced technology in image recognition, has been increasingly applied in coral recognition. However, its performance is still challenged by several issues, such as the imbalance of samples among different coral categories within a dataset and the limitation of data diversity. The former makes the deep learning model more likely to extract features from classes with a large number of samples and, therefore, decreases its ability to recognize small-sample-size corals, which often refer to endangered ones needing to be protected. The latter further reduces the performance of deep learning in recognizing corals with different appearances and are captured in variant environments. To solve these two problems, this study develops a reef-building coral recognition method by integrating a category-quantity adaptive deep data augmentation algorithm and transfer learning. To address the first problem, a category-quantity adaptive deep data augmentation algorithm named DeepSMOTE-F<sub>1</sub> is proposed. This algorithm improves the existing DeepSMOTE by introducing a sample-size determination stagey using an F1-score based evaluation metric. It can adaptively augment the number of samples of each category of corals according to its recognition performance so that the deep learning model can fully learn features from each class of corals. For the second problem, transfer learning is used to further enhance the model's ability to extract features. The experimental results on three widely used public coral recognition datasets, RSMAS, EILAT, and EILAT2 show that the recognition accuracy of the proposed DeepSMOTE-F<sub>1</sub> is improved by 2.88%, 0.39%, and 1.54%, respectively, compared with the traditional DeepSMOTE; and the accuracy of the integrated method is improved by 0.76%, 1.40% and 1.30% compared with the existing deep learning methods for coral recognition.

**Key words:** coral recognition; deep learning; imbalanced dataset; data augmentation; transfer learning.