doi:10. 3969/j.issn. 1672-4623.2023.02.011

基于迁移学习的非约束环境下热带鱼类识别

张　珊[1]，韩　澈[1]，刘薇芳[1]，朱宇鹏[2]
（1. 海南大学旅游学院，海南 海口 570208；2. 海南超图科技有限公司，海南 海口 570220）

摘　要：针对海底非约束环境下视频背景多模、环境变化复杂导致图像识别困难的问题，提出基于迁移学习的非约束环境下热带海域鱼类识别方法。首先构建ResNet深度卷积神经网络；其次引入迁移学习进行网络训练，对比迁移学习前后的识别效果。结果表明，在引入迁移学习下，accuracy和loss指标均优于非迁移学习的情况，并且在训练到150个epoch时，各项指标开始收敛，能够较好地完成非约束环境下热带海域的鱼类识别任务。

关键词：非约束环境；迁移学习；数据增强；鱼类识别；热带海域
中图分类号：P229　　**文献标志码：**B　　**文章编号：**1672-4623（2023）02-0056-06

Tropical Fish Identification Under Unconstrained Environment Based on Transfer Learning

ZHANG Shan[1], HAN Wei[1], LIU Weifang[1], ZHU Yupeng[2]
(1. School of Tourism, Hainan University, Haikou 570208, China; 2. Hainan Chaotu Technology Co., Ltd., Haikou 570220, China)

Abstract: Aiming at the problems that seabed unconstrained environment video often with complex multi-mode background, complex weather variation, which leads to difficulty in image recognition, we proposed a tropical fish identification method under unconstrained environment based on transfer learning. Firstly, we constructed the ResNet deep convolution neural network. And then, we introduced transfer learning for network training, and compared the recognition effect before and after transfer learning. The results show that the indexes of accuracy and loss with transfer learning are better than those without transfer learning. When the training reaches 150 epochs, each index begins to converge, which can better complete the task of tropical fish identification under unconstrained environment.

Key words: unconstrained environment, transfer learning, data enhancement, fish identification, tropical sea region

以往对于海洋鱼类资源监测常通过现场采样法[1]、渔获法[2]、水声学遥测等方法，这些方法不同程度地存在依赖专家经验、破坏生物资源、无法判断鱼的类型等缺点[3]。得益于数据处理技术、评价模型和方法的完善，海洋生物资源监测评价步入精准化、无损化时代[4]。建立水下视频深度学习系统，将是智能化监测海洋鱼类资源最为有效的方法。

计算机视觉捕捉分析技术作为一种非破坏性技术，在监测中可以保持海洋生物资源完整、带来最少的环境扰动，被广泛应用到海洋生物监测研究之中。人们通过摄影、摄像获取海底图像并采用图像处理方法研究海洋生物特征[5-7]。相较于传统方法，利用计算机技术进行海洋生物监测是一种自动、非入侵、经济且有效的方法。

虽然基于图像的目标检测技术近些年得到了广泛的研究，但是基于海底非约束环境下图像的目标检测技术仍然面临很大的挑战。由于鱼类的快速自由运动、天气、光照等外界因素经常变化[8]导致视频影像对比度低、纹理细节模糊，不利于训练网络提取图像特征。

1　鱼类图像识别算法

鱼类图像识别基本流程：输入鱼类图像、选择鱼类特征、构建分类器、分类识别。早期的方法过度依赖人工选择特征，人工选择特征主要依靠专家的先验知识，通过鱼的轮廓、形状、颜色等特征进行检测[9-10]，特征设计中允许出现的参数数量十分有限，难以准确获取高层次特征[5]，从而影响分类识别结果的准确性。随着人工智能技术的发展，深度学习在计算机视觉领域取得了巨大成功，传统的机器学习法正逐渐被基于深度学习的方法替代。深度学习可以从大数据中自动学习特征的表示，可以包含成千上万的参

收稿日期：2021-11-17。
项目来源：国家自然科学基金资助项目（41906176）；海南省自然科学基金资助项目（419QN172）。

数，通过深度神经网络结构将数据从像素级低维特征映射到语义层高维特征，这使得它在提取图像的全局特征和上下文信息方面具有突出优势[11]，为解决传统的计算机视觉问题（如图像分割和关键点检测）带来了新的思路。

2012年后，尤其是以卷积神经网络（convolutional neural network，CNN）为代表的神经网络模型得到了广泛应用[12-14]。CNN是由交替卷积层和子采样层组成的一种流行的深度学习体系结构，由于它能够提取复杂、高维的图像特征，已经被证明在许多计算机视觉应用中有效，如分割[15]、对象识别和分类[16]、目标跟踪[17]。2017年，谷歌和NASA通过CNN技术在开普勒-90海量影像中发现了一颗新行星。Frederic[12]等亦通过CNN从数十万张卫星影像中找到了濒危物种海牛。Spampinato[18]等提出了一种在无约束环境下跟踪鱼类的算法，基于协方差模型描述被检测对象，比较当前对象的协方差矩阵与候选对象的协方差矩阵，通过模板匹配生成轨迹，平均准确率约为80%，有待进一步提高。针对这些问题，Chuang[19]等进一步提出了一种基于可变形多核的跟踪算法，通过对颜色和纹理特征的均值偏移算法，有效地估计核的运动，实现鱼类跟踪，但文中并没有给出具体的测试结果。在此基础上，Jager[20]等通过两阶段图形法，用激活的CNN实现无约束环境下鱼类跟踪，具有较高的跟踪精度和效率，但模型的14个超参数无法自动调整，从而降低了模型的可移植性。Wang[21]等针对视频帧中每条鱼的头部图像，定制CNN并在跟踪过程中对CNN参数进行自适应更新，以适应轻微的光照和鱼的外观变化，该方法被证明对斑马鱼有较好的识别和追踪效果，然而实验仍然局限在水族箱环境内。贾宇霞[22]等通过图像增强和迁移学习在Fish4Knowledge数据集上进行鱼类识别实验，识别精度达到99.63%，但仅限于单个鱼类主体，对无法解决鱼类重叠和遮挡。Pan[23]等构建了多尺度ResNet实现水下目标对象定位，实现了实时识别，但对有遮挡对象和微小对象的识别效果不佳。Zhao[24]等通过改进ResNet，实现在低质量水下环境中的鱼类识别，但该实验对硬件要求较高。

2　数据来源

项目组从2017年起在海南省三亚市蜈支洲岛周围，根据不同礁型、不同材质、不同水深等选取了5处人工礁进行海洋鱼类的拍摄观测，如图1所示。水下电子舱布置在5个观测点，一个舱内布置有两台光电式HB-HM-02水下摄像机从不同角度进行数据采集。视频分辨率为648×480，即宽度和高度，帧速率为50。目前，五套设备已全部布置并运行良好，实现了视频数据和水质监测数据的连续不间断记录和传输。

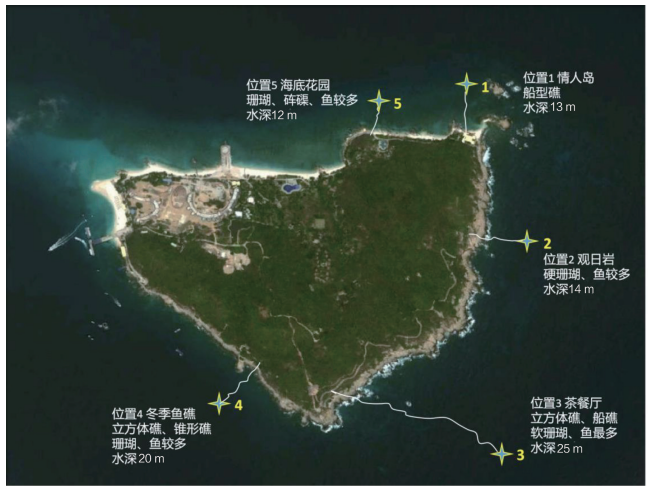


图1　蜈支洲岛水下视频监控布点位置图

10台摄像机每min产生150+MB的视频数据（夜间光照欠佳期间摄像机不拍摄视频），在12 h的监控下，每天产生100 GB以上的视频数据，每年产生35+TB的视频数据，这个地区拥有丰富的鱼类生物资源如图2所示。



图2　蜈支洲岛数据库视频帧视觉效果示意图

从蜈支洲岛数据库中选取不同时间段对24个视频片段，共计3 100个标记视频帧，其中75%作为训练集，25%的数据作为测试集,标签数据使用麻省理工计算机科学与人工智能实验室研发的LabelMe深度学习图像标注工具完成。如表1所示。

表1　蜈支洲岛数据库概况

数据库名	视频数量	视频分辨率	FPS	标记视频帧	训练集样本	测试集样本
蜈支洲岛	24	640×480	50	3 100	2 325	775

3 迁移学习

数据是深度学习的关键，有足够可利用的训练样本是计算机通过学习能够获得好的分类器的前提。大多现有可用数据是针对特定课题而设计的，如手写数字数据库MINIST，图片数据集ImageNet，图像语义理解数据集COCO等，这些开源数据不能完全满足本研究所需的深度学习训练数据要求，因此要针对特定鱼类进行训练数据的补充。然而手动标注图像非常耗时，在对无人驾驶数据集Cityscapes中单个图像的精细像素级的标注平均需要花费1.5 h。此外，由于人的主观因素也容易出现错误标注。因此，项目引入迁移学习来解决目前目标领域仅有少量有标签样本数据的问题。

迁移学习是运用已经训练好的知识对不同但相关领域问题进行求解的一种新的机器学习的方法，以解决目标领域中仅有少量有标签样本数据的问题。以往，针对不同的领域问题需要重新训练深度学习模型，一方面需要耗费大量的时间和计算资源；另一方面，在图像识别领域，图像底层特征相似性高，这些底层特征被组合成为不同的更复杂的特征用于图像分类识别。因此，可以将经过其他大型训练集预训练之后的网络模型作为底层图像特征提取器，从而实现将通用图像特征提取知识迁移到热带鱼类分类识别领域，减少对标注数据的依赖。斯坦福大学吴恩达教授在2016年NIPS会议上说：“迁移学习将会是继监督学习之后的下一个机器学习商业成功的驱动力”。

ImageNet图像数据集始于2009年，是目前世界上图像识别最大的数据库，由美国斯坦福的计算机科学家模拟人类的识别系统建立。包含约1 500万张图片，2.2万个类型，每张图片都经过严格的人工筛选与标记。从2010年起，每年都基于ImageNet举办全世界的大规模视觉知识挑战赛（ILSVRC），世界各地的研究团队在给定的数据集上评估其算法，并在视觉识别任务中争夺更高的准确性，2017年后比赛由Kaggle社区主持。在图像分类上，ImageNet竞赛已经是计算机视觉分类算法事实上的评价标准。Keras是一个由Python编写的开源人工神经网络库，主要提供了基于ImageNet训练的5种开箱即用型的CNN网络——VGG16、VGG19、ResNet50、Inception V3、Xception。与AlexNet、OverFeat还有VGG这些传统顺序型网络架构不同，ResNet50的网络结构依赖于微架构模组，因为使用了全局平均池化，而不是全连接层。尽管ResNet50比VGG16还有VGG19要深，weights却更小。本文使用Keras库中的ResNet50网络，利用其自带已训练好的imagenet作为权重来初始化训练目标模型参数，实现将已经学到知识共享给新模型，然后根据目标任务调整输出层，用热带鱼类图像样本精调网络模型进行训练，提高检测算法的精度和效度，最终得到热带海域鱼类分类识别模型，迁移学习流程如图3所示。

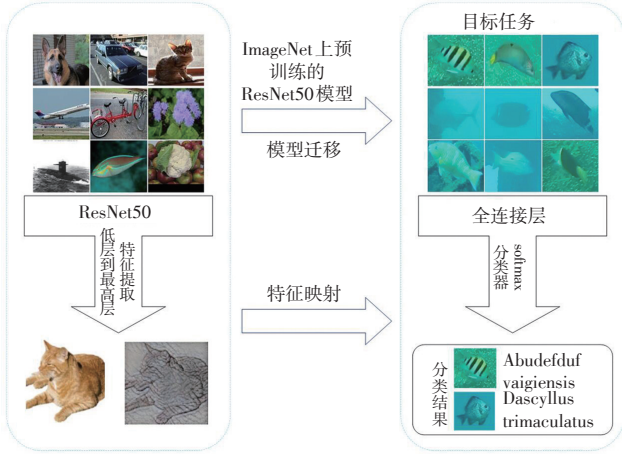


图3 迁移学习流程图

4 模型架构

整个ResNet深度卷积神经网络由初始卷积（Conv）层、若干个Residual Block和最后的全连接层（fully connected，FC）组成。第一个Conv layer不需要进行归一化和ReLU非线性激活。首先按照均值和标准差，对图像进行归一化处理。考虑到梯度下降的训练时间，使用ReLU非线性激活作为激活函数，比tanh激活函数效果更好[28]，然后进行第一次3×3卷积运算，对卷积结果再次进行batch normalization，ReLU非线性激活后再进行第二次3×3卷积运算，执行跳远连接，将第一层的激活迅速传递给第六层后的下一个Residual Block。模型总训练次数（train steps）为300，图片通过一系列3×3的滤波器进行卷积运算，卷积滑动步长固定为1，使用mini_batch对算法性能进行优化。考虑到内存有限，训练集和测试集的batch_size均设置为24，padding size为same进行图像填充。模型总参数23 587 712个，固定参数53 120个，可训练参数13 534 592个，模型结构及参数设置如图4所示。

5 实验及结果分析

为了评估迁移学习，在蜈支洲岛数据集上进行了实验，在网络结构、学习效率、损失函数等保持不变

的情况下，对比了迁移学习前后的各项学习指标。实验使用了一台 Intel(R) Xeon(R) X5675 3.06 GHz 计算机系统型处理器，32 GB 内存和 NVIDA GeForce GTX 1050 Ti GPU。实验所得部分数据如表2所示。

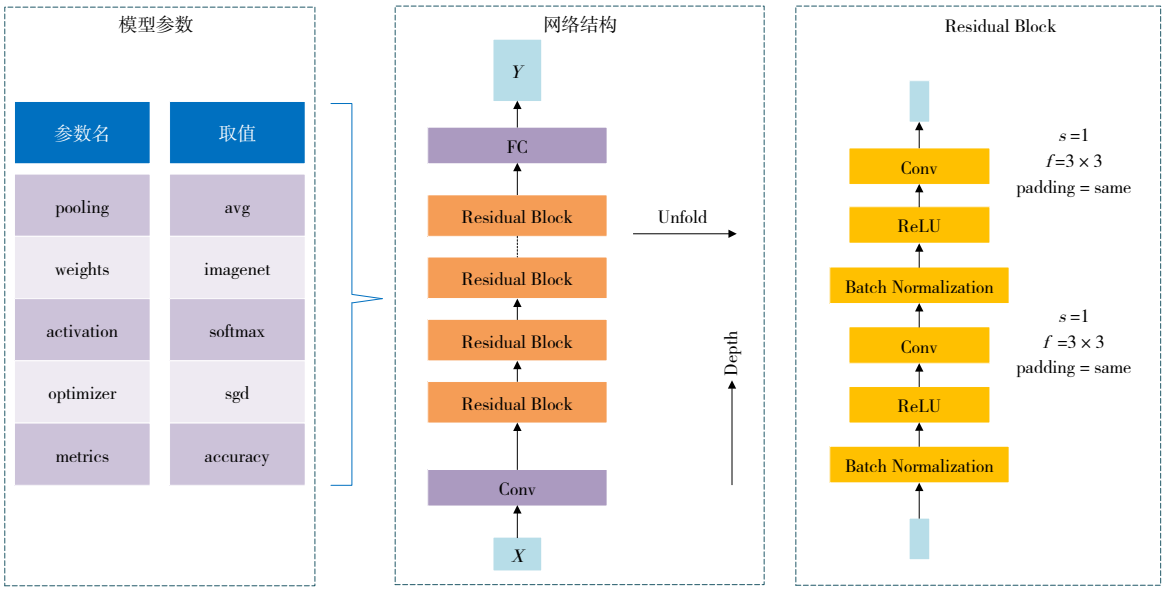


图4　ResNet热带鱼类识别模型架构图

表2　迁移学习前后不同训练次数下的各指标数值

Epoch	Train loss		Validation loss		Train accuracy		Validation accuracy	
	With TL	Without TL	With TL	Without TL	With TL	Without TL	With TL	Without TL
1	1.067 8	1.309 5	0.888 5	1.129 7	0.595 7	0.577 5	0.683 6	0.671 9
51	0.270 0	0.335 2	0.334 9	0.463 0	0.932 9	0.889 1	0.907 8	0.859 4
101	0.218 8	0.272 7	0.307 9	0.399 5	0.956 5	0.910 6	0.924 2	0.875 0
151	0.191 1	0.238 5	0.313 8	0.402 4	0.966 8	0.922 0	0.918 8	0.880 2
201	0.178 7	0.225 8	0.278 8	0.383 7	0.970 0	0.927 3	0.951 6	0.890 6
251	0.170 5	0.218 5	0.277 5	0.380 8	0.974 3	0.924 3	0.962 5	0.895 8
300	0.168 8	0.214 4	0.303 8	0.387 2	0.973 2	0.923 5	0.962 5	0.911 5

train loss 代表训练数据上的损失，使用交叉熵作为损失函数，主要用来衡量训练集上的拟合能力，train loss 值越低，说明模型在训练集上的拟合能力越高。validation loss 代表验证集上的损失，使用交叉熵作为损失函数，主要用来衡量模型在未知数据上的拟合能力，即模型的泛化能力，validation loss 值越低，说明模型的泛化能力越高。由图5可见，在引入 imagenet 已训练好的模型作为网络的初始权重下，train loss 值和 validation loss 值迅速降低，并且随着训练次数的增加，train loss 值和 validation loss 值均保持持续下降的趋势。在训练到150次后，validation loss 值在 0.41 ~ 0.37 区间震荡，随着训练的深入，loss 值进一步震荡下行，说明本方法具有较好的拟合能力和泛化能力。

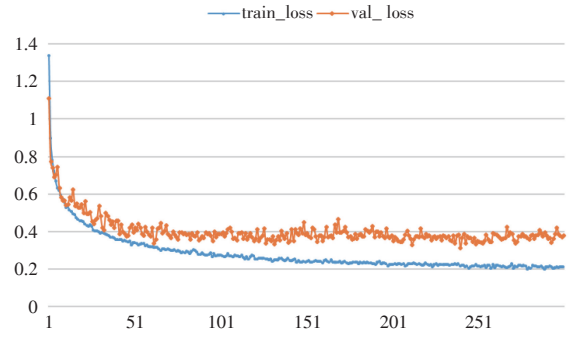


图5　迁移学习训练后的train loss和validation loss变化图

Train accuracy 代表训练数据上的准确率，validation accuracy 代表验证集上的准确率，accuracy 的值越高，说明模型对热带鱼类分类识别的准确度越高。由图6可见，在引入 imagenet 已训练好的模型作为网络的初始权重下，train accuracy 值和 validation accuracy

值迅速升高，并且随着训练次数的增加，train accuracy 值和 validation accuracy 值均保持持续增加的趋势。在训练到150次后，validation accuracy 值在0.85～0.90区间震荡，随着训练的深入，accuracy 值进一步震荡上行，说明本方法具有较好的识别精度。

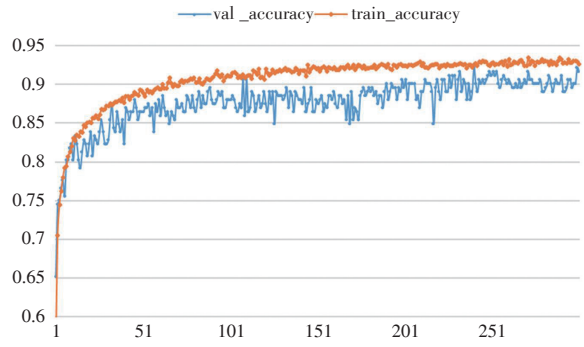


图6 迁移学习训练后的train accuracy和validation accuracy变化图

在保持网络结构和其他超参数相同的情况下，本文对比了使用迁移学习前后的 train loss 和 validation loss 指标的变化。从图7中可见，在使用迁移学习前后，train loss 值均随着训练次数的增加而减小至0.2～0.3左右。在训练起始阶段，loss 会有一个较为明显的下降，当训练到约100次后，train loss 就基本降低到30%以下，validation loss 在0.25～0.45之间震荡。迁移学习后的loss明显低于迁移学习前。

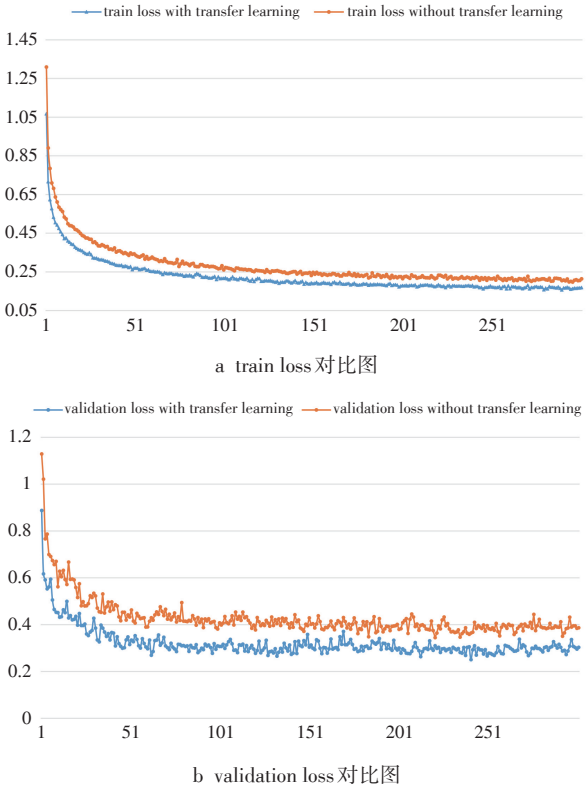


图7 迁移学习前后的train loss和validation loss对比图

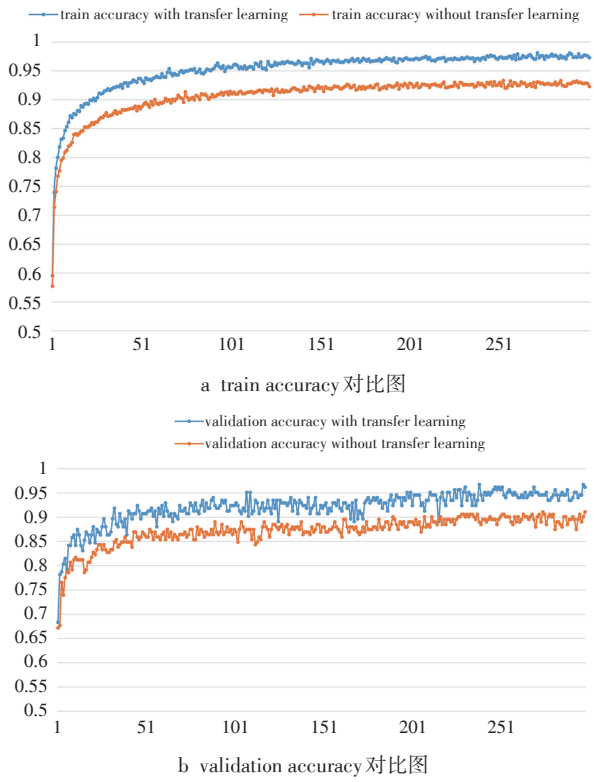


图8 迁移学习前后的train accuracy和validation accuracy对比图

在保持网络结构和其他超参数相同的情况下，本文对比了使用迁移学习前后的 train accuracy 和 validation accuracy 指标的变化。从图8可见，在使用迁移学习前后，train accuracy 值均随着训练次数的增加而增加至0.9左右。在训练起始阶段，accuracy 会有一个较为明显的上升，当训练到约50次后，train accuracy 就基本稳定在0.8以上。迁移学习后的 accuracy 明显高于迁移学习前。

6 结 语

本文以非约束条件下海洋视频监控数据中的热带鱼类为研究对象，构建了基于迁移学习的热带鱼类识别模型，以 ResNet50 作为基础训练网络。结果表明，迁移学习对非约束环境下热带鱼类的计算机识别具有较好的效果，有助于提高识别准确度并缩短学习时间。今后将进一步获取超高清视频数据，优化算法，解决收敛到局部最优而不是全局最优的问题。同时改善硬件条件，进一步提高处理精度。

参考文献

[1] 李永振,史赞荣,艾红,等. 南海珊瑚礁海域鱼类分类多样性大尺度分布格局[J]. 中国水产科学,2011,8(3):619-628

[2] Hodgson G. Reef Check:The first Step in Communi-

ty-based Management[J]. Bulletin of Marine Science,2001,21(9):861-868

[3] 杨红生. 海洋牧场构建原理与实践[M]. 北京:科学出版社,2015

[4] Zhang S,Zhu Y P. Research on the Construction of Underwater Video Surveillance and Water Environment Monitoring Information System for Marine Ranching[J]. Boletín Técnico,2017,55(17):738-742

[5] Hsiao Y H,Chen C C,Lin S I,et al. Real-world Underwater Fish Recognition Identification, Using Sparse Representation[J]. Ecological Informatics,2014,28(23):13-21

[6] 张珊,朱宇鹏. 基于三维视觉的海洋生物区域种类识别仿真[J]. 计算机仿真,2018,35(5):402-406+443

[7] Zhang S,Zhu Y P,Bin Y P. Tropical Coral Reef Fish Identification Method Based on ResNet in Unconstrained Environment of the Seabed[J]. Revista Científica,2019,107(10):1 157-1 164

[8] Toh Y H,Ng T M,Liew B K. Automated Fish Counting Using Image Processing[C]//International Conference on Computational Intelligence and Software Engineering,2009,1-5

[9] Zion B,Alchanatis V,Ostrovsky V,et al. Classification of Guppies Poecilia Reticulata Gender by Computer Vision[J]. Aquacultural Engineering,2008,38(2):97-104

[10] Alsmadi M K A,Omar K B,Noah S A,et al. Fish Recognition Based on the Combination Between Robust Features Selection, Image Segmentation and Geometrical Parameters Techniques Using Artificial Neural Network and Decision Tree[J]. International Journal of Computer Science and Information Security,2016,6(2):215-221

[11] 余凯,贾磊,陈雨强,等. 深度学习的昨天、今天和明天[J]. 计算机研究与发展,2013,50(9):1 799-1 804

[12] Frederic M,Alvarez L M,Hodgson A. Automating Marine Mammal Detection in Aerial Images Captured During Wildlife Surveys:A Deep Learning Approach[C]//Australasian Joint Conference on Artificial Intelligence,2015,379-385

[13] 徐勇,赵俊,过锋,等. 基于BP人工神经网络的大沽河湿地海水水质综合评价[J]. 渔业科学进展,2015,36(5):31-37

[14] 李海涛,刘泰麟,邵泽东,等. 基于决策树与二分分割算法的BP神经网络在赤潮等级预测中的应用研究[J]. 海洋科学,2019,42(9):34-40

[15] Girshick R,Donahue J,Darrell T,et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition,2014,580-587

[16] Szegedy C,Toshev A,Erhan D. Deep Neural Networks for Object Detection[C]//Advances in Neural Information Processing Systems,2013:2 553-2 561

[17] Zhou X,Xie L,Zhang P. Online Object Tracking Based on CNN with Metropolis-hasting Re-sampling[C]//Proceedings of the Acm International Conference on Multimedia,2015:101-113

[18] Spampinato C,Palazzo S,Giordano D,et al. Covariance Based Fish Tracking in Real-life Underwater Environment[C]//International Conference on Computer Vision Theory and Applications,2012:409-414

[19] Chuang M C,Hwang J N,Williams K,et al. Tracking Live Fish from Low Contrast and Low-frame-rate Stereo Ivdeos[J]. IEEE Transactions on Circuits and Systems for Video Technology,2015,25(1):167-79

[20] Jager J,Wolff V,Fricke N K,et al. Visual Fish Tracking: Combining a Two-stage GraphApproach with CNN-features[C]//IEEE Oceans,2017:41-46

[21] Wang S H,Zhao J W,Chen YQ. Robust Tracking of Fish Schools Using CNN for Head Identification[J]. Multimedia Tools and Applications,2017,28(10):86-105

[22] 贾宇霞,樊帅昌,易晓梅. 基于显著性增强和迁移学习的鱼类识别研究[J].渔业现代化,2020,47(1):38-46

[23] Pan T S,Huang H C,Lee J C,et al. Multi-scale ResNet for Real-time Underwater Object Detection[J]. Signal Image and Video Processing,2021,15(2):1-9

[24] Zhao Z X,Liu Y,Sun X D,et al. Composited Fish Net:Fish Detection and Species Recognition from Low-Quality Underwater Videos[J]. IEEE Transactions on Image Processing,2021,30(12):4 719-4 734

第一作者简介：张珊（1981—）博士，副教授，研究方向为智慧海洋、大数据与深度学习。