

基于声呐图像的类别增量学习方法研究

陈鑫哲, 梁红, 徐微雨

(西北工业大学 航海学院, 陕西 西安 710072)

摘 要: 由于声呐图像分辨率低、样本数少, 现有的类别增量学习网络对历史任务目标出现了严重的灾难性遗忘问题, 导致所有任务目标的平均识别率降低。基于生成重放的框架模式, 提出了一种改进的类别增量学习网络, 设计搭建新的深层卷积生成对抗网络取代变分自编码器, 作为生成重放增量网络的重构模型, 提升图像的重构效果; 构建新的卷积神经网络取代多层感知机, 作为生成重放增量网络的识别网络, 提升图像的分类识别性能。结果表明, 改进的生成重放增量网络缓解了历史任务目标的灾难性遗忘问题, 显著提高所有任务目标的平均识别率显著提高。

关 键 词: 声呐图像识别; 生成重放; 类别增量学习

中图分类号: TP391.41

文献标志码: A

文章编号: 1000-2758(2023)02-0303-07

随着水下勘探技术的不断发展与水下环境的逐渐开发, 更多未经训练过的水下目标信息会被获取, 水下目标识别系统的需求也随之增加。目前, 大部分用于水下目标声呐图像的分类识别网络采用的都是一次性输入所有类别样本的训练方法, 这样训练好的分类识别网络只能对学习过的类别样本进行分类识别, 当有新的类别样本需要学习时, 就只能将目前的所有类别样本重新进行训练, 需要耗费大量的时间与算力。

类别增量学习 (class-incremental learning, C-IL)^[1] 要求分类识别网络能不断地从新获取的任务类别样本中学习到新知识, 同时保留历史学习过的任务目标类别知识。但是, 类别增量学习网络在学习了新任务类别样本后, 通常会对历史学习过的任务类别样本出现灾难性遗忘问题^[2], 导致所有任务目标的分类识别准确率严重降低。

为了缓解灾难性遗忘问题, 研究人员提出了各种增量学习方法, 依据方法原理的不同, 可以将目前的增量学习方法分为以下 2 类: 基于正则化 (regularization) 和基于重放 (replay)。基于正则化的代表方法有: Li 等^[3] 提出的不忘学习 (learning without forgetting, LWF), DeepMind 团队提出的弹性权重巩固

固 (elastic weight consolidation, EWC)^[4]。基于重放的代表方法是 Shin 等^[5] 提出的生成重放 (generated replay, GR), 基于生成重放的增量学习方法对类别增量学习总体表现更好, 它对于光学图像取得了较高的类别增量识别准确率, 并且在一定程度上缓解了灾难性遗忘问题。

相比数以百万计的高质量光学图像数据, 由于获取难度高、水下环境噪声干扰大和观测角度有限, 声呐图像存在样本数少、分辨率低和边缘不连续的问题^[6], 直接将光学领域的识别方法应用于声呐图像, 通常表现较差^[7]。因此, GR 增量网络对声呐图像类别增量学习的可行性还待进一步研究。

本文研究 GR 增量网络对于声呐图像类别增量学习的可行性, 提出一种适用于声呐图像类别增量学习的改进 GR 增量网络, 缓解灾难性遗忘问题, 提高声呐图像所有任务目标类别的平均识别准确率。

1 数据集的建立

本文建立的数据集包含前视声呐和侧扫声呐等多种声呐成像设备得到的声呐图像, 依据图像的目标类别将其整理划分为 6 类, 分别是飞机、沉船、瓶

子、轮胎、石块和 水下机器人。采用双线性插值法对声呐图像进行尺寸统一,规范后的图像尺寸为 (112,112),每个目标类别的数据按照 4 : 1 的比例随机分为训练集和测试集,数据集分布如表 1 所示。

表 1 数据集分布						
样本分布	飞机	轮船	瓶子	轮胎	石块	水下机器人
样本数	388	414	110	197	198	310
训练集	311	332	88	158	159	248
测试集	77	82	22	39	39	62

本文将数据集划分为如图 1 所示的 3 个任务目标,依次输入到类别增量学习网络中,作为新的类别知识对网络进行训练。

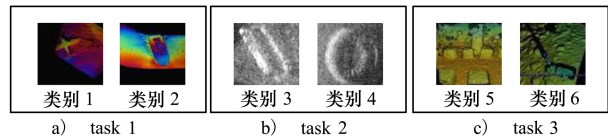


图 1 任务目标

2 GR 增量网络

2.1 GR 增量网络结构

GR 增量网络是由重构模型和识别网络组成的合作双模型架构。利用这种模型架构,网络可使用重构的历史数据和真实当前数据的集合进行联合训练,缓解灾难性遗忘问题,GR 增量网络学习流程如图 2 所示。

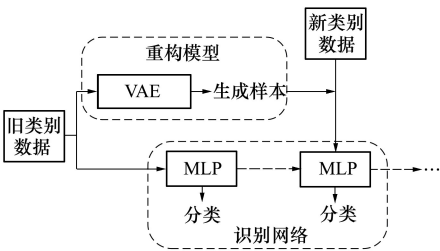


图 2 GR 增量网络学习流程

GR 增量网络在类别增量学习过程中,对历史类别数据进行重构的模型是变分自编码器 (variational autoencoder,VAE)^[8],由编码器、采样模块和解码器组成。编码器将图像数据分布的高维特征映射到低维特征上,这一部分主要是均值和方差

的计算;图像数据经过编码后,采样模块从正态分布中采样数据,并与编码器得到的均值和方差联合计算,从而得出采样变量;获得采样变量后,将其作为解码器的输入,利用解码器进行图像还原,输出生成图像。对图像进行训练的识别网络采用多层感知机 (multilayer perceptron,MLP)^[9] 这一浅层全连接网络。

2.2 GR 增量网络实验分析

本文实验平台为 AMD 5900X CPU、Nvidia RTX 3080 GPU 和 32G 内存的工作站,Pytorch 版本为 1.9,操作系统为 Windows10。针对本文建立的声呐图像数据集,采用 GR 增量网络进行训练。随着任务目标的不断增加,网络对新任务目标和旧任务目标的识别率变化如图 3 所示。

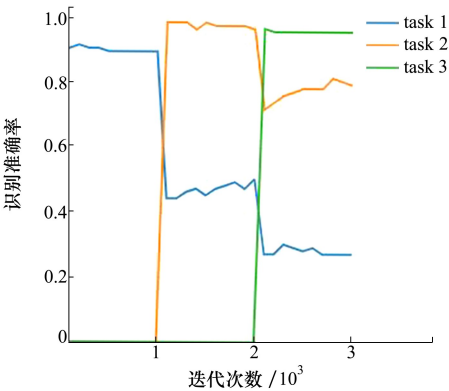


图 3 GR 增量网络的识别率

当第一个任务进入增量网络训练完成后,网络对所有目标类别的识别准确率为 89.9%。当第二个任务作为新类别数据进入增量网络训练完成后,网络对所有目标类别的平均识别率为 73.6%,对历史学习的第一个任务的识别率从 89.9%降为 50.5%。当第三个任务作为新类别数据进入增量网络训练完成后,网络对所有目标类别的平均识别率为 67.6%,对历史学习的旧任务目标的识别率明显下降,其中,对第一个任务的识别准确率降到了 27.3%,对第二个任务的识别准确率降到了 79.6%。结果表明,GR 增量网络的平均识别率较低,对历史任务目标产生了较严重的灾难性遗忘。

3 改进的 GR 增量网络

3.1 网络结构改进原理

由 2.2 节可知,GR 增量网络对于声呐图像的类

别增量学习效果较差,本文对 GR 增量网络的重构模型和识别网络进行改进,以期望获得更好的类别增量学习性能。

GR 增量网络的重构模型 VAE 虽然可以通过编码与解码重构图像,但是这种方式下生成的图像较为模糊^[10]。深层卷积生成对抗网络(deep convolutional generative adversarial network, DCGAN)^[11]由生成器和判别器两部分组成,二者结构近似于卷积神经网络,但是使用带步长的反卷积层或卷积层代替池化层,DCGAN 通过生成器与判别器的生成对抗来重构图像。本文随机选取飞机和沉船的重构图像进行对比,VAE 和 DCGAN 重构的声呐图像分别如图 4a) 和图 4b) 所示。两图的上半部为原始声呐图像,下半部为重构的图像。VAE 重构的图像中,当飞机较小时,图像无法清晰分辨目标,并且所有沉船的重构图像均严重失真。相比而言,DCGAN 重构的飞机与沉船图像轮廓更明显、细节更丰富,大部分重构图像可较为清晰地分辨目标。

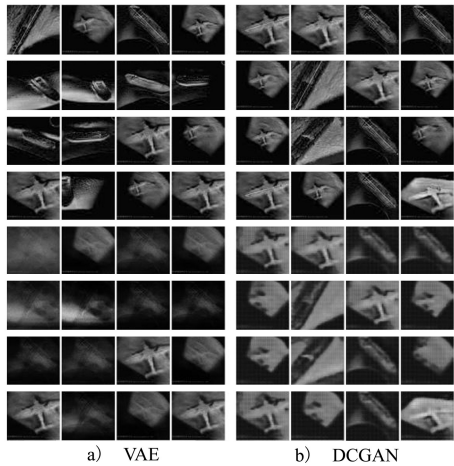


图 4 VAE 和 DCGAN 重构的声呐图像

GR 增量网络的识别网络 MLP 是由多个神经元层组成的全连接网络模型,它早期在手写数字识别和行人检测中获得了不错的效果^[12]。但是,在进行声呐图像类别增量学习训练时,MLP 这种全连接网络模型的参数量是十分巨大的,过大的参数量会导致网络训练过拟合,而且在图像的特征提取过程中,MLP 会将所有图像转换成一维向量,导致其失去空间特征,影响图像的分类识别准确率。卷积神经网络(convolutional neural network, CNN)^[13]利用卷积层和池化层实现了网络结构的局部连接、权值共享和空间下采样,在提取出图像局部空间特征的同时,

有效降低了网络模型训练的参数量,避免训练过拟合,对图像的准确识别起到积极作用。

本文基于水下目标声呐图像的特点,设计搭建新的 DCGAN 和 CNN,并将设计搭建好的 DCGAN 和 CNN 分别替换 GR 增量网络的 VAE 和 MLP,作为网络新的重构模型和识别网络,得到改进的 GR 增量网络。改进的 GR 增量网络结构如图 5 所示。

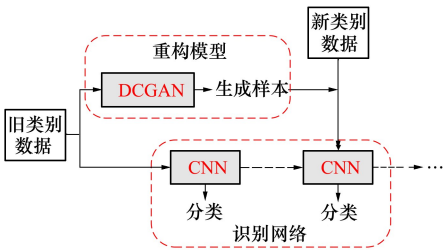


图 5 改进的 GR 增量网络结构

3.2 网络模型搭建

3.2.1 DCGAN 模型搭建

基于生成对抗网络理论与水下目标声呐图像特点,设计搭建 DCGAN 需考虑以下几点:

- 1) 生成器与判别器模型的设置。生成器与判别器要求模型对称,即生成器的反卷积层与判别器的卷积层数目相等,这样才能达到网络可微的目的,从而使网络的损失函数在训练过程中向着一个方向稳定收敛。
- 2) 反卷积层和卷积层数目的设置。由于声呐图像样本数少,应使用较少的层数进行模型搭建,避免因参数过多而导致网络训练出现过拟合。
- 3) 卷积层参数的设计,包括卷积核大小、步长、边界填充数目。声呐图像经过裁剪等预处理操作后,图像的尺寸相较原始图像减小了很多,为了避免网络模型在声呐图像的特征提取过程中丢失大量的有用特征,应使用较小的卷积核,如 3×3 卷积核。设置步长可以压缩卷积层的运算量,加快网络训练的拟合速度,但是步长不宜设置过大,过大的步长会使卷积层提取不到充足的声呐图像特征。边界填充是为了避免带步长的卷积运算错过声呐图像的边缘信息,边界填充的数目需根据输入图像、步长与卷积核来设置,不能超过卷积核的大小,以避免无效运算。
- 4) 网络参数优化算法和激活函数的选取。Adam 算法能基于训练数据迭代更新网络参数,具有计算高效和所需内存少等优势。ReLU、Leaky

ReLU 激活函数在图像处理的卷积神经网络中使用较多,它们能够保持梯度不衰减,从而缓解梯度消失问题。

综合考虑以上几点设计要求,本文设计搭建的 DCGAN 的生成器与判决器模型分别如表 2~3 所示。

表 2 生成器模型

层数	操作名称	参数设置
1	全连接批次标准化	输出维度:7×7×128
	反卷积	反卷积核大小:3×3
2	批次标准化	步长:2
	Leaky ReLU 激活	边界填充数:1
3	反卷积	反卷积核大小:3×3
	批次标准化	步长:2
	Leaky ReLU 激活	边界填充数:1
	反卷积	反卷积核大小:3×3
4	批次标准化	步长:2
	Leaky ReLU 激活	边界填充数:1
	反卷积	反卷积核大小:3×3
	批次标准化	步长:2
5	Leaky ReLU 激活	边界填充数:1

表 3 判别器模型

层数	操作名称	参数设置
1	卷积	卷积核大小:3×3
	批次标准化	步长:2
	Leaky ReLU 激活	边界填充数:1
2	卷积	卷积核大小:3×3
	批次标准化	步长:2
	Leaky ReLU 激活	边界填充数:1
3	卷积	卷积核大小:3×3
	批次标准化	步长:2
	Leaky ReLU 激活	边界填充数:1
4	卷积	卷积核大小:3×3
	批次标准化	步长:2
	Leaky ReLU 激活	边界填充数:1
5	全连接	
	Sigmoid 激活	神经元数目:600
	Dropout	

生成器首先产生随机噪声输入到网络第一层——全连接层,通过全连接层进行尺寸变换,转化为 7×7×128 的特征矩阵,然后该特征矩阵通过 4 层反卷积核大小为 3×3,步长 $s=2$ 的反卷积层,同时,对反卷积层进行 $p=1$ 的边界填充。其中,在每个反卷积层进行反卷积操作前,都使用批次标准化 BatchNorm 和激活函数 Leaky ReLU,将特征归一化

与非线性整合,以加速网络训练和提升训练的稳定性。最终,经过 4 个反卷积层后,网络输出得到与历史图像尺寸(112,112)相同的生成图像。

判别器与生成器形成模型对称,同样设置 4 层卷积核大小为 3×3,步长 $s=2$ 的卷积层,并进行 $p=1$ 的边界填充,经过每一层卷积操作完成特征提取后,对得到的特征进行 BatchNorm 批次标准化以及 Leaky ReLU 激活函数的非线性整合。然后,判别器网络中加入了 Dropout,可在一定程度上避免网络过拟合。最后,将提取到的特征图经过全连接层输出映射为一个值,并将这个值经过 sigmoid 激活函数后,输出该判别器网络对输入图像数据类别的概率值。

3.2.2 CNN 模型搭建

CNN 与 DCGAN 的判别器模型设计搭建原理大体相似,不同的是本文的 CNN 采用 ReLU 激活函数,并且比 DCGAN 的判别器多了池化层的设计,池化层通常位于卷积层之后,用于降低卷积层的尺寸,从而降低网络训练复杂度,但是,池化层在降低网络复杂度的同时会丢失掉图像的许多特征。由于声呐图像分辨率低,应用于声呐图像识别的 CNN 应使用较少的池化层,池化方法可以选取最大池化。本文设计搭建的 CNN 模型如表 4 所示。

表 4 CNN 模型

层数	操作名称	参数设置
1	卷积	卷积核大小:3×3
	批次标准化	步长:2
	ReLU 激活	边界填充数:1
2	最大池化	池化核大小:3×3
		步长:2
3	卷积	卷积核大小:3×3
	批次标准化	步长:2
	ReLU 激活	边界填充数:1
4	卷积	卷积核大小:3×3
	批次标准化	步长:2
	ReLU 激活	边界填充数:1
5	卷积	卷积核大小:3×3
	批次标准化	步长:2
	ReLU 激活	边界填充数:1
6	全连接	
	Sigmoid 激活	神经元数目:600
	Dropout	

3.3 改进的 GR 增量网络实验分析

为验证本文提出的改进的 GR 增量网络的类别

增量学习性能,使用该网络对水下目标声呐图像数据集进行类别增量学习训练,每一个任务目标训练完成后,都用当前学习过的所有类别数据进行测试,得到网络对新任务目标和旧任务目标的识别率变化如图 6 所示。

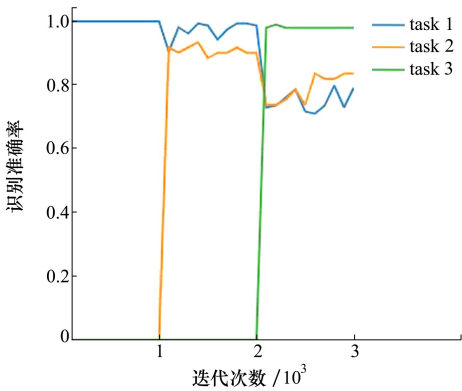


图 6 改进的 GR 增量网络的识别率

将 GR 增量网络与改进的 GR 增量网络的平均识别准确率进行对比,网络改进前后的类别增量识别率结果如表 5 所示。

表 5 增量网络改进前后的类别增量识别率

任务类别	GR 增量网络 识别率/%			改进的 GR 增量网络 识别率/%		
	第一次训练	第二次训练	第三次训练	第一次训练	第二次训练	第三次训练
1	89.9	50.5	27.3	100	98.6	79.3
2		96.8	79.6		90	83.6
3			96			98

采用改进的 GR 增量网络,第一个任务进入增量网络训练完成后,网络对所有目标类别的平均识别率为 100%。当第二个任务作为新类别数据进入增量网络训练完成后,网络对所有目标类别的平均识别率为 94.3%,对历史学习的第一个任务的识别率从 100%降至 98.6%,对第二个任务的识别率为 90%。当第三个任务作为新类别数据进入增量网络训练完成后,对当前所有目标类别的平均识别率为 87%,对历史学习的第一个任务的识别率从 98.6%降到了 79.3%,对历史学习的第二个任务的识别率从 90%降到了 83.6%,对第三个任务的识别率为 98%。改进的 GR 增量网络对当前所有任务目标的

平均识别准确率始终高于 GR 增量网络,平均识别率相比提高了 19.4%。改进的 GR 增量网络对于历史任务目标的平均识别率相比 GR 增量网络提升了 34.7%。

为验证本文提出改进的重构模型和识别网络分别对于声呐图像类别增量学习的作用,将重构模型 VAE 和 DCGAN 与识别网络 MLP 和 CNN 分别进行组合,重构模型和识别网络分别改进前后的平均识别率和参数量如表 6 所示。

表 6 增量网络改进前后的平均识别率和参数量

增量网络结构	平均识别率/%	参数量/ 10^6
VAE+MLP	67.6	41
DCGAN+MLP	79.4	21.5
VAE+CNN	78.8	28.1
DCGAN+CNN	87	8.6

由表 6 可得,相比 VAE+MLP,DCGAN+MLP 的平均识别率提高了 11.8%,参数量降低了 47.5%;VAE+CNN 的平均识别率提升了 11.2%,参数量降低了 31.5%;DCGAN+CNN 的平均识别率提高了 19.4%,参数量降低了 79%。相比单独改进识别网络,单独改进重构模型对 GR 增量网络平均识别率的提高和参数量的降低更加明显,而 DCGAN+CNN 是最佳的 GR 增量网络结构。

4 实验验证与分析

4.1 实验数据集的建立

为验证本文提出的改进 GR 增量网络对水下目标声呐图像类别增量学习的泛化性,通过外场试验建立水下目标声呐图像数据集如表 7 所示,将实验数据集随机划分为图 7 所示的 3 个任务目标。

表 7 实验数据集

样本分布	蛙人	圆柱	拦截网	锚链	T 型管	AUV
样本数	253	253	217	253	253	253
训练集	203	203	174	203	203	203
测试集	50	50	43	50	50	50

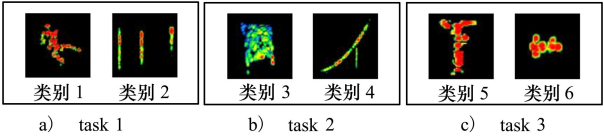


图 7 任务目标

4.2 增量网络泛化性实验

分别采用 GR 增量网络和改进的 GR 增量网络对实验数据集进行类别增量学习训练,得到改进前后的 GR 增量网络对所有任务目标的识别率如表 8 所示。

表 8 增量网络改进前后的类别增量识别率

任务类别	GR 增量网络 识别率/%			改进的 GR 增量网络 识别率/%		
	第一次 训练	第二次 训练	第三次 训练	第一次 训练	第二次 训练	第三次 训练
1	88.9	63.6	49.5	100	95.3	76.1
2		96.8	90.3		93.5	88.5
3			95			96

本文提出的改进的 GR 增量网络相比 GR 增量网络,第一次训练后的平均识别率提高了 11.1%。第二次训练后的平均识别率提高了 14.2%。第三次训练后,对所有任务目标的平均识别率提升了 8.6%。对历史任务目标的平均识别率提升了 18.8%。

对于本文建立的实验数据集,采用经典增量学习方法与本文提出的改进 GR 增量网络进行对比,各种增量学习方法的识别率结果如表 9 所示。

由表 9 可得,基于正则化的 EWC 和 LWF 方法对于声呐图像类别增量学习表现较差,当网络训练

完成后,对所有任务目标平均识别率均低于 45%。相比而言,GR 增量网络总体表现更好,尤其是本文提出的改进 GR 增量网络,对所有任务目标的平均识别率达到 86.9%。

表 9 各种增量学习方法的识别率

增量学习方法	所有任务目标的平均识别率/%
EWC	36.94
LWF	44.63
GR	78.3
改进的 GR	86.9

5 结 论

本文基于深层卷积生成对抗网络与卷积神经网络理论,对 GR 增量网络的重构模型与识别网络进行改进,提出了一种改进的 GR 增量网络,对所有声呐图像任务目标的平均识别率提升了 19.4%,对历史任务目标的平均识别率提升了 34.7%。采用外场实验获得的水下目标声呐图像实验数据集进行了网络泛化性实验,改进的 GR 增量网络对所有任务目标的平均识别率提升了 8.6%,对历史任务目标的平均识别率提升了 18.8%。上述结果表明,本文提出的改进 GR 增量网络对于声呐图像具有良好的类别增量学习性能,较好地缓解了对历史任务目标的灾难性遗忘问题。

参考文献:

[1] LIU Y, SCHIELE B, SUN Q. Adaptive aggregation networks for class-incremental learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 2544-2553

[2] 张智敏. 基于增量学习的分类算法研究[D]. 广州: 华南理工大学, 2010

ZHANG Zhimin. The study of classification based on incremental learning[D]. Guangzhou: South China University of Technology, 2010 (in Chinese)

[3] LI Z, HOIEM D. Learning without forgetting[J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2017, 40(12): 2935-2947

[4] KIRKPATRICK J, PASCANU R, RABINOWITZ N, et al. Overcoming catastrophic forgetting in neural networks[J]. Proceedings of the National Academy of Sciences, 2017, 114(13): 3521-3526

[5] SHIN H, LEE J K, KIM J, et al. Continual learning with deep generative replay[J]. Advances in Neural Information Processing Systems, 2017, 30: 2994-3003

[6] 陈禹乐, 李博, 梁红, 等. 小样本下基于深度学习的声呐图像分类研究[J]. 西北工业大学学报, 2022, 40(4): 739-745

CHEN Yule, LI Bo, LIANG Hong, et al. Research on sonar image few-shot classification based on deep learning[J]. Journal of Northwestern Polytechnical University, 2022, 40(4): 739-745 (in Chinese)

[7] 盛子旗, 霍冠英. 样本仿真结合迁移学习的声呐图像水雷检测[J]. 智能系统学报, 2021, 16(2): 385-392

- SHENG Ziqi, HUO Guanying. Detection of underwater mine target in sidescan sonar image based on sample simulation and transfer learning[J]. CAAI Transactions on Intelligent Systems, 2021, 16(2): 385-392 (in Chinese)
- [8] KINGMA D P, WELING M. Auto-encoding variational Bayes[C] // International Conference on Learning Representations, 2014
- [9] TANG T, DENG C, HUANG G B. Extreme learning machine for multilayer perceptron[J]. IEEE Trans on Neural Networks and Learning Systems, 2016, 27(4): 809-821
- [10] BOND-TAYLOR S, LEACH A, LONG Y, et al. Deep generative modelling: a comparative review of VAEs, GANs, Normalizing Flows, Energy-Based and Autoregressive Models[J]. IEEE Trans Pattern Analysis and Machine Intelligence, 2022, 44(11): 7327-7347
- [11] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks[C] // International Conference on Learning Representations, 2016
- [12] LECUN Y, BENGIO Y, HINTON G E, et al. Deep learning[J]. Nature, 2015, 521(7553): 436-444
- [13] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90

Research on a class-incremental learning method based on sonar images

CHEN Xinzhe, LIANG Hong, XU Weiyu

(School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: Due to the low resolution and the small number of samples of sonar images, the existing class incremental learning networks have a serious problem of catastrophic forgetting of historical task targets, resulting in a low average recognition rate of all task targets. Based on the framework model of generated replay, an improved class incremental learning network is proposed in this paper, and a new deep convolution generative adversarial network is designed and built to replace the variational autoencoder as the reconstruction model of generated replay incremental network to improve the effect of image reconstruction; a new convolution neural network is constructed to replace the multi-layer perception as the recognition network of generated replay incremental network to improve the performance of image classification and recognition. The results show that the improved generated replay incremental network alleviates the problem of catastrophic forgetting of historical task targets, and the average recognition rate for all task targets is significantly improved.

Keywords: sonar images recognition; generated replay; class-incremental learning

引用格式: 陈鑫哲, 梁红, 徐微雨. 基于声呐图像的分类增量学习方法研究[J]. 西北工业大学学报, 2023, 41(2): 303-309
CHEN Xinzhe, LIANG Hong, XU Weiyu. Research on a class-incremental learning method based on sonar images[J]. Journal of Northwestern Polytechnical University, 2023, 41(2): 303-309 (in Chinese)