

基于样本扩充网络的水声目标分类模型优化算法

张博轩¹, 赵天白¹, 常振兴², 蒋翔宇³, 王少博¹

(1. 中国电子科技集团公司第54研究所, 石家庄 050081;

2. 国网电力空间技术有限公司, 北京 102213;

3. 电子科技大学信息与通信工程学院, 成都 610731)

摘要: 水声目标识别是近年来各国的研发热点, 但是由于水声目标难以采集而导致样本数据不足, 严重影响了神经网络的识别效率以及自动化识别装备的水平和性能的发挥; 为此, 提出了一种基于样本扩充网络的水声目标分类模型优化方法, 通过搭建掩模重建的样本扩充网络, 充分利用无标注数据进行训练, 使模型学习到样本的全局高维特征, 再生成样本加入后续的识别模型训练中, 在两次试验过程中, 平均识别准确率从76%提升至80%, 最佳识别准确率从88%提升至96%; 基于实测数据的实验表明, 该方法提升了分类器的准确率、收敛速度以及稳定性。

关键词: 水声目标识别; 样本扩充网络; 循环对抗生成网络; 掩码训练; 梅尔倒谱系数

Optimization Method of Underwater Acoustic Target Classification Model Based on Sample Expansion Network

ZHANG Boxuan¹, ZHAO Tianbai¹, CHANG Zhenxing², JIANG Xiangyu³, WANG Shaobo¹

(1. The 54th Research Institute, China Electronics Technology Group Corporation, Shijiazhuang 050081, China;

2. State Grid Power Space Technology Co., Ltd., Beijing 102213, China;

3. College of Communication and Information Engineering, University of Electronic Science and Technology of China, Chengdu 610731, China)

Abstract: Underwater acoustic target recognition is a research and development hot spot in many countries in recent years. However, it is difficult to collect underwater acoustic targets, resulting in the sample data insufficient, which seriously affects the recognition efficiency of neural network and the level and performance of automatic recognition equipment. Therefore, an optimization method of underwater acoustic target classification model based on the sample expansion network is proposed. Through building the sample expansion network reconstructed by the mask, the unlabeled data is fully utilized to train the model, learn the global high-dimensional features of the samples, and then generate the samples to be added to the subsequent recognition model training. Based on the results of two experiments, the average accuracy of target classification model improves from 76% to 80%, with its maximum accuracy improving from 88% to 96%. Experimental results show that this method improves the accuracy, convergence speed and stability of the classifier.

Keywords: underwater acoustic target recognition; sample expansion network; cycle generative adversarial networks; mask training; mel-scale frequency cepstral coefficients

0 引言

水声目标识别技术是当今世界各国研究的热点和难点, 随着新型低噪声推进器的研发和应用, 如何在复杂的海洋环境中精确发现、定位和识别水声目标是各国海军部门都高度重视的问题。传统的水声目标识别方法主要利用经典信号处理方法提取目标特征并进行匹配, 以此为依据对目标身份进行分类^[1]。这类方法受人为经验影响较大, 自适

应能力差, 在复杂多变的海洋环境中识别准确率较低。因此, 针对水中舰船目标机械辐射噪声信号的人不在回路的识别成为水声信号处理发展的热点, 尤其是如何将基于人工智能方法研发的目标识别算法模型从实验室扩展至实际环境中并依旧取得良好、稳定的性能, 是此类方法研究过程中的重中之重。

将人工智能技术引入水声目标识别系统这一思路尚未取得突破性进展, 无法在海洋实战中实现应用赋能, 其根

收稿日期: 2023-04-24; 修回日期: 2023-06-02。

基金项目: 国家自然科学基金项目(U20B2071)。

作者简介: 张博轩(1992-), 男, 硕士, 工程师。

通讯作者: 赵天白(1993-), 女, 硕士, 工程师。

引用格式: 张博轩, 赵天白, 常振兴, 等. 基于样本扩充网络的水声目标分类模型优化算法[J]. 计算机测量与控制, 2024, 32(4): 143-150.

本原因是深度神经网络模型若想达到对数据的高效、准确的智能化处理, 绝离不开用大量的数据来训练模型参数。但是, 由于绝大部分高价值水声目标现存数据稀少, 敏感目标难以采集, 导致可供训练的声学数据样本数据严重不足, 从而影响了该方法的识别效率。传统的水声目标识别方法不需要大量的声呐信号数据, 如线谱分析法、匹配滤波法等。虽然这些方法在仿真中都取得了较好的识别效果, 但它们对声呐的信噪比或先验因素要求较高, 对不同类型的声信号的辨析能力相差较大且检测速度较慢, 难以应用于实时且噪声较复杂的场景。

在有限样本闭集的前提下, 文献 [2] 等采用支持向量机 (SVM, support vector machine) 对鲸鱼声数据集进行处理, 达到了 94.5% 的多类识别准确率; 文献 [3] 基于 SVM 提出差分进化粒子群混合算法, 调优了核函数和惩罚项参数, 以识别不同水下目标的辐射噪声; 文献 [4] 团队提出梯度提升机 (AdaBoost) 和 SVM 混合模型, 对水面舰船辐射噪声目标信号进行识别, 在样本和特征数量减少 50% 的情况下, 取得相同的识别准确率; 然而, 这些方法在面对样本分布不均衡或总体样本较少的情况时, 性能下降严重, 而这两点是在真实的海洋舰船机械辐射噪声信号侦收过程中普遍存在的问题。因此, 在有效信号采集较少的情况下, 如何让水声目标探测与识别装备保持较高的分类准确率仍是一个极具挑战性的问题。

针对这一问题, 在深度学习领域已有相关思路, 系主要通过通过对数据进行变换、生成等操作来解决样本缺失问题。文献 [5] 提出了生成对抗网络 (GAN, generative adversarial network), 将源域映射到目标域; 文献 [6] 将特征提取器和分类器分别作为标准 GAN (standard generative adversarial network)^[7]、最小二乘 GAN^[8] (least-squares generative adversarial network) 和相对 GAN^[9] (relativistic generative adversarial networks) 中的生成器和判别器, 并用 GAN 的基本训练方式引导其交替更新参数, 从而使各个 GAN 变体模型学习到源域目标域指定特征; 文献 [10] 设计了部分共享参数的特征提取器, 以获得声纹特性的更好表征。文献 [11] 则采用循环一致对抗生成对抗网络 (CycleGAN, cycle generative adversarial networks) 学习源域与目标域无标签数据特征之间的映射, 将源域带标签数据转化至目标域进行模型训练, 从而保证目标域声纹数据能够被准确识别^[12]。

基于上述分析, 在目标的数据样本集总量不足或分布不均时, 如何利用先验信息或现有的少量样本生成大量的、有效的充足样本集以辅助模型训练, 从而提升识别模型的稳定和泛化性能, 是制约深度学习赋能于水声目标识别系统的瓶颈问题, 也是深度神经网络模型在水声目标识别问题中迈向实际应用的必经之路。

为解决该问题, 本文提出一种基于样本扩充网络的水声目标分类模型优化算法。该方法首先对实际侦收的水声信号进行预处理, 提取其梅尔倒谱系数和时频谱图后, 利

用基于 CycleGAN 的样本扩充网络和少量的真实数据实现样本生成, 加入后续的深度学习神经网络训练中, 从而降低后续训练的迭代次数和收敛难度, 并提升模型分类识别的准确率、稳定性与泛化能力。实验表明该方法在声呐信号样本较少的情况下, 通过网络进行样本扩充, 有效提高了神经网络模型的识别准确度, 收敛速度和稳定性^[13]。

1 CycleGAN 网络原理

CycleGAN 是 GAN 网络的一个重要分支, 通过在源域目标域互相映射, 生成更符合需求的样本。CycleGAN 的原理从本质上分析是利用两个相互镜像对称的环形生成对抗网络结构, 通过在源域、目标域的 4 个判别器和两个生成器中建立映射, 以实现样本的传递和循环。并对判别器进行针对性设计, 以适应环形结构带来的改变。

具体地, 该网络由两个生成器和 4 个判别器组成, 多个模块需要不同的约束条件来训练参数。其损失函数如下:

$$L_{full} = L_{adv}^{X \rightarrow Y} + L_{adv}^{Y \rightarrow X} + \lambda_{cyc} (L_{cyc}^{X \rightarrow Y \rightarrow X} + L_{cyc}^{Y \rightarrow X \rightarrow Y}) + \lambda_{id} (L_{id}^{X \rightarrow Y} + L_{id}^{Y \rightarrow X}) + L_{cycleadv}^{X \rightarrow Y \rightarrow X} + L_{cycleadv}^{Y \rightarrow X \rightarrow Y} \quad (1)$$

其中: 前两项是不同的对抗损失, 第三项为身份映射损失, 而第四项为循环一致损失。对抗损失是基础的 GAN 损失, 在对损失值进行极大极小化的博弈过程中提升生成器的生成能力, 以得到符合需要的生成样本。循环一致性损失使网络循环生成样本与原始输入样本的特征保持一致, 从而保证在两个生成器的计算过程中保留全部海况信息以及船只信息。身份保持损失与循环一致性损失相似, 均用以训练生成器。在训练过程中, 需将这四类损失按权重相结合作为最终优化函数, 确保模型向目标方向学习并逐步收敛。

为保证生成样本的多样性和更好的训练网络, 本文通过加入掩码来训练 CycleGAN。在加入掩码后, CycleGAN 的损失函数需要更新为:

$$L_{adv}^{X \rightarrow Y} = E_{y \sim p_y} [\log D_y(y)] + [\log(1 - D_y(G_{X \rightarrow Y}(\text{concat}(\hat{x}, m))))] \quad (2)$$

$$L_{cycleadv}^{X \rightarrow Y \rightarrow X} = E_{x \sim p_x} [\log D'_x(G_{Y \rightarrow X}(\text{concat}(y', m')))] \quad (3)$$

$$L_{cyc}^{X \rightarrow Y \rightarrow X} = E_{x \sim p_x} [\|G_{Y \rightarrow X}(G_{X \rightarrow Y}(\text{concat}(x, m'))) - x\|_1] \quad (4)$$

$$L_{id}^{X \rightarrow Y} = E_{y \sim p_y} [\|G_{X \rightarrow Y}(\text{concat}(\hat{y}, m)) - y\|_1] \quad (5)$$

以上损失函数均为原样本与掩码 m 点乘得到。

2 基于样本扩充网络的识别模型优化方法

本文基于 CycleGAN 搭建了如图 1 所示的样本扩充网络, 其主要包括预处理、扩充网络和分类器 3 个部分。接下来将对模型的每一部分进行详尽的阐述。

2.1 预处理

声信号识别中, 音频信号中的频率成分是随着时间推移而变化的, 无法直接对整段音频提取特征。因此, 本文需要首先对长时间的音频信号进行分帧处理, 利用短时傅里叶变换 (STFT, short-time Fourier transform) 等分析手

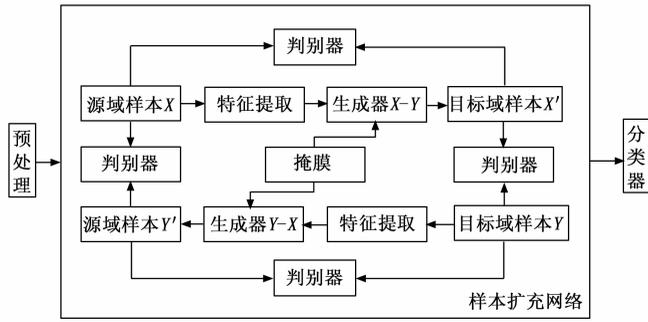


图 1 样本扩充网络框架

段对信号进行解析, 以兼顾在不同维度下短帧信号的时频分辨率。

将长时间信号进行分帧, 本文利用滑动窗加权的方式实现。常用的窗口包括 Kaiser 窗、Blackman 窗、Hanning 窗、Hamming 窗、矩形窗等, 本文采用的窗函数是 Hamming 窗。

窗函数的数学表达式为:

$$w(n) \begin{cases} 0.5 - 0.5 \cos[2\pi n / (N - 1)], & 0 \leq n \leq (N - 1) \\ 0, & \text{else} \end{cases} \quad (6)$$

针对时域信号, 进行了加窗之后的信号输出为:

$$O_i = x_i(n) * w(n) \quad (7)$$

考虑到水声信号为单通道音频, 可以通过谐波分量和冲激分量分解 (HPSS, Harmonic and percussive source separation), 将单通道音频转换为双通道立体声, 其中两个通道分别为谐波分量与冲激分量, 作为新的音频输入供后续特征提取使用。HPSS 的具体步骤如图 2 所示。

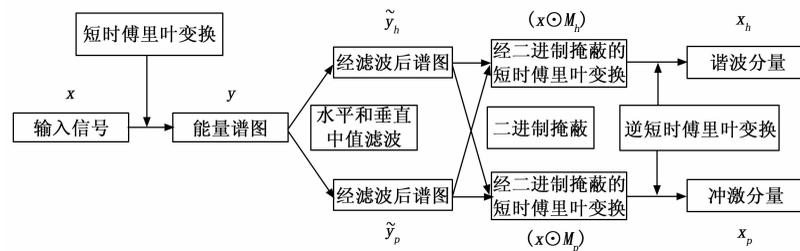


图 2 HPSS 提取步骤

首先对输入的时域音频信号预处理, 先降噪再进行短时傅里叶变换, 变换到频域, 以获得频谱图, 然后对频谱图进行分析, 来识别含有谐波成分的垂直线, 对应的音频为持续音高。并使用中值滤波来提取和增强谐波成分 \tilde{y}_h 和冲激成分 \tilde{y}_p :

$$\begin{aligned} \tilde{y}_h &= medfilt_h(y) \\ \tilde{y}_p &= medfilt_p(y) \end{aligned} \quad (8)$$

而在使用中值滤波时, 我们使用了 01 掩蔽矩阵 (binary masking) 的方式, 对谱图进行处理, 掩蔽矩阵 M_h 和 M_p 的构建方式如下:

$$\begin{aligned} M_h(m, k) &= \begin{cases} 1, & \tilde{y}_h(m, k) \geq \tilde{y}_p(m, k) \\ 0, & \text{else} \end{cases} \\ M_p(m, k) &= \begin{cases} 1, & \tilde{y}_p(m, k) \geq \tilde{y}_h(m, k) \\ 0, & \text{else} \end{cases} \end{aligned} \quad (9)$$

使用掩蔽矩阵 M_h 和 M_p 将 STFT 中所有的时域-频域容器划分到为谐波分量或冲激分量中, 进而从原始的 STFT 结果划分出谐波分量 X_h 和冲激分量 X_p :

$$\begin{aligned} \chi_h &= (\chi \odot M_h) \\ \chi_p &= (\chi \odot M_p) \end{aligned} \quad (10)$$

最后, 通过逆快速傅里叶变换 (iSTFT) 将修改后的频谱图 $X_h \chi_h$ 和 χ_p 反变换回时域, 产生分离的谐波和冲激信号。

利用深度学习进行信号处理时, 直接对经过预处理之后的水声信号进行采样, 再将数据合成一维向量作为深度学习模型的输入的方法不但难以深度挖掘频域信息, 而且采样数据往往受水下噪声的影响较大, 不利于分类器对水中目标的精确分类和识别。

而时频图像包含了水中目标的时域、频域、幅度的三维信息, 更能从多个维度表示出水中目标的特征, 包含的信息量更加丰富, 同时, 作为二维图像, 适合利用卷积神经网络 (CNN, convolutional neural network) 进行特征的提取和分类, 故本文将时频图像和梅尔频率域的倒谱系数谱图作为表征水中目标的数据作为深度学习网络的输入。

对初步处理后的一维时域音频信号求取其梅尔倒谱系数 (MFCC, mel-scale frequency cepstral coefficients)。MFCC 参数是参照人体听觉的临界带效应设计而成, 目前在语音领域被广泛应用, 具有良好实践效果^[14]。

MFCC 谱图的提取流程图如下, 首先对信号进行分帧

加窗, 从而使信号局部平稳, 而后通过离散傅里叶变换获得能量谱。在获得音频的能量谱的同时, 还需要构造一个梅尔滤波器组, 并与能量谱进行点积运算得到梅尔频谱图。梅尔滤波器组的作用是将能量谱转换为更接近人耳机理的梅尔频率, 本文中设置梅尔滤波器的个数为 128。滤波过程的公式如下:

$$E(m) = \sum_{k=1}^{N-1} (p(f) * H_m(f)) \quad (11)$$

其中: N 表示每一帧的总点数, $H_m(f)$ 为梅尔滤波器组系数。

MFCC 与时频图类似, 区别在于时频图为线性谱图, 而 MFCC 的频域轴为非等间距划分, 达到了展现更多低频细节的目的。同时 MFCC 的滤波器组构建也不同于时频图的正弦波, 而是三角滤波器。MFCC 的计算公式如下:

$$E'(m) = \lg E(m) \quad (12)$$

为了适配模型的图像处理性能并提升对比模型的易适配度, 本文选取了包含大量目标时频域信息的 LOFAR 谱作为模型的输入样本^[15]。信号的预处理和特征提取流程如下: 首先对水听器阵列采集的多阵元时域信号进行空域^[16]、

频域滤波, 随后对得到的一维声波序列信号进行滤波得到一组帧序列 $\{L_1(n), L_2(n), \dots, L_k(n)\}$, 对每一帧 $L_k(n)$ 进行标准化和中心化处理:

$$y_k(n) = \frac{L_k(n)}{\max \sum_{1 \leq i \leq N} [L_k(i)]} \quad (13)$$

$$x_k(n) = y_k(n) - \frac{1}{N} \sum_{i=1}^N y_k(i) \quad (14)$$

其中: $y_k(n)$ 代表标准化后的信号, N 代表帧长度, $x_k(n)$ 代表中心化后的信号。将信号的采样序列分为连续的若干帧, 每帧 N 个采样点。接着对每帧信号做归一化和中心化处理, 最后对处理后的每帧信号加窗做短时傅里叶变换得到时频谱图。

2.2 扩充网络

扩充网络由 2 个生成器、4 个判别器组成。生成器可分为 3 个模块: 下采样模块、残差提取模块和上采样模块, 如图 3 所示。

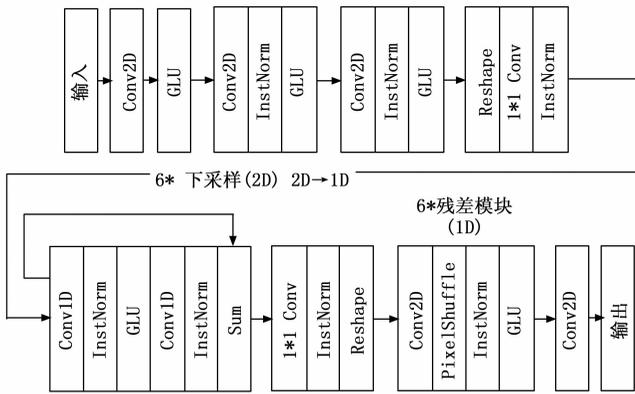


图 3 生成器结构示意图

在上采样模块中, 经过掩码处理的特征向量通过两层卷积层进行特征提取。模块再经过一层 1×1 的卷积层将二维特征转化为一维特征。在残差提取模块中, 网络利用六个相同的残差模块对源域的一维特征进行转化, 保留源域海况信息的同时将源域特征转化为目标域特征。下采样模块通过一层 1×1 的卷积层和三层二维卷积层将一维特征转化为二维的具有目标域特征的特征向量, 其二维尺寸大小与输入的源域合成特征向量一致。具体网络参数见表 1, 其中 n 代表每个批次的样本数量。

4 个判别器结构如图 4 所示。通过一层卷积以及 4 个卷积下采样模块对其进行特征提取, 最后送入一个大小为 1 的输出卷积层, 进行判别。判别器的具体参数见表 2。

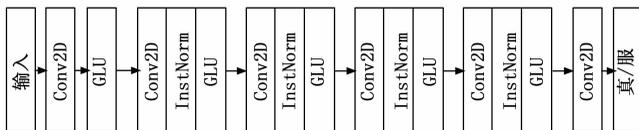


图 4 判别器结构示意图

2.3 分类器

2.3.1 分类器模型

本文使用 KNN^[15]、CNN^[16] 和残差网络^[17] (ResNet, resident network) 3 种基础网络搭建不同分类器。其中 KNN 超参数 $n_neighbors$ 为 9, p 为 1; CNN 共 4 层卷积层, $kernel_size$ 为 3; ResNet18 由 4 个 block 和一个前置卷积层组成, $kernel_size$ 为 5。

表 1 生成器网络结构参数

网络层	卷积核尺寸	输出
输入	—	$[n, 2, 20, 20]$
下采样	—	—
Conv2d, GLU	(5, 15)	$[n, 128, 20, 20]$
Conv2d, InstNorm, GLU	5	$[n, 256, 10, 10]$
Conv2d, InstNorm, GLU	5	$[n, 256, 5, 5]$
Conv1d, InstNorm	1	$[n, 256, 5]$
残差提取	—	—
ResBlock $\times 6$	—	$[n, 256, 5]$
上采样	—	—
Reshape, InstNorm	1	$[n, 1280, 5]$
Conv2d, pixelShuffle, InstNorm, GLU	5	$[n, 256, 10, 10]$
Conv2d, pixelShuffle, InstNorm, GLU	5	$[n, 128, 20, 20]$
Conv2d	(5, 15)	$[n, 1, 20, 20]$

表 2 判别器网络结构参数

网络层	卷积核尺寸	输出
输入	—	$[n, 1, 20, 20]$
下采样	—	—
Conv2d, GLU	(3, 3)	$[n, 128, 20, 20]$
Conv2d, InstNorm, GLU	(3, 3)	$[n, 256, 10, 10]$
Conv2d, InstNorm, GLU	(3, 3)	$[n, 512, 5, 5]$
Conv2d, InstNorm, GLU	(3, 3)	$[n, 1024, 3, 3]$
Conv2d, InstNorm, GLU	(1, 10)	$[n, 1024, 3, 3]$
Conv2d	(1, 3)	$[n, 1, 3, 3]$

在选取模型配置时, 综合了水声目标分类领域的数据集规模和样本特征等因素, 从而选择 ResNet 框架来搭建模型。

与其他卷积神经网络不同, ResNet 的提出是为了解决梯度爆炸和消失问题。随着神经网络的网络层深度不断增加, 但是网络的性能和模型的参数并非线性关系。而在模型层数太深时, 反向传播的梯度信息并不能更新到浅层的参数, 或者梯度太大导致参数不断跳变, 模型失效。

ResNet 的解决方法是引入了“残差学习”(RL, residual learning) 的思想。所谓残差学习, 是指将输入与输出的差值(即“残差”)学习到, 然后将该残差加到输入上, 从而得到输出。这样做的好处是可以让网络更容易地学习到恒等映射 (Identity Mapping), 从而避免了梯度消失和爆

炸的问题。

具体来说, ResNet 通过在网络中添加“残差块”(Residual Block)来实现残差学习。它包含了两个卷积层和跨层连接,负责从输入数据中学习和提取特征。通过两个卷积层来保持输入输出的维度不改变,然后通过跨层连接来规避模型在遇到简单任务时拟合难的问题。通过这种结构,既能让模型的梯度传播到浅层网络,也能在更低的维度空间中持续提高模型的学习能力。

ResNet18 是 ResNet 系列的一个变体。与其他变体相比, ResNet18 的架构较浅,有 18 层。虽然它的层数较少,但仍然具有强大的特征提取能力。

在训练数据有限的情况下, ResNet18 具有较浅的结构,可以帮助减少过度拟合,提高泛化性能。并且 ResNet18 在模型容量(捕捉复杂模式和特征的能力)和模型复杂性(参数和层的数量)之间取得了平衡。它提供了一个合理的深度架构,能够有效地捕捉和学习有用的特征,而不会变得过于复杂或容易过度拟合。

2.3.2 神经网络结构参数的确定

鉴于本文将经过预处理和特征提取的信号谱图作为多模态深度神经网络模型的输入,将根据获得的信号数据决定网络结构的具体参数。

对于 CNN 和 ResNet 来说,其搭建过程中有关的主要参数如下:

- 1) 卷积层的卷积核大小、卷积核个数;
- 2) 激活函数的种类(常用的如 Sigmoid, tanh, relu, leaky relu);
- 3) 网络的层结构(卷积层的个数/全连接层的个数);
- 4) 池化方法的种类;
- 5) Dropout 的概率;
- 6) 有无归一化。

通过神经网络使用的经验可以得到,在调整参数时,重要的是先调整网络的层结构、卷积层的卷积核个数以及激活函数的种类,其他参数虽然也会对性能或多或少地产生影响,但是差异不大,所以首先确定重要参数,然后再对其他参数进行微调即可。

我们根据信号包含信息的丰富程度来调整 CNN 的结构以及每一层的卷积核个数。由于提取的 LOFAR 和 MFCC 谱图具有概括性的特点, CNN 的层数不宜过多,以三层为初始值根据实际实验结果进行微调。若声纳数据中只包含单个识别的目标,我们将用适当的卷积核数量,在增强网络对于多层次信息的提取能力的同时避免过拟合问题。对于池化层,由于水声数据具有较强噪声,而可选取最大池化,以突出信号特征。

考虑到该网络模型在深层特征提取方面的优势,构建多层 ResNet。而适当增加隐藏层向量的维度有利于增强向量对水声目标特征的代表能力,提高分类识别率。同时为了为了避免过拟合,在 ResNet 网络中采用了迁移学习的思想,即将该模型参数的初始值设定为在 ImageNet 数据集经过训

练的模型参数,有利于模型收敛至最优值。对于池化层,选取平均池化,以保留丰富的信号特征。

对于激活函数,常用的激活函数种类包括 Sigmoid, tanh, ReLU, leaky ReLU 等。Sigmoid 和 tanh 激活函数的缺点在于当输入值过大或过小时易出现梯度消失的现象,不利于神经网络的训练,而相比之下,Leaky ReLU 和 ReLU 可以很好地避免这一现象,因此,采取 ReLU 作为神经网络的激活函数。

再确定以上 3 个参数后,对于卷积神经网络的其他参数(如卷积核大小,池化方法等)采用实验微调的方法,寻找在此问题场景下的最佳参数值。

2.3.3 神经网络训练策略及优化方式

对于水中目标识别问题而言,需要搭建的模型往往较为复杂,批量梯度下降法计算量过大,消耗的内存过多,随机梯度下降法易陷入局部最优,从而导致模型过拟合,因此,两种算法均不适用。Adam 算法则将批量梯度下降法和随机梯度下降法二者相结合,利用一阶矩估计和二阶矩估计可以实现学习率的自适应调整。Adam 算法本质上利用梯度的一阶矩估计和二阶矩估计动态调节每个参数的学习率。在经过偏差修正后,每一次迭代都有确定的范围,使得参数比较平稳,善于处理稀疏梯度和非平稳目标。与此同时,算法对内存的需求小,且较适合应用于大规模数据及参数的场景。

3 实验设计

3.1 平台及数据集

本文实验的运行平台:操作系统为 Ubuntu20.04, GPU 为 NVIDIA RTX 3090,内存大小为 16 G。开发语言和平台为 Python3.8 和 Pytorch1.19。

本文实验采用两个数据集,均为舰船噪声信号,每个样本持续 3 秒,采样率为 5 000。数据集 A 包含 6 种不同海洋条件下的 3 类舰船辐射噪声。如表 3 所示,数据集分为训练集和测试集,训练集共有 7 046 个样本,测试集共有 1 459 个样本。训练集为第 1~5 天采集数据,测试集为第 6 天采集数据。

表 3 数据集 A 详细

数据类型	样本数目	
	训练集	测试集
1	2 256	548
2	2 556	551
3	2 234	360

数据集 B 包含 4 类船只在不同海况、航行状态的辐射噪声,单个样本的平均持续时间为 3.3 s。总共有 11 307 个样本,具体情况如表 4 所示。各种海洋条件下的样本均匀分布在整个数据集中。每一类样本按 1:1 的比例分为训练集和测试集。

3.2 网络超参数配置

实验通过对比不同参数设置下的网络性能,得到了最

佳参数配置, 具体为: 生成器和鉴别器的初始层卷积核大小为 5×5 , 卷积层数分别为 12 和 5, 训练批次大小为 64、训练轮次为 160、掩码长度 25。

表 4 数据集 B 详细

数据类型	样本数目	
	训练集	测试集
1	1 988	1 989
2	1 648	1 649
3	1 962	1 962
4	55	56

3.3 评估指标

本实验采用 3 种评价指标来验证算法的性能, 分别为声纹识别准确率、损失函数 $loss$ 值、余弦相似度。

余弦值匹配算法^[18]利用向量空间中两个向量夹角的余弦值作为衡量两个个体间差异的大小。余弦值越接近 1, 就表明夹角越接近 0° , 也就是两个身份特征向量越相似, 越有可能属于同一个目标个体。公式定义为:

$$\cos\theta = \frac{\sum_{i=1}^n (A_i \times B_i)}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} = \frac{A \cdot B}{|A| \times |B|} \quad (15)$$

4 实验结果与分析

本文总共两个独立实验, 分别使用 A、B 两个数据集。实验一旨在验证扩充网络对分类器的整体性能的提升效果。实验二旨在验证扩充网络对分类器识别小样本类的提升效果。

4.1 实验一

首先通过提取舰船噪声的 MFCC 谱图^[19]或 LOFAR 谱图特征, 训练样本扩充网络, 让网络能够生成目标域样本的特征谱图, 再通过反变换生成一维音频信号。本次实验在数据集 A 的三类样本中互相生成对应的声信号样本, 使得三类样本训练集的数据量均为 7 046, 测试集不做变化。

实验中, 根据生成器和判别器 $loss$ 的误差值判断网络的收敛速度^[20]。鉴别器的 $loss$ 曲线呈现先上升后下降趋势, 在 96 轮次后基本收敛。生成器的 $loss$ 曲线总体呈现下降的趋势, 在 120 轮次后基本收敛。

在样本扩充网络训练完成后, 分别对 1, 2, 3 类生成样本, 并与目标类计算 LOFAR 谱图的余弦相似度, 如表 5 所示。

表 5 实验一 余弦相似度

	1 类	2 类	3 类
余弦相似度	0.915	0.931	0.949

在完成类间余弦相似度的验证后, 通过 ResNet18 搭建分类器进行对比实验, 比较分类器的准确率曲线, 验证扩充网络有效性, 分类结果如图 5 所示。

分类器共训练 70 轮次, 自上而下第三条曲线是总准确

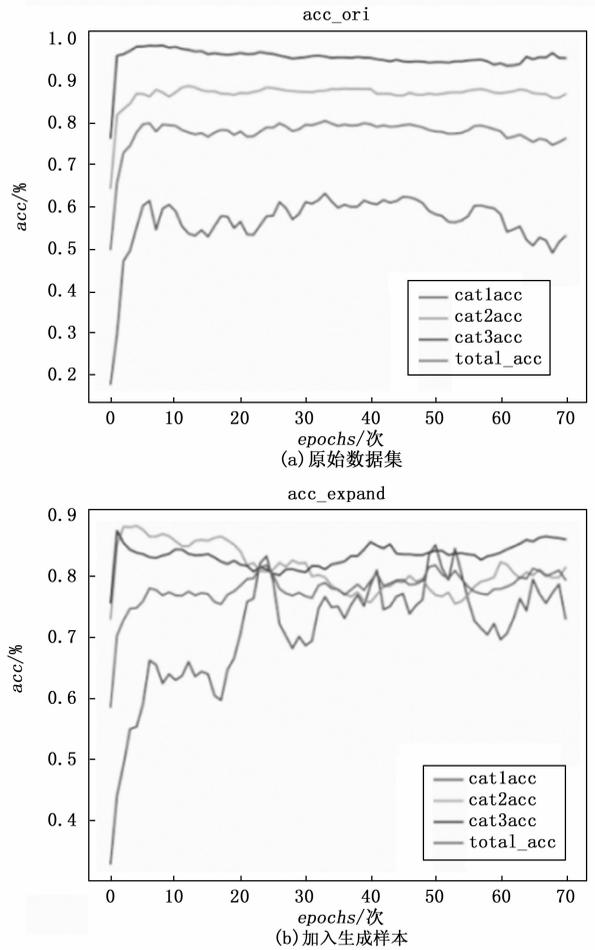


图 5 分类器准确率对比

率。在加入生成样本后, 分类器的总准确率从 76% 提升到 80%。为确保结果的准确性, 实验进一步选取了 3 个不同的分类器进行验证, 分别为 KNN, CNN (4 层卷积), ResNet18, 每个实验重复 5 次, 并在表 6 中给出平均统计结果。

表 6 实验一样本扩充前后分类准确率

谱图特征	分类器	是否加入生成样本训练	准确率/%
MFCC	KNN	Y	48.0
		N	44.0
	CNN	Y	76.6
		N	72.5
	ResNet	Y	65.3
		N	69.6
LOFAR	KNN	Y	52.1
		N	54.7
	CNN	Y	75.9
		N	74.0
	ResNet	Y	82.1
		N	73.8

经过对比实验可以看出, 在不同分类器网络中, 使用

样本扩充网络扩充数据集, 均能提升分类识别效果, 相较于在原始数据集上最佳准确率 74%, 加入生成样本后, 最佳准确率能提升到 82.1%。

4.2 实验二

实验二首先通过提取舰船噪声的 MFCC 谱图/LOFAR 谱图特征, 训练样本扩充网络, 使得网络能够生成目标域样本的 MFCC/LOFAR 谱图。再通过反变换生成声信号样本。与实验一不同的是, 此次实验采用数据集 B 的一半作为训练集, 一半作为测试集, 在样本扩充网络训练完成后, 只针对第 4 类进行扩充。

实验中, 根据生成器和判别器 loss 的误差值判断网络的收敛速度。由于第 4 类样本数量与其他 3 类相差过大, 训练过程中需要多次调整学习率。判别器的 loss 曲线训练时呈现先下降后上升的趋势, 在 300 轮次时基本收敛。生成器的 loss 曲线呈现先上升后下降趋势, 在 450 轮次时基本收敛。

在样本扩充网络训练完成后, 对第 4 类原始样本和生成样本计算余弦相似度, 结果如表 7 所示。

表 7 实验二余弦相似度

	4 类
余弦相似度	0.99

训练好扩充网络后, 通过 ResNet18 搭建分类器进行对比实验, 比较分类器的准确率曲线, 验证扩充网络有效性, 分类结果如图 6 所示。

分类器共训练 30 轮次, 前 4 条曲线是单类的准确率, 自上而下第四条曲线为总准确率。在加入生成样本后, 分类器的综合识别准确率从 88% 提升到了 96%, 且收敛速度加快, 模型也更加稳定。

为确保结果的准确性, 进一步选取了 3 个不同的分类器进行验证, 分别是 KNN, CNN (4 层卷积), ResNet18, 每个实验重复 5 次, 并在表 8 中给出平均结果。

表 8 实验二 样本扩充前后分类准确率

谱图特征	分类器	是否加入生成样本训练	准确率/%	第 4 类准确率/%
LOFAR	KNN	—	Y	92.3
		N	90.5	54.1
	CNN	Y	95.8	75.2
		N	95	61.3
	ResNet	Y	96.5	96.0
		N	97.5	88.0

经过对比实验可以看出, 在不同分类器网络中, 使用样本扩充网络对小样本类进行扩充。相较于在原始数据集上最佳准确率 96.5%, 加入生成样本后, 最佳准确率提升到 97.5%; 针对生成的第 4 类样本, 最佳准确率从 88% 提升到了 96%。

5 结束语

在水声目标识别领域中, 样本分布不均衡或有效样本

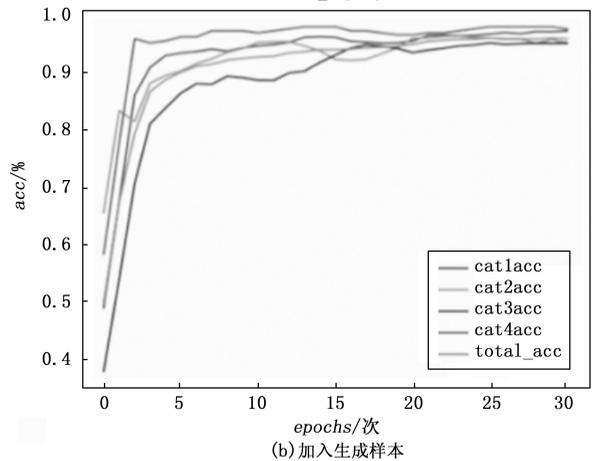
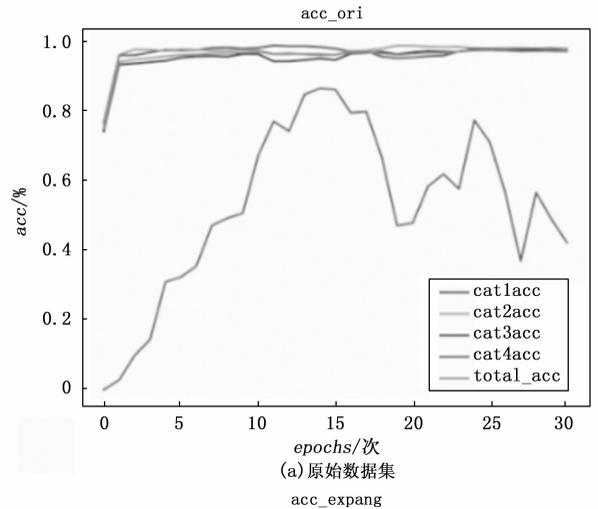


图 6 分类器准确率对比

缺失是一种常见问题。为了解决此问题, 深度学习领域的研究提出了多种方法, 大部分是对模型参数进行配置以实现针对性样本生成。本文提出了一种样本扩充网络, 该算法基于 CycleGAN 网络搭建, 通过学习与目标生成水声数据样本相似环境采集的频谱特征和真实分布, 生成真实样本, 并加入分类器训练中, 提升其性能。

本文设计了基于两个不同水声数据集的对比实验, 通过 3 个分类器对扩充网络进行评估。实验表明, 扩充网络能加速训练和收敛过程, 有效提升分类器的整体性能和对小样本目标类的识别能力。

参考文献:

[1] 戴卫国, 程玉胜, 王易川, 等. 支持向量机对舰船噪声 DEMON 谱的分类识别 [J]. 应用声学, 2010, 29 (3): 206-211.

[2] MCQUAY C, SATTAR F, DRIESSEN P F. Deep learning for hydrophone big data [C] //2017 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM). IEEE, 2017: 1-6.

[3] 孟庆昕, 杨士莪, 于盛齐, 等. 基于波形结构特征和支持向量机的水面目标识别 [J]. 电子与信息学报, 2015, 37 (9):

- 2117 - 2123.
- [4] YANG H, GAN A, CHEN H, et al. Underwater acoustic target recognition using SVM ensemble via weighted sample and feature selection [C] //2016 13th International Bhurban Conference on Applied Sciences and Technology (IBCAST). IEEE, 2016; 522 - 527.
- [5] CORTES C, LAWRENCE N, LEE D, et al. Advances in neural information processing systems 28 [C] //Proceedings of the 29th Annual Conference on Neural Information Processing Systems, 2015.
- [6] BHATTACHARYA G, MONTEIRO J, ALAM J, et al. Generative adversarial speaker embedding networks for domain robust end-to-end speaker verification [C] //2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019; 6226 - 6230.
- [7] ZHOU J, JIANG T, LI L, et al. Training multi-task adversarial network for extracting noise-robust speaker embedding [C] //2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019; 6196 - 6200.
- [8] MAO X, LI Q, XIE H, et al. Least squares generative adversarial networks [C] //Proceedings of the IEEE International Conference on Computer Vision, 2017; 2794 - 2802.
- [9] JOLICOEUR M A. The relativistic discriminator: a key element missing from standard GAN [J]. Arxiv Preprint Arxiv; 1807.00734, 2018; 185 - 187.
- [10] CHEN Z, WANG S, QIAN Y. Adversarial domain adaptation for speaker verification using partially shared network [C] // Interspeech, 2020; 3017 - 3021.
- [11] NIDADAVOLU P S, KATARIA S, VILLALBA J, et al. Low (上接第 105 页)
- [11] 贺波勇, 马鹏斌, 杜卫兵, 等. 部署 retro-GEO 巡视器的月球借力飞行轨道分析 [J]. 中国空间科学技术, 2021, 41 (3): 9 - 15.
- [12] 郭嘉宝, 赵长见, 宋志国. 一种基于动态逆-滑模的变形飞行器姿态控制方法研究 [J]. 航天控制, 2022, 40 (4): 12 - 17.
- [13] 罗忠涛, 夏杭, 卢琨, 等. 超视距雷达中距离-多普勒图的瞬态干扰自动识别方法 [J]. 电子学报, 2021, 49 (7): 1279 - 1285.
- [14] 王春阳, 刘明杰, 曾会勇, 等. 双隐身飞机相干干扰对单脉冲雷达影响及影响有效性研究 [J]. 北京理工大学学报, 2022, 42 (7): 755 - 763.
- [15] 夏悠然, 管军, 易文俊. 基于改进粒子群优化极限学习机的弹丸参数辨识 [J]. 系统工程与电子技术, 2023, 45 (2): 521 - 529.
- [16] 商立群, 李洪波, 侯亚东, 等. 基于特征选择和优化极限学习机的短期电力负荷预测 [J]. 西安交通大学学报, 2022, 56 (4): 165 - 175.
- [17] 李瑞君, 李洁, 何亚雄, 等. 微纳米三坐标测量机三轴垂直度误差测量及建模补偿 [J]. 计量学报, 2022, 43 (4): 452 - resource domain adaptation for speaker recognition using cycle-gans [C] //2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU). IEEE, 2019; 710 - 717.
- [12] 蔡情, 高勇. 一种基于卷积神经网络的快速说话人识别方法 [J]. 无线电工程, 2020, 50 (6): 447 - 451.
- [13] 周晓彦, 王珂, 李凌燕, 等. 基于深度学习的目标检测算法综述 [J]. 电子测量技术, 2017, 40 (11): 89 - 93.
- [14] 赵淑芳, 董小雨. 基于改进的 LSTM 深度神经网络语音识别研究 [J]. 郑州大学学报 (工学版), 2019, 39 (5): 63 - 67.
- [15] HAO Z, BERG A C, MAIRE M, et al. SVM-KNN; discriminative nearest neighbor classification for visual category recognition [C] // 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). IEEE, 2006, 1435 - 1447.
- [16] RAZAVIAN A S, AZIZPOUR H, SULLIVAN J, et al. CNN features off-the-shelf: an astounding baseline for recognition [C] // 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops. IEEE, 2014; 6849 - 6853.
- [17] SZEGEDY C, IOFFE S, VANHOUCHE V, et al. Inception-v4, Inception-ResNet and the impact of residual connections on learning [C] // AAAI Conference on Artificial Intelligence, 2017; 4278 - 4284.
- [18] 夏修臣, 王秀英. 基于余弦相似度的改进 C4.5 决策树算法 [J]. 计算机工程与设计, 2018, 37 (1): 120 - 125.
- [19] 李朝晖, 陈玉莹. 耳蜗模型用于舰船辐射噪声线谱检测研究 [C] //2008 年全国声学学术会议, 上海, 2008; 384 - 385.
- [20] 焦利彬, 王猛. 基于多模态深度学习的流量分类识别方法 [J]. 无线电通信技术, 2021, 47 (2): 215 - 219.
- [18] 马海钊, 陈雪梅, 李光俊, 等. 导管构件的三坐标测量机自动化检测技术的研究与应用 [J]. 锻压技术, 2021, 46 (4): 172 - 179.
- [19] 董大钊, 徐冠华, 高继良, 等. 基于机器视觉的机器人装配位姿在线校正算法 [J]. 浙江大学学报 (工学版), 2021, 55 (1): 145 - 152.
- [20] 王继红, 吴伯彪, 张亚超, 等. 视觉和惯导信息融合小型无人机位姿估计研究 [J]. 中国测试, 2021, 47 (11): 134 - 140.
- [21] 周旭, 鲁墨武, 姜春英, 等. 改进的 PSO-BP 算法在工业机器人末端位姿误差补偿中的应用 [J]. 信息与控制, 2021, 50 (4): 505 - 512.
- [22] 江朝军, 倪江生, 张军, 等. 基于 IMU 及 UWB 的微型弹跳机器人传感器节点位姿检测研究 [J]. 传感技术学报, 2021, 34 (8): 1123 - 1130.
- [23] 王靖, 魏亮, 向文豪, 等. 考虑圆形特征边缘模糊和偏心误差修正的高精度相机标定方法 [J]. 红外与激光工程, 2021, 50 (12): 369 - 379.
- [24] 曹深深, 毕德学, 邹绍昕. H-Bot 型 3D 打印机动力学参数辨识与定位误差补偿 [J]. 天津科技大学学报, 2021, 36 (2): 40 - 45.