



《人工智能与海洋工程》课程报告

基于 ResNet18 网络和 ShipsEar 数据集的水下目标识别

姓 名： 路一阳

学 号： 23111914

日 期： 2024.07



中山大學海洋工程与技术学院
SUN YAT-SEN UNIVERSITY SCHOOL OF OCEAN ENGINEERING AND TECHNOLOGY

基于 ResNet18 网络和 ShipsEar 数据集的水声目标识别

23111914 路一阳

1 研究背景

海洋覆盖地球超过七成的表面积，以其深邃和多样性构成了一个充满神秘和探索价值的广阔空间。随着我国海洋强国建设步伐的加速，对海洋信息的有效获取变得尤为关键。海洋不仅是生命的摇篮，也是全球经济和战略资源的重要基地。水声信号，包括海洋环境噪声、海洋生物的声音以及人类活动产生的噪声，作为水下信息传递的重要载体，其研究对于海洋信息的获取具有不可替代的重要作用。

水声目标识别技术（Underwater Acoustic Target Recognition, UATR），通过处理和分析这些水声信号，能够判断信号的性质和类别，是实现水下信息获取的关键技术之一。这项技术可分为主动识别和被动识别两种方式。主动识别利用主动声纳发射信号并分析回波的特性，具有较高的信噪比，但同时也容易暴露自身位置。相比之下，被动识别通过接收目标自身的噪声进行分析，隐蔽性更好，但受环境影响较大，识别难度增加。

传统的水声目标识别依赖于声纳员的人工识别，这种方法成本高昂，且容易受到个人经验和判断的影响。随着机器学习技术的发展，尤其是深度学习在图像和语音识别等领域取得的巨大成功，其在水声目标识别中的潜力也逐渐显现。深度学习技术能够自动提取特征，减少对人工干预的依赖，提高识别的准确性和效率。然而，水声目标识别面临着复杂海洋环境、多样目标特性、高保密性目标以及数据获取上的缺陷等多重挑战。海洋环境的多变性、目标的多样性以及数据的不完整性都对识别算法的鲁棒性和准确性提出了更高的要求。声纳设备能否准确识别目标并做出决策，不仅在军事侦察和潜艇战等领域具有重要的战略意义，也在海洋生物识别、环境监测等民用领域展现出广阔的应用前景。

综上所述，深入研究水声目标识别方法，不仅对于提升我国海洋科技水平、保障海洋安全具有重要的战略意义，也对于推动海洋经济的可持续发展、加强海洋资源的保护和利用具有深远的实际价值。随着技术的不断进步和创新，水声目标识别技术有望在未来的海洋探索 and 开发中发挥更加关键的作用。

2 问题概述

水下声学目标识别 (UATR) 是一项在水下信号处理领域极具挑战性的任务, 其复杂性主要源于几个关键问题。首先, 标注数据的缺乏限制了传统机器学习算法在这一领域的应用和发展。由于水下环境的特殊性, 获取大量且准确的标注数据非常困难, 这直接影响了识别系统的训练和优化。其次, 水下声波信号的时间空间变化特性, 如温度、盐度、压力等海洋环境因素的变化, 会对信号的传播特性产生显著影响, 导致信号的失真和衰减。此外, 水下环境中存在的各种噪声源, 如海洋生物、船只、海流等, 也会对目标信号造成干扰, 进一步增加了识别的难度。

针对这些挑战, 所选文章提出了一种创新的解决方案, 即基于融合特征和残差网络 (ResNet18) 的新型分类方法。这种方法的核心在于两个关键步骤: 目标特征的提取与融合, 以及网络结构的选取与设计。

在目标特征的提取与融合方面, 文章采用了多种声学特征, 包括 Mel 频率倒谱系数 (MFCC)、Log-Mel 频谱图 (LM) 以及 Chroma、Contrast、Tonnetz 和 Zero-cross ratio 的组合 (CCTZ)。这些特征能够从不同角度捕捉水下声波信号的特性, 通过融合这些特征, 可以提供更全面的信号表示, 从而提高识别系统的鲁棒性和准确性。

在网络结构的选取与设计方面, 文章采用了 16 层的残差网络 (ResNet18), 这是一种深度学习框架, 能够通过残差学习解决深层网络训练中的梯度消失和梯度爆炸问题。ResNet18 的结构设计包括多个卷积层、批量归一化层、ReLU 激活函数和池化层, 能够有效地学习和提取深层次的特征表示。此外, 为了进一步提高识别精度和避免过拟合, 文章还采用了一些策略, 如早停法 (early stopping) 和自适应学习率调整。

综上所述, 这篇文章针对水下声学目标识别面临的挑战, 提出了一种结合了特征融合和深度学习的新方法。通过精心设计的特征提取和网络结构, 这种方法有望显著提高水下声学目标的识别精度, 为水下信号处理领域提供了一种新的研究方向和解决方案。

3 研究方法

本文的研究旨在提高水下声学目标识别（UATR）的准确性和效率上，水下声学目标识别研究过程中面临的主要挑战包括缺乏标注数据、环境因素对声波信号的影响，以及来自其他噪声源的干扰。为了克服这些挑战，文章提出的研究思路如下：

（1）**融合特征的提取**：研究者们首先关注于从水下声波信号中提取有效的特征。这些特征不仅需要能够反映目标的本质属性，还要能够抵抗环境变化和噪声干扰。文章中采用了多种特征提取技术，包括 Mel 频率倒谱系数（MFCC）、Log-Mel 频谱图（LM），以及 Chroma、Contrast、Tonnetz 和 Zero-cross ratio 的组合（CCTZ）。这些特征从不同角度捕捉信号的不同特性，并通过融合这些特征来增强模型对信号的综合理解能力。

（2）**数据增强策略**：为了解决标注数据不足的问题，文章采用了 SpecAugment 这种数据增强方法。SpecAugment 通过在频谱图上进行遮罩操作，模拟了信号在实际环境中可能遇到的部分丢失或损坏的情况，从而增强模型的泛化能力。

（3）**残差网络(ResNet18)的应用**：文章选用了 16 层的残差网络(ResNet18)作为分类器。ResNet18 通过引入残差学习框架，解决了深层网络训练中的梯度消失和梯度爆炸问题，使得网络能够学习到更深层次的特征表示。

（4）**网络结构的设计**：文章对 ResNet18 的结构进行了详细的设计，包括输入层的零填充、卷积层、批量归一化层、ReLU 激活函数、池化层以及残差块的设计。这些设计使得网络能够更有效地捕捉和学习信号的特征；为了避免过拟合并提高模型的分类性能，文章在 ResNet18 的基础上增加了嵌入层，并设计了中心损失函数。这个损失函数旨在增加不同类别之间的距离，同时减少同一类别内部的距离，从而提高分类的准确性。

（5）**训练策略选择**：在训练过程中，文章采用了早停法和自适应学习率调整策略。早停法通过监控验证集损失来提前终止训练，避免过拟合。自适应学习率调整则根据训练进度动态调整学习率，以加速收敛并提高训练效率。

（6）**实验验证与分析**：文章通过在 ShipsEar 数据集上进行实验，验证了所

提方法的有效性。实验结果表明,该方法在水下声学目标识别任务上取得了 0.943 的高准确率, 优于其他现有的分类方法。

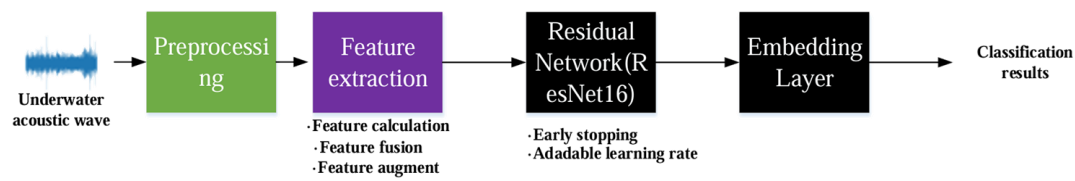


图 1 文章研究思路示意图

3.1 数据预处理

数据预处理是任何机器学习任务的首要步骤, 对于提高模型性能至关重要。在水下声学目标识别中, 由于信号可能具有不同的长度和采样率, 预处理步骤主要包括: 信号分割、重采样、标准化。

信号分割是指使用滑动窗口将水下声学信号分割成固定长度的帧。这种方法不仅适应了水下信号在极短时间内的稳定性, 而且通过重用信号的部分内容, 自然增加了样本数量, 这可以视为一种数据增强手段。

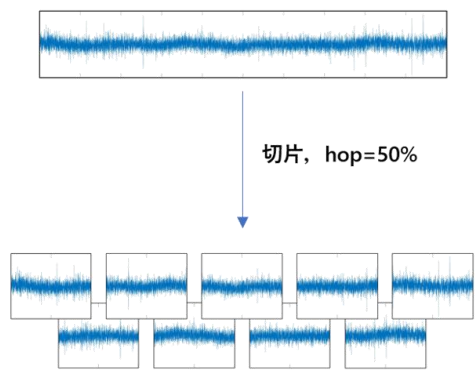


图 2 预处理操作示意图

重采样是指将所有信号帧重采样到统一的采样率, 例如 20,480 Hz, 以确保输入样本的质量和模型计算成本之间的平衡。

标准化是指对信号进行标准化处理, 以消除不同信号之间的量纲和量级差异, 为后续的特征提取和网络训练打下良好的基础。

3.2 特征提取及融合

特征提取是将原始信号转换为能够代表信号特性的数值向量的过程。在本文中，特征提取及融合包括以下几个步骤：特征计算、特征融合、特征增强。通过提取多种声学特征，包括 Log-Mel 频谱图 (LM)、梅尔频率倒谱系数 (MFCC) 以及 Chroma、Contrast、Tonnetz 和 Zero-cross ratio 的组合 (CCTZ)。这些特征从不同角度描述了信号的特性；特征融合是指将不同类型特征融合为一个多维特征向量，以利用它们的互补信息，增强模型对信号的识别能力；使用 SpecAugment 方法对特征进行增强，通过在特征矩阵上应用遮罩操作，模拟信号的不完整性，从而提高模型的泛化性和鲁棒性。

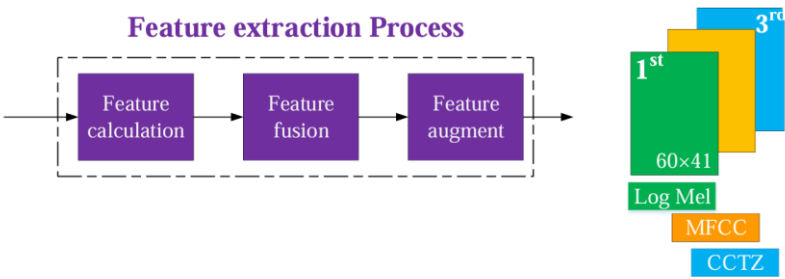


图 3 特征提取流程

3.3 网络构建及训练

网络构建及训练是实现水下声学目标识别的核心环节。本文采用了 ResNet18 网络，并对其进行了适当的修改和训练策略的优化。

文章构建了一个 16 层的残差网络 (ResNet18)，包括卷积层、批量归一化层、ReLU 激活函数、池化层以及残差块。残差神经网络通过构建残差块在神经网络内部进行网络的跳跃连接，解决了传统卷积神经网络因网络深度过大而引起的梯度消失问题。网络的输入是经过预处理和特征提取的多维特征向量，输入数据维度为 $60 \times 41 \times 3$ ；在网络中加入了嵌入层，并使用了中心损失函数来优化嵌入层的参数，使得同一类别的样本在特征空间中更紧凑，不同类别的样本更分散；采用了早停法和自适应学习率调整策略。早停法通过监控验证集的损失来防止过拟合，而自适应学习率调整则根据训练的进展动态调整学习率，以加速收敛；在

ShipsEar 数据集上进行实验验证，通过设置不同的训练集、验证集和测试集比例，评估模型的性能。使用精确度、召回率和 F1 分数等指标来衡量分类系统的性能。

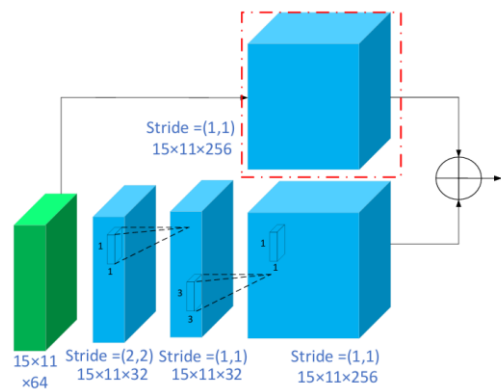


图 4 网络结构示意图

4 改进措施

在不改变文章原本数据集及与预处理方式的情况下，增加以下改进措施。

4.1 数据增强策略

由于研究目标是对水声目标进行分类识别，本文所使用的图像是通过时频分析的方法将原始声目标信号进行维度变换，在尽可能保持原始声信号所包含的特征信息的前提下，将信号由一维序列转换至二维图像。在对所获得的梅尔谱图进行分类识别时，与传统的图像识别任务有明显不同：传统图像识别的数据集一般为实际物体的照片、绘画等可能具有明确边缘、色彩变化、形状等实际特征，而水声信号的梅尔谱图从左至右包含着时间的先后顺序（靠左边的列对应原始信号中较早发生的一段），同时每一列代表着一个窗口长度的时间信号的频谱分布，时序和谱分布都是进行水声目标识别所需要的关键信息，翻转（Flip）、移位（Translate）、旋转（Rotate）等会改变时序，遮蔽（Cut）、改变亮度等会影响频谱信息，缩放（Zoom）会改变梅尔谱图中的信息量。为了解决传统的数据增强方法在声音谱图数据集中适应性差的问题，本文提出了一种针对水声信号的数据增强方法。该方法结合声信号在水声信道中的传播特性，模拟声音信号多径效应，通过构建并调整特征声线参数，获取不同信道条件下的新信号，以实现数据增强目的，此方法基于领域风险最小化原则（Vicinal Risk Minimization，VRM）。

3.3.3.1 多径传播

多径传播是指从声源发射的声信号,由于海水分层介质的折射和海面、海底、障碍物等的反射,沿两个或多个路径到达接收点的传播现象。水声环境复杂多变,声信号在传播过程中到达海面、海底交界面时,由于介质声阻抗不同发生反射;海洋水体温度、盐度、压力的不同形成的声速梯度也会使声线的传播方向发生改变。

本文所使用数据集中所记录数据的采集位置位于浅海声信道,其中声速可以认为均匀分布,如图简单展示了浅海声信道中的多径传播现象。到达接收端水听器的有直达波和经海面、海底反射的反射波。

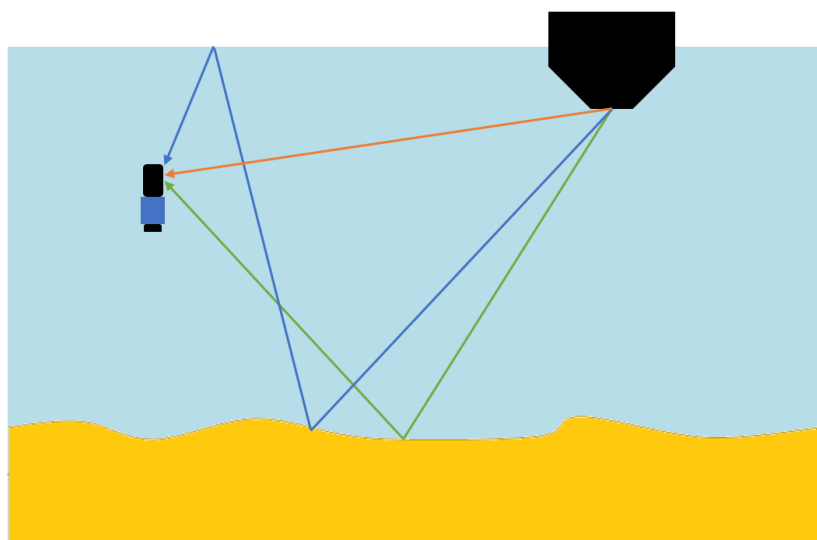


图 5 水声信道多径传播示意图

浅海声信道中海水深度较小,海面与海底间距更小,声音的传播容易受到海水介质上下边界的影响,其多径效应相较于深海信道更加明显,且本文所使用数据集中所记录数据的采集位置位于浅海声信道,所以本文主要考虑声速均匀情况下的多径传播效应。由于不同传播路径的总路程不同,水声信号经由不同路程传播至接受端水听器所用的时长不同,同时传播损耗也不同。一般情况下,传播路程越长,传播所用时长越长,传播损耗越大。

3.3.3.2 模拟声信号多径传播的具体方法

海水中声传播损失是为传播距离 l 和信号频率 f 的函数,可以表示为:

$A(l, f) = l^k [\alpha(f)]^l$ ，若发射功率为 P ，那么接收点功率为 $P / A(l, f)$ 。数据库中记录的数据获取时使用阵列水听器采集，最终记录的数据是经过处理修正后的，可以近似认为是经直达路径到达接收端水听器的、已经计算过传播损耗的直达波。设直达路程为 l_0 ，直达没有时延，则直达波功率为 $P / A(l_0, f)$ 。由于用来增强的源数据其实是直达波，观察数据频谱可以发现，经过海水声吸收后的声信号在频域中，从低频到高频幅值呈现对数降低，低频分量明显，而高频分量极低。因此假设通过其他传播路径所得的反射波信号频谱分布相似，即经过不同传播距离的声信号各频率成分衰减不再因距离的改变而改变，反射波信号与直达波信号幅值上存在比例关系、时间上存在先后关系，而频谱分布相同。这样，便可以由直达波扩展出其他多路信号，各路信号到达接收点功率可以简化为 $P / A(l_n)$ 。水听器接收到的信号是直达波和反射波信号的叠加，由上述假设简化为具有不同时延和衰减系数的多路信号的叠加，可以表示为

$$x(t) = \sum_n \lambda_n s(t - \tau_n)$$

其中， λ_n 为经第 n 条路径传播信号的衰减系数； τ_n 为经第 n 条路径传播信号的时延； $s(t)$ 为接收端所接受到的直达波信号序列，经过水声信道声吸收后其频谱分布已经改变。直达波信号序列还可以表示为

$$x_0(t) = \lambda_0 s(t - \tau_0) = \lambda_0 s(t)$$

其中， $\lambda_0 = 1$ 。由幅值和能量的关系可知， λ_0 相关于 $\sqrt{P / A(l_0, f)}$ ，带入上述假设，即认为 λ_0 相关于 $\sqrt{P / A(l_0)}$ 。对于第 n 条传播路径

$$\lambda_n = \lambda_0 \frac{\sqrt{P / A(l_n)}}{\sqrt{P / A(l_0)}}$$

$$\tau_n = \frac{(l_n - l_0)}{c}$$

其中， c 为海水中声速，本文假设其在浅海信道中均匀分布，即声速恒定，时延 τ_n 是关于第 n 条传播距离 l_n 的线性函数。本数据库采集环境满足中距离传播，根据经验可以认为传播损耗 $TL = 15 \lg l_n$ ，又因为各路信号到达接收点功率为

$P/A(l_n)$ ，可得

$$TL=15\lg l_n = 10\lg A(l_n)$$

$$A(l_n) = l_n^{1.5}$$

最终可得

$$\lambda_n = \lambda_0 \frac{\sqrt{P/A(l_n)}}{\sqrt{P/A(l_0)}} = \left(\frac{l_0}{l_n}\right)^{0.75} = \left(\frac{l_0}{l_0 + c\tau_n}\right)^{0.75}$$

这样建立了时延 τ_n 和衰减系数 λ_n 的关系，通过设置不同的时延 τ_n ，再利用上式求得对应的衰减系数 λ_n ，进而可以将反射波信号表示为：

$$x_n(t) = \lambda_n s(t - \tau_n) = \left(\frac{l_0}{l_0 + c\tau_n}\right)^{0.75} s(t - \tau_n)$$

经过多径传播后，水听器接收到的信号是直达波和反射波信号的叠加，将原始数据视为直达波后，反射波信号可以表示为直达波的函数，再求和获得最终的叠加信号。将叠加所得的信号进行前文介绍过的预处理操作，获得最终的梅尔谱图数据集。基于多径传播的声信号数据增强方法具体实现见算法 1。

算法 1 基于多径传播的声信号数据增强算法

基于多径传播的声信号数据增强算法

Input: 未经切分的声信号时间序列 $s(t)$ 及其对应目标距离 l_0

STEP:

1. 利用随机数函数初始化时延 τ_n 并利用式 66 计算衰减系数 λ_n ;
2. 计算反射波信号时间序列 $x_n(t)$;
3. 将直达波与反射波信号时间序列叠加, 获得总信号 $x(t)$;
4. 对总信号 $x(t)$ 进行信号预处理流程, 获得由经过多径传播模拟后的信号所得的梅尔谱图数据集 $D_{\text{multipath}} = e_1, e_2, \dots, e_m$;
5. 将由原始信号所得的梅尔谱图数据集 $D = d_1, d_2, \dots, d_n$ 与 $D_{\text{multipath}} = e_1, e_2, \dots, e_m$ 合并获得最终的数据集 $D_{\text{augmented}} = d_1, d_2, \dots, d_n, e_1, e_2, \dots, e_m$

END

Output: 梅尔谱图数据集 $D_{\text{augmented}} = d_1, d_2, \dots, d_n, e_1, e_2, \dots, e_m$

4.2 网络结构更新

深度学习网络结构的设计是深度学习方法研究当中最重要的一环, 其核心内容在于对网络结构进行合理设计并优化。标准的深度学习网络结构一般由输入层、隐藏层和输出层三部分组成。输入层主要负责将经过预处理数据输入进网络。输出层一般为全连接层, 将隐藏层的输出数据经过非线性变换映射到数据标签维度上。隐藏层是深度学习网络结构的核心部分, 其关键作用在于对数据集进行特征提取, 建立输入数据和输出数据之间的非线性拟合关系, 通过该非线性拟合关系近似输入和输出之间的数学模型。隐藏层一般由卷积层、池化层、线性整流层和全连接层构成。

4.2.1 卷积神经网络

卷积神经网络作为一种成熟的网络模型被广泛应用于图像处理和语音信号处理等领域。相较于 BP 神经网络，卷积神经网络具有数据卷积和权值共享的特性，可以对数据的深层特征进行充分的提取，从而提高网络回归的精度。

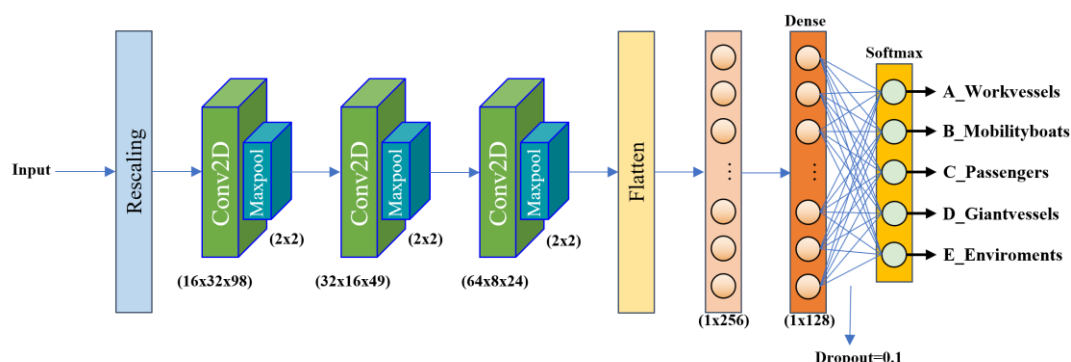


图 6 卷积神经网络结构示意图

本文设计的卷积神经网络的基本结构如上图所示，网络的基本结构由三层卷积层和三层全连接层构成，在每个卷积层后跟随一个最大池化层（Maxpool）来进行数据降维；在卷积模块和全连接模块之间是一个展平层，它将卷积模块输出的多层数据展平至一维，然后与全连接模块进行全连接；所有全连接层和卷积层均选用 ReLU 函数作为激活函数。

同时，全连接模块之后放置了一个随机丢弃层 Dropout，丢弃概率设置为 0.1。随机丢弃层以一定的概率随机对上一层输出的数据进行丢弃，即随机赋零值。同时对留下来的数据进行微调以保证丢弃前后的输出期望保持不变。Dropout 层的存在意义是，在一定程度上抛弃数据的局部特征，让网络着重学习数据的共性特征，从而能够有效的减弱训练过程中出现过拟合的可能性。

卷积神经网络的网络各层的参数设置如下表所示，三个卷积层的卷积核大小均为 3×3 ，卷积核通道数依次设置为 16、32、64，卷积步长均设置为 1，卷积模式均设置为在输入数据矩阵外围增加零矩阵，这样可以确保在前向传播的过程中数据矩阵大小相同，在不改变数据矩阵特征时能够对矩阵的边缘数据特征进行提取。池化层选择最大池化，池化核大小为 2×2 ，步长为 1。数据通过最大池化层后，卷积核中只保留最大值作为输出。三个全连接层的神经元个数分别设置为

256、128。经过全连接层以及随机丢弃之后，网络的输出与数据标签的维度一致，将输出数据映射到了标签空间中。

表 1 卷积神经网络模型参数表

网络类型	卷积核尺寸	输出数据维度	参数数量
输入层		(None, 32, 98, 1)	0
Conv1	$16 \times 3 \times 3$	(None, 32, 98, 32)	16
Max-pooling	2×2	(None, 16, 49, 32)	0
Conv2	$32 \times 3 \times 3$	(None, 16, 49, 32)	4660
Max-pooling	2×2	(None, 8, 24, 32)	0
Conv3	$64 \times 3 \times 3$	(None, 8, 24, 64)	18496
Max-pooling	2×2	(None, 4, 12, 64)	0
Flatten		(None, 3072)	0
Dense1		(None, 256)	786688
Dense2		(None, 128)	32896
Dense3		(None, 5)	645

卷积层的参数主要包括所有卷积核矩阵的权重参数和偏置参数。以第三层卷积层为例，共有 64 个卷积核，对应 64 个偏置参数。每个卷积核的通道数与输入维度相同，均为 32 通道，相应的卷积核的权重参数为 $3 \times 3 \times 32 \times 64$ 。因此，Conv3 卷积层共有 $(3 \times 3 \times 32 + 1) \times 64$ 个参数。全连接层的参数包括相邻两层神经元之间的权重参数和该层神经元所对应的偏置参数。以第三层全连接层为例，该层共有 5 个神经元，对应于水声数据集的 5 个类别。上一层的 128 个神经元与本层神

神经元之间两两建立连接，共有 128×5 个权重参数。因此，FC4 全连接层共有 $(128+1) \times 5$ 个参数。所以，本节所设计的卷积神经网络共有 843,525 个参数，其中全连接层有 820,353 个参数，占参数总量的 97.25%。

4.2.2 残差神经网络

残差神经网络通过构建残差块在神经网络内部进行网络的跳跃连接，解决了传统卷积神经网络因网络深度过大而引起的梯度消失问题，在图像识别领域取得了显著的成绩。

本文所采用的残差神经网络为 Resnet50 残差神经网络。该网络的基本模型如下图所示。网络由五个卷积模块和一个全连接层组成，第一个卷积模块包含一个卷积层和一个最大池化层。其余四个卷积模块分别包含 3、4、6 和 3 个残差模块。每个卷积模块如下图所示包含三个卷积层。四个卷积模块中的第一个残差模块的第二个卷积层在数据宽度方向的卷积步长分别为 1、2、2 和 2。在进行卷积计算时仅选用权重参数，不引入偏置参与计算。每个卷积模块的第一个残差块用 1×1 卷积创建捷径连接，其余残差块均由恒等映射创建捷径连接。Resnet50 网络模型有 $(3+4+6+3) \times 3 + 1$ 个卷积层和 1 个输出层共计 50 层网络结构，这也是其命名原因之一。在卷积神经网络和全连接层之间设置平均池化层用以减小输出数据数量。全连接层共有两层，神经元个数分别为 10245 和 5。选择 ReLU 函数作为网络的非线性激活函数。

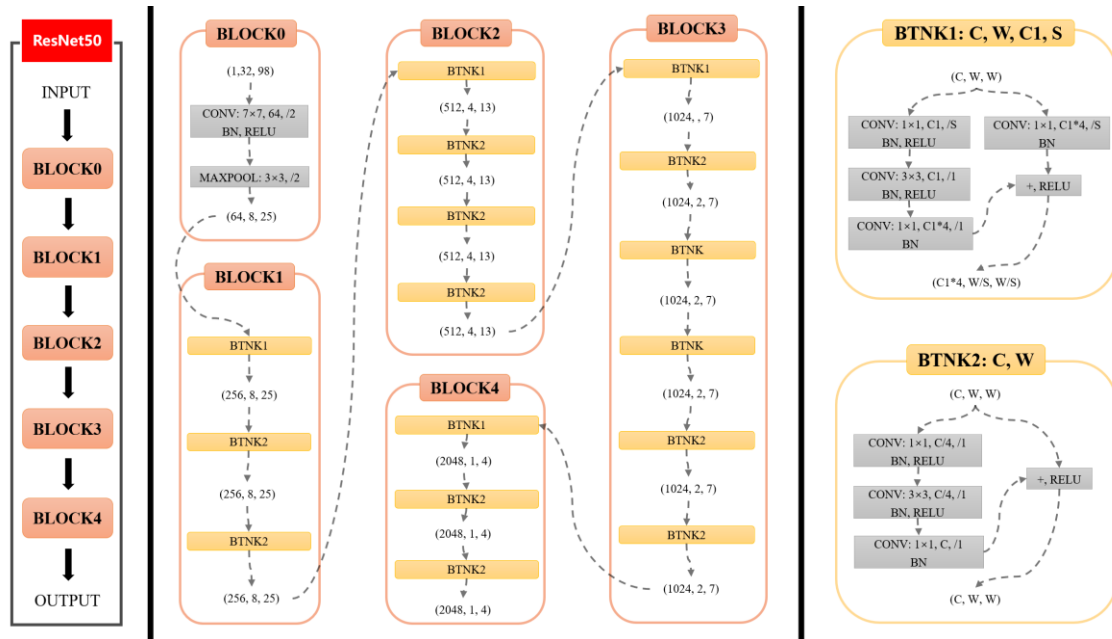


图 7 ResNet50 残差神经网络结构示意图

本文所使用的 Resnet50 残差神经网络由于其结构的复杂性，网络各层的参数设置不进行展示，该网络共使用了 23, 591, 685 个参数，其中可更新参数的有 23, 538, 565 个，占比为 99.77%。

4.2.3 密集卷积网络

密集卷积网络通过建立后面层与所有前面层的“快捷连接”，实现了使用更少的参数构建更加深的网络结构，降低计算成本的同时减缓了梯度消失的问题。

本文所采用的密集卷积网络为 DenseNet121 密集卷积网络。该网络的基本模型如下图所示。网络由 121 层权重层组成，其中有 4 个密集卷积模块，共计 116 层，加上初始输入的 1 个卷积层、3 个过渡层以及最后输出的全连接层，共计 121 层；初始卷积层有 64 个通道数，经过 7×7 卷积将尺寸为 32×98 的输入图片降维至 16×49；密集卷积模块由完全相同的 layer 堆叠而成，layer 的尺寸为：1×1 和 3×3 的两层卷积块（每层卷积块包含批归一化、激活 ReLU、卷积操作）；增长率即特征图通道数设置为 $K = 32$ ；为了提高网络的紧密度，在转换层（transition layers）减少特征图的数量，引入压缩因子 $k = 0.5$ ；通过 1×1 卷积层来减小通道数，2×2 平均池化来减小特征图的尺度。最后经过全局平均池化层以及 1024 维的全连接层采用 Softmax 函数得到最终的分类结果。

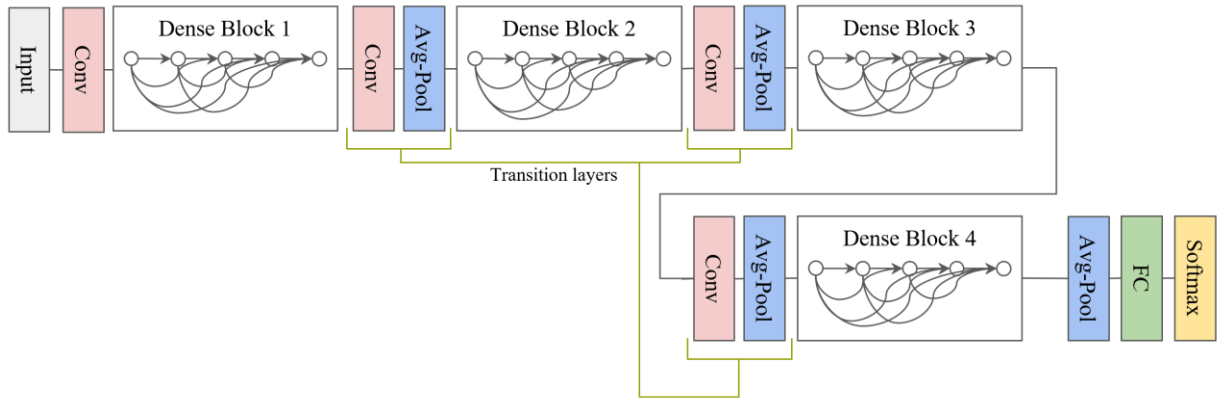


图 8 DenseNet121 密集卷积网络结构示意图

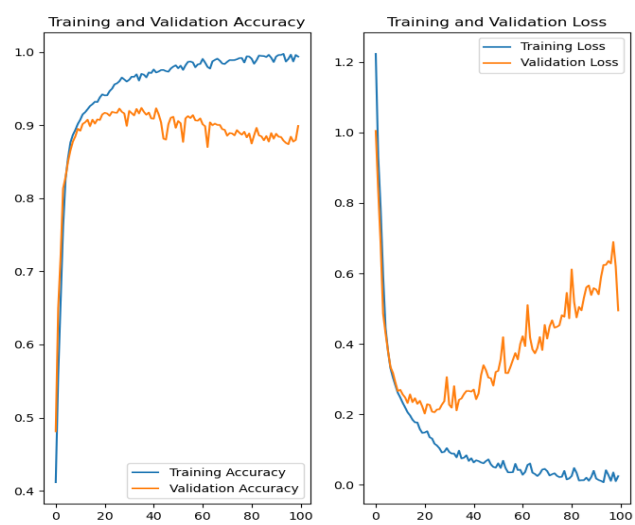
本文所使用 DenseNet121 密集卷积网络由于其结构的复杂性，网络各层的参数设置不进行展示，该网络共使用了 7,036,357 个参数，其中可更新参数的有 6,952,709 个，占比为 98.81%。

5 试验结果

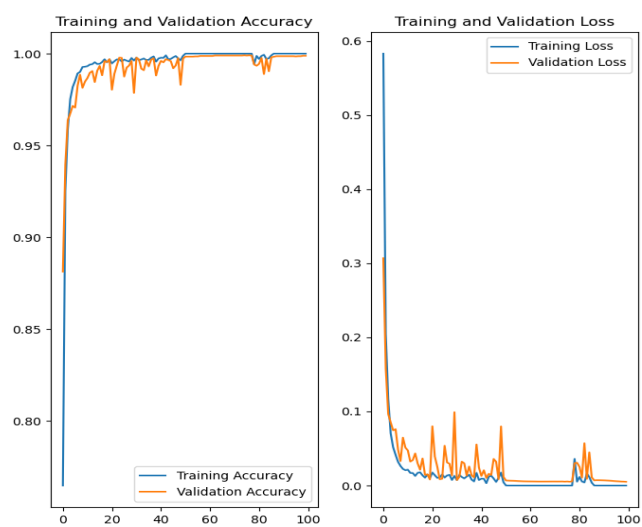
实验结果如下图及下表所示。

表 2 实验结果

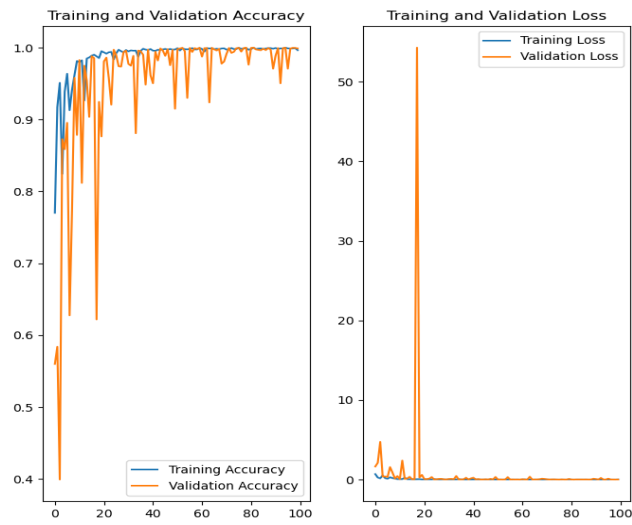
类别	ResNet18	卷积神经网络	ResNet50	DenseNet121
网络层数	8	8	50	121
需更新参数量	8,473,883	2,416,389	23,538,565	6,952,709
验证集准确率 (%)	90.24	98.92	99.21	99.32
训练用时 (s)	1,477.84	407.12	4105.12	3928.01



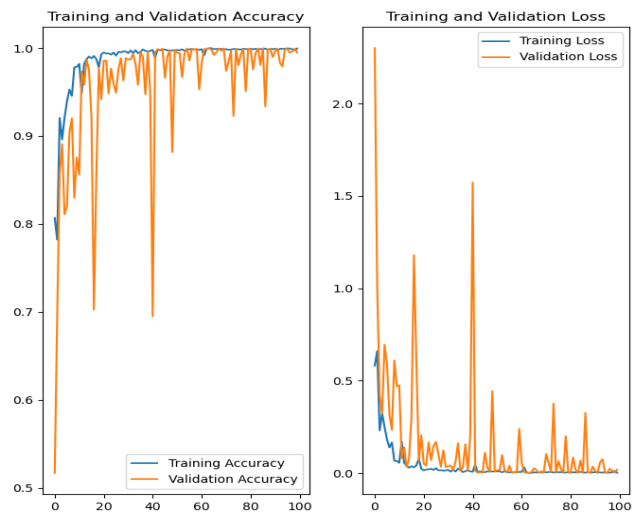
(本文)



(CNN)



(ResNet50)



(DenseNet121)

图 9 迭代过程及结果

所提出模拟多径传播的方法实现了在验证集上相较于原文准确率的有效提升。从网络结构层面上看，改进的卷积神经网络具有最简单的结构，原文网络随之，DenseNet121 网络具有最复杂的结构。ResNet50 网络具有最多的参数量，DenseNet121 网络在结构最为复杂的情况下，参数量相较于 8 层的卷积神经网络并未增加多少，这是符合预期的，密集卷积网络实现了所有前后层之间的直接连接，使得特征在每一层都能被重复利用，因此大大减少了参数量；从分类准确率

上看，DenseNet121 网络和 ResNet50 网络，均优于卷积神经网络，DenseNet121 网络效果最好，但相较于卷积神经网络取得的效果，另两者提升不大，主要是由于分类的准确率已经接近于上限，且网络越复杂需要的训练集数量越大，而本实验中采用同样的数据集进行网络的训练；从训练用时上看，卷积神经网络具有最短的用时，而 ResNet50 网络用时最长，DenseNet121 网络用时略短于 ResNet50 网络，但差距不大，都几乎接近卷积神经网络训练用时的 10 倍。

6 总结展望

主要从特征选取、数据增强方法、网络结构设计三个方面进行了对比试验，在验证了所选文章的主要工作后，提出了相应的改进措施。所提出的数据增强方式从物理特性出发，通过进行物理假设来扩充数据集，成功的增加了样本量，提高了模型准确率；同时本文设计的卷积神经网络在使用增强后的梅尔谱图数据集进行训练后达到了较高的准确率，虽然稍稍逊于 ResNet50 残差网络和 DenseNet121 密集卷积网络，但结构简单，训练时长较短，具有快速、轻量的优势。同时三种网络均相较于原文有效提高了识别准确率。

本文在水声目标识别方法的研究上，取得了初步的成果，但还存在许多后续工作。为了完善和充实对于水声目标识别方法的研究，还可以从以下几个方面入手：

（1）低信噪比条件下的水声目标识别方法研究。本文使用专业人员采集处理的船舶噪声库作为数据来源，具有较高信噪比和质量，因此实验结果较好，而实际应用环境中包含着大量的噪声，因此对于低信噪比条件下的水声目标识别，具有很高的实际价值。

（2）多目标检测技术的研究。水下及水面在同一时间、同一水域可能同时存在着多种噪声源，并且水声信号的中长距离传播可能导致即使多个目标具有较大的时空距离，而在某处水听器同时接受到多种目标辐射噪声的混合信号，因此进行水声多目标检测技术的研究具有广阔的前景。

（3）采用生成式网络进行数据增强。生成对抗网络（Generative Adversarial Network, GAN）以及其众多分支是当下最火的深度学习模型之一，本文曾尝试使用了深度卷积生成对抗网络（Deep Convolutional Generative Adversarial Network,

DCGAN) 来进行水声数据梅尔谱图的生成, 但遗憾的是由于参数选择等多种问题, 所生成的图片效果不佳, 在未来的研究工作中可以继续在此方向继续努力。

(4) 将深度学习领域更加先进的网络与水声目标识别技术的充分融合发展。目前深度学习在图像识别、语言识别等领域发展颇为成熟, 因此本文主要借鉴了图像识别领域相关知识, 同时使用了较为先进的网络结构进行实验, 验证了其极高的性能。如何更好的将更多优秀的模型与水声目标识别相结合, 探索二者之间的关系也是一个不错的方向。

参考文献

- [1] Hong F, Liu C, Guo L, et al. Underwater acoustic target recognition with resnet18 on shipsear dataset[C]//2021 IEEE 4th International Conference on Electronics Technology (ICET). IEEE, 2021: 1240-1244.
- [2] Nii H P, Feigenbaum E A, Anton J J. Signal-to-symbol transformation: HASP/SIAP case study[J]. AI magazine, 1982, 3(2): 23-23.
- [3] Ren H, Chang C I. Automatic spectral target recognition in hyperspectral imagery[J]. IEEE Transactions on Aerospace and Electronic Systems, 2003, 39(4): 1232-1249.
- [4] Maksym J N, Bonner A J, Dent C A, et al. Machine analysis of acoustical signals[J]. Pattern Recognition, 1983, 16(6): 615-625.
- [5] Mohamed N J. Target course recognition using nonsinusoidal look-down radars[J]. IEEE transactions on electromagnetic compatibility, 1994, 36(2): 117-127.
- [6] Rajagopal R, Sankaranarayanan B, Rao P R. Target classification in a passive sonar-an expert system approach[C]//International Conference on Acoustics, Speech, and Signal Processing. IEEE, 1990: 2911-2914.
- [7] 陶笃纯. 噪声和振动谱中线谱的提取和连续谱平滑[J]. 声学学报, 1984, 9(6): 337-344.
- [8] 杨德森, 吴一. 水下目标辐射噪声线谱的分析[J]. 哈尔滨工程大学学报, 1996, 17(1): 38-44.

- [9] 吴国清, 魏学环, 周钢. 提取螺旋桨识别特征的二种途径[J]. 声学学报, 1993, 18(3): 210-216.
- [10] 吴国清, 李靖, 陈耀明, 等. 舰船噪声识别 (I)——总体框架, 线谱分析和提取[J]. 声学学报, 1998, 23(5): 394-400.
- [11] Wang W, Li S, Yang J, et al. Feature extraction of underwater target in auditory sensation area based on MFCC[C]//2016 IEEE/OES China Ocean Acoustics (COA). IEEE, 2016: 1-6.
- [12] 张少康, 田德艳. 水下声目标的梅尔倒谱系数智能分类方法[J]. 应用声学, 2019, 38(2): 267-272.
- [13] 李朝晖, 迟惠生. 听觉外周计算模型研究进展[J]. 声学学报, 2006, 31(5): 449-465.
- [14] Lourens J G. Classification of ships using underwater radiated noise[C]//COMSIG 88@ m_Southern African Conference on Communications and Signal Processing. Proceedings. IEEE, 1988: 130-134.
- [15] Das A, Kumar A, Bahl R. Marine vessel classification based on passive sonar data: the cepstrum - based approach[J]. IET Radar, Sonar & Navigation, 2013, 7(1): 87-93.
- [16] Liu J, He Y, Liu Z, et al. Underwater target recognition based on line spectrum and support vector machine[C]//2014 International Conference on Mechatronics, Control and Electronic Engineering (MCE-14). Atlantis Press, 2014: 79-84.
- [17] Kamal S, Mohammed S K, Pillai P R S, et al. Deep learning architectures for underwater target recognition[C]//2013 Ocean Electronics (SYMPOL). IEEE, 2013: 48-54.
- [18] McDonnell M D, Gao W. Acoustic scene classification using deep residual networks with late fusion of separated high and low frequency paths[C]//ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020: 141-145.
- [19] 李思纯. 基于矢量水听器的目标特征提取与识别技术研究[D]. 哈尔滨工程大学, 2008.
- [20] 杨宏晖, 申昇, 姚晓辉, 等. 用于水声目标特征学习与识别的混合正则化深度置信网络[J]. 西北工业大学学报, 2017, 35(2): 220-225.

- [21] 王强. 基于深度学习理论的水声目标识别方法研究[D]. 西北工业大学, 2018. DOI: 10.27406/d.cnki.gxbgu.2018.000429.
- [22] Hu G, Wang K, Peng Y, et al. Deep learning methods for underwater target feature extraction and recognition[J]. Computational intelligence and neuroscience, 2018.
- [23] Shuguang Wang, Xiangyang Zeng, Robust underwater noise targets classification using auditory inspired time – frequency analysis, Applied Acoustics, Volume 78, 2014, Pages 68–76.