

◇ 李启虎院士八十华诞学术论文 ◇

水下目标多模态深度学习分类识别研究

曾 赛^{1,2†} 杜选民^{1,2}

(1 水声对抗技术重点实验室 上海 201108)

(2 上海船舶电子设备研究所 上海 201108)

摘要 水下目标的分类识别对于水声探测具有重要意义。该文提出一种水下目标多模态深度学习分类识别方法。针对水声信号的一维时域模态和二维频域模态特征建立一种多模态特征融合的深度学习结构,结合长短时记忆网络和卷积神经网络的优点,对一维时域信号和二维频谱信号分别进行并行处理,对输出进行典型相关分析,形成特征融合表示,并利用相邻帧的相关性进行参数优化。利用实测水声信号对算法进行了验证。结果表明:提出的算法对于水下目标识别的精度有显著的提高。

关键词 水下目标识别,长短时记忆网络,卷积神经网络,典型相关分析

中图法分类号: TP391.4 文献标识码: A 文章编号: 1000-310X(2019)04-0589-07

DOI: 10.11684/j.issn.1000-310X.2019.04.016

Multimodal underwater target recognition method based on deep learning

ZENG Sai^{1,2} DU Xuanmin^{1,2}

(1 *Science and Technology on Underwater Acoustic Antagonizing Laboratory, Shanghai 201108, China*)

(2 *Shanghai Marine Electronic Equipment Research Institute, Shanghai 201108, China*)

Abstract Underwater target recognition has great significance for underwater acoustic detection. The multimodal underwater target recognition method was proposed based on deep learning. Due to the time domain features and frequency domain features, a multimodal structure was proposed to incorporate the long short-term memory neural network and convolution neural network. The time domain modal and frequency domain modal were processed respectively, the output of those networks was generated as feature fusion by canonical correlation analysis method. The temporal coherence of adjacent signal frame was utilized to improve the recognition accuracy. The experiments were implemented based on measured underwater acoustic signal. The results show that the proposed method improves the accuracy of underwater target recognition significantly.

Key words Underwater target recognition, Long short-term memory neural network, Convolution neural network, Canonical correlation analysis

2019-02-02 收稿; 2019-05-07 定稿

作者简介: 曾赛 (1989-), 男, 湖北荆州人, 博士研究生, 研究方向: 水下目标特性、水下目标探测、水声信号与信息处理。

† 通讯作者 E-mail: sharememezeng@126.com

0 引言

水声技术的快速发展使得观测数据迅速增长,如何建立基于水声观测数据的水下目标识别方法一直是水声工程领域的研究热点。如今水声探测技术朝着多源、多模态的方向发展,迫切需要寻找多源数据中隐藏的目标本质特性,实现对水下目标的特征提取与分类识别。水下目标的分类识别本质上属于模式识别的范畴,机器学习方法是水下目标识别的主要方法,且目前已经从浅层学习发展为深度学习^[1]。

传统的浅层学习方法本质上是对数据进行非线性特征变换,如多层感知机模型、支持向量机(Support vector machine, SVM)模型^[2]、高斯混合模型^[3]、条件随机场模型^[4]等。浅层学习的前提是特征工程,学习模型一般将特征工程提取的特征向量通过非线性数据映射得到分类边界。因此,浅层学习适合解决特征提取容易、约束具体的分类问题,对于特征工程较难的数据分类效果往往不好,此外,浅层学习的参数无法自适应,因此其自纠错能力欠缺^[5]。

深度学习理论较好地克服了浅层学习依赖特征工程和自适应欠缺的缺点,其能够进行参数自适应学习,且可以表征高维复杂函数,提取目标的深层特征,因此可以获得更多关于目标的有用信息^[6]。深度学习最为典型的模型是长短时记忆(Long short-term memory, LSTM)网络^[7]和卷积神经网络(Convolution neural network, CNN)^[8],LSTM模型适合处理时序数据,如语音信号、自然语言序列等,而CNN模型则适合处理图像信息^[9]。

随着水声工程技术的飞速发展,一方面水声信号数据积累日益增加,另一方面又对水声信号处理方法提出了新的要求,如在水下目标的分类识别方面,采用何种信号处理手段高效地处理积累的水声信号数据是需要研究的问题,目前已有将深度学习方法应用到水下目标分类识别的尝试^[10-14]。但其应用是基于水声信号的单一模态进行处理,对于水声信号而言,一维时域信号和二维时频谱信号均能够为目标识别提供重要信息,将多模态的水声信号进行融合有利于水下目标的分类识别。事实上,多模态信息融合处理方法在视频处理领域应用非常广泛^[15-17]。

基于上述分析,本文提出一种适合水声信号处理的多模态目标识别方法,综合LSTM的语音识别能力和CNN的图像处理能力,建立LSTM和CNN的共享关联表示,充分利用水声信号不同维度的融合特征,提高目标分类识别能力。

1 水下目标多模态深度学习识别理论

水下目标识别的关键是水声信号的特征提取,典型的特征提取有时域信号过零点分布、功率谱特征、线谱特征、听觉域特征、时频谱特征等。由于水声环境的复杂性导致水声信号复杂多变,依靠单一特征模态处理往往难以得到较好的效果,将多模态特征融合利用能够有效增加目标识别能力:即将水声信号的一维时域模态与二维时频谱模态分别作为两种特征模态输入,使用不同的深度网络处理结构进行并行处理,再将这种异构处理得到的高层特征进行关联共享用于目标分类识别。

1.1 CNN深度学习模型

CNN是一种典型的深度学习模型,从20世纪60年代提出以来,CNN模型得到了快速发展并成为时下最为成功和流行的深度学习模型之一。CNN具有稀疏交互、参数共享和等变表示的优点,擅长处理二维图像数据。CNN通过卷积核完成特征的局部提取,然后通过池化降采样完成特征映射,特征逐层抽象提取与映射使得CNN能够得到比浅层学习更为抽象稳定的特征,因此特征提取与映射是CNN最为关键的处理流程。典型的CNN网络结构如图1所示。

CNN处理流程中,每个卷积层 C_l 对 N^{l-1} 层的输入位面 $Z_{1,\dots,N^{l-1}}^{l-1}$ 进行线性滤波,滤波器为卷积核 $K^l \times K^l$,该层卷积层输出位面为 $D^{l-1} \times D^{l-1}$,输出为 Z_{1,\dots,N^l}^l 位面中任选 N^l 层, N^l 在 p^{th} 位面中位置 (x, y) 处的计算值为

$$Z_p^l(x, y) = b_p^l + \sum_q \sum_{s=l}^{K^l} \sum_{t=l}^{K^l} w_{p,q,s,t}^l \times Z_q^{l-1}(i-1+s, j-1+t), \quad (1)$$

式(1)中,偏置 b_p^l 和卷积核权重 $w_{p,q,s,t}^l$ 通过反向传播(Back propagation, BP)算法进行优化。本文利用CNN处理二维时频模态水声数据。

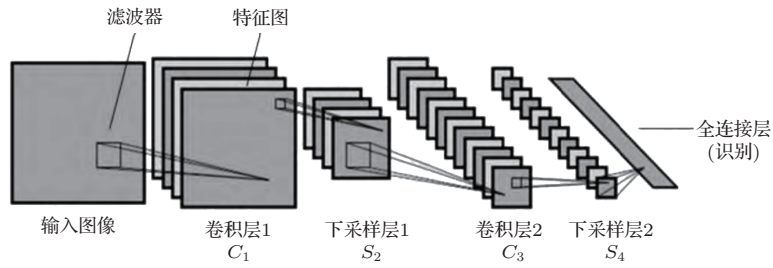


图1 典型的CNN网络结构示意图

Fig. 1 The structure of CNN

1.2 LSTM深度学习模型

LSTM本质上属于循环神经网络(Recurrent Neural Network, RNN),是深度学习领域较为成功的算法。传统的循环神经网络无法在较长时间跨度上对特征进行关联存储,在参数优化过程中面临着梯度消失的局限性,LSTM网络能够有效克服传统RNN网络的缺点,其能够充分利用一维时域数据之间的相关性,形成类似于马尔科夫推理链结构。LSTM通过引入“门”结构实现对信息的有效存储和利用。典型的LSTM单元如图2所示。

在LSTM结构中,“遗忘门”(Forget gate)决定网络从细胞(cell)状态中丢弃何种信息,遗忘门会读取 $a^{(t-1)}$ 和 $x^{(t)}$,然后输出一个介于0到1之间的数值给细胞状态 $c^{(t-1)}$,其中1表示信息完全保留,而0表示信息完全舍去。遗忘门的输出表示为

$$f_{\text{gate}} = \sigma(W_f \cdot [a^{(t-1)}, x^{(t)}] + b_f), \quad (2)$$

式(2)中, W_f 为权值, $a^{(t-1)}$ 为上一个细胞输入, $x^{(t)}$ 为当前细胞输入, σ 表示sigmoid函数, b_f 为偏置。

“更新门”(Update gate)用于决定让多少新信

息加入到当前的细胞状态中,更新的动作分两步完成:首先使用sigmoid层决定哪些信息需要更新,然后利用tanh层提取备用更新内容,最后将两个步骤进行融合。更新门的具体过程表示为

$$i_{\text{gate}} = \sigma(W_i \cdot [a^{(t-1)}, x^{(t)}] + b_i), \quad (3)$$

$$C_{\text{tanh}} = \tanh(W_C \cdot [a^{(t-1)}, x^{(t)}] + b_C), \quad (4)$$

$$C_t = f_{\text{gate}} * C_{t-1} + i_{\text{gate}} * C_{\text{tanh}}. \quad (5)$$

“输出门”(Output gate)用于确定细胞最终的输出信息。过程分为两步:首先是更新门的输出经过sigmoid层决定哪些信息用于输出,然后将该信息通过tanh层得到最终输出。输出门的具体过程表示为

$$O_{\text{gate}} = \sigma(W_O \cdot [a^{(t-1)}, x^{(t)}] + b_O), \quad (6)$$

$$a^{(t)} = O_{\text{gate}} * \tanh(C^t), \quad (7)$$

式(3)~式(7)中, W_i 、 W_O 和 W_C 为权值, b_i 、 b_C 和 b_O 为偏置。通过BP算法更新参数。本文使用LSTM方法处理一维水声时域数据。

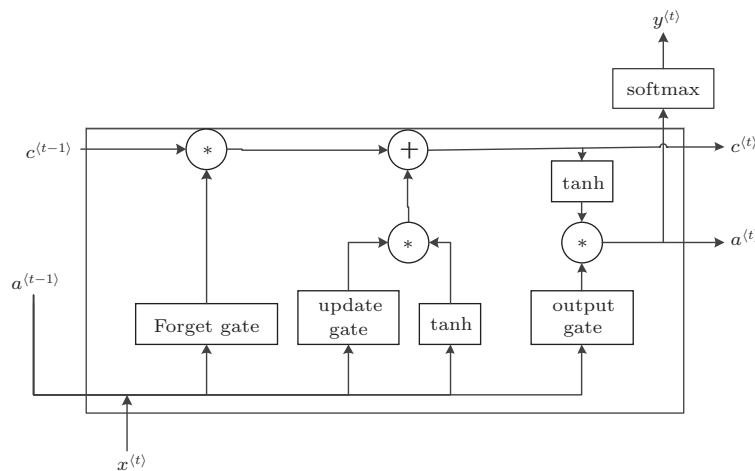


图2 LSTM网络结构示意图

Fig. 2 The structure of LSTM

1.3 融合自编码与关联分析的异构学习

基于特征融合的出发点是将水声信号描述为一维时域特征模态和二维时频模态,模型的输入是时域帧和频谱帧。采用稀疏表示的深度自编码多模态学习方法处理水声信号。深度自编码网络(Deep auto-encoder, DAE)是一种无监督逐层贪心训练算法,其能够从无先验信息的数据中提取数据的高维特征,主要结构为编码器、隐含层和解码器。考虑到水声信号在不同特征模态上的稀疏性,将稀疏理论用于深度自编码算法,即在编码器的隐含层添加约束用于提取高维数据变量的稀疏因子,且保留了输入的非零特征,使算法的鲁棒性更强,有利于增强数据的线性可分性。

融合自编码与关联分析的异构学习模型分为输入层、共享表示层和输出层。输入层中,输入为一维时域模态和二维频谱模态。其中时域模态利用LSTM进行特征训练,二维频谱模态采用CNN进行特征训练。共享表示层中,基于特征融合的思想,使用典型相关分析(Canonical correlation analysis, CCA)将LSTM提取的一维时域特征与CNN提取的二维频谱特征进行融合关联,形成融合特征矩阵。所谓典型相关分析是指将较多变量转换为较少典型变量,通过典型相关系数综合描述多元随机变量之间关系的统计分析方法。在输出层中,给出水声信号的分类识别结果。

2 水声信号特征模态时间相关特性参数优化

在水声信号分析过程中,连续的时间帧或频谱帧中相邻帧大概率包含相同目标的特征,也就是说相邻的信号帧具有相关性,有效挖掘和利用这种相关性便能够显著提高水下目标的分类识别能力,事实上利用时间相关性来提升目标识别准确率的方法已得到研究者的重视。

对水声信号而言,将二维频谱特征模态作为输入,利用CNN算法进行特征提取和分类识别,可以利用相邻帧之间的相关性优化训练参数。若不考虑相邻帧的相关性,CNN从水声信号频谱帧图中随机抽取部分作为训练样本,记为 $\{(x_n, y_n)\}_{n=1, \dots, N}$,其中 x_n 表示输入的频谱图像帧, y_n 为先验标签信息,为优化网络参数,建立似然函数 $L(\theta)$,待优化参数为 θ 。通常利用BP算法对参数进行优化。

$$L(\theta) = - \sum_{n=1}^N \lg P_{\theta}(y_n | x_n). \quad (8)$$

为了提升识别率,考虑利用相邻帧之间的相关性,令 x_1, x_2 表示相邻两帧频谱图,则二者在隐含层中产生的特征映射表示为 $z_{\theta}^l(x_1)$ 和 $z_{\theta}^l(x_2)$,使用 L_1 规范描述二者的相邻关系:

$$\|z_{\theta}^l(x_1) - z_{\theta}^l(x_2)\|_1, \text{ 若 } x_1 \text{ 和 } x_2 \text{ 连续}, \quad (9)$$

$$\max(0, \delta - \|z_{\theta}^l(x_1) - z_{\theta}^l(x_2)\|_1),$$

$$\text{若 } x_1 \text{ 和 } x_2 \text{ 不连续}, \quad (10)$$

其中, δ 表示边缘尺寸,是基于先验选取的超参数。将 x_1, x_2 输入CNN,使二者网络结构共享参数 θ ,得到特征映射 $z_{\theta}^l(x_1)$ 和 $z_{\theta}^l(x_2)$ 后,通过式(9)和式(10)计算出关于 θ 的相关梯度 $L_{\text{coh}}(\theta, x_1, x_2)$,基于BP算法对 θ 进行更新,这样便充分利用了相邻频谱帧之间的相关性。

3 利用时间相关的多模态深度学习算法

基于时间/频谱相关性的多模态深度学习流程如图3所示。

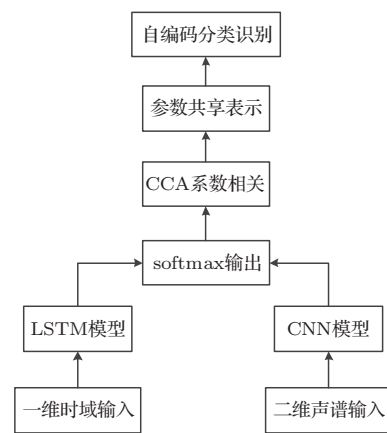


图3 多模态模态网络结构示意图

Fig. 3 The structure of heterogeneous multi-modal network

如图3所示,将水声信号的一维时域特征模态采用LSTM模型处理,将二维频谱特征模态采用CNN模型处理,各自输出的特征映射进行典型相关分析,得到参数共享表示,形成特征融合矩阵,最后利用深度稀疏自编码网络进行目标的分类识别。利用时间相关性的多模态学习模型本质上是对LSTM和CNN的联合概率模型进行优化。具体算法如下:

输入：带标签的训练样本，包含一维时域样本和二维频谱样本

输出：网络优化参数 θ ，连接参数 w ，分类识别标签，识别概率

for 训练样本集

step1: 随机选取 1 个训练样本集中的带标签样本；

step2: 针对一维时域样本输入，执行一次梯度下降迭代优化 LSTM 连接参数 w ；

step3: 针对二维频谱样本输入，若为连续的样本，则执行一次梯度下降迭代优化似然函数 $L(\theta)$ ；

step4: 针对二维频谱样本输入，如为非连续样本，则执行一次梯度下降迭代优化似然函数 $L_{coh}(\theta, x_1, x_2)$ ；

step5: 针对一维时域样本输入，再次执行一次梯度下降，迭代优化 LSTM 中的连接参数 w 。固定连接参数 w ，通过优化 θ 使得 LSTM 和 CNN 的联合概率模型 P 达到最大。

end

直到：迭代训练的误差小于设定的阈值

该算法充分利用水声信号帧之间的相关性，通过不同模态之间的训练迭代交替更新网络参数，能够有效地对多模态特征进行融合，从而有效提高分类识别精度。

4 试验验证

为了验证多模态深度学习方法在水下目标分类识别中的性能，本文利用 5 类合作水声目标辐射噪声数据进行算法验证。其中，3 类为水面舰船，2 类为水下高速航行体。每类目标有 5 段信噪比不同的时域信号，通过人工加噪的方式使得同一目标时域信号的信噪比分别为 12 dB、6 dB、0 dB、-6 dB、-12 dB。信号的采样率为 12 kHz。通常的关于水下目标识别的论文之中，训练集与测试集来自于同一段时域信号，而在本文实验中，并未采用“留出法”将样本集划分为互斥的两类集合，本文的训练集与测试集不是取自同一段时域信号，具体的取法是对每一类目标不同的时间段的 5 段时域信号，选取其中的 M 段时域信号作为训练集，选择 N 段作为测试集，那么训练-测试比为 $M : N$ ，这样便保证了训练集和测试集数据不属于同一段时域信号。

实验一：训练-测试比实验

训练 [12 dB, 0 dB, -6 dB, -12 dB]：测试比 [6 dB] = 4 : 1 时，样本制作方法如下：

(1) 一维时域信号 Dataset：采用分帧处理将时域信号截断成训练样本，信号分帧长度为 60 ms，相邻帧之间重叠率为 50%。每类目标的每段时域信号制作为 6000 个帧训练样本。每类水声目标选取 4 段信号作为样本，1 段信号作为测试样本，则训练集样本为 24000，测试集样本为 6000，训练样本与测试样本比例为 4 : 1。

(2) 二维频谱 Dataset：利用尺度不变特征变换 (Scale-invariant feature transform, SIFT) 制作声谱图，将 10 帧数据做频谱分析得到一张二维频谱样本，二维样本帧重叠率为 80%。与一维时域样本类似，最后生成的训练集总样本数为 42472。测试集样本数为 10668，测试样本与训练样本比例为 4 : 1。

将多模态深度学习模型与 SVM、CNN 以及 LSTM 网络识别方法进行分类识别率比较。其中 SVM 采用高斯核函数，惩罚因子 C 的选择依赖于多次试验，惩罚因子越大，表示分类方法不能容忍出现误差，但是容易出现过拟合，当惩罚因子越小时，表示分类模型对误差的宽容度较强，但容易出现欠拟合，经过多次优化测试，选择惩罚因子为 67.3，训练步数设置为 200。CNN 设置两层卷积层，一层池化层和一层全连接层。激活函数为 ReLU，两层卷积层的节点数分别设置为 16 和 32，卷积核大小分别为 $5 * 5$ 和 $3 * 3$ ，池化层卷积核大小为 $3 * 3$ ，一层全连接层，采用 Softmax 输出层，输出节点为 5，学习率设置为 10×10^{-5} ，epoch 为 10000 步，迭代直到达到设定阈值。LSTM 网络参数中，一维信号进行输入时首先进行归一化处理，动量因子设置 0.95，衰减因子设为 0.1，初始门控设置为 0，采用 L1 正则化方法，激活函数为 sigmoid，学习速率设置为 0.01，输入向量维度为 2，隐含层维度为 32，输出层维度为 1，迭代 10000 步。多模态深度学习模型的两种模态输入参数设置与上述一致，CCA 特征融合时初始迭代因子设置为 2，参数共享后，对特征向量进行深度自编码分类识别，自编码网络选择堆栈自编码 (Stacked auto-encoder, SAE)。采用 Tensorflow 深度学习框架^[18]进行训练。采用英伟达丽台 K4200 GPU 加速，每个模型训练时间达到 6 h。该训练-测试比条件下的结果如表 1 所示，

其中识别率=(测试集中正确分类样本数)/测试集总样本数。

表1 识别率比较

Table 1 Comparison of recognition rate

	一维时域数据集	二维声谱图集
SVM	73.2%	78.2%
CNN	—	85.4%
LSTM	87.2%	—
多模态深度学习模型	96.4%	

由表1可以看出,深度学习方法分类识别率均比传统的机器学习方法高,采用SVM对一维数据集和二维数据进行分类识别的精度分别为73.19%和78.23%,而采用CNN对二维频谱数据集测试的精度达到85.36%,采用LSTM网络对一维时域数据集的分类精度达到87.17%,即采用单一模态深度学习模型的分类识别精度比传统机器学习方法提升至少7%;采用多模态深度学习方法在同样数据集下,分类识别率又优于单一模态的深度学习方法,识别率提升至少8%,误差率只有3.65%。

训练[12 dB, 0 dB, -12 dB]:测试比[-6 dB, 6 dB] = 3 : 2时,样本制作方法与训练测试比为4 : 1时样本制作方法一致,即一维时域信号Dataset训练集样本为18000,测试集样本为12000,训练样本与测试样本比例为3 : 2,二维频谱Dataset训练集总样本数为32004,测试集样本数为21336,测试样本与训练样本比例为3 : 2。四种分类识别方法的参数与训练测试比为4 : 1时分类方法参数基本一致,不同之处在于,在测试比为3 : 2时,SVM惩罚因子为54。该训练-测试比条件下的结果如表2所示。

表2 识别率比较

Table 2 Comparison of recognition rate

	一维时域数据集	二维声谱图集
SVM	63.5%	65.7%
CNN	—	80.3%
LSTM	81.4%	—
多模态深度学习模型	90.6%	

由表2可以看出,深度学习方法分类识别率均比传统的机器学习方法高,采用单一模态深度学习模型的分类识别精度比传统机器学习方法提升至

少7%;采用多模态深度学习方法在同样数据集下,分类识别率又优于单一模态的深度学习方法,识别率提升至少9%,误差率只有9.4%。

通过表1和表2的比较可以看出,在不同的训练-测试比条件下,深度学习模型的分类识别率均优于SVM模型,而多模态深度学习模型的分类识别率又优于单一模态的深度学习模型;此外,随着训练测试比变大,四类分类模型的识别率均下降,这是由于训练样本量降低所导致的。

实验二:测试集信噪比实验

为了验证本文提出的多模态识别方法的鲁棒性,改变测试样本的信噪比,通过识别率的变化判断模型的鲁棒性。实验一中,给出了训练集[12 dB, 0 dB, -6 dB, -12 dB]:测试集[6 dB] = 4 : 1时的识别率,如表1所示。表3给出了训练集[12 dB, 6 dB, 0 dB, -12 dB]:测试集[-6 dB] = 4 : 1时的识别率,表4给出了训练集[12 dB, 6 dB, 0 dB, -6 dB]:测试集[-12 dB] = 4 : 1时的识别率结果。其中,测试集的信噪比依次下降6 dB。实验中参数与实验一基本一致。

表3 识别率比较

Table 3 Comparison of recognition rate

	一维时域数据集	二维声谱图集
SVM	71.5%	73.7%
CNN	—	81.3%
LSTM	85.4%	—
多模态深度学习模型	92.6%	

表4 识别率比较

Table 4 Comparison of recognition rate

	一维时域数据集	二维声谱图集
SVM	63.2%	64.3%
CNN	—	77.4%
LSTM	77.8%	—
多模态深度学习模型	88.5%	

通过表3和表4结果可以看出,随着测试样本的信噪比降低,单一模态深度学习模型识别率依然优于SVM模型,多模态深度学习模型识别率优于单一模态深度学习模型;随着测试集信噪比的降低,四种分类模型的识别率均下降,其中,SVM模型的识别率下降最为剧烈,从6 dB到-12 dB,下降最

大达13.9%，单一深度学习模型识别率下降最大可达9.3%，多模态深度学习模型识别率最大下降小于8%。在测试集为-12 dB时，多模态深度学习模型识别率达到88%以上，依然具有较好的分类识别效果，说明多模态深度学习模型的鲁棒性优于SVM模型和单一模态深度学习模型。

通过实验一和实验二可以看出，在不同测试比和信噪比条件下，多模态深度学习模型识别率均优于传统机器学习模型和单一模态深度学习模型。

5 结论

随着水声科学数据量的积累，迫切需要借鉴机器学习方法对数据处理，如进行水下目标分类识别。本文利用统计机器学习思想，对CNN和LSTM网络深度学习方法进行分析，从数据空间关联和特征融合的角度考虑，建立了水下目标多模态深度学习分类识别方法，充分利用了水声信号不同模态之间的特征，对融合特征进行共享关联表示，并基于信号的时间相关性优化了训练模型。利用实际的水声目标辐射噪声数据对算法进行了验证，将其分类识别性能与传统机器学习方法和单一模态深度学习方法进行了比较，结果表明本文提出的多模态深度学习方法的有效性和优越性，该方法为水下目标分类识别提供了新思路。

参 考 文 献

- [1] Deng L, Yu D. Deep learning: methods and applications[J]. *Foundations and Trends in Signal Processing*, 2014, 7(3/4): 197-387.
- [2] Suykens J A K, Vandewalle J. Least squares support vector machine classifiers[J]. *Neural Processing Letters*, 1999, 9(3): 293-300.
- [3] Zivkovic Z. Improved adaptive Gaussian mixture model for background subtraction[C]//Proc of the 17th Int Conf on Pattern Recognition. Piscataway, NJ: IEEE, 2004: 28-31.
- [4] Quattoni A, Collins M, Darrell T. Conditional random fields for object recognition[C]//Proc of the Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 2004: 1097-1104.
- [5] Arel I, Rose D C, Karnowski T P. Deep machine learning—a new frontier in artificial intelligence research[J]. *IEEE Computational Intelligence Magazine*, 2010, 5(4): 13-18.
- [6] 刘建伟, 刘媛, 罗雄麟. 深度学习研究进展 [J]. *计算机应用研究*, 2014, 31(7): 1921-1930, 1942.
Liu Jianwei, Liu Yuan, Luo Xionglin. The research and progress of deep learning[J]. *Application Research of Computers*, 2014, 31(7): 1921-1930, 1942.
- [7] Goodfellow I, Bengio Y. 深度学习 [M]. 北京: 人民邮电出版社, 2017.
- [8] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. *Science*, 2006, 313(5786): 504-507.
- [9] Hinton G, Deng L, Yu D, et al. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups[J]. *IEEE Signal Processing Magazine*, 2012, 29(6): 82-97.
- [10] Chen J L, Summers J E. Deep neural networks for learning classification features and generative models from synthetic aperture sonar big data[C]//172nd Meeting of the Acoustical Society of America. Honolulu, Hawaii, 28 November-02 December, 2016: 1-13.
- [11] Niu H, Reeves E, Gerstoft P. Source localization in an ocean waveguide using supervised machine learning[J]. *Journal of the Acoustical Society of America*, 2017, 142(3): 1176-1180.
- [12] Huang Z, Xu J, Gong Z, et al. Source localization using deep neural networks in a shallow water environment[J]. *Journal of the Acoustical Society of America*, 2018, 143(5): 2922-2932.
- [13] 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.
Yang Honghui, Shen Sheng, Yao Xiaohui, et al. Underwater acoustic target feature learning and recognition using hybrid regularization deep belief network[J]. *Journal of Northwestern Polytechnical University*, 2017, 35(2): 220-225.
- [14] 王强, 曾向阳. 深度学习方法及其在水下目标识别中的应用 [C]. 中国声学学会水声学分会 2015 年学术会议论文集, 2015.
- [15] Mobahi H, Collobert R, Weston J. Deep learning from temporal coherence in video[C]// Proceedings of the 26th Annual International Conference on Machine Learning. Montreal, Quebec, Canada: ACM, 2009: 737-744.
- [16] 程东阳. 基于无监督特征学习的多媒体内容理解算法研究 [D]. 上海: 上海交通大学, 2014.
- [17] 由清圳. 基于深度学习的视频人脸识别方法 [D]. 哈尔滨: 哈尔滨工业大学, 2012.
- [18] 蒋子阳. TensorFlow 深度学习算法原理与编程实战 [M]. 北京: 中国水利水电出版社, 2018.