



轻量型胶囊网络语音情感识别方法

王 颖^{1*}, 高 胜²

(1. 东北石油大学计算机与信息技术学院 黑龙江 大庆 163318; 2. 东北石油大学机械科学与工程学院 黑龙江 大庆 163318)

【摘要】针对目前语音情感识别模型参数多、运算量大、训练速度慢等问题,提出了一种适用于小数据集、轻量型的网络模型。模型以胶囊网络为基础结构,引入深度可分离卷积模块代替胶囊网络中原有的卷积层以减少计算量。基于迁移学习提取普适的底层图像特征,利用语谱图来微调整个网络,减弱模型在小数据集上的过拟合现象。再利用夹角余弦来计算动态路由结构中向量的相似度,提高动态路由算法性能。实验结果表明,轻量型胶囊网络的识别率和运算速度均优于对比的 7 种深度学习网络模型。

关 键 词 胶囊网络; 深度可分离卷积; 语音情感识别; 迁移学习

中图分类号 TN912.3 **文献标志码** A **doi**:10.12178/1001-0548.2022086

A Speech Emotion Recognition Method Based on Lightweight Capsule Network

WANG Ying^{1*} and GAO Sheng²

(1. School of Computer and Information Technology, Northeast Petroleum University Daqing Heilongjiang 163318;

2. School of Mechanical Science and Engineering, Northeast Petroleum University Daqing Heilongjiang 163318)

Abstract Aiming at the problems of many parameters, large amount of computation and slow training speed of the current speech emotion recognition model, this paper proposes a lightweight network model suitable for small data sets. The model is based on the capsule network, and the deep separable convolution module is introduced to replace the original convolution layer in the capsule network to reduce the amount of calculation. Transfer learning is used to extract the universal underlying image features, and then spectrogram is used to finely tune the over fitting phenomenon of the whole network weakening model on small data sets. The angle cosine is used to calculate the vector similarity in the dynamic routing structure so as to improve the performance of the dynamic routing algorithm. The experimental results show that the recognition rate and operation speed of the lightweight capsule network are better than the seven deep learning network models.

Key words capsule network; depth separable convolution; speech emotion recognition; transfer learning

语音情感识别是建立语音信号与人类情感之间映射关系的一种方法^[1],主要包括情感特征提取和分类模型建立两部分工作。情感特征提取的好坏是影响语音情感识别准确率的关键^[2]。然而由于人类情感的复杂性和多样性,使得传统的语音情感特征提取方法面临一定挑战^[3]。近年来,深度学习依靠其较强的学习非线性样本特征及复杂系统建模的能力,为语音情感识别研究带来了很大进展。文献[4]结合深度神经网络和HMM(hidden Markov model)

方法,在柏林数据集上进行情感识别验证,识别率达到77.92%。文献[5]以TIMIT语音数据库(由德州仪器TI、麻省理工学院MIT和坦福研究院SRI合作构建的声学-音素连续语音语料库)作为数据源,在DNN(deep neural networks)中引入CNN(convolutional neural network)模型,与DNN相比识别准确率获得了较大提升。文献[6]以AlexNet网络模型为基础进行语音信号的情感分析,获得了较好的识别效果。文献[7]提出了一种基于注意力

收稿日期:2022-03-22;修回日期:2023-01-29

基金项目:国家自然科学基金(61720293);国家重点研发计划(2018YFE0196000);黑龙江省自然科学基金(F2018003);黑龙江省博士后专项(LBH-Q20077)

作者简介:王颖(1986-),女,副教授,主要从事深度学习、情感计算、智能优化等方面的研究。

*通信作者:王颖, E-mail: nepuwy@163.com

的长短期记忆网络 (long short term memory, LSTM) 的语音情感识别方法, 分别在 3 个情感语料库上进行了实验, 获得了理想的识别精度。文献 [8] 提出了一种基于深度学习的多模式权重网络模型, 利用语音声学特征、语音转化成文本信息后的语义特征、语音频谱特征建立了 3 种分类器。然后基于这 3 种分类器的分类结果作为权重并构建权重网络分类器以实现语音情感识别。文献 [9] 以 MFCC (Mel frequency cepstrum coefficient)、波形图、语谱图并行作为输入, 分别设计不同的 CNN 网络模型, 然后结合注意力机制获得最终分类结果, 并在柏林情感语料库 (EMO-DB) 和多模态情绪识别数据集 (IEMOCAP) 上验证了所提方法的有效性。文献 [10] 提取语音文件的梅尔频谱系数、色谱图、语谱图、Tonnetz 等参数作为 CNN 网络的输入进行语音情感识别, 并在情感说话/歌唱视听数据集 (RAVDSS)、EMO-DB 以及 IEMOCAP 数据集上进行了实验, 分别得到 71.61% (8 个类别)、86.1% (7 个类别) 和 64.3% (4 个类别) 的准确率。

尽管深度学习的广泛应用使得语音情感识别的性能有了显著提升, 但是深度学习的识别效果需要海量样本数据作为支撑, 如果样本数据不足, 模型会受到过拟合因素的影响且训练难度增加, 识别准确率下降^[1]。此外, 神经元的输入输出均为标量形式导致隐含特征提取不足, 且池化层的引入也会造成有用特征信息的丢失。考虑到传统卷积神经网络的缺点, 文献 [12] 提出了胶囊网络 (capsule network), 它的神经元输入输出改为向量形式, 可以保留确切的空间位置等信息, 与此同时舍弃了池化层结构, 只需较小的数据集就能训练出比较优秀的网络, 已在机械设备故障诊断等领域中初露锋芒^[13-14]。文献 [15] 利用胶囊网络对滚动轴承进行故障诊断, 分类精度上达到了 99.95%, 在收敛速度方面得到较大的提高, 并对网络模型的自适应挖掘高层特征能力进行了验证。文献 [16] 利用胶囊结构构建向量神经元, 引入了 Inception 结构和空间注意力机制进行特征提取, 实现了滚动轴承故障诊断, 在噪声环境下能够得到良好的诊断结果。

本文提出了一种轻量型胶囊网络 (lightweight capsule network, LCN) 进行语音情感识别。以胶囊网络作为主干网络, 利用其矢量化特性和动态路由算法充分挖掘特征空间信息。同时在 LCN 中融合深度分离卷积和参数迁移, 进一步降低了网络计算量、减弱了模型在小数据集上的过拟合现象。通过

试验证明了该方法的适用性, 为快速精准的语音情感识别提供参考。

1 轻量型胶囊网络模型原理

1.1 深度可分离卷积

深度可分离卷积^[17-19] (depthwise separable convolutions) 将通道域和空间域分开处理, 将标准的卷积操作分解成深度卷积 (depthwise convolution) 和逐点卷积 (pointwise convolution) 两步。深度卷积层卷积核的数量与输入特征图的通道数相同, 每个卷积核只和输入的一个通道进行卷积。逐点卷积是利用通道数相同、大小为 1×1 的卷积核将上一层输出的不同通道特征图进行线性组合, 这种分解能够大幅减少计算量和模型参数个数。

1) 标准卷积层是以 $D_F \times D_F \times M$ 作为输入, 以 $D_G \times D_G \times N$ 作为输出, 卷积核为 $D_K \times D_K \times M \times N$, 一个标准卷积层计算消耗为:

$$D_K \times D_K \times M \times N \times D_F \times D_F$$

2) 深度卷积核为 M 个 $D_K \times D_K \times 1$, 计算消耗为:

$$D_K \times D_K \times M \times D_F \times D_F$$

3) 逐点卷积核为 N 个 $1 \times 1 \times M$, 计算消耗为:

$$M \times N \times D_F \times D_F$$

通过二者的比值可以看出模型计算量得到了大幅度减小:

$$\frac{D_K \times D_K \times M \times D_F \times D_F + M \times N \times D_F \times D_F}{D_K \times D_K \times M \times N \times D_F \times D_F} = \frac{1}{N} + \frac{1}{D_K^2}$$

1.2 胶囊网络模型

文献 [20-24] 提出了胶囊网络, 其由胶囊组成, 胶囊是一组神经元, 如图 1 所示。

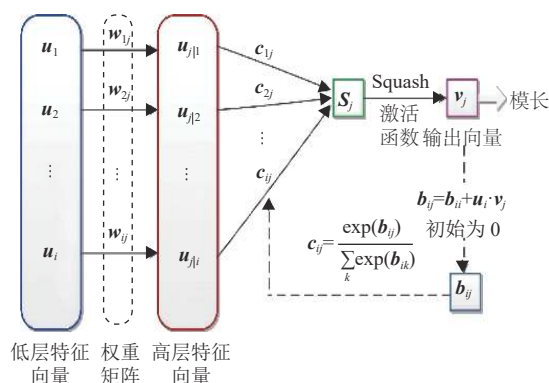


图 1 胶囊结构

网络主要包括卷积层、基础胶囊层和数字胶囊层, 其与卷积神经网络主要存在以下区别: 输入从

标量神经元变为向量胶囊,与标量神经元相比具有更丰富的表达能力;激活函数从 ReLU 改为 Squash 激活函数;池化层被动态路由替代。

低层特征向量 u_i 为输入向量,分别乘上对应的权重矩阵 w_{ij} 进而得到高层特征向量 $u_{j|i}$ 。 $u_{j|i}$ 乘上 c_{ij} 再进行累加求和得到 S_j , 其中 c_{ij} 由动态路由决定。利用 Squash 激活函数对 S_j 进行非线性压缩变换得到输出向量 v_j , Squash 激活函数将模长压缩到 0~1 之间,以表示分类概率,且 v_j 和 S_j 同向,具体计算如下:

$$v_j = \frac{\|S_j\|^2}{1 + \|S_j\|^2} * \frac{S_j}{\|S_j\|}$$

胶囊神经元之间采用动态路由的方式对输入的

特征进行聚类,即根据特征的相似程度对特征进行选择,从而替代池化层。算法过程如下:

1) b_{ij} 为相邻层两个胶囊间的特征传递参数,初始值为 0;

2) 利用 softmax 计算两层胶囊间权重 $c_{ij} =$

$$\frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})};$$

3) 激活向量 $v_j = \text{Squash}(\sum c_{ij} u_{ij})$;

4) 更新权重, $b_{ij} = b_{ij} + u_i v_j$ 。

1.3 轻量型的胶囊网络模型

为减小胶囊网络的计算密集,降低模型参数数量,提高小样本识别准确率,本文提出一种轻量型的胶囊网络,模型结构如图 2 所示。

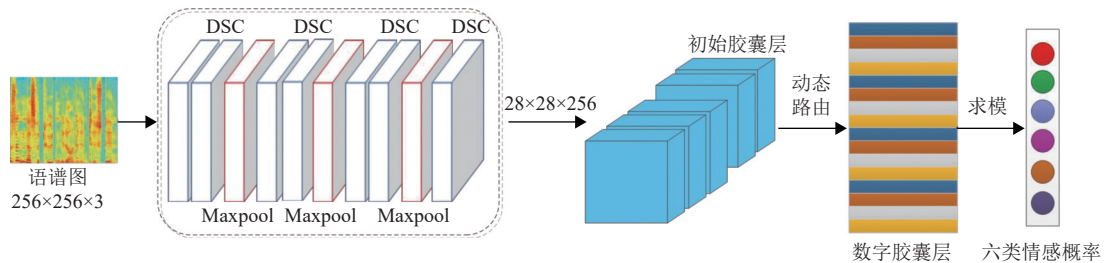


图 2 LCN 模型结构

1) 引入深度可分离卷积结构

引入深度可分离卷积模块替换原有胶囊网络中的标准卷积结构,模块共包括 7 个深度可分离卷积

层和 3 个最大池化层,各层级信息如表 1 所示。DSC 代表深度可分离卷积,“DSC-DC”代表逐通道卷积,“DSC-PC”代表逐点卷积,输入为 256×256×3。

表 1 DSC 模块各层参数

层号	层/卷积核大小	卷积核个数/个	参数个数/个	输出形状
1	DSC-DC 3×3×1	3	3×3×3+1×1×3×64=219	256, 256, 64
	DSC-PC 1×1×3	64		
2	DSC-DC 3×3×1	64	3×3×64+1×1×64×64=4 672	256, 256, 64
	DSC-PC 1×1×64	64		
3	Maxpool 2×2		0	128, 128, 64
4	DSC-DC 3×3×1	64	3×3×64+1×1×64×128=8 768	128, 128, 128
	DSC-PC 1×1×64	128		
5	DSC-DC 3×3×1	128	3×3×128+1×1×128×128=17 536	128, 128, 128
	DSC-PC 1×1×128	128		
6	Maxpool 2×2		0	64, 64, 128
7	DSC-DC 3×3×1	128	3×3×128+1×1×128×256=33 920	64, 64, 256
	DSC-PC 1×1×128	256		
8	DSC-DC 3×3×1	256	3×3×256+1×1×256×256=67 840	64, 64, 256
	DSC-PC 1×1×256	256		
9	Maxpool 2×2		0	32, 32, 256
10	DSC-DC 5×5×1	256	5×5×256+1×1×256×256=71 936	28, 28, 256

2) 改变动态路由权重更新方法

动态路由算法原来的权重更新方法为向量的点乘法, 即 $\mathbf{b}_{ij} = \mathbf{b}_{ij} + \mathbf{u}_i \cdot \mathbf{v}_j$ 。为进一步优化动态路由结构中向量相似度的计算方法, 本文引入夹角余弦来计算向量相似度。余弦相似度的实质是比较几何空间中两个向量的关联程度, 降低对绝对值的敏感性, 适用于离散数据之间距离的测量。夹角余弦越大, 表示两个向量的夹角越小, 相似度越高; 夹角余弦越小, 表示两个向量的夹角越大, 相似度越低。具体计算如下:

$$\mathbf{b}_{ij} = \mathbf{b}_{ij} + \frac{\mathbf{u}_i \cdot \mathbf{v}_j}{\|\mathbf{u}_i\| \|\mathbf{v}_j\|}$$

式中, $\mathbf{u}_i$ 为第 i 个卷积层输出即胶囊网络的输入; $\mathbf{v}_j$ 为胶囊网络输出向量; $\mathbf{b}_{ij}$ 为相邻层两个胶囊间的特征传递参数。

3) 参数迁移

ImageNet 是现有最大的图像识别数据集, 其包含了 1 400 多万张有标注图像。通过对 ImageNet 进行预训练可获得丰富的底层特征, 但 ImageNet 数据与语音情感数据存在较大差异, 因此拟采用迁移学习方法对 ImageNet 预训练模型进行微调, 缓解小样本情感数据易造成过拟合的问题。

本文首先在 ImageNet 源任务上对模型进行预训练, 然后用语音情感数据对预训练获得的模型参数进行微调, 找到能够共享的模型参数, 从而提高识别精度。由于低层网络学习到的图片特征为泛化特征, 如边角、颜色等, 不同的数据集差别不大, 因此这部分参数迁移后可不参与微调^[25]。而高层网络学习的特征能体现出差异性和全局性, 参数迁移后不能冻结, 需进行微调。

1.4 损失函数

利用损失函数调整预测值与实际值的差异, 反向优化模型参数, 提高模型识别准确率。在训练过程中, 间隔损失函数 (margin loss)^[26] 使得胶囊网络倾向于输出一个长向量和若干个短向量, 向量的长度代表胶囊存在的概率。因此, LCN 模型采用间隔损失函数, 其表达式为:

$$L_k = T_k \max(0, m^+ - \|\mathbf{V}_k\|)^2 + \lambda(1 - T_k) \max(0, \|\mathbf{V}_k\| - m^-)^2$$

式中, L_k 表示类别 K 的间隔损失, 取值 1 表示第 K 类存在, 否则取值 0; m^- 取值 0.1, 是 $\|\mathbf{V}_k\|$ 下边界, 避免假阴性; m^+ 取值 0.9, 是 $\|\mathbf{V}_k\|$ 上边界, 避免假阳性; λ 取值 0.5, 为比例系数。总损失为各样例损失之和。

2 实验结果与分析

2.1 数据集

本文的数据源为中国科学院自动化研究所录制的 CASIA 汉语情感语料数据库, 该数据库由 2 男 2 女共 4 位录音人在纯净录音环境下 (信噪比约为 35 dB) 对 500 句文本进行演绎录制而成, 包括 6 类情感 (高兴/joy、悲哀/sadness、生气/anger、惊吓/shock、中性/neutral、害怕/fear)。采样率为 16 kHz, 16 bit 量化, 最终保留其中 9 600 句。为验证 LCN 模型在小样本集上的识别性能, 本实验将数据源分成 10 份, 每份包含 960 句, 其中训练集 800 句, 测试集 160 句。经过 10 次轮换训练及测试, 取平均值作为最终的识别结果。

2.2 数据预处理

对语音数据集进行分帧、加窗和傅里叶变换, 生成语谱图集合:

$$S_p = \sum_{t=0}^{T-1} \text{FFT}(s(t)h(t)e^{-\frac{2\pi i}{T}kt})$$

式中, $s(t)$ 为输入语音信号; T 为窗长; $h(t)$ 为汉明窗函数; FFT 为傅里叶变换函数; $k \in [0, T]$ 。实验中汉明窗长度 T 取 512, 语音帧叠加为 240 个波形点数, 生成语谱图为 $256 \times 256 \times 3$ 的图像。

2.3 实验分析

为检验模型性能, 将 LCN 分别与以下模型进行对比。

1) 普通卷积神经网络 CNN, 模型结构如表 2 所示;

2) VGG-16, 模型结构如表 2;

3) ResNet-50, 模型结构如表 2;

4) 传统胶囊网络 CN, 模型结构如图 1;

5) CN-DSC, 在 4) 的基础上融入深度可分离卷积模块, 即将 4) 中的卷积层替换为 DSC;

6) CN-DSC-PM, 在 5) 的基础上引入参数迁移, 即对 5) 中的 DSC 层引入参数迁移;

7) CN-DSC-DR, 在 5) 的基础上改进动态路由算法。

不同方案下的语音情感识别结果如表 3 所示。

对比模型 1)、2)、3)、4) 发现, ResNet-50 由于网络层数较多、结构复杂使其在准确率和收敛速度上均表现较差。模型 4) 与模型 1)、2)、3) 相比尽管收敛速度没有明显改善, 但在小数据集上的识别准确率得到了一定的提升, 从而也验证了胶囊网

络在小数据集上的有效性。对比模型 4)、5) 发现 DSC 模块的引入使得网络在运算时间、模型参数数量和复杂度方面有了明显提升, 进而提升了情感识别的实时性。对比模型 5)、6) 得出采用参数迁移的方式提高了识别的准确率和训练速度, 有效避免了过拟合现象。模型 6) 首先利用 ImageNet 数据集

进行训练, 然后从训练好的网络模型中选取不同的 DSC 层参数进行迁移, 对不同迁移方案的识别结果进行了进一步对比。通过表 4 数据可以看出迁移前五层效果最佳, 而迁移前六层和前七层准确率反而下降了, 原因可能是层数越高 ImageNet 数据集和 CASIA 数据集的特征差别越明显。

表 2 模型结构

CNN	VGG-16	ResNet-50
Conv : $[3 \times 3, 64] \times 1$ ReLU MP : $[2 \times 2]$, stride = 2	Conv : $[3 \times 3, 64] \times 3$ ReLU MP : $[2 \times 2]$, stride = 2	Conv : $[7 \times 7, 64] \times 1$ MP : $[3 \times 3]$, stride = 2
Conv : $[3 \times 3, 128] \times 1$ ReLU MP : $[2 \times 2]$, stride = 2	Conv : $[3 \times 3, 128] \times 3$ ReLU MP : $[2 \times 2]$, stride = 2	Conv : $\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
Conv : $[3 \times 3, 256] \times 1$ ReLU MP : $[2 \times 2]$, stride = 2	Conv : $[3 \times 3, 256] \times 3$ ReLU MP : $[2 \times 2]$, stride = 2	Conv : $\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$
Conv : $[3 \times 3, 512] \times 1$ ReLU MP : $[2 \times 2]$, stride = 2	Conv : $[3 \times 3, 512] \times 3$ ReLU MP : $[2 \times 2]$, stride = 2	Conv : $\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1\,024 \end{bmatrix} \times 6$
	Conv : $[3 \times 3, 512] \times 3$ ReLU MP : $[2 \times 2]$, stride = 2	Conv : $\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2\,048 \end{bmatrix} \times 3$
FC	FC	FC

表 3 不同方案的实验结果

模型	情感类型						准确率平均值/%	耗时/h
	高兴	悲哀	生气	惊吓	中性	害怕		
CNN	80.55	84.19	85.75	81.44	82.61	84.83	83.23	9.6
VGG-16	79.65	83.63	85.05	80.57	81.82	84.09	82.47	9.8
ResNet-50	78.01	83.71	84.36	80.65	81.23	83.72	81.95	10.7
CN	83.42	88.45	90.04	84.93	86.91	88.74	87.08	9.5
CN-DSC	86.45	92.79	93.16	89.16	90.43	92.58	90.76	7.0
CN-DSC-PM	92.31	93.93	95.41	92.58	93.12	94.32	93.60	5.1
CN-DSC-DR	90.71	93.71	94.28	91.87	92.11	93.59	92.71	5.9
LCN	90.95	95.1	96.62	93.46	94.85	95.14	94.35	4.9

表 4 迁移不同 DSC 层参数结果对比

迁移层	准确率/%	耗时/h
前三层	91.24	5.4
前四层	92.38	5.3
前五层	93.60	5.1
前六层	92.10	5.0
前七层	91.07	5.0

对比模型 5)、7) 得出动态路由算法中权重更新方法的改进增加了对胶囊向量方向的考虑, 获得了更高的识别精度。综合对比模型 LCN 与模型 1)~7), 得出模型 LCN 在小样本集的应用中具有更高的准确率和更快的训练速度。迭代过程中

LCN 在测试集和验证集上的准确率和损失值变化如图 3 所示, 可以看出, 本文模型在迭代次数为 20 时就已收敛且获得了较高的准确率。

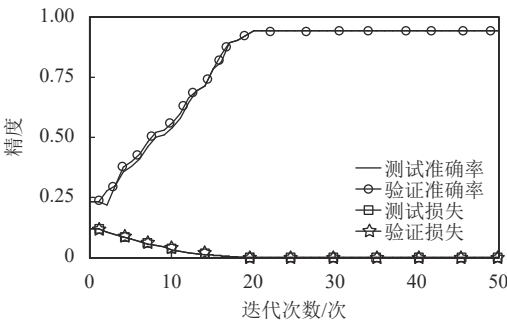


图 3 LCN 在测试集和验证集上准确率和损失值的迭代曲线

3 结束语

本文通过对语音情感识别的研究,提出了一种轻量化胶囊网络识别模型,并在小样本数据上进行了验证。实验结果表明:结合深度可分离卷积的胶囊网络可有效降低模型复杂度;模型参数迁移可以进一步避免微数据集带来的过拟合现象;动态路由算法的改进进一步提高了模型的精度。该方法取得了较好的实验效果,在未来的工作中,将继续研究轻量化模型的构建方法,提高识别的准确度和实时性。

参 考 文 献

- [1] 郑丽. 语音情感识别及其在服务机器人中的应用研究[D]. 长春: 东北师范大学, 2018.
ZHENG L. Research on speech emotion recognition and its application in the service robot[D]. Changchun: Northeast Normal University, 2018.
- [2] 韩文静, 李海峰, 阮华斌, 等. 语音情感识别研究进展综述[J]. 软件学报, 2014, 25(1): 37-50.
HAN W J, LI H F, RUAN H B, et al. Review on speech emotion recognition[J]. Journal of Software, 2014, 25(1): 37-50.
- [3] DAHAKE P P, SHAW K, MALATHI P. Speaker dependent speech emotion recognition using MFCC and support vector machine[C]//International Conference on Automatic Control and Dynamic Optimization Techniques. India: IEEE, 2016: 1080-1084.
- [4] LI L, ZHAO Y, JIANG D, et al. Hybrid deep neural network hidden markov model (DNN-HMM) based speech emotion recognition[C]//Affective Computing and Intelligent Interaction. [S.l.]: IEEE, 2013: 312-317.
- [5] SAINATH T N, VINAYALS O, SENIOR A, et al. Convolutional, long short-term memor, fully connected deep neural networks[C]//IEEE International Conference on Acoustics, Speech and Signal Processing. New York: IEEE, 2015: 4580-4584.
- [6] BADSHAH A M, AHMAD J, RAHIM N, et al. Speech emotion recognition from spectrograms with deep convolutional neural network[C]//Proceedings of the International Conference on Platform Technology and Service. [S.l.]: IEEE, 2017: 1-5.
- [7] XIE Y, LIANG R, LIANG Z, et al. Speech emotion classification using attention-based LSTM[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2019, 27(11): 1675-1685.
- [8] 张晓宇, 张华熊, 高强. 基于深度学习的多模式权重网络语音情感识别[J]. 大连理工大学学报, 2022, 62(5): 526-534.
ZHANG X Y, ZHANG H X, GAO Q. Multi-Modal weighted network for speech emotion recognition based on deep learning[J]. Journal of Dalian University of Technology, 2022, 62(5): 526-534.
- [9] CHO Y I. Modeling speech emotion recognition via attention-oriented parallel CNN encoders[J]. Electronics, 2022, 11: 4047.
- [10] ISSA D, DEMIRCI M F, YAZICI A. Speech emotion recognition with deep convolutional neural networks[J]. Biomedical Signal Processing and Control, 2020, 59: 101894.
- [11] 李靓, 孙存威, 谢凯, 等. 基于深度学习的小样本声纹识别方法[J]. 计算机工程, 2019, 45(3): 262-267.
LI J, SUN C W, XIE K, et al. Small sample voiceprint recognition method based on deep learning[J]. Computer Engineering, 2019, 45(3): 262-267.
- [12] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554.
- [13] HUANG R, LI J, LI W, et al. Deep ensemble capsule network for intelligent compound fault diagnosis using multisensory data[J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(5): 2304-2314.
- [14] WANG Y, NING D, FENG S. A novel capsule network based on wide convolution and multi-scale convolution for fault diagnosis[J]. Applied Sciences-Basel, 2020, 10(10): 3659-3675.
- [15] 董建伟, 王衍学. 基于残差胶囊网络的滚动轴承故障诊断研究[J]. 机电工程, 2021, 8(10): 1292-1298.
DONG J W, WANG Y X. Fault diagnosis of rolling bearing based on residual capsule network[J]. Journal of Mechanical & Electrical Engineering, 2021, 8(10): 1292-1298.
- [16] 孙岩, 彭高亮. 改进胶囊网络的滚动轴承故障诊断方法[J]. 哈尔滨工业大学学报, 2021, 53(1): 23-28.
SUN Y, PENG G L. Improved capsule network method for rolling bearing fault diagnosis[J]. Journal of Harbin Institute of Technology, 2021, 53(1): 23-28.
- [17] KAMAL K C, YIN Z D, WU M Y, et al. Depthwise separable convolution architectures for plant disease classification[J]. Computers and Electronics in Agriculture, 2019, 165(10): 104948.
- [18] HOWARD A, ZHU M L, CHEN B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications[J]. Applied Intelligence, 2019, 50(1): 107-118.
- [19] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: Inverted residuals and linear bottlenecks[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City: IEEE, 2018: 4510-4520.
- [20] 杨平, 苏燕辰, 张振. 基于卷积胶囊网络的滚动轴承故障诊断研究[J]. 振动与冲击, 2020, 39(4): 55-62.

- YANG P, SU Y C, ZHANG Z. A study on rolling bearing fault diagnosis based on convolution capsule network[J]. *Journal of Vibration and Shock*, 2020, 39(4): 55-62.
- [21] 张天柱, 邹承明. 使用模糊聚类的胶囊网络在图像分类上的研究[J]. *计算机科学*, 2019, 46(12): 279-285.
- ZHANG T Z, ZOU C M. Study on image classification of capsule network using fuzzy clustering[J]. *Computer Science*, 2019, 46(12): 279-285.
- [22] 张文凤, 周俊. 基于 Dropout-CNN 的滚动轴承故障诊断研究[J]. *轻工机械*, 2019, 37(2): 62-67.
- ZHANG W F, ZHOU J. Fault diagnosis method of rolling bearing based on Dropout-CNN[J]. *Light Industry Machinery*, 2019, 37(2): 62-67.
- [23] CHEN T, WANG Z, YANG X, et al. A deep capsule neural network with stochastic delta rule for bearing fault diagnosis on raw vibration signals[J]. *Measurement*, 2019, 148(12): 106857.
- [24] CAI W, ZHOU J, WANG J, et al. A novel method for intelligent fault diagnosis of bearing based on capsule neural network[J]. *Complexity*, 2019(7): 1-17.
- [25] 缪裕青, 邹巍, 刘同来, 等. 基于参数迁移和卷积循环神经网络的语音情感识别[J]. *计算机工程与应用*, 2019, 55(10): 135-140.
- MIAO Y Q, ZOU W, LIU T L, et al. Speech emotion recognition model based on parameter transfer and convolutional recurrent neural network[J]. *Computer Engineering and Applications*, 2019, 55(10): 135-140.
- [26] 赵小强, 张和慧. 基于交叉熵的改进 NPE 间歇过程故障检测算法[J]. *控制与决策*, 2021, 36(2): 411-417.
- ZHAO X Q, ZHANG H H. Improved NPE batch process fault detection algorithm based on cross entropy[J]. *Control and Decision*, 2021, 36(2): 411-417.

编辑 税红