

# 基于变压器声纹 Mel 语谱图-ResNet 的铁心松动故障诊断\*

何萍, 李勇, 陈寿龙, 许洪华, 朱雷, 王凌燕  
(国网江苏电力有限公司南京供电分公司, 江苏 南京 210019)

**摘要:** 为实现变压器铁心松动故障的识别, 提出基于变压器声纹的 ResNet 卷积神经网络(CNN)用作铁心松动故障的识别, 分别比较了相同卷积神经网络 ResNet 在交叉熵损失函数(SE-ResNet-Dense)和特征表达的角度空间中最大化分类界限的加性角度裕度损失函数(SE-ResNet-ArcLoss)不同表现效果。通过变压器空载试验采集变压器铁心在额定预紧力、松动 20%、松动 40% 时的噪声信号, 通过离散傅里叶变换将采集的声纹信号生成时频矩阵, 并使用 Mel 滤波器对时频矩阵降维生成尺寸大大缩小的 Mel-语谱图。将采集的噪声信号制作成数据集后输入到两种模型中训练, 最终测试集在模型 SE-ResNet-Dense 上的预测结果为 90.753%, 在模型 SE-ResNet-ArcLoss 上的预测结果为 97.541%。结果验证 SE-ResNet-ArcLoss 最适用于变压器铁心松动故障识别。

**关键词:** 变压器声纹; 铁心松动故障; Mel 语谱图; 卷积神经网络; 故障识别

中图分类号: TM41 文献标志码: A 文章编号: 1673-6540(2022)09-0075-06

doi: 10.12177/emca.2022.084

## Fault Diagnosis of Iron Core Looseness Based on Mel Spectrogram-ResNet with Transformer Voiceprint\*

HE Ping, LI Yong, CHEN Shoulong, XU Honghua, ZHU Lei, WANG Lingyan  
(Nanjing Power Supply Branch, State Grid Jiangsu Electric Power Co., Ltd., Nanjing 210019, China)

**Abstract:** In order to realize the identification of transformer core looseness fault, ResNet convolution neural network (CNN) based on transformer voiceprint is proposed as the identification of core looseness fault. The different performance effects of the same ResNet-CNN in the cross entropy loss function (SE-ResNet-Dense) and the additive angular margin loss function (SE-ResNet-ArcLoss) which maximizes the classification boundary in the angle space of feature expression are compared, respectively. The noise signals of the transformer core at the rated preload, 20% looseness and 40% looseness are collected through the transformer no-load experiment. The collected voiceprint signal is generated into the time-frequency matrix through the discrete Fourier transform, and the Mel filter is used to reduce the dimension of the time-frequency matrix to generate the Mel spectrogram with greatly reduced size. The collected noise signal is made into a data set and input into the two models for training. The prediction result of the final test set on the model SE-ResNet-Dense is 90.753%, and that on the model SE-ResNet-ArcLoss is 97.541%. The results show that SE-ResNet-ArcLoss are most suitable for transformer core looseness identification.

**Key words:** transformer voiceprint; iron core looseness fault; Mel spectrogram; convolutional neural network (CNN); fault identification

收稿日期: 2022-04-12; 收到修改稿日期: 2022-07-21

\* 基金项目: 江苏省电力有限公司重点科技项目(J2021053)

作者简介: 何萍(1980—), 女, 高级工程师, 研究方向为电力系统及其自动化。

## 0 引言

随着经济的不断发展,社会对电能的需求逐年提高,而变压器作为最重要的电力设备之一,其投入运行的数量也在不断增加,但由变压器运行带来的噪声问题也不容忽视<sup>[1]</sup>。国内外学者对变压器噪声产生的机理进行了深入研究。变压器的噪声中含有大量能反映变压器运行状态或变压器故障状态的信息。因此,可以通过数据挖掘技术提取变压器声音信号中的发生变化的特征,就可以对变压器运行状态进行识别。

对于变压器铁心松动故障的诊断,目前的研究主要集中在针对振动信号的分析上。文献[2]提出了利用振动信号中 100 Hz 特征频率分量占总分量(1 kHz 以内)的比值来判断铁心松动故障的方法,指出变压器铁心松动后,50 Hz 分量及其部分倍频分量、基频的倍频分量等新特征频率能够反映故障,能有效识别变压器铁心的松动故障。但是由于振动信号的采集依靠的是振动传感器紧贴在变压器的侧壁上,采集的信号存在干扰,而利用声纹诊断时采用非接触式,采集到的信号更加纯净。在信号处理方面,王丰华等<sup>[3]</sup>提出基于加权降维 Mel 倒谱系数(MFCC)和矢量量化模型(VQ)的变压器声纹识别模型,将提取出的特征值与机器学习相结合的方法,但识别率还需要进一步提高。舒畅等<sup>[4]</sup>提出了基于自适应白噪声完备集成经验模态分解(CEEMDAN)的配电变压器放电故障诊断方法,该方法在放电故障的识别率达到 90% 以上,但是针对的是放电故障。目前利用深度学习对变压器声纹的铁心松动故障研究较少<sup>[5]</sup>。

基于上述分析,针对变压器铁心松动故障识别,本文设计两种声纹识别模型,分别是带注意力机制的 ResNet + 交叉熵损失函数(SE-ResNet-Dense)组成的识别模型和带注意力机制的 ResNet + ArcLoss 损失函数(SE-ResNet-ArcLoss)组成的识别模型。模型的输入均是利用短时傅里叶变换(STFT)提取时频矩阵并通过 Mel 滤波器对其降维后生成的语谱图,比较这两种模型的识别效果,确定最适合变压器铁心松动故障的声纹识别模型。

## 1 噪声信号预处理

### 1.1 语谱图绘制

语谱图能直观地表示语音信号随时间变化的

频谱特性,任一给定频率成分在给定时刻的强弱用相应点的灰度或色调的浓淡来表示<sup>[6]</sup>。正确建立噪声信号时频域的关系,能从中提取到重要的特征量,有助于后期声纹特征的学习。

声纹语谱图的绘制过程包括分帧、加窗和离散傅里叶变换,再计算出每一帧的功率谱,用颜色的深浅表示能量的大小,从而绘制出语谱图<sup>[7]</sup>。取一段变压器运行时采集的 4 s 噪声片段,对截取的片段作分帧处理,分帧长度的选择会影响能否提取到有效的特征量以及特征量的准确性。考虑到变压器噪声较为稳定,因此选择每帧的长度  $N=4\ 096$  为 64 ms(采样频率为 64 kHz),重叠率取 50%。其次,将分帧好的每一帧数据作加窗处理,为减少信号失真现象,本文选择使用汉明窗,表达式如下:

$$\bar{\omega}(n) = \begin{cases} 0.54 - 0.46\cos[2\pi n/(N-1)], & 0 \leq n \leq N-1 \\ 0, & \text{其他} \end{cases} \quad (1)$$

式中: $N$  为批次大小(batchsize); $n$  为类别数。

再对分帧加窗后的每一帧数据作离散傅里叶变换,频矩阵通过下式获得:

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi}{N}kn}, \quad k = 0, 1, 2, \dots, N-1 \quad (2)$$

最后用式(2)计算的功率谱绘制语谱图,此时横坐标表示分帧后的帧数(时间),纵坐标表示频率,其中的颜色深浅表示该时刻该频率下的大小,该大小为功率谱密度。通过短时傅里叶变换生成的语谱图如图 1 所示。

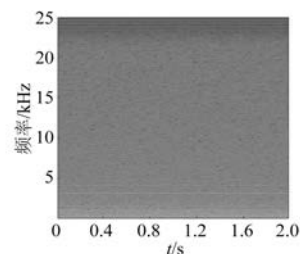


图 1 语谱图

### 1.2 Mel-语谱图降维

通过上述步骤能得到描述变压器运行过程中时域与频域之间的关系,但是生成的语谱图数据尺寸达到 465×2 049(465 代表时间尺度,2 049 代

表频率维度),这样庞大的数据对后续神经网络的训练与识别速度有很大的影响,因此需要对STFT生成的数据进一步压缩。Mel滤波器是根据人耳结构特征,对线性的频段做非线性处理,将低频部分的权重加大,高频部分的权重降低,突出有效频段的作用,因此选择Mel滤波器对 $465 \times 2\ 049$ 的矩阵进行压缩。

Mel频率与实际频率的转化关系为<sup>[8]</sup>

$$f_{\text{Mel}}(f) = 2\ 595 \cdot \log\left(1 + \frac{f}{700}\right) \quad (3)$$

式中: $f_{\text{Mel}}(f)$ 为Mel刻度下的频率; $f$ 为实际频率。

Mel滤波器为一个由 $M$ 个三角形滤波器组成的滤波器组(取 $M=40$ )。其中心频率为 $f(m)$ ,在Mel频率刻度上,滤波器之间的距离是等宽的。该滤波器组的传递函数为<sup>[8]</sup>

$$H_m(k) = \begin{cases} 0, & k < f(m-1) \\ \frac{k-f(m-1)}{f(m)-f(m-1)}, & f(m-1) \leq k \leq f(m) \\ \frac{f(m+1)-k}{f(m+1)-f(m)}, & f(m) < k \leq f(m+1) \\ 0, & k > f(m+1) \end{cases} \quad (4)$$

$f(m)$ 定义为<sup>[8]</sup>

$$f(m) = \left(\frac{N}{f_s}\right) F_{\text{Mel}}^{-1}\left[F_{\text{Mel}}(f_l) + m \frac{F_{\text{Mel}}(f_h) - F_{\text{Mel}}(f_l)}{M+1}\right] \quad (5)$$

式中: $f_h$ 与 $f_l$ 为滤波器滤波频率的上限与下限; $f_s$ 为变压器声纹采样的采样频率, $f_s = 64\ \text{kHz}$ ;  $N$ 为进行STFT时的帧长。

通过使用Mel滤波器对数据进行降维后,时频矩阵从 $465 \times 2\ 049$ 被压缩至 $465 \times 40$ ,大大缩减了数据的尺寸,简化了后续模型训练与识别,生成的Mel-语谱图如图2所示。

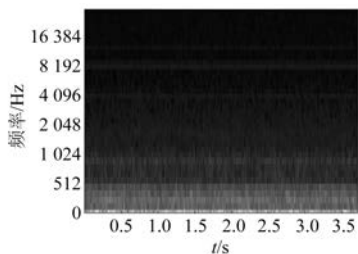


图2 Mel-语谱图

## 2 CNN的损失函数

SE-ResNet-Dense使用的损失函数是softmax损失函数,softmax loss是由softmax和cross entropy loss组合而成的损失函数,其损失函数的表达式为

$$L_1 = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{\mathbf{w}_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^n e^{\mathbf{w}_j^T x_i + b_j}} \quad (6)$$

式中: $\mathbf{w}_{y_i}$ 、 $b_{y_i}$ 均为符号函数,如果样本 $i$ 真实,则 $y = 1$ ,否则为0; $x_i$ 为属于第 $i$ 个样本的嵌入特征, $x_i \in R_d$ ,特征尺寸 $d$ 设置为512; $\mathbf{w}_j$ 为权重矩阵的第 $j$ 列, $\mathbf{w}_j \in R^{d \times n}$ ;  $b_j$ 为偏置项。

式(6)中,log是计算时间复杂度,算法中log级别的时间复杂度均是由于使用了分治思想,这个底数直接由分治的复杂度决定。如果采用二分法,那么就会以2为底数,三分法就会以3为底数,其他亦然。

SE-ResNet-ArcLoss使用的损失函数是additive angular margin loss<sup>[10]</sup>,是在softmax loss的基础上改进而来的。很多研究通过改变损失函数中的参数值,使其特征量在类内能做到紧凑且在类间能做到分离的效果。令偏差项 $b_j = 0$ , $\bar{\mathbf{w}}_{y_i}^T x_i = \|\bar{\mathbf{w}}_{y_i}\| \|x_i\| \cos \theta_{y_i}$ , $\theta_j$ 表示权重 $\bar{\mathbf{w}}_j$ 和特征 $x_i$ 之间的角度,并通过 $L_2$ 正则化,令 $\|\bar{\mathbf{w}}_{y_i}\| = 1$ ,且将 $\|x_i\|$ 缩放到 $s$ ,使得学习过程中得到的嵌入特征分布在半径为 $s$ 的超球面上,此时损失函数表达式为

$$L_2 = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s \cos \theta_{y_i}}}{e^{s \cos \theta_{y_i}} + \sum_{j=1, j \neq y_i}^n e^{s \cos \theta_j}} \quad (7)$$

而additive angular margin loss是在损失函数 $L_2$ 的基础上,在 $x_i$ 与 $\bar{\mathbf{w}}_{y_i}$ 之间的 $\theta$ 角上添加一个角裕度 $m$ ,最终得到的损失函数表达式为

$$L_3 = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s \cos(\theta_{y_i} + m)}}{e^{s \cos(\theta_{y_i} + m)} + \sum_{j=1, j \neq y_i}^n e^{s \cos \theta_j}} \quad (8)$$

## 3 ResNet网络结构

由He等<sup>[11]</sup>提出的基于残差结构卷积神经网络——ResNet,极大地缓解了传统的卷积神经网络

络和全连接网络在训练过程由于网络过深而出现的  
信息丢失、梯度消失和爆炸等问题<sup>[9]</sup>,参数量比  
超分辨率测试序列(VGG)网络减少且准确率更  
高。在网络结构中添加多个由2个(3×3)卷积层  
串接在一起的残差块,整个网络只需要学习输入、  
输出差别的部分,简化了学习目标和难度。

本文构造的两个模型 SE-ResNet-Dense 与  
SE-ResNet-ArcLoss 中选取的卷积结构是  
ResNet34,并在网络中引入注意力机制,再用不同  
的分类器进行分类,两种模型的结构如下表1所  
示。SE-ResNet 网络结构主要由16个 ResNet-  
block 组成,ResNet-block 由2个 kernel 为 3×3、  
stride 为 1×1 的 conv 层和1个 SELayer 构成,通道  
数分别为 64、128、256、512 的 ResNet-block 个数  
分别设置为 3、4、6、3,然后经过 max pool 将特征  
转换为 512 维。

表1 模型的网络结构

名称	SE-ResNet-Dense	SE-ResNet-ArcLoss	输出
输入	(1, 40,465)		—
Conv	(3×3)@64, stride=1		(64, 38, 463)
maxpool	3×3, stride=3		(64, 1, 1)
SElayer 1	$\begin{bmatrix} \text{conv}(3 \times 3) @ 64 \\ \text{conv}(3 \times 3) @ 64 \\ \text{SElayer} \end{bmatrix}$	@ 3	(64, 38, 463)
SElayer 2	$\begin{bmatrix} \text{conv}(3 \times 3) @ 128 \\ \text{conv}(3 \times 3) @ 128 \\ \text{SElayer} \end{bmatrix}$	@ 4	(128, 19, 232)
SElayer 3	$\begin{bmatrix} \text{conv}(3 \times 3) @ 256 \\ \text{conv}(3 \times 3) @ 256 \\ \text{SElayer} \end{bmatrix}$	@ 6	(256, 10, 116)
SElayer 4	$\begin{bmatrix} \text{conv}(3 \times 3) @ 512 \\ \text{conv}(3 \times 3) @ 512 \\ \text{SElayer} \end{bmatrix}$	@ 3	(512, 5, 58)
maxpool	512d		(512, 1, 1)
分类器	Dense(512×256) Dense(256×3)	ArcLoss(512×3)	—
参数量	19,528,689	19,392,221	—

将采集到的4s变压器噪声语音预处理后生  
成的Mel-语谱图时频矩阵化为单通道数据,将矩  
阵尺寸为40×465的数据送入到网络中,ResNet34  
的每个残差块是由2个kernel为3×3, stride=1的

卷积层加注意力机制组成,且通道数逐渐增加,并  
最终用最大池化层将维度转化为512。

而这两种模型的区别在于使用的分类器与损  
失函数的不同,SE-ResNet-Dense 使用的是两个全  
连接层,不断地将数据的维度降至类别数。本文  
是研究变压器的铁心松动故障,采集了铁心3种  
状态的数据,因此最后一层的神经元个数为3,用  
softmax loss 函数来规范模型的权重;SE-ResNet-  
ArcLoss 使用损失函数是 additive angular margin  
loss,将特征向量和权重归一化后,求出其夹角 $\theta$ ,  
计算出 $\cos\theta_j$ ,再通过反余弦 $\arccos(\cos\theta_j)$ 值求出  
 $\theta_{j_j}$ ,即为特征 $x_i$ 与真实权重 $\bar{\omega}_{j_j}$ 之间的夹角,再加  
上角裕度后送入 softmax loss 函数中以此来修正  
权重参数,用 ArcLoss 输出的数据作为分类依据。

## 4 试验及训练结果

### 4.1 数据集

为了验证 Mel 语谱图-ResNet 网络的有效性,  
搭建了变压器铁心松动故障试验平台,采集铁心  
在不同松动程度下的噪声信号。

试验对象为一台 S13-M-200/10 变压器,根据  
国标 GB/T 1094.10—2003 对该变压器噪声测量的  
标准和要求,采用电容式麦克风作为声传感器  
对变压器铁心不同松动情况下的噪声进行测量,  
采样频率为 64 kHz,频率响应为 20 Hz~20 kHz。  
试验选择在比较空旷的变压器厂房进行,避免了  
墙面对声波的反射与折射,如图3所示。

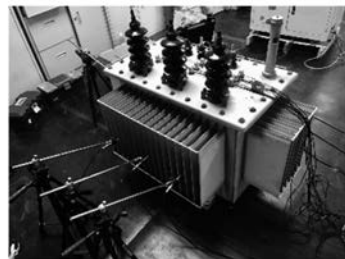


图3 试验环境

将变压器油抽出吊芯,使用扭力扳手先确定  
铁心夹件的额定预紧力,再通过设置不同的预紧  
力来表征变压器铁心的不同松动程度,模拟铁心  
松动故障。铁心压紧程度通过改变螺栓预紧力进  
行模拟如图4所示,分别调整为额定预紧力  
( $F_N$ )、松动 20% ( $0.8 F_N$ )、松动 40% ( $0.6 F_N$ )。

在低压侧加入 400 V 电压,对变压器进行空载运行试验,如图 5 所示,分别测量得出不同状态下的噪声信号长度为 4 s。



图 4 设置铁心预紧力

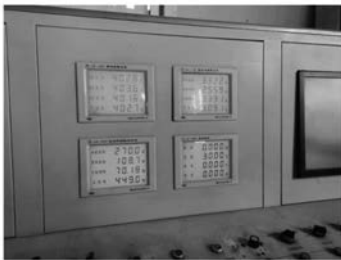


图 5 变压器空载运行

分别采集铁心在额定预紧力下的样本 289 个,松动 20% 时的样本 156 个,松动 40% 时的样本 323 个。其中,将 80% 的数据作为训练样本,剩余 20% 作为测试样本,验证模型的有效性。

#### 4.2 模型训练参数

设置每个模型训练的 epoch=50(训练结果显示在 50 次后模型效果趋向稳定),每训练一个 epoch 时的批处理量 batch\_size=16,优化器选择更适合分类问题的随机梯度下降(SGD),设置其动量为 0.9,权重衰减值设置为  $5 \times 10^{-4}$ ,学习率的设置采用指数衰减法,初始学习率设置为  $10^{-3}$ ,衰减系数为 0.1,即每 10 个 epoch 衰减一次。

#### 4.3 性能指标

给每一段音频生成的语谱图标记,分别送入两个模型中训练,用损失值 loss 和准确率 acc 来评判模型的优劣。损失值 loss 用来描述预测值与实际值之间的差距,acc 表示正确分类的数量占总预测样本的比例。loss 值越小,acc 值越大,则判定该模型越适合变压器铁心松动故障的识别。

#### 4.4 模型测试结果

图 6 所示是这两个模型在训练集上训练的准

确率 acc 随 epoch 变化的曲线图。两个模型从整体来看,均是随着 epoch 的增加准确率逐渐增加。SE-ResNet-Dense 网络结构在训练集上的准确率波动很大,显示出该卷积神经网络的不稳定性,随着 epoch 的增大,准确率在 0.1~1 之间大幅波动。而在卷积神经网络 SE-ResNet-ArcLoss 上的训练效果就相对更稳定,SE-ResNet-ArcLoss 比 SE-ResNet-Dense 稳定性稍好一些,在 epoch=18 以后,准确率在 0.75~1 之间浮动;在 epoch=40 以后,准确率在 0.95~1 之间浮动。

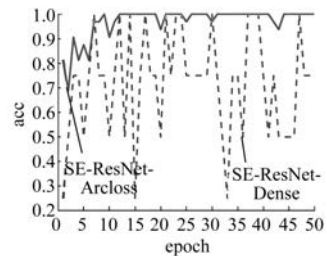


图 6 模型的 acc 曲线

图 7 所示是两个模型在训练集上训练的损失值 loss 随 epoch 变化的曲线图。从图 7 中可以清晰地看出,SE-ResNet-Dense 模型的 loss 值没有随着 epoch 的增加呈现逐渐下降的趋势,说明在训练过程中出现了不同程度的过拟合,但是对于 SE-ResNet-ArcLoss 模型,在 epoch=25 以后,loss 值虽然有波动但已经稳定在 0.5 以下。

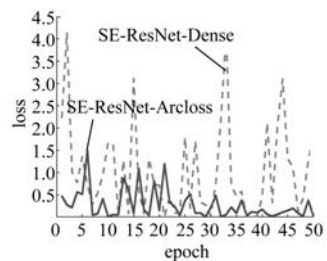


图 7 模型的 loss 值曲线

在测试集上测试的结果显示,SE-ResNet-ArcLoss 的测试 acc 为 97.541%,而 SE-ResNet-Dense 的测试 acc 为 90.753%。综合而言,根据比较准确率 acc、loss 值及参数量,最终找到最适合变压器铁心松动故障识别的模型是 SE-ResNet-ArcLoss 网络。

批尺寸是影响网络最终训练结果的一个重要因素,是指一次送入网络训练的数据量,因此对

SE-ResNet-ArcLoss 网络训练的不同 batchsize 进行测试,测试结果如图 8 所示。从图 8 中可以看出,当 batchsize = 10、20、30 时,训练的准确率均出现了大小不一的波动,而当 batchsize = 60 时,随着 epoch = 40 以后,训练准确率达到 1 并不再变化,因此该网络在训练过程中应将 batchsize 设置为 60。

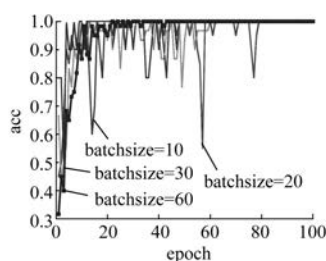


图 8 SE-ResNet-ArcLoss 模型不同 batchsize 训练结果

## 5 结 语

本文提出了一个基于变压器声纹的 Mel 语谱图-ResNet 卷积神经网络的铁心松动故障诊断模型。通过搭建模拟铁心松动故障试验平台,采集不同松动状态下变压器空载运行的噪声信号;利用 Mel 滤波器对离散傅里叶变换下的时频矩阵降维绘制语谱图制作成数据集,输入 ResNet 网络中,比较在 softmax 损失函数和 additive angular margin loss 损失函数上的表现效果,通过准确率和 loss 值比较,计算结果是 SE-ResNet-Dense 模型的准确率为 90.753%,而 SE-ResNet-ArcLoss 模型的准确率可达到 97.541%,同时计算了当批尺寸为 60 时的训练效果最好,因此最适合变压器铁心松动故障的声纹识别模型是 SE-ResNet-ArcLoss。

## 【参 考 文 献】

- [ 1 ] 张建功,王延召,陈豫朝,等. 基于振动和声压测量的特高压变压器声功率估算方法[J]. 高电压技术,2019,45(6): 1843.
- [ 2 ] 王春宁,耿志慧,马宏忠,等. 基于振动的电力变压器铁心松动故障诊断研究[J]. 高压电器,2015,51(12): 59.
- [ 3 ] 王丰华,王邵菁,陈颂,等. 基于改进 MFCC 和 VQ 的变压器声纹识别模型[J]. 中国电机工程学报,2017,37(5): 1535.
- [ 4 ] 舒畅,金潇,李自品,等. 基于 CEEMDAN 的配电变压器放电故障噪声诊断方法[J]. 高电压技术,2018,44(8): 2603.
- [ 5 ] 刘云鹏,罗世豪,王博闻,等. 基于 Mel 时频谱-卷积神经网络的变压器铁芯夹件松动故障声纹模式识别[J]. 华北电力大学学报(自然科学版),2020,47(6): 52.
- [ 6 ] 周欣,李洋锐,马国峰. 声谱图生成方法及实现[J]. 数字技术与应用,2021,39(12): 101.
- [ 7 ] 周梦茜. 电力变压器局部放电超声信号的声纹识别方法研究[D]. 北京:华北电力大学(北京),2021.
- [ 8 ] 徐明月,李喆,孙汉文,等. 基于改进梅尔倒谱系数的 GIS 机械故障诊断方法[J]. 高压电器,2020,56(9): 122.
- [ 9 ] 高盎然,朱永利,张翼,等. 基于边际谱图像和深度残差网络的变压器局部放电模式识别[J]. 电网技术,2021,45(6): 2433.
- [ 10 ] JAMSHAI D U L R. A study on angular softmax[D]. 合肥:中国科学技术大学,2020.
- [ 11 ] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.