

基于音频去干扰自学习神经网络的地下线缆定位系统研究

文/吴清涛¹ 傅晓杰¹ 张晨¹ 白蕾¹ 龚侃² 刘亦颖²

1. 国网上海市电力公司市北供电公司; 2. 上海久隆电力(集团)有限公司

城市地下线缆是城市电力、通信、网络和其他基础设施系统的核心组成部分, 其管理和维护逐渐成为保证城市正常运行的关键。为便捷准确地对地下线缆的故障点进行定位, 提出基于音频去干扰预处理算法和一种引入残差模块的深度神经网络的地下线缆定位方法。首先通过预加重、分帧和加窗等预处理步骤将采集到的原始音频数据转化为规则化的时域音频片段, 然后快速傅里叶变换将其转化到频域, 并提取音频信号的梅尔倒谱系数特征, 最后将音频片段以频谱形式输入进基于VGG网络改进的深度学习神经网络, 从而构建故障音频识别模型。设计不同神经网络模型对故障音频数据集的识别实验, 结果显示改进的VGG网络结合MFCC混合特征向量进行故障音频识别, 能进一步提高识别准确率。

随着城市的快速发展, 地下线缆在现代社会中扮演着至关重要的角色, 是电力、通信、网络和其他基础设施系统的核心组成部分^[1]。尽管地下线缆在各个领域发挥着不可替代的作用, 但其管理和维护也面临着一系列挑战。地下电缆的损坏、故障和定位成为关键问题。因此, 研究和改进地下电缆定位系统至关重要, 以确保电力、通信和城市基础设施的可靠性和稳定性。

现有的地下线缆检修办法一般有: 人工检测、时域反射法、热成像检测、声波检测和高频电流检测等。马晓等人提出一种基于频域反射法的新型高压电缆外破故障定位方法, 对一段30 m长的110 kV电缆上制作铁钉扎入的缺陷, 并进行定位, 验证了定位的准确性^[2]。徐子弘等人提出基于数据驱动的电缆早期故障检测方法, 采用高斯噪声模型来检验所提出的故障检测算法对于实际状态下不确定因素的干扰^[3]。

近年来, 随着神经网络的应用越来越广泛, 越来越多的学者在音频识别领域应用卷积神经网络

(CNN)对音频进行分类和识别。李锋等人提出一种基于车辆音频信号和改进VGG深度学习网络的道路交通车型识别方法, 利用VGG-11网络模型将识别率提高到95%以上。^[4]李伟等人提出加权动态梅尔频率倒谱系数作为声纹识别的特征参数, 利用其静态特性逼近音频的动态特性, 通过实验证明, 在提高识别率的情况下, 大大降低系统的运行时间, 以此实现地下电缆防护。

本文引入音频信号和自学习神经网络进行地下线缆的故障定位。首先通过预加重、分帧和加窗等预处理步骤将采集到的原始音频数据转化为规则化的时域音频片段, 然后快速傅里叶变换将其转化到频域得到梅尔频率图, 将音频片段以频谱形式输入进残差深度学习网络, 经过卷积池化处理在全连接层形成二维矩阵特征集, 同时提取音频信号的梅尔倒谱系数(MFCC)特征向量, 通过多头注意力机制将特征向量与全连接后的神经网络特征集拼接, 最后通过一个softmax函数获得音频类别分类结果, 从而构建地下线缆故障音频识别模型, 其具体流程如

图1所示。

本文的主要研究内容为：①根据采集到的故障音频数据集，进行音频信号预处理；②利用快速傅里叶变换将音频信号转化到频域，对不同场景下的

电缆故障信号特征识别和降噪并提取MFCC混合特征向量；③基于VGG-11改进设计残差神经网络，优化神经网络结构，构建故障音频数据集，并对不同网络模型的故障音频识别准确率。

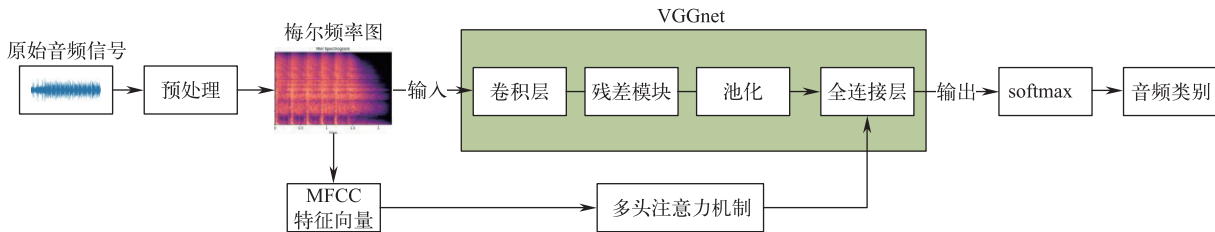


图1 音频去干扰识别整体结构示意图

音频信号的预处理

在时间域中，音频信号的表现形式通常为随时间演变的振幅波形，这种方式直观地揭示了声音的波动和强度的变化。然而，频率变化的规律在时域信号中并不容易直观地观察到。为了有效识别音频信号，需要对信号做预处理以得到短时平稳的音频信号。在自然环境中采集得到的音频信号易出现特征不明、混杂噪声的现象。为了最大限度提高神经网络分辨不同音频信号的性能，应当在制作音频信号训练集之前保证音频信号样本具有持续时长相等、有限振幅等特征，而一般的音频预处理步骤为：预加重、分帧和加窗。

1. 预加重

预加重是音频信号处理中的一个重要步骤，特别是在语音信号处理中。预加重的主要目的是平衡语音信号的频谱，增强高频部分的能量，从而改善后续处理步骤，如分帧、窗函数和傅里叶变换等^[6]。预加重通常通过一个一阶滤波器实现，该滤波器的传递函数为

$$H(z) = 1 - \mu z^{-1} \quad (1)$$

式中， z 为复数频率变量； μ 为预加重因子，通常取值在0.9~1.0之间，本文取0.95。

在时间域，这个滤波器对应于以下的差分方程

$$y[n] = x[n] - \mu x[n-1] \quad (2)$$

式中， $x[n]$ 为输入音频信号； $y[n]$ 为输出音频信号； n 为离散时间指标。

预加重的过程就是将这个滤波器应用于原始的音频信号。这可以通过卷积运算或差分运算实现。预加重后的信号有更平衡的频谱，高频部分的能量

被增强，这可以改善声谱分析的结果，提高语音识别等任务的性能。

2. 分帧

地下电缆的故障声音的持续时间可能会因故障的类型和严重程度而变化。例如，持续的电弧故障可能会产生持续的噪声，而瞬态故障可能只产生短暂的噪声。但故障检测时获得的音频通常为长段音频，其一般是非平稳的，所以通常会将长音频信号分割成多个短时间的帧，假设每一帧内的信号是平稳的。

首先确定帧长和帧移，帧长（Frame Length）决定了每一帧包含的样本数。帧长通常选择在20~30 ms，以确保每一帧内的信号可以看作是平稳的，如图2所示。帧移（Frame Shift）或帧跳（Frame Hop）决定了相邻帧之间的重叠程度。帧移通常小于帧长，这样可以使得相邻的帧有一定的重叠，从而避免帧边界处的信息丢失。然后分割音频信号，从音频信号的开始位置起，按照帧长和帧移的设定，将音频信号分割成多个帧。每一帧包含了帧长所对应的样本数，相邻帧之间的起始点间隔帧移所对应的样本数。之后再对每一帧进行下一步的处理。

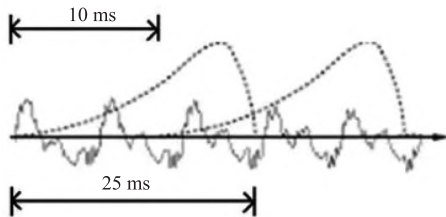


图2 时域信号的帧长与帧移

3. 加窗

加窗在分帧之后进行，其目的是减小帧边界处的不连续性，从而降低在连续的频谱分析中出现频谱泄漏和假频的概率。设 $s(n)$ 为音频信号， $w(n)$ 为窗函数，则加窗运算可表示为

$$S_w(n) = s(n)w(n) \quad (3)$$

常见的窗函数有矩形窗、汉明窗和汉宁窗，三种窗函数中，矩形窗由于帧边界的不连续性，使它在频域中会产生较大的旁瓣而导致频谱泄漏。汉明窗与汉宁窗类似，都在帧的开始和结束处平滑地降到0，但汉明窗并不完全降到0，而是到达一个小的非零值。汉明窗的主瓣宽度和旁瓣级别介于矩形窗和汉宁窗之间。考虑信号的特性，对于地下电缆故障音频信号的检测，需要能准确地检测到故障音频的频率，因此需要选择旁瓣级别较低的窗函数，以减小频谱泄漏。同时，由于故障音频可能突然出现或消失，需要窗函数的主瓣宽度较窄，以提高频率分辨率。综上，选择旁瓣级别较低的汉宁窗作为窗函数，其表达式为^[7]

$$w(n) = \begin{cases} 0.5 \left[1 - \cos\left(\frac{2\pi}{N-1}\right) \right] & 0 \leq n \leq N-1 \\ 0, n = \text{其他值} \end{cases} \quad (4)$$

加窗后的音频帧可以进一步进行傅里叶变换、频谱分析等处理。由于加窗减少了帧边界处的不连续性，可以改善频谱分析的结果，使得频谱更加接近于真实的频率分布。

音频信号的特征提取

音频信号经过预处理后可采用快速傅里叶变换（FFT）方法来进行时频转换。对每一帧应用FFT，将音频信号从时域转换到频域。这可以得到每一帧的频谱，显示出不同频率的能量分布。

1. 快速傅里叶变换

在音频信号处理中，快速傅里叶变换是一个非常重要的工具，它是傅里叶变换的一种高效算法，它利用了信号的对称性和周期性，将傅里叶变换的计算复杂度从 $O(N^2)$ 降低到 $O(N \log N)$ ，其中 N 是信号的长度。这对于处理长音频信号或实时音频处理非常重要。许多音频特征，如梅尔频率倒谱系数（MFCCs）、谱子带能量等，都需要在频域中计

算。FFT提供了一种高效的方法来获取这些特征。

2. 梅尔频率倒谱系数提取

梅尔频谱图是音频处理中常用的一种特征表示方法，它模拟了人耳的听觉感知特性^[8]。通过FFT得到音频信号的频谱可表示为

$$X(l, k) = \sum_{n=0}^{N-1} x(n)w(n-l)e^{-j\frac{2\pi kn}{N}} \quad (5)$$

式中， $X(l, k)$ 为一个关于帧平移量 l 和当前谱线数 k 的二维函数； N 和 n 分别总采样点和当前为第 n 帧。

则其功率谱可表示为

$$S(l, k) = |X(l, k)|^2 \quad (6)$$

将频谱通过梅尔滤波器组，得到每一帧的梅尔滤波器组能量。梅尔滤波器组是一组三角形的滤波器，它们的中心频率被放置在梅尔刻度下均匀分布的位置，对于每一帧音频信号均存在一个对数梅尔能量函数 $E_{mel}(m)$ ，将功率谱与梅尔滤波器组计算得到的对数梅尔能量表达式为^[8]

$$E_{mel}(m) = \ln \left[\sum_{k=1}^N S(l, k) H_m(k) \right], 0 \leq m \leq M \quad (7)$$

经过梅尔滤波后的对数梅尔能力进行离散余弦变换，具体求解表达式为

$$E_{mfcc}(n) = \sum_{m=1}^M E_{mel}(m) \times \cos \left[\frac{\pi n(m-0.5)}{M} \right] \quad (8)$$

式中， $E_{mfcc}(n)$ 函数中 n 表示当前第 n 帧； m 取值满足每组滤波器个数限制。

创建好对应的数组标签后，可以得到MFCC静态特征。利用MFCC静态特征，计算其一阶差分，计算公式为

$$x_i^d = \frac{\sum_{j=1}^v j [E_{mfcc}(i+j) - E_{mfcc}(i-j)]}{2 \sum_{j=1}^v j^2} \quad (9)$$

式中， $v=2$ ，对获得的一阶差分取与静态特征相同的维度，拼接得到二维的MFCC混合特征向量 $V \in R^{S \times F}$ ； S 为音频信号的总帧数； F 选取20为提取的特征量。

自学习卷积神经网络构建

基于VGG-11网络结构，加入深度残差模块和注意力模块对故障点发出的有效音频信号进行识别。深度残差收缩网络是一种面向振动信号的卷积神经网络，可以适用于数据含噪情况，其基本结构由输入层、卷积层、残差模块、注意力模块、全连

接层和输出层组成。

1. VGG网络基本结构

VGG是牛津大学的Visual Geometry课题组提出的深度卷积神经网络模型。该网络证明了增加网络的深度能够在一定程度上影响网络最终的性能。VGG有多种结构,根据层数可分为11~19层的卷积神经网络,可通过加深网络结构提升其性能^[9]。

本文以VGG-11为基础构建卷积神经网络,VGG-11由8个卷积层和3个全连接层组成。输入矩阵经过64个 3×3 的卷积核卷积,并通过一个最大池化层降维完成一个完整的卷积处理,之后依次经过128个卷积核卷积1次,256个卷积核卷积2次,512个卷积核卷积4次,每次卷积完成均采用最大池化层降维。最后经过3个全连接层实现分辨,过程中所用卷积核大小均为 3×3 ,步长为1,最大池化层步长为1,池化窗口为 2×2 。

2. VGG网络改进结构

为适应地下电缆的故障音频识别,本文在VGG-11网络模型的基础上进行了改进,在卷积层

与池化层之间增加了残差模块。并在全连接层之后引入注意力机制,将预处理时提取的MFCC特征向量通过注意力机制导入神经网络,使最终的分辨结果更加准确。

改进的网络结构由8个卷积层、3个全连接层和1个特征拼接层组成,如图3所示。由于音频数据的时长较短和频率范围较窄,可以适当降低网络结构复杂度以提高训练效率,因此输入矩阵首先经过16个卷积核进行卷积两次,之后一次经过32、64、128个卷积核卷积,每个卷积处理后接一个残差层和最大池化层。之后经过3个全连接层后得到大小为 $N \times L_1$ 的二维矩阵特征集,其中 N 为待分辨的故障类型, L_1 为卷积神经网络输出的特征向量维度。同时将预处理时得到的MFCC以及对应的差分特征组合为梅尔特征集,通过多头注意力模块同步处理特征向量,并通过一个多层感知机和线性层输出得到 $N \times L_2$ 的二维矩阵特征集,将两特征集通过拼接层拼接,最后输入softmax函数得出分辨的故障音频类型。

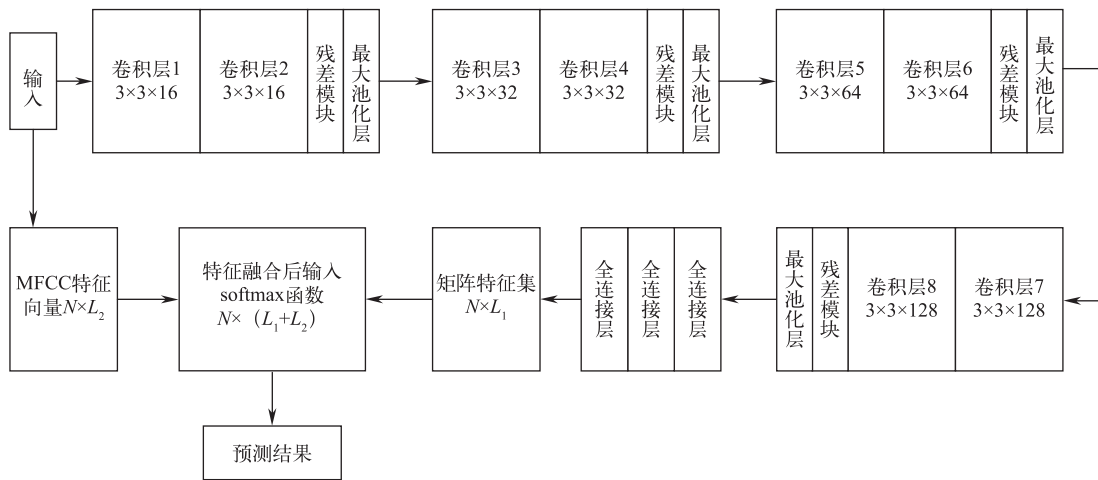


图3 改进的残差神经网络结构示意图

3. VGG网络结构的改进策略

为了解决深度学习网络容易出现过拟合的问题,对VGG网络的卷积层和卷积核进行结构调整。原始的VGG网络模型采用更小的卷积核而耗费更多计算资源,并且使用更多的参数。为降低参数量,网络卷积核的数量由原始的64、128、256、512改为16、32、64、128。由此改进后的网络在结构上仍采用小卷积核代替传统神经网络的大卷积核,在减少参数量的同时保持了模型提取特征的能力。

在原始的VGG网络的卷积层中引入了残差模块。残差模块来源于深度残差收缩网络,该网络是一种面向振动信号的卷积神经网络,可以适用于数据含噪情况。传统网络随着网络深度的增加,训练变得愈加困难。这主要是因为基于随机梯度下降的网络训练过程中误差信号的多层反向传播易产生梯度消失的现象,目前一些特殊的权重初始化策略和批规范化等方法使这个问题得到了改善,但当模型收敛时训练误差没有降低反而升高。尤其是图像

处理方面，往往使用非常多的卷积层、池化层等。每一层都是从前一层提取特征，所以随着层数增加一般会出现退化等问题。残差网络采取跳跃连接的方法避免了深层神经网络带来的一系列问题^[10]。在4组卷积层和最大池化层之间加入残差模块，采取更好的拟合分类函数以获得更高的分类准确度。

在故障音频的识别中，不仅可以通过卷积神经网络提取语谱图的特征信息，还能通过预处理得到故障音频的MFCC以及对应的差分特征，利用多头注意力机制将神经网络提取到的音频特征向量与MFCC混合特征向量进行拼接，可以更加准确地获取故障音频的具体时间段并分辨故障类型。从预处理后的音频信号中得到的MFCC混合特征向量，利用多头注意力（MHA）机制寻找音频信号中显著的故障音频段，采用的注意力机制可以表示为^[11]

$$\text{Att}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (10)$$

通过多头注意力机制得到的输出矩阵送入MLP为主干的前馈网络，其表达式为

$$F_{\text{MLP}}(x) = [\text{ReLU}(xW_1 + b_1)]W_2 + b_2 \quad (11)$$

式中， W_1 、 W_2 为权重矩阵； b_1 、 b_2 为偏置向量；MLP主要由两个线性层和一个ReLU激活函数嵌套组成，最终通过一个全连接层得到二维特征向量 $N \times L_2$ ，为待分辨的故障类型， L_2 为多头注意力机制输出的特征向量维度。

实验设计与结果

1. 实验环境搭建与参数设置

本文实验采用的编程语言为Python3.8。为了验证本文所提出的模型的有效性，在进行神经网络模型训练时，输入的训练集将不仅局限于采集到的故障音频，而是采用AudioSet音频数据集作为训练集，故障音频作为测试集进行实验。卷积神经网络模型在Pytorch框架下训练，硬件CPU为i5-8300H，GPU为NVIDIA GeForce GTX 3060ti。

训练用数据集AudioSet包含从多个视频网站中提取出来的632类音频事件，共2 084 320个经由人工标注的长度为10 s的声音片段样本。数据集总时长约为5 800 h，是目前公开数据集中规模最大、种类最多的通常音频数据集。其标签都是clip-level片段级别的标签，该数据集的音频片段长度以及标签

化特征均与故障音频片段相符。

为比较本文改进的VGG网络相比于其他网络模型在地下线缆故障音频识别上的优势，在PyCharm2023.2平台上分别采用原始VGGNet、ResNet、GoogleNet和改进VGGNet四种网络进行音频识别实验。以改进VGGNet为例。

2. 网络训练与结果分析

实验中将地下线缆故障音频分为开路故障、低阻故障、高阻故障和闪络故障四种不同的故障类型，其故障音频片段的梅尔频谱图以及MFCC特征向量也将不同。在数据集中随机选取每种故障类型的音频各50条作为验证集。其余数据与AudioSet中的音频片段组合（共5 000条）按4:1的比例分别制成训练集和测试集，具体划分见表1。

表1 数据集划分

	故障音频片段	AudioSet音频	总计
训练集	800	3 200	4 000
测试集	200	800	1 000
验证集	50 × 4	0	200
总计	1 200	4 000	5 200

改进的VGG网络模型与其他网络模型经过训练后，对测试集的准确率见表2。对于不同的故障类型，各个网络的分辨能力有所不同，但以原始的VGGNet为例，其对于各类型的地下线缆故障音频的平均预测准确率为92.34%，而改进后的VGGNet对各类型的故障音频平均预测准确率为95.83%。在改进后的VGGNet结果输出后，若与原始故障音频信号提取得来的MFCC混合特征向量进行拼接判决，则可进一步提高模型对音频的识别准确率。而对于混杂在训练集中的AudioSet音频片段，及表2中所示非故障音频，各个卷积神经网络的识别准确率相差无几，表面改进后的VGG网络不仅适用于故障音频识别，还具有一定的泛用性。

表2 不同模型音频识别准确率

网络模型	开路故障	低阻故障	高阻故障	闪络故障	非故障音频
原始VGGNet	0.931 4	0.927 8	0.912 2	0.923 8	0.954 8
ResNet	0.872 3	0.896 5	0.884 1	0.865 1	0.938 1
GoogleNet	0.906 1	0.942 8	0.901 5	0.903 4	0.958 8
改进VGGNet	0.953 7	0.959 8	0.942 6	0.955 6	0.941 7
VGGNet+MFCC混合特征向量	0.967 2	0.964 1	0.958 9	0.966 8	0.948 5

对于本文提出的VGGNet+MFCC混合特征向量故障音频识别方法，可以从混淆矩阵中更加直观地观察到该模型对不同音频的识别情况，如图4所示。从图4中可以看出，在不同类别的故障音频之间仍有小概率出现识别混淆的情况，其中较为明显的是开路故障和闪络故障的混淆，而对于卷积神经网络模型，非故障音频与故障音频之间的区分则更加容易。

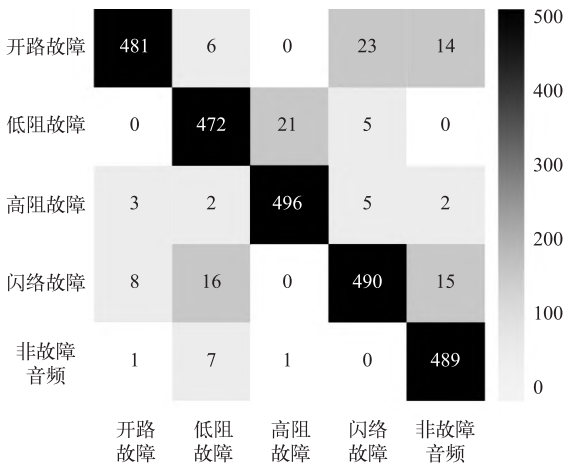


图4 故障音频识别混淆矩阵

结束语

本文从城市地下线缆的维护管理这一现实需求出发，研究基于卷积神经网络和音频特征提取结合的故障音频识别方法。基于VGG网络模型，针对故障音频数据集改进设计包含残差模块的网络结构，并比较不同模型对故障音频数据集的识别准确率。所用方法主要创新点如下：

- 1) 将残差模块引入VGG网络，在传统VGG网络的卷积层间加入残差层，使用残差网络跳跃连接的方法避免深层神经网络训练过程中的梯度消失问题，提高网络的训练效果，加速收敛速度，提高网络的表达能力。
- 2) 利用多头注意力机制将神经网络提取到的音频特征向量与MFCC混合特征向量拼接，可以更

加准确地获取故障音频的具体时间段并分辨故障类型，提高神经网络对故障音频的识别准确率。EA

(收稿日期：2024-06-25)

参考文献

- [1] 杨小军, 龙旭雲, 陈艳华. 浅谈加强城市地下电力电缆管理 [J]. 江西电力, 2022, 46 (6): 44-46, 53.
- [2] 马骁, 刘云龙, 余华兴, 等. 基于频域反射法的高压电缆外破故障定位研究[J]. 四川电力技术, 2020, 43(1): 22-25.
- [3] 徐子弘. 基于降噪自编码器和优化卷积神经网络的地下电缆早期故障检测和识别方法[D]. 广州: 华南理工大学, 2021.
- [4] 李锋, 叶霖, 徐伟, 等. 基于改进VGG深度学习网络的道路交通车型音频识别方法[J]. 广东技术师范大学学报, 2023, 44(3): 17-23.
- [5] 李伟, 曾繁洋, 王博, 等. 基于MFCC加权动态特征组合的声纹识别技术在地下电缆防护的应用[J]. 电力信息与通信技术, 2022, 20(5): 16-22.
- [6] 刘赣俊. 基于神经网络的音频场景识别算法研究[D]. 天津: 天津大学, 2019.
- [7] 刘祯. 基于图神经网络的音频场景识别研究[D]. 济南: 山东师范大学, 2023.
- [8] 王基豪, 周晓彦, 李大鹏, 等. 基于卷积神经网络和Transformer网络的鸟声识别[J]. 声学技术, 2023, 42(05): 675-683.
- [9] 徐科, 姚凌云, 姚静怡, 等. 基于改进VGG-16网络的交通声音事件分类方法研究[J]. 西南大学学报(自然科学版), 2023, 45(12): 145-156.
- [10] 周博言. 基于残差神经网络的代价敏感人脸识别研究[D]. 南京: 南京大学, 2019.
- [11] 顾天麒. 基于多头注意力机制的语音增强算法研究[D]. 南京: 南京邮电大学, 2022.