文章编号:2095-6134(2023)02-0208-09

基于 FVC-CNN 模型的野外车辆声信号分类*

李翔[1,2],王艳[1,2],李宝清[1†]

(1 中国科学院上海微系统与信息技术研究所 微系统技术重点实验室, 上海 201800; 2 中国科学院大学, 北京 100049)

(2020 年 12 月 15 日收稿; 2021 年 4 月 8 日收修改稿)

Li X, Wang Y, Li B Q. Field vehicle signal classification based on FVC-CNN[J]. Journal of University of Chinese Academy of Sciences, 2023, 40(2): 208-216. DOI: 10.7523/j.ucas.2021.0038.

摘 要 针对野外环境下单通道车辆声信号受风噪影响严重、分类性能较低的问题,提出一种基于声阵列 4 通道同步采集信号的一维卷积神经网络模型(FVC-CNN)。该模型借鉴注意力机制加权平均的思想对 Inception 网络结构进行改进,作为输入层有针对性地提取 4 通道声信号多个不同时间尺度的特征,抑制噪声干扰,再根据不同车辆声信号特征分布特点,分别训练 3 个特征提取网络 SWNet、LWNet 和 TNet 来提取相应车辆的特征,最后对提取的特征进行多分支多维度的融合以供分类。在相同数据集上进行验证,实验结果表明,FVC-CNN 模型总识别率可达 94.22%,相较于传统方法识别率提高 14.08%,取得了较好的分类效果。

关键词 野外车辆信号分类;4 通道声阵列输入;Inception 结构;注意力机制;多分支特征提取;多分支多维度特征融合

中图分类号:TN911.7;TP183 **文献标志码**:A **DOI**:10.7523/j.ucas.2021.0038

Field vehicle signal classification based on FVC-CNN

LI Xiang[1,2], WANG Yan[1,2], LI Baoqing[1]

(1 *Science and Technology on Microsystem Laboratory, Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai* 201800, *China*; 2 *University of Chinese Academy of Sciences, Beijing* 100049, *China*)

Abstract Aiming at the problem that single channel vehicle acoustic signal is seriously affected by wind noise and has low classification performance, a one-dimensional convolutional neural network model FVC-CNN (convolutional neural network for field vehicle classification, FVC-CNN) based on four channel synchronous acquisition signal of acoustic array is proposed in this paper. The model uses the idea of weighted average of attention mechanism to improve the structure of Inception network. As the input layer, it extracts the features of four channel acoustic signals with different time scales to suppress noise interference. According to the distribution characteristics of different vehicle acoustic signals, three feature extraction networks, SWNet, LWNet, and TNet, are trained to extract the characteristics of the corresponding vehicle, finally, the extracted features are fused with multi branches and multi dimensions for classification. Verified on the same data set, the experimental results show that the total recognition rate of FVC-CNN model can reach 94.22%, which is 14.08% higher than the traditional method, and the classification effect is better.

* 微系统技术重点实验室基金项目(6142804190304)资助

† 通信作者,E-mail:sinoiot@mail.sim.ac.cn

Keywords　field vehicle signal classification; four channel acoustic array input; Inception structure; attention mechanism; multi branch feature extraction; multi-branch and multi-dimensional feature fusion

野外车辆目标分类研究对战场态势感知、边防侦察、威胁估计和决策具有重要价值。采用视频图像识别车辆是近年来应用最广的车辆识别方法,但视频图像识别检测效果受环境、天气等因素的影响较大,故在野外布设成本较高且识别精度不高;而利用声音信号进行野外目标识别,具有全天候、抗电磁能力强、隐蔽性好等优点,能克服光学、无线电、雷达等现代侦察技术的盲区,且无需实时采集,大幅降低了信号的采集、存储和处理成本。

传统的野外车辆声信号识别主要是基于手工设计的特征提取,如:罗向龙等[1]利用经验模态分解法提取声音信号特征;Kandpal 等[2]对车辆声音信号的时域和频域特征进行分析,并将其送入神经网络分类;Yang 等[3]直接采用离散频谱分析方法提取车辆的声音信号特征,然后参考无线传感网络协议进行分类识别。但手工特征只能提取浅层信息,抽象能力不足,无法提取有区分度的特征对车辆目标进行描述。

近年来,随着深度学习技术的发展,越来越多的研究开始将深度神经网络(deep neural network, DNN)应用于声信号识别。对于声音信号,DNN 能够从原始数据中提取特征,一些基于 DNN 的模型被提出并且表现得比传统的机器学习模型效果更好。然而,DNN 的深度全连接架构对于转换特征并不具备强鲁棒性。一些新的研究发现卷积神经网络(convolutional neural network, CNN)具有强大的通过大量训练数据探索潜在的关联信息能力[4],通过从环境声音中学习类似频谱图的特征[5],将 CNN 应用于环境声分类的几次尝试已经获得了性能提升。

野外车辆识别主要面临如下挑战:1)野外环境下声音信号目标分类性能主要受风噪影响而明显下降;2)不同类型车辆的声音来源不一样,这意味着特征的抽象层次也不一样,需要分别有针对性地提取对应的特征。

针对以上问题,本文设计了一种一维卷积神经网络模型 FVC-CNN(convolutional neural network for field vehicle classification)用于野外车辆识别。FVC-CNN 具有以下特点:1)利用声阵列 4 通道同步采集信号中目标信号具有相关性,背景噪声信号不具有相关性的特性,采用 4 通道输入进行降噪处理;2)为充分利用原始信号中的信息,输入层就采用 Inception 网络结构提取不同尺度的特征,并引入注意力机制,根据不同的损失值(loss)去学习不同尺度特征的权重,使得有效的尺度特征权重大,无效或效果小的尺度特征权重小,以达到去噪和进一步提高特征利用率的效果;3)鉴于轮式车、履带车的不变性特征抽象程度不同,网络结构深度也不同,有针对性地分别设计了不同深度的 3 个分支网络:小型轮式车特征提取网络(neural network for small wheel feature extraction, SWNet)、大型轮式车特征提取网络(neural network for large wheel feature extraction, LWNet)和履带车特征提取网络(neural network for track wheel feature extraction, TNet),并对 3 个特征提取子网络单独设置辅助损失函数,强制学习各目标有区分度的不变性特征;4)对 3 个特征提取子网络中提取的不同车辆关键特征进行多分支多维度的融合,进一步提高特征的表达能力。

1　野外车辆分类卷积神经网络模型

1.1　FVC-CNN 模型结构

针对不同运动目标声阵列信号的特征表示层次和内在规律不同的情况,设计了一种适用于野外车辆声信号分类的一维卷积神经网络模型 FVC-CNN,其模型结构如图 1 所示,图中网络方块上的数字表示每层网络的输出通道数,N 表示最后送入分类器的信号特征类别数。

FVC-CNN 的输入采用 4 通道同步声阵列信号,送入改进的 Inception 网络结构采集不同尺度的特征进行加权平均后送入 3 段不同的特征提取子网络 SWNet、LWNet 和 TNet,然后用特征融合网络 MergeNet 对 SWNet、LWNet 和 TNet 中提取的特征进行多分支多维度的充分融合,在控制特征规模同时使特征的表达能力更强,最后利用 softmax 分类器进行分类。

SWNet 和 LWNet 是用于提取小型和大型轮式车特征而设计的 2 个特征提取子网络。它们由一个个卷积基本单元(即图 1 中有数字标注的网络模块)组成。而卷积基本单元又由一维卷积

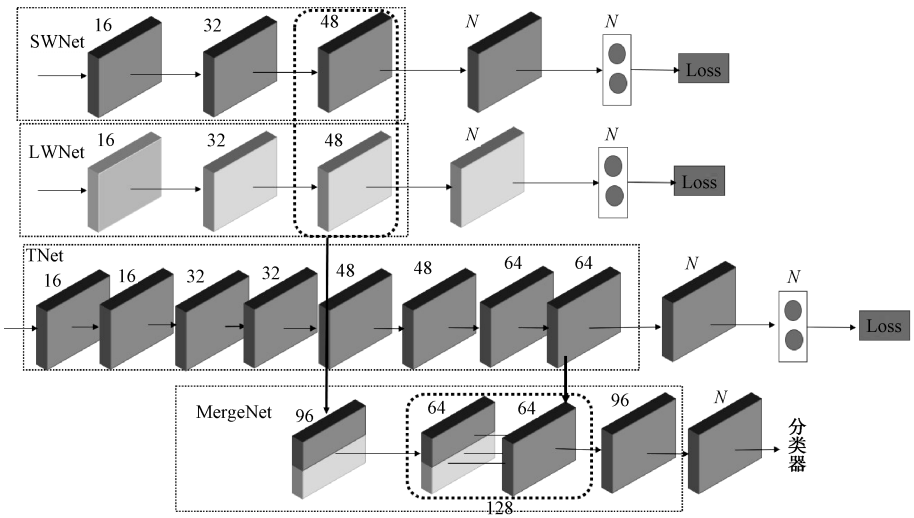


图 1 FVC-CNN 结构图

Fig. 1 Structure of FVC-CNN

层、tanh 激活函数层、批归一化(batch normalize, BN)层和最大值池化层 4 层结构组成。一维卷积层常用于音频信号提取特征,因为其可以很好地识别数据中的简单模式,然后将其用于在更高层中形成更复杂的模式[6];tanh 激活函数层进行特征映射,并在循环过程中不断扩大特征差异[7];BN 层对输出的结果进行归一化,达到加快模型的训练速度和防止过拟合的效果[8];而最后的最大池化层则用来去掉卷积得到的特征图中的冗余信息,并实现特征的降维。

TNet 是为提取履带车特征而设计的一个深层分支网络。运动中的履带车车辆,除发动机产生的声音信号之外,还包含一大部分由履带在运行过程中对地面的振动产生的声音信号以及由于履带运行过程中对地面反复的摩擦碰撞而引起的机械噪声构成的噪声[9]。履带车声音信号所含信息更加繁复,包含更多有区分度的信息,因此首先在之前的浅层网络之后添加 1 个卷积基本单元以加强网络对深层次特征的学习能力,其次在每个卷积基本单元的卷积层之后添加 1 层卷积核大小为 1 的卷积层,这样在本层特征送入下一层次的卷积基本单元前就对其进行进一步的抽象,实现了信息整合,有效地提高了特征的利用率。

MergeNet 首先将 SWNet 和 LWNet 的第 2 个卷积基本单元提取的特征图直接拼接在一块,然后采用一个卷积基本单元对拼接好的特征图进行进一步抽象,提取出新的特征图和 TNet 的最后一个基本单元提取的特征图进行特征叠加,最后再加 1 个卷积基本单元进行降维和信息融合,提取最后的特征送入 softmax 分类器进行分类。

1.2 声阵列 4 通道信号输入

相较于城市与室内环境,野外环境噪声成分比较单一,主要是自然噪声如风声、虫鸣声、鸟叫声等,而其中风噪影响尤为严重,风噪声信号大多是 1 000 Hz 以下的低频信号。图 2(a)和 2(b)分别是野外车辆中轮式车和履带车声信号典型样本的频域幅度谱波形图。由图可知,车辆声信号主要频率成分也在 1 000 Hz 以下,目标信号与噪声信号同处一个频段,故单通道信号难以通过传统的傅里叶变换等方法很好地分离出风噪声,去除风噪干扰。因此本文采用微型十字声阵列采集 4 通道声音信号进行降噪处理。

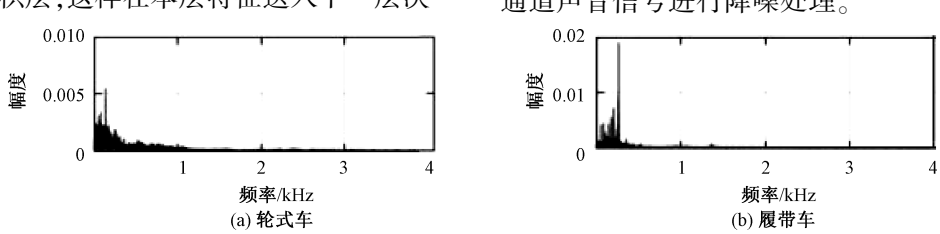


图 2 车辆频域幅度谱波形图

Fig. 2 Amplitude spectrum waveform of vehicle in frequency domain

由文献[10]可知,通过微型十字声阵列采集到的声音信号具有如下特点:

1)由于十字阵列中 4 个阵元传声器具有空间位置差异,各阵元传声器接收的信号相位不同,具有相位差;

2) 在野外环境中,十字阵列的各个阵元接收到的噪声是由风噪声、电路噪声组成的,其中又以风噪声为主要噪声成分,然而无论是风噪声,抑或是电路噪声,都是不相关的,所以,可以认为各阵元获取到的噪声不具有相关性。

设第 i 个阵元传声器接收的信号为

$$y_i(t) = s_i(t) + n_i(t), i = 1,2,3,4. \quad (1)$$

式中:$s_i(t)$ 和 $n_i(t)$ 分别表示的是第 i 个阵元传声器接收到的目标声音信号与噪声信号。又根据文献[11]中的经验公式,当麦克风阵列的各个传声器之间的间距在厘米级时,各阵元的目标声音信号有如下关系

$$s_i(t) \approx s_j(t), i \neq j. \quad (2)$$

令 $y_m(t)$ 表示十字声阵列采集到的声音信号幅度值的平均值,则该值与各阵元接收到的目标声音信号和噪声信号有如下关系

$$\begin{aligned} y_m(t) &= \frac{\sum_{i=1}^{4} y_i(t)}{4} \\ &= \frac{\sum_{i=1}^{4} S_i(t) + \sum_{i=1}^{4} n_i(t)}{4} \\ &\approx s_i(t) + \frac{\sum_{i=1}^{4} n_i(t)}{4}. \end{aligned} \quad (3)$$

根据微型十字声阵列采集到的声音信号特点(式(2))可知 $n_i(t)$ 和 $n_j(t)$ 是不相关的,所以 $y_m(t)$ 的信噪比可以表示为

$$\begin{aligned} \mathrm{SNR}(y_m(t)) &= 10\lg \frac{\mathrm{Power(signal)}}{\mathrm{Power(noise)}} \\ &= 10\lg \frac{\mathrm{Power}(s_1(t))}{\mathrm{Power}(n_1(t))/4} \\ &= 10\lg \frac{\mathrm{Power}(s_1(t))}{\mathrm{Power}(n_1(t))} + 10\lg 4. \end{aligned} \quad (4)$$

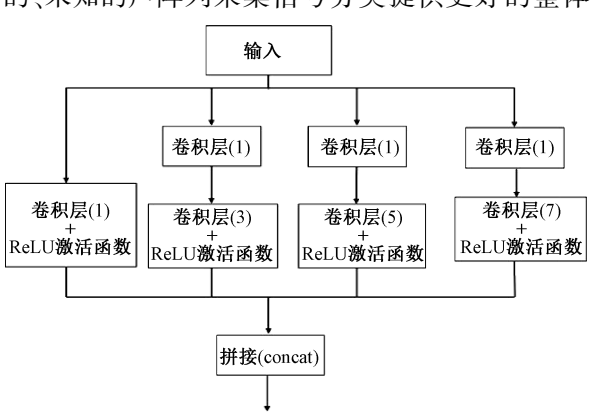
由式(4)可知采用十字声阵列声音信号相关平均法信噪比提高了 6 dB($10\lg 4 = 6$)。此过程是理想情况下针对声阵列 4 通道信号输入采用相关平均法进行降噪处理的结果,然而实际上可能背景噪声具有弱相关性,采用相关平均法的降噪效果不一定最好,但这充分说明 4 通道输入扩大了目标声音信号和噪声信号的差异性,所以本文采用卷积神经网络对 4 通道信号进行自适应学习,尽可能地挖掘 4 通道同步信号的互补信息,最大程度地实现目标信号与背景噪声的解耦合。

为控制网络规模,将降噪网络压缩到特征提取网络之中,相应只需将特征提取网络的输入通道数变为 4 即可。因为特征提取网络的损失函数是判断是否为该目标的误差,网络迭代学习的过程就是减少判断误差的过程,也就是提高目标识别率的过程。这一过程必然要抑制背景噪声干扰,而 4 通道同步信号输入因为空间位置不同导致目标声音信号和背景噪声会产生不同的相位差而提供丰富的互补信息,极大地增强了去背景噪声干扰的效果。

1.3　改进的 Inception 多尺度特征提取层

一般通过增加网络的深度来提高网络的性能。但是直接增加网络的深度,会带来 2 个比较严重的问题:第一是网络规模太大,参数太多容易过拟合,第二是计算资源会急剧增加,即便只有 2 个卷积层连在一起,其计算量也会以幂级增加。所以本文采用 Inception 结构作为输入层来降低网络深度[12],结构图如图 3 所示。

如图 3 所示,Inception 结构对输入分别使用不同大小卷积核进行卷积操作,从而获得不同感受野的特征进行并行计算,最后通过拼接堆叠得到一个单一输出。这样做增加了网络的宽度,同时增加了网络对多尺度的适应性,有利于从含噪的声阵列 4 通道信号中尽可能多地捕获目标信号的特征信息与分布规律,抑制噪声的干扰,为新的、未知的声阵列采集信号分类提供更好的整体



图中括号中的数字代表卷积层卷积核的大小

图 3　Inception 结构图

Fig. 3　Structure of Inception

泛化能力。

鉴于模态分解的思想,可以将车辆目标信号和以风噪为主的背景噪声信号分解成若干个基信号,而这些基信号又可以组合成不同的特征,2 类基信号在不同尺度不同通道特征下的分布是不均匀的,需要加强车辆目标基信号占比多的通道特征,削弱背景噪声基信号占比多的通道特征,因此我们在 Inception 结构中引入注意力机制,自适应地学习特征权重,使得有效的特征向量权重大,无效或效果小的特征向量权重小的方式训练模型达到更好的结果。加入注意力机制改进的 Inception 结构图如图 4 所示。

假设 Inception 结构的输出特征维度为 $W\times C$(W 为特征向量长度,C 为通道数),通过 global pooling 层,拉伸成 $1\times C$,然后经过 2 个 FC 层、ReLU 层和 sigmoid 层得到 C 个概率,与 Inception 的输出相乘,相当于为其输出特征的 C 个通道都赋予一个权重。在训练过程中,自动去除低权重的通道,保留高权重的通道,突出车辆信号的关键通道特征信息。

1.4 分类特征提取

本文中野外车辆主要分为轮式车和履带车 2 类,我们采集的目标声音信号来源主要是车辆在运动过程中产生的噪声。

根据车辆噪声产生方式的不同,通常可以将噪声分成机械噪声和空气动力性噪声[13]。野外

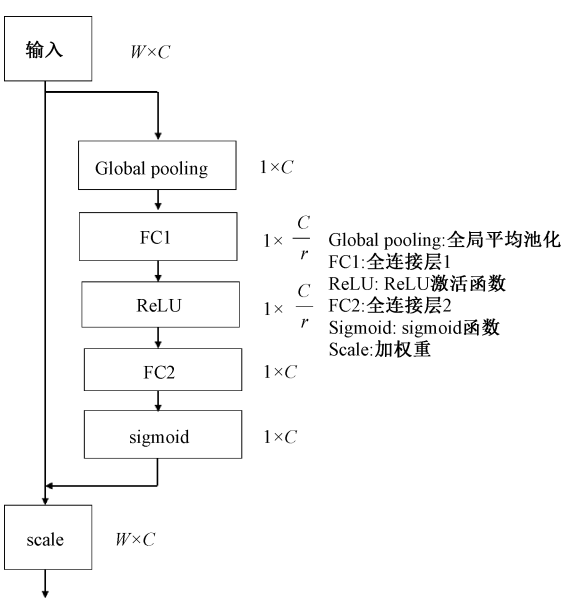


图 4 改进的 Inception 多尺度特征提取层结构图

Fig. 4 Structure of improved concept multi-scale feature extraction layer

车辆的主要噪声源是空气动力性噪声,其中空气动力性噪声主要由车辆的气流噪声、发动机的进气系统以及排气系统发出的噪声等声音构成,其中发动机最主要的噪声源是由发动机进排气系统周期性地排放高温高压产生的,其强度与发动机转速和发动机的汽缸容积成正比。齿轮噪声、活塞敲击声、车辆本身的振动噪声则是构成野外车辆机械噪声的主体[14]。

对于运动中的轮式车车辆,机械噪声信号主要集中在高频部分,在空气中传播的过程中很容易因为距离较远被空气吸收,在近声场中机械噪声为声音信号的主要部分[15];相反地,空气动力性噪声信号则主要集中在低频部分,远声场中占据主要部分,也是本文的主要研究对象。

对于运动中的履带车车辆,除履带车发动机产生的空气动力性噪声以外,履带车的履带在运行过程中也会对地面产生振动和进行反复的摩擦碰撞而生成机械噪声。

如果不用分类子网络分别提取特征,而是直接整体提取特征进行训练,则在训练过程中,网络会偏向于提取更有区分度的履带车声信号特征,导致最终学习到的网络权重使网络中的神经元对轮式车声信号的特征提取不足,整体学习到的特征区分度不足,从而降低大小轮式车的识别准确率。而且由于轮式车和履带车声信号最有区分度特征的抽象程度不同,提取同一层次的特征作为车辆分类的标准会使车辆声信号的特征提取趋于均衡,最后学习到的是总体分类效果最好的特征层次,而不是各自最有区分度的特征层次。

针对此类特征分布具有明显层次差异的分类问题,一个好的特征提取器需要能够不遗漏各种目标特征,并兼顾各类型的特征特点统筹安排,突出有效特征,抑制无效信息。

鉴于此,采用分类子网络进行特征提取的策略,即针对小型轮式车、大型轮式车和履带车 3 种车辆目标分别构建分支网络 SWNet、LWNet 和 TNet 提取各类型运动目标有区分度的特征。由于履带车信号中所含特征信息更为丰富,为充分提取其特征信息,TNet 使用了比 SWNet 和 LWNet 更长的网络。

在训练阶段,对特征提取子网络添加额外的辅助 loss 函数,强制其对大小轮式车辆和履带车目标特征进行进一步挖掘,扩大了提取的大小轮式车特征的差异性和提取的轮式车与履带车特征

的差异性。

因为单分支网络提取的是单一运动目标更具区分度的特征,可看成二分类问题,3 个单分支网络输出均为 0 或 1,判断是否为该目标,故辅助 loss 函数可采用交叉熵损失函数,其表达式为

$$C = -\frac{1}{n}\sum_{i=1}^{n}[y_i \ln\sigma(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{x}_i) + (1-y_i)(1-\ln\sigma(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{x}_i))]. \tag{5}$$

式中:y_i 为信号标签,可取 0 或 1,0 表示是目标信号,1 表示是非目标信号;$\boldsymbol{x}_i$ 表示样本,$\boldsymbol{W}$ 表示网络参数,$\sigma(\cdot)$ 表示激活函数。

为解决数据集各类车辆数据数量不均匀而带来的泛化性能差的问题,采用代价敏感方法,在损失函数中添加权重 ω_i,赋予各个类别不同的错分代价,错分稀有类的样本需要付出更大的代价,即在损失函数中赋予小样本类大的权重,损失函数变为如下形式

$$C = -\frac{1}{n}\sum_{i=1}^{n}\omega_i[y_i \ln\sigma(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{x}_i) + (1-y_i)(1-\ln\sigma(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{x}_i))]. \tag{6}$$

1.5　多分支多维度特征融合

如上所述,在训练阶段将 4 通道声阵列信号分成轮式车和履带车 2 类 3 种目标然后分别用对应的 3 个分类子网络提取其最具代表性的特征,之后将 3 个分支提取的特征进行融合作为最终进行分类的整体特征。本文主要进行 2 个层面的特征融合,一是多分支同一层次的特征拼接,一是直接将低层次特征和高层次特征进行叠加。

前者应用于 SWNet 和 LWNet 的特征融合,因为它们提取的特征都是来源于特征提取网络的第 3 个基本单元的输出,网络抽象层次相同,所以可以直接进行维度拼接。虽然 SWNet 和 LWNet 网络结构相同,但是由于损失函数不同,使 SWNet 和 LWNet 网络的神经元激活程度不同,学习到的网络权重不同,导致最后提取到的特征不同,拼接在一起,扩充了特征丰富性。但是由于同一网络结构提取的特征具有一定的重叠效应,会有一定的冗余信息,故使用一层具有较大卷积核的卷积层进行一次特征提取和降维,提高特征的利用率。

最后再与 TNet 提取的高层次特征进行叠加,共同作为后面层的输入。设 SWNet 和 LWNe 提取的特征拼接后的特征矩阵为 $\boldsymbol{X}_1$,TNet 提取的特征矩阵为 $\boldsymbol{X}_2$,特征层叠加后的输出为 $\boldsymbol{X}_3$,$\boldsymbol{W}$ 为新加的卷积层的权重矩阵,$\boldsymbol{X}_1$、$\boldsymbol{X}_2$、$\boldsymbol{X}_3$ 三者之间的关系如下所示

$$\boldsymbol{X}_3 = \boldsymbol{W}\boldsymbol{X}_1 + \boldsymbol{X}_2. \tag{7}$$

如图 5 所示,一个方块代表对应网络的一个输出通道,表现形式是一个特征向量,4 个深灰色方框泛指有多个轮式车特征向量,4 个浅灰色方块泛指有多个履带车特征向量。直接拼接是将轮式车和履带车特征的全部作为下一层卷积网络的输入,卷积叠加方式则是将轮式车和履带车特征通过卷积叠加充分混合以后再作为下一层卷积网络的输入,图 5(b)中的深灰色方块与浅灰色方块之间的连线则是式(7)的图形化表示,即轮式车的一个特征向量乘以一个权重再与履带车相加。

由于履带车和轮式车特征提取网络卷积层数和卷积核大小均不同,两者特征卷积叠加的过程即是特征组合的过程。设轮式车特征向量的数量为 A,履带车特征向量的数量为 B,则可获得的组合数为 $A\times B$。这样相较于直接拼接,一方面控制了下一层卷积网络输入的规模,另一方面也实现了信息的充分混合,提高了特征利用率,增强了网

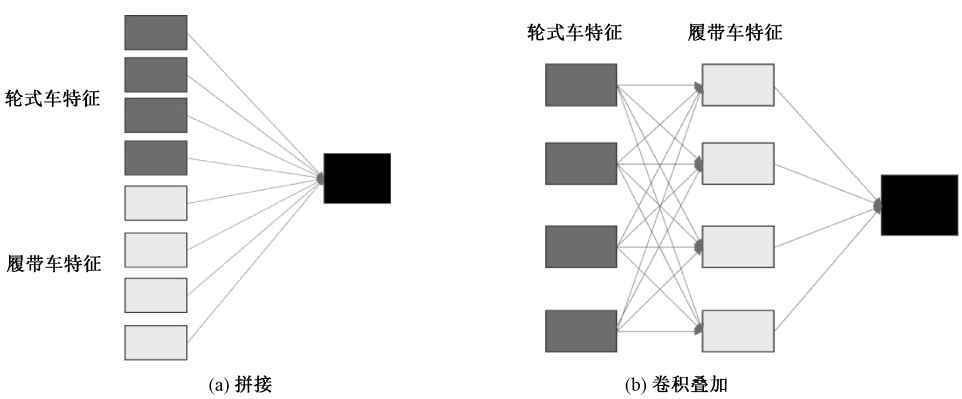


图 5　不同特征融合方式示意图

Fig. 5　Schematic diagram of different feature fusion methods

络的泛化性能。

综上所述,特征融合网络通过拼接、高低层次特征的结构重组和权重共享,在避免低层次特征信息遗失的同时突出了关键特征且抑制了无用特征,增强了特征的表示能力。

2 实验结果与分析

2.1 实验数据集

本文使用实验室自制的数据集验证 FVC-CNN 网络模型对野外车辆目标分类的有效性。实验场景如图 6 所示,车辆在野外一段长为 1 000 m 的硬土路上行驶,十字阵列的声音采集设备放置在距离道路中心位置 60 m 处。

声音采集设备的采样率为 8 192 Hz,一辆车辆行驶一趟采集的数据记为一条数据。为保证野外车辆识别的实时性,将每条数据按照 1 s 时长进行分帧,以每帧的声音信号作为训练样本集。本文主要识别的是小型轮式车(SW),大型轮式车(LW)和履带车(T)3 种车辆目标,然后按照训练集信号条数:测试集信号条数=2∶1 的比例分割总样本集,各个车型数据集如表 1 所示。

实验地点的风力介于 3~4 级间,偶尔会高达 5~6 级,风力数据由风力计采集,整个实验阶段的风力数据统计如表 2 所示。

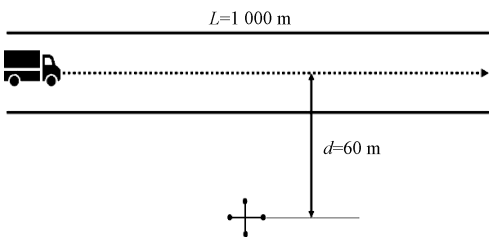


图 6 实验场景图

Fig. 6 Experimental scene map

表 1 实验数据集

Table 1 Experimental data set

	训练集			测试集		
	SW	LW	T	SW	LW	T
条数	96	88	24	48	44	12
帧数	52 120	24 324	29 512	25 975	12 002	14 711

表 2 实验阶段的风力数据统计

Table 2 Wind data statistics of experimental stage

风力/级	1	2	3	4	5
风速/(m/s)	0~0.2	0.3~1.5	1.6~3.3	5.5~7.9	8.0~10.7
数据比例/%	8	11	15	23	30

2.2 实验设置

为了控制网络深度,避免过拟合,首先将每一帧信号平均切成 8 段,得到 1 024 个离散的点,然后进行归一化、预加重、低通滤波操作。最后进行 4 倍降采样,由于采集的是 4 通道信号,所以输入信号变成维度为 4×256 的张量。采用的优化算法是动量为 0.9 的随机小批量梯度下降方法,其中 batch size 为 32,初始学习率为 0.001。实验中学习率是动态调整的,采用分段降低的策略,以第 60 次迭代为分界点将其分为 2 段[1,60]和[61,140],后者学习率降低为前者的 30%。

为增加说服力,用训练集重复进行 3 次训练,再分别用训练出来的参数在测试集上进行测试,得到 3 个识别准确率,取其平均识别准确率作为总识别准确率。

2.3 结果分析

2.3.1 FVC-CNN 模型与其他车辆分类算法的比较

表 3 显示了使用本文提出的 FVC-CNN 模型,采用梅尔倒谱系数(Mel-frequency cepstrum coefficients,MFCC)[16]提取车辆特征,混合高斯模型(Gaussian mixed model,GMM)[17]进行分类的传统方法(MFCC+GMM),时间卷积神经网络(temporal convolutional network,TCN)与改进后的 TCN 网络(M_TCN)[18]以及由 4 个卷积基本单元组成的 CNN 网络(CNN,网络结构和 TNet 一样)在本文数据集上对野外车辆分类的结果,除 FVC-CNN 外其他网络模型都是单通道输入。

由表 3 可看出,传统的 MFCC+GMM 模型在野外条件下车辆识别效果明显不如卷积网络模型,说明卷积网络模型相较于传统的特征提取对信息的利用更加充分,通过监督和学习更加客观地将不同层次的特征提取并整合起来,使最后的形成特征更有区分度。而本文提出的 FVC-CNN 模型则在一维卷积网络模型的基础上针对各类目

表 3 各模型识别准确率

Table 3 Recognition accuracy of each model %

模型	识别准确率 1	识别准确率 2	识别准确率 3	总识别准确率
MFCC+GMM	80.34	79.88	80.19	80.14
CNN	82.55	82.46	82.33	82.50
TCN	83.76	83.63	83.72	83.70
M_TCN	87.38	87.36	87.32	87.35
FVC-CNN	94.26	94.18	94.21	94.22

标最具区分度特征抽象层次不一样的特点设计网络,成功地将总识别率提高到 94.22%,识别效果最好。而且 FVC-CNN 模型的 3 个识别准确率与总识别准确率偏差不超过 0.05%,说明模型收敛性好。

2.3.2　4 通道声阵列信号输入性能评估

1.2 节指出,可利用十字阵列采集的 4 通道同步信号中声源信号具有相关性但风噪不具有相关性的特点,抑制风噪的干扰,增强声源信号。为验证 4 通道降噪的效果,本文设计了一组对比实验,实验结果如表 4 所示。为确保对比的有效性,网络整体结构基本不变,只是输入信号分为单通道信号和 4 通道信号,相应地改变一下网络的输入通道数。

对比表 4 中实验 1 和实验 2 可知,在 CNN 网络结构中 4 通道降噪识别率提高 8.34%;对比表 4 中实验 3 和实验 4 可知,在 TCN 网络结构中 4 通道降噪识别率提高 7.53%;对比表 4 中实验 5 和实验 6 可知,在 FVC-CNN 模型中 4 通道降噪识别率提高 3.79%。在不同网络结构中,4 通道降噪识别率均有提高。之所以相较于 CNN 网络,TCN 网络和 FVC-CNN 模型 4 通道降噪的效果依次降低,是因为 TCN 网络和 FVC-CNN 模型中对目标特征本质的挖掘依次提升,这意味着网络本身降风噪影响的效果越好,加入 4 通道降噪算法的效果就越弱。

2.3.3　改进的 Inception 多尺度特征提取层性能评估

用 4 通道信号检验改进的 Inception 多尺度特征提取层的效果,实验结果如表 5 所示。实验均在 FVC-CNN 模型的基础上进行,对比的 3 个组分别是不用 Inception 网络结构,用原始的 Inception 网络结构和用改进的 Inception 网络结构。

对比表 5 实验 1 与实验 2 可知,加入原始 Inception 后网络识别率提高 1.62%,说明在输入层就先对信号进行多尺度的映射,有效地保留了原信号的特征规律,也能起到抑制噪声干扰的作用;如表 5 实验 2 与实验 3 所示,在此 Inception 基础上引入注意力机制,网络识别率提高 1.32%,注意力机制有针对性地削弱了影响分类的通道在输出特征中所占的比例,即抑制噪声的干扰,提高特征的利用率,最终达到提高分类性能的效果。

表 4　4 通道降噪性能评估

Table 4　Four channel noise reduction performance evaluation　%

实验编号	模型	总识别准确率
1	CNN(单通道)	82.50
2	CNN(4 通道)	90.84
3	TCN(单通道)	83.70
4	TCN(4 通道)	91.23
5	FVC-CNN(单通道)	90.43
6	FVC-CNN(4 通道)	94.22

表 5　改进的 inception 多尺度特征提取层的性能评估

Table 5　Performance evaluation of improved inception multiscale feature extraction layer　%

实验编号	模型	总识别准确率
1	4 通道+未加 Inception	91.28
2	4 通道+原始 Inception	92.90
3	4 通道+改进 Inception	94.22

2.3.4　特征提取子网络性能评估

为检验特征提取子网络的性能,由于大小轮式车的特征提取网络结构相同,只是损失函数不同,导致网络参数不同,所以设计以下一组对比实验,用不同的特征提取网络的输出特征直接接分类器分别做车辆的分类,3 类车辆的识别率测试结果如表 6 所示。

由表 6 可知,3 个特征提取子网络对应的目标识别率最高,说明特征子网络偏向于对相应目标特征的学习,抑制对非相应目标特征的学习,有效地提取了对应目标的具有代表性的特征,并扩大了不同特征提取网络提取特征的差异性。

2.3.5　多分支多维度特征特征融合子网络性能评估

为评估多分支多维度特征特征融合的效果,又设计了一组新的实验进行对比测试,测试结果如表 7 所示。FVC-CNN-LW 和 FVC-CNN-T 分别表示只以 LWNet 和以 TNet 作为特征提取网络进行目标分类(特征提取网络由 FVC-CNN 模型的多分支变为单分支),FVC-CNN-Half 表示 FVC-CNN 模型去掉进行多维度特征融合的 MergeNet 部分,直接将 3 个特征提取网络的输出拼接在一起进行目标分类。

表 6　特征提取子网络识别性能评估

Table 6　Performance evaluation of feature extraction network recognition　%

特征提取子网络	识别率		
	大型轮式车	小型轮式车	履带车
LWNet	98.78	77.07	10.64
SWNet	30.89	98.86	83.13
TNet	34.17	82.36	99.47

表 7 多分支多维度特征特征融合性能评估

Table 7 Performance evaluation of multi-dimensional feature fusion %

实验编号	模型	总识别准确率
1	FVC-CNN-LW	89.68
2	FVC-CNN-T	90.80
3	FVC-CNN-Half	92.38
4	FVC-CNN	94.22

从实验结果可以看出,FVC-CNN-Half 网络直接拼接 3 类目标特征提取子网络输出的特征做分类,使识别准确率提高到 92.38%,相较于只用一种特征提取网络的 FVC-CNN-LW 和 FVC-CNN-T 识别率均有提升。FVC-CNN 网络在 FVC-CNN-Half 网络的基础上添加了多分支多维度特征融合部分,也提高了识别率,说明特征融合网络可以有效地突出车辆目标的有效特征,增强特征表述能力。

3 结束语

针对野外环境下单通道车辆声信号受风噪影响严重、分类性能较低的问题,本文提出基于声阵列的野外车辆信号分类模型 FVC-CNN。该模型可以分为 4 个模块:1)采用 4 通道输入抑制风噪干扰,有效地增强了目标信号;2)输入层采用加入注意力机制的 Inception 网络结构,有针对性地利用不同尺度的特征信息,尽可能地减小背景噪声在输出特征中所占的比例,并与后面的分类特征提取网络相结合,增强相应的目标信号特征在输出特征中所含比例;3)使用 SWNet,LWNet 和 TNet 等 3 个分支网络分别提取 3 种目标的不变性特征,提高了特征的区分度;4)最后将 3 种特征进行 2 个维度的特征融合,增强了特征的表达能力。结果表明,本文提出的 FVC-CNN 模型有效地抑制了风噪干扰,能够提取具有区分度的野外环境下车辆的精细特征,达到 94.22% 的识别准确率,取得了优异的分类效果。因为野外环境下设备计算资源和存储资源的限制,下一步考虑压缩网络参数规模,保证在较小的参数规模下还能维持良好的识别性能。

参考文献

[1] 罗向龙,牛国宏,吴潜蛟,等. 基于经验模态分解和支持向量机的车型声频识别[J]. 应用声学, 2010, 29(3): 178-183.

[2] Kandpal M, Kakar V K, Verma G. Classification of ground vehicles using acoustic signal processing and neural network classifier[C]//2013 INTERNATIONAL CONFERENCE ON SIGNAL PROCESSING AND COMMUNICATION (ICSC). Noida, India:IEEE, 2013: 512-518.

[3] Yang S S, Kim Y G, Choi H. Vehicle identification using discrete spectrums in wireless sensor networks[J]. Journal of Networks, 2008, 3(4): 51-63.

[4] 孙铭堃,梁令羽,汪涵,等. 基于级联卷积网络的面部关键点定位算法[J]. 中国科学院大学学报,2020,37(4):562-569.

[5] Zhang H M, McLoughlin I, Song Y. Robust sound event recognition using convolutional neural networks[C]//2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). South Brisbane, QLD, Australia:IEEE, 2015: 559-563.

[6] 刘莹. 基于深度学习的轨迹预测[D]. 成都:电子科技大学,2019.

[7] 赖策. 卷积神经网络中的激活函数分析[J]. 科学技术创新, 2019 (33): 35-36.

[8] Kalayeh M M, Shah M. Training faster by separating modes of Variation in batch-normalized models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(6): 1483-1500.

[9] 黄琦. 智能传感器侦察网络中的地面目标识别算法研究[D]. 合肥:中国科学技术大学, 2006.

[10] Tiete J, Domínguez F, da Silva B, et al. SoundCompass: a distributed MEMS microphone array-based sensor for sound source localization[J]. Sensors (Basel), 2014, 14(2): 1918-1949.

[11] 黄景昌. 野外监控传感网中运动目标声音信号的探测与识别研究[D]. 北京:中国科学院大学,2015.

[12] Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: IEEE,2015:1-9.

[13] 李光海. 基于聚焦算法的机械振动噪声声源定位的研究和应用[D]. 武汉:武汉理工大学,2015.

[14] 胡利萍. 野外监控传感网声探测器的环境适应性研究[D]. 北京:中国科学院研究生院,2011.

[15] 杨旭. 基于盲源分离的声信号降噪算法研究[D]. 北京:中国科学院大学,2012.

[16] Shi W L, Fan X H. Research on armored vehicle classification based on MFCC and SVM[C]// 2017 3rd IEEE International Conference on Computer and Communications (ICCC). Chengdu, China: IEEE,2017: 1938-1941.

[17] Wei J S, Lv J C, Xie C Z. A new sparse representation classifier (SRC) based on probability judgement rule[C]// 2016 International Conference on Information System and Artificial Intelligence (ISAI). Hong Kong, China: IEEE, 2016: 338-342.

[18] 范裕莹,李成娟,易强,等. 基于改进 TCN 网络的野外运动目标分类[J/OL]. 计算机工程:(2020-09-10)[2021-03-29]. https://doi.org/10.19678/j.issn.1000-3428.0058750.