

文章编号:1007-5321(2021)01-0086-06

DOI:10.13190/j.jbupt.2020-122

# D-2-DenseNet 噪音鲁棒的城市音频分类模型

曹 毅, 黄子龙, 盛永健, 刘 晨, 费鸿博

(江南大学 机械工程学院, 无锡 214122)

**摘要:** 为了提高噪音环境下城市音频分类系统的鲁棒性,提出了一种双特征 2 阶密集卷积神经网络(D-2-DenseNet)噪音鲁棒的城市音频分类模型. 首先介绍了噪音添加和噪音鲁棒处理,阐述了一种双特征互补的算法;然后结合 2 阶密集卷积神经网络与自适应机制提出了一种噪音鲁棒音频分类模型;双特征 2 阶密集卷积神经网络. 模型采用双特征互补自适应算法,可在特征提取与模型训练中更有针对性地提取有效音频信息,降低噪音干扰,以提高噪音鲁棒性. 最后,基于 Dcase2016 数据集开展噪音环境下城市音频分类测试. 实验结果表明,模型分类准确率分别可达 77.12%、75.52%,与基线模型相比,平均分类准确率分别提高了 8.51% 和 10.38%,验证了模型良好的噪音鲁棒性.

**关键词:** 城市音频分类;噪音鲁棒性;双特征互补;2 阶密集卷积神经网络;双特征 2 阶密集卷积神经网络  
**中图分类号:** TP391.42 **文献标志码:** A

## Noise Robust Urban Audio Classification Based on 2-Order Dense Convolutional Network Using Dual Features

CAO Yi, HUANG Zi-long, SHENG Yong-jian, LIU Chen, FEI Hong-bo

(School of Mechanical Engineering, Jiangnan University, Wuxi 214122, China)

**Abstract:** A noise robust urban sound event classification model based on 2-order dense convolutional network using dual features(D-2-DenseNet) is proposed, which aims at the problems of insufficient robustness of current models. Firstly, the brief introduction of the method of noise adding and robust processing is presented. Moreover, a dual feature mutual compensation algorithm and 2-order dense convolutional network is presented. Meanwhile, a noise robust urban sound event classification model based on 2-DenseNet using dual features, i. e. D-2-DenseNet is proposed. Theoretically, D-2-DenseNet combines the advantages of feature compensation and 2-order dense convolutional neural network. The dual feature mutual compensation adaptive algorithm can effectively extract audio information and reduce noise interference to improve noise robustness. Finally, in order to validate advantages of the D-2-DenseNet, this new model is exploited in the urban sound event classification based on Dcase2016 datasets. Under conditions of channel noise and environmental noise, the experiment shows that the accuracy of the network is respectively 77.12% and 75.52%, which has added 8.51% and 10.38% compared with baseline. The noise robustness of D-2-DenseNet are also effectively verified.

**Key words:** sound event classification; noise robust; dual features mutual compensation; 2-order dense convolutional network; 2-order dense convolutional network using dual features

收稿日期: 2020-08-16

**基金项目:** 国家自然科学基金项目(51375209);江苏省“六大人才高峰”计划项目(ZBZZ-012);高等学校学科创新引智计划项目(B18027);江苏省研究生创新计划项目(JNKY20\_1928)

**作者简介:** 曹 毅(1974—),男,教授,硕士生导师, E-mail: caoyi@jiangnan.edu.cn.

声音携带了城市中大量关于生活场景和物理事件的信息,通过深度学习方法智能感知各个声源自动提取上述信息,在构建智慧城市中具有巨大的潜力和运用前景<sup>[1]</sup>. 近年来,基于深度学习开展城市音频分类和对泛化能力的研究取得了飞跃性的进展<sup>[2-3]</sup>. 但是,城市音频分类作为开放环境的系统,若要运用于智慧城市的构建中,必须要考虑各类噪音的干扰,而噪音是影响城市音频分类技术广泛实用化的一个关键问题. 因此,针对噪音环境下城市音频分类模型的鲁棒性研究具有重要的理论与实践意义.

针对音频分类的噪音鲁棒性,国内外诸多学者分别采用特征补偿、模型自适应等算法开展了前瞻性的研究<sup>[4-7]</sup>. 其中,特征补偿主要在特征提取过程中补偿音频的有效信息,以增强抗干扰能力;而模型自适应依靠网络的自适应调整能力,自动变更权值以筛选有效信息,使得模型更有利于分类. 蔡等<sup>[4]</sup>提出了一种子带能量规整感知线性预测系数特征,通过减去子带能量偏差来规整时频能量,能有效补偿语音特征信息,噪音环境下模型的平均识别精度提升了 5.68%. Yu 等<sup>[5]</sup>研究了一种频谱恢复特征补偿算法,以减少心音信号中的噪音分量,并采用了多种训练策略来训练深度神经网络的噪音鲁棒音频分类模型,在不同高斯白噪音下准确率提升了 2.36%~3.87%. Ilyas 等<sup>[6]</sup>基于卷积神经网络开展了模型自适应研究,使用卷积神经网络进行自适应的特征提取和分类,卷积神经网络在噪音环境下分类准确率提高了 4.53%. 易等<sup>[7]</sup>基于迁移学习的模型自适应方法,用干净语音的老师模型指导带噪音语音的学生模型进行训练,模型在噪音环境下平均错误率降低了 7.29%.

尽管已有诸多学者针对音频分类的噪音鲁棒性开展了较为深入的研究,但在已有研究成果中,增强噪音鲁棒性方法相对单一,未能将特征补偿与模型自适应有机结合;而且当前已有模型基于噪音环境下的平均分类准确率仅提升 5% 左右,模型的噪音鲁棒性仍需进一步提高.

基于上述,提出了一种基于双特征 2 阶密集卷积神经网络 (D-2-DenseNet, 2-order dense convolutional network using dual features) 噪音鲁棒的城市音频分类模型. 模型采用双特征互补补偿算法,并结合 2 阶密集卷积神经网络 (2-DenseNet, 2 order dense convolutional network) 和自适应机制,能有效提取音

频信息,降低噪音因素干扰,提高模型噪音鲁棒性. 最后,使用 D-2-DenseNet 模型开展噪音环境下城市音频分类的测试,实验结果验证了 D-2-DenseNet 模型具有良好的噪音鲁棒性.

## 1 噪音环境及噪音鲁棒处理

### 1.1 信道噪音构建

在音频信号传输环节,若存在信道噪音,易引发传输失真的问题,严重的还将制约信号的正常传输. 高斯白噪音 (WGN, white Gaussian noise) 可很好地模拟各类未知的信道噪音,是分析信道噪音的理想模型<sup>[8]</sup>.

在音频信号处理中,音频样本的模拟信号通过采样定理可得数字信号  $S = \{s_1, s_2, \dots, s_n\}$ , 由数字信号可以计算出音频文件的信号功率.

$$P_{\text{sig}} = \sum_{k=1}^n s_k^2 / n \quad (1)$$

添加指定信噪比的 WGN 构建信道噪音模型,信噪比 (SNR, signal-to-noise ratios) 为音频信号功率  $P_{\text{sig}}$  与噪音信号功率  $P_{\text{noi}}$  比值的对数:

$$Q_{\text{SNR}} = 10 \lg (P_{\text{sig}} / P_{\text{noi}}) \quad (2)$$

由式(1)和式(2)可得高斯白噪音的噪音信号功率  $P_{\text{noi}}$  以及数字信号  $S^g = \{s_1^g, s_2^g, \dots, s_n^g\}$ .

将  $S^g$  添加到原始音频信号中,则生成了信道噪音模型下的音频数字信号:

$$S^{\text{mg}} = \{s_1^{\text{mg}}, s_2^{\text{mg}}, \dots, s_n^{\text{mg}}\} = \{s_1 + s_1^g, s_2 + s_2^g, \dots, s_n + s_n^g\} \quad (3)$$

### 1.2 环境噪音的添加

在城市复杂开放环境中,环境噪音(嘈杂人声、车辆、街道等)更是城市音频分类的鲁棒性必须考虑的干扰因素. NoiseX-92 和 Thchs30 数据集提供了专业的环境噪音<sup>[9-10]</sup>,是当前语音识别与音频分类的主要研究对象,将环境噪音添加进城市音频样本中,构建环境噪音模型,进一步开展环境噪音鲁棒性研究.

环境噪音的原始数字信号为  $S^e = \{s_1^e, s_2^e, \dots, s_m^e\}$ , 由于噪音数据集与音频数据集中样本的参数(时长,采样频率,声道数等)的差异性,故将信号处理后输出的环境噪音数字信号  $S^p$  定义为

$$S^p = \{s_1^p, s_2^p, \dots, s_n^p\} = \begin{cases} f_p \{s_1^e, s_2^e, \dots, s_n^e\}, & m \geq n \\ f_p \{s_1^e, s_2^e, \dots, s_m^e, s_1^e, s_2^e, \dots\}, & m < n \end{cases} \quad (4)$$

其中  $f_p$  为信号处理函数。

将处理后的环境噪声添加到原始音频信号中,即生成添加环境噪声的音频数字信号  $S^{mp}$ :

$$S^{mp} = \{s_1^{mp}, s_2^{mp}, \dots, s_n^{mp}\} = \{s_1 + s_1^p, s_2 + s_2^p, \dots, s_n + s_n^p\} \quad (5)$$

### 1.3 噪音鲁棒处理

噪音鲁棒是当前语音识别、声纹识别与音频分类领域中评价噪音环境下模型性能的重要标准<sup>[11]</sup>。上述不仅是评价噪音环境下各模型训练测试效果的基础,同时也是分析量化算法的抗噪性、有效性的指标。因此定义城市音频分类模型的噪音鲁棒性基本概念如下。

1) 模型在干净环境下,具有优秀的分类准确率或识别准确率。

2) 模型出现中、高信噪比噪音干扰时,只能对算法的准确率产生较小影响。

3) 模型出现复杂高噪音干扰时,不可对算法性能产生“灾难性”的影响。

## 2 双特征 2 阶密集卷积神经网络

### 2.1 双特征互补补偿算法

在处理音频时通过信号变换或映射方式,生成代表性的数据来表示音频信号,其定义为特征提取<sup>[12]</sup>。当前主流的音频特征提取方案有:梅尔倒谱系数(MFCC, Mel-scale frequency cepstral coefficients)、伽马通倒谱系数(GFCC, Gammatone frequency cepstral coefficients)、语谱图、Filter Bank 等。

MFCC 作为使用最广泛的语音识别和音频分类特征提取方案,通过人耳听觉特性与梅尔标度频率的非线性对应关系计算出的赫兹频谱特征,在干净音频环境下具有较好的分类准确率。

GFCC 通过模拟耳蜗听觉模型而设计的伽马通滤波器组,在考虑不同的背景噪音和信噪比等因素下,能更好地拟合听觉系统的抗干扰性,故 GFCC 特征具有更好的噪音鲁棒性。

综合上述 2 种特征的优点,提出了一种双特征互补补偿算法,如图 1 所示。算法的主要特点如下。

1) 多通路并行网络的结构框架。MFCC 和 GFCC 2 种特征分别独立输入到特征提取网络中,每个特征提取网络能更独立和更有针对性地提取不同特征的有效音频信息。

2) 特征提取网络处理的信息数据扁平化。其中  $X^{mel}$ 、 $X^{gam}$  分别表示 MFCC 和 GFCC 2 种特征信息数

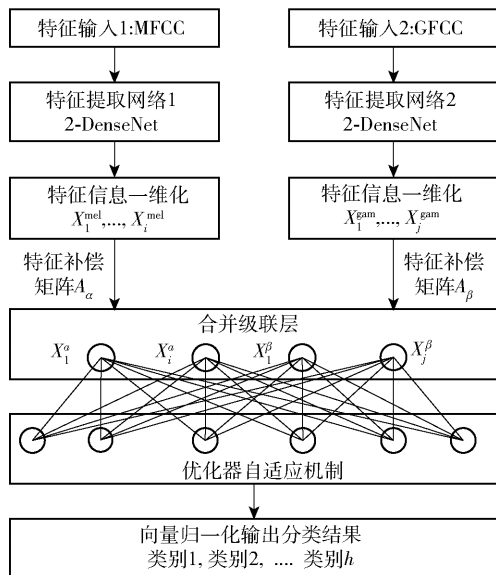


图 1 双特征互补补偿算法结构

据扁平化。分别通过特征补偿矩阵  $A^\alpha$  与  $A^\beta$  在合并级联层进行补偿后,2 种特征信息分别为  $X^\alpha$ 、 $X^\beta$ ;然后在全连接层进行调整,2 种特征的信息相互补偿,使得模型具有良好的噪音鲁棒性。

### 2.2 2-DenseNet

必须指出的是,双特征互补补偿算法中特征提取网络是提取音频有效信息的关键,故为了使特征提取网络高效地提取 2 种特征的有效音频信息,提出了 2-DenseNet<sup>[13]</sup>。2-DenseNet 采用 2 阶有关连接方法,当前层以 2 阶有关连接的方式连接到前 2 层,该方式在避免梯度消失的前提下,有针对性、有规律地重用前 2 层特征图层,能高效地融合前 2 个特征图层的的信息,对于特征提取有显著的优势<sup>[14]</sup>。

#### 2.2.1 2-DenseNet 的前向传播算法

在 2-DenseNet 结构中,前向传播如图 2 所示。 $1 \times 1$  和  $3 \times 3$  的卷积层是一组非线性变化的特征层,用  $X$  表示。每个非线性变化的特征层的输入层被定义为:  $X_1, X_2, \dots, X_l$ 。在 2-DenseBlock 中,第 3 层起,网络卷积变换的特征输出  $U_c$  可定义为

$$U_c = f[\text{BN}(W_{3 \times 3} \otimes f[\text{BN}(W_{1 \times 1} \otimes (X_l, X_{l-1}, X_{l-2}) + B)])] \quad (6)$$

其中:  $(X_l, X_{l-1}, X_{l-2})$  表示当前层通过 2 阶有关连接方式进行通道数合并级联操作,使用前 2 层的特征映射作为当前层的输入,  $W_{3 \times 3}$  和  $W_{1 \times 1}$  表示卷积核大小分别为  $1 \times 1$  和  $3 \times 3$  核函数,  $\text{BN}(\cdot)$  表示批量标准化,  $f[\cdot]$  为 ReLU 激活函数。

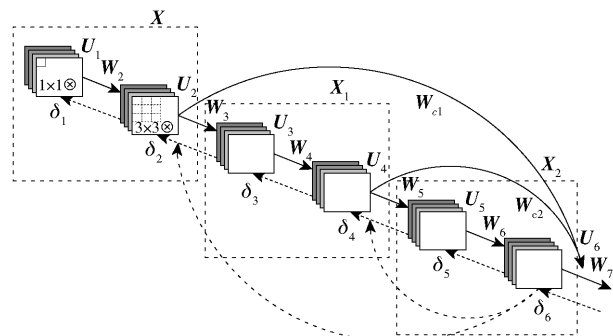


图 2 2-DenseNet 前向传播和反向传播

### 2.2.2 2-DenseNet 的反向传播算法

在 2-DenseNet 特征提取网络训练模型过程中,每个网络层的权重会不断更新直到模型趋于收敛<sup>[14]</sup>.表 1 所示为 2-DenseNet 网络中的反向传播计算.

表 1 6 层 2-DenseBlock 的反向传播各层误差计算

网络层	反向传播计算
输入	$\partial J / \partial W_1 = \delta_1 * W_1 \otimes X$
$L_1$	$\delta_1 = \delta_2 * W_2 \otimes (\partial U_2 / \partial U_1)$
$L_2$	$\delta_2 = \delta_3 * W_3 + \delta_6 * W_{c1}$
$L_3$	$\delta_3 = \delta_4 * W_4 \otimes (\partial U_4 / \partial U_3)$
$L_4$	$\delta_4 = \delta_5 * W_5 + \delta_6 * W_{c2} + \delta_8 * W_{c4}$
$L_5$	$\delta_5 = \delta_6 * W_6 \otimes (\partial U_6 / \partial U_5)$
$L_6$	$\delta_6 = \delta_7 * W_7 \otimes (\partial U_7 / \partial U_6)$
输出	$\partial J / \partial U_9$

表 1 中,  $\delta_1, \delta_2, \dots, \delta_7$  为每一层的误差;每一层的输出定义为  $U_1, U_2, \dots, U_l$ ;  $W_l$  为积层的核函数;  $W_{cl}$  为合并级联层的核函数;  $\otimes$  和  $*$  为卷积和反卷积运算. 从表 1 的推导中可得, 2 阶有关连接层的误差项反向传播回前 2 层, 梯度值可以表示为

$$\frac{\partial J}{\partial W_{l-2}} = (\delta_{l-1} * W_{l-1} + \delta_l * W_{cl} + \delta_{l+1} * W_{c(l+1)}) \otimes X_{l-2} \quad (7)$$

由表 1 和式 (7) 可知, 2-DenseNet 反向传播误差项仅需计算连接层对前 2 层的影响, 与 DenseNet 的密集连接相比, 它更有利于梯度信息计算和加快网络收敛速度, 即网络运行效率更高.

### 2.3 D-2-DenseNet

为了进一步增强模型的噪音鲁棒性, 基于双特征互补补偿的算法, 结合 2-DenseNet 与自适应机制, 提出了一种基于 D-2-DenseNet 噪音鲁棒的城市音频分类模型<sup>[15]</sup>, 网络采用双特征互补补偿自适

应算法, 很好的结合了特征补偿与模型自适应的优点. D-2-DenseNet 分类模型的构建分为以下步骤.

1) 多路并行网络的构建: 音频特征提取所得的 MFCC 和 GFCC 2 种声学特征作为输入, 将 2 种声学特征分别独立输入到特征提取网络中, 输入模型后进行 1 层卷积操作以提取更多有效音频信息特征和 1 次池化处理以压缩图层尺寸.

2) 基于图 2 构建 2-DenseNet 结构: 将特征向量序列输入连续  $k$  个 2-DenseBlock 和过渡层中, 其中 2-DenseBlock 结构是对特征信息进行有针对性地重用, 过渡层中  $1 \times 1$  卷积降维处理、 $2 \times 2$  最大池化进一步压缩尺寸以减少网络参数.

3) 双特征互补补偿算法: 将特征向量序列扁平化为 1 维数据后, 通过特征补偿矩阵  $A$  与网络权值  $W$  在合并级联层进行双特征补偿, 并后接全连接层进行调整权值, 2 种特征的信息相互补偿, 增强模型的噪音鲁棒性.

4) 模型自适应机制: 在网络后端加入优化器和批量数的自适应模块. 首先采用 Adam 优化器训练达到局部最优后, 再用 SGD 优化器进一步的迭代优化以求模型最优; 而批量数设置循环训练, 自适应保存最优模型. 在自适应机制下可进一步提升模型的噪音鲁棒性.

综上可得, D-2-DenseNet 模型的优点如下: ① D-2-DenseNet 以双特征互补补偿算法为基础, 构建了双特征输入的多路并行 2-DenseNet 网络, 能独立和有针对性地提取不同特征的有效音频信息并进行特征补偿; ② D-2-DenseNet 以 2-DenseNet 为特征提取网络, 采用 2 阶有关连接方法, 可更加高效和更具针对性地进行特征重用, 具有一定的自适应能力和较好的收敛速度; ③ D-2-DenseNet 结合双特征互补补偿和 2-DenseNet 的优点, 双特征互补补偿自适应算法, 可在特征提取与模型训练中更有针对性地提取有效音频信息, 使模型具有更好的噪音鲁棒性.

## 3 实验及结果分析

为进一步验证 D-2-DenseNet 模型的有效性, 选用城市音频标准数据集 IEEE AASP 声学场景和事件检测分类挑战赛数据集 Dcase2016<sup>[16]</sup>, 指定信噪比高斯白噪音, 以及 NoiseX-92、Thchs30 数据集的环境噪音, 开展噪音环境下城市音频分类测试. 其中 Dcase2016 数据集训练、验证以及测试样本的个数

划分为:880、290 和 480. 各噪音样本通过信号处理函数  $f_p$  处理后与音频样本的基本参数保持一致,噪音采用随机分布的方式进行添加以保证实验构建噪音模型的有效性. 实验平台配置为:Window10 系统、CPU i7-8750H、显卡 GTX1060、内存 16G; Keras + TensorFlow 作为深度学习框架.

3.1 信道噪音下实验结果分析

为探索 D-2-DenseNet 模型在信道噪音下的鲁棒性,选用 Dcase2016 数据集,测试 D-2-DenseNet、2-DenseNet<sup>[13]</sup>、DenseNet<sup>[14]</sup> 和 VGG<sup>[16]</sup> 模型的分类准确率,其中后 3 个模型,分别在 MFCC 和 GFCC 2 个特征下测试分类准确率,并取 2 种特征下分类准确率较高者与所提模型的分类准确率对比. VGG 模型<sup>[16]</sup>是由牛津大学的 Visual Geometry Group 提出的

一种深层卷积神经网络模型,广泛运用于深度学习的各个领域. 测试结果如表 2 所示. 分析结果如下所示:

- 1) D-2-DenseNet 在干净环境下准确率为 85. 17%,与 VGG、DenseNet 和 2-DenseNet 模型相比,分别提升了 11. 03%、5. 54%与 4. 14% ;
- 2) 随信噪比的减低(噪音干扰增强),各模型准确率也随之降低,准确率平均下降了 22. 97% ;
- 3) 在 MFCC 和 GFCC 2 种特征下,SNR 在 50 ~ 10 dB(中高信噪比)时,噪音的干扰较小,此时各模型准确率在 MFCC 特征下相对更高;而 SNR 在 0 ~ ( - 20) dB(低信噪比)时,由于噪音干扰的增强,各模型在具有更好噪音鲁棒性的 GFCC 特征下准确率相对更高;

表 2 信道噪音下各模型准确率对比 %

模型	干净	SNR/dB							
		50	40	30	20	10	0	- 10	- 20
VGG( MFCC )	74. 14	68. 62	66. 21	63. 59	61. 72	60. 34	55. 23	53. 20	45. 65
VGG( GFCC )	73. 10	66. 90	65. 14	63. 07	61. 36	59. 83	56. 12	53. 78	46. 45
DenseNet( MFCC )	79. 63	73. 10	70. 50	68. 63	66. 41	64. 17	60. 12	57. 65	47. 50
DenseNet( GFCC )	76. 63	71. 03	69. 14	67. 17	65. 52	63. 79	61. 13	58. 56	49. 49
2-DenseNet( MFCC )	81. 03	73. 79	71. 41	69. 46	67. 27	64. 83	61. 23	58. 62	48. 28
2-DenseNet( GFCC )	78. 21	71. 53	69. 71	67. 83	65. 90	64. 57	62. 13	59. 83	49. 78
D-2-DenseNet	85. 17	77. 12	74. 83	72. 41	70. 31	67. 69	64. 12	62. 13	54. 13

4) 在不同信噪比的高斯白噪音环境下,D-2-DenseNet 模型相较于 VGG、DenseNet 和 2-DenseNet 模型,平均分类准确率提高分别为 8. 51%、4. 33% 与 3. 48% .

以上研究表明,D-2-DenseNet 模型在干净环境下具有优秀的准确率,且在不同信噪比的高斯白噪音下,模型的抗噪能力更强、分类准确率更高. 综上可得,在信道噪音下 D-2-DenseNet 模型具有良好的鲁棒性.

3.2 环境噪音下实验结果分析

为探索 D-2-DenseNet 模型在环境噪音下的鲁棒性,采用 Dcase2016 数据集,环境噪音选取 Thchs30 数据集集中的 cafe( - 0. 78 dB)、car( - 17. 46 dB)噪音和 NoiseX-92 数据集集中的 babble( - 16. 40 dB)、factory( - 12. 69 dB)噪音,开展 D-2-DenseNet、2-DenseNet、DenseNet 和 VGG 模型在城市音频分类的鲁棒性研究,实验以 VGG 模型作为基线模型. 实

验结果如表 3 所示. 分析结果如下所示:

表 3 环境噪音下各模型准确率对比 %

模型	cafe	car	factory	babble
VGG	63. 79	61. 72	64. 14	64. 36
DenseNet	64. 48	62. 76	65. 86	67. 93
2-DenseNet	66. 52	64. 48	66. 21	67. 59
D-2-DenseNet	74. 83	72. 07	73. 10	75. 52

- 1) 在 Thchs30 和 NoiseX-92 噪音数据集中,D-2-DenseNet 模型的最佳分类准确率分别为 74. 83% 和 75. 52%,平均分类准确率为 73. 88% ;
- 2) D-2-DenseNet 模型与 VGG、DenseNet、2-DenseNet 模型相比,在 4 种环境噪音下,平均分类准确率分别提高了 10. 38%、8. 62%和 7. 68% .

以上研究证明,D-2-DenseNet 模型在环境噪音下的分类准确率得到了明显提升,该模型对环境噪音抗干扰能力较好,进一步验证了模型在环境噪音

下具有良好的鲁棒性。

## 4 结束语

针对噪音环境下城市音频分类系统的鲁棒性不强的问题,提出了一种双特征2阶密集卷积神经网络噪音鲁棒的城市音频分类模型。采用双特征互补自适应算法,有效融合了MFCC和GFCC2种特征的优点,且采用2阶密集卷积神经网络进行特征提取,在多通路并行的2-DenseNet中进行特征融合与模型自适应调整,模型能有效提取音频信息,降低噪音干扰,以此提高噪音鲁棒性。最后,基于Dcase2016数据集开展噪音环境下城市音频分类的测试。在信道噪音和环境噪音条件下,模型分类准确率分别可达77.12%、75.52%,与基线模型(VGG模型)相比,平均分类准确率分别提高了8.51%和10.38%。噪音环境下的分类准确率得到了明显提升,验证了模型良好的噪音鲁棒性。

## 参考文献:

- [1] 王静远,李超,熊璋,等.以数据为中心的智慧城市研究综述[J].计算机研究与发展,2014,51(2):239-259.  
Wang Jingyuan, Li Chao, Xiong Zhang, et al. Survey of data centric smart city[J]. Journal of Computer Research and Development, 2014, 51(2): 239-259.
- [2] Lim M, Lee D, Hosung P, et al. Convolutional neural network based on audio event classification[J]. KSII Transactions on Internet and Information Systems, 2018, 12: 2748-2759.
- [3] Zhang Xiaohu, Zou Yuexian, Shi Wei. Dilated convolution neural network with leaky ReLU for environmental sound classification acoustics[C]//22nd International Conference on Digital Signal Processing. Washington: IEEE Press, 2017: 123-132.
- [4] 蔡尚,金鑫,高盛翔,等.用于噪音鲁棒性语音识别的子带能量规整感知线性预测系数[J].声学学报,2012,37(6):667-672.  
Cai Shang, Jin Xin, Gao Shengxiang, et al. Sub-band power normalized perceptual linear predictive coefficient for robust automatic speech recognition[J]. ACTA Acustica, 2012, 37(6): 667-672.
- [5] Yu Tsao, Lin Tzu-Hao, Chen Fei, et al. Robust S1 and S2 heart sound recognition based on spectral restoration and multi-style training[J]. Biomedical Signal Processing and Control, 2019, 49: 173-180.
- [6] Ilyas O, Zeynep O, Oghuz F. Noise robust sound event classification with convolutional neural network[J]. Neurocomputing, 2018, 272: 505-512.
- [7] 易江燕,陶建华,刘斌,等.基于迁移学习的噪音鲁棒语音识别声学建模[J].清华大学学报,2018,58(1):55-60.  
Yi Jiangyan, Tao Jianhua, Liu Bin et al. Transfer learning for acoustic modeling of noise robust speech recognition[J]. Journal of Tsinghua University, 2018, 58(1): 55-60.
- [8] 仇景明,曲桦,赵季红.一种鲁棒网络流量分类及新类型发现算法[J].北京邮电大学学报,2020,43(2):40-45.  
Qiu Jingming, Qu Hua, Zhao Jihong. A robust network traffic classification and new type discovery algorithm[J]. Journal of Beijing University of Posts and Telecommunications, 2020, 43(2): 40-45.
- [9] Johnson H. Noise data[EB/OL]. United Kingdom Signal Processing Information Base, 1990[2020-08-16]. <http://spib.linse.ufsc.br/noise.html>.
- [10] Wang Dong, Wu Dalei, Zhu Xiaoyan. THCHS-30: a free Chinese speech corpus[EB/OL]. Beijing: Tsinghua University, 2015(2015-12-10)[2020-08-16]. <http://arxiv.org/abs/1512.01882>. 2015.
- [11] Sumithra M, Prakash D, Yuan X. Adaptive wavelet shrinkage for noise robust speaker recognition[J]. Digital Signal Processing, 2014, 33: 180-190.
- [12] 王雷,王智广.考虑多种特征因素的设计模式自动识别[J].北京邮电大学学报,2017,40(5):36-42.  
Wang Lei, Wang Zhiguang. Automatic design pattern detection on the consideration of multiple characteristic factors[J]. Journal of Beijing University of Posts and Telecommunications, 2017, 40(5): 36-42.
- [13] 曹毅,黄子龙,张威,等. N-DenseNet的城市声音事件分类模型[J].西安电子科技大学学报,2019,46(6):9-16,94.  
Cao Yi, Huang Zilong, Zhang Wei, et al. Urban sound event classification with n-order dense convolutional network[J]. Journal of Xidian University, 2019, 46(6): 9-16, 94.
- [14] Huang Gao, Liu Zhuang, Weinberger K Q. Densely connected convolutional networks[C]//30th IEEE Conference on Computer Vision and Pattern Recognition CVPR 2017. Honolulu: IEEE press, 2017: 2261-2269.
- [15] Huang Zilong, Liu Chen, Fei Hongbo, et al. Urban sound classification based on 2-order dense convolutional network using dual features[J/OL]. Applied Acoustics, 2020, 164[2020-08-16]. <https://doi.org/10.1016/j.apacoust.2020.107243>.
- [16] Mesaros A, Heittola T, Virtanen T. TUT database for acoustic scene classification and sound event detection[C]//24th European Signal Processing Conference. Budapest: IEEE Press, 2016: 3-9.