DOI: 10.13382/j.jemi.B2205628

多尺度卷积的时频域语音分离方法研究*

贾林锋 吴黎明 温腾腾 廖禹韬 高梓皓
(广东工业大学机电工程学院 广州 510006)

摘 要:在进行混合语音分离时,信号时域特征的深度学习语音分离性能优于频域特征。但目前时域特征的语音分离方法在真实噪声环境下的鲁棒性较差,且单一时域特征对分离模型的性能存在局限性。因此,提出一种基于 Conv-TasNet 网络的多特征语音分离方法,融合频域特征与时域特征,提高数据的多维信息。为了进一步提高分离网络性能,引入多尺度卷积块,提高网络对特征的提取能力。在包含真实噪声的实验环境下,所提方法与 Conv-TasNet 模型和最新的时频域融合语音分离基线模型相比,性能分别提高了 0.91 和 0.52 dB,有效提升了语音分离的性能及鲁棒性。

关键词: 语音分离;特征融合;多尺度卷积;时频域特征
中图分类号: TP391.4; TN912.3 **文献标识码**: A **国家标准学科分类代码**: 510.4

Speech separation in time-and-frequency domain based on multi-scale convolution

Jia Linfeng Wu Liming Wen Tengteng Liao Yutao Gao Zihao
(School of Electromechanical Engineering, Guangdong University of Technology, Guangzhou 510006, China)

Abstract: In mixed speech separation, the performance of signal time-domain features is better than that of frequency-domain features. However, the current speech separation methods based on time domain feature have poor robustness in real noise environment, and single time domain feature has limitations on the performance of the separation model. Therefore, a multi-feature speech separation method based on Conv-TasNet network is proposed, which integrates frequency domain features and time domain features to improve multidimensional information of data. In order to further improve the performance of separation network, multi-scale convolution block is introduced to improve the feature extraction ability of network. Compared with the Conv-TasNet model and the latest time-frequency fusion speech separation baseline model, the performance and robustness of the proposed method are improved by 0.91 and 0.52 dB respectively in the experimental environment containing real noise.

Keywords: speech separation; feature fusion; multiscale convolution; time-frequency domain characteristics

0 引 言

语音分离指在复杂声学环境下,从包含随机噪声的混合语音中分离出多个声源,常作为语音识别,音频分类等下游任务的前端[1]。语音分离按信号采集通道可分为多通道和单通道语音分离。多通道语音分离包含了声音的空间信息,通常使用多个麦克风组成的麦克风阵列采集数据。使用单个麦克风的单通道语音分离应用广泛,且从单个信道中分离出多个声源具有重要研究意义。单通道的混合语音分离方法主要从语音信号的频域和时域特征展开研究。传统的语音分离方法主要研究频域特征,如针对噪声估计的谱减法[2],从低秩角度分析的非负矩阵分解[3]以及模拟人类听觉的计算听觉场景分析[4]。近年来,深度学习与语音分离的结合超越了传统语音分离方法[5],但仍存在分离语音排序不定和分离数量未知的问题[6]。针对上述问题,深度聚类[7]、话语排列不变训练(utterance-level permutation invariant training, uPIT)[8]、深度吸引网络[9]从频域角度分析并解决。从自然语言处理等领域对时序信号的进一步研究促进了语音时域分离

收稿日期: 2022-06-27 Received Date: 2022-06-27
*基金项目:国家自然科学基金(61705045)、佛山广工大研究院创新创业人才团队计划项目(20191108)资助

的探索,基于 LSTM 的 TasNet[10] 时域语音分离方法在分离性能上超过了基于深度学习的频域语音分离方法。由于 LSTM 的参数量大及语音分离的精度不一致,Luo 等[11]提出了基于空洞时序卷积的 Conv-TasNet 网络,使用堆叠的卷积块代替 LSTM 学习语音数据的长序列特征。随后,针对语音序列信息的研究,Luo 等[12]提出的双路径循环神经网络(dual-path RNN,DPRNN)和 Subakan 等[13]提出的 SepFormer 都在语音分离性能上有较大的提升,但也提高了语音分离模型的训练难度和模型参数。

对于单个特征的深度学习语音信号分离任务,时域语音分离虽然在性能上优于频域语音分离,但分离带有背景噪声的混合语音信号仍表现出鲁棒性差的问题。因此,在单个时域特征的语音分离任务中结合其他语音特征成为解决模型鲁棒性问题的关键。Yang 等[14]提出基于 TasNet 的时域频域特征融合语音分离网络,通过短时傅里叶变换提取语音频域特征与时域特征拼接,结合聚类算法实现语音分离,验证了时频域特征结合可有效提高语音分离的性能和鲁棒性。Lan 等[15]基于 Conv-TasNet 改进 Yang 的时频域语音分离,验证了不同时频域特征融合机制对语音分离性能的影响,利用不同特征的全局信息进一步提高了时频域结合语音分离的性能。上述工作在纯时域语音分离模型中引入频域特征,尝试解决模型鲁棒性并探讨多特征融合机制。然而,无论是时域和时频域特征,模型的最终性能受制于纯时域网络中堆叠的空洞卷积块结构,其次通过快速傅里叶变换提取的频域特征在时域结合前缺乏进一步的特征提取机制,不能有效利用频域特征的有效信息。

本文提出多尺度卷积的时频域语音分离模型,利用 Conv-TasNet 在分离性能及模型参数上的优势,创新地提出一种频域语音特征与时域语音特征的融合方法,提高不同特征融合后对语音分离任务性能。针对 Conv-TasNet 网络结构中的分离模块对网络性能的局限性,提出改进的多尺度空洞时序卷积块,在整合上下文信息的同时,提高语音数据局部特征的信息权重,有效提取融合特征的高维特征。最后,通过纯语音分离和包含背景噪声的分离鲁棒性实验,验证了本文模型比基线混合语音分离模型具有更好的性能和噪声环境下的语音分离鲁棒性,且提出的时频域融合方法优于最新的时频域融合基线方法。

1　多尺度时频语音分离方法

1.1　数学模型和网络结构

语音的时域信号是时间长度为 T 的一维数据。多人混合语音是指在同一时间段,多个说话人语音重叠,因此得出混合语音的数学表达式:

$$x(t)=\sum_{i=1}^{N}s_i(t) \tag{1}$$

其中,$s_i(t)\in \mathrm{R}^{1\times T}$ 表示混合音乐中每个说话人的干净语音信号。N 为说话人数量,$x(t)$ 表示所有说话人语音叠加的混合语音信号。单通道时域语音分离的目标是从混合语音 $x(t)$ 的时域信号分离出 N 个说话人的语音信号 $s(t)$ 并使用基于掩码矩阵的分离方法。掩码矩阵作为单个说话人在混合语音中的基音权重矩阵,描述了说话人语音在混合语音中的数据分布。Conv-TasNet 网络基于说话人在混合语音中的掩码差异,通过堆叠卷积提取语音特征的掩码矩阵,获取分离后的语音数据,在时域纯语音分离任务中效果显著。

基于多尺度卷积的时频特征语音分离网络结构如图 1 所示。混合语音通过编码器(encode)的时域卷积编码和频域特征提取获得分离网络的两种输入特征。两种语音特征融合并输入分离模块(separator),经过 24 个堆叠时序卷积块特征提取后使用 ReLU 激活函数提高特征非线性。基于说话人数量,激活后的特征通过一维卷积提高特征维度,转换成与说话人数量匹配的语音掩码(mask)。将编码部分输出的特征与语音掩码相乘,获得分离后的语音特征。语音特征通过解码器(Decoder)的反卷积网络重建多个说话人的干净语音信号。

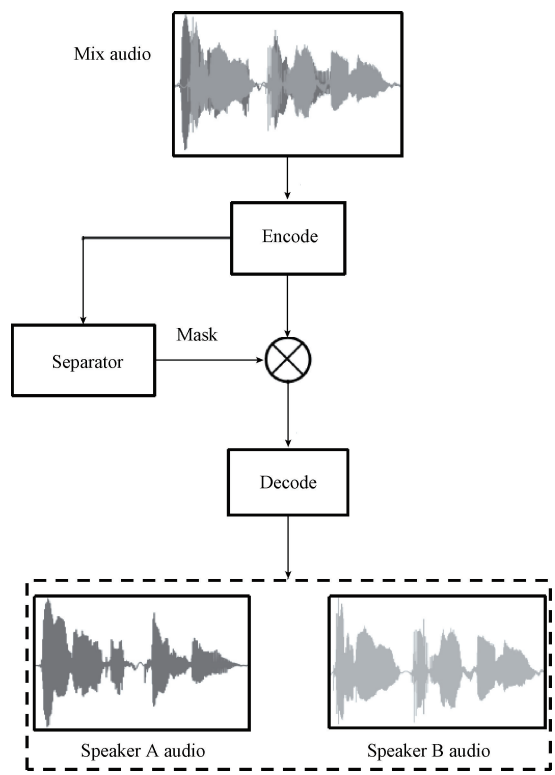

图 1　语音分离网络

Fig. 1　Speech separation network

1.2　语音特征提取与融合

对于时域特征，时域编码器（time_encode）采用一维卷积替代传统的短时傅里叶变换提取语音特征。使用多个固定尺寸的一维卷积核以固定步长滑动卷积，获取混合语音信号的多维时域特征，通过归一化处理，调整数据分布。采用短时傅里叶变换（short-time Fourier transform，STFT）提取频域特征，使用针对频域特征的卷积编码（stft_encode）转换特征维度并归一化，使频域特征和时域特征维度相同，实现融合的特征维度匹配。短时傅里叶变换的窗长和步长决定频域特征长度，影响频域特征的时间分辨率和频率分辨率。设定窗长和步长与时域编码器的卷积参数统一，使频域特征长度上与时域特征长度相近，利于在融合前提高频率特征上采样插值的数据平滑性，但同时也降低了频域特征的频率分辨率。因此使用固定尺寸的时序卷积块（temporal convolutional network，TCN）处理短时傅里叶变换后的频域特征，通过可学习的卷积网络提高频域特征的频率分辨率信息。

采用中期融合机制处理时域信息流和频域信息流。频域特征通过上采样近邻插值方法（upsample）使特征长度与时域特征长度匹配。双特征在对应维度上相加，形成融合特征，使用批归一化[16]统一特征数据分布。特征融合计算公式如下：

$$y = P_{norm}(S_{add}[f_{up}(x_{a\_stft}), x_{a\_time}]) \tag{2}$$

f_{up} 为线性插值的上采样函数。S_{add} 将时域特征流和频域特征流在特征纬度上相加。P_{norm} 为批归一化处理，对每一个训练批次内的融合特征维度进行归一化处理。特征提取及融合网络如图2所示。

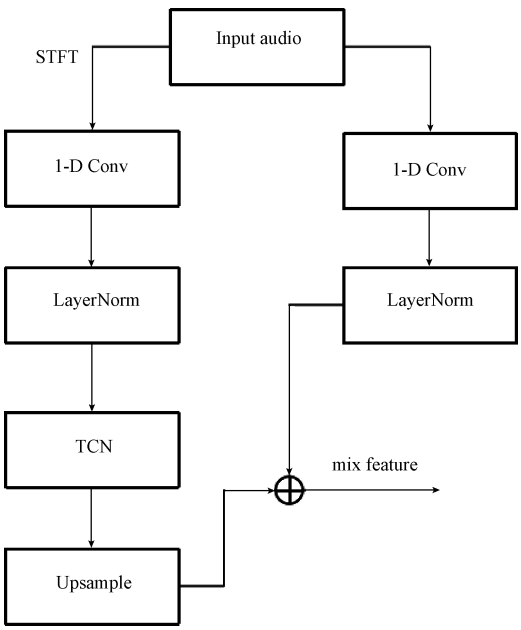

图2　特征提取及融合网络

Fig. 2　Feature extraction and fusion network

1.3　分离网络

编码器提取的高维语音特征通过分离模块进一步提取时域信号的序列信息。分离模块采用堆叠的TCN提取语音信号的序列信息。二维空洞卷积常用于视觉图像领域[17-19]，在不增加计算量的前提下扩大感受野，并通过多种不同尺度的卷积块获取多维信息。语音作为长时间序列特征，计算当前采样点与全局相关信息能提高语音分离的性能。因此，使用一维空洞卷积适用于对混合语音提取特征。

在原始的分离模块中包含了空洞指数不同的TCN模块，使得模型能够学习长序列的上下文信息。随着空洞指数增加，网络提高了对全局特征的学习，但也弱化了局部特征的重要性。为了提高分离模块对时频域特征的学习能力，提出一种多尺度卷积融合的TCN。如图3所示，输入特征经过一维卷积和PReLU激活函数提高特征维度并通过GLN（global layer normalization）进行数据归一化处理。处理后的特征分为两路，一路通过卷积核尺寸为3的一维卷积（1×3 Conv）提取局部特征，另一路通过一维空洞卷积（dilated Conv）提取长序列特征。两路特征相加融合后进行归一化处理。使用激活函数PReLU和GLN处理融合特征。TCN包含跳跃输出（Skip）和残差输出（Res）并通过卷积网络使输出特征维度与输入特征维度一致。所有堆叠块的跳跃输出总和作为分离模块的输出。残差输出将多尺度卷积计算的特征与输入特征相加后作为后续堆叠块的输入，融合特征多维度信息。

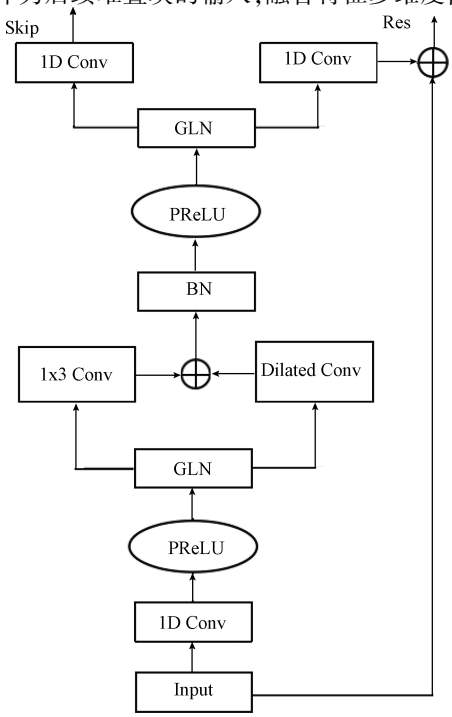

图3　多尺度TCN模块

Fig. 3　Multi-scale TCN module

2　模型验证和分析

2.1　音视频数据集构建

使用开源数据集 GRID[20] 作为模型实验数据集。GRID 数据集包含完整音视频语音数据的说话人有 33 名,其中有 18 名男性和 15 名女性。每个说话人记录了 1 000 条采样率 50 kHz,长度 3 s 的语音数据。随机选取 3 名男性和 3 名女性作为混合语音验证集数据源对象,另外选取 3 名男性和 3 名女性作为混合语音测试集数据源对象,其余的说话人作为混合语音训练集数据源对象。为了使 3 个子集的源对象产生对应的混合语音数据集,在各自子集内随机选择两个不同的说话人的随机语音数据,使用 8 kHz 重采样语音数据,以-5~5 dB 之间的随机信噪比(signal-to-noise ratio,SNR)混合两个语音数据。记录混合语音数据时长,构建 30 h 的混合语音数据集,3 h 的混合语音验证集和 3 h 的混合语音测试集。

GRID 数据集中的语音数据是在单一安静环境下采集的,但在实际语音分离场景中存在不确定的环境噪声。因此,使用开源数据 TUT2016[21] 声学场景数据集构建包含随机噪声的混合语音数据集,作为模型鲁棒性的测试数据集。TUT2016 声学场景数据集包含办公室、餐厅、公园、汽车等 15 个不同真实环境下采集的背景声音,每个场景包含 78 个语音片段。从 TUT2016 数据集中随机选取真实环境语音数据与 GRID 的 3 h 测试数据集按照 AVSpeech[22] 的混合比例生成带环境噪声的混合语音数据。TUT2016 数据集的原始采样率为 44.1 kHz,需重采样到 8 kHz,与 GRID 测试数据集采样率保持一致。

2.2　实验设计和模型参数

为验证多尺度卷积结合时频域融合对纯语音分离的提升效果,使用相同的数据集在本文模型和基线模型 Conv-TasNet 模型上进行了训练。为验证时频域融合方法高效性,对比了近期由 Lan 提出的 GCN 时频特征融合方法。3 个模型设置的超参数相同。

模型的网络结构部分超参数如表 1 所示。频域特征提取使用基于 Python 环境的 Librosa 工具包,计算输入语音的短时傅里叶变换特征。窗函数为汉宁窗,窗函数大小为 16,步长为 8。输出的语音频率特征维度为 9。模型的训练数据长度 3 s,语音采样率为 8 kHz。模型初始学习率设置为 10^{-3},最大训练周期为 100。为了选择最优模型且防止模型过拟合,使用提前停止的模型训练策略。验证集损失值在完成 3 次连续的训练周期后没有提升,模型学习率减半。验证集损失值在模型连续 10 次训练周期后未提升,则提前结束训练。每一个训练周期的数据批大小为 8。使用 Adam[23] 算法优化整个模型。采用尺度不变的信噪比(Si-SNR)[24] 作为训练目标函数以及评估函数,其定义为:

$$\text{starget} = \frac{\langle s_1, s\rangle s}{\| s \|^2} \tag{3}$$

$$\text{noise} = s_1 - \text{starget} \tag{4}$$

$$\text{Si-SNR} = 10\lg\left(\frac{\| \text{starget} \|}{\| \text{noise} \|}\right)^2 \tag{5}$$

式中:s_1 表示语音分离模型中分离出的说话人语音数据,s 表示说话人的干净语音数据。通过 uPIT[8] 计算出分离语音与标签语音损失函数最小的组合,实现分离语音与说话人的身份配对。

表 1　模型主要超参数

Table 1　The main hyperparameters of the model

参数	参数解释	参数值
N	编码器卷积核输出通道数	512
L	编码器卷积核尺寸	16
B	堆叠卷积块输出通道数	128
H	空洞卷积输入通道数	512
P	深度空洞卷积核尺寸	3
aX	时域深度卷积空洞指数	8
mX	频域深度卷积空洞指数	1
R	时域堆叠重复次数	3

2.3　实验结果及分析

选取 3 种方案的最优模型,在测试阶段使用干净语音数据测试集和包含随机环境背景声的带噪语音测试集验证实验结果。对比实验结果如表 2 所示。

表 2　对比实验结果

Table 2　Comparative experimental results

模型	测试集 Si-SNR	噪声 Si-SNR
Conv-TasNet	13.10	8.77
GCN	13.33	9.16
多尺度卷积+时频域语音分离	13.67	9.68

验证集的部分损失函数变化(loss)如图 4 所示,多特征多尺度语音分离算法在无背景噪声环境和随机背景噪声环境下,分别提升了 0.57 和 0.91 dB,比纯时域特征语音分离基线模型的性能更好。对比现有的时频域融合处理算法,在两个环境下分别提高了 0.34 和 0.52 dB,验证时频域特征融合策略优于最新的融合方法。从 3 个模型在验证集的 loss 曲线可以看出,多特征尺度语音分离算法在整个训练过程中,模型性能整体优于两个对比模型。

2.4　消融实验设计及结果分析

为了验证在 Conv-TasNet 基线模型上提出的多尺度

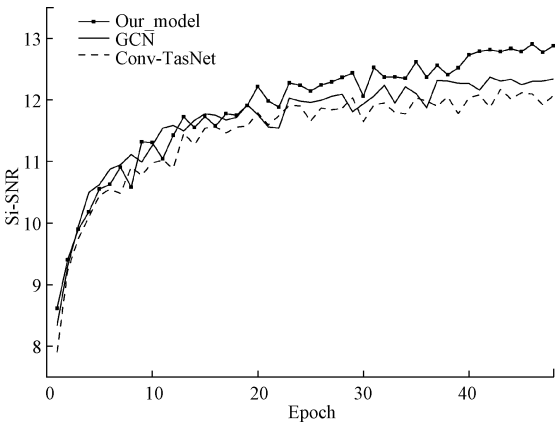

图 4 验证集 loss 变化曲线

Fig. 4 Verification set loss change curve

卷积和多特征融合方法分别对模型性能的影响,设计了模型的消融实验。在同等环境下分别训练了基于 Conv-TasNet 的多尺度卷积模型和时频域模型。同时,引入对数梅尔频谱特征(log-mel),对比 STFT 特征与其他频域特征对基于多尺度卷积的时频语音分离模型性能的影响,实验结果如表 3 所示,消融实验在验证集的部分 loss 曲线如图 5 所示。

表 3 消融实验对比结果

Table 3 Comparison of ablation results

消融变量	测试集 Si-SNR	噪声 Si-SNR
多尺度卷积基线模型(multi_dilated)	13.62	9.48
时频域模型(stft)	13.36	9.53
对数梅尔频谱特征(log-mel)	13.70	9.51

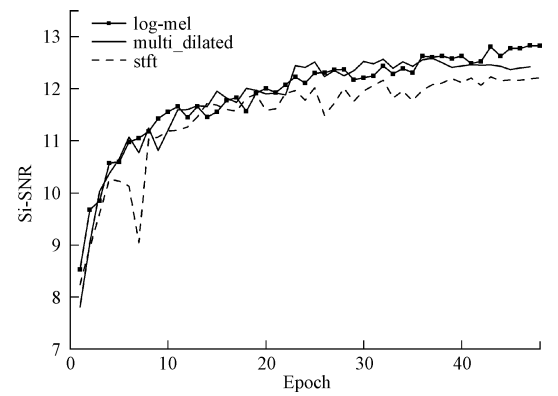

图 5 消融验证集 loss 变化曲线

Fig. 5 Loss curve of ablation validation set

根据消融实验结果,多尺度卷积对纯时域网络的性能提升显著。因为多尺度卷积在基线网络结构上添加了局部特征的采样,使膨胀率大的空洞时序卷积同时获取了采样点局部的特征和长序列特征。对于只添加了频域特征的语音分离网络,在两种环境下都有一定程度的性能提升。结合多尺度卷积对局部特征的重视,通过融合两种方式取得了较大的分离性能提升。对数梅尔频谱是引入具有人声特性的梅尔刻度参与特征提取的语音频域特征,在纯人声的测试集环境中比 STFT 特征的效果稍好,但在噪声环境下的鲁棒性不如 STFT 特征的语音分离模型。

3 结 论

本文提出一种基于多尺度卷积的时频域语音分离方法,解决时域分离网络在噪声环境下鲁棒性差和现有时频域语音分离方法性能不足的问题。在频域特征与时域特征融合前对频域特征进一步地提高时间序列的信息相关性,提高分离模型的鲁棒性。结合多尺度卷积分离模块,使网络同时提取语音长序列信息和局部信息并注重局部特征学习。通过与时域和时频域融合基线模型在噪声环境下的语音分离实验对比,分离评价指标分别提高了 0.91 和 0.52 dB,验证了多尺度卷积的时频域语音分离具有更高的语音分离性能和鲁棒性。

参考文献

[1] 张盛,杨剑鸣. 一种面向自组织麦克风网络的多通道语音分离方法[J]. 信号处理, 2021, 37(5): 757-762.
ZHANG SH, YANG J M. A multichannel speech separation method for ad-hoc microphones[J]. Journal of Signal Processing, 2021, 37(5): 757-762.

[2] 吴礼福,申浩. 掩蔽法减少谱减法去混响中的音乐噪声[J]. 电子测量与仪器学报, 2017, 31(11): 1855-1859.
WU L F, SHEN H. Reducing musical noise in dereverberation of spectral subtraction based on masking method[J]. Journal of Electronic Measurement and Instrumentation, 2017, 31(11): 1855-1859.

[3] MOHAMMADIHA N, SMARAGDIS P, LEIJON A. Supervised and unsupervised speech enhancement using nonnegative matrix factorization[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2013, 21(10): 2140-2151.

[4] WANG D, BROWN G. Computational auditory scene analysis: principles, algorithms and applications[J]. IEEE Transactions on Neural Networks, 2008, 19(1): 199-199.

[5] 徐亮,王晶,杨文镜,等. 基于 Conv-TasNet 的多特征融合音视频联合语音分离算法[J]. 信号处理, 2021, 37(10): 1799-1805.

XU L, WANG J, YANG W J, et al. LUO Yiyu multi feature fusion audio-visual joint speech separation algorithm based on Conv-TasNet[J]. Journal of Signal Processing, 2021, 37(10): 1799-1805.

[6] 黄雅婷,石晶,许家铭,等. 鸡尾酒会问题与相关听觉模型的研究现状与展望[J]. 自动化学报, 2019, 45(2): 234-251.
HUANG Y T, SHI J, XU J M, et al. Research advances and perspectives on the cocktail party problem and related auditory models[J]. Acta Automatica Sinica, 2019, 45(2): 234-251.

[7] HERSHEY J R, CHEN Z, ROUX L J, et al. Deep clustering: Discriminative embeddings for segmentation and separation[J]. 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016: 31-35.

[8] KOLBÆK M, YU D, TAN Z H, et al. Multitalker speech separation with utterance-level permutation invariant training of deep recurrent neural networks[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2017, 25(10): 1901-1913.

[9] CHEN Z, LUO Y, MESGARANI N. Deep attractor network for single-microphone speaker separation[C]. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2017: 246-250.

[10] LUO Y, MESGARANI N. Tasnet: Time-domain audio separation network for real-time, single-channel speech separation[C]. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018: 696-700.

[11] LUO Y, MESGARANI N. Conv-tasnet: Surpassing ideal time-frequency magnitude masking for speech separation[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2019, 27(8): 1256-1266.

[12] LUO Y, CHEN Z, YOSHIOKA T. Dual-path RNN: Efficient long sequence modeling for time-domain single-channel speech separation[C]. ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020.

[13] SUBAKAN C, RAVANELLI M, CORNELL S, et al. Attention is all you need in speech separation[C]. ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2021.

[14] YANG G P, TUAN C I, LEE H Y, et al. Improved Speech Separation with Time-and-Frequency Cross-Domain Joint Embedding and Clustering:, 10.21437/Interspeech.2019-2181[P]. 2019.

[15] LAN T, QIAN Y, LYU Y, et al. Improved speech separation with time-and-frequency cross-domain feature selection[C]. Interspeech, 2021: 3525-3529.

[16] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]. International Conference on Machine Learning, PMLR, 2015: 448-456.

[17] LI C, QIU Z, CAO X, et al. Hybrid dilated convolution with multi-scale residual fusion network for hyperspectral image classification[J]. Micromachines, 2021, 12(5): 545.

[18] 刘琛,王江涛,王明阳. 引入视觉机制的 SSD 网络在摩托车头盔佩戴检测中的应用[J]. 电子测量与仪器学报, 2021, 35(3): 144-151.
LIU CH, WANG J T, WANG M Y. Application of SSD network with visual mechanism in motorcycle helmet wearing detection[J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(3): 144-151.

[19] 何晓云,许江淳,陈文绪. 基于改进 U-Net 网络的眼底血管图像分割研究[J]. 电子测量与仪器学报, 2021, 35(10): 202-208.
HE X Y, XU J CH, CHEN W X. Research on fundus blood vessel image segmentation based on improved U-Net network[J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(10): 202-208.

[20] COOKE M, BARKER J, CUNNINGHAM S, et al. An audio-visual corpus for speech perception and automatic speech recognition[J]. The Journal of the Acoustical Society of America, 2006, 120(5): 2421-2424.

[21] MESAROS A, VIRTANEN T, FAGERLUND E, et al. TUT acoustic scenes 2016 [C]. Development Dataset, 2016.

[22] EPHRAT A, MOSSERI I, LANG O, et al. Looking to listen at the cocktail party[J]. Acm Transactions on Graphics (TOG), 2018, 37: 1-11.

[23] KINGMA D P, BA J. Adam: A method for stochastic optimization[C]. International Conference on Learning Representations. Ithaca, NYarXiv.org, 2014.

[24] ISIK Y, ROUX J L, CHEN Z, et al. Single-channel multi-speaker separation using deep clustering[J]. arXiv preprint arXiv:1607.02173, 2016.

作者简介

贾林锋,2019 年于广州理工学院获得学士学位,现为广东工业大学在读硕士研究生,主要研究方向为智能语音处理、多模态感知。

E-mail:846526132@qq.com

Jia Linfeng received his B. Sc. degree from Guangzhou Institute of Science and Technology in 2019. Now he is a M. Sc. candidate at Guangdong University of Technology. His main research interests include intelligent speech processing and multi-modal sensing.

吴黎明(通信作者),2004 年于华南理工大学获得硕士学位,现为广东工业大学教授,主要研究方向为信号处理、智能感知。

E-mail:jkyjs@gdut.edu.cn

Wu Liming (Corresponding author) received his M. Sc. degree from South China University in 2004. Now he is a professor at Guangdong University of Technology. His main research interests include signal processing and intelligent sensing.