

# 利用生成噪声提高语音增强方法的泛化能力

袁文浩, 娄迎曦, 梁春燕, 夏 斌

(山东理工大学计算机科学与技术学院, 山东淄博 255000)

**摘 要:** 如何提高对未知噪声类型的泛化能力是有监督语音增强方法中亟待解决的重要问题, 通过对大量不同类型噪声进行建模, 深度神经网络成为了解决该问题的有效手段. 为了进一步提高基于深度神经网络的语音增强方法的泛化能力, 本文基于生成式对抗网络(Generative Adversarial Networks, GAN)设计了能够由真实噪声数据生成新的噪声类型的 NoiseGAN; 通过在训练集中增加生成噪声类型, 提高训练集噪声类型的多样性, 从而达到提高语音增强模型泛化能力的目的. 不同结构的网络下的语音增强实验结果表明, 本文提出的 NoiseGAN 能够生成新的噪声类型, 具备提高训练集噪声类型多样性的能力, 有效提高了语音增强模型在未知噪声类型下的泛化能力.

**关键词:** 语音增强; 生成式对抗网络; 泛化能力; 深度神经网络

**中图分类号:** TN912.3      **文献标识码:** A      **文章编号:** 0372-2112 (2019)04-0791-07

**电子学报 URL:** <http://www.ejournal.org.cn>

**DOI:** 10.3969/j.issn.0372-2112.2019.04.004

## Improving Generalization Ability of Speech Enhancement Approaches Using Generated Noise

YUAN Wen-hao, LOU Ying-xi, LIANG Chun-yan, XIA Bin

(College of Computer Science and Technology, Shandong University of Technology, Zibo, Shandong 255000, China)

**Abstract:** How to improve the generalization ability to unknown noise types is an important problem to be solved urgently in supervised speech enhancement approaches. By modeling a large number of types of noise, the deep neural network (DNN) becomes an effective way to solve this problem. In order to further improve the generalization ability of speech enhancement approaches based on DNN, this paper designs NoiseGAN based on Generative Adversarial Networks (GAN) to generate new noise types from real noise data. By adding generated noise to training set, the diversity of noise types in training set is increased, and thereby the generalization ability of speech enhancement model is improved. The results of speech enhancement experiments under different structures of networks show that the proposed NoiseGAN can generate new noise types, increase the diversity of noise types in training set, and effectively improve the generalization ability of speech enhancement models under unknown noise types.

**Key words:** speech enhancement; generative adversarial networks; generalization ability; deep neural network

### 1 引言

语音增强是噪声环境下语音信号处理的必要环节<sup>[1]</sup>. 传统语音增强方法一般基于语音和噪声统计特性的某种假设, 限制了其对具有不同统计特性的噪声的适用性, 不能有效跟踪非平稳噪声的统计特性并进行处理, 无法适用于复杂多变的噪声环境<sup>[2,3]</sup>. 有监督语音增强方法通过训练学习得到更多关于语音和噪声特性的先验信息, 提高了对非平稳噪声的处理能力<sup>[4]</sup>. 然而, 与有监督学习方法的其他应用场景一样, 制约有

监督语音增强方法应用的主要问题是其对未经训练的新噪声类型的泛化能力.

近几年, 随着深度学习技术的发展, 基于深度神经网络的语音增强成为了解决该问题的有效途径, 通过建立包含大量不同类型噪声信息的模型, 基于深度神经网络的语音增强方法显著提升了未知噪声下的语音增强性能<sup>[5,6]</sup>. 由于深度神经网络本身具有强大的学习能力, 因此提高基于深度神经网络的语音增强方法泛化能力的最直接方法, 是提高训练集中噪声类型的多样性. 通过采用包含 104 类和 115 类不同噪声的训练

集,文献[7]和文献[8]训练的 DNN 提高了对未知噪声的处理能力;文献[9]与文献[10]通过对 10000 种不同类型噪声建立 DNN 模型来提高对未知噪声的泛化能力,主客观实验结果表明采用包含更多噪声类型的训练集能显著提高未知噪声下的语音可懂度.然而,真实噪声数据的现场收集,过程非常复杂且成本较高.因此,如何通过分析已有噪声数据的特征分布,来人为构造与生成更多类型的噪声,具有重要的研究意义.例如,与直接增加训练集噪声类型数量的方法不同,文献[11]用对有限种类的噪声施加不同的扰动项的方式来提高噪声特性的多样性,实验结果表明该方法同样能有效提高未知噪声下的语音增强性能,证明了通过生成噪声来提高泛化能力的可行性;但是,该方法过分依赖于经验公式和经验参数的设置,很难推广到噪声种类较多的训练集中.

近年来,生成式对抗网络在图像生成中表现出了很好的效果,通过在训练过程中不断调整生成器 G 和判别器 D, GAN 能够对真实数据样本的分布规律进行估计,并依据此分布生成新的样本<sup>[12]</sup>.受 GAN 在图像生成中的成功应用启发,我们尝试运用 GAN 来生成新的噪声,以此提高训练集噪声类型的多样性,从而达到提高语音增强方法对未经训练噪声类型泛化能力的目的.

## 2 NoiseGAN 模型

语音或噪声信号经过短时傅里叶变换 (Short-Time Fourier Transform, STFT) 后,在二维时频域,其相邻帧和相邻频带间的相关性,与图像中相邻像素间的相关性非常类似.因此,通过将真实噪声的短时傅里叶变换后的时频域特征作为真实样本数据,借鉴在图像生成中具有良好效果的 DCGAN (Deep Convolutional Generative Adversarial Networks)<sup>[13]</sup>,我们构造了能够生成不同类型噪声的 NoiseGAN,如图 1 所示.

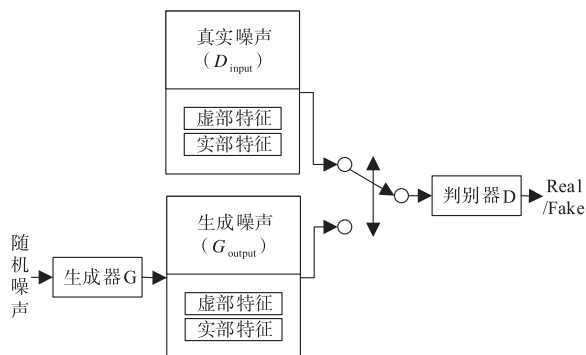


图1 NoiseGAN结构图

### 2.1 输入与输出

用于表示语音和噪声信号的常见时频域特征是幅度谱和对数谱,但是由于相位信息的缺失,使用信号的

幅度谱或对数谱并不能直接重构时域信号.因此,为了方便由生成噪声的时频域特征重构其时域信号,本文直接使用真实噪声信号短时傅里叶变换后的虚部和实部作为真实样本输入.设真实噪声的时域信号为  $n$ ,其在  $l$  帧的短时傅里叶变换形式为

$$N(k, l) = N_r(k, l) + j \cdot N_i(k, l) \quad (1)$$

其中,  $k = 1, 2, \dots, K$  是频带序号,  $N_r(k, l)$  和  $N_i(k, l)$  分别是实部和虚部分量.类似于图像中的 RGB 三个通道,将实部和虚部特征矩阵作为两个通道输入判别器 D,

$$D_{\text{input}}(:, :, 1) = \begin{bmatrix} N_r(1,1) & N_r(1,2) & \cdots & N_r(1,L) \\ N_r(2,1) & N_r(2,2) & \cdots & N_r(2,L) \\ \vdots & \vdots & \vdots & \vdots \\ N_r(K,1) & N_r(K,2) & \cdots & N_r(K,L) \end{bmatrix} \quad (2)$$

$$D_{\text{input}}(:, :, 2) = \begin{bmatrix} N_i(1,1) & N_i(1,2) & \cdots & N_i(1,L) \\ N_i(2,1) & N_i(2,2) & \cdots & N_i(2,L) \\ \vdots & \vdots & \vdots & \vdots \\ N_i(K,1) & N_i(K,2) & \cdots & N_i(K,L) \end{bmatrix} \quad (3)$$

其中,  $L$  是以帧数表示的噪声段长度.这样,生成器 G 的输出将是生成噪声的实部和虚部,

$$G_{\text{output}}(:, :, 1) = \begin{bmatrix} \hat{N}_r(1,1) & \hat{N}_r(1,2) & \cdots & \hat{N}_r(1,L) \\ \hat{N}_r(2,1) & \hat{N}_r(2,2) & \cdots & \hat{N}_r(2,L) \\ \vdots & \vdots & \vdots & \vdots \\ \hat{N}_r(K,1) & \hat{N}_r(K,2) & \cdots & \hat{N}_r(K,L) \end{bmatrix} \quad (4)$$

$$G_{\text{output}}(:, :, 2) = \begin{bmatrix} \hat{N}_i(1,1) & \hat{N}_i(1,2) & \cdots & \hat{N}_i(1,L) \\ \hat{N}_i(2,1) & \hat{N}_i(2,2) & \cdots & \hat{N}_i(2,L) \\ \vdots & \vdots & \vdots & \vdots \\ \hat{N}_i(K,1) & \hat{N}_i(K,2) & \cdots & \hat{N}_i(K,L) \end{bmatrix} \quad (5)$$

在得到生成噪声第  $l$  帧的实部与虚部后,第  $l$  帧的生成噪声的时域信号可以通过短时傅里叶逆变换 (Inverse STFT, ISTFT) 得到

$$\hat{n}(l) = \text{ISTFT}(\hat{N}_r(k, l) + j \cdot \hat{N}_i(k, l)) \quad (6)$$

另外,为了进一步增加生成噪声的多样性,可以先将部分生成噪声的实部和虚部的频带进行倒置,即

$$\hat{M}_r(k, l) = \hat{N}_r(K+1-k, l) \quad (7)$$

$$\hat{M}_i(k, l) = \hat{N}_i(K+1-k, l) \quad (8)$$

然后再重构生成噪声的时域信号

$$\hat{n}(l) = \text{ISTFT}(\hat{M}_r(k, l) + j \cdot \hat{M}_i(k, l)) \quad (9)$$

### 2.2 真实噪声数据

为了得到真实噪声样本数据,我们将来自俄亥俄州立大学 Perception and Neurodynamics 实验室的 100 类

噪声的采样频率均转换为  $8\text{kHz}$ <sup>[14]</sup>, 然后采用短时傅里叶变换计算其时频域特征, 其中帧长设为  $32\text{ms}$  (256 点), 帧移设为  $16\text{ms}$  (128 点), 得到的实部和虚部特征维度均为 129. 由于这 100 类噪声所对应的噪声段具有不同的长度, 其中最短的噪声段在短时傅里叶变换后的帧长为  $L = 50$ , 因此为了适应网络的输入和输出, 我们将所有噪声特征的帧长都截取为  $L = 50$ , 相应的生成器 G 的输出和判别器 D 的输入维度均为  $50 \times 129 \times 2$ .

### 2.3 生成器 G 与判别器 D

依据网络输入、输出的特征维度, 参考 DCGAN 的网络结构设计 NoiseGAN, 其中生成器 G 的输入是 100 维的随机噪声, 经过全连接层和 reshape 后得到  $7 \times 17 \times 256$  的特征图, 再经过三次 Batch Normalization (BN)、

非线性激活 (ReLU) 和反卷积 (deconv) 的组合操作后得到 2 层  $50 \times 129$  的输出, 分别对应 50 帧生成噪声的实部和虚部特征. 生成器 G 的具体结构如图 2 所示, 其中三次反卷积的滤波器个数分别为 128、64 和 2, 滤波器的大小均为  $5 \times 5$ , 卷积步长均为 2.

判别器 D 的输入是 50 帧的真实噪声或生成噪声的实部和虚部特征, 经过三次卷积 (conv)、Batch Normalization (BN) 和非线性激活 (LeakyReLU) 的组合操作后得到 32 层  $7 \times 17$  的特征图, 再通过一个全连接层和一个非线性激活 (Sigmoid) 得到对噪声类型真假的判断. 判别器 D 的具体结构如图 3 所示, 其中卷积滤波器的个数分别为 8、16 和 32, 滤波器的大小均为  $5 \times 5$ , 卷积步长均为 2.

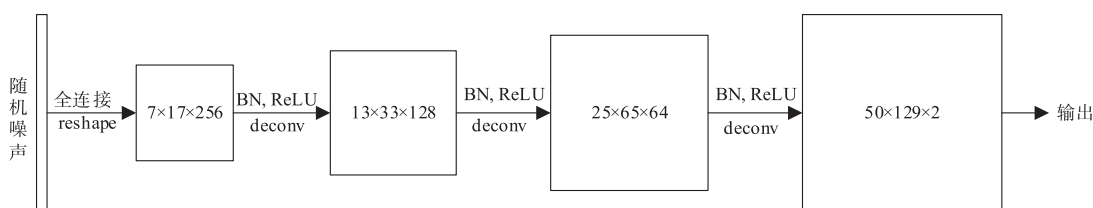


图2 生成器G结构图

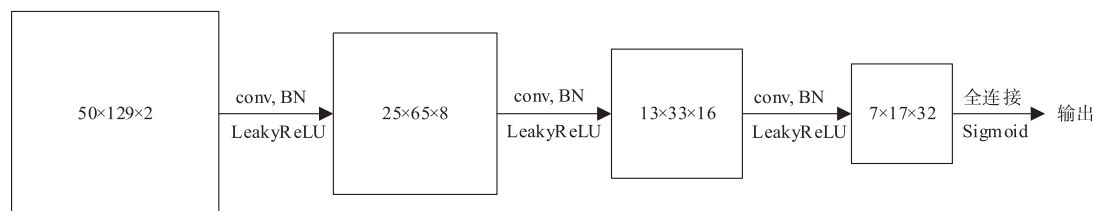


图3 判别器D结构图

### 2.4 生成噪声样例

图 4 给出了 NoiseGAN 迭代 200 次后, 由生成器 G 生成的两段噪声的语谱图, 可以看到, 两段噪声在时频域具有截然不同的能量分布特性, 证明我们设计的 NoiseGAN 确实具有生成不同类型噪声的能力.

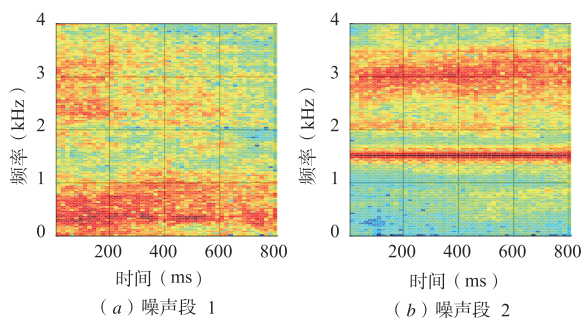


图4 NoiseGAN生成噪声的语谱图

## 3 语音增强实验与结果分析

### 3.1 实验配置

为了更加客观的测试生成噪声对训练集噪声类型

多样性和语音增强方法泛化能力的影响, 我们设计了三个包含不同数量噪声类型的训练集, 训练集中所采用的纯净语音来自 TIMIT 语音数据库的 Training 集<sup>[15]</sup>, 所用的噪声数据来自俄亥俄州立大学 Perception and Neurodynamics 实验室的 100 类真实噪声, 以及由 NoiseGAN 得到的 50 段生成噪声, 其中由式 (6) 和式 (9) 生成的噪声各占一半. 三个训练集采用不同的噪声数据集成, 其中, 训练集 1 由 50 类真实噪声合成, 训练集 2 由 100 类真实噪声合成, 训练集 3 由 100 类真实噪声和 50 类生成噪声合成. 三个训练集均包含 50000 段含噪语音 (约 40 小时), 每段含噪语音采用如下方法合成: 从 TIMIT 语音库的 Training 集的 4620 段纯净语音中随机选取 1 段, 并从 50/100/150 类噪声中随机选取 1 类, 然后将该类噪声的随机截取片段按照  $-10\text{dB}$ 、 $-5\text{dB}$ 、 $0\text{dB}$ 、 $5\text{dB}$  和  $10\text{dB}$  5 种全局信噪比中的随机 1 种混入语音中.

测试集中所采用的纯净语音来自 TIMIT 语音库的 Core test 集, 噪声数据则来自 Noisex92 噪声库<sup>[16]</sup>, 为了检验语音增强方法的泛化能力, 测试集中的 8 类噪声

是与训练集噪声完全不同的未知噪声,分别是 Factory1、Factory2、Babble、Buccaneer1、Buccaneer2、Destroyer engine、Destroyer operations、HF channel 噪声. 将 Core test 集的 192 段语音与 8 类噪声的随机截取片段分别按照  $-5\text{dB}$ 、 $0\text{dB}$  和  $5\text{dB}$  的全局信噪比进行混合,得到共包含 4608 ( $192 \times 3 \times 8$ ) 段含噪语音的测试集.

为了更加准确地评估不同训练集对语音增强模型泛化能力的影响,我们设计了三种不同的网络结构,并在每种网络结构下采用不同训练集训练了三个不同的语音增强模型. 所有网络的训练均采用 MVN (Mean and Variance Normalization) 处理后的对数功率谱 (Logarithmic Power Spectra, LPS) 作为输入输出特征,采用 MSE (Mean Square Error) 作为代价函数,输入窗长均设为 15 帧,mini-batch 的大小均为 1024,迭代次数均为 20,训练工具采用的是微软的 Cognitive Toolkit<sup>[17]</sup>. 本文采用的语音增强模型的一般结构如图 5 所示.

通过对未知噪声条件下的语音增强性能进行比较,来分析不同训练集和不同模型下的泛化能力. 本文采用三种客观指标进行语音增强性能的评价,

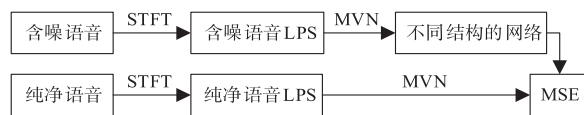


图5 本文采用的语音增强模型的一般结构

包括:评价增强语音质量的 PESQ (Perceptual Evaluation of Speech Quality)<sup>[18]</sup>;评价增强语音可懂度的 STOI (Short Time Objective Intelligibility)<sup>[19]</sup>;以及评价增强语音信噪比的分段信噪比 (Segmental SNR, SegSNR).

### 3.2 DNN-baseline 下的语音增强性能

首先对由三个训练集训练得到的三个基线 DNN 的语音增强性能进行比较,本文的 DNN-baseline 具有 5 个隐层,每个隐层有 1024 个节点,激活函数为 ELU. 图 6 (a) ~ (c) 分别给出了测试集经三个 DNN-baseline 处理后增强语音的平均 PESQ、STOI 和 SegSNR. 可见,在三种不同指标下,训练集 3 对应的 DNN-baseline 在三种不同信噪比下都取得了最好的语音增强性能,其次是训练集 2 对应的 DNN-baseline,而与训练集 1 对应的 DNN-baseline 在三种指标中都是表现最差的.

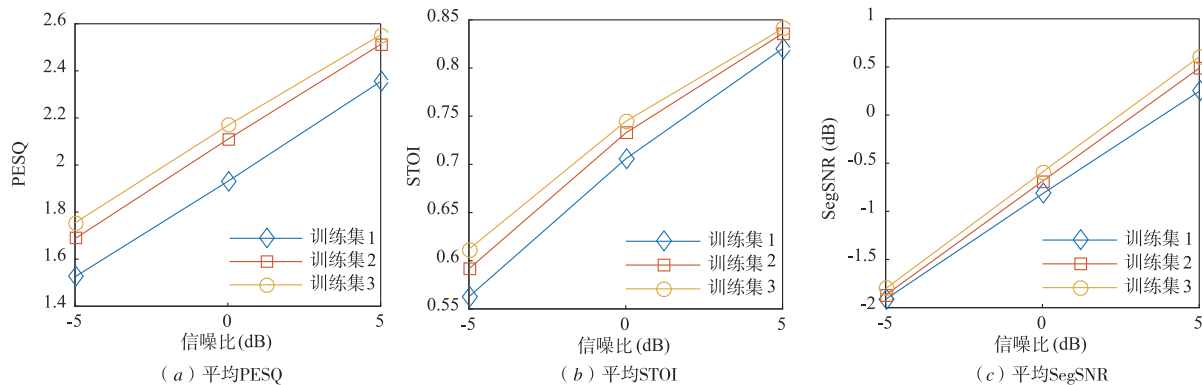


图6 DNN-baseline下不同训练集的语音增强性能

### 3.3 DNN-dropout 下的语音增强性能

为了减少模型本身产生的过拟合对泛化能力的影响,我们在 DNN-baseline 的基础上采用 Dropout 技术得到 DNN-dropout. DNN-dropout 同样具有 5 个隐层,每个隐层有 1024 个节点,在每个隐层后面都紧接一个 Dropout 层,丢弃率设为 0.5,激活函数为 ELU. 分别为三个训练集训练 DNN-dropout,并用其对测试集进行语音增强,图 7 (a) ~ (c) 给出了增强语音的平均 PESQ、STOI 和 SegSNR. 可见,与 DNN-baseline 下的语音增强性能一致,在三种不同指标下,训练集 3 对应的 DNN-dropout 在三种不同信噪比下都取得了最佳的语音增强性能.

为了验证 Dropout 技术对模型泛化能力的影响,图 8 分别给出了三个训练集下 DNN-baseline 和 DNN-dropout 的语音增强性能的对比. 可见,在三种客观指标下,不管采用何种训练集,DNN-dropout 相比由相同训练集

训练得到的 DNN-baseline 都具有更好的语音增强性能;而当采用相同网络结构时,不管是 DNN-baseline 还是 DNN-dropout,随着训练集中噪声类型的增加,语音增强性能都会得到相应提高;其中,由训练集 3 训练得到的 DNN-dropout 在三种指标中都取得了最好的语音增强性能. 上述结论表明扩大训练集噪声多样性和采用 Dropout 技术对泛化能力的提升是叠加的,在建立语音增强模型时可以同时予以考虑.

### 3.4 DCNN 下的语音增强性能

近期研究表明,DCNN (Deep Convolutional Neural Network) 相比于 DNN 在语音增强任务中具有更好的性能. 下面通过对由三个训练集训练得到的 DCNN 的语音增强性能进行比较,分析生成噪声对相对复杂模型泛化能力的影响. 所采用的 DCNN 包含 3 个卷积层和 2 个全连接层,具体的网络结构如图 9 所示,其中卷积层

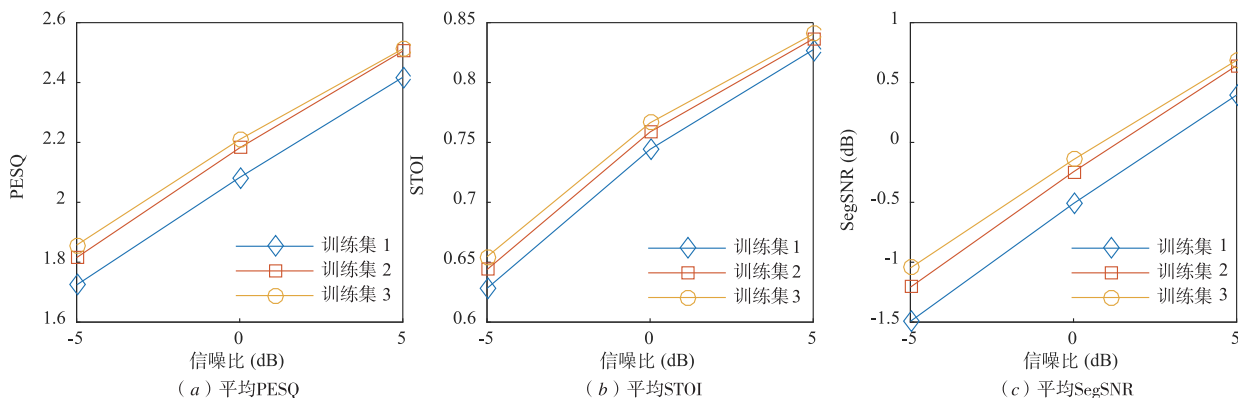


图7 DNN-dropout下不同训练集的语音增强性能

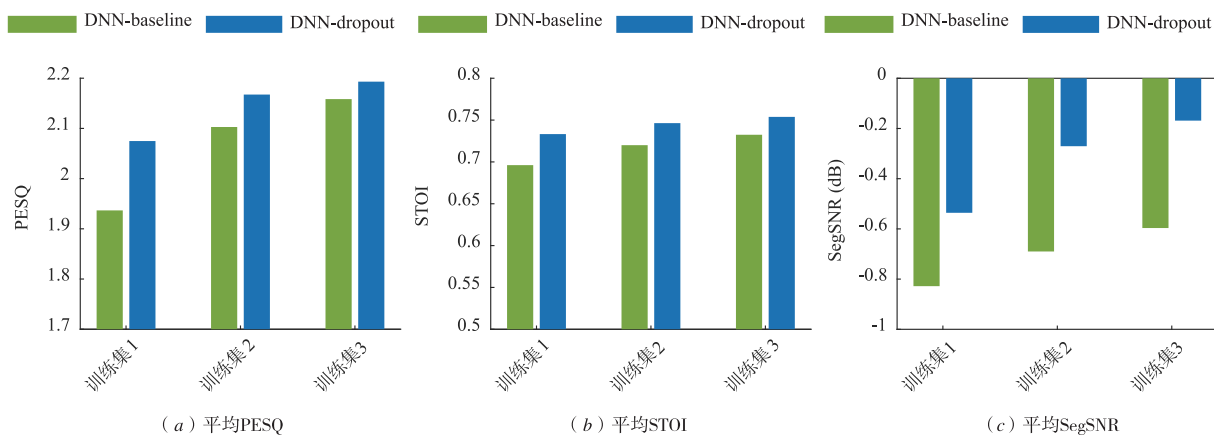


图8 DNN-baseline和DNN-dropout在不同训练集下的语音增强性能对比

的步长为 1,池化层的步长为 2.

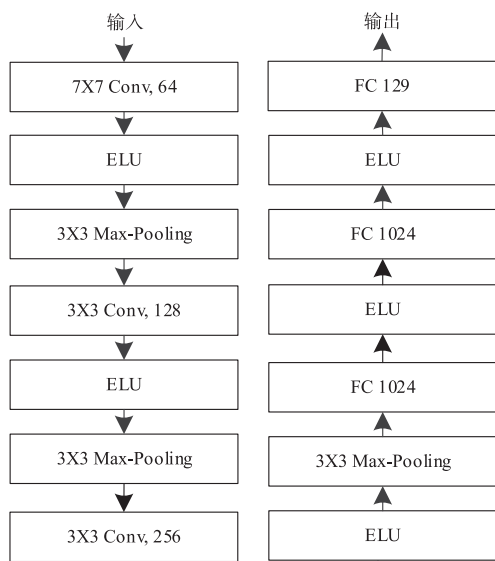


图9 本文所用DCNN的结构框图

图 10(a) ~ (c) 分别给出了与三个训练集对应的三个 DCNN 增强后语音的平均 PESQ、STOI 和 SegSNR. 可见,在相对复杂的 DCNN 模型下,训练集 3 对应的模

型仍然在三种不同指标下的三种不同信噪比中都取得了最佳的语音增强性能.

综合上述实验结果,在从简单到复杂的三种不同网络结构下,都是训练集中包含更多噪声类型的模型具有更好的语音增强性能(训练集 3 > 训练集 2 > 训练集 1),进一步证明了提高训练集噪声类型的多样性是提高未知噪声下语音增强性能的有效手段;而由训练集 3 训练得到的模型在三种不同网络结构下都具有最好的语音增强性能,则表明由本文提出的 NoiseGAN 生成的新噪声能够提高训练集噪声类型的多样性.

#### 4 总结

扩大训练集噪声类型规模是提高有监督语音增强方法泛化能力的重要手段,考虑到真实噪声收集和过程中较高的成本和复杂程度,本文基于图像生成中的 DCGAN,设计了能够基于真实环境噪声生成新的噪声类型的 NoiseGAN. 在采用不同网络结构的语音增强实验中,通过在训练集中增加由 NoiseGAN 生成的噪声类型,训练后的模型在未知噪声类型下表现出了更好的语音增强性能,增强后的语音具有更高的语音质量和更好的可懂度;这表明本文提出的 NoiseGAN 具有

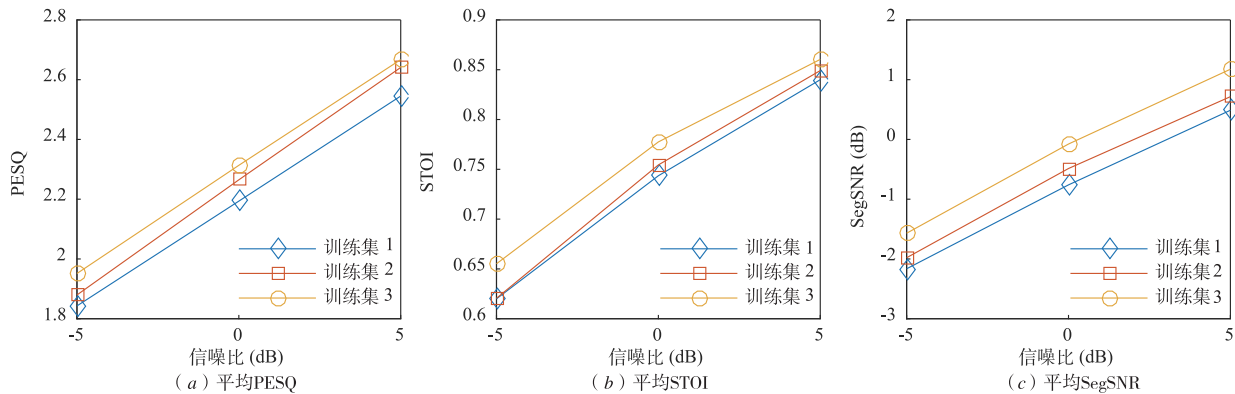


图10 DCNN下不同训练集的语音增强性能

生成新噪声类型的能力,能够有效提高训练集噪声类型的多样性,从而提高语音增强模型对未知噪声类型的泛化能力。

#### 参考文献

- [1] LOIZOU P C. Speech Enhancement: Theory and Practice [M]. UK: CRC Press, 2013.
- [2] 孟宪波, 鲍长春. 基于最小控制 GARCH 模型的噪声估计算法[J]. 电子学报, 2016, 44(3): 747-752.  
MENG Xian-bo, BAO Chang-chun. Noise estimate algorithm based on minima controlled GARCH model[J]. Acta Electronica Sinica, 2016, 44(3): 747-752. (in Chinese)
- [3] 何玉文, 鲍长春, 夏丙寅, 等. 基于 AR-HMM 在线能量调整的语音增强方法[J]. 电子学报, 2014, 42(10): 1991-1997.  
HE Yu-wen, BAO Chang-chun, XIA Bing-yin, et al. Online energy adjustment using AR-HMM for speech enhancement [J]. Acta Electronica Sinica, 2014, 42(10): 1991-1997. (in Chinese)
- [4] 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.
- [5] 刘文举, 聂帅, 梁山, 等. 基于深度学习语音分离技术的研究现状与进展[J]. 自动化学报, 2016, 42(6): 819-833.  
LIU Wen-Ju, NIE Shuai, LIANG Shan, et al. Deep learning based speech separation technology and its developments [J]. Acta Automatica Sinica, 2016, 42(6): 819-833. (in Chinese)
- [6] XU Y, DU J, DAI L R, et al. An experimental study on speech enhancement based on deep neural networks[J]. IEEE Signal Processing Letters, 2014, 21(1): 65-68.
- [7] XU Y, DU J, DAI L R, et al. A regression approach to speech enhancement based on deep neural networks[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2015, 23(1): 7-19.
- [8] XU Y, DU J, HUANG Z, et al. Multi-objective learning and mask-based post-processing for deep neural network based speech enhancement[A]. Proceedings of Sixteenth Annual Conference of the International Speech Communication Association[C]. Dresden: ISCA, 2015. 1508-1512.
- [9] WANG Y, CHEN J, WANG D L. Deep Neural Network Based Supervised Speech Segregation Generalizes to Novel Noises Through Large-Scale Training[R]. Ohio State University Columbus, 2015.
- [10] CHEN J, WANG Y, YOHO S E, et al. Large-scale training to increase speech intelligibility for hearing-impaired listeners in novel noises[J]. The Journal of the Acoustical Society of America, 2016, 139(5): 2604-2612.
- [11] CHEN J, WANG Y, WANG D L. Noise perturbation for supervised speech separation[J]. Speech Communication, 2016, 78: 1-10.
- [12] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[A]. Proceedings of Advances in Neural Information Processing Systems[C]. US: NIPS, 2014. 2672-2680.
- [13] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks[A]. International Conference on Learning Representations[C]. US: ICLR, arXiv: 1511.06434v2.
- [14] HU G. 100 Nonspeech Environmental Sounds, 2004[OL]. <http://web.cse.ohio-state.edu/pnl/corpus/HuNonspeech/HuCorpus.html>, 2004.
- [15] GAROFOLO J S, LAMEL L F, FISHER W M, et al. TIMIT Acoustic-Phonetic Continuous Speech Corpus [CD]. Philadelphia: Linguistic Data Consortium, 1993.
- [16] VARGA A, STEENEKEN H J M. Assessment for automatic speech recognition; II. NOISEX-92: A database and an experiment to study the effect of additive noise on

- speech recognition systems [J]. *Speech Communication*, 1993, 12(3): 247 – 251.
- [17] YU D, EVERSOLE A, SELTZER M, et al. An Introduction to Computational Networks and the Computational Network Toolkit [R]. Tech Rep MSR, Microsoft Research, 2014.
- [18] RIX A W, BEERENDS J G, HOLLIER M P, et al. Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs [A]. *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing* [C]. US: IEEE, 2001. 749 – 752.
- [19] TAAL C H, HENDRIKS R C, HEUSDENS R, et al. An algorithm for intelligibility prediction of time-frequency weighted noisy speech [J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2011, 19 (7): 2125 – 2136.

#### 作者简介



袁文浩 (通信作者) 男, 1985 年出生, 山东寿光人. 2013 年毕业于华东理工大学获博士学位, 现为山东理工大学计算机科学与技术学院讲师. 主要研究方向为语音信号处理、语音增强.

E-mail: why\_sdut@126.com



娄迎曦 女, 1996 年出生, 山东聊城人. 2018 年毕业于青岛理工大学琴岛学院, 现为山东理工大学计算机科学与技术学院硕士研究生. 主要研究方向为语音信号处理、语音增强.

E-mail: 1804224373@qq.com