文章编号：2095-2163(2021)02-0019-04 中图分类号：TP183 文献标志码：J

基于升降编解码全卷积神经网络语音增强技术

孙立辉，曹丽静，张竟雄
（河北经贸大学 信息技术学院，石家庄 050061）

摘 要： 步兵战车强噪声背景下由于强背景噪声的存在，既影响了口令识别的正确率，又降低了指挥所后台监听的清晰度，为了提高语音质量，本文对口令数据进行增强处理。为此，本文提出了一种基于升降编解码全卷积神经网络（Increase Decrease Encoder Decode Convolution Neural Network，IDEDCNN）的语音增强算法，该算法将输入语音信号通过预处理，获取其傅里叶幅度谱特征，并将连续 8 帧的语音信号作为网络的输入，通过编码器来对相邻多帧语音信号建模以提取上下文信息，利用解码器挖掘当前待增强语音帧和上下文信息之间的联系，从而实现语音增强的目的。通过实验证明了该算法能够实现较好的语音增强效果。

关键词： 噪声估计；语音增强；全卷积神经网络

Speech enhancement technology based on lift codec Full Convolutional Neural Network

SUN Lihui, CAO Lijing, ZHANG Jingxiong
(School of Information Technology, Hebei University of Economics and Business, Shijiazhuang 050061, China)

【Abstract】 Due to the presence of strong background noise in the background of infantry fighting vehicles, the accuracy of password recognition is not only affected, but also the clarity of background monitoring of command post is reduced. In order to improve the voice quality, this paper carries out enhanced processing of password data. To this end, this paper puts forward a lift decoding the convolutional Neural Network (happens Decrease Encoder Decode Convolution Neural Network, IDEDCNN), which is the speech enhancement algorithm. In this algorithm, the input speech signal is preprocessed, the Fourier amplitude spectrum features are obtained, and eight adjacent frames of speech signal are taken as network input, model of adjacent frames of voice signal is modeled through the use of the encoder to extract context information. The decoder is used to mine the connection between the speech frame and the context information so as to realize the purpose of speech enhancement. Experimental results show that this algorithm can achieve better speech enhancement effect.

【Key words】 noise estimation; speech enhancement; FCNN

0 引 言

随着军事化训练的自动化，实现对综合采集的战士口令数据的识别，对评估战士的训练效果具有重要意义。在战车训练过程中要对采集的战士口令数据进行后台监听以及口令识别操作。但是由于战车强噪声背景的存在，导致目前的算法无法实现较好的口令识别效果，因此，有必要增强口令数据，从而提高监听效果和口令识别准确率。

神经网络具有强大的学习能力，能够很好地实现语音增强的效果。文献[1]提出利用冗余卷积编码器解码器网络结构学习有噪声语音光谱和干净语音光谱之间的映射，解决了助听器中存在的噪声问题，提高了语音的清晰度。文献[2]通过将新的网络建立到编码器和译码器上，增加基于卷积的短时傅里叶变换层（STFT）和逆 STFT 层来模拟 STFT 的正逆操作，得到了较好的语音增强效果。文献[3]并没有直接对时域信号进行处理，而是将信号转换为频域上的信号，并且使用增强 STFT 幅度和干净 STFT 之间的平均绝对误差损失来训练 CNN，该方法避免了无效 STFT 问题，实验结果表明该算法能够完成增强的目的。

本文提出了一种基于升降编解码全卷积神经网络（Increase Decrease Encoder Decode Convolution Neural Network，IDEDCNN）的语音增强算法，该算法将输入语音信号通过预处理，获取其傅里叶幅度谱特征，并将连续 8 帧的语音信号作为网络的输入，通过编码器来对相邻多帧语音信号建模以提取上下文信息，利用解码器挖掘当前待增强语音帧和上下文信息之间的联系，从而实现语音增强的目的。通过

作者简介：孙立辉（1970-），男，博士，教授，主要研究方向：计算机视觉、机器学习；曹丽静（1994-），女，硕士研究生，主要研究方向：语音增强、深度学习；张竟雄（1996-），男，硕士研究生，主要研究方向：计算机视觉、深度学习。
通讯作者：曹丽静 Email：1443160967@qq.com

收稿日期：2021-01-04

实验证明了该算法能够实现较好的语音增强效果。

1 步兵战车环境下语音增强问题描述

步兵战车强噪声背景下的语音数据是由战士的口令数据 s 和发动机等背景噪声 d 组成的带噪数据 y，即：

$$y = s + d \,, \tag{1}$$

步兵战车环境下的语音增强目标就是输入带噪语音数据 y，得到 s 的较为准确的估计值 s'。为了完成步兵战车背景下战士语音数据增强的任务，在网络的训练阶段使网络学习含噪语音特征和干净语音特征之间的映射关系，即：

$$s' = f(y) \,, \tag{2}$$

在增强阶段利用训练好的模型获得估计的干净语音信号。步兵战车环境下战士语音口令数据增强系统如图 1 所示。

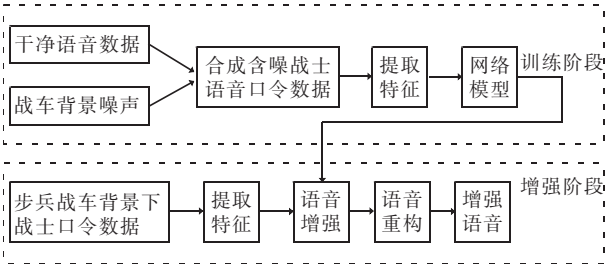

图 1 语音增强系统

Fig. 1 Speech enhancement system

2 升降编解码全卷积神经网络

本文通过实验验证直接利用全卷积神经网络结构实现步兵战车环境下战士语音口令数据的增强，无法实现较大跨度的增强效果，提高语音的质量。受 Lee 等人[1]利用 R-CED(R-Convolution Encode Decode)网络实现了助听器语音数据的增强，本文提出了另外一种卷积网络体系结构，即升降编解码全卷积神经网络(Increase Decrease Encoder Decode Convolution Neural Network，IDEDCNN)来解决步兵战车环境下战士语音口令数据增强。升降编解码全卷积神经网络结构如图 2 所示。

步兵战车背景下战士语音口令增强网络的输入为 129 * 8 的 STFT 矢量，网络是重复的卷积、归一化和 ReLu 激活函数组成，网络深度为 15 个卷积层，实验训练轮数 16 轮，学习率最初设置为 $a = 0.0015$，并且当验证损失在 4 次训练不变时，学习率依次下降为 $a/2$，$a/3$，$a/4$ 来进行训练，损失函数为交叉熵，为了验证本文提出网络结构的可行性，与 FCN 结构进行对比，2 种网络结构见表 1。

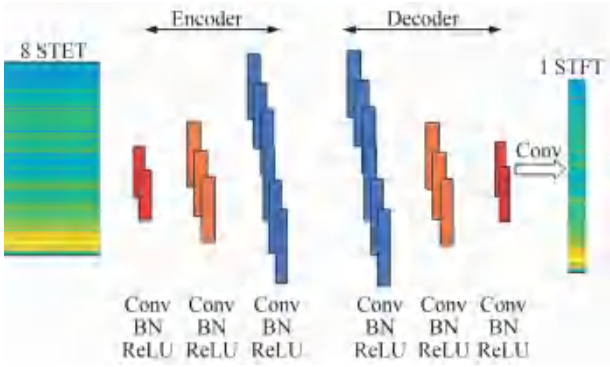

图 2 IDEDCNN 网络结构

Fig. 2 IDEDCNN network structure

表 1 网络结构

Tab. 1 Network structure

网络名称	网络结构	卷积层数量
FCN	(Conv,BN,ReLU) * 15	25-23-21-19-15-14-12-8-12-14-15-19-21-23-25-1
IDEDCNN	(Conv,BN,ReLU) * 15	8-12-14-15-19-21-23-25-23-21-19-15-14-12-8-1

3 实验与结果分析

3.1 数据集

步兵战车环境下战士语音口令数据增强分为训练和增强两个阶段。对此拟做阐释分述如下。

(1)训练数据集。实验数据集分为训练集、测试集和验证集，干净数据为 Common Voice，噪声数据是步兵训练场上采集的各种战车的背景噪声，并且在 0 dB 信噪比时随机添加噪声来增强鲁棒性测试集。训练集共计 5 000 个语音数据段，测试集 200 个语音数据段，实验中 1% 的数据集作为验证集。

(2)增强数据集。增强阶段输入含噪语音口令数据，进行特征提取后输入到预训练好的模型中，进行增强和语音重构后，获得增强后的数据集。数据集共计 3 300 条步兵战车强噪声背景下战士语音口令数据。

3.2 预处理和参数选取

将输入的音频数据进行降采样操作，降到 8 kHz，通过 256 点短时傅里叶变换(32 ms 汉明窗口)计算得到频谱矢量，窗口移动长度为 8 ms，并且通过对称移除信号操作，将 256 点的短时傅里叶(the short-time Fourier transform，STFT))向量简化为 129 点。

通过预处理操作，获得的网络输入特征是由 8

个连续的 STFT 向量组成，并且输入特征都进行了标准化，使其均值和单位方差均为 0。由于语音增强系统是逐帧进行语音增强，因此文中解码器最终只输出当前待增强语音的干净语音特征估计，即只输出一帧，因此输出特征为 129 * 1 的向量，并且进行标准化使其均值和单位方差都为 0。

3.3　优化

为了提高语音的质量，减小噪声过估计，保证噪声估计的鲁棒性，进行了优化，具体如下。

由网络计算出语音存在概率估计值，求出频点间平均功率谱：

$$S_f(\lambda,k)=\begin{cases}\dfrac{\sum_{i=-L_w}^{L_w} b(i)I(\lambda-i,k)\mid Y(\lambda,k)\mid^2}{\sum_{i=-L_w}^{L_w} b(i)I(\lambda-i,k)}, & if \sum_{i=-L_w}^{L_w} I(\lambda-i,k)\neq 0,\\ \sum_{i=-L_w}^{L_w} b(i)\mid Y(\lambda-1,k-i)\mid^2, & otherwise.\end{cases} \tag{3}$$

其中，L_w 是进行帧间平滑的连续帧数；$b(i)$ 为归一化窗函数，$\sum_{i=-u}^{u} b(i)=1$；$\mid Y(\lambda,k)\mid^2$ 为带噪语音第 λ 帧，第 k 频点的功率谱；$I(\lambda-i,k)$ 为 CNN 估计的语音存在概率。

然后进行平滑得到新的功率谱：

$$S(\lambda,k)=\partial S(\lambda-1,k)+(1-\partial)S(\lambda,k) \tag{4}$$

其中，∂ 为谱平滑因子；$S(\lambda,k)$ 为第 λ 帧，第 k 频点处平滑后的功率谱。此时在当前帧以前的 M 帧内搜索平滑后功率谱中的最小值 $S_{\min}(\lambda,k)$，其数学公式可写为：

$$S_{\min}(\lambda,k)=\min\{S(\lambda',k)\mid \lambda-M+1\leq\lambda'\leq\lambda\}, \tag{5}$$

最后，再利用语音存在概率计算噪声估计更新因子 $\partial_d(\lambda,k)$，即：

$$\partial_d(\lambda,k)=\beta+(1-\beta)I(\lambda,k), \tag{6}$$

其中，β 为平滑因子。

进而，得到噪声功率谱估计的更新：

$$\widehat{\sigma_D^2}(\lambda,k)=\alpha_d(\lambda,k)\,\widehat{\sigma_D^2}(\lambda-1,k)+[1-\alpha_d(\lambda,k)]S_{\min}(\lambda,k). \tag{7}$$

3.4　实验与分析

在训练阶段，通过将战士语音口令数据进行特征提取后，输入到对应的网络模型后，通过多次训练得到战士语音口令增强模型，增强阶段将采集的实弹环境下战士口令数据输入到训练模型中进行增强并且重构后得到增强后的数据。通过实验验证了与 FCN 网络相比，本文提出的网络结构能够实现很好的语音增强效果，提高了语音的质量和可懂度。图 3 为带噪语音口令数据波形，图 4 为 FCN 增强后的语音口令数据波形，图 5 为 IDEDCNN 增强后的语音口令数据波形。

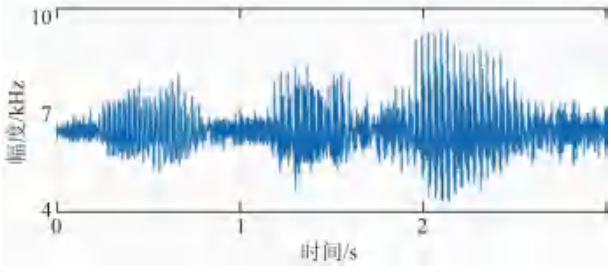

图 3　带噪语音口令数据波形

Fig. 3　Noise speech password data waveform

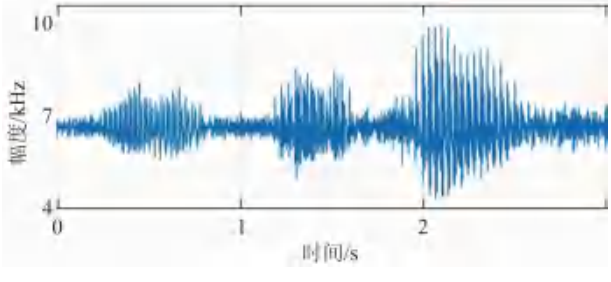

图 4　FCN 增强后语音口令数据波形

Fig. 4　FCN enhanced voice password data waveform

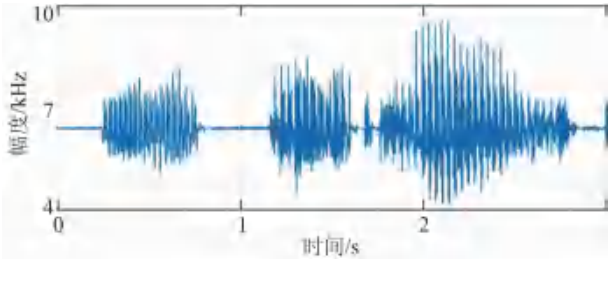

图 5　IDEDCNN 增强后语音口令数据波形

Fig. 5　IDEDCNN enhanced voice password data waveform

4　结束语

本文设计了基于升降编解码卷积神经网络结构实现步兵战车环境下战士语音口令数据增强，与传

统的全卷积神经网络相比,该网络结构在编码阶段滤波器数量逐渐增多,从而获取数据更高维特征,解码阶段压缩特征,并且为了保持语音数据上下文之间的联系,网络的输入为相邻 8 帧的数据。通过与传统全卷积神经网络结构相比,本文提出的网络结构能够实现更好的增强效果。但是由于战车强噪声的极其不稳定,增强结果仍然会存在噪声残留,接下来会继续分析如何更好降低战车强噪声背景下的语音增强,从而实现更好的识别工作。

参考文献

[1] PARK S R, LEE J W. A fully Convolutional Neural Network for speech enhancement [C]//INTERSPEECH 2017. Stockholm, Sweden:ISCA, 2017:1993-1997.

[2] ZHU Yuanyuan, XU Xu, YE Zhongfu. FLGCNN: A novel fully convolutional neural network for end-to-end monaural speech enhancement with utterance-based objective functions [J]. Applied Acoustics,2020,170(2):107511.

[3] PANDEY A, WANG DeLiang. A new framework for CNN-based speech enhancement in the time domain[J]. IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP),2019,27(7):1179-1188.

[4] TAN Ke, CHEN Jitong, WANG DeLiang. Gated residual networks with Dilated Convolutions for monaural speech enhancement[J]. IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP),2019,27(1):189-198.

[5] 彭川. 基于深度学习的语音增强算法研究与实现[D]. 成都:电子科技大学,2020.

[6] 张明亮,陈雨. 基于全卷积神经网络的语音增强算法[J]. 计算机应用研究,2020,37(S1):135-137.

[7] JIA Hairong, WANG Weimei, MEI Shulin. Combining adaptive sparse NMF feature extraction and soft mask to optimize DNN for speech enhancement[J]. Applied Acoustics,2021,171:107666.

[8] YU Hongjiang, ZHU Weiping, CHAMPAGNE B. Speech enhancement using a DNN-augmented colored-noise Kalman filter[J]. Speech Communication,2020,125(2):142-151.

[9] 王师琦,曾庆宁,龙超,等. 语音增强与检测的多任务学习方法研究[J/OL]. 计算机工程与应用:1-8[2020-11-26]. https://kns. cnki. net/kcms/detail/11. 2127. TP. 20201126. 0923. 004. html.

[10] 房慧保,马建芬,田玉玲,等. 基于感知相关代价函数的深度学习语音增强[J]. 计算机工程与设计,2020,41(11):3212-3217.

[11] 郑展恒,曾庆宁. 语音增强算法的研究与改进[J]. 现代电子技术,2020,43(21):27-30.

[12] 袁文浩,时云龙,胡少东,等. 一种基于时频域特征融合的语音增强方法[J/OL]. 计算机工程:1-10[2020-11-26]. https://doi. org/10. 19678/j. issn. 1000-3428. 0059354.

[13] 张行,赵馨. 基于神经网络噪声分类的语音增强算法[J]. 中国电子科学研究院学报,2020,15(9):880-885,893.

[14] 范珍艳,庄晓东,李钟晓. 基于变换域稀疏度量的多级 FrFT 语音增强[J]. 计算机工程与设计,2020,41(9):2574-2584.

[15] 田玉静,左红伟,王超. 语音通信降噪研究[J/OL]. 应用声学:1-11[2020-07-22]. http://kns. cnki. net/kcms/detail/11. 2121. O4. 20200721. 1827. 008. html.

[16] 袁文浩,胡少东,时云龙,等. 一种用于语音增强的卷积门控循环网络[J]. 电子学报,2020,48(7):1276-1283.

[17] 龚杰,冯海泓,陈友元,等. 利用波束形成和神经网络进行语音增强[J]. 声学技术,2020,39(3):323-328.

[18] 李劲东. 基于深度学习的单通道语音增强研究[D]. 呼和浩特:内蒙古大学,2020.

[19] 张宇飞. 基于深度神经网络和循环神经网络的语音增强方法研究[D]. 绵阳:中国工程物理研究院,2020.

[20] 蓝天,彭川,李森,等. 单声道语音降噪与去混响研究综述[J]. 计算机研究与发展,2020,57(5):928-953.

[21] 孔德廷. 一种改进的基于对数谱估计的语音增强算法[J]. 声学技术,2020,39(2):208-213.

[22] 高登峰,杨波,刘洪,等. 多特征全卷积网络的地空通话语音增强方法[J]. 四川大学学报(自然科学版),2020,57(2):289-296.

[23] 王文益,伊雪. 基于改进语音存在概率的自适应噪声跟踪算法[J]. 信号处理,2020,36(1):32-41.

[24] 吴庆贺,吴海锋,沈勇,等. 工业噪声环境下多麦状态空间模型语音增强算法[J]. 计算机应用,2020,40(5):1476-1482.

[25] DANIEL M, TAN Zhenghua, SIGURDUR S, et al. Deep-learning-based audio-visual speech enhancement in presence of Lombard effect[J]. CoRR abs/1905. 12605,2019.

[26] SALEEMN, KHATTAK M I, PEREZ E V. Spectral phase estimation based on deep neural networks for single channel speech enhancement[J]. Journal of Communications Technology and Electronics,2019,64(12):1372-1382.

[27] 董胡,徐雨明,马振中,等. 基于小波包与自适应维纳滤波的语音增强算法[J]. 计算机技术与发展,2020,30(1):50-53.

(上接第 18 页)

feedback control design for singularly perturbed systems with pole placement constraints: An LMI approach[J]. IEEE Transactions on Fuzzy Systems,2006,14(3):361-371.

[14] 陈金香,杨卫东. 模糊奇异摄动系统动态输出反馈 H∞ 控制[J]. 信息与控制,2008,37(5):581-587.

[15] LIN K J, LI T H S. Stabilization of uncertain singularly perturbed systems with pole-placement constraints[J]. IEEE Transactions on Circuits and Systems II: Express Briefs,2006,53(9): 916-920.

[16] MOGHADAM M G, BEHESHTI M T H. On output feedback multiobjective control for singularly perturbed systems [J]. Mathematical Problems in Engineering, 2010,2011:1-28.

[17] HU Jianchen, DING Bochao. Dynamic Output feedback predictive control with one free control move for the Takagi-Sugeno model with bounded disturbance [J]. IEEE Transactions on Fuzzy Systems,2019,27(3):462-473.

[18] XUE Yanmei, ZHENG Bochao, YU Xinghuo. Robust sliding mode control for T-S fuzzy systems via quantized state feedback [J]. IEEE Transactions on Fuzzy Systems, 2018, 26(4): 2261-2272.

[19] TANG Xiaoming, DENG Li, LIU Na, et al. Observer-based output feedback MPC for T-S fuzzy system with data loss and bounded disturbance [J]. IEEE Transactions on Cybernetics, 2019, 49(6): 2119-2132.

[20] YANG C Y, ZHANG Q L. Multiobjective control for T-S fuzzy singularly perturbed systems [J]. IEEE Transactions on Fuzzy Systems,2009,17(1):104-115.