

基于高低频带对数能量谱比贝叶斯决策的语音端点检测

张子丞 谭志苇 张晨瑞 王旋 刘晓璇 俞一彪

苏州大学电子信息学院 江苏 苏州 215006

(494725223@qq.com)

摘要 在分析语音信号与噪声在高低频带的能量谱表现特征基础上,提出一种低信噪比条件下采用高低频带对数能量谱比贝叶斯决策的语音端点检测方法。首先根据样本计算语音信号和背景噪声在高低两个不同频带的对数能量谱比值,依据最大似然估计得到两类信号对数能量谱比的统计分布,并基于贝叶斯决策准则推导最佳判决阈值。信号输入时,逐帧计算高低频带对数能量谱比并与判决阈值进行比较来进行语音和背景噪声的分类判决,从而实现语音信号的端点检测。实验结果表明,与传统的双门限检测法和谱熵检测法相比,提出的方法在较低信噪比条件下能更加准确地检测语音端点,明显提高了端点检测的准确率和速度。

关键词: 语音端点检测;对数能量谱比;贝叶斯决策;低信噪比

中图法分类号 TN912.34

Speech Endpoint Detection Based on Bayesian Decision of Logarithmic Power Spectrum Ratio in High and Low Frequency Band

ZHANG Zi-cheng, TAN Zhi-wei, ZHANG Chen-ru, WANG Xuan, LIU Xiao-xuan and YU Yi-biao

School of Electronic and Information Engineering, Soochow University, Suzhou, Jiangsu 215006, China

Abstract Based on the analysis of the power spectrum of speech signal and noise in high and low frequency band, a speech endpoint detection method under low SNR based on Bayesian decision of logarithmic power spectrum ratio in high and low frequency band is proposed. Firstly, the logarithm power spectrum ratio of speech signal and background noise in two different frequency bands is calculated respectively, and the statistical distribution is obtained according to the maximum likelihood estimation, and the optimal decision threshold is derived based on Bayesian decision criterion. When the signal is input, the log energy spectrum ratio of high and low frequency bands is calculated frame by frame and it is compared with the decision threshold to classify the speech and background noise, so as to realize the endpoint detection of speech signal. The experimental results show that, compared with the traditional double threshold detection method and spectral entropy detection method, the proposed method can detect speech endpoint more accurately under the condition of low SNR, and significantly improve the accuracy and speed of endpoint detection.

Keywords Speech endpoint detection, Logarithmic power spectrum ratio, Bayesian decision, Low SNR

1 引言

语音是人类最自然的信息载体,很可能成为未来人机交互的主要方式,而语音端点检测是语音分析、语音合成、语音编码、说话人识别中的一个重要环节。语音端点检测,指的是在有环境噪声的情况下,从语音信号中确定出语音的起始点和结束点位置的操作。端点检测性能的好坏会影响后续处理的效果,准确检测出语音信号的起止点可以在很大程度上降低后续声学特征的提取以及模型建立的数据量和运算量,有利于提高系统的总体性能。研究表明,即使在安静的环境中,语音识别系统一半以上的识别错误都与端点检测有关。因此,语音信号端点的检测方法,尤其是低信噪比条件下的端点检测,一直是语音信号处理领域的研究热点^[1]。传统的端点检测方法主要利用了语音信号在时域上的特征参数,如过零率^[2]、短时能量^[3]、相关函数^[4]等,同时也有基于频域参数的端点检测方法,如谱熵法^[5]等。然而,这些传统的方法对含有

噪声的语音信号端点检测或多或少存在漏检、错检等误判的情况。例如,短时能量和过零率、平均幅度差等方法比较适合无噪声或者噪声很小的情况,而谱熵法能够应用于噪声环境,但在低信噪比条件下性能快速下降。

本文提出了一种基于高低频带对数能量谱比贝叶斯决策的端点检测方法(Logarithmic Power Spectrum Ratio, LPSR),与传统的谱熵法和短时能量和过零率方法相比较,其在语音端点检测的准确性和检测效率方面都得到了较为明显的提升。

2 语音端点检测经典方法

经典的方法主要有基于短时能量和过零率的双门限方法和谱熵法。

2.1 双门限法

双门限法利用短时平均能量和平均过零率参数进行语音端点检测^[6]。汉语的韵母中都包含元音,其能量较大,所以可

以根据短时平均能量判断韵母的位置;而声母是辅音,它们的频率较高,相应的短时平均过零率较大。因此,结合短时能量和过零率这两个参数可以分别确定语音的声母和韵母部分,从而判断出完整的汉语音节起始和终止位置,完成语音端点检测。

短时平均能量通过对含噪语音信号 $x(n)$ 加窗分帧进行计算,设第 i 帧长为 L ,则得到该帧语音信号 $x_i(n)$,每一帧的短时能量和过零率分别如式(1)和式(2)所示^[7]。

$$E_i = \sum_{n=0}^{L-1} x_i^2(n) \quad (1)$$

$$Z_i = \sum_{n=0}^{L-1} |\text{sgn}[x_i(n)] - \text{sgn}[x_i(n-1)]| \quad (2)$$

其中, $\text{sgn}[\cdot]$ 是符号函数。

$$\text{sgn}(x) = \begin{cases} 1, & x \geq 0 \\ -1, & x < 0 \end{cases} \quad (3)$$

双门限端点检测法通过设置合理的短时能量和过零率阈值 T_E 及 T_Z , 结合语音时长进行端点判决。

2.2 谱熵法

“熵”这个字来源于统计热力学,是紊乱程度的度量^[8],指一个随机事件的不确定性的平均度量,是一个重要的物理概念。熵在多个领域都有重要应用,1948年 Shannon 把熵的概念引入信息论中表示随机事件的平均信息量,而 Shen 等将谱熵成功用于端点检测^[9]。

设含噪语音信号时域波形为 $x(n)$,加窗分帧后对每一帧语音信号进行快速傅里叶变换,对其幅度谱求平方则得到每个频率分量处的能量谱 $P(k)$, $k=0,1,\dots,N-1$,则每个频率分量的归一化谱概率密度函数定义如式(4)所示。

$$p_i(k) = \frac{P_i(k)}{\sum_{l=0}^{N-1} P_i(l)} \quad (4)$$

其中, $p_i(k)$ 为第 i 帧第 k 个频率分量对应的概率密度; $P_i(k)$ 为第 k 个频率分量的能量谱; N 为 FFT 长度。

每个语音帧的短时谱熵定义为:

$$H_i = - \sum_{k=0}^{N-1} p_i(k) \log p_i(k) \quad (5)$$

根据信息理论可以证明,一个离散随机事件等概率分布时熵达到极大值,表明等概分布时信源的平均不确定性为最大。对于式(5)而言,当 $p_i(k) = 2/N, \forall k$ 时,谱熵 H_i 最大。

对于噪声来说,它的归一化谱概率密度函数分布比较均匀,所以它的谱熵值就大;而对于语音信号而言,由于频谱具有共振峰频谱特性,它的归一化谱概率密度函数分布不均匀,语音信号的谱熵通常来说都低于噪声的谱熵,所以可以利用这一特性,通过设置合理的阈值 T_H 从带噪语音信号中提取语音的端点^[10]。

3 高低频带对数能量谱比及端点检测

语音信号端点检测本质是一个二分类判决问题,核心是要提取能够有效表达和区分语音和背景噪声的特征参数,并基于这个特征进行分类。基于语音信号和噪声在高低两个频带的能量谱所表现的不同特性,本文提出一种新的低信噪比条件下的语音端点检测方法。

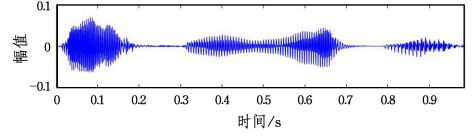
3.1 高低频带能量谱特征分析

对一段普通的语音信号进行分帧,求快速傅里叶变换,进一步计算能量谱,则信号的时域波形和能量谱图如图 1 所示。

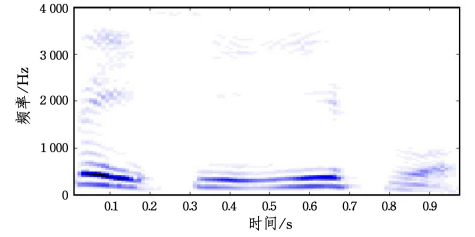
可以从图中明显地看到,语音信号的能量谱呈现出高低频带之间的明显差异,并且能量主要聚集在 1500 Hz 以下的第一、二共振峰,而在高频带(3000~4000 Hz),语音信号的能量谱相对衰减很大,其特征表现为式(6)所示的高低频带对数能量谱比值远小于 0。

$$PR = \text{abs} \left(\log \left(\frac{\text{高频带能量谱}}{\text{低频带能量谱}} \right) \right) \ll 0 \quad (6)$$

其中, $\text{abs}(\cdot)$ 为绝对值函数。



(a) 语音信号波形



(b) 语谱图

图 1 一段含噪语音信号及其语谱图

Fig. 1 Noisy speech signal and its spectrogram

然而对于噪声而言,它的低频带能量谱与高频带能量谱表现得很接近,相对语音信号的能量谱来说波动很小,高低频带对数能量谱比值随时间的变化近似为一条水平直线,特别是在噪声为白噪声的情况下更是如此,即有 $PR \approx 0$ 。

以上分析说明,采用高低频带对数能量谱比值可以有效区分语音信号与背景噪声,实现低信噪比条件下的语音端点检测。

3.2 对数能量谱比的统计分布及贝叶斯决策

无论是语音信号还是背景噪声,不同帧之间的对数能量谱比 PR 会在一定范围内随机变化,一般可以认为是围绕一个均值以一定方差变化的正态分布。因此,可以利用事先采集得到的语音信号和背景噪声样本,采用最大似然估计方法来估计各自的概率密度函数。

正态分布的均值和方差的最大似然估计公式如式(7)和式(8)所示,由此可估计得到语音信号和背景噪声各自的概率密度函数,如式(9)和式(10)所示。

$$m = \frac{1}{N} \sum_{i=0}^{N-1} PR_i \quad (7)$$

$$\sigma = \frac{1}{N} \sum_{i=0}^{N-1} (PR_i - m)^2 \quad (8)$$

$$p_s(x) = \frac{1}{\sqrt{2\pi}\sigma_s} \exp\left(-\frac{(x-m_s)^2}{2\sigma_s^2}\right) \quad (9)$$

$$p_n(x) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{(x-m_n)^2}{2\sigma_n^2}\right) \quad (10)$$

其中, N 为实际样本的总帧数。

基于对数能量谱比进行语音端点检测的问题可以认为是一个依据 PR 特征的二分类问题,由于 PR 特征的随机性,采用贝叶斯分类决策方法是比较合理的,即依据最大后验概率准则进行分类决策。在先验概率一致的前提下,最大后验概率转化为最大条件概率准则,即:若某帧的高低频带对数能量谱比 PR 使得 $p_s(PR) > p_n(PR)$, 则该帧判决为语

音帧,反之判决为背景噪声。

3.3 端点检测算法

基于以上分析,下面给出语音端点检测的高低频带对数能量谱比贝叶斯决策算法的具体步骤。

(1) 短时分帧处理

设输入信号为 $x(n)$,采用局部重叠移位法对其加短时窗 $w(n)$ 分帧处理得到各短时帧信号 $x_i(n)$,如下:

$$x_i(n) = w(n) \cdot x((i-1) \cdot inc + n), 1 \leq n \leq L, 1 \leq i \leq F \quad (11)$$

其中, L 为帧长, inc 为帧移长度, F 为总帧数。

(2) 计算短时能量谱

对每一个短时帧信号 $x_i(n)$ 计算傅里叶变换,并进一步计算能量谱。

$$X_i(\omega) = FT(x_i(n)) \quad (12)$$

$$P_i(\omega) = X_i(\omega) X_i^*(\omega) \quad (13)$$

(3) 计算高低频带对数能量谱比

得到每一帧的能量谱之后,分别计算低频带 $[\omega_1, \omega_2]$ 与高频带 $[\omega_3, \omega_4]$ 的平均能量谱 P_L 和 P_H ,并进一步计算高低频带对数能量谱比 PR 。

$$P_L^i = \frac{1}{\omega_2 - \omega_1} \int_{\omega_1}^{\omega_2} P_i(\omega) d\omega \quad (14)$$

$$P_H^i = \frac{1}{\omega_4 - \omega_3} \int_{\omega_3}^{\omega_4} P_i(\omega) d\omega \quad (15)$$

$$PR_i = \text{abs} \left(\log \left(\frac{P_H^i}{P_L^i} \right) \right), 1 \leq i \leq F \quad (16)$$

(4) 设置对数能量谱比阈值

将贝叶斯分类决策规则转化成阈值判决规则,根据当前帧对数能量谱比 PR 和阈值直接进行分类决策,在实际应用时,相比计算后验概率可以有效减少处理时间。

贝叶斯分类决策的判决面为 $p_s(x) = p_n(x)$,根据式(9)和式(10)推导可得到理论上具有最小错误率的贝叶斯判决阈值:

$$T_{pr} = \frac{(m_s \sigma_n^2 - m_n \sigma_s^2) - \sigma_s \sigma_n \sqrt{(m_s - m_n)^2 + \ln \left(\frac{\sigma_s}{\sigma_n} \right)^2}}{\sigma_n^2 - \sigma_s^2} \quad (17)$$

该阈值是根据统计分布事先确定的,位于 (m_n, m_s) 之间,并不需要在端点检测时反复计算。

(5) 分类决策

根据式(16)得到的对数能量谱比和式(17)的判决阈值,最大后验概率贝叶斯分类决策可转化为如下的阈值分类决策。

$$y = \begin{cases} \text{语音帧}, & PR > T_{pr} \\ \text{背景噪声}, & PR < T_{pr} \end{cases} \quad (18)$$

实际算法应用时,还应该结合语音时长的一般统计规律做平滑处理,将一些孤立语音帧和噪声帧平滑掉。

4 实验结果与分析

4.1 实验环境与条件

在 Windows 操作系统下,使用了 Matlab2017 为仿真平台进行语音信号端点检测的实验分析。实验数据在标准录音室采集录制,包含了多名不同性别的说话人的 21 个单词或短语音样本,采集的原始声音文件可以近似为纯净的语音数

据,语音内容为日常生活中出现的名词,采样频率为 8000 Hz、16 比特,以 wav 文件格式存储。实验中,短时窗采用哈密窗,长度为 25ms。对以上纯净语音信号进行加噪处理,分别加入了 SNR 为 10 dB, 20 dB 和 30 dB 的噪声。对基于短时能量和过零率的双门限法、谱熵法和本文提出的高低频带对数能量谱比贝叶斯决策法(LPSR)这 3 种端点检测方法进行比较。

实验结果通过端点检测准确率来评价。检测准确率定义为总帧数减去错误帧数后占总帧数的百分比^[11],即:

$$\text{准确率} = \frac{\text{总帧数} - \text{错误帧数}}{\text{总帧数}} \times 100\% \quad (19)$$

4.2 结果与分析

为了检验本文提出算法的正确性并将其与其他经典算法进行比较,实验分别采用双门限法、谱熵法和对数能量谱比贝叶斯决策法(LPSR)对 21 组语音样本进行处理并分析,得到了 63 组端点检测的结果。图 2 给出了 SNR=10 dB 条件下语音信号样本“布鞋”的端点检测情况。

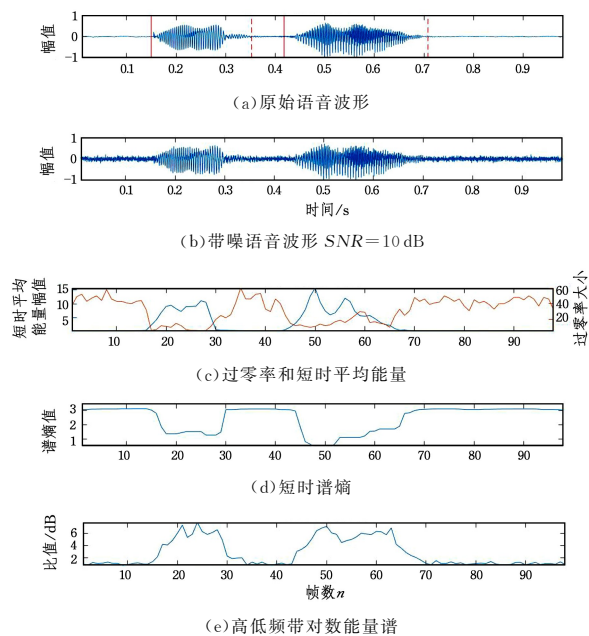


图 2 语音信号“布鞋”时域波形及 3 种方法端点检测对比图

Fig. 2 Time domain waveform of speech signal “cloth shoes” and comparison of three endpoint detection methods

从图 2 中可以看出,在信噪比为 10dB 的情况下,3 种方法都能够较为准确地将语音信号与噪声信号区分开来,完成端点检测。根据准确率的计算公式可得,谱熵法的准确率为 84.44%,双门限法的准确率为 84.44%,LPSR 方法的准确率为 93.33%,LPSR 有较大优势。观察图 1 可以发现,在对每个音节的起始点位置进行判断时,谱熵法会过多地将噪声片段当作语音信号,提前认定为端点,而双门限法则会部分语音信号视作噪声,提前结束语音段,两种方法在检测精度上都存在一定程度上误差;而 LPSR 法则较为准确并完整地提取了语音片段。

图 3 给出了 SNR=20 dB 条件下语音信号样本“钉子”的信号波形及 3 种不同方法的端点检测情况。从图中同样可以较为明显地看到 LPSR 方法相较于其他两种方法可以更加准确、完整地检测出语音的端点位置。

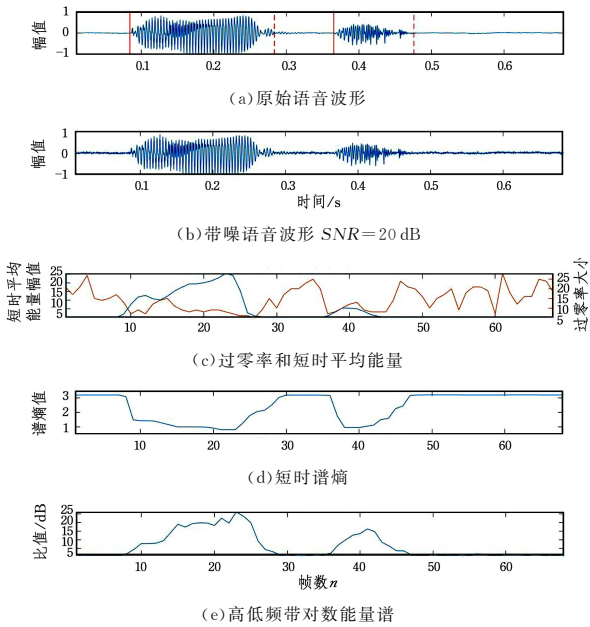


图3 语音信号“钉子”时域波形和3种方法端点检测对比图

Fig. 3 Time domain waveform of speech signal “nail” and comparison of three endpoint detection methods

在 $SNR=10\text{ dB}$ 条件下, 3 种端点检测方法对全部 21 种语音样本的检测准确率统计结果如表 1 所列。可以看到, 全部 21 个测试语音信号中, 总体上高低频带对数能量谱比贝叶斯决策法 LPSR 的性能较好, 短时能量和过零率双门限法、谱熵法和本文提出的 LPSR 法的端点检测平均准确率分别是 84.72% 、 85.80% 和 90.76% 。不仅如此, LPSR 方法具有较高检测准确率的同时, 准确率方差也最小, 因此稳定性和普遍适用性更强。

表 1 不同端点检测方法在 10 dB 条件下的检测结果Table 1 Test results of different endpoint detection methods at 10 dB

语音	端点检测方法		
	谱熵	双门限	LPSR
报纸	97.50	57.50	77.50
本子	63.64	84.85	81.82
玻璃类	95.92	100.00	95.92
布鞋	84.44	84.44	93.33
充电器	94.00	100.00	92.00
窗帘	97.56	92.68	100.00
床单	100.00	93.18	95.45
钉子	82.14	89.29	100.00
废电器	94.34	94.34	86.79
废弃衣服	91.67	90.00	93.33
废铁	87.88	78.79	87.88
剪刀	90.70	76.74	90.70
金属类	86.36	93.18	93.18
金属瓶盖	86.44	89.83	94.92
酒瓶	94.00	86.00	92.00
毛毯	70.45	79.55	90.91
玩具汽车	85.11	81.91	90.43
桌布	94.34	81.13	88.68
纸箱	87.23	87.23	87.23
调味瓶	29.31	51.72	85.19
鞋盒	88.68	86.79	88.68
平均值	85.80	84.72	90.76
方差	0.0243	0.0141	0.0030

此外, 表 2 可知, 在信噪比为 20 dB 时, 谱熵法的准确率为 81.67% , 双门限法的准确率 88.39% , LPSR 法的准确率为 92.66% ; 在信噪比为 30 dB 时, 谱熵法的准确率为 85.07% , 双门限法的准确率 74.67% , LPSR 法的准确率为 92.79% 。可以看出, 在以上 3 种信噪比情况下, LPSR 法的准确率高于另外两种, 并随信噪比增加呈现稳定上升趋势, 而双门限法和谱熵法的性能存在一定波动。由表 3 同样可以看到, LPSR 方法相对双门限法和谱熵法的准确率方差更低, 也就是在 3 种信噪比条件下都更加稳定和具有普遍适用性。

表 2 3 种方法在不同信噪比时的平均准确率

Table 2 Average accuracy of three methods with different SNR (单位: %)

方法	10 dB	20 dB	30 dB
谱熵法	85.80	81.67	85.07
双门限法	84.72	88.39	74.67
LPSR	90.76	92.66	92.79

表 3 3 种方法在不同信噪比时的准确率方差

Table 3 Accuracy variance of three methods with different SNR

方法	10 dB	20 dB	30 dB
谱熵法	0.0243	0.0256	0.0076
双门限法	0.0141	0.0059	0.0186
LPSR	0.0030	0.0024	0.0037

结束语 本文对低信噪比条件下的语音端点检测算法进行了研究, 针对现有端点检测方法在噪声干扰情况下存在失效和不稳定现象, 提出了一种采用高低频带对数能量谱比贝叶斯决策的语音端点检查方法(LPSR)。实验验证了提出的 LPSR 方法是有效的, 并且其相对于传统的短时能量过零率方法和谱熵法具有较大的优势, 在检测准确率和准确率方差两个指标上都有较大提升。

研究表明: 对数能量谱比作为一种特征应用于低信噪比条件下的语音端点检测是有效的, 基于贝叶斯决策准则推导的对数能量谱比判决阈值能够有效实现判决分类, 提出的 LPSR 方法相对于传统方法能够较好地实现低信噪比条件下的语音端点检测。

参考文献

- [1] CHEN Y Y, JIAN L. Speech endpoint detection based on maximum entropy spectral estimation and time-frequency characteristics[J]. Computer Applications and Software, 2017, 34 (11): 91-96.
- [2] SUN H F, LONG H, SHAO Y B, et al. Speech music classification algorithm based on zero crossing rate and spectrum [J]. Journal of Yunnan University (Natural Science Edition), 2019, 41(5): 925-931.
- [3] CHEN H H, XU P. Sound endpoint detection method based on spectral subtraction and short-term energy [J]. Mechanical Manufacturing and Automation, 2016, 45(3): 191-192, 209.
- [4] CHEN Z W, ZENG Q N, XIE X M, et al. Speech endpoint detection method based on autocorrelation function [J]. Computer Engineering and Application, 2018, 54(6): 216-221, 256.
- [5] WANG R D, CHAI P Q. An improved speech endpoint detection method based on spectral entropy [J]. Information and Control, 2004(1): 77-81.
- [6] RABINER L R, SAMBUR M R. An Algorithm for Determining

the Endpoints of Isolated Utterances[M]. John Wiley & Sons, 1975.

- [7] FEI Y Q, WANG Y J, XIA Y L. Research on speech endpoint detection algorithm [J]. Automation Technology and Application, 2017, 36(8): 98-102.
- [8] LI Y, CHENG L F, ZHANG P L. A speech endpoint detection method based on improved spectral entropy [J]. Computer Science, 2016, 43(S2): 233-236.
- [9] WANG L, LI C R. An improved endpoint detection method based on adaptive spectral entropy [J]. Computer Simulation, 2010, 27(12): 373-375, 395.
- [10] WANG S, PU B M, LI X Z, et al. Wiener filter speech enhancement algorithm based on improved energy entropy ratio [J]. Computer System Applications, 2017, 26(11): 124-131.
- [11] WU P, ZHANG X B, DING W. Speech endpoint detection based

on spectral subtraction and Short-term zero entropy method [J]. Electroacoustic Technology, 2018, 42(3): 55-59.



ZHANG Zi-cheng, born in 1999, undergraduate. His main research interests include digital signal processing and speech processing.



YU Yi-biao, born in 1962, Ph.D, professor. His main research interests include speech and image processing, pattern recognition and multimedia system.

(上接第 20 页)

- [6] ZHOU F Y, JIN L N, DONG J. Review of Convolutional Neural Network[J]. Chinese Journal of Computers, 2017, 40(6): 1229-1251.
- [7] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet Classification with Deep Convolutional Neural Networks[C]// Proceedings of the 25th International Conference on Neural Information Processing Systems Inc. Lake Tahoe: Curran Associates Inc, 2012: 1097-1105.
- [8] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition[EB/OL]. (2015-04-10)[2020-07-05]. <https://arxiv.org/pdf/1409.1556.pdf>.
- [9] SZEGEDY C, LIU W, JIA Y, et al. Going Deeper with Convolutions[C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE Press, 2015: 1-9.
- [10] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE Press, 2016: 770-778.
- [11] REDMON J, DIVVALA S, GIRSHICK R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE Press, 2016: 779-788.
- [12] GIRSHICK R, DONAHUE J, DARRELL T, et al. Region-Based Convolutional Networks for Accurate Object Detection and Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(1): 142-158.
- [13] KIM Y. Convolutional Neural Networks for Sentence Classification[EB/OL]. (2014-09-03)[2020-07-05]. <https://arxiv.org/pdf/1408.5882.pdf>.
- [14] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional Networks for Biomedical Image Segmentation[C]// Proceedings of the 18th International Conference on Medical Image Computing and Computer Assisted Intervention. Munich: Springer, 2015: 234-241.
- [15] LONG J, SHELHAMER E, DARRELL T. Fully Convolutional

Networks for Semantic Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(4): 640-651.

- [16] CHEN P, BAI X, LIU W. Vehicle Color Recognition on Urban Road by Feature Context[J]. IEEE Transactions on Intelligent Transportation Systems, 2014, 15(5): 2340-2346.
- [17] ZHANG X. Object color recognition based on convolutional neural network[D]. Hangzhou: Zhejiang University, 2015.
- [18] CHEN P. Research on vehicle color recognition in natural scenes [D]. Wuhan: Huazhong University of Science and Technology, 2016.
- [19] CHEN H C. A Method of vehicle color recognition based on deep convolutional neural networks[J]. Journal of The Hebei Academy of Sciences, 2017, 34(2): 1-6.
- [20] GUAN D Y, JU M, AN L H. Research on vehicle color recognition technology based on convolution neural network[J]. Journal of Shandong Jianzhu University, 2018, 149(1): 25-31.
- [21] REDMON J, FARHADI A. YOLOv3: An Incremental Improvement[EB/OL]. (2018-04-08)[2020-07-05]. <https://arxiv.org/pdf/1804.02767.pdf>.



ZHOU Xin, born in 1974, Ph.D, associate professor. His main research interests include computer vision and intelligent transportation.



CHEN Yuan-yuan, born in 1983, Ph.D, associate professor. Her main research interests include deep neural networks and so on.