第 45 卷 第 3 期 计 算 机 工 程 2019 年 3 月
Vol. 45 No. 3 Computer Engineering March 2019

·多媒体技术及应用· 文章编号：1000-3428(2019)03-0262-06 文献标志码：A 中图分类号：TN912.34

基于深度学习的小样本声纹识别方法

李 靓[1a]，孙存威[1b]，谢 凯[1a]，贺建飚[2]

（1. 长江大学 a. 电子信息学院；b. 计算机科学学院，湖北 荆州 434023；2 中南大学 信息科学与工程学院，长沙 410083）

摘 要： 利用小样本声纹作为训练集训练卷积神经网络（CNN）时，网络不能达到较好的收敛状态，从而导致识别率较低。为此，提出一种新的声纹识别方法。利用深度 CNN 提取潜在的声纹特征，在 CNN 训练过程中采用基于凸透镜成像原理的图像增多算法解决小样本训练样本不足的问题，并在卷积过程中引入快速批量归一化（FBN）方法以提高网络收敛速度、缩短训练时间。在包含 630 人的 TIMIT 语音数据库中进行训练、验证和测试，结果表明，FBN-Alexnet 网络比 Alexnet 网络训练时间缩短 48.2%，与 GMM、GMM-UBM 及 GMM-SVM 方法相比，该方法识别率分别提高7.3%、2.2%、2.8%。

关键词： 声纹识别；深度学习；FBN-Alexnet 网络；小样本；快速批量归一化；图像增多算法

中文引用格式： 李靓，孙存威，谢凯，等. 基于深度学习的小样本声纹识别方法[J]. 计算机工程，2019，45(3)：262-267，272.

英文引用格式： LI Jing，SUN Cunwei，XIE Kai，et al. Small sample voiceprint recognition method based on deep learning[J]. Computer Engineering，2019，45(3)：262-267，272.

Small Sample Voiceprint Recognition Method Based on Deep Learning

LI Jing[1a]，SUN Cunwei[1b]，XIE Kai[1a]，HE Jianbiao[2]

（1a. School of Electronic and Information；1b. School of Computer Science，Yangtze University，Jingzhou，Hubei 434023，China；
2. College of Information Science and Engineering，Central South University，Changsha 410083，China）

【Abstract】 When training Convolutional Neural NetWork（CNN）with small sample voiceprints as training set，the network cannot reach a good convergence state，which results in low recognition rate. So，this paper proposes a new voiceprint recognition method. The proposed method uses deep CNN to extract the rich and latent features of voiceprint，which improves the voiceprint recognition rate. In order to solve the problem that small sample cannot train the CNN，this paper proposes an image increasing algorithm based on the principle of convex lens imaging. At the same time，the Fast Batch Normalization（FBN）is introduced in the convolutional process，which improves the speed of the network convergence and shortens the training time. Select a TIMIT speech database containing voices of 630 speakers for training，validating and testing. Experimental results show that，compared with the GMM，GMM-UBM，and GMM-SVM algorithms，the proposed method improves the recognition rate by 7.3%，2.2%，and 2.8% and compared with the original network，the training time of the FBN-Alexnet network is reduced by 48.2%. It means that it is an effective method for voiceprint recognition of small samples.

【Key words】 voiceprint recognition；deep learning；FBN-Alexnet network；small sample；Fast Batch Normalization（FBN）；image increasing algorithm

DOI： 10.19678/j.issn.1000-3428.0049975

0 概述

声纹识别是根据人特有的语音特征识别说话人身份的一种生物特征识别技术[1-2]。在传统机器学习方法中，特征选择是识别精度的关键。目前常见的特征提取包括 Mel 频率倒谱系数[3]、线性预测倒谱系数[4]、i-supervector[5]，机器学习方法则主要有 HMM-UBM、GMM、GMM-SVM。这些方法提取的特征具有单一性，导致识别精度低，尤其在噪音背景下的识别效果较差。

随着深度学习时代的到来，卷积神经网络（Convolutional Neural Network，CNN）在图片识别领

基金项目： 国家自然科学基金（61272147）；湖北省教育厅项目（B2015446）；长江大学青年基金（2016cqn10）；大学生创新创业计划基金（2017009）。

作者简介： 李 靓（1996—），男，硕士研究生，主研方向为语音信号处理、图像处理；孙存威（通信作者），硕士研究生；谢 凯，教授、博士生导师；贺建飚，副教授。

收稿日期： 2018-01-04 **修回日期：** 2018-02-05 **E-mail：** SCW_1501@163.com

域取得了较大的进展[6]。相较于传统方法,CNN 避免了手工提取特征表征能力不足的问题。近年来语音识别领域也引入了 CNN[7-9],用 CNN 直接学习语谱图,相比传统提取语音特征方式,减少了在时域和频域上的信息损失。同时,由于 CNN 局部连接和权值共享的特点,使得 CNN 具有平移不变性,因此能够克服语音信号本身多样性的问题。

基于深度学习的声纹识别模型通过大量语音数据训练时,可自动学习丰富的声学特征(频谱、基音、共振峰等),提高了声纹识别率,但是对小样本的声纹识别并不理想。因为训练好一个深层的 CNN 需要大量的训练样本,学习数百万个网络参数[10],仅用小样本声纹作为训练集训练 CNN,网络并不能达到较好的收敛状态,从而导致声纹识别率低。

针对目前研究人员较少利用深度学习解决小样本声纹识别率低的问题,本文提出一种深度模型下的小样本声纹识别方法。由于在实际工作中很难获得大量的声纹数据,本文给出一种基于凸透镜成像的图像增多算法。此外,在网络训练过程中,由于存在网络层数较多、网络参数巨大、训练耗时以及网络拟合问题,本文引入快速批量归一化(Fast Batch Normalization,FBN)方法,以在训练 FBN-Alexnet 网络时加速网络收敛。

1　小样本声纹识别算法

本文算法流程如图 1 所示。

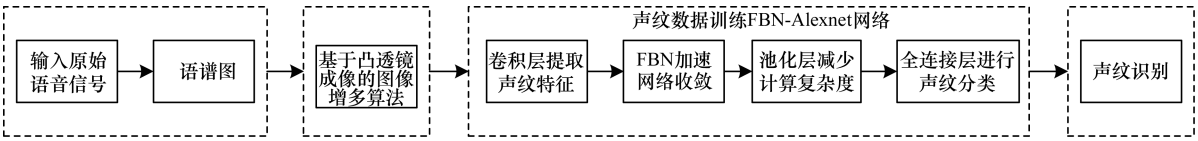

图 1　基于深度模型的小样本声纹识别流程

1.1　原始语音信号预处理

在使用音频训练或测试模型之前,由于语音信号具有短时不变性的特点,本文对一段语音信号 $x(t)$ 进行分帧,转化为 $x(m,n)$(m 为帧的个数,n 为帧长),通过短时傅里叶变换得 $X(m,n)$,将 $X(m,n)$ 经 $Y(m,n)$($Y(m,n) = X(m,n) \times X(m,n)'$)变换得到周期图。根据时间将 m 变换为刻度 M,根据频率将 n 变换为刻度 N,将(M,N,$10 \times \mathrm{lb}\ 10Y(m,n)$)画成二维图(即语谱图)。由原始语音信号生成的语谱图如图 2 所示。

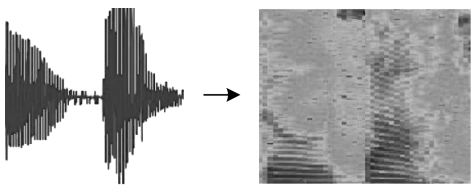

图 2　语音信号-语谱图转换

1.2　基于凸透镜成像的图像增多算法

本文提出的图像增多算法采用凸透镜成像原理,通过改变光谱图的大小获得更多的训练数据。

将预处理得到的语谱图放在 P 点位置,根据凸透镜成像原理:

1)当 P 点离透镜的距离大于 F 小于 $2F$(F 为透镜焦距)时,获得的图像比原始图像大,如图 3(a)所示。

2)当 P 点离透镜的距离为 $2F$ 时,获得的图像与原始图像一样大,如图 3(b)所示。

3)当 P 点离透镜的距离大于 $2F$ 时,获得的图像比原始图像小,如图 3(c)所示。

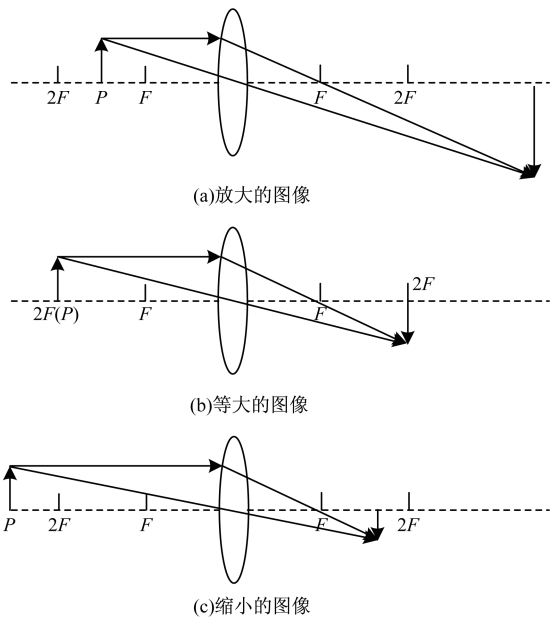

图 3　凸透镜成像原理

通过 3 种变换可获得多张图像,并将所有图像尺度归一化为 227 × 227 作为 CNN 的输入。

1.3　FBN-Alexnet 网络

Alexnet[11,12]网络结构和 FBN-Alexnet 网络结构分别如图 4、如图 5 所示。在图 5 中,S 表示步幅,Pad 表示补白。

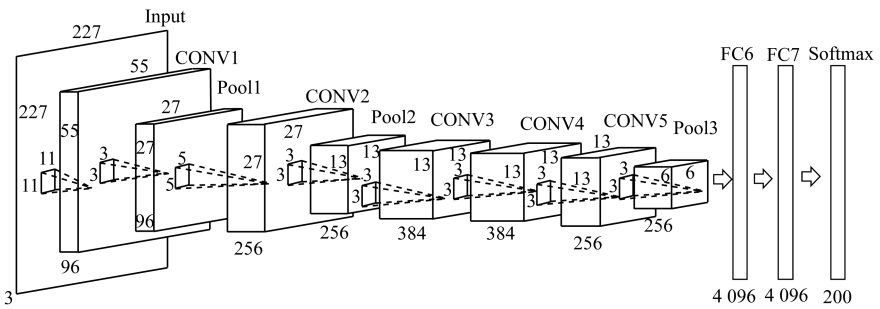

图 4 Alexnet 网络结构

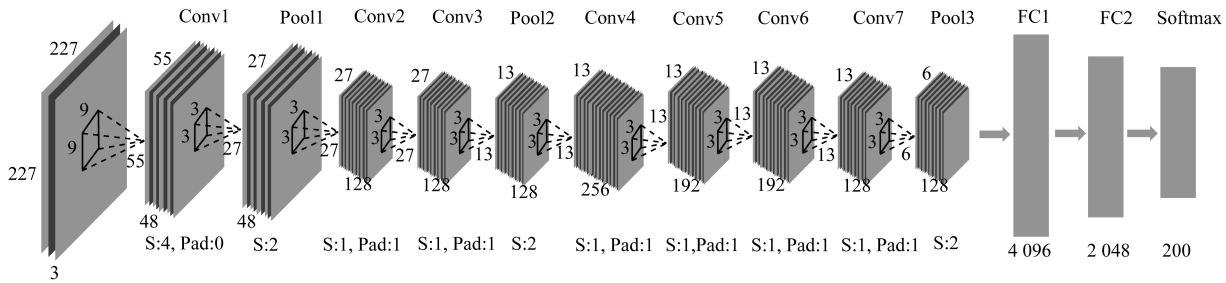

图 5 FBN-Alexnet 网络结构

本文使用的 FBN-Alexnet 网络相比 Alexnet 有以下改进：

1)通过改变卷积核的数量和大小，减少全连接层的节点数，FBN-Alexnet 网络相比 Alexnet 网络降低了计算消耗。具体改变为：

(1)用大小为 9×9 的卷积核替换 11×11 的卷积核(CONV1→Conv1)。

(2)将大小为 5×5 的卷积核分解为 2 层 3×3 的卷积核(CONV2→Conv2 和 Conv3)。

(3)减少每一层网络的特征图数。

(4)增加一层卷积层(Conv7)。

(5)将第 2 层全连接层的节点数减半。

2)在实验过程中，笔者发现通过 Alexnet 网络中的局部响应归一化进行参数初始化，对网络训练效果较小，于是去除局部响应归一化并且在 ReLU 激活函数之前加入 FBN 层，对数据进行归一化，加快网络的收敛速度。

整个网络的训练过程如下：

(1)正向传播学习网络

输入特征在每层中的神经元计算公式如下：

$$net^{(l+1)} = \boldsymbol{W}^{(l+1)}\boldsymbol{x}^{(l)} + \boldsymbol{b}^{(l+1)} \tag{1}$$

$$\boldsymbol{x}^{(l+1)} = s(FBN(net^{(l+1)})) \tag{2}$$

其中，$\boldsymbol{x}^{(l)}$ 是第 l 层的向量输出，$\boldsymbol{x}^{(l+1)}$ 是第 $l+1$ 层的向量输出，$\boldsymbol{W}^{(l+1)}$ 是层间线性系数组成的矩阵，$\boldsymbol{b}^{(l+1)}$ 是第$(l+1)$层的偏差组成的向量，$s(\cdot)$是激活函数，$FBN(\cdot)$为本文引入的 FBN 算法。

设某层神经元的输出包含 t 维数据，$net=\{e^{(1)}, e^{(2)},\cdots,e^{(t)}\}$，在每个维度中，独立应用归一化算法。样本容量为 s 的小批量数据样本，表示为：$B_e=\{e_1, e_2,\cdots,e_s\}$，样本数据归一化后的结果为：$B_g=\{g_1, g_2,\cdots,g_s\}$，$g_i(i\in[1,s])$服从 $N(0,1)$分布，FBN 算法具体执行见算法 1。

算法 1 FBN 算法

输入 小批量数据样本 B_e

输出 归一化后的样本数据 B_g

1. 小批量均值为：$\mu=\frac{1}{s}\sum_{i=1}^{s}e_i$

2. 样本方差为：$\sigma^2=\frac{1}{s}\sum_{i=1}^{s}(e_i-\mu)^2$

3. 归一化后的样本值为：$g_i=\frac{e_i-\mu}{\sigma}$

4. 更新全局平均值为：$\mu_B=(1-\gamma)\mu_B+\gamma\mu$

5. 更新全局方差为：$\sigma_B^2=(1-\gamma)\sigma_B^2+\gamma\sigma^2$

在算法 1 中，初始化参数值 $\mu_B=0$，$\sigma_B^2=1$，$\gamma=0.01$。通过引入动量 γ 实现全局均值和方差的更新，在测试阶段 μ_B 及 σ_B^2 取最后训练得到的值。

在反向传播过程中，对于损失函数 L，应用链式规则，归一化层的反向传播梯度由式(3)~式(5)决定。

$$\frac{\partial L}{\partial \sigma^2}=-\frac{1}{2}\sum_{i=1}^{s}\frac{\partial L}{\partial g_i}(e_i-\mu)(\sigma^2)^{-\frac{3}{2}} \tag{3}$$

$$\frac{\partial L}{\partial \mu}=\sum_{i=1}^{s}\frac{\partial L}{\partial g_i}\cdot\frac{-1}{\sigma}+\frac{\partial L}{\partial \sigma^2}\cdot\frac{-2\sum_{i=1}^{s}(e_i-\mu)}{s} \tag{4}$$

$$\frac{\partial L}{\partial e_i}=\frac{\partial L}{\partial g_i}\cdot\frac{1}{\sigma}+\frac{\partial L}{\partial \sigma^2}\cdot\frac{2(e_i-\mu)}{s}+\frac{1}{s}\cdot\frac{\partial L}{\partial \mu} \tag{5}$$

在网络训练过程中任一层网络的参数变化都会

引起后续神经网络各层输入的分布变化,导致神经网络必须不断地适应新的数据分布,这就要求更细致地调整参数、使用更小的学习率去训练,并且由于在激活操作中存在非线性饱和问题,使得网络训练更加困难。在算法 1 中,将神经网络各层输入数据归一化为标准正态分布,能够有效解决上述问题,进而降低网络训练时间、加速网络收敛。

(2)BP 算法反向传播调整网络

为了提高网络的自适应性,利用 BP 算法[13]反向调整参数。由于方差损失函数权重更新过慢,本文采用交叉熵代价函数[14],其优点是:误差大,网络参数更新快;误差小,网络参数更新慢。对于含 N 个声纹的样本集 $x=\{(x^{(1)},y^{(1)}),(x^{(2)},y^{(2)}),\cdots,(x^{(N)},y^{(N)})\}$,交叉熵代价函数定义为:

$$\Gamma(\theta)=-\frac{1}{N}\sum_{i=1}^{N}[y^{(i)}\mathrm{lb}(o^{(i)})+(1-y^{(i)})\mathrm{lb}(1-o^{(i)})] \tag{6}$$

其中,N 表示声纹训练样本容量,$o^{(i)}$ 表示输入 $x^{(i)}$ 对应的实际输出,$y^{(i)}$ 表示第 i 组数据对应的类别标记,$y^{(i)}\in\{1,2,\cdots,k\}$,$k$ 是声纹类别的数目,卷积层参数 w 和 b 的反向传播梯度由式(7)、式(8)决定。

$$\frac{\partial}{\partial w^{(i)}}\Gamma(\theta)=\frac{1}{N}\sum_{i=1}^{N}x^{(i)}(o^{(i)}-y^{(i)}) \tag{7}$$

$$\frac{\partial}{\partial b^{(i)}}\Gamma(\theta)=\frac{1}{N}\sum_{i=1}^{N}(o^{(i)}-y^{(i)}) \tag{8}$$

采用梯度下降法更新 CNN 网络参数,使得网络输出层误差函数值达到最小。ρ 为学习率,每层参数 $w^{(i)}$ 和 $b^{(i)}$ 更新计算公式如下:

$$w^{(i)}=w^{(i)}-\rho\frac{\partial}{\partial w^{(i)}}\Gamma(\theta) \tag{9}$$

$$b^{(i)}=b^{(i)}-\rho\frac{\partial}{\partial b^{(i)}}\Gamma(\theta) \tag{10}$$

1.4　声纹识别

对于输入 $x=(x_1,x_2,\cdots,x_k)$,在第 i 类的概率 P_i 计算公式如下:

$$P_i=\frac{\exp(Z_i)}{\sum_{i=1}^{k}\exp(Z_i)},i=1,2,\cdots,k \tag{11}$$

其中,Z_i 是 Softmax 层的输入,P_i 是 Softmax 层的输出。最大 P_i 对应的声纹类别即为识别结果。

2　实验与结果分析

2.1　实验平台

本文实验是在操作系统为 UBUNTU1404,GPU 为 NVIDIA GEFORCE GTX 1050,内存大小为16 GB,软件平台为 PYTHON3. 5、TENSORFLOW 1.2.1,界面软件为跨平台的 Qt 的机器上实现的。图 6 所示为应用本文算法开发的一款小样本声纹识别软件操作界面。

图 6　声纹识别软件操作界面截图

此外,笔者还开发了一款基于声纹开锁的智能信报箱,如图 7 所示。

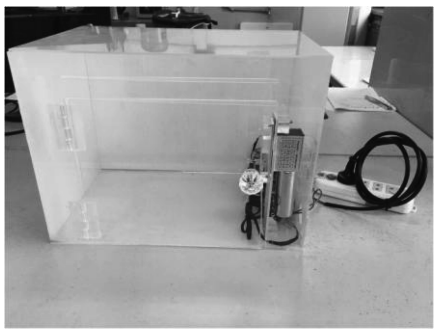

图 7　智能声纹信报箱实物图

智能信报箱的实物连接如图 8 所示,主要由 1-树莓派(自带 WiFi 模块)、2-电子锁、3-电源适配器、4-继电器组成。

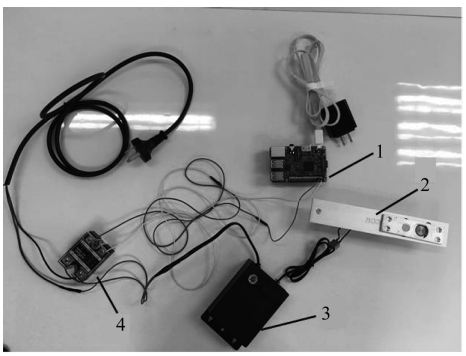

图 8　智能声纹信报箱的实物连接图

智能信报箱的原理图如图 9 所示,在移动设备上采集语音信号,并在服务器上执行特征提取和识别。通过服务器和移动端之间的通信以及服务器和树莓派之间的通信,打开智能声纹信报箱,用户可以

在手机上完成整个过程。

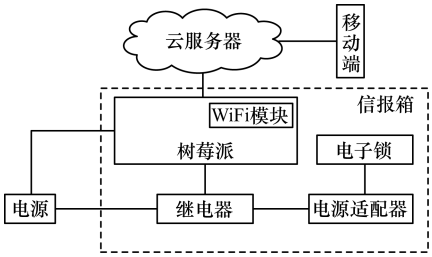

图 9 智能声纹信报箱的原理图

2.2 实验训练集与测试集

本文实验数据集来源于美国国家标准技术局的 TIMIT 数据库,其中包括了 630 个来自美国不同区域的人(每人 10 句话)。在 630 人中任选 430 人构成训练集,剩余 200 人作为验证集和测试集。截取每人10 个2 s 的 wav 格式语音片段,对每个语音片段对应生成一张语谱图,并将每张语谱图根据图像增多算法生成 50 张,即训练集包括 430 × 50 张语谱图,在剩余 200 × 50 张语谱图中,验证集∶测试集 =7∶3。

2.3 实验测量参数

记识别正确的声纹数为 N_r,测试总声纹数为 N_n,声纹识别率 R 计算公式如下:

$$R=\frac{N_r}{N_n} \tag{12}$$

此外,实验还需要测量网络训练时间、Loss 函数值。

2.4 声纹识别方法对比实验

本文方法与常用于语音识别的 GMM[15]、GMM-SVM[15]、GMM-UBM[16]方法进行对比实验,声纹识别结果如表 1 所示。

表 1 4 种方法识别率的比较 %

识别方法	识别率
GMM 方法	91.3
GMM-SVM 方法	95.8
GMM-UBM 方法	96.4
FBN-Alexnet 方法	98.6

从实验结果可以看出,基于深度学习的声纹识别率高于传统识别模型。现阶段基于深度学习的声纹识别,模型通过对大量数据训练,自动学习数据的潜在特征,包括 MFCC 特征、解剖学声学特征(倒频谱、共振峰)、韵律特征、通道信息等。而传统的识别模型学习语音的单一特征,很难保证提取的语音特征的质量,甚至可能会丢失一些重要特征。与传统识别模型相比,深度模型能提取更多潜在的声学特征,从而提高了声纹识别率。

2.5 训练样本数对识别率的影响

在实验中,采用 FBN-Alexnet 网络模型来测试训练样本数对识别率产生的影响,将上述 430 人的声纹数据(人均 10、20、30、40、50 个语谱图)分别作为训练集,将剩余 200 人的声纹数据(人均 10、20、30、40、50 个语谱图)作为验证和测试样本,验证集∶测试集 =7∶3。不同训练样本容量的识别率结果如图 10 所示。

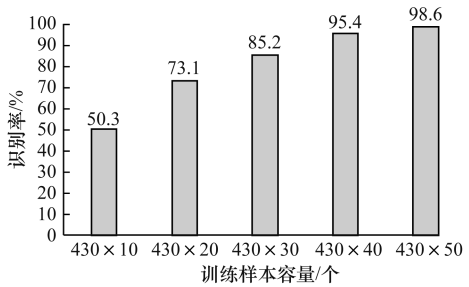

图 10 不同训练样本容量的声纹识别率

实验结果表明,随着训练样本容量的增加,FBN-Alexnet 网络模型的识别率呈现上升趋势,结果符合模式识别的规律。训练样本越少,识别率越低,原因在于网络并没有达到收敛状态。当样本容量达到一定数目时,网络到达收敛状态,识别率可达 98%以上。

2.6 迭代次数对识别率的影响

本文实验采用上述 430 × 50 张语谱图训练 Alexnet 网络和 FBN-Alexnet 网络,在相同的训练集下比较 FBN 对网络收敛速度的影响。FBN-Alexnet 网络的训练迭代次数不可避免地会对识别率有一定影响,不同迭代次数条件下的识别率结果如表 2 所示。

表 2 迭代次数对识别率的影响

迭代次数	识别率/%
50	90.8
100	96.1
150	98.1
200	98.6

由表 2 可知,随着迭代次数的增加,声纹识别率逐渐提高。当网络达到收敛状态时,随着迭代次数增加,识别率趋于稳定。

图 11 是训练过程中网络的损失函数输出值(Loss 值)的变化,反映了网络是否正确收敛。从图 11可以看出,随着网络训练的进行,Loss 值越来越小;刚开始训练时,Loss 值下降的速度快,但随着训练迭代的进行,Loss 值下降的速度趋于平稳,并且波动性也较小。

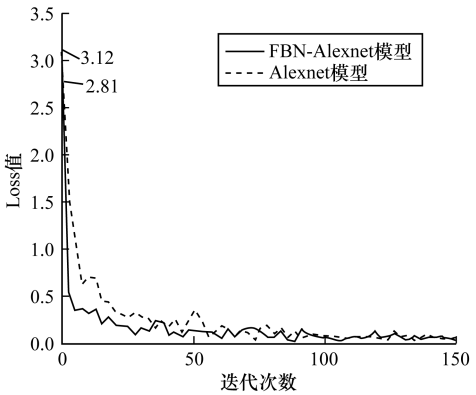

图 11　迭代次数对 Loss 值变化的影响

2.7　网络训练时间的对比实验

本文实验设定 Alexnet 基础学习率为 0.01,FBN-Alexnet 的学习率为 0.05,网络整体代价误差(Loss)为0.01。5 次网络训练时间的平均值如表 3 所示。

表 3　2 种模型训练时间的比较　　h

模型	训练时间
Alexnet	4.27
FBN-Alexnet	2.21

实验结果表明,在网络训练过程中,FBN 发挥了良好的作用。FBN-Alexnet 的网络训练时间比原 Alexnet 网络减少 48.2% 。这是由于 FBN 操作将数据归一化到零均值和单位方差,能够加速收敛,并用小批量的样本方差和均值代替总的样本方差和均值,降低了计算量,而且通过改进网络结构模型(改变卷积核的大小和数量,减少第 2 层全连接层的节点数),进一步减少了网络的运算消耗,缩短了网络的训练时间。

2.8　结果分析

本文方法摒弃了传统的语音识别框架,采用基于语谱图和 FBN-Alexnet 神经网络的声纹识别方法。该方法的优势是语谱图充分包含了说话人的语音特点和 FBN-Alexnet 神经网络能够自动提取潜在的声纹特征,相比传统的声纹识别方法只能获取单一特征,能够显著提高声纹识别率。

本文采用 CNN 架构以提升网络收敛速度。由于网络参数变化通常会导致网络各层的输入数据分布发生变化,造成网络不同层的不同维度间的数据所需要的学习率不一样。在训练网络时,通常需要选取最小的学习率来进行训练,从而防止网络过拟合,保证梯度的正常下降。FBN 操作将数据归一化到零均值和单位方差,即使用较大的学习率也能保证梯度的正常下降。此外,通过改进网络模型,进一步缩短了网络训练时间。

3　结束语

本文提出一种深度模型下的小样本声纹识别方法,将小样本语谱图通过图像增多算法生成多张语谱图,解决在实际应用中声纹数据不足的问题。利用声纹数据训练 FBN-Alexnet 神经网络,并在卷积过程中加入 FBN,加快网络收敛速度,缩短网络训练时间。该方法利用深度 FBN-Alexnet 神经网络提取声纹潜在的特征,提高了声纹识别率。在此基础上,通过采用交叉熵损失函数自适应地调整网络参数,使网络模型更适用于小样本声纹数据集,进一步提高声纹识别率,解决了传统语音识别模型和小样本声纹识别率低的问题。实验结果表明,该方法比 Alexnet 网络训练时间缩短了 48.2% ,识别率超过 98% 。

参考文献

[1] SALEEM M M, HANSEN J H L. A discriminative unsupervised method for speaker recognition using deep learning[C]//Proceedings of IEEE International Workshop on Machine Learning for Signal Processing. Washington D. C., USA:IEEE Press,2016:1-5.

[2] 陈锦飞,徐欣. 基于梅尔频率倒谱系数与动态时间规整的安卓声纹解锁系统[J]. 计算机工程,2017,43(2):201-205.

[3] BAE H S,LEE H J,LEE S G. Voice recognition based on adaptive MFCC and deep learning[C]//Proceedings of IEEE Conference on Industrial Electronics and Applications. Washington D. C., USA: IEEE Press, 2016:1542-1546.

[4] AZMY M M. Classification of lung sounds based on linear prediction cepstral coefficients and support vector machine[C]//Proceedings of Applied Electrical Engineering and Computing Technologies. Washington D. C., USA:IEEE Press,2015:1-5.

[5] 林舒都,邵曦. 基于 i-vector 和深度学习的说话人识别[J]. 计算机技术与发展,2017,27(6):66-71.

[6] CIRESAN D D, MEIER U, MASCI J, et al. Flexible, high performance convolutional neural networks for image classification[C]//Proceedings of the International Joint Conference on Artificial Intelligence. Palo Alto,USA:AAAI Press,2011:1237-1242.

[7] ABDEL-HAMID O,MOHAMED A R,JIANG H,et al. Convolutional neural networks for speech recognition[J]. IEEE/ACM Transactions on Audio Speech and Language Processing,2014,22(10):1533-1545.

[8] HUANG J T,LI J,GONG Y. An analysis of convolutional neural networks for speech recognition[C]//Proceedings of IEEE International Conference on Acoustics,Speech and Signal Processing. Washington D. C., USA: IEEE Press, 2015:4989-4993.

[9] ZHANG Y,PEZESHKI M,BRAKEL P,et al. Towards end-to-end speech recognition with deep convolutional neural networks[EB/OL]. [2017-11-12]. https://arxiv.org/abs/1701.02720.

(下转第 272 页)

表 5 2 种算法切片平均 PSNR 对比 dB

转码序列	PSNR		ΔPSNR
	x264 算法	本文算法	
480p60	42.80	42.96	0.16
360p60	41.60	41.63	0.03
240p60	40.63	40.55	-0.08
平均值	—	—	0.04

3 结束语

传统的码率控制算法应用于自适应码率视频直播时,视频切片的码率稳定难以控制,这使得自适应算法判断失准、客户端视频易出现卡顿的状况。本文提出基于切片结构的比特分配和码率控制算法,分别在切片级别和帧级别进行固定分配和权值分配,并利用预测模型调整每行的量化参数,实现单帧大小的准确控制。实验结果表明,使用该算法编码出的视频切片码率波动比 x264 算法减少 76%,PSNR 提升 0.04 dB。下一步将改进比特分配算法的内容自适应性,根据不同的视频内容动态调整权值系数,以达到优化视频质量的目的。

参考文献

[1] STOCKHAMMER T. Dynamic adaptive streaming over HTTP--: standards and design principles [C]// Proceedings of ACM Conference on Multimedia Systems. New York, USA: ACM Press, 2011: 133-144.

[2] THANG T C, LE H T, PHAM A T, et al. An evaluation of bitrate adaptation methods for HTTP live streaming [J]. IEEE Journal on Selected Areas in Communications, 2014, 32(4): 693-705.

[3] STOCKHAMMER T, SODAGAR I. MPEG DASH: the enabler standard for video delivery over theinternet [J]. SMPTE Motion Imaging Journal, 2012, 121(5): 40-46.

[4] YIN X, JINDAL A, SEKAR V, et al. A control-theoretic approach for dynamic adaptive video streaming over HTTP [J]. ACM SIGCOMM Computer Communication Review, 2015, 45(5): 325-338.

[5] SODAGAR I. The MPEG-DASH standard for multimedia streaming over the internet [J]. IEEE Multimedia, 2011, 18(4): 62-67.

[6] YAMAGISHI K, HAYASHI T. Parametric quality-estimation model for adaptive-bitrate streaming services [J]. IEEE Transactions on Multimedia, 2017, 19(7): 1545-1557.

[7] SEUFERT M, EGGER S, SLANINA M, et al. A survey on quality of experience of HTTP adaptive streaming [J]. IEEE Communications Surveys and Tutorials, 2015, 17(1): 469-492.

[8] SHUAI Y, GORIUS M, HERFET T. Low-latency dynamic adaptive video streaming [C]// Proceedings of IEEE International Symposium on Broadband Multimedia Systems and Broadcasting. Washington D. C., USA: IEEE Press, 2014: 1-6.

[9] HUANG T Y, JOHARI R, MCKEOWN N, et al. A buffer-based approach to rate adaptation: evidence from a large video streaming service [J]. ACM SIGCOMM Computer Communication Review, 2015, 44(4): 187-198.

[10] JULURI P, TAMARAPALLI V, MEDHI D. SARA: segment aware rate adaptation algorithm for dynamic adaptive streaming over HTTP [C]// Proceedings of International Conference on Communication Workshop. Washington D. C., USA: IEEE Press, 2015: 1765-1770.

[11] LI B, LI H, LI L, et al. (λ) domain rate control algorithm for high efficiency video coding [J]. IEEE Transactions on Image Processing, 2014, 23(9): 3841-3854.

[12] YISHUT, QIANG S, YANWEI L, et al. Average bit-rate algorithm optimization for rate control of X264 [J]. Journal of Computer Applications, 2013, 33(3): 680-683.

[13] XIANG G, JIA H, YANG M, et al. A novel adaptive quantization method for videocoding [J]. Multimedia Tools and Applications, 2018, 77(12): 14817-14840.

[14] 田一姝,沈强,刘延伟,等. X264 的平均比特率控制算法优化[J]. 计算机应用,2013,33(3):680-683.

[15] 李维,杨付正,任鹏. 考虑视频内容的 H.265/HEVC 帧层码率分配算法[J]. 通信学报,2015,36(9):76-81.

编辑 樊丽娜

(上接第 267 页)

[10] OQUAB M, BOTTOU L, LAPTEV I, et al. Learning and transferring mid-level image representations using convolutional neural networks [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2014: 1717-1724.

[11] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C]// Proceedings of International Conference on Neural Information Processing Systems. Red Hook, USA: Curran Associates Inc., 2012: 1097-1105.

[12] LIU X, KAN M, WU W. et al. VIPLFaceNet: an open source deep face recognition SDK [J]. Frontiers of Computer Science, 2017, 11(2): 208-218.

[13] BEIGY H, MEYBODI M R. Adaptation of parameters of BP algorithm using learning automata [C]// Proceedings of Brazilian Symposium on Neural Networks. Washington D. C., USA: IEEE Press, 2000: 24-31.

[14] KLINE D M, BERARDI V L. Revisiting squared-error and cross-entropy functions for training neural network classifiers [J]. Neural Computing and Applications, 2005, 14(4): 310-318.

[15] 赵立辉,毛竹,霍春宝,等. 基于 GMM-SVM 的说话人识别系统研究[J]. 工矿自动化,2014,40(5):49-53.

[16] 周国鑫,高勇. 基于 GMM-UBM 模型的说话人辨识研究[J]. 无线电工程,2014,44(12):14-17.

编辑 刘盛龄