2014年 14t h 国际控制、自动化与系统会议（ICCAS 20 14）
10月22日至韩国京畿道 KINTEX，25、20 14

（最佳论文奖 d)

音频库：用于音频事件识别的高级声学信号表示

Tushar Sandhan、Sukanya Sonowal 和 Jin Young Cho i

韩国首尔国立大学电气与计算机工程系，韩国首尔
151-742{tushar, sukanya, jychoi}@snu.ac.kr

摘要：自动音频事件识别对于促进人机交互发挥着关键作用，并广泛应用于工业自动化、控制和监控系统。音频事件由复杂且相互谐波纠缠的声学模式构成。音频识别主要依赖低层和中层特征；这些特征已展现出识别能力，但计算成本高、语义含义有限。本文提出了一种计算效率更高的音频识别新框架。音频库是一种新型音频高层表示，由具有区分性的音频检测器构成，每个检测器在频率—时间空间中表示一个音频类别。我们使用非负矩阵分解降低所得特征向量的维度，同时保留其判别能力和丰富的语义信息。采用多种分类器（SVM、神经网络、高斯过程分类和 K 近邻）获得的高音频识别性能，证明了所提方法的有效性。

关键词： 音频事件分类、非负矩阵分解、特征构建、高斯过程 s.

1. 引言

音频事件识别（AER）或分类是听觉场景分析的一个子领域，旨在无需人力的情况下自动理解音频数据。近年来，该领域受到了广泛关注，因为除了在音频检索、索引和自动标注方面具有直接应用外，它还有望在感知型接口中发挥关键作用，例如会议室中的计算机或机器人辅助、工业自动化和过程控制、在多样化环境中工作的移动机器人 [1]，人类行为识别 [17] 以及视频监控 [3]，[13]。例如，在视频监控系统中，将音频信息与视频结合使用，可以提升系统对场景的理解能力，尤其是在昏暗的照明条件下或摄像机视野之外的区域；否则，仅使用视频进行的活动识别将因缺乏感知信息而无法实现。在会议室助理机器人的情况下，检测或分类与人类存在相关的某些音频事件，例如椅子移动、开门、键盘打字等，可能有助于机器人检测并识别房间内发生的活动。在这些场景中，整个控制与自动化系统的性能高度依赖于其正确分类音频事件的能力。

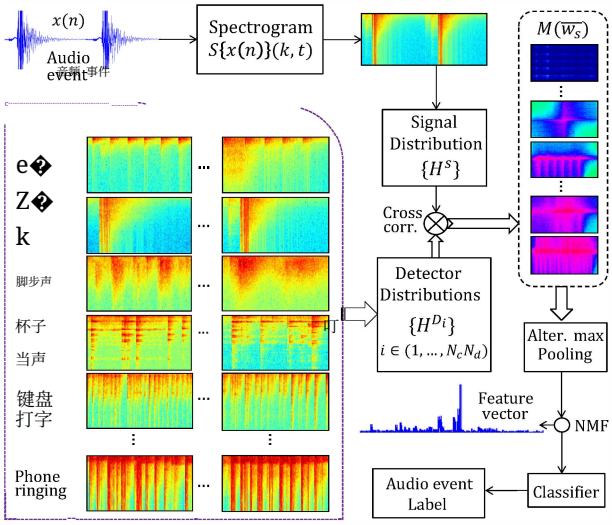


图1所提出的音频事件分类（AEC）框架概览。音频库（第2节）是音频事件的高层表示。它存储了一组音频检测器（第2.2 节），用于表示频谱图空间（第2.1节）中的每个音频类别。它通过拼接来自每个检测器的最大池化互相关（检测器与信号分布之间）模式，生成特征向量（第 2.3 节）。

以往大多数音频事件识别工作 [2], [4], [5], [10] 都使用标准特征，例如梅尔频率倒谱系数（MFCC）、过零率、子带能量、短时能量及其统计属性，如均值、标准差和熵。但这些低层和中层特征所能捕获的声学语义有限，因此生成的表示判别能力不足，从而搭建

对于稳健的音频识别而言，它们都不是明显的赢家。在本文中，我们提出了音频库，这是一种针对音频事件的新型高级表示，其语义信息丰富，并且本质上具有很强的判别性。音频库探索了一组具有独特性的音频检测器，这些检测器最终如同高维“音频空间”的基底，从而提供语义信息丰富的高级表示。图1概述了基于音频库的特征构建过程和音频事件识别框架。

978-89-93215-06-995560114/$15 ©ICROS

授权使用仅限于：印度理工学院坎普尔分校。于 2023 年 4 月 5 日 06:03:33 UTC 从 IEEE Xplore 下载。使用限制适用。

低层特征和中层特征无法全面表示音频数据中潜在的谐波结构。因此，传统的音频识别方法尝试通过利用隐马尔可夫模型 (HMM) [22] 和高斯混合模型 (GMM) [20]的数据结构来揭示隐藏的声学模式。这些模型需要大量训练数据。将其他领域的方法用于音频事件识别（AER），例如基于图的模型[24] 和稀疏表示 [21]，并不直接，而且计算成本也很高，尤其是对于高维特征向量而言。相比之下，所提出的音频库特征表示具有足够的判别能力，即使使用简单的分类器（如 k 近邻 [18]），也能产生良好的识别结果。

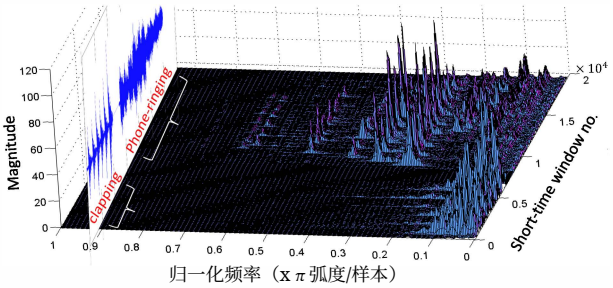


图 2 两种不同的音频事件（拍手声、电话铃声）及其对应的频谱图。两种信号呈现出不同的频率模式，其中电话铃声类别的频率成分高得多。

我们应用了多种学习算法，以基于这些音频库特征衡量音频分类的准确率。此外，我们还通过改变库检测器的数量和训练数据量进行了多项实验，以评估所提出框架对各种参数的依赖性。我们的主要贡献包括：提出一种利用音频检测器构建的全新、鲁棒的高层特征表示，并通过采用多种学习算法开展实验研究，评估该表示对分类任务的适用性。我们使用 UPC-TALP 数据库 [9] 进行音频事件识别，该数据库包含一组彼此独立的会议室声学事件。它是一个平衡数据集 [7]，即每个音频类别都有足够数量的样本。实验结果表明，与传统特征提取方法相比，所提出的框架能够取得更高的 AER率。

2. 音频库表示

音频信号由复杂的声学模式组成，这些模式可能会在信号中重复数次。因此，音频库试图利用这一事实，将音频表示为许多声学检测器所收集的输出，每个检测器都会产生一个相关模式。尽管乍看之下，音频库似乎与物体库[16], 有关，但音频分类问题显然不同于图像分类。对于后者而言，显然，图像由树木、水、山脉、人、建筑物等基本物体构成。因此，物体库明确由独立的物体检测器组成。但对于音频信号而言，构成音频的原子声学模式很难彼此分离。因此，我们采用基于模板的音频检测器，其中具有代表性的音频样本本身充当模板。为了增强对响度（振幅）的鲁棒性，我们使用特征分布相似性度量（巴塔查亚系数）来匹配音频检测器。通过在频域中分析信号，可以实现对音高（音频的基频）的不变性。为了考虑音色（音高品质的差异），我们从音频空间中采样具有代表性的模板。

2.1 特征提取

音频的频谱内容会随时间变化，因此对整个信号应用离散傅里叶变换（DFT）无法揭示频谱内容的变化（非平稳信号）。但在短时间内，音频可以被视为平稳的。因此，我们的特征空间由音频频谱图构成，它是短时信号的功率谱集合，如下所示，

$$S\{x(n)\}(k,t) = \left| \sum_{n=0}^{N-1} x(n+tM)\, w(n)\, e^{-j\frac{2\pi}{N}nk} \right|^2, \quad (1)$$

其中，$x(n)$ 是音频信号，$w(n)$ 是长度为 N 的（短时）分析窗（例如汉明窗），k 是频率 bin 的索引，t 是时间帧的索引，而 M 是分帧步长（相邻两帧之间间隔的样本数）。人类听觉系统能够分辨各种音频事件，因为它们呈现出典型的时变频率模式。图 2 展示了两种不同音频事件的频谱图和信号波形。

2.2 选择检测器组

音频库在选择检测器类型方面提供了很大的灵活性；事实上，可以同时使用不同类型的检测器（采用不同的特征）。在我们的实现中，我们使用独特的信号频谱图作为检测器，因为它能够利用单个模板有效定位事件、计算效率高（使用快速傅里叶变换 [23]）、具备音高分析能力，并且可自然地解释为将音频分解为空间—时间频率。为了捕捉具有判别性的音频模式，各库检测器必须彼此不同。

令 xi 为音频（事件）类别 $c \in$ {I,⋯ ,Nc} 中的样本。然后，对于每个 c, 我们使用 k 均值聚类 [25]，从 { S {xf} | $i = 1, 2, \ldots$ } 中选取 Nd 个代表作为检测器 {Di }f1d 。为增强对音色的鲁棒性，我们还从音频空间（来自同一事件但音质不同的音频）中采样检测器，并要求检测器大小足够小，以至于其中只包含底层音频事件的少数几次出现。所有检测器都以频谱图的形式存储，因此生成大小为 N $_D$ = Nc × Nd 的音频库。

授权许可使用仅限于：印度理工学院坎普尔分校。于2023年4月5日06:03:33（UTC）从 IEEE Xplore 下载。相关限制适用。

2.3 音频库特征向量

直观而言，音频库特征向量就是所有 N^{D} 个检测器对信号特征 $S\{x\}$. 的检测响应的串接。要获得大尺寸音频特征中的检测响应，需要将检测器放置在各个位置，并计算检测器 D 与信号特征 $S\{x\}$ 在对应位置上的频率分布（直方图 H^{D} 和 H^{s}）之间的相似度。设该相似度为 $sim\big(H^{s}(\bar{w}_d),\ H^{D}(\bar{w}_d)\big)$，其中 $\bar{w}_d = (k,t)$ 遍历检测器的频率—时间支持范围。信号特征的每个位置 $\bar{w}_s = (k,t)$ 处的全局相似度匹配度量 M (w s)， 通过对检测器上的各个相似度度量求和得到：

$$M(\bar{w}_s) = \sum_{\bar{w}_d} sim\big(H^{s}(\bar{w}_d), H^{D}(\bar{w}_d - \bar{w}_s)\big). \tag{2}$$

我们使用巴氏系数 [6] 作为相似度度量，原因有二：首先，它将取值限制在 0 和1，之间，1 表示完全匹配；其次，它易于高效实现。对于两个直方图 HI 和H2， 每个直方图有 B 个箱，且b 表示箱索引，巴氏系数定义如下：

$$sim(H^{1}, H^{2}) = \sum_{b=1}^{B} \sqrt{H_b^{1}, H_b^{2}}. \tag{3}$$

现在，将度量式 (3) 代入全局匹配度量式 (2)，并重新排列求和顺序，可得：

$$M(\bar{w}_s) = \sum_{b} \sum_{\bar{w}_d} \sqrt{H_b^{s}(\bar{w}_d)} \sqrt{H_b^{D}(\bar{w}_d - \bar{w}_s)}, \tag{4}$$

这相当于直方图各个箱之间的互相关，如下所示：$M\ (\bar{w}\ _s)$ $= \sum_b \sqrt{H_b^{s}} \star \sqrt{H_b^{D}}$. 这些二维相关可利用傅里叶变换的卷积定理在频域中高效计算 [8], , 将计算开销较大的相关运算替换为相对简单的逐点乘法，具体如下：

$$\sqrt{\widetilde{H}_b^{D}}\}\Big\}, \tag{5}$$

其中，$\widetilde{H}\ _b^{D}$ 是翻转后的检测器分布，$F\{\circ\}$ 和$F^{-1}\{\cdot\}$ 分别表示傅里叶变换及其逆变换。傅里叶变换通过使用快速傅里叶变换算法高效地实现。音频库特征向量通过使用交替最大池化操作，从全局相关模式! vI ($\bar{w}_s$), 中构建。在此操作中，输出相关模式随后被划分为$\{2^{0},\ 2^{2}, 2^{4}\}$ 个相等部分，然后将所有部分中的最大值一个接一个地拼接起来，从而为每个输入音频信号生成$N_d \times N_c \times 21$ 维特征向量。交替最大池化操作（如图3所示）能够高效且紧凑地捕捉相关性输出中的潜在变化。因此，它试图寻找最大相关的频率模式。

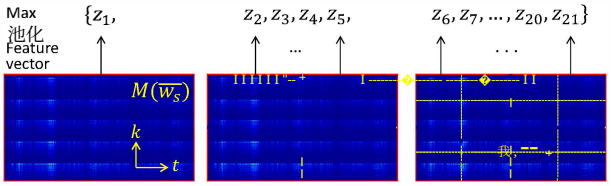


图 3 交替最大池化操作从每个检测器的相关性输出中提取特征向量。

2.4 非负矩阵分解（NMF）

NMF 是一种信号表示方法，用于在降维后的空间 [11] 中进行抗噪特征提取。其目标是最小化原始信号与其近似信号之间的距离。给定一个非负的$m \times n$ 矩阵X, NMF 将其分解为非负的$m \times k$ 矩阵 W 和非负的$k \times n$矩阵H ，使以下重构误差最小化，

$$J(W,H) = \|X - WH\|_F^2, \tag{6}$$

其中，$\|\cdot\|_F$ 表示矩阵的弗罗贝尼乌斯范数。W 的列向量是基向量，而H 中各列的值表示 W 的列向量所作的贡献，换言之，它表示信号值在基矩阵 W 上的分解。因此，NMF 通过仅允许基向量的加性组合，找到了基于“部件”的非负数据线性表示。可以根据（6）使用 [15] 提出的方法学习 W，然后在 W 固定的情况下学习 H 。在我们的工作中，X 是最大池化特征向量构成的矩阵（见图 3）。随后，将特征在基矩阵上的分解视为新的编码特征，并将其作为分类器的输入。

3. 实验

3.1 数据集描述与实验设置

UPC-TALP数据集[9] 包含在会议室环境中发生的14 类音频事件。这些事件的录制方式确保它们在时域上不重叠，录音采用 44.1kHz采样频率进行。表 1 展示了数据库中存在的 14 个音频类别及其标注标签。在我们的工作中，我们没有使用“未知”类别进行训练和评估，因为该类别意义不大，并且包含静音和噪声的出现情况。“开门”和“关门”在感知上听起来相似（因为两者都会产生相同的摔门声），因此我们将它们合并为统一的“门动作”类别，共有 12 1个样本。因此，我们最终得到 12 个类别，并基于这些类别报告音频事件识别结果。

在特征提取过程中，音频样本被降采样至 $\frac{1}{4}$ ，并采用 50% 的帧重叠进行 256 点离散傅里叶变换（DFT）。

授权许可使用仅限于：印度理工学院坎普尔分校。于 2023 年 4 月 05日06:03:33 UTC 从 IEEE Xplore 下载。适用相关限制。

表 1 UPC-TALP 数据库中的音频事件及其音频样本数量和总时长（秒）。

事件名称	#样本 \ 时间	事件名称	#样本 \ 时间
敲门声 (kn)	50 \ 64	钥匙串叮当声（kj）	65 \ 146
开门（do）	60 \ 66	键盘输入错误（kt）	66 \ 194
关门（ds）	61 \ 78	电话铃声 (pr)	116 \ 315
脚步声 (st)	73 \ 250	掌声 (ap)	60 \ 212
移动椅子（em）	76 \ 216	咳嗽（co）	65 \ 85
杯子叮当声（cj）	64 \ 187	笑声 (la)	64 \ 118
文书工作 (pw)	84 \ 300	未知 (un)	126 \ 89

用于计算频谱图。使用各种分类器进行 AER 分析时，从每个类别中随机选取 60% 的样本作为训练数据，其余 40% 的样本用于测试。测试阶段不会再次使用音频库中的任何音频。我们进行了大量实验（将在后续章节中介绍），以验证所提出特征的鲁棒性。

3.2 使用不同分类器的音频库

我们分别将几种分类器应用于音频库特征，以评估它们在音频事件识别中的适用性和判别能力。我们使用了 k近邻（kNN）、一对其余支持向量机（SVM-A）、一对一支持向量机[19]（SVM-O）、高斯过程 [12]（GP）和神经网络 [26]（NN）进行分类。对于 k近邻，k 设为 5。对于 SVM-A，我们使用径向基函数核，其中$C = 150$，$\sigma = 75$；对于 SVM-O，则$C = 100$，$\sigma = 60$。这些参数通过五折交叉验证确定。

高斯过程由其均值和协方差函数完全确定。GP 分类器的均值函数设为常数，并使用各向同性平方指数核作为协方差函数。我们使用 GPML 工具箱 [14], 实现了 GP，该工具箱可学习 GP 超参数并预测类别标签。由于 GP 分类器本质上是二分类器，我们采用一对一方法将其扩展用于多分类。对于神经网络，隐藏层的数量任意设为 300。

对于每个分类器，我们都进行了 5 次识别实验。也就是说，在 5 次运行中，每次都将数据随机分配到训练集和测试集，然后应用分类器。图 A 展示了使用音频库特征获得的平均 AER 准确率，每个柱形顶部的误差棒表示标准差。我们可以看到，对于 k近邻这类简单的分类器，AER 的表现相当不错；而对于更.....

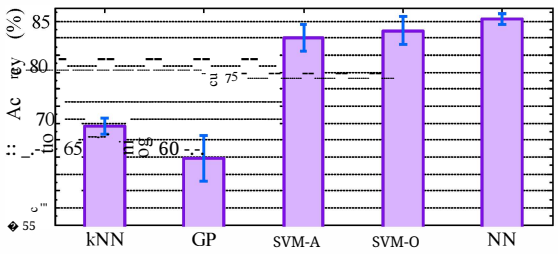


图 4 使用不同分类器的音频库进行音频事件识别。

表 2 在 5 次实验运行中，不同分类器对音频库特征进行仿真的平均仿真时间，其中音频库大小固定为 $N_D = 48$

分类器	k-NN	SVM-A	SVM-O	GP	NN
仿真时间 （秒）	0.49	0.89	46.44	180	10.72

与 SVM 和神经网络等复杂分类器相比，性能显著更好。然而，对于高斯过程，准确率可能会因特征维度过高而下降。

计算时间：表 2 显示了上述各分类器在测试和训练中所需的总仿真时间。显而易见，k近邻由于结构简单，所需时间最少。另一方面，SVM-O 和高斯过程的速度明显较慢，因为这两种分类器采用一对一方式实现，其计算复杂度会随着类别数量呈平方级增长。SVM-A 的仿真时间相当短（与 k近邻相当），而神经网络分类器则稍慢一些。因此，若根据仿真时间和识别准确率对这 5 种分类器进行比较，SVM-A 显然胜出，因为它在速度几乎不受影响的情况下，性能相当出色。

探索 kNN分类器：通过将 k 从 1 到 100 每隔 10 取值一次，评估采用 kNN分类器时 AER 性能对参数k取值的依赖性。对于每个 k，记录 10 次识别性能，并在图 S 中报告平均性能，同时用误差棒标示相应的标准差。从图中可以看出，随着 k 增大，识别性能会下降。因此，k 值越小，AER 准确率越高。

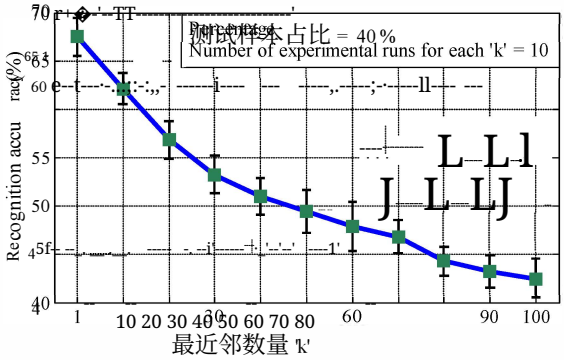


图 5 使用音频库特征的 kNN分类器的平均识别准确率，其中最近邻数 (k) 从 1 到 100 不等。均值和标准差分别用标记和误差棒表示。

探索神经网络：我们还分析了神经网络分类器中隐藏层数量对识别性能的影响。隐藏层数量从 10 变化到1000，并通过与 k近邻实验类似的方式，对 10 次运行结果取平均值，测量每个隐藏层数值下的识别性能。

授权许可使用仅限：印度理工学院坎普尔分校。于 2023 年 4 月 5 日 06:03:33 UTC 从 IEEE Xplore 下载。适用相关限制。

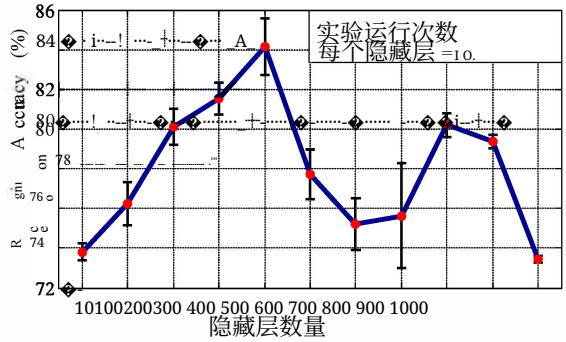
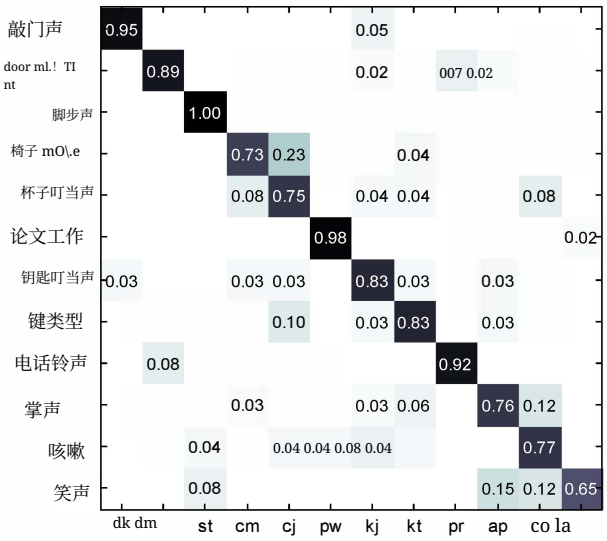


图6 使用音频库特征向量的神经网络分类器的平均识别准确率，其中隐藏层数量从10变化到1000。圆形标记表示10次运行所得的平均值，误差棒表示标准差。

图 6展示了每个隐藏层数值所获得的 AER准确率。结果表明，当隐藏层数量在 300到 400 之间时，识别性能相当出色。对于其他隐藏层数量，性能也没有显著下降。

3.3 与其他方法的比较

音频库（大小为 $N_D = 4S$）的特征与使用相同SVM-O分类器的 3组传统特征进行比较。特征“set-A”由32 维（32-D）对数滤波器组能量、过零率（ZCR）以及信号拼接成一个34维特征向量组成。特征“set-B”由使用快速傅里叶变换系数 [S], 、ZCR和信号能量构成的2050 维特征向量组成。特征“set-C”由 13 维梅尔频率倒谱系数 [4]组成。表3显示了所达到的平均 AER率



F图7 使用 Audio B 的混淆矩阵（AER率） ank.

表3 传统特征集与音频库特征的对比。

方法	set-A [10]	set-B [8]	set-C [4]	Audio Bank
音频事件识别率（%）	51.44	44.16	73.38	84.59

在实验的5 次运行中，音频库特征的性能显著优于传统特征。其每类性能以图7中混淆矩阵的形式呈现。对角线上的高数值表明所提出特征具有类间可辨识性。该方法对“椅子移动”类别的识别稍有混淆，将其中一些样本误分类为“杯子叮当声”。该方法对“脚步声”实现了完美的识别率，并且对“敲门声”“纸张处理”和“电话铃声”类别的准确率超过 90%。

3.4 训练数据规模变化

训练数据量在决定分类性能方面起着至关重要的作用。训练数据越多，泛化性能就越好。因此，将每个类别的训练数据量从 10% 调整到 90%（相应地，每个类别的测试数据量从 90% 变为 10%）；所得 AER 性能汇总于图 S 中。图中还显示了神经网络分类器的不稳定性（这是由于其对参数初始化的依赖性较强）。随着训练数据量增加，SVM-A 和 SVM-O 的性能逐渐提升。可以看到，即使每个类别的训练数据仅占 40%，AER率也超过 50%。这表明所提出的特征具有良好的判别能力和类别泛化能力。

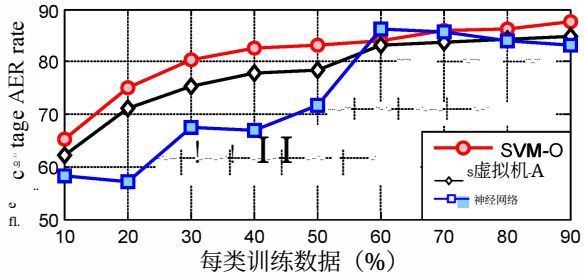


图 S 不同训练数据量下三个分类器的音频事件识别（AER）率。

3.5 音频库大小变化

所提出的框架非常通用，只需向现有音频库中添加更多新的distinctive检测器，即可轻松适应不同的音频事件识别（AER）设置。然而，音频库越大并不一定会带来更好的性能，因为随着库检测器数量(N_D) 增加，音频库特征向量的维度也会增加，而由于维度灾难，AER率可能会受到影响。为了评估这些影响，我们通过逐步改变音频库大小来执行AER。音频库由每个音频中的 distinctive N_d 检测器组成

授权许可使用仅限于：印度理工学院坎普尔分校。于 2023 年 4 月 5 日 06:03:33（UTC）从 IEEE Xplore 下载。使用受限。

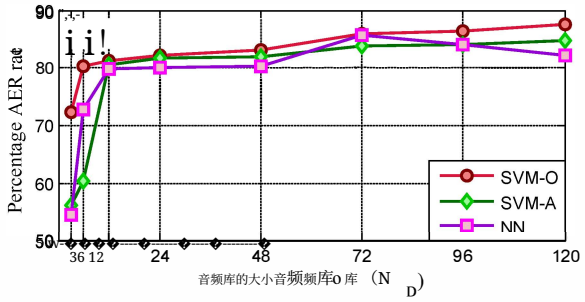


图9分析音频事件识别率受音频库大小的影响。定义符号如下：小型音频库（检测器 N_D = ），较小型音频库 $^{3}(N_D)$，大型音频库 $= 12(N_D)$。 > 50

类别（参见第 2.2 节）。我们将音频库的大小从 3 调整到 120。图 9 展示了 AER率（使用每个类别 60% 的训练数据）与音频库大小之间的关系。对于 神经网络 分类器，由于维度灾难以及 神经网络 分类器的不稳定性（参见第 3.3 节），随着音频库增大，AER率略有下降。微型音频库 ($N_D = 3$ 或 6) 的检测器数量少于类别总数（ 12 个），因此生成的特征向量无法表示某些类别，导致微型音频库的表现不佳。但即使是小型音频库 (N_D =12，即每个类别一个检测器)，也达到了与大型音频库 ($N_D > 50$) 相当的准确率（80%）。音频库的大小对其性能影响如此之小，这一出人意料的稳定性可归因于它能够在音频空间中表示事件，从而生成具有组稀疏特性的高层表示。对于特定类别的音频事件，只有与之密切相关的检测器会产生较高的检测响应；而语义上关系较远的检测器则会保持休眠状态。因此，即使使用小型音频库，生成的组稀疏且语义丰富的特征也足以实现较高的 AER性能。

4. 结论

我们提出了音频库（Audio Bank），一种面向音频事件识别（AER）的新型高层高效音频表示，旨在探索音频数据中存在的底层谐波结构。它由具有区别性的音频检测器组成，在频率—时间空间中分别表示每个音频类别。通过对大量较小的音频检测器进行适当池化，该方法在区分音频事件方面产生了优于低层特征的特征。音频库大小所带来的特征稳定性，以及使用多种分类器（SVM、神经网络、高斯过程和 k-NN）所获得的高音频事件识别性能，表明了所提出方法的有效性。

参考文献

[I] C. Miiller，JI. Biel、E. Kim 和 D，Rosario，“面向汽车应用的语音重叠声学事件检测”，载于 Interspeech 论文集，第 2590-2593 页，2008 年。

[2] C. Clavel、T. Ehrette 和 G. Richard，《基于音频的监控系统中的事件检测》，IEEE 国际多媒体与博览会会议，2005年。[3] T. Sandhan、T. Srivastava、A. Sethi 和 J. Choi，《通过揭示低频模式对监控视频中的异常事件检测进行无监督学习》，IEEE国际图像与视频计算会议，新西兰（IVCNZ），2013 年。[4] A. Mesaros、T. Heittola、A. Eronen和 T. Virtanen，《真实录音中的声学事件检测》，欧洲信号处理会议，2010 年。[5] P. K. Atrey、M.Maddage和M. S. Kankanhalli，《用于多媒体监控的基于音频的事件检测》，IEEE国际声学、语音与信号处理会议（ICASSP），2006年。[6] S. H. Chao，《概率密度函数之间距离/相似性度量的综合调查》，《应用科学中的数学模型与方法国际期刊》，2007 年。[7] T. Sandhan 和 J.Choi，《通过部分引导的混合采样处理不平衡数据集以进行模式识别》，IEEE 国际模式识别会议（ICPR），2014 年。[8] A.Y.Oppenheim 和 R. W. Schafer，《离散时间信号处理》，PrenticeHall，第 2版，1999 年。[9] A. Temko、R. Malkin、C. Zieger、D. Macho、C.Nadeu 和M. Omologo，《声学事件检测与分类系统的 CLEAR 评估》，《用于人类感知的多模态技术》，第311–322 页，Springer，2007 年。[10] A. Temko和C.Nadeu，《会议室环境中的声学事件检测》，《模式识别快报》，第 30卷，第 14期，第1281–1288 页，2009 年。[II] D. D.Lee 和 H. S.Seung，《通过非负矩阵分解学习物体的组成部分》，《自然》，第 401 卷，1999年。[12] R. M. Neal，《使用高斯过程先验进行回归和分类》，《贝叶斯统计 6》，牛津大学出版社，1998年。[13] T. Sandhan、Y. Yoo、H. Yoo、S. Yun 和 M. Byeon，《通过轨迹的过采样时间序列表示进行多任务学习，以实现交通运动模式识别》，第 11 届IEEE国际高级视频与基于信号的监控会议（AVSS），2014 年。[14] C. E.Rasmussen和 H. Nickisch，《用于机器学习的高斯过程（gpml）工具箱》，《机器学习研究期刊》，第II 卷，第 3011–3015页，2010 年。[15] J. Mairal、F.Bach、J. Ponce 和 G.Sapiro，《矩阵分解与稀疏编码的在线学习》，《机器学习研究期刊》，第II 卷，第 19–60 页，2010 年。[16] L. J. Li、H. Su、E.P.Xing和F. F. Li，《物体库：用于场景分类和语义特征稀疏化的高层次图像表示》，《神经信息处理系统会议论文集》，第2 卷，第 3 期，第 5 页，2010 年。[17] T.Sandhan 和 J. Choi，《用于动作和手势识别的频谱图与多特征联合稀疏表示》，IEEE 国际图像处理会议（ICIP），2014 年。[18] T. Cover 和P. Hart，《最近邻模式分类》，《IEEE 信息论汇刊》，第 13卷，第 1 期，第 21–27 页，1967 年。[19] C. Cortes 和 Y.Vapnik，《支持向量机》，《机器学习》，第20卷，第 3 期，第 273–297 页，1995 年。[20] D.Reynolds，《高斯混合模型》，《生物统计学百科全书》，第 659–663 页，2009 年。[21] T.Sandhan、H. Chang 和 J. Choi，《用于指纹识别的抽象拉东轮廓》，IEEE国际图像处理会议（ICIP），2013年。[22] S.R. Eddy，《隐马尔可夫模型》，《结构生物学当前观点》，第 6卷，第 3 期，第 361–365页，1996年。[23] J. S. Walker，《快速傅里叶变换》，第2 版，CRC Press，1996 年 8 月。[24] Y.Yoo、T.Sandhan、J.Choi、S. Kim，《面向大量蛋白质家族的同时聚类与基序建模》，IEEE 国际生物信息学与生物医学会议（BIBM），2013 年。[25] J.A. Hartigan和 M. A. Wong，《算法 AS 136：一种 k均值聚类算法》，《应用统计学》，第 100–108 页，1979 年。[26] T. M. Hagan、H. B.Demuth 和 M.H.Beale，《神经网络设计》，波士顿：Pws Pub。, 1996

授权许可使用仅限于：印度理工学院坎普尔分校。下载时间：2023年4月05日06:03:33 UTC，来源为 IEEE Xplore。限制条款适用。