

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 9 月 17 日 (17.09.2020)



(10) 国际公布号
WO 2020/181998 A1

- (51) 国际专利分类号:
G10L 25/78 (2013.01) *G10L 25/51* (2013.01)
G10L 25/30 (2013.01)
- (21) 国际申请号: PCT/CN2020/077189
- (22) 国际申请日: 2020 年 2 月 28 日 (28.02.2020)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201910179592.0 2019年3月11日 (11.03.2019) CN
- (71) 申请人: 江苏大学 (JIANGSU UNIVERSITY) [CN/CN]; 中国江苏省镇江市学府路 301 号, Jiangsu 212013 (CN)。
- (72) 发明人: 毛启容 (MAO, Qirong); 中国江苏省镇江市学府路 301 号, Jiangsu 212013 (CN)。 陈静

静 (CHEN, Jingjing); 中国江苏省镇江市学府路 301 号, Jiangsu 212013 (CN)。 高利剑 (GAO, Lijian); 中国江苏省南京市清江南路, Jiangsu 210019 (CN)。 黄多林 (HUANG, Duolin); 中国江苏省镇江市学府路 301 号, Jiangsu 212013 (CN)。 张飞飞 (ZHANG, Feifei); 中国北京市海淀区中关村东路 95 号中国科学院自动化研究所, Beijing 100190 (CN)。

(74) 代理人: 北京德崇智捷知识产权代理有限公司 (JW IP LAW FIRM); 中国北京市朝阳区酒仙桥路 14 号 A5 楼 7 层 701, Beijing 100015 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS,

(54) Title: METHOD FOR DETECTING MIXED SOUND EVENT ON BASIS OF FACTOR DECOMPOSITION OF SUPERVISED VARIATIONAL ENCODER

(54) 发明名称: 一种基于监督变分编码器因素分解的混合声音事件检测方法

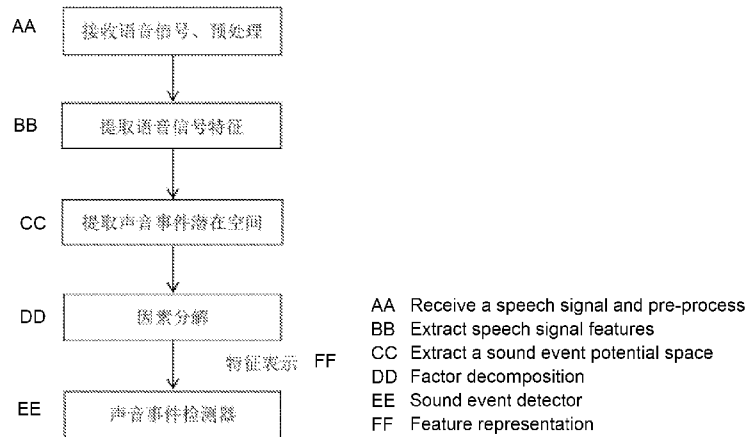


图 1

(57) Abstract: A method for detecting a mixed sound event on the basis of the factor decomposition of a supervised variational encoder, comprising the following steps: receiving a speech signal and performing pre-processing on the speech signal; extracting pre-processed speech signal features; using a supervised variational automatic encoder to extract a potential attribute space of sound events; using a factor decomposition method to decompose various factors that make up a mixed sound, and then learning to obtain a feature representation related to each specific sound event; and then using a corresponding sound event detector to detect whether a specific sound event occurs. The factor decomposition learning method is used to solve the problem in which the accuracy of sound event detection is low when there are many types of sound events in the mixed sound, thereby effectively improving the accuracy of sound event detection in real scenes, and may also be used for tasks such as speaker recognition.

JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

(57) 摘要: 一种基于监督变分编码器因素分解的混合声音事件检测方法, 包括如下步骤: 接收语音信号, 并对语音信号进行预处理; 提取预处理后的语音信号特征; 使用监督变分自动编码器提取声音事件潜在属性空间; 使用因素分解方法分解构成混合声音的各种因素, 进而学习得到每个特定声音事件相关的特征表示; 再使用对应的声音事件检测器检测特定声音事件是否发生。采用因素分解学习的方法解决混合声音中声音事件类别较多的情况下, 声音事件检测准确率不高的问题, 有效提高真实场景声音事件检测的准确度, 还可用于说话人识别等任务。

一种基于监督变分编码器因素分解的混合声音事件检测方法

技术领域

本发明涉及语音信号处理、模式识别等领域，特别涉及一种关于变分自动编码器和因素分解方法的声音事件检测方法。

背景技术

多类别声音事件检测是指从一个混有多种声音的事件当中，检测出每种事件是否发生。与传统少类别声音事件检测相比，在现实领域的适用性更广，在医学场景监听、交通场景声音事件检测等领域有着广阔的应用前景和实际意义。

传统的多类别声音事件检测方法主要是采用语音识别和模板匹配的思想，例如，使用混合高斯模型和以梅尔频率倒谱系数为特征的隐马尔可夫模型，或者是使用非负矩阵分解来表示每一种事件，并将其与声音事件词典进行匹配；然而，这种传统方法中的手工特征并不能完全表示不同的声音事件。最近，引入带有瓶颈层的深度神经网络来学习多类别声音事件检测的瓶颈特征，取得了很好的结果，但是准确率不是很高。无监督特征表示学习在捕获数据生成因子方面取得了不错的进展，然而如果直接用于多类别声音事件检测，则会为所有的声音事件学习到同样的一组特征，这可能会导致性能的下降，也就是说，这组特征对于多类别声音事件没有足够的辨别能力。尽管目前很多方法已经通过特征学习取得了一些新的进展，但是目前仍然没有解决如何通过因素分解的方法进行多类别声音事件检测，这正是现实环境中声音事件检测的重中之重。

发明内容

本发明提供一种因素分解方法，使得分解出的特征不受与检测任务无关的因素干扰，分解出的特征只针对每一个特定的声音事件，从而解决多类别声音事件检测在真实环境当中准确率不高的问题，提高检测的准确度。

为了解决以上技术问题，本发明首先对语音信号进行预处理、提取特征，然后通过监督变分编码器提取声音事件潜在属性空间，再通过因素分解的方法学习到每个特定声音事件的特征表示，然后使用对应的声音事件检测器检测特定声音事件是否发生。

具体技术方案如下：

一种基于监督变分编码器因素分解的混合声音事件检测方法，包括下列步骤：

步骤一，对语音信号进行预处理；

步骤二，提取预处理后的语音信号特征；

步骤三，使用监督变分自动编码器提取声音事件潜在属性空间；

步骤四，使用因素分解方法分解构成混合声音的各种因素，进而学习得到每个特定声音事件的特征表示；

步骤五，使用对应的声音事件检测器检测特定声音事件是否发生。

进一步，所述步骤一具体为：将语音信号按照固定的帧长度进行分帧，帧与帧之间有重叠部分。

进一步，所述步骤二具体为：提取预处理后语音信号的梅尔频率倒谱系数。

进一步，所述步骤三中声音事件潜在属性空间具体为：将输入的语音信号特征压缩到低维高斯分布中。

进一步，所述步骤四中特定声音事件的特征表示 $z_k^* = a_k * z$ ，其中 a_k 为声音事件潜在属性空间的注意力权重， z 为声音事件潜在属性空间。

进一步，所述步骤五中对应的声音事件检测器采用深度神经网络作为检测器网络。

本发明具有有益效果：与传统的多类别声音事件检测相比，该种基于监督变分编码器因素分解的混合声音事件检测方法，引入特征表示学习，学习到声音事件潜在属性空间，能够处理现实场景当中多类别声音事件情况下的检测工作；另一个优势就是该方法引入了一个生成模型-变分自动编码器，这样就可以生成更多的训练数据，从而通过数据增强的方法提高检测准确率。该方法还有可用于各种识别任务，如说话人检测等。

附图说明

图 1 是基于监督变分编码器因素分解的混合声音事件检测方法的流程图。

图 2 是实施例中注意力机制的说明示意图。

具体实施方式

下面结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整的描述。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都是本发明保护的范围。

参见图 1，是本发明提供的一个实施例的基于因素分解的声音事件检测方法的具体流程，该方法包括如下步骤：

步骤一，接收语音信号，并对语音信号进行预处理：主要是将语音信号按照固定的帧长度进行分帧，帧与帧之间有重叠部分，即存在帧内重叠。

步骤二，提取预处理后的语音信号特征

提取预处理后的语音信号特征是指提取语音信号每一帧的 MFCC（梅尔频率倒谱系数）特征，并将 5 帧信号作为一个样本，5 帧信号对应着连续的不同时刻，所以每个样本包含了时域信息。

步骤三，使用监督变分自动编码器提取声音事件潜在属性空间

用长短期记忆网络将输入的 5 帧语音信号特征 X 压缩到低维高斯分布当中去，该高斯分布的均值和方差分别为 μ 和 σ ；通过公式计算声音事件潜在属性空间 z ，其公式如下：

$$z = (\mu + \sigma \odot \varepsilon) \quad (1)$$

其中 ε 是服从与均值为 0、方差为 1 的正态分布的随机数；因为每个样本包含 5 帧语音信号的特征， z 就包含时域信息，这也是选择长短期记忆网络来处理语音信号特征的最主要原因，长短期记忆网络能够处理时域信息，并且将其长期保存在网络内，大大降低梯度消失和梯度爆炸的可能性。

步骤四，使用因素分解方法分解构成混合声音的各种因素，进而学习得到每个特定声音事件相关的特征表示

如图 2 所示，在声音事件潜在属性空间运用注意力机制，避免将输入序列编码作为一个固定长度的潜在向量，从而提供更大的灵活性；要为每一个声音事件类型设计一个注意力层，共有 K 个声音事件类型，所以共设计了 K 个注意力层，使用 *soft max* 函数对声音事件潜在属性空间进行激活后，则可获取到声音事件潜在属性空间的注意力权重 a_k ，其计算公式为：

$$a_k = \text{soft max}_k(z) \quad (2)$$

计算特定声音事件相关的特征表示 z_k^* ，其计算公式如下：

$$z_k^* = a_k * z \quad (3)$$

通常合理地假设声音事件的出现是互相独立的，也就是说 z_k^* 是相互独立的，那么就可以计算后验分布与先验分布之间的 KL(Kullback-Leibler)散度，其计算公式如下：

$$\begin{aligned}
& D_{KL}(q_\phi(z^*|x) \| p_\theta(z_k^*)) \\
&= \sum_{k=1}^K D_{KL}(q_\phi(z_k^*|x) \| p_\theta(z_k^*)) \\
&= \sum_{k=1}^K \frac{1}{2I} \sum_{i=1}^I \left[(\mu_k^{*(i)})^2 + (\sigma_k^{*(i)})^2 - 1 - \log(\mu_k^{*(i)})^2 \right] \quad (4) \\
&:= \sum_{k=1}^K D_{KL}^{*(k)}(\phi; x, z^*)
\end{aligned}$$

其中, i 代表第 i 个样本, $\mu_k^{*(i)}$ 和 $\sigma_k^{*(i)}$ 分别是 $z_k^{*(i)}$ 的均值和方差, 对于每一个特征表示 z_k^* 来说, 后验分布 $q_\phi(z_k^*|x)$ 应该与先验分布 $p_\theta(z_k^*)$ 相匹配, $p_\theta(z_k^*)$ 服从于均值为 0、方差为 1 的标准正态分布, 其中 $i=1 \cdots I$, I 表示总的样本数, $k=1 \cdots K$; 该散度作为因素分解损失函数的第一部分。

步骤五, 使用对应的声音事件检测器检测特定声音事件是否发生

用对应的声音事件检测器检测特定声音事件是否发生, 是指为每一个特定的声音事件类型构造一个声音事件检测器, 用二分类函数 *sigmoid* 来检测对应的声音事件发生的概率, 从而判断该事件是否发生, 其方法为:

$$p(y_k) = \text{Detector}_k(z_k^*) \quad (5)$$

Detector 即为构造的声音事件检测器, 每一个声音事件检测器对应一个 z_k^* , 检测器是一个以 *sigmoid* 函数作为输出的多层感知器。

所有的检测器都用一个二值交叉熵损失作为损失函数来进行训练:

$$L_{AED}^{(k)}(\theta'; y) = - \sum_{i=1}^I [y_k^{(i)} \log(p(y_k^{(i)})) + (1 - y_k^{(i)}) \log(1 - p(y_k^{(i)}))] \quad (6)$$

其中, $y_k^{(i)}$ 代表第 i 个样本的真实值, 为 1 或者 0; $p(y_k^{(i)})$ 是第 i 个样本被识别为第 k 个声音事件的可能性。该损失函数作为因素分解损失函数的第二部分。

综上, 本发明实施例提出的总的特定事件因素分解损失函数为:

$$L_{disent}(\phi, \theta'; x, y, z) = \frac{1}{K} \sum_{k=1}^K [\beta D_{KL}^{*(k)}(\phi; x, z^*) + L_{AED}^{(k)}(\theta'; y)] \quad (7)$$

其中, β 衡量每一个声音事件的潜在表示的因素分解程度。

此外, 实施例还训练了一个解码器来通过声音事件潜在属性空间 z 来对输入的语音信号特征进行重构, 以确保潜在属性空间 z 捕获到了数据生成因子, 其损失函数为:

$$L_{recons}(\theta, \phi; x, z) := E_{q_{\phi}(z^*|x)} [\log p_{\theta}(x|z^*)] \quad (8)$$

E 表示采用均方误差损失函数。

定义最后的总的损失函数为：

$$L_{s-\beta-VAE}(\theta, \phi, \theta'; x, y, z) = L_{recons}(\theta, \phi; x, z) + \lambda L_{disent}(\phi, \theta'; x, y, z) \quad (9)$$

其中， λ 是衡量声音事件检测和重构任务的权重因子。

实施例选用 2 个广泛使用的声音事件检测基准数据库来进行实验评估：TUT2017 和 Freesound，同时实施例还在 TIMIT 数据集上进行说话人识别的评估。为比较实施例方法与其他方法的性能，在每一个数据集上，将实施例方法与当下最先进的方法（普通深度神经网络 DNN、长短期记忆网络 LSTM、增强拓扑结构的联合神经进化网络 J-NEAT、卷积-循环神经网络 CRNN、身份向量 i-Vector）进行对比，从而证明实施例所提算法的有效性。在所有实验当中，实施例采用两种评价指标，分别是 F1 得分和错误率(ER)，其计算公式分别为：

$$F1 = \frac{2 \cdot \sum_{k=1}^K TP(k)}{2 \cdot \sum_{k=1}^K TP(k) + \sum_{k=1}^K FP(k) + \sum_{k=1}^K FN(k)} \quad (10)$$

其中， $TP(k)$ 是真正， $FP(k)$ 是假正， $FN(k)$ 是假负；

$$ER = \frac{\sum_{k=1}^K S(k) + \sum_{k=1}^K D(k) + \sum_{k=1}^K I(k)}{\sum_{k=1}^K N(k)} \quad (11)$$

其中， $N(k)$ 是总样本个数， $S(k)$ 、 $D(k)$ 、 $I(k)$ 分别是替换、删除和插入的个数。

(1) TUT2017 数据集

TUT2017 数据集包含了各种各样街道场景下的声音，音量大小各不相同，这个数据集与人类活动和真实交通场景最为密切相关。

表 1 采用不同方法后的 F1 得分和错误率(ER)

方法	F1(%)	ER
DNN	42.80	0.9358
LSTM	43.22	0.9031
J-NEAT	44.90	0.8979

CRNN	41.70	0.7914
监督变分自动编码器	45.86	0.8259

从表 1 的实验结果中,可以看出,实施例的基于监督变分编码器因素分解的方法取得了最高的 F1 得分,与此同时,还保持着非常有竞争力的 ER。在国际声音事件检测大赛 DCASE2017 当中,J-NEAT 方法取得了最高的 F1 得分,但 ER 排第 15 位;CRNN 方法取得了最好的 ER,但 F1 得分排第 11 位。作为比较,实施例的基于监督变分编码器因素分解的方法取得了最高的 F1 得分,并且在 ER 上排到了第 4 位。

(2) Freesound 数据集

Freesound 数据集是从用户上传的音频样本当中提取出来的声音事件数据库,包含了 28 种声音事件,用来评估在复杂程度逐渐增加的情况下,实施例所提出的算法的性能。

表 2 不同声音事件类别数目下的 F1 得分和错误率(ER)

声音事件 类别数目	方法	F1(%)	ER
5	DNN	74.14	0.4215
	CRNN	78.58	0.3544
	监督变分自动编码器	89.49	0.1954
10	DNN	69.26	0.4028
	CRNN	74.48	0.2700
	监督变分自动编码器	87.19	0.2006
15	DNN	62.77	0.5698
	CRNN	72.86	0.3789
	监督变分自动编码器	84.58	0.2096
20	DNN	41.39	0.7254
	CRNN	71.30	0.3917
	监督变分自动编码器	85.09	0.1999

从表 2 的实验结果中,可以看出,随着声音事件类别的增加,DNN 和 CRNN 方法的 F1 得分快速下降,而实施例所提算法 F1 得分的下降速度则较为缓慢。DNN 和 CRNN 方法的 ER 错误率快速增加,而实施例所提算法的 ER 错误率则缓慢增加。由此可以看出:实施例所提出的算法,最大的优势就是其可以处理现实场景中多类别的声音事件检测问题,这也是其它的方法所不擅长的地方。

(3) TIMIT 数据集

TIMIT 数据集总共包含了 6300 条语音，来自 630 个人，每个人 10 条语音。TIMIRT 数据集中的每一条语音都只源自一个说话人，将其用来评估实施例提出的算法对于混合语音说话人识别的性能。

表 3 不同方法在 TIMIT 数据集上说话人识别的 F1 得分和错误率(ER)

方法	F1(%)	ER
监督变分自动编码器	0.8120	0.3049
i-Vector	0.7338	0.4255

从表 3 的实验结果中，可以看到 i-Vector 方法的 F1 得分为 73.38%，ER 错误率为 0.4255；而实施例的方法 F1 得分为 81.20%，ER 错误率为 0.3049，实施例的方法比 i-Vector 方法性能更好。

从上面的验证结果可以看出，实施例提出的方法为各种各样的声音事件检测和识别任务提供了一个通用的框架。

以上实验结果表明：与其它的算法相比，实施例所采用的基于监督变分编码器因素分解的声音事件检测方法可以有效解决在多类别声音事件情况下，检测准确率不高的问题，提高准确度；同时，还为声音事件检测和识别任务提供了一个通用的框架。

以上所述是本发明的优选实施方式，应当指出，对于本技术领域的普通技术人员，在不脱离本发明原理的前提下，还可以做出若干改进和润饰，这些改进和润饰也视为本发明的保护范围。

权 利 要 求 书

1.一种基于监督变分编码器因素分解的混合声音事件检测方法，其特征在于，包括下列步骤：

步骤一，对语音信号进行预处理；

步骤二，提取预处理后的语音信号特征；

步骤三，使用监督变分自动编码器提取声音事件潜在属性空间；

步骤四，使用因素分解方法分解构成混合声音的各种因素，进而学习得到每个特定声音事件的特征表示；

步骤五，使用对应的声音事件检测器检测特定声音事件是否发生。

2.根据权利要求1所述的基于监督变分编码器因素分解的混合声音事件检测方法，其特征在于，所述步骤一具体为：将语音信号按照固定的帧长度进行分帧，帧与帧之间有重叠部分。

3.根据权利要求1所述的基于监督变分编码器因素分解的混合声音事件检测方法，其特征在于，所述步骤二具体为：提取预处理后语音信号的梅尔频率倒谱系数。

4.根据权利要求1所述的基于监督变分编码器因素分解的混合声音事件检测方法，其特征在于，所述步骤三中声音事件潜在属性空间具体为：将输入的语音信号特征压缩到低维高斯分布中。

5.根据权利要求1所述的基于监督变分编码器因素分解的混合声音事件检测方法，其特征在于，所述步骤四中特定声音事件的特征表示 $z_k^* = a_k * z$ ，其中 a_k 为声音事件潜在属性空间的注意力权重， z 为声音事件潜在属性空间。

6.根据权利要求1所述的基于监督变分编码器因素分解的混合声音事件检测方法，其特征在于，所述步骤五中对应的声音事件检测器采用深度神经网络作为检测器网络。

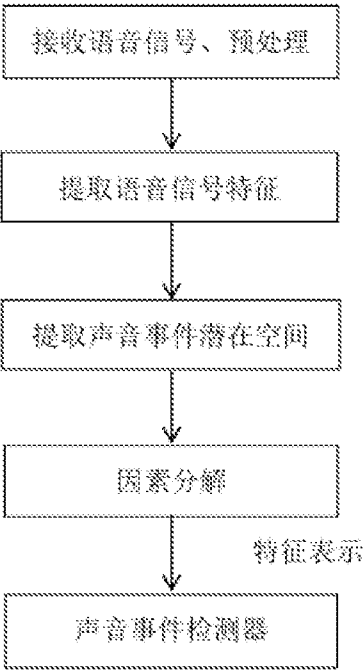


图 1

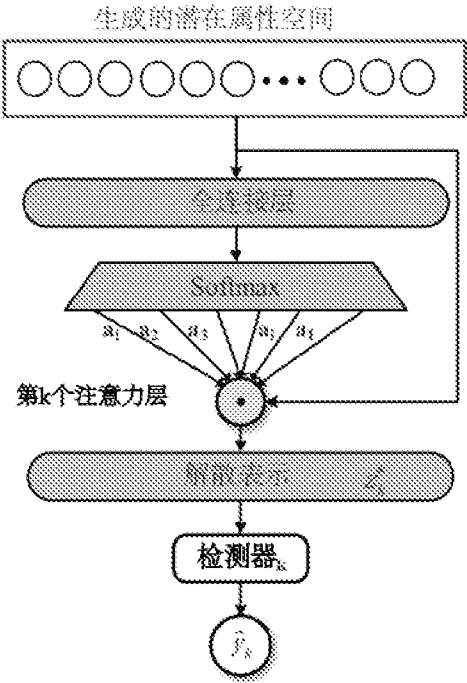


图 2

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/077189

A. CLASSIFICATION OF SUBJECT MATTER

G10L 25/78(2013.01)i; G10L 25/30(2013.01)i; G10L 25/51(2013.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNKI, CNPAT, EPODOC, WPI: 因素分解, 监督, 卷积神经网络, 长短时记忆, 长短期记忆, LSTM, 深度神经网络, 声音, 事件, 检测, 分类, 监测, 高斯分布, 注意力, 权重, 自动编码器, factorization, supervision, convolutional neural network, long-term and short-term memory, deep neural network, sound, event, detection, classification, monitoring, Gaussian distribution, attention, weight, autoencoder.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 110070895 A (JIANGSU UNIVERSITY) 30 July 2019 (2019-07-30) claims 1-6	1-6
Y	CN 106251860 A (ZHANG, Aiyang et al.) 21 December 2016 (2016-12-21) description, paragraphs [0003]-[0043] and [0093]	1-6
Y	CN 104021373 A (JIANGSU UNIVERSITY) 03 September 2014 (2014-09-03) description, paragraphs [0004]-[0006] and [0015]-[0072]	1-6
A	JP 2015057630 A (NIPPON TELEGRAPH & TELEPHONE CORP.) 26 March 2015 (2015-03-26) entire document	1-6
A	CN 104795064 A (FUZHOU UNIVERSITY) 22 July 2015 (2015-07-22) entire document	1-6
A	US 2017372725 A1 (PINDROP SECURITY INC.) 28 December 2017 (2017-12-28) entire document	1-6



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance
 “E” earlier application or patent but published on or after the international filing date
 “L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
 “O” document referring to an oral disclosure, use, exhibition or other means
 “P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
 “X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
 “Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
 “&” document member of the same patent family

Date of the actual completion of the international search

28 April 2020

Date of mailing of the international search report

03 June 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration
 No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
 100088
 China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2020/077189

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	110070895	A	30 July 2019	None			
CN	106251860	A	21 December 2016	CN	106251860	B	11 February 2020
CN	104021373	A	03 September 2014	WO	2015180368	A1	03 December 2015
				CN	104021373	B	15 February 2017
JP	2015057630	A	26 March 2015	JP	6235938	B2	22 November 2017
CN	104795064	A	22 July 2015	WO	2016155047	A1	06 October 2016
				CN	104795064	B	13 April 2018
US	2017372725	A1	28 December 2017	US	10141009	B2	27 November 2018
				WO	2018005620	A1	04 January 2018
				US	2019096424	A1	28 March 2019

国际检索报告

国际申请号

PCT/CN2020/077189

A. 主题的分类 G10L 25/78(2013.01)i; G10L 25/30(2013.01)i; G10L 25/51(2013.01)i 按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类																							
B. 检索领域 检索的最低限度文献(标明分类系统和分类号) G10L 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用)) CNKI, CNPAT, EPDOC, WPI: 因素分解, 监督, 卷积神经网络, 长短时记忆, 长短期记忆, LSTM, 深度神经网络, 声音, 事件, 检测, 分类, 监测, 高斯分布, 注意力, 权重, 自动编码器, factorization, supervision, convolutional neural network, long-term and short-term memory, deep neural network, sound, event, detection, classification, monitoring, Gaussian distribution, attention, weight, autoencoder.																							
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>PX</td> <td>CN 110070895 A (江苏大学) 2019年 7月 30日 (2019 - 07 - 30) 权利要求1-6</td> <td>1-6</td> </tr> <tr> <td>Y</td> <td>CN 106251860 A (张爱英 等) 2016年 12月 21日 (2016 - 12 - 21) 说明书第[0003]-[0043]、[0093]段</td> <td>1-6</td> </tr> <tr> <td>Y</td> <td>CN 104021373 A (江苏大学) 2014年 9月 3日 (2014 - 09 - 03) 说明书第[0004]-[0006]、[0015]-[0072]段</td> <td>1-6</td> </tr> <tr> <td>A</td> <td>JP 2015057630 A (NIPPON TELEGRAPH & TELEPHONE CORP.) 2015年 3月 26日 (2015 - 03 - 26) 全文</td> <td>1-6</td> </tr> <tr> <td>A</td> <td>CN 104795064 A (福州大学) 2015年 7月 22日 (2015 - 07 - 22) 全文</td> <td>1-6</td> </tr> <tr> <td>A</td> <td>US 2017372725 A1 (PINDROP SECURITY INC.) 2017年 12月 28日 (2017 - 12 - 28) 全文</td> <td>1-6</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	PX	CN 110070895 A (江苏大学) 2019年 7月 30日 (2019 - 07 - 30) 权利要求1-6	1-6	Y	CN 106251860 A (张爱英 等) 2016年 12月 21日 (2016 - 12 - 21) 说明书第[0003]-[0043]、[0093]段	1-6	Y	CN 104021373 A (江苏大学) 2014年 9月 3日 (2014 - 09 - 03) 说明书第[0004]-[0006]、[0015]-[0072]段	1-6	A	JP 2015057630 A (NIPPON TELEGRAPH & TELEPHONE CORP.) 2015年 3月 26日 (2015 - 03 - 26) 全文	1-6	A	CN 104795064 A (福州大学) 2015年 7月 22日 (2015 - 07 - 22) 全文	1-6	A	US 2017372725 A1 (PINDROP SECURITY INC.) 2017年 12月 28日 (2017 - 12 - 28) 全文	1-6
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																					
PX	CN 110070895 A (江苏大学) 2019年 7月 30日 (2019 - 07 - 30) 权利要求1-6	1-6																					
Y	CN 106251860 A (张爱英 等) 2016年 12月 21日 (2016 - 12 - 21) 说明书第[0003]-[0043]、[0093]段	1-6																					
Y	CN 104021373 A (江苏大学) 2014年 9月 3日 (2014 - 09 - 03) 说明书第[0004]-[0006]、[0015]-[0072]段	1-6																					
A	JP 2015057630 A (NIPPON TELEGRAPH & TELEPHONE CORP.) 2015年 3月 26日 (2015 - 03 - 26) 全文	1-6																					
A	CN 104795064 A (福州大学) 2015年 7月 22日 (2015 - 07 - 22) 全文	1-6																					
A	US 2017372725 A1 (PINDROP SECURITY INC.) 2017年 12月 28日 (2017 - 12 - 28) 全文	1-6																					
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																							
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																							
国际检索实际完成的日期 2020年 4月 28日		国际检索报告邮寄日期 2020年 6月 3日																					
ISA/CN的名称和邮寄地址 中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		授权官员 王玥 电话号码 86-(10)- 53962545																					

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2020/077189

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	110070895	A	2019年 7月 30日	无	
CN	106251860	A	2016年 12月 21日	CN	106251860 B 2020年 2月 11日
CN	104021373	A	2014年 9月 3日	WO	2015180368 A1 2015年 12月 3日
				CN	104021373 B 2017年 2月 15日
JP	2015057630	A	2015年 3月 26日	JP	6235938 B2 2017年 11月 22日
CN	104795064	A	2015年 7月 22日	WO	2016155047 A1 2016年 10月 6日
				CN	104795064 B 2018年 4月 13日
US	2017372725	A1	2017年 12月 28日	US	10141009 B2 2018年 11月 27日
				WO	2018005620 A1 2018年 1月 4日
				US	2019096424 A1 2019年 3月 28日