



(12)发明专利

(10)授权公告号 CN 109147810 B

(45)授权公告日 2019.11.26

(21)申请号 201811155530.8

G10L 21/0232(2013.01)

(22)申请日 2018.09.30

G10L 25/30(2013.01)

(65)同一申请的已公布的文献号

G06N 3/04(2006.01)

申请公布号 CN 109147810 A

G06N 3/08(2006.01)

(43)申请公布日 2019.01.04

审查员 李海龙

(73)专利权人 百度在线网络技术(北京)有限公司

地址 100085 北京市海淀区上地十街10号  
百度大厦

(72)发明人 成学军

(74)专利代理机构 北京鸿德海业知识产权代理  
事务所(普通合伙) 11412

代理人 袁媛

(51)Int.Cl.

G10L 21/02(2013.01)

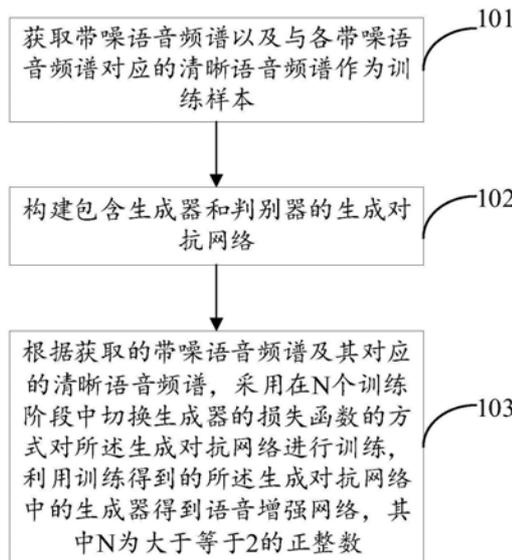
权利要求书4页 说明书12页 附图2页

(54)发明名称

建立语音增强网络的方法、装置、设备和计算机存储介质

(57)摘要

本发明提供了一种建立语音增强网络的方法、装置、设备和计算机存储介质,所述方法包括:获取带噪语音频谱以及与各带噪语音频谱对应的清晰语音频谱作为训练样本;构建包含生成器和判别器的生成对抗网络;根据获取的带噪语音频谱及其对应的清晰语音频谱,采用在N个训练阶段中切换生成器的损失函数的方式对所述生成对抗网络进行训练,利用训练得到的所述生成对抗网络中的生成器得到语音增强网络,其中N为大于等于2的正整数。本发明能够增强生成对抗网络训练收敛的稳定性,从而提升基于生成对抗网络的语音增强网络的性能,进而实现提高语音识别准确性的目的。



1. 一种建立语音增强网络的方法,其特征在于,所述方法包括:

获取带噪语音频谱以及与各带噪语音频谱对应的清晰语音频谱作为训练样本;

构建包含生成器和判别器的生成对抗网络;

根据获取的带噪语音频谱及其对应的清晰语音频谱,采用在N个训练阶段中切换生成器的损失函数的方式对所述生成对抗网络进行训练,利用训练得到的所述生成对抗网络中的生成器得到语音增强网络,其中N为大于等于2的正整数;

其中,所述在N个训练阶段中切换生成器的损失函数包括:在N个训练阶段中分别将生成器的损失函数切换为生成器的第一损失函数或者第二损失函数中的一个。

2. 根据权利要求1所述的方法,其特征在于,所述生成器的网络结构可以为深度神经网络、循环神经网络或者长短时记忆网络中的一种,所述判别器的网络结构可以为深度神经网络、循环神经网络或者长短时记忆网络中的一种。

3. 根据权利要求1所述的方法,其特征在于,所述采用在N个训练阶段中切换生成器的损失函数的方式对所述生成对抗网络进行训练包括:

采用在两个训练阶段中切换生成器的损失函数的方式对所述生成对抗网络进行训练,所述两个训练阶段包括第一训练阶段以及第二训练阶段。

4. 根据权利要求3所述的方法,其特征在于,在所述第一训练阶段对所述生成对抗网络进行训练包括:

将带噪语音频谱输入生成器,得到生成器的输出结果;

在迭代次数小于等于预设的迭代次数时,根据生成器的输出结果及其对应的清晰语音频谱,得到生成器的第一损失函数;

根据所述生成器的第一损失函数更新生成器的网络结构中的参数;

在所述第二训练阶段对所述生成对抗网络进行训练包括:

将带噪语音频谱输入生成器,得到生成器的输出结果;

在迭代次数大于预设的迭代次数之后,将生成器的输出结果及其对应的清晰语音频谱输入判别器,根据判别器的输出结果分别得到判别器的损失函数以及生成器的第二损失函数;

根据所述判别器的损失函数以及生成器的第二损失函数,分别更新判别器以及生成器的网络结构中的参数,直至生成对抗网络收敛。

5. 根据权利要求3所述的方法,其特征在于,在所述第一训练阶段对所述生成对抗网络进行训练包括:

将带噪语音频谱输入生成器,得到生成器的输出结果;

在迭代次数小于等于预设的迭代次数时,将生成器的输出结果及其对应的清晰语音频谱输入判别器,根据判别器的输出结果得到生成器的第二损失函数;

根据所述生成器的第二损失函数,更新生成器的网络结构中的参数;

在所述第二训练阶段对所述生成对抗网络进行训练包括:

将带噪语音频谱输入生成器,得到生成器的输出结果;

在达到预设的迭代次数之后,根据生成器的输出结果及其对应的清晰语音频谱,得到生成器的第一损失函数;

将所述生成器的输出结果输入判别器,根据判别器的输出结果得到判别器的损失函

数；

根据所述判别器的损失函数以及生成器的第一损失函数，分别更新判别器以及生成器的网络结构中的参数，直至生成对抗网络收敛。

6. 根据权利要求1所述的方法，其特征在于，所述生成器的第一损失函数通过以下公式计算得到：

$$L_{G1} = E_{x,y \sim p_{data}(x,y)} [\|y - G(x)\|_1]$$

在公式中： $L_{G1}$ 表示生成器的第一损失函数； $x$ 表示带噪语音频谱； $y$ 表示与带噪语音频谱对应的清晰语音频谱； $p_{data}$ 表示训练样本； $E$ 表示从训练样本中取样； $G(x)$ 表示输入为带噪语音频谱时生成器的输出结果。

7. 根据权利要求1所述的方法，其特征在于，所述生成器的第二损失函数通过以下公式计算得到：

$$L_{G2} = E_{x \sim p_{data}(x)} [\log(1 - D(G(x)))] + f(epoch) E_{x,y \sim p_{data}(x,y)} [\|y - G(x)\|_1]$$

在公式中： $L_{G2}$ 表示生成器的第二损失函数； $x$ 表示带噪语音频谱； $y$ 表示与带噪语音频谱对应的清晰语音频谱； $p_{data}$ 表示训练样本； $E$ 表示从训练样本中取样； $G(x)$ 表示输入为带噪语音频谱时生成器的输出结果； $D(G(x))$ 表示当输入为生成器的输出结果时判别器的输出结果； $f(epoch)$ 表示与迭代次数 $epoch$ 成反比关系的函数。

8. 根据权利要求1所述的方法，其特征在于，所述方法还包括：

将所述语音增强网络与语音识别系统组合，为所述语音识别系统提供清晰语音频谱。

9. 一种建立语音增强网络的装置，其特征在于，所述装置包括：

样本获取单元，用于获取带噪语音频谱以及与各带噪语音频谱对应的清晰语音频谱作为训练样本；

网络构建单元，用于构建包含生成器和判别器的生成对抗网络；

网络训练单元，用于根据获取的带噪语音频谱及其对应的清晰语音频谱，采用在 $N$ 个训练阶段中切换生成器的损失函数的方式对所述生成对抗网络进行训练，利用训练得到的所述生成对抗网络中的生成器得到语音增强网络，其中 $N$ 为大于等于2的正整数；

其中，所述网络训练单元在 $N$ 个训练阶段中切换生成器的损失函数时，具体执行：在 $N$ 个训练阶段中分别将生成器的损失函数切换为生成器的第一损失函数或者第二损失函数中的一个。

10. 根据权利要求9所述的装置，其特征在于，所述生成器的网络结构可以为深度神经网络、循环神经网络或者长短时记忆网络中的一种，所述判别器的网络结构可以为深度神经网络、循环神经网络或者长短时记忆网络中的一种。

11. 根据权利要求9所述的装置，其特征在于，所述网络训练单元采用在 $N$ 个训练阶段中切换生成器的损失函数的方式对所述生成对抗网络进行训练时，具体执行：

采用在两个训练阶段中切换生成器的损失函数的方式对所述生成对抗网络进行训练，所述两个训练阶段包括第一训练阶段以及第二训练阶段。

12. 根据权利要求11所述的装置，其特征在于，所述网络训练单元在所述第一训练阶段对所述生成对抗网络进行训练时，具体执行：

将带噪语音频谱输入生成器，得到生成器的输出结果；

在迭代次数小于等于预设的迭代次数时,根据生成器的输出结果及其对应的清晰语音频谱,得到生成器的第一损失函数;

根据所述生成器的第一损失函数更新生成器的网络结构中的参数;

所述网络训练单元在所述第二训练阶段对所述生成对抗网络进行训练时,具体执行:

将带噪语音频谱输入生成器,得到生成器的输出结果;

在迭代次数大于预设的迭代次数之后,将生成器的输出结果及其对应的清晰语音频谱输入判别器,根据判别器的输出结果分别得到判别器的损失函数以及生成器的第二损失函数;

根据所述判别器的损失函数以及生成器的第二损失函数,分别更新判别器以及生成器的网络结构中的参数,直至生成对抗网络收敛。

13. 根据权利要求11所述的装置,其特征在于,所述网络训练单元在所述第一训练阶段对所述生成对抗网络进行训练时,具体执行:

将带噪语音频谱输入生成器,得到生成器的输出结果;

在迭代次数小于等于预设的迭代次数时,将生成器的输出结果及其对应的清晰语音频谱输入判别器,根据判别器的输出结果得到生成器的第二损失函数;

根据所述生成器的第二损失函数,更新生成器的网络结构中的参数;

所述网络训练单元在所述第二训练阶段对所述生成对抗网络进行训练时,具体执行:

将带噪语音频谱输入生成器,得到生成器的输出结果;

在达到预设的迭代次数之后,根据生成器的输出结果及其对应的清晰语音频谱,得到生成器的第一损失函数;

将所述生成器的输出结果输入判别器,根据判别器的输出结果得到判别器的损失函数;

根据所述判别器的损失函数以及生成器的第一损失函数,分别更新判别器以及生成器的网络结构中的参数,直至生成对抗网络收敛。

14. 根据权利要求9所述的装置,其特征在于,所述网络训练单元通过以下公式计算得到所述生成器的第一损失函数:

$$L_{G1} = E_{x,y \sim P_{data}(x,y)} [\|y - G(x)\|_1]$$

在公式中: $L_{G1}$ 表示生成器的第一损失函数; $x$ 表示带噪语音频谱; $y$ 表示与带噪语音频谱对应的清晰语音频谱; $P_{data}$ 表示训练样本; $E$ 表示从训练样本中取样; $G(x)$ 表示输入为带噪语音频谱时生成器的输出结果。

15. 根据权利要求9所述的装置,其特征在于,所述网络训练单元通过以下公式计算得到所述生成器的第二损失函数:

$$L_{G2} = E_{x \sim P_{data}(x)} [\log(1 - D(G(x)))] + f(\text{epoch}) E_{x,y \sim P_{data}(x,y)} [\|y - G(x)\|_1]$$

在公式中: $L_{G2}$ 表示生成器的第二损失函数; $x$ 表示带噪语音频谱; $y$ 表示与带噪语音频谱对应的清晰语音频谱; $P_{data}$ 表示训练样本; $E$ 表示从训练样本中取样; $G(x)$ 表示输入为带噪语音频谱时生成器的输出结果; $D(G(x))$ 表示当输入为生成器的输出结果时判别器的输出结果; $f(\text{epoch})$ 表示与迭代次数 $\text{epoch}$ 成反比关系的函数。

16. 根据权利要求9所述的装置,其特征在于,所述装置还包括:

网络组合单元,用于将所述语音增强网络与语音识别系统组合,为所述语音识别系统提供清晰语音频谱。

17.一种计算机设备,包括存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序,其特征在于,所述处理器执行所述程序时实现如权利要求1~8中任一项所述的方法。

18.一种计算机可读存储介质,其上存储有计算机程序,其特征在于,所述程序被处理器执行时实现如权利要求1~8中任一项所述的方法。

## 建立语音增强网络的方法、装置、设备和计算机存储介质

### 【技术领域】

[0001] 本发明涉及语音识别技术,尤其涉及一种建立语音增强网络的方法、装置、设备和计算机存储介质。

### 【背景技术】

[0002] 噪声环境下的语音识别一直是目前语音识别领域急需解决的难题,目前的主流方法都是在语音识别系统前加一个语音增强网络。目前为止,生成对抗网络(Generative Adversarial Network, GAN)是作为语音增强网络最新的增强方法。通过研究发现,采用现有的训练方式训练生成对抗网络时,虽然能够加速生成对抗网络训练的收敛,但是会导致生成对抗网络收敛的不稳定,会让生成对抗网络中的生成器生成过于清晰的语音频谱,导致现有的基于生成对抗网络的语音增强网络会损失语音频谱中某些细微却重要的信息。

### 【发明内容】

[0003] 有鉴于此,本发明提供了一种建立语音增强网络的方法、装置、设备和计算机存储介质,用于增强生成对抗网络训练收敛的稳定性,从而提升基于生成对抗网络的语音增强网络的性能,进而实现提高语音识别准确性的目的。

[0004] 本发明为解决技术问题所采用的技术方案是提供一种建立语音增强网络的方法,所述方法包括:获取带噪语音频谱以及与各带噪语音频谱对应的清晰语音频谱作为训练样本;构建包含生成器和判别器的生成对抗网络;根据获取的带噪语音频谱及其对应的清晰语音频谱,采用在N个训练阶段中切换生成器的损失函数的方式对所述生成对抗网络进行训练,利用训练得到的所述生成对抗网络中的生成器得到语音增强网络,其中N为大于等于2的正整数。

[0005] 根据本发明一优选实施例,所述生成器的网络结构可以为深度神经网络、循环神经网络或者长短时记忆网络中的一种,所述判别器的网络结构可以为深度神经网络、循环神经网络或者长短时记忆网络中的一种。

[0006] 根据本发明一优选实施例,所述采用在N个训练阶段中切换生成器的损失函数的方式对所述生成对抗网络进行训练包括:采用在两个训练阶段中切换生成器的损失函数的方式对所述生成对抗网络进行训练,所述两个训练阶段包括第一训练阶段以及第二训练阶段。

[0007] 根据本发明一优选实施例,在所述第一训练阶段对所述生成对抗网络进行训练包括:将带噪语音频谱输入生成器,得到生成器的输出结果;在迭代次数小于等于预设的迭代次数时,根据生成器的输出结果及其对应的清晰语音频谱,得到生成器的第一损失函数;根据所述生成器的第一损失函数更新生成器的网络结构中的参数;在所述第二训练阶段对所述生成对抗网络进行训练包括:将带噪语音频谱输入生成器,得到生成器的输出结果;在迭代次数大于预设的迭代次数之后,将生成器的输出结果及其对应的清晰语音频谱输入判别器,根据判别器的输出结果分别得到判别器的损失函数以及生成器的第二损失函数;根据

所述判别器的损失函数以及生成器的第二损失函数,分别更新判别器以及生成器的网络结构中的参数,直至生成对抗网络收敛。

[0008] 根据本发明一优选实施例,在所述第一训练阶段对所述生成对抗网络进行训练包括:将带噪语音频谱输入生成器,得到生成器的输出结果;在迭代次数小于等于预设的迭代次数时,将生成器的输出结果及其对应的清晰语音频谱输入判别器,根据判别器的输出结果得到生成器的第二损失函数;根据所述生成器的第二损失函数,更新生成器的网络结构中的参数;在所述第二训练阶段对所述生成对抗网络进行训练包括:将带噪语音频谱输入生成器,得到生成器的输出结果;在达到预设的迭代次数之后,根据生成器的输出结果及其对应的清晰语音频谱,得到生成器的第一损失函数;将所述生成器的输出结果输入判别器,根据判别器的输出结果得到判别器的损失函数;根据所述判别器的损失函数以及生成器的第一损失函数,分别更新判别器以及生成器的网络结构中的参数,直至生成对抗网络收敛。

[0009] 根据本发明一优选实施例,所述生成器的第一损失函数通过以下公式计算得到:

$$[0010] \quad L_{G1} = E_{x,y \sim p_{data}(x,y)} [\|y - G(x)\|_1]$$

[0011] 在公式中: $L_{G1}$ 表示生成器的第一损失函数; $x$ 表示带噪语音频谱; $y$ 表示与带噪语音频谱对应的清晰语音频谱; $p_{data}$ 表示训练样本; $E$ 表示从训练样本中取样; $G(x)$ 表示输入为带噪语音频谱时生成器的输出结果。

[0012] 根据本发明一优选实施例,所述生成器的第二损失函数通过以下公式计算得到:

$$[0013] \quad L_{G2} = E_{x \sim p_{data}(x)} [\log(1 - D(G(x)))] + f(epoch) E_{x,y \sim p_{data}(x,y)} [\|y - G(x)\|_1]$$

[0014] 在公式中: $L_{G2}$ 表示生成器的第二损失函数; $x$ 表示带噪语音频谱; $y$ 表示与带噪语音频谱对应的清晰语音频谱; $p_{data}$ 表示训练样本; $E$ 表示从训练样本中取样; $G(x)$ 表示输入为带噪语音频谱时生成器的输出结果; $D(G(x))$ 表示当输入为生成器的输出结果时判别器的输出结果; $f(epoch)$ 表示与迭代次数 $epoch$ 成反比关系的函数。

[0015] 根据本发明一优选实施例,所述方法还包括:将所述语音增强网络与语音识别系统组合,为所述语音识别系统提供清晰语音频谱。

[0016] 本发明未解决技术问题所采用的技术方案是提供一种建立语音增强网络的装置,所述装置包括:样本获取单元,用于获取带噪语音频谱以及与各带噪语音频谱对应的清晰语音频谱作为训练样本;网络构建单元,用于构建包含生成器和判别器的生成对抗网络;网络训练单元,用于根据获取的带噪语音频谱及其对应的清晰语音频谱,采用在 $N$ 个训练阶段中切换生成器的损失函数的方式对所述生成对抗网络进行训练,利用训练得到的所述生成对抗网络中的生成器得到语音增强网络,其中 $N$ 为大于等于2的正整数。

[0017] 根据本发明一优选实施例,所述生成器的网络结构可以为深度神经网络、循环神经网络或者长短时记忆网络中的一种,所述判别器的网络结构可以为深度神经网络、循环神经网络或者长短时记忆网络中的一种。

[0018] 根据本发明一优选实施例,所述网络训练单元采用在 $N$ 个训练阶段中切换生成器的损失函数的方式对所述生成对抗网络进行训练时,具体执行:采用在两个训练阶段中切换生成器的损失函数的方式对所述生成对抗网络进行训练,所述两个训练阶段包括第一训练阶段以及第二训练阶段。

[0019] 根据本发明一优选实施例,所述网络训练单元在所述第一训练阶段对所述生成对

抗网络进行训练时,具体执行:将带噪语音频谱输入生成器,得到生成器的输出结果;在迭代次数小于等于预设的迭代次数时,根据生成器的输出结果及其对应的清晰语音频谱,得到生成器的第一损失函数;根据所述生成器的第一损失函数更新生成器的网络结构中的参数;所述网络训练单元在所述第二训练阶段对所述生成对抗网络进行训练时,具体执行:将带噪语音频谱输入生成器,得到生成器的输出结果;在迭代次数大于预设的迭代次数之后,将生成器的输出结果及其对应的清晰语音频谱输入判别器,根据判别器的输出结果分别得到判别器的损失函数以及生成器的第二损失函数;根据所述判别器的损失函数以及生成器的第二损失函数,分别更新判别器以及生成器的网络结构中的参数,直至生成对抗网络收敛。

[0020] 根据本发明一优选实施例,所述网络训练单元在所述第一训练阶段对所述生成对抗网络进行训练时,具体执行:将带噪语音频谱输入生成器,得到生成器的输出结果;在迭代次数小于等于预设的迭代次数时,将生成器的输出结果及其对应的清晰语音频谱输入判别器,根据判别器的输出结果得到生成器的第二损失函数;根据所述生成器的第二损失函数,更新生成器的网络结构中的参数;所述网络训练单元在所述第二训练阶段对所述生成对抗网络进行训练时,具体执行:将带噪语音频谱输入生成器,得到生成器的输出结果;在达到预设的迭代次数之后,根据生成器的输出结果及其对应的清晰语音频谱,得到生成器的第一损失函数;将所述生成器的输出结果输入判别器,根据判别器的输出结果得到判别器的损失函数;根据所述判别器的损失函数以及生成器的第一损失函数,分别更新判别器以及生成器的网络结构中的参数,直至生成对抗网络收敛。

[0021] 根据本发明一优选实施例,所述网络训练单元通过以下公式计算得到所述生成器的第一损失函数:

$$[0022] \quad L_{G1} = E_{x,y \sim P_{data}(x,y)} [\|y - G(x)\|_1]$$

[0023] 在公式中: $L_{G1}$ 表示生成器的第一损失函数; $x$ 表示带噪语音频谱; $y$ 表示与带噪语音频谱对应的清晰语音频谱; $P_{data}$ 表示训练样本; $E$ 表示从训练样本中取样; $G(x)$ 表示输入为带噪语音频谱时生成器的输出结果。

[0024] 根据本发明一优选实施例,所述网络训练单元通过以下公式计算得到所述生成器的第二损失函数:

$$[0025] \quad L_{G2} = E_{x \sim P_{data}(x)} [\log(1 - D(G(x)))] + f(epoch) E_{x,y \sim P_{data}(x,y)} [\|y - G(x)\|_1]$$

[0026] 在公式中: $L_{G2}$ 表示生成器的第二损失函数; $x$ 表示带噪语音频谱; $y$ 表示与带噪语音频谱对应的清晰语音频谱; $P_{data}$ 表示训练样本; $E$ 表示从训练样本中取样; $G(x)$ 表示输入为带噪语音频谱时生成器的输出结果; $D(G(x))$ 表示当输入为生成器的输出结果时判别器的输出结果; $f(epoch)$ 表示与迭代次数 $epoch$ 成反比关系的函数。

[0027] 根据本发明一优选实施例,所述装置还包括:网络组合单元,用于将所述语音增强网络与语音识别系统组合,为所述语音识别系统提供清晰语音频谱。

[0028] 利用本发明所提供的技术方案,通过在多个训练阶段中切换生成器的损失函数的方式对生成对抗网络进行训练,能够增强生成对抗网络训练收敛的稳定性,从而提升基于生成对抗网络的语音增强网络的性能,进而实现提高语音识别准确性的目的。

**【附图说明】**

[0029] 图1为本发明一实施例提供的一种建立语音增强网络的方法流程图；

[0030] 图2为本发明一实施例提供的一种建立语音增强网络的装置结构图；

[0031] 图3为本发明一实施例提供的计算机系统/服务器的框图。

**【具体实施方式】**

[0032] 为了使本发明的目的、技术方案和优点更加清楚，下面结合附图和具体实施例对本发明进行详细描述。

[0033] 在本发明实施例中使用的术语是仅仅出于描述特定实施例的目的，而非旨在限制本发明。在本发明实施例和所附权利要求书中所使用的单数形式的“一种”、“所述”和“该”也旨在包括多数形式，除非上下文清楚地表示其他含义。

[0034] 应当理解，本文中使用的术语“和/或”仅仅是一种描述关联对象的关联关系，表示可以存在三种关系，例如，A和/或B，可以表示：单独存在A，同时存在A和B，单独存在B这三种情况。另外，本文中字符“/”，一般表示前后关联对象是一种“或”的关系。

[0035] 取决于语境，如在此所使用的词语“如果”可以被解释成为“在……时”或“当……时”或“响应于确定”或“响应于检测”。类似地，取决于语境，短语“如果确定”或“如果检测(陈述的条件或事件)”可以被解释成为“当确定时”或“响应于确定”或“当检测(陈述的条件或事件)时”或“响应于检测(陈述的条件或事件)”。

[0036] 图1为本发明一实施例提供的一种建立语音增强网络的方法流程图，如图1中所示，所述方法包括：

[0037] 在101中，获取带噪语音频谱以及与各带噪语音频谱对应的清晰语音频谱作为训练样本。

[0038] 在本步骤中，获取多个带噪语音频谱以及与各带噪语音频谱对应的清晰语音频谱作为训练样本，所获取的清晰语音频谱即为不带噪的语音频谱。

[0039] 可以理解的是，本步骤对所获取的语音频谱的帧长、帧移以及语音频谱中所包含的声学特征不进行限定。

[0040] 在102中，构建包含生成器和判别器的生成对抗网络。

[0041] 在本步骤中，构建包含生成器和判别器的生成对抗网络。

[0042] 本发明所使用的生成对抗网络包括生成器和判别器。其中，生成器的职责是生成与真实样本尽可能相似的生成样本，判别器的职责则是尽可能区分真实样本和生成样本。通过生成器和判别器之间对抗博弈的方式训练整个生成对抗网络，使得生成器输出的生成样本的真实性尽可能高，从而使得判别器无法区分由生成器所得到的输出为生成样本还是真实样本。

[0043] 其中，生成器的网络结构可以为深度神经网络、循环神经网络或者长短时记忆网络中的一种，判别器的网络结构可以为深度神经网络、循环神经网络或者长短时记忆网络中的一种，本发明对此不进行限定。

[0044] 在103中，根据获取的带噪语音频谱及其对应的清晰语音频谱，采用在N个训练阶段中切换生成器的损失函数的方式对所述生成对抗网络进行训练，利用训练得到的所述生成对抗网络中的生成器得到语音增强网络，其中N为大于等于2的正整数。

[0045] 通常理解,在训练生成对抗网络时,一般采用交替训练生成器和判别器的方式,也就是在现有的训练过程中,生成器和判别器的损失函数不发生变化,直至生成对抗网络收敛。而在本发明中,不再使用传统的交替训练生成器和判别器的方式,而是采用在N个训练阶段中切换生成器的损失函数的方式,对生成对抗网络中的生成器和判别器进行分阶段训练,其中N为大于等于2的正整数。

[0046] 为了不使说明书中的内容过于冗杂,在下文中仅描述采用两个训练阶段以及三个训练阶段对生成对抗网络进行训练的方式,其他的四个训练阶段、五个训练阶段等的训练方式可以根据本文的描述内容推理得到。

[0047] (一)根据获取的带噪语音频谱及其对应的清晰语音频谱,采用在两个训练阶段中切换生成器的损失函数的方式对生成对抗网络进行训练。

[0048] 1、第一种训练方式

[0049] 在第一训练阶段训练生成对抗网络时,可以采用以下方式:将带噪语音频谱输入生成器,得到生成器的输出结果;在迭代次数小于等于预设的迭代次数时,根据生成器的输出结果及其对应的清晰语音频谱,得到生成器的第一损失函数;根据所得到的第一损失函数更新生成器的网络结构中的参数。也就是说,在该种训练方式中的第一训练阶段只根据第一损失函数来更新生成器的网络结构中的参数,而不更新判别器的网络结构中的参数。

[0050] 其中,可以使用以下公式计算生成模型的第一损失函数:

$$[0051] \quad L_{G1} = E_{x,y \sim p_{data}(x,y)} [\|y - G(x)\|_1]$$

[0052] 在公式中: $L_{G1}$ 表示生成器的第一损失函数; $x$ 表示带噪语音频谱; $y$ 表示与带噪语音频谱对应的清晰语音频谱; $p_{data}$ 表示训练样本; $E$ 表示从训练样本中取样; $G(x)$ 表示输入为带噪语音频谱时生成器的输出结果。

[0053] 在第二训练阶段训练生成对抗网络时,可以采用以下方式:将带噪语音频谱输入经过第一训练阶段训练所得到的生成器,得到生成器的输出结果;在迭代次数大于预设的迭代次数之后,将生成器的输出结果及其对应的清晰语音频谱输入判别器,根据判别器的输出结果分别得到判别器的损失函数以及生成器的第二损失函数;根据判别器的损失函数以及生成器的第二损失函数,分别更新生成器以及判别器的网络结构中的参数,直至生成对抗网络收敛。

[0054] 其中,可以使用以下公式计算判别器的损失函数:

$$[0055] \quad L_D = E_{y \sim p_{data}(y)} [\log D(y)] + E_{x \sim p_{data}(x)} [\log(1 - D(G(x)))]$$

[0056] 在公式中: $L_D$ 表示判别器的损失函数; $x$ 表示带噪语音频谱; $y$ 表示与带噪语音频谱对应的清晰语音频谱; $p_{data}$ 表示训练样本; $E$ 表示从训练样本中取样; $G(x)$ 表示输入为带噪语音频谱时生成器的输出结果; $D(G(x))$ 表示当输入为生成器的输出结果时判别器的输出结果; $D(y)$ 表示当输入为清晰语音频谱时判别器的输出结果。

[0057] 可以使用以下公式计算生成器的第二损失函数:

$$[0058] \quad L_{G2} = E_{x,y \sim p_{data}(x,y)} [\log(1 - D(G(x)))] + f(epoch) E_{x,y \sim p_{data}(x,y)} [\|y - G(x)\|_1]$$

[0059] 在公式中: $L_{G2}$ 表示生成器的第二损失函数; $x$ 表示带噪语音频谱; $y$ 表示与带噪语音频谱对应的清晰语音频谱; $p_{data}$ 表示训练样本; $E$ 表示从训练样本中取样; $G(x)$ 表示输入为带噪语音频谱时生成器的输出结果; $D(G(x))$ 表示当输入为生成器的输出结果时判别器的

输出结果;  $f(\text{epoch})$  表示与  $\text{epoch}$  (迭代次数) 相关的函数。

[0060] 在本实施例中,  $f(\text{epoch})$  与迭代次数呈反比关系, 即随着迭代次数的增加, 生成器的第二损失函数中的  $E_{x,y \sim p_{\text{data}}(x,y)}[\|y - G(x)\|_1]$  部分的权重逐渐降低。

[0061] 可以理解的是,  $f(\text{epoch})$  的形式可以为  $1/\text{epoch}$ , 也可以为  $10/(1+e^{\text{epoch}})$ 。

[0062] 对上述过程进行举例说明, 若预设的迭代次数为5次, 则, 在第6次迭代之前, 仅利用生成器的第一损失函数对生成器的网络结构中的参数进行更新; 从第6次迭代开始, 使用生成器的第二损失函数对生成器的网络结构中的参数进行更新, 同时使用判别器的损失函数对判别器的网络结构的参数进行更新, 直至生成器的第二损失函数以及判别器的损失函数收敛时, 即认为对生成对抗网络的训练完成。

[0063] 2、第二种训练方式

[0064] 在第一训练阶段训练生成对抗网络时, 可以采用以下方式: 将带噪语音频谱输入生成器, 得到生成器的输出结果; 在迭代次数小于等于预设的迭代次数时, 将生成器的输出结果及其对应的清晰语音频谱输入判别器, 根据判别器的输出结果得到生成器的第二损失函数; 根据生成器的第二损失函数, 更新生成器的网络结构中的参数。

[0065] 相应地, 在第二训练阶段训练生成对抗网络时, 还可以采用以下方式: 将带噪语音频谱输入经过第一训练阶段训练所得到的生成器, 得到生成器的输出结果; 在达到预设的迭代次数之后, 根据生成器的输出结果及其对应的清晰语音频谱, 得到生成器的第一损失函数; 将生成器的输出结果输入判别器, 根据判别器的输出结果得到判别器的损失函数; 根据所得到的第一损失函数更新生成器的网络结构中的参数, 根据判别器的损失函数更新判别器的网络结构中的参数, 直至生成对抗网络收敛。

[0066] 可以理解的是, 第二种方式中计算生成器的第一损失函数、第二损失函数以及判别器的损失函数的所使用的公式与第一种方式中的一致。

[0067] 对上述过程进行举例说明, 若预设的迭代次数为20次, 则, 在第21次迭代之前, 利用生成器的第二损失函数对生成器的网络结构中的参数进行更新; 从第21次迭代开始, 利用生成器的第一损失函数对生成器的网络结构中的参数进行更新, 同时使用判别器的损失函数对判别器的网络结构的参数进行更新, 直至生成器的第一损失函数以及判别器的损失函数收敛时, 即认为对生成对抗网络的训练完成。

[0068] 3、第三种训练方式

[0069] 在第一训练阶段训练生成对抗网络时, 可以采用以下方式: 将带噪语音频谱输入生成器, 得到生成器的输出结果; 在迭代次数小于等于预设的迭代次数时, 将生成器的输出结果及其对应的清晰语音频谱输入判别器, 根据判别器的输出结果得到判别器的损失函数以及生成器的第二损失函数; 根据判别器的损失函数以及生成器的第二损失函数, 更新生成器以及判别器的网络结构中的参数。

[0070] 相应地, 在第二训练阶段训练生成对抗网络时, 还可以采用以下方式: 将带噪语音频谱输入经过第一训练阶段训练所得到的生成器, 得到生成器的输出结果; 在达到预设的迭代次数之后, 根据生成器的输出结果及其对应的清晰语音频谱, 得到生成器的第一损失函数; 根据所得到的第一损失函数更新生成器的网络结构中的参数, 直至第一损失函数收敛, 即认为对生成对抗网络的训练完成。

[0071] 对上述过程进行举例说明, 若预设的迭代次数为100次, 则, 在第101次迭代之前, 利

用生成器的第二损失函数对生成器的网络结构中的参数进行更新,利用判别器的损失函数对判别器的网络结构的参数进行更新;从第101次迭代开始,利用生成器的第一损失函数对生成器的网络结构中的参数进行更新,直至第一损失函数收敛,即认为对生成对抗网络的训练完成。

[0072] (二)根据获取的带噪语音频谱及其对应的清晰语音频谱,采用在三个训练阶段中切换生成器的损失函数的方式对生成对抗网络进行训练。

[0073] 可以采用以下方式在三个训练阶段中切换生成器的损失函数的方式,对生成对抗网络进行训练:

[0074] 在第一训练阶段,当迭代次数小于等于预设的第一迭代次数时,使用生成器的第一损失函数更新生成器的网络结构的参数;在第二训练阶段,当迭代次数大于预设的第一迭代次数且小于等于预设的第二迭代次数时,使用生成器的第二损失函数更新生成器的网络结构的参数;在第三训练阶段,当迭代次数大于预设的第二迭代次数时,使用生成器的第一损失函数更新生成器的网络结构的参数,并同时使用判别器的损失函数更新判别器的网络结构的参数,直至生成对抗网络收敛。

[0075] 还可以采用以下方式在三个训练阶段中切换生成器的损失函数的方式,对生成对抗网络进行训练:

[0076] 在第一训练阶段,当迭代次数小于等于预设的第一迭代次数时,使用生成器的第二损失函数更新生成器的网络结构的参数;在第二训练阶段,当迭代次数大于预设的第一迭代次数且小于等于预设的第二迭代次数时,使用生成器的第一损失函数更新生成器的网络结构的参数,并同时使用判别器的损失函数更新判别器的网络结构的参数;在第三训练阶段,当迭代次数大于预设的第二迭代次数时,使用生成器的第二损失函数更新生成器的网络结构的参数,直至第二损失函数收敛,即认为生成对抗网络收敛。

[0077] 通过以上内容的描述,其他的在三个训练阶段中切换生成器的损失函数的训练方式、在四个训练阶段中切换生成器的损失函数的训练方式、在五个训练阶段中切换生成器的损失函数的训练方式等的具体过程可推理得到。

[0078] 在对生成对抗网络训练完成后,将生成对抗网络中的生成器取出,作为语音增强网络。利用该语音增强网络,能够根据输入的带噪语音频谱得到清晰语音频谱。在将该语音增强网络与语音识别系统组合后,该语音增强网络便能够向语音识别系统输入清晰语音频谱,从而提高在语音识别系统在噪声环境下识别语音的准确性。

[0079] 图2为本发明一实施例提供的一种建立语音增强网络的装置结构图,如图2中所示,所述装置包括:样本获取单元21、网络构建单元22、网络训练单元23以及网络组合单元24。

[0080] 样本获取单元21,用于获取带噪语音频谱以及与各带噪语音频谱对应的清晰语音频谱作为训练样本。

[0081] 样本获取单元21获取多个带噪语音频谱以及与各带噪语音频谱对应的清晰语音频谱作为训练样本。其中,样本获取单元21所获取的清晰语音频谱即为不带噪的语音频谱。

[0082] 可以理解的是,对样本获取单元21所获取的语音频谱的帧长、帧移以及语音频谱中所包含的声学特征不进行限定。

[0083] 网络构建单元22,用于构建包含生成器和判别器的生成对抗网络。

[0084] 网络构建单元22构建包含生成器和判别器的生成对抗网络。

[0085] 网络构建单元22所构建的生成对抗网络包括生成器和判别器。其中,生成器的职责是生成与真实样本尽可能相似的生成样本,判别器的职责则是尽可能区分真实样本和生成样本。通过生成器和判别器之间对抗博弈的方式训练整个生成对抗网络,使得生成器输出的生成样本的真实性尽可能高,从而使得判别器无法区分由生成器所得到的输出为生成样本还是真实样本。

[0086] 其中,生成器的网络结构可以为深度神经网络、循环神经网络或者长短时记忆网络中的一种,判别器的网络结构可以为深度神经网络、循环神经网络或者长短时记忆网络中的一种,本发明对此不进行限定。

[0087] 网络训练单元23,根据获取的带噪语音频谱及其对应的清晰语音频谱,采用在N个训练阶段中切换生成器的损失函数的方式对所述生成对抗网络进行训练,利用训练得到的所述生成对抗网络中的生成器得到语音增强网络,其中N为大于等于2的正整数。

[0088] 通常理解,在训练生成对抗网络时,一般采用交替训练生成器和判别器的方式,训练过程中生成器和判别器的损失函数不发生变化,直至生成对抗网络收敛。而在本发明中,不再使用传统的交替训练生成器和判别器的方式,而是由网络训练单元23采用在N个训练阶段中切换生成器的损失函数的方式,对生成对抗网络中的生成器和判别器进行分阶段训练,其中N为大于等于2的正整数。

[0089] 为了不使说明书中的内容过于冗杂,在下文中仅描述采用两个训练阶段以及三个训练阶段对生成对抗网络进行训练的方式,其他的四个训练阶段、五个训练阶段等的训练方式可以根据本文的描述内容推理得到。

[0090] (一)根据获取的带噪语音频谱及其对应的清晰语音频谱,网络训练单元23采用在两个训练阶段中切换生成器的损失函数的方式对生成对抗网络进行训练。

[0091] 1、第一种训练方式

[0092] 网络训练单元23在第一训练阶段训练生成对抗网络时,可以采用以下方式:将带噪语音频谱输入生成器,得到生成器的输出结果;在迭代次数小于等于预设的迭代次数时,根据生成器的输出结果及其对应的清晰语音频谱,得到生成器的第一损失函数;根据所得到的第一损失函数更新生成器的网络结构中的参数。也就是说,网络训练单元23在该种训练方式中的第一训练阶段只根据第一损失函数来更新生成器的网络结构中的参数,而不更新判别器的网络结构中的参数。

[0093] 网络训练单元23在第二训练阶段训练生成对抗网络时,可以采用以下方式:将带噪语音频谱输入经过第一训练阶段训练所得到的生成器,得到生成器的输出结果;在迭代次数大于预设的迭代次数之后,将生成器的输出结果及其对应的清晰语音频谱输入判别器,根据判别器的输出结果分别得到判别器的损失函数以及生成器的第二损失函数;根据判别器的损失函数以及生成器的第二损失函数,分别更新生成器以及判别器的网络结构中的参数,直至生成对抗网络收敛。

[0094] 2、第二种训练方式

[0095] 网络训练单元23在第一训练阶段训练生成对抗网络时,可以采用以下方式:将带噪语音频谱输入生成器,得到生成器的输出结果;在迭代次数小于等于预设的迭代次数时,将生成器的输出结果及其对应的清晰语音频谱输入判别器,根据判别器的输出结果得到生

成器的第二损失函数;根据生成器的第二损失函数,更新生成器的网络结构中的参数。

[0096] 相应地,网络训练单元23在第二训练阶段训练生成对抗网络时,还可以采用以下方式:将带噪语音频谱输入经过第一训练阶段训练所得到的生成器,得到生成器的输出结果;在达到预设的迭代次数之后,根据生成器的输出结果及其对应的清晰语音频谱,得到生成器的第一损失函数;将生成器的输出结果输入判别器,根据判别器的输出结果得到判别器的损失函数;根据所得到的第一损失函数更新生成器的网络结构中的参数,根据判别器的损失函数更新判别器的网络结构中的参数,直至生成对抗网络收敛。

[0097] 3、第三种训练方式

[0098] 网络训练单元23在第一训练阶段训练生成对抗网络时,可以采用以下方式:将带噪语音频谱输入生成器,得到生成器的输出结果;在迭代次数小于等于预设的迭代次数时,将生成器的输出结果及其对应的清晰语音频谱输入判别器,根据判别器的输出结果得到判别器的损失函数以及生成器的第二损失函数;根据判别器的损失函数以及生成器的第二损失函数,更新生成器以及判别器的网络结构中的参数。

[0099] 相应地,网络训练单元23在第二训练阶段训练生成对抗网络时,还可以采用以下方式:将带噪语音频谱输入经过第一训练阶段训练所得到的生成器,得到生成器的输出结果;在达到预设的迭代次数之后,根据生成器的输出结果及其对应的清晰语音频谱,得到生成器的第一损失函数;根据所得到的第一损失函数更新生成器的网络结构中的参数,直至第一损失函数收敛,即认为对生成对抗网络的训练完成。

[0100] (二)根据获取的带噪语音频谱及其对应的清晰语音频谱,网络训练单元23采用在三个训练阶段中切换生成器的损失函数的方式对生成对抗网络进行训练。

[0101] 网络训练单元23可以采用以下方式在三个训练阶段中切换生成器的损失函数的方式,对生成对抗网络进行训练:

[0102] 在第一训练阶段,当迭代次数小于等于预设的第一迭代次数时,网络训练单元23使用生成器的第一损失函数更新生成器的网络结构的参数;在第二训练阶段,当迭代次数大于预设的第一迭代次数且小于等于预设的第二迭代次数时,网络训练单元23使用生成器的第二损失函数更新生成器的网络结构的参数;在第三训练阶段,当迭代次数大于预设的第二迭代次数时,网络训练单元23使用生成器的第一损失函数更新生成器的网络结构的参数,并同时使用判别器的损失函数更新判别器的网络结构的参数,直至生成对抗网络收敛。

[0103] 网络训练单元23还可以采用以下方式在三个训练阶段中切换生成器的损失函数的方式,对生成对抗网络进行训练:

[0104] 在第一训练阶段,当迭代次数小于等于预设的第一迭代次数时,网络训练单元23使用生成器的第二损失函数更新生成器的网络结构的参数;在第二训练阶段,当迭代次数大于预设的第一迭代次数且小于等于预设的第二迭代次数时,网络训练单元23使用生成器的第一损失函数更新生成器的网络结构的参数,并同时使用判别器的损失函数更新判别器的网络结构的参数;在第三训练阶段,当迭代次数大于预设的第二迭代次数时,网络训练单元23使用生成器的第二损失函数更新生成器的网络结构的参数,直至第二损失函数收敛,即认为生成对抗网络收敛。

[0105] 通过以上内容的描述,网络训练单元23其他的在三个训练阶段中切换生成器的损失函数的训练方式、在四个训练阶段中切换生成器的损失函数的训练方式、在五个训练阶

段中切换生成器的损失函数的训练方式等的具体过程可推理得到。

[0106] 在网络训练单元23对生成对抗网络训练完成后,将生成对抗网络中的生成器取出,作为语音增强网络。利用该语音增强网络,能够根据输入的带噪语音频谱得到清晰语音频谱。

[0107] 网络组合单元24,用于将所述语音增强网络与语音识别系统组合。

[0108] 网络组合单元24在将通过网络训练单元23所得到的语音增强网络与语音识别系统组合后,该语音增强网络便能够向语音识别系统输入清晰语音频谱,从而提高在语音识别系统在噪声环境下识别语音的准确性。

[0109] 如图3所示,计算机系统/服务器012以通用计算设备的形式表现。计算机系统/服务器012的组件可以包括但不限于:一个或者多个处理器或者处理单元016,系统存储器028,连接不同系统组件(包括系统存储器028和处理单元016)的总线018。

[0110] 总线018表示几类总线结构中的一种或多种,包括存储器总线或者存储器控制器,外围总线,图形加速端口,处理器或者使用多种总线结构中的任意总线结构的局域总线。举例来说,这些体系结构包括但不限于工业标准体系结构(ISA)总线,微通道体系结构(MAC)总线,增强型ISA总线、视频电子标准协会(VESA)局域总线以及外围组件互连(PCI)总线。

[0111] 计算机系统/服务器012典型地包括多种计算机系统可读介质。这些介质可以是任何能够被计算机系统/服务器012访问的可用介质,包括易失性和非易失性介质,可移动的和不可移动的介质。

[0112] 系统存储器028可以包括易失性存储器形式的计算机系统可读介质,例如随机存取存储器(RAM)030和/或高速缓存存储器032。计算机系统/服务器012可以进一步包括其它可移动/不可移动的、易失性/非易失性计算机系统存储介质。仅作为举例,存储系统034可以用于读写不可移动的、非易失性磁介质(图3未显示,通常称为“硬盘驱动器”)。尽管图3中未示出,可以提供用于对可移动非易失性磁盘(例如“软盘”)读写的磁盘驱动器,以及对可移动非易失性光盘(例如CD-ROM,DVD-ROM或其它光介质)读写的光盘驱动器。在这些情况下,每个驱动器可以通过一个或者多个数据介质接口与总线018相连。存储器028可以包括至少一个程序产品,该程序产品具有一组(例如至少一个)程序模块,这些程序模块被配置以执行本发明各实施例的功能。

[0113] 具有一组(至少一个)程序模块042的程序/实用工具040,可以存储在例如存储器028中,这样的程序模块042包括——但不限于——操作系统、一个或者多个应用程序、其它程序模块以及程序数据,这些示例中的每一个或某种组合中可能包括网络环境的实现。程序模块042通常执行本发明所描述的实施例中的功能和/或方法。

[0114] 计算机系统/服务器012也可以与一个或多个外部设备014(例如键盘、指向设备、显示器024等)通信,在本发明中,计算机系统/服务器012与外部雷达设备进行通信,还可与一个或者多个使得用户能与该计算机系统/服务器012交互的设备通信,和/或与使得该计算机系统/服务器012能与一个或多个其它计算设备进行通信的任何设备(例如网卡,调制解调器等等)通信。这种通信可以通过输入/输出(I/O)接口022进行。并且,计算机系统/服务器012还可以通过网络适配器020与一个或者多个网络(例如局域网(LAN),广域网(WAN)和/或公共网络,例如因特网)通信。如图所示,网络适配器020通过总线018与计算机系统/服务器012的其它模块通信。应当明白,尽管图中未示出,可以结合计算机系统/服务器012

使用其它硬件和/或软件模块,包括但不限于:微代码、设备驱动器、冗余处理单元、外部磁盘驱动阵列、RAID系统、磁带驱动器以及数据备份存储系统等。

[0115] 处理单元016通过运行存储在系统存储器028中的程序,从而执行各种功能应用以及数据处理,例如实现本发明实施例所提供的方法流程。

[0116] 上述的计算机程序可以设置于计算机存储介质中,即该计算机存储介质被编码有计算机程序,该程序在被一个或多个计算机执行时,使得一个或多个计算机执行本发明上述实施例中所示的方法流程和/或装置操作。例如,被上述一个或多个处理器执行本发明实施例所提供的方法流程。

[0117] 随着时间、技术的发展,介质含义越来越广泛,计算机程序的传播途径不再受限于有形介质,还可以直接从网络下载等。可以采用一个或多个计算机可读的介质的任意组合。计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质。计算机可读存储介质例如可以是——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件,或者任意以上的组合。计算机可读存储介质的更具体的例子(非穷举的列表)包括:具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、光纤、便携式紧凑磁盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本文件中,计算机可读存储介质可以是任何包含或存储程序的有形介质,该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。

[0118] 计算机可读的信号介质可以包括在基带中或者作为载波一部分传播的数据信号,其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式,包括——但不限于——电磁信号、光信号或上述的任意合适的组合。计算机可读的信号介质还可以是计算机可读存储介质以外的任何计算机可读介质,该计算机可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。

[0119] 计算机可读介质上包含的程序代码可以用任何适当的介质传输,包括——但不限于——无线、电线、光缆、RF等等,或者上述的任意合适的组合。

[0120] 可以以一种或多种程序设计语言或其组合来编写用于执行本发明操作的计算机程序代码,所述程序设计语言包括面向对象的程序设计语言——诸如Java、Smalltalk、C++,还包括常规的过程式程序设计语言——诸如“C”语言或类似的设计语言。程序代码可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中,远程计算机可以通过任意种类的网络——包括局域网(LAN)或广域网(WAN)连接到用户计算机,或者,可以连接到外部计算机(例如利用因特网服务提供商来通过因特网连接)。

[0121] 利用本发明所提供的技术方案,通过在多个训练阶段中切换生成器的损失函数的方法对生成对抗网络进行训练,能够增强生成对抗网络训练收敛的稳定性,从而提升基于生成对抗网络的语音增强网络的性能,进而实现提高语音识别准确性的目的。

[0122] 在本发明所提供的几个实施例中,应该理解到,所揭露的系统,装置和方法,可以通过其它的方式实现。例如,以上所描述的装置实施例仅仅是示意性的,例如,所述单元的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式。

[0123] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0124] 另外,在本发明各个实施例中的各功能单元可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现,也可以采用硬件加软件功能单元的形式实现。

[0125] 上述以软件功能单元的形式实现的集成的单元,可以存储在一个计算机可读取存储介质中。上述软件功能单元存储在一个存储介质中,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备等)或处理器(processor)执行本发明各个实施例所述方法的部分步骤。而前述的存储介质包括:U盘、移动硬盘、只读存储器(Read-Only Memory,ROM)、随机存取存储器(Random Access Memory,RAM)、磁碟或者光盘等各种可以存储程序代码的介质。

[0126] 以上所述仅为本发明的较佳实施例而已,并不用以限制本发明,凡在本发明的精神和原则之内,所做的任何修改、等同替换、改进等,均应包含在本发明保护的范围之内。

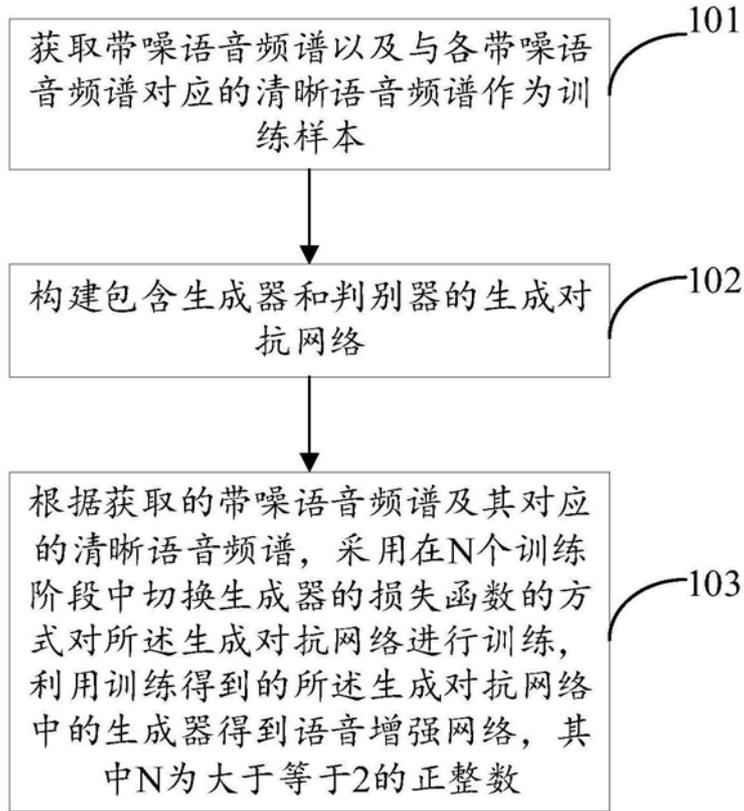


图1

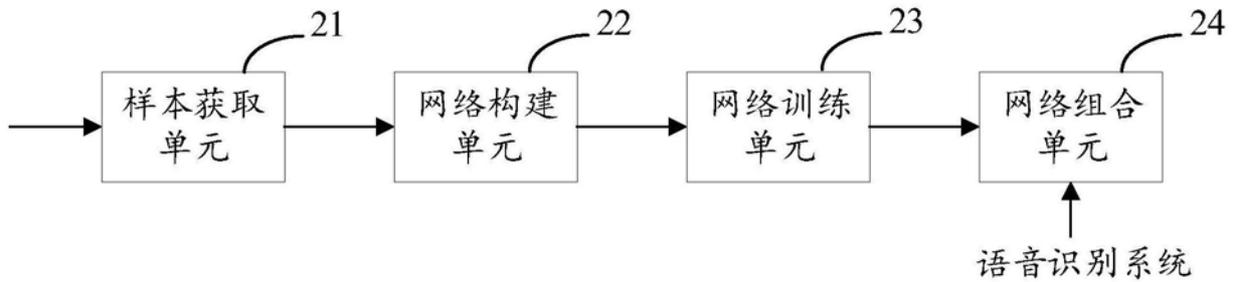


图2

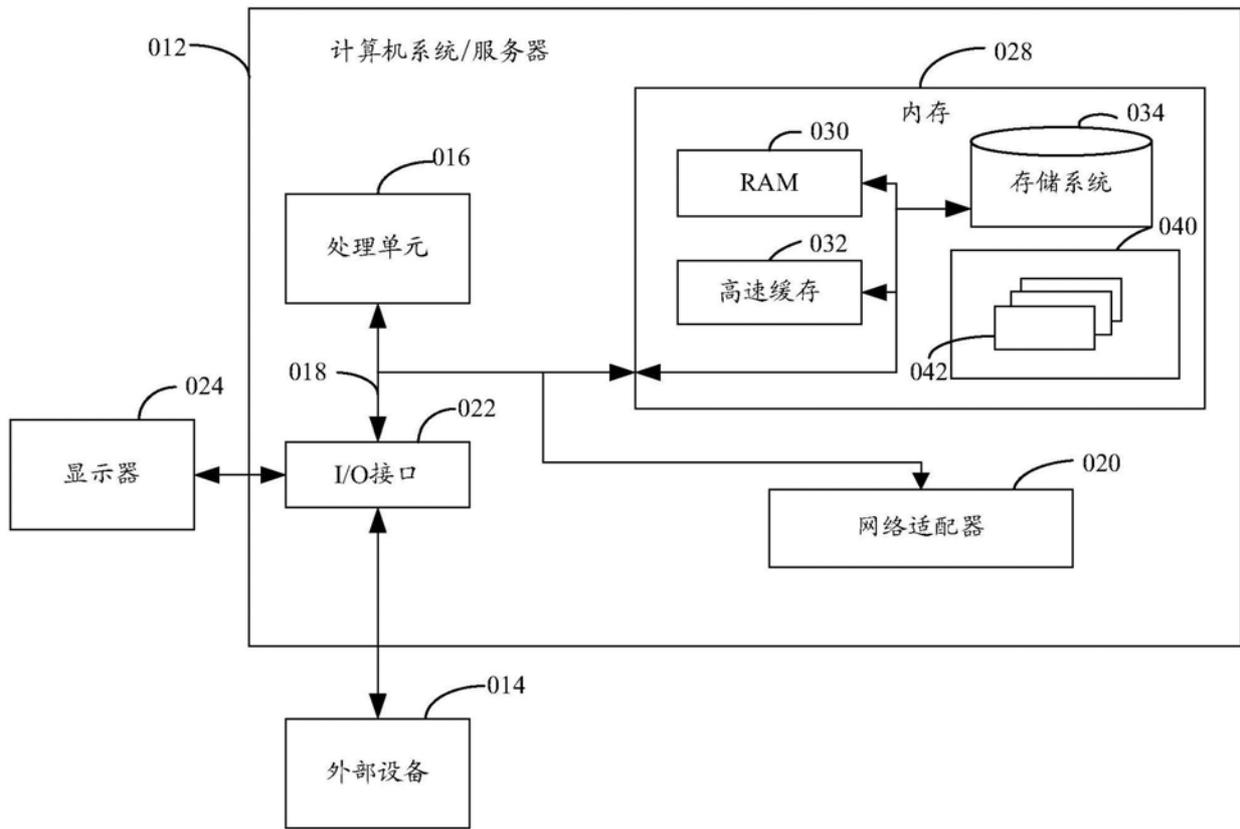


图3