(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 112908312 B
(45)授权公告日 2022.06.24

(21)申请号 202110132066.6

(22)申请日 2021.01.30

(65)同一申请的已公布的文献号
申请公布号 CN 112908312 A

(43)申请公布日 2021.06.04

(73)专利权人 云知声智能科技股份有限公司
地址 100096 北京市海淀区西三旗建材城内1幢一层101号
专利权人 厦门云知芯智能科技有限公司

(72)发明人 许东星 曹昊 周雷 梁家恩

(51)Int.Cl.
G10L 15/06(2013.01)
G10L 15/16(2006.01)
G10L 15/22(2006.01)

(56)对比文件
CN 110534098 A,2019.12.03
CN 110265040 A,2019.09.20
US 2010185444 A1,2010.07.22
CN 111179944 A,2020.05.19
CN 109545218 A,2019.03.29
CN 105895105 A,2016.08.24
CN 110619889 A,2019.12.27
CN 110473526 A,2019.11.19
JP 2020140029 A,2020.09.03
惠琳等.短时频谱通用背景模型群联合韵律的年龄语音转换.《声学学报》.2017,(第06期),
方尔庆等.基于视听信息的自动年龄估计方法.《软件学报》.2011,(第07期),

审查员 王玥

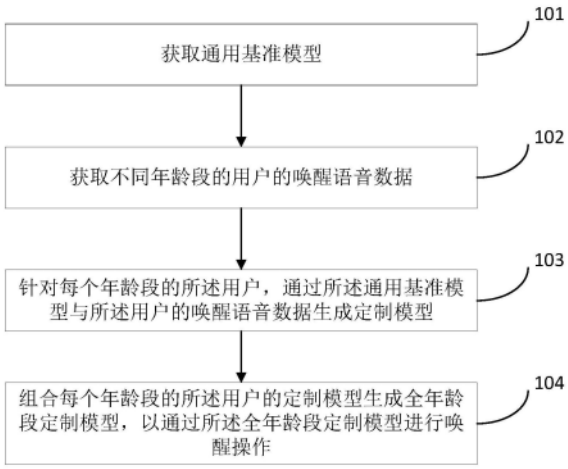
权利要求书1页 说明书4页 附图2页

(54)发明名称

一种提高唤醒性能的方法和设备

(57)摘要

本发明涉及一种提高唤醒性能的方法和设备，该方法包括：获取通用基准模型，其中，所述通用基准模型是通过带标注的语音训练神经网络得到的；获取不同年龄段的用户的唤醒语音数据；针对每个年龄段的所述用户，通过所述通用基准模型与所述用户的唤醒语音数据生成定制模型；组合每个年龄段的所述用户的定制模型生成全年龄段定制模型，以通过所述全年龄段定制模型进行唤醒操作。通过对不同年龄段的人群定制不同的神经网络模型，可以有针对性地提高不同年龄段的模型区分能力，提高不同年龄段人群唤醒性能；另外组合成的全年龄段定制模型，且本方案可以复用大部分神经网络计算，对计算资源消耗增加有限。



CN 112908312 B

1.一种提高唤醒性能的方法,其特征在于,包括:

获取通用基准模型,其中,所述通用基准模型是通过带标注的语音训练神经网络得到的;

获取不同年龄段的用户的唤醒语音数据;

针对每个年龄段的所述用户,通过所述通用基准模型与所述用户的唤醒语音数据生成定制模型;

组合每个年龄段的所述用户的定制模型生成全年龄段定制模型,以通过所述全年龄段定制模型进行唤醒操作。

2.根据权利要求1所述的方法,其特征在于,所述年龄段包括:儿童年龄段、中青年人年龄段、老年人年龄段;

所述唤醒语音数据覆盖所述用户的全年龄段。

3.根据权利要求1所述的方法,其特征在于,所述定制模型是基于所述用户的唤醒语音数据对所述通用基准模型进行迁移学习得到的。

4.根据权利要求1所述的方法,其特征在于,所述神经网络包括共享层与迁移层;

不同年龄段的所述用户的定制模型的共享层相同,且迁移层参数不同。

5.根据权利要求4所述的方法,其特征在于,所述全年龄段定制模型通过声学分进行唤醒操作;

所述声学分由每个年龄段的所述定制模型中的所述迁移层的输出分数进行加权得到。

6.一种提高唤醒性能的设备,其特征在于,包括:

第一获取模块,用于获取通用基准模型,其中,所述通用基准模型是通过带标注的语音训练神经网络得到的;

第二获取模块,用于获取不同年龄段的用户的唤醒语音数据;

定制模块,用于针对每个年龄段的所述用户,通过所述通用基准模型与所述用户的唤醒语音数据生成定制模型;

组合模块,用于组合每个年龄段的所述用户的定制模型生成全年龄段定制模型,以通过所述全年龄段定制模型进行唤醒操作。

7.根据权利要求6所述的设备,其特征在于,所述年龄段包括:儿童年龄段、中青年人年龄段、老年人年龄段;

所述唤醒语音数据覆盖所述用户的全年龄段。

8.根据权利要求6所述的设备,其特征在于,所述定制模型是基于所述用户的唤醒语音数据对所述通用基准模型进行迁移学习得到的。

9.根据权利要求6所述的设备,其特征在于,所述神经网络包括共享层与迁移层;

不同年龄段的所述用户的定制模型的共享层相同,且迁移层参数不同。

10.根据权利要求9所述的设备,其特征在于,所述全年龄段定制模型通过声学分进行唤醒操作;

所述声学分由每个年龄段的所述定制模型中的所述迁移层的输出分数进行加权得到。

一种提高唤醒性能的方法和设备

技术领域

[0001] 本发明涉及语音识别技术领域，具体涉及一种提高唤醒性能的方法和设备。

背景技术

[0002] 在语音交互系统中，唤醒是很重要的交互入口。唤醒需要实时处理语音，不仅需要唤醒率高，误唤醒率低，同时对内存和功耗等系统资源也要求苛刻。目前的唤醒系统，通常采用神经网络作为声学模型，将声学模型的输出输入唤醒系统，得到唤醒输出及唤醒输出的置信度。最终根据预先设定好的阈值，得到唤醒结果。

[0003] 但是由于资源限制，唤醒采用的神经网络参数通常都较小，采用统一的神经网络模型，对不同年龄段人群的声学输出区分度不足，导致最终唤醒性能不佳。

发明内容

[0004] 本发明提供一种提高唤醒性能的方法和设备，能够解决现有技术中唤醒性能不佳的技术问题。

[0005] 本发明解决上述技术问题的技术方案如下：

[0006] 本发明实施例提出了一种提高唤醒性能的方法，包括：

[0007] 获取通用基准模型，其中，所述通用基准模型是通过带标注的语音训练神经网络得到的；

[0008] 获取不同年龄段的用户的唤醒语音数据；

[0009] 针对每个年龄段的所述用户，通过所述通用基准模型与所述用户的唤醒语音数据生成定制模型；

[0010] 组合每个年龄段的所述用户的定制模型生成全年龄段定制模型，以通过所述全年龄段定制模型进行唤醒操作。

[0011] 在一个具体的实施例中，所述年龄段包括：儿童年龄段、中青年人年龄段、老年人年龄段；

[0012] 所述唤醒语音数据覆盖所述用户的全年龄段。

[0013] 在一个具体的实施例中，所述定制模型是基于所述用户的唤醒语音数据对所述通用基准模型进行迁移学习得到的。

[0014] 在一个具体的实施例中，所述神经网络包括共享层与迁移层；

[0015] 不同年龄段的所述用户的定制模型的共享层相同，且迁移层参数不同。

[0016] 在一个具体的实施例中，所述全年龄段定制模型通过声学分进行唤醒操作；

[0017] 所述声学分由每个年龄段的所述定制模型中的所述迁移层的输出分数进行加权得到。

[0018] 本发明实施例还提出了一种提高唤醒性能的设备，包括：

[0019] 第一获取模块，用于获取通用基准模型，其中，所述通用基准模型是通过带标注的语音训练神经网络得到的；

[0020] 第二获取模块，用于获取不同年龄段的用户的唤醒语音数据；

[0021] 定制模块，用于针对每个年龄段的所述用户，通过所述通用基准模型与所述用户的唤醒语音数据生成定制模型；

[0022] 组合模块，用于组合每个年龄段的所述用户的定制模型生成全年龄段定制模型，以通过所述全年龄段定制模型进行唤醒操作。

[0023] 在一个具体的实施例中，所述年龄段包括：儿童年龄段、中青年人年龄段、老年人年龄段；

[0024] 所述唤醒语音数据覆盖所述用户的全年龄段。

[0025] 在一个具体的实施例中，所述定制模型是基于所述用户的唤醒语音数据对所述通用基准模型进行迁移学习得到的。

[0026] 在一个具体的实施例中，所述神经网络包括共享层与迁移层；

[0027] 不同年龄段的所述用户的定制模型的共享层相同，且迁移层参数不同。

[0028] 在一个具体的实施例中，所述全年龄段定制模型通过声学分进行唤醒操作；

[0029] 所述声学分由每个年龄段的所述定制模型中的所述迁移层的输出分数进行加权得到。

[0030] 本发明的有益效果是：

[0031] 通过对不同年龄段的人群定制不同的神经网络模型，可以有针对性地提高不同年龄段的模型区分能力，提高不同年龄段人群唤醒性能；另外组合成的全年龄段定制模型，且本方案可以复用大部分神经网络计算，对计算资源消耗增加有限。

附图说明

[0032] 图1为本发明实施例提供的一种提高唤醒性能的方法的流程示意图；

[0033] 图2为本发明实施例提供的一种具体的提高唤醒性能的方法的流程示意图；

[0034] 图3为本发明实施例提供的一种提高唤醒性能的方法中的通用基准模型训练示意图；

[0035] 图4为本发明实施例提供的一种提高唤醒性能的方法中单个年龄段定制模型训练流程图；

[0036] 图5为本发明实施例提供的一种提高唤醒性能的方法中将不同年龄段的定制模型组合成全年龄段定制模型的示意图；

[0037] 图6为本发明实施例提供的一种提高唤醒性能的设备的结构示意图。

具体实施方式

[0038] 以下结合附图对本发明的原理和特征进行描述，所举实例只用于解释本发明，并非用于限定本发明的范围。

[0039] 实施例1

[0040] 本发明实施例提供的一种提高唤醒性能的方法，如图1或2所示，包括以下步骤：

[0041] 步骤101、获取通用基准模型，其中，所述通用基准模型是通过带标注的语音训练神经网络得到的；具体的，如图3所示，利用大量带标注的语音，训练一个神经网络，作为通用基准模型。

[0042] 步骤102、获取不同年龄段的用户的唤醒语音数据;具体的,所述年龄段包括:儿童年龄段、中青年人年龄段、老年人年龄段;所述唤醒语音数据覆盖所述用户的全年龄段。

[0043] 具体的,通过发起唤醒语音采集任务,将目标客户群体按照年龄分为不同的年龄段,分别为儿童、中青年人、老年人。语音采集需要覆盖不同的年龄段,例如儿童为0-18岁,中青年人为18-55岁,老年人为大于55岁的年龄段。

[0044] 步骤103、针对每个年龄段的所述用户,通过所述通用基准模型与所述用户的唤醒语音数据生成定制模型;

[0045] 具体的,所述定制模型是基于所述用户的唤醒语音数据对所述通用基准模型进行迁移学习得到的。

[0046] 进一步的,所述神经网络包括共享层与迁移层;

[0047] 不同年龄段的所述用户的定制模型的共享层相同,且迁移层参数不同。

[0048] 具体的,如图4所示,分别为每个年龄段的用户定制一个模型,这里,每个年龄段的定制模型由通用基准模型通过迁移学习得到。每个年龄段的定制模型与通用基准模型结构相同,差别主要表现在迁移层参数不同。

[0049] 步骤104、组合每个年龄段的所述用户的定制模型生成全年龄段定制模型,以通过所述全年龄段定制模型进行唤醒操作。

[0050] 具体的,如图5所示,将不同年龄段的定制模型组合成全年龄段定制模型。4利用全年龄段定制模型进行唤醒检测。唤醒引擎利用全年龄段定制模型,声学分由每个年龄段迁移层输出分数加权得到。

[0051] 所述全年龄段定制模型通过声学分进行唤醒操作;

[0052] 所述声学分由每个年龄段的所述定制模型中的所述迁移层的输出分数进行加权得到。

[0053] 本方案通过对不同年龄段的人群定制不同的神经网络模型,可以有针对性地提高不同年龄段的模型区分能力,提高不同年龄段人群唤醒性能;另外组合成的全年龄段定制模型,可以复用大部分神经网络计算(共享层计算占整个NN计算的90%以上),对计算资源消耗增加有限

[0054] 实施例2

[0055] 本发明实施例2还公开了一种提高唤醒性能的设备,如图6所示,包括:

[0056] 第一获取模块201,用于获取通用基准模型,其中,所述通用基准模型是通过带标注的语音训练神经网络得到的;

[0057] 第二获取模块202,用于获取不同年龄段的用户的唤醒语音数据;

[0058] 定制模块203,用于针对每个年龄段的所述用户,通过所述通用基准模型与所述用户的唤醒语音数据生成定制模型;

[0059] 组合模块204,用于组合每个年龄段的所述用户的定制模型生成全年龄段定制模型,以通过所述全年龄段定制模型进行唤醒操作。

[0060] 在一个具体的实施例中,所述年龄段包括:儿童年龄段、中青年人年龄段、老年人年龄段;

[0061] 所述唤醒语音数据覆盖所述用户的全年龄段。

[0062] 在一个具体的实施例中,所述定制模型是基于所述用户的唤醒语音数据对所述通

用基准模型进行迁移学习得到的。

[0063] 在一个具体的实施例中，所述神经网络包括共享层与迁移层；

[0064] 不同年龄段的所述用户的定制模型的共享层相同，且迁移层参数不同。

[0065] 在一个具体的实施例中，所述全年龄段定制模型通过声学分进行唤醒操作；

[0066] 所述声学分由每个年龄段的所述定制模型中的所述迁移层的输出分数进行加权得到。

[0067] 本发明涉及一种提高唤醒性能的方法和设备，该方法包括：获取通用基准模型，其中，所述通用基准模型是通过带标注的语音训练神经网络得到的；获取不同年龄段的用户的唤醒语音数据；针对每个年龄段的所述用户，通过所述通用基准模型与所述用户的唤醒语音数据生成定制模型；组合每个年龄段的所述用户的定制模型生成全年龄段定制模型，以通过所述全年龄段定制模型进行唤醒操作。通过对不同年龄段的人群定制不同的神经网络模型，可以有针对性地提高不同年龄段的模型区分能力，提高不同年龄段人群唤醒性能；另外组合成的全年龄段定制模型，且本方案可以复用大部分神经网络计算，对计算资源消耗增加有限。

[0068] 以上，仅为本发明的具体实施方式，但本发明的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本发明揭露的技术范围内，可轻易想到各种等效的修改或替换，这些修改或替换都应涵盖在本发明的保护范围之内。因此，本发明的保护范围应以权利要求的保护范围为准。

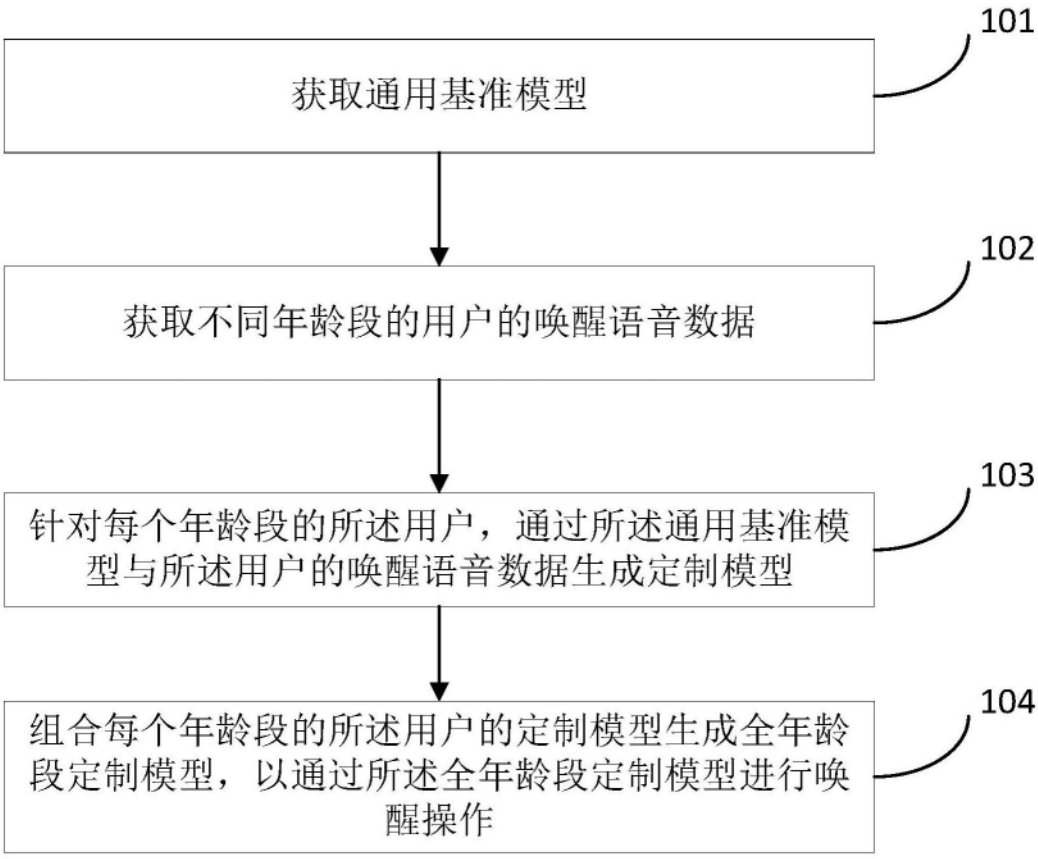

101
获取通用基准模型
102
获取不同年龄段的用户的唤醒语音数据
103
针对每个年龄段的所述用户，通过所述通用基准模型与所述用户的唤醒语音数据生成定制模型
104
组合每个年龄段的所述用户的定制模型生成全年龄段定制模型，以通过所述全年龄段定制模型进行唤醒操作

图1

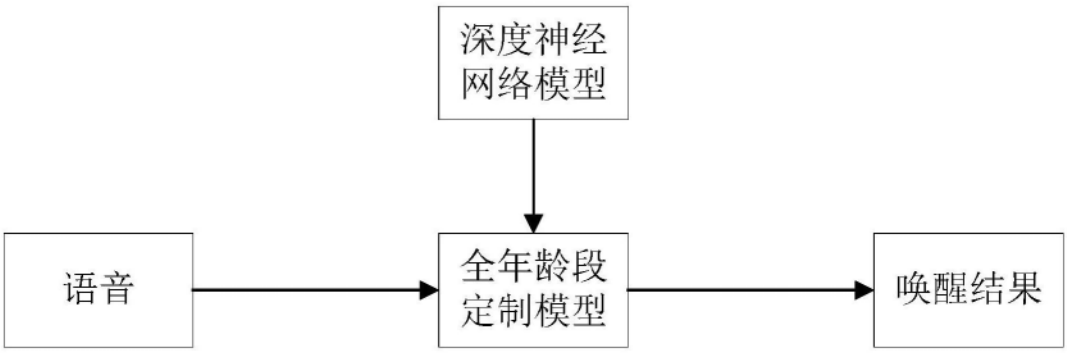

深度神经网络模型
语音
全年龄段定制模型
唤醒结果

图2

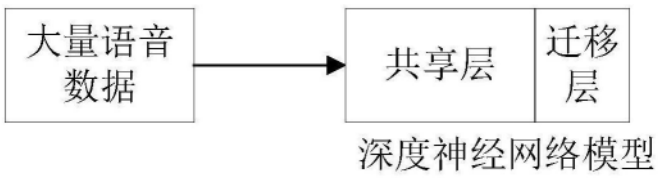

大量语音数据
共享层
迁移层
深度神经网络模型

图3

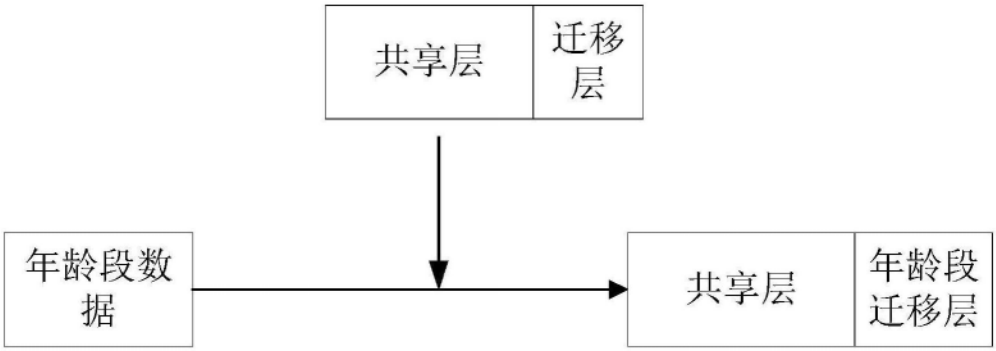

共享层
迁移层
年龄段数据
共享层
年龄段迁移层

图4

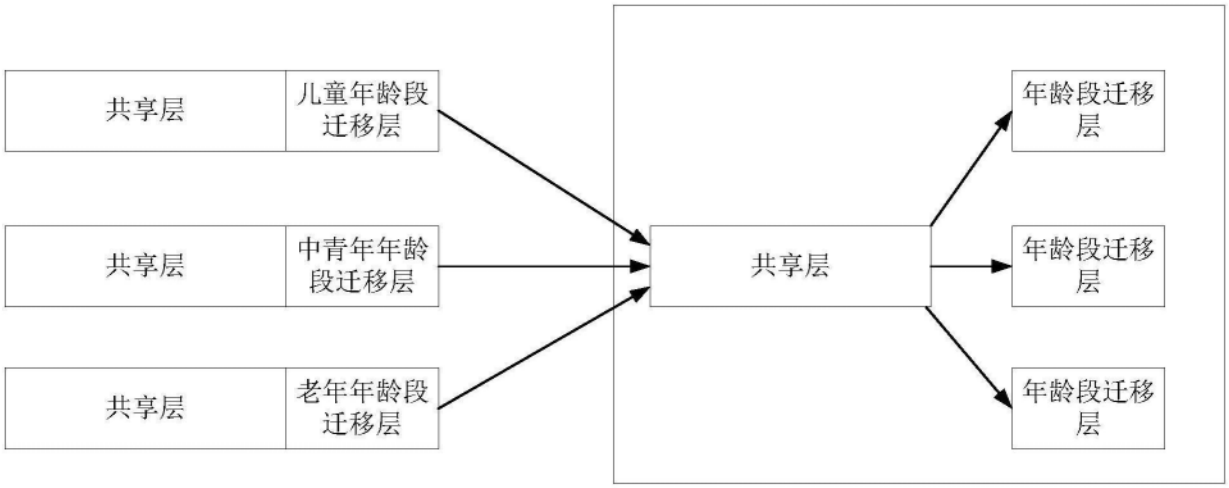

共享层
儿童年龄段迁移层
共享层
中青年年龄段迁移层
共享层
老年年龄段迁移层
共享层
年龄段迁移层
年龄段迁移层
年龄段迁移层

图5

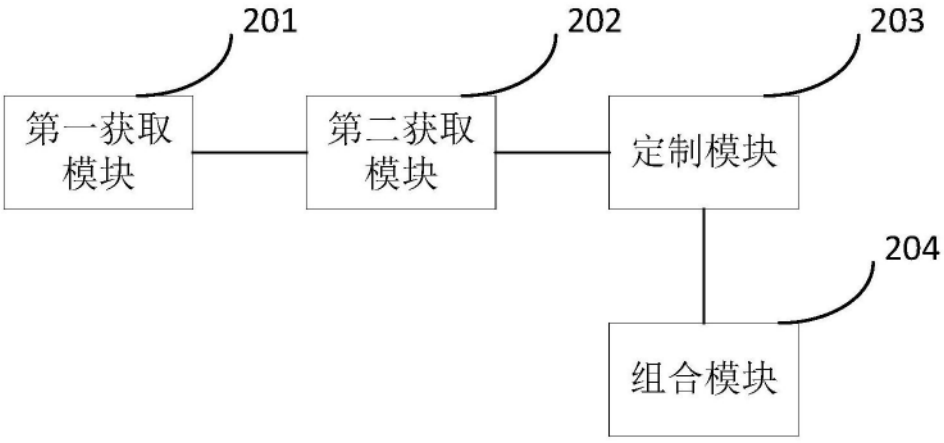

201
202
203
第一获取模块
第二获取模块
定制模块
204
组合模块

图6