



(12) 发明专利申请

(10) 申请公布号 CN 119404204 A

(43) 申请公布日 2025. 02. 07

(21) 申请号 202380046481.7

(22) 申请日 2023.05.23

(30) 优先权数据

2022-100193 2022.06.22 JP

(85) PCT国际申请进入国家阶段日

2024.12.11

(86) PCT国际申请的申请数据

PCT/JP2023/019081 2023.05.23

(87) PCT国际申请的公布数据

W02023/248676 JA 2023.12.28

(71) 申请人 松下电器(美国)知识产权公司

地址 美国加利福尼亚

(72) 发明人 大毛胜统 中尾武寿

(74) 专利代理机构 永新专利商标代理有限公司
72002

专利代理师 徐殿军

(51) Int.Cl.

G06Q 10/06 (2006.01)

G06Q 50/04 (2006.01)

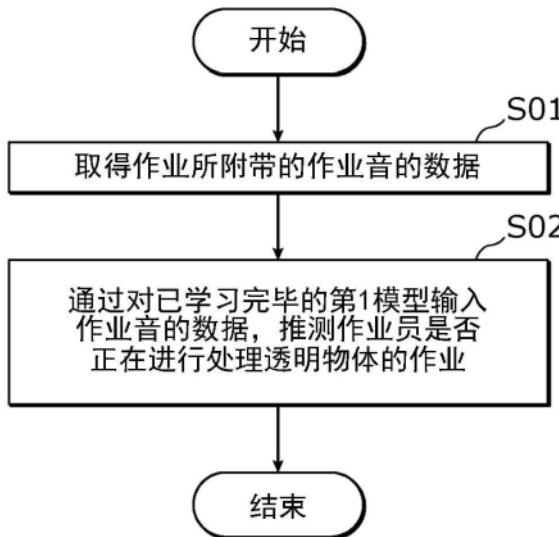
权利要求书2页 说明书22页 附图31页

(54) 发明名称

推测方法及推测装置

(57) 摘要

推测方法是由计算机进行的推测作业员的作业的推测方法,计算机取得收集到的作业所附带的作业音的数据(S01),通过对已学习完毕的第1模型输入作业音的数据,推测作业员是否正在进行处理透明物体的作业(S02)。



1. 一种推测方法,是由计算机进行的推测作业员的作业的推测方法,其中,
所述计算机取得收集到的所述作业所附带的作业音的数据;
所述计算机通过对已学习完毕的第1模型输入所述作业音的数据,推测所述作业员是否正在进行处理透明物体的作业。
2. 如权利要求1所述的推测方法,其中,
所述计算机取得与所述作业音的数据对应的、显现有进行所述作业的所述作业员的图像的数据;
所述计算机通过对已学习完毕的第2模型输入所述图像的数据,推测所述作业员是否正在进行处理所述透明物体的作业;
所述计算机基于使用所述第1模型的推测结果及使用所述第2模型的推测结果,推测所述作业员是否正在进行处理所述透明物体的作业。
3. 如权利要求1所述的推测方法,其中,
所述计算机取得与所述作业音的数据对应的、显现有进行所述作业的所述作业员的图像的数据;
所述计算机通过将所述作业音的数据和所述图像的数据输入到所述第1模型中,推测所述作业员是否正在进行处理所述透明物体的作业。
4. 如权利要求1~3中任一项所述的推测方法,其中,
所述计算机基于从所述第1模型输出的所述作业音的特征量与预先保存在存储部中的处理所述透明物体的作业的作业音的特征量之间的类似度,推测所述作业员是否正在进行处理所述透明物体的作业。
5. 如权利要求1~3中任一项所述的推测方法,其中,
所述计算机进一步基于从所述第1模型输出的所述作业音的特征量相对于预先保存在存储部中的处理所述透明物体的作业的作业音的特征量的类似度、以及相对于预先保存在所述存储部中的可能被误推测为所述作业员正在进行处理所述透明物体的作业的作业音的特征量的类似度,推测所述作业员是否正在进行处理所述透明物体的作业。
6. 如权利要求5所述的推测方法,其中,
所述计算机在从所述第1模型输出的所述作业音的特征量相对于处理所述透明物体的作业的所述作业音的特征量的所述类似度比相对于可能被误推测为所述作业员正在进行处理所述透明物体的作业的所述作业音的特征量的所述类似度高的情况下,推测为所述作业员正在进行处理所述透明物体的作业。
7. 如权利要求5所述的推测方法,其中,
所述计算机在将处理与所述透明物体不同的非透明物体的作业的作业音的数据输入到所述第1模型中而得到的处理所述非透明物体的作业的所述作业音的特征量与处理所述透明物体的作业的所述作业音的特征量之间的类似度高于阈值的情况下,将处理所述非透明物体的作业的所述作业音判定为可能被误推测的作业音,该可能被误推测的作业音是可能被误推测为处理所述透明物体的作业的作业音的作业音;
所述计算机将处理所述非透明物体的作业的所述作业音的特征量作为所述可能被误推测的作业音的特征量保存至所述存储部。
8. 如权利要求1~3中任一项所述的推测方法,其中,

所述作业音的数据包含非可听频带的声音的数据。

9. 一种推测装置, 推测作业员的作业, 其具有:

取得部, 取得收集到的所述作业所附带的作业音的数据; 以及
推测部, 通过对已学习完毕的第1模型输入所述作业音的数据, 推测所述作业员是否正在进行处理透明物体的作业。

10. 一种程序, 其中,

用来使计算机执行权利要求1~3中任一项所述的推测方法。

推测方法及推测装置

技术领域

[0001] 本公开涉及推测作业员的作业的推测方法等。

背景技术

[0002] 作为用于工厂的生产性提高的第一阶段,将与作业员进行的作业有关的数据自动地收集而进行作业的分类,按每个分类测量作业所需要的时间。例如,在专利文献1中,公开了通过根据在多个摄像条件下拍摄的图像识别作业中处理的物体(例如透明物体等)来进行作业的分类的技术。

[0003] 现有技术文献

[0004] 专利文献

[0005] 专利文献1:日本特开2018—017653号公报

发明内容

[0006] 发明要解决的课题

[0007] 但是,在专利文献1所记载的技术中,即使使摄像条件变化,在物体的透明度高的情况下,或者在物体的光的折射率或反射率的变化小的情况下,该物体的识别精度低。因此,在专利文献1所记载的技术中,存在不能精度良好地推测处理透明性高的物体(以下称作透明物体)的作业的情况。

[0008] 所以,本公开提供一种能够精度良好地推测处理透明物体的作业的推测方法等。

[0009] 用来解决课题的手段

[0010] 有关本公开的一个技术方案推测方法,是由计算机进行的推测作业员的作业的推测方法,所述计算机取得收集到的所述作业所附带的作业音的数据;所述计算机通过对已学习的第1模型输入所述作业音的数据,推测所述作业员是否正在进行处理透明物体的作业。

[0011] 发明效果

[0012] 根据本公开,能够精度良好地推测处理透明物体的作业。

附图说明

[0013] 图1是表示实施方式的推测系统的功能结构的一例的框图。

[0014] 图2是表示实施方式的推测系统的动作例1的流程图。

[0015] 图3是示意地表示图2的步骤S02中的流程的一例的图。

[0016] 图4是表示收集到的作业音的特征量与处理透明物体的作业的作业音的特征量的类似度的曲线图的图。

[0017] 图5是表示在验证例1中对1小时的作业音以时间序列进行分析的结果的图。

[0018] 图6是用来说明由验证例3执行的袋作业的推测方法的图。

[0019] 图7是表示神经网络的架构的一例的图。

- [0020] 图8是表示进行两个分类的推测的情况下的正确回答率的计算方法的图。
- [0021] 图9是表示验证例3的两个分类的推测结果及正确回答率的图。
- [0022] 图10是表示进行3个分类的推测的情况下的正确回答率的计算方法的图。
- [0023] 图11是表示验证例3的3个分类的推测结果及正确回答率的图。
- [0024] 图12是表示基于输入数据的组合的两个分类的推测方法及正确回答率的计算方法的图。
- [0025] 图13是表示验证例3的基于输入数据的组合的两个分类的推测结果及正确回答率的图。
- [0026] 图14是表示使用图像AI的推测方法及动作例1的推测方法的推测精度的比较结果的图。
- [0027] 图15是用来说明使用作业音的数据的推测结果与使用图像的数据的推测结果的差异的图。
- [0028] 图16是用来说明实施方式的推测系统的动作例2的流程的概要的图。
- [0029] 图17是表示实施方式的推测系统的动作例2的流程图。
- [0030] 图18是表示实施方式的推测系统的动作例2的变形例1的流程图。
- [0031] 图19是示意地表示执行动作例2的变形例1的流程的推测部的结构例1的图。
- [0032] 图20是用来说明由结构例1执行的袋作业的推测方法的图。
- [0033] 图21是示意地表示执行动作例2的变形例1的流程的推测部的结构例2的图。
- [0034] 图22是示意地表示执行动作例2的变形例1的流程的推测部的结构例3的图。
- [0035] 图23是表示图像子网络的架构的一例的图。
- [0036] 图24是表示声音子网络的架构的一例的图。
- [0037] 图25是表示融合层的架构的一例的图。
- [0038] 图26是表示分类网络的架构的一例的图。
- [0039] 图27是表示对比学习网络的架构的一例的图。
- [0040] 图28是示意地表示执行动作例2的变形例2的流程的推测部的结构例的图。
- [0041] 图29是表示推测部误推测为作业员正在进行处理透明物体的作业时的作业音的例子图。
- [0042] 图30A是实施方式的推测系统的动作例3的流程图。
- [0043] 图30B是表示将可能被误推测的作业音的特征量预先登记的动作的一例的流程图。
- [0044] 图31是表示有关其他实施方式的推测系统的功能结构的一例的框图。

具体实施方式

[0045] (作为本公开的基础的见解)

[0046] 作为用于工厂的生产性提高的第一阶段,自动地收集与作业员进行的作业有关的数据而进行作业的分类,按每个分类测量作业所需要的时间。由此,用户能够掌握作业员在哪个作业中需要时间,所以能够设立作业计划以便作业员能够更有效率地进行作业。

[0047] 以往,将作业员进行的作业的状况用照相机进行摄影,通过识别作业员处理的物体来进行作业的分类。例如,在专利文献1中,根据改变摄像条件而拍摄的多个图像识别透

明物体,分类为作业员正在进行处理透明物体的作业。但是,对于透明性高的物体(即透明物体),即使使摄像条件变化,在其透明度高的情况或该物体的光的折射率或反射率的变化小的情况下也难以通过影像识别该物体。因此,在专利文献1所记载的技术中,存在不能精度良好地推测处理透明物体的作业的情况。

[0048] 所以,要求以下方法:通过即使在作业中处理的物体的透明性高的情况或该物体的光的折射率或反射率的变化小的情况下也精度良好地识别该物体、精度良好地推测进行处理透明物体的作业,能够将作业员的作业精度良好地分类。

[0049] 此外,在以往的方法中,主要着眼于用照相机对静止状态的物体进行摄影来识别该物体。对此,本申请的发明人发现,通过对作业所附带的作业音(换言之,随着作业产生的声音)进行拾音,即使在由作业员进行的作业中透明物体移动或变形,也能够精度良好地推测处理透明物体的作业。

[0050] (本公开的概要)

[0051] 有关本公开的一个技术方案的例1的推测方法,是由计算机进行的推测作业员的作业的推测方法,其中,所述计算机取得收集到的所述作业所附带的作业音的数据;所述计算机通过对已学习完毕的第1模型输入所述作业音的数据,推测所述作业员是否正在进行处理透明物体的作业。

[0052] 由此,执行推测方法的装置由于使用以作业音的数据为输入、输出是否是处理透明物体的作业的第1模型,所以能够精度良好地推测处理透明物体的作业。

[0053] 此外,有关本公开的一个技术方案的例2的推测方法在例1的推测方法中,也可以是,所述计算机取得与所述作业音的数据对应的、显现有进行所述作业的所述作业员的图像的数据;所述计算机通过对已学习完毕的第2模型输入所述图像的数据,推测所述作业员是否正在进行处理所述透明物体的作业;所述计算机基于使用所述第1模型的推测结果及使用所述第2模型的推测结果,推测所述作业员是否正在进行处理所述透明物体的作业。另外,使用第1模型的推测结果是由第1模型根据作业音的数据推测的推测结果,使用第2模型的推测结果是由第2模型根据图像的数据推测的推测结果。

[0054] 由此,执行推测方法的装置基于由第1模型根据作业音的数据推测出的推测结果以及由第2模型根据图像的数据推测出的推测结果,推测作业员是否正在进行处理透明物体的作业。因此,执行推测方法的装置与仅使用作业音的数据推测的情况相比能够精度更好地推测处理透明物体的作业。

[0055] 此外,有关本公开的一个技术方案的例3的推测方法在例1的推测方法中,也可以是,所述计算机取得与所述作业音的数据对应的、显现有进行所述作业的所述作业员的图像的数据;所述计算机通过将所述作业音的数据和所述图像的数据输入到所述第1模型中,推测所述作业员是否正在进行处理所述透明物体的作业。

[0056] 由此,执行推测方法的装置由于使用以作业音的数据及对应于作业音的图像的数据为输入、输出是否是处理透明物体的作业的第1模型,所以与仅使用作业音的数据进行推测的情况相比,能够精度更好地推测处理透明物体的作业。

[0057] 此外,有关本公开的一个技术方案的例4的推测方法在例1~例3的任一项的推测方法中,也可以是,所述计算机基于从所述第1模型输出的所述作业音的特征量与预先保存在存储部中的处理所述透明物体的作业的作业音的特征量之间的相似度,推测所述作业员

是否正在进行处理所述透明物体的作业。

[0058] 由此,执行推测方法的装置由于基于从第1模型输出的作业音的特征量与处理透明物体的作业的作业音的特征量的类似度,推测作业员是否正在进行处理透明物体的作业,所以能够精度良好地推测处理透明物体的作业。

[0059] 此外,有关本公开的一个技术方案的例5的推测方法在例1~例4的任一项的推测方法中,也可以是,所述计算机进一步基于从所述第1模型输出的所述作业音的特征量相对于预先保存在存储部中的处理所述透明物体的作业的作业音的特征量的类似度、以及相对于预先保存在所述存储部中的可能被误推测为所述作业员正在进行处理所述透明物体的作业的作业音的特征量的类似度,推测所述作业员是否正在进行处理所述透明物体的作业。

[0060] 由此,执行推测方法的装置通过将第1模型输出的作业音的特征量与处理透明物体的作业的作业音的特征量的类似度和从第1模型输出的可能被误推测为作业音的特征量的作业音的特征量的类似度比较,能够减少误推测的发生。因而,执行推测方法的装置即使仅使用作业音的数据也能够精度良好地推测处理透明物体的作业。

[0061] 此外,有关本公开的一个技术方案的例6的推测方法在例5的推测方法中,也可以是,所述计算机在从所述第1模型输出的所述作业音的特征量相对于处理所述透明物体的作业的所述作业音的特征量的所述类似度比相对于可能被误推测为所述作业员正在进行处理所述透明物体的作业的所述作业音的特征量的所述类似度高的情况下,推测为所述作业员正在进行处理所述透明物体的作业。

[0062] 由此,执行推测方法的装置能够减少误推测的发生,所以即使仅使用作业音的数据也能够精度良好地推测处理透明物体的作业。

[0063] 此外,有关本公开的一个技术方案的例7的推测方法在例5或例6的推测方法中,也可以是,所述计算机在将处理与所述透明物体不同的非透明物体的作业的作业音的数据输入到所述第1模型中而得到的处理所述非透明物体的作业的所述作业音的特征量与处理所述透明物体的作业的所述作业音的特征量之间的类似度高于阈值的情况下,将处理所述非透明物体的作业的所述作业音判定为可能被误推测的作业音,该可能被误推测的作业音是可能被误推测为处理所述透明物体的作业的作业音;所述计算机将处理所述非透明物体的作业的所述作业音的特征量作为所述可能被误推测的作业音的特征量保存至所述存储部。

[0064] 由此,执行推测方法的装置能够基于处理非透明物体的作业的作业音的特征量与处理透明物体的作业的作业音的特征量的类似度,精度良好地判定该处理非透明物体的作业的作业音是否是可能被误推测为正在进行处理透明物体的作业的作业音。因此,执行推测方法的装置能够将被误推测的可能性比较高的作业音的特征量保存到存储部中。因而,执行推测方法的装置通过使用保存在存储部中的可能被误推测的作业音的特征量,能够减少误推测的发生,所以即使仅使用作业音的数据也能够精度良好地推测处理透明物体的作业。

[0065] 此外,有关本公开的一个技术方案的例8的推测方法在例1~例7的任一项的推测方法中,也可以是,所述作业音的数据包含非可听频带的声音的数据。

[0066] 由此,执行推测方法的装置使用包含从可听频带的声音到非可听频带的声音的作

业音的数据推测作业员是否正在进行处理透明物体的作业。这样,通过作业音的数据包含非可听频带的声音,作业音的数据中的成为误推测的原因的来自环境的杂音变少,所以执行推测方法的装置能够提高处理透明物体的作业的推测精度。进而,执行推测方法的装置能够基于比仅使用可听频带的声音的数据的情况多的信息来推测作业员是否正在进行处理透明物体的作业。因而,执行推测方法的装置能够精度更好地推测处理透明物体的作业。

[0067] 此外,有关本公开的一个技术方案的例9的推测装置是推测作业员的作业的推测装置,具有:取得部,取得收集到的所述作业所附带的作业音的数据;以及推测部,通过对已学习完毕的第1模型输入所述作业音的数据,推测所述作业员是否正在进行处理透明物体的作业。

[0068] 由此,推测装置由于使用以作业音的数据为输入、输出是否是处理透明物体的作业的第1模型,所以能够精度良好地推测处理透明物体的作业。

[0069] 此外,有关本公开的一个技术方案的例10的程序是用来使计算机执行例1~例8的任一项的推测方法的程序。

[0070] 由此,能够使用计算机起到与上述的推测方法同样的效果。

[0071] 另外,这些包含性或具体的技术方案也可以由系统、方法、装置、集成电路、计算机程序或计算机可读的CD-ROM(Compact Disc Read Only memory,只读光盘)等记录介质实现,也可以由系统、方法、装置、集成电路、计算机程序及记录介质的任意的组合实现。

[0072] 以下,一边参照附图一边对本公开的实施方式具体地进行说明。在以下的实施方式中表示的数值、形状、材料、构成要素、构成要素的配置位置及连接形态、步骤、步骤的顺序等是一例,不旨在限定权利要求的范围。此外,关于以下的实施方式的构成要素中的在表示最上位概念的独立权利要求中没有记载的构成要素,设为任意的构成要素进行说明。此外,各图并不一定严密地进行图示。在各图中,对于实质上相同的结构赋予相同的标号,有将重复的说明省略或简略化的情况。

[0073] 此外,在本公开中,平行及垂直等表示要素间的关系性的用语以及矩形等表示要素的形状的用语及数值不仅表示严格的意思,而意味着实质上同等的范围,例如也包含几个百分点左右的差异。

[0074] (实施方式)

[0075] 以下,一边参照附图一边对实施方式具体地进行说明。

[0076] [1.概要]

[0077] 首先,对实施方式的推测系统的概要进行说明。图1是表示实施方式的推测系统200的功能结构的一例的框图。

[0078] 推测系统200是推测作业员的作业的系统。推测系统200取得例如由拾音装置10收集到的作业所附带的作业音,通过将作业音的数据输入到已学习完毕的第1模型132(以下,也简称作第1模型132)来推测作业员是否正在进行处理透明物体的作业。

[0079] 推测系统200例如可以将由推测装置100推测出的推测结果显示在信息终端50的显示部上并提示给用户。由此,用户能够参照推测结果而掌握处理透明物体的作业及处理非透明物体的作业所需要的时间,此外,用户能够参照推测结果而设立作业员的作业计划,所以能够提高作业空间80中的作业的效率。

[0080] 作业所附带的作业音包含伴随着作业产生的声音。作业音例如是使作业员处理的

物体移动或变形时产生的声音。作业例如是零件的拾取、清扫作业、检验或打包等。作业空间80例如是指在制造工厂或物流仓库等中作业员进行作业的空间。

[0081] 此外,透明物体是透明性高的物体,例如由合成树脂或玻璃等透明性高的材料构成。透明性高是指,例如在该物体是片状的情况下、或在由片状体构成的情况下片的雾度(haze)小于0.5%,在该物体是平板状或块状的情况下、或在由平板状或块状体构成的情况下,光的折射率为1.30以上1.70以下。透明物体例如是容器、袋、缓冲材或零件等。

[0082] 作为合成树脂,例如可以是聚氯乙烯树脂等乙烯树脂、聚碳酸酯树脂、聚酯树脂、聚萘二甲酸乙二醇酯树脂、聚乙烯树脂、聚丙烯树脂、聚酰亚胺树脂、聚苯乙烯树脂、聚氨酯树脂、丙烯酸树脂、氟树脂等。另外,构成透明性高的物体的材料并不限于上述的例子,例如也可以包含微细纤维状纤维素等天然聚合物。

[0083] 另外,推测系统200可以取得由摄像装置20拍摄的显现有进行作业的作业员的图像的数据,将所取得的图像的数据和作业音的数据输入到第1模型132中,由此来推测作业员是否正在进行处理透明物体的作业,也可以基于将图像的数据输入到已学习完毕的第2模型133(以下,也简称作第2模型133)中而得到的推测结果与将作业音的数据输入到第1模型132中而得到的推测结果来推测作业员是否正在进行处理透明物体的作业。图像的数据与作业音的数据相对应。

[0084] [2. 结构]

[0085] 接着,一边参照图1一边对实施方式的推测系统200的结构进行说明。推测系统200例如具有拾音装置10、摄像装置20、信息终端50和推测装置100。拾音装置10及摄像装置20设置在作业员进行作业的空间(作业空间80)中,与信息终端50及推测装置100经由通信连接。另外,图1所示的推测系统200的结构只不过是一例,并不限定于该例。

[0086] [拾音装置10]

[0087] 拾音装置10例如收集作业员的作业所附带的作业音。拾音装置10例如设置在作业空间80中。拾音装置10能够收集从可听频带到非可听频带的声音。可听频带是能够由人耳感知的频带,非可听频带是不能由人耳感知的频带。非可听频带的声音例如是20kHz以上的频带的声音。拾音装置10更具体地讲是麦克风,例如既可以是(Micro Electro Mechanical Systems,微机电系统)麦克风,也可以是激光麦克风。

[0088] 拾音装置10例如在是激光麦克风的情况下,能够收集比通常的麦克风频带更宽的声音。此外,激光麦克风由于不像通常的麦克风那样具有振动板,所以在电磁波、高温或高热等环境下也能够拾音。

[0089] 在图1中,示出了推测系统200具有1个拾音装置10的例子,但也可以具有两个以上的拾音装置10。此外,拾音装置10也可以是指向性的麦克风。由此,拾音装置10不易收集周围的杂音等成为噪声的声音,所以能够高灵敏度地收集作业音。

[0090] 拾音装置10将收集到的声音(作业音)转换为电信号,向推测装置100输出。另外,拾音装置10也可以对收集到的作业音的数据赋予时间戳和自身的识别号码,向推测装置100输出。

[0091] [摄像装置20]

[0092] 摄像装置20例如拍摄显现有进行作业的作业员的图像。该图像的数据与由拾音装置10收集到的作业音的数据对应。即,摄像装置20也可以与拾音装置10连动动作,例如通过

对所取得的数据(作业音的数据及图像的数据)赋予时间戳,将作业音的数据与图像的数据建立对应。此时,例如摄像装置20也可以对图像数据赋予自身的识别号码。摄像装置20例如设置在作业空间80中。摄像装置20例如是RGB照相机,但也可以包含距离数据。

[0093] 摄像装置20将所拍摄的图像的数据向推测装置100输出。

[0094] [信息终端50]

[0095] 信息终端50是用户使用的信息终端,例如是个人计算机或平板电脑终端等。信息终端50将由推测装置100推测出的推测结果显示在显示部上。此外,信息终端50受理由用户输入的指示,将该指示向拾音装置10、摄像装置20及推测装置100发送。

[0096] [推测装置100]

[0097] 推测装置100是推测作业员的作业的装置。推测装置100例如通过取得由拾音装置10收集到的作业所附带的作业音的数据并将该作业音的数据输入到已学习完毕的第1模型132中,推测作业员是否正在进行处理透明物体的作业。

[0098] 例如,如图1所示,推测装置100具有通信部110、信息处理部120、存储部130、模型生成部140和输入受理部150。推测装置100例如是服务器装置。另外,在图1的例子中,推测装置100具有第2模型133,但并不一定需要具有第2模型133。以下,对推测装置100的各结构进行说明。

[0099] [通信部110]

[0100] 通信部110是用于推测装置100与拾音装置10及摄像装置20进行通信的通信电路(通信模组)。通信部110具有用来经由广域通信网络进行通信的通信电路(通信模组),但也可以具有用来经由局部通信网络进行通信的通信电路(通信模组)。通信部110例如是进行无线通信的无线通信电路,但也可以是进行有线通信的有线通信电路。另外,关于通信部110进行的通信的通信规格不作特别限定。

[0101] [信息处理部120]

[0102] 信息处理部120进行与推测装置100有关的各种信息处理。更具体地讲,例如信息处理部120取得由拾音装置10收集到的作业音的数据(例如,作业音的电信号),进行与作业员是否正在进行处理透明物体的作业的推测有关的各种信息处理。此外,例如信息处理部120也可以取得由摄像装置20拍摄的、显现有进行作业的作业员的图像的数据,进行与作业员是否正在进行处理透明物体的作业的推测有关的各种信息处理。信息处理部120既可以使用作业音的数据进行作业的推测,也可以使用作业音的数据和图像的数据进行作业的推测。信息处理部120具体而言,具有取得部121和推测部122。取得部121及推测部122的功能通过构成信息处理部120的处理器或微型计算机执行存储在存储部130中的计算机程序来实现。

[0103] [取得部121]

[0104] 取得部121例如取得由拾音装置10收集到的作业音的数据。作业音的数据是作业员的作业所附带的声音,例如是伴随着由作业员进行的作业而产生的声音。此外,取得部121例如取得由摄像装置20拍摄的、与作业音的数据对应的显现有进行作业的作业员的图像的数据。作业音的数据既可以是对由拾音装置10收集到的作业音的电信号进行傅里叶变换而制作的频谱的图像,也可以是时间序列的数值数据。

[0105] [推测部122]

[0106] 如果取得部121取得了作业音的数据,则推测部122根据作业音的数据,推测作业员是否正在进行处理透明物体的作业。推测部122例如通过对已学习完毕的第1模型132(以下称作第1模型132)输入作业音的数据,推测作业员是否正在进行处理透明物体的作业。具体而言,例如,推测部122基于从第1模型132输出的作业音的特征量与预先保存在存储部130(例如,存储部130内的特征量数据库131)中的处理透明物体的作业的作业音的特征量的类似度,推测作业员是否正在进行处理透明物体的作业。更具体地讲,例如推测部122可以将作业音的数据输入到第1模型132中,计算由第1模型132提取出的处理透明物体的作业的作业音的特征量与预先保存在存储部130中的处理透明物体的作业的作业音的特征量的类似度,在计算出的类似度为规定值(即阈值)以上的情况下,推测为作业员正在进行处理透明物体的作业。另外,并不限于该例,推测部122也可以使用根据作业音的数据直接输出作业员是否正在进行处理透明物体的作业的推测结果的模型。

[0107] 此外,如果取得部121取得了与作业音的数据对应的、显现有进行作业的作业员的图像的数据,则推测部122也可以根据作业音的数据及图像的数据推测作业员是否正在进行处理透明物体的作业。具体而言,例如推测部122通过将作业音的数据与对应于作业音的数据的显现有进行作业的作业员的图像的数据输入到第1模型132中,从而推测作业员是否正在进行处理透明物体的作业。关于第1模型132的详细情况在后面叙述。

[0108] 此外,例如在推测装置100具有已学习完毕的第2模型133的情况下,如果取得部121取得了上述的图像的数据,则推测部122通过将该图像的数据输入到第2模型133中,推测作业员是否正在进行处理透明物体的作业。此时,推测部122通过将由取得部121取得的显现于该图像的数据中的作业员进行的作业的作业音的数据输入到第1模型132中,推测作业员是否正在进行处理透明物体的作业。并且,推测部122基于使用第2模型133根据图像的数据推测出的推测结果和使用第1模型132根据作业音的数据推测出的推测结果,推测作业员是否正在进行处理透明物体的作业。

[0109] 此外,推测部122例如可以判定由拾音装置10收集到的作业音是否是可能被误推测为处理透明物体的作业的作业音的作业音。具体而言,例如在将与处理透明物体不同的非透明物体的作业的作业音的数据输入到第1模型132中而得到的处理非透明物体的作业的作业音的特征量与处理透明物体的作业的作业音的特征量的类似度高于规定值(即阈值)的情况下,推测部122将该处理非透明物体的作业的作业音判定为可能被推测部122误推测为处理透明物体的作业的作业音的作业音。并且,推测部122将被判定为可能被误推测的作业音的作业音的特征量保存在存储部130的特征量数据库131(特征量DB)中。

[0110] 另外,在特征量数据库131中,可以保存预先保存的处理透明物体的作业的作业音的特征量。关于特征量数据库131在后面说明。

[0111] [存储部130]

[0112] 存储部130是存储用于信息处理部120执行各种信息处理的专用的应用程序等的存储装置。例如,在存储部130中,储存有特征量数据库131、第1模型132和第2模型133。此外,存储部130例如由HDD(Hard Disk Drive,硬盘驱动)实现,但也可以由半导体存储器实现。

[0113] 特征量数据库131储存预先提取出的作业音的特征量。该特征量可以作为嵌入(例如张量、矩阵等)、嵌入向量或方差表现等而由数值或数值的组合表现。例如,在特征量数据

库131中,可以储存处理透明物体的作业所附带的作业音的特征量及可能被误推测为作业人员正在进行处理透明物体的作业的作业音的特征量。此外,特征量数据库131可以储存预先提取出的图像的特征量。例如,特征量数据库131可以储存显现有进行处理透明物体的作业的作业员的图像的特征量(具体而言,表示显现在图像中的透明物体的特征量)。

[0114] 第1模型132例如是由模型生成部140生成的已学习完毕的模型。第1模型132例如以作业音的数据为输入,输出作业人员是否正在进行处理透明物体的作业。更具体地讲,第1模型132例如提取输入的作业音数据的特征量,计算提取出的特征量与预先保存在存储部130中的处理透明物体的作业的作业音的特征量的类似度,在计算出的类似度为规定值以上的情况下,推测为作业人员正在进行处理透明物体的作业。此外,例如第1模型132可以还以与作业音的数据对应的、显现有进行作业的作业员的图像的数据为输入,输出作业人员是否正在进行处理透明物体的作业。更具体地讲,第1模型132例如可以提取输入的图像数据的特征量,计算提取出的特征量与预先保存在存储部130中的显现有进行处理透明物体的作业的作业员的图像的特征量的类似度,在计算出的类似度为规定值以上的情况下,推测为作业人员正在进行处理透明物体的作业。

[0115] 第2模型133是由模型生成部140生成的已学习完毕的模型。第2模型133例如以与作业音的数据对应的、显现有进行作业的作业员的图像的数据为输入,输出作业人员是否正在进行处理透明物体的工作。更具体地讲,第2模型133例如可以提取输入的图像数据的特征量,计算提取出的特征量与预先保存在存储部130中的显现有进行处理透明物体的作业的作业员的图像的特征量的类似度,在计算出的类似度为规定值以上的情况下,推测为作业人员正在进行处理透明物体的作业。

[0116] 另外,第1模型132及第2模型133也可以提取输入数据的特征量,输出提取出的特征量。

[0117] 第1模型132及第2模型133具体而言是神经网络模型,例如可以是卷积神经网络(CNN)、循环神经网络(RNN)或LSTM(Long—Short Term Memory,长短期记忆)。

[0118] [模型生成部140]

[0119] 模型生成部140例如通过使用教师数据进行机械学习,生成第1模型132及第2模型133。例如,模型生成部140通过机械学习,生成以作业音的数据为输入、输出作业人员是否正在进行处理透明物体的作业的声音识别模型(以下,也称作音频子网络)。此外,例如模型生成部140可以通过机械学习,进一步生成以对应于作业音的数据的、显现有进行作业的作业员的图像的数据为输入、输出作业人员是否正在进行处理透明物体的作业的图像识别模型(以下,也称作影像子网络)。第1模型132例如既可以是声音识别模型,也可以是包括声音识别模型和图像识别模型的模型。对第1模型132输入的作业音的数据例如既可以是频谱的图像,也可以是时间序列的数值数据。作业音的数据可以包含非可听频带的声音的数据。

[0120] 此外,模型生成部140可以通过机械学习,生成以图像的数据为输入、输出表示显现在图像中的透明物体的特征量的图像识别模型(例如,第2模型133)。

[0121] 如上述那样,声音识别模型例如提取输入的作业音的数据的特征量,计算提取出的特征量与预先保存在存储部130中的处理透明物体的作业的作业音的特征量的类似度,在计算出的类似度为规定值以上的情况下,推测为作业人员正在进行透明作业。此外,图像识别模型例如提取输入的图像的数据的特征量,计算提取出的特征量与预先保存在存储部

130中的显现有进行处理透明物体的作业的作业员的图像的特征量的类似度,在计算出的类似度为规定值以上的情况下,推测为作业员正在进行处理透明物体的作业。另外,包括声音识别模型和图像识别模型的模型基于使用这两个模型的推测结果,推测作业员是否正在进行处理透明物体的作业。

[0122] 模型生成部140也可以通过将已学习完毕的模型储存到存储部130中,将第1模型132及第2模型133更新。模型生成部140例如通过处理器执行存储在存储部130中的程序来实现。

[0123] 另外,第1模型132及第2模型133可以提取输入数据的特征量,输出提取出的特征量。

[0124] [输入受理部150]

[0125] 输入受理部150是受理由使用推测装置100的用户进行的操作输入的输入接口。输入受理部150具体而言由触摸面板显示器等实现。例如,在输入受理部150搭载有触摸面板显示器的情况下,触摸面板显示器作为显示部(未图示)及输入受理部150发挥功能。另外,输入受理部150并不限于触摸面板显示器,例如也可以是键盘、定点设备(例如,触控笔或鼠标)或硬件按钮等。此外,输入受理部150在受理通过声音进行的输入的情况下,也可以是麦克风。

[0126] [3.动作例]

[0127] 接着,对实施方式的推测系统200的动作例进行说明。

[0128] [动作例1]

[0129] 首先,一边参照图2一边对实施方式的推测系统200的动作例1具体地进行说明。图2是表示实施方式的推测系统200的动作例1的流程图。

[0130] 虽然在图2中没有表示,但在推测系统200,例如拾音装置10收集作业员的作业所附带的作业音,将收集到的作业音的数据向推测装置100输出。

[0131] 推测装置100的取得部121取得由拾音装置10收集到的作业音的数据(S01),将所取得的作业音的数据向推测部122输出。

[0132] 接着,推测装置100的推测部122通过将作业音的数据输入到已学习完毕的第1模型132中,推测作业员是否正在进行处理透明物体的作业(S02)。

[0133] 以下,对步骤S02具体地进行说明。图3是示意地表示图2的步骤S02的流程的一例的图。例如,推测部122将从取得部121取得的作业时的声音数据(即作业音的数据)分割为每隔规定的时间(例如2秒)的数据,将分割后的数据输入到声音识别模型(例如第1模型132)中。此时,如图3所示,可以在将作业音的数据输入到声音识别模型中之前进行标准化等前处理。声音识别模型从输入的作业音的数据中提取处理透明物体的作业的作业音的特征量。在这里,将由声音识别模型提取出的特征量称作评价对象的特征量即评价声音特征量。

[0134] 接着,推测部122计算表示从声音识别模型输出的评价声音特征量与事先登记在存储部130中的处理透明物体的作业的作业音(在这里,称作目的声音)的特征量即登记特征量以何种程度类似的类似度,输出计算出的类似度。

[0135] 图4是表示收集到的作业音的特征量与处理透明物体的作业的作业音的特征量的类似度的曲线图的图。在图4中,还记载了用户对由摄像装置20拍摄的图像以目视进行确

认,将作业员正在进行处理透明物体的作业的时间(在这里,称作作业时间)与作业员并未进行处理透明物体的作业的时间(在这里,称作非作业时间)进行区别所得的结果。此外,图中的虚线表示类似度的阈值。例如,在由声音识别模型提取出的作业音的特征量相对于事先登记的处理透明物体的作业的作业音的特征量的类似度为阈值(这里是30)以上的情况下推测为作业员正在进行处理透明物体的作业。如图4所示,作业时间与非作业时间的差由类似度的得分表示。例如,如果收集到在处理透明物体(例如,塑料袋、缓冲材等)的作业中产生的声音,则类似度的得分上升。另一方面,例如在透明物体置于作业台之上但作业员没有接触透明物体的时间中,由于没有收集到从透明物体发出的声音,所以没有计算出类似度的得分。

[0136] [动作例1的验证例1]

[0137] 接着,对验证了动作例1的作业的推测精度的验证例1进行说明。在验证例1中,对1小时的作业音以时间序列进行了分析。图5是表示在验证例1中对1小时的作业音以时间序列进行了分析的结果的图。在验证例1中,透明物体是透明的塑料袋(以下称作透明袋),计算将作业员的作业所附带的作业音的数据输入到第1模型132(例如,图3的声音识别模型)中而得到的作业音的特征量与事先登记的处理透明袋的作业所附带的作业音的特征量的类似度。另外,在验证例1中收集的作业音的数据是可听频带的声音的数据,但也可以包含非可听频带的声音的数据。

[0138] 与图4同样地,用户对图像以目视进行确认而手动赋予表示正在进行处理透明袋的作业(以下,也称作袋作业)的正确答案标签。在图5的例子中,将作业员没有与透明袋接触但透明袋存在于作业台之上的状态和作业员正在将商品装袋的状态设为“袋作业”而赋予正确答案标签。另一方面,将作业员正在登记文件的状态、拆除包装的作业等设为不是处理透明袋的作业(即,非袋作业)。

[0139] 此外,图5所示的图像的特征量的类似度示出了表示在使用图像识别模型提取出的图像中显现的透明袋的特征量与表示在预先登记的图像中显现的透明袋的特征量的类似度。

[0140] 如图5所示,在产生了从透明袋发出的声音以外的声音的情况下,类似度的得分上升。在验证例1中,由声音识别模型得到的作业的认识精度,正确回答率是28%,错误回答率是5%。

[0141] 另外,在动作例1的验证例1中,说明了计算类似度而推测透明物体的第1模型132的一例和其动作的流程的一例,但并不限于该例。例如,第1模型132也可以是输入作业音的数据而直接推测(换言之输出)是否是处理透明物体的作业模型。以下,说明第1模型132的另一例和其动作的流程的一例。

[0142] [动作例1的验证例2]

[0143] 接着,对动作例1的验证例2进行说明。在动作例1的验证例2中,说明第1模型132是输入作业音的数据而直接输出是否是处理透明物体的作业的推测结果的模型的例子。图6是用来说明由验证例2执行的袋作业的推测方法的图。图6所示的神经网络是第1模型132的一例。

[0144] 首先,对在袋作业的推测中使用的神经网络的学习进行说明。

[0145] 模型生成部140作为学习数据而使用作业音的频谱的图像或与作业音对应的(即,

在与收集到作业音的时刻相同的时刻拍摄的) 显现有作业员的图像数据。此外,模型生成部140作为教师数据,使用对学习数据赋予了作业员是否正在进行袋作业(换言之,袋作业的有无)的两个分类、或在存在袋作业的情况下包含该袋的种类(例如,大袋、小袋等)的3个分类的标签的数据。模型生成部140通过学习决定神经网络的参数。

[0146] 接着,推测部122使用在学习时决定的参数,进行通过神经网络的推理。例如,推测部122将想要进行作业分类的数据(作业音的数据或图像的数据)输入到神经网络中,输出袋作业的有无的两个分类或在存在袋作业的情况下对袋的种类进行分类的3个分类的推测结果。

[0147] 图7是表示图6所示的神经网络的架构的一例的图。在图7的例子中,神经网络由于输入数据是图像,所以具有卷积层,但例如在输入数据是时间序列的数值数据的情况下可以不具有卷积层。另外,图7的例子只不过是一例,并不限于此。

[0148] [关于两个分类的推测]

[0149] 首先,对袋作业的有无的两个分类的推测进行说明。图8是表示进行两个分类的推测的情况下的正确回答率的计算方法的图。神经网络使用带有存在袋作业及无袋作业的标签的数据作为教师数据而进行了学习。正确回答率(%)由图8所示的计算式计算。在图9中表示推测结果及正确回答率。

[0150] 图9是表示动作例1的验证例2的两个分类的推测结果及正确回答率的图。图9的(a)表示对神经网络输入的作业音的数据是可听频带的声音的数据的情况下的两个分类的推测结果及正确回答率,图9的(b)表示作业音的数据是包含非可听频带的声音的宽频带音的数据的情况下的两个分类的推测结果及正确回答率。袋作业1是处理纵及横分别为约10cm的聚乙烯制的袋的作业,袋作业2是处理纵及横分别为约30cm的聚乙烯制的袋的作业。如图9的(a)及图9的(b)所示,如果作为输入数据而使用包含非可听频带的声音的宽频带的声音的作业音的数据,则与使用可听频带的声音的作业音的数据相比正确回答率较高。因而,确认了作业音的数据是比可听频带的声音的数据频带宽的声音的数据时能够精度更好地推测作业员的作业。

[0151] [关于3个分类的推测]

[0152] 接着,对在存在袋作业的情况下对袋的种类进行分类的3个分类的推测进行说明。图10是表示进行3个分类的推测的情况下的正确回答率的计算方法的图。神经网络使用带有在存在袋作业的情况下表示袋的种类的标签和无袋作业的标签的数据作为教师数据进行了学习。正确回答率(%)由图10所示的计算式计算。在图11中表示推测结果及正确回答率。

[0153] 图11是表示动作例1的验证例2的3个分类的推测结果及正确回答率的图。图11的(a)表示对神经网络输入的作业音的数据是可听频带的声音的数据的情况下的3个分类的推测结果及正确回答率,图11的(b)表示作业音的数据是包含非可听频带的声音的宽频带音的数据的情况下的推测结果及正确回答率。如图11的(a)及图11的(b)所示,如果作为输入数据而使用包含非可听频带的声音的宽频带的声音的作业音的数据,则与使用可听频带的声音的作业音的数据相比正确回答率较高。因而,确认了作业音的数据是比可听频带的声音的数据频带宽的声音的数据时,能够精度更好地推测作业员的作业。

[0154] [关于基于输入数据的组合的两个分类的推测]

[0155] 接着,对基于输入数据的组合进行的袋作业的有无的两个分类的推测进行说明。图12是表示基于输入数据的组合进行的两个分类的推测方法及正确回答率的计算方法的图。图12的(a)表示推测结果的分类方法,图12的(b)表示推测结果与标签的对应关系。在图12的(a)中,分类A表示能够在输入数据是(i)图像数据及(ii)图像的数据+宽频带音的数据的至少任一个的情况下,按照存在袋作业的标签正确推测作业。此外,分类D表示能够在输入数据是上述(i)及(ii)的至少任一个的情况下,按照无袋作业的标签正确推测作业。正确回答率(%)由图12的(b)所示的计算式计算。在图13中表示推测结果及正确回答率。

[0156] 图13是表示动作例1的验证例2的基于输入数据的组合进行的两个分类的推测结果及正确回答率的图。图13的(a)表示对神经网络输入的输入数据是图像的数据的情况下的两个分类的推测结果及正确回答率,图13的(b)表示输入数据是图像的数据及宽频带的作业音的数据的情况下的推测结果及正确回答率。如图13的(a)及图13的(b)所示,如果作为输入数据而使用宽频带的作业音的数据,则与仅使用图像的数据的情况相比,正确回答率较高。因而,确认了与对神经网络输入的输入数据仅是图像的数据相比,是图像数据及宽频带的作业音的数据时能够精度更好地推测作业员的作业。

[0157] [动作例1的验证例3]

[0158] 接着,对动作例1的验证例3具体地进行说明。在验证例1中,在作业的推测中使用可听频带的作业音,而在验证例3中,在使用包含非可听频带的声音的作业音的数据这一点上与验证例1不同。进而,在验证例3中,比较了使用包含非可听频带的声音的作业音的数据实施动作例1所记载的推测方法的情况(称作本方法)下的推测精度与实施使用图像AI(换言之,影像AI)的推测方法的情况下的推测精度。另外,图像AI使用通常的图像AI。将结果表示在图14中。

[0159] 图14是表示使用图像AI的推测方法的推测精度与本方法的推测精度的比较结果的图。在图14中,标签的栏的“1”表示带有表示正在进行处理透明袋的作业(即袋作业)的标签(即正确答案标签)，“0”表示不带有正确答案标签(即,非袋作业)。此外,图像AI及本方法的栏的“1”表示推测为正在进行袋作业,“0”表示推测为并未进行袋作业。这里,确认了在标签的栏中记载的“0”及“1”是否与基于图像AI及本方法的推测结果一致。结果,图像AI的推测精度是0%,本方法的推测精度是72%。

[0160] 验证例3的结果确认了,通过使用包含非可听频带的声音的作业音的数据进行作业的推测,与可听频带的声音相比,处理透明物体的作业的推测精度提高。此外确认了,通过并用图像AI(即,图像识别模型)和声音识别模型,与仅使用图像AI推测作业的情况相比推测精度提高。

[0161] [动作例2]

[0162] 接着,一边参照图15、图16及图17一边对实施方式的推测系统200的动作例2具体地进行说明。图15是用来说明使用作业音的数据的推测结果与使用图像的数据的推测结果的差异的图。图16是用来说明实施方式的推测系统200的动作例2的流程的概要的图。图17是表示实施方式的推测系统200的动作例2的流程图。在动作例2中,以与动作例1不同的点为中心进行说明,关于共同的步骤将说明省略或简化。

[0163] 首先,对动作例2的由来进行说明。例如,如图15所示,用户在图像中以目视确认作业员的作业而判定存在袋作业的区间(袋作业区间),确认了基于目视的判定结果、基于作

业音的袋作业的推测结果(作业音相对于袋作业音的类似度)和基于图像的袋作业的推测结果的差异。关于袋作业的次数,在仅使用图像数据推测了作业的情况下也能够进行计数。但是,例如声音的类似度的得分比基于图像的推测先产生反应(上升)。此外,例如因为透明袋被文件等遮挡而未显现在图像中,所以即使使用图像的数据也不能推测出袋作业,但由于存在在透明袋时发出的声音(透明袋声音),所以声音的类似度的得分产生反应(上升)。

[0164] 这样,在基于图像的袋作业的推测中,在图像中未显现有透明袋的情况下,有时推测不出作业员正在进行袋作业。因此,通过将基于图像的袋作业的推测与基于作业音的袋作业的推测组合来推测袋作业,能够精度更好地推测袋作业。

[0165] 接着,说明动作例2的流程的概要。例如,如图16所示,在动作例2中,如果取得与作业音的数据对应的图像的数据,则推测装置100的取得部121向以图像的数据为输入的推测系统200输入图像的数据。例如,该推测系统200如图3所示,进行输入的图像数据的尺寸的调整或标准化等前处理,输入神经网络(例如图像识别模型)而基于输出的图像的特征量,计算与表示显现在图像中的透明袋的特征量的类似度。此外,在动作例2中,如果取得了作业音的数据,则推测装置100的取得部121向以作业音的数据为输入的推测系统200输入作业音的数据。例如,该系统如图3所示,进行输入的作业音的数据的标准化等前处理,输入神经网络(例如声音识别模型)而基于输出的作业音的特征量,计算与袋作业音的特征量的类似度。并且,通过这些由推测系统200得到的推测结果组合,输出推测结果。

[0166] 接着,一边参照图17一边对动作例2进行说明。虽然在图17中没有示出,但拾音装置10收集作业员的作业所附带的作业音,将收集到的作业音的数据向推测装置100输出。此外,摄像装置20拍摄与由拾音装置10收集到的作业音对应的(即,在相同的时刻拍摄的)显现有进行作业的作业员的图像,将所拍摄的图像的数据向推测装置100输出。另外,在图像中,当作业员正在进行处理透明物体的作业时,在图像中与作业员一起显现出透明物体(在这里是透明袋)。

[0167] 接着,如果取得了作业员的作业所附带的作业音的数据(S01),则推测装置100的取得部121将所取得的作业音的数据向推测部122输出。接着,推测部122通过将作业音的数据输入到第1模型132中,推测作业员是否正在进行处理透明物体的作业(S02)。具体而言,例如,推测部122在由第1模型132提取出的特征量与预先保存在存储部130中的处理透明物体的作业的作业音的特征量的类似度为规定值(即阈值)以上的情况下,推测为作业员正在进行处理透明物体的作业。

[0168] 此外,如果取得了与作业音的数据对应的、显现有进行作业的作业员的图像的数据(S03),则推测装置100的取得部121将所取得的图像的数据向推测部122输出。接着,推测部122通过将图像的数据输入到第2模型133中,推测作业员是否正在进行处理透明物体的作业(S04)。具体而言,例如,推测部122在由第2模型133提取出的显现有进行透明作业的作业员的图像的特征量与预先保存在存储部130中的显现有处理透明物体的作业员的图像的特征量的类似度为规定值(即阈值)以上的情况下,推测为作业员正在进行处理透明物体的工作。

[0169] 接着,推测部122基于使用第1模型132根据作业音的数据推测出的推测结果及使用第2模型133根据图像的数据推测出的推测结果,推测作业员是否正在进行处理透明物体

的作业(S05)。具体而言,例如推测部122在由第1模型132提取出的作业音的特征量与预先保存在存储部130中的处理透明物体的作业的作业音的特征量的类似度为规定值(阈值)以上,并且由第2模型133提取出的图像的特征量与预先保存在存储部130中的显现有处理透明物体的作业员的图像的特征量的类似度为规定值(阈值)以上的情况下,推测为作业员正在进行处理透明物体的作业。

[0170] [动作例2的变形例1]

[0171] 在动作例2中,说明了基于对第1模型132输入作业音的数据而得到的特征量和对第2模型133输入图像的数据而得到的特征量来推测作业员是否正在进行处理透明物体的作业的例子。在动作例2的变形例1中,通过在动作例1的验证例2中所说明的直接推测是否是处理透明物体的作业的第1模型132的例子,基于对第1模型132输入作业音的数据及图像的数据而得到的作业音的特征量及图像的特征量,推测作业员是否正在进行处理透明物体的作业。

[0172] 图18是表示实施方式的推测系统200的动作例2的变形例1的流程图。如图18所示,推测装置100的取得部121取得由拾音装置10收集到的作业音的数据(S01),将所取得的数据向推测部122输出。此外,推测装置100的取得部121取得由摄像装置20拍摄的与作业音的数据对应的、显现有进行作业的作业员的图像的数据(S03),将所取得的数据向推测部122输出。

[0173] 接着,推测部122基于将作业音的数据和图像的数据输入第1模型132而得到的作业音的特征量及图像的特征量,推测作业员是否正在进行处理透明物体的作业(S06)。

[0174] [执行动作例2的变形例1的流的推测部122的结构例1]

[0175] 接着,对执行动作例2的变形例1的流的推测部122的结构例1进行说明。图19是示意地表示执行动作例2的变形例1的流的推测部122的结构例1的图。图20是用来说明由结构例1执行的袋作业的推测方法的图。

[0176] 如图19所示,推测部122具有嵌入向量制作部、作业分类部和袋作业的确定部。嵌入向量制作部具有以图像的数据为输入而提取图像的特征量的图像子网络、以声音(在这里是作业音)的数据为输入而提取声音特征量(在这里是作业音的特征量)的声音子网络和融合层。

[0177] 如图19所示,神经网络例如可以具有图像子网络和声音子网络。这样的神经网络可以是第1模型132。此外,声音子网络可以是第1模型132,图像子网络可以是第2模型133。

[0178] 如图20所示,模型生成部140作为学习数据而使用图像的数据和作业音的数据,作为教师数据而使用赋予了学习数据是否类似的标签的数据。模型生成部140通过学习来决定神经网络的参数。作业音的数据是包含可听频带的声音或非可听频带的声音的宽频带的声音的数据。作业音的数据例如可以是 257×199 像素的频谱。图像的数据例如可以是 224×224 像素的数据。另外,模型生成部140可以对融合层进行迁移学习。

[0179] 接着,推测部122通过使用了在学习时决定的参数的融合层制作嵌入向量。接着,推测部122将嵌入向量输入到作业分类部中,基于从Softmax层输出的概率值来确定袋作业。

[0180] [执行动作例2的变形例1的流的推测部122的结构例2]

[0181] 接着,对执行动作例2的变形例1的流的推测部122的结构例2进行说明。图21是

示意地表示执行动作例2的变形例1的流程的推测部122的结构例2的图。在结构例1中,作业分类部具有分类网络和Softmax层,而在结构例2中,作业分类部具有对比学习网络。对比学习是不进行赋予标签,利用将数据彼此比较的机制而能够直接学习庞大的数据的自监督学习之一。在对比学习中,以类似的数据接近、不同的数据远离的方式学习特征量。

[0182] [执行动作例2的变形例1的流程的推测部122的结构例3]

[0183] 接着,对执行动作例2的变形例1的流程的推测部122的结构例3进行说明。图22是示意地表示执行动作例2的变形例1的流程的推测部122的结构例3的图。在结构例1及结构例2中,融合层配置在分类网络之前,而在结构例3中,融合层配置在分类网络之后。

[0184] [关于图像子网络及声音子网络的架构的例子]

[0185] 接着,对图像子网络及声音子网络的架构的例子进行说明。图23是表示图像子网络的架构的一例的图。图24是表示声音子网络的架构的一例的图。如图23及图24所示,由于作为输入数据的图像的数据及作业音的数据的尺寸不同的情况多,所以图像子网络及声音子网络的各层的尺寸也可以不相同,但只要这些子网络的最终层的尺寸相同即可。

[0186] [关于融合层的架构的例子]

[0187] 图25是表示融合层的架构的一例的图。如图25所示,从图像子网络输出的数据和从声音子网络输出的数据输入到结合层中,在学习时和推理时得到不同的输出。

[0188] [关于分类网络的架构的例子]

[0189] 图26是表示分类网络的架构的一例的图。如图26所示,分类网络的第一层的尺寸例如在配置在图像子网络之后的情况下是与从图像子网络的最终层输出的数据相同的尺寸,在配置在声音子网络之后的情况下是与从声音子网络的最终层输出的数据相同的尺寸。此外,分类网络的第一层的尺寸例如在分类网络配置在融合层之后的情况下是与从融合层的最终层输出的数据相同的尺寸。

[0190] [关于对比学习网络的架构的例子]

[0191] 图27是表示对比学习网络的架构的一例的图。如图27所示,对比学习网络的第一层的尺寸例如与从嵌入向量制作部输出的嵌入向量的尺寸相同。对比学习网络用作迁移学习。但是,作为损失函数而例如使用下述式1。

[0192] [式1]

$$[0193] \quad l_{i,j} = -\log \frac{\exp(\text{sim}(z_i, z_k)/\tau)}{\sum_{k=1}^{2N} \mathbf{1}_{k \neq i} \exp(\text{sim}(z_i, z_k)/\tau)} \quad (\text{式1})$$

[0194] 这里, $\text{sim}(x, y)$ 是计算类似度的函数,例如可以使用余弦类似度。 z_i, z_j 是对应的嵌入向量,例如可以分别使用图像的数据及宽频带的作业音的数据的嵌入向量。 τ 是调整参数。

[0195] 上述式1的损失函数在两个嵌入向量的类似度大时变大,在类似度小时变小。

[0196] [执行动作例2的变形例2的流程的推测部122的结构例]

[0197] 图28是示意地表示执行动作例2的变形例2的流程的推测部122的结构例的图。对于动作例2的变形例2的流程,一边参照图18所记载的动作例2的变形例1的流程一边进行说明。在动作例2的变形例1中,取得图像的数据而作为输入数据使用,而在动作例2的变形例2中,在图18的步骤S03中代替由摄像装置20拍摄的图像的数据,可以使用由测距传感器等取

得的距离数据作为输入数据。在此情况下,如图28所示,推测部122代替图像子网络而具有距离子网络。另外,其他的结构例并不限于图28的例子,例如可以如动作例2的变形例1的结构例2或结构例3那样将融合层的配置位置变更,作业分类部可以具有对比学习网络。

[0198] [动作例3]

[0199] 接着,一边参照图29、图30A及图30B一边对实施方式的推测系统200的动作例3进行说明。图29是表示由推测部122误推测为作业员正在进行处理透明物体的作业时的作业音的例子。图30A是实施方式的推测系统200的动作例3的流程图。图30B是表示将可能被误推测的作业音的特征量预先登记的动作的一例的流程图。

[0200] 在动作例3中,将处理透明袋的作业所附带的作业音称作透明袋音,将处理非透明袋(换言之,不处理透明袋)的作业所附带的作业音称作非透明袋音。此外,将处理透明袋的作业称作袋作业。

[0201] 首先,一边参照图29一边对可能被误推测的作业音(以下,也称作误推测对象音)进行说明。在图29的例子中,类似度的阈值例如是25,推测部122在作业音相对于透明袋音的类似度为阈值以上的情况下,推测为作业员正在进行处理透明袋的作业(袋作业)。此时,存在基于将塑料袋打开的声音、从搁架将袋取出的声音等处理透明袋的作业所附带的作业音而正确地推测作业员的作业的情况,但也存在尽管并未进行袋作业但被误推测为正在进行袋作业的情况。例如,基于通过橡皮筋捆绑盒子的声音、将盒子或袋收纳到推车的下段的声音或移动中的条码扫描的声音等那样的透明袋音以外的作业音(即非透明袋音),误推测为作业员正在进行袋作业。

[0202] 为了减少这样的误推测,推测部122计算非透明袋音的特征量与预先登记的透明袋音的特征量的类似度,在该类似度高于阈值的情况下,将非透明袋音判定为误推测对象音,保存到存储部130中。在动作例3中,推测部122将预先登记的可能被误推测的作业音(以下,也称作误推测对象音)的特征量和透明袋音的特征量从存储部130读入,将作业音的特征量与这些特征量的类似度比较,推测作业员是否正在进行袋作业。

[0203] 接着,一边参照图30A一边对动作例3进行说明。虽然没有图示,但推测装置100的取得部121取得由拾音装置10收集到的作业音的数据,将所取得的数据向推测部122输出。

[0204] 推测部122将所取得的作业音的数据输入到声音识别模型中(S11),根据输入的作业音的数据检测声音,提取输入特征量(S12)。

[0205] 接着,推测部122使用声音识别模型提取作业音(以下称作输入音)的特征量(声音特征量)(S13)。接着,推测部122从存储部130将透明袋音的特征量和误推测对象音的特征量读入(S14)。

[0206] 接着,推测部122在类似度计算(S15)中,计算透明袋音与输入音的类似度以及误推测对象音与输入音的类似度。

[0207] 接着,推测部122判定透明袋音与输入音的类似度是否高于误推测对象音与输入音的类似度(S16),在判定为高于的情况下(S16中为“是”),判定透明袋音与输入音的类似度是否高于阈值(S17)。推测部122在判定为透明袋音与输入音的类似度高于阈值的情况下(S17中为“是”),判定为输入音是透明袋音(S18)。由此,推测部122基于输入音(作业音)的特征量,推测为作业员正在进行处理透明袋的作业。

[0208] 另一方面,推测部122在步骤S16中判定为透明袋音与输入音的类似度不高于误推

测对象音与输入音的类似度的情况下(S16中为“否”),判定为输入音不是透明袋音(S19)。此外,推测部122在步骤S17中判定为透明袋音与输入音的类似度不高于阈值的情况下(S17中为“否”),判定为不是透明袋音(S19)。由此,推测部122基于输入音(作业音)的特征量,推测为作业员正在进行不处理透明袋的作业。

[0209] 接着,一边参照图30B一边对将在动作例3中使用的误推测对象音的特征量预先保存在存储部130中的动作例进行说明。虽然没有图示,但推测装置100的取得部121取得由拾音装置10取得的作业音的数据,将所取得的数据向推测部122输出。此时,由取得部121取得的作业音的数据是不处理透明袋的作业所附带的作业音。

[0210] 接着,推测部122将所取得的作业音的数据输入到声音识别模型中(S21),根据输入的作业音的数据检测声音,提取输入特征量(S22)。

[0211] 接着,推测部122使用声音识别模型提取作业音(以下,称作输入音)的特征量(声音特征量)(S23)。接着,推测部122从存储部130将透明袋音的特征量读入(S24)。

[0212] 接着,推测部122在类似度计算(S25)中计算透明袋音与输入音的类似度。

[0213] 接着,推测部122判定透明袋音与输入音的类似度是否高于阈值(S26),在判定为高于该阈值的情况下(S26中为“是”),将该输入音判定为误推测对象音(S27)。并且,推测部122将收集到的声音(作业音)的特征量作为误推测对象音的特征量保存到存储部130中(S29)。另一方面,推测部122在判定为透明袋音与输入音的类似度不高于阈值的情况下(S26中为“否”),判定为该输入音不是误推测对象音(S28)。

[0214] [5.效果等]

[0215] 如以上的说明,有关本实施方式的推测方法是由计算机(例如,推测装置100)进行的推测作业员的作业的推测方法,计算机取得收集到的作业所附带的作业音的数据(图2的S01),通过对已学习完毕的第1模型132输入作业音的数据,推测作业员是否正在进行处理透明物体的作业(图2的S02)。

[0216] 由此,执行推测方法的装置(例如,推测装置100)由于使用以作业音的数据为输入而输出是否是处理透明物体的作业的第1模型132,所以能够精度良好地推测处理透明物体的作业。

[0217] 例如,在有关本实施方式的推测方法中,计算机(例如,推测装置100)取得与作业音的数据对应的、显现有进行作业的作业员的图像的数据(图17的S03),对已学习完毕的第2模型133输入图像的数据,从而推测作业员是否正在进行处理透明物体的作业(图17的S04),基于使用第1模型132的推测结果及使用第2模型133的推测结果,推测作业员是否正在进行处理透明物体的作业(图17的S05)。另外,使用第1模型132的推测结果是由第1模型132根据作业音的数据推测出的推测结果,使用第2模型133的推测结果是由第2模型133根据图像的数据推测出的推测结果。

[0218] 由此,执行推测方法的装置(例如,推测装置100)基于由第1模型132根据作业音的数据推测出的推测结果和由第2模型133根据图像的数据推测出的推测结果,推测作业员是否正在进行处理透明物体的作业。因此,执行推测方法的装置与仅使用作业音的数据进行推测的情况相比能够精度更好地推测处理透明物体的作业。

[0219] 例如,在有关本实施方式的推测方法中,计算机(例如,推测装置100)取得与作业音的数据对应的、显现有进行作业的作业员的图像的数据(图18的S03),通过将作业音的数

据和图像的数据输入到第1模型132中,推测作业员是否正在进行处理透明物体的作业(S06)。

[0220] 由此,执行推测方法的装置(例如,推测装置100)使用以作业音的数据及对应于作业音的图像的数据为输入、输出是否是处理透明物体的作业的第1模型132,因此与仅使用作业音的数据进行推测的情况相比能够精度更好地推测处理透明物体的作业。

[0221] 例如,在有关本实施方式的推测方法中,计算机(例如,推测装置100)基于从第1模型132输出的作业音的特征量与预先保存在存储部130(例如,图1的特征量数据库131)中的处理透明物体的作业的作业音的特征量的类似度,推测作业员是否正在进行处理透明物体的作业。

[0222] 由此,执行推测方法的装置(例如,推测装置100)基于从第1模型132输出的作业音的特征量与处理透明物体的作业的作业音的特征量的类似度,推测作业员是否正在进行处理透明物体的作业,所以能够精度良好地推测处理透明物体的作业。

[0223] 例如,在有关本实施方式的推测方法中,计算机(例如,推测装置100)基于从第1模型132输出的作业音的特征量相对于预先保存在存储部130(例如,特征量数据库131)中的处理透明物体的作业的作业音的特征量的类似度(换言之,第1类似度),以及相对于预先保存在存储部130(例如,特征量数据库131)中的可能被误推测为作业员正在进行处理透明物体的作业的作业音(例如,图30A的误推测对象音)的特征量的类似度(换言之,第2类似度),推测作业员是否正在进行处理透明物体的作业(图30A的S16~S19)。

[0224] 由此,执行推测方法的装置(例如,推测装置100)通过比较从第1模型132输出的作业音的特征量与处理透明物体的作业的作业音的特征量的类似度(第1类似度),和从第1模型132输出的作业音的特征量与可能被误推测的作业音的特征量的类似度(第2类似度),从而能够减少误推测的发生。因而,执行推测方法的装置即使仅使用作业音的数据也能够精度良好地推测处理透明物体的作业。

[0225] 例如,在有关本实施方式的推测方法中,计算机(例如,推测装置100)在从第1模型132输出的作业音的特征量相对于处理透明物体的作业的作业音的特征量的上述的类似度(上述的第1类似度)高于相对于可能被误推测为作业员正在进行处理透明物体的作业的作业音(图30A的误推测对象音)的特征量的上述的类似度(上述的第2类似度)的情况下(图30A的S16中为“是”),推测为作业员正在进行处理透明物体的作业。

[0226] 由此,执行推测方法的装置(例如,推测装置100)由于能够减少误推测的发生,所以即使仅使用作业音的数据,也能够精度良好地推测处理透明物体的作业。

[0227] 例如,在有关本实施方式的推测方法中,计算机(例如,推测装置100)在将处理与透明物体不同的非透明物体的作业的作业音的数据输入到第1模型132中而得到的处理非透明物体的作业的作业音的特征量与处理透明物体的作业的作业音的特征量的类似度(换言之,第3类似度)高于阈值的情况下(图30B的S26中为“是”),将该处理非透明物体的作业的作业音判定为可能被误推测为处理透明物体的作业的作业音的作业音(即误推测对象音)(图30B的S27),将处理非透明物体的作业的作业音的特征量作为可能被误推测的作业音的特征量保存到存储部130(例如,特征量数据库131)中(图30B的S29)。

[0228] 由此,执行推测方法的装置(例如,推测装置100)基于处理非透明物体的作业的作业音的特征量与处理透明物体的作业的作业音的特征量的类似度(第3类似度),能够精度

良好地判定该处理非透明物体的作业的作业音是否是可能被误推测为正在进行处理透明物体的作业的作业音。因此,执行推测方法的装置能够将被误推测的可能性比较高的作业音的特征量保存到存储部130中。因而,执行推测方法的装置通过使用保存在存储部130中的可能被误推测的作业音的特征量,能够减少误推测的发生,所以即使仅使用作业音的数据,也能够精度良好地推测处理透明物体的作业。

[0229] 在有关本实施方式的推测方法中,作业音的数据也可以包含非可听频带的声音的数据。

[0230] 由此,执行推测方法的装置(例如,推测装置100)使用包括从可听频带的声音到非可听频带的声音的作业音的数据推测作业员是否正在进行处理透明物体的作业。这样,由于作业音的数据包含非可听频带的声音,作业音的数据中的成为误推测的原因的来自环境的杂音变少,所以执行推测方法的装置能够提高处理透明物体的作业的推测精度。进而,执行推测方法的装置能够基于比仅使用可听频带的声音的数据的情况多的信息推测作业员是否正在进行处理透明物体的作业。因而,执行推测方法的装置能够精度更好地推测处理透明物体的作业。

[0231] 此外,有关本实施方式的推测装置100是推测作业员的作业的推测装置,具有:取得部121,取得收集到的作业所附带的作业音的数据;以及推测部122,通过对已学习完毕的第1模型132输入作业音的数据,推测作业员是否正在进行处理透明物体的作业。

[0232] 由此,推测装置100由于使用以作业音的数据为输入、输出是否是处理透明物体的作业的第1模型132,所以能够精度良好地推测处理透明物体的作业。

[0233] 此外,有关本实施方式的程序是用来使计算机执行上述的推测方法的程序。

[0234] 由此,能够使用计算机起到与上述的推测方法同样的效果。

[0235] (其他实施方式)

[0236] 以上,对实施方式进行了说明,但本公开并不限定于上述的实施方式。

[0237] 图31是表示有关其他实施方式的推测系统的功能结构的一例的框图。有关实施方式的推测系统200说明了推测装置100是服务器装置的例子,但推测装置100也可以不是服务器装置。例如,在有关其他实施方式的推测系统200a,推测装置100a也可以是个人计算机等的固定安置型的计算机装置。推测装置100a在具有显示部160这一点上与推测装置100不同。以下,仅说明不同的点。

[0238] [显示部160]

[0239] 显示部160例如显示推测结果。显示部160例如是显示包含字符等的图像信息的显示装置,例如是包括液晶(LC)面板或有机EL(Electro Luminescence)面板等作为显示设备的显示器。

[0240] 另外,推测装置100a例如可以具有拾音部和摄像部,可以在作业空间80设置1个以上。具有拾音部和摄像部既可以是与拾音装置10及摄像装置20以有线或无线通信连接的形态,也可以是具有拾音装置10及摄像装置20的单一的装置。并且,推测装置100a例如也可以与服务器装置或用户的信息终端通信连接。在此情况下,推测装置100a可以将推测结果在存储部130中保存规定期间(例如,1天、几天、一周等),将推测结果对服务器装置或信息终端输出,也可以每次推测就输出推测结果。服务器装置可以是云服务器。此外,信息终端可以是个人计算机等固定安置型的计算机装置,也可以是平板电脑终端等便携型的计算机装

置。

[0241] 此外,例如在上述的实施方式中,推测系统200、200a分别由多个装置实现,但也可以作为单一的装置实现。此外,在系统由多个装置实现的情况下,推测系统200、200a分别具有的多个构成要素怎样分配给多个装置都可以。此外,例如也可以是能够与推测系统200或200a通信的服务器装置具有信息处理部120包含的多个构成要素。

[0242] 例如,关于上述实施方式的装置间的通信方法不作特别限定。此外,在装置间的通信中也可以经由未图示的中继装置。

[0243] 此外,在上述实施方式中,可以将特定的处理部执行的处理由其他的处理部执行。此外,可以将多个处理的顺序变更,也可以并行执行多个处理。

[0244] 此外,在上述实施方式中,各构成要素也可以通过执行适合于各构成要素的软件程序实现。各构成要素也可以通过CPU或处理器等程序执行部将记录在硬盘或半导体存储器等的记录介质中的软件程序读出并执行来实现。

[0245] 此外,各构成要素也可以由硬件实现。例如,各构成要素也可以是电路(或集成电路)。这些电路既可以作为整体构成1个电路,也可以分别是不同的电路。此外,这些电路分别既可以是通用的电路,也可以是专用的电路。

[0246] 此外,本公开的全部或具体的技术方案也可以由系统、装置、方法、集成电路、计算机程序或计算机可读的CD-ROM等的记录介质实现。此外,也可以由系统、装置、方法、集成电路、计算机程序及记录介质的任意的组合实现。

[0247] 例如,本公开也可以作为推测装置100等的计算机执行的推测方法实现,也可以作为使计算机执行这样的推测方法的程序实现。此外,本公开也可以作为用来使通用的计算机作为上述实施方式的推测装置100动作的程序实现。本公开也可以作为记录有这些程序的计算机可读的非易失性的记录介质实现。

[0248] 除此以外,对各实施方式施以本领域技术人员想到的各种变形得到的形态或在不脱离本公开的主旨的范围中将各实施方式的构成要素及功能任意地组合而实现的形态也包含在本公开中。

[0249] 工业实用性

[0250] 根据本公开,由于能够精度良好地推测处理透明物体的作业,所以能够正确地掌握作业时间等,能够实现工厂或物流等现场的作业的高效化。

[0251] 标号说明

[0252] 10 拾音装置

[0253] 20 摄像装置

[0254] 50 信息终端

[0255] 80 作业空间

[0256] 100、100a推测装置

[0257] 110 通信部

[0258] 120 信息处理部

[0259] 121 取得部

[0260] 122 推测部

[0261] 130 存储部

- [0262] 131 特征量数据库
- [0263] 132第1模型
- [0264] 133第2模型
- [0265] 140 模型生成部
- [0266] 150 输入受理部
- [0267] 160 显示部
- [0268] 200、200a推测系统

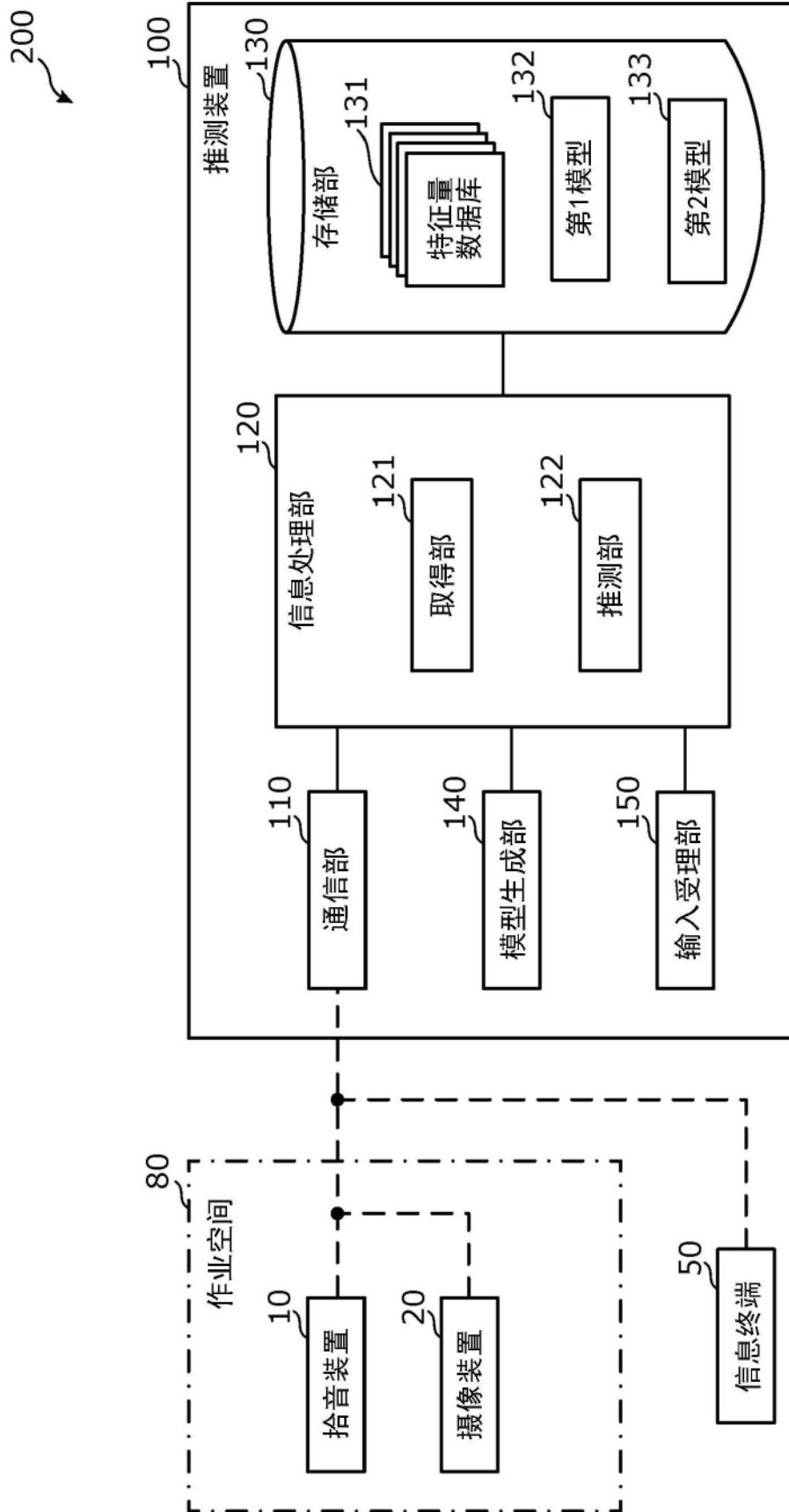


图1

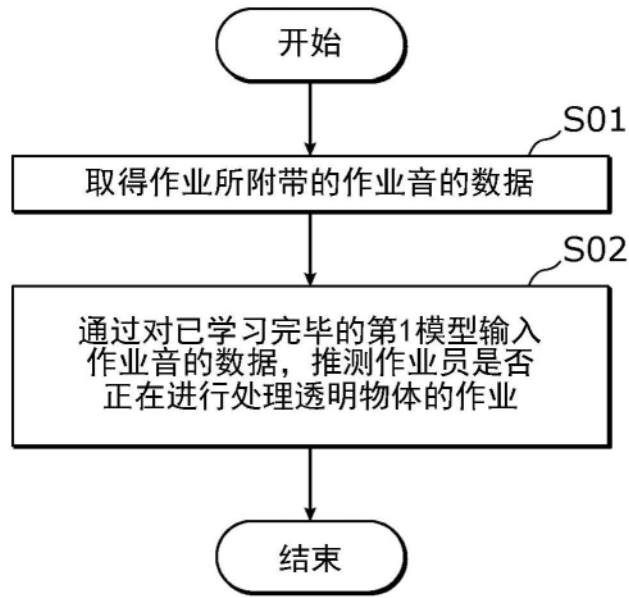


图2

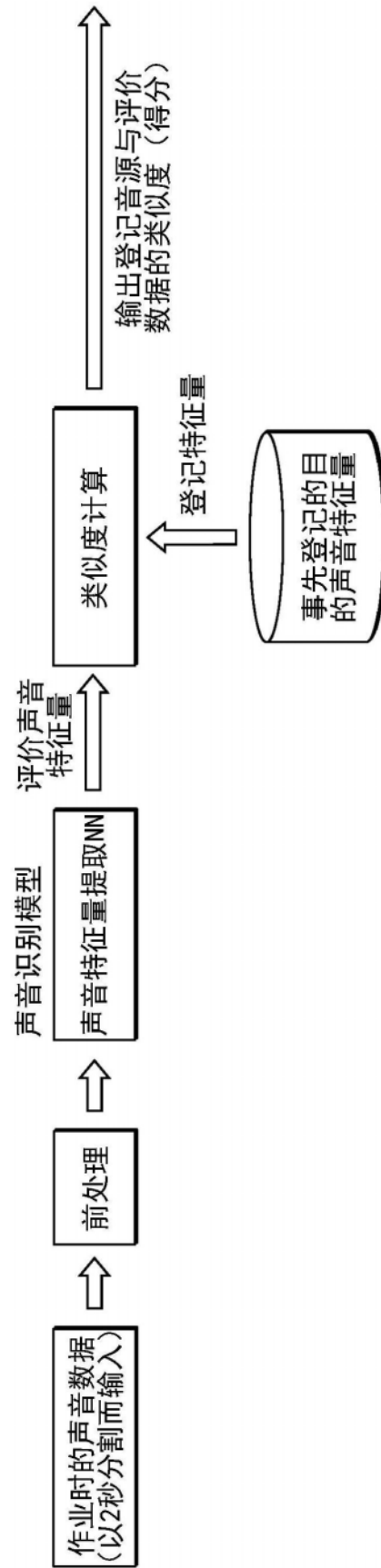


图3

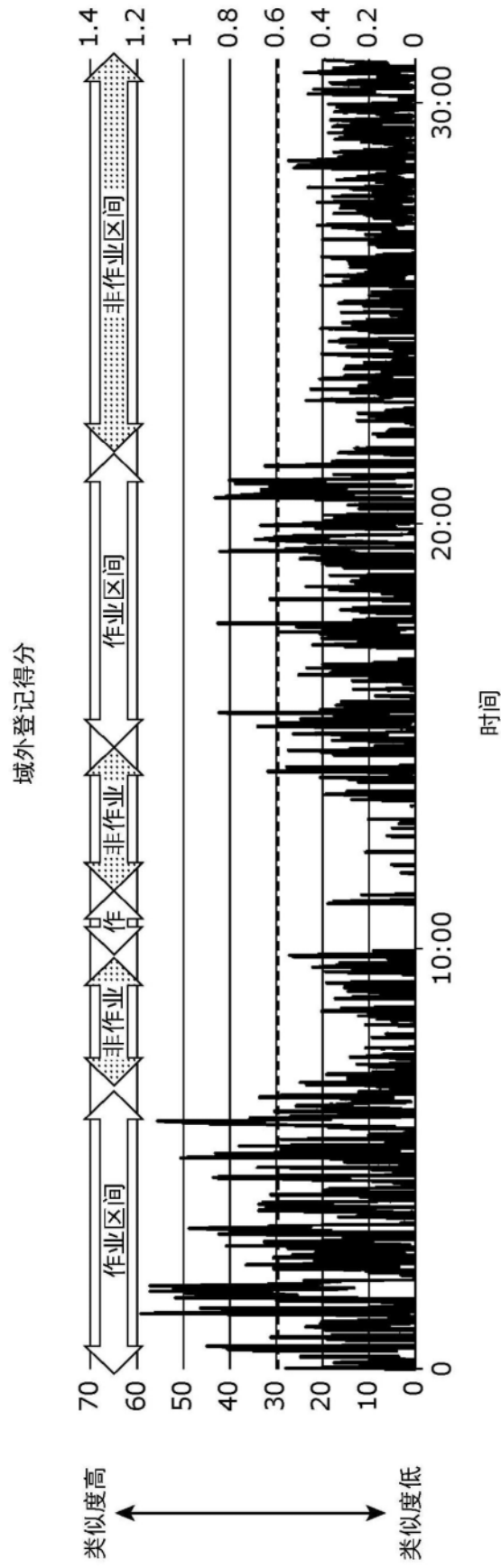


图4

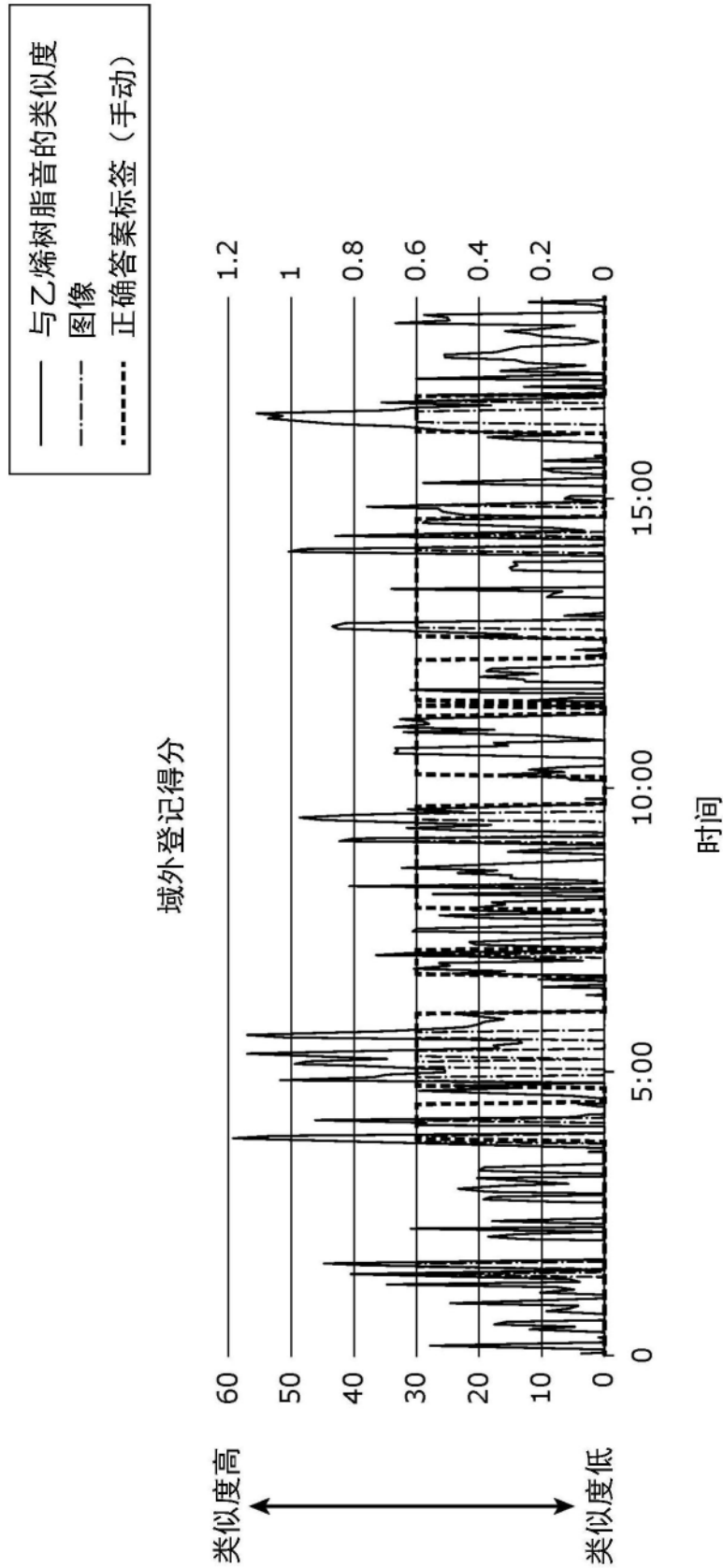


图5

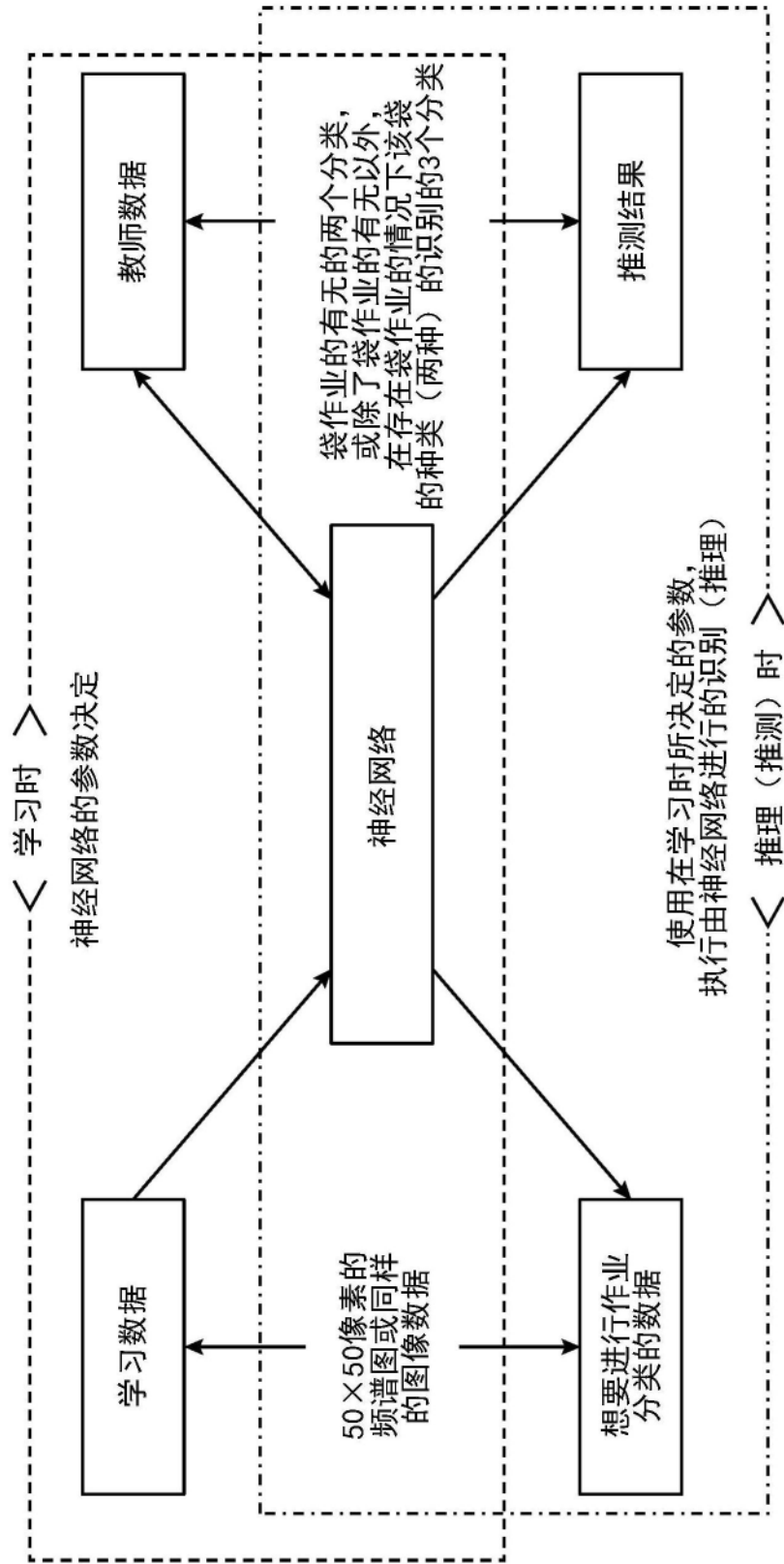


图6

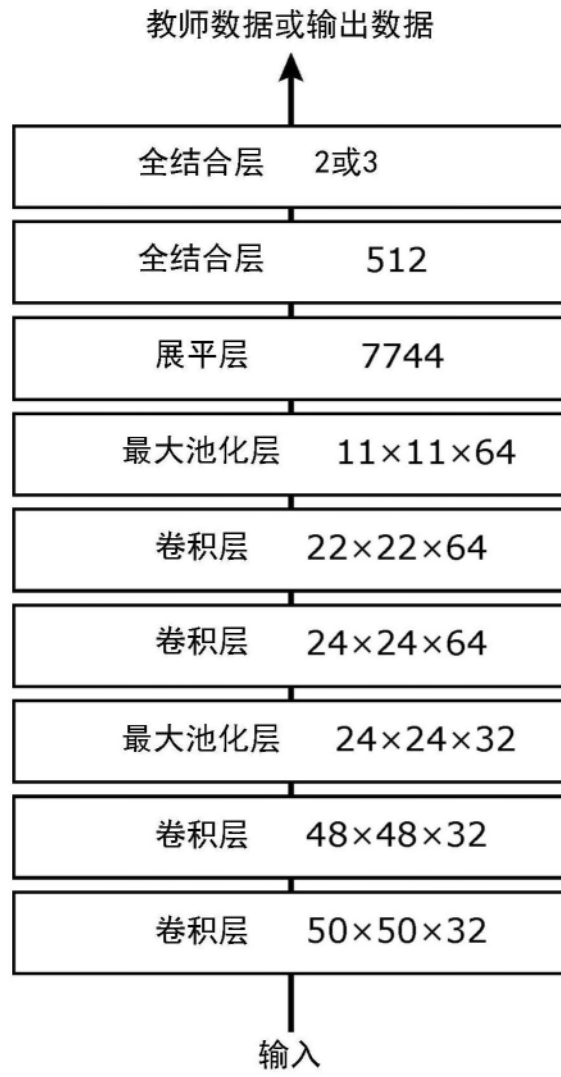


图7

		推测结果		
		存在袋作业	无袋作业	合计
标签	存在袋作业	A	B	A+B
	无袋作业	C	D	C+D
	合计	A+C	B+D	A+B+C+D

$$\text{正确回答率(\%)} = \frac{A+D}{A+B+C+D} \times 100$$

图8

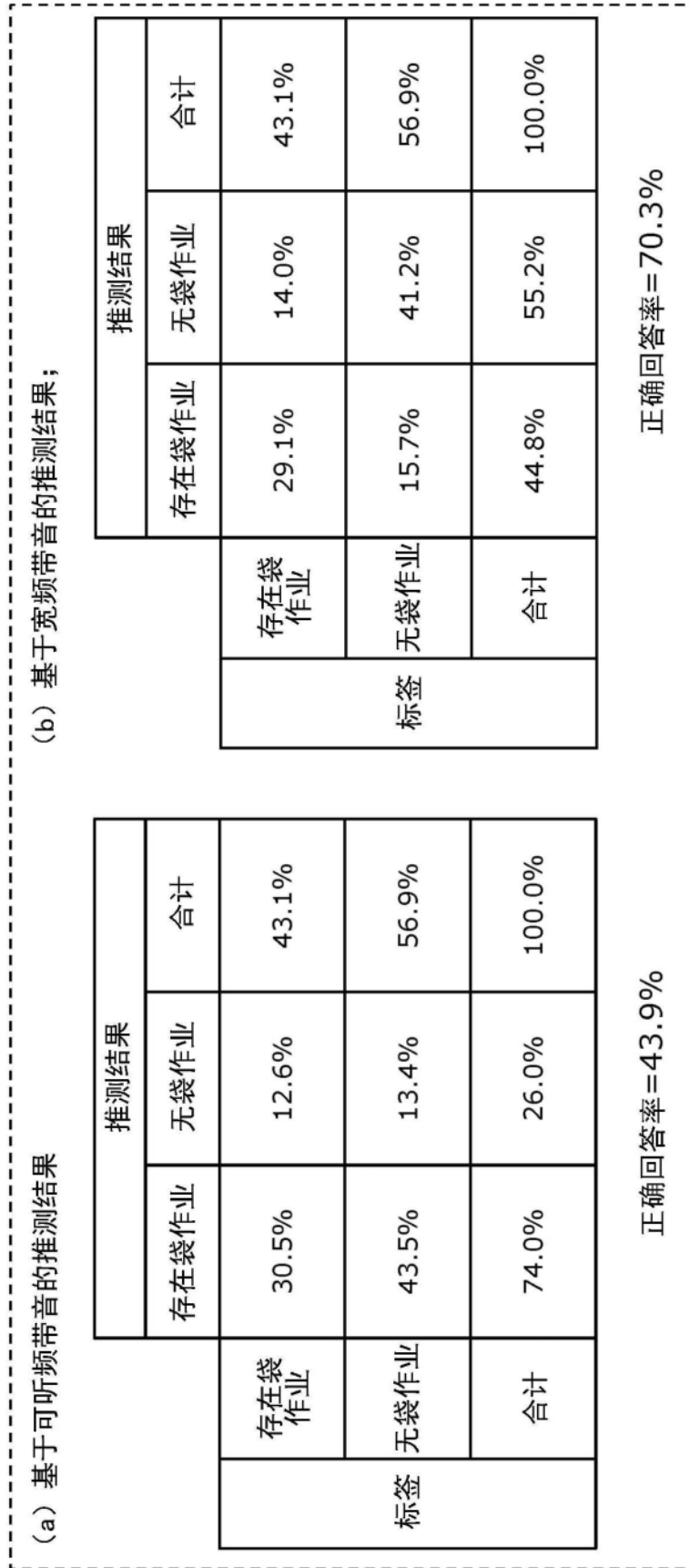


图9

		推测结果			
		袋作业1	袋作业2	无袋作业	合计
标签	袋作业1	A	B	C	A+B+C
	袋作业2	D	E	F	D+E+F
	无袋作业	G	H	I	G+H+I
	合计	A+D+G	B+E+H	C+F+I	A+B+C+ D+E+F+ G+H+I

$$\text{正确回答率(\%)} = \frac{A+E+I}{A+B+C+D+E+F+G+H+I} \times 100$$

图10

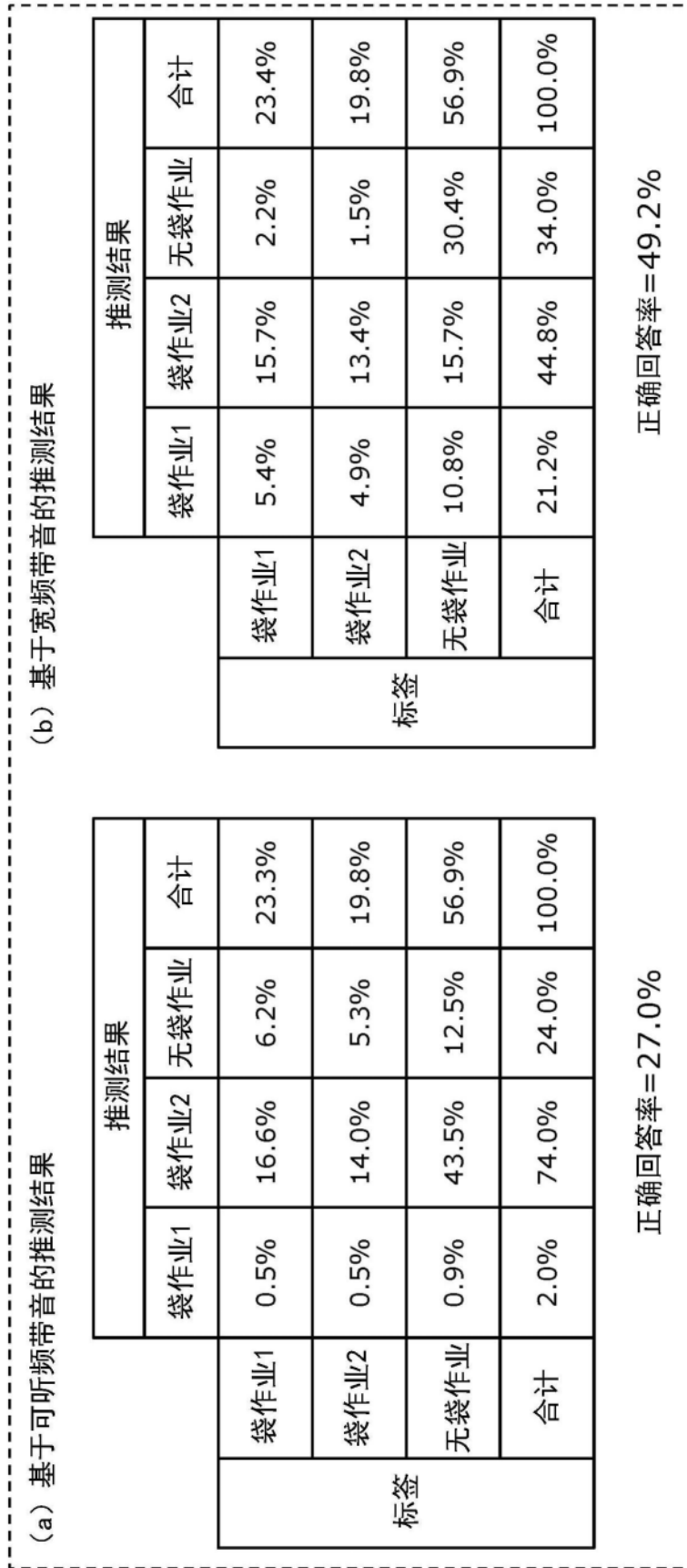


图11

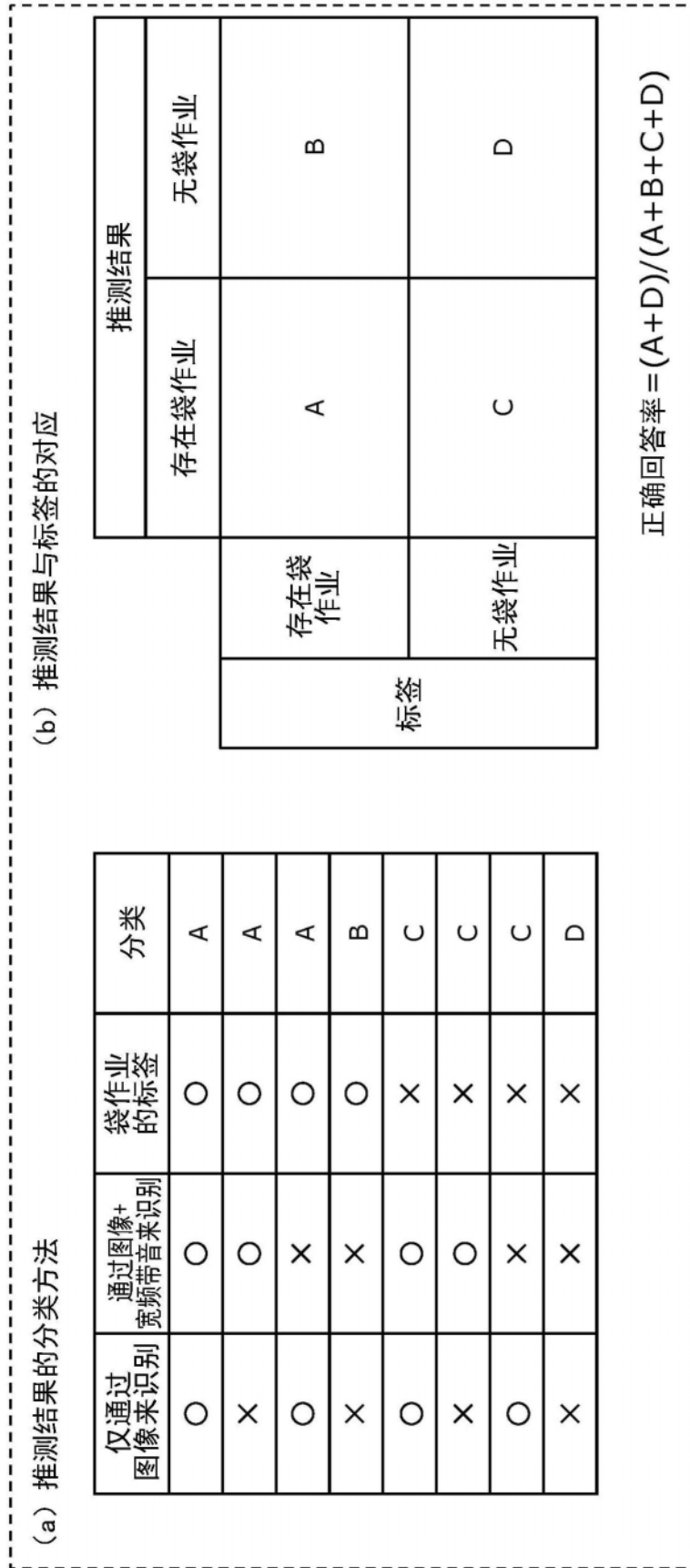


图12

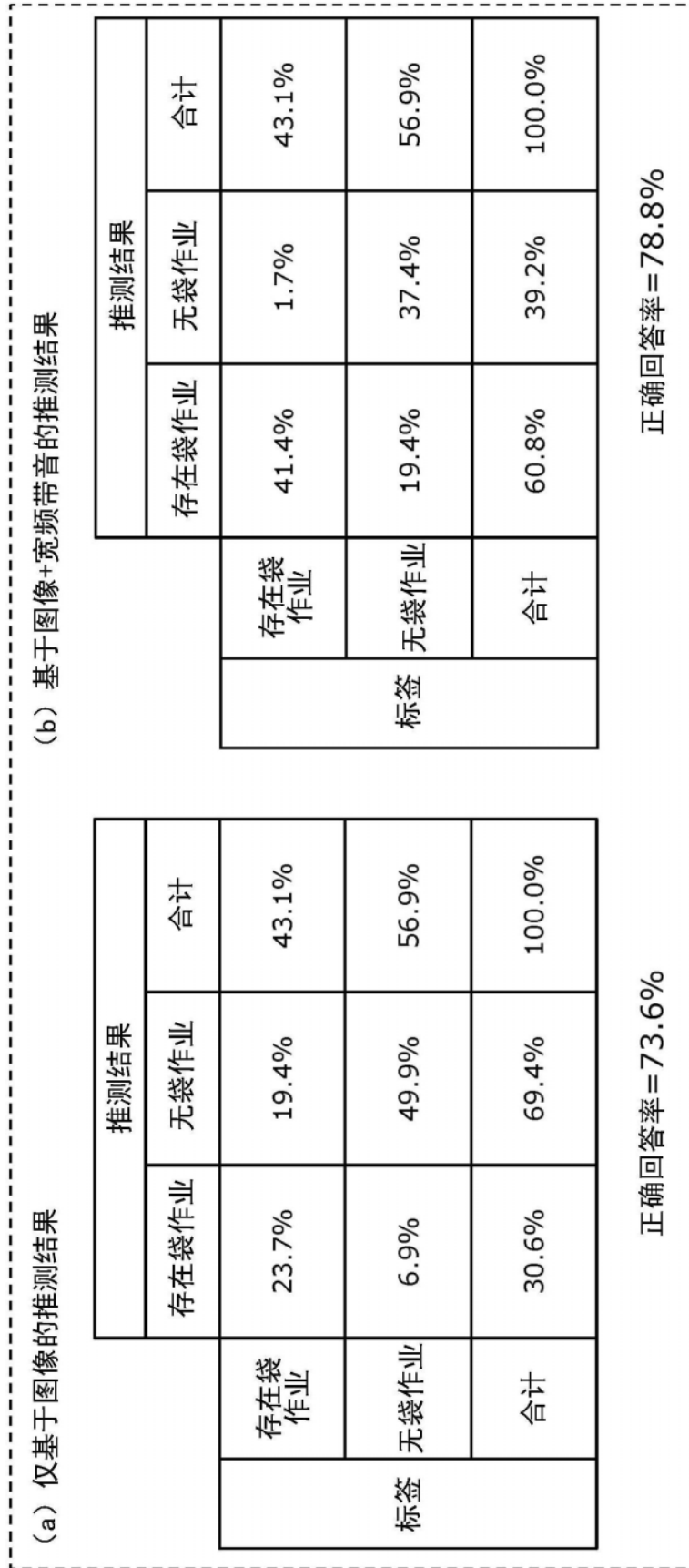


图13

场景	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	精度	
标签	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	1	1	1	
图像AI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0%
本方法	0	0	0	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	72%

图14

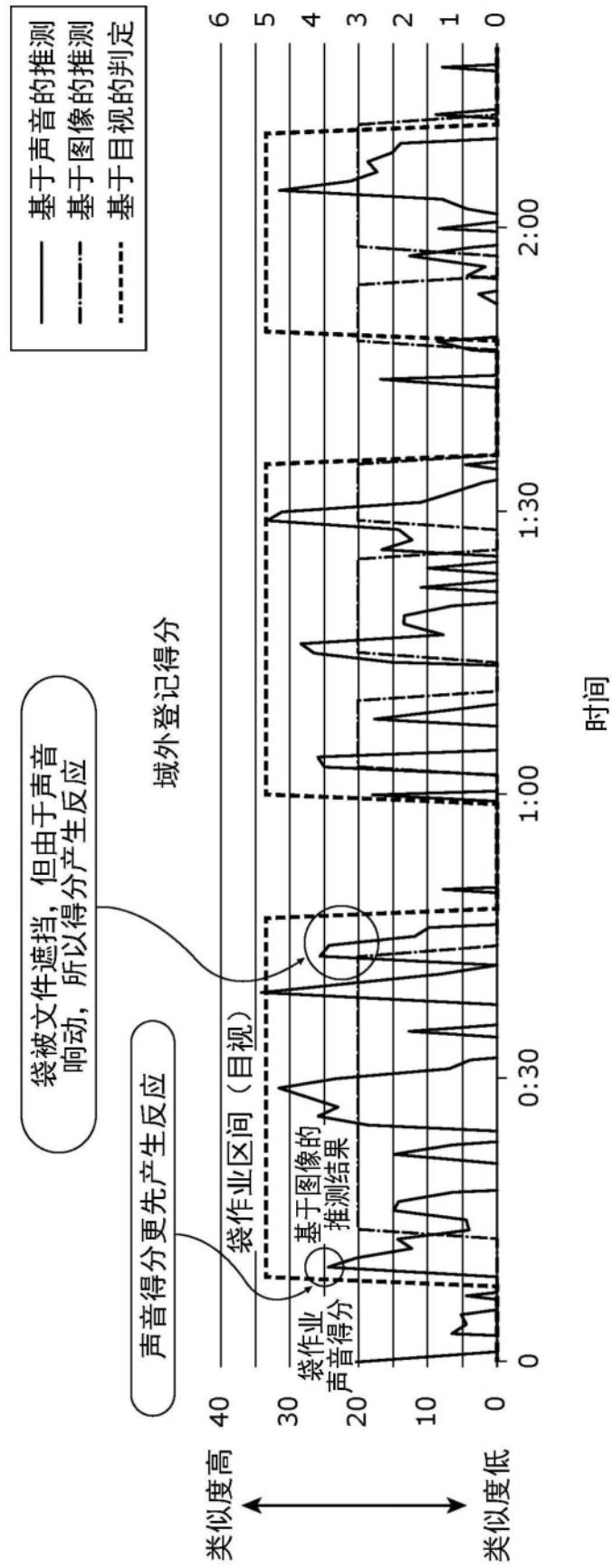


图15

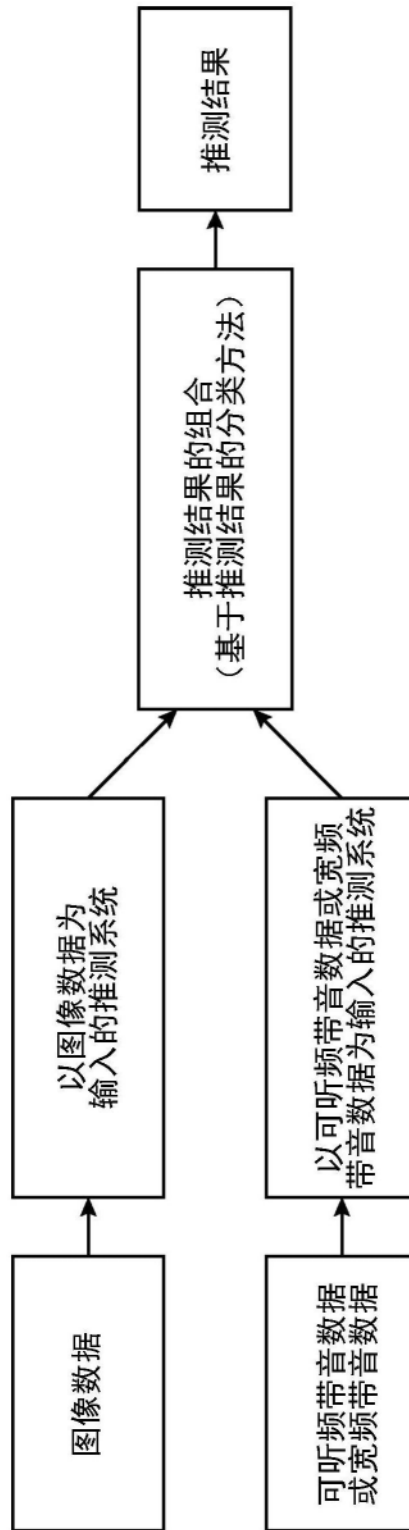


图16

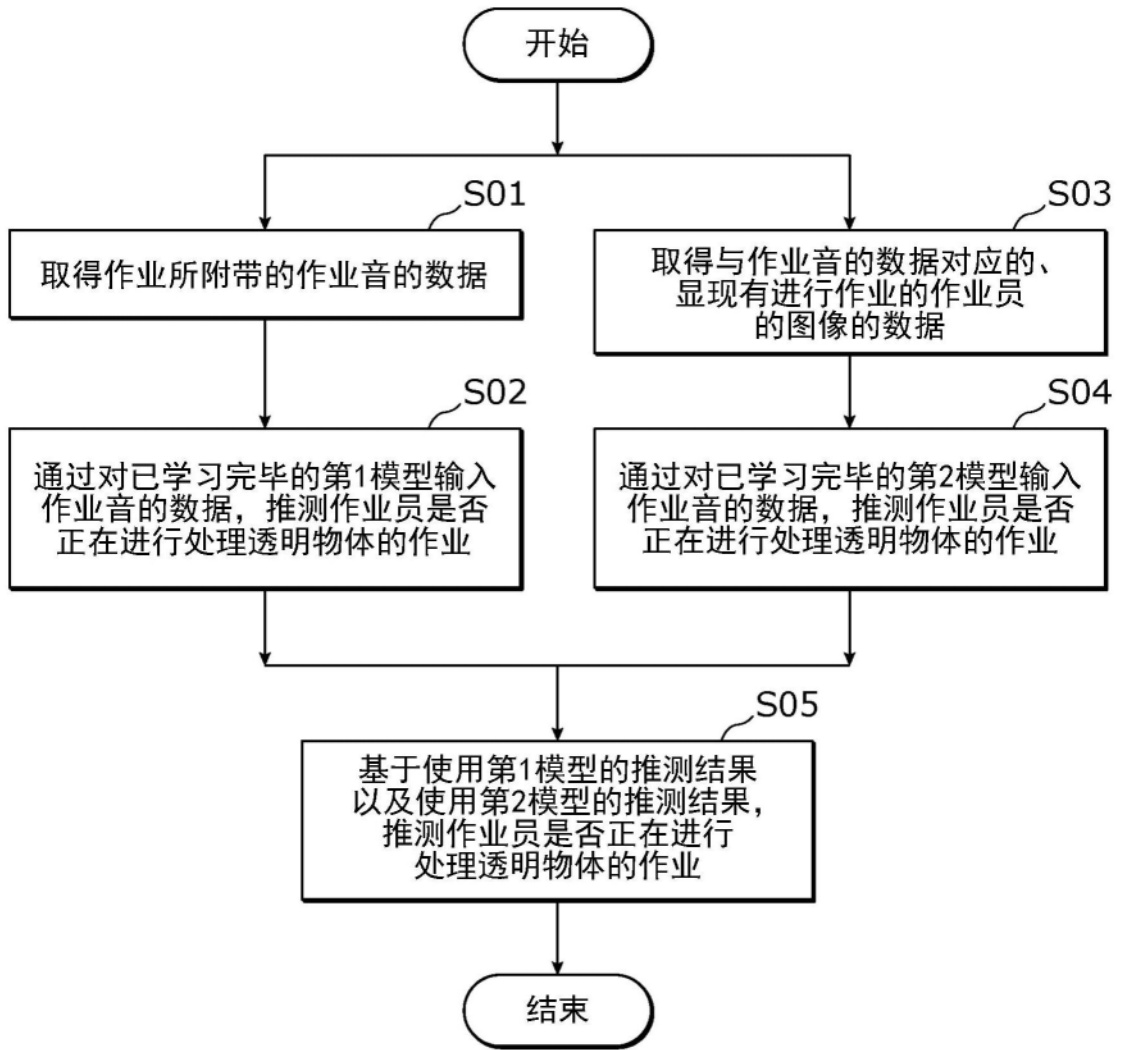


图17

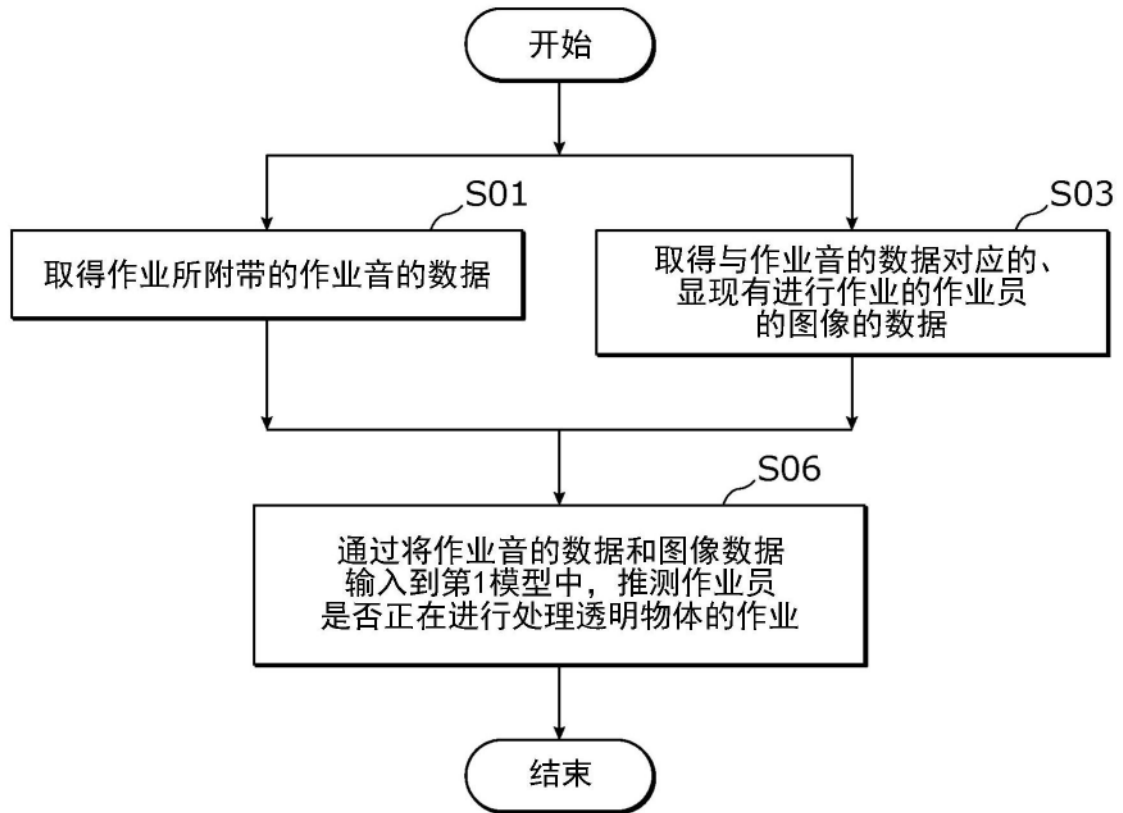


图18

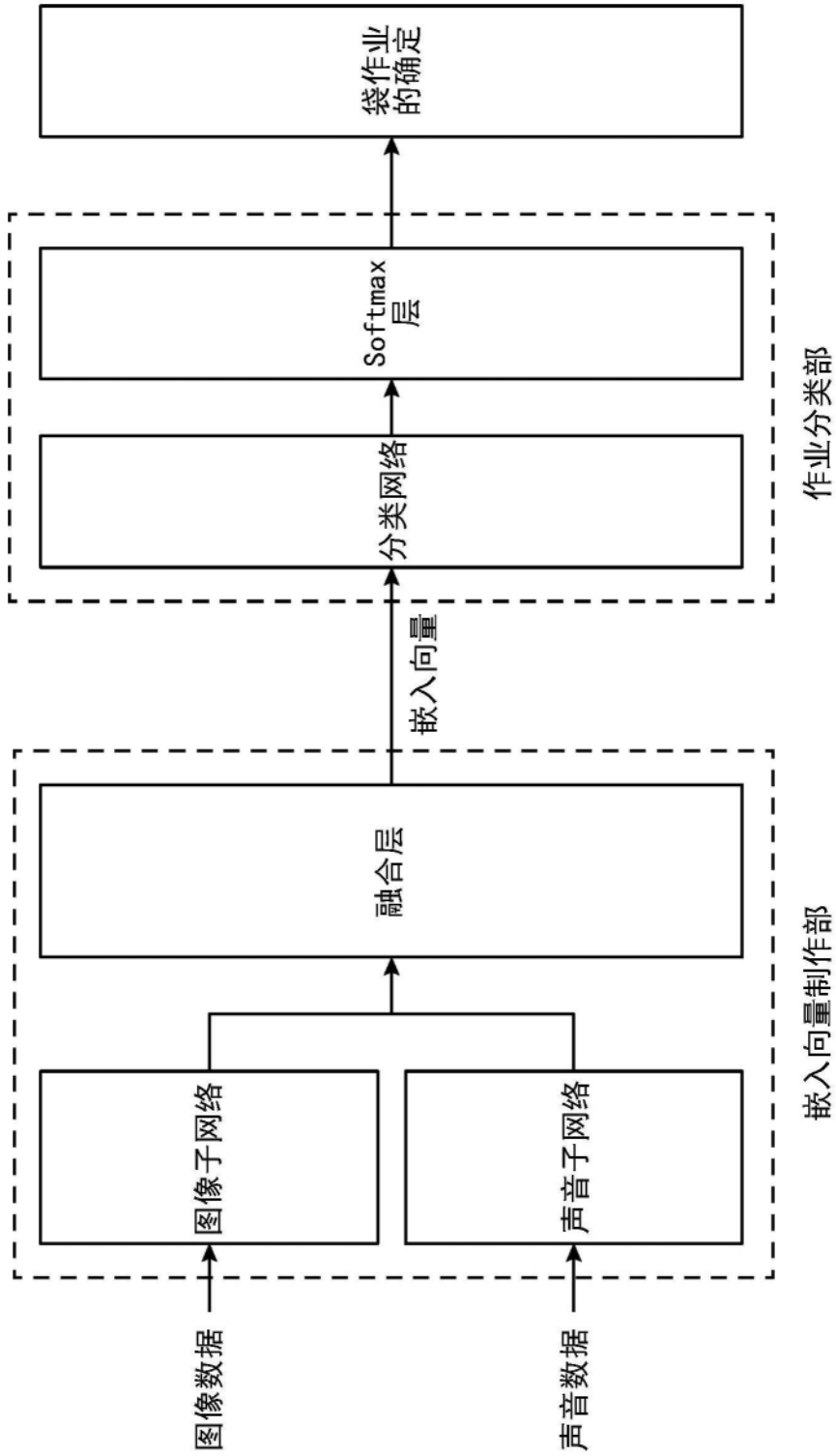


图19

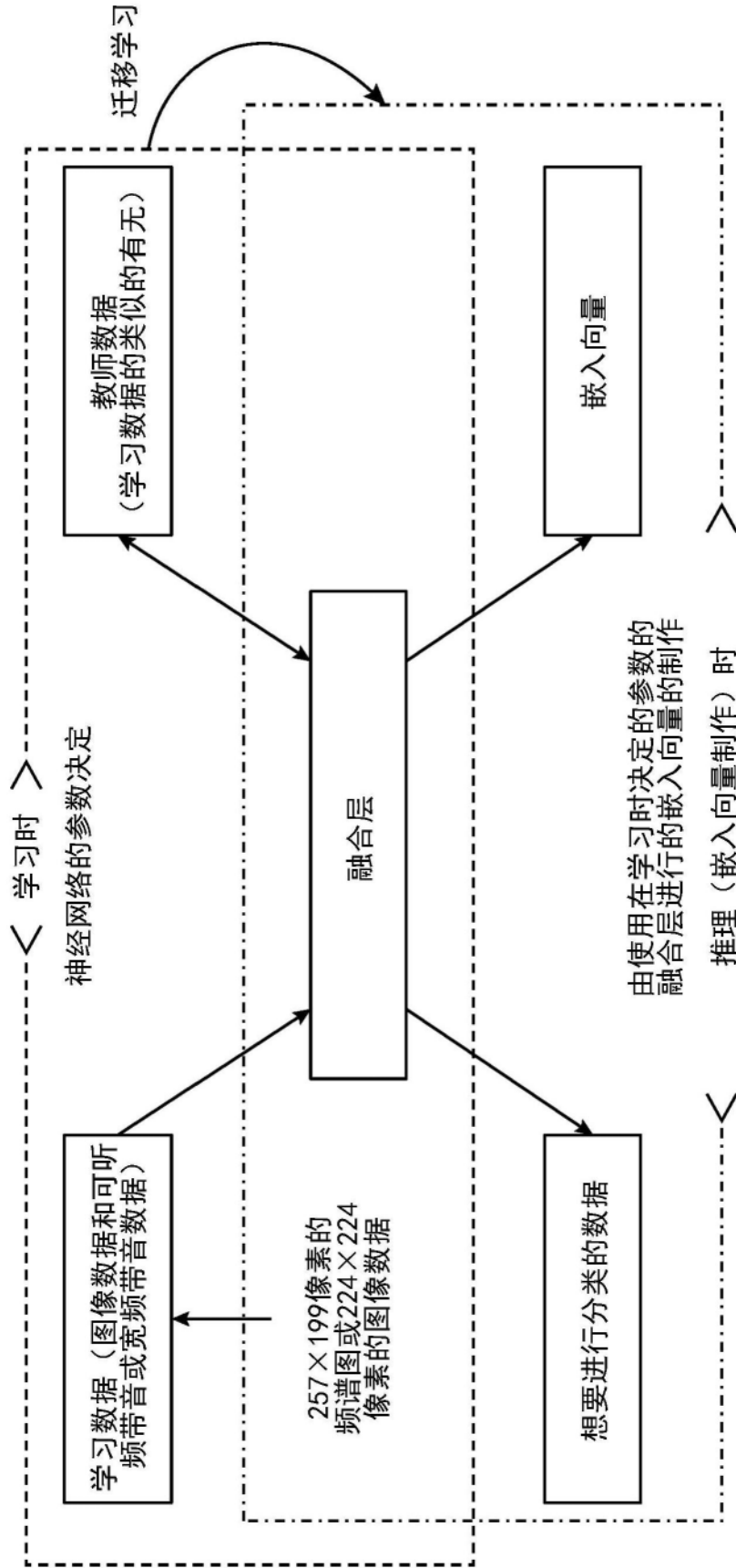


图20

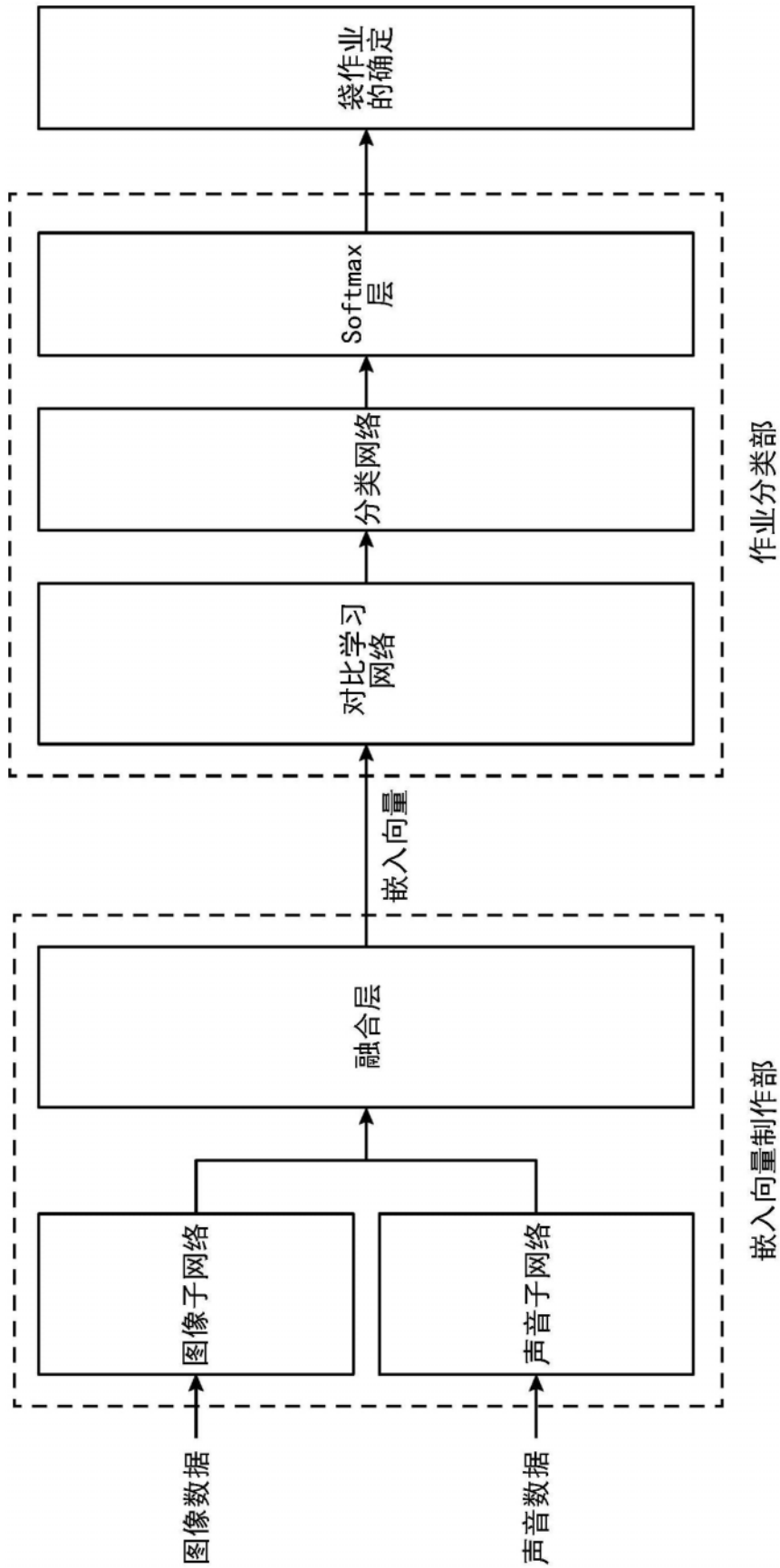


图21

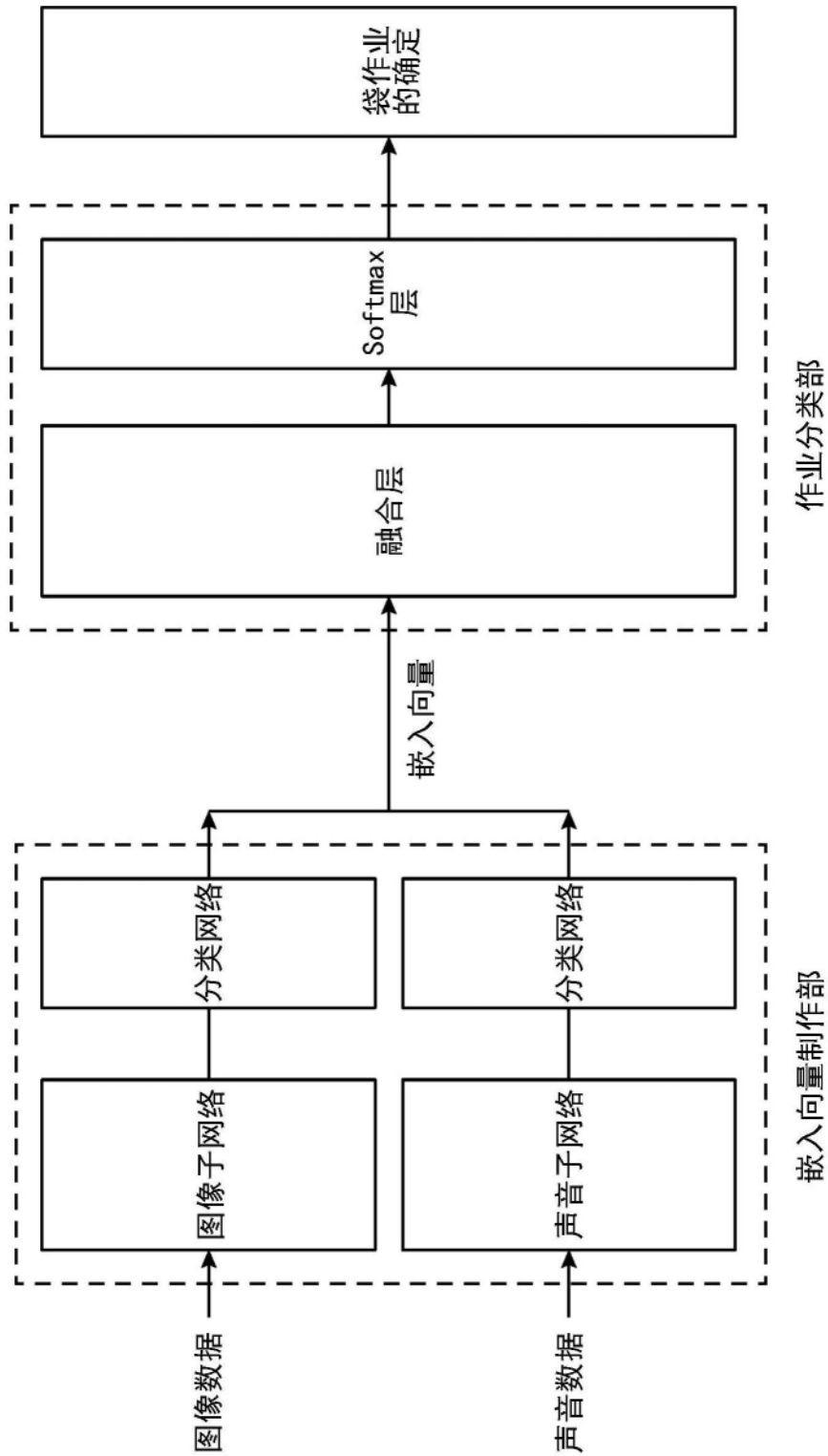


图22

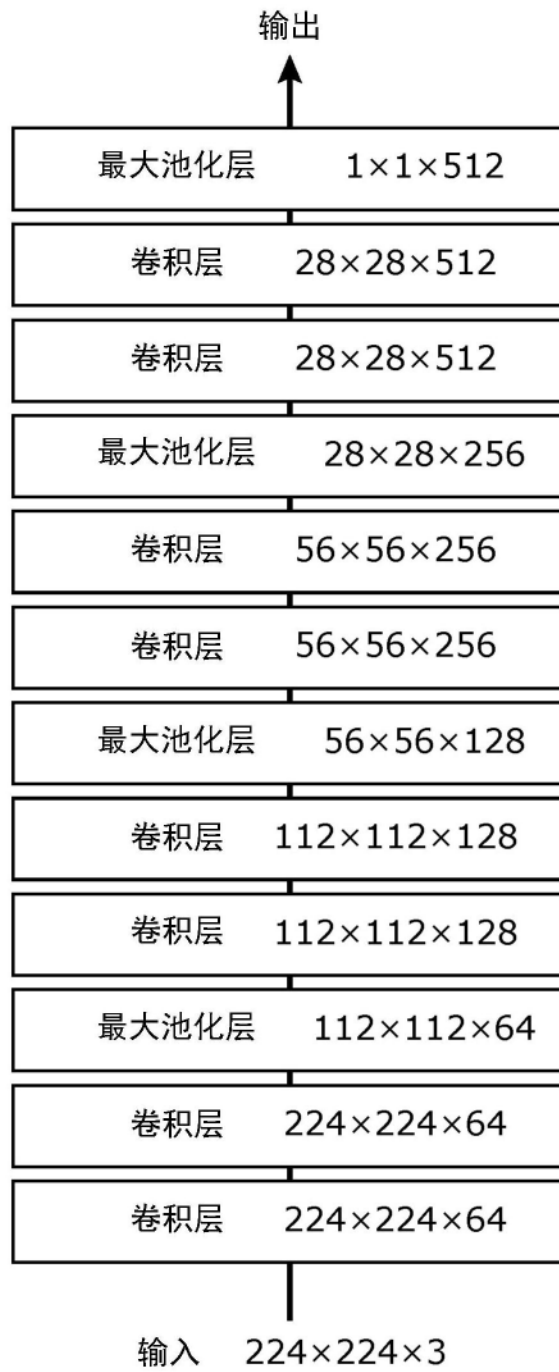


图23

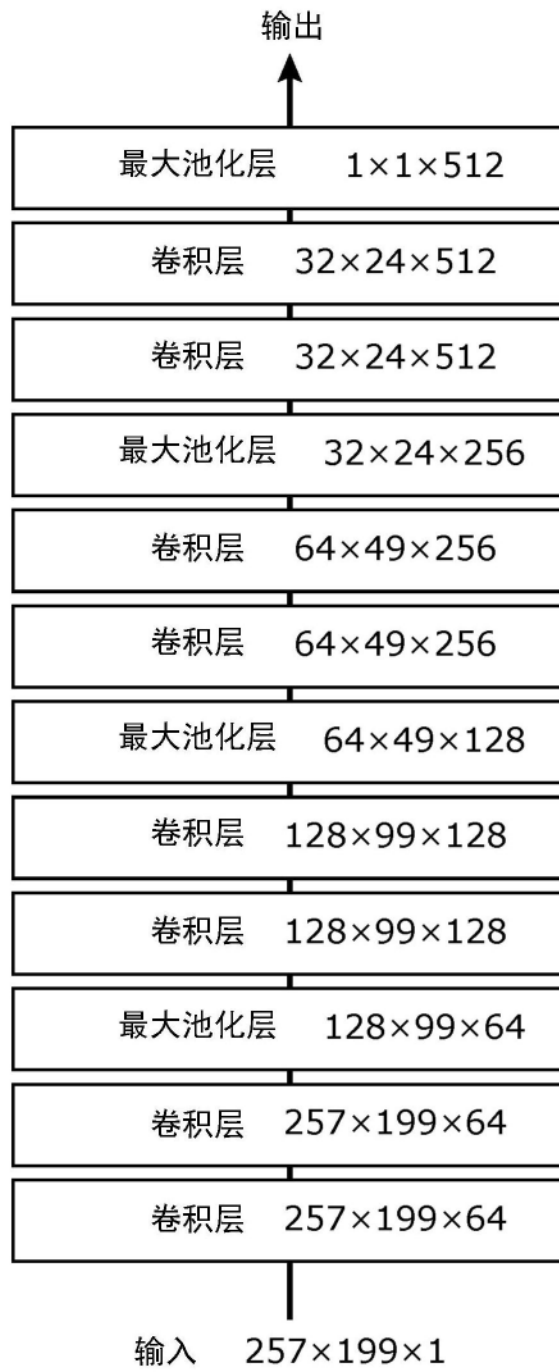


图24

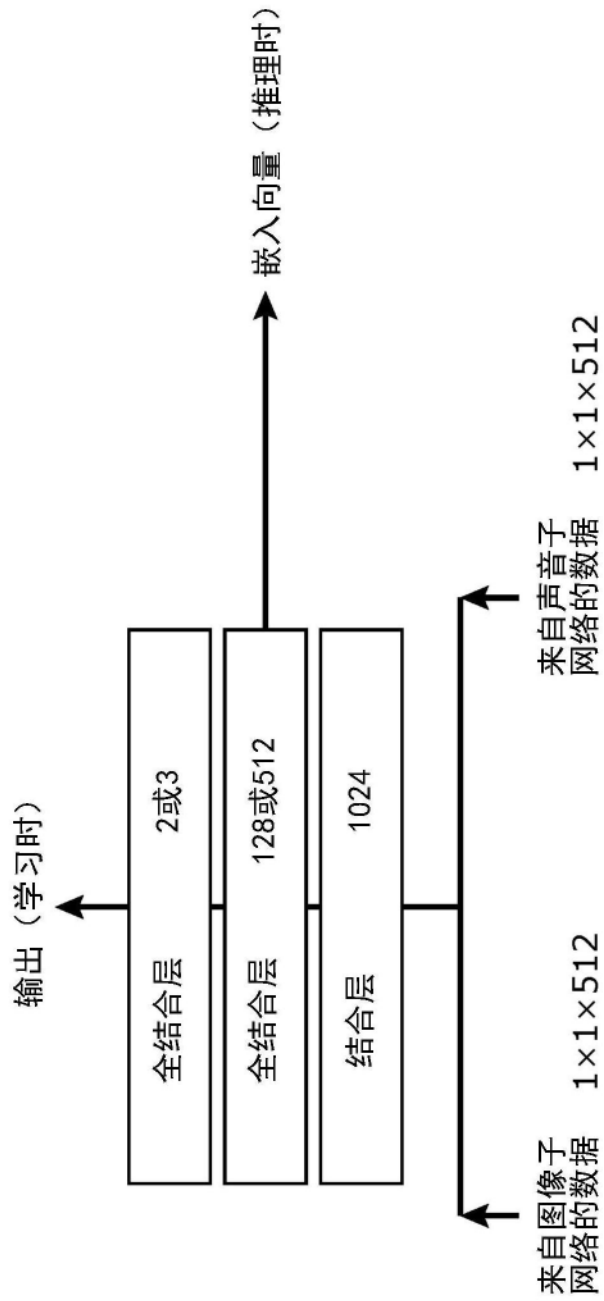


图25

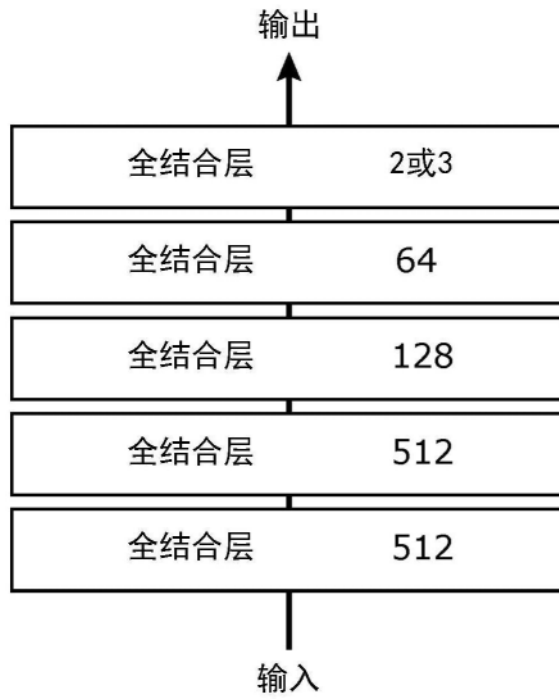


图26

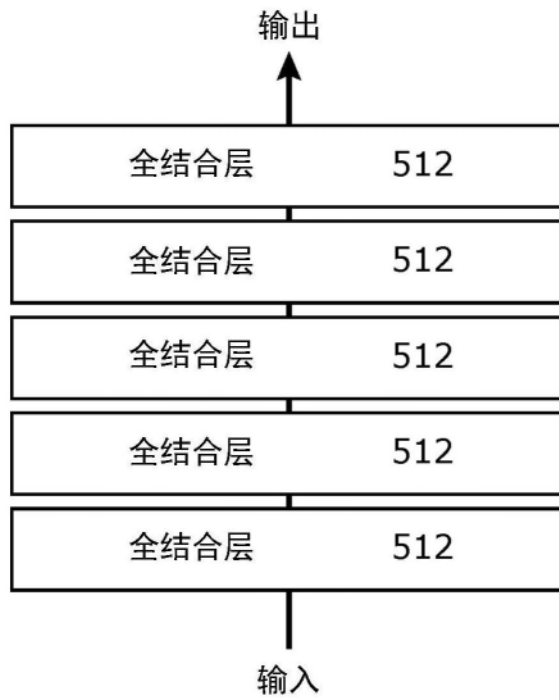


图27

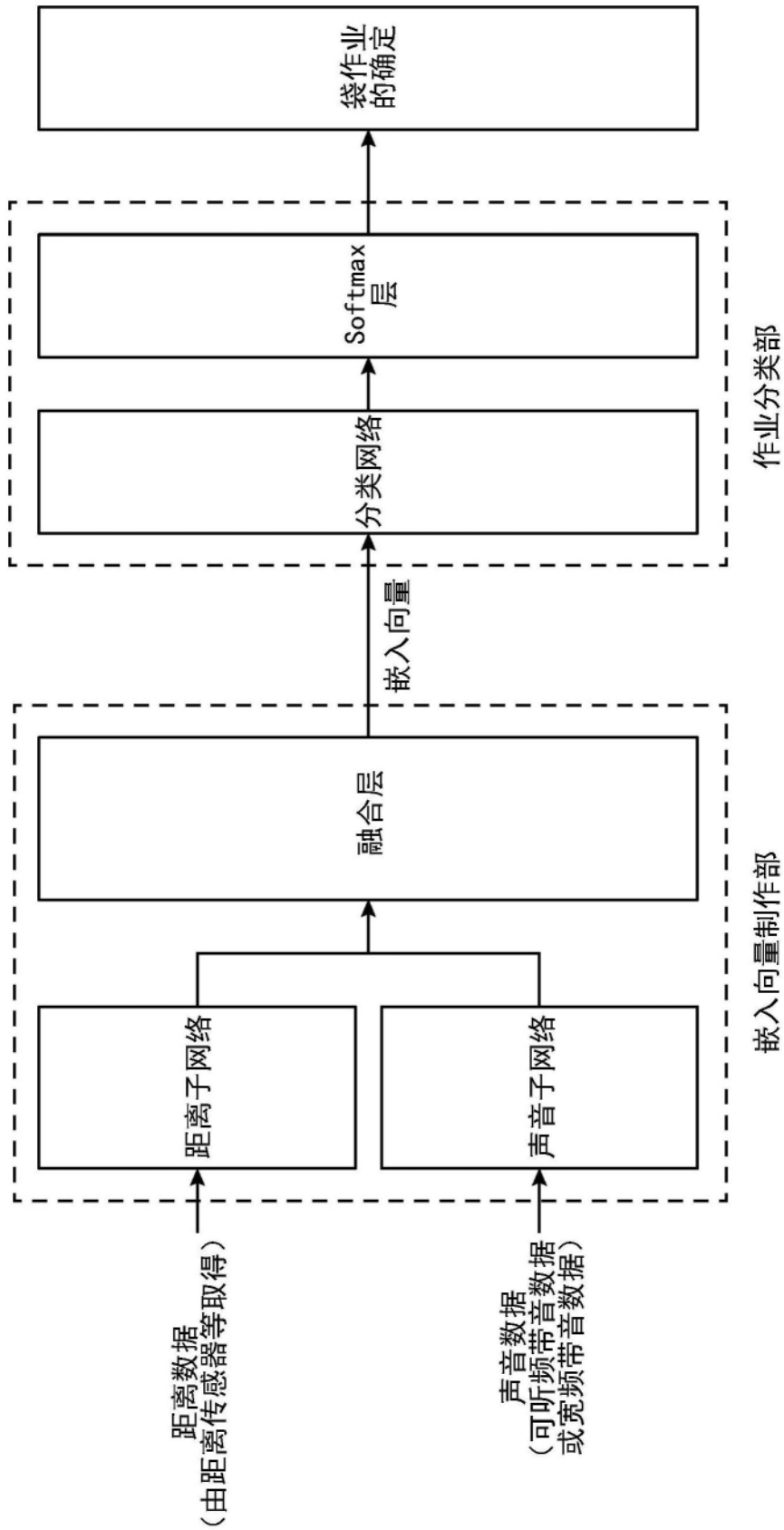


图28

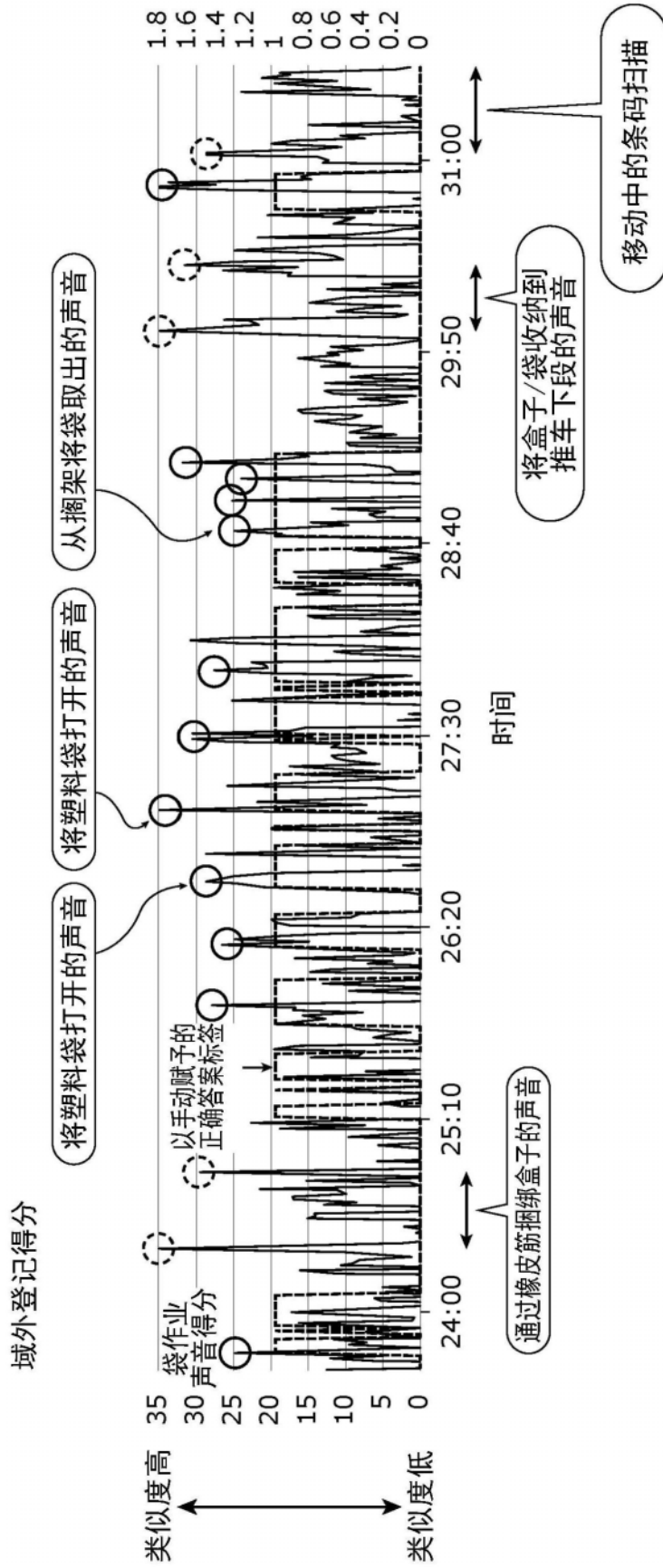


图29

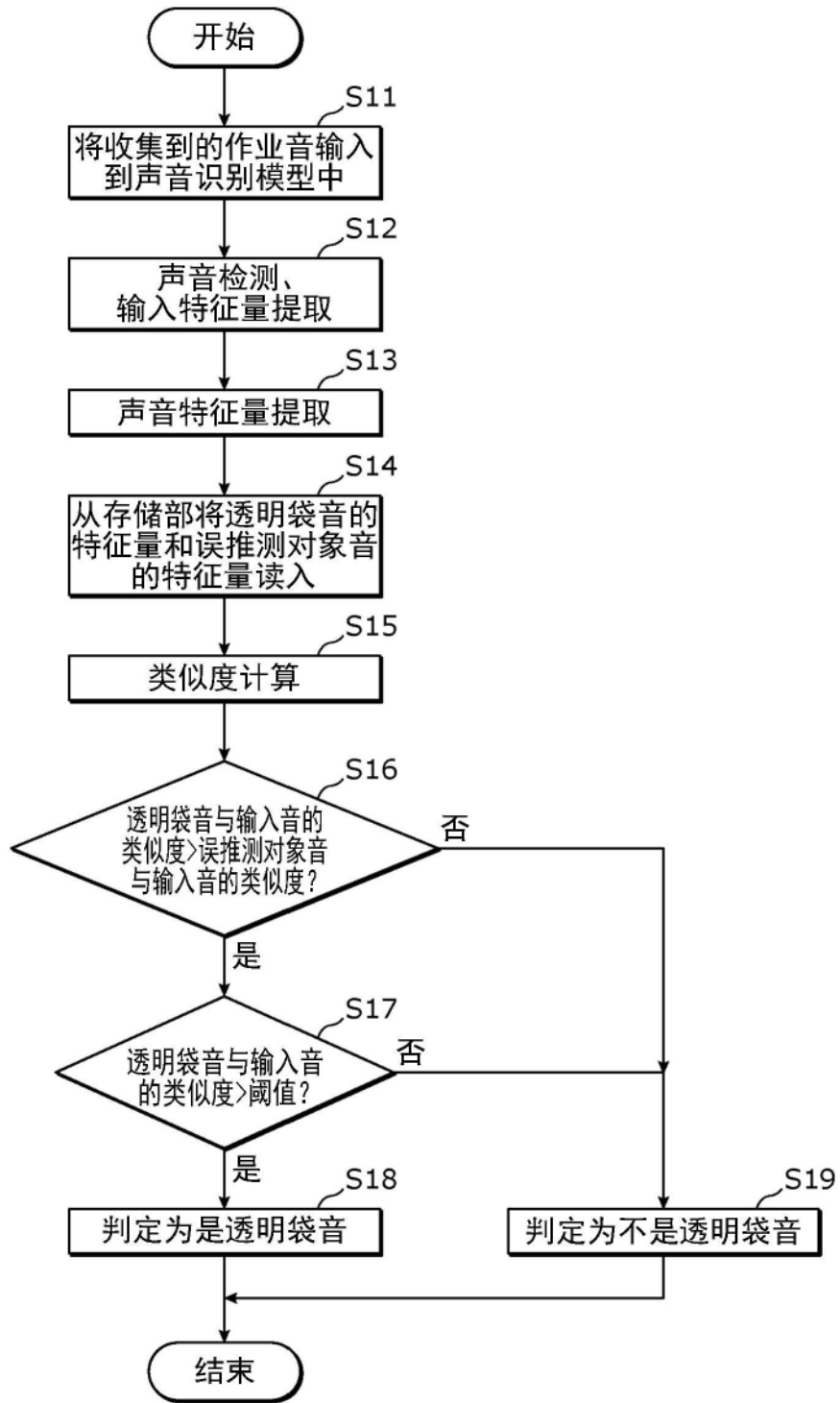


图30A

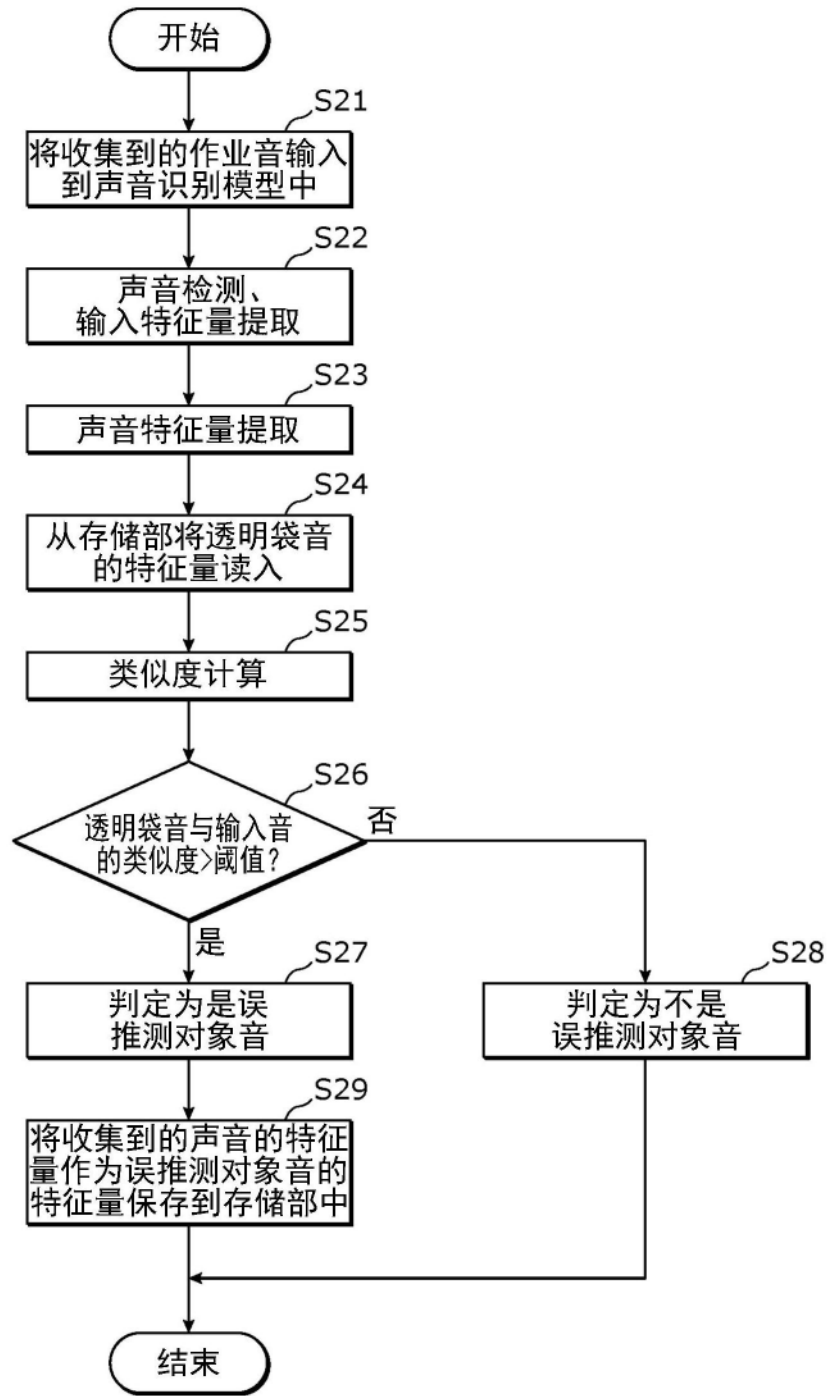


图30B

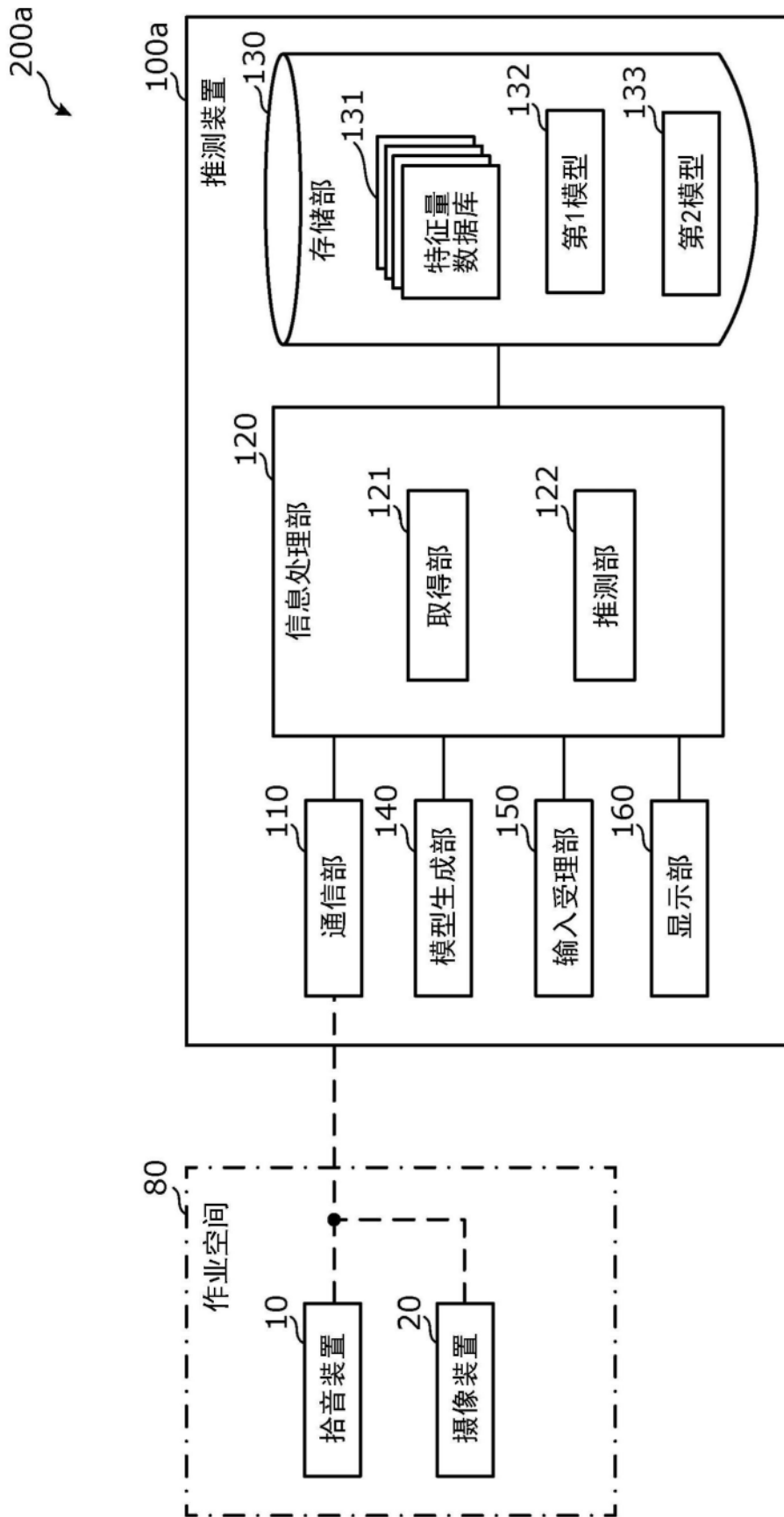


图31