



(12) 发明专利申请

(10) 申请公布号 CN 114420163 A

(43) 申请公布日 2022. 04. 29

(21) 申请号 202210055284.9

G06N 3/04 (2006.01)

(22) 申请日 2022.01.18

G06K 9/62 (2022.01)

(71) 申请人 小米汽车科技有限公司

G06F 3/14 (2006.01)

地址 100176 北京市大兴区北京经济技术
开发区科创十街15号院5号楼6层618
室

G06V 10/764 (2022.01)

G06V 10/82 (2022.01)

(72) 发明人 闫志勇 丁翰林 王永庆 张俊博
王育军

(74) 专利代理机构 北京英创嘉友知识产权代理
事务所(普通合伙) 11447

代理人 张岩龙

(51) Int. Cl.

G10L 25/51 (2013.01)

G06V 20/59 (2022.01)

G06N 3/08 (2006.01)

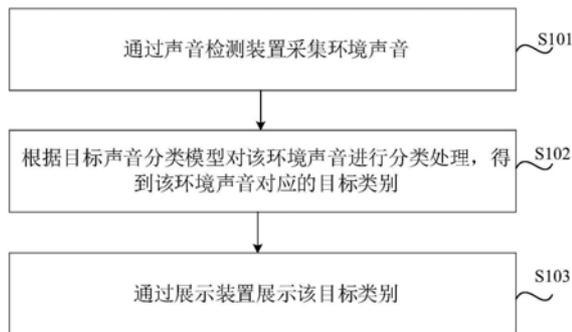
权利要求书3页 说明书17页 附图4页

(54) 发明名称

声音识别方法、装置、存储介质、电子设备及
车辆

(57) 摘要

本公开涉及一种声音识别方法、装置、存储
介质、电子设备及车辆。该方法包括：通过声音检
测装置采集环境声音；根据目标声音分类模型对
该环境声音进行分类处理，得到该环境声音对应
的目标类别；通过展示装置展示所述目标类别。
这样，可以通过对环境声音进行声音识别从而实
现对周边环境物体的全面准确的识别和分类，提
高了物体检测的可靠性。



1. 一种声音识别方法,其特征在于,所述方法包括:
通过声音检测装置采集环境声音;
根据目标声音分类模型对所述环境声音进行分类处理,得到所述环境声音对应的目标类别;
通过展示装置展示所述目标类别。
2. 根据权利要求1所述的方法,其特征在于,所述通过展示装置展示所述目标类别包括:
确定所述目标类别对应的目标图像;
通过所述展示装置展示所述目标图像。
3. 根据权利要求2所述的方法,其特征在于,所述确定所述目标类别对应的目标图像包括:
根据类别图像对应关系,确定所述目标类别对应的目标图像,所述类别图像对应关系包括所述目标类别与所述目标图像的对应关系。
4. 根据权利要求2所述的方法,其特征在于,所述通过所述展示装置展示所述目标图像包括:
在所述展示装置的预设区域展示所述目标图像。
5. 根据权利要求2所述的方法,其特征在于,所述展示装置包括车辆的外反光镜、内反光镜和中控屏幕中的一种或多种。
6. 根据权利要求5所述的方法,其特征在于,在所述展示装置包括车辆的外反光镜的情况下,所述通过所述展示装置展示所述目标图像包括:
在检测到副驾驶座位有乘客乘坐的情况下,通过车辆两侧的外反光镜展示所述目标图像;或者,
在未检测到副驾驶座位有乘客乘坐的情况下,通过车辆驾驶员侧的外反光镜展示所述目标图像。
7. 根据权利要求1所述的方法,其特征在于,所述声音检测装置为一个或多个,所述声音检测装置设置在车辆的任意一侧或多侧的外反光镜内,所述环境声音为所述车辆周边的环境声音。
8. 根据权利要求7所述的方法,其特征在于,在所述声音检测装置为多个的情况下,所述多个声音检测装置分别设置在车辆的左侧外反光镜内和右侧外反光镜内。
9. 根据权利要求1所述的方法,其特征在于,所述根据目标声音分类模型对所述环境声音进行分类处理,得到所述环境声音对应的目标类别包括:
将所述环境声音输入所述目标声音分类模型,得到一个或多个第一候选类别,以及所述环境声音与每个第一候选类别的第一目标相似度;
根据所述第一目标相似度从所述第一候选类别中确定所述目标类别。
10. 根据权利要求9所述的方法,其特征在于,所述根据所述第一目标相似度从所述第一候选类别中确定所述目标类别包括:
将所述第一目标相似度从大到小排名前N位,且所述第一目标相似度大于或等于预设相似度阈值的第一候选类别,作为第二候选类别;
根据所述第二候选类别确定所述目标类别。

11. 根据权利要求10所述的方法,其特征在於,所述第二候选类别为多个,所述根据所述第二候选类别确定所述目标类别包括:

根据预设类别对应关系,确定每个第二候选类别与其他第二候选类别之间的类别关系;所述预设类别对应关系包括任意两个第二候选类别之间的类别关系,所述类别关系包括混淆关系和同类关系;

根据所述第二候选类别和类别关系,确定所述目标类别。

12. 根据权利要求1至11中任一项所述的方法,其特征在於,所述目标声音分类模型为根据目标神经网络模型训练后得到的,所述目标神经网络模型是对预设神经网络模型进行训练,并对训练后的预设神经网络模型进行模型压缩得到的模型。

13. 根据权利要求12所述的方法,其特征在於,所述目标声音分类模型是通过以下方式训练得到的:

获取用于训练的多个样本声音和每个所述样本声音对应的样本类别;

根据所述样本声音和所述样本类别对预设神经网络模型执行预设训练步骤,得到第一待定模型;

对所述第一待定模型进行模型压缩,得到目标神经网络模型;

根据所述样本声音和所述样本类别对所述目标神经网络模型执行所述预设训练步骤,得到第二待定模型;

根据所述第二待定模型,确定所述目标声音分类模型。

14. 根据权利要求13所述的方法,其特征在於,所述对所述第一待定模型进行模型压缩,得到目标神经网络模型包括:

根据所述第一待定模型的预设数目个卷积层,获取所述目标神经网络模型;其中,所述预设数目小于所述第一待定模型的卷积层总层数。

15. 根据权利要求13所述的方法,其特征在於,所述预设训练步骤包括:

循环执行模型训练步骤对目标模型进行训练,直至根据所述样本类别和预测类别确定训练后的目标模型满足预设停止迭代条件,所述目标模型包括预设神经网络模型或者所述目标神经网络模型,所述预测类别为所述样本声音输入该训练后的目标模型后输出的类别;

所述模型训练步骤包括:

获取所述样本声音与多个所述样本类别的第一样本相似度;

根据所述第一样本相似度,从多个所述样本类别中确定所述样本声音对应的预测类别;

在根据所述样本类别和所述预测类别确定训练后的目标模型不满足预设停止迭代条件的情况下,根据所述样本类别和所述预测类别确定目标损失值,根据所述目标损失值更新所述目标模型的参数,得到训练后的目标模型,并将该训练后的目标模型作为新的目标模型。

16. 根据权利要求15所述的方法,其特征在於,所述获取所述样本声音与多个所述样本类别的第一样本相似度包括:

将所述样本声音按照预设周期进行特征提取,得到多个周期的样本特征;

根据多个周期的样本特征,获取所述样本声音与多个所述样本类别的第一样本相似

度。

17. 根据权利要求16所述的方法,其特征在于,所述根据多个周期的样本特征,获取所述样本声音与多个所述样本类别的第一样本相似度包括:

针对每个周期的样本特征,获取该样本特征对应的第一特征编码;并根据所述第一特征编码,得到该样本特征与多个样本类别的第二样本相似度;

根据多个样本特征的所述第二样本相似度,计算得到所述样本声音与多个所述样本类别的第一样本相似度。

18. 根据权利要求13所述的方法,其特征在于,所述样本类别包括所述目标类别和非目标类别。

19. 根据权利要求13所述的方法,其特征在于,所述根据所述第二待定模型,确定所述目标声音分类模型包括:

对所述第二待定模型的模型参数进行模型量化处理,得到所述目标声音分类模型。

20. 一种声音识别装置,其特征在于,所述装置包括:

声音采集模块,被配置为通过声音检测装置采集环境声音;

声音分类模块,被配置为根据目标声音分类模型对所述环境声音进行分类处理,得到所述环境声音对应的目标类别;

展示模块,被配置为通过展示装置展示所述目标类别。

21. 一种电子设备,其特征在于,包括:

处理器;

用于存储处理器可执行指令的存储器;

其中,所述处理器被配置为执行权利要求1至19中任一项所述方法的步骤。

22. 一种计算机可读存储介质,其上存储有计算机程序指令,其特征在于,该程序指令被处理器执行时实现权利要求1至19中任一项所述方法的步骤。

23. 一种车辆,其特征在于,所述车辆包括权利要求21所述的电子设备。

声音识别方法、装置、存储介质、电子设备及车辆

技术领域

[0001] 本公开涉及人工智能技术领域,具体地,涉及一种声音识别方法、装置、存储介质、电子设备及车辆。

背景技术

[0002] 随着人工智能技术在车辆技术领域的应用,车辆的自动驾驶或辅助驾驶技术得到了广泛发展。在相关技术中,主要以雷达或摄像头获取周边物体的信息,并通过图像处理实现自动驾驶或辅助驾驶。而雷达或摄像头都会存在检测盲区,在某些特定场景下存在环境检测可靠性较低的问题。

发明内容

[0003] 为克服相关技术中存在的上述问题,本公开提供一种声音识别方法、装置、存储介质、电子设备及车辆。

[0004] 根据本公开实施例的第一方面,提供一种声音识别方法,所述方法包括:

[0005] 通过声音检测装置采集环境声音;

[0006] 根据目标声音分类模型对所述环境声音进行分类处理,得到所述环境声音对应的目标类别;

[0007] 通过展示装置展示所述目标类别。

[0008] 可选地,所述通过展示装置展示所述目标类别包括:

[0009] 确定所述目标类别对应的目标图像;

[0010] 通过所述展示装置展示所述目标图像。

[0011] 可选地,所述确定所述目标类别对应的目标图像包括:

[0012] 根据类别图像对应关系,确定所述目标类别对应的目标图像,所述类别图像对应关系包括所述目标类别与所述目标图像的对应关系。

[0013] 可选地,所述通过所述展示装置展示所述目标图像包括:

[0014] 在所述展示装置的预设区域展示所述目标图像。

[0015] 可选地,所述展示装置包括车辆的外反光镜、内反光镜和中控屏幕中的一种或多种。

[0016] 可选地,在所述展示装置包括车辆的外反光镜的情况下,所述通过所述展示装置展示所述目标图像包括:

[0017] 在检测到副驾驶座位有乘客乘坐的情况下,通过车辆两侧的外反光镜展示所述目标图像;或者,

[0018] 在未检测到副驾驶座位有乘客乘坐的情况下,通过车辆驾驶员侧的外反光镜展示所述目标图像。

[0019] 可选地,所述声音检测装置为一个或多个,所述声音检测装置设置在车辆的任意一侧或多侧的外反光镜内,所述环境声音为所述车辆周边的环境声音。

[0020] 可选地,在所述声音检测装置为多个的情况下,所述多个声音检测装置分别设置在车辆的左侧外反光镜内和右侧外反光镜内。

[0021] 可选地,所述根据目标声音分类模型对所述环境声音进行分类处理,得到所述环境声音对应的目标类别包括:

[0022] 将所述环境声音输入所述目标声音分类模型,得到一个或多个第一候选类别,以及所述环境声音与每个第一候选类别的第一目标相似度;

[0023] 根据所述第一目标相似度从所述第一候选类别中确定所述目标类别。

[0024] 可选地,所述根据所述第一目标相似度从所述第一候选类别中确定所述目标类别包括:

[0025] 将所述第一目标相似度从大到小排名前N位,且所述第一目标相似度大于或等于预设相似度阈值的第一候选类别,作为第二候选类别;

[0026] 根据所述第二候选类别确定所述目标类别。

[0027] 可选地,所述第二候选类别为多个,所述根据所述第二候选类别确定所述目标类别包括:

[0028] 根据预设类别对应关系,确定每个第二候选类别与其他第二候选类别之间的类别关系;所述预设类别对应关系包括任意两个第二候选类别之间的类别关系,所述类别关系包括混淆关系和同类关系;

[0029] 根据所述第二候选类别和类别关系,确定所述目标类别。

[0030] 可选地,所述目标声音分类模型为根据目标神经网络模型训练后得到的,所述目标神经网络模型是对预设神经网络模型进行训练,并对训练后的预设神经网络模型进行模型压缩得到的模型。

[0031] 可选地,所述目标声音分类模型是通过以下方式训练得到的:

[0032] 获取用于训练的多个样本声音和每个所述样本声音对应的样本类别;

[0033] 根据所述样本声音和所述样本类别对预设神经网络模型执行预设训练步骤,得到第一待定模型;

[0034] 对所述第一待定模型进行模型压缩,得到目标神经网络模型;

[0035] 根据所述样本声音和所述样本类别对所述目标神经网络模型执行所述预设训练步骤,得到第二待定模型;

[0036] 根据所述第二待定模型,确定所述目标声音分类模型。

[0037] 可选地,所述对所述第一待定模型进行模型压缩,得到目标神经网络模型包括:

[0038] 根据所述第一待定模型的预设数目个卷积层,获取所述目标神经网络模型;其中,所述预设数目小于所述第一待定模型的卷积层总层数。

[0039] 可选地,所述预设训练步骤包括:

[0040] 循环执行模型训练步骤对目标模型进行训练,直至根据所述样本类别和预测类别确定训练后的目标模型满足预设停止迭代条件,所述目标模型包括预设神经网络模型或者所述目标神经网络模型,所述预测类别为所述样本声音输入该训练后的目标模型后输出的类别;

[0041] 所述模型训练步骤包括:

[0042] 获取所述样本声音与多个所述样本类别的第一样本相似度;

- [0043] 根据所述第一样本相似度,从多个所述样本类别中确定所述样本声音对应的预测类别;
- [0044] 在根据所述样本类别和所述预测类别确定训练后的目标模型不满足预设停止迭代条件的情况下,根据所述样本类别和所述预测类别确定目标损失值,根据所述目标损失值更新所述目标模型的参数,得到训练后的目标模型,并将该训练后的目标模型作为新的目标模型。
- [0045] 可选地,所述获取所述样本声音与多个所述样本类别的第一样本相似度包括:
- [0046] 将所述样本声音按照预设周期进行特征提取,得到多个周期的样本特征;
- [0047] 根据多个周期的样本特征,获取所述样本声音与多个所述样本类别的第一样本相似度。
- [0048] 可选地,所述根据多个周期的样本特征,获取所述样本声音与多个所述样本类别的第一样本相似度包括:
- [0049] 针对每个周期的样本特征,获取该样本特征对应的第一特征编码;并根据所述第一特征编码,得到该样本特征与多个样本类别的第二样本相似度;
- [0050] 根据多个样本特征的所述第二样本相似度,计算得到所述样本声音与多个所述样本类别的第一样本相似度。
- [0051] 可选地,所述样本类别包括所述目标类别和非目标类别。
- [0052] 可选地,所述根据所述第二待定模型,确定所述目标声音分类模型包括:
- [0053] 对所述第二待定模型的模型参数进行模型量化处理,得到所述目标声音分类模型。
- [0054] 根据本公开实施例的第二方面,提供一种声音识别装置,所述装置包括:
- [0055] 声音采集模块,被配置为通过声音检测装置采集环境声音;
- [0056] 声音分类模块,被配置为根据目标声音分类模型对所述环境声音进行分类处理,得到所述环境声音对应的目标类别;
- [0057] 展示模块,被配置为通过展示装置展示所述目标类别。
- [0058] 可选地,所述展示模块,被配置为确定所述目标类别对应的目标图像;通过所述展示装置展示所述目标图像。
- [0059] 可选地,所述展示模块,被配置为根据类别图像对应关系,确定所述目标类别对应的目标图像,所述类别图像对应关系包括所述目标类别与所述目标图像的对应关系。
- [0060] 可选地,所述展示模块,被配置为在所述展示装置的预设区域展示所述目标图像。
- [0061] 可选地,所述展示装置包括车辆的外反光镜、内反光镜和中控屏幕中的一种或多种。
- [0062] 可选地,在所述展示装置包括车辆的外反光镜的情况下,所述展示模块,被配置为在检测到副驾驶座位有乘客乘坐的情况下,通过车辆两侧的外反光镜展示所述目标图像;或者,在未检测到副驾驶座位有乘客乘坐的情况下,通过车辆驾驶员侧的外反光镜展示所述目标图像。
- [0063] 可选地,所述声音检测装置为一个或多个,所述声音检测装置设置在车辆的任意一侧或多侧的外反光镜内,所述环境声音为所述车辆周边的环境声音。
- [0064] 可选地,在所述声音检测装置为多个的情况下,所述多个声音检测装置分别设置

在车辆的左侧外反光镜内和右侧外反光镜内。

[0065] 可选地,所述声音分类模块,被配置为将所述环境声音输入所述目标声音分类模型,得到一个或多个第一候选类别,以及所述环境声音与每个第一候选类别的第一目标相似度;根据所述第一目标相似度从所述第一候选类别中确定所述目标类别。

[0066] 可选地,所述声音分类模块,被配置为将所述第一目标相似度从大到小排名前N位,且所述第一目标相似度大于或等于预设相似度阈值的第一候选类别,作为第二候选类别;根据所述第二候选类别确定所述目标类别。

[0067] 可选地,所述第二候选类别为多个,所述声音分类模块,被配置为根据预设类别对应关系,确定每个第二候选类别与其他第二候选类别之间的类别关系;所述预设类别对应关系包括任意两个第二候选类别之间的类别关系,所述类别关系包括混淆关系和同类关系;根据所述第二候选类别和类别关系,确定所述目标类别。

[0068] 可选地,所述目标声音分类模型为根据目标神经网络模型训练后得到的,所述目标神经网络模型是对预设神经网络模型进行训练,并对训练后的预设神经网络模型进行模型压缩得到的模型。

[0069] 可选地,所述装置还包括模型训练模块;所述模型训练模块,被配置为:

[0070] 获取用于训练的多个样本声音和每个所述样本声音对应的样本类别;

[0071] 根据所述样本声音和所述样本类别对预设神经网络模型执行预设训练步骤,得到第一待定模型;

[0072] 对所述第一待定模型进行模型压缩,得到目标神经网络模型;

[0073] 根据所述样本声音和所述样本类别对所述目标神经网络模型执行所述预设训练步骤,得到第二待定模型;

[0074] 根据所述第二待定模型,确定所述目标声音分类模型。

[0075] 可选地,所述模型训练模块,被配置为根据所述第一待定模型的预设数目个卷积层,获取所述目标神经网络模型;其中,所述预设数目小于所述第一待定模型的卷积层总层数。

[0076] 可选地,所述预设训练步骤包括:

[0077] 循环执行模型训练步骤对目标模型进行训练,直至根据所述样本类别和预测类别确定训练后的目标模型满足预设停止迭代条件,所述目标模型包括预设神经网络模型或者所述目标神经网络模型,所述预测类别为所述样本声音输入该训练后的目标模型后输出的类别;

[0078] 所述模型训练步骤包括:

[0079] 获取所述样本声音与多个所述样本类别的第一样本相似度;

[0080] 根据所述第一样本相似度,从多个所述样本类别中确定所述样本声音对应的预测类别;

[0081] 在根据所述样本类别和所述预测类别确定训练后的目标模型不满足预设停止迭代条件的情况下,根据所述样本类别和所述预测类别确定目标损失值,根据所述目标损失值更新所述目标模型的参数,得到训练后的目标模型,并将该训练后的目标模型作为新的目标模型。

[0082] 可选地,所述模型训练模块,被配置将所述样本声音按照预设周期进行特征提取,

得到多个周期的样本特征;根据多个周期的样本特征,获取所述样本声音与多个所述样本类别的第一样本相似度。

[0083] 可选地,所述模型训练模块,被配置针对每个周期的样本特征,获取该样本特征对应的第一特征编码;并根据所述第一特征编码,得到该样本特征与多个样本类别的第二样本相似度;根据多个样本特征的所述第二样本相似度,计算得到所述样本声音与多个所述样本类别的第一样本相似度。

[0084] 可选地,所述样本类别包括所述目标类别和非目标类别。

[0085] 可选地,所述模型训练模块,被配置对所述第二待定模型的模型参数进行模型量化处理,得到所述目标声音分类模型。

[0086] 根据本公开实施例的第三方面,提供一种电子设备,包括:

[0087] 处理器;

[0088] 用于存储处理器可执行指令的存储器;

[0089] 其中,所述处理器被配置为执行本公开第一方面所提供的声音识别方法的步骤。

[0090] 根据本公开实施例的第四方面,提供一种计算机可读存储介质,其上存储有计算机程序指令,该程序指令被处理器执行时实现本公开第一方面所提供的声音识别方法的步骤。

[0091] 根据本公开实施例的第五方面,提供一种车辆,该车辆包括本公开第三方面所提供的电子设备。

[0092] 本公开的实施例提供的技术方案可以包括以下有益效果:通过声音检测装置采集环境声音;根据目标声音分类模型对该环境声音进行分类处理,得到该环境声音对应的目标类别;通过展示装置展示所述目标类别。这样,可以通过对环境声音进行声音识别从而实现对外围环境物体的全面准确的识别和分类,从而解决摄像头或雷达检测出现盲区的问题,提高了物体检测的可靠性。进一步地,为了能够在车辆端或设备端进行声音检测,可以通过模型压缩降低目标声音分类模型的复杂度,同时又通过两次训练保障了训练后的目标声音分类模型的声音分类准确度,从而可以将目标声音分类模型部署在车辆端或设备端,提高了声音识别和分类的及时性。

[0093] 应当理解的是,以上的一般描述和后文的细节描述仅是示例性和解释性的,并不能限制本公开。

附图说明

[0094] 此处的附图被并入说明书中并构成本说明书的一部分,示出了符合本公开的实施例,并与说明书一起用于解释本公开的原理。

[0095] 图1是根据一示例性实施例示出的一种声音识别方法的流程图。

[0096] 图2是根据一示例性实施例示出的一种声音检测装置设置在车辆外反光镜上的示意图。

[0097] 图3是根据一示例性实施例示出的一种目标声音分类模型的训练方法的流程图。

[0098] 图4是根据图1所示实施例示出的一种S102步骤的流程图。

[0099] 图5是根据一示例性实施例示出的一种声音识别装置的框图。

[0100] 图6是根据一示例性实施例示出的另一种声音识别装置的框图。

[0101] 图7是根据一示例性实施例示出的电子设备的框图。

[0102] 图8是根据一示例性实施例示出的一种车辆的框图。

具体实施方式

[0103] 这里将详细地对示例性实施例进行说明,其示例表示在附图中。下面的描述涉及附图时,除非另有表示,不同附图中的相同数字表示相同或相似的要素。以下示例性实施例中所描述的实施方式并不代表与本公开相一致的所有实施方式。相反,它们仅是与如所附权利要求书中所详述的、本公开的一些方面相一致的装置和方法的例子。

[0104] 首先,对本公开的应用场景进行说明。本公开可以应用于声音识别场景,例如基于声音识别的车辆自动驾驶或辅助驾驶、智能家居监控、健康检测、机器产线残次品筛选、工业设备故障检测等场景。以车辆自动驾驶为了,在相关技术中,主要以雷达或摄像头获取周边物体的信息,并通过图像处理实现自动驾驶或辅助驾驶。而雷达或摄像头都会存在检测盲区,例如,激光雷达可以对车身周边几米距离的物体进行定位,无法对超过范围的远处移动物体进行定位;在采用车身摄像头进行视觉识别的情况下,同样存在一定的视觉检测盲区,例如距离较远的视频模糊导致无法识别,或者是在摄像头被遮挡无法检测的情况下,很难通过摄像头实现周边物体的准确识别,这样会导致环境检测可靠性降低,进而会影响车辆自动驾驶的可靠性。

[0105] 需要说明的是,本公开以车辆自动驾驶的应用场景为例进行说明,但并不局限于该应用场景,例如,在基于声音识别的智能家居监控、健康检测、机器产线残次品筛选、工业设备故障检测等场景均可以使用本公开提供的方法。

[0106] 为了解决上述问题,本公开提供了一种声音识别方法、装置、存储介质、电子设备及车辆,可以通过声音检测装置采集环境声音;根据目标声音分类模型对环境声音进行分类处理,得到该环境声音对应的目标类别,从而解决摄像头或雷达检测出现盲区的问题,提高了物体检测的可靠性。

[0107] 下面结合具体实施例对本公开进行说明。

[0108] 图1是根据一示例性实施例示出的一种声音识别方法,如图1所示,该方法可以包括:

[0109] S101、通过声音检测装置采集环境声音。

[0110] 示例地,该声音检测装置可以包括一个或多个声音传感器,例如电动麦克风、电容麦克风、或者MEMS (Micro-Electro-Mechanical System,微机电系统) 麦克风。

[0111] 在不同的应用场景下,该声音检测装置的安装位置可以不同,例如,在车辆自动驾驶或辅助驾驶场景下,该声音检测装置可以设置在车辆的车身外侧任意一个或多个位置,例如车辆两侧车身位置、车辆两侧车窗位置、车辆前脸位置、车辆后脸位置、车顶位置或者车辆外后视镜位置等,通过该声音检测装置可以采集车辆周边的环境声音。在智能家居监控场景下,该声音检测装置可以安装在家庭中的每个房间,通过该声音检测装置可以采集每个房间的环境声音。

[0112] S102、根据目标声音分类模型对该环境声音进行分类处理,得到该环境声音对应的目标类别。

[0113] 其中,该目标声音分类模型可以是根据样本声音对通用的声音分类模型进行训练

后得到的。

[0114] S103、通过展示装置展示该目标类别。

[0115] 其中,该展示装置可以包括图像展示装置(例如显示屏)、声音展示装置(例如蜂鸣器或发声器)。

[0116] 示例地,在车辆自动驾驶或辅助驾驶场景下,该展示装置可以包括车辆的外反光镜、内反光镜和中控屏幕中的一种或多种,通过该展示装置可以展示目标类别对应的目标图像,以便提示用户在环境中出现目标类别。其中,外反光镜可以包括车辆两侧的两个外反光镜。

[0117] 进一步地,该展示装置也可以包括车载音响装置,通过该车载音响装置可以给出目标类别对应的目标声音。

[0118] 采用上述方法,通过声音检测装置采集环境声音;根据目标声音分类模型对该环境声音进行分类处理,得到该环境声音对应的目标类别,并通过展示装置展示该目标类别。这样,可以通过对环境声音进行声音识别从而实现了对周边环境物体的全面准确的识别和分类,从而解决摄像头或雷达检测出现盲区的问题,提高了物体检测的可靠性。

[0119] 进一步地,上述S103步骤可以通过以下方式展示目标类别:

[0120] 首先,确定目标类别对应的目标图像。

[0121] 示例地,可以根据类别图像对应关系,确定该目标类别对应的目标图像,该类别图像对应关系包括目标类别与目标图像的对应关系。例如,目标类别“人”对应的目标图像为“人形图像”;目标类别“动物”对应的目标图像为“四足动物图像”;目标类别“救护车”对应的目标图像为“救护车辆图像”等。

[0122] 然后,通过展示装置展示目标图像。

[0123] 示例地,可以在展示装置的预设区域展示目标图像。

[0124] 在该展示装置为外反光镜的情况下,可以在任意一侧或多侧的外反光镜的预设区域展示该目标图像。该预设区域可以是侧边区域,例如该预设区域可以是外反光镜的上侧区域、下侧区域、左侧区域或右侧区域中的一个或多个。

[0125] 这样,通过展示装置,可以展示目标类别对应的目标图像,从而准确地提示用户出现目标类别,以便辅助用户进行相应的应急处理。

[0126] 需要说明的是,若该展示装置包括车辆的外反光镜、内反光镜和中控屏幕中的多个,多个展示装置可以互为备份,提高图像展示的可靠性,避免由于某个展示装置故障导致无法展示目标图像。

[0127] 进一步地,在展示装置包括车辆的外反光镜的情况下,可以通过以下方式展示目标图像:

[0128] 在检测到副驾驶座位有乘客乘坐的情况下,通过车辆两侧的外反光镜展示该目标图像;或者,

[0129] 在未检测到副驾驶座位有乘客乘坐的情况下,通过车辆驾驶员侧的外反光镜展示该目标图像。

[0130] 这样,可以通过车辆载客情况,在相应的外反光镜上展示目标图像,以便提示用户检测到相应的目标类别。

[0131] 在本公开的另一实施例中,上述声音检测装置可以为一个或多个,该声音检测装

置可以设置在车辆的任意一侧或多侧的外反光镜内,上述环境声音为车辆周边的环境声音。

[0132] 图2是根据一示例性实施例示出的一种声音检测装置设置在车辆外反光镜上的示意图,如图2所示,该车辆外反光镜包括反光镜片201、反光镜外壳202和声音检测装置203,该声音检测装置203可以安装在反光镜片201和反光镜外壳202之间。该声音检测装置可以包括一个或多个声音传感器,例如电动麦克风、电容麦克风、或者MEMS麦克风。

[0133] 这样,既可以方便地收集车辆外部的环境声音,又不会影响车辆的外观,同时还可以避免声音检测装置受到日晒雨淋导致受损,提高声音检测装置的使用寿命。

[0134] 进一步地,若声音检测装置为多个,则多个声音检测装置可以分别设置在车辆的左侧外反光镜内和右侧外反光镜内。

[0135] 需要说明的是,该左侧外反光镜可以为驾驶员侧的外反光镜,右侧外反光镜可以为副驾驶座位侧的外反光镜;也可以是相反的,该左侧外反光镜可以为副驾驶座位侧的外反光镜,右侧外反光镜可以为驾驶员侧的外反光镜。本公开对此不作限定。

[0136] 这样,通过多个声音检测装置可以分别检测得到车辆两侧的声音强度,从而可以确定环境声音的来源方向。

[0137] 进一步地,还可以根据该来源方向确定目标反光镜,并通过该目标反光镜展示目标图像。

[0138] 例如,若左侧外反光镜的声音检测装置检测的声音强度大于右侧外反光镜的声音检测装置检测的声音强度,则可以确定环境声音的来源方向为车辆的左侧,将左侧外反光镜作为目标反光镜,并将目标图像展示在该目标反光镜上。

[0139] 再例如,若右侧外反光镜的声音检测装置检测的声音强度大于左侧外反光镜的声音检测装置检测的声音强度,则可以确定环境声音的来源方向为车辆的右侧,将右侧外反光镜作为目标反光镜,并将目标图像展示在该目标反光镜上。

[0140] 这样,可以通过展示目标图像的外反光镜提示用户目标类别可能出现的方向。

[0141] 进一步地,车辆的左侧外反光镜和右侧外反光镜内可以分别设置多个声音检测装置。例如,该多个声音检测装置可以为四个,其中两个设置在车辆左侧外反光镜的外壳和反光镜片之间,另外两个设置在车辆右侧外反光镜的外壳和反光镜片之间。可以对每个声音检测装置进行防震处理和/或防水处理。例如,可以在每个声音检测装置的外面包裹橡胶囊,以防水和防震,并能够降低风噪。

[0142] 进一步地,该声音检测装置可以包括数据接口,该数据接口可以通过线缆与车辆的车载声音模块相连接,以便将检测到的环境声音通过数据接口传输至车载声音模块,以便车载声音模块对环境声音进行分类处理。

[0143] 另外,该声音检测装置还可以包括电源接口和时钟接口,该电源接口和时钟接口同样可以通过线缆与车辆的车内模块相连接,以便对该声音检测装置提供电源和时钟。

[0144] 在本公开的另一实施例中,上述目标声音分类模型可以为根据目标神经网络模型训练后得到的,该目标神经网络模型是对预设神经网络模型进行训练,并对训练后的预设神经网络模型进行模型压缩得到的模型。

[0145] 需要说明的是,相关技术中对声音进行分类识别的人工智能模型可以采用大型化的复杂的神经网络模型,这种大型模型需要运行在服务器上,并且对服务器的硬件要求较

高,可以部署在云端服务器,但难以部署在车辆端或设备端。然而,通过云端进行声音识别的实时性不够,影响声音辅助的自动驾驶功能的可靠性和及时性。在本实施例中,该目标声音分类模型为根据目标神经网络模型训练后得到的,该目标神经网络模型是对预设神经网络模型进行训练,并对训练后的预设神经网络模型进行模型压缩得到的模型。这样,可以降低目标声音分类模型的复杂度,降低模型对硬件的依赖程度,从而可以将该目标声音分类模型部署到车辆端或设备端。

[0146] 采用上述方法,通过声音检测装置采集环境声音;根据目标声音分类模型对该环境声音进行分类处理,得到该环境声音对应的目标类别;其中,该目标声音分类模型为根据目标神经网络模型训练后得到的,该目标神经网络模型是对预设神经网络模型进行训练,并对训练后的预设神经网络模型进行模型压缩得到的模型。这样,可以通过对环境声音进行声音识别从而实现对外围环境物体的全面准确的识别和分类,从而解决摄像头或雷达检测出现盲区的问题,提高了物体检测的可靠性。并且,为了能够在车辆端或设备端进行声音检测,可以通过模型压缩降低目标声音分类模型的复杂度,同时又通过两次训练保障了训练后的目标声音分类模型的声音分类准确度,从而可以将目标声音分类模型部署在车辆端或设备端,提高了声音识别和分类的及时性。

[0147] 图3是根据一示例性实施例示出的一种目标声音分类模型的训练方法的流程图,如图3所示,该训练方法可以包括:

[0148] S301、获取用于训练的多个样本声音和每个样本声音对应的样本类别。

[0149] 示例地,可以从公开的声音数据库中获取用于训练的多个样本声音,并对每个样本声音标注样本类别;也可以从视频数据库中获取视频数据,并从视频数据很重抽取音频数据作为样本声音,然后对每个样本声音标注样本类别。该样本类别可以是声音的类别,例如警报声、人声、哭声、车辆喇叭声、车辆急刹车声等等。

[0150] 进一步地,该样本类别可以是样本声音的强标签,也可以是弱标签。在该样本类别为强标签的情况下,需要标注样本声音中出现的样本类别以及该样本类别出现的起始和结束时间;在该样本类别为弱标签的情况下,可以只标注样本声音中出现的样本类别,无需标注具体的起始和结束时间。使用弱标签可以减少人工标注的工作量,提高样本获取效率。

[0151] 进一步地,上述样本类别可以包括目标类别和非目标类别。该目标类别包括通过目标声音分类模型对环境声音进行处理后得到的多个目标类别,该非目标类别表征除了上述目标类别之外的其他类别,也就是目标声音分类模型不输出的类别。

[0152] 示例地,该目标声音分类模型用于车辆自动驾驶场景,预期的对环境声音进行分类处理后输出的目标类别可以包括“警报声、人声、哭声、车辆喇叭声和车辆急刹车声”,若样本类别中仅包括这些目标类别,由于实际的环境声音中包括的声音类别较多,因此应用这些样本类别进行训练后的目标声音分类模型会存在误识别的问题,例如,环境的风声与哭声有一定的相似度,若不专门设计风声的样本进行训练,会导致训练后的目标声音分类模型误将风声识别为哭声。而在本实施例中,可以通过增加非目标类别的样本声音对模型进行训练,从而可以对目标类别之前的声音进行吸收,使得训练后的模型学到更精细的特征,提升模型的抽象能力,进一步提升目标类别和非目标类别的区分性,也就提高了声音识别的准确性。

[0153] S302、根据样本声音和样本类别对预设神经网络模型执行预设训练步骤,得到第

一待定模型。

[0154] 示例地,可以根据傅里叶变换和梅尔滤波器,对样本声音进行特征提取,得到样本音频特征,该样本音频特征可以包括FBANK特征、MFCC特征或者PNCC特征等。例如,可以针对样本声音,每20ms计算一个1024点的傅里叶变换,窗长为64ms,然后经过64个梅尔滤波器组得到64维的FBANK特征。

[0155] 可以将样本音频特征输入预设神经网络模型执行预设训练步骤,该预设神经网络模型可以为相关技术中的卷积神经网络,该预设训练步骤可以是相关技术中的卷积神经网络训练步骤。

[0156] 该预设神经网络模型可以是移动端卷积神经网络的模型,例如MobileNet等,该移动端卷积神经网络可以包括N层卷积层,N可以为大于或等于5的任意正整数,例如N可以为10或16。

[0157] S303、对该第一待定模型进行模型压缩,得到目标神经网络模型。

[0158] 示例地,可以根据该第一待定模型的预设数目个卷积层,获取该目标神经网络模型;其中,该预设数目小于该第一待定模型的卷积层总层数。

[0159] S304、根据该样本声音和该样本类别对该目标神经网络模型执行上述预设训练步骤,得到第二待定模型。

[0160] 示例地,该第一待定模型的卷积层总层数为N,该预设数目可以为M,M小于N,例如N为10,M可以为5。

[0161] 这样,根据训练后得到的第一待定模型的前M层卷积层的参数作为目标神经网络模型的卷积层的初始化参数。然后对目标神经网络模型执行上述预设训练步骤,得到第二待定模型。

[0162] S305、根据该第二待定模型,确定该目标声音分类模型。

[0163] 示例地,可以将该第二待定模型作为目标声音分类模型。

[0164] 采用上述方式,根据多个样本声音和每个样本声音对应的样本类别,对预设神经网络模型执行预设训练步骤,得到第一待定模型,对该第一待定模型进行模型压缩,得到目标神经网络模型;在根据样本声音和样本类别对该目标神经网络模型执行预设训练步骤,得到第二待定模型;根据该第二待定模型,确定目标声音分类模型。这样,通过模型压缩和两次训练,既可以降低训练后得到的目标声音分类模型的复杂度,也能够确保该目标声音分类模型对声音分类的准确性,从而使得到的目标声音分类模型更加精简和高效,降低对硬件的依赖程度,从而降低车载端或设备端部署的难度。

[0165] 进一步地,在上述S305步骤中,还可以对该第二待定模型的模型参数进行模型量化处理后,得到该目标声音分类模型。

[0166] 其中,该模型量化处理可以包括模型参数压缩,示例地,可以对该模型的参数量化至预设位数,该预设位数可以是8bit或16bit,例如,将所有的浮点参数量化压缩至整数型参数,这样,可以在保证模型性能基本不变的情况下,进一步减少模型的尺寸,降低模型的运算功耗,得到的目标声音分类模型更加精简和高效,降低对硬件的依赖程度,从而降低车载端或设备端部署的难度。

[0167] 进一步地,上述预设训练步骤可以包括以下方式:

[0168] 循环执行模型训练步骤对目标模型进行训练,直至根据该样本类别和预测类别确

定训练后的目标模型满足预设停止迭代条件,该目标模型包括预设神经网络模型或者目标神经网络模型,该预测类别为该样本声音输入训练后的目标模型后输出的类别。

[0169] 需要说明的是,上述预设停止迭代条件可以是现有技术中常用的停止迭代的条件,例如样本类别和预测类别的相似度差异小于预设相似度差异阈值等条件,本公开对此不作限定。

[0170] 上述模型训练步骤包括:

[0171] S11、获取该样本声音与多个该样本类别的第一样本相似度。

[0172] 示例地,将该样本声音按照预设周期进行特征提取,得到多个周期的样本特征;然后根据多个周期的样本特征,获取该样本声音与多个该样本类别的第一样本相似度。

[0173] 示例地,上述样本声音可以是大于5秒的任意样本音频数据,上述预设周期可以是20毫秒至2秒之间的任意时间,例如,预设周期可以是1秒或500毫秒,可以将该样本音频数据按照预设周期进行分割,对分割后的音频片段进行特征提取,可以得到每个分割后音频片段的样本特征。

[0174] 然后,根据多个周期的样本特征,获取该样本声音与多个该样本类别的第一样本相似度可以包括以下相似度获取方式一和相似度获取方式二中的任意一种,其中:

[0175] 相似度获取方式一可以包括以下步骤:

[0176] 首先,针对每个周期的样本特征,获取该样本特征对应的第一特征编码。

[0177] 其次,根据多个样本特征的第一特征编码计算得到样本声音的第二特征编码。

[0178] 最后,根据第二特征编码获取该样本声音与多个该样本类别的第一样本相似度。

[0179] 示例地,可以将多个第一特征编码的平均值作为第二特征编码,也就是在输出层之前的嵌入层(Embedding)取平均,之后再根据取平均后的第二特征编码获取该样本声音与多个该样本类别的第一样本相似度。例如,可以计算得到该第二特征编码与每个样本类别对应的样本特征编码的相似度,并将该相似度作为该样本声音与每个样本类别的第一样本相似度。

[0180] 采用该方式获取第一样本相似度,并对模型进行训练后得到的目标声音分类模型,对与样本声音的时长基本一致的环境声音的识别准确度较高,为了提高环境声音识别的实时性和准确性,在采用该相似度获取方式一进行训练的模型中,可以获取时长较短的样本声音进行训练。

[0181] 相似度获取方式二可以包括以下步骤:

[0182] 首先,针对每个周期的样本特征,获取该样本特征对应的第一特征编码;并根据该第一特征编码,得到该样本特征与多个样本类别的第二样本相似度。

[0183] 然后,根据多个样本特征的第二样本相似度,计算得到样本声音与多个样本类别的第一样本相似度。

[0184] 示例地,可以将每个周期的样本特征输入卷积层,得到该样本特征对应的第一特征编码,计算得到该第一特征编码与每个样本类别对应的样本特征编码的相似度,并将该相似度作为该样本特征与每个样本类别的第二样本相似度。

[0185] 然后,针对每个样本类别,计算样本声音的多个样本特征与该样本类别的第二样本相似度的平均值,作为该样本声音与该样本类别的第二样本相似度。

[0186] 采用该方式获取第一样本相似度,对用于训练的样本声音的长度要求不高,由于

对样本声音分段后的音频进行了相似度计算,因此,即使是时间较长的样本声音,也可以提高训练后的目标声音识别模型对环境声音识别的及时性和准确性。

[0187] S12、根据该第一样本相似度,从多个该样本类别中确定该样本声音对应的预测类别。

[0188] S13、在根据该样本类别和该预测类别确定训练后的目标模型不满足预设停止迭代条件的情况下,根据该样本类别和该预测类别确定目标损失值,根据该目标损失值更新该目标模型的参数,得到训练后的目标模型,并将该训练后的目标模型作为新的目标模型。

[0189] 同样地,上述预设停止迭代条件可以是现有技术中常用的停止迭代的条件,例如样本类别和预测类别的相似度差异小于预设相似度差异阈值等条件,本公开对此不作限定。

[0190] 这样,通过上述预设训练步骤,可以对目标模型进行训练,可以提高训练后的目标模板对环境声音进行识别后得到的目标类别的准确性。

[0191] 图4是根据图1所示实施例示出的一种S102步骤的流程图,如图4所示,上述S102步骤可以包括:

[0192] S1021、将该环境声音输入该目标声音分类模型,得到一个或多个第一候选类别,以及该环境声音与每个第一候选类别的第一目标相似度;

[0193] S1022、根据第一目标相似度从第一候选类别中确定目标类别。

[0194] 示例地,可以将该第一目标相似度从大到小排名前N位的第一候选类别作为目标类别;也可以将该第一目标相似度大于或等于预设相似度阈值的第一候选类别作为目标类别;也可以将第一目标相似度从大到小排名前N位,且该第一目标相似度大于或等于预设相似度阈值的第一候选类别,作为目标类别。

[0195] 在本公开的另一实施例中,上述S1022步骤可以包括:

[0196] 首先,将该第一目标相似度从大到小排名前N位,且该第一目标相似度大于或等于预设相似度阈值的第一候选类别,作为第二候选类别。

[0197] 然后,根据该第二候选类别确定该目标类别。

[0198] 其中,该第二候选类别可以为一个或多个。

[0199] 若该第二候选类别为一个,可以直接将该第二候选类别作为目标类别。

[0200] 若该第二候选类别为多个,可以直接将该多个第二候选类别作为目标类别;也可以将第一目标相似度最大的第二候选类别作为目标类别。

[0201] 进一步地,若该第二候选类别为多个,还可以通过以下方式确定目标类别:

[0202] 首先,根据预设类别对应关系,确定每个第二候选类别与其他第二候选类别之间的类别关系。

[0203] 其中,预设类别对应关系包括任意两个第二候选类别之间的类别关系,该类别关系包括混淆关系和同类关系。混淆关系用于表征两个第二候选类别之间为非同类的易混淆的类别,例如“风声”和“哭声”;同类关系用于表征两个第二候选类别之间为相同场景下的类别,例如“哭声”和“人声”。

[0204] 然后,根据第二候选类别和类别关系,确定该目标类别。

[0205] 例如,若多个第二候选类别中只包括同类关系的第二候选类别,则可以直接将多个第二候选类别作为目标类别,或者将第一目标相似度最大的第二候选类别作为目标类

别。

[0206] 再例如,若多个第二候选类别中包括混淆关系的第二候选类别,则可以计算该多个第二候选类别的混淆系数,在该混淆系数小于或等于预设混淆门限的情况下,将多个第二候选类别作为目标类别,或者将第一目标相似度最大的第二候选类别作为目标类别;而在该混淆系数大于预设混淆门限的情况下,可以不输出目标类别。

[0207] 其中,该混淆系数可以表征类别关系为混淆关系的数目在多个第二候选类别的类别关系的总数中所占的比例。示例地,该多个第二候选类别为5个,每个两个第二候选类别之间存在类别关系,该类别关系的总数为6;其中类别关系为混淆关系的数目为3,则该混淆系数可以为0.5。上述预设混淆门限可以为0.7,这样,该混淆系数小于预设混淆门限,因此,可以将该多个第二候选类别作为目标类别,或者将第一目标相似度最大的第二候选类别作为目标类别。

[0208] 通过该方式,可以根据识别出的候选类别的混淆关系确定模型的识别准确性,在混淆关系满足预设条件(混淆系数小于或等于预设混淆门限)的情况下,确定模型的识别准确性满足条件,使得获取到的目标类别更为准确。

[0209] 图5是根据一示例性实施例示出的一种声音识别装置500的框图,如图5所示,该装置500可以包括:

[0210] 声音采集模块501,被配置为通过声音检测装置采集环境声音;

[0211] 声音分类模块502,被配置为根据目标声音分类模型对该环境声音进行分类处理,得到该环境声音对应的目标类别;

[0212] 展示模块503,被配置为通过展示装置展示该目标类别。

[0213] 可选地,该展示模块503,被配置为确定该目标类别对应的目标图像;通过该展示装置展示该目标图像。

[0214] 可选地,该展示模块503,被配置为根据类别图像对应关系,确定该目标类别对应的目标图像,该类别图像对应关系包括该目标类别与该目标图像的对应关系。

[0215] 可选地,该展示模块503,被配置为在该展示装置的预设区域展示该目标图像。

[0216] 可选地,该展示装置包括车辆的外反光镜、内反光镜和中控屏幕中的一种或多种。

[0217] 可选地,在该展示装置包括车辆的外反光镜的情况下,该展示模块503,被配置为在检测到副驾驶座位有乘客乘坐的情况下,通过车辆两侧的外反光镜展示该目标图像;或者,在未检测到副驾驶座位有乘客乘坐的情况下,通过车辆驾驶员侧的外反光镜展示该目标图像。

[0218] 可选地,该声音检测装置为一个或多个,该声音检测装置设置在车辆的任意一侧或多侧的外反光镜内,该环境声音为该车辆周边的环境声音。

[0219] 可选地,在该声音检测装置为多个的情况下,该多个声音检测装置分别设置在车辆的左侧外反光镜内和右侧外反光镜内。

[0220] 可选地,该声音分类模块502,被配置为将该环境声音输入该目标声音分类模型,得到一个或多个第一候选类别,以及该环境声音与每个第一候选类别的第一目标相似度;根据该第一目标相似度从该第一候选类别中确定该目标类别。

[0221] 可选地,该声音分类模块502,被配置为将该第一目标相似度从大到小排名前N位,且该第一目标相似度大于或等于预设相似度阈值的第一候选类别,作为第二候选类别;根

据该第二候选类别确定该目标类别。

[0222] 可选地,该第二候选类别为多个,该声音分类模块502,被配置为根据预设类别对应关系,确定每个第二候选类别与其他第二候选类别之间的类别关系;该预设类别对应关系包括任意两个第二候选类别之间的类别关系,该类别关系包括混淆关系和同类关系;根据该第二候选类别和类别关系,确定该目标类别。

[0223] 可选地,该目标声音分类模型为根据目标神经网络模型训练后得到的,该目标神经网络模型是对预设神经网络模型进行训练,并对训练后的预设神经网络模型进行模型压缩得到的模型。

[0224] 图6是根据一示例性实施例示出的另一种声音识别装置的框图,如图6所示,该装置还可以包括模型训练模块601,该模型训练模块601,被配置为:

[0225] 获取用于训练的多个样本声音和每个该样本声音对应的样本类别;

[0226] 根据该样本声音和该样本类别对预设神经网络模型执行预设训练步骤,得到第一待定模型;

[0227] 对该第一待定模型进行模型压缩,得到目标神经网络模型;

[0228] 根据该样本声音和该样本类别对该目标神经网络模型执行该预设训练步骤,得到第二待定模型;

[0229] 根据该第二待定模型,确定该目标声音分类模型。

[0230] 可选地,该模型训练模块601,被配置为根据该第一待定模型的预设数目个卷积层,获取该目标神经网络模型;其中,该预设数目小于该第一待定模型的卷积层总层数。

[0231] 可选地,该模型训练模块601,被配置为循环执行模型训练步骤对目标模型进行训练,直至根据该样本类别和预测类别确定训练后的目标模型满足预设停止迭代条件,该目标模型包括预设神经网络模型或者该目标神经网络模型,该预测类别为该样本声音输入该训练后的目标模型后输出的类别;

[0232] 该模型训练步骤包括:

[0233] 获取该样本声音与多个该样本类别的第一样本相似度;

[0234] 根据该第一样本相似度,从多个该样本类别中确定该样本声音对应的预测类别;

[0235] 在根据该样本类别和该预测类别确定训练后的目标模型不满足预设停止迭代条件的情况下,根据该样本类别和该预测类别确定目标损失值,根据该目标损失值更新该目标模型的参数,得到训练后的目标模型,并将该训练后的目标模型作为新的目标模型。

[0236] 可选地,该模型训练模块601,被配置为将该样本声音按照预设周期进行特征提取,得到多个周期的样本特征;根据多个周期的样本特征,获取该样本声音与多个该样本类别的第一样本相似度。

[0237] 可选地,该模型训练模块601,被配置为针对每个周期的样本特征,获取该样本特征对应的第一特征编码;并根据该第一特征编码,得到该样本特征与多个样本类别的第二样本相似度;根据多个样本特征的该第二样本相似度,计算得到该样本声音与多个该样本类别的第一样本相似度。

[0238] 可选地,该样本类别包括该目标类别和非目标类别。

[0239] 可选地,该模型训练模块601,被配置为对该第二待定模型的模型参数进行模型量化处理,得到该目标声音分类模型。

[0240] 关于上述实施例中的装置,其中各个模块执行操作的具体方式已经在有关该方法的实施例中进行了详细描述,此处将不做详细阐述说明。

[0241] 综上所述,采用本公开上述实施例中的装置,通过声音检测装置采集环境声音;根据目标声音分类模型对该环境声音进行分类处理,得到该环境声音对应的目标类别;其中,该目标声音分类模型为根据目标神经网络模型训练后得到的,该目标神经网络模型是对预设神经网络模型进行训练,并对训练后的预设神经网络模型进行模型压缩得到的模型。这样,可以通过对环境声音进行声音识别从而实现了对周边环境物体的全面准确的识别和分类,从而解决摄像头或雷达检测出现盲区的问题,提高了物体检测的可靠性。并且,为了能够在车辆端或设备端进行声音检测,可以通过模型压缩降低目标声音分类模型的复杂度,同时又通过两次训练保障了训练后的目标声音分类模型的声音分类准确度,从而可以将目标声音分类模型部署在车辆端或设备端,提高了声音识别和分类的及时性。

[0242] 本公开还提供一种计算机可读存储介质,其上存储有计算机程序指令,该程序指令被处理器执行时实现本公开提供的声音识别方法的步骤。

[0243] 图7是根据一示例性实施例示出的电子设备900的框图。例如,电子设备900可以是移动电话,计算机,数字广播终端,消息收发设备,游戏控制台,平板设备,医疗设备,健身设备,个人数字助理、路由器、车载终端等。

[0244] 参照图7,电子设备900可以包括以下一个或多个组件:处理组件902,存储器904,电力组件906,多媒体组件908,音频组件910,输入/输出(I/O)接口912,传感器组件914,以及通信组件916。

[0245] 处理组件902通常控制电子设备900的整体操作,诸如与显示,电话呼叫,数据通信,相机操作和记录操作相关联的操作。处理组件902可以包括一个或多个处理器920来执行指令,以完成上述声音识别方法的全部或部分步骤。此外,处理组件902可以包括一个或多个模块,便于处理组件902和其他组件之间的交互。例如,处理组件902可以包括多媒体模块,以方便多媒体组件908和处理组件902之间的交互。

[0246] 存储器904被配置为存储各种类型的数据以支持在电子设备900的操作。这些数据的示例包括用于在电子设备900上操作的任何应用程序或方法的指令,联系人数据,电话簿数据,消息,图片,视频等。存储器904可以由任何类型的易失性或非易失性存储设备或者它们的组合实现,如静态随机存取存储器(SRAM),电可擦除可编程只读存储器(EEPROM),可擦除可编程只读存储器(EPROM),可编程只读存储器(PROM),只读存储器(ROM),磁存储器,快闪存储器,磁盘或光盘。

[0247] 电力组件906为电子设备900的各种组件提供电力。电力组件906可以包括电源管理系统,一个或多个电源,及其他与为电子设备900生成、管理和分配电力相关联的组件。

[0248] 多媒体组件908包括在所述电子设备900和用户之间的提供一个输出接口的屏幕。在一些实施例中,屏幕可以包括液晶显示器(LCD)和触摸面板(TP)。如果屏幕包括触摸面板,屏幕可以被实现为触摸屏,以接收来自用户的输入信号。触摸面板包括一个或多个触摸传感器以感测触摸、滑动和触摸面板上的手势。所述触摸传感器可以不仅感测触摸或滑动动作的边界,而且还检测与所述触摸或滑动操作相关的持续时间和压力。在一些实施例中,多媒体组件908包括一个前置摄像头和/或后置摄像头。当电子设备900处于操作模式,如拍摄模式或视频模式时,前置摄像头和/或后置摄像头可以接收外部的多媒体数据。每个前置

摄像头和后置摄像头可以是一个固定的光学透镜系统或具有焦距和光学变焦能力。

[0249] 音频组件910被配置为输出和/或输入音频信号。例如,音频组件910包括一个麦克风(MIC),当电子设备900处于操作模式,如呼叫模式、记录模式和语音识别模式时,麦克风被配置为接收外部音频信号。所接收的音频信号可以被进一步存储在存储器904或经由通信组件916发送。在一些实施例中,音频组件910还包括一个扬声器,用于输出音频信号。

[0250] I/O接口912为处理组件902和外围接口模块之间提供接口,上述外围接口模块可以是键盘,点击轮,按钮等。这些按钮可包括但不限于:主页按钮、音量按钮、启动按钮和锁定按钮。

[0251] 传感器组件914包括一个或多个传感器,用于为电子设备900提供各个方面的状态评估。例如,传感器组件914可以检测到电子设备900的打开/关闭状态,组件的相对定位,例如所述组件为电子设备900的显示器和小键盘,传感器组件914还可以检测电子设备900或电子设备900一个组件的位置改变,用户与电子设备900接触的存在或不存在,电子设备900方位或加速/减速和电子设备900的温度变化。传感器组件914可以包括接近传感器,被配置用来在没有任何的物理接触时检测附近物体的存在。传感器组件914还可以包括光传感器,如CMOS或CCD图像传感器,用于在成像应用中使用。在一些实施例中,该传感器组件914还可以包括加速度传感器,陀螺仪传感器,磁传感器,压力传感器或温度传感器。

[0252] 通信组件916被配置为便于电子设备900和其他设备之间有线或无线方式的通信。电子设备900可以接入基于通信标准的无线网络,例如Wi-Fi,2G、3G、4G、5G、NB-IOT、eMTC、或其他6G等,或它们的组合。在一个示例性实施例中,通信组件916经由广播信道接收来自外部广播管理系统的广播信号或广播相关信息。在一个示例性实施例中,所述通信组件916还包括近场通信(NFC)模块,以促进短程通信。例如,在NFC模块可基于射频识别(RFID)技术,红外数据协会(IrDA)技术,超宽带(UWB)技术,蓝牙(BT)技术和其他技术来实现。

[0253] 在示例性实施例中,电子设备900可以被一个或多个应用专用集成电路(ASIC)、数字信号处理器(DSP)、数字信号处理设备(DSPD)、可编程逻辑器件(PLD)、现场可编程门阵列(FPGA)、控制器、微控制器、微处理器或其他电子元件实现,用于执行上述声音识别方法。

[0254] 在示例性实施例中,还提供了一种包括指令的非临时性计算机可读存储介质,例如包括指令的存储器904,上述指令可由电子设备900的处理器920执行以完成上述声音识别方法。例如,所述非临时性计算机可读存储介质可以是ROM、随机存取存储器(RAM)、CD-ROM、磁带、软盘和光数据存储设备等。

[0255] 在另一示例性实施例中,还提供一种计算机程序产品,该计算机程序产品包含能够由可编程的装置执行的计算机程序,该计算机程序具有当由该可编程的装置执行时用于执行上述声音识别方法的代码部分。

[0256] 图8是根据一示例性实施例示出的车辆的框图,如图8所示,该装置可以包括上述电子设备900。

[0257] 本领域技术人员在考虑说明书及实践本公开后,将容易想到本公开的其它实施方案。本申请旨在涵盖本公开的任何变型、用途或者适应性变化,这些变型、用途或者适应性变化遵循本公开的一般性原理并包括本公开未公开的本技术领域中的公知常识或惯用技术手段。说明书和实施例仅被视为示例性的,本公开的真正范围和精神由下面的权利要求指出。

[0258] 应当理解的是,本公开并不局限于上面已经描述并在附图中示出的精确结构,并且可以在不脱离其范围进行各种修改和改变。本公开的范围仅由所附的权利要求来限制。

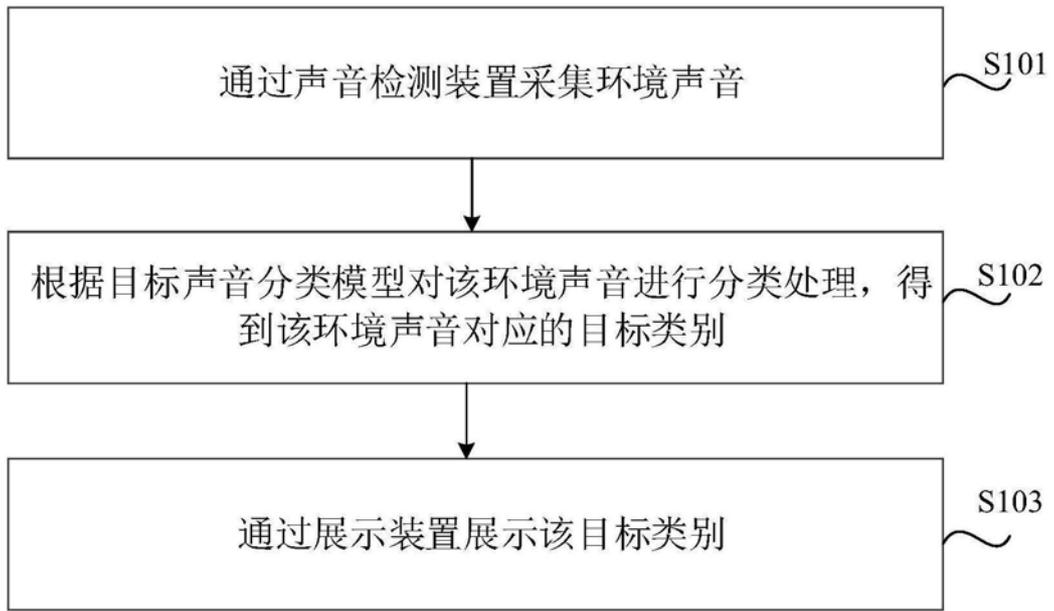


图1

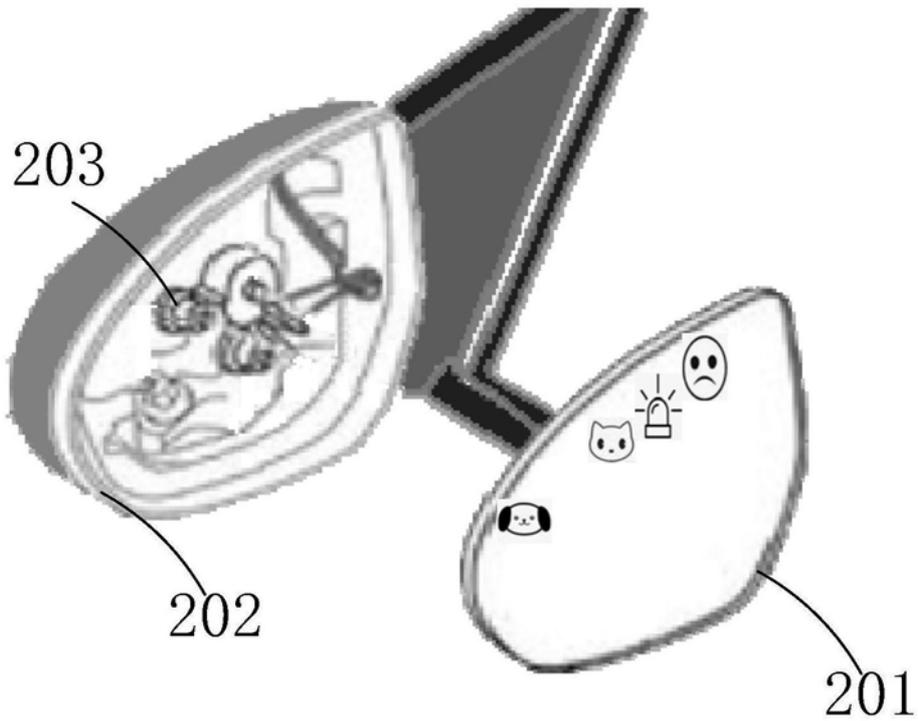


图2

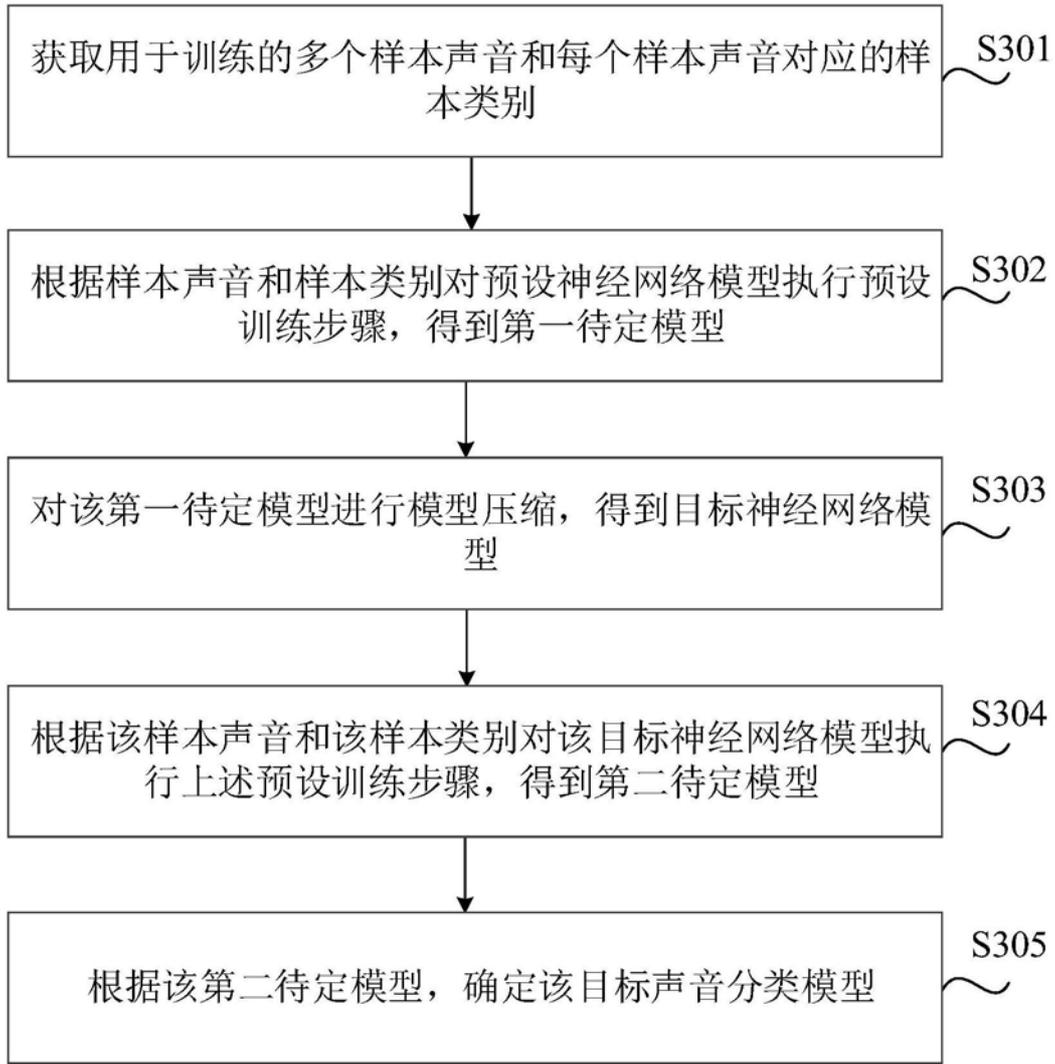


图3

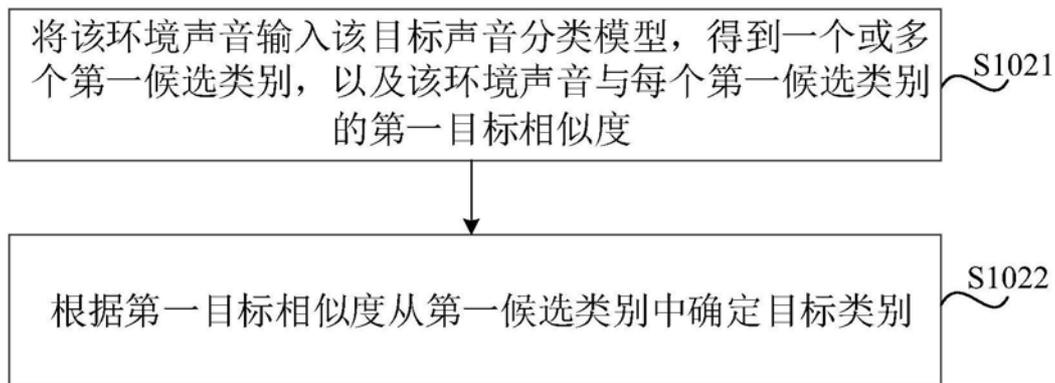


图4

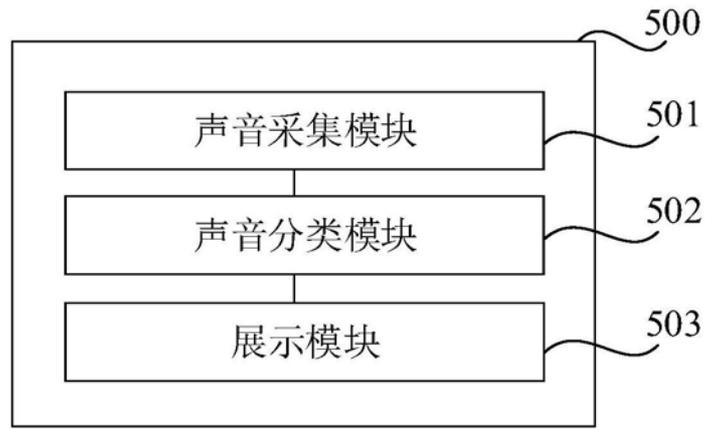


图5

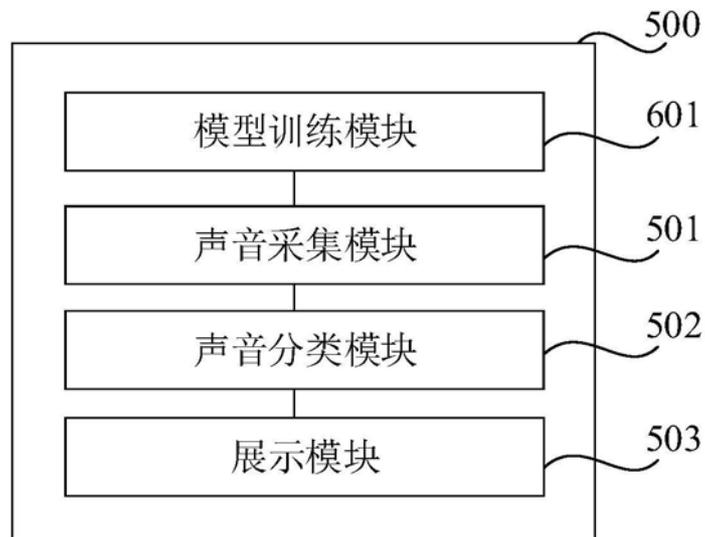


图6

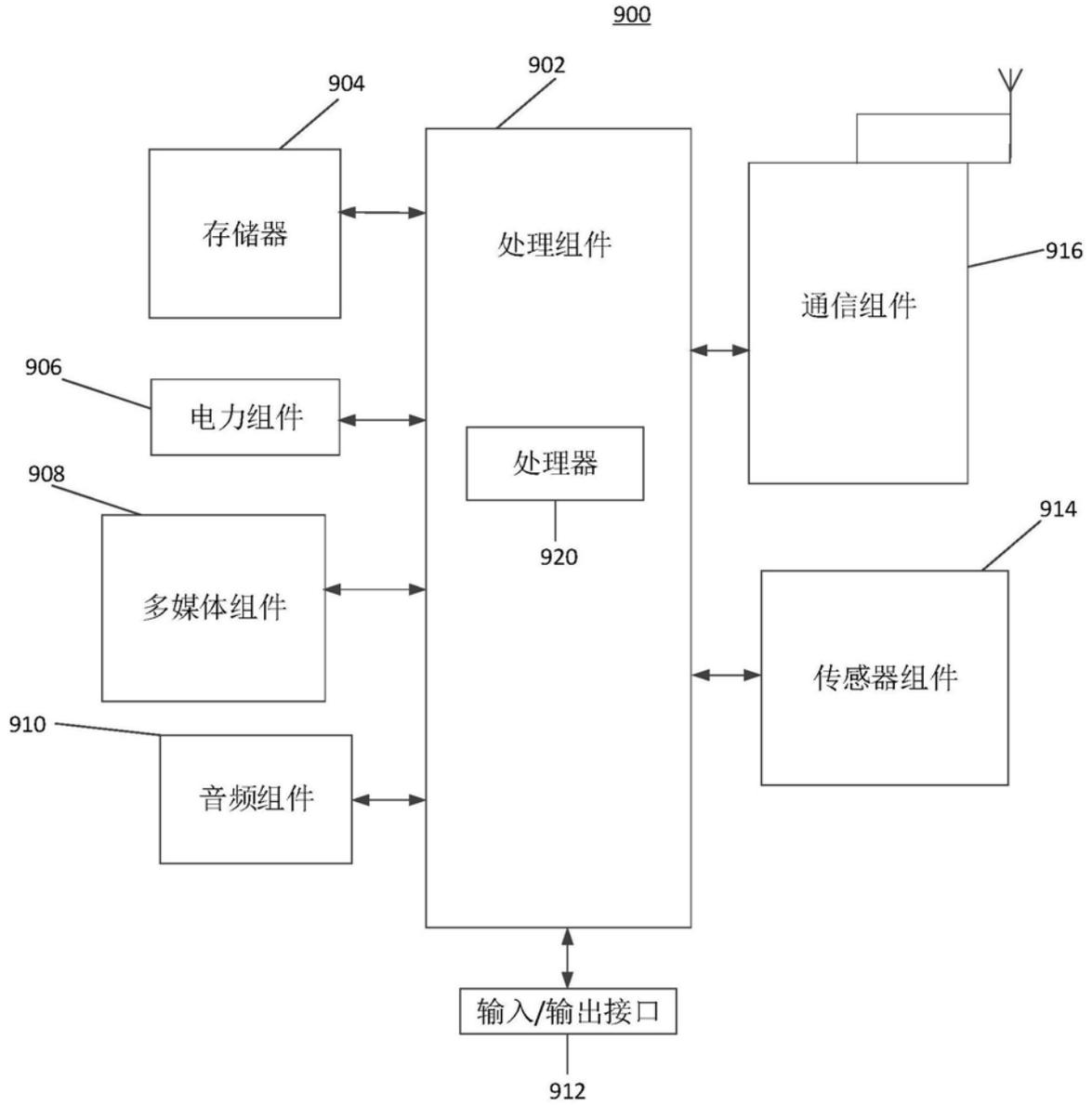


图7

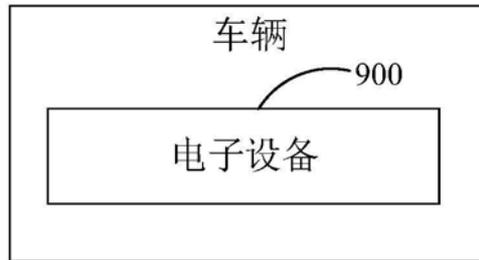


图8