

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2024年12月19日 (19.12.2024)



(10) 国际公布号
WO 2024/254861 A1

- (51) 国际专利分类号:
B60W 30/09 (2012.01)
- (21) 国际申请号: PCT/CN2023/100778
- (22) 国际申请日: 2023年6月16日 (16.06.2023)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (71) 申请人: 深圳引望智能技术有限公司 (SHENZHEN YINWANG INTELLIGENT TECHNOLOGIES CO., LTD.) [CN/CN]; 中国广东省深圳市龙岗区坂田街道万科城社区华为公司华为总部办公楼101, Guangdong 518129 (CN)。
- (72) 发明人: 苏鹏(SU, Peng); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。李世勇(LI, Shiyong); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。黄青虬(HUANG, Qingqiu); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。许春景(XU, Chunjing); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

- 叶超强(YE, Chaoqiang); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。
董思远(DONG, Siyuan); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。
- (74) 代理人: 广州三环专利商标代理有限公司 (SCIHEAD IP LAW FIRM); 中国广东省广州市越秀区先烈中路80号汇华商贸大厦1508室, Guangdong 510070 (CN)。
- (81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(54) Title: INTELLIGENT DRIVING METHOD AND APPARATUS

(54) 发明名称: 一种智能驾驶方法及装置

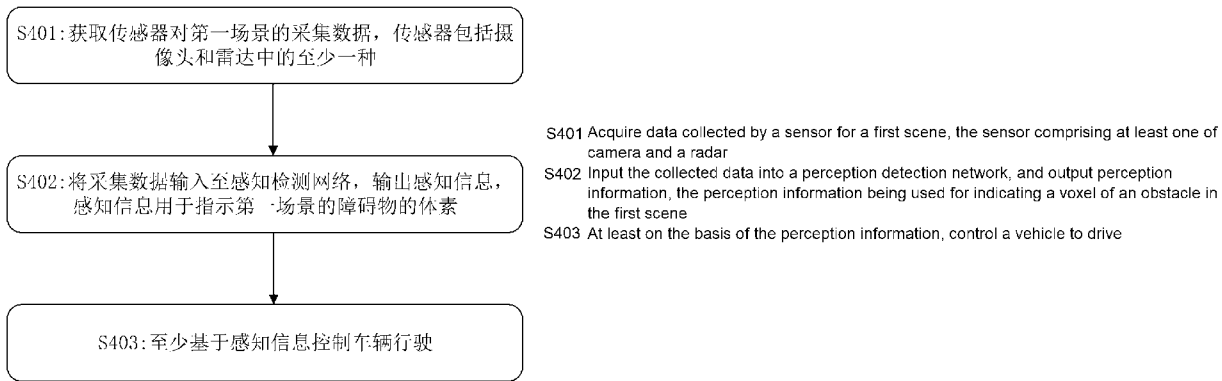


图 4

(57) Abstract: Disclosed are an intelligent driving method and apparatus, the method comprising: acquiring data collected by a sensor of a vehicle for a scene, the sensor comprising at least one of a camera and a radar; inputting the collected data into a perception detection network, and outputting perception information, the perception information being used for indicating a voxel of an obstacle in a first scene; and at least on the basis of the perception information, controlling the vehicle to drive. In this way, the perception capability of the vehicle for the surrounding environment can be enhanced, thereby improving the accuracy in detecting an obstacle, and preventing collisions.

(57) 摘要: 公开了一种智能驾驶方法及装置, 该方法包括: 获取车辆的传感器对场景的采集数据, 该传感器包括摄像头和雷达中的至少一种; 将采集数据输入至感知检测网络, 输出感知信息, 感知信息用于指示第一场景的障碍物的体素; 至少基于感知信息控制该车辆行驶。如此, 能增强车辆对周围环境的感知能力, 有利于提高障碍物的检测准确率, 避免碰撞的发生。

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

本国际公布:

— 包括国际检索报告(条约第21条(3))。

一种智能驾驶方法及装置

技术领域

本申请涉及智能驾驶领域，尤其涉及一种智能驾驶方法及装置。

背景技术

自动驾驶车辆的感知系统对周围环境的感知能力与车辆的安全行驶息息相关。

当前有车辆的感知系统采用纯视觉的方式对周围环境中的障碍物进行检测，这种方式对训练材料（例如白名单）的依赖度高，感知系统必须经过对障碍物的训练学习后才能识别出该障碍物。

当前也有车辆的感知系统采用激光雷达或毫米波雷达对周围环境中的障碍物进行检测，但该方式的检测容易受到天气的影响，例如雨、雪天气下对障碍物的检测准确率低。

发明内容

本申请公开了一种智能驾驶方法及装置，能够增强车辆对周围物体的感知能力，有利于提高障碍物的检测准确率，避免碰撞的发生。

第一方面，本申请提供了一种智能驾驶方法，所述方法包括：获取传感器对第一场景的采集数据，所述传感器包括摄像头和雷达中的至少一种；将所述采集数据输入至感知检测网络，输出感知信息，所述感知信息用于指示所述第一场景的障碍物的体素；至少基于所述感知信息控制车辆行驶。

这里，障碍物是指车辆在行驶过程中不期望与之发生碰撞的实体，该实体可以是静态的，也可以是动态的。其中，静态的实体例如可以是道路中的纸箱、道路施工牌、道路分界栏杆、土堆、轮胎、侧翻的车辆、躺着的人、动物、道路旁的建筑物、树木、停放的车辆、道路指示牌、电线杆、路边隔离带等具有体积和质量的静止物体；动态的实体例如可以是行人（例如行走的行人、骑自行车的行人等）、动物、车辆、载物的车辆（例如装载了纸箱、树枝或其他货物）等具有体积和质量的运动物体。

可以理解，本申请并不限定障碍物在物理世界中的呈现形态，以车辆为例，其可以是行驶或停车时的轮胎着地的形态，也可以是经过碰撞事故后的处于翻倒的形态，也可以是车辆的后箱装载了货物（例如树枝、纸箱等）时的形态，还可以是多节车连接时的形态。这里，也不限定障碍物为车辆时该车辆的种类，车辆的种类例如可以是轿车、货车、客车、挂车、非完整车辆、摩托车、自行车等。

这里，雷达包括激光雷达、毫米波雷达中的至少一种。

示例性地，第一场景可以理解为车辆在行驶过程中车辆上的传感器可以探测到的环境空间。可以理解，车辆在行驶过程中，每个时刻可以对应一个场景，多个时刻对应的场景包括这多个时刻中各个时刻对应的场景。

示例性地，感知检测网络是基于传感器数据集和采用 4D 重建（即包括动、静态目标的时空重建）生成的该传感器数据集对应的标签信息进行训练获得。该标签信息用于在感知检

测网络的训练过程中为感知检测网络提供其预测结果的真值信息。可以理解，4D重建可以实现在时间维度上描述三维空间内实体对象的变化。

示例性地，除了可以基于感知信息控制车辆行驶，还可以结合导航地图信息、高精地图信息、路侧设备以及周围其车辆广播的交通实况信息等中的至少一项控制车辆的行驶。

示例性地，该方法可以应用于车辆或者车辆内用于智能驾驶控制的组件（例如芯片或者集成电路）。该车辆配置有自动驾驶系统，这里，自动驾驶系统并不局限于完全自动驾驶系统、高度自动驾驶系统、有条件自动驾驶系统、或部分自动驾驶系统等，本领域技术人员可以理解，提供智能驾驶的非完全人工驾驶系统都可以涵盖在本概念之下。

上述方法中，采用纯视觉或者视觉与雷达结合的方式采集场景的数据，通过感知检测网络对场景的数据进行处理以输出指示了障碍物的体素的感知信息，能够增强车辆对周围物体的感知能力，实现了与语义类别无关的障碍物的感知，提高了场景中障碍物的检测的泛化能力以及准确率。另外，基于感知信息控制车辆的驾驶，可以提高车辆行车的安全性。

在第一方面的一种可能的实现方式中，所述方法还包括：基于所述感知信息显示所述障碍物，所述障碍物以多边形框进行标记；和/或基于所述感知信息显示所述障碍物的体素，所述障碍物的体素以多边形框进行标记。

示例性地，多边形框可以是二维，也可以是三维的。

示例性地，对障碍物或者障碍物的体素进行显示时，可以通过不同的颜色区分当前时刻下动态的障碍物和静态的障碍物，也可以通过在动态的障碍物上另外显示一个箭头来区分动态的障碍物和静态的障碍物，动态的障碍物上的箭头用于指示该动态的障碍物的运动方向。

实施上述实现方式，以多边形框标记障碍物更加贴合障碍物本身的形状，通过对障碍物和/或障碍物的体素进行呈现，用户能清晰直观地了解到车辆当前时刻对周围环境的感知情况。

在第一方面的一种可能的实现方式中，所述感知信息包括以下信息的至少一项：所述第一场景的体素的占据状态、所述第一场景的体素的速度信息、所述第一场景的体素的可见状态和所述障碍物对应的多边形框的角点信息；其中，所述障碍物对应的多边形框与所述障碍物的体素关联。

这里，体素的可见状态例如可以分为“可见”和“不可见”。例如，车辆在当前时刻所处的场景中，如场景中的某个体素在当前时刻未被车辆上的任何一个传感器（包括摄像头和雷达）的观测信号所触及，则该体素的可见状态为不可见；如果该体素被该车辆上的至少一个传感器的观测信号触及，则该体素的可见状态为可见。

这里，体素的占据状态例如可以分为“占据”或“空”（即未被占据）。例如，车辆在当前时刻所处的场景中，如场景中的某个体素在该场景所在的物理世界中对应的空间位置上存在物理实体，则该体素的占据状态为占据；如该体素在该场景所在的物理世界中对应的空间位置上不存在物理实体，则体素的占据状态为空。可以理解，空气不是物理实体。

示例性地，障碍物对应的多边形框与所述障碍物的体素关联可以理解为：障碍物对应的多边形框的角点信息基于该障碍物的体素的索引信息获得。障碍物对应的多边形框的角点信息例如可以是采用凸包算法基于障碍物的体素的索引信息计算获得。

实施上述实现方式，基于体素的可见状态可以知晓当前场景下的车辆的视线盲区，基于体素的占据状态可以知晓当前场景下车辆应避免处于“占据”状态的体素所在的区域以避免碰撞的发生，基于障碍物对应的多边形框的角点信息可以快速定位场景中的障碍物，基于体素

的速度信息以及障碍物对应的边形框的角点信息可以确定当前场景下障碍物的速度信息。

在第一方面的一种可能的实现方式中，所述感知信息还用于指示所述第一场景的路面的体素，所述至少基于所述感知信息控制车辆行驶，包括：至少根据所述感知信息，生成所述第一场景的路面几何信息；根据所述路面几何信息，调整所述车辆内的悬架。

示例性地，路面几何信息用于指示第一场景的路面状况（例如路面是否有坑洼、路面是否有凸起等）。

实施上述实现方式，基于感知信息车辆可以提前获取车辆前方的路面状况，在监测到前方路面有起伏时，车辆有足够的时间可以及时调整车辆的悬架，以使车辆在行驶过程中尽可能始终保持水平且平稳的状态，减少因路面起伏带来的振动感，提高了乘坐车辆的舒适性。

在第一方面的一种可能的实现方式中，所述至少基于所述感知信息控制车辆行驶，包括：至少根据所述感知信息，调整所述车辆的行驶路径，所述调整后的行驶路径不途径所述障碍物的体素所在的区域。

实施上述实现方式，可以避免车辆在行驶过程与障碍物发生碰撞，有利于提高车辆行驶的安全性。

在第一方面的一种可能的实现方式中，所述采集数据包括图像数据和点云数据，所述感知检测网络包括图像特征提取网络、点云特征提取网络、特征融合网络和输出网络，其中，

所述图像特征提取网络，用于提取所述图像数据的3D图像特征；

所述点云特征提取网络，用于提取所述点云数据对应的体素的点云特征；

所述特征融合网络，用于根据所述3D图像特征和所述点云数据对应的体素的点云特征进行融合，获得所述第一场景的体素的融合特征；

所述输出网络，用于处理所述第一场景的体素的融合特征并输出所述感知信息。

这里，区别于点云数据对应的体素，在本申请中，第一场景的体素是指经过特征融合网络融合后的体素。

实施上述实现方式，通过车辆上的多模态传感器的原始数据实现对周围环境的障碍物的感知，融合了不同传感器的优势（例如提供了图像的纹理语义信息、提供了点云的深度信息），有利于增强车辆对周围环境的感知能力，提高了障碍物检测的泛化能力和精度。

在第一方面的一种可能的实现方式中，所述方法还包括：将文本查询信息和所述障碍物的体素的融合特征输入至属性识别网络，输出所述障碍物的类别信息；所述文本查询信息用于请求查询类别；显示所述障碍物的类别信息；其中，所述障碍物的体素的融合特征基于所述障碍物对应的多边形框的角点信息和所述第一场景的体素的融合特征确定，所述障碍物对应的多边形框与所述障碍物的体素关联。

示例性地，文本查询信息用于请求查询Q种类别，假设某场景中障碍物的类别的数量为P，其中，Q、P为正整数且Q为大于P。也就是说，属性识别网络实际支持识别的类别的数量大于任一场景中障碍物的类别，如此，能确保属性识别网络对任一场景中的障碍物进行类别识别时避免出现遗漏。

示例性地，障碍物的体素的融合特征基于所述障碍物对应的多边形框的角点信息和所述第一场景的体素的融合特征确定是指：由于障碍物对应的多边形框的角点信息与该障碍物的体素的索引信息对应，基于该障碍物的体素的索引信息可以从所述第一场景的体素的融合特征中确定该障碍物的体素的融合特征。

这里，障碍物对应的多边形框与所述障碍物的体素关联是指：障碍物对应的多边形框的角点信息基于该障碍物的体素的索引信息获得。

示例性地，属性识别包括文本编码网络和属性解码网络，其中，文本编码网络用于提取文本查询信息的词向量特征；属性解码网络用于根据该词向量特征和障碍物的体素的融合特征输出该障碍物的类别信息。

实施上述实现方式，在感知检测网络的基础上，通过部署属性识别网络使得车辆在行驶过程中不仅可以检测到周围环境中的障碍物，还可以识别出障碍物的类别，实现了车辆不仅看的到物还能看的懂物。

在第一方面的一种可能的实现方式中，所述方法还包括：获取车辆的多条规划路径；将所述车辆的多条规划路径和所述第一场景的体素的融合特征输入至路径评估网络，输出所述多条规划路径的推荐系数和所述多条规划路径中的推荐路径，所述推荐路径与所述多条规划路径的推荐系数关联；显示所述推荐路径。

示例性地，路径评估网络包括路径编码网络、特征交互网络和评估输出网络，其中，路径编码网络用于提取多条规划路径中每条规划路径的路径特征；特征交互网络用于根据每条规划路径的路径特征和第一场景的体素的融合特征获得每条规划路径的风险特征；评估输出网络用于根据这多条规划路径的风险特征输出这多条规划路径的推荐系数和这多条规划路径中的推荐路径。

示例性地，推荐路径为多条规划路径中最高推荐系数对应的规划路径。

示例性地，规划路径的推荐系数可以基于该规划路径的风险系数、舒适度和通行效率中的至少一项获得。其中，该规划路径的风险系数与该规划路径与障碍物（包括可见的障碍物以及当前处于盲区的障碍物）的距离、该规划路径与道路中其他交通参与者的路径是否有冲突（例如当前时刻或者未来时刻是否会发生碰撞）等因素中的至少一项有关，该规划路径的通行效率与该规划路径的长度、该规划路径对应的预估通行时长、该规划路径途径的红绿灯的数量、该规划路径所在的可行驶区域的面积等因素中的至少一项有关，该规划路径的舒适度与该规划路径的转向加速度的大小及转向频率、该规划路径的加速度的变化率、该规划路径途径的路面的平整度、该规划路径途径的红绿灯的数量、该规划路径所在的道路的类型、该规划路径途径区域是否阴凉等因素中的至少一项有关。

示例性地，在其他因素不变的情况下，该规划路径的风险系数越低，则该规划路径的推荐系数越高；在其他因素不变的情况下，该规划路径的舒适度越高，则该规划路径的推荐系数越高；在其他因素不变的情况下，该规划路径的通行效率越高，则该规划路径的推荐系数越高。

实施上述实现方式，在感知检测网络的基础上，通过部署属性识别网络可以实现路径推荐，有利于提高车辆行车的安全性和舒适性。

第二方面，本申请提供了一种用于智能驾驶的系统，所述系统包括：感知检测网络，用于根据传感器对第一场景的采集数据输出感知信息，所述感知信息用于指示所述第一场景的障碍物的体素；所述传感器包括摄像头和雷达中的至少一种；属性识别网络，用于根据文本查询信息和所述障碍物的体素的融合特征输出所述障碍物的类别信息，所述障碍物的体素的融合特征基于所述障碍物对应的多边形框的角点信息和所述第一场景的体素的融合特征确定，所述障碍物对应的多边形框与所述障碍物的体素关联，所述第一场景的体素的融合特征为所

述感知检测网络基于从所述采集数据中提取的 3D 图像特征和体素的点云特征中的至少一项进行时间和/或空间上的融合获得；路径评估网络，用于根据多条规划路径和所述第一场景的体素的融合特征输出所述多条规划路径的推荐系数和所述多条规划路径中的推荐路径，所述推荐路径与所述多条规划路径的推荐系数关联。

示例性地，该系统可以部署在车辆或者车辆内用于智能驾驶控制的组件，该组件例如可以是芯片或者集成电路。车辆具体可以参考上述第一方面对车辆的叙述，在此不再赘述。

上述方法中，通过感知检测网络可以增强用于智能驾驶的系统对周围环境的感知能力，从而能避免该系统的部署端与障碍物发生碰撞，提高了该系统的安全性；通过属性识别网络使得该系统在感知到障碍物的基础上还能识别出障碍物的类别，提高了该系统的智能性；通过路径评估网络能实现低风险路径的推荐，为智能出行提供了方便。

下述第二方面的任一特征的有益效果可以参考上述第一方面相应特征的有益效果的描述，在此不再赘述。

在第二方面的一种可能的实现方式中，所述感知信息包括以下信息的至少一项：所述第一场景的体素的占据状态、所述第一场景的体素的速度信息、所述第一场景的体素的可见状态和所述障碍物对应的多边形框的角点信息；其中，所述障碍物对应的多边形框与所述障碍物的体素关联。

在第二方面的一种可能的实现方式中，所述采集数据包括图像数据和点云数据，所述感知检测网络包括图像特征提取网络、点云特征提取网络、特征融合网络和输出网络，其中，

所述图像特征提取网络，用于提取所述图像数据的 3D 图像特征；

所述点云特征提取网络，用于提取所述点云数据对应的体素的点云特征；

所述特征融合网络，用于根据所述 3D 图像特征和所述点云数据对应的体素的点云特征进行融合，获得所述第一场景的体素的融合特征；

所述输出网络，用于处理所述第一场景的体素的融合特征并输出所述感知信息。

在第二方面的一种可能的实现方式中，所述属性识别网络包括文本编码网络和属性解码网络，其中，所述文本编码网络，用于提取文本查询信息的词向量特征；所述属性解码网络，用于根据所述词向量特征和所述障碍物的体素的融合特征输出所述障碍物的类别信息。

在第二方面的一种可能的实现方式中，所述路径评估网络包括路径编码网络、特征交互网络和评估输出网络，其中，

所述路径编码网络，用于提取多条规划路径中每条规划路径的路径特征；

所述特征交互网络，用于根据每条规划路径的路径特征和所述第一场景的体素的融合特征获得所述每条规划路径的风险特征；

所述评估输出网络，用于根据所述多条规划路径的风险特征输出所述多条规划路径的推荐系数和所述多条规划路径中的推荐路径。

第三方面，本申请提供了一种用于智能驾驶的装置，该装置包括：接收单元，用于获取传感器对第一场景的采集数据，所述传感器包括摄像头和雷达中的至少一种；处理单元，用于将所述采集数据输入至感知检测网络，输出感知信息，所述感知信息用于指示所述第一场景的障碍物的体素；该处理单元，还用于至少基于所述感知信息控制车辆行驶。

在第三方面的一种可能的实现方式中，该装置还包括显示单元，显示单元用于基于所述感知信息显示所述障碍物，所述障碍物以多边形框进行标记；和/或基于所述感知信息显示所

述障碍物的体素。

在第三方面的一种可能的实现方式中，所述感知信息包括以下信息的至少一项：所述第一场景的体素的占据状态、所述第一场景的体素的速度信息、所述第一场景的体素的可见状态和所述障碍物对应的多边形框的角点信息；其中，所述障碍物对应的多边形框与所述障碍物的体素关联。

在第三方面的一种可能的实现方式中，所述感知信息还用于指示所述第一场景的路面的体素，所述处理单元具体用于：至少根据所述感知信息，生成所述第一场景的路面几何信息；根据所述路面几何信息，调整所述车辆内的悬架。

在第三方面的一种可能的实现方式中，所述处理单元具体用于：至少根据所述感知信息，调整所述车辆的行驶路径，所述调整后的行驶路径不途径所述障碍物的体素所在的区域。

在第三方面的一种可能的实现方式中，所述采集数据包括图像数据和点云数据，所述感知检测网络包括图像特征提取网络、点云特征提取网络、特征融合网络和输出网络，其中，

所述图像特征提取网络，用于提取所述图像数据的3D图像特征；

所述点云特征提取网络，用于提取所述点云数据对应的体素的点云特征；

所述特征融合网络，用于根据所述3D图像特征和所述点云数据对应的体素的点云特征进行融合，获得所述第一场景的体素的融合特征；

所述输出网络，用于处理所述第一场景的体素的融合特征并输出所述感知信息。

在第三方面的一种可能的实现方式中，处理单元还用于：将文本查询信息和所述障碍物的体素的融合特征输入至属性识别网络，输出所述障碍物的类别信息；所述文本查询信息用于请求查询类别；显示单元还用于显示所述障碍物的类别信息；其中，所述障碍物的体素的融合特征基于所述障碍物对应的多边形框的角点信息和所述第一场景的体素的融合特征确定，所述障碍物对应的多边形框与所述障碍物的体素关联。

在第三方面的一种可能的实现方式中，接收单元还用于：获取车辆的多条规划路径；处理单元还用于：将所述车辆的多条规划路径和所述第一场景的体素的融合特征输入至路径评估网络，输出所述多条规划路径的推荐系数和所述多条规划路径中的推荐路径，所述推荐路径与所述多条规划路径的推荐系数关联；显示单元还用于：显示所述推荐路径。

第四方面，本申请提供了一种用于智能驾驶的装置，该装置包括处理器和存储器，其中，存储器用于存储程序指令；所述处理器调用所述存储器中的程序指令，使得装置执行第一方面或者第一方面的任一可能的实现方式中的方法。

第五方面，本申请提供了一种车辆，该车辆包括如上述第二方面或者第二方面的任一可能的实现方式的系统，或者包括如上述第三方面或者第三方面的任一可能的实现方式的装置，或者包括上述第四方面的装置。

第六方面，本申请提供了一种计算机可读存储介质，包括计算机指令，当所述计算机指令在被处理器运行时，实现上述第一方面或者第一方面的任一可能的实现方式中的方法。

第七方面，本申请提供了一种计算机程序产品，当该计算机程序产品被处理器执行时，实现上述第一方面或者第一方面的任一可能的实施例中的所述方法。该计算机程序产品，例如可以为一个软件安装包，在需要使用上述第一方面的任一种可能的设计提供的方法的情况下，可以下载该计算机程序产品并在处理器上执行该计算机程序产品，以实现第一方面或者第一方面的任一可能的实施例中的所述方法。

附图说明

- 图 1 是本申请实施例提供的一种通信系统的示意图；
图 2 是本申请实施例提供的一种用于智能驾驶的感知模型的系统示意图；
图 3 是本申请实施例提供的一种感知检测网络的特征提取的示意图；
图 4 是本申请实施例提供的一种智能驾驶方法的流程图；
图 5 是本申请实施例提供的一些场景示意图；
图 6A 是本申请实施例提供的一种以多边形框标记场景中的障碍物的示意图；
图 6B 是本申请实施例提供的一种障碍物的体素的显示示意图；
图 6C 是本申请实施例提供的一种显示了场景中路面的体素的示意图；
图 7A 是本申请实施例提供的一种感知检测网络的训练方法的流程图；
图 7B 是本申请实施例提供的一种感知检测网络的训练过程示意图；
图 8 是本申请实施例提供的一种芯片硬件结构示意图；
图 9A 是本申请实施例提供的一种计算装置的结构示意图；
图 9B 是本申请实施例提供的一种训练装置的结构示意图；
图 10 是本申请实施例提供的一种处理设备的结构示意图。

具体实施方式

需要说明的是，本申请中采用诸如“第一”、“第二”的前缀词，仅仅为了区分不同的描述对象，对被描述对象的位置、顺序、优先级、数量或内容等没有任何限定作用。例如，被描述对象为“字段”，则“第一字段”和“第二字段”中“字段”之前的序数词并不限制“字段”之间的位置或顺序，“第一”和“第二”并不限制其修饰的“字段”是否在同一个消息中，也不限制“第一字段”和“第二字段”的先后顺序。再如，被描述对象为“等级”，则“第一等级”和“第二等级”中“等级”之前的序数词并不限制“等级”之间的优先级。再如，被描述对象的数量并不受前缀词的限制，可以是一个或者多个，以“第一设备”为例，其中“设备”的数量可以是一个或者多个。此外，不同前缀词修饰的对象可以相同或不同，例如，被描述对象为“设备”，则“第一设备”和“第二设备”可以是同一个设备、相同类型的设备或者不同类型的设备；再如，被描述对象为“信息”，则“第一信息”和“第二信息”可以是相同内容的信息或者不同内容的信息。总之，本申请实施例中用于区分描述对象的前缀词的使用不构成对所描述对象的限制，对所描述对象的陈述参见权利要求或实施例中上下文的描述，不应因为使用这种前缀词而构成多余的限制。

需要说明的是，本申请实施例中采用诸如“a1、a2、……和 an 中的至少一项（或至少一个）”等的描述方式，包括了 a1、a2、……和 an 中任意一个单独存在的情况，也包括了 a1、a2、……和 an 中任意多个的任意组合情况，每种情况可以单独存在。例如，“a、b 和 c 中的至少一项”的描述方式，包括了单独 a、单独 b、单独 c、a 和 b 组合、a 和 c 组合、b 和 c 组合，或 abc 三者组合的情况。

为了便于理解，下面先对本申请实施例可能涉及的相关术语等进行介绍。

(1) 自动驾驶

自动驾驶又可以称为智能驾驶或辅助驾驶，是车辆智能化发展的重要方向，随着感知技术的发展以及芯片能力的提升，智能驾驶为人们提供了越来越多的丰富的驾驶功能，逐渐实现不同级别的驾驶体验。自动机工程师学会（society of automotive engineers, SAE）提供了一

种驾驶自动化分级标准，包括驾驶等级 L0 至 L5，其中 L0 级为无自动化，由人类驾驶者全权操作车辆，在行驶过程中可以得到驾驶系统的警告或辅助，例如自动紧急制动（autonomous emergency braking, AEB），盲点检测（blind spot monitoring, BSM）或车道偏离报警（lane departure warning, LDW）等。L1 级为驾驶支援，驾驶操作由人类驾驶者和驾驶系统共同完成，驾驶系统可以通过驾驶环境对方向盘或加减速操作提供驾驶支援，其他的驾驶操作由人类驾驶员进行，例如自适应巡航控制（adaptive cruise control, ACC）或车道保持辅助/支持（lane keep assistance/support, LKA/LKS）等；L2 级为部分自动化，通过驾驶环境对方向盘和加减速中的多项提供驾驶支援，其他的驾驶动作由人类驾驶员进行，例如结合了自适应巡航控制（adaptive cruise control, ACC）和车道保持辅助（lane keep assistance, LKA）的跟车功能；L3 级为有条件自动化，可以由驾驶系统完成所有的驾驶操作，但人类驾驶员需要在适当的时候应答驾驶系统的请求，即人类驾驶员需要做好接管驾驶系统的准备；L4 级为高度自动化，可以由驾驶系统完成所有的驾驶操作，人类驾驶员不一定需要对驾驶系统的请求作出应答，例如在道路和环境条件允许的情况下（比如封闭的园区、高速公路、城市道路或固定的行车线路等）人类驾驶员可以不接管驾驶；L5 级为完全自动化，在各种人类驾驶员可以应对的道路和环境条件下的驾驶操作均可以由驾驶系统自主完成。可见，L0 至 L2 的级别，驾驶系统主要为驾驶员提供支持，驾驶员仍然需要做好驾驶监督，根据需要进行转向、制动或加速以保证安全。L3 至 L5 级别，驾驶系统可以代替驾驶员完成所有的驾驶操作，L3 级别下，驾驶员要做好接管驾驶的准备，L4 和 L5 级别驾驶系统可以实现部分条件和所有条件下的完全驾驶，驾驶员可以选择是否接管。

以上分级是一种示例，随着技术的演进或者在不同国家或地区的规定不同，以上分级可以变化，例如，中国工业和信息化部提出的车辆自动化分级包括在车辆驾驶自动化的 6 个等级，其中 0-2 级为驾驶辅助，系统辅助人类执行动态驾驶任务，驾驶主体仍为驾驶员；3-5 级为自动驾驶，系统在设计运行条件下代替人类执行动态驾驶任务，当功能激活时，驾驶主体是系统。各级名称及定义如下：0 级驾驶自动化（应急辅助，emergency assistance）系统不能持续执行动态驾驶任务中的车辆横向或纵向运动控制，但具备持续执行动态驾驶任务中的部分目标和事件探测与响应的能力。1 级驾驶自动化（部分驾驶辅助，partial driver assistance）系统在其设计运行条件（或称为设计运行范围 ODD）下持续地执行动态驾驶任务中的车辆横向或纵向运动控制，且具备与所执行的车辆横向或纵向运动控制相适应的部分目标和事件探测与响应的能力。2 级驾驶自动化（组合驾驶辅助，combined driver assistance）系统在其设计运行条件下持续地执行动态驾驶任务中的车辆横向和纵向运动控制，且具备与所执行的车辆横向和纵向运动控制相适应的部分目标和事件探测与响应的能力。3 级驾驶自动化（有条件自动驾驶，conditionally automated driving）系统在其设计运行条件下持续地执行全部动态驾驶任务。4 级驾驶自动化（高度自动驾驶，highly automated driving）系统在其设计运行条件下持续地执行全部动态驾驶任务并自动执行最小风险策略。5 级驾驶自动化（完全自动驾驶，fully automated driving）系统在任何可行驶条件下持续地执行全部动态驾驶任务并自动执行最小风险策略。其中，横向控制主要用于车辆转向的控制，例如，控制方向盘扭矩或角度以控制车辆的方向；纵向控制主要用于车辆的速度控制，例如控制制动踏板、加速踏板、或档位等以控制车辆的加/减速、刹车等。

无论采用何种分级方式，本申请实施例的描述可以适用于以上需要部分或全部参与车辆

驾驶的自动驾驶系统。

(2) 障碍物

在本申请实施例中，障碍物是指车辆在行驶过程中不期望与之发生碰撞的实体，该实体可以是静态的，也可以是动态的。其中，静态的实体例如可以是道路中的纸箱、道路施工牌、道路分界栏杆、土堆、轮胎、侧翻的车辆、躺着的人、动物、道路旁的建筑物、树木、停放的车辆、道路指示牌、电线杆、路边隔离带等具有体积和质量的静止物体；动态的实体例如可以是行人（例如行走的行人、骑自行车的行人等）、动物、车辆、载物的车辆（例如装载了纸箱、树枝或其他货物）等具有体积和质量的运动物体。

(3) 场景

场景是指车辆在行驶过程中车辆上的传感器可以探测到的环境空间。可以理解，车辆在行驶过程中，每个时刻对应一个场景，多个时刻对应的场景包括这多个时刻中各个时刻对应的场景。

(4) 体素

体素(voxel)，也可以称为立体像素或体积元素。体素是三维空间上分割的最小单位，类似于二维空间的最小单位-像素。通过体素可以对3D空间进行网格划分并赋予每个网格特征，在此情况下，体素表示三维空间中规则网格上的值，基于体素相对于其他体素的位置可以推断该体素的定位。

下面将结合附图，对本申请实施例中的技术方案进行描述。

参见图1，图1是本申请实施例提供的一种通信系统的示意图。如图1所示，该系统包括网络侧设备和车辆，其中，网络侧设备与车辆之间以无线的方式进行通信。

这里，网络侧设备是具有计算能力的设备。网络侧设备例如可以是部署在网络侧的服务器（例如用于智能驾驶处理的服务器），或者为该服务器中的组件或者芯片。在一些可能的实施例中，网络侧设备也可以是由多个服务器组成的系统级设备或者计算设备集群。网络侧设备可以部署在云环境或者边缘环境中，本申请实施例不做具体限定。

这里，车辆是指配置有自动驾驶系统的车辆。自动驾驶系统并不局限于完全自动驾驶系统、高度自动驾驶系统、有条件自动驾驶系统、或部分自动驾驶系统等，本领域技术人员可以理解，提供智能驾驶的非完全人工驾驶系统都可以涵盖在本概念之下。

示例性地，依据车辆的动力来源的不同，车辆例如可以是新能源车辆或传统车辆等，其中，传统车辆是指燃油类车辆，例如可以是汽油车辆、柴油车辆等，新能源车辆例如可以是电动车辆（electric vehicle, EV）、混合动力车辆（hybrid electric vehicle, HEV）、增程式电动车辆（range extended EV）、插电式混合动力车辆（Plug-in HEV）、燃料电池车辆或其他新能源车辆，在此不作具体限定。

其中，车辆上部署有摄像头和雷达，其中，摄像头用于采集车辆当前周围环境的图像数据，雷达用于采集车辆当前周围环境的点云数据。雷达包括激光雷达Lidar、毫米波雷达Radar等中的至少一种。基于摄像头在车辆的安装位置，摄像头例如可以分为前视摄像头、环视摄像头、后视摄像头和侧视摄像头等；基于摄像头的结构划分，摄像头例如可以分为单目摄像头、双目摄像头、广角摄像头等。这里，本申请实施例不限定车辆配置的摄像头的数量，出于安全考虑，车辆上的摄像头需要能采集到车身周围360度的图像数据。

示例性地，网络侧设备上部署有感知模型，网络侧设备使用训练数据对感知模型进行训练，其中，训练数据包括从数据源设备（例如，采集车队）获取传感器数据，传感器数据包括车载摄像头采集的图像数据以及车载雷达采集的点云数据。网络侧设备将感知模型训练好后可以将训练好后的感知模型提供给车辆使用。感知模型的训练过程具体可参考下述方法实施例中相应内容的叙述，在此不再赘述。

进一步地，车辆可以从网络侧设备获取感知模型（即训练好的感知模型）。在车辆的行驶过程中，车辆通过自身搭载的传感器（例如摄像头、雷达等）对距离自车一定范围内的环境（或称为场景）进行数据采集获得采集数据，该采集数据例如包括针对该场景采集的图像数据以及点云数据，车辆使用该感知模型对采集数据进行处理以输出该场景的感知信息，感知信息用于指示该场景内的障碍物的体素，车辆至少基于感知信息可以控制自身的行驶。

感知模型具体可参考下述图 2 实施例的相关叙述，在此不再赘述。

在图 1 所示系统，网络侧设备与车辆之间的通信可使用蜂窝通信技术，例如 2G 蜂窝通信，例如全球移动通信系统（global system for mobile communication, GSM）、通用分组无线业务（general packet radio service, GPRS）；或者 3G 蜂窝通信，例如宽带码分多址（wideband code division multiple access, WCDMA）、时分同步码分多址接入（time division-synchronous code division multiple access, TS-SCDMA）、码分多址接入（code division multiple access, CDMA），或者 4G 蜂窝通信，例如长期演进（long term evolution, LTE）、LTE-车联网无线通信技术（vehicle to everything, V2X），PC5 通信，或者 5G 蜂窝通信，例如新空口（new radio, NR）-V2X PC5 通信，或者其他演进的蜂窝通信技术。无线通信系统也可利用非蜂窝通信技术，如 Wi-Fi 与无线局域网（wireless local area network, WLAN）通信，在此不作具体限定。

可以理解，图 1 仅为示例性架构图，但不限定图 1 所示系统包括的网元的数量。虽然图 1 未示出，但除图 1 所示的功能实体外，图 1 还可以包括其他功能实体。另外，本申请实施例提供的方法可以应用于图 1 所示的通信系统，当然本申请实施例提供的方法也可以适用其他通信系统，本申请实施例对此不予限制。

参见图 2，图 2 是本申请实施例提供的一种用于智能驾驶的感知模型的系统示意图。

在图 2 中，感知模型包括感知检测网络，感知检测网络用于根据传感对场景的采集数据（例如包括图像数据和点云数据）输出感知信息，感知信息用于指示该场景的障碍物的体素。在一些可能的实施例中，感知信息还用于指示该场景内的路面的体素。感知信息可以用于辅助车辆的驾驶。

下面介绍感知检测网络的框架。

一种实现方式中，在采集数据包括图像数据和点云数据的情况下，感知检测网络包括图像特征提取网络、点云特征提取网络、特征融合网络和输出网络，其中，图像特征提取网络用于从图像数据中提取该图像数据的 3D 图像特征并将该特征输出至特征融合网络，点云特征提取网络用于从点云数据中提取该点云数据对应的体素的点云特征并将该特征输出至特征融合网络，特征融合网络用于对图像数据的 3D 图像特征和点云数据对应的体素的点云特征进行融合，获得对应场景的体素的融合特征并将该特征输出至输出网络，输出网络根据该对应场景的体素的融合特征进行预测，输出该场景的感知信息。

示例性地，图 2 中的特征融合网络可以只执行空间上的特征融合。例如，图像数据为第

t时刻摄像头采集的图像，点云数据为第t时刻雷达采集的数据，则特征融合网络只需对第t时刻该图像数据的3D图像特征和第t时刻该点云数据对应的体素的点云特征进行空间上的融合。

示例性地，图2中的特征融合网络可以执行空间和时间上的特征融合。例如，图像数据为摄像头在n个时刻采集的图像数据，点云数据为雷达在这n个时刻采集的点云数据，则特征融合网络可以先对n个时刻中每个时刻对应的3D图像特征和该时刻对应的体素的点云特征进行空间上的融合获得每个时刻对应的体素的空间融合特征，再将n个时刻中各时刻对应的体素的空间融合特征进行时间上的融合。

在一些可能的实施例中，如采集数据只包括图像数据，则图2所示感知检测网络内的点云特征提取模块可以缺省，如果图像数据为摄像头在n个时刻采集的图像数据，则特征融合网络对n个时刻中各时刻对应的3D图像特征进行时间上的融合即可。

在一些可能的实施例中，如果采集数据只包括点云数据，则图2所示感知检测网络内的图像特征提取模块可以缺省，如点云数据为雷达在n个时刻采集的点云数据，则特征融合网络对n个时刻中各时刻对应的体素的点云特征进行时间上的融合即可。

示例性地，特征融合网络可以采用循环神经网络（recurrent neural network, RNN）或者循环卷积神经网络（recurrent CNN, RCNN）的网络结构，CNN例如可以是长期短记忆网络（long short-term memory networks, LSTM）、门控循环单元网络（gated recurrent unit network, GRU）等。

进一步地，图像特征提取网络包括相机主干网络和立体转换网络，其中，相机主干网络用于提取图像数据的2D图像特征，立体转换网络用于将图像数据的2D图像特征转换为图像数据的3D图像特征。这里，立体转换网络可以实现将2D图像特征转换为车体坐标系下的3D图像特征，而从雷达的点云数据中提取出的特征本身就是车体坐标下的3D特征，如此，方便后续特征融合网络对来自不同传感器的特征进行特征融合，有利于消除多模态传感器之间的异构差异。

示例性地，图像数据的2D图像特征包括但不限于该图像数据的颜色特征、形状特征、纹理特征和空间关系特征等。

示例性地，相机主干网络可以采用卷积神经网络（convolutional neural networks, CNN）（例如残差网络 Resnet）、变换 transformer 网络、视觉变换（vision transformer, ViT）网络、或者其他主干网络的网络结构。立体转换网络可以采用变换 transformer 网络或者举起投掷射击（lift-splat-shoot, LSS）网络的网络结构。

进一步地，点云特征提取网络包括雷达编码网络和点主干网络，其中，雷达编码网络用于对点云数据进行体素化处理以建立点云数据中的点与体素之间的对应关系，从而获得点云数据对应的体素的特征，点主干网络用于根据点云数据对应的体素的特征提取点云数据对应的体素的点云特征（即3D特征）。在一些可能的实施例中，雷达编码网络和点主干网络可以合并为一个网络，以提取点云数据对应的体素的点云特征，在此不作具体限定。在一些可能的实施例中，在算力支持的情况下，雷达编码网络和点主干网络也可以合并为一个网络，该网络用于提取点云数据对应的体素的点云特征。

示例性地，雷达编码网络可以采用体素特征编码（voxel feature encoding, VFE）网络或者支柱特征编码（pillar feature encoding, PFE）网络等的网络结构。点主干网络可以采用卷

积神经网络（例如 U-Net）或变换 transformer 网络等的网络结构。

输出网络即为感知检测网络的检测头。输出网络包括至少一个头网络，输出网络中头网络的数量基于输出网络输出的感知信息中预测结果的种类数确定。示例性地，如图 2 所示，感知信息包括体素的占据状态、体素的速度信息、体素的可见状态和障碍物对应的多边形框的角点信息，其中，障碍物对应的多边形框与障碍物的体素关联，由此可以看出，感知信息中包含 4 种预测结果，故输出网络包括四个头网络，分别为头网络 1、头网络 2、头网络 3 和头网络 4、其中，头网络 1 用于输出障碍物对应的多边形框的角点信息，头网络 2 用于输出体素的占据状态，头网络 3 用于输出体素的速度信息，头网络 4 用于输出体素的可见状态。

这里，体素的可见状态是指：车辆在当前时刻所处的场景中，如场景中的某个体素在当前时刻未被车辆上的任何一个传感器（包括摄像头和雷达）的观测信号所触及，则该体素的可见状态为不可见；如果该体素被至少一个传感器的观测信号触及，则该体素的可见状态为可见。

示例性地，体素的占据状态是指：车辆在当前时刻所处的场景中，如场景中的某个体素在该场景所在的物理世界中对应的空间位置上存在实体，则该体素的占据状态为占据；如该体素在该场景所在的物理世界中对应的空间位置上不存在实体，则体素的占据状态为空（即未被占据）。这里，实体可以理解为具有一定体积和质量的物体。可以理解，空气不是实体。

示例性地，输出网络中的任一头网络可以采用卷积神经网络 CNN 或者变换 transformer 网络的网络结构。这里，输出网络中不同头网络的内部网络结构可以相同也可以不同，可以理解，不同头网络对相同的输入特征的处理方式不同。

在一些可能的实施例，为了减少算力的消耗，在图 2 所示的感知检测网络中，还可以在特征融合网络与输出网络之间设置神经采样网络，即特征融合网络将该场景的体素的融合特征输出至神经采样网络，神经采样网络根据场景的体素中所在区域的重要度采用不同分辨率对该场景的体素的融合特征进行处理，例如，该场景中区域一的重要度大于该场景中区域二的重要度，则以第一分辨率对该区域一中体素的融合特征进行处理，以第二分辨率对区域二中的体素的融合特征进行处理，其中，第一分辨率大于第二分辨率。如此，神经采样网络可以实现对场景中关键区域的体素进行细粒度的处理，以及对场景中非关键区域的体素进行粗粒度的处理，如此，可以大大节省算力，有利于提高感知检测网络的数据处理效率，也有利于降低硬件的部署成本。

示例性地，上述区域一和区域二满足下述条件中的至少一项时，区域一的重要度大于区域二的重要度：

- (1) 区域一距离车辆的距离小于区域二距离车辆的距离；
- (2) 区域一内障碍物的数量大于区域二内障碍物的数量；
- (3) 区域一内动态的障碍物的数量大于区域二内动态的障碍物的数量；和
- (4) 区域一内障碍物的体积大于区域二内障碍物的体积。

示例性地，神经采样网络可以采用神经网络、多层感知器（multi-layer perceptron, MLP）或者变换 transformer 网络的网络结构。

为了更清楚地显示感知检测网络的特征提取的流程，参见图 3，图 3 是本申请实施例提供的一种感知检测网络的特征提取的示意图。在图 3 中，基于 n 个摄像头采集的图像数据经过上述相机主干网络可以提取出图像数据的 2D 图像特征，图像数据的 2D 图像特征经过立体

转换网络可以提取出该图像数据的 3D 图像特征，雷达采集的点云数据经过雷达编码网络可以提取出点云数据对应的体素的特征（即 3D 特征），点云数据对应的体素的特征经过点主干网络可以提取出点云数据对应的体素的点云特征（即 3D 特征），上述图像数据的 3D 图像特征和点云数据对应的体素的点云特征经过特征融合网络的融合输出体素的融合特征，最后，体素的融合特征经过上述输出网络分别输出体素的占据状态、体素的速度信息、体素的可见状态和障碍物对应的多边形框的角点信息。

可以理解，图 3 只是对感知检测网络的特征提取过程的一种示例，并不限定感知检测网络中特征提取流程仅为图 3 所示。

在一些可能的实施例中，感知模型还包括属性识别网络，属性识别网络可以用于识别障碍物的类别。示例性地，属性识别网络用于根据文本查询信息和障碍物的体素的融合特征输出障碍物的类别信息，其中，文本查询信息用于请求查询类别，障碍物的体素的融合特征基于障碍物对应的多边形框的角点信息和场景的体素的融合特征确定，障碍物对应的多边形框与该障碍物的体素关联。由图 2 可以知晓，障碍物对应的多边形框的角点信息来自感知检测网络中的输出网络（具体为输出网络中的头网络 1），场景的体素的融合特征为感知检测网络中的特征融合网络的输出。

这里，障碍物对应的多边形框与该障碍物的体素关联可以理解为：障碍物对应的多边形框的角点信息基于该障碍物的体素的索引信息获得。障碍物对应的多边形框的角点信息例如可以是图 2 中的头网络 1 基于学习到的规则对该障碍物的体素的索引信息进行预测获得，也可以是采用凸包算法基于障碍物的体素的索引信息计算获得该障碍物对应的多边形框的角点信息，在此不作具体限定。

这里，障碍物对应的多边形框可以是二维的，也可以是三维的，在此不作具体限定。

一种实现方式中，属性识别网络包括文本编码网络和属性解码网络，其中，文本编码网络，用于提取文本查询信息的词向量特征；属性解码网络，用于根据该词向量特征和障碍物的体素的融合特征输出该障碍物的类别信息。

这里，文本查询信息用于请求查询 Q 种类别，假设某场景中障碍物的类别的数量为 P ，其中， Q 、 P 为正整数且 Q 为大于 P 。也就是说，属性识别网络实际支持识别的类别的数量大于任一场景中障碍物的类别，如此，能确保属性识别网络对任一场景中的障碍物进行类别识别时避免出现遗漏。

例如，在类别的推理过程中，文本查询信息例如包括“是车吗”、“是行人吗”、“是电线杆吗”、“是道路指示牌吗”、“是道路分界栏杆吗”、……等 K 条文本查询信息，属性识别网络中的文本编码网络对 K 条文本查询信息进行特征提取获得每条文本查询信息对应的词向量特征，其中，每条文本查询信息对应的词向量特征可以表征该文本查询信息指示的类别的图像语义特征，以属性识别网络中的属性解码网络对障碍物 1 的类型识别为例，障碍物 1 为场景中的任意一个障碍物，属性解码网络将障碍物 1 的体素的融合特征与 K 条文本查询信息中每条文本查询信息对应的词向量特征进行相似度计算，确定与障碍物 1 的体素的融合特征相似度最高的词向量特征对应的类别为障碍物 1 的类别，从而可以输出该障碍物 1 的类别信息。

这里，障碍物 1 的体素的融合特征基于障碍物 1 对应的多边形框的角点信息和场景的体素的融合特征确定可以是：由于障碍物 1 对应的多边形框的角点信息与障碍物 1 的体素的索引信息对应，基于障碍物 1 的体素的索引信息可以从场景的体素的融合特征确定障碍物 1 的

体素的融合特征。

示例性地，文本编码网络、属性解码网络均可以采用卷积神经网络或者变换 transformer 网络的网络结构。可以理解，文本编码网络、属性解码网络可以根据自身的功能自适应调整网络的相关参数。

在一些可能的实施例中，感知模型还包括路径评估网络，路径评估网络可以用于为车辆确定推荐路径。示例性地，路径评估网络用于根据车辆的多条规划路径和场景的体素的融合特征输出这多条规划路径的推荐系数以及这多条规划路径中的推荐路径。

一种实现方式中，路径评估网络包括路径编码网络、特征交互网络和评估输出网络，其中，路径编码网络用于提取车辆的多条规划路径中每条规划路径的路径特征；特征交互网络用于根据每条规划路径的路径特征和场景的体素的融合特征获得每条规划路径的风险特征；评估输出网络用于根据这多条规划路径的风险特征输出这多条规划路径的推荐系数和这多条规划路径中的推荐路径。

示例性地，规划路径的推荐系数可以基于该规划路径的风险系数、舒适度和通行效率中的至少一项获得。其中，该规划路径的风险系数与该规划路径与障碍物（包括可见的障碍物以及当前处于盲区的障碍物）的距离、该规划路径与道路中其他交通参与者的路径是否有冲突（例如当前时刻或者未来时刻是否会发生碰撞）等因素中的至少一项有关，该规划路径的通行效率与该规划路径的长度、该规划路径对应的预估通行时长、该规划路径途径的红绿灯的数量、该规划路径所在的可行驶区域的面积等因素中的至少一项有关，该规划路径的舒适度与该规划路径的转向加速度的大小及转向频率、该规划路径的加速度的变化率、该规划路径途径的路面的平整度、该规划路径途径的红绿灯的数量、该规划路径所在的道路的类型、该规划路径途径区域是否阴凉等因素中的至少一项有关。

示例性地，推荐路径为这多条规划路径中最高推荐系数对应的规划路径。

示例性地，路径编码网络可以采用卷积神经网络、变换 transformer 网络、图神经网络（graph neural network, GNN）、图卷积神经网络（graph convolution neural networks, GCNNs）的网络结构。特征交互网络可以采用图神经网络或者变换 transformer 网络的网络结构。评估输出网络可以采用神经网络或多层感知器 MLP 的网络结构。

可以理解，图 2 所示的感知模型的框架只是本申请是实施例给出的一种可行的示例，并不应对感知模型的框架构成限定。

示例性地，在感知模型包括感知检测网络、属性识别网络和路径评估网络的情况下，感知检测网络、属性识别网络和路径评估网络的训练可以是分开的，例如先训练感知检测网络，感知检测网络训练完成后依次训练属性识别网络和路径评估网络。感知检测网络、属性识别网络和路径评估网络的训练也可以是同时进行，在此不作具体限定。感知模型中各个网络的训练过程可参考下述实施例中的相应内容的叙述，在此不再赘述。

参见图 4，图 4 是本申请实施例提供的一种智能驾驶方法的流程图。该方法可以应用于上述图 1 中的车辆或者车辆上用于自动驾驶控制的组件（例如芯片或集成电路等），该车辆上至少部署有上述感知检测网络。该方法包括但不限于以下步骤：

S401：获取传感器对第一场景的采集数据，传感器包括摄像头和雷达中的至少一种。

这里，第一场景可以理解为车辆在行驶过程中传感器可以探测到的环境空间。

这里，传感器部署在车辆上。其中，基于摄像头在车辆的安装位置，摄像头例如可以分为前视摄像头、环视摄像头、后视摄像头和侧视摄像头等。雷达包括激光雷达、毫米波雷达中的至少一项。本申请实施例不限定车辆上配置的摄像头的数量以及雷达的数量。

其中，摄像头用于采集图像数据，雷达用于采集点云数据，故上述采集数据包括图像数据和点云数据中的至少一种。

示例性地，车辆上可以配置多个摄像头，不同摄像头的视场角不同，这多个摄像头的视场角可以覆盖以车辆为中心的 360 度的视野范围。示例性地，多个摄像头中相邻摄像头的视场角范围可以存在部分重叠，如此，同一环境空间内的数据可以同时被多个传感器采集到，有利于提高数据观测的置信度。

示例性地，传感器包括摄像头和雷达，假设车辆上摄像头的数量为 m ， m 个摄像头对第一场景进行图像数据的采集，假设每个时刻每个摄像头采集一张图像，即意味着每个时刻对应的采集数据均包括摄像头采集的 m 张图像对应的图像数据和雷达采集的点云数据。

S402：将采集数据输入至感知检测网络，输出感知信息，感知信息用于指示第一场景的障碍物的体素。

这里，障碍物是指车辆在行驶过程中不期望与之发生碰撞的实体，该实体可以是静态的，也可以是动态的。其中，静态的实体例如可以是道路中的纸箱、道路施工牌、道路分界栏杆、土堆、轮胎、侧翻的车辆、躺着的人、动物、道路旁的建筑物、树木、停放的车辆、道路指示牌、电线杆、路边隔离带等具有体积和质量的静止物体；动态的实体例如可以是行人（例如行走的行人、骑自行车的行人等）、动物、车辆、载物的车辆（例如装载了纸箱、树枝或其他货物）等具有体积和质量的运动物体。

这里，感知检测网络为部署在车端的已训练好的感知检测网络。感知检测网络用于根据传感器对第一场景的采集数据输出感知信息。例如，感知检测网络为图 1 所示的网络侧设备基于传感器数据集和采用 4D 重建生成的该传感器数据集对应的标签信息进行训练获得。传感器数据集对应的标签信息可以是网络侧设备采用自监督方式基于该传感器数据集进行 4D 重建生成，该标签信息用于在感知检测网络的训练过程中为感知检测网络提供其预测结果的真值信息。

例如，感知检测网络的预测任务包括预测体素的占据状态、体素的速度信息、体素的可见状态以及障碍物对应的多边形框的角点信息这四种预测任务，在感知检测网络的训练过程中，假设感知检测网络当前的输入数据为 t 时刻的图像数据和 t 时刻的点云数据，则感知检测网络对该输入数据执行上述四种预测任务的处理并输出预测的感知信息（即预测结果），相应地，标签信息包括该 t 时刻的图像数据和 t 时刻的点云数据二者对应的预测结果的真值信息。

一种实现方式中，采集数据包括图像数据和点云数据，感知检测网络包括图像特征提取网络、点云特征提取网络、特征融合网络和输出网络，在此情况下，感知检测网络的处理过程例如可以参考下述步骤 A1-A4：

A1：图像特征提取网络提取该图像数据的 3D 图像特征；

A2：点云特征提取网络提取该点云数据对应的体素的点云特征；

A3：特征融合网络根据所上述 3D 图像特征和点云数据对应的体素的点云特征进行融合，获得第一场景的体素的融合特征；

A4：输出网络处理第一场景的体素的融合特征并输出感知信息。

这里，感知检测网络的推理过程具体可参考上述图 2 实施例中对感知检测网络的叙述，上述图像特征提取网络、点云特征提取网络、特征融合网络和输出网络参考上述图 2 实施例中相应内容的叙述，在此不再赘述。可以理解，上述示例不对感知检测网络的框架构成限制。

在本申请实施例中，感知信息包括以下信息的至少一项：第一场景的体素的占据状态、第一场景的体素的速度信息、第一场景的体素的可见状态和第一场景的障碍物对应的多边形框的角点信息；其中，第一场景的障碍物对应的多边形框与该障碍物的体素关联。

例如，参见图 2 所示的感知检测网络的框架，可知输出网络包括四个头网络，每个头网络对应一种预测任务，在此情况下，感知信息包括第一场景的体素的占据状态、第一场景的体素的速度信息、第一场景的体素的可见状态和第一场景的障碍物对应的多边形框的角点信息。

这里，体素的占据状态可以分为两种，即“占据”和“空”。有关体素的占据状态可参考前述对体素的占据状态的相关叙述，在此不再赘述。

例如，场景的体素 1 与该场景所在的物理世界中的车辆 A 对应，则体素 1 的占据状态为“占据”；场景的体素 2 与该场景所在的物理世界中的空气对应，则体素 2 的占据状态为“空”。

这里，体素的可见状态也可以两种，即“可见”和“不可见”。有关体素的可见状态可参考前述对体素的可见状态的相关叙述，在此不再赘述。

示例性地，体素的可见状态是可以变化的。参见图 5，图 5 是本申请实施例提供的一些场景示意图。图 5 的 (1) 示出了 t1 时刻对应的场景 1，在图 5 的 (1) 中，车辆 1 为主车（即上述感知检测网络部署在车辆 1 上），可以看出，车辆 1、车辆 2 和车辆 3 位于同一车道，且车辆 2 当前正在执行换道操作，假设车辆 2 的车辆体形大于前方车辆 3 的车辆体形，导致从车辆 1 的视角看车辆 3 被车辆 2 完全遮挡，车辆 3 位于车辆 1 的视线盲区内，故在 t1 时刻车辆 1 上的任意一个传感器的观测信号均无法触及车辆 3 的体素，因此，车辆 1 输出的感知信息中，车辆 2 的体素在 t1 时刻的可见状态为“可见”但车辆 3 的体素在 t1 时刻的可见状态均为“不可见”。图 5 的 (2) 示出了 t2 时刻对应的场景 2，可以看出，车辆 2 当前已完成换道操作，假设车辆 2 和车辆 3 均出现在车辆 1 的传感器的采集视野范围内，即意味着车辆 2 的体素以及车辆 3 的体素在 t2 时刻均可以被车辆 1 上的至少一个传感器的观测信号触及，因此，车辆 1 输出的感知信息中，车辆 2 的体素在 t2 时刻的可见状态为“可见”且车辆 3 的体素在 t2 时刻的可见状态也为“可见”。由此也可以看出，将多个时刻的采集数据输入至感知检测网络，不仅可以补齐观测信息，也有利于从多方位、多角度更真实地对场景所在的物理世界进行还原。

S403：至少基于感知信息控制车辆行驶。

这里，控制车辆的行驶包括以下操作中的至少一项：变换车道、调整行驶速度、调整行驶路径、开启警示灯和调整车辆的悬架。如此，车辆至少基于感知信息方便实时决策，以提高自身行驶时过程中的安全性。

一种实现方式中，至少基于感知信息控制车辆行驶，包括：至少根据感知信息，调整车辆的行驶路径，其中，调整后的行驶路径不途径障碍物的体素所在的区域。

例如，根据感知信息可以确定车辆当前的行驶路径在当前时刻以及未来时刻与该场景中相应时刻下的障碍物的体素是否会发生碰撞，在预测到有碰撞的情况下，可以及时调整车辆当前的行驶路径，使得调整后的行驶路径不途径障碍物的体素所在的区域，如此，可以避免车辆在行驶过程与障碍物发生碰撞，有利于提高车辆行驶的安全性。

一种实现方式中，感知信息还用于指示第一场景的路面的体素，则至少基于感知信息控制车辆行驶，包括：至少根据感知信息，生成第一场景的路面几何信息；根据该路面几何信息，调整车辆内的悬架。

这里，路面几何信息用于指示第一场景的路面状况（例如路面是否有坑洼、路面是否有凸起等），基于感知信息车辆可以提前获取车辆前方的路面状况，在监测到路面有起伏时，车辆有足够的时间可以及时调整车辆的悬架，以使车辆在行驶过程中尽可能始终保持水平且平稳的状态，减少因路面起伏带来的振动感，提高了乘坐车辆的舒适性。

示例性地，基于感知信息控制车辆行驶，还可以是：根据感知信息确定场景中的盲区和盲区内的障碍物信息（例如该障碍物的速度信息、该障碍物对边的多边形框的角点信息等），当车辆接近盲区时，基于盲区内的障碍物信息控制车辆减速、停车或转向。这里，不限定盲区内的障碍物的种类、呈现形态等，且盲区内的障碍物可能是静态的，也可能是动态，在此不作具体限定。如此，当车辆接近当前时刻场景中的盲区时，控制车辆处于减速状态或停车状态或转向状态，能避免车辆与盲区内的障碍物发生碰撞，提高了车辆行车的安全性。

这里，盲区例如为感知信息中当前时刻可见状态为“不可见”的体素所在的区域。示例性地，盲区包括当前时刻车辆上的传感器的观测信号本可以触及的区域中因其他障碍物遮挡未能触及的区域和传感器本身的探测盲区。

这里，障碍物的速度信息例如可以基于该障碍物的体素的速度信息获得。

在一些可能的实施例，车辆除了可以基于车辆自身输出的感知信息控制车辆的行驶，还可以结合导航地图信息、高精地图信息、路侧设备广播的交通实况信息以及周围其他车辆广播的交通实况信息等中的至少一项控制车辆的行驶。这里，路侧设备例如可以是路侧单元（road side unit, RSU）、多接入边缘计算（multi-access edge computing, MEC）或者传感器等装置，或者是这些装置内部的组件或者芯片，也可以是由 RSU 和 MEC 组成的系统级设备，或者是由 RSU 和传感器组成的系统级设备，还可以是由 RSU、MEC 和传感器组成的系统级设备。

可选地，在一些可能的实施例，上述智能驾驶方法还包括：基于感知信息显示第一场景的障碍物，其中，第一场景的障碍物以多边形框进行标记；和/或基于感知信息显示第一场景的障碍物的体素。

示例性地，可以在车辆的显示装置上呈现障碍物或者障碍物的体素。例如，显示装置可以是车端设备的车机平板、车载显示器、抬头显示（head up display, HUD）系统或者增强抬头显示 AR-HUD 系统等，在此不作具体限定。

参见图 6A，图 6A 是本申请实施例提供的一种以多边形框标记场景中的障碍物的示意图。图 6A 显示了当前时刻自车所在场景中的障碍物，其中，障碍物以多边形框进行标记。图 6A 中处于中心下方的车辆为自车，可以看出，该场景下自车周围环境中的障碍物被用多边形框进行了标记显示，基于多边形框的形状可以看出该场景的障碍物至少包括车辆、建筑等。示例性地，多边形框可以是二维的，也可以是三维的。以图 6A 中自车右侧距离自车最近的多边形框为例，该多边形框以 2D 显示时，该多边形框可以由 1 组角点信息指示的 10 个角点连接而成；该多边形框以 3D 显示时，该多边形框可以由多组角点信息指示的角点连接而成，其中，每组角点信息指示 10 个角点。在一些可能的实施例，对于场景中动态的障碍物，还可以在该动态的障碍物对应的多边形框上添加箭头，该箭头表示该障碍物为动态的障碍物且

该箭头的方向指示了该障碍物的运动方向，该箭头的长度表示该障碍物的速度的大小。可以理解，图 6A 仅为某个时刻车辆所在场景的障碍物的标记显示的一种示例，并不应对车辆所在场景的障碍物的标记显示构成限定。

参见图 6B，图 6B 是本申请实施例提供的一种障碍物的体素的显示示意图。图 6B 示出了当前时刻自车所在场景中的障碍物的体素，可以看出，障碍物的体素由该场景中的多个体素组成，体素可以理解为图 6B 中最小单元的立体方格。示例性地，在图 6B 中，可以通过不同的颜色将动态的障碍物和静态的障碍物进行区分显示（即可以通过不同颜色区分不同速度的障碍物），也可以通过不同的颜色将不同的障碍物进行区分，在此不作具体限定。可以理解，图 6B 仅为某个时刻车辆所在场景的障碍物的体素的一种显示示例，并不应对车辆所在场景的障碍物的体素的显示构成限定。

在一些可能的实施例中，还可以显示车辆当前所在场景中路面的体素。参见图 6C，图 6C 是本申请实施例提供的一种显示了场景中路面的体素的示意图。图 6C 不仅对当前时刻该场景中障碍物的体素进行了显示，还对当前时刻该场景中路面的体素也进行了显示，如此，基于图 6C 可以看出前方路面的起伏程度。可以理解，图 6C 仅为某个时刻车辆所在场景的障碍物的体素以及路面的体素的一种显示示例，并不应对车辆所在场景中障碍物的体素以及路面的体素的显示构成限定。

在一些可能的实施例中，车辆上除了部署有上述感知检测网络，还可以部署属性识别网络，其中，属性识别网络用于识别障碍物的类别。如此，车辆在行驶过程中不仅可以检测到周围环境中的障碍物，还可以识别出障碍物的类别，实现了车辆不仅看的到物还能看的懂物。

进一步地，上述智能驾驶方法还包括：获取文本查询信息；将文本查询信息和障碍物的体素的融合特征输入至属性识别网络，输出该障碍物的类别信息；其中，文本查询信息用于请求查询类别；显示该障碍物的类别信息；其中，该障碍物的体素的融合特征基于该障碍物对应的多边形框的角点信息和第一场景的体素的融合特征确定，该障碍物对应的多边形框与该障碍物的体素关联。其中，障碍物对应的多边形框的角点信息和第一场景的体素的融合特征均来自感知检测网络，进一步地，结合上述图 2 所示的感知检测网络可以知晓，障碍物对应的多边形框的角点信息来自感知检测网络中的输出网络，第一场景的体素的融合特征来自上述感知检测网络中的特征融合网络。此实施例具体可参考上述图 2 实施例对属性识别网络的相关说明，为了说明书的简洁，在此不再赘述。

在一些可能的实施例中，在获得上述感知信息后，还可以结合摄像头内配置的检测算法的检测结果、雷达内配置的检测算法的检测结果或其他模型的检测结果进行进一步融合处理，如此，当同一障碍物通过多种不同的方式均能感知到的情况下，则该检测出该障碍物的置信度也更高。

在一些可能的实施例中，车辆上除了部署有上述感知检测网络，还可以部署路径评估网络，其中，路径评估网络用于为车辆推荐最低风险路径。如此，有利于提高驾驶的安全性和驾驶决策的准确率。

进一步地，上述智能驾驶方法还包括：获取车辆的多条规划路径；将车辆的多条规划路径和第一场景的体素的融合特征输入至路径评估网络，输出这多条规划路径的推荐系数和这多条规划路径中的推荐路径，其中，推荐路径与这多条规划路径的推荐系数关联；显示所述推荐路径。可以知晓，第一场景的体素的融合特征来自感知检测网络，结合上述图 2 对感知

检测网络的叙述可以知第一场景的体素的融合特征由感知检测网络中的特征融合网络提供。这里，多条规划路径由车辆生成，例如车辆基于导航地图信息生成多条规划路径。此实施例具体可参考上述图 2 实施例对路径评估网络的相关说明，为了说明书的简洁，在此不再赘述。

示例性地，路径评估网络输出的推荐路径包括多条规划路径中的至少两条规划路径，在此情况下，在人机共驾的场景下，还可以用于用户推荐该推荐路径，从用户接收反馈信息，反馈信息用于指示用户从该至少两条规划路径中选择的路径，并控制自身车辆沿着用户选择的路径行驶。

可以理解，推荐路径包括的规划路径的数量为多个的情况下，可以理解为推荐路径包括的各规划路径的推荐系数相近或相同，但有的规划路径是耗时最短的路径，有的规划路径是舒适度最高的路径，有的规划路径是距离最短的路径等，在此情况下，可供用户根据自身的需求自由选择，为用户提供了良好的乘车体验感。

在一些可能的实施例中，车辆上也可以同时部署上述感知检测网络、属性识别网络和路径评估网络，相应描述可以参考相应实施例的描述，在此不再赘述。

可以看到，实施本申请实施例，通过在车辆在部署上述感知检测网络，能够增强车辆对周围环境的感知能力，使得车辆在行驶过程中能感知到周围的障碍物，从而能避免碰撞的发生，提高了车辆的安全性。另外，通过在车辆上部署上述属性识别网络，车辆在感知到障碍物的基础上，还能识别出障碍物的类别，提高了车辆的智能性。

参见图 7A，图 7A 是本申请实施例提供的一种感知检测网络的训练方法的流程图。该方法可以应用于上述图 1 所示的网络侧设备或者网络侧设备内的组件（例如芯片或集成电路等）。以图 2 所示的感知检测网络为例，感知检测网络包括图像特征提取网络、点云特征提取网络、特征融合网络和输出网络，其中，输出网络包括多个头网络。该方法包括但不限于以下步骤：

S701：在每次训练过程中，通过图像特征提取网络对一批次传感器数据中每个时刻的图像数据进行特征提取，获得 K 个时刻的图像数据的 3D 图像特征。

示例性地，每次训练过程中使用的一个批次的传感器数据位于传感器数据集中。

例如，一个批次的传感器数据包括 K 个时刻的图像数据，其中，K 个时刻的图像数据来自车辆的至少一个摄像头，K 为正整数。

具体地，该一个批次的传感器数据中 K 个时刻的图像数据输入至图像特征提取网络，图像特征提取网络基于每个时刻的图像数据得到该时刻的图像数据的 3D 图像特征，故图像特征提取网络将获得 K 个时刻的图像数据的 3D 图像特征，并将其输出至特征融合网络。

这里，图像特征提取网络的框架具体可参考图 2 实施例中相应内容的叙述，在此不再赘述。

S702：在每次训练过程中，通过点云特征提取网络对一批次传感器数据中每个时刻的点云数据进行特征提取，获得 K 个时刻的点云数据对应的体素的点云特征。

这里，一个批次的传感器数据还包括这 K 个时刻的点云数据，其中，这 K 个时刻的点云数据来自该车辆的至少一个雷达。

具体地，该一个批次的传感器数据中 K 个时刻的点云数据输入至点云特征提取网络，点云特征提取网络基于每个时刻的点云数据得到该时刻的点云数据对应的体素的点云特征，故点云特征提取网络将获得 K 个时刻的点云数据对应的体素的点云特征，并将其输出至特征融

合网络。

S703: 通过特征融合网络对 K 个时刻的图像数据的 3D 图像特征和 K 个时刻的点云数据对应的体素的点云特征进行特征融合, 获得 K 个时刻的场景的体素的融合特征。

这里, 场景的体素是指经过特征融合网络执行特征融合后的体素。

示例性地, 特征融合网络根据每个时刻的图像数据的 3D 图像特征和该时刻的点云数据对应的体素的点云特征进行空间上的融合, 获得该时刻的场景的体素的空间融合特征; 特征融合网络再根据 K 个时刻的场景的体素的空间融合特征进行时间上的融合, 获得 K 个时刻的场景的体素的融合特征, 在此情况下, 每个时刻的场景的体素的融合特征可以称为该时刻的场景的体素的时空融合特征。

S704: 通过输出网络中的每个头网络根据 K 个时刻的场景的体素的融合特征输出 K 个预测结果, 其中, K 个预测结果中的每个预测结果对应一个时刻的场景。

其中, K 个时刻的场景的体素的融合特征与输出网络中的每个头网络对应, 输出网络中的每个头网络执行一种预测任务。

例如, 图 2 所示的输出网络包含头网络 1、头网络 2、头网络 3 和头网络 4, 其中, 头网络 1 用于预测场景的障碍物对应的多边形框的角点信息, 头网络 2 用于预测场景的体素的占据状态, 头网络 3 用于预测场景的体素的速度信息, 头网络 4 用于预测场景的体素的可见状态。

具体地, 在输出网络中, 每个头网络基于每个时刻的场景的体素的融合特征进行预测, 获得该时刻的场景对应的预测结果, 从而每个头网络可以获得 K 个预测结果。

以图 2 所示的输出网络中的头网络 1 为例, 假设 K 个时刻包括 t_1 时刻、 t_2 时刻、.....、 t_K 时刻, 头网络 1 基于 t_1 时刻的场景的体素的融合特征输出预测结果 1, 预测结果 1 包括 t_1 时刻的场景中障碍物对应的多边形框的角点信息; 头网络 1 基于 t_2 时刻的场景的体素的融合特征输出预测结果 2, 预测结果 2 包括 t_2 时刻的场景中障碍物对应的多边形框的角点信息;, 如此, 头网络 1 基于 K 个时刻的场景的体素的融合特征将输出 K 个预测结果。

S705: 根据该批次的传感器数据对应的标签信息和输出网络中每个头网络输出的 K 个预测结果获得输出网络中每个头网络的损失值。

其中, 该批次的传感器数据对应的标签信息是通过 4D 重建基于该批次的传感器数据生成。该批次的传感器数据对应的标签信息用于为感知检测网络提供感知检测网络对该批次的传感器数据的预测结果对应的真值信息。

以图 2 所示的输出网络为例, 则该批次的传感器数据对应的标签信息包括 K 个时刻的场景的障碍物对应的多边形框的角点真值信息、这 K 个时刻的场景的体素的占据状态真值信息、这 K 个时刻的场景的体素的速度真值信息和这 K 个时刻的场景的体素的可见状态真值信息, 其中, K 个时刻的场景的障碍物对应的多边形框的角点真值信息与上述头网络 1 输出的 K 个预测结果对应, 这 K 个时刻的场景的体素的占据状态真值信息与上述头网络 2 输出的 K 个预测结果对应, 这 K 个时刻的场景的体素的速度真值信息与上述头网络 3 输出的 K 个预测结果对应, 以及这 K 个时刻的场景的体素的可见状态真值信息与上述头网络 4 输出的 K 个预测结果对应。

以头网络 1 的损失值的计算为例, 基于上述标签信息中 K 个时刻的场景的障碍物对应的多边形框的角点真值信息和头网络 1 输出的 K 个预测结果, 获得头网络 1 的损失值。示例性

地，可以先根据每个时刻的场景的障碍物对应的多边形框的角点真值信息和头网络 1 输出的 K 个预测结果中该时刻的预测结果（即包括该时刻的场景中障碍物对应的多边形框的角点信息）获得头网络 1 在该时刻的损失值，然后再根据头网络 1 在 K 个时刻中各时刻的损失值获得头网络 1 的损失值。同理，输出网络中的其他头网络也可以采用此方式进行自身头网络的损失值的计算，如此，可以获得该输出网络中每个头网络的损失值。

S706：对输出网络中各个头网络的损失值进行加权，获得每次训练过程对应的一个损失值；利用该损失值对感知检测网络中的参数进行更新。

这里，输出网络中每个头网络的权重可以是用户自定义设置。

在得到每次训练过程对应的一个损失值后，利用该损失值对感知检测网络（例如输出网络中的每个头网络+特征融合网络+图像特征提取网络+点云特征提取网络）中的参数进行更新。

可以理解，图 7A 为上述感知检测网络单独进行训练的一种示例，并不限定感知检测网络的训练流程仅为图 7A 所示形式。在一些可能的实施例中，上述感知检测网络中输出网络中的各个头网络也可以单独训练的，在此情况下，上述 S706 不是必要执行步骤。在一些可能的实施例中，感知检测网络还可以与神经辐射场 NeRF 网络进行联合训练，以进一步提高检测的精准率以及训练效率，在此不作具体限定。

实施本申请实施例，每次训练过程中一个批次的传感器数据对应的标签信息无需通过人工标注生成，不仅节省了人力的消耗，也提高了标签信息的获取效率。感知检测网络采用自监督训练这种方式，能够学习到基于任一时刻车辆的输入数据可以准确预测该时刻下车辆所在场景中的感知信息。

参见图 7B，图 7B 是本申请实施例提供的一种感知检测网络的训练过程示意图。

如图 7B 所示，将传感器数据集中一个批次的传感器数据输入至感知检测网络中的图像特征提取网络和点云特征提取网络，图像特征提取网络对该批次的传感器数据中的图像数据进行特征提取，获得该图像数据的 3D 图像特征，点云特征提取网络对该批次的传感器数据中的点云数据进行特征提取，获得该点云数据对应的体素的点云特征，感知检测网络中的特征融合网络对来自图像特征提取网络的该图像数据的 3D 图像特征和来自点云特征提取网络的该点云数据对应的体素的点云特征进行特征融合，获得场景的体素的融合特征并将其输入至感知检测网络中的输出网络中，输出网络中的每个头网络基于场景的体素的融合特征获得对应的预测结果。基于输出网络中每个头网络输出的预测结果和该批次的传感器数据对应的标签信息获得输出网络中每个头网络的损失值，将输出网络中各头网络的损失值进行加权，该次训练过程对应的一个损失值，并利用该损失值进行反向传播，实现依次更新上述输出网络、特征融合网络、特征提取网络（包括图像特征提取网络和点云特征提取网络）的参数。可以理解，图 7B 仅为感知检测网络的训练过程的一种示意，并不限定感知检测网络的训练过程仅为图 7B 所示，例如输出网络中的各个头网络也可以单独训练。

这里，感知检测网络的训练具体可参考上述图 7A 实施例的相关叙述，此处不赘述。

在一些可能的实施例中，在完成对感知检测网络的训练后，可以对属性识别网络进行训练。下面来说明属性识别网络的训练过程。

以图 2 所示的属性识别网络为例，属性识别网络包括文本编码网络和属性解码网络。示

例性地，文本编码网络可以直接使用训练好的词向量特征提取器（或者采用文本-图像的预训练学习获得），本申请实施例可以仅对属性解码网络进行训练。该训练好的词向量特征提取器可以基于输入的文本查询信息提取该文本查询信息的词向量特征，且该文本查询信息的词向量特征可以表征该文本查询信息指示的类别的图像语义特征。

示例性地，采用文本-图像的预训练获得文本编码网络的过程可以是：获取海量的文本-图像训练数据，其中，文本-图像训练数据包括多个文本-图像数据组，每个文本-图像数据组包括指示了类别信息的文本信息和该文本信息对应的图像，例如，文本-图像训练组 1 包括指示了车的文本信息和车的图像，将文本-图像训练数据中的文本信息输入至文本编码器以分别提取每条文本信息的词向量特征，将文本-图像训练数据中的文本信息中的图像输入至图像编码器以分别提取每张图像的图像特征，基于使得属于同一文本-图像数据组的文本信息的词向量特征与图像的图像特征尽可能接近、而属于不同的文本-图像数据组的文本信息的词向量特征与图像的图像特征尽可能远离这一训练思想调整文本编码器的参数和图像编码器的参数，则训练好的文本编码器可以直接作为上述属性识别网络中的文本编码网络使用。

对于属性解码网络的训练，例如可以是：文本编码网络基于接收的多条文本查询信息向属性解码网络输入每条文本查询信息的词向量特征，属性解码网络根据该多条文本查询信息的词向量特征和感知检测网络提供的障碍物的体素的融合特征预测每个障碍物的类别信息，根据预测的每个障碍物的类别信息和每个障碍物的类别标注信息获得属性解码网络该次训练的损失值，最后基于属性解码网络该次训练的损失值反向更新属性解码网络的参数。

在一些可能的实施例中，在完成对感知检测网络的训练后，可以对路径评估网络进行训练。下面来说明路径评估网络的训练过程。

以图 2 所示的路径评估网络为例，路径评估网络包括路径编码网络、特征交互网络和评估输出网络。路径评估网络的训练过程例如可以是：获取路径训练数据，路径训练数据包括车辆在上述 K 个时刻范围内规划的多条路径以及这多条路径的推荐系数标注信息，将该多条路径输入至路径编码网络，路径编码网络将提取的每条路径的路径特征输出至特征交互网络，特征交互网络根据每条路径的路径特征和上述 K 个时刻的场景的体素的融合特征（来自感知检测网络）输出每条路径的风险特征，评估输出网络基于该多条路径的风险特征输出多条路径的预测推荐系数并基于这多条路径的预测推荐系数从这多条路径中确定预测的推荐路径，根据这多条路径的预测推荐系数和这多条路径的推荐系数标注信息获得这多条路径的损失值，其中，这多条路径的损失值是基于该多条路径中各条路径的损失值加权获得，最后基于该多条路径的损失值更新路径评估网络中的参数。

可以看出，上述感知检测网络、属性识别网络和路径评估网络的训练是分开独立进行的。在一些可能的实施例中，感知检测网络、属性识别网络和路径评估网络也可以联合训练，在此情况下，每次训练过程对应的损失值是基于属性识别网络在该次训练过程中的损失值、感知检测网络在该次训练过程中的损失值（即属性解码网络在该次训练过程中的损失值）和路径评估网络在该次训练过程中的损失值进行加权获得，最后可以该次训练过程中的损失值分别更新上述感知检测网络、属性识别网络中的属性解码网络以及上述路径评估网络内的参数。

请参见图 8，图 8 是本申请实施例提供的一种芯片硬件结构示意图，可以用于执行本申请实施例中的智能驾驶方法和/或训练方法。

如图 8 所示，神经网络处理器（neural-networks processing unit, NPU）80 作为协处理器挂载到主 CPU（Host CPU）上，由 Host CPU 分配任务，以执行前述实施例中的智能驾驶方法或者前述实施例中的训练方法的相关过程。

NPU 的核心部分为运算电路 803，控制器 804 控制运算电路 803 提取存储器（权重存储器或输入存储器）中的数据并进行运算。

在一些实现中，运算电路 803 内部包括多个处理单元（process engine, PE）。在一些实现中，运算电路 803 是二维脉动阵列。运算电路 803 还可以是一维脉动阵列或者能够执行例如乘法和加法这样的数学运算的其它电子线路。在一些实现中，运算电路 803 是通用的矩阵处理器。

举例来说，假设有输入矩阵 A，权重矩阵 B，输出矩阵 C。运算电路从权重存储器 802 中取矩阵 B 相应的数据，并缓存在运算电路中每一个 PE 上。运算电路从输入存储器 801 中取矩阵 A 数据与矩阵 B 进行矩阵运算，得到的矩阵的部分结果或最终结果，保存在累加器（accumulator）808 中。

向量计算单元 807 可以对运算电路的输出做进一步处理，如向量乘，向量加，指数运算，对数运算，大小比较等等。例如，向量计算单元 807 可以用于神经网络中非卷积/非 FC 层的网络计算，如池化（pooling），批归一化（batch normalization），局部响应归一化（local response normalization）等。

在一些实现种，向量计算单元 807 能将经处理的输出的向量存储到统一存储器 806。例如，向量计算单元 807 可以将非线性函数应用到运算电路 803 的输出，例如累加值的向量，用以生成激活值。在一些实现中，向量计算单元 807 生成归一化的值、合并值，或二者均有。在一些实现中，处理过的输出的向量能够用作到运算电路 803 的激活输入，例如用于在神经网络中的后续层中的使用。

统一存储器 806 用于存放输入数据以及输出数据。

存储单元访问控制器 805（direct memory access controller, DMAC）将外部存储器中的输入数据搬运到输入存储器 801 和/或统一存储器 806、将外部存储器中的权重数据存入权重存储器 802，以及将统一存储器 806 中的数据存入外部存储器。

总线接口单元（bus interface unit, BIU）810，用于通过总线实现主 CPU、DMAC 和取指存储器 809 之间进行交互。

与控制器 804 连接的取指存储器（instruction fetch buffer）809，用于存储控制器 804 使用的指令。

控制器 804，用于调用取指存储器 809 中缓存的指令，实现控制该运算加速器的工作过程。

一般地，统一存储器 806，输入存储器 801，权重存储器 802 以及取指存储器 809 均为片上（On-Chip）存储器，外部存储器为该 NPU 外部的存储器，该外部存储器可以为双倍数据率同步动态随机存储器（double data rate synchronous dynamic random access memory, DDR SDRAM）、高带宽存储器（high bandwidth memory, HBM）或其他可读可写的存储器。

参见图 9A，图 9A 是本申请实施例提供的一种计算装置的结构示意图，计算装置 30 包括接收单元 310 和处理单元 312。该计算装置 30 可以通过硬件、软件或者软硬件结合的方式

来实现。

其中，接收单元 310，用于获取传感器对第一场景的采集数据，传感器包括摄像头和雷达中的至少一种；处理单元 312，用于将采集数据输入至感知检测网络，输出感知信息，感知信息用于指示第一场景的障碍物的体素；处理单元 312 还用于至少基于感知信息显示障碍物的体素。

在一些可能的实施例中，计算装置 30 还包括显示单元 314（图未示），显示单元 314 用于：基于感知信息显示上述障碍物，该障碍物以多边形框进行标记；和/或，基于感知信息显示上述障碍物的体素。

该计算装置 30 可用于实现图 4 实施例所描述的方法。在图 4 实施例中，接收单元 310 可用于执行 S401，处理单元 312 可用于执行 S402 和 S403。

参见图 9B，图 9B 是本申请实施例提供的一种训练装置的结构示意图，训练装置 40 包括编码单元 410、解码单元 412 和更新单元 414。该训练装置 40 可以通过硬件、软件或者软硬件结合的方式来实现。

其中，编码单元 410 用于在每次训练过程中，通过图像特征提取网络对一批次传感器数据中每个时刻的图像数据进行特征提取，获得 K 个时刻的图像数据的 3D 图像特征，K 为正整数；在每次训练过程中，通过点云特征提取网络对一批次传感器数据中每个时刻的点云数据进行特征提取，获得 K 个时刻的点云数据对应的体素的点云特征；以及通过特征融合网络对 K 个时刻的图像数据的 3D 图像特征和 K 个时刻的点云数据对应的体素的点云特征进行特征融合，获得 K 个时刻的场景的体素的融合特征；解码单元 412，用于通过输出网络中的每个头网络根据 K 个时刻的场景的体素的融合特征输出 K 个预测结果，其中，K 个预测结果中的每个预测结果对应一个时刻的场景；更新单元 414 用于根据该批次的传感器数据对应的标签信息和输出网络中每个头网络输出的 K 个预测结果获得输出网络中每个头网络的损失值；以及对输出网络中各个头网络的损失值进行加权，获得每次训练过程对应的一个损失值；并利用该损失值对感知检测网络中的参数进行更新。

该训练装置 40 可用于实现图 7A 实施例所描述的方法。在图 7A 实施例中，编码单元 410 可用于执行 S701-S703，解码单元 412 可用于执行 S704，更新单元 414 可用于执行 S705 和 S706。

应理解，以上装置（例如计算装置 30 和训练装置 40）中各单元的划分仅是一种逻辑功能的划分，实际实现时可以全部或部分集成到一个物理实体上，也可以物理上分开。此外，装置中的单元可以以处理器调用软件的形式实现；例如装置包括处理器，处理器与存储器连接，存储器中存储有指令，处理器调用存储器中存储的指令，以实现以上任一种方法或实现该装置各单元的功能，其中处理器例如为通用处理器，例如中央处理单元（central processing unit, CPU）或微处理器，存储器为装置内的存储器或装置外的存储器。或者，装置中的单元可以以硬件电路的形式实现，可以通过对硬件电路的设计实现部分或全部单元的功能，该硬件电路可以理解为一个或多个处理器；例如，在一种实现中，该硬件电路为专用集成电路（application-specific integrated circuit, ASIC），通过对电路内元件逻辑关系的设计，实现以上部分或全部单元的功能；再如，在另一种实现中，该硬件电路为可以通过可编程逻辑器件

(programmable logic device, PLD)实现,以现场可编程门阵列(field programmable gate array, FPGA)为例,其可以包括大量逻辑门电路,通过配置文件来配置逻辑门电路之间的连接关系,从而实现以上部分或全部单元的功能。以上装置的所有单元可以全部通过处理器调用软件的形式实现,或全部通过硬件电路的形式实现,或部分通过处理器调用软件的形式实现,剩余部分通过硬件电路的形式实现。

在本申请实施例中,处理器是一种具有信号的处理能力的电路,在一种实现中,处理器可以是具有指令读取与运行能力的电路,例如中央处理单元(central processing unit, CPU)、微处理器、图形处理器(graphics processing unit, GPU)(可以理解为一种微处理器)、或数字信号处理器(digital signal processor, DSP)等;在另一种实现中,处理器可以通过硬件电路的逻辑关系实现一定功能,该硬件电路的逻辑关系是固定的或可以重构的,例如处理器为专用集成电路(application-specific integrated circuit, ASIC)或可编程逻辑器件(programmable logic device, PLD)实现的硬件电路,例如FPGA。在可重构的硬件电路中,处理器加载配置文档,实现硬件电路配置的过程,可以理解为处理器加载指令,以实现以上部分或全部单元的功能的过程。此外,还可以是针对人工智能设计的硬件电路,其可以理解为一种ASIC,例如神经网络处理单元(neural network processing unit, NPU)、张量处理单元(tensor processing unit, TPU)、深度学习处理单元(deep learning processing unit, DPU)等。

可见,以上装置中的各单元可以是被配置成实施以上方法的一个或多个处理器(或处理电路),例如:CPU、GPU、NPU、TPU、DPU、微处理器、DSP、ASIC、FPGA,或这些处理器形式中至少两种的组合。

此外,以上装置中的各单元可以全部或部分可以集成在一起,或者可以独立实现。在一种实现中,这些单元集成在一起,以片上系统(system-on-a-chip, SOC)的形式实现。该SOC中可以包括至少一个处理器,用于实现以上任一种方法或实现该装置各单元的功能,该至少一个处理器的种类可以不同,例如包括CPU和FPGA, CPU和人工智能处理器, CPU和GPU等。

参见图10,图10是本申请实施例提供的一种处理设备的结构示意图。如图10所示,处理设备50包括:处理器501、通信接口502、存储器503和总线504。处理器501、存储器503和通信接口502之间通过总线504通信。应理解,本申请不限定处理设备50中的处理器、存储器的个数。

一种实现方式中,处理设备50为车辆内用于自动驾驶控制的组件(例如芯片或集成电路等)。其中,该车辆配置有自动驾驶系统,这里,自动驾驶系统并不局限于完全自动驾驶系统、高度自动驾驶系统、有条件自动驾驶系统、或部分自动驾驶系统等,本领域技术人员可以理解,提供智能驾驶的非完全人工驾驶系统都可以涵盖在本概念之下。

另一种实现方式中,处理设备50可以是网络侧设备。网络侧设备是具有计算能力的设备。网络侧设备例如可以是部署在网络侧的服务器(例如用于智能驾驶处理的服务器),或者为该服务器中的组件或者芯片。在一些可能的实施例中,网络侧设备也可以是由多个服务器组成的系统级设备。网络侧设备可以部署在云环境或者边缘环境中,在此不作具体限定。

总线504可以是外设部件互连标准(peripheral component interconnect, PCI)总线或扩展工业标准结构(extended industry standard architecture, EISA)总线等。总线可以分为地址总

线、数据总线、控制总线等。为便于表示，图8中仅用一条线表示，但并不表示仅有一根总线或一种类型的总线。总线504可包括在处理设备50各个部件（例如，存储器503、处理器501、通信接口502）之间传送信息的通路。

处理器501可参考上述实施例中对处理器的相关描述，在此不再赘述。

存储器503用于提供存储空间，存储空间中可以存储操作系统和计算机程序等数据。存储器503可以是随机存取存储器(random access memory, RAM)、可擦除可编程只读存储器(erasable programmable read only memory, EPROM)、只读存储器(read-only memory, ROM)、或便携式只读存储器(compact disc read memory, CD-ROM)等中的一种或者多种的组合。存储器503可以单独存在，也可以集成于处理器501内部。

通信接口502可用于为处理器501提供信息输入或输出。或者可替换的，该通信接口502可用于接收外部发送的数据和/或向外部发送数据，可以为包括诸如以太网电缆等的有线链路接口，也可以是无无线链路（如Wi-Fi、蓝牙、通用无线传输等）接口。或者可替换的，通信接口502还可以包括与接口耦合的发射器（如射频发射器、天线等），或者接收器等。

在一些可能的实施例中，当处理设备50还包括显示器505。显示器505与处理器501通过总线504连接或耦合。显示器505可以用于显示第一场景的多边形实例。显示器505可以是显示屏，显示屏可以是液晶显示器(liquid crystal display, LCD)、有机或无机发光二极管(organic light-emitting diode, OLED)、有源矩阵有机发光二极管面板(active matrix/organic light emitting diode, AMOLED)等。显示器505也可以是车机平板、车载显示器、抬头显示(head up display, HUD)系统或增强抬头显示AR-HUD系统等。

该处理设备50中的处理器501用于读取存储器503中存储的计算机程序，用于执行前述的方法，例如图4或图7A所描述的方法。

在一种可能的设计方式中，处理设备50可为执行图4所示方法的执行主体中的一个或多个模块，该处理器501可用于读取存储器中存储的一个或多个计算机程序，用于执行以下操作：

通过接收单元310获取传感器对第一场景的采集数据，传感器包括摄像头和雷达中的至少一种；

将采集数据输入至感知检测网络，输出感知信息，感知信息用于指示第一场景的障碍物的体素；至少基于感知信息显示障碍物的体素。

在另一种可能的设计方式中，处理设备50可为执行图7A所示方法的执行主体中的一个或多个模块，该处理器501可用于读取存储器中存储的一个或多个计算机程序，用于执行以下操作：

通过编码单元410在每次训练过程中，通过图像特征提取网络对一批次传感器数据中每个时刻的图像数据进行特征提取，获得K个时刻的图像数据的3D图像特征，K为正整数；在每次训练过程中，通过点云特征提取网络对一批次传感器数据中每个时刻的点云数据进行特征提取，获得K个时刻的点云数据对应的体素的点云特征；以及通过特征融合网络对K个时刻的图像数据的3D图像特征和K个时刻的点云数据对应的体素的点云特征进行特征融合，获得K个时刻的场景的体素的融合特征；

通过解码单元 412 通过输出网络中的每个头网络根据 K 个时刻的场景的体素的融合特征输出 K 个预测结果，其中，K 个预测结果中的每个预测结果对应一个时刻的场景；

通过更新单元 414 根据该批次的传感器数据对应的标签信息和输出网络中每个头网络输出的 K 个预测结果获得输出网络中每个头网络的损失值；以及对输出网络中各个头网络的损失值进行加权，获得每次训练过程对应的一个损失值；并利用该损失值对感知检测网络中的参数进行更新。

在本文上述的实施例中，对各个实施例的描述都各有侧重，某个实施例中并没有详细描述的部分，可以参见其他实施例的相关描述。另外，在本申请的各个实施例中，如果没有特殊说明以及逻辑冲突，各个实施例之间的术语和/或描述具有一致性、且可以相互引用，不同的实施例中的技术特征根据其内在的逻辑关系可以组合形成新的实施例。

需要说明的是，本领域普通技术人员可以看到上述实施例的各种方法中的全部或部分步骤是可以通程序来指令相关的硬件来完成，该程序可以存储于一计算机可读存储介质中，存储介质包括只读存储器（read-only memory, ROM）、随机存储器（random access memory, RAM）、可编程只读存储器（programmable read-only memory, PROM）、可擦除可编程只读存储器（erasable programmable read only memory, EPROM）、一次可编程只读存储器（one-time programmable read-only memory, OTPROM）、电子抹除式可复写只读存储器（electrically-erasable programmable read-only memory, EEPROM）、只读光盘（compact disc read-only memory, CD-ROM）或其他光盘存储器、磁盘存储器、磁带存储器、或者能够用于携带或存储数据的计算机可读的任何其他介质。

本申请的技术方案本质上或者说做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式体现出来，该计算机程序产品存储在一个存储介质中，包括若干指令用以使得一个设备（可以是个人计算机，服务器，或者网络设备、机器人、单片机、芯片、机器人等）执行本申请各个实施例所述方法的全部或部分步骤。

权利要求书

1. 一种智能驾驶方法，其特征在于，所述方法包括：
获取传感器对第一场景的采集数据，所述传感器包括摄像头和雷达中的至少一种；
将所述采集数据输入至感知检测网络，输出感知信息，所述感知信息用于指示所述第一场景的障碍物的体素；
至少基于所述感知信息控制车辆行驶。
2. 根据权利要求1所述的方法，其特征在于，所述方法还包括：
基于所述感知信息显示所述障碍物，所述障碍物以多边形框进行标记；和/或
基于所述感知信息显示所述障碍物的体素。
3. 根据权利要求1或2所述的方法，其特征在于，所述感知信息包括以下信息的至少一项：
所述第一场景的体素的占据状态、所述第一场景的体素的速度信息、所述第一场景的体素的可见状态和所述障碍物对应的多边形框的角点信息；
其中，所述障碍物对应的多边形框与所述障碍物的体素关联。
4. 根据权利要求1-3任一项所述的方法，其特征在于，所述感知信息还用于指示所述第一场景的路面的体素，所述至少基于所述感知信息控制车辆行驶，包括：
至少根据所述感知信息，生成所述第一场景的路面几何信息；
根据所述路面几何信息，调整所述车辆内的悬架。
5. 根据权利要求1-4任一项所述的方法，其特征在于，所述至少基于所述感知信息控制车辆行驶，包括：至少根据所述感知信息，调整所述车辆的行驶路径，所述调整后的行驶路径不途径所述障碍物的体素所在的区域。
6. 根据权利要求1-5任一项所述的方法，其特征在于，所述采集数据包括图像数据和点云数据，所述感知检测网络包括图像特征提取网络、点云特征提取网络、特征融合网络和输出网络，其中，
所述图像特征提取网络，用于提取所述图像数据的3D图像特征；
所述点云特征提取网络，用于提取所述点云数据对应的体素的点云特征；
所述特征融合网络，用于根据所述3D图像特征和所述点云数据对应的体素的点云特征进行融合，获得所述第一场景的体素的融合特征；
所述输出网络，用于处理所述第一场景的体素的融合特征并输出所述感知信息。
7. 根据权利要求6所述的方法，其特征在于，所述方法还包括：
将文本查询信息和所述障碍物的体素的融合特征输入至属性识别网络，输出所述障碍物的类别信息；所述文本查询信息用于请求查询类别；
显示所述障碍物的类别信息；

其中，所述障碍物的体素的融合特征基于所述障碍物对应的多边形框的角点信息和所述第一场景的体素的融合特征确定，所述障碍物对应的多边形框与所述障碍物的体素关联。

8. 根据权利要求6或7所述的方法，其特征在于，所述方法还包括：

获取车辆的多条规划路径；

将所述车辆的多条规划路径和所述第一场景的体素的融合特征输入至路径评估网络，输出所述多条规划路径的推荐系数和所述多条规划路径中的推荐路径，所述推荐路径与所述多条规划路径的推荐系数关联；

显示所述推荐路径。

9. 一种用于智能驾驶的系统，其特征在于，所述系统包括：

感知检测网络，用于根据传感器对第一场景的采集数据输出感知信息，所述感知信息用于指示所述第一场景的障碍物的体素；所述传感器包括摄像头和雷达中的至少一种；

属性识别网络，用于根据文本查询信息和所述障碍物的体素的融合特征输出所述障碍物的类别信息，所述障碍物的体素的融合特征基于所述障碍物对应的多边形框的角点信息和所述第一场景的体素的融合特征确定，所述障碍物对应的多边形框与所述障碍物的体素关联，所述第一场景的体素的融合特征为所述感知检测网络基于从所述采集数据中提取的3D图像特征和体素的点云特征中的至少一项进行时间和/或空间上的融合获得；

路径评估网络，用于根据多条规划路径和所述第一场景的体素的融合特征输出所述多条规划路径的推荐系数和所述多条规划路径中的推荐路径，所述推荐路径与所述多条规划路径的推荐系数关联。

10. 根据权利要求9所述的系统，其特征在于，所述感知信息包括以下信息的至少一项：

所述第一场景的体素的占据状态、所述第一场景的体素的速度信息、所述第一场景的体素的可见状态和所述障碍物对应的多边形框的角点信息；

其中，所述障碍物对应的多边形框与所述障碍物的体素关联。

11. 根据权利要求9或10所述的系统，其特征在于，所述采集数据包括图像数据和点云数据，所述感知检测网络包括图像特征提取网络、点云特征提取网络、特征融合网络和输出网络，其中，

所述图像特征提取网络，用于提取所述图像数据的3D图像特征；

所述点云特征提取网络，用于提取所述点云数据对应的体素的点云特征；

所述特征融合网络，用于根据所述3D图像特征和所述点云数据对应的体素的点云特征进行融合，获得所述第一场景的体素的融合特征；

所述输出网络，用于处理所述第一场景的体素的融合特征并输出所述感知信息。

12. 根据权利要求9-11任一项所述的系统，其特征在于，所述属性识别网络包括文本编码网络和属性解码网络，其中，

所述文本编码网络，用于提取文本查询信息的词向量特征；

所述属性解码网络，用于根据所述词向量特征和所述障碍物的体素的融合特征输出所述障碍物的类别信息。

13. 根据权利要求 9-12 任一项所述的系统，其特征在于，所述路径评估网络包括路径编码网络、特征交互网络和评估输出网络，其中，

所述路径编码网络，用于提取多条规划路径中每条规划路径的路径特征；

所述特征交互网络，用于根据每条规划路径的路径特征和所述第一场景的体素的融合特征获得所述每条规划路径的风险特征；

所述评估输出网络，用于根据所述多条规划路径的风险特征输出所述多条规划路径的推荐系数和所述多条规划路径中的推荐路径。

14. 一种用于智能驾驶的装置，其特征在于，所述装置包括：

接收单元，用于获取传感器对第一场景的采集数据，所述传感器包括摄像头和雷达中的至少一种；

处理单元，用于将所述采集数据输入至感知检测网络，输出感知信息，所述感知信息用于指示所述第一场景的障碍物的体素；

所述处理单元，还用于至少基于所述感知信息控制车辆行驶。

15. 一种计算装置，其特征在于，所述计算装置包括存储器和处理器，所述存储器用于存储程序指令；在所述处理器执行所述存储器中的程序指令时，所述计算装置执行如权利要求 1-8 中任一项所述的方法。

16. 一种车辆，其特征在于，所述车辆包括上述如权利要求 9-13 任一项所述的系统，或者，包括如权利要求 14 或 15 所述的装置。

17. 一种计算机可读存储介质，其特征在于，所述计算机可读存储介质存储有程序指令，所述程序指令用于实现权利要求 1-8 任一项所述的方法。

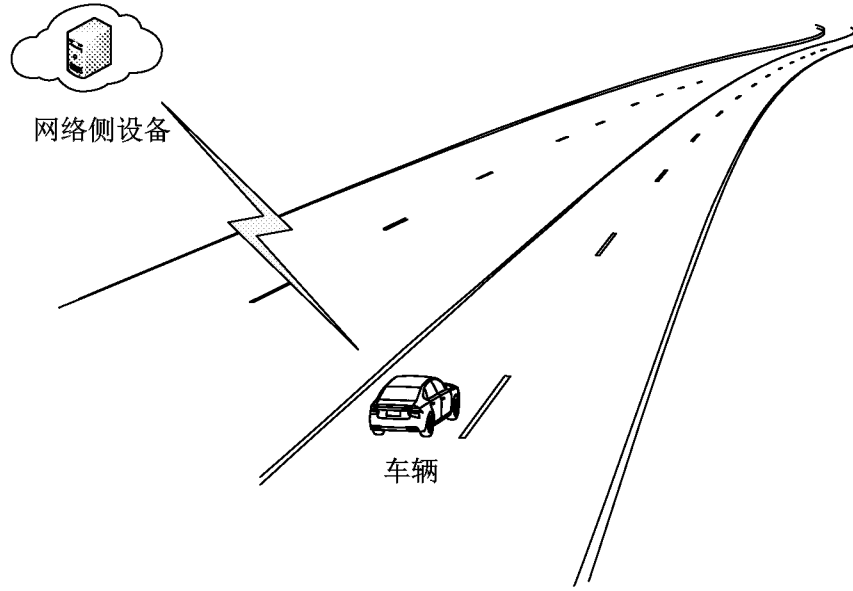


图 1

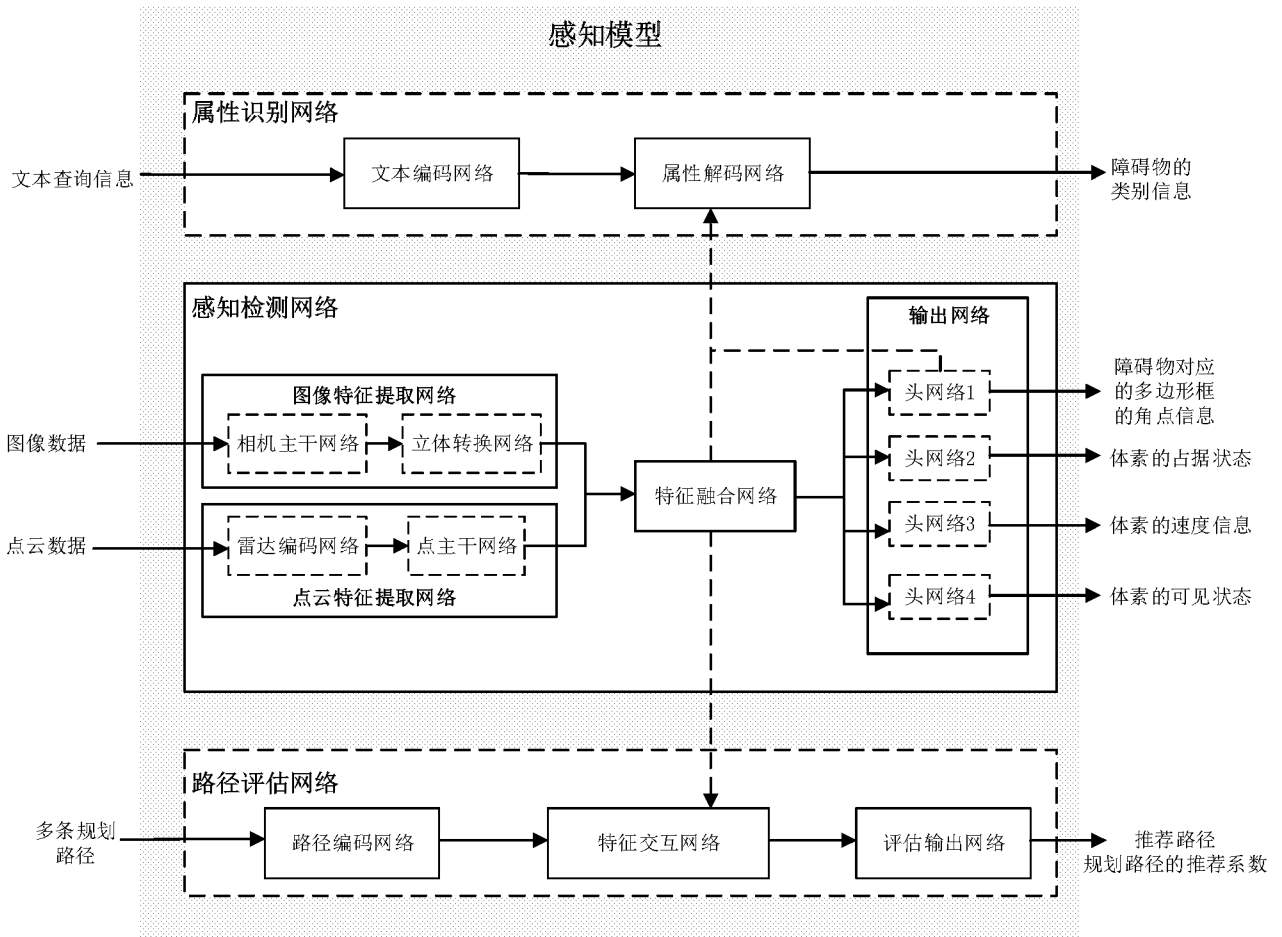


图 2

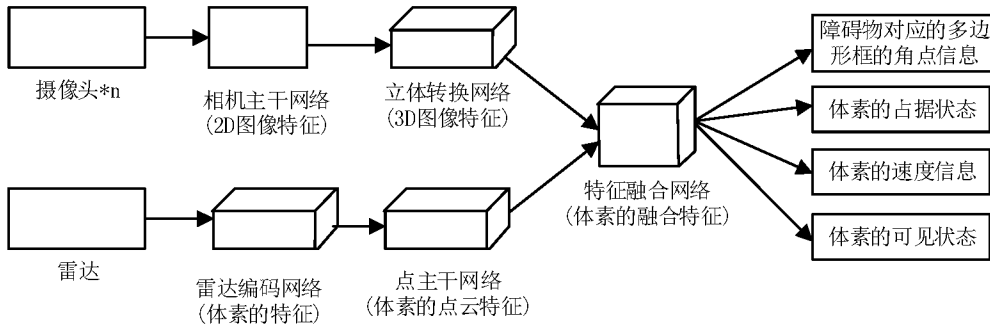


图 3

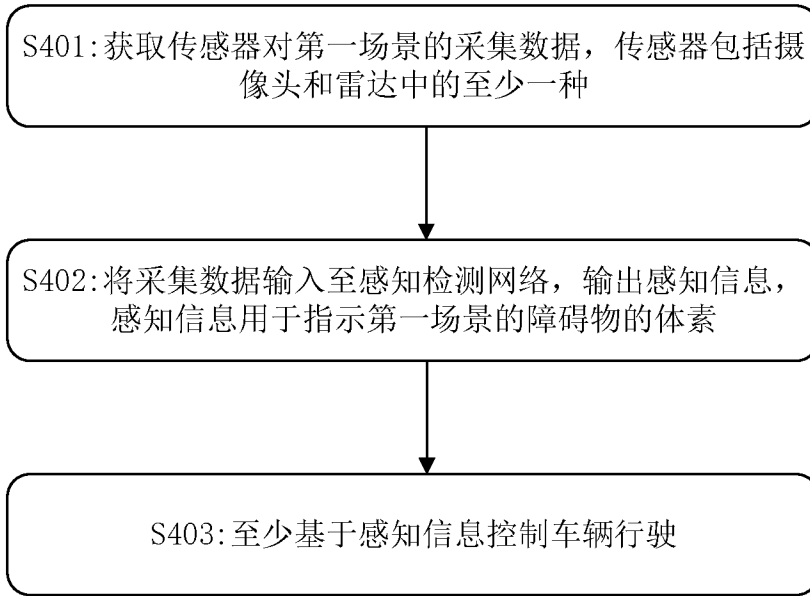


图 4

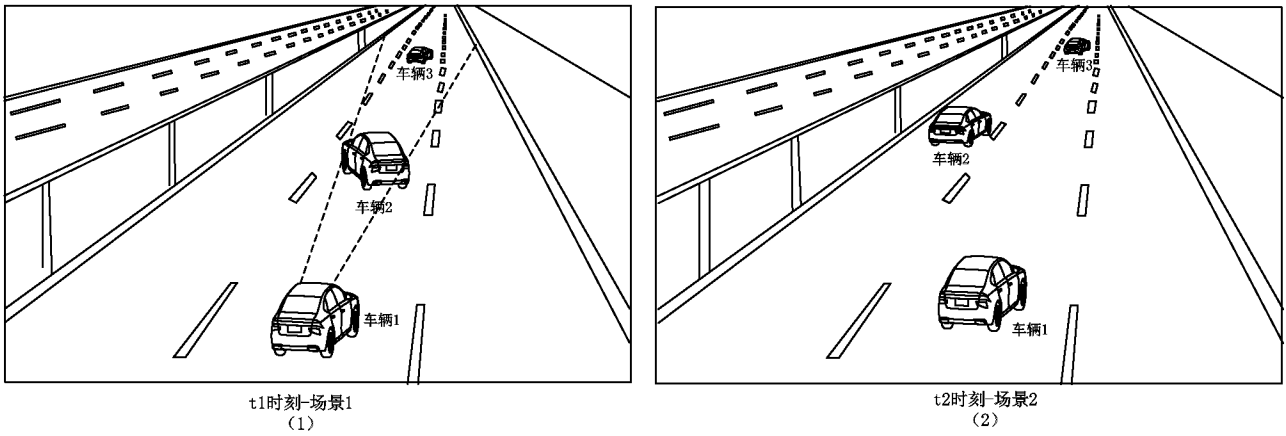


图 5

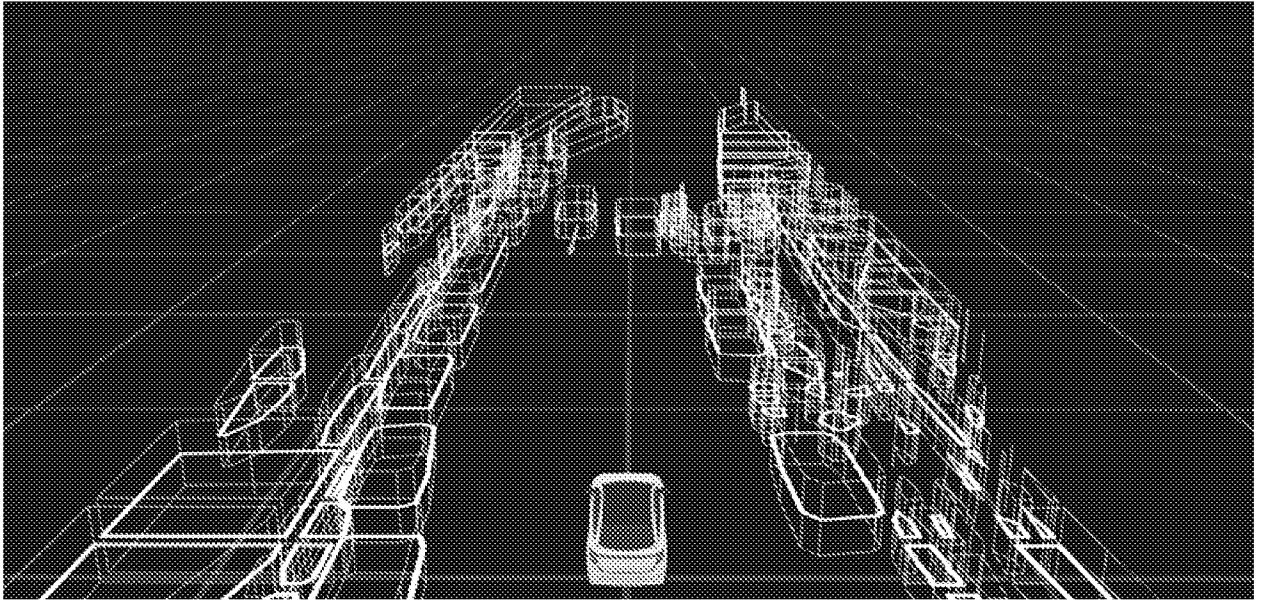


图 6A

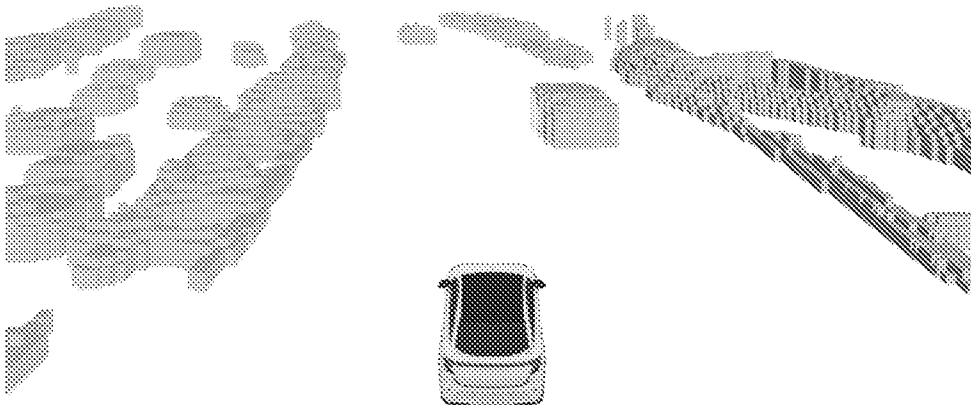


图 6B

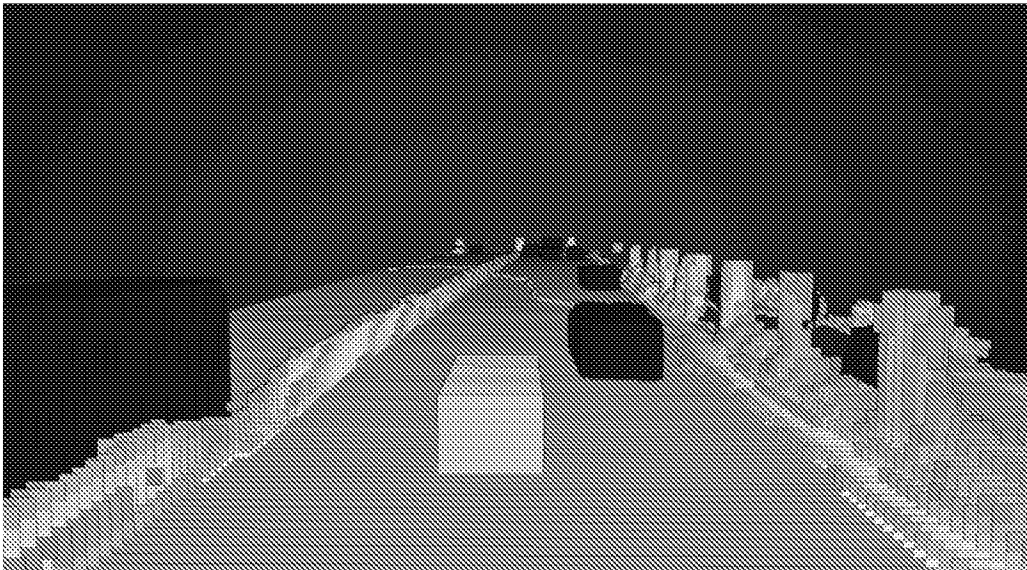


图 6C

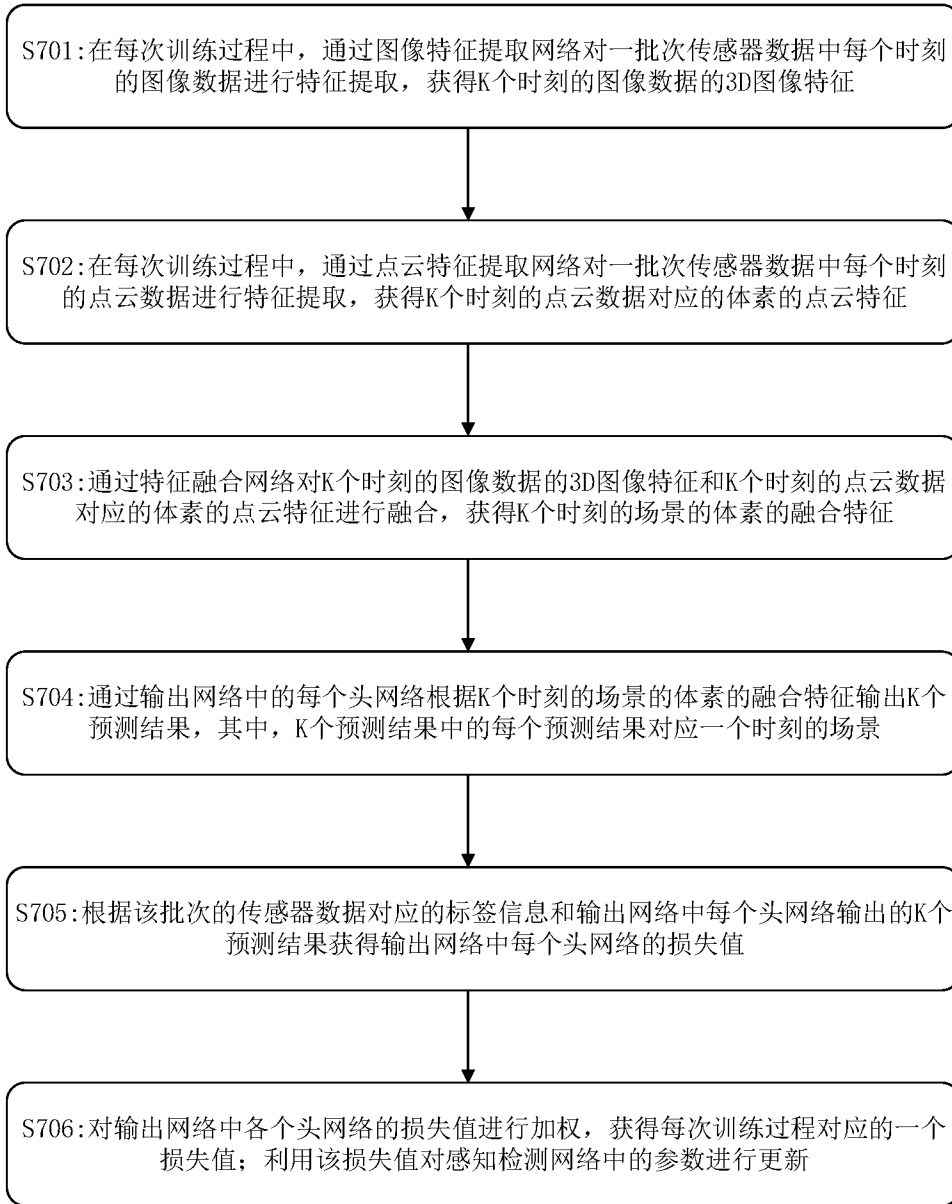


图 7A

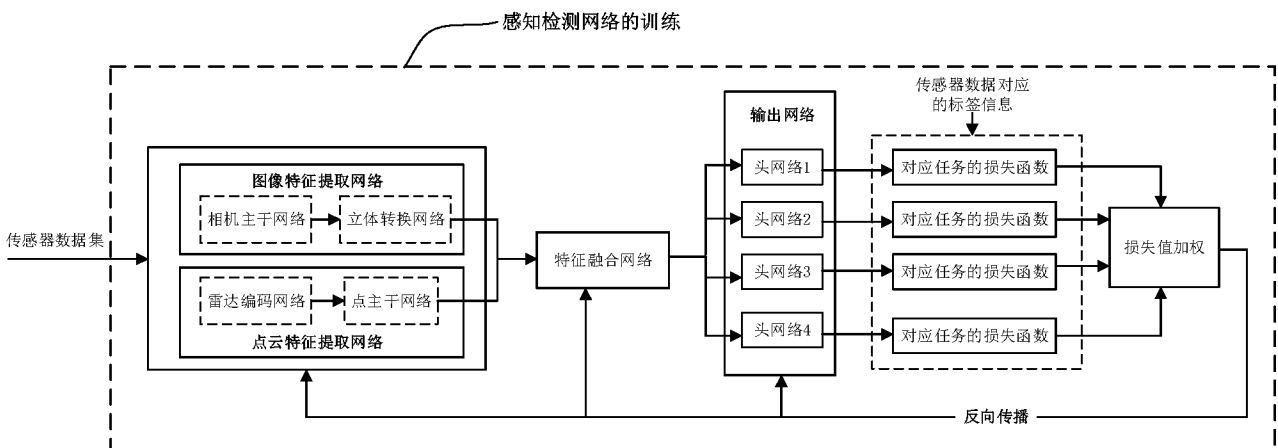
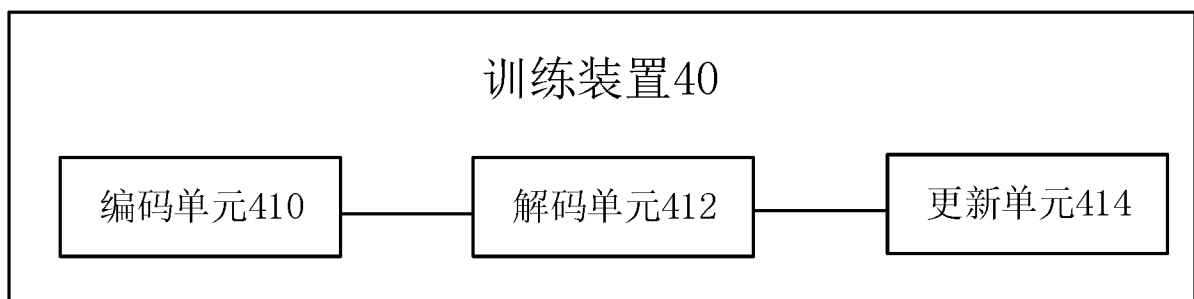
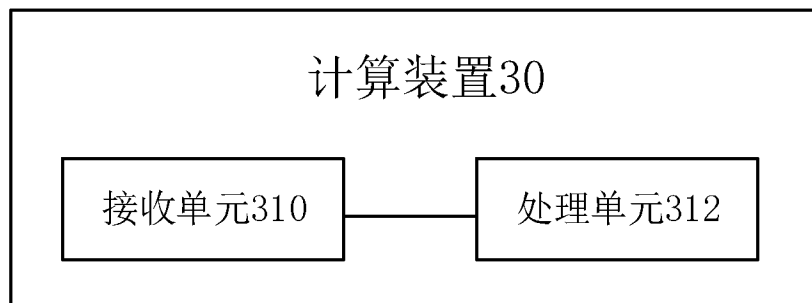
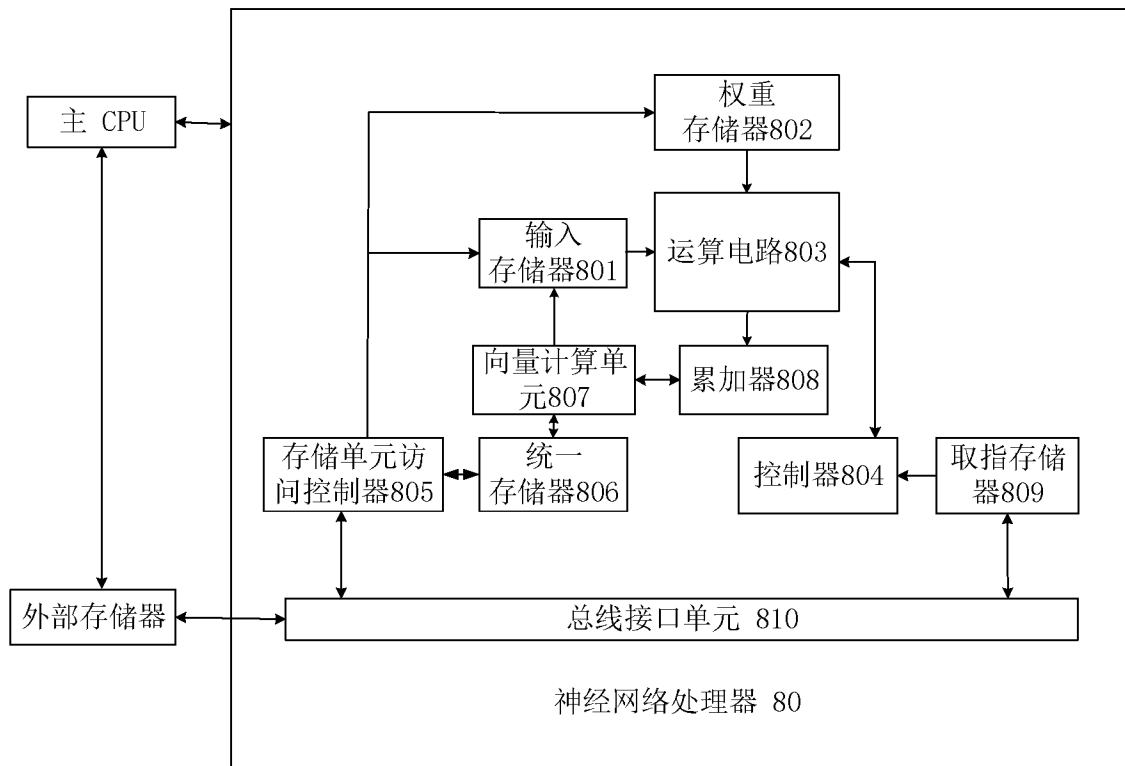


图 7B



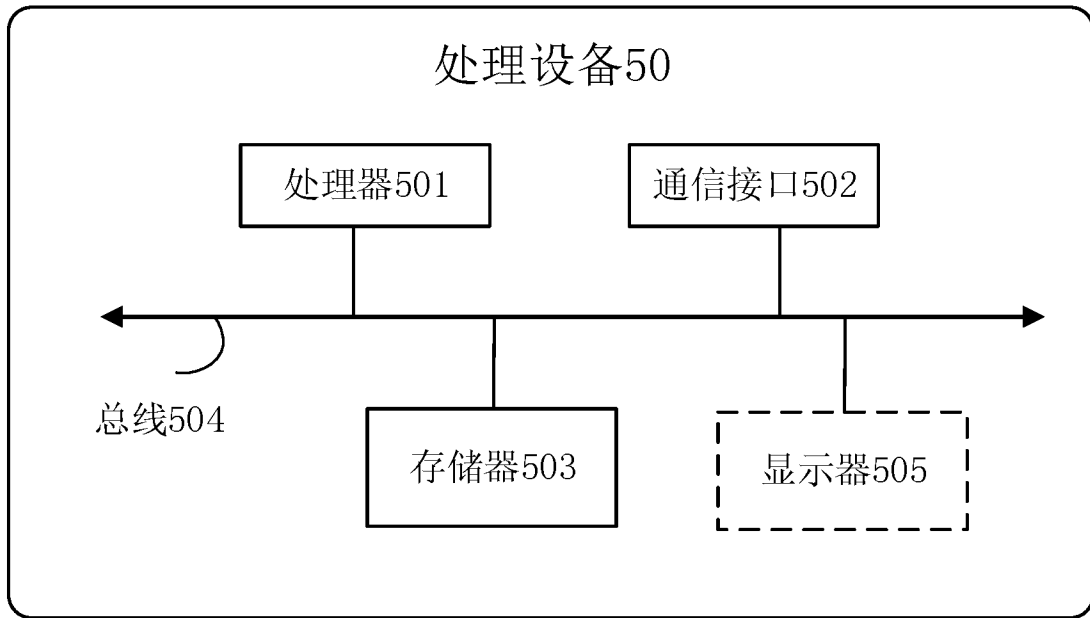


图 10

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/100778

A. CLASSIFICATION OF SUBJECT MATTER B60W30/09(2012.01)j According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC: B60W Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) CNABS, CNTXT, ENTXTC, CJFD, CNKI: 自动驾驶, 智能驾驶, 障碍物, 摄像头, 雷达, 路径, 推荐, 图像, 点云, 融合, 文本; USTXT, EPTXT, GBTXT, WOTXT, ENTXT: automatic drive, obstacle, camera, radar, path, recommend, picture, point cloud, fusion, text.		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 112703144 A (HUAWEI TECHNOLOGIES CO., LTD.) 23 April 2021 (2021-04-23) description, paragraphs 51-221, and figures 1-7	1-5, 14-17
Y	CN 112703144 A (HUAWEI TECHNOLOGIES CO., LTD.) 23 April 2021 (2021-04-23) description, paragraphs 51-221, and figures 1-7	6, 8
Y	CN 109145680 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.) 04 January 2019 (2019-01-04) description, paragraphs 30-89, and figures 1-3	6, 8
X	CN 113859267 A (GUANGZHOU XPENG AUTOMATIC DRIVING TECHNOLOGY CO., LTD.) 31 December 2021 (2021-12-31) description, paragraphs 59-176, and figures 1-9	1-5, 14-17
Y	CN 113859267 A (GUANGZHOU XPENG AUTOMATIC DRIVING TECHNOLOGY CO., LTD.) 31 December 2021 (2021-12-31) description, paragraphs 59-176, and figures 1-9	6, 8
A	CN 111665852 A (CHINA FAW CO., LTD.) 15 September 2020 (2020-09-15) entire document	1-17
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "D" document cited by the applicant in the international application "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 26 February 2024		Date of mailing of the international search report 29 February 2024
Name and mailing address of the ISA/CN China National Intellectual Property Administration (ISA/CN) China No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088		Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No. PCT/CN2023/100778

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	112703144	A	23 April 2021	WO	2022133684	A1	30 June 2022
				EP	4261092	A1	18 October 2023

CN	109145680	A	04 January 2019	US	2018365503	A1	20 December 2018
				US	10685245	B2	16 June 2020
				CN	109145680	B	27 May 2022

CN	113859267	A	31 December 2021	CN	113859267	B	25 August 2023

CN	111665852	A	15 September 2020	CN	111665852	B	06 September 2022

KR	20230079855	A	07 June 2023	KR	102585557	B1	06 October 2023

<p>A. 主题的分类</p> <p>B60W30/09(2012.01)i</p> <p>按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类</p>																																				
<p>B. 检索领域</p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p>IPC: B60W</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))</p> <p>CNABS, CNTXT, ENTXTC, CJFD, CNKI:自动驾驶, 智能驾驶, 障碍物, 摄像头, 雷达, 路径, 推荐, 图像, 点云, 融合, 文本; USTXT, EPTXT, GBTXT, WOTXT, ENTXT: automatic drive, obstacle, camera, radar, path, recommend, picture, point cloud, fusion, text.</p>																																				
<p>C. 相关文件</p> <table border="1"> <thead> <tr> <th>类型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>CN 112703144 A (华为技术有限公司) 2021年4月23日 (2021 - 04 - 23) 说明书第51-221段及附图1-7</td> <td>1-5、14-17</td> </tr> <tr> <td>Y</td> <td>CN 112703144 A (华为技术有限公司) 2021年4月23日 (2021 - 04 - 23) 说明书第51-221段及附图1-7</td> <td>6、8</td> </tr> <tr> <td>Y</td> <td>CN 109145680 A (百度在线网络技术(北京)有限公司) 2019年1月4日 (2019 - 01 - 04) 说明书第30-89段及附图1-3</td> <td>6、8</td> </tr> <tr> <td>X</td> <td>CN 113859267 A (广州小鹏自动驾驶科技有限公司) 2021年12月31日 (2021 - 12 - 31) 说明书第59-176段及附图1-9</td> <td>1-5、14-17</td> </tr> <tr> <td>Y</td> <td>CN 113859267 A (广州小鹏自动驾驶科技有限公司) 2021年12月31日 (2021 - 12 - 31) 说明书第59-176段及附图1-9</td> <td>6、8</td> </tr> <tr> <td>A</td> <td>CN 111665852 A (中国第一汽车股份有限公司) 2020年9月15日 (2020 - 09 - 15) 全文</td> <td>1-17</td> </tr> <tr> <td>A</td> <td>KR 20230079855 A (WISE AUTOMOTIVE CORP) 2023年6月7日 (2023 - 06 - 07) 全文</td> <td>1-17</td> </tr> </tbody> </table> <p><input type="checkbox"/> 其余文件在C栏的续页中列出。 <input checked="" type="checkbox"/> 见同族专利附件。</p> <p>* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “D” 申请人在国际申请中引证的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件</p> <table border="1"> <tr> <td>国际检索实际完成的日期</td> <td>国际检索报告邮寄日期</td> </tr> <tr> <td>2024年2月26日</td> <td>2024年2月29日</td> </tr> <tr> <td>ISA/CN的名称和邮寄地址</td> <td>授权官员</td> </tr> <tr> <td>中国国家知识产权局 中国北京市海淀区蓟门桥西土城路6号 100088</td> <td>张华</td> </tr> <tr> <td></td> <td>电话号码 (+86) 020-28958775</td> </tr> </table>			类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	X	CN 112703144 A (华为技术有限公司) 2021年4月23日 (2021 - 04 - 23) 说明书第51-221段及附图1-7	1-5、14-17	Y	CN 112703144 A (华为技术有限公司) 2021年4月23日 (2021 - 04 - 23) 说明书第51-221段及附图1-7	6、8	Y	CN 109145680 A (百度在线网络技术(北京)有限公司) 2019年1月4日 (2019 - 01 - 04) 说明书第30-89段及附图1-3	6、8	X	CN 113859267 A (广州小鹏自动驾驶科技有限公司) 2021年12月31日 (2021 - 12 - 31) 说明书第59-176段及附图1-9	1-5、14-17	Y	CN 113859267 A (广州小鹏自动驾驶科技有限公司) 2021年12月31日 (2021 - 12 - 31) 说明书第59-176段及附图1-9	6、8	A	CN 111665852 A (中国第一汽车股份有限公司) 2020年9月15日 (2020 - 09 - 15) 全文	1-17	A	KR 20230079855 A (WISE AUTOMOTIVE CORP) 2023年6月7日 (2023 - 06 - 07) 全文	1-17	国际检索实际完成的日期	国际检索报告邮寄日期	2024年2月26日	2024年2月29日	ISA/CN的名称和邮寄地址	授权官员	中国国家知识产权局 中国北京市海淀区蓟门桥西土城路6号 100088	张华		电话号码 (+86) 020-28958775
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求																																		
X	CN 112703144 A (华为技术有限公司) 2021年4月23日 (2021 - 04 - 23) 说明书第51-221段及附图1-7	1-5、14-17																																		
Y	CN 112703144 A (华为技术有限公司) 2021年4月23日 (2021 - 04 - 23) 说明书第51-221段及附图1-7	6、8																																		
Y	CN 109145680 A (百度在线网络技术(北京)有限公司) 2019年1月4日 (2019 - 01 - 04) 说明书第30-89段及附图1-3	6、8																																		
X	CN 113859267 A (广州小鹏自动驾驶科技有限公司) 2021年12月31日 (2021 - 12 - 31) 说明书第59-176段及附图1-9	1-5、14-17																																		
Y	CN 113859267 A (广州小鹏自动驾驶科技有限公司) 2021年12月31日 (2021 - 12 - 31) 说明书第59-176段及附图1-9	6、8																																		
A	CN 111665852 A (中国第一汽车股份有限公司) 2020年9月15日 (2020 - 09 - 15) 全文	1-17																																		
A	KR 20230079855 A (WISE AUTOMOTIVE CORP) 2023年6月7日 (2023 - 06 - 07) 全文	1-17																																		
国际检索实际完成的日期	国际检索报告邮寄日期																																			
2024年2月26日	2024年2月29日																																			
ISA/CN的名称和邮寄地址	授权官员																																			
中国国家知识产权局 中国北京市海淀区蓟门桥西土城路6号 100088	张华																																			
	电话号码 (+86) 020-28958775																																			

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2023/100778

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	112703144	A	2021年4月23日	WO	2022133684	A1	2022年6月30日
				EP	4261092	A1	2023年10月18日
CN	109145680	A	2019年1月4日	US	2018365503	A1	2018年12月20日
				US	10685245	B2	2020年6月16日
				CN	109145680	B	2022年5月27日
CN	113859267	A	2021年12月31日	CN	113859267	B	2023年8月25日
CN	111665852	A	2020年9月15日	CN	111665852	B	2022年9月6日
KR	20230079855	A	2023年6月7日	KR	102585557	B1	2023年10月6日