



(12) 发明专利申请

(10) 申请公布号 CN 113065590 A

(43) 申请公布日 2021.07.02

(21) 申请号 202110326528.8

G06T 5/50 (2006.01)

(22) 申请日 2021.03.26

G06T 5/00 (2006.01)

(71) 申请人 清华大学

G06N 3/08 (2006.01)

地址 100084 北京市海淀区清华园1号

G06N 3/04 (2006.01)

G01S 17/89 (2020.01)

(72) 发明人 张新钰 李骏 王力 兰天蔚

李志伟 卢一倩

(74) 专利代理机构 北京三聚阳光知识产权代理有限公司 11250

代理人 张建国

(51) Int. Cl.

G06K 9/62 (2006.01)

G06K 9/00 (2006.01)

G06T 7/80 (2017.01)

G06T 7/38 (2017.01)

G06T 7/11 (2017.01)

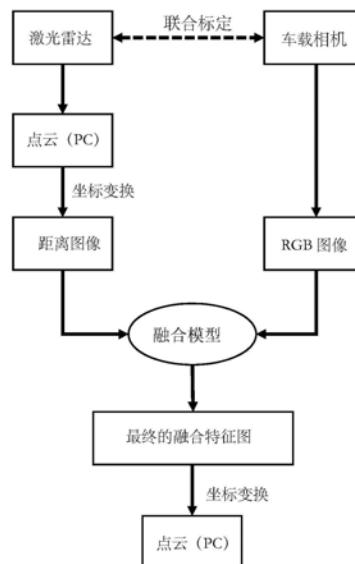
权利要求书2页 说明书7页 附图3页

(54) 发明名称

一种基于注意力机制的视觉与激光雷达多模态数据融合方法

(57) 摘要

本发明公开了一种基于注意力机制的视觉与激光雷达多模态数据融合方法,所述方法包括:同时采集自动驾驶车辆的激光雷达点云数据以及相机RGB图像;对激光雷达点云数据以及相机RGB图像进行预处理;通过球坐标变换将点云数据转换成距离图像;将距离图像和RGB图像输入预先建立和训练好的距离图像融合模型,输出最终的融合特征图;所述距离图像融合模型采用自注意机制对距离图像和RGB图像进行融合;将最终的融合特征图与带有坐标信息的距离图像连接到一起,通过空间坐标转换从二维图像还原到空间点云的表示形式。本发明的方法使用自注意机制有效地融合了多源信息,使两种数据优势互补,提高了3D目标检测算法的准确性。



1. 一种基于注意力机制的视觉与激光雷达多模态数据融合方法,所述方法包括:
 同时采集自动驾驶车辆的激光雷达点云数据以及相机RGB图像;
 对激光雷达点云数据以及相机RGB图像进行预处理;
 通过球坐标变换将点云数据转换成距离图像;
 将距离图像和RGB图像输入预先建立和训练好的距离图像融合模型,输出最终的融合特征图;所述距离图像融合模型采用自注意机制对距离图像和RGB图像进行融合;
 将最终的融合特征图与带有坐标信息的距离图像连接到一起,通过空间坐标转换从二维图像还原到空间点云的表示形式。

2. 根据权利要求1所述的基于注意力机制的视觉与激光雷达多模态数据融合方法,其特征在于,所述对激光雷达点云数据以及相机RGB图像进行预处理;具体包括:
 对激光雷达点云数据以及相机RGB图像进行标定和对齐;
 根据点云数据与RGB图像的对应信息,对RGB图像进行裁剪,只保留图像下半部分三分之一的尺寸。

3. 根据权利要求2所述的基于注意力机制的视觉与激光雷达多模态数据融合方法,其特征在于,所述通过球坐标变换将点云数据转换成距离图像;具体包括:
 将原始的激光雷达点云数据表示为 $[N, 4]$ 的矩阵,其中, N 为一帧场景中的点云数量,每一个点的四维特征为 $[x, y, z, i]$, x, y, z 表示空间坐标, i 表示激光反射强度;
 将点云数据通过球坐标变换转换成距离图像,具体变换公式为:

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \frac{1}{2}[1 - \arctan(y, x)\pi^{-1} \times w \\ [1 - (\arcsin(z, r) + f_{down})f^{-1} \times h \end{pmatrix}$$

其中, u, v 代表点云转换后在距离图像中的像素坐标; f 是激光雷达的俯仰角, f_{down} 为中心点向下的角度, $f = f_{down} + f_{up}$, f_{up} 为中心点向上的角度; w 和 h 分别表示转换后的距离图像的宽度和高度; r 代表点距离激光雷达的中心的绝对距离:

$$r = \sqrt{x^2 + y^2 + z^2}$$

距离图像的每一个像素点对应空间中的一个点,包含了点云数据的5个特征: x, y, z 、激光反射强度 i 和深度 r 。

4. 根据权利要求3所述的基于注意力机制的视觉与激光雷达多模态数据融合方法,其特征在于,所述距离图像融合模型包括:encoder模块、第一注意力特征融合模块、第二注意力特征融合模块和decoder模块,其中,encoder模块和decoder模块是对称的;

所述encoder模块包括:对距离图像进行处理的第一Range降采样子模块、对距离图像进行处理的第二Range降采样子模块、对RGB图像进行处理的RGB降采样子模块以及对融合图像进行处理的降采样子模块;其中,第一Range降采样子模块和第二Range降采样子模块均包含一层空洞卷积单元、一层批归一化单元以及一层激活函数单元;RGB降采样子模块包含一层卷积单元、一层批归一化单元以及一层激活函数单元;降采样子模块包含一层卷积单元、一层批归一化单元以及一层激活函数单元;

所述第一注意力特征融合模块和第二注意力特征融合模块均包括:距离图特征提取单元、第一RGB图像特征提取单元、第二RGB图像特征提取单元、注意力图生成单元、融合向量

生成单元以及融合特征图生成单元;距离图特征提取单元、第一RGB图像特征提取单元和第二RGB图像特征提取单元均由两层卷积层组成;

所述距离图特征提取单元,用于提取距离图像的Range特征,输出至注意力图生成单元;

所述第一RGB图像特征提取单元,用于提取RGB图像的RGB特征,输出至注意力图生成单元;

所述注意力图生成单元,用于将Range特征矩阵转换为Range向量,将RGB特征矩阵转换为RGB向量,然后将Range向量和RGB向量进行叉乘得到注意力图,输出至融合向量生成单元;

所述第二RGB图像特征提取单元,用于提取RGB图像的RGB特征,输入融合向量生成单元;

所述融合向量生成单元,用于将RGB特征转换成向量,与注意力图进行叉乘得到融合向量,输出至融合特征图生成单元;

所述融合特征图生成单元,用于将融合向量进行重新拼接,形成融合特征图;

decoder模块,用于将融合特征图恢复到原始输入图像的尺寸,包括两个上采样子模块,每个上采样子模块都是由一层反卷积、一层批归一化以及一层激活层组成。

5. 根据权利要求4所述的基于注意力机制的视觉与激光雷达多模态数据融合方法,其特征在于,所述方法还包括:利用数据集KITTI对距离图像融合模型进行训练的步骤。

6. 根据权利要求5所述的基于注意力机制的视觉与激光雷达多模态数据融合方法,其特征在于,所述将距离图像和RGB图像输入预先建立和训练好的距离图像融合模型,输出最终的融合特征图;具体包括:

将距离图像R输入第一Range降采样子模块,输出距离图像R1;将RGB图像I输入第一RGB降采样子模块,输出与距离图像R1相同尺寸的RGB图像I1;

将距离图像R1和RGB图像I1输入第一注意力特征融合模块,输出融合特征图RI1;

将距离图像R1输入第二Range降采样子模块,输出距离图像R2,将融合特征图RI1输入降采样子模块,输出与距离图像R2相同尺寸的融合特征图RI2;

将距离图像R2和融合特征图RI2输入第二注意力特征融合模块,输出融合特征图RI3;

将融合特征图RI3输入Decoder模块,输出与原始的距离图像R尺寸相同的最终的融合特征图,其通道数为8。

7. 一种计算机设备,包括存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序,其特征在于,所述处理器执行所述计算机程序时实现权利要求1-6任一项所述的方法。

8. 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计算机程序,所述计算机程序当被处理器执行时使所述处理器执行权利要求1-6任一项所述的方法。

一种基于注意力机制的视觉与激光雷达多模态数据融合方法

技术领域

[0001] 本发明涉及自动驾驶技术领域,提出一种基于注意力机制的视觉与激光雷达多模态数据融合方法。

背景技术

[0002] 近些年来,无人驾驶技术领域发展迅速,在无人驾驶技术中感知技术是不可或缺的一环。无人驾驶车辆依赖激光雷达、相机、轮速里程计等传感器对环境进行感知,如车道线检测、3D目标检测都属于无人驾驶技术的基本功能。环境感知可以为汽车反馈道路信息,以便对车辆的驾驶行为提供更为精准的控制,从而提高驾驶的安全性和舒适性。

[0003] 作为自动驾驶(辅助驾驶)系统的基础功能之一,3D目标检测可以认为是环境感知中极为重要的一环,主要是检测场景中的车辆、行人等交通参与者。该功能为自动驾驶车辆后续的规划和控制提供空间障碍物的分布信息,关系到驾驶安全性问题,如果出现错误则有可能造成非常严重的后果。

[0004] 当前,国内外均对3D目标检测技术有了一定的研究,从硬件角度出发,针对3D目标检测任务的数据输入来源主要来自于相机和激光雷达这两类传感器。相机作为针孔模型的感光元器件,会提供类似于人眼直观感受的RGB图像信息;而激光雷达作为一种主动采集式的传感器,能够通过发射接收激光,得到场景中目标的真实空间坐标信息。

[0005] 从领域当前的研究现状来看,这两种传感器都广泛应用于3D目标检测。然而,这两种传感器都有着很明显的缺陷。对于相机来说,RGB图像作为一种稠密的2D表现方式,通过像素位置和像素数值来表达空间信息,压缩了Z轴的深度信息,这会导致3D目标检测任务中难以预测物体的空间绝对位置;另外,相机镜头还对光线的变化十分敏感,时常会出现过曝或者过暗的情形,在这种场景下采集到的数据难以辨识,缺失有效信息。而对于激光雷达来说,虽然可以得到目标精确的空间位置信息,但是激光雷达作为主动式传感器,只能接受由传感器发射的激光信号,无法接受物体通过可见光反射的信号,所以无法得到物体的颜色纹理等特征信息;再者,激光的波长十分短,这又导致激光雷达的数据对于特殊天气十分敏感,比如在雨雪天气下得到的激光雷达点云通常会出现很多噪点。

[0006] 综上所述,如果使用图像作为主要的输入信息,在光照变化等条件导致图像信息缺失的时候,检测系统将会出现严重的失灵,此外图像中缺乏空间距离信息也造成了检测的短板。激光雷达点云虽然可以克服光照变化的问题,但点云的稀疏性和无序性也限制了其检测能力。因此,针对单纯使用车载相机或激光雷达时检测能力不足的问题,一种新颖的思路是考虑结合更多的传感器,并进行数据融合以提高对环境感知的能力。但现阶段大量的实验表明,如果只是简单的对激光雷达获得的点云信息和相机获得的RGB图像做像素级融合,不仅不会对感知算法的精度提升,反而会起到负面作用,通常都会在一定程度上降低感知精度。

发明内容

[0007] 本发明的目的在于克服上述技术缺陷,提出了一种基于注意力机制的视觉和激光雷达多模态数据融合方法,可以有效地通过RGB特征提升点云的信息量,该过程是一种点云特征增强的方式,使用增强后的点云做3D目标检测可以验证该方法的有效性。

[0008] 为实现上述目的,本发明的实施例1提出了一种基于注意力机制的视觉与激光雷达多模态数据融合方法,所述方法包括:

[0009] 同时采集自动驾驶车辆的激光雷达点云数据以及相机RGB图像;

[0010] 对激光雷达点云数据以及相机RGB图像进行预处理;

[0011] 通过球坐标变换将点云数据转换成距离图像;

[0012] 将距离图像和RGB图像输入预先建立和训练好的距离图像融合模型,输出最终的融合特征图;所述距离图像融合模型采用自注意机制对距离图像和RGB图像进行融合;

[0013] 将最终的融合特征图与带有坐标信息的距离图像连接到一起,通过空间坐标转换从二维图像还原到空间点云的表示形式。

[0014] 作为上述方法的一种改进,所述对激光雷达点云数据以及相机RGB图像进行预处理;具体包括:

[0015] 对激光雷达点云数据以及相机RGB图像进行标定和对齐;

[0016] 根据点云数据与RGB图像的对应信息,对RGB图像进行裁剪,只保留图像下半部分三分之一的尺寸。

[0017] 作为上述方法的一种改进,所述通过球坐标变换将点云数据转换成距离图像;具体包括:

[0018] 将原始的激光雷达点云数据表示为 $[N, 4]$ 的矩阵,其中, N 为一帧场景中的点云数量,每一个点的四维特征为 $[x, y, z, i]$, x, y, z 表示空间坐标, i 表示激光反射强度;

[0019] 将点云数据通过球坐标变换转换成距离图像,具体变换公式为:

$$[0020] \quad \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \frac{1}{2} [1 - \arctan(y, x) \pi^{-1} \times w \\ [1 - (\arcsin(z, r) + f_{down}) f^{-1} \times h \end{pmatrix}$$

[0021] 其中, u, v 代表点云转换后在距离图像中的像素坐标; f 是激光雷达的俯仰角, f_{down} 为中心点向下的角度, $f = f_{down} + f_{up}$, f_{up} 为中心点向上的角度; w 和 h 分别表示转换后的距离图像的宽度和高度; r 代表点距离激光雷达的中心的绝对距离:

$$[0022] \quad r = \sqrt{x^2 + y^2 + z^2}$$

[0023] 距离图像的每一个像素点对应空间中的一个点,包含了点云数据的5个特征: x, y, z 、激光反射强度 i 和深度 r 。

[0024] 作为上述方法的一种改进,所述距离图像融合模型包括:encoder模块、第一注意力特征融合模块、第二注意力特征融合模块和decoder模块,其中,encoder模块和decoder模块是对称的;

[0025] 所述encoder模块包括:对距离图像进行处理的第一Range降采样子模块、对距离图像进行处理的第二Range降采样子模块、对RGB图像进行处理的RGB降采样子模块以及对融合图像进行处理的降采样子模块;其中,第一Range降采样子模块和第二Range降采样子模块均包含一层空洞卷积单元、一层批归一化单元以及一层激活函数单元;RGB降采样子模

块包含一层卷积单元、一层批归一化单元以及一层激活函数单元；降采样子模块包含一层卷积单元、一层批归一化单元以及一层激活函数单元；

[0026] 所述第一注意力特征融合模块和第二注意力特征融合模块均包括：距离图特征提取单元、第一RGB图像特征提取单元、第二RGB图像特征提取单元、注意力图生成单元、融合向量生成单元以及融合特征图生成单元；距离图特征提取单元、第一RGB图像特征提取单元和第二RGB图像特征提取单元均由两层卷积层组成；

[0027] 所述距离图特征提取单元，用于提取距离图像的Range特征，输出至注意力图生成单元；

[0028] 所述第一RGB图像特征提取单元，用于提取RGB图像的RGB特征，输出至注意力图生成单元；

[0029] 所述注意力图生成单元，用于将Range特征矩阵转换为Range向量，将RGB特征矩阵转换为RGB向量，然后将Range向量和RGB向量进行叉乘得到注意力图，输出至融合向量生成单元；

[0030] 所述第二RGB图像特征提取单元，用于提取RGB图像的RGB特征，输入融合向量生成单元；

[0031] 所述融合向量生成单元，用于将RGB特征转换成向量，与注意力图进行叉乘得到融合向量，输出至融合特征图生成单元；

[0032] 所述融合特征图生成单元，用于将融合向量进行重新拼接，形成融合特征图；

[0033] decoder模块，用于将融合特征图恢复到原始输入图像的尺寸，包括两个上采样子模块，每个上采样子模块都是由一层反卷积、一层批归一化以及一层激活层组成。

[0034] 作为上述方法的一种改进，所述方法还包括：利用数据集KITTI对距离图像融合模型进行训练的步骤。

[0035] 作为上述方法的一种改进，所述将距离图像和RGB图像输入预先建立和训练好的距离图像融合模型，输出最终的融合特征图；具体包括：

[0036] 将距离图像R输入第一Range降采样子模块，输出距离图像R1；将RGB图像I输入第一RGB降采样子模块，输出与距离图像R1相同尺寸的RGB图像I1；

[0037] 将距离图像R1和RGB图像I1输入第一注意力特征融合模块，输出融合特征图RI1；

[0038] 将距离图像R1输入第二Range降采样子模块，输出距离图像R2，将融合特征图RI1输入降采样子模块，输出与距离图像R2相同尺寸的融合特征图RI2；

[0039] 将距离图像R2和融合特征图RI2输入第二注意力特征融合模块，输出融合特征图RI3；

[0040] 将融合特征图RI3输入Decoder模块，输出与原始的距离图像R尺寸相同的最终的融合特征图，其通道数为8。

[0041] 本发明的实施例2提供了一种计算机设备，包括存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序，所述处理器执行所述计算机程序时实现上述的方法。

[0042] 本发明的实施例3提出了一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，所述计算机程序当被处理器执行时使所述处理器执行上述的方法。

[0043] 本发明的优势在于：

[0044] 1、本发明的方法使用自注意机制有效地融合了多源信息,使两种数据优势互补,提高了3D目标检测算法的准确性;其中由于激光点云PC和RGB数据表达形式差别过大,难以直接融合,因此使用了球坐标变换作为连接源信息的桥梁;

[0045] 2、本发明的方法可以增加3D目标检测精度,尤其是对行人、骑行的人等小目标得到明显的精度提升。

附图说明

[0046] 为了更清楚的说明本发明,下面将对本发明中所需要使用的附图做一简单地介绍,显而易见地,下面描述中的附图是本发明的一些实施例,对于本发明的一些实施例,对于本领域普通技术人员,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0047] 图1为本发明的实施例1提供的基于自注意机制的视觉与激光雷达多模态数据融合方法的流程图;

[0048] 图2为本发明实施例1提供的距离图像融合(RI-Fusion)模型结构的示意图;

[0049] 图3为本发明实施例1提供的注意力特征融合(RI-Attention)模块示意图。

具体实施方式

[0050] 下面结合附图对本发明的技术方案进行详细说明。

[0051] 如图1所示,本发明的实施例1提供了一种基于注意力机制的视觉与激光雷达多模态数据融合方法,包括如下步骤:

[0052] 步骤1)采集自动驾驶场景下的激光雷达点云数据以及相机RGB图像,对两个数据完成标定和对齐;并根据点云数据与RGB图像的对应信息,对RGB图像做一定的修正;

[0053] 由于激光雷达和相机的成像原理不同,采集到的数据所在的空間不一定完全重合,通常点云投影在RGB图像上只在图像的下三分之一有分布,对于上面绝大部分的图像上是没有点云的,所以对RGB图像进行裁剪,只保留图像下面三分之一尺寸。

[0054] 步骤2)将激光雷达点云数据表示转换成距离图的表示;

[0055] 步骤2-1)将原始的激光雷达点云数据表示为 $[N, 4]$ 的矩阵,其中, N 为一帧场景中的点云数量,每一个点的四维特征 $[x, y, z, i]$,分别是空间 x, y, z 坐标和激光反射强度 i ;

[0056] 步骤2-2)将点云数据通过球坐标变换转换成一张2D的距离图像,具体变换公式为:

$$[0057] \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \frac{1}{2}[1 - \arctan(y, x)\pi^{-1} \times w \\ [1 - (\arcsin(z, r) + f_{down})f^{-1} \times h \end{pmatrix} \quad (1)$$

[0058] 其中, u, v 代表点云转换后在距离图像中的像素坐标, x, y 和 z 表示每一个点的三维坐标; f 是激光雷达的俯仰角, f_{down} 为中心点向下的角度, $f = f_{down} + f_{up}$, f_{up} 为中心点向上的角度; w 和 h 分别表示转换后的距离图的宽度和高度; r 代表点距离激光雷达的中心的绝对距离:

$$[0059] r = \sqrt{x^2 + y^2 + z^2} \quad (2)$$

[0060] 由此,就可以得到激光雷达数据的Range表示。转换后的每一个像素点对应空间中

的一个点,包含了点云的x、y、z、激光反射强度i和深度r总共五维信息。将PC转换后得到的2D图像称为距离图像(Range Image);

[0061] 步骤3) 利用距离图像融合模型对步骤2)的距离图像和步骤3)的修正后的RGB图像进行融合,输出融合特征图,即增强后的点云数据;

[0062] 图2所示为本发明的距离图像融合(RI-Fusion)模型结构图,该步骤主要包括网络结构设计、模型训练和模型测试三个部分。需要说明的是,距离图像融合(RI-Fusion)模型的搭建、训练、测试的硬件和软件平台为Ubuntu16.04+RTX3090+pytorch1.7+cuda11.1+python3.7。

[0063] 1) 距离图像融合模型(RI-Fusion)的结构包括:encoder模块、两个注意力特征融合模块和decoder模块,encoder模块和decoder模块是对称的。encoder模块侧重于从输入数据中学习2D信息的特征,需要注意的是这里同时输入RGB和Range图都过相同的卷积神经网络进行特征提取,但这两部分的卷积层参数是不共享的。将提取到的特征送入注意力特征融合(RI-Attention)模块进行特征融合。decoder模块则使用反卷积进行上采样,将encoder模块学习的特征以及两源信息融合到的特征反向还原到原始的尺度。

[0064] encoder模块分别包括对距离图像(Range)进行处理的一个Range降采样子模块、对RGB图像进行处理的RGB降采样子模块以及对融合图像进行处理的降采样子模块。其中Range降采样子模块包含一层空洞卷积单元、一层批归一化单元以及一层激活函数单元;RGB降采样子模块包含一层卷积单元、一层批归一化单元以及一层激活函数单元,降采样子模块包含一层卷积单元、一层批归一化单元以及一层激活函数单元。

[0065] 图3所示为本发明的注意力特征融合模块。该模块基于transformer的思想,借此通过Range信息关注RGB中更关键的信息。

[0066] 该模块的具体实现方式:

[0067] 首先确保输入的两类信息尺度一致,所以首先对RGB做自适应池化将RGB调整到与Range图像相同的尺寸。

[0068] Range图像和RGB图像分别通过两层卷积层得到Range特征和RGB特征,然后将两部分特征从矩阵转换为向量并进行叉乘得到注意力图;

[0069] RGB图像再一次通过两层卷积层的到新的RGB特征,然后将其转换成向量后与注意力图进行叉乘,再将得到的新的向量重新拼接成特征图。

[0070] decoder模块将高维语义特征恢复到原始输入大小,具体由两个上采样子模块构成,每个上采样子模块都是由一层反卷积、一层批归一化以及一层激活层组成。

[0071] 如图2所示, $5 \times 48 \times 512$ (5代表通道数:x、z、y、强度和深度)的距离图像R输入第一Range降采样子模块,输出 $6 \times 24 \times 256$ 的距离图像R1; $3 \times 48 \times 512$ (3代表通道数:R、G和B)的RGB图像I输入RGB降采样子模块,输出 $16 \times 24 \times 256$ 的RGB图像I1; $6 \times 24 \times 256$ 的距离图像R1和 $16 \times 24 \times 256$ 的RGB图像I1输入第一注意力特征融合模块,输出 $16 \times 24 \times 256$ 的融合特征图RI1;

[0072] $16 \times 24 \times 256$ 的距离图像R1输入第二Range降采样子模块,输出 $32 \times 12 \times 128$ 的距离图像R2, $16 \times 24 \times 256$ 的融合特征图RI1输入降采样子模块,输出 $32 \times 12 \times 128$ 的融合特征图RI2; 将 $32 \times 12 \times 128$ 的距离图像R2和 $32 \times 12 \times 128$ 的融合特征图RI2输入第二注意力特征融合模块,输出 $32 \times 12 \times 128$ 的融合特征图RI3;

[0073] $32 \times 12 \times 128$ 的融合特征图RI3经过第一上采样模块,变成 $16 \times 24 \times 256$ 的融合特征图,经过第二上采样模块,变成 $8 \times 48 \times 512$ 的最终融合特征图RI-Final,8为融合后图像的通道数。

[0074] 2) 建立训练集和验证集

[0075] 由于算法的验证要保证公开有效,所以在公开数据集KITTI进行相关的实验。

[0076] 数据集KITTI中的点云可以表示为 $[N, 4]$ 的矩阵,其中,N为一帧场景中的点云数量,每一个点的四维特征 $[x, y, z, i]$,分别是空间x、y、z坐标和激光反射强度i;

[0077] 将点云数据通过球坐标变换转换成一张2D的距离图,具体变换公式为:

$$[0078] \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \frac{1}{2} [1 - \arctan(y, x) \pi^{-1} \times w \\ [1 - (\arcsin(z, r) + f_{down}) f^{-1} \times h \end{pmatrix} \quad (1)$$

[0079] 其中,u、v代表点云转换后在距离图中的像素坐标,x、y和z表示每一个点的三维坐标;在KITTI数据集中激光雷达俯仰角f为 30° ,向下的角度 f_{down} 约为 25° ,向上的角度 f_{up} 约为 5° ;w和h分别表示转换后的距离图的宽度和高度;实验中根据KITTI数据集的统计结果设置为 512×48 ;由此,就可以得到激光雷达数据的Range表示。

[0080] 根据点云数据与RGB图像的对应信息,对RGB图像做一定的修正;

[0081] 由于本实验是在KITTI公开数据集上进行的测试,所以这里以KITTI数据举例说明。由于激光雷达和相机的成像原理不同,采集到的数据所在的空间不一定完全重合,通常点云投影在RGB图像上只在图像的下三分之一有分布,对于上面绝大部分的图像上是没有点云的,所以对RGB图像进行裁剪,只保留图像下面三分之一尺寸。

[0082] 3) 对模型进行训练

[0083] 利用训练集对距离图像融合模型进行训练。

[0084] 步骤4) 将距离图像融合模型输出的最终的融合特征图与带有xyz坐标信息的Range图连接到一起,通过空间坐标转换反向从2D图像还原到空间点云,表示为 $[N, 12]$ 的矩阵,其中后面的8维向量为RGB特征。

[0085] 将增强后的点云数据作为3D目标检测算法的输入,可以验证特征增强的有效性。

[0086] 本发明的实施例2还可提供的一种计算机设备,包括:至少一个处理器、存储器、至少一个网络接口和用户接口。该设备中的各个组件通过总线系统耦合在一起。可理解,总线系统用于实现这些组件之间的连接通信。总线系统除包括数据总线之外,还包括电源总线、控制总线和状态信号总线。其中,用户接口可以包括显示器、键盘或者点击设备(例如,鼠标,轨迹球(track ball)、触感板或者触摸屏等。

[0087] 可以理解,本申请公开实施例中的存储器可以是易失性存储器或非易失性存储器,或可包括易失性和非易失性存储器两者。其中,非易失性存储器可以是只读存储器(Read-Only Memory,ROM)、可编程只读存储器(Programmable ROM,PROM)、可擦除可编程只读存储器(Erasable PROM,EPROM)、电可擦除可编程只读存储器(Electrically EPROM,EEPROM)或闪存。易失性存储器可以是随机存取存储器(Random Access Memory,RAM),其用作外部高速缓存。通过示例性但不是限制性说明,许多形式的RAM可用,例如静态随机存取存储器(Static RAM,SRAM)、动态随机存取存储器(Dynamic RAM,DRAM)、同步动态随机存取存储器(Synchronous DRAM,SDRAM)、双倍数据速率同步动态随机存取存储器(Double Data Rate SDRAM,DDRSRAM)、增强型同步动态随机存取存储器(Enhanced SDRAM,ESDRAM)、同步

连接动态随机存取存储器 (Synchlink DRAM,SLDRAM) 和直接内存总线随机存取存储器 (Direct Rambus RAM,DRRAM)。本文描述的存储器旨在包括但不限于这些和任意其它适合类型的存储器。

[0088] 在一些实施方式中,存储器存储了如下的元素,可执行模块或者数据结构,或者他们的子集,或者他们的扩展集:操作系统和应用程序。

[0089] 其中,操作系统,包含各种系统程序,例如框架层、核心库层、驱动层等,用于实现各种基础业务以及处理基于硬件的任务。应用程序,包含各种应用程序,例如媒体播放器(Media Player)、浏览器(Browser)等,用于实现各种应用业务。实现本公开实施例方法的程序可以包含在应用程序中。

[0090] 在本上述的实施例中,还可通过调用存储器存储的程序或指令,具体的,可以是应用程序中存储的程序或指令,处理器用于:执行实施例1的方法的步骤。

[0091] 实施例1的方法可以应用于处理器中,或者由处理器实现。处理器可能是一种集成电路芯片,具有信号的处理能力。在实现过程中,上述方法的各步骤可以通过处理器中的硬件的集成逻辑电路或者软件形式的指令完成。上述的处理器可以是通用处理器、数字信号处理器(Digital Signal Processor,DSP)、专用集成电路(Application Specific Integrated Circuit,ASIC)、现场可编程门阵列(Field Programmable Gate Array,FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。可以实现或者执行实施例1中的公开的各方法、步骤及逻辑框图。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。结合实施例1所公开的方法的步骤可以直接体现为硬件译码处理器执行完成,或者用译码处理器中的硬件及软件模块组合执行完成。软件模块可以位于随机存储器,闪存、只读存储器,可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器,处理器读取存储器中的信息,结合其硬件完成上述方法的步骤。

[0092] 可以理解的是,本发明描述的这些实施例可以用硬件、软件、固件、中间件、微码或其组合来实现。对于硬件实现,处理单元可以实现在一个或多个专用集成电路(Application Specific Integrated Circuits,ASIC)、数字信号处理器(Digital Signal Processing,DSP)、数字信号处理设备(DSP Device,DSPD)、可编程逻辑设备(Programmable Logic Device,PLD)、现场可编程门阵列(Field-Programmable Gate Array,FPGA)、通用处理器、控制器、微控制器、微处理器、用于执行本申请所述功能的其它电子单元或其组合中。

[0093] 对于软件实现,可通过执行本发明的功能模块(例如过程、函数等)来实现本发明技术。软件代码可存储在存储器中并通过处理器执行。存储器可以在处理器中或在处理器外部实现。

[0094] 实施例3

[0095] 本发明实施例3还可提供一种非易失性存储介质,用于存储计算机程序。当该计算机程序被处理器执行时可以实现上述方法实施例1中的各个步骤。

[0096] 最后所应说明的是,以上实施例仅用以说明本发明的技术方案而非限制。尽管参照实施例对本发明进行了详细说明,本领域的普通技术人员应当理解,对本发明的技术方案进行修改或者等同替换,都不脱离本发明技术方案的精神和范围,其均应涵盖在本发明的权利要求范围当中。

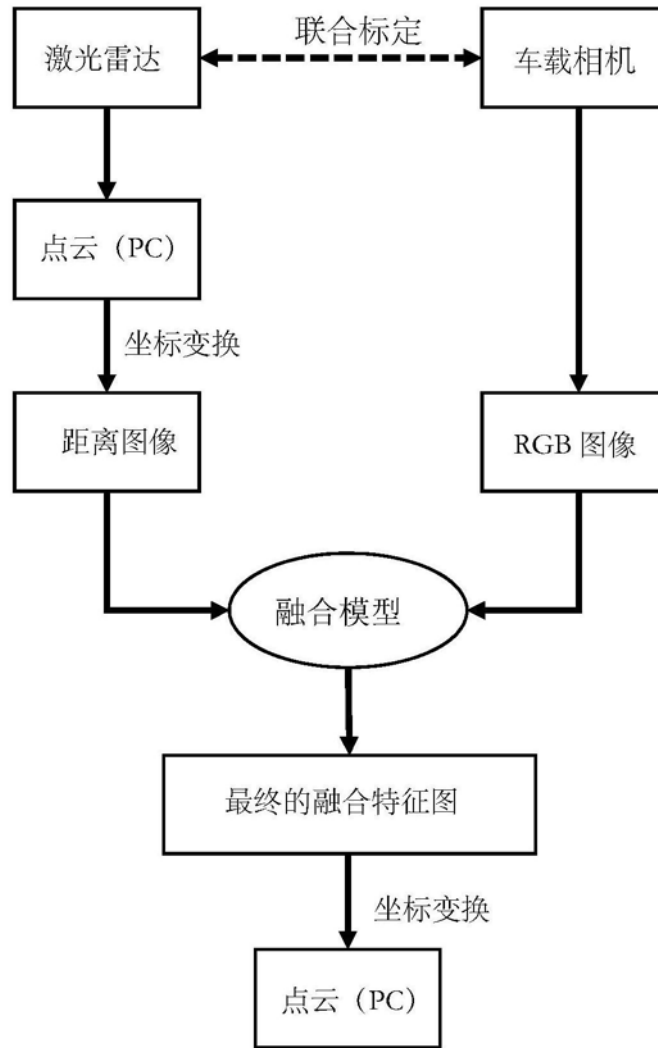


图1

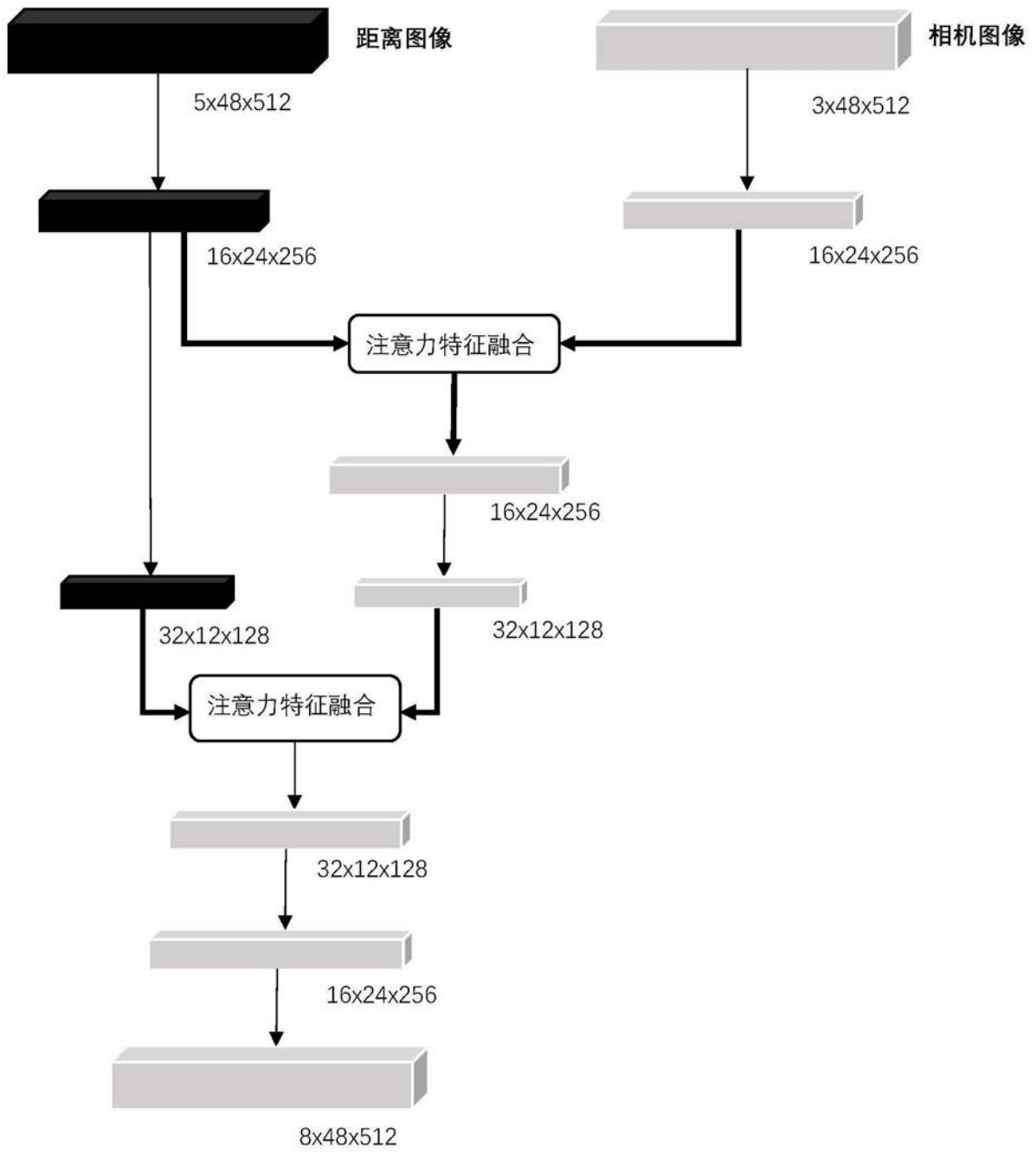


图2

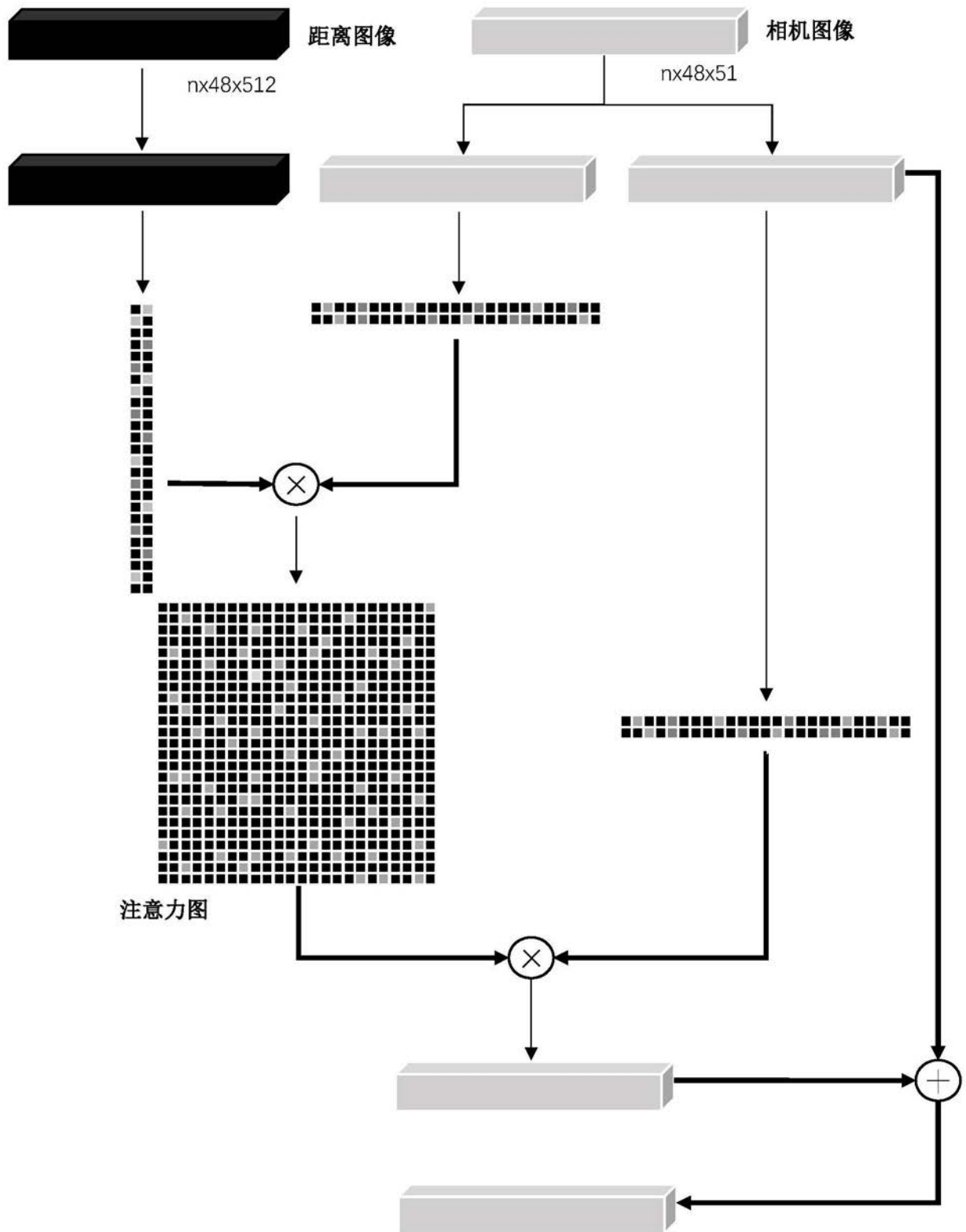


图3