



(12) 发明专利申请

(10) 申请公布号 CN 111723721 A

(43) 申请公布日 2020.09.29

(21) 申请号 202010542533.8

(22) 申请日 2020.06.15

(71) 申请人 中国传媒大学

地址 100024 北京市朝阳区定福庄东街1号

(72) 发明人 严明 李中全 林茜茜 金立标

(74) 专利代理机构 北京市恒有知识产权代理事

务所(普通合伙) 11576

代理人 郭文浩 尹文会

(51) Int. Cl.

G06K 9/00 (2006.01)

G06K 9/62 (2006.01)

G06N 3/04 (2006.01)

G06N 3/08 (2006.01)

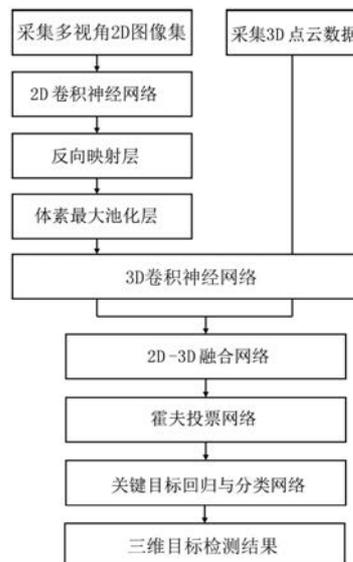
权利要求书3页 说明书11页 附图2页

(54) 发明名称

基于RGB-D的三维目标检测方法、系统及装置

(57) 摘要

本发明属于目标检测技术领域,具体涉及了一种基于RGB-D的三维目标检测方法、系统及装置,旨在解决现有技术无法兼顾效率和3D目标检测准确率的问题。本发明包括:对待检测目标的2D图像集进行特征提取,并反向映射到3D空间;将待检测目标的3D点云数据进行体素划分,结合2D图像的映射体素,通过3D卷积神经网络进行特征编码;进行2D图像纹理特征和3D点云数据几何特征的聚合;通过霍夫投票网络获取目标特征簇集合;通过目标回归与分类网络获取目标包围框,作为三维目标检测结果。本发明将2D图像数据反向映射到3D空间,融合3D点云几何特征和2D图像纹理特征,提高了3D目标检测的准确率,同时通过霍夫投票方法生成预选区域,保证了3D目标检测的效率。



1. 一种基于RGB-D的三维目标检测方法,其特征在于,该三维目标检测方法包括:

步骤S10,获取待检测目标的2D图像集以及3D点云数据;

步骤S20,通过2D卷积神经网络提取所述2D图像集中每一个图像的特征,获得对应的特征图集;

步骤S30,通过反向映射方法将所述特征图集中每一个特征图的特征反向映射到3D空间,获得第一体素集合;

步骤S40,进行所述3D点云数据的体素划分,获得第二体素集合,并通过3D卷积神经网络进行所述第一体素集合、第二体素集合的特征编码,获得2D图像纹理特征和3D点云数据几何特征;

步骤S50,通过2D-3D聚合网络进行所述2D图像纹理特征和3D点云数据几何特征的聚合,获得2D-3D聚合特征;

步骤S60,通过霍夫投票网络对所述2D-3D聚合特征进行投票,并通过KNN法聚类,获得目标特征簇集合;

步骤S70,通过目标回归与分类网络分别获取所述目标特征簇集合中每一个簇对应的目标包围框,作为三维目标检测结果。

2. 根据权利要求1所述的基于RGB-D的三维目标检测方法,其特征在于,所述2D卷积神经网络,其训练中引入局部损失函数 L_{mask} :

$$L_{\text{mask}} = g(x) \log(p(x)) + (1-g(x)) \log(1-g(x))$$

其中, $g(x)$ 代表真实值, $p(x)$ 代表预测值, x 为像素的位置。

3. 根据权利要求1所述的基于RGB-D的三维目标检测方法,其特征在于,步骤S30中“通过反向映射方法将所述特征图集中每一个特征图的特征反向映射到3D空间”,其方法为:

$$R^{-1}M^{-1}s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} x_{2D-3D} \\ y_{2D-3D} \\ z_{2D-3D} \end{bmatrix} + R^{-1}t$$

其中, $M = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$ 代表采集2D图像的相机的内部参数矩阵, f_x 和 f_y 代表相机的

焦距, c_x 和 c_y 分别代表相机中心点在x轴和y轴上的平移量; R 和 t 分别代表采集2D图像的相机

到采集3D点云数据的雷达之间的旋转矩阵和平移向量, $[R, t] = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix}$,

r_{11} 、 r_{12} 、 r_{13} 、 r_{21} 、 r_{22} 、 r_{23} 、 r_{31} 、 r_{32} 、 r_{33} 分别代表旋转矩阵的参数, t_1 、 t_2 、 t_3 分别代表x、y、z轴三

个方向的平移向量; R^{-1} 和 M^{-1} 分别代表对矩阵 R 和 M 求逆; s 为预先设定的常数, $\begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$ 代表图像

上的像素坐标, $s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = M[R, t] \begin{bmatrix} x_{2D-3D} \\ y_{2D-3D} \\ z_{2D-3D} \\ 1 \end{bmatrix}$, $\begin{bmatrix} x_{2D-3D} \\ y_{2D-3D} \\ z_{2D-3D} \end{bmatrix}$ 代表2D像素映射到3D空间后的世

界坐标。

4. 根据权利要求1所述的基于RGB-D的三维目标检测方法,其特征在于,步骤S30之后还设置有第一体素池化的步骤,其方法为:

分别通过平均池化层计算所述第一体素集合中每一个第一体素对应的特征的平均值,并以所述平均值作为所述第一体素的特征。

5. 根据权利要求1所述的基于RGB-D的三维目标检测方法,其特征在于,所述霍夫投票网络,其训练过程中引入损失函数 $L_{vote-reg}$:

$$L_{vote-reg} = \frac{1}{M_{pos}} \sum_i \|\Delta x_i - \Delta x_i^*\| \mathbb{I}[s_i \text{ on object}]$$

其中, $s_i = [x_i; f_i]$ 代表第*i*个2D-3D聚合特征, x_i 代表第*i*个2D-3D聚合特征的欧几里得空间坐标, f_i 代表第*i*个2D-3D聚合特征的特征编码; Δx_i 代表第*i*个2D-3D聚合特征的欧几里得空间坐标的偏移, Δx_i^* 代表 x_i 到关键目标空间中心的真实距离; $\mathbb{I}[s_i \text{ on object}]$ 代表判断 s_i 是否在目标表面; M_{pos} 代表关键目标表面点云的数量。

6. 根据权利要求1所述的基于RGB-D的三维目标检测方法,其特征在于,步骤S70之前还设置有目标特征簇调整的步骤,其方法为:

步骤T10,对于目标特征簇集合中每一个目标特征簇,获取其投票位置,并将所述投票位置转换为局部归一化坐标系统:

$$z'_i = (z_i - z_j) / r$$

其中, z'_i 和 z_i 分别代表归一化后和归一化前的投票位置, z_j 代表投票簇中心点, r 代表聚类半径;

步骤T20,基于所述局部归一化坐标系统,结合目标特征簇中每一个特征,通过预设的目标特征簇调整方法进行目标特征簇调整。

7. 根据权利要求6所述的基于RGB-D的三维目标检测方法,其特征在于,所述预设的目标特征簇调整方法为:

$$p(C) = MLP_2 \left\{ \max_{i=1, \dots, n} \{MLP_1([z'_i; h_i])\} \right\}$$

其中, $p(C)$ 代表目标特征簇集合, h_i 代表归一化后的投票位置 z'_i 的特征, n 为目标特征簇中特征的数量, h_i 为投票位置 z'_i 对应的特征, MLP_1 代表多层感知神经网络1, MLP_2 代表多层感知神经网络2。

8. 一种基于RGB-D的三维目标检测系统,其特征在于,该三维目标检测系统包括输入模块、特征提取模块、反向映射模块、点云划分模块、特征编码模块、特征聚合模块、霍夫投票模块、分类与回归模块、输出模块;

所述输入模块,配置为获取待检测目标的2D图像集并发送至特征提取模块以及获取待检测目标的3D点云数据并发送至点云划分模块;

所述特征提取模块,配置为通过2D卷积神经网络提取所述2D图像集中每一个图像的特征,获得对应的特征图集;

所述反向映射模块,配置为通过反向映射方法将所述特征图集中每一个特征图的特征反向映射到3D空间,获得第一体素集合;

所述点云划分模块,配置为进行所述3D点云数据的体素划分,获得第二体素集合;

所述特征编码模块,配置为通过3D卷积神经网络进行所述第一体素集合、第二体素集合的特征编码,获得2D图像纹理特征和3D点云数据几何特征;

所述特征聚合模块,配置为通过2D-3D聚合网络进行所述2D图像纹理特征和3D点云数据几何特征的聚合,获得2D-3D聚合特征;

所述霍夫投票模块,配置为通过霍夫投票网络对所述2D-3D聚合特征进行投票聚类,获得目标特征簇集合;

所述分类与回归模块,配置为通过目标回归与分类网络分别获取目标特征簇集合中每一个簇对应的目标包围框,作为三维目标检测结果;

所述输出模块,配置为输出获取的三维目标检测结果。

9. 一种存储装置,其中存储有多条程序,其特征在于,所述程序适于由处理器加载并执行以实现权利要求1-7任一项所述的基于RGB-D的三维目标检测方法。

10. 一种处理装置,包括

处理器,适于执行各条程序;以及

存储装置,适于存储多条程序;

其特征在于,所述程序适于由处理器加载并执行以实现:

权利要求1-7任一项所述的基于RGB-D的三维目标检测方法。

基于RGB-D的三维目标检测方法、系统及装置

技术领域

[0001] 本发明属于目标检测技术领域,具体涉及了一种基于RGB-D的三维目标检测方法、系统及装置。

背景技术

[0002] 环境感知问题不仅是计算机视觉中的一个重要问题,更影响着自动驾驶和智能机器人产业。在环境感知问题中,面向二维图像数据和面向三维点云数据的感知技术均取得了空前的发展。基于二维图像的Faster-Rcnn神经网络算法和基于三维点云数据的PointRCNN等技术广泛地用在了工业领域上。

[0003] 在三维目标检测方面有许多常见的方法:第一种是投影法,将3D点云映射到2D图像中,并结合图像的特征提取技术,比较典型的方法是椎体映射法。这种方法在操作庞大的三维点云数据时,因为无法有效地在大型场景的点云数据中定位对象,因此将3D点云数据映射到2D空间上,利用成熟的2D对象检测器,属于多传感器信息融合的方法[1][2]。但是随着工业要求的日益提高,单一数据类型提供的信息量已经无法满足精度的要求。相机传感器是将三维世界的场景映射到二维图像上,在映射过程中会损失对象的深度信息,并且相机传感器会受到光线的影响,强光、逆光和不良天气都会导致相机传感器的精度下降。雷达点云传感器通过扫描外部世界获取几何信息和深度信息,但是无法获取纹理信息,因为点云数据的稀疏性导致较远目标和较小目标难以表示。第二种是体素切割法,这是一种单一传感器的方法,将点云数据按照等间距切割,整个场景被划分成等体积的空间,每个空间中点云的数量相同,使点云数据成为规则性的数据集,以每个空间为单位使用深度学习技术为该单位的点云数据编码,所有的空间共享神经网络中的参数。这种方法最大的缺点是体素划分过程中会对点云做随机采样,造成某些关键目标更加稀疏,早期也有单独为某个体素空间的点云数据做特征编码,但是效果并不好。第三种是RPN法,主要利用单一传感器方法,第一阶段利用3D卷积网络为点云特征编码,在编码的基础上利用RPN方法分割出前景点和背景点,并且提出一些预选框。第二阶段将每个建议集合的点转换为规范坐标,从而更好地学习局部空间特征,并与第一阶段每个点的全局语义特征相结合,进行准确的box细化和置信度预测,这种方法虽然准确率足够高,但是效率非常慢。

[0004] 以下文献是与本发明相关的技术背景资料:

[0005] [1]朱明、王佳荣,一种基于多传感器信息融合的三维目标检测方法及装置,201912,CN201911269486.8.

[0006] [2]程洪、林子彘、许成凤、赵洋,一种基于多模态深度学习的无人车目标检测方法,201903,CN201811388553.3.

发明内容

[0007] 为了解决现有技术中的上述问题,即现有技术无法兼顾效率和3D目标检测准确率的问题,本发明提供了一种基于RGB-D的三维目标检测方法,该三维目标检测方法包括:

[0008] 步骤S10,获取待检测目标的2D图像集以及3D点云数据;

[0009] 步骤S20,通过2D卷积神经网络提取所述2D图像集中每一个图像的特征,获得对应的特征图集;

[0010] 步骤S30,通过反向映射方法将所述特征图集中每一个特征图的特征反向映射到3D空间,获得第一体素集合;

[0011] 步骤S40,进行所述3D点云数据的体素划分,获得第二体素集合,并通过3D卷积神经网络进行所述第一体素集合、第二体素集合的特征编码,获得2D图像纹理特征和3D点云数据几何特征;

[0012] 步骤S50,通过2D-3D聚合网络进行所述2D图像纹理特征和3D点云数据几何特征的聚合,获得2D-3D聚合特征;

[0013] 步骤S60,通过霍夫投票网络对所述2D-3D聚合特征进行投票,并通过KNN法聚类,获得目标特征簇集合;

[0014] 步骤S70,通过目标回归与分类网络分别获取所述目标特征簇集合中每一个簇对应的目标包围框,作为三维目标检测结果。

[0015] 在一些优选的实施例中,所述2D卷积神经网络,其训练中引入局部损失函数 L_{mask} :

[0016] $L_{\text{mask}} = g(x) \log(p(x)) + (1-g(x)) \log(1-g(x))$

[0017] 其中, $g(x)$ 代表真实值, $p(x)$ 代表预测值, x 为像素的位置。

[0018] 在一些优选的实施例中,步骤S30中“通过反向映射方法将所述特征图集中每一个特征图的特征反向映射到3D空间”,其方法为:

$$[0019] \quad R^{-1}M^{-1}s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} x_{2D-3D} \\ y_{2D-3D} \\ z_{2D-3D} \end{bmatrix} + R^{-1}t$$

[0020] 其中, $M = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$ 代表采集2D图像的相机的内部参数矩阵, f_x 和 f_y 代表相

机的焦距, c_x 和 c_y 分别代表相机中心点在x轴和y轴上的平移量; R 和 t 分别代表采集2D图像的相机到采集3D点云数据的雷达之间的旋转矩阵和平移向量,

$$[R, t] = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix}, r_{11}, r_{12}, r_{13}, r_{21}, r_{22}, r_{23}, r_{31}, r_{32}, r_{33} \text{ 分别代表旋转矩阵的}$$

参数, t_1, t_2, t_3 分别代表x、y、z轴三个方向的平移向量; R^{-1} 和 M^{-1} 分别代表对矩阵 R 和 M 求逆; s

为预先设定的常数, $\begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$ 代表图像上的像素坐标, $s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = M[R, t] \begin{bmatrix} x_{2D-3D} \\ y_{2D-3D} \\ z_{2D-3D} \\ 1 \end{bmatrix}, \begin{bmatrix} x_{2D-3D} \\ y_{2D-3D} \\ z_{2D-3D} \end{bmatrix}$ 代

表2D像素映射到3D空间后的世界坐标。

[0021] 在一些优选的实施例中,步骤S30之后还设置有第一体素池化的步骤,其方法为:

[0022] 分别通过平均池化层计算所述第一体素集合中每一个第一体素对应的特征的平均值,并以所述平均值作为所述第一体素的特征。

[0023] 在一些优选的实施例中,所述霍夫投票网络,其训练过程中引入损失函数 $L_{vote-reg}$:

$$[0024] \quad L_{vote-reg} = \frac{1}{M_{pos}} \sum_i \|\Delta x_i - \Delta x_i^*\| \quad \mathbb{I}[s_i \text{ on object}]$$

[0025] 其中, $s_i = [x_i; f_i]$ 代表第*i*个2D-3D聚合特征, x_i 代表第*i*个2D-3D聚合特征的欧几里得空间坐标, f_i 代表第*i*个2D-3D聚合特征的特征编码; Δx_i 代表第*i*个2D-3D聚合特征的欧几里得空间坐标的偏移, Δx_i^* 代表 x_i 到关键目标空间中心的真实距离; $\mathbb{I}[s_i \text{ on object}]$ 代表判断 s_i 是否在目标表面; M_{pos} 代表关键目标表面点云的数量。

[0026] 在一些优选的实施例中,步骤S70之前还设置有目标特征簇调整的步骤,其方法为:

[0027] 步骤T10,对于目标特征簇集合中每一个目标特征簇,获取其投票位置,并将所述投票位置转换为局部归一化坐标系统:

$$[0028] \quad z'_i = (z_i - z_j) / r$$

[0029] 其中, z'_i 和 z_i 分别代表归一化后和归一化前的投票位置, z_j 代表投票簇中心点, r 代表聚类半径;

[0030] 步骤T20,基于所述局部归一化坐标系统,结合目标特征簇中每一个特征,通过预设的目标特征簇调整方法进行目标特征簇调整。

[0031] 在一些优选的实施例中,所述预设的目标特征簇调整方法为:

$$[0032] \quad p(C) = MLP_2 \left\{ \max_{i=1, \dots, n} \{MLP_1([z'_i; h_i])\} \right\}$$

[0033] 其中, $p(C)$ 代表目标特征簇集合, h_i 代表归一化后的投票位置 z'_i 的特征, n 为目标特征簇中特征的数量, h_i 为投票位置 z'_i 对应的特征, MLP_1 代表多层感知神经网络1, MLP_2 代表多层感知神经网络2。

[0034] 本发明的另一方面,提出了一种基于RGB-D的三维目标检测系统,该三维目标检测系统包括输入模块、特征提取模块、反向映射模块、点云划分模块、特征编码模块、特征聚合模块、霍夫投票模块、分类与回归模块、输出模块;

[0035] 所述输入模块,配置为获取待检测目标的2D图像集并发送至特征提取模块以及获取待检测目标的3D点云数据并发送至点云划分模块;

[0036] 所述特征提取模块,配置为通过2D卷积神经网络提取所述2D图像集中每一个图像的特征,获得对应的特征图集;

[0037] 所述反向映射模块,配置为通过反向映射方法将所述特征图集中每一个特征图的特征反向映射到3D空间,获得第一体素集合;

[0038] 所述点云划分模块,配置为进行所述3D点云数据的体素划分,获得第二体素集合;

[0039] 所述特征编码模块,配置为通过3D卷积神经网络进行所述第一体素集合、第二体素集合的特征编码,获得2D图像纹理特征和3D点云数据几何特征;

[0040] 所述特征聚合模块,配置为通过2D-3D聚合网络进行所述2D图像纹理特征和3D点云数据几何特征的聚合,获得2D-3D聚合特征;

[0041] 所述霍夫投票模块,配置为通过霍夫投票网络对所述2D-3D聚合特征进行投票聚类,获得目标特征簇集合;

[0042] 所述分类与回归模块,配置为通过目标回归与分类网络分别获取目标特征簇集合

中每一个簇对应的目标包围框,作为三维目标检测结果;

[0043] 所述输出模块,配置为输出获取的三维目标检测结果。

[0044] 本发明的第三方面,提出了一种存储装置,其中存储有多条程序,所述程序适于由处理器加载并执行以实现上述的基于RGB-D的三维目标检测方法。

[0045] 本发明的第四方面,提出了一种处理装置,包括处理器、存储装置;所述处理器,适于执行各条程序;所述存储装置,适于存储多条程序;所述程序适于由处理器加载并执行以实现上述的基于RGB-D的三维目标检测方法。

[0046] 本发明的有益效果:

[0047] (1) 本发明基于RGB-D的三维目标检测方法,将获取的多视角2D图像信息反向映射到3D空间,并与获取的3D点云数据融合,既避免了相机将三维世界的场景映射到二维图像上时损失深度信息,又避免了雷达无法获取纹理信息,从而因为点云数据的稀疏性导致较远目标和较小目标难以表示。本发明综合考虑目标物体的2D图像纹理特征和3D点云数据几何特征,在各种环境下,尤其是复杂环境下,大大提升了困难3D目标检测的精度和准确率,同时在进行中等难度和简单级别的3D目标检测中,也保持了较高的精度和准确率。

[0048] (2) 本发明基于RGB-D的三维目标检测方法,采用霍夫投票方法生成预选区域,解决RPN方法在3D领域中效率比较低的问题,在保证3D目标检测的准确率的同时提高检测效率。

附图说明

[0049] 通过阅读参照以下附图所作的对非限制性实施例所作的详细描述,本申请的其它特征、目的和优点将会变得更明显:

[0050] 图1是本发明基于RGB-D的三维目标检测方法的三维目标检测流程示意图;

[0051] 图2是本发明基于RGB-D的三维目标检测方法一种实施例的PR曲线平滑处理示意图;

[0052] 图3是本发明基于RGB-D的三维目标检测方法一种实施例的与PointRCNN算法检测结果对比图。

具体实施方式

[0053] 下面结合附图和实施例对本申请作进一步的详细说明。可以理解的是,此处所描述的具体实施例仅用于解释相关发明,而非对该发明的限定。另外还需要说明的是,为了便于描述,附图中仅示出了与有关发明相关的部分。

[0054] 需要说明的是,在不冲突的情况下,本申请中的实施例及实施例中的特征可以相互组合。下面将参考附图并结合实施例来详细说明本申请。

[0055] 本发明提供一种基于RGB-D的三维目标检测方法,本方法使用一种反向映射层将2D图像数据映射到3D空间中,与3D点云数据相结合,这样可以融合3D点云的几何特征和2D图像的纹理特征。本发明还采用霍夫投票方法生成预选区域,解决RPN方法在3D领域中效率比较低的问题,在保证效率的同时提高3D目标检测的准确率。

[0056] 本发明的一种基于RGB-D的三维目标检测方法,该三维目标检测方法包括:

[0057] 步骤S10,获取待检测目标的2D图像集以及3D点云数据;

[0058] 步骤S20,通过2D卷积神经网络提取所述2D图像集中每一个图像的特征,获得对应的特征图集;

[0059] 步骤S30,通过反向映射方法将所述特征图集中每一个特征图的特征反向映射到3D空间,获得第一体素集合;

[0060] 步骤S40,进行所述3D点云数据的体素划分,获得第二体素集合,并通过3D卷积神经网络进行所述第一体素集合、第二体素集合的特征编码,获得2D图像纹理特征和3D点云数据几何特征;

[0061] 步骤S50,通过2D-3D聚合网络进行所述2D图像纹理特征和3D点云数据几何特征的聚合,获得2D-3D聚合特征;

[0062] 步骤S60,通过霍夫投票网络对所述2D-3D聚合特征进行投票聚类,获得目标特征簇集合;

[0063] 步骤S70,通过目标回归与分类网络分别获取所述目标特征簇集合中每一个簇对应的目标包围框,作为三维目标检测结果。

[0064] 为了更清晰地对本发明基于RGB-D的三维目标检测方法进行说明,下面结合图1对本发明实施例中各步骤展开详述。

[0065] 本发明一种实施例的基于RGB-D的三维目标检测方法,包括步骤S10-步骤S70,各步骤详细描述如下:

[0066] 步骤S10,获取待检测目标的2D图像集以及3D点云数据。

[0067] 本发明中利用相应的传感器采集数据,相机负责采取2D图像,激光雷达负责采集3D点云数据。在一些场景中,可能会选用双目相机甚至更多的相机来采集图像,因此获得的2D图像也不止一个。

[0068] 步骤S20,通过2D卷积神经网络提取所述2D图像集中每一个图像的特征,获得对应的特征图集。

[0069] 2D卷积神经网络,其网络结构包括但不限于VGG网络结构、GoogleNet网络结构、MobileNet网络结构。在本发明中,2D卷积神经网络主要是用来区分2D图像的前景点和背景点,为了使网络能够达到预期的功能,本发明在进行2D卷积神经网络的训练过程中,引入了局部损失函数 L_{mask} ,如式(1)所示:

[0070] $L_{\text{mask}} = g(x) \log(p(x)) + (1-g(x)) \log(1-g(x))$ 式(1)

[0071] 其中, $g(x)$ 代表真实值, $p(x)$ 代表预测值, x 为像素的位置。

[0072] 由于2D图像集中可能存在多张图像,分别通过2D卷积神经网络进行每一张2D图像的特征提取。进行每一张2D图像的特征提取的神经网络共享网络参数。

[0073] 步骤S30,通过反向映射方法将所述特征图集中每一个特征图的特征反向映射到3D空间,获得第一体素集合。

[0074] 反向映射的作用是为了使2D图像数据映射到三位维空间和3D点云数据相连接。本发明一个实施例中,相机和激光雷达在同一采集设备上,两个设备的参数已知且两设备之间的旋转矩阵可逆。利用3D数据和2D数据的之间的唯一映射关系将2D数据反向映射到3D空间中,如式(2)所示:

$$[0075] \quad R^{-1}M^{-1}s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} x_{2D-3D} \\ y_{2D-3D} \\ z_{2D-3D} \end{bmatrix} + R^{-1}t \quad \text{式 (2)}$$

[0076] 其中,M代表采集2D图像的相机的内部参数矩阵,R和t分别代表采集2D图像的相机到采集3D点云数据的雷达之间的旋转矩阵和平移向量, R^{-1} 和 M^{-1} 分别代表对矩阵R和M求逆,

s为预先设定的常数, $\begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$ 代表图像上的像素坐标。

[0077] 采集2D图像的相机的内部参数矩阵M,其具体表达方式如式(3)所示:

$$[0078] \quad M = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad \text{式 (3)}$$

[0079] 其中, f_x 和 f_y 代表相机的焦距, c_x 和 c_y 分别代表相机中心点在x轴和y轴上的平移量。

[0080] 采集2D图像的相机到采集3D点云数据的雷达之间的旋转矩阵和平移向量之间的关系如式(4)所示:

$$[0081] \quad [R, t] = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \quad \text{式 (4)}$$

[0082] 其中, r_{11} 、 r_{12} 、 r_{13} 、 r_{21} 、 r_{22} 、 r_{23} 、 r_{31} 、 r_{32} 、 r_{33} 分别代表旋转矩阵的参数, t_1 、 t_2 、 t_3 分别代表x、y、z轴三个方向的平移向量。

[0083] 常数s和 $\begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$ 之间的关系如式(5)所示:

$$[0084] \quad s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = M[R, t] \begin{bmatrix} x_{2D-3D} \\ y_{2D-3D} \\ z_{2D-3D} \\ 1 \end{bmatrix} \quad \text{式 (5)}$$

[0085] 其中, $\begin{bmatrix} x_{2D-3D} \\ y_{2D-3D} \\ z_{2D-3D} \end{bmatrix}$ 代表2D像素映射到3D空间后的世界坐标,1用来保证矩阵相乘对齐。

[0086] 步骤S30之后还设置有第一体素池化的步骤,其方法为:

[0087] 分别通过平均池化层计算所述第一体素集合中每一个第一体素对应的特征的平均值,并以所述平均值作为所述第一体素的特征。

[0088] 由于在实际应用中,数据采集设备中可能会存在多个相机,因此会有多个图像数据,各图像数据之间可以实现特征的相互补充,但同时也带来了体素中特征数量过多,后续计算效率低的问题。为了解决这个问题,本发明提出了一种体素池化层。该层会单独计算每个图像与体素的映射关系。每个体素可能有多个不同角度的RGB特征图映射后的3D数据,通过平均池化操作来计算每个特征通道的最大响应,作为代表该体素的特征。

[0089] 步骤S40,进行所述3D点云数据的体素划分,获得第二体素集合,并通过3D卷积神经网络进行所述第一体素集合、第二体素集合的特征编码,获得2D图像纹理特征和3D点云数据几何特征。

[0090] 3D卷积神经网络通过特征编码层对3D点云划分的体素以及2D数据反向映射的体素做编码,特征编码是为了数据的分层特征。

[0091] 非空体素集如式(6)所示:

$$[0092] \quad V = \{p_i = [x_i, y_i, z_i, r_i]^T \in \mathbb{R}^4\}_{i=1\dots t} \quad \text{式(6)}$$

[0093] 其中,非空体素集包含 $t \leq T$ 个点云数据 p_i ,每个点云数据 p_i 包含点云的X、Y、Z坐标 x_i 、 y_i 、 z_i 和相对应的反射率 r_i 。

[0094] 首先,计算整个非空体素集V中所有点的质心的局部均值 (v_x, v_y, v_z) 。

[0095] 然后,利用其他点云与均值的偏差来扩充输入特征,扩充后的特征如式(7)所示:

$$[0096] \quad V_{in} = \{\hat{p}_i = [x_i, y_i, z_i, r_i, x_i - v_x, y_i - v_y, z_i - v_z]^T \in \mathbb{R}^7\}_{i=1\dots t} \quad \text{式(7)}$$

[0097] 特征扩充以后,将 \hat{p}_i 输入到一组全连接神经网络(FCN, Fully Connected Network),通过FCN去聚合点云的特征为每个体素编码,编码后的结果为 $f_i \in \mathbb{R}^m$ 。

[0098] FCN为本发明一个实施例中选用的进行体素特征编码的3D卷积神经网络,整个FCN包含线性层、批量均值层(BN, Batch normalization)和线性激活层(Rectified Linear Unit)。在其他的实施例中,也可以根据需要进行选择其他的网络,本发明在此不一一详述。

[0099] 得到 f_i 以后利用最大池化层去获取聚合特征信息 \tilde{f} 。

[0100] 最后,将 f_i 和 \tilde{f} 信息串联获得最后的输出 $f_i^{out} = [f_i^T, \tilde{f}^T]^T$ 。

[0101] 所有的非空体素都会以同样的方式通过特征编码层,并且共享所有参数。由于输出特性结合了点方向特性和局部聚合特性,因此叠加特征编码层对体素中的点交互进行编码,并使最终的特性表示能够学习描述性的形状信息。

[0102] 步骤S50,通过2D-3D聚合网络进行所述2D图像纹理特征和3D点云数据几何特征的聚合,获得2D-3D聚合特征。

[0103] 聚合网络主要用于聚合3D数据和2D数据的特征,在3D点云数据和2D反向映射数据经过特征学习网络以后,首先,将两种数据的最后输出 V_{3d}^{out} 和 V_{2d}^{out} 做串联,聚合结果为

$V_{mid}^{out} = [V_{3d}^{out}, V_{2d}^{out}]$;然后,将结果通过卷积中间层去结合3D几何信息和2D纹理信息,获得2D-3D聚合特征。

[0104] 本发明一个实施例中,卷积中间层主要应用3D卷积,包括BN层和ReLU激活层。

[0105] 步骤S60,通过霍夫投票网络对所述2D-3D聚合特征进行投票,并通过KNN法聚类,获得目标特征簇集合。

[0106] 本发明一个实施例中,霍夫投票网络基于深度神经网络构建,投票模块的投票网络由共享权重的全连接网络组成,同时包含BN层和ReLU激活函数层。

[0107] 投票模块的输入是2D-3D聚合特征点 $\{s_i\}_{i=1}^M$,如式(8)所示:

[0108] $s_i = [x_i; f_i]$ 式 (8)

[0109] 其中, $x_i \in \mathbb{R}^3$ 是聚合特征点的欧几里得空间坐标, 而 $f_i \in \mathbb{R}^C$ 是对应的特征编码。

[0110] 投票模块的输出是欧几里得坐标空间的偏移 $\Delta x_i \in \mathbb{R}^3$ 和特征偏移 $\Delta f_i \in \mathbb{R}^C$ 。

[0111] 最后的投票如式 (9) 所示:

[0112] $v_i = [y_i; g_i]$ 式 (9)

[0113] 其中, $y_i = x_i + \Delta x_i, g_i = f_i + \Delta f_i$ 。

[0114] 本发明的霍夫投票网络, 其训练过程中引入损失函数 $L_{vote-reg}$, 如式 (10):

[0115] $L_{vote-reg} = \frac{1}{M_{pos}} \sum_i \|\Delta x_i - \Delta x_i^*\| \mathbb{I}[s_i \text{ on object}]$ 式 (10)

[0116] 其中, $s_i = [x_i; f_i]$ 代表第 i 个 2D-3D 聚合特征, x_i 代表第 i 个 2D-3D 聚合特征的欧几里得空间坐标, f_i 代表第 i 个 2D-3D 聚合特征的特征编码; Δx_i 代表第 i 个 2D-3D 聚合特征的欧几里得空间坐标的偏移, Δx_i^* 代表 x_i 到关键目标空间中心的真实距离; $\mathbb{I}[s_i \text{ on object}]$ 代表判断 s_i 是否在目标表面; M_{pos} 代表关键目标表面点云的数量。必须保证点云在关键目标的表面, 如果不在表面表示该点云离关键目标的中心比较远, 那么该点云参与计算是没有意义的。

[0117] 投票和 2D-3D 聚合特征的张量表示法是一样的, 但不再基于物体表面, 投票可能存在于空的空间中或关键目标的内部。

[0118] 聚类主要是在投票基础上求出投票簇, 即目标特征簇。本发明一个实施例中, 主要根据空间邻近性进行均匀采样和分组, 利用 KNN 方法进行聚类。

[0119] 步骤 S70 之前还设置有目标特征簇调整的步骤, 其方法为:

[0120] 步骤 T10, 对于目标特征簇集合中每一个目标特征簇, 将其记作 $C = \{w_i\}$, 其中 $i = 1, 2, \dots, n$, 其中心为 w_j 。 $w_i = [z_i; h_i]$, $z_i \in \mathbb{R}^3$ 为投票的位置, 而 $h_i \in \mathbb{R}^C$ 为投票的特征, 将所述投票位置转换为局部归一化坐标系统, 如式 (11) 所示:

[0121] $z'_i = (z_i - z_j) / r$ 式 (11)

[0122] 其中, z'_i 和 z_i 分别代表归一化后和归一化前的投票位置, z_j 代表投票簇中心点, r 代表聚类半径。

[0123] 步骤 T20, 基于所述局部归一化坐标系统, 结合目标特征簇中每一个特征, 通过预设的目标特征簇调整方法进行目标特征簇调整。

[0124] 预设的目标特征簇调整方法如式 (12) 所示:

[0125] $p(C) = MLP_2 \left\{ \max_{i=1, \dots, n} \{MLP_1([z'_i; h_i])\} \right\}$ 式 (12)

[0126] 其中, $p(C)$ 代表目标特征簇集合, h_i 代表归一化后的投票位置 z'_i 的特征, n 为目标特征簇中特征的数量, h_i 为投票位置 z'_i 对应的特征, MLP_1 代表多层感知机网络 1, MLP_2 代表多层感知机网络 2。

[0127] 上述公式可以表示为每个目标特征簇中的投票先通过 MLP_1 处理, 然后通过最大池化合并成单个的特征向量, 再然后传递给 MLP_2 将来自不同投票的信息进一步合并。最后的结果 p 表示为一个多维向量, 其中包含客观评分、边界框参数和语义分类评分。

[0128] 步骤S70,通过目标回归与分类网络分别获取所述目标特征簇集合中每一个簇对应的目标包围框,作为三维目标检测结果。

[0129] 本发明对比现有的三维目标检测算法在高识别率的基础上大大的提升了效率,并且提升了在困难环境中算法的准确率。为了评价本发明方法的性能,引入KITTI评价标准。评价标准主要是计算真实3D包围框与预测3D包围框之间的IoU(intersection-over-union,检测评价函数),其计算方法如式(13)所示:

$$[0130] \quad IoU(b_1, b_2) = \frac{A(b_1 \cap b_2)}{A(b_1 \cup b_2)} \quad \text{式(13)}$$

[0131] 其中, $A(\cdot)$ 表示面积, b_i 表示真实值或预测包围框。这种IoU主要是计算真实值和预测值之间的重合程度,重合度越高对应的值越大。

[0132] 同时KITTI中包含了三个等级:简单、中等和困难,具体的划分方法如表1所示:

[0133] 表1

	包围框的高	遮挡程度	截断程度
简单	≥ 40 像素	$\leq 30\%$	$\leq 15\%$
中等	≥ 25 像素 & < 40 像素	$> 30\%$ & $\leq 70\%$	$\leq 50\%$ & $> 15\%$
困难	< 25 像素	$> 70\%$	$> 50\%$

[0134] 为了全面评价算法的性能和鲁棒性还需要预测率和召回率,其计算方法如式(14)、式(15)所示:

$$[0135] \quad Precision = \frac{TP}{TP+FP} \quad \text{式(14)}$$

$$[0136] \quad Recall = \frac{TP}{TP+FN} \quad \text{式(15)}$$

[0137] 其中, Precision为准确率, Recall为召回率。TP(true positive)为真正性数据, FP(false positive)为假正性数据, FN(false negative)假负性数据。在KITTI数据集中车辆的阈值大于等于70%为TP,而行人和自行车要求大于等于50%为TP。

[0138] 单类别中使用IoU、精确率和召回率就可以表示算法的性能,但是在多类别中还需要引入一个指标AP(Average Precision)即平均精度,是单目标检测中衡量检测精度的一个重要标准。计算方法主要是利用准确率和召回率计算PR(Precision, Recall)曲线,并对曲线求积分获得AP值。实际情况下并不会对PR曲线求积分,而是对PR曲线进行平滑处理,如图2所示,为本发明基于RGB-D的三维目标检测方法一种实施例的PR曲线平滑处理示意图,即PR曲线上的每个点精确率的值取该点右侧最大的精确率的值。AP值的计算方法如式(16)所示:

$$[0139] \quad AP = \frac{1}{11} \sum_{0,0.1\dots 1.0} P_{smooth}(i)$$

$$[0140] \quad \text{其中, } P_{smooth}(i) = \max_{i' \geq i} P(i').$$

[0141] 本发明对本发明方法以及现有一些算法在各种不同环境下作了测试,其结果如表2所示:

[0142] 表2

方法	AP (IoU=0.7 Class=Car)		
	简单	中等	困难
MV3D	71.09	62.98	56.56
F-PointNet	81.20	65.46	63.56
SECOND	83.13	76.48	69.10
VoxelNet	77.47	65.46	62.85
PointRCNN	85.94	77.67	76.30
本发明方法	84.88	80.77	77.44

[0143] 对比表2中的数据可以看出,本发明方法在中等目标识别任务中一下与当前表现最好的PointRCNN算法相比AP值相提高了3.1,在困难目标识别任务中的目标检测下本发明方法的AP值高了1.14。困难和中等环境下目标检测的AP性能提高的原因是本发明综合了目标场景中的纹理特征和几何特征,所以在条件相对苛刻的环境下表现更为稳定。

[0144] 如图3所示,为本发明基于RGB-D的三维目标检测方法一种实施例的与PointRCNN算法检测结果对比图,图3中左边的图像是PointRCNN的检测结果,其上半部是将三维检测结果映射到图像上的效果图,下半部是在点云数据集下的三维检测结果(为了观察,将点云转换为鸟瞰图);图3中右边是本发明方法的检测结果,其上半部分是图像检测结果的效果图,下半部分是点云检测结果鸟瞰效果图。

[0145] 根据图3可以看出,PointRCNN和本发明方法均可以可以检测出简单目标,而中等和较难的目标PointRCNN可以检测出两个——最右侧车辆左侧的两个车辆;而发明方法中等和较难目标可以检测出三个——最右侧车辆左侧的三个车辆。可以得出结论,聚合2D纹理信息和3D点云的几何信息可以提高算法在困难目标检测任务中的能力。

[0146] 本发明第二实施例的基于RGB-D的三维目标检测系统,该三维目标检测系统包括输入模块、特征提取模块、反向映射模块、点云划分模块、特征编码模块、特征聚合模块、霍夫投票模块、分类与回归模块、输出模块;

[0147] 所述输入模块,配置为获取待检测目标的2D图像集并发送至特征提取模块以及获取待检测目标的3D点云数据并发送至点云划分模块;

[0148] 所述特征提取模块,配置为通过2D卷积神经网络提取所述2D图像集中每一个图像的特征,获得对应的特征图集;

[0149] 所述反向映射模块,配置为通过反向映射方法将所述特征图集中每一个特征图的特征反向映射到3D空间,获得第一体素集合;

[0150] 所述点云划分模块,配置为进行所述3D点云数据的体素划分,获得第二体素集合;

[0151] 所述特征编码模块,配置为通过3D卷积神经网络进行所述第一体素集合、第二体素集合的特征编码,获得2D图像纹理特征和3D点云数据几何特征;

[0152] 所述特征聚合模块,配置为通过2D-3D聚合网络进行所述2D图像纹理特征和3D点云数据几何特征的聚合,获得2D-3D聚合特征;

[0153] 所述霍夫投票模块,配置为通过霍夫投票网络对所述2D-3D聚合特征进行投票聚

类,获得目标特征簇集合;

[0154] 所述分类与回归模块,配置为通过目标回归与分类网络分别获取目标特征簇集合中每一个簇对应的目标包围框,作为三维目标检测结果;

[0155] 所述输出模块,配置为输出获取的三维目标检测结果。

[0156] 所属技术领域的技术人员可以清楚地了解到,为描述的方便和简洁,上述描述的系统的具工作过程及有关说明,可以参考前述方法实施例中的对应过程,在此不再赘述。

[0157] 需要说明的是,上述实施例提供的基于RGB-D的三维目标检测系统,仅以上述各功能模块的划分进行举例说明,在实际应用中,可以根据需要而将上述功能分配由不同的功能模块来完成,即将本发明实施例中的模块或者步骤再分解或者组合,例如,上述实施例的模块可以合并为一个模块,也可以进一步拆分成多个子模块,以完成以上描述的全部或部分功能。对于本发明实施例中涉及的模块、步骤的名称,仅仅是为了区分各个模块或者步骤,不视为对本发明的不当限定。

[0158] 本发明第三实施例的一种存储装置,其中存储有多条程序,所述程序适于由处理器加载并执行以实现上述的基于RGB-D的三维目标检测方法。

[0159] 本发明第四实施例的一种处理装置,包括处理器、存储装置;处理器,适于执行各条程序;存储装置,适于存储多条程序;所述程序适于由处理器加载并执行以实现上述的基于RGB-D的三维目标检测方法。

[0160] 所属技术领域的技术人员可以清楚地了解到,为描述的方便和简洁,上述描述的存储装置、处理装置的具体工作过程及有关说明,可以参考前述方法实施例中的对应过程,在此不再赘述。

[0161] 本领域技术人员应该能够意识到,结合本文中所公开的实施例描述的各示例的模块、方法步骤,能够以电子硬件、计算机软件或者二者的结合来实现,软件模块、方法步骤对应的程序可以置于随机存储器(RAM)、内存、只读存储器(ROM)、电可编程ROM、电可擦除可编程ROM、寄存器、硬盘、可移动磁盘、CD-ROM、或技术领域内所公知的任意其它形式的存储介质中。为了清楚地说明电子硬件和软件的可互换性,在上述说明中已经按照功能一般性地描述了各示例的组成及步骤。这些功能究竟以电子硬件还是软件方式来执行,取决于技术方案的特定应用和设计约束条件。本领域技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能,但是这种实现不应认为超出本发明的范围。

[0162] 术语“第一”、“第二”等是用于区别类似的对象,而不是用于描述或表示特定的顺序或先后次序。

[0163] 术语“包括”或者任何其它类似用语旨在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者设备/装置不仅包括那些要素,而且还包括没有明确列出的其它要素,或者还包括这些过程、方法、物品或者设备/装置所固有的要素。

[0164] 至此,已经结合附图所示的优选实施方式描述了本发明的技术方案,但是,本领域技术人员容易理解的是,本发明的保护范围显然不局限于这些具体实施方式。在不偏离本发明的原理的前提下,本领域技术人员可以对相关技术特征作出等同的更改或替换,这些更改或替换之后的技术方案都将落入本发明的保护范围之内。

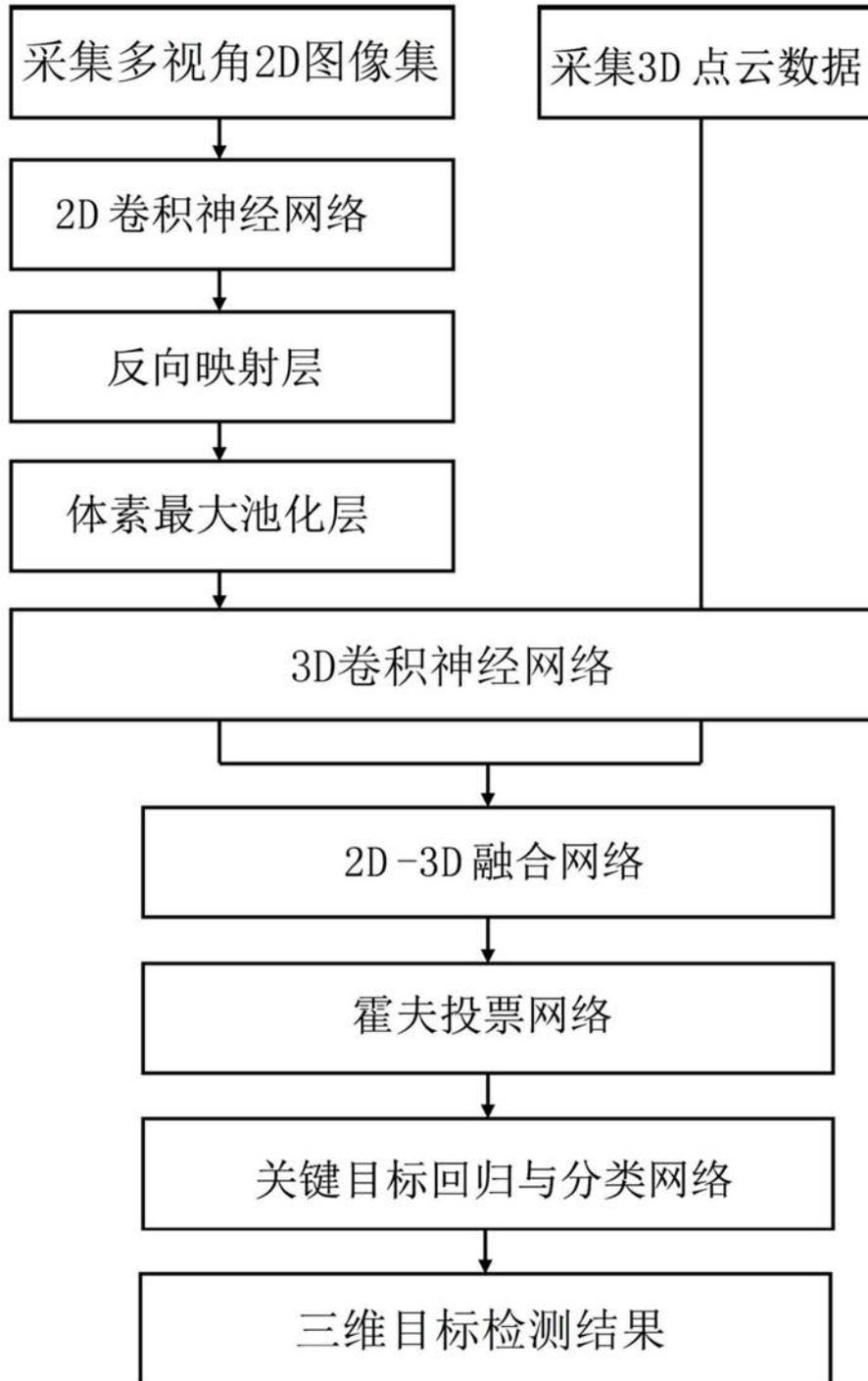


图1

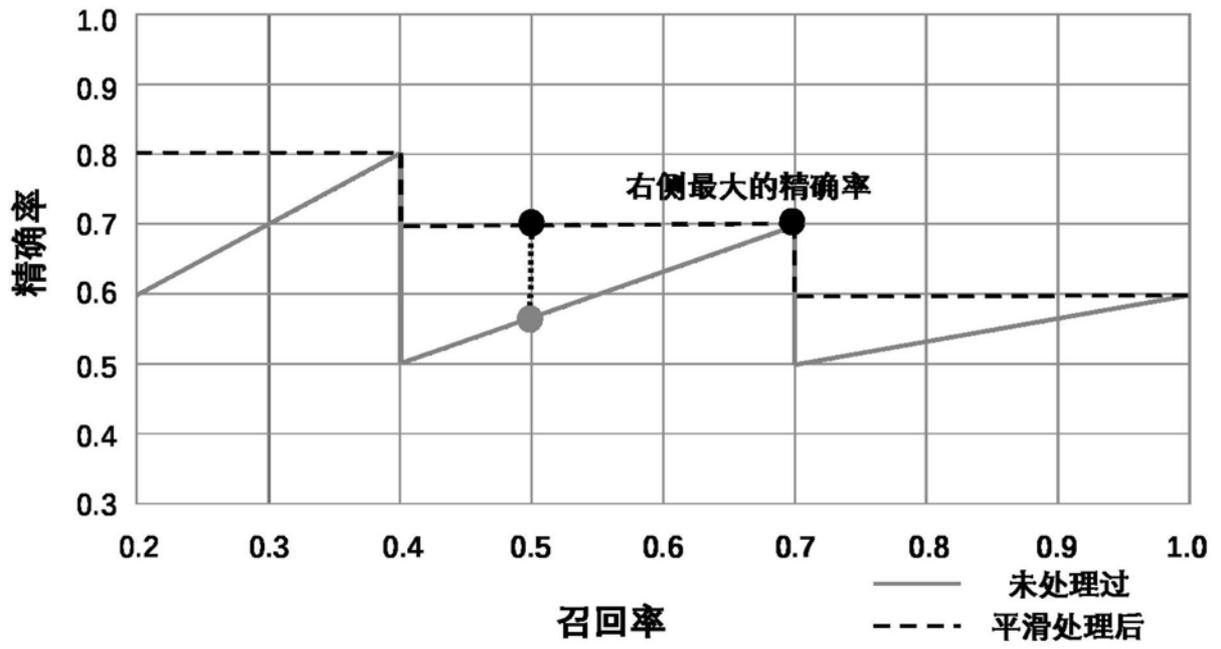


图2

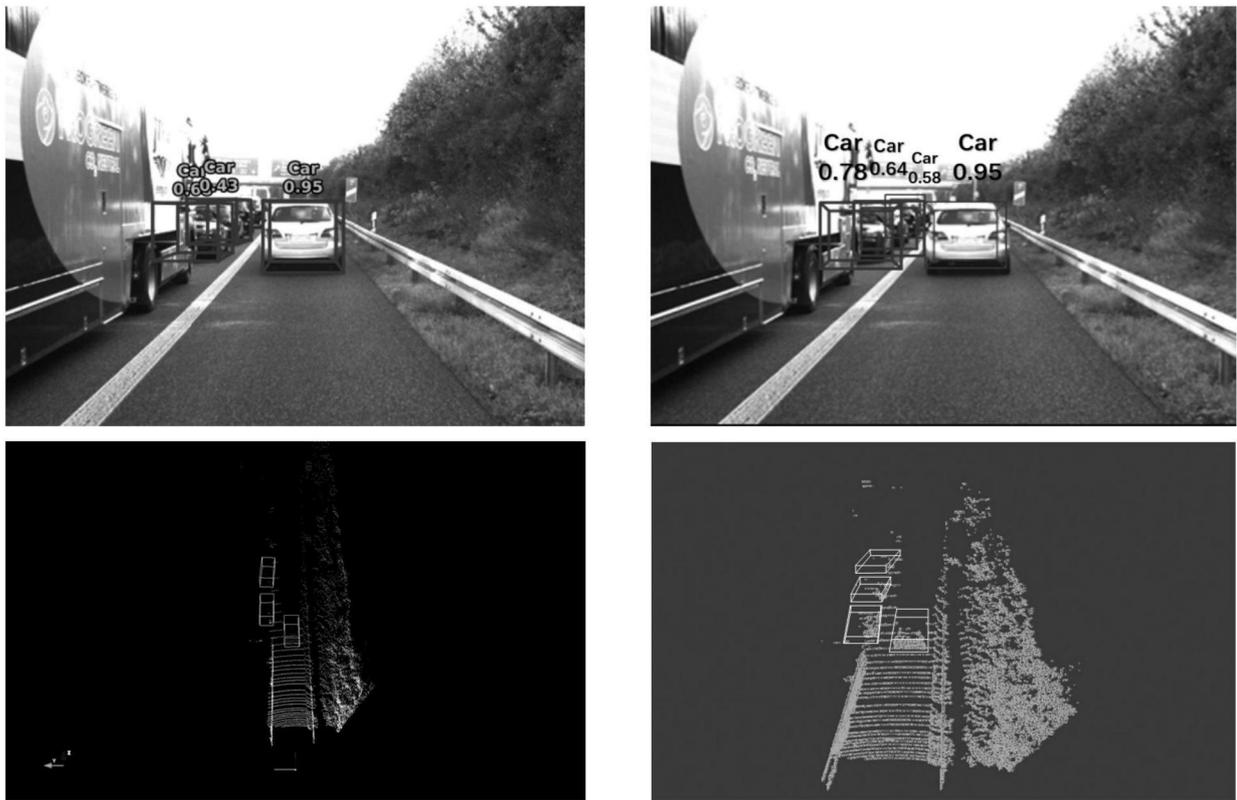


图3