(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 118644848 A
(43)申请公布日 2024.09.13

(21)申请号 202410668633.3

(22)申请日 2024.05.28

(71)申请人 天津理工大学
地址 300384 天津市西青区宾水西道391号

(72)发明人 周冕 吴少清 田雪媛

(74)专利代理机构 天津耀达律师事务所 12223
专利代理师 邵洪军

(51)Int.Cl.
G06V 20/64(2022.01)
G06V 10/26(2022.01)
G06V 10/40(2022.01)
G06V 10/80(2022.01)
G06V 10/764(2022.01)
G06V 10/56(2022.01)

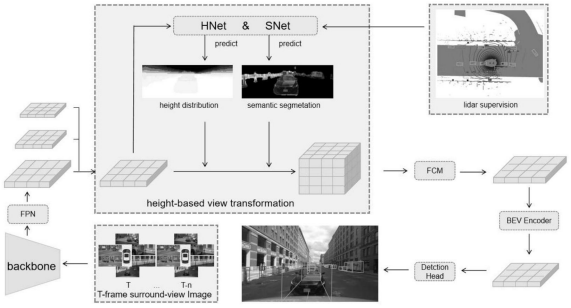
权利要求书2页 说明书5页 附图3页

(54)发明名称

一种面向自动驾驶的基于BEV表征的纯视觉三维目标检测方法

(57)摘要

本发明涉及一种面向自动驾驶的基于BEV表征的纯视觉三维目标检测方法,属于计算机视觉技术领域,方法包括:从环视图像中获取图像特征;预测图像特征的高度分布;对图像特征进行语义分割;根据高度和语义信息将图像特征投影到预定义好的三维空间得到三维采样点的特征;将三维采样点的特征压缩到bev层面进行特征融合;定义三维目标检测的任务头,构建整体的三维目标检测模型;使用检测模型对新的输入数据进行三维目标检测,识别场景中物体的类别,确定物体的精确位置、方向、尺寸和速度;本发明实现二维图像特征和三维采样点位置更准确的对应关系,可以获得更准确的三维特征;过滤掉大量无用的的背景特征,可以提高检测的准确度和效率。



CN 118644848 A

1.一种面向自动驾驶的基于BEV表征的纯视觉三维目标检测方法,其特征在于,包括:

从环视图像中获取图像特征;

预测图像特征的高度分布;

对图像特征进行语义分割;

根据高度和语义信息将图像特征投影到预定义好的三维空间得到三维采样点的特征;

将三维采样点的特征压缩到bev层面进行特征融合;

定义三维目标检测的任务头,构建整体的三维目标检测模型;

使用检测模型对新的输入数据进行三维目标检测,识别场景中物体的类别,确定物体的精确位置、方向、尺寸和速度。

2.根据权利要求1所述的面向自动驾驶的基于BEV表征的纯视觉三维目标检测方法,其特征在于,还包括:

采用图像增强策略对图像进行增强用于训练三维目标检测模型;

采用BEV空间增强策略对BEV特征进行增强用于训练三维目标检测模型;

采用时间融合策略引入多帧图像用于训练三维目标检测模型。

3.根据权利要求1所述的面向自动驾驶的基于BEV表征的纯视觉三维目标检测方法,其特征在于:

从环视图像中获取图像特征的具体步骤为:

将六张环视图像输入到backbone模块,再经过FPN模块,再融合部分特征得到图像特征。

4.根据权利要求1所述的面向自动驾驶的基于BEV表征的纯视觉三维目标检测方法,其特征在于:

预测图像特征的高度分布的具体步骤为:

利用激光雷达信息作为监督,通过HNet模块,预测每个图像特征单元的高度的分布。

5.根据权利要求1所述的面向自动驾驶的基于BEV表征的纯视觉三维目标检测方法,其特征在于:

对图像特征进行语义分割的具体步骤为:

利用激光雷达信息作为监督,通过SNet模块对得到的图像特征进行语义分割,判断每个图像特征单元属于前景物体还是背景信息。

6.根据权利要求1所述的面向自动驾驶的基于BEV表征的纯视觉三维目标检测方法,其特征在于:

根据高度和语义信息将图像特征投影到预定义好的三维空间得到三维采样点特征的具体步骤为:

先根据语义过滤掉背景的图像特征,再根据高度分配每个图像特征单元的权重,再将处理过的图像特征转换到预定义好的三维空间的位置,得到每个三维采样点的特征。

7.根据权利要求1所述的面向自动驾驶的基于BEV表征的纯视觉三维目标检测方法,其特征在于:

将三维采样点的特征压缩到bev层面进行特征融合的具体步骤为:

将三维采样点的特征先压缩到bev表征,再经过一个bev encoder模块进行特征融合,得到bev特征。

8.根据权利要求1所述的面向自动驾驶的基于BEV表征的纯视觉三维目标检测方法,其特征在于:

定义三维目标检测的任务头,构建整体的三维目标检测模型的具体步骤为:

三维目标检测的任务头包括一个分类网络,用于预测每个检测框内物体的类别,和一个回归网络用于预测每个物体的三维边界框,包括其在三维空间中的位置、尺寸、方向和速度。

9.根据权利要求2所述的面向自动驾驶的基于BEV表征的纯视觉三维目标检测方法,其特征在于:

采用图像增强策略对图像进行增强用于训练感知网络模型具体包括:

对输入图像进行随机裁剪,随机翻转,随机旋转;

采用BEV空间增强策略对BEV特征进行增强用于训练感知网络模型具体包括:

对输入BEV特征进行随机翻转。

10.根据权利要求2所述的面向自动驾驶的基于BEV表征的纯视觉三维目标检测方法,其特征在于:

采用时间融合策略引入多帧图像用于训练感知网络模型的具体步骤为:

先将多帧图片对齐到当前帧,然后将多帧图片和当前帧图片同时输入网络进行特征提取,再同时投影到预定义好的三维空间得到三维采样点特征。

一种面向自动驾驶的基于BEV表征的纯视觉三维目标检测方法

技术领域

[0001] 本发明属于计算机视觉技术领域,涉及一种面向自动驾驶的基于BEV的纯视觉三维目标检测方法。

背景技术

[0002] 三维目标检测在自动驾驶领域是一个关键的任务,是自动驾驶系统的核心组成部分。它解决了如何在复杂的环境中准确识别和定位物体的问题,对于实现安全和有效的自动驾驶至关重要。三维目标检测的输入通常是来自车载传感器的数据,如激光雷达(LiDAR)扫描或摄像头图像。输出是检测到的场景中的物体的类别、位置、方向、尺寸和速度。鸟瞰视图(BEV)是周围场景的统一表示,它从上方观察场景,适合于自动驾驶任务,所以很多三维目标检测会采用BEV的表征方式。

[0003] 但是,目前学术界提出的基于BEV的纯视觉三维目标检测方法存在以下弊端:1.只使用二维的图像作为输入信息,去预测三维世界中的物体,缺少了一个维度的信息,会导致同一个二维图像特征点均等的对应了相机射线经过的所有的三维位置,这明显是不合理的。2.三维目标检测本身只关注场景中的前景物体,而不关注背景信息,但目前的三维目标方法把所有二维特征点都投影到三维空间去,导致其中混杂有大量的无用的背景信息。这不仅会影响前景物体的检测的精度,同时也造成了计算资源的浪费。

发明内容

[0004] 为了解决目前基于BEV的三维目标检测存在的二维到三维特征投影位置不准确,投影特征混杂大量背景信息的问题,本发明公开了一种面向自动驾驶的基于BEV的纯视觉三维目标检测方法。

[0005] 实现发明目的的技术方案如下:一种面向自动驾驶的基于BEV的纯视觉三维目标检测方法,包括:

[0006] 从环视图像中获取图像特征;

[0007] 预测图像特征的高度分布;

[0008] 对图像特征进行语义分割;

[0009] 根据高度和语义信息将图像特征投影到预定义好的三维空间得到三维采样点的特征;

[0010] 将三维采样点的特征压缩到bev层面进行特征融合;

[0011] 定义三维目标检测的任务头,构建整体的三维目标检测模型;

[0012] 使用检测模型对新的输入数据进行三维目标检测,识别三维场景中物体的类别,确定物体的精确位置、方向、尺寸和速度;

[0013] 采用图像增强策略对图像进行增强用于训练三维目标检测模型;

[0014] 采用BEV空间增强策略对BEV特征进行增强用于训练三维目标检测模型;

[0015] 采用时间融合策略引入多帧图像用于训练三维目标检测模型；

[0016] 优选地，从环视图像中获取图像特征的具体步骤为：

[0017] 将六张环视图像输入到backbone模块，再经过FPN模块，再融合部分特征得到图像特征；

[0018] 优选地，预测图像特征的高度分布的具体步骤为：

[0019] 利用激光雷达信息作为监督，通过HNet模块，预测每个图像特征单元的高度的分布；

[0020] 优选地，对图像特征进行语义分割的具体步骤为：

[0021] 利用激光雷达信息作为监督，通过SNet模块对得到的图像特征进行语义分割，判断每个图像特征单元属于前景物体还是背景信息；

[0022] 优选地，根据高度和语义信息将图像特征投影到预定义好的三维空间得到三维采样点特征的具体步骤为：

[0023] 先根据语义过滤掉背景的图像特征，再根据高度分配每个图像特征单元的权重，再将处理过的图像特征转换到预定义好的三维空间的位置，得到每个三维采样点的特征；

[0024] 优选地，将三维采样点的特征压缩到bev层面进行特征融合的具体步骤为：

[0025] 将三维采样点的特征先压缩到bev表征，再经过一个bev encoder模块进行特征融合，得到bev特征；

[0026] 优选地，定义三维目标检测的任务头，构建整体的三维目标检测模型的具体步骤为：

[0027] 三维目标检测的任务头包括一个分类网络，用于预测每个检测框内物体的类别，和一个回归网络用于预测每个物体的三维边界框，包括其在三维空间中的位置、尺寸、方向和速度。

[0028] 优选地，采用图像增强策略对图像进行增强用于训练三维目标检测模型具体包括：

[0029] 对输入图像进行随机裁剪，随机翻转，随机旋转；

[0030] 优选地，采用BEV空间增强策略对BEV特征进行增强用于训练三维目标检测模型具体包括：

[0031] 对输入BEV特征进行随机翻转。

[0032] 优选地，采用时间融合策略引入多帧图像用于训练三维目标检测模型的具体步骤为：

[0033] 先将多帧图片对齐到当前帧，然后将多帧图片和当前帧图片同时输入网络进行特征提取，再同时投影到预定义好的三维空间得到三维采样点特征；

[0034] 技术效果

[0035] 与现有技术相比，本发明至少具有如下有益效果：

[0036] 1.实现二维图像特征和三维采样点位置更准确的对应关系，可以获得更准确的三维特征。

[0037] 2.过滤掉大量无用的的背景特征，可以提高检测的准确度和效率。

附图说明

[0038] 为了更清楚地说明本申请实施例的技术方案,下面将对实施例中所需要使用的附图作简单地介绍:

[0039] 图1为本发明的三维目标检测方法的流程示意图;

[0040] 图2为本发明的三维目标检测方法中的图像特征转换的示意图;

[0041] 图3为本发明的高度预测和语义掩码的可视化图片;

[0042] 图4为本发明的可视化结果。

具体实施方式

[0043] 下面通过实施例并结合说明书附图对本发明做进一步说明。需要说明是,本发明还可以采用其他不同于在此描述的其他方式来实施,因此,本发明的保护范围并不受下面公开的具体实施例的限制。

[0044] 本发明的一个具体实施例,如图1,公开了一种面向自动驾驶的基于BEV的纯视觉三维目标检测方法,具体方法如下:

[0045] 步骤1:从环视图像中获取图像特征;

[0046] 步骤11:将六张环视图像(分别为前视相机图像,左前相机图像,左后相机图像,右前相机图像,右后相机图像,后视相机图像)先经过图像增强,这里的图像增强包括随机裁剪,随机翻转,随机旋转,再经过BEV空间增强,包括随机翻转,再输入到backbone模块(使用resnet50模型)进行图像特征的提取,获得原始的4层图像特征;

[0047] 步骤12:原始的四层图像特征经过FPN模块,获得融合后的多尺度图像特征S1,S2,S3;

[0048] 步骤13:将三层特征逐层上采样并拼接得到融合的单层图像特征F_2d,公式为F_2d=S1↑4+(S1↑2+S2)↑2+S3,其中“↑2”表示上采样2倍,“↑4”表示上采样4倍,“+”表示拼接操作;

[0049] 步骤2:将图像特征F_2d输入到HeightNet模块,预测每个图像特征单元的高度的分布H;

[0050] 步骤21:把ego坐标系下的激光雷达点(Ego_3d)先使用相机的外参矩阵[R|t]转换到相机坐标系下(Cam_3d),其中转换公式为Cam_3d=[R|t]Ego_3d,获得该激光雷达点在相机坐标系下相对于相机的高度真值H_3d,计算公式为H_3d=Cam_3d[1],相机坐标系y轴垂直向下,值代表距离相机的距离,即高度H_3d;

[0051] 步骤22:再使用相机的内参矩阵K把激光雷达点转换到图像坐标系确认该激光雷达点在图像坐标系下的位置(Pix_2d),转换公式为Pix_2d=K Cam_3d,至此获得了每个图像特征单元在相机坐标系下相对于相机高度的真值H_3d,作为预测每个图像特征单元在相机坐标系下的高度分布的监督,用于训练Height Net;

[0052] 步骤23:用训练好的HeightNet预测每个图像特征单元在相机坐标系下的高度分布H;

[0053] 步骤231:HeightNet是一个由多层感知机构成的回归网络,输入是图片特征单元的特征,输出是该图片特征单元的高度分布,也就是该图片特征单元处于各个高度的可能性;

[0054] 步骤3:将图像特征输入到SemanticNet模块,判断每个图像特征单元属于前景物体还是背景信息;

[0055] 步骤31:把激光雷达点(Ego_3d)转换到到图像坐标系(Pix_2d),转换公式为Pix_2d＝K[R|t]Ego_3d,用激光雷达自带的语义信息作为语义分割的监督,用于训练SemanticNet;

[0056] 步骤32:用训练好的Semantic Net来学习每个图像特征单元的语义信息,区分这个单元是前景物体和背景信息。

[0057] 步骤321:SemanticNet是一个由多层感知机构成的二分类网络,输入是图片特征单元的特征,输出是该图片特征单元属于前景还是背景的分类结果;

[0058] 步骤4:如图2,根据高度和语义信息将图像特征投影到预定义好的三维空间得到三维采样点的特征;

[0059] 步骤41:在三维空间中预定义256x 256大小的bev网格,每个bev格子垂直向上采10个采样点;

[0060] 步骤42:每个采样点根据相机的内外参数投影到图像的对应位置;

[0061] 步骤421:每个采样点(ref_3d)先转换到相机坐标系(cam_3d),转换公式为cam_3d＝[R|t]ref_3d,并在相机坐标系下计算采样点相对于相机的高度h_3d,计算公式为h_3d＝cam_3d[1];

[0062] 步骤422:采样点再从相机坐标系转换到图像坐标系(pix_2d),转换公式为pix_2d＝K cam_3d,也就投影到了图像的对应位置上。

[0063] 步骤43:如果投影到的图像的对应位置的图像特征单元经过Semantic Net预测是背景信息,则此特征单元的语义系数fc＝0;

[0064] 步骤44:如果投影到的图像的对应位置的图像特征单元经过Semantic Net预测是前景物体,则此特征单元的语义系数fc＝1,此特征单元的高度系数fh＝H[h_3d-1];

[0065] 步骤45:将此图像特征单元的特征F_2d与语义系数fc和高度系数fh相乘,获得此三维采样点的特征F_3d,公式为F_3d＝fc*fh*F_2d。

[0066] 图2中,分别以深色和浅色采样点为例:两个深色采样点投影到图片,对应的图片特征点是背景信息,则语义系数为0,也就是说不在此处提取语义信息;两个浅色采样点投影到图片,对应的图片特征点是前景物体,则语义系数为1,再根据投影到图片的点的图片特征的高度分布,两个浅色采样点的高度系数分别为α_i和α_j,分别将图像特征与语义系数和高度系数相乘,获得两个浅色采样点的特征。

[0067] 图3示出了高度预测和语义掩码的可视化图片,其中其中左起第一张图是高度预测的可视化图,颜色越深代表高度越高,第二张图是某个前景物体的像素点的高度分布,第三张图是原图,第四张图是语义掩码的可视化图,只保留需要被检测的前景物体,比如车子,人。如图可以看到本发明的模型对于高度的预测和语义的分割都非常的准确。

[0068] 步骤5:将三维采样点的特征压缩到bev层面进行特征提取和融合;

[0069] 步骤51:把三维采样点的特征通过Z*C的操作压缩到BEV层面,公式为HW(ZC)＝HWZC;

[0070] 步骤52:把融合后的bev特征经过一个bev编码器进行特征的提取。

[0071] 步骤6:定义三维目标检测的任务头,构建整体的三维目标检测模型;

[0072] 步骤61:三维目标检测的任务头使用一个分类网络预测每个检测框内物体的类别;

[0073] 步骤62:使用一个回归网络预测每个物体的三维边界框,包括其在三维空间中的位置、尺寸、方向和速度。

[0074] 步骤7:使用检测模型对新的输入数据进行三维目标检测,识别三维场景中物体的类别,确定物体的精确位置、方向、尺寸和速度。

[0075] 图4示出了本发明的可视化结果。从可视化图片可以看到本发明的模型可以准确的检测出三维场景中的前景物体,比如图中三维边界框里的轿车,人,自行车和卡车。

[0076] 以上所述,仅为本发明较佳的具体实施方式,但本发明的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本发明揭露的技术范围内,可轻易想到的变化或替换,都应涵盖在本发明的保护范围之内。

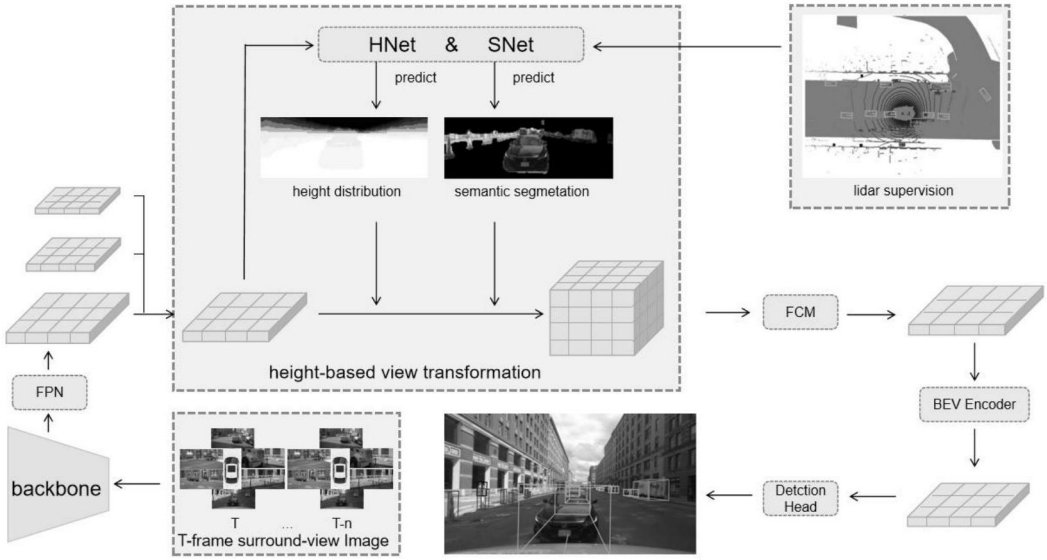

HNet & SNet
predict
predict
height distribution
semantic segmetation
lidar supervision
FPN
backbone
T
...
T-n
T-frame surround-view Image
height-based view transformation
FCM
BEV Encoder
Detction Head

图1

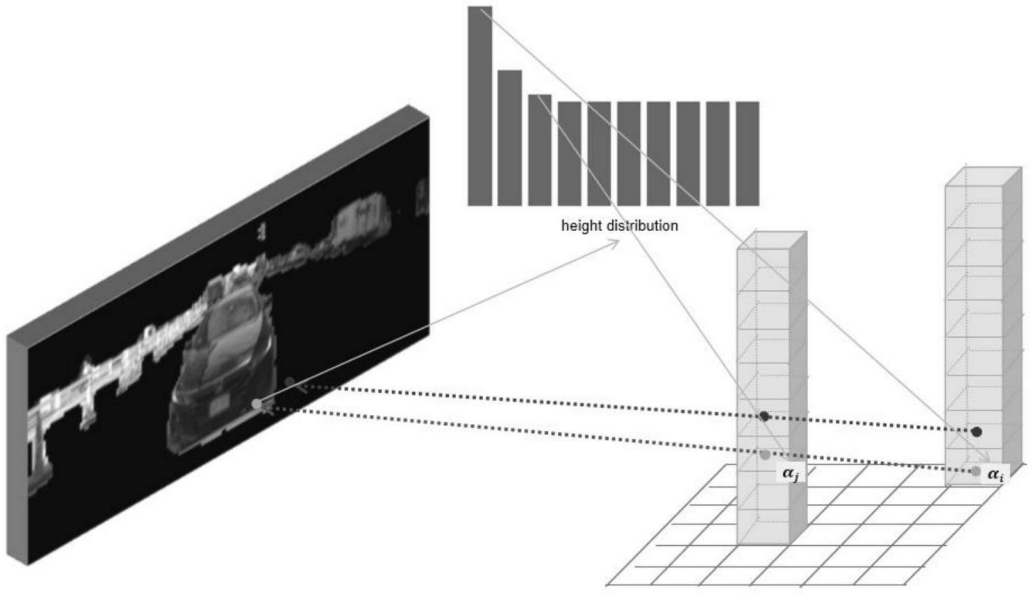

height distribution
α_j
α_i

图2

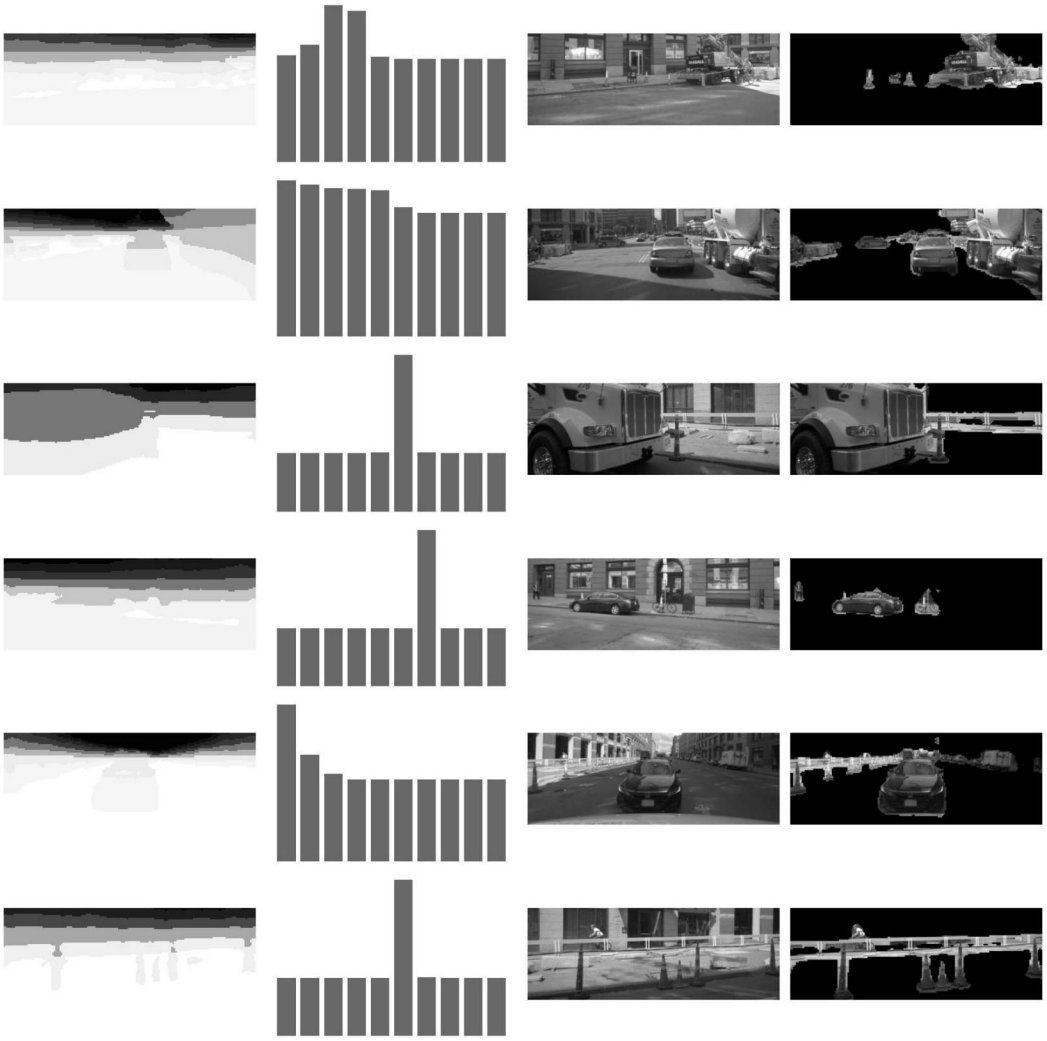


图3


CAM_FRONT
CAM_FRONT_RIGHT
CAM_BACK_RIGHT
CAM_BACK
CAM_BACK_LEFT
CAM_FRONT_LEFT

图4