文章编号：2095-2163(2022)10-0023-08 中图分类号：TP242 文献标志码：A

毫米波雷达与相机两级融合的 3D 目标检测方法研究

侯志斌[1]，陆 峰[2]，娄静涛[2]，朱 愿[2]
（1 陆军军事交通学院 学员五大队，天津 300161；2 陆军军事交通学院 军事交通运输研究所，天津 300161）

摘 要： 多传感器融合是当前自动驾驶领域感知系统的重点。针对在恶劣复杂天气场景下目标以及远距离小目标检测效果差的问题，提出一种基于毫米波雷达与相机两级融合的 3D 目标检测方法。该方法先在数据层将毫米波雷达条码化处理与相机信息融合建立三通道图像，然后输入到加入注意力机制的特征提取网络中进行初级检测；在特征层采取截锥体数据关联的方式，将毫米波雷达信息与初级检测结果进行关联，进一步提升检测精度。实验结果显示，在大型自动驾驶数据集 nuScenes 下的 mini 集对融合网络进行评估，相比基准网络 Centerfusion，平均精度（*mAP*）提升了1.09%，检测指标（*NDS*）提升了1.21%。

关键词： 传感器融合；3D 目标检测；神经网络；雷达；相机

Research on 3D object detection method based on two-level fusion of millimeter-wave radars and cameras

HOU Zhibin [1], LU Feng[2], LOU Jingtao[2], ZHU Yuan [2]
(1 Fifth Brigade of Cadets, Army Military Transportation University, Tianjin 300161, China;
2 Military Transportation Research Institute, Army Military Transportation University, Tianjin 300161, China)

【Abstract】 Multi-sensor fusion is the focus of current perception systems in the field of autonomous driving. Aiming at the problem of poor detection effect of targets in severe and complex weather scenarios and long-distance small targets, a 3D object detection method based on two-level fusion of millimeter-wave radars and cameras is proposed. First, in the data layer, millimeter wave radar barcode processing and camera information are fused to create a three-channel image, which is input into the feature extraction network with attention mechanism for primary detection. Then, at the feature layer, the millimeter-wave radar information is associated with the primary detection results by frustum-based method to further improve the detection accuracy. The experimental results show that evaluated with mini-set under the large-scale autonomous driving dataset nuScenes, compared with the benchmark network Centerfusion, the average precision (*mAP*) is improved by 1.09%, and the detection index (*NDS*) is improved by 1.21%.

【Key words】 sensor fusion; 3D object detection; neural networks; radars; cameras

0 引 言

感知是自动驾驶系统的重要组成模块，而 3D 目标检测是自动驾驶感知模块的重要内容。尤其是对自动驾驶下游任务，发挥着重要作用。由于采用单一传感器均存在一些缺陷，因此多模态融合是当前研究重点。目前来看，现有传感器融合方法大多集中在激光雷达与摄像机融合上，但在雪、雨、雾霾、沙尘暴等恶劣天气条件以及远距离目标下，激光雷达与相机融合方案的检测质量会大幅下降。在当前技术水平下，开展相机与毫米波雷达融合策略方法研究是一套低成本且应对恶劣环境下目标检测的更鲁棒方案。

国内外对毫米波雷达与相机融合的目标检测方法已经做了一定研究。如：Nabati 等人[1]提出了 RRPN 网络，通过仿照图像检测中的 RPN 网络，将毫米波雷达信息投影到图像坐标系中，提出了基于毫米波雷达点云的预设 *ROI*，再进行检测，减少了锚框数量，提升了检测速度，但整个过程中并未解决毫米波雷达信息投影到图像坐标系上存在噪声及高度误差问题。Meyer 等人[2]提出了将毫米波雷达点云转为鸟瞰视角，点云直接输入到 CNN 网络中来进行目标检测。而问题在于一帧毫米波雷达点云过于稀疏，且 CNN 直接作用于点云会产生较多噪点，影响检测精度。高洁等人[3]在目标跟踪框架中，提出将上一帧图像检测结果与当前帧雷达建立图像与雷达点的关联，实现雷达预分类；再利用目标跟踪框架来实现同一雷达点关联，找出属于上一时刻目标在当前时刻的量测，利用 RRPN 建立候选区域，从而得到当前目标检测结果，但同样未考虑毫米波雷达高度信息不准的问题。Nabati 等人[4]提出了

基金项目：军队学科专业建设项目。

作者简介：侯志斌(1990-)，男，硕士研究生，主要研究方向：智能车辆、多模态融合感知。

收稿日期：2022-06-11

Centerfusion 网络,通过毫米波雷达与相机融合进行 3D 目标检测。首先由单目检测结果建立 3D ROI,然后在特征层运用截锥体的毫米波雷达点云与单目初级检测结果建立关联,进行二次回归,补充图像特征,提升检测水平。而问题是,仅在特征层融合毫米波雷达点云信息,会使整体网络框架比较依赖单目 3D 目标检测结果,而单目进行目标检测存在固有缺陷,从而影响最终检测质量。

为此,本文在 Centerfusion 网络的基础上进行改进,提出了毫米波雷达相机两级融合的 3D 目标检测网络,将毫米波雷达信息和图像分别在数据级、特征级两级进行融合,以弥补毫米波雷达投影到图像坐标中高度信息不准以及单模态目标检测存在的不足,提升 3D 目标检测精度以及在复杂天气条件下或对远距离小目标检测的鲁棒性。

1 两级融合的 3D 目标检测网络框架

本节将主要介绍雷达和相机传感器二级融合的 3D 目标检测框架。首先,在输入端将毫米波点云信息进行预处理后与相机建立数据层融合,生成三通道图像附加雷达信息;采用加入注意力机制的 CenterNet 网络[4]作为基于中心的目标检测网络,进行初级检测,回归出目标的属性、三维位置、方向和尺寸等初级三维检测结果,克服了相机单模态目标检测存在的固有缺陷,提升了小目标、模糊目标、以及不利气候条件下的检测精度;然后参照文献[5]中方法再进行特征层融合,使用截锥体机制将雷达检测与其对应对象的中心点相关联,并利用雷达和图像特征,进一步估计深度、速度、旋转和属性来提升初步检测精度,网络结构如图 1 所示。

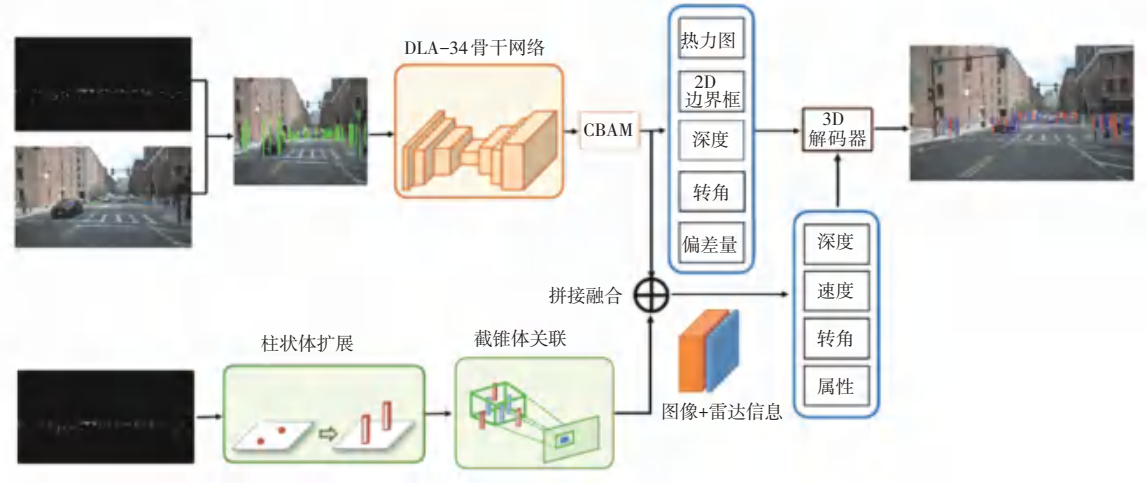


图 1 两级融合的 3D 目标检测网络结构图

Fig. 1 Structure diagram of 3D object detection network with two-level fusion

1.1 毫米波雷达与相机信息数据层融合

毫米波雷达和相机对目标的检测是相互独立的,各自的测量数据也基于不同坐标系。因此,在进行信息融合前,需将雷达和相机测量的目标数据转换到相同的坐标系中,需对不同传感器的目标数据进行空间配准。毫米波雷达与相机涉及到 5 个不同坐标系之间的转换,坐标系之间的关系如图 2 所示。

本文基于数据集开展研究,因此可通过数据集中相机内外参数[6],将毫米波雷达信息投影到图像坐标系上。为解决毫米波雷达投影到图像坐标系下高度信息不准的问题,改进使用文献[7]中方法,将毫米波雷达信息进行条码化改进处理。将其扩展为 2.5 m 红色线段,以确保在图像坐标系下,将检测物体(汽车、卡车、摩托车、自行车和行人等)进行覆盖。雷达数据以像素宽度 2 映射到图像平面,使相机像素与毫米波雷达信息建立基本。雷达回波的特征作为像素值投影到三通道图像中,在不存在雷达回波的图像像素位置,将投影雷达通道值设置为 0。输入图像转为附加有毫米波雷达信息的三通道图像,如图 3 所示。同时为解决毫米波雷达稀疏的问题,本文将 6 个雷达周期共同融合到本文的数据格式中,来增加雷达数据的密度。

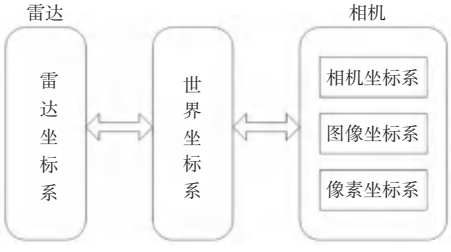


图 2 坐标系关系示意图

Fig. 2 Diagram of coordinate system relationship

（a）毫米波雷达原始信息　　（b）毫米波雷达条码化处理

图 3　毫米波雷达点云条码化处理示意图

Fig. 3　Schematic diagram of barcode processing of millimeter wave radar point cloud

1.2　初级检测网络

1.2.1　加入空间通道注意力机制的关键点检测网络

初级检测使用 CenterNet 框架作为基础网络，DLA－34 网络[5]作为骨干网络。为提取三通道雷达图像信息中雷达投影信息，本文在骨干网络末端加入空间通道注意力模块[8] CBM 和 SAM，对卷积特征的通道和空间建立注意力机制。其中，通道注意力模块 CBM 结构如图 4 所示。

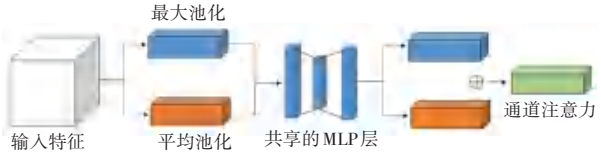


图 4　CBM 通道注意力模块

Fig. 4　CBM channel attention module

上述方法的数学推导见式(1)：

$$\boldsymbol{M}_c(\boldsymbol{F}) = \sigma(MLP(AvgPool(\boldsymbol{F})) + MLP(MaxPool(\boldsymbol{F}))) = \sigma(\boldsymbol{W}_1(\boldsymbol{W}_0(\boldsymbol{F}^c_{avg})) + \boldsymbol{W}_1(\boldsymbol{W}_0(\boldsymbol{F}^c_{\max}))) \tag{1}$$

其中，$\boldsymbol{F}$ 为输入特征，经过并行的平均池化层和最大池化层后，得到 2 个多通道 1×1 维度特征图后，再将其分别送入一个 2 层 MLP 网络中。将 MLP 输出的特征进行张量内对应元素（element－wise）相加，再经过激活操作，生成通道特征 $\boldsymbol{M}_c$，最后将 $\boldsymbol{M}_c$ 和输入特征做张量内对应元素相乘，作为通道注意力模块。

之后，将 CBM 注意力模块输出作为 SAM 注意力模块输入，建立空间注意力机制。空间注意力模块 SAM 结构如图 5 所示。

上述方法的数学推导见式(2)：

$$\boldsymbol{M}_s(\boldsymbol{F}) = \sigma(f^{7\times7}([AvgPool(\boldsymbol{F}); MaxPool(\boldsymbol{F})])) = \sigma(f^{7\times7}([\boldsymbol{F}^s_{avg}; \boldsymbol{F}^s_{\max}])) \tag{2}$$

其中，$\boldsymbol{F}$ 为输入特征图。首先做一个基于通道的全局最大池化和全局平均池化，得到 2 个 $H\times W\times1$ 的特征图，将这 2 个特征图基于通道做通道拼接，并经过一个 7×7 卷积操作，降维为 1 个通道，即 $H\times W\times1$；再经过激活函数生成空间注意力特征，最后将该特征与模块输入做乘法，得到最终生成特征。

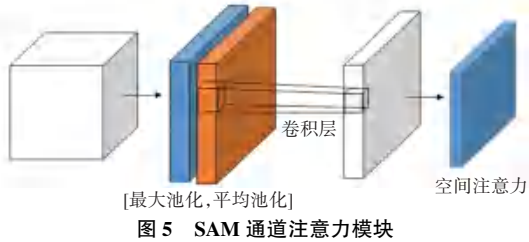


图 5　SAM 通道注意力模块

Fig. 5　SAM channel attention module

将附加有关联条码化雷达信息的三通道图像 $I_{I+R}\in R^{W\times H\times3}$ 作为输入。为防止雷达投影到三通道图像导致完全覆盖三通道信息，影响网络泛化水平，建立投影权重系数 α。经过实验，当 $\alpha=0.6$ 时检测结果最佳。作为超参，则三通道图像为：

$$I_{I+R} = \alpha\times radar + (1-\alpha)\times image \tag{3}$$

关键点热力图输出为：

$$\hat{Y}\in[0,1]^{\frac{W}{R}\times\frac{H}{R}\times C} \tag{4}$$

其中，W、H 是图像的宽和高；R 是下采样率；C 是检测对象类别。$\hat{Y}_{x,y,c}=1$ 表示在图像上检测到的为关键点，而 $\hat{Y}_{x,y,c}=0$ 表示背景。输出热力图如图 6 所示。

图 6　输出热力图

Fig. 6　Output heat map

在网络训练阶段,参照文献[9],则标注物体关键点位置信息为 $p_i \in R^2$, 检测对象类别为 C, 经过下采样对应的关键点 $\tilde{p} = \lfloor \frac{p}{R} \rfloor$。通过高斯核滤波 $Y_{xyc} = \exp\left(-\frac{(x-\tilde{p}_x)^2 + (y-\tilde{p}_y)^2}{2\sigma_p^2}\right)$ (σ_p 是对象大小自适应标准偏差),将标注的物体关键点显示在热力图 $Y \in [0,1]^{\frac{W}{R}\times\frac{H}{R}\times C}$ 上,对于图像上给定的标注数据 p_0、p_1 等,建立 focal 损失函数[9]:

$$L_k = \frac{1}{N}\sum_{xyc}\begin{cases}(1-\hat{Y}_{xyc})^{\alpha}\log(\hat{Y}_{xyc}) & Y_{xyc}=1\\ (1-Y_{xyc})^{\beta}(\hat{Y}_{xyc})^{\alpha}\log(1-\hat{Y}_{xyc}) & \text{otherwise}\end{cases} \tag{5}$$

其中,N 是图像中关键点数量,参照文献[4]中研究结果,在实验中将 $\alpha=2$、$\beta=4$ 定义为超参数。为恢复由下采样引起的误差,对每个预测的关键点增加一个局部偏移量,偏移量建立 L_1 损失函数:$L_{off} = \frac{1}{N}\sum_{p}\left|\hat{O}_{\tilde{p}} - \left(\frac{p}{R} - \tilde{p}\right)\right|$,偏移量仅在预测关键点 $\tilde{p}$ 生效。

设 $(x_1^{(k)}, y_1^{(k)}, x_2^{(k)}, y_2^{(k)})$ 是种类为 C_k 的物体 k 的边界框位置坐标,其中心点位于 $\left(\frac{x_1^{(k)}+x_2^{(k)}}{2}, \frac{y_1^{(k)}+y_2^{(k)}}{2}\right)$,每一个对象其尺寸 $s_k = (x_2^{(k)} - x_1^{(k)}, y_2^{(k)} - y_1^{(k)})$。通过检测网络预测所有关键点 $\hat{Y}$,对尺寸建立与关键点偏差量相近的损失函数 L_1:

$$L_{size} = \frac{1}{N}\sum_{k=1}^{N}\left|\hat{S}_{p_k} - s_k\right| \tag{6}$$

故热力图生成总的损失函数为:

$$L_{det} = L_k + \lambda_{size}L_{size} + \lambda_{off}L_{off} \tag{7}$$

1.2.2 通过关键点进行 3D 目标检测

3D 目标检测是对生成的每个关键点附加 3 个属性(深度、3D 尺寸、方向),并为每个属性增加一个单独的检测头。深度信息对于每个关键点是一个标量,其很难直接回归,将深度作为关键点预测的一个额外的输出通道 $\hat{D} \in [0,1]^{\frac{W}{R}\times\frac{H}{R}}$,该通道使用了 2 个卷积层和 1 个 *ReLU* 模块,参照文献[10],对输出预测深度做出变换,即 $d = 1/\sigma(\hat{d}) - 1$,这里 σ 是 *sigmoid* 函数,训练时建立 L_1 损失函数:

$$L_{dep} = \frac{1}{N}\sum_{k=1}^{N}\left|\frac{1}{\sigma(\hat{d}_k)} - 1 - d_k\right| \tag{8}$$

其中,d_k 是标注信息(g_t)的绝对深度,以 m 为单位。

3D 尺寸是 3 个标量值,使用单独的检测头作为 3 个通道 $\hat{\Gamma} \in R^{\frac{W}{R}\times\frac{H}{R}\times 3}$ 直接回归得出。同深度类似,训练时建立 L_1 损失函数:

$$L_{dim} = \frac{1}{N}\sum_{k=1}^{N}\left|\hat{\gamma}_k - \gamma_k\right| \tag{9}$$

其中,γ_k 是标注物体的高、宽、长,以 m 为单位。

目标方向是一个标量,直接回归比较困难。参考文献[11]中方法,将方向解码为 2 个 *bin*、8 个标量,即每个 *bin* 有 4 个值,预测的 *bin* 表示为 $\hat{\alpha} = [\hat{b}_1, \hat{a}_1, \hat{b}_2, \hat{a}_2]$。$\hat{b}_1$、$\hat{b}_2$ 用作 *softmax* 分类,$\hat{a}_1$、$\hat{a}_2$ 用于回归到每个 *bin* 中的角度,预测角度通过如下公式解码:

$$\hat{\theta} = \arctan 2(\hat{a}_{j1}, \hat{a}_{j2}) + m_j \tag{10}$$

训练时建立 L_1 损失函数为:

$$L_{ori} = \frac{1}{N}\sum_{k=1}^{N}\sum_{i=1}^{2}(softmax(\hat{b}_i, c_i) + c_i\left|\hat{a}_i - a_i\right|) \tag{11}$$

1.3 毫米波雷达与相机信息特征层融合

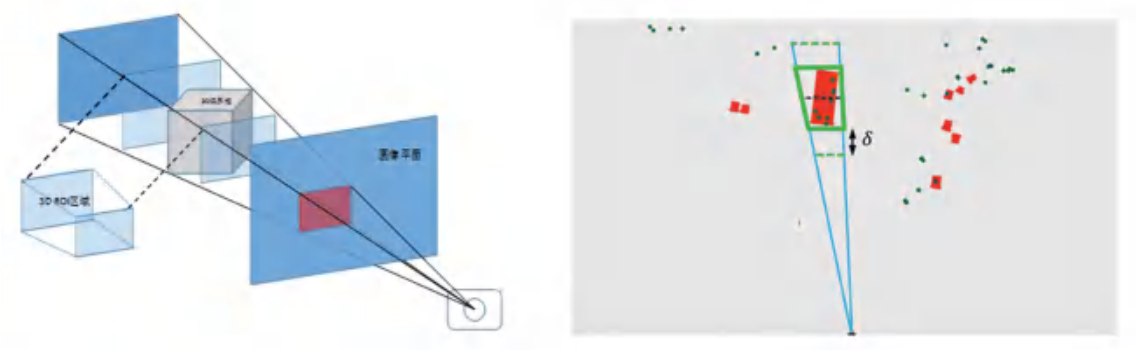
经过初级检测网络,生成了目标的热力图、2D 目标尺寸、3D 目标尺寸、深度、方向、偏差等。为进一步提升精度,需在特征层进行二次融合。

1.3.1 雷达关联

参照文献[4]中截锥体关联方法,在特征层将毫米波雷达点云扩展为垂直柱体,为解决高度不准确问题,使用初级检测中生成的边界框(bboxing)及其回归的深度和目标尺寸来创建一个 3D 兴趣区域(3D RoI)截锥体,并忽略截锥体之外的任何点。为消除多检测关联问题,在此 RoI 内有多个毫米波雷达点云,本文将最近的点作为对应于这个对象的雷达检测,如图 7 所示。其中,图 7(a)为基于对象的 3D 边界框生成截锥体的兴趣区域,图 7(b)为鸟瞰视角下的截锥体关联机制示意图。

1.3.2 雷达特征提取

在雷达信号与其对应目标关联后,使用雷达信号中的深度和速度为图像,创建互补特征。其中,对于每一个与物体相关的雷达信号,都会生成 (d, v_x, v_y) 三个以物体的 2D 边界框为中心的热力图通道。热力图的宽度和高度与对象的二维边界框成比例,热图值是标准化的物体深度 d,也是在自车坐标系中径向速度(V_x 和 V_y)的 X 和 Y 分量:



（a）基于对象的 3D 边界框生成截锥体的兴趣区域　　（b）截锥体关联机制鸟瞰图

图 7　截锥体关联方法示意图

Fig. 7　Schematic diagram of frustum correlation method

$$F^{j}_{x,y,i}=\frac{1}{M_i}\begin{cases}f_i & |x-c^j_x|\leqslant\alpha\, w^j \text{ and}\\ & |y-c^i_y|\leqslant\alpha\, h^j\\ 0 & \text{otherwise}\end{cases}\tag{12}$$

其中，$i\in 1,2,3$ 是特征通道；M_i 是规范化因子；f_i 是 (d,v_x,v_y) 的特征值；c^j_x 和 c^i_y 是第 j 个对象在图像上中心点的 x、y 坐标；w^j、h^j 是第 j 个对象 2D 边界框的宽度和高度。

如果 2 个对象具有重叠的热图区域，则深度值较小的对象占主导地位，因为只有最近的对象在图像中才完全可见。

生成的热力图作为额外通道连接到图像特征，这些特征作为二次回归输入，重新估算对象的三维信息、以及速度和类别。与初级检测相比，经过特征融合后，有助于从雷达特征中学习更高层次的特征，最后将生成值解码为 3D 边界框。3D 边界框从初级检测器获得 3D 尺寸，并从二次回归中得到估计的深度、速度、转角和类别。

2　实验分析与对比验证

2.1　数据集

本文使用 nuScenes 数据集[12]进行模型训练及测试。该数据集是第一个携带毫米波雷达信息的自动驾驶场景数据集，其中涵盖了在波士顿和新加坡采集的 1 000 个场景的数据，是目前最大的具有三维目标标注信息的自动驾驶汽车多传感器数据集。其传感器配置上含有 6 个摄像头、5 个雷达和 1 个激光雷达，所有这些都具有全 360°视野。传感器参数见表 1。

表 1　nuScenes 数据集传感器参数表

Tab. 1　Sensor parameters of nuScenes dataset

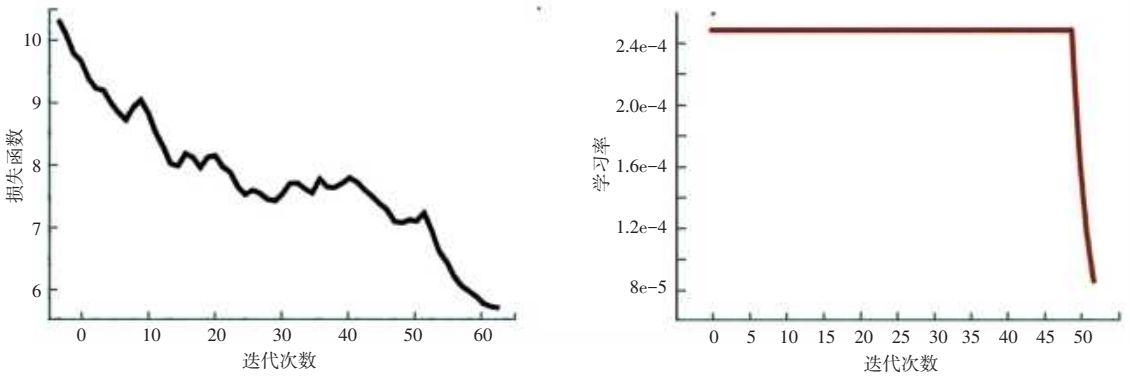
传感器	参数
6 个相机	RGB 相机、12 Hz 采样率、1 600 * 900 分辨率、CMOS 传感器
1 个雷达	32 线、20 线雷达，20 Hz 采样率，采样距离 70 m，360°水平视场，-30°到 10°的垂直视场，正负 2 cm 精度误差
5 个毫米波雷达	采样范围小于 250 m，77 GHz、FMCW、13 Hz 捕获频率

2.2　实验设置

本文采取网络骨干为 DLA-34 的 CenterNet 网络进行训练。训练时采取 Centerfusion 提供的预训练模型进行训练，同时在不同位置加入注意力机制进行性能对比实验。实验平台的操作系统为 ubuntu16.04，并带有型号为 GeForce GTX 1050 的 GPU。

训练阶段共迭代 60 个 *epoch*，训练批次大小设置为 2，初始学习率为 2.4e-4，同时采用学习率衰减策略，训练 50 个 *epoch* 后学习率下降 10%。三通道图像输入到网络前进行随机左右翻转、随机移位等数据加强。测试阶段，采用 60 个 *epoch* 的训练权重，来对本文方法进行测试。

以下实验均使用单个 GPU 完成。由于完整数据集较大，本文仿真主要通过 nuScenes 的 v1.0-mini 数据集进行训练，重点测试改进的网络检测精度。v1.0-mini数据集是由整个数据集中抽取出的 10 个场景组成，其中训练样本为 14 065 个，测试样本为 6 019 个，训练收敛曲线如图 8 所示。



(a) 回归任务损失曲线 (b) 学习率变化曲线

图 8 训练过程收敛曲线

Fig. 8 Convergence curve of training process

2.3 3D 目标检测数据对比

以 Centerfusion 作为基准网络,为确保训练及测试数据相一致,用 nuScenes v1.0-mini 对 Centerfusion 重新进行训练及测试,测试集选用数据集中的"scene-0103"、"scene-0916" 两个场景作为 mini-test 集,并与本文方法进行比较。表 2 中列出了对 Centernet(3d)、Centerfusion 和本文方法进行 3D 目标检测性能的比较结果。可以看出,在 mini 集进行训练、在 mini-test 集进行测试后,检测分数(*NDS*)上升了近 1.21%。图 9 展示了 Centerfusion 和本文方法 *NDS* 的收敛过程。

表 2 3D 检测性能对比表

Tab. 2 3D detection performance comparison table

方 法	检测分数 (*NDS*) ↑	平均精度 (*mAP*) ↑	平均平移误差 (*mATE*) ↓	平均比例误差 (*mASE*) ↓	平均方位误差 (*mAOE*) ↓	平均速度误差 (*mAVE*) ↓	平均属性误差 (*mAAE*) ↓
Centernet	0.461 8	0.442 4	0.589 6	0.398 4	0.496 7	0.840 8	0.268 1
Centerfusion	0.515 3	0.455 9	**0.552 5**	0.390 8	0.433 7	0.489 4	0.259 6
The proposed	**0.527 4**	**0.466 8**	0.568 9	**0.389 2**	**0.413 7**	**0.431 7**	**0.256 3**

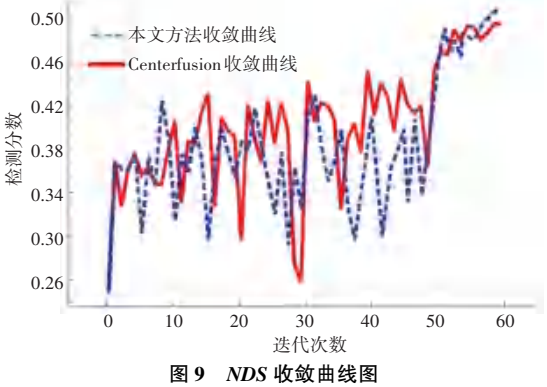


图 9 *NDS* 收敛曲线图

Fig. 9 *NDS* convergence curve

由图 9 中可见,随着训练迭代次数的增多,本文方法与 Centerfusion 均呈现抖动上升趋势,在训练 60 个迭代周期后,本文网络 *NDS* 指标明显高出约 0.03。

nuScenes v1.0-mini 数据集中 7 类物体检测的平均精度结果见表 3。由表 3 可见,在测试集中,本文方法在巴士、行人、摩托车、自行车等的检测精度均高于 Centerfusion 检测结果。尤其是对于自行车的检测精度上,相比提升了近 40%。

表 3 3D 目标检测对象精度对比表

Tab. 3 Object accuracy comparison table of 3D target detection

方法	检测精度 (*AP*)						
	轿车	货车	巴士	行人	摩托车	自行车	交通锥
Centerfusion	0.336	**0.336**	0.427	0.359	0.463	0.136	**0.213**
The proposed	0.336	0.281	**0.469**	**0.363**	**0.484**	**0.569**	0.187

2.4　通道空间注意力机制对比实验

本文采取 2 种注意力机制 CBM、SAM 的对比实验，主要对比 CBM、SAM 加入位置及初始网络权重等在网络中发挥的作用。实验中，分别在骨干网络中的基本模块和骨干网络末端加入空间通道注意力机制。如图 10 所示，在骨干网络中加入空间通道注意力机制，使用预训练模型，新增注意力机制模块默认使用 kaiming 初始化网络权重，在训练 180 个迭代周期后，实验结果检测精度（*NDS*）仅为 0.209 4，效果并不理想。

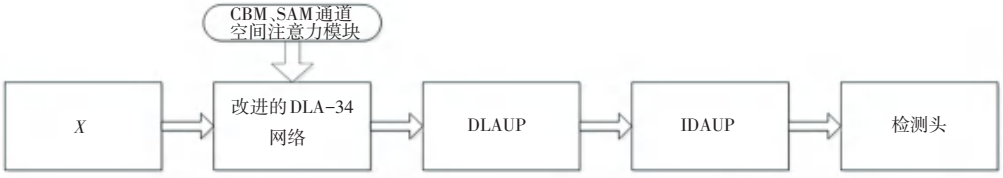


图 10　骨干网络中加入注意力机制示意图

Fig. 10　Schematic diagram of adding attention mechanism to the backbone network

将空间通道注意力机制加入骨干网络末端，如图 11 所示。首先，在冻结改进的 DLA-34、DLAUP 上采样层、IDAUP 融合网络层后，训练 60 个迭代周期，然后再联合训练 60 个周期，*NDS* 即上升至 0.527 4。实验得出结论是：注意力机制在迁移学习方法下，加入到骨干网络末端和检测头相比于骨干网络中效果更优。

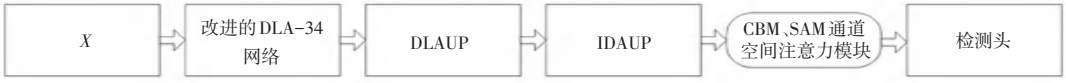


图 11　骨干网络末加入注意力机制示意图

Fig. 11　Schematic diagram of adding attention mechanism at the end of backbone network

2.5　实验结果可视化分析

本文对基础网络模型和毫米波雷达与相机两级融合的网络模型的检测效果的可视化比较结果如图 12、图 13 所示。从可视化效果可以看出：2 种方法均能实现较好的 3D 目标检测效果，但本文的方法对远距离小目标漏检率低，且具有更强的鲁棒性；相比来看，本文方法的 3D 边界框更加准确，在一些特定场景中，误检率明显降低。

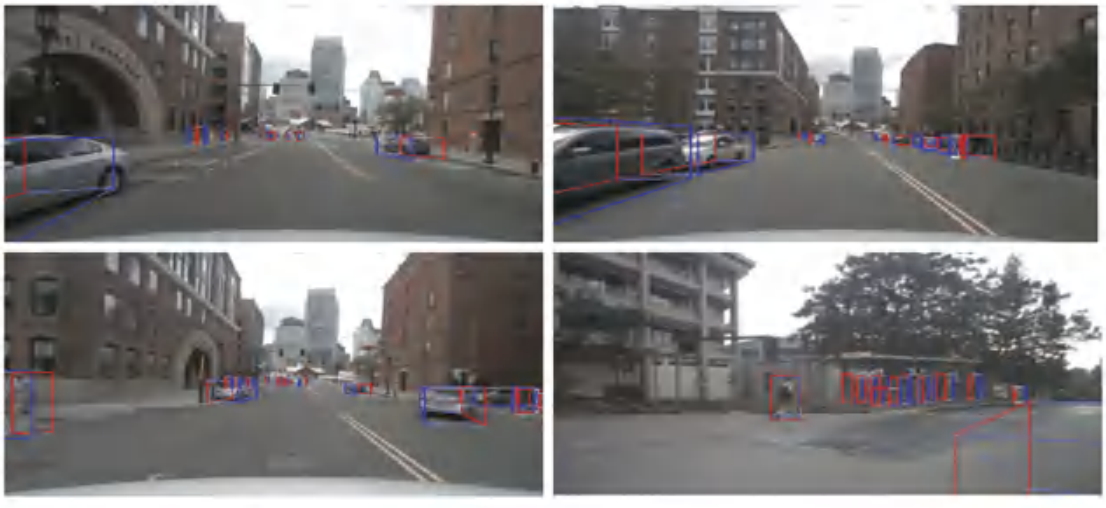


图 12　Centerfusion 可视化效果图

Fig. 12　Centerfusion visualization

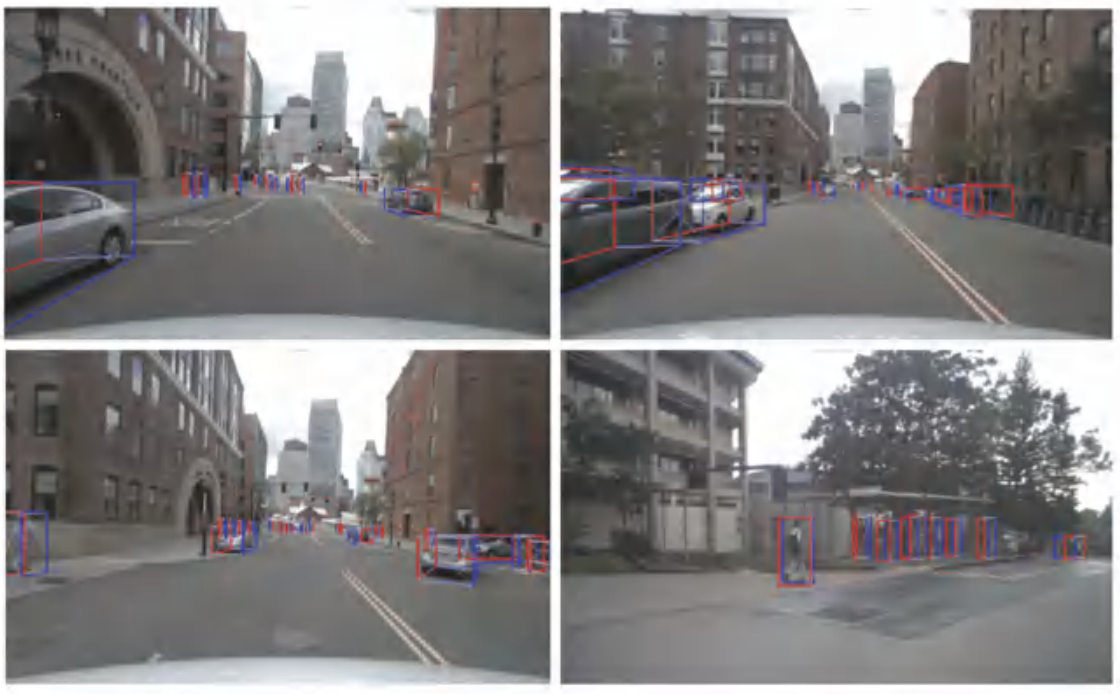


图 13 两级融合网络可视化效果图

Fig. 13 Visual renderings of two-level fused network

3 结束语

本文在毫米波雷达和相机特征层融合网络 Centerfusion 的基础上进行改进,针对原算法在一阶段未考虑单目检测固有缺陷的问题,提出了一种毫米波雷达与相机两级融合的 3D 目标检测算法,将雷达点云信息进行处理后,在数据层和特征层均进行融合;同时在一阶段中心点检测网络中加入了注意力机制。实验证明,本文方法相比原算法在复杂恶劣天气条件下以及对远距离小目标的检测效果上均有提升,在大型自动驾驶数据集 nuScenes3D 检测基准上,评估了本文提出的方法,相比 Centerfusion 检测分数 (*NDS*) 有了一定提升。

参考文献

[1] NABATI R, QI H. RRPN: Radar region proposal network for object detection in autonomous vehicles [C]//2019 IEEE International Conference on Image Processing (ICIP). Taipei: IEEE,2019:3093-3097.

[2] MEYER M, KUSCHK G. Deep learning based 3d object detection for automotive radar and camera[C]//2019 16th European Radar Conference(EuRAD). Paris:IEEE,2019:133-136.

[3] 高洁,朱元,陆科. 基于雷达和相机融合的目标检测的方法[J]. 计算机应用,2021,41(11):3242-3250.

[4] NABATI R, QI H. Centerfusion: Center-based radar and camera fusion for 3d object detection[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. IEEE, 2021:1527-1536.

[5] NOBIS F, GEISSLINGER M, WEBER M. et al. A deep learning-based radar and camera sensor fusion architecture for object detection [C]//2019 Sensor Data Fusion: Trends, Solutions, Applications(SDF). Bonn, Germany:IEEE,2019:1-7.

[6] YU F, WANG D, SHELHAMER E, et al. Deep layer aggregation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.Salt Lake City,USA: IEEE,2018:2403-2412.

[7] ZHANG Zhengyou. A flexible new technique for camera calibration [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2000,22(11):1330-1334.

[8] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision(ECCV). Munich, Germany:dblp, 2018:3-19.

[9] LAW H, DENG J. Cornernet:Detecting objects as paired keypoints [C]//Proceedings of the European Conference on Computer Vision(ECCV). Munich, Germany:dblp, 2018:734-750.

[10] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss ofr dense object detection [C]//Proceedings of the IEEE International Conference on Computer Vision.Venice:IEEE,2017:2980-2988.

[11] EIGEN D , PUHRSCH C, FERGUS R. Depth map prediction from a single image using a multi-scale deep network [C]// Advances in Neural Information Processing Systems. Montreal, Quebec, Canada,:NIPS Foundation ,2014:2366-2374.

[12] MOUSAVIAN A, ANGUELOV D, FLYNN J, et al. 3d bounding box estimation using deep learning and geometry[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA:IEEE,2017:7074-7082.

[13] CAESAR H, BANKITI V, LANG A H, et al. Nuscenes: A multimodal dataset ofr atuonomous driving[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA:IEEE,2020:11621-11631.