

doi: 10.19562/j.chinasae.qcgc.2024.12.014

基于点云时空特征补偿网络的智能网联车辆 协同感知*

张名芳, 刘颖, 马健, 何焱, 王力

(北方工业大学, 城市道路交通智能控制技术北京市重点实验室, 北京 100144)

[摘要] 为克服网络延迟对协同感知准确率的影响,同时提高点云特征表达能力,本文提出了一种基于时空特征补偿网络的智能网联车辆点云协同感知方法。首先,采用点-体柱特征提取方法对点云原始数据进行处理,将扫描点局部邻域特征与体柱特征图进行拼接;然后,设计基于PredRNN算法的时间延迟补偿模块,对接收到的来自周围网联车辆的历史帧点云特征进行预测,实现两车点云特征同步;其次,利用空间特征融合补偿模块聚合跨车辆点云特征,通过双向多尺度特征金字塔网络融合多分辨率特征,输出车辆目标几何尺寸和航向角等信息;最后,在V2V4real数据集和自制数据集上的测试结果表明,该方法的检测精度优于经典协同感知算法,且对不同网络延迟时间具有较好的适应性,推理时间满足实时性要求。

关键词: 智能网联车辆; 协同感知; 时空特征补偿; 点云; 特征融合

Collaborative Perception Based on Point Cloud Spatio-Temporal Feature Compensation Network for Intelligent Connected Vehicles

Zhang Mingfang, Liu Ying, Ma Jian, He Ye & Wang Li

North China University of Technology, Beijing Key Lab of Urban Intelligent Traffic Control Technology, Beijing 100144

[Abstract] In order to overcome the influence of network latency on the cooperative perception accuracy and simultaneously improve the point cloud feature expression capability, a cooperative perception method based on point cloud spatio-temporal feature compensation network for intelligent connected vehicles is proposed. Firstly, the point-to-pillar feature extraction method is used to process the raw point cloud data, and the local neighborhood features of the laser points are then spliced with pillar feature maps. Secondly, the temporal latency compensation module based on the PredRNN algorithm is designed to predict the point cloud features of historical frames received from the surrounding connected vehicles, so as to achieve the synchronization of point cloud features from two vehicles. Thirdly, the spatial feature fusion compensation module is utilized to aggregate the inter-vehicle point cloud features, and multi-resolution features are fused through the bidirectional multi-scale feature pyramid network. The output includes vehicle target geometry size, heading angle and other information. Finally, the test results on the V2V4real dataset and the self-collected dataset demonstrate that the detection accuracy of the proposed method is superior to classical cooperative perception algorithms. Furthermore, it exhibits good adaptability to various latency cases and the inference process meets the real-time requirements.

Keywords: intelligent connected vehicles; collaborative perception; spatio-temporal feature compensation; point cloud; feature fusion

* 国家重点研发计划项目(2022YFB4300400)和北京市教育委员会科学研究计划项目(KM202210009013)资助。

原稿收到日期为2024年05月11日,修改稿收到日期为2024年06月19日。

通信作者:张名芳,副教授,博士,E-mail:mingfang@ncut.edu.cn。

前言

准确感知动态道路交通环境是智能网联车辆实现可靠决策与运动规划的前提^[1]。目前常用的环境感知传感器包括摄像头、毫米波雷达和激光雷达。由于摄像头容易受光照影响,毫米波雷达、2.5D激光雷达探测角度有限,使得3D激光雷达凭借全方位深度扫描优势在环境感知领域得到广泛应用^[2-3]。

现阶段已有的3D点云环境感知方法^[4-5]大多针对单车智能,按照体素或BEV表征点云,并采用深度学习算法提取道路交通参与者的多模态特征。这类方法过于依赖单车装载的感知传感器性能。随着V2X技术快速发展^[6],部分学者提出采用协同感知技术弥补单车感知缺陷,如距离受限和盲区遮挡等^[7]。协同感知技术根据数据融合方式可分为前期融合、中期融合和后期融合^[8]。前期融合指对各网联车辆的原始点云数据进行融合。Chen等^[9]首次提出基于原始级别数据融合的协同感知方法,将深度学习算法与Cooper算法相结合进行点云目标检测。实验结果表明,该算法能够扩展感知区域、提高检测准确性,但原始传感器数据量过大导致传输过程中占用大量带宽,实时性较差。后期融合即决策级融合,指各网联车辆在本地完成感知处理后直接接收周围车辆感知结果,进行结果级融合。Hurl等^[10]提出一种后期融合方法用于检测周围车辆;Yu等^[11]选用PointPillars作为3D检测器,将路侧基础设施和车载传感器的检测结果进行融合。这类方法直接传输本地生成的检测结果,虽能减小数据传输量,但过于依赖本地检测,可能使得融合过程中无法有效感知远距离或遮挡目标。

中期融合即特征级融合,指先对各智能网联车辆采集到的数据提取中间特征再进行特征传输。相比于前、后期融合,中期融合方法能更好地保留本地有价值的信息同时减少冗余传输。F-Cooper^[12]采用体素特征编码提取点云体素特征,并通过maxout操作对传输的体素特征进行融合,该方法相较前期融合计算成本更低且实时性更好。V2VNet^[13]利用卷积块提取点云特征并转换成鸟瞰图进行压缩传输,利用空间感知神经网络融合多个周围车辆传输来的特征及自车感知到的特征,该方法能较好地平衡检测准确率和传输带宽要求,但易受网络延迟影响。V2X-ViT^[14]采用PointPillar网络提取点云伪图像特

征,并利用视觉Transformer网络自适应融合车辆和路侧基础设施之间的异构特征,通过自适应延迟感知位置编码解决网络延迟期间运动引起的目标不对齐现象。王秉路等^[15]提出了一种多模态协同感知方法,通过注意力机制融合多车点云特征与多车图像特征检测周围车辆,但该方法没有考虑通信延迟的影响;为改善中期融合方法对网络延迟的适应性并提高点云特征表达能力,本文提出一种基于点云时空特征补偿网络的智能网联车辆协同感知方法,本文主要工作和贡献如下:

(1)构建点-体柱特征提取网络用于提取点云局部空间特征与全局结构特征,在不增加计算成本的前提下,扩大感受野并丰富特征细粒度信息;

(2)提出基于PredRNN的时间延迟补偿模块,用于捕捉点云特征序列时间依赖性并预测特征状态变化,减小因网络延迟引起的特征滞后效应;

(3)根据车间相对位姿关系设计空间特征融合补偿模块,融合多车视角下的多分辨率空间特征信息,有效扩大检测区域范围并提高不同尺度目标感知精度。

1 智能网联车辆协同感知方法

本文所提协同感知方法的整体结构如图1所示。首先将周围车辆采集的点云坐标转换到自车坐标系下,通过构建点-体柱特征提取网络提取点云细粒度特征,并利用特征压缩模块降维传输给自车;然后对来自周围车辆的环境特征进行逆压缩,采用PredRNN算法设计时间特征延迟补偿模块,通过历史特征序列预测未来时刻特征;接着利用跨车辆特征聚合模块进行多源特征对齐与拼接,进一步通过多分辨率特征提取网络提升不同尺寸目标检测鲁棒性,结合浅层和深层特征优势融合多尺度特征,获取具有丰富语义的特征表达,最后将处理后的特征输入至包含全连接层的检测头,生成目标位置、航向角等检测结果。

1.1 坐标转换

为融合自车 CAV_1 和周围车辆 CAV_2 采集到的环境点云信息,须进行坐标转换,使两车坐标系对齐。将 CAV_1 和 CAV_2 的本地笛卡尔坐标系分别记为 $X_1O_1Y_1$ 和 $X_2O_2Y_2$,如图2所示。当 CAV_2 进入有效通信范围时, CAV_1 将自车位姿信息发送给 CAV_2 ,结合两车相对位姿信息计算变换矩阵 T ,利用变换矩阵 T

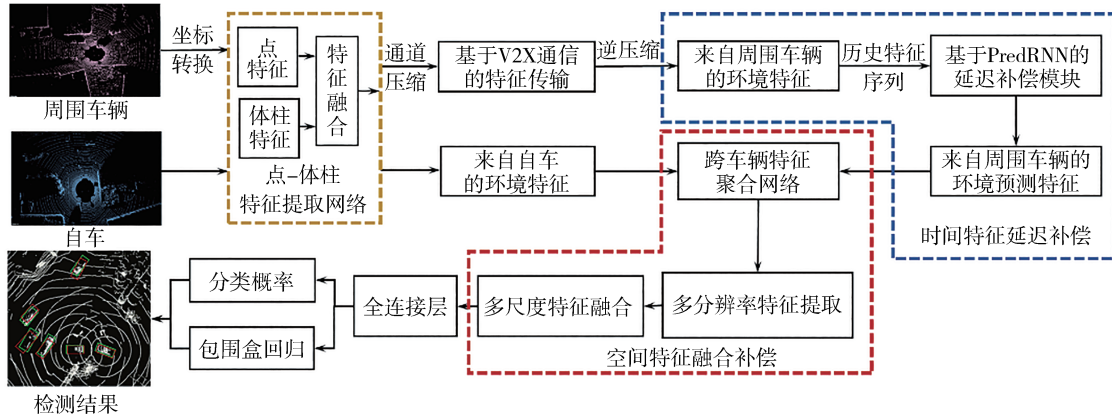


图1 协同感知方法框架

将 CAV_2 采集的点云 P_2 投影到 CAV_1 坐标系下,得到转换后的点云 P'_1 , 具体为

$$\theta = \theta_1 - \theta_2 \quad (1)$$

$$T = \begin{bmatrix} \cos \theta & -\sin \theta & d_x \\ \sin \theta & \cos \theta & d_y \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

$$P'_1 = T \cdot P_2 \quad (3)$$

式中: θ_1 和 θ_2 分别为 CAV_1 和 CAV_2 的航向角; θ 为两车之间的航向角夹角; d_x 和 d_y 分别为 CAV_2 坐标系下两车之间的横纵向距离。

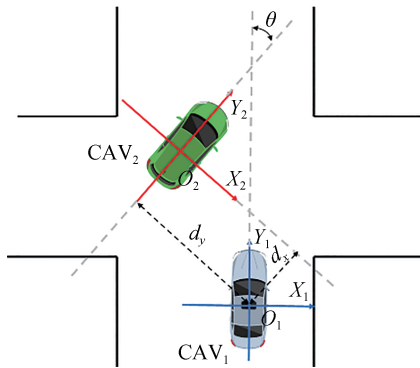


图2 坐标系转换

1.2 点-体柱特征提取网络

1.2.1 点特征提取

由于特征学习常用的最远点采样、K近邻算法时间复杂度较高以及随机采样算法存在采样不均匀问题,为此本文提出快速覆盖感知特征学习层进行点特征学习,以简化中心点采样和邻近点查询过程。为降低远距离点云稀疏影响,将笛卡尔坐标系下坐标转换后的扫描点 (x, y, z) 转换到以各车几何中心为原点的极坐标系下,记为 (ρ, θ, z) 。以极坐标系原

点为中心,沿纵向高度 $[h_{\min}, h_{\max}]$ 、径向半径 R 划分圆柱体兴趣区域,根据径向间隔 $\Delta\rho$ 、圆周间隔 $\Delta\theta$ 、纵向间隔 Δz 分割圆柱体,则各扫描点的体素索引为 $(\lfloor \rho/\Delta R \rfloor, \lfloor \theta/\Delta\theta \rfloor, \lfloor z/\Delta z \rfloor)$, 其中 $\lfloor \cdot \rfloor$ 表示向下取整操作。将带有体素索引的无序点云输入至点特征提取网络。

特征学习过程具体分为3个步骤:圆柱形覆盖感知采样、分组、点特征聚合,如图3所示。图中绿色网格代表非空体素 O_v ,黄色网格代表中心体素 O_c ,蓝色虚线框表示占据 3×3 网格的中心体素的邻近体素,橙色实心点为采样点 p_i ,橙色空心点为 p_i 的邻近点,蓝色空心点表示其他扫描点,橙色虚线框表示采样点感受野。

首先利用贪婪算法从非空体素 O_v 中随机采样中心体素 $O_c = \{v_c^i | i = 1, \dots, m_1\}$, 将未被采样的体素记为候选体素 $O_p = \{v_p^j | j = 1, \dots, m_2\}$, 对每个中心体素进行随机采样得到采样点 p_i , 降低点云处理运算量同时保留点云关键信息。中心体素采样流程具体为:遍历每个候选体素 v_p^j , 对 v_c^i 发起挑战,比较覆盖范围的收益值 A 和代价值 L , 若 $A > L$, 则用 v_p^j 替换对应的 v_c^i , 反之则不替换。

假设 $\delta(x) = \begin{cases} 1, & x = 0 \\ 0, & x \neq 0 \end{cases}$, 则收益值 A 计算如下:

$$A = \sum_{v \in \pi(v_p)} \delta(c_v) \quad (4)$$

式中: $\pi(v_p)$ 为候选体素 v_p 的邻近体素; c_v 为体素 v 的邻近体素中 v_c 的数量。 $\delta(c_v)$ 的含义为: $c_v = 0$, 没有 v_c 覆盖体素 v , 而 v_p 覆盖到了体素 v , 因此收益值增加 1; 反之则不增加。

代价值 L 计算如下:

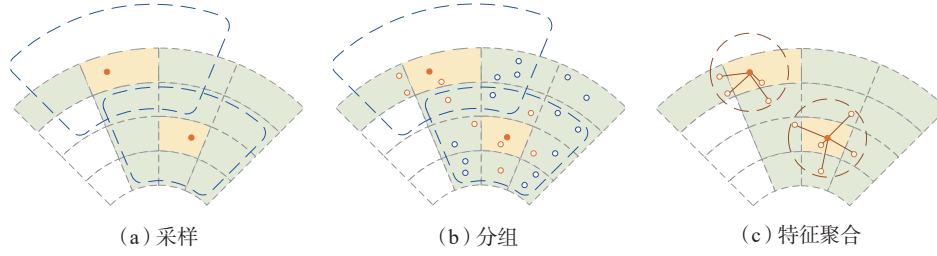


图3 点特征提取

$$L = \sum_{v \in \pi(v)} \delta(c_v - 1) \quad (5)$$

采样完成后,以每个采样点为中心,只从采样点邻近体素中根据欧氏距离搜寻 K 个邻近点。相比于传统KNN算法依赖于无序点云的三维坐标来确定 K 个邻近点,该方法能够通过体素索引快速搜寻到邻近点,显著降低计算成本。

接着利用多层感知机将 K 个邻近点 $\{p_i^k = \{x_i^k, f_i^k\} | k = 1, 2, \dots, K\}$ 的特征聚合到采样点 $p_i = \{x_i, f_i\}$,学习得到局部邻域特征 f_i :

$$f_i = \max_{k=1, \dots, K} \text{pool}(\text{MLP}(\text{concat}(x_i^k - x_i, f_i^k))) \quad (6)$$

式中: \max_{pool} 为最大池化操作; MLP 为多层感知机; concat 表示两个向量的串联操作。

1.2.2 体柱特征提取

为保证后续跨车辆特征聚合一致性,将各车几何中心作为中心点,以 CAV_1 运动方向 $[y_{\min}, y_{\max}]$ 、水平方向 $[x_{\min}, x_{\max}]$ 为兴趣区域,划分大小均匀的网格,网格尺寸设为 m ,采用PointPillar方法^[16]对网格内垂直方向无高度限制的体柱初步提取点云特征。考虑到点云分布不均匀可能导致体柱间点云数量差异大,创建尺寸为 (D, P, N) 的密集张量,其中 D 为维度, P 为非空体柱数量, N 为各体柱内点数阈值。若体柱内点数大于 N ,则对该体柱内的点进行随机采样;反之则进行零填充。接着利用一个线性层、BatchNorm和ReLU激活函数对密集张量进行扩充,生成尺寸为 (C, P, N) 的张量,其中 C 表示通道数。再对每个体柱进行池化得到尺寸为 (C, P) 的张量。最后,根据每个体柱的索引值将生成的 (C, P) 张量映射到二维平面上,生成尺寸为 (H, W, C) 的伪图像特征 f ,其中 H 和 W 分别为高度和宽度:

$$W = (x_{\max} - x_{\min})/m; H = (y_{\max} - y_{\min})/m \quad (7)$$

1.2.3 点-体柱特征融合

点特征提取方法尽管感受野更大,但计算成本较高;体柱特征提取方法运算效率高,但过多的卷积

操作会导致信息丢失,细粒度特征精度降低。因此,本文结合这两类方法的优点提出点-体柱特征融合方法。

首先计算点特征提取方法采样得到的点 $p_i(x_i, y_i, z_i)$ 在体柱特征图 $f(H, W, C)$ 中的索引:

$$a = \left\lfloor \frac{x_i}{m} \right\rfloor; b = \left\lfloor \frac{y_i}{m} \right\rfloor \quad (8)$$

将特征图 $f(H, W, C)$ 的网格 (a, b) 内所有采样点纳入索引集合 S_{ab} 中:

$$S_{ab} = \{p_i(x_i, y_i, z_i) \rightarrow (a, b)\} \quad (9)$$

利用最大池化操作对网格 (a, b) 内的采样点进行特征聚合:

$$f_{ab} = \max_{k \in S_{ab}} \text{pool}(f_i) \quad (10)$$

式中 f_{ab} 表示网格 (a, b) 内采样点特征的最大池化结果。

将池化后的采样点特征 f_{ab} 拼接到体柱特征图对应网格的通道维度上,并通过多层感知机将拼接后的特征转换到原始维度,得到融合后的特征 $f_u \in \mathbb{R}^{(H \times W \times C)}$:

$$f_u = \text{MLP}(\text{concat}(f(a, b, C_{ab}), f_{ab})) \quad (11)$$

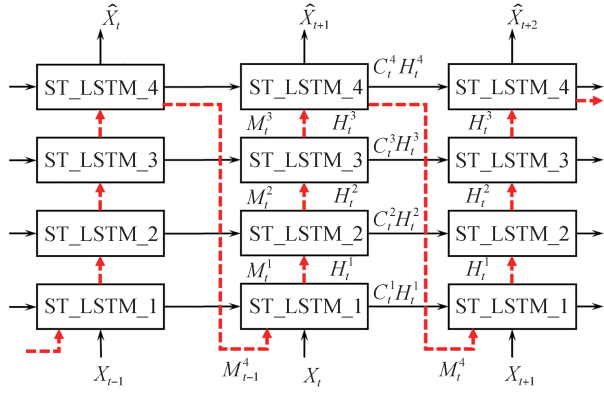
式中 $f(a, b, C_{ab})$ 表示网格 (a, b) 的通道特征。

为尽可能降低网络传输过程所需带宽,在通道维度上逐步压缩特征图,采用 1×1 的卷积核将特征 f_u 的通道数量 C 降维到 C' ,压缩后的特征 f_v 的尺寸为 (H, W, C') ,网联环境下 CAV_2 将特征 f_v 通过网络发送到 CAV_1 , CAV_1 接收到特征后采用 1×1 的卷积核进行反向操作,将特征 f_v 的尺寸投影回 (H, W, C) 。

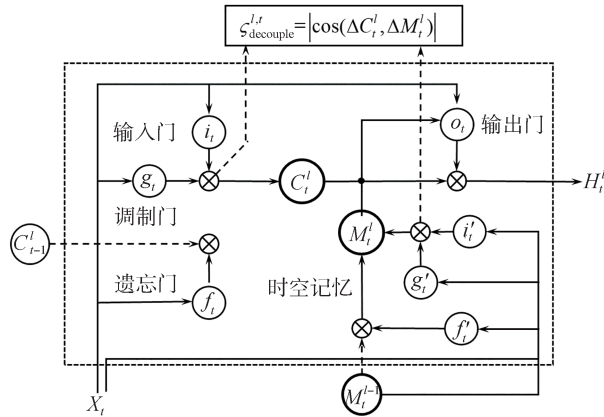
1.3 基于PredRNN的时间延迟补偿模块

由于V2V通信网络传输数据存在延迟,使得 CAV_1 接收到的 CAV_2 点云特征具有时间滞后效应,无法准确描述当前帧 CAV_2 点云真实状态。考虑到PredRNN算法^[17]采用时空LSTM替代传统LSTM单元,具有较好的时空序列数据预测能力,本文设计基于PredRNN算法的时间延迟补偿模块,对 CAV_2 点云

历史帧特征进行预测,实现两车点云状态时空同步。该模块的网络结构如图4(a)所示,通过引入锯齿状的信息传播方式,使得记忆状态信息 C_t 能在同层中沿时间方向横向传递,捕捉时间长期依赖关系,并使新的记忆信息 M_t 在网络中沿锯齿状方向上下流动,可保留空间短期动态信息。时空 LSTM 网络结构如图4(b)所示,该网络在 LSTM 基础上加入时空记忆模块 M ,用于层间时空信息传递,并与原始记忆模块 C 共用同一个输出门,将两种记忆信息相结合,能够有效对时空序列数据中物体的运动状态进行建模。



(a) PredRNN算法结构图



(b) ST-LSTM网络结构

图4 PredRNN网络结构

时空 LSTM 的计算公式为

$$g_t = \tanh(W_{xg} \cdot X_t + W_{hg} \cdot H_{t-1}^l) \quad (12)$$

$$i_t = \sigma(W_{xi} \cdot X_t + W_{hi} \cdot H_{t-1}^l) \quad (13)$$

$$f_t = \sigma(W_{xf} \cdot X_t + W_{hf} \cdot H_{t-1}^l) \quad (14)$$

$$C_t^l = f_t \odot C_{t-1}^l + i_t \odot g_t \quad (15)$$

$$g_t' = \tanh(W'_{xg} \cdot X_t + W'_{mg} \cdot M_{t-1}^l) \quad (16)$$

$$i_t' = \sigma(W'_{xi} \cdot X_t + W'_{mi} \cdot M_{t-1}^l) \quad (17)$$

$$f_t' = \sigma(W'_{xf} \cdot X_t + W'_{mf} \cdot M_{t-1}^l) \quad (18)$$

$$M_t^l = f_t' \odot M_{t-1}^l + i_t' \odot g_t' \quad (19)$$

$$o_t = \sigma(W_{xo} \cdot X_t + W_{ho} \cdot H_{t-1}^l + W_{co} \cdot C_t^l + W_{mo} \cdot M_t^l) \quad (20)$$

$$H_t^l = o_t \odot \tanh(W_{1 \times 1} \cdot [C_t^l, M_t^l]) \quad (21)$$

式中: C_t^l 和 M_t^l 分别为原始时间记忆模块和时空记忆模块的输出; g_t, i_t, f_t 分别为原始时间记忆模块的调制门、输入门和遗忘门; g_t', i_t', f_t' 分别为时空记忆模块的调制门、输入门和遗忘门; o_t 为输出门; H_t^l 和 M_t^l 为细胞状态; \tanh 表示 \tanh 激活函数; σ 表示 sigmoid 激活函数; W 表示权重系数; “ \odot ” 表示矩阵乘法。

1.4 空间特征融合补偿模块

1.4.1 跨车辆特征聚合

CAV₂ 传输的特征经时间延迟补偿后生成预测特征,根据两车相对位置关系聚合对齐预测特征与 CAV₁ 获取的本地环境特征,生成聚合特征。具体过程如下:假设全局坐标系下 CAV₁ 和 CAV₂ 的中心坐标分别为 (x_1, y_1) 、 (x_2, y_2) ,将两辆车获取的本地环境特征均转换成尺寸为 (H, W, C) 的伪图像特征,如图5所示。

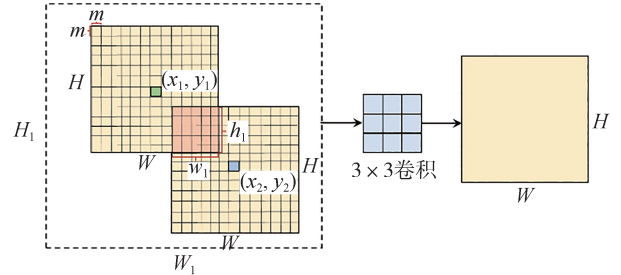


图5 跨车辆特征聚合

为满足不同时刻特征聚合一致性要求,划定高度 H_1 、宽度 W_1 的候选聚合区域 ($H_1 > H, W_1 > W$),若两车位置关系同时满足 $|x_1 - x_2|/m < W, |y_1 - y_2|/m < H$ 则判定两伪图像特征间有重叠区域,计算得到重叠区域的宽 w_1 和高 h_1 为

$$w_1 = W - \frac{|x_1 - x_2|}{m} \quad (22)$$

$$h_1 = H - \frac{|y_1 - y_2|}{m} \quad (23)$$

接着对重叠区域采用 Maxout 操作进行通道级别对齐聚合,即各通道选择较大值进行非线性融合,同时保留非重叠区域特征数据,采用零元素填充候选聚合区域中空白部分,生成尺寸为 (H_1, W_1, C) 的聚合

特征。为使输入和输出特征图的尺寸一致,采用 3×3 卷积将聚合特征的尺寸 (H_1, W_1, C) 压缩回 (H, W, C) 。

1.4.2 多分辨率特征提取

考虑到单一尺度特征难以同时捕获目标细节和全局语义信息,本文通过构建多分辨率特征提取网络全面感知不同尺寸的障碍物目标,网络结构如图6所示。输入的聚合特征先通过1个Conv2D(256, 128, 3, 1)卷积层调整通道数,再通过3个Conv2D

(128, 128, 3, 1)卷积层提取浅层特征 F_1 ,用于捕捉目标细粒度特征,提高小尺度目标感知精度;然后通过3轮提取操作依次得到深层特征 F_2, F_3, F_4 ,尺寸分别为原始特征 $1/2, 1/4, 1/8$,其中每轮操作由1个轻量级协调注意力模块(coordinate attention, CA)、1个步长为2的池化层、4个Conv2D(256, 256, 3, 1)卷积层构成。生成的多分辨率特征 F_1, F_2, F_3, F_4 涵盖不同尺度特征信息。

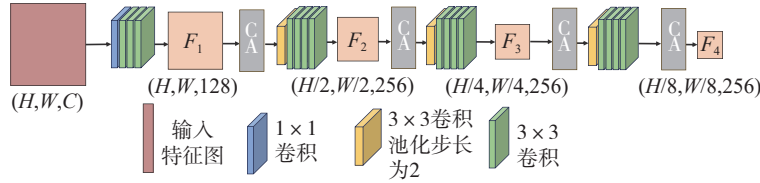


图6 多分辨率特征提取网络

本文所提CA模块能够同时考虑各通道间关系和空间位置信息,首先分别沿水平、垂直方向对输入特征进行池化,得到相应方向第 c 个通道的嵌入特征 F_c^h 和 F_c^w :

$$F_c^h(h) = \frac{1}{W} \sum_{0 \leq j < W} x_c(h, j) \quad (24)$$

$$F_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w) \quad (25)$$

式中: $F_c^h(h)$ 表示高度为 h 处第 c 个通道沿水平方向的特征集合; $F_c^w(w)$ 表示宽度为 w 处第 c 个通道沿垂直方向的特征集合; $x_c(h, j)$ 表示高度为 h 、宽度为 j 处第 c 个通道沿水平方向的输入特征; $x_c(j, w)$ 表示高度为 j 、宽度为 w 处第 c 个通道沿垂直方向的输入特征。然后将水平、垂直方向嵌入特征 F^h 和 F^w 进行拼接,生成横纵向特征全局感受野,通过 1×1 卷积降低通道数,再进行批量归一化处理和非线性激活,接着利用sigmoid激活函数计算得到水平、垂直方向的注意力权重向量集合。

1.4.3 多尺度特征融合

为有效结合浅层和深层特征优势,采用双向多尺度特征金字塔网络对不同分辨率特征进行融合,网络结构如图7所示。对深层特征 F_2, F_3, F_4 分别进行1个 1×1 卷积操作,得到通道数与 F_1 相同的特征 F'_2, F'_3, F'_4 。接着构建两层双向金字塔特征提取网络^[18]将特征向量 (F_1, F'_2, F'_3, F'_4) 转换成特征融合向量 (D_1, D_2, D_3, D_4) ,该网络通过横向连接捕获两层特征间的关联性,在浅层和深层特征之间直接传递语义信息时,每层网络自下而上结合高层语义特征和

低层位置信息进行下采样,同时通过跳跃连接使输入特征参与融合过程。为避免特征分辨率对检测精度造成影响,采用快速归一化卷积融合操作为不同分辨率特征增加权重,卷积融合计算如下:

$$O = \sum_i \frac{w_i}{\varepsilon + \sum_j w_j} \cdot I_i \quad (26)$$

式中: I_i 为输入特征; O 为加权特征融合后的输出值; w_i 和 w_j 为可学习权重; $\varepsilon=0.001$ 。

最后将提取到的多分辨率特征 (D_1, D_2, D_3, D_4) 输入至检测头进行分类和回归,输出车辆目标置信度、包围盒中心点坐标、几何尺寸以及航向角等参数信息。

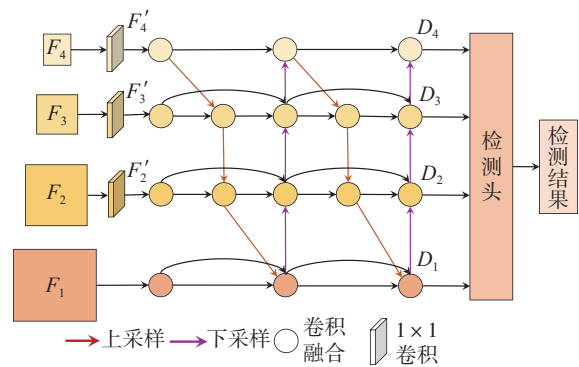


图7 双向多尺度特征金字塔网络

2 实验

2.1 环境搭建

本文硬件平台搭载 NVIDIA RTX 3080 显卡和

Intel I7 CPU,实验环境为Ubuntu 20.04系统,基于Pytorch搭建算法框架。所有模型的训练均采用Adam优化器。设置点特征提取参数 $h_{\min}=2.5\text{ m}$, $h_{\max}=2\text{ m}$, $R=50\text{ m}$, $\Delta\rho=0.1\text{ m}$, $\Delta z=0.2\text{ m}$, $\Delta\theta=0.5^\circ$;设置体柱特征提取参数 $x_{\min}=-40\text{ m}$, $x_{\max}=40\text{ m}$, $y_{\min}=-50\text{ m}$, $y_{\max}=50\text{ m}$ 。

2.2 数据集

V2V4real数据集^[19]是真实世界采集的用于协同感知的首个大规模数据集,由配置Velodyne VLP-32激光雷达的两辆车同时采集得到。本文选取该数据集进行模型训练和测试,其中训练集1.3万帧,测试集0.3万帧。并利用两辆均配备RS-Helios-32P激光雷达的车辆在北京市某园区内同步采集点云构建自制数据集(图8),包括直线道路、交叉口和弯道等场景,包含1.1万帧数据。

2.3 评价指标

(1)交并比(intersection over union, IoU):表示目标检测边界框与真实边界框的交、并集面积之比。通过边界框重叠程度量化单目标框检测准确性,IoU



图8 自制数据集采集平台

值越大表明重叠区域越大,将高于IoU阈值的目标检测框视为正确检测,否则为错误检测。

(2)平均精度(average precision, AP):表示精确率-召回率曲线下所围成的面积。本文通过计算不同IoU阈值下目标检测平均精度AP值来量化目标检测算法性能,AP值越接近于1,则目标检测效果越好。

2.4 检测结果

2.4.1 典型场景下算法性能测试

为验证所提算法对远距离、短暂或部分遮挡场景的适应性,将所提算法与单车感知效果进行对比测试,结果如图9所示。图中绿色矩形框表示目标

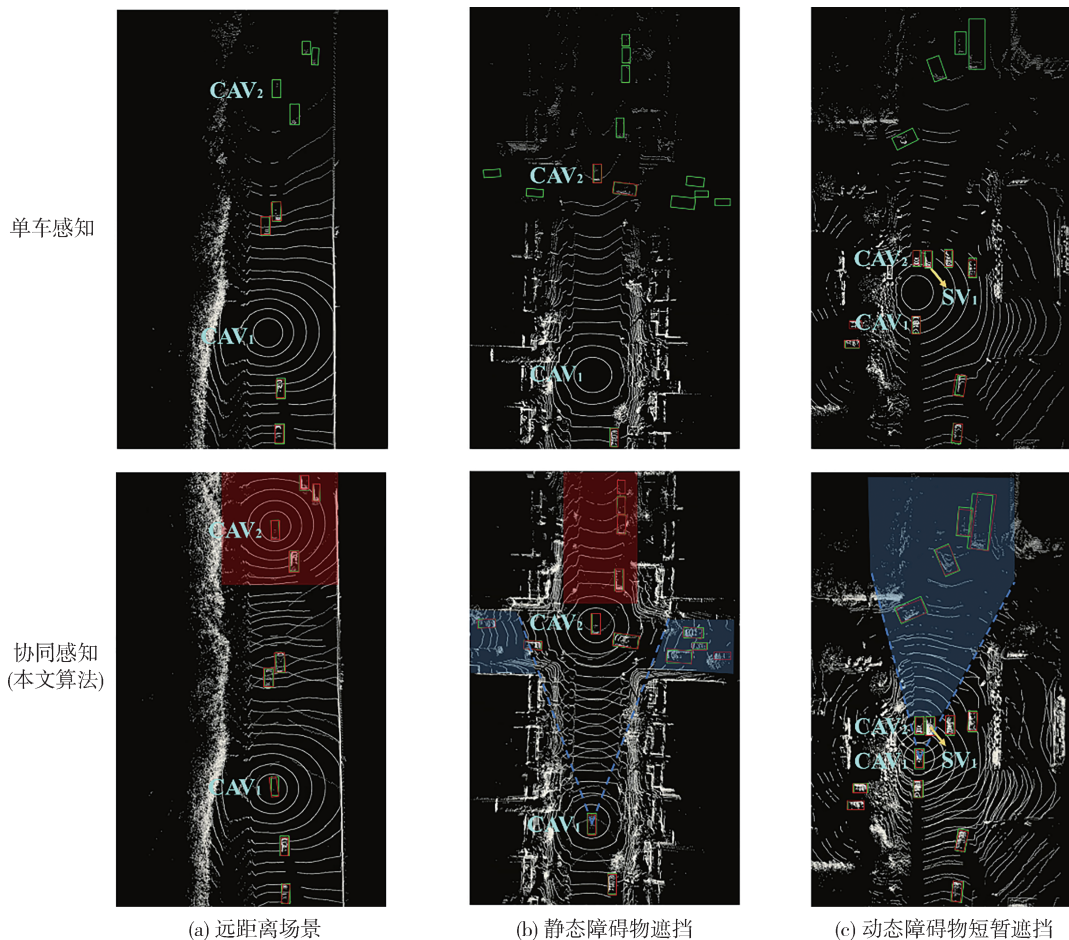


图9 典型场景算法性能测试

真实边界框,红色框表示目标检测边界框, SV_1 表示周围某障碍物车辆,蓝色、红色区域分别表示遮挡、探测距离受限引起的 CAV_1 点云检测盲区。可以看出,仅依靠 CAV_1 单车激光雷达无法检测远距离或遮挡区域内的车辆目标,所提算法通过协同共享 CAV_1 和 CAV_2 提取到的周围环境点云特征,融合补偿多车视角下的多分辨率空间特征信息,能够有效扩大检测区域范围,并较好地克服遮挡影响,准确检测到周围运动车辆目标。

2.4.2 无通信延迟场景对比实验

不考虑通信延迟影响的情况下,利用V2V4real和自制数据集对比已有融合算法与所提算法的协同感知性能,结果如表1所示。由表1可知,IoU阈值为0.5或0.7时,各协同感知算法的检测平均精度均明

显高于无融合(单车感知)算法,表明通过协同感知能更全面准确地感知周围车辆目标。相比于F-Cooper、V2VNet和CoBEVT中期融合算法,所提算法在 $AP@IoU=0.5$ 上升了1.7%~6.4%,在 $AP@IoU=0.7$ 上提升了1.8%~6.7%,这是因为所提点-体柱特征提取网络能够较大程度保留点云空间特征信息,有效提高检测精度,而另外3种算法利用体柱特征网络或3D卷积网络进行特征提取,都会损失点的位置信息;尽管所提时间延迟补偿模块中较多的卷积操作占用了部分计算资源,但本文算法的推理时间也仅略多于F-Cooper算法,仍满足车载环境实时性要求。V2XViT中期融合算法采用Transformer网络模型融合体柱特征,检测精度与所提算法较为接近,但该模型参数量大,实时性较差。

表1 不同融合算法性能对比

算法	融合方式	V2V4real		自制数据集		推理时间/ms
		$AP@IoU=0.5$	$AP@IoU=0.7$	$AP@IoU=0.5$	$AP@IoU=0.7$	
无融合 ^[18]	无融合	0.653	0.484	0.606	0.402	23
Pointpillar ^[16]	晚期融合	0.736	0.645	0.727	0.620	25
Cooper ^[9]	早期融合	0.823	0.731	0.812	0.725	2 263
F-Cooper ^[12]	中期融合	0.846	0.773	0.840	0.741	34
V2Vnet ^[13]	中期融合	0.842	0.801	0.845	0.809	58
CoBEVT ^[22]	中期融合	0.872	0.822	0.862	0.804	61
V2XViT ^[14]	中期融合	0.894	0.837	0.882	0.834	102
本文算法	中期融合	0.926	0.868	0.879	0.856	48

2.4.3 有通信延迟场景对比实验

为进一步验证有通信延迟场景下所提算法性能,定量分析网络延迟时间 T_d 对算法性能的影响,并与其它中期融合算法对比。表2为不同延迟时间场景下算法性能对比结果。可以看出:F-Cooper和CoBEVT算法检测精度较差,原因在于这两种算法均未考虑通信延迟的影响;V2VNet、V2XViT算法由于分别采用分段线性速度模型和时间信息编码策略降低通信延迟的影响,使得性能稍优于前两种算法;本文算法在3种延迟时间场景下性能表现均优于其它算法,表明所提算法能够有效地解决由通信延迟引起的数据不同步问题,检测结果更准确。为定性分析有通信延迟场景下所提算法的协同感知性能,利用V2V4real数据集进行测试,结果如图10所示。图中 SV_1 - SV_4 为周围运动车辆, SV_5 - SV_{10} 为静止车辆,绿色矩形框表示目标真实边界框,红色框表示目标检测边界框。可以看出:3种协同感知算法在不同延迟时间条件下均能准确地检测静止车辆;随着

通信延迟时间增大,V2VNet和V2XViT算法对运动车辆的位置检测误差逐渐增大,而所提算法的检测误差稳定保持在0.5 m范围内,原因在于所提算法利用PredRNN锯齿状时空记忆流传播结构和带记忆解耦的时空LSTM单元捕捉学习时空序列数据的短期动态信息和长期依赖关系,能够有效补偿通信延迟影响,算法鲁棒性更好。

2.4.4 消融实验

为验证点-体柱特征提取网络模块对所提算法协同感知整体性能的影响,对特征提取模块进行消融实验,将该模块分别替换为PointPillar^[16]、Second^[20]网络,选取V2V4real数据集进行性能评估,结果如表3所示。可知,相比于其他两种网络,本文通过点-体柱特征提取网络将平均检测精度分别提高了2.4%和1.4%,推理时间较为接近,这是由于本文通过结合扫描点局部细粒度信息与体柱全局结构信息有效提升算法精度,同时利用覆盖感知采样进行点特征提取降低运算量,使得推理时间满足实时

表2 不同延迟时间条件下算法性能对比

算法	V2V4real(AP@IoU=0.5/0.7)			自制数据集(AP@IoU=0.5/0.7)		
	100 ms	200 ms	300 ms	100 ms	200 ms	300 ms
F-Cooper ^[12]	0.697/0.413	0.524/0.346	0.305/0.173	0.587/0.406	0.484/0.323	0.302/0.143
CoBEVT ^[22]	0.725/0.443	0.538/0.352	0.446/0.223	0.622/0.423	0.509/0.368	0.422/0.195
V2VNet ^[13]	0.765/0.480	0.563/0.387	0.484/0.235	0.675/0.462	0.533/0.337	0.475/0.232
V2XViT ^[14]	0.776/0.503	0.601/0.437	0.491/0.368	0.671/0.484	0.584/0.389	0.483/0.306
本文算法	0.798/0.541	0.713/0.524	0.538/0.426	0.726/0.528	0.636/0.432	0.532/0.372

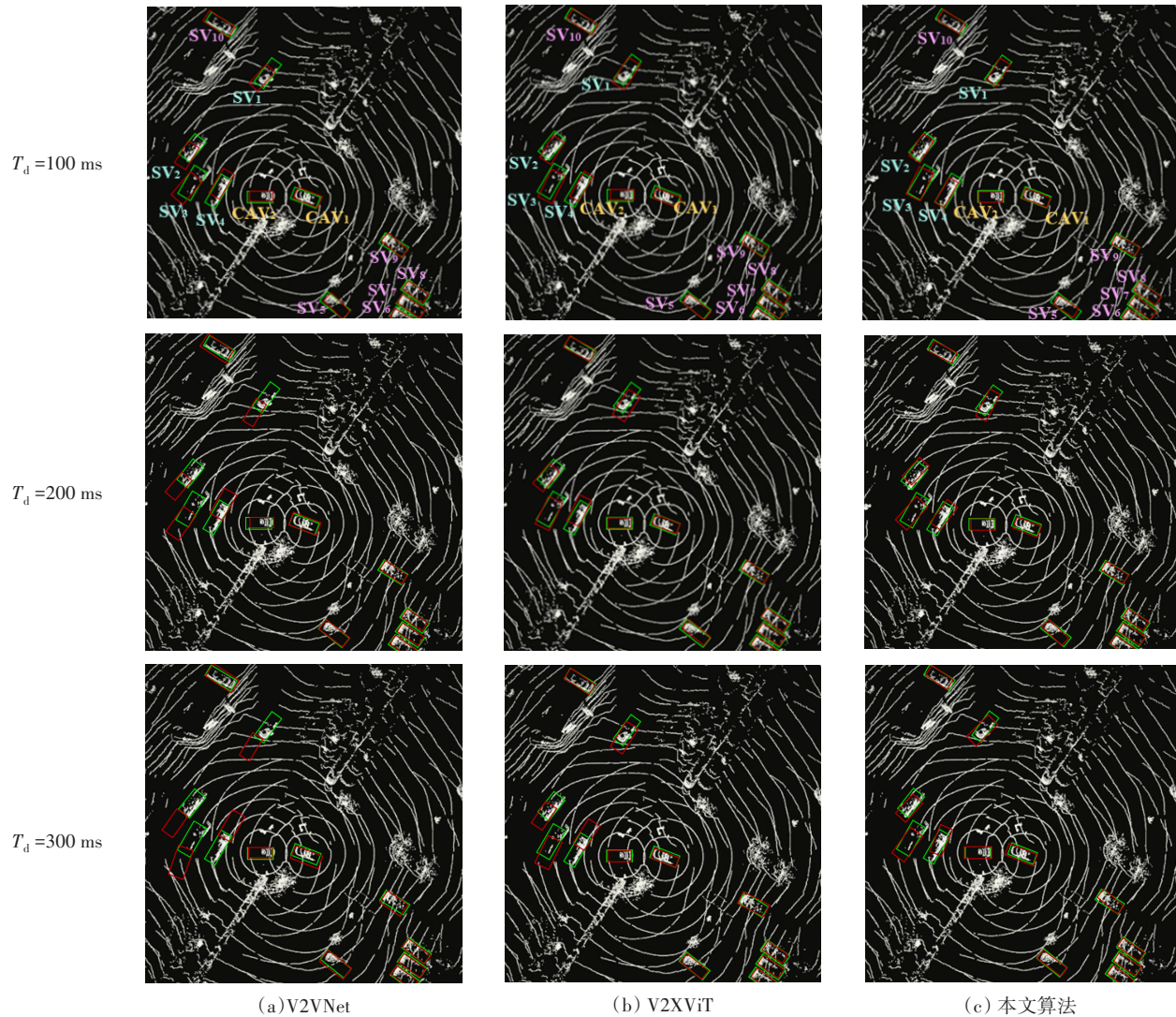


图10 不同网络延迟时间条件下各算法协同感知效果

性需求。

表3 不同特征提取网络消融研究

特征提取网络	AP@IoU=0.5	AP@IoU=0.7	推理时间/ms
PointPillar ^[16]	0.896	0.853	45
Second ^[20]	0.904	0.865	52
点-体柱	0.926	0.868	48

为进一步验证所提时间延迟补偿模块对协同感知算法整体性能的影响,对延迟补偿模块进行消融实验,将该模块替换为Transformer^[21]、ConvLSTM^[22],并与无延迟补偿模块进行对比,选取V2V4real数据集进行算法验证,测试网络延迟时间 T_d 分别为100、200和300ms时的各算法性能,结果如表4所示。由

表4可知,无延迟补偿模块时,算法平均检测精度随延迟时间增大出现明显下降,相比于延迟100 ms,延迟300 ms时AP@IoU=0.5和AP@IoU=0.7分别降低了39%和23.1%,而加入PredRNN延迟补偿模块后所提算法平均精度分别仅降低了24.9%和10.9%;Transformer模型作为延迟补偿模块时算法在不同延迟条件下的检测精度均略高于所提算法,原因是Transformer采用自注意力机制捕捉输入历史帧和当前帧的时空依赖关系,能够更为准确地预

测未来特征数据,但该模型参数数量过大且占用过多计算资源,导致实时性差;ConvLSTM作为延迟补偿模块时算法在不同延迟条件下的检测平均精度比所提算法低4.8%~10.3%,这是因为ConvLSTM网络训练过程中顶层单元会忽略底层记忆状态信息,导致算法检测精度较低。因此,本文采用PredRNN算法作为时间延迟补偿模块,既能有效减小因网络延迟引起的检测误差,能满足协同感知实时性要求。

表4 不同时间延迟补偿模块消融研究

延迟补偿模块	100 ms		200 ms		300 ms		推理时间/ms
	AP@IoU=0.5	AP@IoU=0.7	AP@IoU=0.5	AP@IoU=0.7	AP@IoU=0.5	AP@IoU=0.7	
无	0.707	0.423	0.535	0.348	0.317	0.192	36
Transformer ^[21]	0.804	0.563	0.685	0.541	0.549	0.431	215
ConvLSTM ^[22]	0.732	0.445	0.623	0.484	0.493	0.329	58
本文算法	0.798	0.541	0.713	0.524	0.538	0.426	48

为探究空间特征融合模块对所提算法的影响,对该模块进行消融实验,将其替换为FuseBEVT^[23]和FPN^[24]网络,测试结果如表5所示。可以看出,相比于注意力机制融合网络FuseBEVT和金字塔网络FPN,所提算法的平均检测精度分别提升了5.2%和3.9%,说明本文通过引入双向多尺度特征融合网络能够有效聚合多分辨率空间特征信息,增强了网络对点云局部细粒度信息与全局上下文信息的理解能力。

表5 不同特征融合网络消融研究

融合网络	AP@IoU=0.5	AP@IoU=0.7	推理时间/ms
FuseBEVT ^[23]	0.902	0.815	50
FPN ^[24]	0.883	0.807	43
本文算法	0.926	0.868	48

3 结论

提出一种基于点云时空特征补偿网络的智能网联车辆协同感知算法。首先采用点-体柱特征提取网络对点云数据进行特征化处理,然后通过基于PredRNN算法的时间特征延迟补偿模块预测车辆未来时刻特征状态;最后基于空间特征融合补偿模块融合不同分辨率的点云特征。实验结果表明:所提点-体柱特征提取方法和空间特征融合补偿模块能够更充分提取点云信息,提高远距离、短暂遮挡场景下的目标检测性能;时间延迟特征补偿模块能够有

效缓解通信延迟带来的滞后效应。

参考文献

- [1] 金祖亮, 隗寒冰, LIU Z, 等. 基于局部窗口交叉注意力的轻量级语义分割[J]. 汽车工程, 2023, 45(9): 1617-1625.
JIN Z L, WEI H B, LIU Z, et al. Lightweight semantic segmentation method based on local window cross attention[J]. Automotive Engineering, 2023, 45(9): 1617-1625.
- [2] 李勇滔, 孙晨旭, 郑伟光, 等. 基于毫米波雷达与视觉融合的碰撞预警[J]. 汽车工程, 2023, 45(9): 1666-1676.
LI Y T, SUN C X, ZHENG W G, et al. Collision warning based on fusion of millimeter wave radar and vision[J]. Automotive Engineering, 2023, 45(9): 1666-1676.
- [3] 金字锋, 陶重彝. 基于Transformer的融合信息增强3D目标检测算法[J]. 仪器仪表学报, 2023, 44(12): 297-306.
JIN Y F, TAO C B. Fusion information enhanced method based on transformer for 3D object detection[J]. Chinese Journal of Scientific Instrument, 2023, 44(12): 297-306.
- [4] 张名芳, 吴禹峰, 王力, 等. 基于金字塔特征融合的二阶段三维点云车辆检测[J]. 交通运输系统工程与信息, 2022, 22(5): 107-116.
ZHANG M F, WU Y F, WANG L, et al. Pyramid-feature-fusion-based two-stage vehicle detection via 3D point cloud[J]. Journal of Transportation Systems Engineering and Information Technology, 2022, 22(5): 107-116.
- [5] AN P, DUAN Y C, HUANG Y L, et al. SP-Det: leveraging saliency prediction for voxel-based 3D object detection in sparse point cloud[J]. IEEE Transactions on Multimedia, 2023, 26: 2795-2808.
- [6] 张瑞士, 魏中华. 车路协同交叉口应用综述[J]. 交通工程, 2022, 22(2): 1-6.

- ZHANG R S, WEI Z H. A review of the application of CVIS in intersection [J]. *Journal of Transportation Engineering*, 2022, 22(2): 1–6.
- [7] MENG Z L, XIA X, XU R S, et al. HYDRO-3D: hybrid object detection and tracking for cooperative perception using 3D LiDAR [J]. *IEEE Transactions on Intelligent Vehicles*, 2023, 8(8): 4069–4080.
- [8] ARNOLD E, DIANATI M, TEMPLE R D, et al. Cooperative perception for 3D object detection in driving scenarios using infrastructure sensors [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(3): 1852–1864.
- [9] CHEN Q, TANG S H, YANG Q, et al. Cooper: cooperative perception for connected autonomous vehicles based on 3D point clouds [C]. 2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS). IEEE, 2019: 514–524.
- [10] HURL B, COHEN R, CZARNECKI K, et al. Trupercept: trust modelling for autonomous vehicle cooperative perception from synthetic data [C]. 2020 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2020: 341–347.
- [11] YU H B, LUO Y Z, SHU Y Y, et al. Dair-V2X: a large-scale dataset for vehicle–infrastructure cooperative 3D object detection [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 21361–21370.
- [12] CHEN Q, MA X, TANG S H, et al. F-cooper: feature based cooperative perception for autonomous vehicle edge computing system using 3D point clouds [C]. Proceedings of the 4th ACM/IEEE Symposium on Edge Computing, 2019: 88–100.
- [13] WANG T H, MANIVASAGAM S, LIANG M, et al. V2VNet: vehicle-to-vehicle communication for joint perception and prediction [C]. *Computer Vision–ECCV 2020: 16th European Conference*, 2020: 605–621.
- [14] XU R S, XIANG H, TU Z Z, et al. V2X-ViT: vehicle-to-everything cooperative perception with vision transformer [C]. *European Conference on Computer Vision*, 2022: 107–124.
- [15] 王秉路, 靳杨, 张磊, 等. 基于多传感器融合的协同感知方法 [J]. *雷达学报*, 2024, 13(1): 87–96.
- WANG B L, JIN Y, ZHANG L, et al. Collaborative perception method based on multisensor fusion [J]. *Journal of Radars*, 2024, 13(1): 87–96.
- [16] LANG A H, VORA S, CAESAR H, et al. Pointpillars: fast encoders for object detection from point clouds [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 12697–12705.
- [17] WANG Y, WU H, ZHANG J, et al. PredRNN: a recurrent neural network for spatiotemporal predictive learning [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 45(2): 2208–2225.
- [18] TAN M, PANG R, LE Q V. EfficientDet: scalable and efficient object detection [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10781–10790.
- [19] XU R, XIA X, LI J, et al. V2V4real: a real-world large-scale dataset for vehicle-to-vehicle cooperative perception [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 13712–13722.
- [20] YAN Y, MAO Y, LI B. Second: sparsely embedded convolutional detection [J]. *Sensors*, 2018, 18(10): 3337.
- [21] YE X, BILODEAU G A. VPTR: efficient transformers for video prediction [C]. 2022 26th International Conference on Pattern Recognition (ICPR). IEEE, 2022: 3492–3499.
- [22] LIN Z H, LI M M, ZHENG Z B, et al. Self-attention convlstm for spatiotemporal prediction [C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020: 11531–11538.
- [23] XU R S, TU Z D, XIANG H, et al. CoBEVT: cooperative bird's eye view semantic segmentation with sparse transformers [J]. arXiv preprint arXiv:2207.02202, 2022.
- [24] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2017: 2117–2125.