面向多智能体协同感知的基于注意力的特征融合

Ahmed N. Ahmed*、Siegfried Mercelis*、Ali Anwar*
*电子与信息通信技术工程系，IDLab - 安特卫普大学 - imec，比利时安特卫普
ahmed.ahmed@uantwerpen.be

arXiv:2305.02061v1 [cs.MA] 3 May 2023

摘要—**在智能交通系统（ITS）领域，协同感知已成为一种颇具前景的方法：通过使多个智能体交换信息，克服单体感知的局限，从而增强其态势感知能力。协同感知克服了单体传感器的局限，使互联智能体能够感知超出其视线范围和视场的环境。然而，协同感知的可靠性在很大程度上取决于数据聚合策略和通信带宽，而这两者必须应对网络资源有限所带来的挑战。为提高目标检测精度并缓解网络资源受限的问题，我们提出一种采用图注意力网络（GAT）的中期协同感知方案。该方法设计了一种基于注意力的聚合策略，用于融合多个互联智能体之间交换的中间表示。该方法能够在通道和空间层级上自适应地突出中间特征图中的重要区域，从而提高目标检测精度。我们提出一种采用基于注意力架构的特征融合方案，并通过与其他最先进的协同感知方法进行比较，对结果进行定量评估。我们使用V2XSim数据集 [1]验证了所提出的方法。研究结果表明，该方法能够有效实现中期协同感知，在提高目标检测平均精度的同时减少网络资源使用。**

索引词—**协同感知，特征图，注意力，聚合**

一、引言

近年来，深度学习的进步显著提升了智能交通系统（ITS）的功能和性能。这些进步有望提高交通效率、减少交通事故并节约资源 [2]。感知是智能交通系统中的一个基础模块，可使智能体识别、解读并表征感官输入，从而实现态势感知。自动驾驶是智能交通系统中研究广泛的主题之一，主要依赖单体感知，即车辆通过自身传感器感知周围环境。然而，此类系统的可靠性取决于车载感知单元的稳健性[2]–[4]。尽管单体感知已取得显著进展 [2], [4], [5], ，开发可靠的感知系统仍然是一项挑战。这主要是由传感器特性、障碍物遮挡、光照和恶劣天气条件等问题造成的。此外，车辆的感知范围通常受限于其传感器的探测范围，导致视场中出现盲区，难以提供全面的感知能力。自动驾驶车辆的信息。因此，自动驾驶车辆未能及时发现迫在眉睫的危险。

单车感知的局限性促使车联网（V2X）通信技术应运而生，成为一种颇具前景的解决方案 [6]。车联网通信促进车辆与其他实体之间的通信和数据共享，包括车车通信（V2V）和车路通信（V2I）。研究人员在车联网领域开发了协同感知技术，通过引入额外视角并融合从多个互联智能体接收的数据，将单车感知升级为协同感知，从而提升自动驾驶车辆的感知能力。这使车辆能够看得更远、更清晰，甚至穿透遮挡 [6]–[8]。

一般而言，根据参与者之间交换的信息类型，协同感知技术可分为三类：早期融合、中间融合和后期融合。早期融合是将各智能体的原始传感器数据融合为一个整体视角 [9]–[11]。该方法能够全面、细致地呈现环境，但需要大量通信带宽。另一方面，后期融合是在处理原始传感器数据后，融合各智能体的感知输出 [12], [13]。这种方法的带宽效率较高，但融合结果可能存在噪声且不完整。中间融合则是提取各智能体上每种传感器模态的数据特征，并将其聚合为单一表示 [7], [8], [14]–[16]。与早期融合相比，该方法的带宽效率更高，并有望兼顾通信带宽效率与感知能力的提升。然而，设计不佳的中期协同策略可能导致特征抽象和融合过程中的信息损失，从而限制感知能力的提升。

本文提出了一种新颖方法，用于为互联智能体之间的多智能体协同感知设计有效的中期协同策略。具体而言，我们提出了一种基于图注意力特征融合的神经网络，其中图中的每个节点表示从各个智能体接收到的中间表示，每条边则反映两个智能体之间的成对协作。此外，我们引入通道注意力机制和空间注意力机制，引导模型识别相关特征，从而增强特征图融合。

我们的主要贡献在于开发了这一新模型，它有望通过利用协同感知显著提升自动驾驶车辆的感知能力。本研究的贡献可概括如下：

- 我们提出了一种用于目标检测的中间层协同感知方法，该方法利用来自邻近车辆和基础设施的信息。
- 我们开发了一种聚合方法，将中间特征结合起来，并采用空间与通道注意力模块来学习聚合特征之间的关系。该方法通过注意力模块引导网络关注“什么”和“哪里”，从而增强了网络的表征能力。
- 我们使用 V2X-Sim 2.0 [1], 这一大规模车车通信和车联网数据集验证了所提出的方法。该数据集包含来自车辆和基础设施的激光雷达数据。实验结果表明，所提出的方法在目标检测任务中有效，性能优于现有方法。

本文其余部分的结构如下。第二节总结并分析协同感知融合方法。第三节介绍所提出的方法，包括其主要组成部分和特征融合策略。第四节介绍实验设置、评估指标和基线模型。第五节简要讨论数据集和评估指标，并将所提出框架的结果与基线方法进行比较。最后，第六节总结本文。

II. 相关工作

在本节中，我们全面回顾协同感知方法，并系统总结感知网络中的协作模块。

早期协同感知-早期协作方法通过根据智能体之间共享的位姿信息，将所有激光雷达点云投影到自车的坐标系中，从而在网络的输入层级融合原始传感器数据。随后，自车汇总所有接收到的点云，并将其输入检测器以实现协作感知。Cooper [9] 提出使用共享的原始点云数据进行协作感知，并引用了一种用于低密度点云数据目标检测的神经网络。 [10]，中描述了另一种方法，该方法涉及一个新的空间变换方程，利用拼接来融合传感器数据。在 Cooper 的研究基础上， [11] 采用原始点云融合方法，其中每辆车都对原始点云进行压缩以减小其大小，然后共享一条协作感知消息（CPM），其中包含压缩后的点云以及环境中的位置信息。在接收到来自邻近智能体的传感器数据后，点云会被变换到接收车辆的坐标系中，并与自车数据及接收到的数据进行拼接。随后，将生成的聚合点云输入感知网络进行处理。

中期协同感知 -中期协同方法是一种融合技术，它结合每个智能体的预测模型生成的中间特征，从而允许传输并随后融合这些特征，以产生感知结果。为此，Chen 等人 [14] 提出了基于激光雷达的 F-Cooper 方法，用于协同目标检测。该方法比较了体素特征融合（VFF）和空间特征融合（SFF）技术，并研究了共享部分特征图通道的影响。尽管后续研究已将注意力机制这一流行的深度学习技术应用于特征融合，但 Who2com [16]利用注意力机制，在选择与之通信的车辆的同时，尽可能减少带宽使用。相比之下，When2com [15] 利用注意力机制确定通信情境，并使用基于拼接的规则融合多辆车辆的特征图。DiscoNet[8]利用矩阵值注意力分数突出显示包含重要信息的区域，以进行特征图融合，并采用知识蒸馏（KD）训练模型，使其接近输入共享的特征图。V2VNet[7] 使用图神经网络（GNN）表示多辆车辆的拓扑结构，并通过相邻节点之间的消息传递执行融合，其中两个卷积神经网络会在特征图融合前纠正误差。具体而言，一个卷积神经网络补偿时间延迟，另一个则学习变换过程，以隐式纠正定位误差。

后期协同感知 -后期协同是指在各智能体分别处理自身观测后，通常在输出空间中合并感知结果的概念。Miller 等人 [12] 利用这一技术为自动驾驶车辆设计了感知与定位系统，重点解决通信链路中断和延迟相关的问题。Allig 等人 [13]研究了共享的检测到的目标的时空对齐，并建议使用非预测发送方状态进行变换，从而无需进行发送方运动补偿。尽管后期协同具有更高的通信带宽效率，但它容易受到定位误差的影响；由于本地观测不完整，这可能导致较大的估计误差和噪声。

总之，通过比较三种协作策略，我们发现，就通信带宽使用量而言，中期协同效率更高，并且能够实现高精度的感知结果，优于后期协同策略[6]。然而，在实践中实现中期协同感知面临着重大的算法挑战，例如，如何选择最有价值且精简的特征以供智能体之间传输，以及如何融合接收特征以提升感知能力。

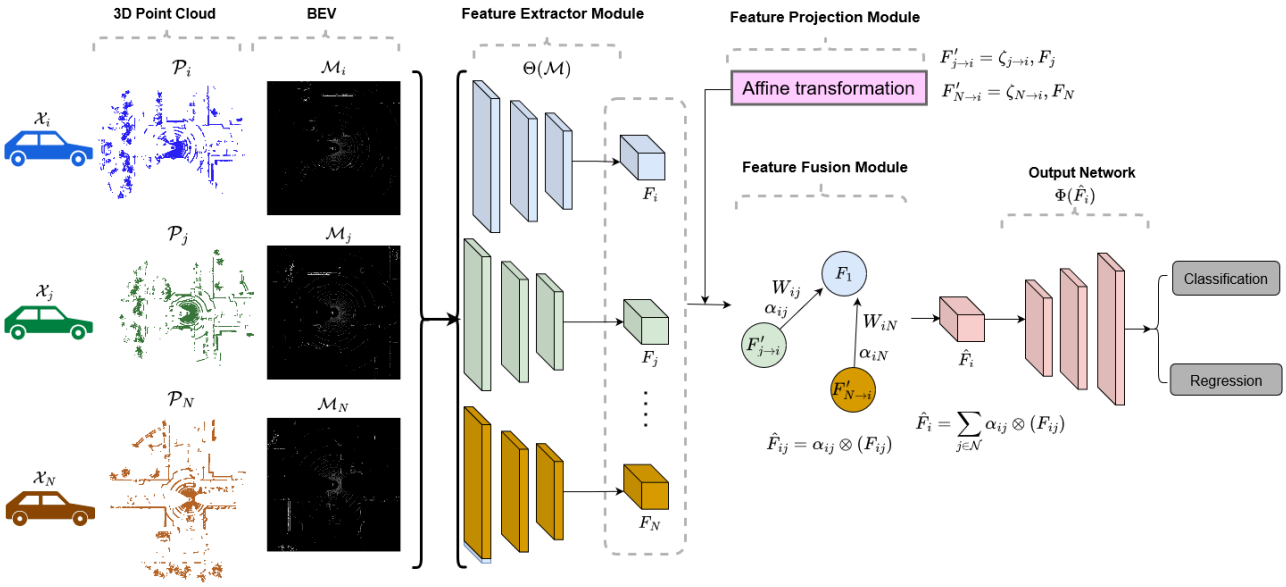


图 1. 所提出方法的整体架构始于每个智能体 $\mathcal{X}$ 将其点云 $\mathcal{P}$ 转换为鸟瞰图（BEV）地图 $\mathcal{M}$。共享特征提取器 $\mathcal{E}$ 处理 BEV 地图以获取特征图 F。随后，在与自车智能体的特征图聚合 $\mathcal{A}$ 之前，将接收到的特征图转换到自车智能体的坐标系中，并利用通道注意力机制和空间注意力机制生成更新后的表示 $\hat{} F_i$。然后将更新后的表示输入输出网络 $\mathcal{D}$ 以执行目标检测。

要应对这些挑战，需要开发高效且有效的特征选择与融合方法，以优化协同感知过程中的性能。

III. 方法论

本文提出了一种用于多模态特征融合的协同融合架构，该架构利用注意力机制有选择地组合从多个智能体获得的中间表示。所提出的方法包括三个阶段：特征提取、基于注意力的特征融合和输出预测网络。在特征提取阶段，使用点云数据提取相关特征。随后，将生成的特征图传递到基于注意力的特征聚合阶段，该阶段借助注意力机制，从不同智能体获得的各个中间表示中有选择地融合信息量最高的特征。最后，在输出预测阶段，利用融合后的特征预测用于目标检测的边界框。

A. 特征提取

每个智能体上部署的激光雷达传感器首先采集三维点云 $\mathcal{X}\{i, j..N\}$，其中 $\{i, j, .., N\}$ 是已连接的智能体。随后，每个智能体根据原始点云生成本地二值鸟瞰图（BEV）地图 $\mathcal{M} \in \mathbb{R}^{W_m \times H_m}$，其中 $W_{\mathcal{M}}$ 和 $H_{\mathcal{M}}$ 分别表示本地 BEV 地图的宽度和高度，如图 1 所示。

随后，第 i 个智能体 $\mathcal{M}_i$ 的 BEV 会经过特征提取器，以提取信息丰富的特征 BEVprovided by the agent，并获得特征图 $F_i \leftarrow \Theta(\mathcal{M}_i)$，其中 $\Theta(\cdot)$ 为特征提取器，以及 $F_i \in \mathbb{R}^{W \times H \times C}$，其中 W, H 和 C 是特征维度。特征提取器的目标是转换输入数据并提取可供处理的特征，同时保留原始数据中的信息。所提出的方法考虑采用同质中期协同感知进行目标检测，其中多个智能体共享相同的模型组件，因此，所有智能体都共享相同的 $\Theta(\cdot)$ 设计。此外，所提出方法的主要目标是改进特征图融合与聚合策略，并将我们的聚合方法与最先进水平的模型进行比较。因此，我们使用 DicoNet $\Theta_{DiscoNet}(\cdot)$ 和 V2VNet $\Theta_{V2VNet}(\cdot)$ 的特征提取器对聚合方法进行基准测试，以便独立考察聚合策略的性能，不受特征提取器影响。

B. 特征投影

由于各智能体的地理位置和朝向各不相同，因此需要将生成的特征图变换到接收方的坐标系中进行特征融合。仿射变换能够保持共线性和距离比例，可用于此目的 [17]。仿射变换是一类特殊的射影变换，不会将任何对象从仿射空间移至无穷远平面，也不会反向移动。为此，每个智能体都会传输其特征图 F以及位姿信息 ζ，其中位姿信息表示该智能体的位置和朝向，形式为 (X, Y, Z, roll, yaw, pitch)。第 i 个智能体接收到来自第 j 个智能体的 F_j 和 ζ 后，会依据图 1 所示接收方和发送方智能体的自车位姿 ζ_i 和 ζ_j，通过变换 $\zeta_{j\to i}$ 将 F_j 投影到其坐标系中。智能体的变换后特征图

j, $F'_{j\to i}$ 随后表示在与从第 i 个智能体获取的特征图 F_i 相同的坐标系中。将从相邻智能体接收的所有特征图变换到第 i 个智能体的坐标系后，自车特征图和变换后的特征图即可进入聚合阶段进行处理，从而得到整合所有智能体信息的最终表示。

C. 特征融合

先前的研究已广泛探讨注意力在特征融合中的重要性，相关内容见[18]。注意力不仅能够引导关注，还能增强兴趣对象的表征能力。在我们提出的特征融合策略中，受具有代表性的基于注意力的特征融合模块 SENet [19], MS-CAM [20] 和 CBAM [21],启发，我们旨在利用通道注意力和空间注意力，发挥各自的优势，以增强表征能力，并融合多个互联智能体的中间表示。注意力机制会聚焦于对融合至关重要的特征，抑制不必要的特征，并自适应地选择关键区域。特征融合的第一步是聚合感兴趣的特征图，即自车的特征图 F_i 和邻近智能体经过变换的特征图 $F'_{j\to i}$，如下所示：

$$F_{ij} = F_i \uplus F'_{j\to i} \tag{1}$$

其中， $\uplus$ 表示聚合过程，可以采用拼接或加法，如表 I 所示，本文稍后将对此进行详细说明。

通道注意力 为生成通道注意力图，我们利用特征的通道间关系，其中，特征图的每个通道都充当一个特征检测器 [22]。我们采用通道注意力，因为它着重确定输入聚合特征图中“什么”具有重要意义。给定一个中间特征 F_{ij}，计算通道注意力时，使用全局平均池化（GAP）压缩聚合特征图的空间维度。随后，通道注意力操作会重新校准各通道的权重，以确定哪些内容需要关注。为此，我们使用编码器-解码器逐点卷积（PwConv） [23]生成通道注意力图 α_{ch}。该操作对输入特征图应用 1×1 卷积滤波器，从而调整通道特征，同时保持特征图的空间维度不变。通道注意力权重$\alpha_{ch}\in\mathbb{R}^C$ 计算如下：

$$\alpha_{ch} = \delta(\mathbf{L_{ch}}(\delta\mathbf{G_{ch}}(GAP(F_{ij})))) \tag{2}$$

其中， $GAP(F_{ij}) \in \mathbb{R}^C$ 表示聚合特征图 F_{ij}的全局平均池化。 δ 表示修正线性单元（ReLU） [24]。通道注意力特征图由一系列 PWConv 构成的自编码器计算得出，其中， G_{ch} 是基于 PWConv 的编码器，表示降维层； L_{ch} 是基于 PWConv 的解码器，表示升维层； r 是通道缩减比，如图 2 所示。我们可以可以看到，通道注意力将每个大小为 $H\times W$ 的特征图压缩为一个标量 $\alpha_{ch}\in R$。

空间注意力 除了通道注意力外，我们的特征融合模型还引入了空间注意力，它充当自适应的空间区域选择机制，引导模型决定应将注意力集中在“哪里”。我们利用特征的空间间关系生成空间注意力图。我们采用基于自编码器的 PwConv 操作（如图 2 所示），通过结合跨通道信息和空间信息来提取有用特征，并生成空间注意力图 α_{sp}。由于 PwConv 每次仅使用核 1×1 对单个空间位置进行操作，因此即使对于非常大的特征图，也能高效地完成计算。这使其非常适用于空间注意力机制，因为在这类机制中，需要为特征图中的每个空间位置计算注意力权重。由此，聚合模块能够沿着两个主要维度——通道轴和空间轴——突出重要特征。

在聚合后的特征图 F_{ij} 上，我们应用基于自编码器的 PWConv（如图 2 所示），以生成空间注意力图 $\alpha_{sp}\in\mathbb{R}^{C\times H\times W}$，其计算方式如下：

$$\alpha_{sp} = \beta(\delta(\mathbf{L_{sp}}(\beta(\delta(\mathbf{G_{sp}}(F_{ij})))))) \tag{3}$$

其中， β 表示批归一化（BN）， [25], ， $\uplus$表示拼接或求和操作。

我们分别引入通道注意力模块和空间注意力模块，如图 2 所示，使每个分支能够分别学习在通道轴和空间轴上“关注什么”以及“关注哪里”，随后将注意力图进行组合，以充分利用这些学习结果。随后，将分别计算得到的通道注意力特征图 α_{ch} 和空间注意力特征图 α_{sp} 相加，并通过 sigmoid 函数 σ ，生成 $\alpha(F_{ij})\in\mathbb{R}^{C\times H\times W}$，表示总注意力权重：

$$\alpha_{ij} = \sigma(\alpha_{ch}\oplus\alpha_{sp}) \tag{4a}$$

$$\hat{F}_{ij} = \alpha_{ij}\otimes(F_{ij}) \tag{4b}$$

对于接收到的每个特征图，都会重复这一过程，并计算第 i 个智能体 $F_i\in\mathbb{R}^{C\times H\times W}$ 更新后的融合特征图，计算公式如下：

$$\hat{F}_i = \sum_{j\in\mathcal{N}}\alpha_{ij}\otimes(F_{ij}) \tag{5}$$

D. 输出网络

协作和特征图聚合后，每个智能体都会将更新后的特征图传递给输出网络，以生成最终检测结果。解码后的特征图为 $\hat{}y$, $\hat{}x\leftarrow\Phi(\hat{}F_i)$， $\Phi(\cdot)$由卷积层组成，用于对前景和背景类别进行分类并回归边界框。根据第 III-A 节讨论的推理过程，我们采用与 DiscoNet $\Phi_{DiscoNet}$ [8] 和 V2VNet Φ_{V2VNet} [7] 相同的输出网络来生成最终检测结果。

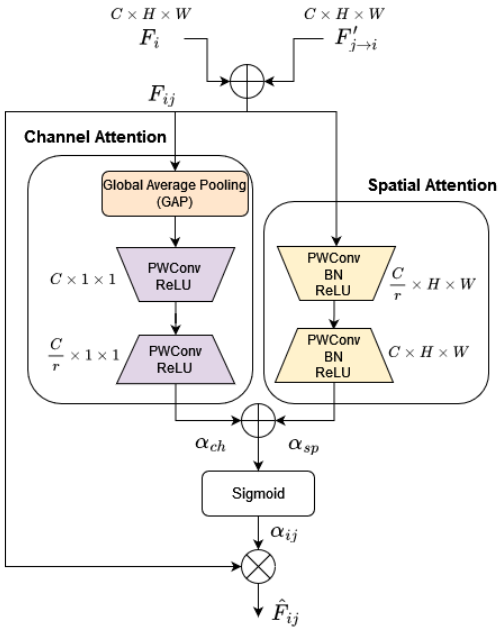


图2. 所提出的通道-空间注意力特征融合示意图。

IV. 实现细节

数据集 V2X-Sim 2.0 [1]是一个基于车联网仿真的数据集，由 SUMO [26]和 Carla [27]集成构建。该仿真利用 SUMO 生成逼真的交通流数据，并使用 Carla 获取位于同一地理区域内的多辆车辆的传感器数据流，例如激光雷达数据。该数据集包含 10,000 帧，分布在 100 个不同的场景中，每个场景由 100 帧组成。数据集划分为训练集、验证集和测试集，分别包含 8,000/1,000/1,000 帧。每帧均由多个智能体（包括车辆和路侧单元（RSU））采集的数据构成。该数据集包含 37,200 个训练样本、5,000 个验证样本和 5,000 个测试样本。我们的实验评估在两种场景下进行：不使用（w/o）和使用（w/）RSU，如表 II 所示。在这项工作中，我们仅关注车辆检测，并报告测试集上的结果。

评估 我们将模型总损失函数定义为分类损失 L_{cls} 与回归损失 L_{reg} 的加权和，如下所示：

$$L_{total} = \rho L_{cls} + \psi L_{reg} \tag{6}$$

其中， ρ 和 ψ 是定义分类损失和回归损失权重的超参数。对于分类损失 L_{cls} ，我们采用二元交叉熵损失来监督前景-背景分类：

$$L_{cls}(y, \hat{y}) = -(ylog(\hat{y}) + (1-y)log(1-\hat{y})) \tag{7}$$

其中， y 是输入的真实标签（0 或 1），$\hat{y}$ 是输入属于类别 1 的预测概率。对于回归损失 L_{reg}，使用 $smooth_{L_1}$ 损失来监督边界框回归：

$$smooth_{L_1}(x) = \begin{cases} 0.5(\hat{x}-x)^2 & \text{if } |\hat{x}-x| < 1 \\ |\hat{x}-x| - 0.5 & \text{otherwise,} \end{cases} \tag{8}$$

其中， x 是边界框单个坐标的预测值与真实值之间的差异。为了计算整个边界框的平滑L1损失，通常会分别计算每个坐标（例如，x坐标、y坐标、宽度和高度）的损失，然后将它们相加。对于边界框回归，平滑L1损失是一个不错的选择，因为与L2损失等其他损失函数相比，它对异常值不那么敏感，因此能够产生更稳定、更准确的预测。

表I 我们提出的方法的架构。其中，“聚合操作”表示式1中的 $\uplus$ 特征。“层”表示特征图维度缩减（另见图2、式2和式3）

模型基座	实验 Name	聚合操作 ($\uplus$)	层氏
DiscoNet	FF_D Sum64	Sum	256、128、64
	FF_D Sum32	Sum	256、128、64、32
	FF_D–拼接64	拼接	512, 256, 128, 64
	FF_D–拼接32	拼接	512、256、128、64、32。
V2VNet	FF_V–拼接64	拼接	512、256、128、64
	FF_V–拼接32	拼接	512、256、128、64、32。

基线模型 我们的目标是评估所提出的协同感知策略相对于现有基线协同感知模型的有效性，这些模型分为无协同、早期协作、中间协作和后期协同。对于无协同模型，我们采用下界模型 [1], ，该模型是一种单智能体感知模型，在无协作的情况下处理单视角点云。对于早期协作模型，我们采用上界模型 [1], ，该模型融合了所有互联智能体的原始点云数据。中间协作模型包括 DiscoNet [8],、V2VNet [7],、F-Cooper [14], 、When2com [15], 和 Who2com [16],，这些模型均基于神经网络中间特征的通信。对于后期协同模型，我们考虑余下界模型 [1],，该模型融合了不同单智能体感知模型的输出预测。为了与基线模型进行比较，我们选取交并比（IoU）阈值 0.5 和 0.7 来计算平均精度阈值。

实现细节如 III-A 和 III-D 节所述，我们采用了与前文 III-A 节讨论的 DicoNet [8] 和车车通信 [7]相同的特征提取器和输出网络。此外，所有方法均采用相同的设置进行训练，以确保性能提升源于我们提出的协作方式。

表 II V2X-Sim 2.0 上 BEV 检测的定量结果。AP 表示在交并比为 0.5 和 0.7 时，有无 RSU 的平均感知结果。Δ_{size} 表示与最先进水平的中期协同方法 V2VNet 和 DiscoNet 相比，该模型的尺寸增加% 及减少幅度 ↑ 和 ↓。

方法	交并比为 = 0.5 时的 AP		交并比为 = 0.7 时的 AP		模型尺寸	Δ_{size} %	Δ_{size} %
	不含 RSU	带 RSU	无 RSU	带 RSU	(Mb)	V2VNet	DiscoNet
下界 [1]	49.9	46.96	44.21	42.33	90.5	↓ 50.11	↓ 25.45
余下界 [1]	43.99	42.98	39.1	38.26	90.5	↓ 50.11	↓ 25.45
上界 [1]	70.43	77.08	67.04	72.57	90.5	↓ 50.11	↓ 25.45
When2com [15]	44.02	46.39	39.89	40.32	247.5	↑ 51.32	↑ 126.11
When2com* [15]	45.35	46.39	40.45	41.43	247.5	↑ 51.32	↑ 126.11
Who2com [16]	44.02	46.39	39.89	40.32	-	-	-
Who2com* [16]	45.35	48.28	40.45	41.13	-	-	-
V2VNet [7]	68.35	72.08	62.83	65.85	181.4	-	↑ 49.42
DiscoNet [8]	**69.03**	72.87	63.44	66.4	**121.4**	↓ 33.07	-
FF_D Sum64（我们的方法）	68.14	71.81	62.36	65.86	122.5	↓ 32.47	↑ 0.91
FF_D Sum32（我们的方法）	68.56	**72.96**	**63.48**	**65.94**	122.7	↓ 32.36	↑ 1.07
FF_D 拼接64（我们的方法）	68.97	71.61	62.88	64.41	122.7	↓ 30.76	↑ 3.46
FF_D 拼接32（我们的方法）	68.50	72.25	63.32	63.74	126	↓ 30.54	↑ 3.79
FF_V 拼接64（我们的方法）	67.53	70	61.55	63.52	126.6	↓ 30.20	↑ 4.28
FF_V 拼接32（我们的方法）	68.46	70.94	63.10	63.15	126.6	↓ 30.70	↑ 3.52

采用该策略，而非对模型或架构进行其他修改。此外，我们在同一中间特征层进行协作。我们使用 Adam 优化器 [28]，初始学习率为 10^{-3}，并每训练 10 个 epoch 以 0.1 的因子逐步衰减。所有模型均在 Tesla V100 上训练。

V. 结果与讨论

表 II 和表 I 展示了内部方法与基线方法在 AP（@ IoU 0.5 和 0.7）方面的比较。如第 III-C 节所述，我们采用了基于卷积的编码器-解码器注意力机制来计算注意力权重，从接收的特征图中识别重要区域，以增强自车智能体的感知。在这方面，表 I 展示了所采用的两种编码器-解码器层深度，即 (256, 128, 64) 和 (256, 128, 64, 32)，用于研究更深的网络对学习到的注意力权重、整体模型性能和模型大小的影响。值得注意的是，$\uplus$ 符号（见式 1）用于表示聚合操作，即表 I 所示的拼接或加法操作，以考察并比较这些操作对初始聚合阶段的影响。表 II 所示的实验结果表明，$FF_D\ Sum64$ 模型在保持模型参数数量较少的同时，实现了最高的检测精度，这与 [18] 提出的假设一致，即特征加法能够带来更好的融合效果。此外，可以观察到，更深的编码器-解码器能够带来更高的 AP，因为网络中的深层能够访问输入的更高层次、更抽象的表示，从而使网络能够学习特征之间更复杂的模式和关系。这是因为深层能够组合在较早层中学习到的特征，形成能够捕捉输入数据更复杂特征的高层次表示，而这对于学习注意力权重至关重要。然而，我们也发现，网络加深（超过表 I 所示的测试深度）会导致 AP 降低，这是由于梯度消失问题；该问题发生在梯度信号过小时，无法在多个层之间传播。这使得网络早期层的权重难以更新，从而导致收敛效果不佳。

结果显然表明，在所有情况下，协作都能显著提升性能，优于无协同（下界）。在各种协作方法中，上界（早期协作）的表现优于其他所有方法，但也带来了共享数据大小方面的挑战：每次扫描约为 860kb，而协同下界和中期协同每次扫描分别约为 5kb 和 260kb。中期协同方法的表现优于协同下界。就地形和道路类别而言，本文提出的方法（表 I）、V2VNet、DiscoNet 和上界的表现相近。与仅使用车车通信相比，车车通信和车路通信（配备 RSU）通常能通过增加视角提升感知性能；表 II 所示的各中期模型均体现了这一点，这得益于可用于训练模型的数据增多。

所提出的基于注意力的协作采用了精心设计的通道和空间注意力融合，在中间模型中取得了极具竞争力的性能。表 II 中的结果表明，所提出的协作通常能够超越 V2VNet 的性能，例如在 $FF_D\ Sum64$、$FF_D\ Sum32$、$FF_D\ Concat64$ 和 $FF_D\ Concat32$ 中。尽管如此，我们提出的方法始终能够显著减少参数预算，正如“δ_{size}% ($V2VNet$)”列所示，其差值最大为 33.07%，最小为 30.20%。与 DiscoNet 相比，采用 DiscoNet 基础 FF_D 的方法取得了相近或略优的性能，例如在 $FF_D\ Sum64$ 中。

VI. 结论

本文提出了一种基于通道-空间注意力的方法，用于在协同感知中实现中间融合。

感知。通过对多种最先进的多智能体协作感知方法进行基准测试，展示了所提出方法的有效性。此外，使用V2X-Sim 2.0这一大规模多智能体三维目标检测数据集对实验结果进行了验证。所提出的方法取得了与最先进方法相当的结果，并且在某些情况下，在模型参数数量或平均精度方面超越了这些方法。

致谢

本研究得到弗兰德斯研究基金会（FWO）项目编号1S90022N的资助。本研究将在由欧盟委员会根据协议编号952189资助的Horizon 2020 5G-Blueprint项目中使用5G网络进行验证。

参考文献

[1] Yiming Li、Dekun Ma、Ziyan An、Zixun Wang、Yiqi Zhong、Siheng Chen 和 Chen Feng。V2X-Sim：用于自动驾驶的多智能体协同感知数据集与基准。*IEEE*机器人与自动化快报，7(4):10914–10921，2022。[2] Li-Hua Wen 和 Kang-Hyun Jo。面向自动驾驶的基于深度学习的感知系统：全面综述。神经计算，2022。[3] Yuxiao Zhang、Alexander Carballo、Hanting Yang 和 Kazuya Takeda。恶劣天气条件下自动驾驶车辆的感知与传感：综述。*ISPRS*摄影测量与遥感杂志，196:146–177，2023。[4] Simegnew Yihunie Alaba 和 John E Ball。面向自动驾驶的基于深度学习的图像三维目标检测。*IEEE*传感器期刊，2023。[5] Daniel Bogdoll、Enrico Eisen、Maximilian Nitsche、Christin Scheib 和 J Marius Z¨ollner。面向自动驾驶的道路未知物体多模态检测。载于*2022*年*IEEE*系统、人和控制论国际会议（*SMC*），第325–332页。IEEE，2022。[6] Guangzhen Cui、Weili Zhang、Yanqiu Xiao、Lei Yao 和 Zhanpeng Fang。车联网环境下自动驾驶的协同感知技术：综述。传感器，22 5535，2022 15( ): 。[7] Tsun-Hsuan Wang、Sivabalan Manivasagam、Ming Liang、Bin Yang、Wenyuan Zeng 和 Raquel Urtasun。V2VNet：用于联合感知与预测的车对车通信。载于《计算机视觉—*ECCV 2020*：第*16*届欧洲计算机视觉会议》，*2020*年*8*月*23*日至*28*日，英国格拉斯哥，论文集，第二部分*16*，第605–621页。施普林格，2020。[8] Yiming Li、Shunli Ren、Pengxiang Wu、Siheng Chen、Chen Feng 和 Wenjun Zhang。为多智能体感知学习蒸馏协作图。神经信息处理系统进展，34:29541–29552，2021。[9] Qi Chen、Sihai Tang、Qing Yang 和 Song Fu。Cooper：基于三维点云的联网自动驾驶车辆协作感知。载于*2019*年*IEEE*第*39*届分布式计算系统国际会议（*ICDCS*），第514–524页。IEEE，2019。[10] Eduardo Arnold、Mehrdad Dianati、Robert de Temple 和 Saber Fallah。利用基础设施传感器在驾驶场景中进行三维目标检测的协作感知。*IEEE*智能交通系统汇刊，23(3):1852–1864，2020。[11] Ahmed N Ahmed、Ian Ravijts、Jens de Hoog、Ali Anwar、Siegfried Mercelis 和 Peter Hellinckx。面向联网车辆的联合感知方案。载于*2022*年*IEEE*传感器会议，第1–4页。IEEE，2022。[12] Aaron Miller、Kyungzun Rim、Parth Chopra、Paritosh Kelkar 和 Maxim Likhachev。用于协作驾驶的协作感知与定位。载于*2020*年*IEEE*国际机器人与自动化会议（*ICRA*），第1256–1262页。IEEE，2020。[13] Christoph Allig 和 Gerd Wanielik。协作感知中感知信息的对齐。载于*2019*年*IEEE*智能车辆研讨会（*IV*），第1849–1854页。IEEE，2019。[14] Qi Chen、Xu Ma、Sihai Tang、Jingda Guo、Qing Yang 和 Song Fu。F-cooper：面向自动驾驶车辆边缘计算系统的基于三维点云的特征协作感知。载于第*4*届*ACM/IEEE*边缘计算研讨会论文集，第88–100页，2019。[15] Yen-Cheng Liu、Junjiao Tian、Nathaniel Glaser 和 Zsolt Kira。When2com：通过通信图分组实现多智能体感知。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第 4106–4115 页，2020 年。[16] Yen-Cheng Liu、Junjiao Tian、Chih-Yao Ma、Nathan Glaser、Chia-Wen Kuo 和 Zsolt Kira。Who2com：通过可学习握手机制通信实现协同感知。载于*2020 IEEE*国际机器人与自动化会议（*ICRA*），第 6876–6883 页。IEEE，2020 年。[17] Toru Wakahara、Yoshimasa Kimura 和 Akira Tomono。使用全局仿射变换相关性识别灰度字符的仿射不变方法。*IEEE*模式分析与机器智能汇刊，23(4):384–395，2001 年。[18] Meng-Hao Guo、Tian-Xing Xu、Jiang-Jiang Liu、Zheng-Ning Liu、Peng-Tao Jiang、Tai-Jiang Mu、Song-Hai Zhang、Ralph R Martin、Ming-Ming Cheng 和 Shi-Min Hu。计算机视觉中的注意力机制：综述。《计算视觉媒体》，8 368，2022 3 331( ): – 。[19] Jie Hu、Li Shen 和 Gang Sun。压缩与激励网络。载于*IEEE*计算机视觉与模式识别会议论文集，第 7132–7141 页，2018 年。[20] Yimian Dai、Fabian Gieseke、Stefan Oehmcke、Yiquan Wu 和 Kobus Barnard。注意力特征融合。载于*IEEE/CVF*冬季计算机视觉应用会议论文集，第 3560–3569 页，2021 年。[21] Sanghyun Woo、Jongchan Park、Joon-Young Lee 和 In So Kweon。CBAM：卷积块注意力模块。载于欧洲计算机视觉会议论文集（*ECCV*），第 3–19 页，2018 年。[22] Matthew D Zeiler 和 Rob Fergus。卷积网络的可视化与理解。载于计算机视觉 ——*ECCV 2014*：第 *13* 届欧洲计算机视觉会议，瑞士苏黎世，*2014* 年 *9* 月 *6–12* 日，论文集，第一部分 *13*，第 818–833 页。施普林格，2014 年。[23] Binh-Son Hua、Minh-Khoi Tran 和 Sai-Kit Yeung。逐点卷积神经网络。载于*IEEE*计算机视觉与模式识别会议论文集，第 984–993 页，2018 年。[24] Vinod Nair 和 Geoffrey E Hinton。修正线性单元改进受限玻尔兹曼机。载于第 *27* 届国际机器学习会议（*ICML-10*）论文集，第 807–814 页，2010 年。[25] Sergey Ioffe 和 Christian Szegedy。批归一化：通过减少内部协变量偏移加速深度网络训练。载于国际机器学习会议，第 448–456 页。PMLR，2015 年。[26] Daniel Krajzewicz、Jakob Erdmann、Michael Behrisch 和 Laura Bieker。SUMO——城市交通仿真的最新进展与应用。系统与测量进展国际期刊，5(3&4)，2012 年。[27] Alexey Dosovitskiy、German Ros、Felipe Codevilla、Antonio Lopez 和 Vladlen Koltun。CARLA：开放式城市驾驶模拟器。载于机器人学习年度会议论文集，第 1–16 页，2017 年。[28] Diederik P Kingma 和 Jimmy Ba。Adam：一种随机优化方法。*arXiv* 预印本 *arXiv:1412.6980*，2014 年。