Collaborative Perception Review

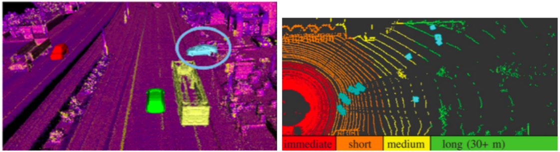
Zhengyi Li

State Key Laboratory of Automotive Simulation and Control, Jilin University

2022 年 11 月 10 日

1. Backgroud
2. Related Work
3. Baseline
4. Reference

Background



(a) The occlusion issue. [14]
(b) The long range issue. [33]

图 1: 单车感知的两个局限性：a) 遮挡，b) 长距离衰减

- 感知模块是自动驾驶车辆中最重要的模块之一，单车感知能力已经取得了非常大的进步，但在一定程度上已经达到了瓶颈

Background

- 为了持续提高车辆的感知能力，打破感知的局限性，在通讯技术的支持下，协同感知收到了广泛的关注 [1]

Related Work

目前，相关工作有一定的数量和规模

- 上海交通大学MediaBrain发表综述文章–*Collaborative Perception for Autonomous Driving: Current Status and Future Trend* [1] 并提出了 *V2VNet, DiscoNet, Who2Comm, When2Comm, Where2Comm, SyncNet* 等一系列模型
- 法国 ESIGELEC 发表综述文章–*Survey on Cooperative Perception in an Automotive Context* [2]
- UCLA Mobility Lab近两年在 *NeurIPS*、*ICLR*、*ICRA* 等顶会发表并开源工作，包括数据集 *OPV2V* [3]、开源框架 *OpenCDA* [4]、模型 *V2X-ViT* [5] 等
- University of North Texas 的 Qi Chen 等先后提出了 Cooper, F-Cooper 模型 [6, 7]

Related Work

[1] 介绍了协同感知的基本概念，协同方式、总结了过程中的关键问题和应用，并讨论了一些行业面临的问题和挑战.
协作方式和关键问题，是这篇文章写作的两条主线，也是该实验室几项工作的主线。

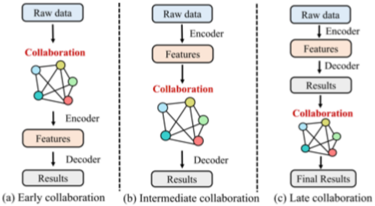


图 2: 三种协同方式

如图2所示，协作方式可以主要分为：

- *Early Collaboration*--早期融合 [6, 8, 7]
  - 方法简单，可实施性高
  - 占据大量的通信带宽
- *Late Collaboration*--后期融合
  - 所需的通信资源较少
  - 容易受到丢包、延迟等噪声影响
- *Intermediate Collaboration*--中期融合 [9, 10, 11, 12, 13]
  - 在效果和通信带宽之间做了比较好的权衡

Related Work I

[1] 介绍了协同感知过程中的几种关键组成部分

- 协作图:
  - 协作图是构建整个过程的重要工具，在一些工作中，将车辆或路侧设备作为图的节点，通信关系作为图的边，对过程进行建模。
  - 针对于协作图的研究，是 MediaBrain 几篇工作的主要研究内容，例如 [9, 12, 10, 11, 13] 都是针对协作图进行研究，优化目标为: **使用更少的通信带宽来达到更强的感知能力。**
- 信息对齐:
  - 由于多个智能体信息的时空不对称性，信息对齐是协同感知重要的组成部分之一
  - MediaBrain 提出了 *SyncNet* [14]，补偿了通信延时对整个模型造成的影响

Related Work II

- UCLA-Mobility Lab 提出的 *V2X-ViT* [5] 将通信延迟作为 Transformer [15] 结构中的 *Position Embedding* 输入，对通信延时进行补偿
- 信息融合：
  - 收到多个智能体信息后，单车的信息融合是整个过程中最重要的环节。
  - *CommNet* [16] 将收到的信息通过均值操作合并到一起，以 *VAIN* [17] 为代表的注意力机制在得到的广泛的应用

Related Work

Related Work Summary:

- [1, 2] 两篇综述中，相关工作大多均为 2020 年后发表
- *Intermediate Collaboration* 是学术研究中使用最广泛的一种方式
- MediaBrain 与 UCLA Mobility Lab 的工作最为先进和成体系，但两者侧重点不同
- MediaBrain 与 UCLA-Mobility Lab 工作均开源数据集和开放源码

MediaBrain – V2X-Sim [18]

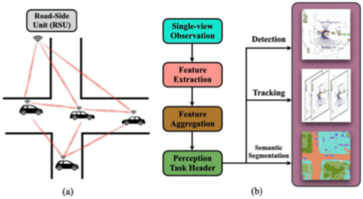


Fig. 1. (a) Intersection for vehicle-to-everything (V2X) communication. (b) Workflow of multi-agent collaborative perception with intermediate-/feature-based strategy. We benchmark collaborative object detection, multi-object tracking, and semantic segmentation in the bird's eye view (BEV).

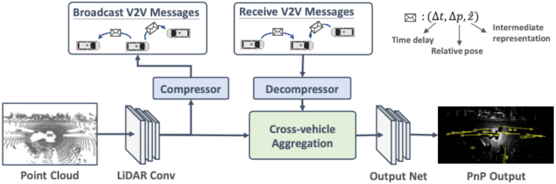
V2X-Sim 是基于仿真场景的，协同感知数据集，包括：

- 路侧传感器和车辆传感器数据
- 多类型传感器数据
- 多种感知任务的 Ground Truth 标签

Basebone:

- Encoder
- Collaboration Graph
- Decoder
- Output Header

MediaBrain – V2VNet [9]



V2VNet 是一种利用 *GNN* 聚合信息的感知预测 (*P&P*) 模型

Key Words: *GNN*, *ConvRNN* *V2VNet* 主要分为三个阶段：

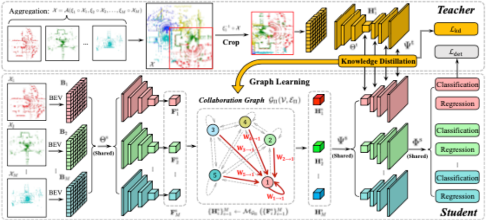
- 数据预处理 (*CNNs*) 与压缩 (*SOTA* 算法)
- 信息融合：
  - 信息解压缩
  - 使用 *GNN* 进行信息融合，以满足不同车辆的空间位置与时间的变化
- 结果输出： *End-to-End* 输出感知与预测结果

MediaBrain – DiscoNet [12]

[12] 提出了一种基于知识蒸馏 [19] 的可训练的协作图，使用矩阵表示，矩阵上的值表示节点之间传递消息的权重

- 教师模型使用 *Early Collaboration*，使用全局信息进行感知
- 学生模型分为四个部分：
  - 特征编码：将 *BEV* 经过 *CNNs* 得到 *Feature Map*
  - 特征压缩：使用 $1 \times 1 conv$ 进行特征压缩
  - 协作图构建：
    - 信息传递
    - 得到图注意力
    - 信息合并
  - 解码并输出结果：使用 $CNNs$ 解码并输出结果

MediaBrain – DiscoNet [12]

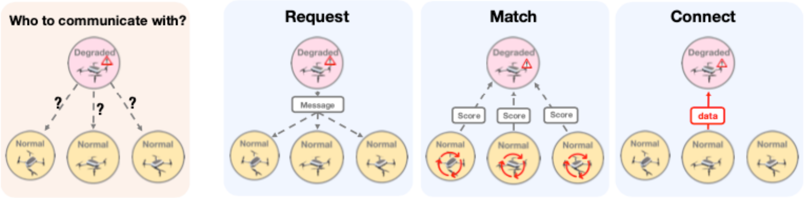


- Message Transmission: 每个智能体将压缩后的 $FeatureMap$ 传输给其他智能体
- Message Attention:
  - 智能体 i 接收智能体 j 的信息 F_j，并基于位置，将信息 F_j 变换到 $agent_i$ 的坐标系下 $F_{j\to i} = \Gamma_{j\to i}(F_i)$
  - 使用边编码器将边编码 $W_{j\to i} = \Pi(F_{j\to i}, F_i)$
  - 使用 $1\times 1 convs$ 降维得到 $Attention$ 数值
- Message Aggregation: 每个智能体根据 $Attention$ 权重得到合并后的 $FeatureMap$

MediaBrain – Who2com [10]

[10] 提出了一种多阶段、握手机制 (Multi-stage Handshake Communication) 的通信方式，当智能体发送请求，其他智能体接收信息并计算匹配程度，自动选择一个或多个智能体进行协同，大大减少了带宽的消耗，包含三个部分：

- *Request*
- *Match*
- *Connect*



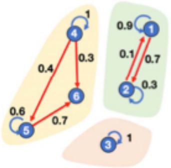
MediaBrain – Who2com [10]

- Request: $Agent_i$ 发送压缩后的 $Request: \nu_i = G_m^j(\tilde{x}_i; \theta_m)$
- Match: 匹配分数 $s_{ij} = \Phi(\nu_i, \kappa_j), \kappa_j = G_k^j(x_j; \theta_k)$
  - 其中，$\Phi = \nu_j^T W_a \kappa_i$，应是根据 [15] 中 $Self_Attention$ 计算相关度的方法
- Connect: 连接并传送信息

训练过程中不以最优匹配为目标，以最大化感知能力为目标，克服了该问题数据集缺少的问题

MediaBrain – When2comm [11]

When2com 在 *Who2com* 的基础上，增加了对通信触发的判断（类似于事件触发），进一步减少了通信带宽的消耗。



$$\bar{M} = \begin{pmatrix} 0.9 & 0.1 & 0 & 0 & 0 & 0 \\ 0.7 & 0.3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0.4 & 0.6 & 0 \\ 0 & 0 & 0 & 0.3 & 0.7 & 0 \end{pmatrix}$$

在 *Who2com* 的三段上，*When2com* 基于 $Self - Attention$ 和 $Cross - Attention$ 生成一个矩阵，表示每个智能体接收不同其他智能体的权重，当

$$m_{i,i} = 1$$

时，$Agent_i$ 不需要接收其他智能体的信息

MediaBrain – SyncNet

UCLA-Mobility Lab – OPV2V

UCLA-Mobility Lab – OpenCDA

UCLA-Mobility Lab – V2X-ViT

UCLA-Mobility Lab – Bridging—

Reference I

[1] Shunli Ren, Siheng Chen, and Wenjun Zhang.
Collaborative perception for autonomous driving: Current status and future trend.
In *Proceedings of 2021 5th Chinese Conference on Swarm Intelligence and Cooperative Control*, pages 682–692. Springer, 2023.

[2] Antoine Caillot, Safa Ouerghi, Pascal Vasseur, Rémi Boutteau, and Yohan Dupuis.
Survey on cooperative perception in an automotive context.
IEEE Transactions on Intelligent Transportation Systems, 2022.

[3] Runsheng Xu, Hao Xiang, Xin Xia, Xu Han, Jinlong Li, and Jiaqi Ma.
Opv2v: An open benchmark dataset and fusion pipeline for perception with vehicle-to-vehicle communication.
In *2022 International Conference on Robotics and Automation (ICRA)*, pages 2583–2589. IEEE, 2022.

[4] Runsheng Xu, Yi Guo, Xu Han, Xin Xia, Hao Xiang, and Jiaqi Ma.
Opencda: an open cooperative driving automation framework integrated with co-simulation.
In *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, pages 1155–1162. IEEE, 2021.

[5] Runsheng Xu, Hao Xiang, Zhengzhong Tu, Xin Xia, Ming-Hsuan Yang, and Jiaqi Ma.
V2x-vit: Vehicle-to-everything cooperative perception with vision transformer.
arXiv preprint arXiv:2203.10638, 2022.

[6] Qi Chen, Sihai Tang, Qing Yang, and Song Fu.
Cooper: Cooperative perception for connected autonomous vehicles based on 3d point clouds.
In *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*, pages 514–524. IEEE, 2019.

Reference II

[7] Qi Chen, Xu Ma, Sihai Tang, Jingda Guo, Qing Yang, and Song Fu.
F-cooper: Feature based cooperative perception for autonomous vehicle edge computing system using 3d point clouds.
In *Proceedings of the 4th ACM/IEEE Symposium on Edge Computing*, SEC '19, page 88–100, New York, NY, USA, 2019. Association for Computing Machinery.

[8] Eduardo Arnold, Mehrdad Dianati, Robert de Temple, and Saber Fallah.
Cooperative perception for 3d object detection in driving scenarios using infrastructure sensors.
IEEE Transactions on Intelligent Transportation Systems, 23(3):1852–1864, 2022.

[9] Tsun-Hsuan Wang, Sivabalan Manivasagam, Ming Liang, Bin Yang, Wenyuan Zeng, and Raquel Urtasun.
V2vnet: Vehicle-to-vehicle communication for joint perception and prediction.
In *European Conference on Computer Vision*, pages 605–621. Springer, 2020.

[10] Yen-Cheng Liu, Junjiao Tian, Chih-Yao Ma, Nathan Glaser, Chia-Wen Kuo, and Zsolt Kira.
Who2com: Collaborative perception via learnable handshake communication.
In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6876–6883. IEEE, 2020.

[11] Yen-Cheng Liu, Junjiao Tian, Nathaniel Glaser, and Zsolt Kira.
When2com: Multi-agent perception via communication graph grouping.
In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, pages 4106–4115, 2020.

[12] Yiming Li, Shunli Ren, Pengxiang Wu, Siheng Chen, Chen Feng, and Wenjun Zhang.
Learning distilled collaboration graph for multi-agent perception.
Advances in Neural Information Processing Systems, 34:29541–29552, 2021.

Reference III

[13] Yue Hu, Shaoheng Fang, Zixing Lei, Yiqi Zhong, and Siheng Chen.
Where2comm: Communication-efficient collaborative perception via spatial confidence maps.
In *Neural Information Processing Systems (NeurIPS)*, December 2022.

[14] Zixing Lei, Shunli Ren, Yue Hu, Wenjun Zhang, and Siheng Chen.
Latency-aware collaborative perception.
arXiv preprint arXiv:2207.08560, 2022.

[15] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin.
Attention is all you need.
In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[16] Sainbayar Sukhbaatar, arthur szlam, and Rob Fergus.
Learning multiagent communication with backpropagation.
In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016.

[17] Yedid Hoshen.
Vain: Attentional multi-agent predictive modeling.
In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[18] Yiming Li, Dekun Ma, Ziyan An, Zixun Wang, Yiqi Zhong, Siheng Chen, and Chen Feng.
V2x-sim: A virtual collaborative perception dataset and benchmark for autonomous driving.
2022.

Reference IV

[19] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean.
Distilling the knowledge in a neural network (2015).
arXiv preprint arXiv:1503.02531, 2, 2015.

Thanks!